UNIVERSITY OF OSLO

Master thesis

Development of a large-scale and high-throughput serum proteomics workflow

Kristian Andersen Sundby

Molecular Biology and Biochemistry
xxx 60 credits

Department of Biosciences
The Faculty of Mathematics and Natural Sciences

Development of a large-scale and high-throughput serum proteomics workflow

Kristian Andersen Sundby

Acknowledgements

Abstract

Proteomics is the study of a specimens entire protein complement at a given time and condition. Mass spectrometry-based «bottom-up» proteomics, where the mass-to-charge ratio of peptide fragments are analyzed in order to identify the precursor proteins, is the most commonly utilized approach for large-scale or high-throughput protein sample analysis.

The human plasma proteome is a complex mixture of true plasma proteins and those resulting from leakage into the blood by the surrounding tissue. This makes proteome analysis of plasma and serum of immense interest, both for research of samples collected in biobanks and clinical studies for discovery of biomarkers. However, its complexity also composes a great challenge to reasonable coverage of the entire plasma proteome. This is predominantly caused by the dynamic range of proteins, with over 10 orders of magnitude between those of most and least abundance. Research into improving serum proteomics has been ventured for more than 20 years without much success.

This project aims to develop an approach for large-scale and high-throughput analysis of the serum proteome, in order to improve the number of identifiable serum proteins using mass spectrometry. In this study, three objectives for this purpose were explored. First, depletion of the abundant proteins albumin and immunoglobulin gamma was attempted using the depletion kits Albumin/IgG Removal Kit (Pierce™), AlbuSorb™ PLUS, AlbuVoid™ LC-MS On-Bead, Albuvoid™ PLUS, and BioMag® ProMax. Second, methods for fractionation of the serum proteome were examined by using the Cohn process of ethanol protein precipitation, protein size exclusion chromatography, zinc chloride protein precipitation, and High-pH reversed-phase peptide fractionation by liquid chromatography (PepSwift™) and spin column (Pierce™). Third, cysteine-containing peptide enrichment was explored with the reactive group pyridyl disulfide in order to reduce the complexity of the analyte. The affinity beads BcMag™ Thiol-Activated Magnetic Beads, BcMag™ Long Thiol-Activated Magnetic Beads, High Capacity Acyl-rac S3™ Capture Beads, and Thiopropyl Sepharose™ 6B Affinity Resin were employed for this objective.

Of the tested depletion kits, the Albumin/IgG depletion kit identified the most proteins, but not more than without depletion. 234 proteins were identified with this kit when cold acetone was used to precipitate the depleted sample before protein digestion. More promising results were accomplished through fractionation. The most promising of the protein fractionation methods were size exclusion chromatography, with a tentative 574 identified proteins, and zinc chloride precipitation, which identified 464 serum proteins. Spin column-based HpH-RP peptide fractionation identified 406 proteins, proving the most successful of the two. The results from $ZnCl_2$ precipitation being accomplished with only two fractions made this approach the most favorable. Enrichment of cysteine-containing peptides could not be successfully adapted to serum by the techniques used here, but further investigation with alternative methodologies could prove more lucrative.

From what was accomplished here, it would seem that fractionation is the best approach when seeking to identify as many serum proteins as possible, particularly through $ZnCl_2$ protein precipitation with only two fractions. Yet, still more optimizations can be made to improve the yield further.

Abbreviations

$\sum$ total.

m/z mass to charge ratio.

pK_a acid dissociation constant.

2-DE two-dimensional gel electrophoresis.

6C-CysPAT iodoacetamido-LC-phosphonic acid.

ABC ammonium bicarbonate.

ACN acetonitrile.

AcOH acetic acid.

Apo apolipoprotein.

AVBB AlbuVoid Binding Buffer.

AVEB AlbuVoid Elution Buffer.

AVWB AlbuVoid Wash Buffer.

BB1 Binding Buffer 1.

BSA bovine serum albumin.

CB cibacron Blue 3GA.

CF complement factor.

CID collision-induced dissociation.

COFRADIC combined fractional diagonal chromatography.

CSF cerebrospinal fluid.

ctrl control.

Da dalton.

DDA data dependent acquisition.

DIA data independent acquisition.

DNA deoxyribonucleic acid.

DTT dithiothreitol.

EDTA ethylenediaminetetraacetate.

ESI electrospray ionization.

ETD electron transfer dissociation.

EtOH ethanol.

Fab fragment antigen-binding.

Fc fragment crystallizable.

FDA Food and Drug Administration.

FDR false discovery rate.

GE gel electrophoresis.

HCD higher-energy C-trap dissociation.

HDL high-density lipoprotein.

HGP Human Genome Project.

HILIC hydrophilic interaction liquid chromatography.

HPG-ALD dendritic polyglycerol aldehyde polymers.

HpH high-pH.

HPLC high performance liquid chromatography.

HPP Human Proteome Project.

HSA human serum albumin.

HT high-throughput.

HUPO Human Proteome Organization.

IAA iodoacetamide.

ICAT isotope-coded affinity tag.

ID identification.

IDL intermediate-density lipoprotein.

IEF isoelectric focusing.

IEX ion exchange chromatography.

Ig immunoglobulin.

IgG immunoglobulin G.

LC liquid chromatography.

LDL low-density lipoprotein.

LpH low-pH.

MALDI matrix-assisted laser desorption ionization.

MQ milli-Q.

MS mass spectrometry.

MS/MS tandem mass spectrometry.

Mw molecular weight.

NaOAC sodium acetate.

NIH National Institutes of Health.

Nrich N-terminal peptide enrichment method.

NTA nitrilotriacetic acid.

OW organic wash.

PAMPS poly(acrylamidomethylpropyl sulfonate).

PASEF parallel accumulation-serial fragmentation.

PBS phosphate-buffered saline.

PCR polymerase chain reaction.

PDADMAC poly(diallyldimethylammonium chloride).

PEA proximity extension assay.

pI isoelectric point.

PMACo poly(dimethylamine-*co*-epichlorohydrin-*co*-ethylenediamine).

PPT protein precipitation.

PSA prostate specific antigen.

PTM post-translational modification.

Q quadropole.

rcf relative centrifugal force.

RNA ribonucleic acid.

RP reversed-phase.

SAX strong anion exchange chromatography.

SC spectral count.

SCX strong cation exchange chromatography.

SDS-PAGE sodium dodecyl sulfate polyacrylamide electrophoresis.

SEC size exclusion chromatography.

SOMAmer slow off-rate modified aptamer.

SPE solid phase extraction.

TAILS terminal amine isotopic labeling of substrates.

TCA trichloroacetic acid.

TCEP tris(2-carboxyethyl)Phosphine Hydrochloride.

TEA triethylamine.

TFA trifluoroacetic acid.

TIMS trapped ion mobility spectrometry.

TMT tandem mass tag.

TOF time of flight.

tris-HCl tris(hydroxymethyl)aminomethane - hydrochloride.

UniParc UniProt Archive.

UniProt Universal Protein Resource.

UniProtKB UniProt Knowledge base.

UniRef UniProt reference clusters.

UV ultraviolet.

VIS visible spectrophotometry.

VLDL very-low-density lipoprotein.

ZnCl$_2$ zinc chloride.

α-G α-globulin.

β-G β-globulin.

List of Figures

List of Tables

Table of Contents

1 Introduction

1.1 Proteomics

The word «Proteome» was introduced by Marc Wilkins et al. in 1996, defined as the entire protein complement of a genome, cell, tissue, or organism at a given time point under defined conditions (*113*). Although all cells of an organism share the same genome, their individual protein expression varies greatly depending on cell type, what tissue/organ they reside in, and their function and activity. Furthermore, the complexity of the proteomes is increased by alternative splicing of the messenger ribonucleic acids (RNAs) and post-translational modifications (PTMs). Protein expression will also change in response to different stimuli and stressors, and with time as they progress through the cell cycle. Thus, proteomics becomes the study of a specimens complete array of proteins and its dynamic fluctuations.

1.1.1 Bottom-Up Proteomics

When complex samples undergo large-scale or high-throughput analysis, it is standard to use a «bottom-up» proteomics workflow (*50*). Here, proteins are identified from their peptide fragments, usually produced enzymatically from the original sample, in contrast to mass spectrometry-based «top-down» proteomics where whole proteins are directly analyzed. The typical steps involved in bottom-up proteomics are presented in Figure 1.1, with examples of various sample preparation steps. The center course shows the standard approach that all samples typically undergo, whilst the branching paths are options that may be utilized depending on the sample and objective. During sample preparation, the protein sample is digested into peptides and purified. Adjustments to the protein or peptide sample can also be performed, such as depletion, fractionation, or enrichment. The sample is then separated by liquid chromatography (LC) and analyzed using online coupling of the LC to a mass spectrometer. Then the resulting mass spectral data is analyzed in order to identify the proteins in the sample. Protein identification in bottom-up proteomics is limited to the amino acid sequences present in protein databases, which often excludes PTMs and protein isoforms. In the top-down approach, intact proteins are fragmented directly inside the mass spectrometer, requiring a high-performance instrument with a powerful fragmentation technique. This allows for characterization of PTMs and isoforms, but is only suitable for single or low complexity protein solutions. As such, the bottom-up approach is more suitable for protein identification, while top-down is useful for full structural analysis of single proteins (*13, 19*).

Mass spectrometry (MS)-based proteomics is a versatile, well established method that provides highly detailed information about the identified proteins. Recent non-MS-based proteomic techniques have also been developed, with high-throughput and multiplex analysis (*79*). One such technique is the aptamer-based assay SomaScan (*17*). This technique uses slow off-rate modified aptamers (SOMAmers), which are single-stranded deoxyribonucleic acid (DNA) affinity reagents that bind to the epitopes of target proteins. Using DNA as tags for protein binding allows for the use of DNA detection methods. Another technique is proximity extension assay (PEA) by Olink Proteomics AB (Uppsala, Sweden) (*110*). PEA uses paired antibodies with attached DNA oligonucleotides, which hybridize upon contact with a matching antibody, creating a unique DNA-barcode for each protein. The barcodes can then be quantified, for example through polymerase chain reaction (PCR) amplification.
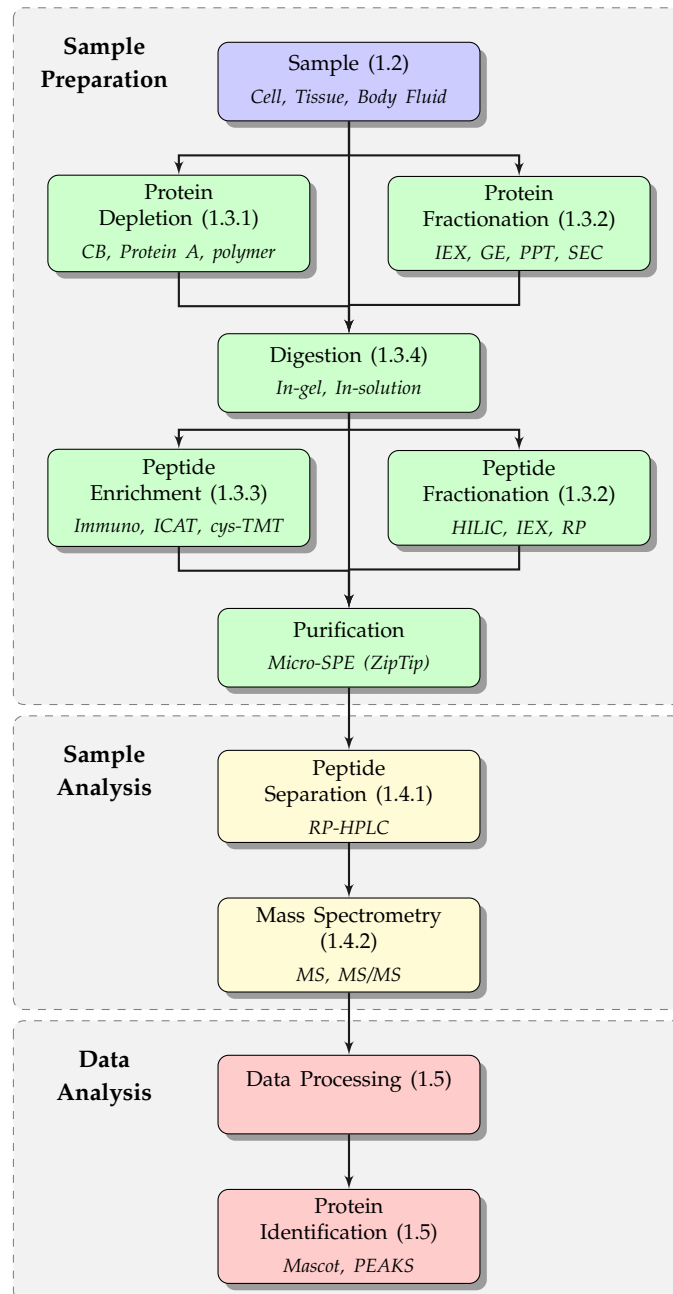
Figure 1.1: **Common steps in bottom-up proteomics workflow.** The proteins can origin from different sample types (1.2). Abundant proteins can be removed by protein depletion (1.3.1), or complex proteomes can be divided by protein fractionation 1.3.2). Digestion of the proteins produce peptides (1.3.4), which can also be divided by peptide fractionation (1.3.2), or specific peptides can be enriched by peptide enrichment (1.3.3). Typically, the peptides are cleaned up and purified, followed by peptide separation (1.4.1) using RP-HPLC and analysis by mass spectrometry (1.4.2). Subsequently, data processing, and the use of database search engine (e.g., Mascot, PEAKS) are used for protein identification (1.5). Typical examples are shown in italic. The numbers in brackets indicate in which part of the thesis the methodology is described.

These affinity-based methods can reveal a high coverage of the proteome with more than a few thousand proteins, similar to what is currently possible with MS-based proteomics. The non-MS-based proteomics methods are particularly interesting for the analysis of proteomes with a high dynamic range, as discussed later for blood proteins. However, a major weakness

is inadequate specificity and cross-reactivity of the antibodies, as well as being limited to target proteins. MS-based proteomics, on the other hand, is hypothesis free, meaning all proteins in a given sample are readily detectable, given their availability in a protein database.

1.1.2 Biomarkers

Proteomics can be immensely useful in the discovery and detection of biomarkers. The U.S. Food and Drug Administration (FDA) and the National Institutes of Health (NIH) defines a biomarker as «a defined [biological] characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention.» (*10*). Several types of biomarkers exist, depending on the purpose of the measurement, including biomarkers for detecting disease, establishing a prognosis, and monitoring the disease progression and treatment response. In the field of proteomics, the presence or amount of a protein or peptide can serve as a biomarker (*73*). As of 2013, the FDA had approved 109 protein biomarkers for use in clinical tests (*5*). In later years, an average of < 1.5 new tests per year have been approved. Only a limited number of medical conditions have specific biomarkers established that are able to give high confidence prognostics. An example of a FDA approved protein marker, frequently used in the clinic for prostate cancer screening, is prostate specific antigen (PSA). Unfortunately, PSA has a very high false-positive rate, where only 30 - 45% of subjects with $4 - 10\,\text{ng/mL}^{-1}$ PSA detected in serum by immunoassay indeed have prostate cancer (*99*).

Many biomarker candidates have been explored in various studies in recent years. By searching PubMed (*88*) with the query «proteomics AND biomarker AND mass spectrometry», 156 clinical trials between the years 2013 - 2022 appear. The journey from discovery to approval and clinical use, however, is steep. The primary limitation of biomarker approval lies in the process of validation. This requires a clearly outlined aim, well-described sampling and experimental methodology, large sample size, befitting selection of subjects, and relevant statistical applications (*73*). Most of the approved protein biomarker tests are immunoassay based, which requires prior knowledge of the protein. Since MS-based analysis is hypothesis free, it enables the discovery of new biomarkers. This is of particular interest in diagnostic and clinical medicine. However, in order for biomarker discovery to be achievable, the given proteome must first be well characterized.

1.2 Mapping of The Human Proteome

In 2010, the Human Proteome Project (HPP) (*66*) was launched by the Human Proteome Organization (HUPO). This project has the goal of mapping the entire human proteome by uncovering the protein products of the protein-coding genes profiled by the Human Genome Project (HGP) initiated in 1990 (*108*). The latest update from 2021 indicates that 18 357 proteins, that is 92.6%, of these predicted human proteins have been detected at the protein level based on MS analysis (*78*). An integral part of this project is the Peptide Atlas (*30*), which hosts a compendium of MS identified peptides.

1.2.1 Body fluids

A biofluid is any fluid that a biological organism excretes or secretes. For humans, this includes cerebrospinal fluid (CSF), blood, saliva, tears, and urine, to name the best studied body fluids. These are listed in Table 1.1 with the approximate protein concentration and the number of total proteins that have been identified. They are predominantly comprised of water, but also contain proteins, lipids, salts, metabolites, sometimes cells, and more. Biofluids can be collected through various means, ranging from noninvasive methods for fluids such as saliva,

tears and urine, to microinvasive techniques for blood, and highly invasive methods as for CSF. Yet obtaining biofluids is generally less invasive than tissue samples or other types of specimens (*54, 103, 114*).

Table 1.1: **Comparison of protein concentration and abundance in human body fluids.** List of best studied human body fluids with the approximate protein concentration and the number of total proteins identified in each.

Body fluid	Protein concentration (mg/mL)	No. of protein IDs	Reference
Cerebrospinal fluid	0.15 – 0.45	3 769	(*32, 118*)
Blood (Plasma)	60 – 80	4 395	(*31, 65*)
Saliva	0.5 – 2.0	5 635	(*97, 112*)
Tears	6.0 – 11	1 526	(*1, 124*)
Urine	0.01 – 10	6 085	(*3, 122*)

Cerebrospinal fluid surrounds the brain and spinal cord, providing protection from mechanical damage, waste handling, and support. It is derived from blood, filtered at the blood-brain barrier, but has a much lower protein concentration. Typically, it is sampled for the identification of biomarkers that indicate neurological diseases such as Alzheimer's disease (*92, 121*).

Saliva is produced by salivary glands lining the oral mucosa and plays a role in food digestion and defense against pathogens. It also contains non-salivary components from food, nasal secretions, and microorganisms, among others. Saliva is used in research and diagnostics of oral and other systemic diseases (*75, 97*).

Tears are a combination of secretions from the lacrimal gland, cornea, and goblet cells, among others. They create a film over the ocular surface that helps focus light, and in addition protects the eye from drying out and against pathogens. The main use of tears in proteomics is to detect protein biomarkers that are related to various eye diseases (*1, 124*).

Urine is the body fluid that is the least invasive to collect in large amounts and is routinely used to study bladder and renal diseases. Blood is filtered in the kidney to produce urine, which is then stored in the bladder before excretion. Waste expulsion, particularly of nitrogenous substances such as urea, is the main function of urine. Homeostasis is also maintained through regulation of the water and salt concentration (*2, 3*).

1.2.2   The Blood Proteome

Blood is a complex mixture of cells, proteins, and other components. The cellular components of blood are mainly the erythrocytes involved in oxygen transport, leukocytes involved in the immune system, and thrombocytes involved in wound healing (*89*). One of the main functions of the circulatory system is transporting substances, such as metabolites, gasses, ions, and waste, to and from all parts of the body. By its very nature, this far-reaching network thus accumulates the proteome of virtually all tissues, making blood possibly the most complex body fluid in the field of proteomics (*63*). Blood is for this reason highly prominent in clinical research and practice. Typically, the liquid portion of blood is collected, either as plasma or serum. Plasma is produced when whole blood is treated with an anticoagulant. Serum is collected after the blood is allowed to clot from the supernatant after centrifugation and clot removal (*65*). Serum has a similar protein content to that of plasma. The main difference being

that clotting factors, such as fibrinogen, are removed with the clot. In addition, there may be proteins that bind nonspecifically to these clotting proteins that are therefore also lost. Nevertheless, this could also be advantageous, as proteins of low abundance may become more available (55).

As of 2021, the Human Plasma PeptideAtlas encompasses 4,395 canonical plasma proteins (31). It is estimated that about 51% of these are true plasma proteins, 25% stem from tissue leakage, 18% are receptor ligands and the remaining 6% are other deviant secretions (4). True plasma proteins are those secreted into the blood that serve their primary function there. This includes proteins such as those involved in the immune system, transport, homeostasis, etc. 99% of the total protein mass in plasma is made up of only the 22 most abundant proteins. This comprehends the following proteins: albumin, α-1-acid glycoprotein, α-1-antitrypsin, α-2-macroglobulin, apolipoprotein (A-I, A-II, A-IV, B-100), ceruloplasmin, complement factor (B, C3, C4, C8, H), fibrinogen, haptoglobin, hemopexin, immunoglobulin (A, G, M), plasminogen, and transferrin (6), also listed in Table 1.2 with their molecular weight (Mw) in Dalton (Da).

Table 1.2: **Molecular weight (Mw) of the 22 most abundant plasma proteins (89).** The proteins are grouped as α-Globulin (red), apolipoprotein (orange), β-Globulin (yellow), complement factor (green), immunoglobulin (purple) and other (blue).

α-globulin (α-G)	Mw (kDa)	Apolipo-protein	Mw (kDa)	Complement Factor	Mw (kDa)
α-1-acid glycoprotein	42	A-I	28	B	93
α-1-antitrypsin	51	A-II	9	C3	185
α-2-macroglobulin	720	A-IV	43	C4	190
Ceruloplasmin	120	B-100	513	C8	151
Haptoglobin	43			H	155
β-globulin (β-G)		Immunoglobulin		Other	
Hemopexin	49	A	150/50[*]	Albumin	66
Plasminogen	90	G	150/50[*]	Fibrinogen	340
Transferrin	76	M	960/63[*]		

[*] total protein / heavy-chain fragment.

Human serum albumin (HSA) makes up more than 50% of the plasma proteome of the average adult, thus becoming the most abundant protein found in plasma. It is built up of 585 amino acids that form three repeating domains, as shown in Figure 1.2a (26), and has a molecular weight of 66 kDa. This multifunctional protein is an important regulator of the oncotic pressure and fatty acid transport in plasma. It also harbors several enzymatic properties, binds several different metabolites and drugs, can act as a chaperone molecule, and possesses anti-oxidant and anti-coagulant activity (86).

Apolipoproteins (Apo) make up the protein part of lipoproteins, which are lipid and protein complexes. In plasma their function is to transport the hydrophobic lipid molecules to their target destination, thus playing a key role in cholesterol homeostasis. There exists five categories of lipoproteins: chylomicron, very-low-density lipoprotein (VLDL), intermediate-density lipoprotein (IDL), low-density lipoprotein (LDL), and high-density lipoprotein (HDL), listed by increasing protein and decreasing lipid content. Apolipoproteins A-I, A-II, and B-100 are all primarily synthesized in the liver and A-IV in the intestine. A-I and A-II are major components of HDLs, A-IV of chylomicrons, whilst B-100 is the main protein of LDLs, but are also found in VLDLs and chylomicrons.

Complement factors (CFs) are a group of structurally and functionally different plasma proteins that all take part in the complement system, which accompanies the immune system in its antibacterial activity. These factors take part in both the adaptive and innate immune response, and their activation leads to signaling of inflammatory cells and destruction of pathogens. The components of the classical pathway, which works in tandem with the adaptive immune system, are labeled with the capital letter C followed by a number in order of their historical discovery. Components of the alternative pathway are given other capital letters.

Globulins are a large family of diverse proteins that are divided into four fractions by electrophoresis: α_1, α_2, β, and γ. The plasma proteins that belong to the α- and β-fractions perform various enzymatic and transport functions. Of the 22 most abundant proteins mentioned previously, α-1-acid glycoprotein and α-1-antitrypsin are α_1-globulins, α-2-macroglobulin, ceruloplasmin, and haptoglobin are α_2-globulins, whilst hemopexin, plasminogen, and transferrin belong to the β-globulins.

γ-globulins, otherwise known as immunoglobulins (Igs), are collectively the most abundant proteins in plasma. These are the antibodies produced by B lymphocytes of the adaptive immune system that can be either presented on their cell surface or secreted in soluble form. The protein structure, seen in Figure 1.2b, consists of two light- and two heavy-chains, bound by disulfide bonds and noncovalent associations to form a Y-shaped dimer. Each prong of the Y-shape consists of four domains: one variable and one constant region of a light- and heavy-chain. These are known as the fragment antigen-binding (Fab) regions and is where interactions with antigens occur. The stem of the Y-shape, known as the fragment crystallizable (Fc) region, is made up of two constant regions of the two heavy-chains, and mainly functions in immune response activation. Five classes of Igs exist: A, D, E, G and M (*49, 89*). Of these, immunoglobulin G (IgG) is the most abundant, amounting to between 10 - 20% of the plasma protein mass (*109*). IgG has a molecular weight of 150 kDa, while its heavy-chain fragment weights 50 kDa and the light-chains around 25 kDa.

(a) Human serum albumin (1BJ5)   (b) Immunoglobulin G (1HZH)
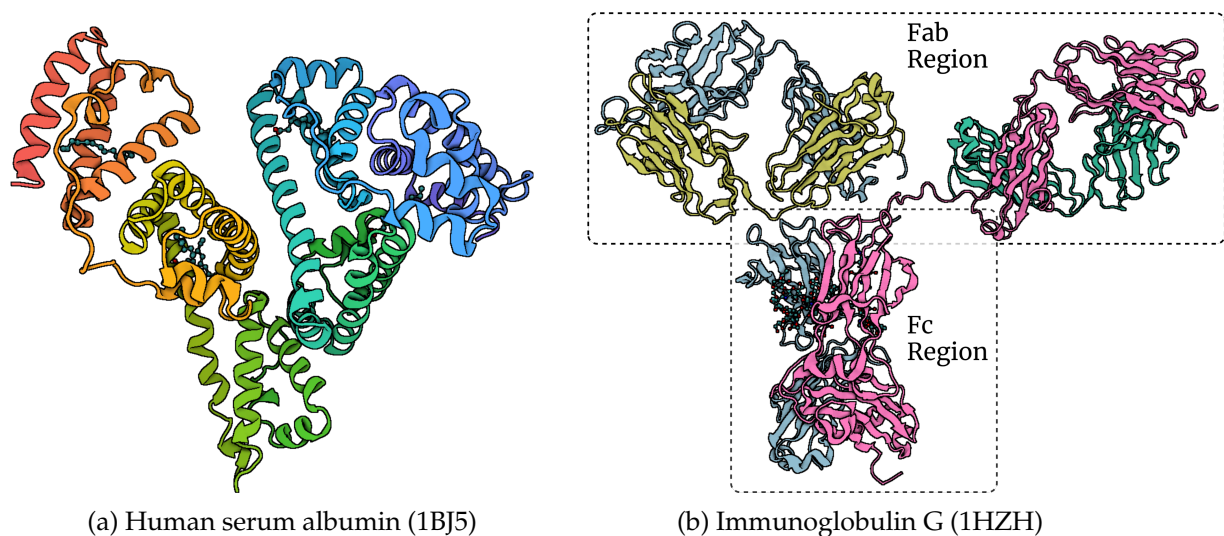
Figure 1.2: **Protein structure of HSA and IgG.** (a) Three domains of HSA in orange, green, and blue. (b) IgG with heavy-chain (light blue/pink) and light-chain (yellow/green), with the Fab and Fc region framed. (*Created with BioRender*).

1.2.3 Dynamic Range of Blood Proteins

The dynamic range, the ratio between the least and most abundant proteins, can for many proteomes be called «the bane of proteomics». This is the case for blood plasma, where the dynamic range can be over 10 orders of magnitude (*6*). The primary contributors to this large difference in range are the 22 proteins presented before, albumin and immunoglobulins being the biggest offenders. Since these make up ∼99% of the plasma proteome mass, reversely the remaining 1% thus contains the over 4,000 other proteins known to be present in plasma. This gradient of protein concentration is illustrated in Figure 1.3. As a result, low abundance proteins become overshadowed by the aforementioned highly abundant proteins, and their presence may go undetected. The cellular proteome, for example, has a much more balanced dynamic range. Present-day, when using a 1-hour LC gradient to separate peptides for most cell lines, the number of detectable proteins can be ≥ 5,000. However, analysis of serum peptides under the same conditions and with the same instrument typically only yields around 400 proteins.
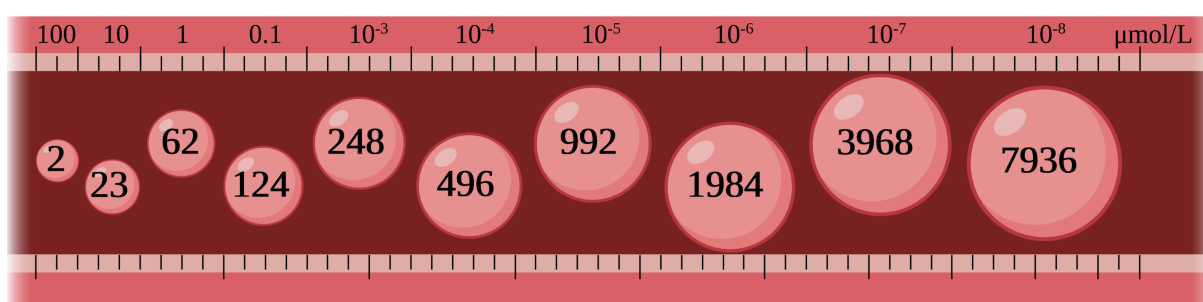
Figure 1.3: **Distribution of plasma proteins by concentration.** Number of plasma proteins, shown in circles, placed at the extrapolated concentration (µmol/L) that they appear in plasma based on data from Hortin et al. (*53*). From left to right the number of proteins increase over 10 orders of increasing magnitude. (*Created with BioRender*).

Furthermore, the complexity of the plasma/serum proteome, although a great boon in regard to research prospects, makes it a challenge to analyze. Besides true plasma proteins, tissue leakage proteins from virtually all tissues contribute largely to the plasma proteome. Consequently, the entire human proteome, that is more than 20,000 proteins as predicted from the human genome (*31*), could also be present in blood. Yet the complexity of the proteome still escalates when taking alternative splicing of the messenger RNA and PTMs into account. Not to mention the variability between individuals, and their respective fluctuant proteomes. It may be that the plasma proteomes at different parts of the circulation system, say the arteries versus the veins, are intrinsically divergent. Rather than having one fluctuating but uniform proteome, plasma is made up of a gradient of sub-proteomes (*6*).

1.3 Sample Preparation of Blood Proteins

A major goal in sample preparation of plasma/serum is combating the dynamic range problem. This can be achieved in numerous ways, typically by either depletion of abundant proteins, fractionation of the sample, or enrichment that favors less abundant proteins.

1.3.1 Depletion of Abundant Proteins

The most common method of dealing with abundant proteins in serum samples is by depleting them from the sample. By reducing the amount of these proteins, the less abundant proteins can become more readily available for detection. Immunoaffinity depletion involves specific

binding of proteins to antibodies which are attached to beads. Examples for commercial products are the Multiple Affinity Removal System (MARS-6, 7, and 14) (Agilent), Seppro® IgY kits (7, 12, and 14) (Millipore), Pierce™ depletion kits (2 and 12) (ThermoFisher), and ProteoPrep® kits (2 and 20) (Sigma). With these kits, up to the 20 most abundant proteins can be depleted. However, the high cost and limited sample loading capacity of the immobilized antibodies of immunoaffinity depletion is problematic for large-scale proteome analyses (*123*).

Depletion of Albumin and Immunoglobulins

The main target proteins for depletion are albumin and/or immunoglobulin gamma (*125*). Depletion of HSA or IgG can be performed by using beads equipped with functional chemistry that allow for ligand binding, typically agarose. Agarose is a polysaccharide consisting of repeating units of D-galactose and 3,6-anhydro-L-galactose. It is derived from red algae and used for the separation of biomolecules in various methods (*46*). Magnetic beads are superparamagnetic microbeads made of agarose with ingrained iron oxides. When exposed to a magnetic field, the beads become magnetic and congregate together (*102*). The bound ligands can then be separated from the remaining analytes. Depending on the binding affinity of the beads, the abundant proteins can either bind to the beads or remain in the solution. The use of magnetic beads has the advantage of being relatively easy to perform, without the need of centrifugation. However, it has been shown that the different kinds of beads are binding different classes of proteins nonspecifically (*105*).

Albumin can be depleted using cibacron Blue 3GA (CB), a widely utilized dye-ligand affinity chromatography tool in protein purification. Cibacron Blue 3GA interacts with proteins through a combination of hydrogen bonding, electrostatic force, and hydrophobic interactions. These interactions are facilitated by triazine, sulfonic, and aromatic groups on the cibacron Blue 3GA molecule, the structure is shown in Figure 1.4 (*68*). This method has low cost and has high-binding capacity. However, low binding specificity of cibacron Blue 3GA can also capture more than 100 other proteins in addition to albumin (*33*). Among these, the most abundant proteins absorbed are apolipoproteins, fibronectin, immunoglobulins, plasma protease inhibitors, ATP-binding proteins, fibrinogens, and complement factors.
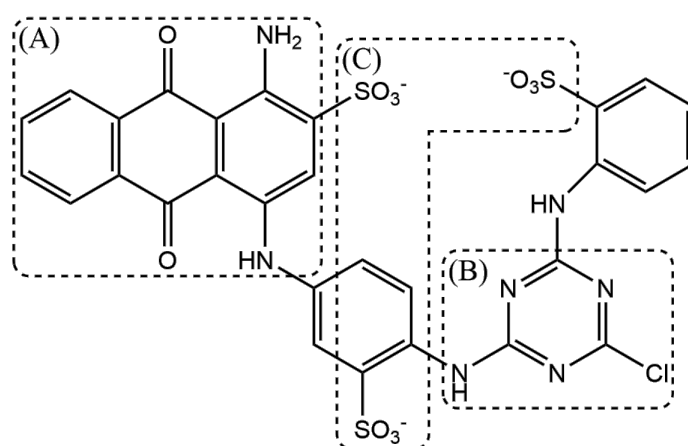
Figure 1.4: **Chemical structure of cibacron Blue 3GA.** The groups facilitating protein binding are (A) an aromatic part, (B) triazine, and (C) sulfonic groups.

Several different polymeric materials have also been developed for albumin depletion. One example of such are polyelectrolytes, which are made up of repeating charged ionizable groups. Examples of polyelectrolytes are presented in Figure 1.5. The binding properties of these poly-

mers depend on the charge, as well as the composition of the polymer and architecture of the cross-links. Proteins with oppositely charged surface residues will interact with and bind to the polyelectrolytes. This makes it possible to separate the bound proteins from the supernatant, which can be either albumin or the other proteins. However, polyeletrolytes also suffer from unspecific binding (*96*).
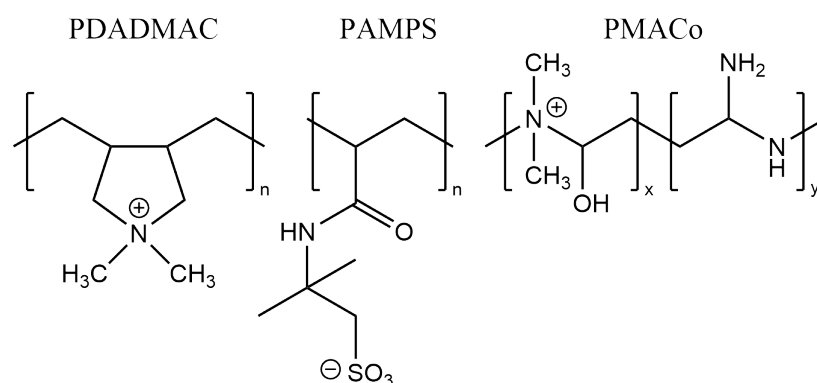
Figure 1.5: **Molecular structure of polyelectrolytes:** poly(diallyldimethylammonium chloride) (PDADMAC), poly(acrylamidomethylpropyl sulfonate) (PAMPS), and poly(dimethylamine-*co*-epichlorohydrin-*co*-ethylenediamine) (PMACo).

Protein A and G are bacterial cell wall receptors that are commonly used for depleting IgG. They both bind to the Fc region of the antibody with high affinity, and may also bind the Fab region, as shown in Figure 1.6. Protein A is found on the cell wall of *Staphylococcus aureus*, while certain strains of *Streptococci* produce Protein G (*29*, *87*). Protein G is shown to bind both IgG and HSA (*98*), has superior binding properties toward monoclonal antibodies, and can bind to antibodies of a broader range of mammalian species than Protein A (*9*). Protein A, on the other hand, can bind to a broader spectrum of immunoglobulin classes (*87*). These proteins can be used alongside the previously mentioned methods for the concurrent depletion of HSA and IgG.

(a) Protein A : IgG Fc complex (5U4Y)   (b) Protein G : IgG Fab complex (1IGC)
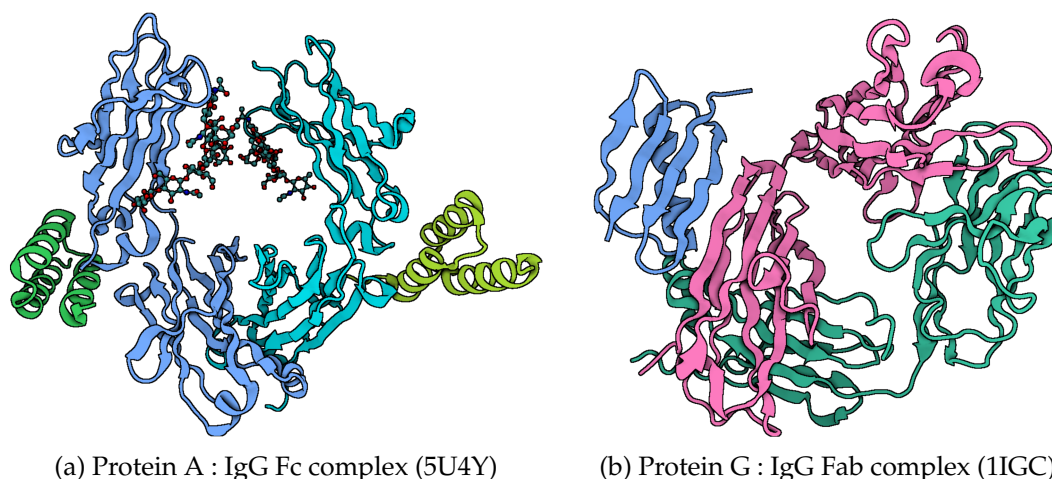
Figure 1.6: **Binding site of Protein A and G:** (a) Protein A B-domain (green) bound to the Fc region of IgG (blue). (b) Protein G (blue) bound to the heavy-chain constant domain (pink) of the Fab region (pink/green). (*Created with BioRender*).

There are several other methods of depleting the sample of HSA and IgG, e.g., immunodepletion (*115*), hydrogel particles (*100*), and combinatorial peptide ligand libraries (*45*). However,

they all suffer from the downside of co-depletion. Both albumin and immunoglobulins are known to bind other proteins, so these may also be removed from the sample during their depletion (*44*). Furthermore, the depletion is often incomplete because these proteins are far too abundant. HSA especially will often remain as the most abundant protein in a sample even after depletion (*125*).

1.3.2 Fractionation

An alternative to depletion of abundant proteins is fractionation of the sample into subproteomes. Here, various techniques can be used to separate proteins or peptides into several fractions based on different characteristics, such as amino acid composition, isoelectric point, size, or hydrophobicity. This could theoretically result in the abundant proteins being isolated into one fraction, allowing for better determination of the less abundant proteins, or at least reduce the complexity of each fraction.

Protein Precipitation

A frequently used protein fractionation method is precipitation. Different proteins precipitate under different conditions, depending on the pH, protein concentration, temperature, and ionic strength of the solution. In principle, precipitation occurs when the interaction between the solved protein and the buffer solution is disturbed. The proteins aggregate either by protein unfolding or surface dehydration due to the hydrophobic effect and electrostatic forces. By adding a precipitation reagent to a protein sample, typically with increasing concentrations, the precipitated proteins at each step can be separated into different fractions. Alternatively, the proteins that remain in solution can be separated from those that have precipitated. Common reagents to use for precipitation are acids, metal ions, organic solvents, or salts. Examples of each are shown in Table 1.3 (*82*, *93*).

Table 1.3: **Examples of the most commonly used reagents for protein precipitation.**

Precipitants	Typically used reagents
Acids	Trichloroacetic acid
Metal Ions	Calcium (Ca^{2+}), Copper (Cu^{2+}), Zinc (Zn^{2+})
Organic Solvents	Acetone, acetonitrile, ethanol, methanol
Salts	Ammonium sulphate

Protein solubility is affected by the solutions pH. At the pH where the protein has no net surface charge, its isoelectric point (pI), their solubility will be at a minimum, often resulting in their precipitation. Also, extremely low pH by the use of acids such as trichloroacetic acid (TCA) can cause proteins to precipitate (*20*, *76*, *82*).

Metal ions can also be used to precipitate proteins. Divalent metal ions form complexes between protein monomers by selectively binding to amino acid side chains on the proteins surface. This aggregation continues until macroscopic clusters are formed, resulting in precipitation. The rate of aggregation is shown to increase with the number of metal binding sites on the proteins surface, as well as the metal ion concentration. However, as formation of complexes depends on the presence of unbound species, the latter is only to a certain point (*56*). Some typically used metal ions are calcium, copper, and zinc. Zinc and copper have been shown to bind the side chains of histidine and cysteine residues (*35*), while calcium has an affinity towards oxygen atoms, and binds to those both in the peptide backbone and amino acid side chains (*25*).

Organic solvents cause proteins to precipitate by dehydration. The ethanol fractionation method for plasma proteins was pioneered by Cohn et al. in 1946 (*21*). The intent was to purify albumin, and other plasma proteins, to be used as a substitute for human plasma in blood transfusion. This process is based on the concentration of ethanol, where protein solubility is reduced with the addition of water miscible organic liquids, as well as pH, conductivity, and temperature. To prevent protein denaturation, ethanol should be kept below 0°C. Illustrated in Figure 1.7, the major components of plasma proteins are separated into fractions as follows: fibrinogen in Fraction I, γ- and β-globulins in Fraction II+II, α-globulins in Fraction IV-1, remnant α- and β-globulins in Fraction IV-4, and lastly albumin in Fraction V, which has the highest solubility and lowest isoelectric point. Albumin can further be purified by washing the precipitate in the last fraction (*21*).

```
Plasma
    ├── I: Fibrinogen
    ├──── II + III: γ- and β-globulins
    ├────── IV-1: α-globulins
    ├──────── IV-4: α- and β-globulins
    └────────── V: Albumin
```
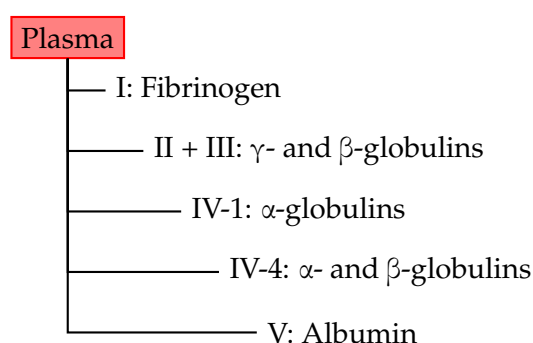
Figure 1.7: **Major plasma proteins in each fraction from the Cohn process of plasma fractionation (*21*).** Plasma can be fractionated into fibrinogen (I), γ- and β-globulins (II+III), α-globulins (IV-1), α- and β-globulins (IV-4), and albumin (V) using the Cohn process.

Precipitation with salts, commonly called salting out, is one of the oldest precipitation methods and still widely used today. Treatment with high salt concentrations disrupts the hydrogen bonds between the solvent and protein, causing them to precipitate. One advantage of salt precipitation is that the proteins retain their native conformation, but the sample must be desalted before further analysis can commence. Ammonium sulphate is a salt that is commonly used in protein precipitation and has also been applied to albumin depletion (*12*, *76*, *93*).

Protein Chromatography and Gel Electrophoresis

Several alternative methods of fractionation exist, including gel electrophoresis (GE), ion exchange chromatography (IEX), and size exclusion chromatography (SEC) (*34*). Gel electrophoresis can be used to separate proteins by size (sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE)), charge (isoelectric focusing (IEF)), or both with two-dimensional gel electrophoresis (2-DE). Migration through the gel commences when an electric field is applied, as negatively charged proteins migrate toward the anode and positively charged proteins toward the cathode. During IEF, a pH gradient will cause the amphoteric proteins to align themselves according to their pI. At that point they become neutrally charged, and thus the electric field no longer affects them. For SDS-PAGE the proteins are saturated in SDS buffer, which gives them a constant negative charge. They then migrate through the porous gel where larger proteins are deferred more than smaller proteins. 2-DE is accomplished by first performing IEF, then incubating the gel in SDS buffer before applying it to an SDS-PAGE gel (*40*). Gel electrophoresis provides a manner to visualize the sample prior to MS by staining the gel. However, it is labor intensive and in-gel digestion becomes necessary.

Size exclusion chromatography separates proteins based on molecular weight. The mobile

phase runs continuously through the matrix of the stationary phase, where fractions are created based on the travel time of the proteins through the column. Smaller proteins are retained by entering into the pores of the matrix beads, resulting in the larger proteins eluting faster. Depending on the pore size of the packing materials, different sized species can be discriminated (*14*). Ion exchange chromatography, on the other hand, separates proteins by surface charge. The stationary phase contains charged functional groups to which proteins with a net opposite charge bind. A salt or pH-gradient is then applied as the mobile phase, which causes the proteins to elute starting with those with the weakest ionic interaction to the column matrix. The strength of the ionic interaction depends on the functional groups used for the matrix, as well as the number and spacing of the protein charges (*37*).

Peptide Fractionation

It is also possible to fractionate the peptide sample after protein digestion. This can be achieved by different chromatographic methods. Hydrophilic interaction liquid chromatography (HILIC), ion exchange chromatography, and reversed-phase (RP) chromatography are the most commonly used methods for peptide fractionation.

HILIC separates analytes according to their hydrophilicity, using a hydrophilic stationary phase and a highly organic mobile phase. The elution gradient increases in polarity, by raising the concentration of aqueous solvent, causing the most hydrophobic peptides to elute first (*16*).

Ion exchange chromatography is also used to separate peptides, following the same principle as for proteins. In strong cation exchange chromatography (SCX), the peptides protonated basic amino acid residues interact with the sulfonate groups in the stationary phase, while deprotonated acidic amino acid residue interact with quaternary amino groups in strong anion exchange chromatography (SAX) (*71*). A pH- or salt-gradient is then used to elute the peptides in order of their charge.

In contrast to HILIC, RP chromatography separates peptides by hydrophobicity. It employs a hydrophobic or nonpolar stationary phase, typically alkyl chains linked to silica, and hydrophilic or polar mobile phase. Here the elution gradient is made with increasing amounts of organic solvent to the starting polar solvent. As such, the more hydrophobic peptides bind to the resin and the hydrophilic peptides elute first (*13*). This can be performed with a low-pH (LpH) or high-pH (HpH) gradient.

Peptide separation by HpH-RP chromatography can be accomplished either by using high performance liquid chromatography (HPLC) or resin-based spin columns. The stationary phase of RP chromatography is less polar than the mobile phase, allowing for the retention of analytes through hydrophobic interactions. The C18 silica-based reversed-phase material is the most commonly used stationary phase and can be made into beads for resin-based fractionation or made into columns for HPLC. Polar solvents, such as alcohols, acetonitrile, or water, are used to elute analytes from the least to most hydrophobic (*107*). Using an HPLC allows for the automatic sample separation and fraction collection, but necessitates instrument maintenance. It also enables the use of an online procedure, although this requires advanced instrumental setup and that the mobile phases are compatible. It is simpler, though more laborious, to perform offline HPLC separation (*43*). By using offline HpH-RP HPLC peptide fractionation, Bekker-Jensen et al., 2017 (*8*) were able to identify more than 12,000 protein groups from the proteome of the HeLa cervix carcinoma cell line which is comparable to next generation RNA-sequencing technology. Resin-based spin columns are more labor intensive, and can only be performed offline, but do not rely on the use of an HPLC instrument.

Before MS analysis, the peptide sample is commonly separated using RP-HPLC with a low-pH gradient. As such, it is optimal to use a peptide fractionation method that has high orthogonality to low pH RP-HPLC with this final dimension of separation. Orthogonality is driven by the separation characteristics of each method. Peptides being separated by charge in SCX/SAX versus hydrophobicity in RP promotes orthogonality. HILIC and RP use opposing aspects, hydrophilic and hydrophobic interactions, respectively. As pH affects the charge of basic and acidic amino acid residues, the high-pH to low-pH gradient alters binding affinity. Yeung et al., 2020 (*117*) found that the overall orthogonality of the previously mentioned methods to LpH-RP increases as follows: SCX < HILIC < HpH-RP < SAX. In terms of both orthogonality and resolution, HpH-RP yielded the best result. It is also worth keeping in mind the importance of the compatibility of the eluents between the first and second dimension separation. When a salt-gradient is used for SCX/SAX, the samples will require desalting before the LC-MS can commence (*71*). HILIC is also unsuitable for direct coupling to RP-LC-MS, as each fraction will have high concentrations of organic content (*16*). The direct combination of samples from HpH-RP and LpH-RP would lead to the formation of salts. However, both HILIC and HpH-RP can be performed with volatile solvents that can easily be exchanged by evaporation (*18*).

Whether protein-level or peptide-level separation is preferable depends on the objective. The main difference between the two approaches is that proteins are more sensitive to precipitation and degradation, which may be undesirable. Peptide separation, on the other hand, will result in multiple fractions containing peptides from the same parent protein which can reduce the number of proteins identified (*34*). However, the biggest disadvantage to any fractionation method is that it increases the mass spectrometry run time.

1.3.3 Reduced sample complexity

High-throughput (HT) analysis is highly desirable for the analysis of serum and plasma samples, because large sample collections exist, e.g., in biobanks. The aim would be the analysis of multiple samples per day by the same method. Automation of analytical methods, both sample preparation and analysis, is the main approach to enable high-throughput (*107*). However, the identified number of proteins must also be as high as possible, which is very challenging with plasma and serum. To achieve this, an option would be to reduce the sample complexity by enrichment of specific proteins or peptides. Here, proteins or peptides with distinct attributes can be intentionally acquired through specific enrichment techniques.

A commonly used method is immunoenrichment (*115*), using antibodies that specifically captures target proteins. The disadvantage with such techniques is that only known proteins can be targeted. Instead, enrichment methods targeting broader protein categories, such as glycoproteins (*120*) or phosphoproteins (*57*), can be performed. However, this approach is best suited for studies in which the targeted attribute is of interest. For hypothesis-free studies the ideal is to find an enrichment method that does not discriminate against unknown protein species. This is perhaps possible by targeting peptides from protein digests for enrichment. To achieve this, methods have been developed to enrich the terminal peptides of the proteins or specific amino acids containing peptides.

Terminomics

One way to enrich specific peptides is by targeting the protein end terminus, a field known as terminomics. This can be accomplished by either targeting the N-terminal amine group, N-terminomics, of the carboxyl group at the C-terminus, C-terminomics (*90*). These targets are especially suitable when studying proteolysis or other terminal PTMs. Theoretically, the full human proteome could be identified with around 20,000 N-terminal or C-terminal peptides.

If both approaches are combined, only 40,000 peptides can be generated and could be used to identify every human protein with two peptides. In practice, it has not yet been shown that these numbers were even approximately reached.

In general, the amine group at the N-terminus is much more reactive than the carboxyl group at the C-terminus. Furthermore, aspartic acid and glutamic acid have a carboxyl group in the side chain, and it is not possible to differentiate these from the C-terminus by labelling reaction. In contrast, the N-terminal amine and the amine of lysines can be differentially labelled because of differences of these sites in pH (84). Consequently, several methods for N-terminomics have been developed in contrast to few approaches for C-terminomics. N-terminal combined fractional diagonal chromatography (COFRADIC) (42) and N-terminal amine isotopic labeling of substrates (N-TAILS) (61), both illustrated in Figure 1.8, have been the most used methods for N-terminomics, but variants and other novel methods have been published as well (11).
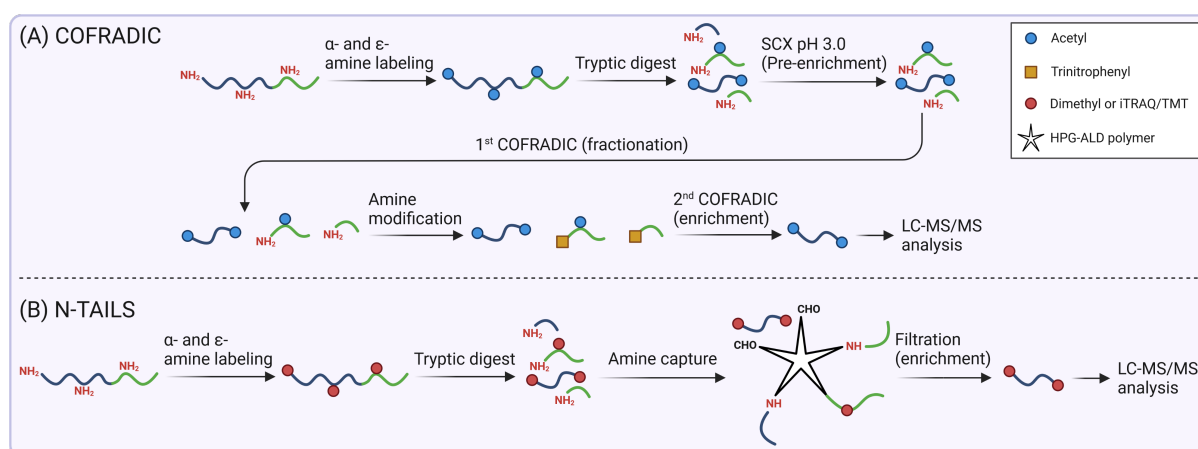
Figure 1.8: **Schematic representation of leading strategies for negative enrichment of N-termini (36).** (A) COFRADIC and (B) N-TAILS. Enrichment of N-terminal peptides is achieved by primary amine protection of intact proteins, followed by protein digestion and removal of internal and C-terminal peptides. (*Created with BioRender*).

Most of the methods are based on negative selection, i.e., by removing other types of peptides. The original N-terminal COFRADIC approach included cysteine alkylation, amine groups acetylation, followed by digestion of the proteins with trypsin which only cleaves after arginines. As a result, the N-terminal peptides of the proteins are acetylated (in vitro and in vivo) and the remaining internal peptides contain a free amine group. Then, this peptide mixture is separated by RP-C18 LC and fractions are collected. Next, the fractions are treated with a reagent which adds a very hydrophobic group on the free amines, but not on the acetylated peptides from the N-termini. Modified internal peptides elute later from the column, whereas the N-terminal peptides will have the same retention times when using a second RP-C18 LC run with the same conditions. For the original TAILS approach, dimethylation was used to block amine groups, followed by digestion with trypsin. Internal peptides were covalently attached to dendritic polyglycerol aldehyde polymers (HPG-ALD) and these polymers were removed by centrifugation. The N-terminal peptides remain in solution and can further be analyzed by LC-MS. Around 1,000 – 2,000 identified N-terminal peptides were reported using N-TAILS (83, 119). Recent further developments of TAILS for N-terminomics include the N-terminal peptide enrichment method (Nrich) which enabled the identification of more than 5,000 N-terminal peptides from 100 μg of human cell lysate (59). A scale down tip-based version of it (tipNrich) revealed 177 proteins from 1 μL of plasma (64). The first study of applying N-terminomics on serum and plasma proteomics obtained 222 proteins ranging over six orders of magnitude in abundance (111). There is so far a huge gap between the theoretical

number of identifiable protein termini and the experimental obtained results.

Enrichment of specific amino acid-containing peptides

A given amino acid can also be a possible target for enrichment. The optimal amino acid residue should be present in most proteins, but with low abundance to limit the number of peptides. If an average abundant amino acid such as lysine or arginine would be used, then the number of tryptic peptides would be halved, which is still quite high. Besides being a rare amino acid, it must have a chemical feature that allows it to bind with high affinity and specificity to a ligand of choice or to react with a chemical reagent. Therefore, potential candidates are cysteine, histidine, tyrosine, and tryptophan, due to being among the rarest occurring amino acids, as shown in Figure 1.9.
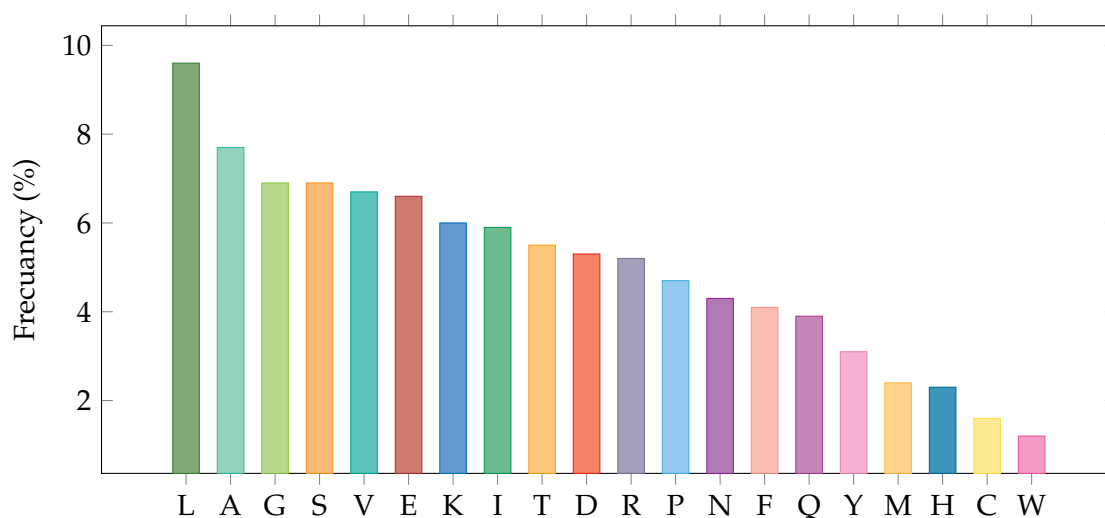
Figure 1.9: **Frequency (%) of amino acids in proteins (*94*).** The 20 essential amino acids are sorted from highest (left) to lowest (right) frequency they appear in proteins.

The chemical properties of the amino acid side chains, seen in Figure 1.10, can be used to capture residue-containing peptides. This can be achieved, for instance, by using thiol-, imidazole-, phenole-, or indole-affinity resin/tags or chemical derivatization reactions, for cysteine, histidine, tyrosine, or tryptophan, respectively. However, chemical derivatization techniques are sparse for histidine, tyrosine, and tryptophan.
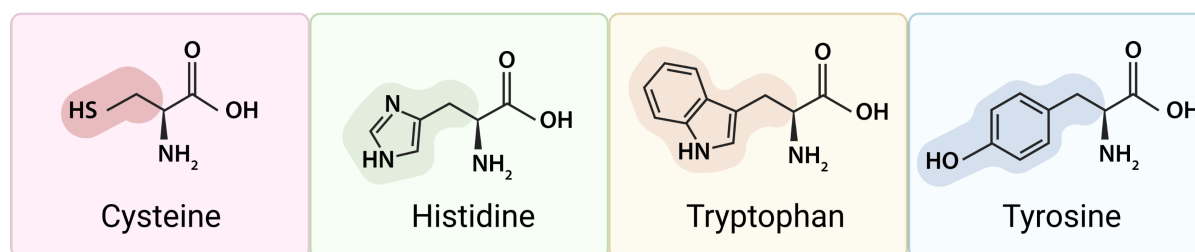
Figure 1.10: **Chemical structure of cysteine, histidine, tryptophan, and tyrosine.** (*Created with BioRender*).

Selective enrichment of tryptophan-containing peptides has been obtained by reversible malondialdehyde derivatization and solid-phase capture on hydrazine beads (*39*). However, aldehydes can also react with amines and the indol group of tryptophans is susceptible to oxidation reactions. The side chain of tyrosine has a phenolic group that is reactive in

electrophilic substitutions due to its hydroxyl group. Sulfur-triazole chemistry has been described for tyrosine-containing peptides (*48*). However, specific derivatization of the hydroxyl groups from serine, threonine, and tyrosines is challenging. Histidine is an effective nucleophilic catalyst because its side chain contains an imidazole group, a tertiary amine and strong base with acid dissociation constant (pK_a) near 7. Imidazole has a high affinity toward metal ions, such as copper, iron, and zinc. Still, the enrichment of single histidines with metal ion-based affinity material is not very specific. Specific chemical derivatization of histidine residues is also difficult as with reagents such as diethylpyrocarbonate which reacts with several other amino acids beside histidine (*80*). Thus, cysteine seems to be the most promising amino acid for selective enrichment because of the high reactivity of the thiol group on its side chain. In fact, a high diversity of derivatization reactions of cysteine have been established (*27*). In its ionized form, thiolate anion (-S$^-$), it is a powerful nucleophile with a pK_a of 8.7. The thiolate anion group is able to participate in many chemical reactions. The ability of cysteines to form disulfide bonds by oxidation makes it a critical antioxidant barrier of the cell. Many proteins rely on disulfide bonds, formed between two cysteine residues, to maintain their folded conformation. Although stable covalent bonds, a disulfide bond will spontaneously react with ionized thiol groups and can easily be broken by reducing reagents.

Enrichment Techniques for Cysteine-containing Peptides

The amine group of lysine and the thiol group of cysteine are the most reactive functional groups of the proteinogenic amino acids. Cysteines are rare amino acids, but present in 97% of all human proteins (*52*). Therefore, different approaches for the enrichment of cysteine containing peptides have been established, either on reversible or irreversible reactions.

(a) Thiopropyl Sepharose 6B

(b) Thiol reaction
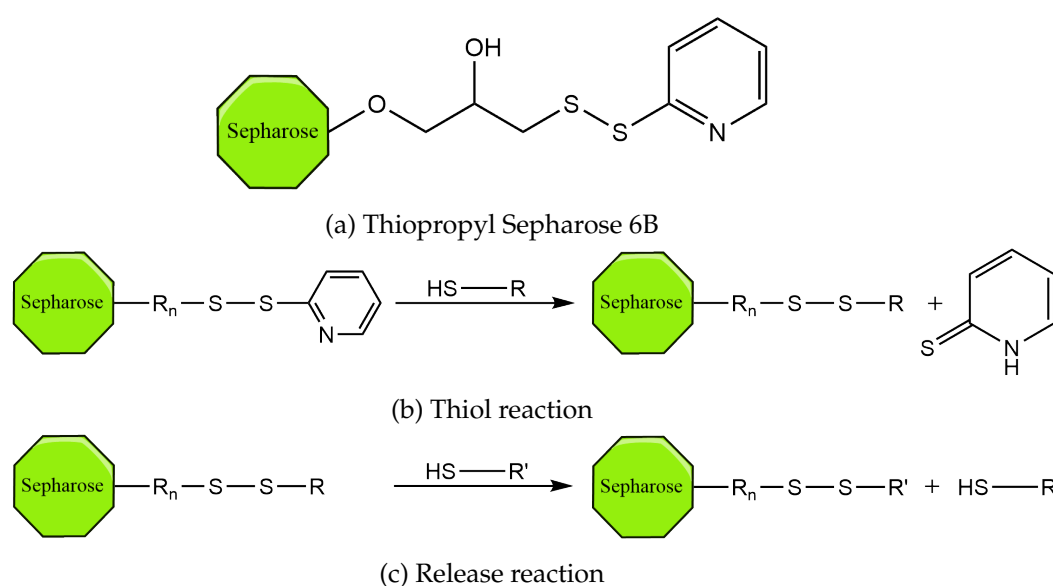
(c) Release reaction

Figure 1.11: **Chemical structure of cysteine-containing peptide enrichment reaction.** (a) Thiopropyl Sepharose 6B, (b) Thiol reaction with activated thiopropyl sepharose 6B, and (c) Second thiol reaction, upon which the first thiol is released. R represents a cysteine-containing peptide.

As with depletion or fractionation, it is also possible to use resin, polymer- or magnetic bead-based methods for enrichment. This is achieved as described previously, but with the reactive group for the peptide target used instead, immobilized onto the bead matrix of choice. For example, Figure 1.11b illustrated the reaction between a thiol group (HS-R) on the side chain of cysteine with the pyridyl disulfide reactive group attached to sepharose-based resin. Also

shown is the reverse reaction in 1.11c, where the bound peptide is released upon surplus addition of other thiol-containing reagents such as DTT.

Stable isotope labeling, often used in quantitative proteomics, is a labeling technique that can be combined with peptide enrichment. This method takes advantage of the difference in mass between isotopes to distinguish between samples. The isotope-coded affinity tag (ICAT) is the first isotopic tag which was employed for quantitative analysis of proteins (*47*). ICAT was further developed to the cleavable ICAT (cICAT) tag (*51*), as presented in Figure 1.12. cICAT consists of an affinity tag, a reactive group, and an acid-cleavable linker group, connecting the biotin moiety with the sulfhydryl reactive isotope tag. The reactive group binds specifically to sulfhydryl groups, found on the side chain of cysteine residues. The biotin affinity tag allows for the isolation of the bound cysteine-containing peptides. The linker region was originally developed with deuterium/hydrogen, as seen in the figure, but were later redesigned to be either heavy with 9 carbon-13 atoms, or light, containing 9 carbon-12 atoms. Two labeled samples, heavy and light, can be combined, analyzed by MS and compared relative to one another on intact peptide level (MS1) considering the intensities of cysteine-containing peptide pairs with a mass difference of 9 Da.
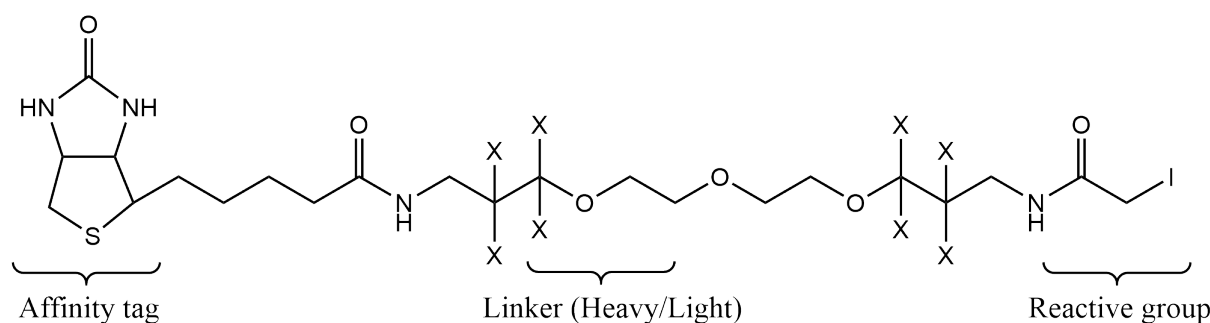
Figure 1.12: **Chemical structure of cICAT reagent (47).** cICAT consists of an affinity tag (biotin), a linker region where X = deuterium (Heavy) or hydrogen (Light), and a thiol-specific reactive group.

For isobaric labeling, the tandem mass tag (TMT) was developed by Thompson et. al (*104*) in 2002. These tags consist of a mass reporter, a mass normalizer, and a reactive group. An example is shown in Figure 1.13. The reactive group binds to target peptides. TMT reactive groups that label either amines of lysines and the peptide terminus or the thiol group of cysteines exist. The mass reporter is coupled to the mass normalizer through a linker, which is cleaved during collision-induced dissociation (CID). Thus, fragmentation (MS2 level) of the peptide moiety can be used for identification of the peptide whereas the reporter ions with a characteristic mass to charge ratio (*m/z*) ratio can be used for quantification of the peptide. The overall mass of the tag is balanced by the mass normalizer. These tags use the heavy isotopes ^{13}C or ^{15}N in the reporter and balancer group. Counterbalancing of the mass reporter and mass normalizer enables multiplexing with up to 18-plex for lysine labeling (*67*) and up to 6-plex for cysteine labeling (*85*), allowing for multiple samples to be analyzed simultaneously (*28*). In fact, automated TMT-16 plex plasma proteomics has been published (*41*). Nevertheless, using iodoTMT would even allow the reduction of the sample complexity. In combination with the analysis of six samples simultaneously with the 6-plex iodoTMT reagent, it is a promising strategy for HT proteome analysis of serum and plasma. Alternatively, cysteine-containing peptides could be enriched with e.g., iodoacetamido-LC-phosphonic acid (6C-CysPAT) labeling, tryptic digestion, labeling with amine-specific TMT reagents, enrichment by Fe^{3+}-nitrilotriacetic acid (NTA) and LC-MS analysis which would allow up to 18-plex analysis

(*69*). However, the TMT reagents are rather expensive and the application of iodoTMT or 6C-CysPAT to serum/plasma has not been reported as of now.

(a) IodoTMTzero

(b) Multiplex reagents for iodoTMT
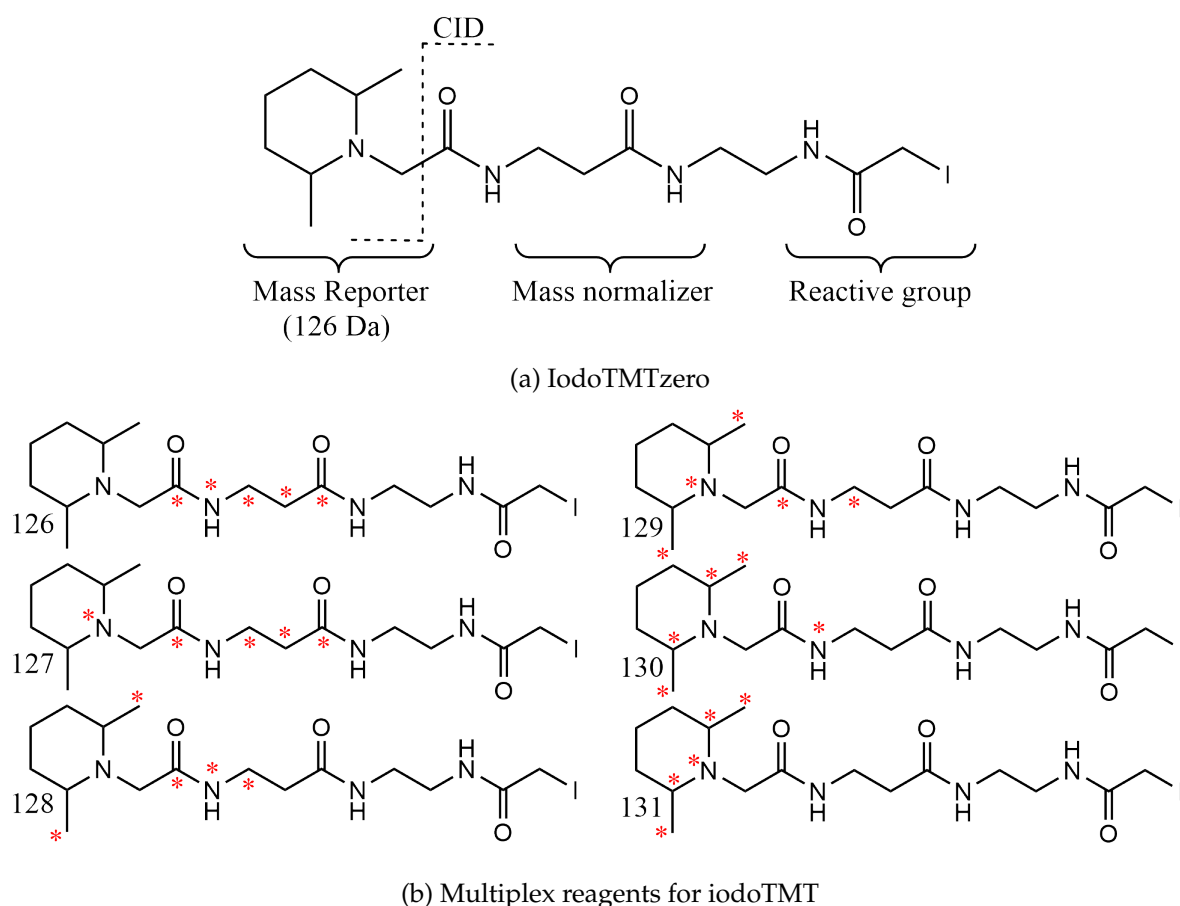
Figure 1.13: **Chemical structure of (a) the Thermo Scientific iodoTMT™zero label reagents and (b) multiplex reagents for iodoTMT.** IodoTMTzero consists of a mass reporter, mass normalizer, and a thiol-specific reactive group. When intact, the molecular weight of iodoTMTzero is 452.33 Da.

1.3.4 Protein Digestion and Purification

When using a bottom-up proteomics workflow, the protein sample must first be digested into peptides before analysis by MS. This is most commonly achieved by enzymatic digestion. The most specific endoproteinases are presented in Table 1.4. An endoproteinase cleaves a protein at one side of certain amino acids, making it possible to obtain peptides of a certain length and with known terminal amino acids. The less specific endoproteinases produce more and shorter peptides. This makes MS analysis more complex as smaller peptides are hard to annotate and the peptides may overlap in sequence (*101*).

Of the endoproteinases listed in Table 1.4, trypsin is the by far the most frequently used in proteomics. Trypsin cleaves peptide bonds on the C-terminal side of lysine (Lys) and arginine (Arg) with high specificity. This produces peptide fragments with at least two positive charges: the Lys/Arg side chain at the C-terminus and the amino group at the N-terminus. Since these residues are relatively abundant and distributed well in most proteins the resulting peptides are on average between 7 - 20 amino acids in length. This makes trypsin highly suitable for use in combination with CID/HCD-MS (*15*). However, trypsin is less suitable when dealing with proteins that have an overabundant amount or lack of Lys/Arg residues.

Table 1.4: **Examples of the most commonly used enzymes for protein digestion in proteomics.** Includes the amino acid specificity, which terminal side the cleavage is made, and restriction conditions to the proteolytic reaction.

Enzyme	Specificity	Cleavage side	Restriction[*]
Arg-C	Arginine	C-term	
Asp-N	Aspartic acid	N-term	Proline
Chymotrypsin	Leucine/Phenylalanine/Methionine/Tryptophan/Tyrosine	C-term	Proline
Glu-C	Glutamic acid/Aspartic acid	C-term	Proline
Lys-C	Lysine	C-term	Proline
Lys-N	Lysine	N-term	
LysargiNase	Arginine/Lysine	N-term	
Trypsin	Arginine/Lysine	C-term	Proline

[*] When the specific amino acid is followed by this residue, the protease may not cleave.

If the protein sample has been separated by gel electrophoresis, the digestion may be conducted directly by in-gel digestion (*95*). The main advantage of gel electrophoresis protein separation prior to digestion is that the gel can be stored safely for a long while, and that specific bands of proteins can be selected and cut out. Otherwise, it is easier to perform in-solution digestion.

1.4 Sample Analysis

1.4.1 Liquid Chromatography

Before mass spectrometry can commence for proteome analysis, the generated peptides are typically separated by high performance liquid chromatography. As mentioned previously, the most common approach is to resolve the peptides based on hydrophobicity using reversed-phase C18 columns. The stationary phase retains the peptides, which are then eluted by the mobile organic solvent phase. Applying a gradient of solvent ratios changes the polarity of the mobile phase gradually, allowing for the peptides to be separated. By reversed-phase LC, the hydrophobic peptides attain a longer retention time than the more hydrophilic. Both the time and elevation of the gradient can be optimized for a given sample and desired peptide partition. This is done in order to not overload the MS detector with all peptides at once. Consequently, more mass spectra can be recorded to achieve a higher resolution. Although this can be carried out offline, it is preferred to couple the two systems with an interface for online LC-MS analysis (*93, 107*).

1.4.2 Mass Spectrometry

Mass spectrometry is an analytical technique that measures the mass-to-charge ratio of ionized analytes in the gas phase. The main components of mass spectrometer are a source, an analyzer, and a detector. A schematic overview of these can be seen in Figure 1.14. The source functions as an ionization chamber that converts the sample to a stream of gas phase ions. A magnetic or electric field is sustained in the analyzer, ensuring uniform velocity amongst the incoming ions. With this, the ions are separated based on their *m/z* ratio, which is recorded by the detector (*93, 107*).
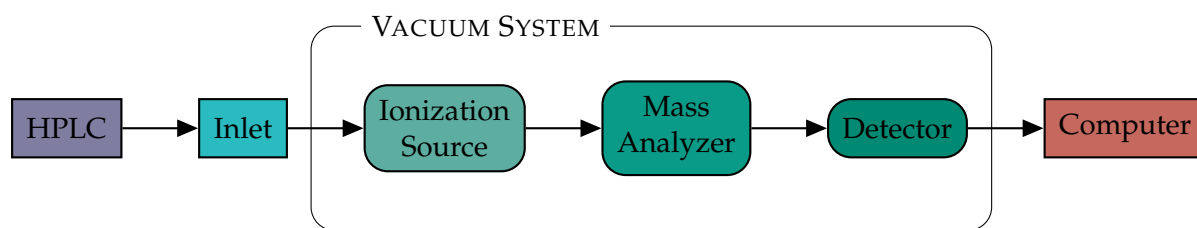
Figure 1.14: **General components of a mass spectrometer.** The components are in order of operation from left to right, including an introductory HPLC and connecting inlet, ionization source, mass analyzer and detector surrounded by a vacuum system, and a computer to which the data is sent.

Analytes can be ionized using beams of electrons, photons or neutral atoms, thermal heating, or electric fields. The two most widely used ionization techniques in MS-based proteomics are matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). MALDI was developed by Franz Hillenkamp and Michael Karas in 1985 (*60*). Here, the sample is embedded in a matrix, which is bombarded by high-energy ultraviolet (UV) laser radiation. This causes the matrix to evaporate, and a proton to be transferred to the sample. MALDI is mainly used in proteomics for single protein analysis after 2-DE protein separation. ESI, developed by John B. Fenn in 1989 (*38*), is an ionization technique based on evaporation driven by high voltage. The sample is converted into a spray of charged liquid droplets in an electric field, from which the solvent is evaporated. Ions are liberated from the droplets through Coulomb explosion. This mode generates species with multiple charges and is the preferred method for bottom-up proteomics because ESI can be directly coupled to an LC (*93, 116*).

Different mass analyzers vary in speed, sensitivity, resolution, mass and dynamic range, and ion transmission. The main types of analyzers for macromolecules are the quadropole (Q), ion trap, orbitrap, and time of flight (TOF) analyzers. A Q is made up of two pairs of rods connected by electricity (opposite) or voltage (adjacent) that run parallel to the ion bean. Specific *m/z* ratios are then selected by varying the radiofrequency of voltage, where all others will collide with the rods. An ion trap analyzer forces the ions into specific orbits depending on their *m/z* ratio by applying an electrostatic field. The ions may then be selectively ejected towards the detector by adjusting the field. The Orbitrap functions similarly, trapping ions in orbit around a central electrode spindle. The frequencies of the ions harmonic oscillations are then measured to determine their *m/z* values. In TOF analyzers, the *m/z* ratio is calculated from the time it takes for the accelerated ions to arrive at the detector at the end of a long tube (*50, 93*). In MS-based proteomics, mostly hybrid instruments are used, i.e., the combination of two different analyzers (Q-TOF, Q-orbitrap, and ion trap-orbitrap). An overview about MS instruments used for blood proteomics was published recently (*123*).

In proteomics it is common to apply tandem mass spectrometry (MS/MS). During MS1 the *m/z* ratio of the intact peptide is measured. Peptides are then passed on to further fragmentation in the collision cell. There exists two general approaches to data acquisition in proteomics: data dependent acquisition (DDA) and data independent acquisition (DIA). DDA is the most commonly used mode, which selects precursor peptides based on the intensity of their MS1 spectra for fragmentation. This mode has a bias towards abundant ions, but this can be corrected for by preventing repeated acquisition. DIA sends all precursor peptides on to fragmentation, alleviating the drawback of DDA, but making the resulting data more complex (*79*). The resulting ions are measured in MS2, producing a fragmentation pattern that is unique for each peptide. Fragment ions are classified as a, b, or c when derived from the N-terminal, and x, y, or z when derived from the C-terminal of the peptide, dependent on which bond

of the peptide backbone was fragmented, as shown in Figure 1.15. There are several types of fragmentation techniques. Collision-induced dissociation (CID) (*58*) and higher-energy C-trap dissociation (HCD) (*77*) are most commonly used, where fragmentation is caused by impact with an inert gas producing mainly b- and y-ions. Chemical fragmentation can be obtained by electron transfer dissociation (ETD) (*72*), where c- and z-ions are generated when electrons are transferred to the peptide through radical anions. The resulting raw mass spectra are then processed to extract the mass information.
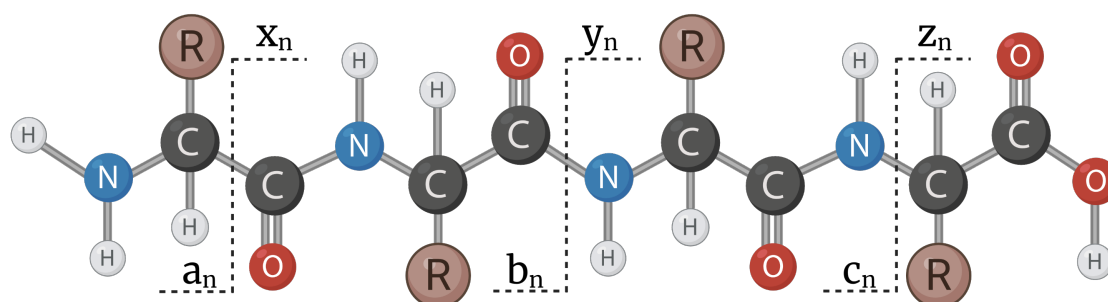
Figure 1.15: **MS2 fragment ions.** N-terminal fragment ions are named a, b, or c and C-terminal fragment ions are named x, y, or z, where «n» represents the number of amino acid residues in the molecule. (*Created with BioRender*).

1.5   Data Analysis

In order to identify proteins, the mass spectra must be compared to a protein sequence database. These are collections of peptide sequences from several sources that range in content, quality, annotation, and how the sequences have been discovered. One of the leading protein databases available is the Universal Protein Resource (UniProt) consortium (*22*). This consortium is comprised of three parts: The UniProt Knowledge base (UniProtKB), a combination of the unreviewed TrEMBL and reviewed Swiss-Prot databases, UniProt reference clusters (UniRef), a collection of sequence clusters, and UniProt Archive (UniParc), an archive of all known unique sequences. TrEMBL consists of a large array of sequences derived from translations of gene sequences and are computer annotated. The Swiss-Prot database (*7*) is manually curated by experts, which makes it a high-quality, albeit smaller, database without redundant protein sequences. Protein sequences in the Swiss-Prot database are also highly annotated, with information about the proteins function, structure, PTMs, expression, etc. As of 2022, UniProtKB has recorded over 227 million sequences. The human Swiss-Prot database was fully curated a few years ago.

Matching the MS data up against a protein database is performed using a search engine. Andromeda (*23*), Mascot (*81*), MSFragger (*62*), PEAKS (*70*), SEQUEST, and X!Tandem (*24*) are examples of such search engines that are commonly used. They use a protein identification approach called theoretical matching, which uses *in silico* digestion of the protein database to generate theoretical fragmentation patterns. These are then used as a basis to compare the experimental fragment masses. First, the statistical probability between the theoretical peptide masses and recorded data is calculated by the algorithm and reported. A match is made based on a set identification threshold, as well as variable modifications and protease miscleavages. A fitting score is then calculated. The resulting identified peptides are then used to generate a list of precursor proteins, which use a decoy database to calculate a false discovery rate (FDR). Database search results can then be quantified using software tools, such as MaxQuant/Perseus (*106*), PEAKS (*70*), and Scaffold (*91*).

2 Aim of The Project

This project has a bipartite aim. The first aim was to develop an approach for large-scale proteome analysis of serum proteins, that is, to improve the number of identified proteins in serum. The second aim was to establish a method for high-throughput proteome analysis of serum proteins in order to analyze more samples in a shorter amount of time.

Three main objectives were investigated. 1. depletion of abundant proteins, 2. protein and/or peptide fractionation for the first aim and 3. reduction of the sample complexity by enrichment of cysteine-containing peptides for the second aim. Several sample preparation methods subject to these objectives were investigated to find an experimental strategy for improved serum proteome analysis.

3 Materials and Methods

3.1 General

3.1.1 Acetone Precipitation

Protein samples in solution were precipitated in 4x the samples volume of cold acetone, briefly vortexed, and stored overnight at -32°C. They were then centrifuged for 30 minutes at 4°C, 16,000 relative centrifugal force (rcf). The supernatant was discarded, and the pellets were left to dry for a short while.

3.1.2 In-Solution Protein Digestion

The protein samples (10 µL serum) were diluted with 50 µL freshly made 6 M urea, 100 mM ammonium bicarbonate (ABC), and 2.5 µL 200 mM dithiothreitol (DTT), 0.1 M tris(hydroxymethyl)aminomethane - hydrochloride (tris-HCl), pH 8, was added before incubating for 30 minutes at 30°C. 7.5 µL freshly made 200 mM iodoacetamide (IAA) was then added to each sample, before 1 hour incubation in the dark. Then, 10 µL 200 mM DTT was added and the samples were once again incubated for 30 minutes at 30°C. Lastly, the samples were diluted with 200 µL 50 mM ammonium bicarbonate before adding 5 µg Trypsin Gold (Promega Corporation) in 20 µL 20 mM ammonium bicarbonate to each sample. They were then incubated overnight, for at least 16 hours, at 37°C, after which 5 µL 50% formic acid was added to stop the digestion reaction.

3.1.3 Micro-solid phase extraction (SPE)

10 µL organic wash (OW) [0.1% formic acid:acetonitrile (1:2)] was pipetted into microfuge tubes, one per sample. The ZipTip® µ-C18 (Millipore corporation, Billerica) was then conditioned, first with 10 µL OW and then 10 µL 0.1% formic acid. 50 µL sample was then loaded twice onto the ZipTip µ-C18, before washing twice with 10 µL HPLC grade water and then eluting the sample in one of the prepared tubes. This was repeated for the entire sample. The ZipTip µ-C18 was conditioned again before the next sample. After completing all samples, they were dried at 30°C for 15 minutes using a SpeedVac.

3.1.4 Liquid Chromatography - Mass Spectrometry (LC-MS)

Dry peptide samples were dissolved in 4 µL 0.1% formic acid and transferred to microvials. 2 µL of the samples were analyzed by LC-MS using a timsTOF Pro (Bruker Daltonik, Bremen, Germany) which was coupled online to a nanoElute nanoflow liquid chromatography system (Bruker Daltonik, Bremen, Germany) via a CaptiveSpray nanoelectrospray ion source. The peptides were separated on a reversed-phase C18 column (25 cm x 75 µm, 1.5 µm, PepSep (Bruker Daltonics, Bremen, Germany)). Mobile phase A contained water with 0.1% formic acid, and acetonitrile with 0.1% formic acid was used as mobile phase B. The peptides were separated by a gradient from 0 - 35% of mobile phase B over 60 min at a flow rate of 300 nL/min at a column temperature of 50°C. MS acquisition was performed in DDA-parallel accumulation-serial fragmentation (PASEF) mode. The capillary voltage was set to 1.5 kV with a mass range of 100 to 1700 m/z. The number of PASEF ranges was set to 20 with a total cycle time of 1.16 s, charge up to 5, target intensity of 20,000, intensity threshold of 1,750, and active exclusion with release after 0.4 min. An inversed reduced trapped ion mobility spectrometry (TIMS) (1/k0) of 0.85 - 1.40 Vs/cm2 was used with a range time of 100 ms, an accumulation

time of 100 ms, a duty cycle of 100%, and a ramp rate of 9.51 Hz. Precursors for DDA were fragmented with an ion mobility-dependent collision energy, which was linearly increased from 20 to 59 eV.

3.1.5 Data Analysis

The spectral peaks were analyzed using Mascot Daemon search engine (Matrix Science, version 2.7.0.1) against the SwissProt human database. The following search parameters were utilized: enzyme: trypsin, maximum missed cleavages: 1, peptide charge: 2+ and 3+, peptide tolerance: 15 ppm, MS/MS tolerance 0.03 Da, fixed modifications: carbamidomethylation on cysteine, variable modifications: acetylation at N-terminus and oxidation on methionine. Scaffold (version 5.1.2, Proteome Software™, Inc., Portland, OR 97219, Oregon, USA) was then used to validate the peptide and protein identifications assigned by Mascot. The peptide identification threshold was set to 95.0% FDR and the protein identification threshold to 99.0%.

3.2 Depletion of Abundant Proteins

For the depletion of albumin and/or IgG, five different depletion kits were used: The Albumin/IgG Removal Kit (Pierce™, Thermo Scientific), AlbuSorb™ PLUS, AlbuVoid™ LC-MS On-Bead, Albuvoid™ PLUS (Biotech Support group, Monmouth (JCT, NJ, USA), and BioMag® ProMax (Bangs Laboratories, Fishers, IN, USA). An overview of the depletion kits can be seen in Table 3.1. Depletion was performed using 10 µL human serum protein sample (Sigma-Aldrich, H4522), followed by in-solution or on-bead digestion, then micro-solid phase extraction (SPE) and LC-MS. Before digestion, the samples were divided in two, where one fraction was precipitated with acetone and the other was not. Each depletion method was performed in accordance with the protocol of the manufacturer using the provided buffers, with minor adjustments to make them more comparable. First a test-run of each kit was performed, from which samples were taken to perform SDS-PAGE. Three replicates of each kit were then carried out.

Table 3.1: **Depletion kit overview:** A list of all depletion kits with their depletion target, capturing agent, and company.

Depletion Kit	Target	Capturing agents	Company
The Albumin/IgG Removal Kit	Albumin & IgG	Cibacron Blue 3GA /Protein A	Pierce™ Thermo Scientific
AlbuSorb™ PLUS	Albumin & IgG	Polyelectrolytes /Protein A	Biotech Support group
AlbuVoid™ LC-MS On-Bead	Albumin	Polyelectrolytes	Biotech Support group
Albuvoid™ PLUS	Albumin & IgG	Polyelectrolytes /Protein A	Biotech Support group
BioMag® ProMax	Albumin or IgG	Magnetic beads	Bangs Laboratories

3.2.1 The Albumin/IgG Removal Kit (Pierce™)

170 µL immobilized cibacron Blue 3GA/Protein A gel was transferred to spin columns placed in microcentrifuge collection tubes, which were then centrifuged at 10,000 x rcf for 1 minute to get rid of the storage buffer. 10 µL serum samples was diluted to 75 µL with Binding/Wash Buffer [25mM Tris, 25 mM NaCl, 0.01% NaN$_3$; pH7.5] before being transferred to the columns. After vortexing briefly, they were then incubated for 10 minutes at room temperature while

shaking. Next, the columns were centrifuged at 10,000 x rcf for 1 minute, after which the filtrate was reapplied to the resin and set to incubate for another 10 minutes. Filtrate containing sample depleted of HSA and IgG was then collected by centrifugation at 10,000 x rcf for 1 minute. Another 75 µL Binding/Wash Buffer was added to the columns, which were then centrifuged at 10,000 x rcf for 1 minute, and the filtrates were pooled together.

3.2.2 AlbuSorb™ PLUS Albumin and IgG Depletion Kit

60 mg AlbuSorb™ PLUS powder was weighed out into microfuge spin-filters. The powder was then conditioned by adding 400 µL Binding Buffer 1 (BB1) [0.05M K_2HPO_4 Dibasic, pH 7.5], before vortexing for 3 minutes and centrifuging at 1,000 x rcf for 2 minutes, discarding the filtrate. This was performed twice. Then 250 µL BB1 and 20 µL (test run) or 10 µL (replicate run) serum samples were applied to the columns. After 10 minutes incubation while shaking, the columns were then centrifuged at 9,000 x rcf for 4 minutes, collecting the filtrate containing depleted serum.

3.2.3 AlbuVoid™ PLUS Albumin and IgG Depletion Kits

IgG Depletion 60 mg NuGel™ Protein A beads were weighed out into 0.45µ SpinX® centrifuge tube filters, then 400 µL Buffer 1 (undisclosed composition) was added before vortexing for 3 minutes and centrifuging for 3 minutes at 2 300 rcf. 20 µL (test run) or 10 µL (replicate run) serum diluted in 250 µL Buffer 1 was added onto the beads and vortexed for 10 minutes before centrifugation at 9,300 rcf for 4 minutes, collecting the filtrate. Then 100 µL Buffer 1 was added to the beads, which were then vortexed for 10 minutes and centrifuged at 9,300 rcf for 4 minutes, combining the two filtrates.

Albumin Depletion 25 mg AlbuVoid™ beads were weighed out into 0.45µ SpinX centrifuge tube filters before adding 250 µL AlbuVoid Binding Buffer (AVBB), pH 6.0, (undisclosed composition). The tubes were then vortexed for 5 minutes and centrifuged at 2,300 rcf for 3 minutes discarding the supernatant, twice. The combined filtrated from the previous IgG depletion diluted with 175 µL AVBB was then added to the beads, before being vortexed for 10 minutes and centrifuged at 9,300 rcf for 4 minutes, collecting the HSA containing flowthrough for SDS-PAGE. Following this, 250 µL AVBB was added prior to vortexing for 5 minutes and centrifuging for 4 minutes at 9,300 rcf, then again with 250 µL AlbuVoid Wash Buffer (AVWB), pH 7.0, (undisclosed composition). One sample, to be used for SDS-PAGE, was eluted by adding 200 µL AlbuVoid Elution Buffer (AVEB), vortexing for 10 minutes, and centrifuging for 4 minutes at 9,300 rcf. The other samples proceeded to On-Bead digestion.

On-Bead Digestion 5 µL 200 mM DTT mixed with 100 µL AVWB was added to the beads before vortexing for 10 minutes and incubating for 30 minutes at 60°C. After the samples were cooled to room temperature, 100 µL 20 mM IAA was added and the samples were incubated for 45 minutes in the dark. The tubes were then centrifuged at 9,300 rcf for 4 minutes, discarding the filtrate, and rinsed with 500 µL 50% acetonitrile. 1 µg Trypsin Gold dissolved in 20 µL 20 mM ammonium bicarbonate and 180 µL AVWB was then transferred to the beads, before incubation at 37°C overnight. The albumin/IgG depleted filtrate was collected by centrifuging at 9,300 rcf for 4 minutes, then adding 150 µL 10% formic acid to the beads, vortexing for 10 minutes and centrifuging again, combining the filtrates.

3.2.4 AlbuVoid™ LC-MS On-Bead Albumin Depletion Kit

25 mg AlbuVoid™ beads were weighed out into 0.45µ SpinX centrifuge tube filters. 125 µL Binding Buffer AVBB (undisclosed composition) was added before vortexing for 5 minutes,

followed by centrifugation at 800 rcf and discarding of the filtrate. This was performed twice. 20 µL (test run) or 10 µL (replicate run) serum was diluted in 100 µL AVBB and transferred to the spin-tubes. The spin-tubes were then vortexed for 10 minutes before being centrifuged at 9,300 rcf for 5 min, collecting the HSA filtrate for SDS-PAGE. Next, the beads were washed twice with 250 µL AVWB (undisclosed composition), vortexed for 5 minutes, and centrifuged at 9,300 rcf for 4 minutes. For SDS-PAGE, one sample was eluted with 200 µL AVEB, vortexing for 10 minutes and centrifuging for 4 minutes at 9,300 rcf. To the rest, 95 µL AVWB and 5 µL 200 mM DTT was then added before vortexing for 10 minutes and incubation at 60°C for 30 minutes. After cooling, 80 µL AVWB and 20 µL 200 mM IAA was added before incubating in the dark at room temperature for 45 minutes. The samples were then centrifuged at 9,300 rcf for 5 minutes and the supernatant was discarded. Finally, 1 µg Trypsin Gold dissolved in 20 µL 20 mM ammonium bicarbonate and 80 µL AVWB was added to the beads and incubated at 37°C overnight. The depleted peptide filtrate was retained by centrifugation at 9,300 rcf for 5 minutes. Then 150 µL 10% formic acid was added to the beads, which were then vortexed for 10 minutes and centrifuged at 9,300 rcf for 5 minutes, and the filtrates were combined.

3.2.5   BioMag® ProMax Albumin and IgG Removal Kits

Albumin Depletion   35 µL Binding/Wash buffer (undisclosed composition) and 10 µL serum were transferred to a microcentrifuge tube and mixed thoroughly, then 50 µL BioMag® ProMax re-suspended particles were added, and the samples were vortexed for 10 minutes. The particles were pelleted using a magnetic rack (Invitrogen DYNAL®, Thermo Scientific) and the albumin containing supernatant collected. Then the particles were washed 3 times by re-suspending them in 500 µL Binding/Washing buffer and pelleted via magnetic separation to remove the supernatant. Finally, 100 µL Elution Buffer was added to the beads, which were then vortexed for 10 minutes before pelleting and collecting the depleted supernatant.

IgG Depletion   59 µL ProMax Serum IgG Removal Binding/Wash Buffer (undisclosed composition) was transferred to a microcentrifuge tube. Then 1 µL serum, or the entire albumin depleted sample, was added, along with 40 µL BioMag® ProMax particles. The samples were then vortexed for 10 minutes before pelleting by magnetic separation, and the IgG depleted fractions were collected.

3.2.6   Sodium dodecyl sulfate polyacrylamide electrophoresis

After serum depletion, 10 µL of the respective depleted filtrates were mixed with 10 µL Laemmli Sample Buffer (Bio-Rad Laboratories, Inc.). 50 µL sample buffer was applied to the beads of the Albumin/IgG removal kit (Pierce), AlbuSorb PLUS, AlbuVoid PLUS IgG and BioMag ProMax IgG depletion kits and eluted in accordance with the last step of their respective protocols. For the AlbuVoid LC-MS On-Bead, AlbuVoid PLUS, and BioMag ProMax albumin removal kits, 5 µL albumin containing filtrate was mixed with 5 µL sample buffer. 1 µL serum was added to 10 µL sample buffer per kit. Then all samples were heated at 95°C for 3 minutes and briefly spun down. 10 µL of the samples from each kit were loaded onto separate NuPAGE™ 4-12% Bis-Tris gels (Novex™, Life Technologies), along with 3 µL Precision Plus Protein™ Standard Dual Color marker (Bio-Rad Laboratories, Inc.). The SDS-PAGE running chambers were then filled with NuPAGE™ MOPS SDS Running Buffer (Novex™, Life Technologies) diluted to 1x concentration, and run at 200V for 35 minutes. The gels were then stained with Quick Coomassie® Stain (Protein Ark, Calibre Scientific UK) for 2 hours, and washed in MQ water overnight.

3.3   Fractionation Methods

Three protein fractionation and two peptide fractionation methods were explored: ethanol protein precipitation, size exclusion chromatography, and zinc chloride (ZnCl$_2$) protein precipitation, and HpH-RP peptide fractionation using PepSwift™ Monolithic Capillary HPLC column or Pierce™ spin column fractionation kit (Thermo Scientific). An overview of all the fractionation methods is shown in Table 3.2. Three replicates of 10 µL serum was used as starting material for all methods. The protein fractionation methods were performed prior to in-solution protein digestion with Trypsin Gold followed by micro-SPE, while the peptide fractionation methods were performed after. All samples were purified by micro-SPE before LC-MS analysis.

Table 3.2: **Fractionation methods overview.** A list of the sample type, methods, base of separation, number of fractions, and relevant company for all fractionation methods.

Sample	Method	Base of Separation	No. of fractions	Company
Protein	Ethanol precipitation	Organic solvent dehydration	7	NA
	Size exclusion chromatography	Molecular weight	20	ÄKTA™ (GE HealthCare Life Sciences)
	Zinc chloride precipitation	Metal-ion aggregation	2 - 4	NA
Peptide	High-pH reversed-phase liquid chromatography	Hydrophobicity	15	PepSwift™ (Thermo Scientific)
	High-pH reversed-phase spin column	Hydrophobicity	8	Pierce™ (Thermo Scientific)

3.3.1   The Cohn Process of Protein Fractionation

Figure 3.1 shows the overview of the ethanol (EtOH) concentration and pH used in order to extract each of the fractions. First, solutions with the desired final ethanol concentration and approximate pH were derived using Hydrion™ Microfine™ pH test paper [2.9 - 5.2, 4.9 - 6.9, and 5.5 - 8.0], resulting in the solutions shown in Table 3.3 below.

Table 3.3: **Precipitation solutions per sample for the Cohn process fractionation method.** The total volume of each precipitation solution, with the content (µL) of ethanol (EtOH), sodium acetate (NaOAC), acetate buffer, and milli-Q (MQ) water used.

Solution No.	EtOH	4M NaOAC	Acetate Buffer*	MQ	Total Volume
1	40 µL	100 µL		350 µL	490 µL
2	115 µL				115 µL
3			30 µL	210 µL	240 µL
4	315 µL	2 µL			317 µL
5			15 µL		15 µL
6	50 µL			450 µL	500 µL

* 10M acetic acid (AcOH)/4M NaOAC (1:1)

Next, to 10 µL serum, solution no. 1 was added and incubated at 4°C for 30 minutes. The sample was then centrifuged at -4°C, 16,000 rcf, for 15 minutes, and the supernatant transferred to a new tube. This process was repeated sequentially by adding solution no. 2 - 5 to the supernatant of the previous step. The pellet produced from solution no. 5 was dissolved in solution no. 6, and the final fractions were collected. The supernatant fractions V-1 and -2 were precipitated with acetone, as described previously, before in-solution protein digestion.
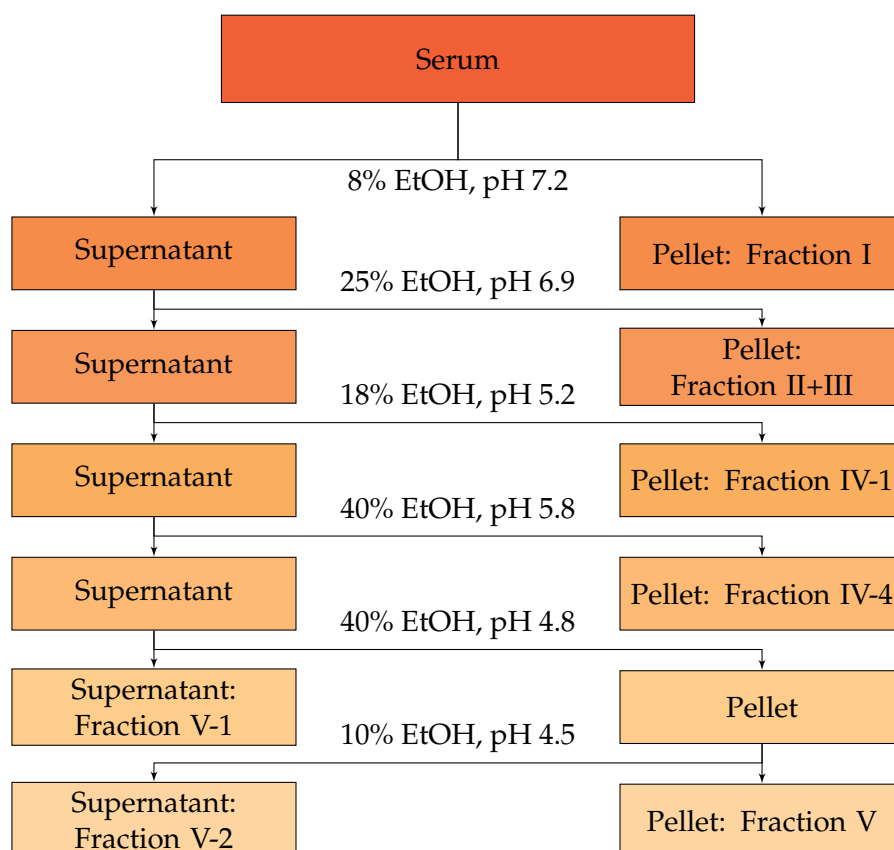
Figure 3.1: **The Cohn process of serum protein fractionation method overview (*21*).** The concentration of ethanol and pH used at each step in fractionation of serum, with the resulting fraction names adhering to those used in the Cohn process.

3.3.2 Size Exclusion Chromatography with ÄKTA™ Protein Fractionation

Size exclusion chromatography was performed using a Superdex® 200 Increase 10/300 GL column (GE HealthCare Life Sciences, US) on the ÄKTA™ chromatography (GE HealthCare Life Sciences, US). 1 L Tris buffer [50 mM tris-HCl, 1 mM NaCl, pH 7.5] was degassed for 15 minutes, before being used to condition the column for 30 minutes. 10 µL serum sample was then diluted to 20 µL with the Tris buffer, before being injected into the pre-column. The sample was fractionated over a run time of 30 minutes, with 0.5 mL fraction size, while measuring UV of 280 nm wavelength. Fractions were collected from time point 6 to 26 minutes, 40 in total. In order to reduce the number of fractions, two and two sequential fractions were combined. The fractions were then evaporated using a SpeedVac until about approximately 200 µL remained, then precipitated with cold acetone before in-solution protein digestion.

3.3.3 Zinc Chloride Sequential Protein Fractionation

Three procedures for $ZnCl_2$ sequential protein fractionation was performed, one with three increasing concentrations of $ZnCl_2$, the next with the two highest concentrations, and the final

with only the highest concentration. For the first procedure, 10 µL 0.002 mM $ZnCl_2$ was added to 10 µL serum and incubated for 30 minutes at room temperature. The sample was then centrifuged at 16,000 rcf for 10 minutes and the supernatant was transferred into a new tube. The pellet was washed with 20 µL 0.002 mM $ZnCl_2$ and centrifuged at 16,000 rcf for 2 minutes, twice. To the supernatant, 20 µL 0.2 mM $ZnCl_2$ was added before incubating at room temperature for 30 minutes. The sample was then centrifuged, and the pellet washed with 0.2 mM $ZnCl_2$ as described previously. The resulting supernatant was then incubated for 30 minutes at room temperature with 40 µL 20 mM $ZnCl_2$, and again centrifuged and washed. This resulted in three pelleted and one supernatant fraction. The next procedure started with adding 10 µL 0.2 mM $ZnCl_2$ to 10 µL serum, followed by adding 20 µL 20 mM $ZnCl_2$ to the resulting supernatant, giving two pellets and one supernatant fraction. In the last procedure 10 µL serum was incubated with 10 µL 20 mM $ZnCl_2$ directly, resulting in one pellet and one supernatant fraction. All of the final supernatant fractions were then precipitated in cold acetone before in-solution protein digestion.

3.3.4 High-pH Reversed-Phase Liquid Chromatography (PepSwift™) Peptide Fractionation

The digested peptide samples were dissolved in 4 µL 0.1% ammonium hydroxide/99.9% HPLC grade water and transferred to micro-vials. 2 µL of each sample was then fractionated by HPLC (Ultimate 3000, Dionex) using a PepSift™ Monolithic Capillary HPLC column (Thermo Scientific) and a high-pH gradient as presented in Figure 3.2, where buffer A is 0.1% ammonium hydroxide/99.9% HPLC grade water and buffer B is 90% acetonitrile/10% buffer A. The fractions were sampled every 2 minutes from time point 11 minutes until 41 minutes into the gradient. They were then dried using SpeedVac at 30°C for 10 minutes.
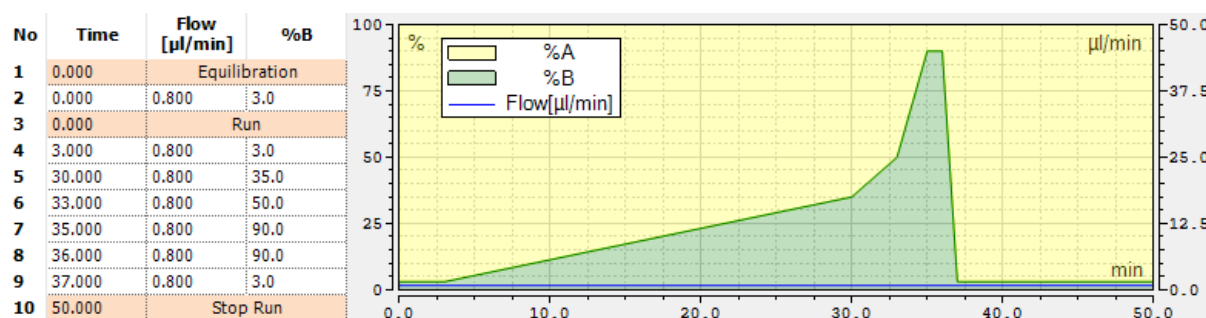
Figure 3.2: **Flow gradient for high-pH reverse-phase HPLC peptide fractionation.** Time, flow (µl/min), and % of buffer B is shown in the table to the right. Buffer B (green) is increased gradually to 35% from minute 3 - 30, then to 50% at minute 33, and finally to 90% at minute 35, before being lowered back to 3% at minute 36.

3.3.5 High-pH Reversed-Phase Spin Column (Pierce™ Kit) Peptide Fractionation

The provided reversed-phase fractionation spin columns were placed in a sample tube and centrifuged at 5,000 rcf for 2 minutes to discard the packaging solution. The columns were then conditioned, first twice with 300 µL acetonitrile and then twice with 300 µL 0.1% trifluoroacetic acid (TFA), centrifuging at 5,000 rcf for 2 minutes and discarding the flowthrough between each wash. Elution solutions were prepared according to Table 3.4, which is sufficient for the fractionation of three samples. Digested peptide samples were dissolved in 300 µL 0.1% TFA solution, loaded onto the column, and centrifuged at 3,000 rcf for 2 minutes to elute the «flowthrough» fraction. The beads were then washed with 300 µL water and centrifuged at 3,000 rcf for 2 minutes. The column was placed into a new sample tube, 300 µL elution solution

was loaded onto the beads, and then they were centrifuged at 3,000 rcf for 2 minutes to collect the fraction. This was repeated for every elution solution, proceeding from fraction no. 1 - 8. The final fractions were then dried using a SpeedVac at 45°C until they were completely dry.

Table 3.4: **Elution solutions for HpH-RP spin column (Pierce™) peptide fractionation kit.** Fraction number with the corresponding amount of acetonitrile (ACN) (in % and µL) and triethylamine (TEA) (µL) used.

Fraction No.	ACN (%)	ACN (µL)	0.1% TEA (µL)
1	5.0	50	950
2	7.5	75	925
3	10.0	100	900
4	12.5	125	875
5	15.0	150	850
6	17.5	175	825
7	20.0	200	800
8	50.0	500	500

* 10M AcOH/4M NaOAC (1:1)

3.4 Reduced complexity: Cysteine-containing Peptide Enrichment

Four methods of cysteine-containing peptide enrichment were investigated: BcMag™ Thiol-Activated Magnetic Beads, BcMag™ Long-Arm Thiol-Activated Magnetic Beads (Bioclone Inc., San Diego, CA, USA), High Capacity Acyl-rac S3™ Capture Beads (Nanocs, Newy York, NJ, USA), and Thiopropyl Sepharose™ 6B Affinity Resin (Sigma-Aldrich, Oslo, Norway), listed in Table 3.5. The preliminary investigation was performed using 1 µL 100 pmol bovine serum albumin (BSA), which was first digested with simple protein digestion (3.4.1), then purified by micro-SPE before cysteine-containing peptide enrichment, purified again by micro-SPE and analyzed by LC-MS. The most promising method was then carried out on three replicates of 10 µL serum, using in-solution protein digestion (3.1.2).

Table 3.5: **Cysteine-containing peptide enrichment overview.** A list of all cysteine-containing peptide enrichment kits with their bead type, reactive group, and company.

Enrichment kit	Bead type	Reactive group	Company
BcMag™ Thiol-Activated (R_3) beads	Magnetic Beads	pyridyl disulfide	Bioclone Inc
BcMag™ Long-Arm Thiol-Activated ($R_1 7$) beads	Magnetic Beads	pyridyl disulfide	Bioclone Inc
High Capacity Acyl-rac S3™ Capture Beads	Agarose	pyridyl disulfide	Nanocs
Thiopropyl Sepharose™ 6B Affinity Resin	Sepharose	pyridyl disulfide	Sigma-Aldrich

3.4.1 Simple Protein Digestion

The protein samples were dissolved in 10 µL 10 mM tris(2-carboxyethyl)Phosphine Hydrochloride (TCEP) and incubated for 30 minutes at room temperature. Then, 20 µL 50 mM ammonium bicarbonate was added, followed by 10 µL of 0.5 µg Trypsin Gold in 20 mM am-

monium bicarbonate. The samples were then incubated for 2 hours at 37°C, before adding 2.5 µL 50% formic acid.

3.4.2 BcMag™ Thiol-Activated Magnetic Beads

30 mg BcMag™ thiol-activated or long-arm thiol-activated magnetic beads were re-suspended in 1 mL coupling buffer [0.1 M sodium phosphate, 5 mM ethylenediaminetetraacetate (EDTA), pH 7.0] and vortexed for 2 minutes. The beads were then magnetically separated, and the supernatant was removed. This was repeated twice. The dried protein sample was dissolved in 1 mL coupling buffer, and incubated with the beads at room temperature, 800 rpm, for 1 hour. Next, the beads were magnetically separated, and the supernatant discarded. The beads were then washed with 1 mL coupling buffer four times. Then, 8 mg L-cysteine in 1 mL coupling buffer was added to the beads, and they were incubated at room temperature, 800 rpm, for 1 hour. After removing the supernatant by magnetic separation, the beads were washed with 1 mL 1M NaCl twice and then 1 mL water twice. The beads were then incubated with 0.5 mL phosphate-buffered saline (PBS) and 10 µL 100 mM DTT at room temperature, 800 rpm, for 30 minutes. The beads were magnetically separated, and the supernatant collected. Finally, 0.5 mL 0.1% TFA/80% acetonitrile was added to the beads, vortexed briefly before magnetic separation, and the supernatant was collected and combined with the previous one.

3.4.3 High Capacity Acyl-rac S3™ Capture Beads

Cysteine enrichment with High Capacity Acyl-rac S3™ Capture Beads was performed using the BcMag™ Thiol-Activated Magnetic Beads method, only with half the amount of each solution, and the Thiopropyl Sepharose™ 6B Affinity Resin method, as described next. 20 µL S3™ capture beads were used for both methods.

3.4.4 Thiopropyl Sepharose™ 6B Affinity Resin

In order to reduce disulfide formations, dried sample was first dissolved in 20 µL coupling buffer [50 mM tris-HCl, 21 mM EDTA, pH 7.5], 2 µL 100 mM DTT was added, and then incubated at 37°C for 1 hour. Meanwhile, 30 mg Thiopropyl Sepharose™ 6B resin was rehydrated in 1 mL water for 15 minutes, vortexed briefly and left for another 10 minutes. Then, 0.5 mL from the top of the supernatant was discarded, the resin was re-suspended, transferred to a 0.22 µm cellulose acetate spin filter, and centrifuged at 1,000 rcf for 30 seconds. The resin was washed with 0.5 mL water twice, then coupling buffer twice, centrifuging at 1,000 rcf for 30 seconds and discarding the flow through between each. Once the sample finished incubating, 80 µL coupling buffer was added before transferring to the spin column and incubated at room temperature, 900 rpm, for 1 hour. To collect the non-cysteine-containing peptides, the spin column was placed into a new collection tube and centrifuged at 1,500 rcf for 1 minute. The resin was then washed with 0.5 mL of: washing buffer [50 mM tris-HCl, 1 mM EDTA, pH 8] twice, 1M NaCl twice, 0.1% TFA/80% acetonitrile twice, washing buffer twice, and coupling buffer twice, centrifuging at 1,500 rcf for 30 seconds and discarding the flow through between each. Next, 100 µL 20 mM DTT was added to the resin, which was then incubated at room temperature, 900 rpm, for 30 minutes. Cysteine-containing peptides were then collected in a new collection tube by centrifuging at 1,500 rcf for 1 minute. This was repeated twice, first with 100 µL 20 mM DTT then 100 µL 80% acetonitrile, now with 10 minutes incubation time. 32 µL 1M IAA was added to the pooled elutes and incubated for 1 hour at room temperature in the dark.

4 Results

4.1 Depletion of Abundant Proteins

Depletion was performed to remove abundant proteins from the sample, so that the less abundant proteins would become more readily detected. This was done using depletion kits: the Albumin/IgG removal kit (Pierce™), AlbuSorb™ PLUS, AlbuVoid™ LC-MS On-Bead, AlbuVoid™ PLUS, and BioMag® ProMax HSA/IgG. The AlbuSorb/Void PLUS kits use a combination of polyelectrolytes and Protein A to deplete HSA and IgG, respectively, while AlbuVoid LC-MS only uses polyelectrolytes to deplete albumin. Protein A was also used by the Albumin/IgG removal kit, which instead employs cibacron Blue 3GA to deplete albumin. The BioMag kits separate albumin or IgG from serum by magnetic separation.
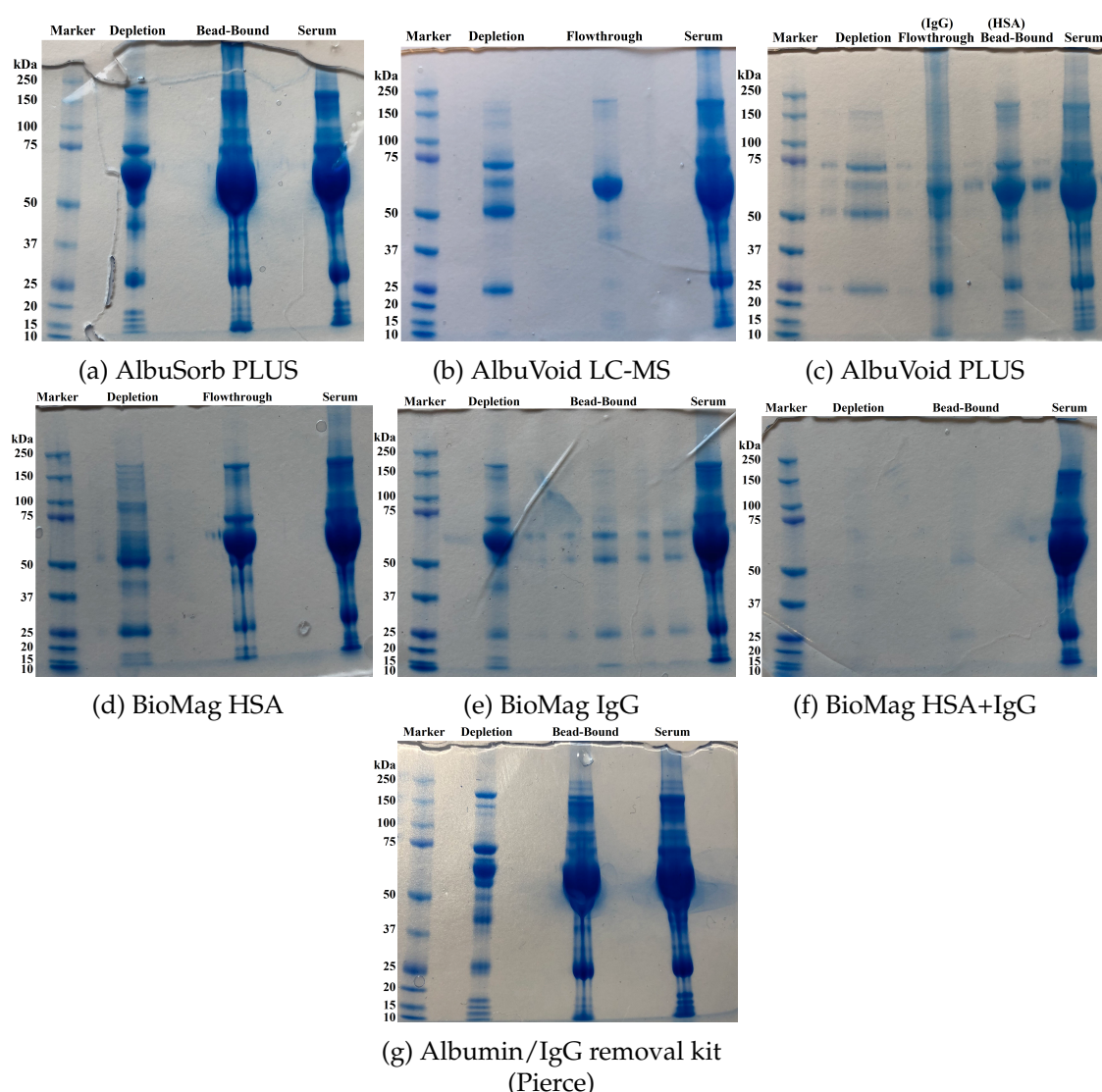
Figure 4.1: **SDS-PAGE gels of depletion kits:** (a) AlbuSorb PLUS, (b) AlbuVoid LC-MS, (c) AlbuVoid PLUS, (d) BioMag ProMax HSA, (e) IgG, and (f) HSA+IgG removal kit, and (g) the Albumin/IgG removal kit (Pierce). Labeled from left to right are Marker (Precision Plus Protein Standard Dual Color Marker), Depletion (depleted sample), Bead-Bound/Flowthrough (HSA and/or IgG containing sample), and Serum (undepleted control).

First, a test run of each kit was done, from which SDS-PAGE was performed. The resulting gels are shown in Figure 4.1. Undepleted serum is shown in the far-right lane of each gel, where the large, oversaturated band between 50 - 75 kDa arose from HSA (66 kDa). At around 150 kDa and 25 kDa visible bands containing IgG and its light-chain, respectively, were seen. However, the expected band from IgG heavy-chain at 50 kDa was overshadowed by the HSA band. The lanes in the middle of each gel, marked «Bead-Bound» or «Flowthrough», show the HSA and/or IgG-containing sample obtained either from the beads after eluting the depleted sample or the flowthrough, respectively. Here, a band containing HSA was clearly seen for the Albumin/IgG removal kit (4.1g), AlbuSorb (4.1a), AlbuVoid LC-MS (4.1b), AlbuVoid PLUS (4.1c), and BioMag HSA (4.1d) depletion kits. A 150 kDa band from IgG was also seen in these lanes, as well as faintly in BioMag IgG (4.1e). This band was stronger in the HSA- than the IgG-flowthrough of AlbuVoid PLUS. Meanwhile, the 50 kDa IgG band was visible in the IgG-flowthrough of AlbuVoid PLUS and BioMag IgG, and faintly in BioMag HSA+IgG (4.1f). The light-chain band at 25 kDa was also clearly visible in this lane of all gels, but faintly so in AlbuVoid LC-MS, and slightly stronger in the AlbuVoid IgG- than HSA-flowthrough. For the depleted samples, in the left-most lanes (excluding the protein marker), bands of albumin were clearly visible in the Albumin/IgG removal kit, AlbuSorb, and BioMag IgG, and more faintly in AlbuVoid LC-MS, AlbuVoid PLUS, and BioMag HSA. IgG bands were clearly observed in all gels, where the 50 kDa band was particularly visible in AlbuVoid LC-MS, AlbuVoid PLUS and BioMag HSA. The depleted samples of all kits also retained other proteins, indicated by bands of other molecular weight in each lane. However, the bands of both the depletion and flowthrough of BioMag HSA+IgG were barely visible.
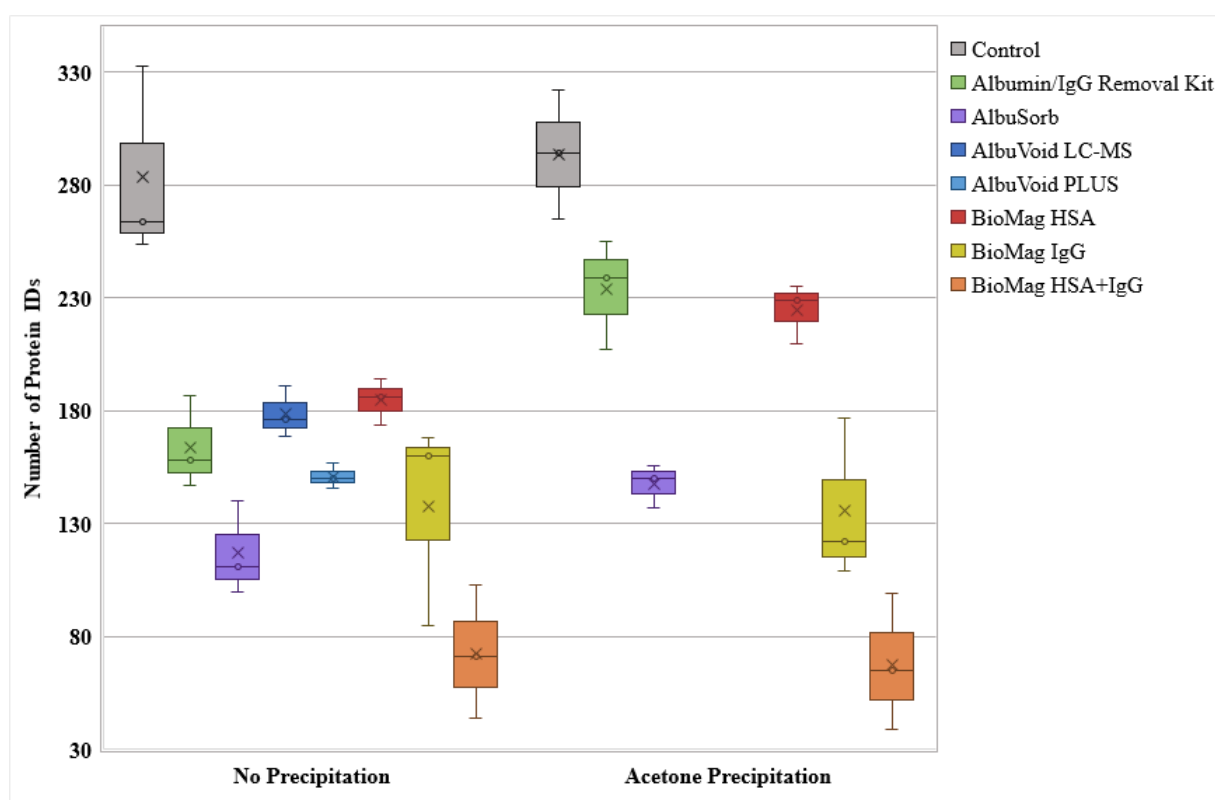
Figure 4.2: **Depletion: Boxplot diagram with the number of identified proteins between no precipitation and acetone precipitation.** The plot includes the control (gray) and each depletion kit: the Albumin/IgG removal kit (green), AlbuSorb (pruple), AlbuVoid LC-MS (dark blue), AlbuVoid PLUS (light blue), BioMag HSA (red), IgG (yellow), and HSA+IgG (orange).

After completing the initial test runs of each kit, three replicates of each were performed and analyzed by LC-MS in sequence. Table 4.1 shows the MS2 spectral data from the LC-MS analysis of the depletion kit replicates along with a control (ctrl) of undepleted serum. These values represent the average between the three replicates. The table contains the total spectral count (SC) and number of identified proteins of each sample. Also presented are the differences between the samples when protein precipitation (PPT) with cold acetone was performed before digestion or not. None of the depletion kits had more proteins identified than the control, and the total SC were also lower. This was the case regardless of whether the samples were precipitated or not. For the undepleted serum control, there was very little difference between the «No Precipitation» and «Acetone Precipitation» samples. When it came to the depletion kits, the Albumin/IgG removal kit, AlbuSorb, and BioMag HSA had higher total SCs and protein identifications (IDs), while the BioMag IgG and HSA+IgG had less. This excluded both AlbuVoid methods, where precipitation was not possible as the samples were digested on-bead. Figure 4.2, which shows a boxplot diagram of all the protein IDs, also illustrates this. Although the BioMag IgG and HSA+IgG had less protein IDs in the acetone precipitated samples, the boxplot diagram makes it clear that these values largely overlapp.

Table 4.1: **Depletion: MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) for all depletion kits both with and without acetone precipitation. Highlighted highest protein IDs among depletion kits (red, underline), and highest % SC of each column (orange).

| | Ctrl | Albumin/IgG removal kit | Albu Sorb | AlbuVoid | | BioMag | | |
				LC-MS	PLUS	HSA	IgG	Both
No Precipitation								
Total SC	15,519	7,351	3,606	4,963	3,589	6,746	4,686	1,754
Protein IDs	284	164	117	179	151	185	138	73
% SC α-G	16%	20%	19%	12%	11%	17%	25%	27%
% SC Apo	7%	11%	7%	12%	8%	7%	9%	5%
% SC β-G	6%	9%	9%	8%	10%	6%	6%	8%
% SC CF	16%	12%	15%	17%	23%	19%	12%	13%
% SC HSA	17%	14%	14%	9%	7%	8%	17%	19%
% SC IgG	5%	5%	5%	7%	7%	9%	3%	4%
% SC Other Ig	10%	8%	9%	11%	11%	11%	7%	7%
% SC Other	23%	22%	23%	22%	25%	24%	22%	17%
Acetone Precipitation								
Total SC	16,800	1,2002	5,444			9,616	5,058	1,822
Protein IDs	294	234	148			225	136	68
% SC α-G	15%	22%	20%			14%	21%	24%
% SC Apo	8%	11%	7%			3%	7%	5%
% SC β-G	6%	9%	10%			5%	8%	9%
% SC CF	15%	10%	10%			16%	10%	12%
% SC HSA	18%	14%	14%			8%	25%	25%
% SC IgG	5%	3%	5%			11%	3%	4%
% SC Other Ig	13%	10%	13%			24%	8%	8%
% SC Other	21%	22%	21%			18%	18%	12%

The percentage of MS2 spectral counts (% SC) of some of the most abundant proteins mentioned in section 1.2.2 of the introduction and the remaining other proteins are also shown in Table 4.1, and are presented graphically in Figure 4.3. These values show how many of the identified peptide spectra stem from each protein/protein group. Undepleted serum had about 16/15% α-globulin (α-G), 7/8% apolipoprotein (Apo), 6% β-globulin (β-G), 16/15% complement factor (CF) proteins, 17/18% HSA, 5% IgG, 10/13% other Igs, and 23/21% other, less abundant proteins in the non/acetone precipitated samples, respectively. Compared to the control, all kits had a lower HSA percentage, with the exception of BioMag IgG and HSA+IgG which had near equal (no PPT) or more (acetone PPT) HSA spectra. When it came to the immunoglobulins, only the Albumin/IgG removal kit (acetone PPT), BioMag IgG and HSA+IgG had fewer IgG. These, as well as AlbuSorb, also had less other immunoglobulins. For the remaining abundant proteins, AlbuVoid LC-MS and PLUS, and BioMag HSA had less α-globulins, all acetone precipitated BioMag kits and non-precipitated BioMag HSA+IgG had the same or less apolipoproteins, BioMag HSA had less β-globulins (acetone PPT), while the Albumin/IgG removal kit, AlbuSorb, BioMag IgG and HSA+IgG had less complement factors. Otherwise, the content of these groups were the same or higher in the depletion kit. Lastly, for the remaining proteins not included in the previously mentioned groups, the depletion kits which contained more of these less abundant proteins were the Albumin/IgG removal kit, AlbuVoid PLUS, and BioMag HSA.
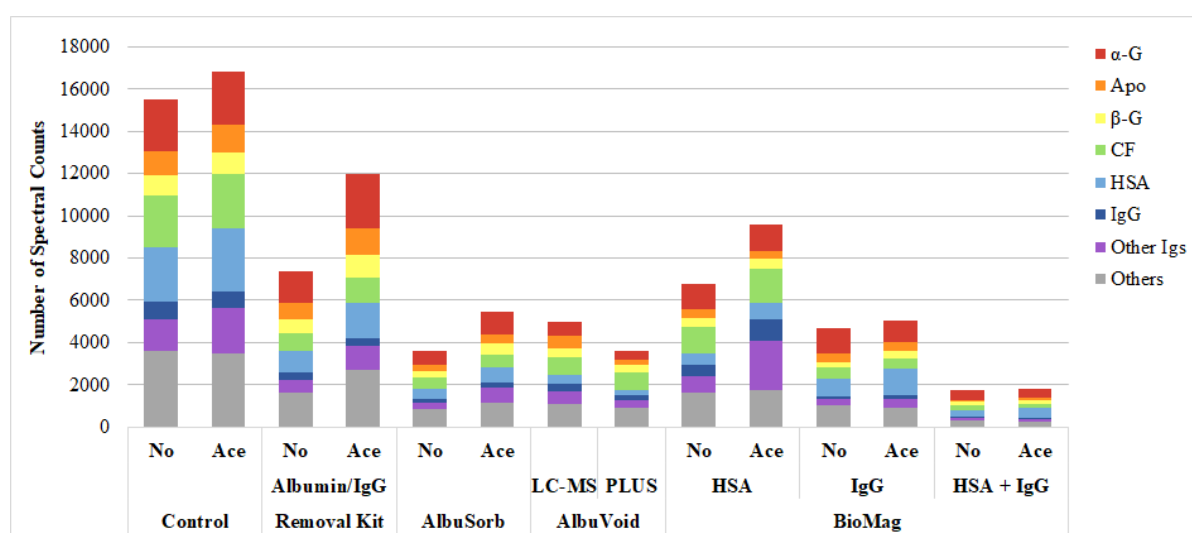
Figure 4.3: **Depletion: Protein MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (light blue), IgG (dark blue), other Igs (purple) and Others (gray) in the tested depletion kits.

What may be of interest are the unique proteins that were only found when utilizing a certain depletion kit. For each depletion kit, Figure 4.4 shows how many of the identified proteins, with > 2 MS2 spectral counts, were also found in the control, and how many were only found in the control or when using the kit. Here, more unique proteins were identified in the control than in any of the kits. The kits with the most unique proteins compared to the control were the Albumin/IgG removal kit and BioMag HSA. When comparing the kits also to each other, the number (shown in brackets) of uniquely identified proteins decreased for all samples.

The comparison of the number of unique proteins identified, with > 2 MS2 spectral counts, between the respective samples when precipitated in cold acetone or not is shown in Figure 4.5. For the control, as well as in depletion kit BioMag HSA+IgG, more unique proteins were identified in the non-precipitated samples. For all of the other kits, excluding the kits with

«On-Bead» digestion, there were more unique proteins in the acetone precipitated samples.
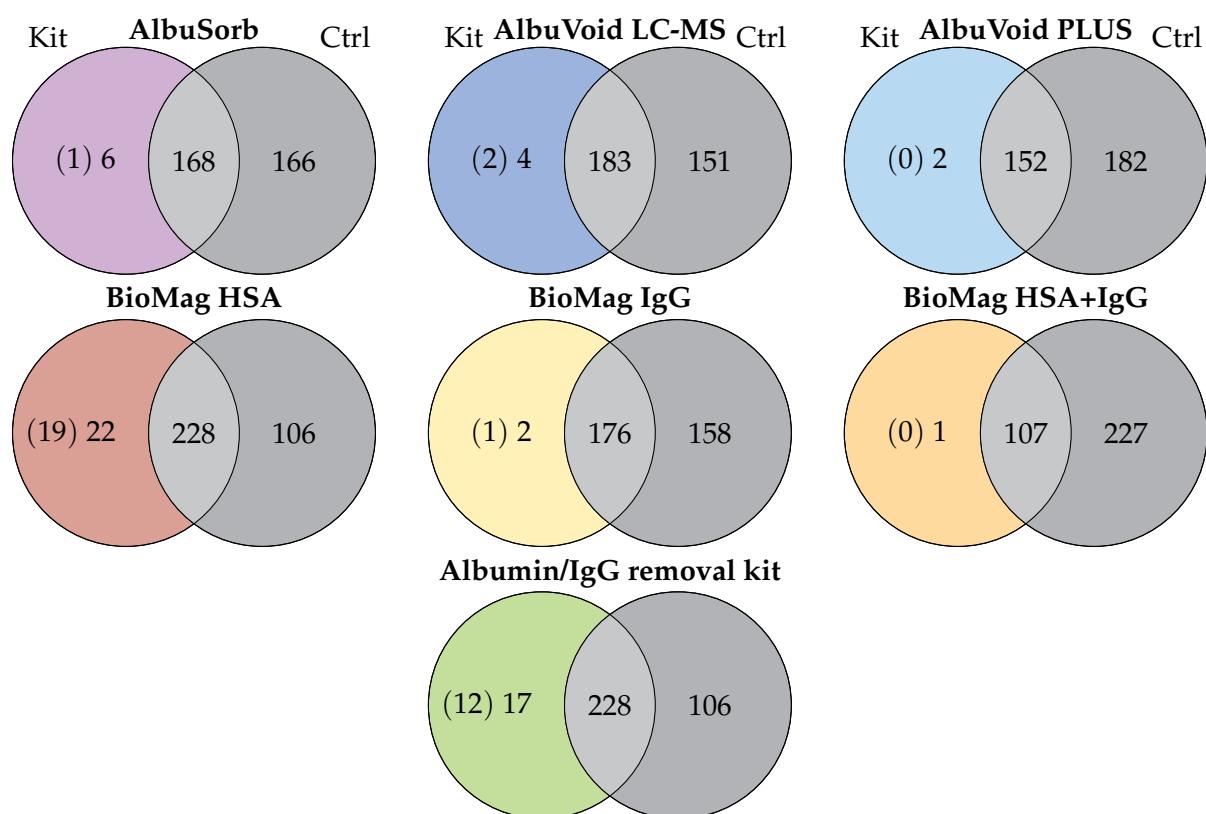
Figure 4.4: **Depletion: Kit versus Control.** The number of proteins identified in each depletion kit compared to the undepleted serum control. Unique proteins (>2 MS2 spectral counts) identified only in the kit (left), control (right) or in both (middle). The number of unique proteins found only with the given kit in brackets.
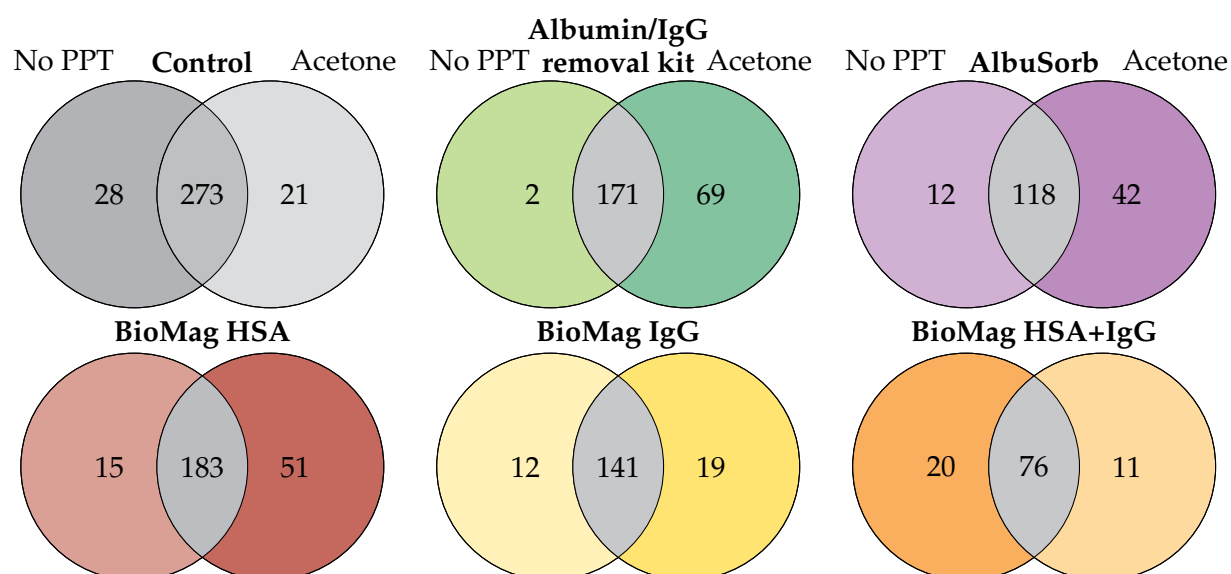
Figure 4.5: **Depletion: No protein precipitation versus Acetone precipitation.** Unique proteins (>2 MS2 spectral counts) identified only without precipitation (left), with acetone precipitation (right) or in both (middle).

4.2 Fractionation Methods

After the investigation of the depletion kits was completed, protein fractionation methods ethanol precipitation, size exclusion chromatography, and ZnCl$_2$ precipitation, and peptide fractionation method HpH-RP LC (PepSwift) and HpH-RP spin column (Pierce kit) were executed. These methods were performed separately, and were therefore firstly compared to their respective controls, analyzed simultaneously as the methods fractions, before being compared to each other. The MS2 spectral data from each individual fraction were also merged into one file, giving the total ($\sum$) values of each method. As with the depletion kits, three replicates of each fractionation method were performed and, unless otherwise stated, the following results are the average of the respective replicates.

4.2.1 Ethanol Protein Precipitation

An attempt to adapt the Cohn process of plasma protein fractionation for micro-volumes of serum was made. Ethanol precipitation uses increasing concentration of the organic solvent ethanol, as well as variable pH, dehydrating the proteins thus causing them to precipitate from the solution into fractions. The concentrations and pH are shown in Figure 3.1 using the solutions in Table 3.3. The results of applying this approach are shown in Table 4.2, which includes the total MS2 spectral counts, protein IDs, and the SC percentage of abundant proteins in each fraction and the total fractions. The number of MS2 spectral counts in each fraction of the protein groups are also represented graphically in Figure 4.6. The fractions were named in concurrence with those given by Cohn et. al., where fractions I, I+III, IV-1, IV-4, and V are the pelleted fractions mentioned in the original method, while V-1 and V-2 are additional supernatant fractions.

Table 4.2: **Ethanol protein precipitation: MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) for fractions of the Cohn process.

	Ctrl	$\sum$	Fraction I	II+III	IV-1	IV-4	V-1	V	V-2
Total SC	13,393	4,479	3,601	1,618	1,782	2,977	1,812	761	276
Total IDs	173	176	137	97	97	138	109	64	29
% SC α-G	13%	10%	16%	14%	14%	12%	12%	15%	13%
% SC Apo	11%	9%	15%	16%	15%	11%	4%	7%	4%
% SC β-G	8%	7%	7%	7%	7%	10%	7%	15%	11%
% SC CF	17%	18%	15%	18%	18%	20%	7%	6%	2%
% SC HSA	15%	16%	14%	14%	14%	7%	29%	21%	41%
% SC Ig	18%	18%	16%	15%	15%	17%	23%	21%	19%
% SC Other	18%	21%	17%	16%	18%	23%	19%	17%	9%

Comparing first the fraction total ($\sum$) to the control, although there were less MS2 spectral counts in the total fraction, there was about the same number of proteins identified. The percentage of protein groups were also very similar, the biggest difference being 3% more α-globulins in the control and 3% more other proteins in the total fraction. Fractions I, II+III, and IV-1 had between 14 - 18% of all protein groups, except β-globulins which were 7%. Fraction IV-4 contained ≥20% complement factors and other proteins, 10 - 20% apolipoproteins, α-globulins and immunoglobulins, and < 10% albumin and β-globulins. The remaining three

fractions contained mostly HSA and IgG, intermediate percentages of α- and β- globulins, and other proteins, and little apolipoproteins and complement factors.

Figure 4.6: **Ethanol protein precipitation: Protein MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (blue), Igs (purple) and Others (gray) in each ethanol precipitation fraction. Above each fraction is the number of protein IDs.

Table 4.3: **Ethanol protein precipitation: Protein distribution.** The percentage of total spectral count (SC), protein IDs, and protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) distributed in each fraction. Highlighted highest percentage of each row (orange).

	Fraction						
	I	II+III	IV-1	IV-4	V-1	V	V-2
Total SC	28%	13%	14%	23%	14%	6%	2%
Protein IDs	79%	56%	56%	79%	63%	37%	17%
% SC α-G	32%	13%	14%	21%	12%	6%	2%
% SC Apo	36%	17%	17%	21%	5%	3%	1%
% SC β-G	24%	11%	12%	27%	12%	11%	3%
% SC CF	29%	15%	16%	31%	6%	2%	0%
% SC HSA	25%	12%	13%	10%	27%	8%	6%
% SC Ig	26%	11%	12%	22%	19%	7%	2%
% SC Other	25%	11%	13%	30%	14%	5%	1%

While Table 4.2 gives the percentage of each fraction that was made up of an abundant protein group, Table 4.3 gives the percentage distribution of a protein group between the fractions. Most of the total SC and identified proteins were found in fraction I and IV-4, while very little was found in V-2. A high percentage, more than 20%, of all protein groups were found in fraction I. In addition, α-, β-, and immunoglobulins were largely in fraction IV-4, where the most apolipoproteins, complement factors and other proteins were too. Most of the albumin was located in fraction V-1. Between 10 - 20% of each protein was found in fractions II+III, IV-1, and V-1, with the exception of apolipoproteins and complement factors that were < 10% in the latter. More than 10% of β-globulins were also in fraction V. Otherwise, there was < 10% of a

protein group in the other fractions. This can also be inferred from Figure 4.6.

Table 4.4: **Ethanol protein precipitation: Unique proteins.** The number of proteins identified (with >2 MS2 spectral counts) exclusively in each ethanol precipitation fraction, along with the human UniProt entry name.

	No. of	Accession
I	10	APOC4, CHLE, CO4A, HV349, HV373, HV551, IC1, KAIN, KV224, LCAT, LV310, PLF4, RPGF5
IV-4	6	ACTB, CD14, MASP1, PLSL, PROC, QSOX1
V-1	2	ECM1, LV746

In accordance with the previous results, there were unique proteins identified with > 2 MS2 spectral counts exclusively in fraction I, then IV-4, and lastly V-1, presented in Table 4.4. None were not also found in the control.

4.2.2 Size Exclusion Chromatography

Size exclusion chromatography separates proteins based on their molecular weight, by moving them through a stationary phase of porous matrix that deters smaller molecules from passing through. This results in larger molecules eluting first. SEC was carried out using the ÄKTA system with a 10/300 GL column. The resulting chromatograms from the three replicates, run sequentially, are shown in Figure 4.7 and 4.8. The images show the UV (280 nm) spectra of each run, along with the collected fractions numbered 1 - 20 in order of elution. This means that the largest proteins were expected to be in the first fractions, while the smaller proteins were expected to be in the latter. Peak intensity corresponds to the amount of protein that eluted at that point in time. From the chromatography, most of the proteins eluted between fraction 1 - 8, with the highest intensity peak coinciding with fraction 7. There was also a pronounced peak around fraction 16 - 17. All three replicates were near identical in terms of peak formation and location, implying a high degree of reproducibility.
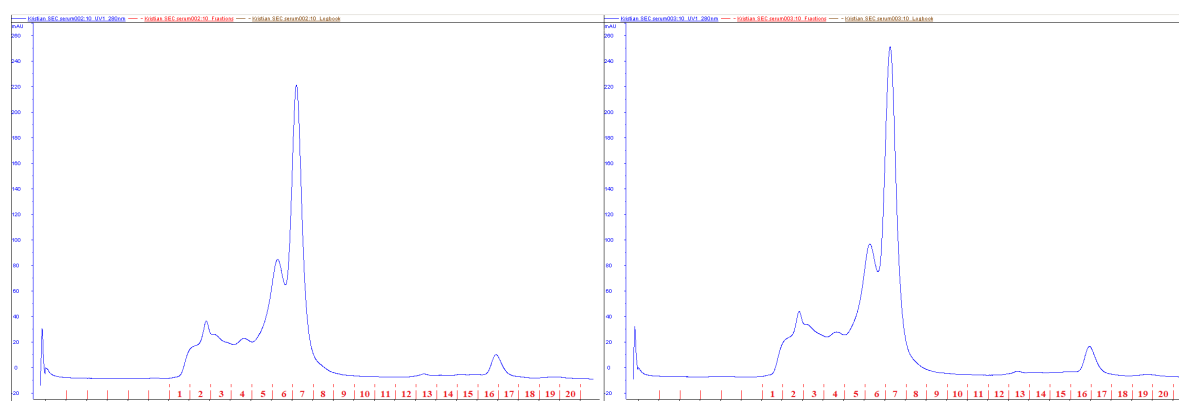
Figure 4.7: **Size exclusion chromatography: UV (280 nm) chromatograms of serum fractionation.** Chromatogram of replica B (left) and C (right) with the corresponding fractions numbered in red.
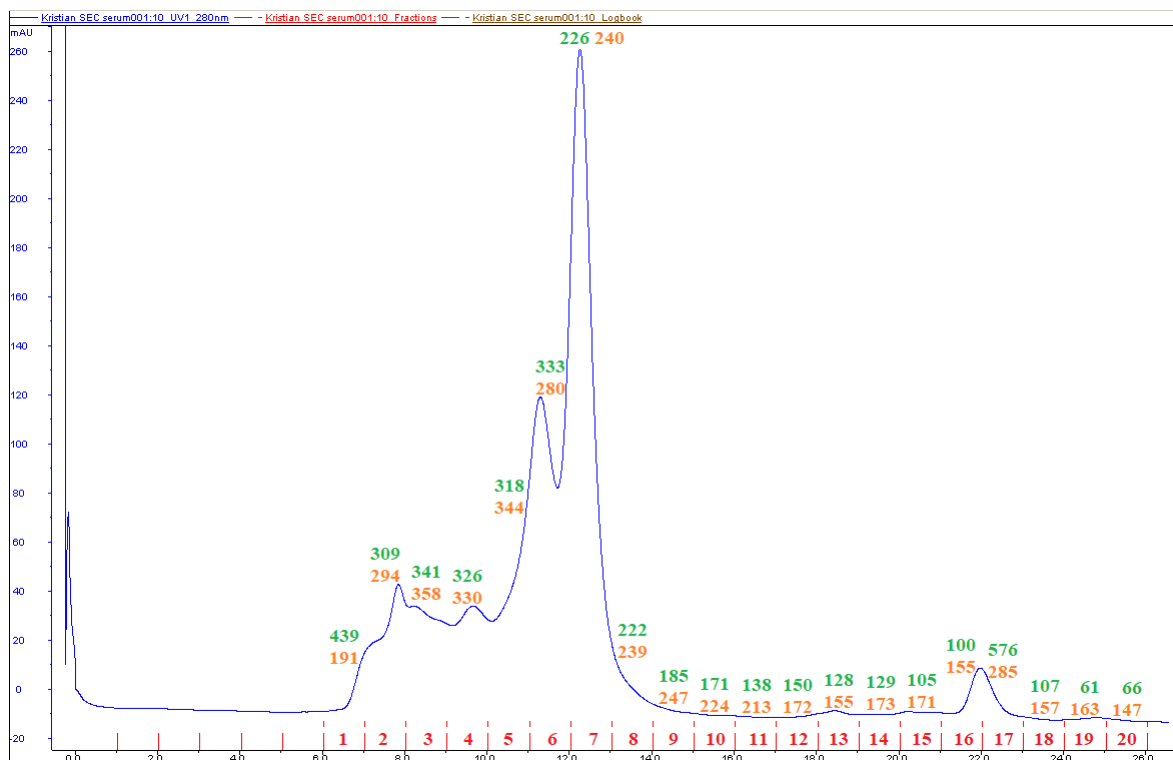
Figure 4.8: **Size exclusion chromatography: UV (280 nm) chromatograms of serum fractionation.** Chromatogram of replica A with the corresponding fractions numbered in red. The chromatogram includes the number of protein IDs from replica B (green) and C (orange) found in the corresponding fraction.

Due to the high amount of fractions, each replicate was analyzed by LC-MS separately along with individual controls. The samples from replicate A were lost during analysis because of low LC-MS performance, so these results are not included. Furthermore, the results from the MS2 spectral data analysis of replicate B and C are presented separately, owing to the difference in identified proteins in some of the fractions. These MS2 spectral data from the 20 fractions are shown in Table 4.6 and 4.7, respectively, and together in Figure 4.9. Table 4.5 shows the total ($\sum$) and respective controls from both replica, as well as the combined fractions «9-14» and «15-20» of replica B and the resulting 10-fraction total $\sum$(10). Fractions «9-14» and «15-20» were made by pooling the denominated fractions and analyzing them by LC-MS after the individual fractions had been analyzed. For the 20-total fractions, almost 1,000 proteins were identified in replica B, while 574 proteins were identified in replica C. This was about 3 and 2 times that of their controls, respectively. Possibly the first fraction of replica B contained protein contamination from a previous sample. Thus, the protein IDs of replica C seem to be more reliable. Combining the last 12 fractions of replica B into two gave ∼100 less proteins than the 20-total. The total MS2 spectral counts were also proportionately high. Percentage SCs were similar in the three totals, where the majority of the spectra stem from other, less abundant proteins. In comparison to their control, replica B with 20 fractions gave more HSA, Igs and less-abundant proteins, B(10) had more α-globulins as well, while replica C gave only more Igs.

Table 4.5: **Size exclusion chromatography: MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of SEC totals (∑s), control, and combined fractions «9-14» and «15-20». Highlighted highest protein IDs among the fractions (dark green, underline), and highest % SC of each column (light green).

Replica:	B (20)		C (20)				B (10)
	ctrl	∑	ctrl	∑	9 - 14	15 - 20	∑(10)
Total SC	17,008	119,079	12,596	158,645	5,318	3,459	96,401
Total IDs	354	996	291	574	294	400	899
% SC α-G	19%	19%	19%	18%	16%	9%	20%
% SC Apo	8%	4%	8%	5%	3%	2%	4%
% SC β-G	7%	6%	7%	6%	10%	5%	5%
% SC CF	18%	14%	18%	16%	9%	3%	16%
% SC HSA	9%	13%	12%	16%	23%	15%	10%
% SC Ig	13%	18%	14%	17%	15%	9%	18%
% SC Other	26%	27%	23%	21%	24%	56%	27%
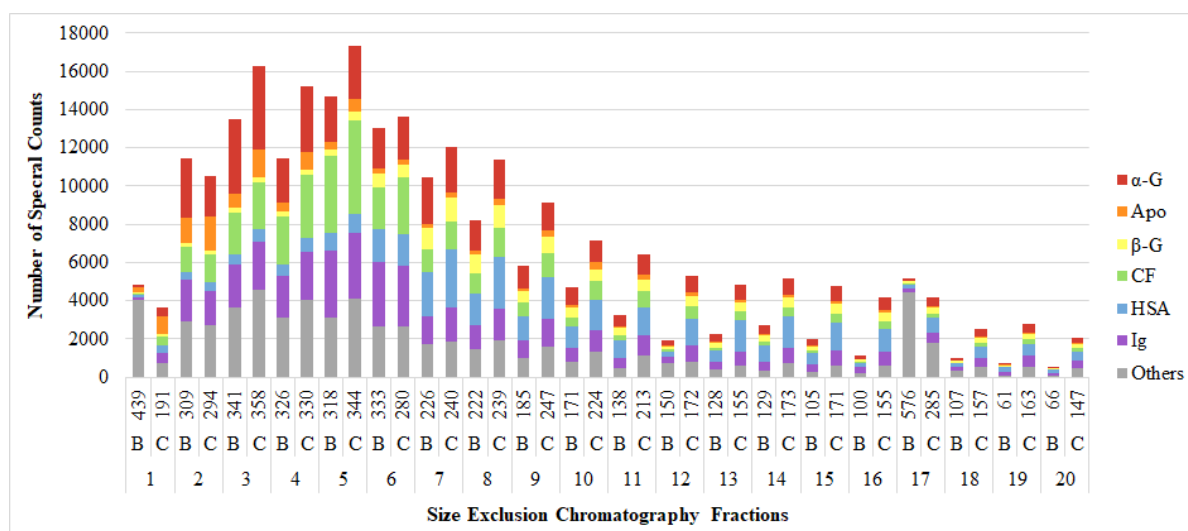
Figure 4.9: **Size exclusion chromatography: Protein MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (blue), Igs (purple) and Others (gray) in each SEC fraction, with the number of protein IDs below each bar.

Looking at the fractions for replica B in Table 4.6, the one with the most identified proteins was fractions 17 with near 576 IDs, then fraction 1 and «15-20» with ≥ 400. Fractions 2 - 6 had > 300, 7, 8 and «9-14» > 200, 9 - 16 and 18 ≥ 100, and 19 and 20 < 100 protein IDs. Fractions with the highest percentage of α-globulins were 2, 3, and 7, fraction 5 had more complement factors, albumin was the most abundant protein in fractions 9 - 11, 13 - 14, and 19, while 6, 16, and 20 had the most of immunoglobulins. For the remaining fractions, the largest percentage of spectra arose from other proteins. This was especially true for fractions 1 and 17, where 83% and 85% are less abundant proteins, respectively.

Table 4.6: **Size exclusion chromatography: Replica B MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of size exclusion chromatography replicate B. Highlighted highest protein IDs among the fractions (dark green, underline), and highest % SC of each column (light green).

Replica B Fractions									
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Total SC 4,850	11,441	13,481	11,427	14,716	13,056	10,461	8,224	5,814	4,735
Total IDs 439	309	341	326	318	333	226	222	185	171
% SC α-G 3%	27%	29%	20%	16%	17%	23%	20%	20%	20%
% SC Apo 5%	11%	5%	4%	3%	2%	2%	2%	3%	3%
% SC β-G 1%	2%	2%	2%	2%	5%	11%	12%	10%	11%
% SC CF 1%	12%	16%	22%	27%	17%	12%	13%	12%	10%
% SC HSA 3%	3%	4%	5%	6%	14%	22%	20%	21%	24%
% SC Ig 3%	19%	16%	19%	24%	26%	14%	15%	16%	16%
% SC Other 83%	26%	27%	27%	21%	20%	16%	18%	18%	16%
11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Total SC 3,259	1,908	2,288	2,743	2,016	1,165	5,188	1,011	766	566
Total IDs 138	150	128	129	105	100	576	107	61	66
% SC α-G 19%	14%	19%	18%	18%	18%	3%	14%	16%	15%
% SC Apo 2%	3%	2%	2%	2%	3%	1%	2%	3%	4%
% SC β-G 12%	7%	11%	12%	12%	8%	2%	9%	10%	9%
% SC CF 8%	5%	5%	6%	5%	8%	1%	4%	4%	4%
% SC HSA 27%	16%	28%	32%	31%	20%	4%	20%	30%	26%
% SC Ig 17%	18%	16%	16%	19%	23%	5%	18%	27%	29%
% SC Other 15%	37%	18%	13%	14%	21%	85%	33%	9%	13%

Moving on to the MS2 spectral data of replicate C in Table 4.7, where the last twelve fractions were not pooled. Unlike replicate B, the fractions with the most proteins identified in replicate C were 3 then 5, with 358 and 344 IDs, respectively. Fractions 3 - 5 had > 300, fractions 2, 6 - 11 and 17 - 11 had > 200, and the remaining fractions identified > 100 proteins. Of these identified proteins, the fractions where the highest percentage of MS2 spectral counts were from less abundant proteins were 2 - 4 and 17, the latter having had the highest percentage of 43% other proteins. For the remaining fractions, 1 had most apolipoproteins, 5 had most complement factors, 6 had most immunoglobulins, and the rest had most albumin.

Table 4.7: **Size exclusion chromatography: Replica C MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of size exclusion chromatography replicate C. Highlighted highest protein IDs among the fractions (dark green, underline), and highest % SC of each column (light green).

	Replica C Fractions									
	1	2	3	4	5	6	7	8	9	10
Total SC	3,662	10,503	16,310	15,248	17,359	13,604	12,030	11,414	9,144	7,156
Total IDs	191	294	358	330	344	280	240	239	247	224
% SC α-G	14%	20%	27%	23%	16%	16%	20%	18%	16%	16%
% SC Apo	24%	17%	9%	6%	4%	2%	2%	3%	4%	5%
% SC β-G	5%	2%	2%	2%	2%	5%	10%	11%	10%	9%
% SC CF	13%	14%	15%	22%	28%	22%	13%	13%	14%	14%
% SC HSA	10%	4%	4%	5%	6%	12%	25%	24%	24%	22%
% SC Ig	15%	17%	16%	17%	20%	23%	15%	14%	16%	16%
% SC Other	20%	26%	28%	26%	24%	20%	15%	17%	18%	18%
	11	12	13	14	15	16	17	18	19	20
Total SC	6,406	5,314	4,805	5,154	4,780	4,206	4,189	2,513	2,792	2,080
Total IDs	213	172	155	173	171	155	285	157	163	147
% SC α-G	16%	17%	16%	16%	17%	17%	12%	16%	16%	14%
% SC Apo	4%	4%	2%	2%	2%	2%	2%	3%	3%	4%
% SC β-G	9%	10%	10%	10%	11%	11%	7%	10%	9%	9%
% SC CF	14%	12%	9%	10%	10%	10%	6%	7%	9%	9%
% SC HSA	23%	26%	35%	32%	31%	28%	19%	26%	22%	23%
% SC Ig	17%	16%	15%	15%	16%	17%	12%	17%	19%	20%
% SC Other	18%	15%	13%	15%	13%	15%	43%	22%	20%	22%

The number of unique proteins identified only in one fraction of replicate B or C, the most abundant listed in Table 4.8, coincide somewhat with the fractions with the most overall protein IDs. Replica B had the most unique proteins identified in fractions 1 and 17 when looking at the 20-fraction method. These values were much higher than any of the other fractions. When the latter half of the fractions were combined, the number of unique IDs almost doubled in fraction 1, while the combined «15-20» becomes 1/8 of fraction 17 alone. Replica C also had the most uniquely identified proteins in fraction 17, where the number of unique protein IDs was 50, but only 4 in fraction 1. Combining the unique proteins identified in fraction 2 - 8 resulted in ~11 IDs for both replicas. Most fractions had < 5 unique proteins, and many had zero. Highlighted in green are the proteins that were also not found in the control.

Table 4.8: Size exclusion chromatography: Unique proteins. The number of proteins identified (with >2 MS2 spectral counts) exclusively in each fraction of SEC replica B 20 fractions (20), 10 fractions (10) and C, along with the human UniProt entry name of the four most abundant. Proteins that were not found in the control either are highlighted in green.

		Replica B 20 fractions		Replica B 10 fractions		Replica C 20 fractions
	No.	Accession	No.	Accession	No.	Accession
1	55	HSP72, IDE, INVO, PLEC	94	ACTN4, EPIPL, FAS, IDE	4	BASI, MUC24, RPN1, ZNT7
2-8	10	C163A, IGM, S1A7A, SPB4	11	C163A, IGM, S1A7A, SPB4	11	CA2D1, CSF1R, G6PE, IGLC7
9-14	0		4	CAH1, CAH2, IGLC7, TYB4	2	FHR3, IBP4
15-16	2	CAH6, MUC5B	20	ARF4, C1TC, EF1A2, MUC5B, PDIA4, SND1, TCPB, XPO2	0	
17	156	EF1A2, PDIA4, TCPB, XPO2			50	EF2, FAS, LMNA, MYH9
18-20	0				0	

Table 4.9: Size exclusion chromatography: Protein distribution. The distribution percentage of total spectral count (SC), protein IDs, and protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other). Highlighted highest % SC of each row (green), wrapped to include both halves.

Fractions	1	2	3	4	5	6	7	8	9	10
Total SC	2%	7%	10%	10%	11%	9%	8%	7%	6%	5%
Total IDs	33%	51%	62%	57%	60%	49%	42%	42%	43%	39%
% SC α-G	2%	7%	15%	12%	10%	7%	8%	7%	5%	4%
% SC Apo	10%	21%	17%	10%	8%	4%	3%	4%	4%	4%
% SC β-G	2%	2%	3%	3%	4%	6%	13%	12%	9%	6%
% SC CF	2%	6%	10%	13%	20%	12%	6%	6%	5%	4%
% SC HSA	1%	2%	2%	3%	4%	6%	12%	11%	9%	6%
% SC Ig	2%	7%	9%	9%	13%	12%	7%	6%	5%	4%
% SC Other	2%	8%	14%	12%	12%	8%	6%	6%	5%	4%
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Total SC	4%	3%	3%	3%	3%	3%	3%	2%	2%	1%
Total IDs	37%	30%	27%	30%	30%	27%	50%	27%	28%	26%
% SC α-G	4%	3%	3%	3%	3%	2%	2%	1%	2%	1%
% SC Apo	3%	2%	1%	1%	1%	1%	1%	1%	1%	1%
% SC β-G	6%	5%	5%	5%	5%	5%	3%	2%	3%	2%
% SC CF	4%	2%	2%	2%	2%	2%	1%	1%	1%	1%
% SC HSA	6%	5%	7%	6%	6%	5%	3%	3%	2%	2%
% SC Ig	4%	3%	3%	3%	3%	3%	2%	2%	2%	2%
% SC Other	3%	2%	2%	2%	2%	2%	5%	2%	2%	1%

Table 4.9 shows the percentage of each protein group distributed among the 20 fractions of replicate C. Most of the total spectra were found in fractions 3 - 5 (~13%), whilst $\geq 5\%$ were in fractions 2, 6 - 9, and 17, and $< 5\%$ in the rest. For the abundant protein groups, apolipoproteins were mostly found in fraction 2 and 3 (17 - 21%), α-globulins in fraction 3 and 4 (12 - 15%), complement factors in fraction 4 and 5 (13 - 20%), immunoglobulins in fraction 5 and 6 (12 - 13%), and albumin and β-globulins in fractions 7 and 8 (11 - 12% and 12 - 13%, respectively). The majority, $> 10\%$ of the less abundant proteins were located in fraction 3 - 5.

4.2.3 Zinc Chloride Protein Precipitation

Precipitation with zinc chloride works on the bases of metal-ions binding to and forming soluble aggregates with proteins. Three sequences of $ZnCl_2$ precipitation were performed: one using 20 mM $ZnCl_2$ resulting in 2 fractions, one using 0.2 mM then 20 mM $ZnCl_2$ resulting in 3 fractions, and one using 0.02 mM, 0.2 mM and 20 mM $ZnCl_2$ resulting in 4 fractions. The respective sequences were thus labeled 2, 3, and 4, and the fractions as P (Pellet) and S (Supernatant) in the following tables. Numbering of the P fractions, 1 up to 3, are in order of highest to lowest $ZnCl_2$ concentration used. So, for example, fraction P1 of all sequences was obtained using the highest (20 mM) concentration of $ZnCl_2$. The MS2 spectral data of all fractions, as well as the total ($\sum$) of each sequence, is shown in Table 4.10. These data are also illustrated in Figure 4.10. Starting with the sequence totals, all had approximately a hundred more identified proteins than the control. The sequence with the most protein IDs was 2 fractions, but the values of the other sequences were quite similar. Meanwhile, the 4-fraction sequence had the most MS2 spectral counts total. The percentage of protein groups were similar between sequences 3 and 4, and between sequence 2 to that of the control. All three sequences had more of the less abundant proteins than the control, and more α-globulins, more or the same of apolipoproteins, and the same of β-globulins. They all had less albumin and complement factors as well. For immunoglobulins, the 2-fraction sequence had less than the control while the others had more.
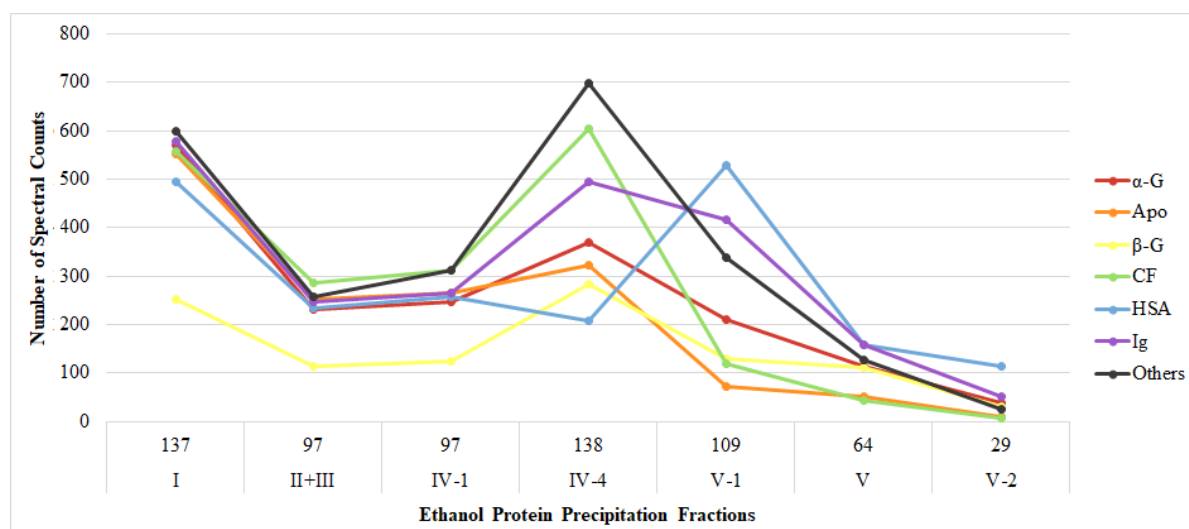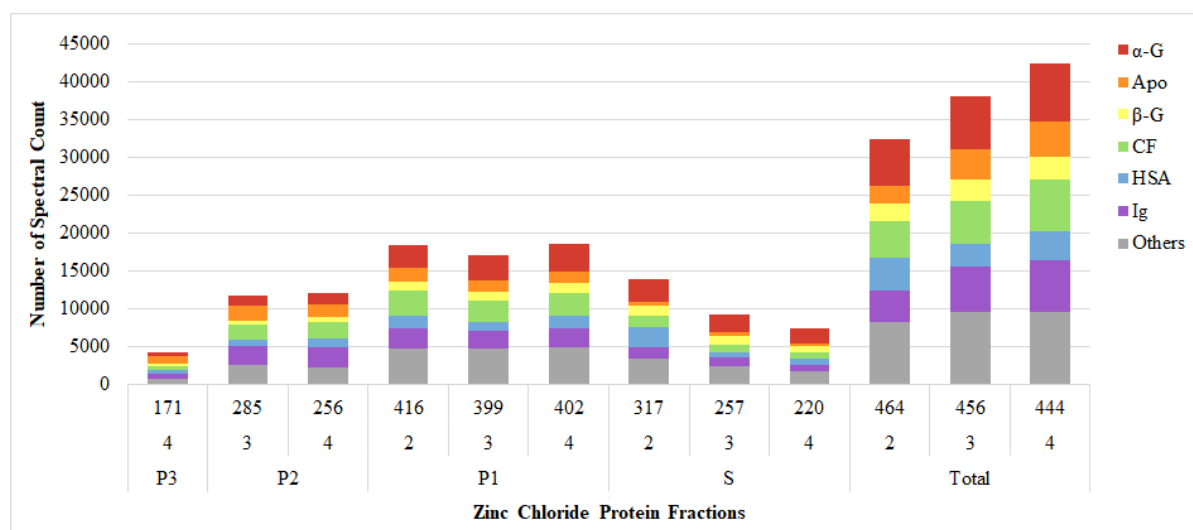

Figure 4.10: **ZnCl$_2$ protein precipitation: Protein MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (blue), Igs (purple) and Others (gray) in each $ZnCl_2$ precipitation fraction, with the number of protein IDs below each bar.

For the individual fractions, seen in the lower half of Table 4.10, ~400 proteins were identified in the last, or only, P fraction in all sequences (P1). Sequence 3 and 4 had the second most protein IDs in the preceding P fraction (P2), followed by the S fraction, with

between 200 - 300 IDs. Sequence 2, fraction S had > 300 IDs, while P3 of sequence 4 had < 200 IDs. All the P1 fractions and 2-S had the highest percentage (> 20%) of other, less abundant proteins. Meanwhile, 3-S and 3-P2 had the same amount of α-globulins or immunoglobulins, respectively, 4-S had more α-globulins, 4-P2 more immunoglobulins, and 4-P3 more apolipoproteins.

Table 4.10: **ZnCl$_2$ protein precipitation: MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of the ZnCl$_2$ protein precipitation methods. Highlighted highest protein IDs among the fractions (purple, underline), and highest % SC of each column (pink).

	Ctrl	Σ 2	Σ 3	Σ 4
Total SC	17,375	32,329	37,965	42,268
Total IDs	360	464	456	444
% SC α-G	17%	19%	18%	18%
% SC Apo	7%	7%	10%	11%
% SC β-G	7%	7%	7%	7%
% SC CF	17%	15%	15%	16%
% SC HSA	16%	14%	8%	9%
% SC Ig	15%	13%	16%	16%
% SC Other	21%	25%	25%	23%

	2		3			4			
	P1	S	P2	P1	S	P3	P2	P1	S
Total SC	18,410	13,919	11,766	17,000	9,199	4,257	12,125	18,505	7,381
Total IDs	416	317	285	399	257	171	256	402	220
% SC α-G	17%	22%	11%	19%	26%	11%	13%	20%	26%
% SC Apo	10%	4%	17%	9%	4%	26%	14%	8%	5%
% SC β-G	6%	9%	5%	7%	13%	6%	5%	6%	11%
% SC CF	18%	11%	16%	17%	11%	14%	19%	17%	12%
% SC HSA	9%	19%	8%	7%	8%	11%	9%	8%	10%
% SC Ig	15%	11%	21%	14%	12%	17%	21%	14%	12%
% SC Other	26%	25%	21%	28%	26%	16%	19%	27%	23%

Looking at the percentage of the protein group distribution between the fractions in Table 4.11, the total MS2 spectral distribution was the same as for the protein IDs as mentioned previously. That is, most spectra were counted in fraction P1, and so on. Moving on to the protein groups, most of the albumin content was in 2-S or P1 of sequence 3 and 4. α-globulins had a 50/50 split in sequence 2, but were retained mostly in P1 of the others, whilst β-globulins were more in fraction S of 2 and 3, and P1 of 4. Apolipoproteins and immunoglobulins were mainly in 2-S or 3/4-P2, and both complement factors and the less abundant proteins were mostly in fraction P1 of all sequences.

Table 4.11: **ZnCl$_2$ protein precipitation: Protein distribution.** The percentage of total spectral count (SC), protein IDs, and protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) distributed in each fraction. Highlighted highest percentage of each row per sequence (pink).

| | 2 | | 3 | | | 4 | | | |
	P1	S	P2	P1	S	P3	P2	P1	S
Total SC	57%	43%	31%	45%	24%	10%	29%	44%	17%
Protein IDs	90%	68%	63%	88%	56%	39%	58%	91%	50%
% SC α-G	50%	50%	19%	46%	34%	6%	20%	48%	25%
% SC Apo	78%	22%	51%	39%	10%	23%	37%	32%	8%
% SC β-G	48%	52%	20%	39%	41%	8%	22%	41%	29%
% SC CF	68%	32%	33%	50%	17%	9%	33%	45%	13%
% SC HSA	39%	61%	32%	42%	26%	12%	29%	40%	19%
% SC Ig	65%	35%	43%	39%	18%	10%	39%	37%	14%
% SC Other	58%	42%	26%	49%	25%	7%	23%	52%	18%

Table 4.12 contains the number of unique proteins, with > 2 MS2 spectral counts, in each fraction of one sequence (left) and between the different sequences (right). The fractions with the most unique proteins identified was P1, with more than 60 unique IDs for all sequences. After this came the S fractions, and last was P3 with only one unique protein. Overall, most unique IDs were found in sequence 2, then 3, and lastly 4.

Table 4.12: **ZnCl$_2$ protein precipitation: Unique proteins.** The number of proteins identified (with >2 MS2 spectral counts) exclusively in each fraction (left) and sequence (right), along with the human UniProt entry name of the four most abundant. Proteins that were not found in the control either are highlighted in pink.

		Between Fractions		Between Sequences	
		No.	Accession	No.	Accession
2	**P1**	88	BIP, F13A, FCGBP, PROP	13	CADH6, FA20C, FGL1, GSHR, LV208, SODC, TRPC5, TRPM2
	S	27	CAH1, LYVE1, PTPRJ, VNN1		
3	**P2**	8	BPIB1, CD9, DEF1, ITA2B	4	CD9, IPSP, ITA2B, MEGF9
	P1	67	NCHL1, PLGB, PLSL, TENX		
	S	27	DPEP, DPP4, ICAM1, LYVE1		
4	**P3**	0		2	MARCO, ROBO4
	P2	1	MARCO		
	P1	73	IGLC7, NCHL1, PLGB, PLSL		
	S	15	CAH1, DPP4, TARSH, TPIS		

When all of the proteins observed in each pellet fraction (P1, P2, and P3) and supernatant fraction (S) of all three sequences were combined, the ∑ viewed in Table 4.13 was achieved. Also displayed here are the number of unique proteins that were only observed in the pellet *or* supernatant, along with the percentage of both detected in each respective sequence and all sequences at the same time. For the combined pellet fraction, most of the total proteins were observed in sequence 4, while most of the unique proteins were observed in sequence 2.

Meanwhile, sequence 2 had the most observed proteins in the total of fraction S, while sequence 3 had a higher percentage of unique proteins.

Table 4.13: **ZnCl$_2$ protein precipitation: Unique protein percentage.** Number of total ($\sum$) proteins and the percentage of unique (with >2 MS2 spectral counts) IDs found in the combined pellet (P) or supernatant (S) fractions, and the percentage of which were observed in each sequence, 2, 3, and 4, and in all sequences simultaneously (All). Highlighted highest percentage of total and unique proteins (pink).

		$\sum$	2	3	4	All
P	Total	531	90%	90%	91%	67%
	Unique	109	73%	62%	55%	36%
S	Total	437	90%	72%	69%	48%
	Unique	26	62%	73%	46%	27%

4.2.4 High-pH Reversed-Phase Liquid Chromatography (PepSwift™)

(a) Replicate A

(b) Replicate B

(c) Replicate C
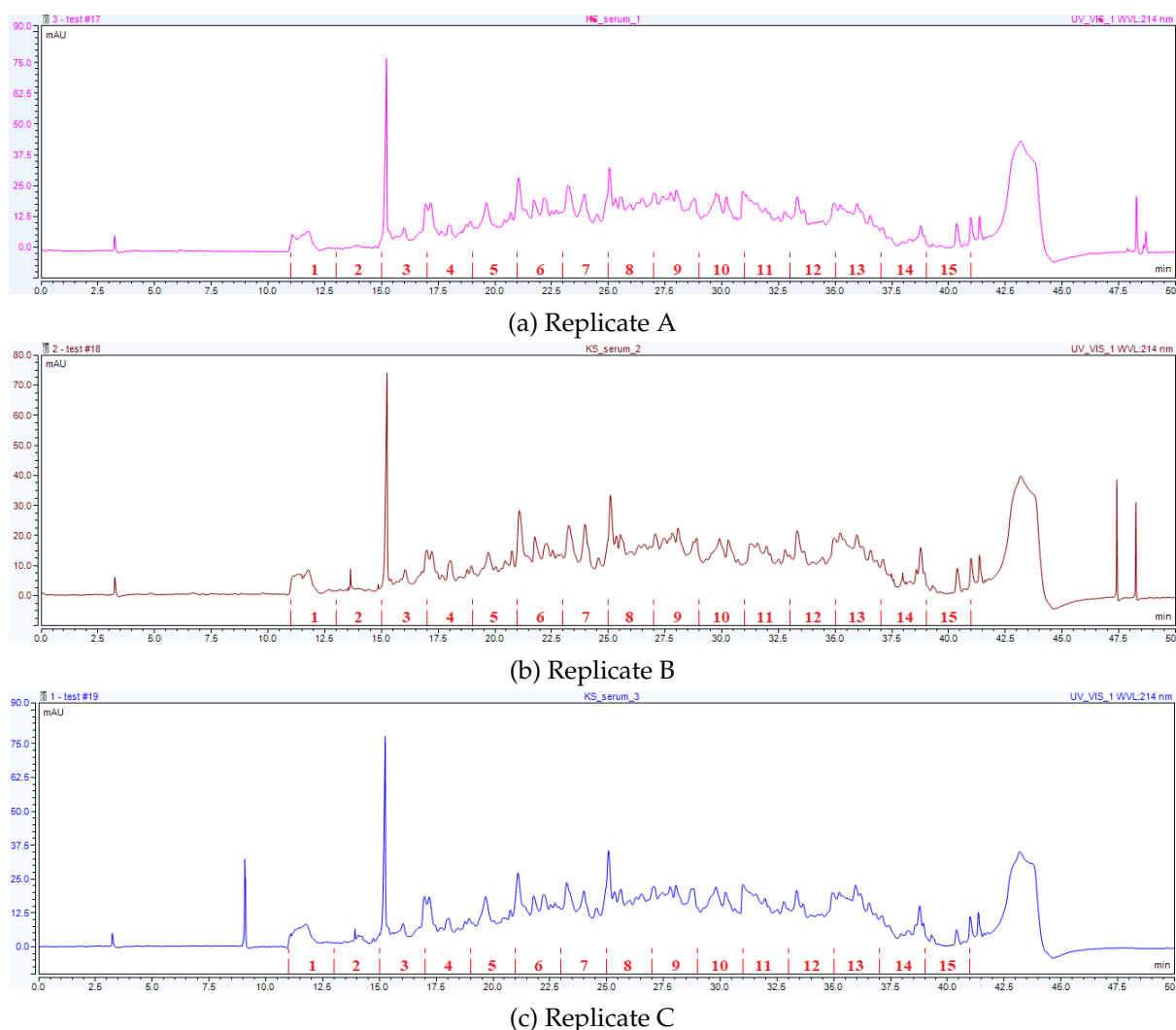
Figure 4.11: **HpH-RP LC (PepSwift): UV-VIS 214 nm chromatograms of serum peptide fractionation.** Chromatogram of replica A (top, pink), B (middle, brown) and C (bottom, blue) with the corresponding fractions numbered in red.

HpH-RP employs a non-polar stationary phase and polar mobile phase with an increasing gradient of organic solvent to separate peptides based on hydrophobicity. This was performed here using a monolithic C18 PepSwift™ column on a HPLC system. The UV-visible spectrophotometry (VIS) (214 nm) chromatograms from each replicate can be seen in Figure 4.11, illustrating the time points at which the 15 fractions (two minutes per fraction) were collected. In these chromatograms, peak intensity corresponds to the amount of peptide collected in each fraction. The similarities in the three chromatograms point to a high degree of reproducibility.

After peptide separation, each fraction was analyzed by LC-MS, and the resulting MS2 spectral data was compiled in Table 4.14. This table includes the total MS2 spectral counts and protein IDs, along with the MS2 spectral count percentage of abundant protein groups of each fraction and the fraction total ($\sum$). There were slightly more protein IDs in the HpH-RP LC total, even if there were more MS2 spectral counts in the control. As for the protein groups, HpH-RP LC identified more apolipoprotein, β-globulin, immunoglobulin, as well as other, less abundant protein spectra than the control.

Table 4.14: **HpH-RP LC (PepSwift): MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of HpH-RP LC fractionation method. Highlighted highest protein IDs among the fractions (dark blue, underline), and highest % SC of each column (light blue).

		Fractions							
	ctrl	1	2	3	4	5	6	7	8
Total SC	15,330	105	253	899	1,161	1,297	1,439	1,564	1,417
Total IDs	296	18	39	106	124	132	137	137	130
% SC α-G	19%	11%	16%	12%	15%	11%	13%	14%	17%
% SC Apo	9%	10%	6%	15%	18%	13%	10%	9%	9%
% SC β-G	7%	2%	3%	4%	6%	10%	15%	12%	9%
% SC CF	17%	10%	13%	13%	12%	11%	11%	11%	12%
% SC HSA	14%	45%	31%	20%	14%	16%	11%	10%	8%
% SC Ig	16%	7%	13%	16%	15%	21%	21%	22%	23%
% SC Other	19%	15%	18%	19%	22%	21%	20%	23%	22%
	$\sum$	9	10	11	12	13	14	15	
Total SC	11,387	1,283	665	333	227	107	22	7	
Total IDs	299	129	90	56	41	22	9	3	
% SC α-G	14%	15%	15%	18%	20%	21%	14%	0%	
% SC Apo	11%	10%	14%	11%	9%	9%	9%	0%	
% SC β-G	9%	7%	8%	8%	9%	8%	5%	0%	
% SC CF	12%	12%	15%	17%	17%	14%	9%	0%	
% SC HSA	12%	7%	9%	11%	11%	17%	36%	57%	
% SC Ig	22%	28%	20%	16%	19%	14%	14%	0%	
% SC Other	22%	23%	24%	23%	18%	19%	23%	43%	

In the individual fractions, most proteins were identified in fractions 5 - 8, with 130 or more protein IDs in each. The number of protein IDs diminished in both directions from these fractions, where fraction 3, 4, and 9 had > 100 IDs, and the remainder < 100. Fraction 15 only had 3 protein IDs. This is also show in Figure 4.12.
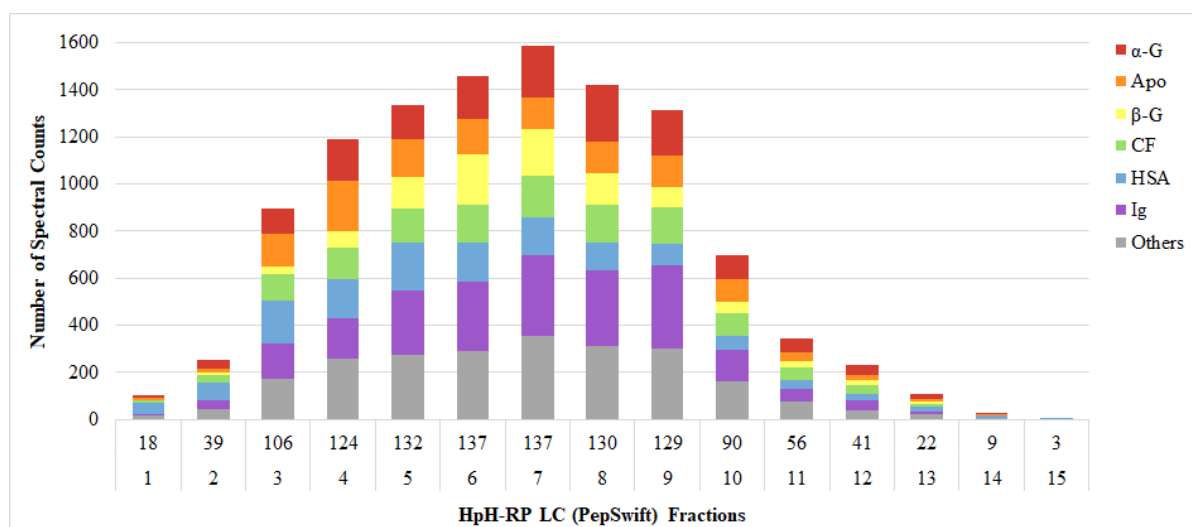
Figure 4.12: **HpH-RP LC (PepSwift): Protein MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (blue), Igs (purple) and Others (gray) in each fraction, with the number of protein IDs below each bar.

The fractions unique protein content was also examined, to see whether it was possible that all peptides of one protein would be exclusively eluted in one fraction. This is shown in Table 4.15. Here, all fractions had ≤ 5 unique protein IDs, whereas most only had 1.

Table 4.15: **HpH-RP LC (PepSwift): Unique proteins.** The number of proteins identified (with >2 MS2 spectral counts) exclusively in each HpH-RP LC fraction, along with the human UniProt entry name. Proteins that were not found in the control either are highlighted in blue.

	No.	Accession
2	1	SAA2
3	3	C1RL, HV320, HV374
4	2	1433Z, IBP2
6	1	MMRN1
7	1	KV240
9	5	KV113, LV321, LV537, MA2A1, PTGDS
11	1	CO4A

4.2.5 High-pH Reversed-Phase Spin Column (Pierce™ Kit)

High-pH reversed-phase peptide fractionation was also accomplished using Pierce™ spin column kit. With this method, 8 fractions were collected, along with the flowthrough after loading the column and a wash using water. MS2 spectral data from LC-MS analysis of these fractions are presented in Table 4.16, together with the total ($\sum$) of all fractions and a control. They can also be viewed in Figure 4.13.

Table 4.16: **HpH-RP spin column (Pierce kit) MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of HpH-RP spin column peptide fractionation kit. Both individual and concatenated fractions are included. Highlighted highest protein IDs among the fractions (dark blue, underline), and highest % SC of each column (light blue).

| | F* | W* | 1 | Fraction | | | | | | |
				2	3	4	5	6	7	8
Total SC	1,434	251	3,485	5,468	8,930	10,275	11,622	11,389	9,937	11,891
Total IDs	89	22	158	172	220	226	231	232	221	240
% SC α-G	26%	27%	15%	21%	20%	20%	20%	20%	19%	19%
% SC Apo	6%	8%	9%	12%	9%	9%	10%	11%	10%	8%
% SC β-G	7%	10%	7%	6%	7%	7%	8%	7%	8%	6%
% SC CF	14%	8%	20%	15%	13%	14%	14%	14%	15%	18%
% SC HSA	23%	34%	18%	16%	18%	16%	14%	13%	14%	12%
% SC Ig	13%	5%	12%	10%	12%	14%	13%	13%	15%	15%
% SC Other	12%	9%	19%	22%	22%	20%	20%	20%	20%	22%

| | Fraction | | Concatenation | | | | |
	ctrl	Σ	Σ	1+5	2+6	3+7	5+8
Total SC	13,697	72,998	45,300	11,074	10,238	11,264	12,723
Total IDs	253	406	396	297	282	291	310
% SC α-G	18%	20%	20%	19%	20%	20%	21%
% SC Apo	9%	10%	9%	10%	12%	8%	8%
% SC β-G	7%	7%	6%	8%	6%	6%	6%
% SC CF	17%	15%	15%	14%	14%	14%	16%
% SC HSA	13%	15%	13%	13%	12%	15%	13%
% SC Ig	15%	13%	15%	16%	14%	15%	16%
% SC Other	20%	21%	22%	21%	23%	22%	20%

* F = Flowthrough, **W = Wash

The total proteins identified using HpH-RP spin column fractionation was 406, which was about 63% more than in the control. There was likewise more MS2 spectral counts in the fractions total. Among the protein groups, the percentage of albumin, α-globulins, apolipoproteins, and the less abundant proteins were higher than in the control, β-globulins were equal, and complement factors and immunoglobulins were lower. Of the individual fractions, the most proteins IDs were in fraction 8, with 240 IDs. Following this, fractions 6 and 5 had > 230 IDs, 4, 3, and 7 ≥ 220 IDs, and 1 and 2 > 150. The flowthrough and wash had the lowest number of identified proteins, with 89 and 22 IDs, respectively. These two fractions also had a majority (> 20%) of albumin and α-globulin MS2 spectral content. In fraction 1, complement factors were the most abundant spectra, but also had high albumin and less abundant protein counts. Fractions 2 - 8 had the highest percentage (≥ 20%) of the less abundant proteins, but also higher α-globulin percentage.

Also presented in Table 4.16 are the concatenate of fractions 1 + 5, 2 + 6, 3 + 7, and 5 + 8. These concatenates were made by merging the MS2 spectral data of the respective fractions together. Doing this gave a more even distribution of protein IDs in each fraction, about 300 IDs in each, but less total IDs than the fraction total. The protein group percentiles were similar between

the concatenation and fraction totals, yet the concatenation total had higher percentages of immunoglobulins and less abundant proteins, and lower albumin, apolipoprotein, and β-globulin values. For the concatenation fractions, α-globulins and less abundant proteins were also here the most abundant protein groups for all, at about 20%.
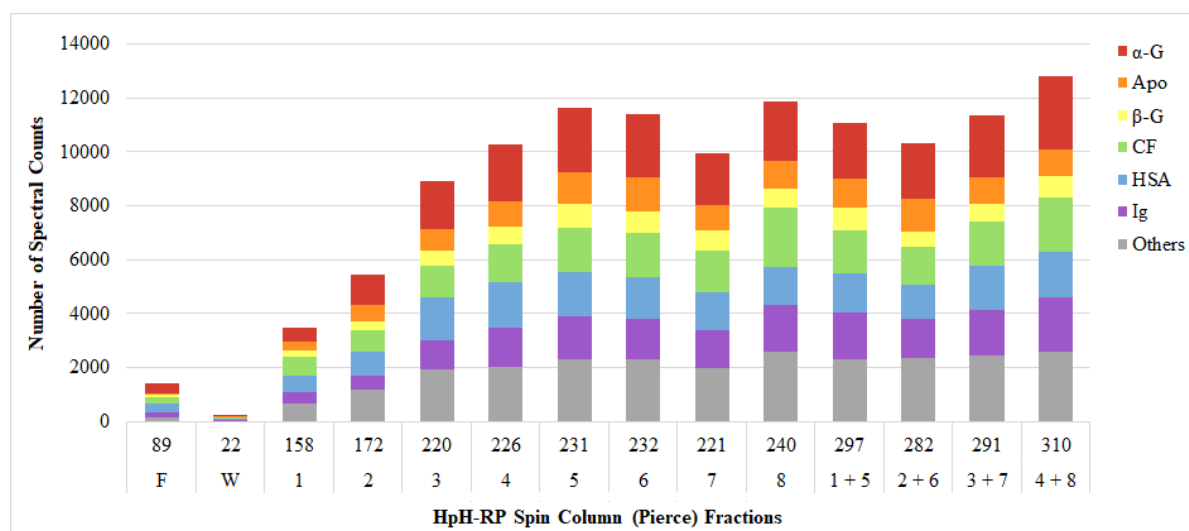
Figure 4.13: **HpH-RP spin column (Pierce kit): Protein group MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (blue), Igs (purple) and Others (gray) in the flowthrough (F), wash (W), and each fraction 1 - 8, with the number of protein IDs below each bar.

Table 4.17: **HpH-RP spin column (Pierce kit): Unique proteins.** The number of proteins identified (with >2 MS2 spectral counts) exclusively in each fraction and concatenate (right), along with the human UniProt entry name of the four most abundant. Proteins that were not found in the control either are highlighted in blue.

		Fractions			Fractions			Concatenate
	No.	Accession		No.	Accession*		No.	Accession*
F	10	DIC, HV205, HV323, PTBP1	**W**	0				
1	12	LV545, PF4V, PLGB, TPM3	**5**	3	CAD13, PCP, PDIA6		7	CFA43, EGLN, MMRN1, PDIA6
2	4	EST1, IBP4, RTN4R, S38AA	**6**	2	H2BFS, IGM		4	BLMH, MTPN, SEPR, URP2
3	3	KV240, LDHA, LITD1	**7**	4	KV139, MINP1, NCDN, WFDC3		4	ANGL3, BCAM, PECA1, TBA1A
4	3	CILP2, LEG3, LIPE	**8**	17	CRP, HV313, HV70D, SPB3		8	ACE, IPSP, LAMP1, SPB3

Moving on to look at the uniquely identified proteins in each fraction, seen in Table 4.17. Unlike with the HpH-RP LC (PepSwift) peptide fractionation method, here, two fractions, 1 and 8, had > 10 unique protein IDs. Of particular interest was the relatively high amount of unique IDs in the flowthrough fraction.

4.2.6 Coupled Zinc Chloride Protein Precipitation & High-pH Reversed-Phase spin column (Pierce™ kit) Peptide Fractionation

Based on the results from the protein and peptide fractionation, it was decided to pursue further increase of identified proteins by coupling the protein fractionation method of $ZnCl_2$ precipitation (sequence 2) with the HpH-RP spin column (Pierce Kit) peptide fractionation method. In other words, doing $ZnCl_2$ precipitation before protein digestion, then doing HpH-RP using the spin column kit on the resulting peptides. This gave 16 fractions per replica, where each fraction was labeled P1/S and 1 - 8 in accordance with their $ZnCl_2$ and HpH-RP spin column fractions, respectively. Results from the LC-MS analysis of these fractions, along with the total ($\sum$) of all 16 fractions combined, are presented in Table 4.18. Figure 4.14 also shows this data, but only for the fractions. Where the $ZnCl_2$ precipitation method identified ~30% more proteins, and HpH-RP spin column ~60% more than their respective control, combining them resulted in just under 300 protein IDs, which was about ~30% more than the control. The total distribution of the protein groups was here more similar to that of $ZnCl_2$ precipitation than HpH-RP spin column peptide fractionation. Compared to the control, the total had less α-globulins, apolipoproteins, albumin and immunoglobulin, and more of the less abundant proteins. In contrast to lone $ZnCl_2$ precipitation, a higher percentage of complement factors were observed. Interestingly, the difference to the control in percentage MS2 spectral counts of less abundant proteins was greater here (+10%) than for either $ZnCl_2$ precipitation (+4%) or HpH-RP spin column (+1%).
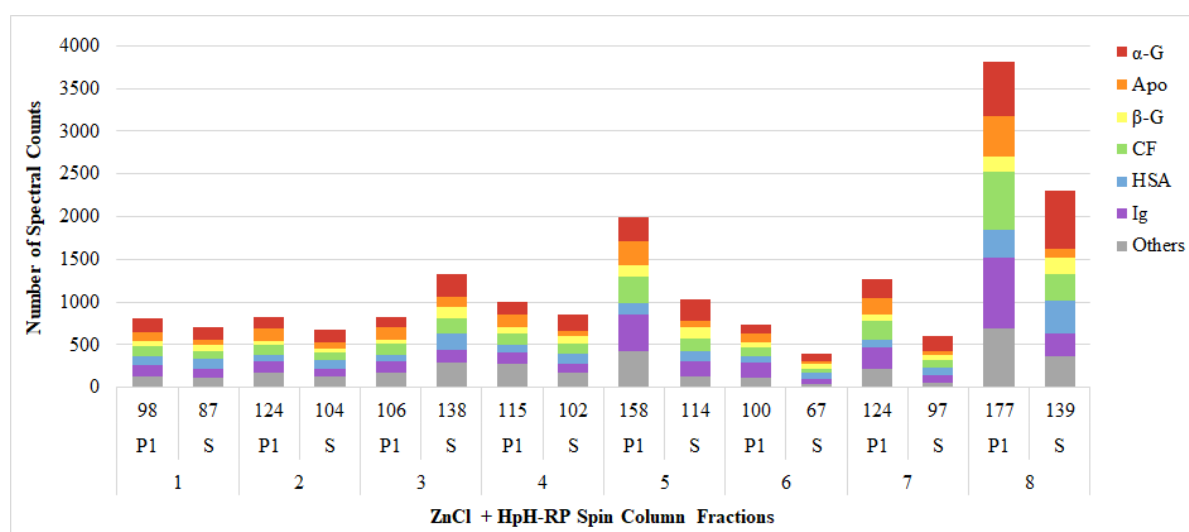
Figure 4.14: **$ZnCl_2$ precipitation + HpH-RP spin column (Pierce kit): Protein group MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (blue), Igs (purple) and Others (gray) in each $ZnCl_2$ (P1 & S) and HpH-RP spin column (1 - 8) fractions, with the number of protein IDs below each bar.

The fractions with the most protein IDs were P1-8 and -5, with 177 and 158 IDs, respectively. Of the S fractions, both S-3 and S-8 had $\geq$ 130 IDs. Fractions with > 110 protein IDs were P1-2, -4, -7, and S-5. S-1, -6, -7, and P1-1 had <100 IDs, whilst the others had about 100 identified proteins. The most abundant protein group with $\geq$ 20% MS2 spectral counts was α-globulin in all S fractions, except S-3 where there were more of the less abundant proteins. In the P1 fractions, P1-1 also had more α-globulins, whilst 2 - 5 had most «others», and 6 - 8 had more immunoglobulins.

Table 4.18: **ZnCl$_2$ precipitation + HpH-RP spin column (Pierce kit): MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of coupled ZnCl$_2$ protein precipitation and HpH-RP spin column peptide fractionation methods. Highlighted highest protein IDs among the fractions (purple, underline), and highest % SC of each column (pink).

HpH-RP spin column Fraction:	1	2	3	4	5	6	7	8
Ctrl			**ZnCl$_2$ Fraction P1**					
Total SC	14,836	814	828	826	1,002	1,989	738	1,259	3,807
Total IDs	219	98	124	106	115	158	100	124	177
% SC α-G	19%	21%	17%	15%	16%	14%	15%	17%	17%
% SC Apo	10%	12%	19%	17%	15%	14%	15%	14%	12%
% SC β-G	7%	8%	5%	6%	7%	6%	8%	6%	5%
% SC CF	15%	15%	14%	15%	13%	16%	14%	18%	18%
% SC HSA	11%	13%	9%	9%	8%	6%	9%	7%	8%
% SC Ig	17%	16%	16%	17%	13%	21%	23%	20%	22%
% SC Other	20%	15%	21%	20%	28%	22%	16%	17%	18%
Σ			**ZnCl$_2$ Fraction S**					
Total SC	13,878	704	676	1,331	864	1,137	442	675	2,456
Total IDs	296	87	104	138	102	114	67	97	139
% SC α-G	16%	22%	22%	20%	21%	21%	22%	26%	27%
% SC Apo	9%	9%	11%	10%	8%	7%	7%	7%	4%
% SC β-G	7%	9%	5%	9%	10%	11%	11%	9%	8%
% SC CF	17%	13%	14%	13%	13%	13%	12%	12%	13%
% SC HSA	10%	16%	14%	14%	14%	11%	16%	12%	16%
% SC Ig	15%	15%	13%	11%	11%	15%	15%	14%	11%
% SC Other	30%	16%	20%	22%	20%	12%	7%	8%	15%

Table 4.19: **ZnCl$_2$ precipitation + HpH-RP spin column (Pierce kit): Protein distribution.** The percentage of total spectral count (SC), protein IDs, and protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) distributed in each fraction (left) compared to the ZnCl$_2$ 2-sequence protein fractionation from Table 4.11 (right). Highlighted highest percentage of each row per method (pink).

	Coupled		ZnCl$_2$	
	P1	**S**	**P1**	**S**
Total SC	57%	43%	57%	43%
Total IDs	93%	81%	90%	68%
% SC α-G	49%	51%	50%	50%
% SC Apo	73%	27%	78%	22%
% SC β-G	47%	53%	48%	52%
% SC CF	63%	37%	68%	32%
% SC HSA	44%	56%	39%	61%
% SC Ig	68%	32%	65%	35%
% SC Other	55%	45%	58%	42%

The distribution of proteins between the two $ZnCl_2$ fractions here, as seen in Table 4.19 to the left, was almost the same as for the two-sequence precipitation results on the right. Here, a higher percentage of proteins were identified in both fractions, while the total MS2 spectral count distribution remained the same. For the protein groups, α-, β- and immunoglobulins had a greater difference between the two fractions than previously, whilst the other groups were more evenly distributed.

Table 4.20: **$ZnCl_2$ precipitation + HpH-RP spin column (Pierce kit): Unique proteins.** The number of proteins identified (with >2 MS2 spectral counts) exclusively in each individual $ZnCl_2$ fraction (top), HpH-RP spin column fraction (middle), and total $ZnCl_2$ fraction (bottom), along with the human UniProt entry name of the four most abundant.

	P1		S	
	Nr. of	Accession	Nr. of	Accession
Between $ZnCl_2$ fractions				
1	4	APMAP, COMP, FHR1, FIBB	2	EF1A2, IGHG2
2	9	BTD, C1R, FHR1, HV374	5	EF1A2, HBA, RET4, RS23
3	6	APOA, HGFL, IGHG2, TSP1	9	A2GL, EF1A2, HPTR, LG3BP
4	10	FIBG, HABP2, TSP1, VWF	5	ATRN, EF1A2, LG3BP, TFR1
5	28	FA10, IGHG2, IGKC, TSP1	4	ATRN, HV313, LG3BP, RS23
6	9	APOE, C1R, HEP2, IGHG2	2	IC1, LUM
7	16	CO4A, IGHG2, LBP, PCYOX	4	IC1, PHLD, PZP, THBG
8	31	C1QB, C4BPA, FIBG, IGL1	2	A2GL, TAGL2
Between HpH-reversed-phase (RP) fractions				
1	0		0	
2	3	HV374, IBP3, LMAN2	2	HV108, HV309
3	0		2	HV335, KVD21
4	0		0	
5	2	F13A, TLN1	1	IGL1
6	0		0	
7	1	C1RL	0	
8	8	ATRN, CALL5, CD14, KV127	6	CALL5, CD14, CO4A, LBP
Total	2	COMP, FHR1	0	

Unique proteins are once again shown in Table 4.20. Here, the two $ZnCl_2$ precipitation fractions from each HpH-RP fraction were compared at the top of the table, as in P1-1 versus S-1, and so on. Below this, the eight HpH-RP fractions from each $ZnCl_2$ fraction were compared. To begin with the pairwise $ZnCl_2$ fractions, the P1 fraction had more unique proteins for all but fraction 3. This was especially true for P1-5 and P1-8. When the HpH-reversed-phase (RP) fractions were compared to each other, fraction 8 of both P1 and S had the most unique proteins identified. At the very bottom of the table, the total of the P1 and S fractions were correlated. Here, P1 had 2 unique IDs whilst S had zero, which was considerably less than the 88 unique P1 and 27 unique S proteins from the preceding $ZnCl_2$ precipitation. None of the uniquely identified proteins presented here were not also present in the control as well.

4.2.7 Comparison of Fractionation Methods

To compare the different fractionation methods, a side-by-side comparison of the overall results was required. Figure 4.15 shows Venn diagrams that compare the total number of proteins identified using each method with their respective controls. To the left in each diagram are the number of protein IDs not found in the control, along with the constituent percentage of the total IDs. In the center are the shared protein IDs, identified both with and without using the method. To the right are the number and percentage of protein IDs only found in the control. All methods identified more unique proteins than their control, aside from ethanol protein precipitation and the coupled $ZnCl_2$ precipitation + HpH-RP spin column (Pierce kit) method. The latter was also the method with the highest percentage of unique proteins exclusively identified in the control, at 20%, where the second most was HpH-RP LC (PepSwift) with 12%. The remainder had 0 - 3% unique control protein IDs. Meanwhile, the three size exclusion chromatography totals all had $\geq$ 50% protein IDs exclusive to the method.
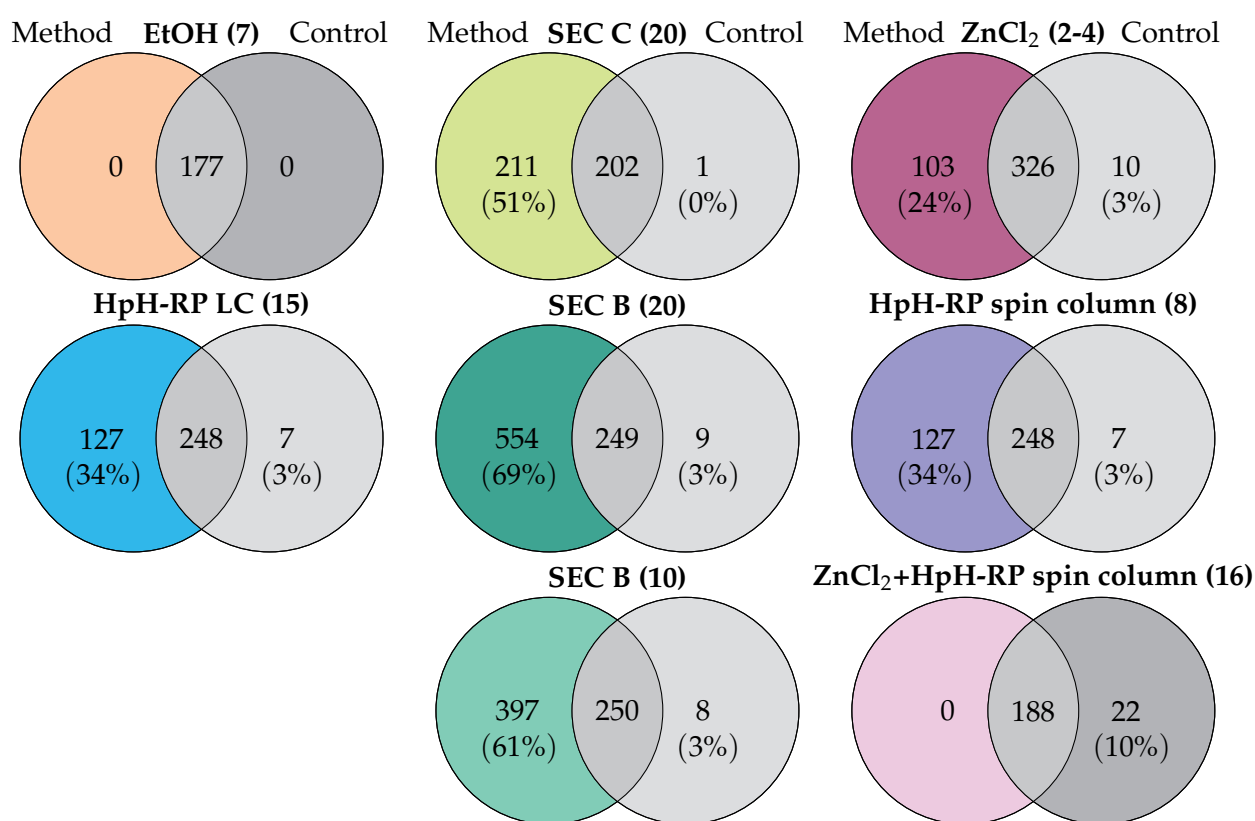
Figure 4.15: **Fractionation: Method versus Control.** The number of proteins identified in each fractionation method compared to the undepleted serum control, with the number of fractions in brackets next to method name. Unique proteins (>2 MS2 spectral counts) identified only using fractionation (left), control (right) or in both (middle), with the percentage of unique proteins in brackets under number of unique IDs.

The total number of proteins identified in each method is shown as a boxplot diagram in Figure 4.16. This image shows the difference in number of protein IDs between each method, the contrast to their respective control, and the variation in the replicates of one method. Most proteins were identified using size exclusion chromatography. After SEC comes all three sequences of $ZnCl_2$ protein precipitation, followed by HpH-RP spin column (Pierce kit). HpH-RP LC and coupled $ZnCl_2$ precipitation + HpH-RP spin column identified about the same amount of proteins, whilst the least amount of proteins were identified using ethanol protein precipitation. In regard to the improvement in protein IDs compared to their control,

the largest difference was seen again using SEC. After which the HpH-RP spin column, then ZnCl$_2$ precipitation, followed by coupled ZnCl$_2$ + HpH-RP spin column were the most different from their control. Neither ethanol precipitation, nor HpH-RP LC were significantly improved. Regarding the variance in replicates, the most consistent results were achieved with coupled ZnCl$_2$ + HpH-RP spin column and ethanol precipitation. HpH-RP LC and ZnCl$_2$ precipitation also gave fairly consistent values, followed by HpH-RP spin column. Meanwhile, the size exclusion chromatography results were highly variable.
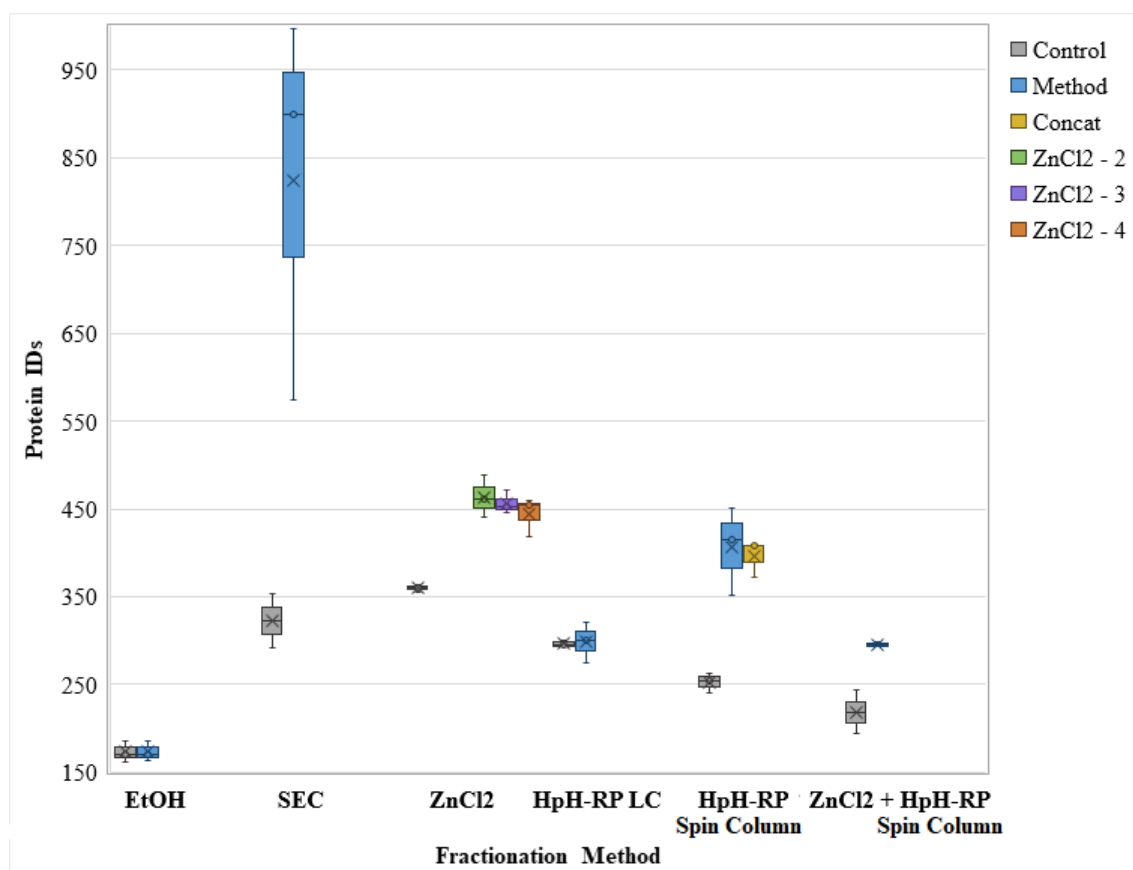
Figure 4.16: **Fractionation: Boxplot diagram showing the number of identified proteins in each fractionation method:** Controls (gray), fractionation method (blue), concatenate (yellow) and ZnCl$_2$ sequence 2 (green), 3 (purple) and 4 (orange).

4.3 Cysteine-containing Peptide Enrichment

Finally, enrichment of cysteine-containing peptides was pursued using four thiol-reactive bead-resins: BcMag™ Thiol-Activated Magnetic Beads (BcMag Thiol), BcMag™ Long-arm Thiol-Activated Magnetic Beads (BcMag long), High Capacity Acyl-rac S3™ Capture Beads (S3), and Thiopropyl Sepharose™ 6B Affinity Resin (S6B). These beads had pyridyl disulfide reactive groups that target the thiol-group on cysteine side-chains, thus making it possible to separate the cysteine-containing peptides from the non-cysteine containing peptides. By doing this, the complexity of the sample can be reduced.

4.3.1 Single Protein Analysis (BSA Test)

First, each bead type was tested on bovine serum albumin (BSA). Since no protocol was given for the S3 capture beads, the BcMag magnetic bead (BM) method and the thiopropyl sepharose

6B (S6B) method were both tested for this resin. Performing one replicate of each method, the samples with cysteine-containing BSA peptides from each, along with a control of BSA without enrichment, were analyzed by LC-MS. Figure 4.17 shows the resulting sequence coverage from each sample. Identified peptide sequences are highlighted in yellow, where the enclosed cysteine (and methionine) residues are in green. Here, no cysteine-containing peptides were observed in the control, but some were in all of the methods.
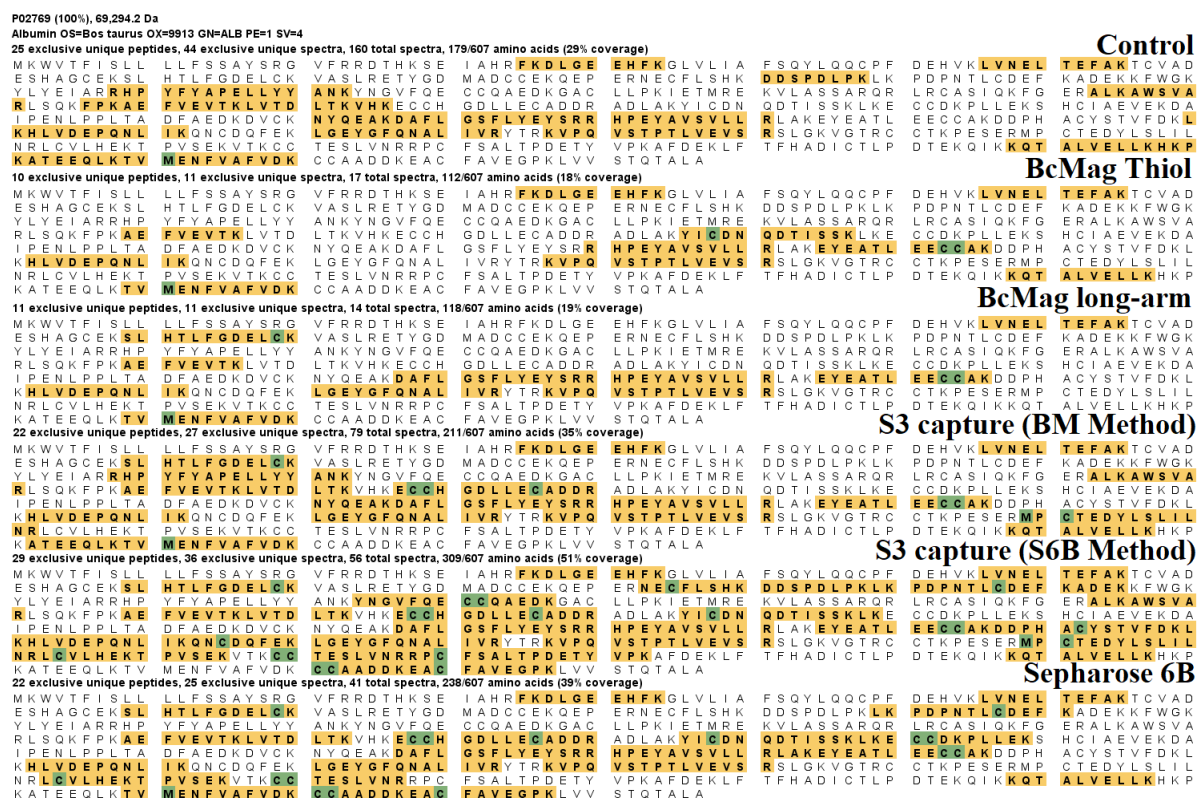
Figure 4.17: **Single protein analysis: Sequence coverage.** Covered sequence of BSA (yellow) with cysteine/methionine (green) between control, BcMag thiol and long-arm, S3 capture, and sepharose 6B.

Table 4.21: **Single protein analysis: Cysteine-containing peptide percentage.** Number of total BSA MS2 spectral counts and the percentage of cysteine-containing peptides found in the control, BcMag thiol and long-arm, S3 capture, and sepharose 6B. Highlighted highest percentage of cysteine-containing peptides (yellow).

	Total Spectra	%Cysteine-containing
BSA control	160	~0%
BcMag thiol	17	12%
BcMag long	14	14%
S3 (BcMag)	79	5%
S3 (S6B)	56	57%
Sepharose 6B	41	49%

The number of total MS2 spectral counts, and the percentage of these that contain cysteine, are shown in Table 4.21. In accordance with Figure 4.17, the control had no observed cysteine-containing peptides. The methods with the highest amount of cysteine-containing peptides

were sepharose 6B and S3 capture using the S6B method, where about 50% of the identified MS2 spectra contained cysteine. Both BcMag magnetic bead methods had few MS2 spectral counts and < 15% of these contained cysteine. S3 capture with the BM method gave only 5% cysteine-containing peptides, despite having the most MS2 spectral counts of the methods.

4.3.2 Complex Proteome Analysis (Serum)

Based on the results from the single protein analysis, Thiopropyl sepharose 6B affinity resin was selected to go forward with serum trials. First, the same approach was applied to three replicates of serum. Samples from both the flowthrough, with non-cysteine-containing peptides, and the enrichment, with cysteine-containing peptides, were analyzed. These are labeled «Non» and «Cys», respectively, in the following results. The MS2 spectral data from these replicates are labeled «S6B». After these initial replicates, several changes to the procedure described in section 3.4.4 were examined. One analysis of each revision was performed. The ratio between serum and beads, labeled «Bead ratio», was examined with 50, 70, and 90 mg beads, compared to the initial 30 mg used. To the 70 mg bead test, extra elution steps, labeled «Elution», with TCEP, 80% ACN, and urea were performed. 50 mg beads were then used in the subsequent tests. An extra micro-SPE purification step, labeled «ZipTip», was added after the initial incubation in order to remove DTT before adding the sample to the resin. Finally, samples were taken from the washing steps, labeled «Wash», with washing buffer (WB1), 1M NaCl, 0.1% TFA/80% ACN (ACN), washing buffer again (WB2), and coupling buffer (CB). The following results give an overview of the outcome from the initial replicates along with the ensuing tests.

Table 4.22: **Complex proteome analysis: Cysteine-containing peptide percentage.** Number of HSA MS2 spectra and the percentage of cysteine-containing peptides found in the control, bead-ratio, ziptip, wash, and elution tests of thiopropyl sepharose 6B. Highlighted highest percentage of cysteine-containing peptides per test (yellow).

Spectra:	Flowthrough		Enrichment	
	Total	% Cys	Total	% Cys
ctrl	1280	57%		
S6B	213	34%	132	40%
50mg	842	58%	23	52%
70mg	521	56%	70	50%
90mg	545	57%	50	42%
ZipTip	238	53%	25	36%
Wash	427	41%	47	45%

	Wash		Elution		
	Total	% Cys		Total	% Cys
Wash Buffer 1	187	51%	TCEP	94	45%
NaCl	133	44%			
0.1%TFA/80%ACN	81	46%	80%ACN	90	50%
Wash Buffer 2	92	41%			
Coupling Buffer	61	59%	Urea	76	49%

The percentage of the HSA spectra that contain cysteine residues can be viewed in Table 4.22 along with the total number of HSA spectra. In all cases, there were more MS2 spectral counts in the control samples than the enriched samples, and more in the flowthrough than the

enrichment. Cysteine-containing peptides were 57% in the control. For the initial «S6B» run the number of cysteine-containing peptides was also higher in the enrichment, but only at 40%. For the following tests, more cysteine-containing peptides were found in the flowthroughs, (∼55%), except for the final wash test that had more in the enrichment (45%). Samples from the washing steps, taken between the corresponding flowthrough and enrichment samples, had MS2 spectral counts that fall between these. The percentage of cysteine-containing peptides were also higher in the first, third, and last washing step than the enrichment. The additional elution samples from the 70 mg beads had slightly more MS2 spectral counts, but with the same or less percentage of cysteine-containing peptides.

ALBU_HUMAN (100%), 69,366.9 Da
Serum albumin OS=Homo sapiens OX=9606 GN=ALB PE=1 SV=2
83 exclusive unique peptides, 153 exclusive unique spectra, 1273 total spectra, 547/609 amino acids (90% coverage)

```
MKWVTFISLL  FLFSSAYSRG  VFRRDAHKSE  VAHRFKDLGE  ENFKALVLIA  FAQYLQQCPF  EDHVKLVNEV  TEFAKTCVAD
ESAENCDKSL  HTLFGDKLCT  VATLRETYGE  MADCCAKQEP  ERNECFLQHK  DDNPNLPRLV  RPEVDVMCTA  FHDNEETFLK
KYLYEIARRH  PYFYAPELLF  FAKRYKAAFT  ECCQAADKAA  CLLPKLDELR  DEGKASSAKQ  RLKCASLQKF  GERAFKAWAV
ARLSQRFPKA  EFAEVSKLVT  DLTKVHTECC  HGDLLECADD  RADLAKYICE  NQDSISSKLK  ECCEKPLLEK  SHCIAEVEND
EMPADLPSLA  ADFVESKDVC  KNYAEAKDVF  LGMFLYEYAR  RHPDYSVVLL  LRLAKTYETT  LEKCCAAADP  HECYAKVFDE
FKPLVEEPQN  LIKQNCELFE  QLGEYKFQNA  LLVRYTKKVP  QVSTPTLVEV  SRNLGKVGSK  CCKHPEAKRM  PCAEDYLSVV
LNQLCVLHEK  TPVSDRVTKC  CTESLVNRRP  CFSALEVDET  YVPKEFNAET  FTFHADICTL  SEKERQIKKQ  TALVELVKHK
PKATKEQLKA  VMDDFAAFVE  KCCKADDKET  CFAEEGKKLV  AASQAALGL
```

(a) HSA sequence

Sequence Coverage	Protein	Category	Bio Sample	%Cov
	Serum albumin	Sepharose 6B	S6B_Non_A	50%
	Serum albumin	Sepharose 6B	S6B_Non_B	45%
	Serum albumin	Sepharose 6B	S6B_Non_C	62%
	Serum albumin	Sepharose 6B	S6B_Non_Merge	63%
	Serum albumin	Sepharose 6B	S6B_Cys_A	60%
	Serum albumin	Sepharose 6B	S6B_Cys_B	36%
	Serum albumin	Sepharose 6B	S6B_Cys_C	22%
	Serum albumin	Sepharose 6B	S6B_Cys_Merge	60%
	Serum albumin	Sepharose 6B	S6B-Merge	67%
	Serum albumin	Bead_Ratio	50mg_Non	87%
	Serum albumin	Bead_Ratio	50mg_Cys	21%
	Serum albumin	Bead_Ratio	50mg_Merge	87%
	Serum albumin	Bead_Ratio	70mg_Non	85%
	Serum albumin	Bead_Ratio	70mg_Cys	52%
	Serum albumin	Bead_Ratio	70mg_Merge	85%
	Serum albumin	Bead_Ratio	90mg_Non	85%
	Serum albumin	Bead_Ratio	90mg_Cys	42%
	Serum albumin	Bead_Ratio	90mg_Merge	85%
	Serum albumin	Elution	Elute_TCEP	72%
	Serum albumin	Elution	Elute_ACN	66%
	Serum albumin	Elution	Elute_Urea	63%
	Serum albumin	Wash	Wash_Non	81%
	Serum albumin	Wash	Wash_Cys	37%
	Serum albumin	Wash	Wash_WB1	75%
	Serum albumin	Wash	Wash_NaCl	68%
	Serum albumin	Wash	Wash_ACN	53%
	Serum albumin	Wash	Wash_WB2	61%
	Serum albumin	Wash	Wash_CB	46%
	Serum albumin	Wash	Wash_Merge	81%
	Serum albumin	Extra_ZipTip	ZipTip_Non	68%
	Serum albumin	Extra_ZipTip	ZipTip_Cys	31%
	Serum albumin	Extra_ZipTip	ZipTip_Merge	68%
	Serum albumin	Control	Ctrl	90%

(b) Sequence coverage overview.

Figure 4.18: **Complex proteome analysis: Overview of sequence coverage using thiopropyl sepharose 6B.** Covered HSA amino acid sequence (yellow) and cysteine/methionine (green) for the flowthrough (Non), enrichment (Cys), and combined (Merge) bio samples of initial replicates (S6B), bead ratio test (50mg, 70mg, and 90mg), elution test (TCEP, ACN, and Urea), washing test (WB1, NaCl, ACN, WB2, and CB), and zipTip test (ZipTip).

Table 4.23: **Complex proteome analysis: MS2 spectral data analysis.** Total spectral count (SC), protein IDs, and percentage MS2 spectral counts (% SC) of protein groups α-globulin (α-G), apolipoprotein (Apo), β-globulin (β-G), complement factor (CF), human serum albumin (HSA), immunoglobulin G (IgG), other immunoglobulins (Igs) and less abundant proteins (Other) of cysteine-containing peptide enrichment in serum with thiopropyl Sepharose 6B. Flowthrough (Non), enrichment (Cys), and total (∑) samples of initial replicates (S6B), elution test (TCEP, ACN, and Urea), zipTip test (ZipTip), bead ratio test (50mg, 70mg, and 90mg), and washing test (WB1, NaCl, ACN, WB2, and CB). Highlighted highest protein IDs among the tests (orange, underline), and highest % SC of each column (yellow).

		S6B			Elution			ZipTip		
	Ctrl	**Non**	**Cys**	**∑**	**TCEP**	**ACN**	**Urea**	**Non**	**Cys**	**∑**
Total SC	12,701	2,064	971	3,042	440	383	272	2,804	121	2,936
Total IDs	291	87	61	94	57	57	46	188	31	192
% SC α-G	17%	13%	10%	12%	26%	27%	26%	18%	18%	18%
% SC Apo	9%	8%	6%	7%	8%	6%	6%	10%	13%	10%
% SC β-G	7%	12%	11%	12%	9%	9%	10%	8%	13%	8%
% SC CF	18%	16%	11%	14%	10%	8%	6%	16%	4%	15%
% SC HSA	10%	10%	14%	11%	21%	23%	28%	8%	21%	9%
% SC Ig	14%	17%	22%	19%	13%	13%	14%	17%	17%	17%
% SC Other	26%	25%	26%	25%	13%	12%	10%	22%	14%	22%

Bead Ratio		90 mg			70 mg			50 mg		
		Non	**Cys**	**∑**	**Non**	**Cys**	**∑**	**Non**	**Cys**	**∑**
Total SC		6,539	163	6,740	6,613	245	8,013	8,241	63	8,314
Total IDs		238	16	239	235	29	242	246	17	247
% SC α-G		15%	32%	15%	17%	32%	18%	16%	19%	16%
% SC Apo		10%	12%	10%	10%	9%	10%	10%	10%	10%
% SC β-G		9%	14%	9%	8%	13%	8%	8%	11%	8%
% SC CF		18%	2%	18%	19%	3%	17%	18%	5%	18%
% SC HSA		8%	31%	9%	8%	30%	11%	10%	37%	10%
% SC Ig		16%	7%	16%	16%	10%	15%	16%	14%	16%
% SC Other		24%	2%	23%	23%	3%	21%	23%	5%	23%

			Wash					
	Non	**WB1**	**NaCl**	**ACN**	**WB2**	**CB**	**Cys**	**∑**
Total SC	3,832	1,629	814	400	461	216	160	7,648
Total IDs	213	160	116	66	74	45	36	217
% SC α-G	19%	16%	13%	13%	16%	6%	4%	16%
% SC Apo	7%	8%	8%	6%	5%	0%	1%	7%
% SC β-G	10%	12%	15%	17%	14%	15%	18%	12%
% SC CF	14%	13%	10%	4%	6%	1%	1%	12%
% SC HSA	11%	11%	16%	20%	20%	28%	29%	14%
% SC Ig	13%	14%	11%	10%	13%	3%	1%	13%
% SC Other	26%	26%	27%	31%	26%	47%	45%	27%

Figure 4.18 shows an overview of the sequence coverage of HSA in each test. 4.18a illustrates the amino acid sequence of HSA from the control. An overview of all of the tests is then shown in 4.18b. In both cases, the covered sequence is marked in yellow, while cysteine and methionine residues are colored green. As stated previously, the control sample, containing serum

that had not undergone enrichment, had a higher percentage coverage than all other samples. Other samples with > 80% sequence coverage were the total (merge) and flowthrough of all three bead ratio tests and the wash. In all cases the flowthrough «Non» samples had a higher sequence coverage than the respective enriched «Cys», with the exception of «S6B» replicate A.

Moving on to the other proteins identified in these Thiopropyl sepharose 6B tests, Table 4.23 shows the total MS2 spectral counts, protein IDs, and percentage MS2 spectral counts for all protein groups in serum. Very few proteins were identified in the enriched samples («Cys») the highest one being 61 protein IDs in the average of the initial «S6B» replicates. In all cases, more proteins were identified in the flowthrough samples («Non»), and yet still more in the control. Not even the combined total ($\sum$) of the flowthrough and enriched sample gave more protein IDs than the control. In the case of the bead ratio test, using 70 mg beads gave the most protein IDs in the enriched sample, however, more proteins were identified in the following elution steps for all reagents. The extra ZipTip step only increased the number of protein IDs in the flowthrough and total, but not the enriched sample. This was the same for the respective samples in the wash test. Here, all of the samples from the washing steps had more proteins than the final enriched sample, but less than the flowthrough. In regard to the protein groups, all «S6B» and Wash samples, as well as the remaining flowthrough, totals, and the control, had the highest percentage of the less abundant proteins. Most of these values were between 20 - 30%, whilst for the wash, the last step «CB» and the enrichment had > 40%. The remaining enriched samples had a low percentage of less abundant proteins, where it was $\leq$ 5% for all Bead-ratio enrichment samples, whilst the albumin and α-globulins were $\geq$ 20% in most. All additional elution steps also had the highest percentage of albumin and α-globulins. This can also be seen in Figure 4.19.
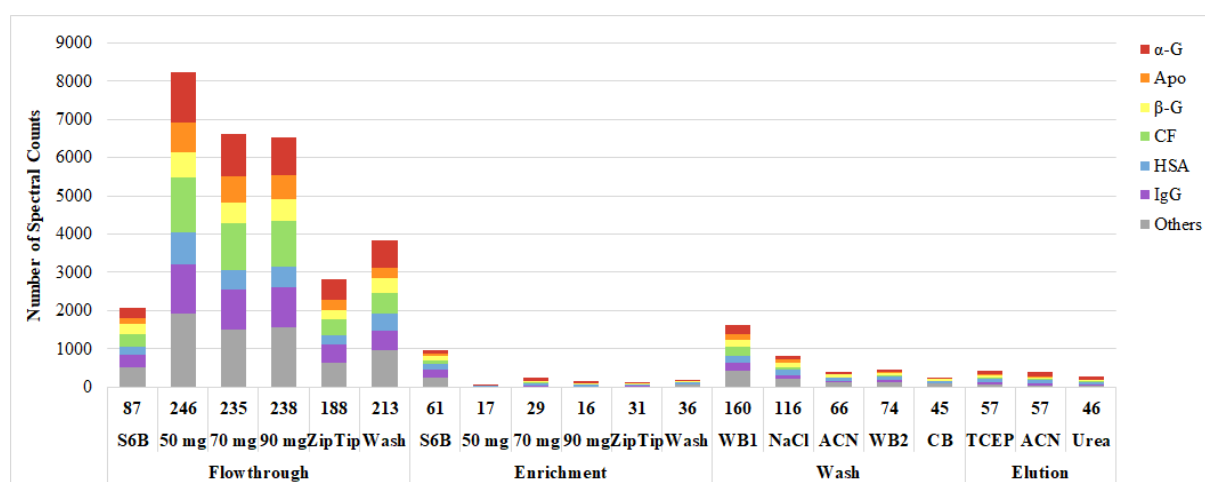
Figure 4.19: **Complex proteome analysis: Protein MS2 spectral counts.** The number of MS2 spectral counts of the protein groups α-G (red), Apo (orange), β-G (yellow), CF (green), HSA (blue), Igs (purple) and Others (gray) of cysteine-containing peptide enrichment in serum with thiopropyl Sepharose 6B. Flowthrough and enrichment of initial replicates (S6B), bead ratio test (50mg, 70mg, and 90mg), zipTip test (ZipTip), and wash, along with elution test (TCEP, ACN, and Urea), washing steps (WB1, NaCl, ACN, WB2, and CB).

5 Discussion

5.1 Depletion of Abundant Proteins

Albumin and immunoglobulin G were chosen as the targets for depletion due to being the most and second-most abundant protein groups in serum and plasma. With this in mind, six depletion kits specifically targeting one or both of these proteins were selected. The Albumin/IgG removal kit and AlbuSorb both bind the depletion targets, rather than the serum proteome, using cibacron Blue 3GA and polyelectrolyte beads, respectively. Both also utilize Protein A. Two kits, AlbuVoid PLUS and AlbuVoid LC-MS, are on-bead digestion methods where the serum proteome binds to the bead matrix, while the depletion targets are blocked from binding, and remain bound to the beads while digestion is performed. The BioMag kits are magnetic beads that use magnetic separation to easily remove the albumin containing supernatant, or the bead-bound IgG from the serum proteome. The ensuing discussion shall determine which of these depletion kits performed the best, in terms of successfully depleting the target proteins and yielding the highest number of identified proteins.

SDS-PAGE

Overall, the results from the SDS-PAGE gels in Figure 4.1 showed depletion of the intended target in most of the kits. Due to limited instructions on how to couple the BioMag HSA and BioMag IgG methods together, the test run was done by taking 1 µL of the resulting HSA depleted sample through to IgG depletion. As seen by the barely visible bands in the SDS-PAGE, this resulted in very poor yield. So, for the replicates, the entire sample from the HSA depletion was added to the BioMag IgG kit beads, on the basis that the depletion would have removed enough proteins from the sample so as not to overload the IgG kit beads. Therefore, the remainder of the discussion about the SDS-PAGE gels will exclude the results from this kit.

In the «Bead-bound/Flowthrough» samples, large bands of HSA were observed in the Albumin/IgG removal kit and AlbuSorb, and less so, but still visibly there, in the other albumin depleting kits. These two kits also had large amounts of other proteins in the bead-bound sample, showing to a high degree of co-depletion, likely caused by these proteins binding to the column resin. The kit with the least amount of co-depletion was the AlbuVoid LC-MS kit, which had a clear visible band of albumin and only faint bands of other protein. BioMag IgG, which only depletes IgG, had almost no albumin in its bead-bound sample, as expected, but instead a visible 50 kDa IgG and 25 kDa light-chain band, though only a slight 150 kDa band. Clear IgG bands were also observed in the bead-bound/flowthrough sample of the Albumin/IgG removal kit, AlbuSorb, and AlbuVoid PLUS, especially the 25 kDa band, but also in AlbuVoid PLUS HSA-flowthrough, AlbuVoid LC-MS, and BioMag HSA despite these kits not being marketed as being able to deplete IgG. This could be due to co-depletion, as IgG is a highly abundant protein in serum, the chance of some being expelled along with the flowthrough is rather high.

In the depleted samples, large amounts of albumin still remained in the AlbuSorb depletion. Some was also present in the Albumin/IgG removal kit depletion, while very little was left in AlbuVoid PLUS, AlbuVoid LC-MS, and BioMag HSA. This indicates that the latter kits were more efficient at removing albumin from serum. As expected, albumin was also present in the BioMag IgG depletion and visible bands of IgG at 150, 50, and 25 kDa were also observed. These bands were also visible in the other depleted samples, regardless of whether the kit

promised to deplete IgG or not. The 150 and 25 kDa band in the depleted samples could also be due to the other immunoglobulins that are of about the same size. Of the IgG depleting kits, BioMag IgG was the most efficient at removing the 50 kDa heavy-chain fragments, then the Albumin/IgG removal kit, and lastly AlbuSorb. Although bands of other proteins can be observed in the depleted sample of all kits, they were the clearest in the Albumin/IgG removal kit and AlbuSorb, and the least in AlbuVoid PLUS.

So far, from the gel results alone, it would seem that the AlbuVoid kits, particularly AlbuVoid LC-MS, had the most successful depletion in terms of the least amount of co-depletion and retained albumin. This could be related to their method of binding the entire proteome and excluding albumin, as the BioMag HSA also had little remnant albumin in the depletion and the AlbuVoid PLUS IgG-flowthrough contained more co-depleted proteins. On the other hand, the Albumin/IgG removal kit seemed to have had the most clear protein bands from other proteins in its depletion, whiles also having little retained albumin. Perhaps this could be related to difficulties in elution of the bead-bound proteome from the former kits, or loss of sample in the various washing steps.

LC-MS Data

LC-MS analysis of the samples, illustrated in Figure 4.2, showed that all depletion kits had lower total protein identifications than the control. As follows, the number of unique proteins identified was much greater in the control than any of the kits, both when compared one-by-one and all together, as seen in Figure 4.4. The kit with the highest protein yield was the Albumin/IgG removal kit, followed by BioMag HSA, both when precipitating with cold acetone. These were also the kits with the most unique protein identifications. This result was in accordance with the results from SDS-PAGE gels, which appeared to have had high amounts of protein bands in comparison to the other kits.

AlbuVoid PLUS and AlbuVoid LC-MS both involved on-bead digestion. As a result, no acetone precipitation between depletion and digestion could be performed. Whether it would be better to compare the results from these kits to the «No Precipitation» or «Acetone Precipitation» results of the others is hard to say. On the one hand, the on-bead digestion was performed after removing the flowthrough of abundant protein, meaning there was no remnant buffer that could disturb the digestion reaction, similar to the samples precipitated with acetone. On the other hand, the process of precipitating the proteins with acetone could have had other effects on the sample that the on-bead digestion cannot account for. As such, these results will be compared to both sets of samples, or rather to the «best case» results from each kit. Doing this places AlbuVoid LC-MS and PLUS as the third and fourth highest in protein identifications, respectively.

Despite higher total MS2 spectral counts, the yield of proteins in BioMag IgG and BioMag HSA+IgG were the two lowest. As mentioned previously, the poor yield of BioMag coupled HSA+IgG could be due to inadequate transmission between the two kits. Rather than simply evaporating the HSA depleted sample, acetone precipitation could have been better suited for substituting the buffers. Although one would assume that remnant elution buffer would cause poor binding of IgG to the bead resin, rather than having the opposite effect. The yield of BioMag IgG depletion alone was also rather low, so the cause could lie in this step of the coupled depletion. Perhaps increasing the bead-to-sample ratio in both cases could provide better results. Or, if the poor yield was due to unspecific binding of proteins to the matrix, then altering the binding/wash buffer could have an effect. The composition of the buffers supplied with the BioMag depletion kits are undisclosed, so it would necessitate developing a «home brew».

As the BioMag HSA depletion kit gave a relatively high yield when precipitated with acetone, but retains a large amount of immunoglobulins, it would be of interest to investigate whether improving the coupled method would result in a high return of less abundant proteins.

Depletion of the intended abundant proteins seemed to have been successful in most of the kits for the spectral data as well. When looking at the MS2 spectral counts in Table 4.1, all kits, save for BioMag IgG and HSA+IgG, had decreased percentage of HSA in comparison to the control. As was indicated by the SDS-PAGE results, both AlbuVoid kits and BioMag HSA were the most successful at depleting albumin. AlbuVoid PLUS and BioMag HSA (no precipitation) also had correspondingly higher percentages of the less abundant proteins, whilst all others had about the same as the control. For depletion of IgG, BioMag IgG and acetone precipitated Albumin/IgG removal kit were the only ones that seemed to have actually depleted this protein. This was also reflected in the BioMag HSA+IgG results. It could be that for the other kits, where IgG depletion happened either before or simultaneously as HSA depletion, the over-abundance of albumin blocked IgG from binding to the bead matrix.

Since these depletion kits focus on the depletion of albumin and/or IgG, one could expect that the other abundant proteins would thus comparatively be enriched. According to the application report for AlbuVoid PLUS (74), this kit binds apolipoproteins and heavily glycosylated proteins poorly whilst being especially suited for enrichment of the complement factors. This was reflected in the high %SC of complement factors and the low %SC of α-globulins when compared to the control. The levels of apolipoprotein MS2 spectral counts, however, were 1% higher. AlboVoid LC-MS, on the other hand, only had lower α-globulin content, and had a greater apolipoprotein content than the control. Enrichment of α-globulins seems to have occurred in the case of the Albumin/IgG removal kit, AlbuSorb, BioMag IgG, and BioMag-HSA+IgG. The enrichment or depletion of these other abundant protein groups was likely due to the specific composition of the bead matrix of each individual kit, but could also be a coincidence.

Like with the gels, it would appear that the AlbuVoid kits and BioMag HSA were better at removing albumin from the sample, whilst more other proteins were retained when using the Albumin/IgG removal kit.

Acetone Precipitation

The use of cold acetone to precipitate the proteins from their solution before protein digestion seemed to have had little effect on the pure serum «control» sample. Both total MS2 spectral counts and protein identifications, as well as the distribution of the protein groups, had only minor differences, and the number of unique proteins in each were also about the same. This was not surprising as the serum solution is very concentrated, so it makes up only a fraction of the total digestion buffer volume. However, looking at the boxplot diagram in Figure 4.2, the plot for acetone precipitated control was more symmetrical than the one without precipitation.

In regard to the depleted samples, the Albumin/IgG removal kit, AlbuSorb, and BioMag HSA all showed significantly improved protein identifications and total MS2 spectral counts when using acetone precipitation. By removing the depletion buffer from the sample prior to digestion, not only will the concentration of digestion reagents be higher, but it also negates the possibility of unfavorable conditions and contamination. This was likely the cause of the improved results when using precipitation with acetone. The protein identifications were about the same for precipitation and no precipitation in the case of BioMag IgG and HSA+IgG. This may be partially due to the fact that the final volume of the BioMag IgG depleted sample was the lowest of the tested kit. Also prevalent is the fact that these two methods had quite poor

yields regardless, and a larger spread of values as seen in the boxplot diagram. This points to the uncertainty of these results in general, as discussed before. As seen in Figure 4.5, the Albumin/IgG removal kit and AlbuSorb also identified much more unique proteins in the acetone precipitated samples in comparison to BioMag HSA+IgG.

The percentage of other, less abundant proteins remained about the same in the Albumin/IgG removal kit and AlbuSorb samples, but was diminished in favor of albumin or other immunoglobulins in the BioMag kits. For the other groups, the BioMag kits also saw a decrease in many of these, excluding albumin and immunoglobulin. In the Albumin/IgG removal kit and AlbuSorb the percentage of complement factors decreased, and other immunoglobulins increased, whilst the other groups remained more or less constant.

It would seem that the overall effect of acetone precipitation was an increase in MS2 spectral counts and identified proteins when applied to the sample resulting from a depletion kit. This increase appears to be uniform for all of the included protein groups and does not favor one group in particular over the others. That being said, an increase in protein identifications overall does therefore also mean an increase in the less abundant proteins identified.

In conclusion, Pierce Albumin/IgG removal kit, having identified the most proteins, had the most favorable result overall. Originally, the most promising depletion kit was to be attempted optimized for better protein yield. This was to be done by, for example, testing the bead-to-sample ratio to grant better protein capture, increasing incubation time, and altering the elution conditions to diminish co-depletion. However, this was not investigated further.

5.2 Fractionation Methods

The various fractionation methods were chosen for exploration on the basis of what was practical feasible, in regard to the available equipment, and what seemed promising or interesting based on previous research. First, the discussion starts with the contents of each fractionation method, then considering the success of the fractionation of the proteins, before comparing the total results between each method.

5.2.1 Protein Fractionation

Each protein fractionation method chosen to audit here separate proteins based on different characteristics: ethanol precipitation by organic solvents interactions, size exclusion chromatography by size, and $ZnCl_2$ precipitation by metal-ion binding. As a result, the methods can be expected to yield different results in terms of how the abundant proteins separate. The sub-goal of this objective was to find out if abundant proteins could be separated into distinct fractions, so that the less abundant proteins retained in the remaining fractions can be more readily identified. This would also be beneficial in the case of identifying specific proteins of interest.

Ethanol Precipitation Fractions

From the Cohn et. al. (*21*) method, the fractions were expected to contain a majority of fibrinogen in Fraction I, γ- and β-globulins in Fraction II+III, α-globulins in Fraction IV-1, remnant α- and β-globulins in Fraction IV-4, and albumin in Fraction V. The two supernatant fractions (V-1 and V-2) would then contain either other, less abundant proteins or very little protein at all. Almost no fibrinogen was detected in any of the fractions, nor the control, which is to be expected as this method was designed for use on plasma, not serum where such clotting

factors have been removed. Unlike fibrinogen, the other aforementioned proteins were present to some degree in all fractions. As seen in Table 4.2 Fraction I contained a high percentage of all protein groups, as well as high protein identifications in total. This was likely due to it being the first precipitation step, in which the high concentration of proteins present in the still pure serum can happen to precipitate simply due to the sudden change in ionic strength that comes with the introduction of the organic solution. α-, β- and γ-immunoglobulins were otherwise observed at a high percentage in fraction IV-4. Only remnants of globulins, not already precipitated in fractions II+III and IV-1, were expected to appear in this fraction. Although still present in the former fractions, it was to a significantly lower degree. Fraction IV-4 contained the highest total protein identifications and, like fraction I, had a high percentage of most protein groups, including apolipoproteins, complement factors, the already mentioned globulins, and other, less abundant proteins. As this fraction was obtained by raising the supernatants ethanol concentration to 40%, this large elevation was likely the cause of the observed high protein numbers. Albumin, however, was largely retained in fraction V-1. This was the first supernatant fraction, where it would be expected to find the remnant proteins not already precipitated, whilst albumin was expected to precipitate in the pellet of fraction V. Albumin not precipitating from the supernatant could be due to inadequate difference in the solutions pH, as the ethanol concentration remained the same as for the previous step. V-2, the last fraction, was more of a washing step to further purify fraction V and should hardly contain any proteins. In this regard, the results match the expectation.

The inconsistencies in the results presented here with what was expected from this method could largely be explained as being due to the range in sample quantity. Cohn's process was designed to be used on liters-worth of plasma to purify albumin in an industrial setting. This means that the precipitation buffers can be more readily introduced to the plasma under constant stirring, as not to cause local excesses that would result in acute precipitation, and the pH could be monitored in solution. Temperatures could also be kept below 0°C to avoid denaturation of the proteins. Here, only microliters of serum was used. This made both monitoring of pH and buffer infusion challenging. As no composition beyond the ethanol concentration and pH were given, the exact molarity and pH of the buffers had to be studied. Furthermore, maintaining a temperature below 0°C during sample handling between centrifugation was not possible with the available equipment. Since protein denaturation was not of great concern, as the proteins were to be denatured later during digestion regardless, this was not considered to be of absolute importance. However, protein denaturation alters the proteins interaction with the solution as well as other proteins. As more hydrophobic regions on the protein are exposed, this could lead to protein aggregation and precipitation prior to what was intended. If further development of this technique for small quantities of serum is intended, establishing a mode for temperature preservation would be necessary. Improvements to the precipitation buffers would also have to be made, especially in regard to the solutions pH. In conclusion, neither the protein fractionation nor the number of identified proteins improved the analysis of the serum proteome using sequential ethanol precipitation.

Size Exclusion Chromatography Fractions

From the UV chromatograms in Figure 4.7 and 4.8 it appears that the size exclusion of the three replicates was successful, in that clearly defined peaks appeared at different time points. Since the number of proteins identified with replicate B and C diverged greatly, the overall reproducibility of this method was difficult to determine. However, since the UV chromatograms for all three replicates were very similar, the discrepancies between replicate B and C are more likely to stem from post SEC occurrences. This is especially likely considering the large number of fractions necessitated the replications be digested and analyzed by LC-MS individually, rather than simultaneously as for the other fractionation methods. The majority of the proteins eluted

before the half-way point of the run, and the peaks here were poorly separated. This could be improved by using a different column with a different pore size. The highest peak intensities indicate that the majority of the highly abundant proteins can be expected in these fractions.

In regard to the differences between replicate B and C beyond simply the merged total number, despite replica B having identified more total proteins, most of the replica C fractions had more identified proteins than the corresponding replicate B fraction. The only exceptions were fraction 1, 2, 6, and 17, 1 and 17 being the most dramatic. Fraction 17 had a high number of protein identifications in both replicates, which was expected as it likely corresponds to the pronounced peak towards the end of the SEC UV chromatograms, but only in B was it the highest with double the value of most of the other fractions. Meanwhile fraction 1 revealed the lowest number of protein identification of replicate C, as was expected from the low peak height in the UV chromatograms, where it was the second highest in replicate B. Looking closer at the unique protein identifications in each fraction, many of the proteins only identified in fraction 1 or 17 appear to be intracellular proteins, as one would expect to find when analyzing a cell lysate. Although intracellular proteins will be present in blood as part of the tissue leakage proteome, the extent here was questionable. Analysis of HeLa cells is routinely performed when the LC-MS system undergoes maintenance or conditioning, as a way to estimate performance. It is possible that fraction 1 of sample B was contaminated with the HeLa proteome. This was probably due to remnants left in the column or needle from a prior injection, not sufficiently removed by the preceding blank injection. By subtracting the number of proteins identified in fraction 1 of sample B from the 20-fraction total, the new total then becomes closer to the total of replica C, which supports this theory. In view of this, the results from replica B will be exempt from further discussion, as replica C is likely to be more accurate.

Besides albumin, which had a molecular weight of 66 kDa, the protein groups contain proteins that vary in size and can therefore be expected to appear in several fractions. Of the proteins included in this grouping, the most abundant proteins of each are listed in Table 1.2, and the molecular weight of these will be the focus. Starting with the apolipoproteins, these were found predominantly among the first fractions, namely 2 and 3. This was likely due to the 512 kDa apolipoprotein B-100, which is the most abundant of the four listed. The α-globulins appear in 3 and 4, then complement factors in 4 and 5. Since amid the most abundant proteins are α-2-macroglobulin (720 kDa) and ceruloplasmin (120 kDa) of α-globulins, and complement factor C3 (185 kDa) and C3 (190 kDa), this was as expected. The immunoglobulins appear mostly in fractions 5 and 6, but also a lot were in 3 and 4. Most of the immunoglobulins have a molecular weight of 150 kDa when intact, which would coincide to their fractionation in relation to the other proteins mentioned. Albumin (66 kDa) and the β-globulins (50 - 90 kDa) were mainly in fraction 7 and 8. Fraction 7 was also the fraction which in the UV chromatogram showed the highest peak intensity, thus it matches the expected result of albumin, the most abundant protein, being in this fraction. Lastly, most of the less abundant proteins have also eluted in the same fractions as the abundant groups, but a relatively high percentage was also seen in fraction 17. As this is the most diverse group in terms of its members, and therefore probably also size, the more even spread among the fractions was to be expected. All of the fractions contained some proteins from all of the groups, with most having a high percentage in two fractions that tail of at either side. This could be due to the poor separation of the peaks seen in the UV chromatograms, which would cause a lot of overlap between the fractions in this region, but was not surprising since the groups encompass proteins of several sizes. This was also probably true for albumin which, due to its abundance, likely had an extended elution from the SEC column. In addition, it may also have co-eluted with other proteins that it can bind to and degradation product of albumin in blood might exist.

From this it would appear that the fractionation of serum proteins by size was mostly successful. However, with the loss of replicate A and the possible contamination of replicate B, further tests would have to be made in order to draw a definitive conclusion on this methods success. Furthermore, as analyzing 20 samples is far too many to make this method high-throughput. Optimizing the fraction number, as was attempted with replicate B, would be necessary. In conjunction with this, columns of different pore-size, or perhaps different modes of size exclusion all together, should also be examined in order to achieve better separation of the most abundant proteins.

In conclusion, it was shown that protein fractionation can improve the number of identified proteins in serum. The highest number of protein identification has been obtained with SEC, but it required the analysis of many fractions.

ZnCl$_2$ Precipitation Fractions

Protein precipitation using Zn^{2+} ions has been shown to be very effective in plasma (*82*) and that it can particularly be used to precipitate immunoglobulins at low molarity (*35*). With ZnCl$_2$ precipitation, three sequences of increasing concentrations were tested. Sequence 2 used 20 mM ZnCl$_2$ to produce two fractions, sequence 3 used 0.2 and 20 mM ZnCl$_2$ to produce three fractions, and sequence 4 used 0.002, 0.2, and 20 mM ZnCl$_2$ to produce four fractions. The 3 and 4 sequence methods were done in order from the lowest to highest ZnCl$_2$ concentration. Here, the focus of the discussion will be comparing these three sequences, while also looking at the individual fractions, as illustrated in Figure 4.10. In all three sequences, the fractions with the most total protein identifications, as well as the most unique protein identifications and the highest percentage of less abundant proteins, was the pellet fraction obtained when using 20 mM of ZnCl$_2$ (P1). This was to be expected, as it was the strongest concentration of ZnCl$_2$ used in this experiment. Following this, the 0.2 mM ZnCl$_2$ pellet (P2) had the most proteins and the highest percentage of the less abundant protein distribution, but the final supernatant had the most unique protein identifications. The latter observation was probably due to higher overlap in the proteins precipitating between P2 and P1. Fraction P3, from 0.002 mM ZnCl$_2$, had the least proteins in all aspects. From this it would appear that the proteins that precipitated by 20 mM ZnCl$_2$ in the 2-sequence method have divided themselves between the pelleted fractions of sequence 3 and 4, with a lot of overlap between pelleted fractions. Also, less of the proteins were retained in the supernatant of these multi-step sequences. This was likely due to the prolonged precipitation process gradually depleting the supernatant of proteins, thus making it so that more proteins were available for precipitation when the final 20 mM ZnCl$_2$ was introduced.

Precipitation of the protein by ZnCl$_2$ was dependent on the interaction between the zinc-ion with histidine and cysteine residues, primarily on the proteins exposed surface. Proteins with higher amounts of these residues are thus expected to appear in the pelleted sample, with the highest appearing in the pellets formed using the lowest ZnCl$_2$ concentration. As with size exclusion, the bulk of each protein group was likely to be comprised mainly of the most abundant proteins of each, the ones mentioned in the introduction section 1.2.2. These may have different amounts of histidine/cysteine and were likely distributed among the different fractions. Considering the pelleted fractions as a whole compared to their respective supernatant fractions, the apolipoproteins, complement factors, and immunoglobulin were mostly present in the pellets of each sequence. In contrast, the α- and β-globulins were more uniformly spread between the pellets and the supernatant, pointing towards greater diversity in the residue composition of these proteins. This is to be expected, as these two groups are, besides the less abundant proteins which had a similar distribution, the most heterogeneous. Albumin, on the other hand, would be expected to have a higher divergence in fraction distribution than what was seen

here. This was likely due to the extreme abundance of this protein that makes it appear in all fractions regardless. The distribution of proteins between pellet fractions in sequence 3 and 4 mostly favors the P1 fraction, which was to be expected from the higher total protein identifications. However, apolipoproteins and immunoglobulins are, in both cases, more present in fraction P2. This could be due to a higher histidine/cysteine content in these proteins, or as a result of them being more abundant than the other groups and, like albumin, thus precipitate sooner.

The overall results from the three sequences were similar. However, the 2-sequence method, in terms of total and unique protein yield, was slightly better than the 3 and 4 sequence method. Also when combining the pelleted fractions of 3 and 4 was this the case. Although sequence 4 had more total proteins in the combined pellet, and sequence 3 had more unique proteins in the supernatant, the differences here were minor. In other words, nothing was gained from first precipitation with 0.002 mM and/or 0.2 mM $ZnCl_2$. As such, the two-sequence method was the best option, because it requires the least LC-MS analyses without reducing the protein identifications significantly. Going forward, different concentrations of $ZnCl_2$ could be tried to examine if even better yields or more defined partition of proteins would be possible. In conclusion, generation of two fractions by $ZnCl_2$ precipitation was an easy and cost-effective approach to increase the number of protein identifications in serum.

Summary Protein Fractionation

Results from the protein fractionation methods tested here were quite varied. Ethanol precipitation yielded very low total protein identifications, on par with the depletion methods discussed before. Considering the minor difference to its control, both in terms of total protein and unique protein identifications, as well as what was discussed previously, it was deemed not worth to further explore this approach.

SEC revealed the most total protein identifications in comparison, and revealed the greatest improvement in comparison to the control. Nevertheless, the large number of fractions of this method requires a lot of LC-MS instrument time.

The total number of protein identifications for $ZnCl_2$ precipitation was significantly better than that of its control. Furthermore, the number of identified proteins was the second highest after SEC, also in comparison to the peptide fractionation methods that will be discussed next. As the number of fractions for this method can be kept at two, the precipitation reagent is very cheap, and no special instrumentation is required, this method could be considered an easy method to improve serum proteome coverage.

5.2.2 Peptide Fractionation

Unlike the protein fractionation methods, which all separated proteins on the basis of different characteristics, both peptide fractionation methods used high-pH reversed-phase C18 to separate peptides by hydrophobicity. The major differences between these two techniques were that either the use of an HPLC instrument is required versus the usage of spin columns, and that the number of fractions can be different.

Since MS2 spectral counts were made on the basis of the measured *m/z* ratios of peptide fragments, the total MS2 spectral counts will be approximately equivalent to the number of peptides in each fraction. Thus, total MS2 spectral counts will be the main focus of this part of the discussion, rather than total protein identifications. When it comes to proteins, the fractions that had the most peptides, i.e., the most MS2 spectral counts, can also be expected to have the

most protein identifications. However, the distribution of proteins, as mentioned in the results section 4.2.4, will not necessarily be affected by the peptide distribution. As these methods separate peptides by hydrophobicity, it can be that some proteins would generate more hydrophobic peptides, and therefore be detected mainly in the latter fractions. This is, for example, the case for membrane proteins. However, as many of the proteins found in serum are globular, which have both external hydrophilic and internal hydrophobic regions, this is unlikely to be observed. Rather, the proteins are likely to be distributed somewhat evenly between the fractions, as a result of the distribution of their different peptide fragments.

Even if a distribution of proteins as for the protein fractionation methods cannot be expected, the higher resolving power of reversed-phase C18 material might lead to identifying a high percentage of less abundant proteins. Alternatively, it may be that the ratio of abundant to less abundant protein-related peptides are more equal, and thus overshadowing as in the control would be avoided.

HpH-RP (PepSwift) Fractions

One advantage of HPLC-based peptide fractionation is that the separation of peptides can be visualized by UV chromatography prior to being analyzed by LC-MS. By comparing the gradient seen in Figure 3.2 with the UV-VIS chromatogram in Figure 4.11, the peptides appear to begin eluting after buffer B had reached 10%. With the LC gradient used here, all peptides were expected to have eluted after 40 minutes. It appears that most peptides were likely to be present in the fractions between 3 - 13. The chromatograms from all three replicates appeared similar in regard to the peak formation, which points to a high degree of reproducibility.

Fractions with over 1,000 MS spectral counts were found near the upper middle, from fraction 4 to 9, which was also where most of the highest peaks in the chromatogram were seen. As expected, few peaks were observed in the first and last two fractions.

Here, the group of other proteins was the highest or second-highest in percentage for all fractions. The percentage of the other groups varies somewhat from fraction to fraction, especially toward the tail ends of the run, as the total protein identified in these were quite low. Towards the middle of the fractions, where the number of MS2 spectral counts and also protein identifications were more even, the percentages of the protein groups also become more even. It is possible that combining fractions 1 - 3 and 10 - 15 would yield similar results. Thus, reducing the number of fractions could be performed without loss of protein identifications. Further trials where this is attempted would have to be performed. Changing the LC gradient and perhaps the fraction size could also be points of interest for optimization. It should be pointed out that the fractions for this method were collected manually, as in every two minutes the drop that had formed at the end of the collection column was captured in a sample tube. This means that there were likely some inaccuracies in sampling time points, though this should only be a matter of seconds. Nevertheless, an automatic collector could improve reproducible sampling. However, the major drawback of this method was certainly that a nanoLC system was used with a monolithic column, and therefore, the loading capacity was limited. Changing to an LC instrument and column with higher loading capacity could assuredly increase the number of protein identifications from serum.

HpH-RP Spin Column (Pierce kit) Fractions

HpH-RP spin column peptide fractionation being resin-based does not require an LC instrument and is therefore cheaper and easier to perform. On the other hand, the selection of fractions can be more easily refined using an LC. Here, as seen in Figure 4.13, total MS2 spectral

counts were the lowest in the first three fractions, while ∼10,000 spectral counts were obtained in the last five fractions. More likely than there being very few hydrophilic peptides, that had not already been lost in the flowthrough, was that the increase in acetonitrile concentration between fraction 1 and 2 was simply not steep enough to yield as many peptides as in the remaining fractions. It is possible that skipping the first elution step with 5% ACN and going straight to 7.5% ACN would be just as effective. The percentage of each protein group was rather similar between the eight fractions and was also very similar to the control. This points to an even distribution of the peptides from these proteins.

A substantial number of peptides were detected in the flowthrough, which was most likely a result of overloading the column. This could be improved by increasing the bead-to-sample ratio. Very few peptides being eluted in the wash indicates that binding to the bead-resin was accomplished. However, unique proteins were identified in the flowthrough. Despite the low protein identifications, it may therefore be worthwhile to collect this fraction. As will be discussed in the summary of peptide fractionation, these likely stem from proteins of short amino acid sequence.

Pairwise concatenation of the MS2 spectral data from the first and last four fractions was performed in order to evaluate whether this would save run time by cutting the number of fractions in half. This resulted in 10 fewer protein identifications and ∼30,000 lower MS2 spectral counts in total. Since the protein identifications were only slightly decreased, it could be worth-wile to try performing this HpH-RP method with half the number of elution buffers.

Summary Peptide Fractionation

Since the number of proteins identified in the respective control for these two peptide fractionation methods were quite similar, the total number of identified proteins by each method are comparable. Here, HpH-RP using the spin column kit identified more proteins than by LC, and the contrast to its control was greater. Typically, one would expect that increasing the number of fractions could yield better results simply from the reduced complexity, but this was not the case here.

The percentage of protein groups in each fraction was even more stable in HpH-RP spin column kit than in the HpH-RP LC method. This was likely due to the higher number of protein identifications in each HpH-RP spin column kit fraction, and the more even distribution of MS2 spectral counts. Whether this was solely due to the reduced number of fractions would have to be investigated further, preferably by reducing the number of fractions in the HpH-RP LC method.

What was somewhat surprising was that some of the fractions, particularly in the spin column-based method, contain unique proteins with more than 2 MS2 spectral counts. These are most likely proteins of low abundance and with short amino acid sequences. Proteins with short sequences will only produce a few peptides, perhaps only one, which would have the same degree of hydrophobicity and thus be expected to elute in the same fraction. If the protein is simultaneously of low abundance, then the chance of the peptides exclusively eluting in one fraction is even greater.

5.2.3   Coupled Protein and Peptide Fractionation

For the final part of the fractionation method study, one protein and one peptide fractionation method were performed in tandem to examine whether their combination would improve their

individual results. For this, 2-sequence ZnCl$_2$ protein precipitation and HpH-RP spin column kit peptide fractionation were chosen. ZnCl$_2$ was chosen over SEC primarily due to the low number of fractions, as well as also having a high yield in protein identifications. HpH-RP spin column kit was chosen both due to the better overall result, but also due to the HPLC system used for the HpH-RP LC method no longer being operational.

As seen in Table 4.19, the distribution of proteins between the collective ZnCl$_2$ pellet and supernatant fraction was very similar to that of 2 sequence ZnCl$_2$ alone. This means that the following peptide fractionation did not have an overall effect on the protein distribution. In fact, with the low number of unique proteins identified, the opposite is more likely true. This was also supported by the higher percentage of proteins identified in the combined supernatant. The percentage of protein groups for the method total was also more similar to ZnCl$_2$ than HpH-RP spin column kit, while protein group percentages in each fraction was more similar to their respective HpH-RP spin column kit fraction, which was to be expected.

In conclusion, no improvement was obtained by coupling these protein and peptide fractionation methods together. Rather, it only resulted in an increased number of samples. Perhaps other combinations would be more effective, especially if optimization steps of each were made in conjunction to their coupling. Here, even the protein fractionation methods could be combined, as they separate proteins based on different characteristics it is feasible that this could improve the overall proteins identified. In any case, the number of fractions for each respective method, besides ZnCl$_2$ which already works well with 2 fractions, would have to be reduced.

5.3 Cysteine-containing Peptide Enrichment

Finally, the complexity of the serum proteome was attempted reduced through enrichment of cysteine-containing peptides. Two magnetic bead types, BcMag thiol and long-arm thiol, Acyl-rac S3 capture, and Thiopropyl Sepharose 6B affinity resin, all with pyridyl disulfide reactive group, underwent preliminary trials using bovine serum albumin. The most promising of these were then applied to serum.

Single Protein Analysis (BSA Test)

Initial tests of the four bead-alternatives using BSA seemed promising, as shown in Table 4.21. All of the kits had cysteine-containing peptides in the enriched sample. The highest percentage was achieved with acyl-rac S3 capture beads using the S6B method, followed closely by the thiopropyl sepharose 6B resin. BcMag thiol-activated and long-arm magnetic beads, however, and S3 capture with the same method, yielded quite low cysteine-containing peptide percentages. In the case of the BcMag beads, this could be due to the low concentration of BSA used in this test contra the volume of coupling buffer. Alternatively, using cysteine for eluting the peptides, rather than DTT as in the S6B method, could also be less effective. It is possible that rather than outcompeting the peptides by binding to the beads, they instead formed disulfide bonds with other cysteine residues. Further experimentation with these beads would have to be performed to exclude these possibilities. The poor yield from the S3 capture beads with the BcMag method was most likely due to the method being unsuitable for this bead-type.

Surprisingly, no cysteine-containing peptides were observed in the BSA control, which had only been digested and nothing else. This may have been due to the simplified protein digestion performed, as described in materials and methods section 3.4.1. With this in mind, it may be that the BcMag magnetic beads could show improved results by simply extending the

incubation time of the protein digestion.

From this, one bead-alternative was chosen to use for enrichment in serum. Thiopropyl sepharose 6B affinity resin was selected, even though S3 capture (using the S6B method) resulted in more MS2 spectral counts and a higher percentage of cysteine-containing peptides. This choice was made considering the amount of material at hand, on the assumption that several tests might have to be performed with serum to optimize the technique.

Complex Proteome Analysis (Serum)

Since BSA and HSA are homologous proteins, one would expect to see similar results in their enrichment of cysteine-containing peptides. However, despite better sequence coverage, the percentage of cysteine-containing peptides, as seen in Table 4.22, remains largely the same, or worse, in the enriched sample. Meanwhile the percentage of cysteine-containing peptides was generally higher in both the flowthrough and the control. Furthermore, there were significantly less total MS2 spectral counts in the enriched sample as well. Lower MS2 spectral counts in the enriched sample was to be expected, as the goal of peptide enrichment was to reduce the complexity of the sample and thus get fewer peptides, with hopefully lower dynamic range. However, this did not lead to higher protein identifications, as Table 4.23 illustrates. Additionally, the total MS2 spectral counts and protein identifications when combined with the flowthrough were also much lower than the control. It could therefore be that most of the cysteine residues on the peptides have not bound with the reactive groups of the beads, or that the peptides were otherwise lost during the procedure. This could be due to several possible poor reactive conditions.

First, the bead-to-sample ratio was increased to see if the cause was an overload of sample. This increased the percentage of cysteine-containing peptides in both the flowthrough and the enrichment. The number of proteins identified increased significantly in the flowthrough, but became even lower in the enrichment of all bead amounts.

Then an extra purification step was applied to remove DTT from the sample prior to applying it to the beads. As DTT is a reducing agent that was first added to break possible disulfide-bond formations between peptides, it was speculated that this could have caused the reaction with the bead-matrix to fail. But here the percentage of cysteine-containing peptides was only better than the initial trial for the flowthrough sample. The number of proteins identified in the enrichment remained low as well.

After this, additional elution steps were added to see if the opposite was the cause, that the cysteine-containing peptides were not releasing. This produced more total MS2 spectral counts and protein identifications than the prior enriched sample, but adding these together with the enrichment and flowthrough did not yield any significant improvement to the total number of identified proteins. The percentage of cysteine-containing peptides also remained the same as in the enrichment.

Lastly, samples from the washing steps, performed after collecting the flowthrough and before releasing the enriched sample, were collected to investigate if peptides were being lost here. For several of the washing steps, the percentage of cysteine-containing peptides was higher than the preceding flowthrough and following enrichment. There were also more proteins identified and MS2 spectral counts than in the enrichment for all washing steps. However, like with the elution trial, no significant increase in the identified proteins was gained.

From these results, it is difficult to conclude why this method was not yielding better results. It could be that the entire protocol needs to be reworked. Some sample loss is expected to occur in every step of a procedure, but one would then expect to regain this loss when analyzing the samples from these steps. The most likely candidate, as it consistently yielded the highest protein identifications, is that much of the sample was lost in the flowthrough. If enough of the peptides were eluted here, this would mean the complexity of the sample would be more or less unchanged, and thus the total proteins identified would also remain low. The thiol group of cysteine is very reactive, so a possibility is that, as the reactivity of the thiol group is pH-dependent, unsuitable pH in the buffers might have led to poor reactivity. These avenues could be followed-up if experiments with this bead type were to be repeated. Otherwise, it may be that these beads were unsuitable to use with a proteome as complex as serum, and that other methods, like the ones mentioned in the introduction section 1.3.3, might be more suitable.

6   Conclusion and Future Prospects

Serum and plasma MS-based proteome analysis is probably the most challenging task in proteomics. It is of huge interest because of the many existing samples from e.g., biobanks, but the dynamic range of protein abundance with more than 10 orders of magnitude makes it very difficult to perform with a reasonable coverage of the human proteome. Despite huge efforts in the last two decades, progress had been very limited and was mostly achieved by improved LC-MS instrumentation. Currently, non-mass spectrometry-based proteomics techniques (e.g., the proximity extension essay (OLINK) (*79*, *110*) and the aptamer-based proteomics assay (SomaScan) (*17*)) can reveal a much higher coverage of the serum and plasma proteome with more than a few thousand proteins. However, MS-based proteomics is established all over the world, more versatile than the OLINK and SomaScan technology, and can be used to obtain much more detailed information about the proteins, such as post-translational modification. Therefore, it is highly desirable to develop competitive approaches for MS-based plasma and serum proteomics.

By use of depletion kits, the first objective was to decrease the amount of abundant proteins, specifically albumin and immunoglobulins, from the serum sample in order to better detect the less abundant proteins. The results from this part concluded that most proteins were identified without depletion, also when accounting for other abundant proteins, and that removal of albumin and immunoglobulins was largely unsatisfactory using the depletion kits. A slight improvement was made when precipitating with cold acetone. Of the depletion kits examined in this project, Albumin/IgG removal kit, when precipitating with cold acetone, showed the most overall potential. Although less effective at depleting albumin/IgG than some of the other kits, the ultimate goal of depletion of abundant proteins is to increase the chance of detecting other, less abundant proteins. However, as the number of proteins identified was still less than that of undepleted serum, optimization of this kit would be required. Investigating different bead-to-sample ratios, incubation time, and elution conditions of this method, could provide better results.

Fractionation methods were explored as the second objective to achieve the overall goal of improving protein detection by separating the more and less abundant proteins into different fractions. Out of the methods tested here, SEC and $ZnCl_2$ precipitation showed most potential for fractionation of proteins and the HpH-RP spin column kit showed most potential for fractionation of peptides. The results for size exclusion chromatography revealed the highest

number of protein identifications. The greatest disadvantage of these methods is that they produce several fractions, which required increased LC-MS analysis time. In this regard, $ZnCl_2$ precipitation with two fractions seemed to be an attractive approach.

Finally, the third objective was to reduce the complexity of the sample by cysteine-containing peptide enrichment to diminish the peptide count from more abundant proteins and increase the chance of detecting the less abundant ones. The promising results with a single protein (BSA) showed the potential of this method. However, the adaptation of the approach to an as complex proteome as serum was not successful. This will require more experiments for in-depth evaluation of each step or investigations of alternative methodologies such as using iodoTMT reagents.

Three objectives were pursued here. Firstly, depletion of the abundant proteins comprehending serum albumin and immunoglobulins did not improve proteome coverage of serum. Secondly, fractionation approaches for proteins with SEC and $ZnCl_2$ precipitation and for peptides with HpH-RP spin column showed promising results with increased protein identification. Thirdly, the enrichment of cysteine-containing peptides to reduce the number of analytes worked well with a single protein (BSA), but could not be successfully adapted to serum.

References

(*1*)   Aass, C., Norheim, I., Eriksen, E. F., Thorsby, P. M. and Pepaj, M. (2015). Single unit filter-aided method for fast proteomic analysis of tear fluid. *Anal. Biochem. 480*, 1–5.

(*2*)   Adachi, J., Kumar, C., Zhang, Y., Olsen, J. V. and Mann, M. (2006). The human urinary proteome contains more than 1500 proteins, including a large proportion of membrane proteins. *Genome Biol. 7*.

(*3*)   Aitekenov, S., Gaipov, A. and Bukasov, R. (2021). Review: Detection and quantification of proteins in human urine. *Talanta. 223*, 121718.

(*4*)   Anderson, N. L. (2010). The Clinical Plasma Proteome: A Survey of Clinical Assays for Proteins in Plasma and Serum. *Clin. Chem. 56*, 177–185.

(*5*)   Anderson, N. L., Ptolemy, A. S. and Rifai, N. (2013). The Riddle of Protein Diagnostics: Future Bleak or Bright? *Clin Chem. 59*, 194–197.

(*6*)   Anderson, N. L. and Anderson, N. G. (2002). The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *MCP. 1*, 845–867.

(*7*)   Bairoch, A. and Boeckmann, B. (1992). The SWISS-PROT protein sequence data bank. *Nucleic Acids Res. 20*, 2019–2022.

(*8*)   Bekker-Jensen, D. B., Kelstrup, C. D., Batth, T. S., Larsen, S. C., Haldrup, C., Bramsen, J. B., Sørensen, K. D., Høyer, S., Ørntoft, T. F., Andersen, C. L., Nielsen, M. L. and Olsen, J. V. (2017). An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst. 4*, 587–599.e4.

(*9*)   Bendayan, M. and Garzon, S. (1988). Protein G-gold complex: comparative evaluation with protein A-gold for high-resolution immunocytochemistry. *JHC. 36*, 597–607.

(*10*) BEST (Biomarkers, EndpointS, and other Tools) Resource FDA-NIH Biomarker Working Group, https://www.ncbi.nlm.nih.gov/books/NBK326791/ (accessed 12th Apr. 2023).

(*11*) Bogaert, A. and Gevaert, K. (2020). Protein amino-termini and how to identify them. *Expert Rev. Proteom. 17*, 581–594.

(*12*) Bollineni, R. C., Guldvik, I. J., Grönberg, H., Wiklund, F., Mills, I. G. and Thiede, B. (2015). A differential protein solubility approach for the depletion of highly abundant proteins in plasma using ammonium sulfate. *Analyst. 140*, 8109–8117.

(*13*) Boone, C. and Adamec, J. In *Proteomic Profiling and Analytical Chemistry (Second Edition)*, Ciborowski, P. and Silberring, J., Eds., Second Edition; Elsevier: Boston, 2016, pp 175–191.

(*14*) Brusotti, G., Calleri, E., Colombo, R., Massolini, G., Rinaldi, F. and Temporini, C. (2018). Advances on Size Exclusion Chromatography and Applications on the Analysis of Protein Biopharmaceuticals and Protein Aggregates: A Mini Review. *Chromatographia. 81*, 3–23.

(*15*) Burkhart, J. M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A. and Zahedi, R. P. (2012). Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics. 75*, 1454–1462.

(*16*) Buszewski, B. and Noga, S. (2012). Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *ABC. 402*, 231–247.

(*17*) Candia, J., Daya, G. N., Tanaka, T., Ferrucci, L. and Walker, K. A. (2022). Assessment of variability in the plasma 7k SomaScan proteomics assay. *Sci. Rep. 12*, 17147.

(*18*) Cao, Z., Tang, H.-Y., Wang, H., Liu, Q. and Speicher, D. W. (2012). Systematic Comparison of Fractionation Methods for In-depth Analysis of Plasma Proteomes. *J. Proteome Res. 11*, 3090–3100.

(*19*) Chait, B. T. (2006). Mass Spectrometry: Bottom-Up or Top-Down? *Science. 314*, 65–66.

(*20*) Chen, Y.-Y., Lin, S.-Y., Yeh, Y.-Y., Hsiao, H.-H., Wu, C.-Y., Chen, S.-T. and Wang, A. H.-J. (2005). A modified protein precipitation procedure for efficient removal of albumin from serum. *Electrophoresis. 26*, 2117–2127.

(*21*) Cohn, E. J., Strong, L. E., Hughes, W. L., Mulford, D. J., Ashworth, J. N., Melin, M. and Taylor, H. L. (1946). Preparation and Properties of Serum and Plasma Proteins. IV. A System for the Separation into Fractions of the Protein and Lipoprotein Components of Biological Tissues and Fluids[1a,b,c,d]. *J. Am. Chem. Soc. 68*, 459–475.

(*22*) Consortium, T. U. (2022). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res. 51*, D523–D531.

(*23*) Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V. and Mann, M. (2011401). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res. 10*, 1794–1805.

(*24*) Craig, R. and Beavis, R. C. (2003). A method for reducing the time required to match protein sequences with tandem mass spectra. *RCM. 17*, 2310–2316.

(*25*) Creighton, T. E., *The Biophysical Chemistry of Nucleic Acids & Proteins*; Helvetian Press: 2010, pp 231–258, 480, 535.

(*26*) Curry, S., Mandelkow, H., Brick, P. and Franks, N. (1998). Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nat. Struct. Biol. 5*, 827–835.

(*27*) Dai, J., Wang, J., Zhang, Y., Lu, Z., Yang, B., Li, X., Cai, Y. and Qian, X. (2005). Enrichment and Identification of Cysteine-Containing Peptides from Tryptic Digests of Performic Oxidized Proteins by Strong Cation Exchange LC and MALDI-TOF/TOF MS. *Anal. Chem. 77*, 7594–7604.

(*28*) Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D. F., Burkhard, P. R. and Sanchez, J.-C. (2008). Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal. Chem. 80*, 2921–2931.

(*29*) Derrick, J. P. and Wigley, D. B. (1994). The Third IgG-Binding Domain from Streptococcal Protein G: An Analysis by X-ray Crystallography of the Structure Alone and in a Complex with Fab. *JMB. 243*, 906–918.

(*30*) Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S. N. and Aebersold, R. (2006). The PeptideAtlas project. *Nucleic Acids Res. 34*, D655–D658.

(*31*) Deutsch, E. W., Omenn, G. S., Sun, Z., Maes, M., Pernemalm, M., Palaniappan, K. K., Letunica, N., Vandenbrouck, Y., Brun, V., Tao, S.-c., Yu, X., Geyer, P. E., Ignjatovic, V., Moritz, R. L. and Schwenk, J. M. (2021). Advances and Utility of the Human Plasma Proteome. *J. Proteome Res. 20*, 5241–5263.

(*32*) Deutsch, E. W., Sun, Z., Campbell, D., Kusebauch, U., Chu, C. S., Mendoza, L., Shteynberg, D., Omenn, G. S. and Moritz, R. L. (2015). State of the Human Proteome in 2014/2015 As Viewed through PeptideAtlas: Enhancing Accuracy and Coverage through the AtlasProphet. *J. Proteome Res. 14*, 3461–3473.

(*33*) Di Girolamo, F. and Righetti, P. G. (2011). Plasma proteomics for biomarker discovery: A study in blue. *Electrophoresis. 32*, 3638–3644.

(*34*) Dowell, J. A., Frost, D. C., Zhang, J. and Li, L. (2008). Comparison of two-dimensional fractionation techniques for shotgun proteomics. *Anal. Chem. 80*, 6715–6723.

(*35*) Dutra, G., Komuczki, D., Jungbauer, A. and Satzer, P. (2020). Continuous capture of recombinant antibodies by ZnCl2 precipitation without polyethylene glycol. *Eng. Life Sci. 20*, 265–274.

(*36*) Eckhard, U., Marino, G., Butler, G. S. and Overall, C. M. (20163). Positional proteomics in the era of the human proteome project on the doorstep of precision medicine. *Biochimie. 122*, 110–118.

(*37*) Fekete, S., Beck, A., Veuthey, J.-L. and Guillarme, D. (2015). Ion-exchange chromatography for the characterization of biopharmaceuticals. *J. Pharm. Biomed. Anal. 113*, 43–55.

(*38*) Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science. 246*, 64–71.

(*39*) Foettinger, A., Leitner, A. and Lindner, W. (20079). Selective enrichment of tryptophan-containing peptides from protein digests employing a reversible derivatization with malondialdehyde and solid-phase capture on hydrazide beads. *J. Proteome Res. 6*, 3827–3834.

(*40*) Friedman, D. B., Hoving, S. and Westermeier, R. In *Guide to Protein Purification, 2nd Edition*, Burgess, R. R. and Deutscher, M. P., Eds.; Methods in Enzymology, Vol. 463; Academic Press: 2009, pp 515–540.

(*41*) Gaun, A., Lewis Hardell, K. N., Olsson, N., O'Brien, J. J., Gollapudi, S., Smith, M., McAlister, G., Huguet, R., Keyser, R., Buffenstein, R. and McAllister, F. E. (2021205). Automated 16-Plex Plasma Proteomics with Real-Time Search and Ion Mobility Mass Spectrometry Enables Large-Scale Profiling in Naked Mole-Rats and Mice. *J. Proteome Res. 20*, 1280–1295.

(*42*) Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R. and Vandekerckhove, J. (20035). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol. 21*, 566–569.

(*43*) Gilar, M., Olivova, P., Daly, A. E. and Gebler, J. C. (2005). Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci. 28*, 1694–1703.

(*44*) Granger, J., Siddiqui, J., Copeland, S. and Remick, D. (2005). Albumin depletion of human plasma also removes low abundance proteins including the cytokines. *Proteomics. 5*, 4713–4718.

(*45*) Guerrier, L., Righetti, P. G. and Boschetti, E. (2008). Reduction of dynamic protein concentration range of biological extracts for the discovery of low-abundance proteins by means of hexapeptide ligand library. *Nat. Protoc. 3*, 883–890.

(*46*) Gustavsson, P.-E. and Son, P.-O., *Monolithic Polysaccharide Materials*; Monolithic Materials - Preparation, Properties and Applications, Vol. 67; Elsevier: Boston, 2003, pp 121–141.

(*47*) Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol. 17*, 994–999.

(*48*) Hahm, H. S., Toroitich, E. K., Borne, A. L., Brulet, J. W., Libby, A. H., Yuan, K., Ware, T. B., McCloud, R. L., Ciancone, A. M. and Hsu, K.-L. (20202). Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. *Nat. Chem. Biol. 16*, 150–159.

(49)   Hall, J. E. and Hall, M. E., *Guyton and Hall textbook of medical physiology*, 14th edition., ISBN:9780323597128; Elsevier: Philadelphia, 2021.

(50)   Han, X., Aslanian, A. and Yates, J. R. (2008). Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol. 12*, 483–490.

(51)   Hansen, K. C., Schmitt-Ulms, G., Chalkley, R. J., Hirsch, J., Baldwin, M. A. and Burlingame, A. (2003). Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable 13C-Isotope-coded Affinity Tag and Multidimensional Chromatography. *MCP. 2*, 299–314.

(52)   Hoch, D. G., Abegg, D. and Adibekian, A. (2018501). Cysteine-reactive probes and their use in chemical proteomics. *ChemComm. 54*, 4501–4512.

(53)   Hortin, G. L., Sviridov, D. and Anderson, N. L. (2008). High-Abundance Polypeptides of the Human Plasma Proteome Comprising the Top 4 Logs of Polypeptide Abundance. *Clin. Chem. 54*, 1608–1616.

(54)   Hu, S., Loo, J. A. and Wong, D. T. (2006). Human body fluid proteome analysis. *Proteomics. 6*, 6326–6353.

(55)   Ignjatovic, V., Geyer, P. E., Palaniappan, K. K., Chaaban, J. E., Omenn, G. S., Baker, M. S., Deutsch, E. W. and Schwenk, J. M. (2019). Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J. Proteome Res. 18*, 4085–4097.

(56)   Iyer, H. V. and Przybycien, T. M. (1996). A Model for Metal Affinity Protein Precipitation. *J. Colloid Interface Sci. 177*, 391–400.

(57)   Jaros, J. A., Guest, P. C., Ramoune, H., Rothermundt, M., Leweke, F. M., Martins-de-Souza, D. and Bahn, S. (2012). Clinical use of phosphorylated proteins in blood serum analysed by immobilised metal ion affinity chromatography and mass spectrometry. *J. Proteomics. 76*, 36–42.

(58)   Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T. and Watson, J. T. (1987). Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem. 59*, 2621–2625.

(59)   Ju, S., Kwon, Y., Kim, J.-M., Park, D., Lee, S., Lee, J.-W., Hwang, C.-S. and Lee, C. (2020505). iNrich, Rapid and Robust Method to Enrich N-Terminal Proteome in a Highly Multiplexed Platform. *Anal. Chem. 92*, 6462–6469.

(60)   Karas, M., Bachmann, D. and Hillenkamp, F. (1985). Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem. 57*, 2935–2939.

(61)   Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L. J., Kizhakkedathu, J. N. and Overall, C. M. (20103). Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol. 28*, 281–288.

(62)   Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. and Nesvizhskii, A. I. (20175). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods. 14*, 513–520.

(63)   Lan, J., Núñez Galindo, A., Doecke, J., Fowler, C., Martins, R. N., Rainey-Smith, S. R., Cominetti, O. and Dayon, L. (2018). Systematic Evaluation of the Use of Human Plasma and Serum for Mass-Spectrometry-Based Shotgun Proteomics. *J. Proteome Res. 17*, 1426–1435.

(64)   Lee, S., Ju, S., Kim, S. J., Choi, J.-O., Kim, K., Kim, D., Jeon, E.-S. and Lee, C. (2021). tipNrich: A Tip-Based N-Terminal Proteome Enrichment Method. *Anal. Chem. 93*, 14088–14098.

(65) Leeman, M., Choi, J., Hansson, S., Storm, M. U. and Nilsson, L. (2018). Proteins and antibodies in serum, plasma, and whole blood—size characterization using asymmetrical flow field-flow fractionation (AF4). *ABC. 410*, 4867–4873.

(66) Legrain, P. et al. (2011). The human proteome project: current state and future direction. *MCP. 10*, M111.009993.

(67) Li, J., Cai, Z., Bomgarden, R. D., Pike, I., Kuhn, K., Rogers, J. C., Roberts, T. M., Gygi, S. P. and Paulo, J. A. (2021507). TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing. *J. Proteome Res. 20*, 2964–2972.

(68) Liang-Schenkelberg, J., Fieg, G. and Waluga, T. (2017). Molecular Insight into Affinity Interaction between Cibacron Blue and Proteins. *Ind. Eng. Chem. Res. 56*, 9691–9697.

(69) Liu, X., Rossio, V., Gygi, S. P. and Paulo, J. A. (2023407). Enriching Cysteine-Containing Peptides Using a Sulfhydryl-Reactive Alkylating Reagent with a Phosphonic Acid Group and Immobilized Metal Affinity Chromatography. *J. Proteome Res. 22*, 1270–1279.

(70) Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A. and Lajoie, G. (2003). PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *RCM. 17*, 2337–2342.

(71) Mielczarek, P. and Silberring, J. In *Proteomic Profiling and Analytical Chemistry (Second Edition)*, Second Edition; Elsevier: Boston, 2016, pp 63–99.

(72) Mikesh, L. M., Ueberheide, B., Chi, A., Coon, J. J., Syka, J. E., Shabanowitz, J. and Hunt, D. F. (2006). The utility of ETD mass spectrometry in proteomic analysis. *BBA. 1764*, 1811–1822.

(73) Mischak, H. et al. (2010). Recommendations for Biomarker Identification and Qualification in Clinical Proteomics. *Sci. Transl. Med. 2*, 46ps42.

(74) mkuruc@biotechsupportgroup.com Evaluating Different Windows of Observation Solves The Many Challenges of Serum Proteomics, AlbuVoid™ PLUS & AlbuSorb™ PLUS, tech. rep., 2019.

(75) Mohamed, R., Campbell, J.-L., Cooper-White, J., Dimeski, G. and Punyadeera, C. (2012). The impact of saliva collection and processing methods on CRP, IgE, and Myoglobin immunoassays. *CTM. 1*.

(76) Novák, P. and Havlíček, V. In *Proteomic Profiling and Analytical Chemistry (Second Edition)*, Ciborowski, P. and Silberring, J., Eds., Second Edition; Elsevier: Boston, 2016, pp 51–62.

(77) Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S. and Mann, M. (2007). Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods. 4*, 709–712.

(78) Omenn, G. S. et al. (2021). Progress Identifying and Analyzing the Human Proteome: 2021 Metrics from the HUPO Human Proteome Project. *J. Proteome Res. 20*, 5227–5240.

(79) Palstrøm, N. B., Matthiesen, R., Rasmussen, L. M. and Beck, H. C. (2022). Recent Developments in Clinical Plasma Proteomics—Applied to Cardiovascular Research. *Biomedicines. 10*, 162.

(80) Pan, X., Kirsch, Z. J. and Vachet, R. W. (2022118). Distinguishing Histidine Tautomers in Proteins Using Covalent Labeling-Mass Spectrometry. *Anal. Chem. 94*, 1003–1010.

(81) Perkins, D. N., Pappin, D. J. C., Creasy, D. M. and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis. 20*, 3551–3567.

(82) Polson, C., Sarkar, P., Incledon, B., Raguvaran, V. and Grant, R. (2003). Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B. 785*, 263–275.

(83) Prudova, A., Gocheva, V., Auf dem Keller, U., Eckhard, U., Olson, O. C., Akkari, L., Butler, G. S., Fortelny, N., Lange, P. F., Mark, J. C., Joyce, J. A. and Overall, C. M. (2016809). TAILS N-Terminomics and Proteomics Show Protein Degradation Dominates over Proteolytic Processing by Cathepsins in Pancreatic Tumors. *Cell Rep. 16*, 1762–1773.

(84) Qin, H., Wang, F., Zhang, Y., Hu, Z., Song, C., Wu, R., Ye, M. and Zou, H. (2012625). Isobaric cross-sequence labeling of peptides by using site-selective N-terminus dimethylation. *ChemComm. 48*, 6265–6267.

(85) Qu, Z., Meng, F., Bomgarden, R. D., Viner, R. I., Li, J., Rogers, J. C., Cheng, J., Greenlief, C. M., Cui, J., Lubahn, D. B., Sun, G. Y. and Gu, Z. (2014703). Proteomic quantification and site-mapping of S-nitrosylated proteins using isobaric iodoTMT reagents. *J. Proteome Res. 13*, 3200–3211.

(86) Raoufinia, R., Mota, A., Keyhanvar, N., Safari, F., Shamekhi, S. and Abdolalizadeh, J. (2016). Overview of Albumin and Its Purification Methods. *Adv. Pharm. Bull 6*, 495–507.

(87) Saha, K., Bender, F. and Gizeli, E. (2003). Comparative Study of IgG Binding to Proteins G and A: Nonequilibrium Kinetic and Binding Constant Determination with the Acoustic Waveguide Device. *Anal. Chem. 75*, 835–842.

(88) Sayers, E. W. et al. (2021). Database resources of the national center for biotechnology information. *Nucleic Acids Res. 50*, D20–D26.

(89) Schaller, J., Gerber, S., Kämpfer, U., Lejon, S. and Trachsel, C., *Human blood plasma proteins structure and function*; John Wiley & Sons: Chichester, West Sussex, England ; 2008.

(90) Schilling, O., Barre, O., Huesgen, P. F. and Overall, C. M. (2010). Proteome-wide analysis of protein carboxy termini: C terminomics. *Nat. Methods. 7*, 508–U33.

(91) Searle, B. C. (2010). Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics. 10*, 1265–1269.

(92) Segal, M. B. (1993). Extracellular and cerebrospinal fluids. *JIMD. 16*, 617–638.

(93) Sheehan, D., *Physical biochemistry : principles and applications*, 2nd ed.; Wiley Blackwell: Chichester, 2009; Chapter 4: Mass Spectrometry & 7: Hydrodynamic Methods.

(94) Shen, S., Kai, B., Ruan, J., Torin Huzil, J., Carpenter, E. and Tuszynski, J. A. (2006). Probabilistic analysis of the frequencies of amino acid pairs within characterized protein sequences. *Physica A. 370*, 651–662.

(95) Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc. 1*, 2856–2860.

(96) Sieberz, J., Cinar, E., Wohlgemuth, K. and Schembecker, G. (2017). Clarification of a monoclonal antibody with cationic polyelectrolytes: Analysis of influencing parameters. *Biochem. Eng. J. 122*, 60–70.

(97) Sivadasan, P., Gupta, M. K., Sathe, G. J., Balakrishnan, L., Palit, P., Gowda, H., Suresh, A., Kuriakose, M. A. and Sirdeshmukh, R. (2015). Human salivary proteome — a resource of potential biomarkers for oral cancer. *J. Proteomics. 127*, 89–95.

(98) Sjöbring, U., Björck, L. and Kastern, W. (1989). Protein G genes: structure and distribution of IgG-binding and albumin-binding domains. *Mol. Microbiol. 3*, 319–327.

(99) Sokoll, L. J., Sanda, M. G., Feng, Z., Kagan, J., Mizrahi, I. A., Broyles, D. L., Partin, A. W., Srivastava, S., Thompson, I. M., Wei, J. T., Zhang, Z. and Chan, D. W. (2010). Early Detection Research Network Study of [-2]proPSA: Improving Prostate Cancer Detection and Correlating with Cancer Aggressiveness. *CEBP. 19*, 1193–1200.

(100) Such-Sanmartín, G., Ventura-Espejo, E. and Jensen, O. N. (2014). Depletion of Abundant Plasma Proteins by Poly(N-isopropylacrylamide-acrylic acid) Hydrogel Particles. *Anal. Chem. 86*, 1543–1550.

(101) Switzar, L., Giera, M. and Niessen, W. M. A. (2013). Protein Digestion: An Overview of the Available Techniques and Recent Developments. *J. Proteome Res. 12*, 1067–1077.

(102) Tadic, M., Kralj, S., Jagodic, M., Hanzel, D. and Makovec, D. (2014). Magnetic properties of novel superparamagnetic iron oxide nanoclusters and their peculiarity under annealing treatment. *Appl. Surf. Sci. 322*, 255–264.

(103) Thomas, S. L., Thacker, J. B., Schug, K. A. and Maráková, K. (2021). Sample preparation and fractionation techniques for intact proteins for mass spectrometric analysis. *J. Sep. Sci. 44*, 211–246.

(104) Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T. and Hamon, C. (2003). Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem. 75*, 1895–1904.

(105) Trinkle-Mulcahy, L., Boulon, S., Lam, Y. W., Urcia, R., Boisvert, F.-M., Vandermoere, F., Morrice, N. A., Swift, S., Rothbauer, U., Leonhardt, H. and Lamond, A. (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol. 183*, 223–239.

(106) Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M. and Cox, J. (20169). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods. 13*, 731–740.

(107) Vas, G., Nagy, K. and Vékey, K., *Medical applications of mass spectrometry*, 1st ed.; Elsevier: Amsterdam ; 2008, pp 37–59, 61–89, 93–137.

(108) Venter, J. C. et al. (2001). The Sequence of the Human Genome. *Science. 291*, 1304–1351.

(109) Vidarsson, G., Dekkers, G. and Rispens, T. (2014). IgG subclasses and allotypes: from structure to effector functions. *Front. Immunol. 5*, 195–209.

(110) Wik, L., Nordberg, N., Broberg, J., Björkesten, J., Assarsson, E., Henriksson, S., Grundberg, I., Pettersson, E., Westerberg, C., Liljeroth, E., Falck, A. and Lundberg, M. (2021). Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *MCP. 20*, 100168.

(111) Wildes, D. and Wells, J. A. (2010309). Sampling the N-terminal proteome of human blood. *PNAS. 107*, 4561–4566.

(112) Wilhelm, M. et al. (2014). Mass-spectrometry-based draft of the human proteome. *Nature. 509*.

(113) Wilkins, M. R., Sanchez, J.-C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F. and Williams, K. L. (1996). Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It. *BGER. 13*, 19–50.

(114) Wu, C., Duan, J., Liu, T., Smith, R. D. and Qian, W.-J. (2016). Contributions of immunoaffinity chromatography to deep proteome profiling of human biofluids. *J. Chromatogr. B. 1021*, 57–68.

(115) Wu, C., Duan, J., Liu, T., Smith, R. D. and Qian, W.-J. (2016). Contributions of immunoaffinity chromatography to deep proteome profiling of human biofluids. *J. Chromatogr. B. 1021*, 57–68.

(116) Yates, J. R., Ruse, C. I. and Nakorchevsky, A. (2009). Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng. 11*, 49–79.

(*117*) Yeung, D., Mizero, B., Gussakovsky, D., Klaassen, N., Lao, Y., Spicer, V. and Krokhin, O. V. (2020). Separation Orthogonality in Liquid Chromatography–Mass Spectrometry for Proteomic Applications: Comparison of 16 Different Two-Dimensional Combinations. *Anal. Chem. 92*, 3904–3912.

(*118*) Yoshihara, T., Zaitsu, M., Ito, K., Hanada, R., Chung, E., Yazawa, R., Sakata, Y., Furusho, K., Tsukikawa, H., Chiyoda, T., Matsuki, S. and Irie, S. (2021). Cerebrospinal Fluid Protein Concentration in Healthy Older Japanese Volunteers. *IJERPH. 18*.

(*119*) Zelanis, A., Oliveira, A. K., Prudova, A., Huesgen, P. F., Tashima, A. K., Kizhakkedathu, J., Overall, C. M. and Serrano, S. M. T. (2019906). Deep Profiling of the Cleavage Specificity and Human Substrates of Snake Venom Metalloprotease HF3 by Proteomic Identification of Cleavage Site Specificity (PICS) Using Proteome Derived Peptide Libraries and Terminal Amine Isotopic Labeling of Substrates (TAILS) N-Terminomics. *J. Proteome Res. 18*, 3419–3428.

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

(*121*) Zhang, Y., Guo, Z., Zou, L., Yang, Y., Zhang, L., Ji, N., Shao, C., Sun, W. and Wang, Y. (2015). A comprehensive map and functional annotation of the normal human cerebrospinal fluid proteome. *J. Proteomics. 119*, 90–99.

(*122*) Zhao, M., Li, M., Yang, Y., Guo, Z., Sun, Y., Shao, C., Li, M., Sun, W. and Gao, Y. (2017). A comprehensive analysis and annotation of human normal urinary proteome. *Sci. Rep. 7*.

(*123*) Zhao, Y., Xue, Q., Wang, M., Meng, B., Jiang, Y., Zhai, R., Zhang, Y., Dai, X. and Fang, X. (2023). Evolution of Mass Spectrometry Instruments and Techniques for Blood Proteomics. *J. Proteome Res. 22*, 1009–1023.

(*124*) Zhou, L. and Beuerman, R. W. (2017). The power of tears: how tear proteomics research could revolutionize the clinic. *Expert Rev. Proteom. 14*, 189–191.

(*125*) Zolotarjova, N., Martosella, J., Nicol, G., Bailey, J., Boyes, B. E. and Barrett, W. C. (2005). Differences among techniques for high-abundant protein depletion. *Proteomics. 5*, 3304–3313.