Natural Products and Medicinal Chemistry Research Group, Department of Pharmacy
Faculty of Health Sciences, UiT – The Arctic University of Norway

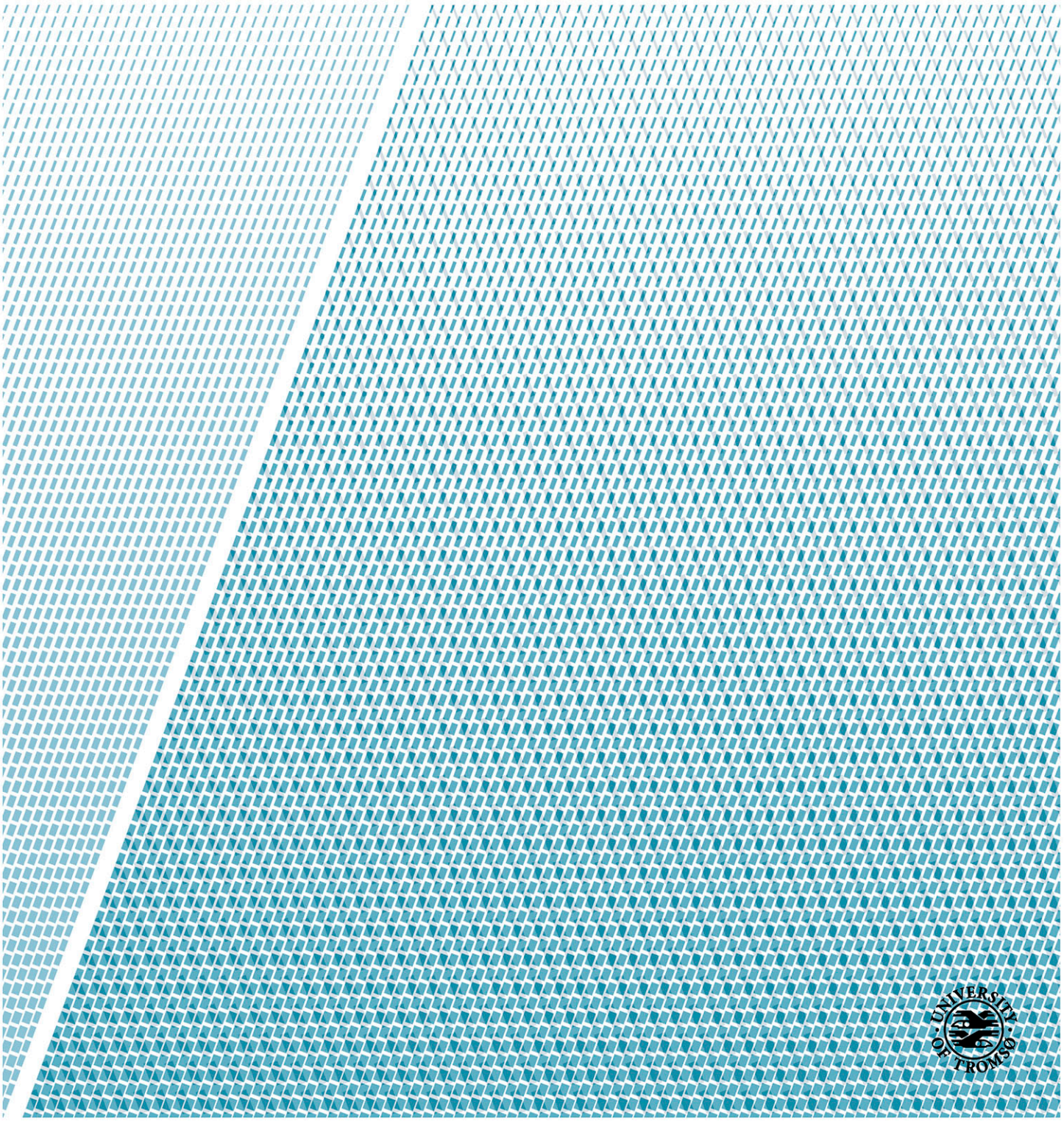
Identification of N-terminal and C-terminal peptides in proteomics

—

Christoffer Ågnes Sivertsen

Supervisor: Associate Professor Terkel Hansen
Co-supervisors: Postdoc. Yvonne Pasing,
PhD candidate Armin Schniers

Thesis for the degree Master of Pharmacy, May 2018



MASTER THESIS FOR THE DEGREE MASTER OF PHARMACY

Identification of N-terminal and C-terminal peptides in proteomics

BY

CHRISTOFFER ÅGNES SIVERTSEN

SUPERVISORS

Associate professor Terkel Hansen

Postdoc Yvonne Pasing

PhD candidate Armin Schniers

Natural Products and Medicinal Chemistry Research Group

Department of Pharmacy

Faculty of Health Sciences

UiT – The Arctic University of Norway

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations

Abbreviation	Explanation
Ac-Trp	N-Acetyl Tryptophan
ACN	Acetonitrile
ALD-beads	Polymer-bound 4-Benzyloxy-2,6-dimethoxybenzaldehyde-beads
BCA	Bicinchoninic Acid
BSA	Bovine Serum Albumin
Chymo	Chymotrypsin
COFRADIC	Combined Fractional Diagonal Chromatography
DTT	Dithiothreitol
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
FA	Formic Acid
FASP	Filter-Aided Sample Preparation
FDR	False Discovery Rate
HPLC	High-Performance LC
IAA	Iodoacetamide
IL-6	Interleukin-6
LC	Liquid Chromatography
Lys-C	Endoproteinase Lys-C
MES	2-(N-Morpholino)-ethanesulfonic acid monohydrate
MS	Mass Spectrometry/Spectrometer
MS/MS	Tandem MS
N,N-DMEDA	N,N-Dimethylethylenediamine
NHS	N-hydroxysuccinimide
NHS-beads	NHS Mag Sepharose™
nLC	Nano-scale LC
OSM	Oncostatin M
ProC-TEL	Profiling of Protein C-termini by Enzymatic Labelling
PTM	Post Translational Modification
RP-HPLC	Reverse-phase HPLC
RT	Room temperature
SDC	Sodium Deoxycholate
TAILS	Terminal Amine Isotopic Labelling of Substrates
TEAB	Tetraethylammonium bromide
TFA	Trifluoroacetic acid
TQMS	Triple Quadrupole MS
UPLC	Ultra-Performance LC

Acknowledgements

This project was carried out from September 2017 to May 2018 at the Natural Products and Medicinal Chemistry research group, Department of Pharmacy, University of Tromsø – The Arctic University of Norway, under the supervision of associate professor Terkel Hansen, postdoc Yvonne Pasing and PhD candidate Armin Schniers.

I would like to extend my deepest gratitude to the following persons:

Terkel – for the introduction to the world of research and proteomics, for all your time, guidance, help, and input during these months. And for sharing the in the frustration nothing worked for reasons (still) unknown.

Armin – for showing me the ropes in the lab, helping me with PEAKS (and MaxQuant, even though I ended up not using it), and for your valuable input throughout the project.

Jack Ansgar-Bruun, and Tromsø University Proteomics Platform – for analysing my samples more or less immediately, whenever I handed them over.

Terje Vasskog – for helping set up the UPLC-MS method for my model systems.

Further, I would like to thank my classmates — past and present — for all the good times had, and all the fond memories gained during the years at the university. Hopefully there are, at the very least, a few more to come.

A very special thank you goes out to my former classmate Anne Christine, with whom I now share a home, a life, and two cats. Without you, I would surely be lost. Your very presence is soothing and uplifting. Thank you for all your support. I now believe I understand what you went through this time last year.

And lastly, I want to thank my family and friends back in Fauske for the continued support and encouragement in both academic and personal matters — past, present, and (hopefully) future.

Abstract

Background: Identifying modifications made to terminal parts of proteins are very useful in understanding diseases and other out of the ordinary biological states. This thesis has focused on developing methods for enriching N-terminal and C-terminal peptides from a complex protein mixture, so that analysis of these samples can give better, more comprehensive and more reproducible results.

Materials and methods: This thesis applies a bottom-up proteomics work-flow approach to develop and compare methods for enrichment of N-terminal peptides, using two digestion enzymes, three sample clean-up methods, and two enrichment mediums in different combinations. One method for C-terminal enrichment was developed, with basis in the method that gave the best results from N-terminal enrichment.

Results and discussion: None of the N-terminal enrichment methods improved the number of terminal peptides compared to the control samples. However, the results suggest that trypsin should be the enzyme of choice when enriching for N-terminal
The method for enrichment of C-terminal peptides was developed with basis in the method that gave the best results for the N-terminal enrichment. This method yielded only one terminal peptide, which is far lower than expected based on existing literature.

Conclusion: Due to one or more unknown factors that are not under control, none of the methods developed in this thesis improves the number of terminal peptides. These methods are clearly in need of further investigation, as similar methods have produced good results

.

1 Introduction

1.1 Background

Proteomics is the study the of proteome, or the proteins that can be found in an organism. Through proteomics we hope to further understand both structure and function of the various proteins that can be found in the cell(s) of an organism, and by extent the cell and the organism itself[1].

1.2 Protein Biology

The central dogma of biology states that once genetic information has been translated to a protein, it cannot be translated back to DNA/RNA or replicated to create a new protein[2], as illustrated in Figure 1-1[3].

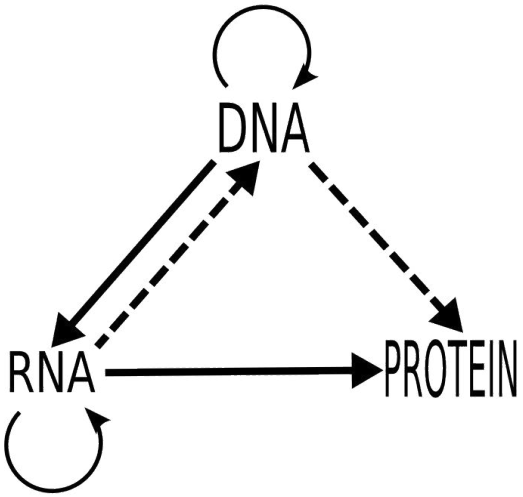


Figure 1-1: The Central Dogma of Molecular Biology

There are only 20 amino acids commonly coded for in the DNA. All of these are α-amino acids, meaning that they have an amine and a carboxylic acid connected to the same carbon — the α-carbon. These three constituents make up the main chain of the amino acid, and they are the same for all the 20 common amino acids. What makes each amino acid unique is its side chain. The side chain can be small and simple, a hydrogen in glycine — the smallest amino acid — or it can be large and complex, like the indole ring system found in tryptophan — the largest of the common amino acids. All the common amino acids, except for proline, have the same

generic structure. Proline has a cyclic structure, where the side chain is attached to both the α-carbon and the amine, making it the only common amino acid without a primary amine[4]. Both proline and the generic amino acid structure can be seen in Figure 1-2. The main chain of the amino acids can be joined together by a condensation reaction, amine to carboxyl-carbon, which forms a peptide. As more and more amino acids are added to the peptide, it gains specific properties and starts folding itself to gain secondary and tertiary structures, and becomes a polypeptide or a protein. The primary structure is the amino acid sequence itself. Separate polypeptide/protein chains can also join together, by covalent or non-covalent bonds, and create a quaternary structure[4].

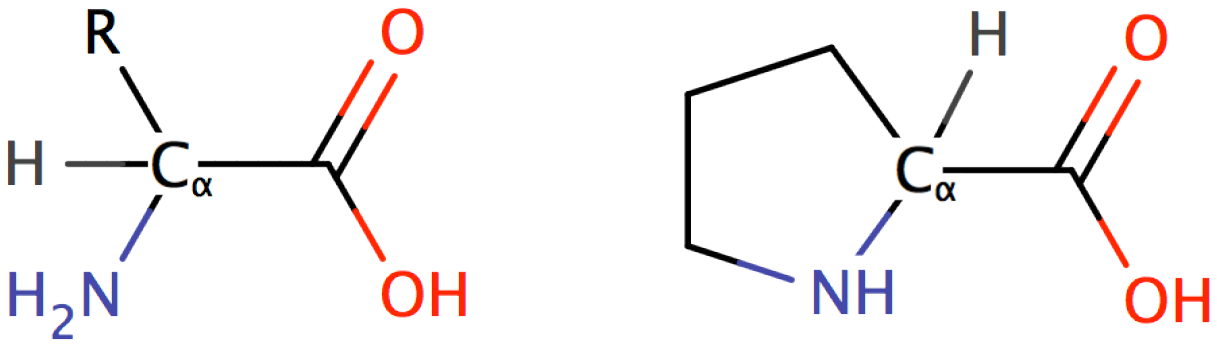


Figure 1-2: Generic α-amino acid structure on the left. Proline on the right.

After translation, proteins may be modified in order to change or activate their effects. This is often achieved by removing certain amino acids, or by adding different functional groups to them. These post translational modifications (PTMs) are important in the regulation of the activity of proteins, and can be important in various diseases and biological states[1].

1.2.1 Information found in terminal peptides

Changes made in the terminal parts of proteins often mean a change in location, conformation or activity for the protein.

An example of this is the protein oncostatin M (OSM), a cytokine in the interleukin-6 subfamily, with activities related to e.g. tumour growth, haematopoiesis, inflammation and bone growth[5–7].

The first 25 amino acids on the N-terminal side of OSM form a hydrophobic signal peptide. This peptide signals for secretion of the protein, and is cleaved off during the secretion[8]. On the C-terminal side, OSM has a hydrophilic pro-peptide consisting of 30 amino acids that is removed during maturation of the protein. The presence of this pro-peptide is shown to have

little effect on the receptor binding activity of OSM, with both mature protein and pro-OSM having equal affinity to the receptor, but the presence of this pro-peptide greatly lowers the inhibitory activity of OSM[9].

Another example is interleukin-6 (IL-6). Like OSM, IL-6 has a hydrophobic signal peptide on the N-terminal side of the protein which is removed during secretion of the protein. Interestingly, the location of the cleavage site on IL-6 seems to be important in regard to further modifications. When the signal peptide had been cleaved off N-terminally to the 28th amino acid (an alanine), IL-6 had O-linked glycosylations, while IL-6 with the new N-terminal at the 30th amino acid (a valine), had either N-linked glycosylation or both O- and N-linked glycosylations[10].

These two proteins are only two of numerous examples of regulation of protein activity through modification of the N- and C-terminus.

1.3 Top-down and bottom-up proteomics approaches

With the availability of gene and genome sequence databases and instrument improvements, the use of Mass Spectrometry (MS) has become an increasingly common way to approach proteomics[11].

In MS-based proteomics there are two main approaches: top-down and bottom-up.

In short, the top-down approach uses MS to fragment entire proteins and gives information on the molecular weight for both protein and the peptide fragments. This could yield information on the primary structure of the protein, as well as its modifications. However, the top-down approach can be difficult to apply to complex samples, such as whole proteome samples, and usually requires some sort of "simplification" through e.g. fractionation.[12] The technical limitations of MS-instruments also puts a cap on protein size that can be ionized and fragmented, which makes analysis of large proteins difficult.

The bottom-up approach digests proteins enzymatically before analysis by MS. The total protein coverage is typically lower than what can be seen in the top-down approach, and some information on PTMs can be lost[12]. However, since many of the peptide fragments are unique for the parent protein, identification and quantification of the proteins present in a sample can be successful, though skewed in favour of abundant proteins.

Due to the nature of this thesis, enrichment of terminal peptides, the bottom-up proteomics is a rather obvious choice of approach. Figure 1-3 shows a typical workflow for bottom-up proteomics, using Proteome Discoverer software for data analysis. Other data analysis software options (such as e.g. PEAKS Studio or MaxQuant) are available.

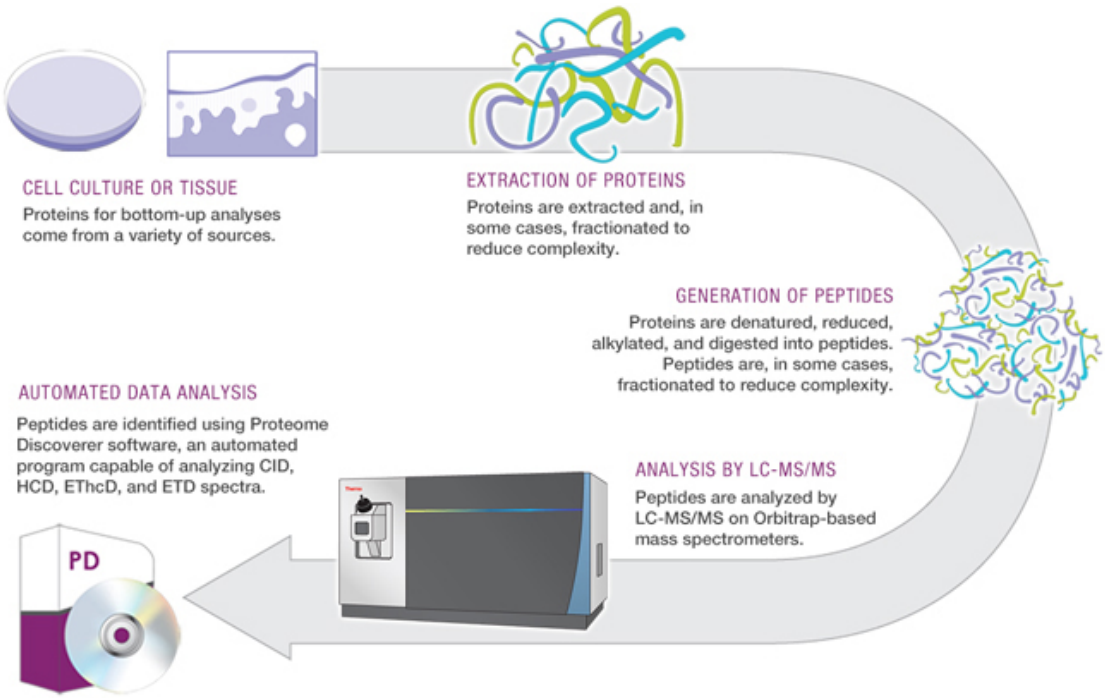


Figure 1-3: Typical workflow for Bottom-up proteomics (image from http://www.planetorbitrap.com/bottom-up-proteomics)

1.3.1 Terminomics

As mentioned in 1.2.1 the terminal parts of proteins can contain a lot of valuable information. Identifying and quantifying these peptides and the changes made to them could help with the understanding of the biological processes involved in diseases and other "out-of-the-ordinary" biological states. Terminomics, or positional proteomics, is a field within proteomics that deals with these parts of the proteins. Terminomics can be further divided into N-terminomics and C-terminomics which focuses on different ends of the protein.

Combined fractional diagonal chromatography (COFRADIC) was one of first techniques utilize a negative selection approach to analyse N-terminal peptides. The method, developed by Gevaert et al., labels α- and ϵ-amine on protein level and fractionates the resulting peptide mixture by reverse-phase HPLC (RP-HPLC). After fractionation, the free α-amines of internal and C-terminal peptides are labelled using 2,4,6-trinitrobenzensulfonic acid, a highly

hydrophobic molecule. The fractions are then separated with a second RP-HPLC run using the same settings and parameters, removing the internal and C-terminal peptides from the fraction, before analysis by LC-MS/MS[13].

Terminal amine isotopic labelling of substrates (TAILS), is another negative selection technique used for enrichment and analysis of N-terminal peptides. As with COFRADIC, TAILS labels α- and ϵ-amines on a protein label, typically using formaldehyde or isobaric tags. Unlabelled, internal and C-terminal peptides are then removed using an amine reactive enrichment agent, and the samples can be analysed using LC-MS/MS.

On the C-terminal side, the development of techniques has been somewhat slower. Lower reactivity of the carboxy-groups and the lack of basic amino acids on the neo-N-termini of C-terminal peptides complicates the C-terminomic approach. However, both COFRADIC[14] and TAILS[15] have been modified to work with C-terminal peptides, and new methods are emerging.

A method published in 2017 by Duan & Xu, termed Profiling of Protein C-termini by Enzymatic Labelling (ProC-TEL). ProC-TEL uses a positive enrichment approach labelling, using carboxypeptidase Y to label a methyl-esterified terminal carboxy-group with biotin. The proteins are then separated using gel electrophoresis; before reduction and alkylation of disulphide bonds, and digestion. After being extracted from the gel, biotinylated peptides are bound to an enrichment medium while unlabelled peptides are removed. C-terminal peptides can then be eluted from the enrichment medium and analysed using MS/MS.

1.4 Protein and peptide processing

1.4.1 Protein denaturation

To improve efficiency of the digestion enzyme, the proteins should be denatured to increase the number of cleavage sites available. This commonly done using anionic detergents, such as sodium dodecyl sulphate (SDS) or sodium deoxycholate (SDC). The detergent interferes with the hydrophobic interactions within the protein, causing the protein to lose the its tertiary structure[17]. If the protein sample is prepared from cell cultures, the detergent should be added before cell lysis, as it has lysing properties, as well as solubilizing properties which separates hydrophobic proteins from its lipid rich environment.

After unfolding, a disulphide reducing agent — such as dithiothreitol (DTT) or 2-mercaptoethanol — can access disulphide bridges normally buried in the interior of the protein, and reduce these to further denature the protein. After reduction of the disulphide bridges, the sulfhydryls can be alkylated, using an alkylation agent such as iodoacetamide, to prevent reformation of disulphide bonds.

1.4.2 Protein digestion

A handful of proteases are available for use in proteomics, each with a more or less specific cleavage site[19]. This thesis will focus on the serine proteases trypsin and chymotrypsin.

Trypsin and chymotrypsin have identical mechanisms of action (shown in Figure 1-4), but have different specificities. Due to a negatively charged aspartate in the 189-position, trypsin has a specificity for long, positively charged amino acids (arginine, lysine) when these are not immediately followed by proline. Chymotrypsin has a charge neutral serine in this position, which results in specificity for hydrophobic amino acids[20], such as e.g. tryptophan and phenylalanine.

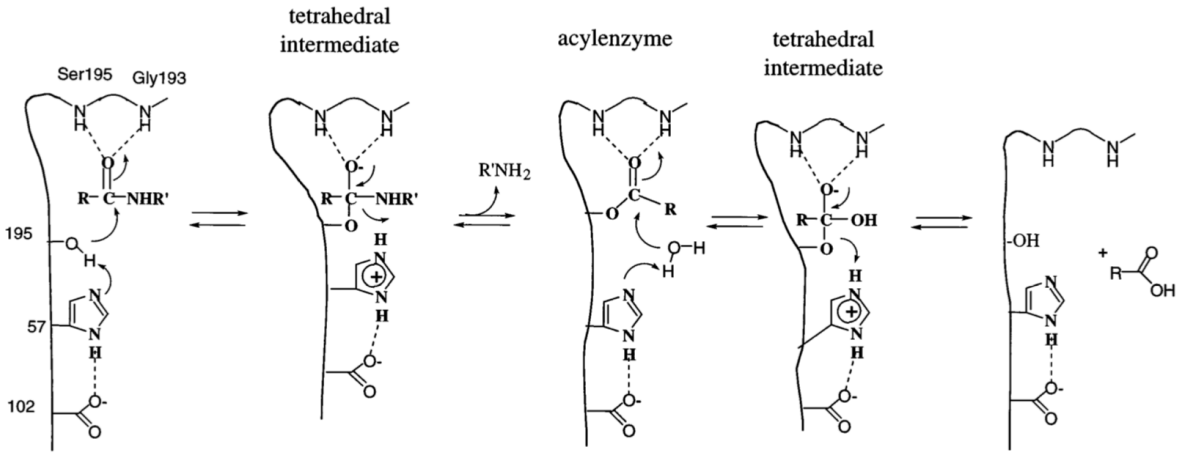


Figure 1-4: Serine protease mechanism of action

1.4.3 Peptide labelling

In order to protect the desired peptides from removal during the enrichment process, the reactive groups need to be deactivated. Amines can be protected in several different ways, depending on the purpose of the experiment. Some examples are methylation, acetylation or reaction with 1-Flouro-2,4-dinitrobenzen.

Carboxy groups can be protected using e.g. ethanolamine or N,N-dimethylethylenediamine.

1.5 Enrichment of terminal peptides

After digestion of proteins in whole proteome samples, the resulting peptide mixture is rather complex. The two terminal peptides constitute only a small part of the entire protein and will therefore be present in low concentrations compared to internal peptides. To reduce complexity and increase the proportion of terminal peptides, the peptide mixture must be enriched[21].

N-termini positive selection enrichment focuses on labelling α-amines in a way that allows α-amine containing peptides to be retained, while unwanted peptides are removed from the peptide mixture. E.g. one could label lysine ϵ-amines using O-methylisourea before labelling protein α-amines using biotin. After digestion, the peptide solution can be passed through immobilized streptavidin, retaining biotin labelled peptides. With the non-biotin labelled peptides removed, the biotin-streptavidin bond can be reduced, and the labelled peptides eluted[22]. Known problems with positive selection enrichment are the difficulty in differentiating between α- and ϵ-amine when labelling amines, without added steps; as well as trouble labelling N-termini with naturally occurring modifications, such as acetylation.

N-termini negative selection enrichment, which this thesis will comprehend, labels both α- and ϵ-amines on protein level, preventing further reactions. After digestion the enrichment medium binds to the α-amine of neo-N-termini, and is then removed so that only peptides with inactivated amines remain[23,24]. E.g. one could label free amines using reductive dimethylation and then digest the proteins in the sample. After digestion, HPG-ALD polymer can be added to the sample along with a reductive agent, such as cyanoborohydride, to bond with the amine group of neo-N-termini. The polymer can then be removed through filtration, leaving only dimethylated peptides in the sample[25].

The enrichment of C-termini is challenging due to the lower reactivity of carboxy-groups compared to amine-groups[26], but the principles are the same as with N-terminal enrichment: positive enrichment binds desired peptides to the enrichment medium, and negative selection blocks free carboxy-groups and binds the Neo-C-termini of undesired peptides to the enrichment medium.

2 Aims of the Study

The aim of this master thesis is to:

- Develop and compare different methods for N-terminal enrichment.
- Using the best method from N-terminal enrichment to develop a method for enrichment of C-terminal peptides.
- Isolation and identification of N-terminal and C-terminal peptides from a complex mixture of proteins, using different enrichment mediums.

3 Materials

3.1 Chemicals

Table 3-1: List of chemicals

Chemical	Catalogue #	Manufacturer
1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC)	03450-5G	Sigma-Aldrich
2-(N-Morpholino)ethanesulfonic acid monohydrate (MES)	69892-25G	Sigma
α-Chymotrypsin from bovine pancreas (chymo)	C7762-25MG	Sigma
Acetone	32201-1L	Sigma-Aldrich
Acetonitrile (ACN)	34851-2.5L	Sigma-Aldrich
Amine-terminated magnetic particles (NH_2-beads)	I7643-5ML	Sigma
Ammonium acetate PA	1.1116-1	KEBO Lab
Benzylamine	13180	Fluka
Calcium chloride	21074-1KG	Sigma-Aldrich
Dithiothreitol (DTT)	D5545-5G	Sigma-Aldrich
Endoproteinase Lys-C (Lys-C)	unknown	unknown
Ethanol	24106-2.5L-R	Sigma-Aldrich
Formaldehyde solution, 37%	252549-1L	Sigma-Aldrich
Formic acid ≥ 98% (FA)	33015-1L	Sigma-Aldrich
Glycine HCl	A1067,1000	AppliChem
Guanidine hydrochloride	50950-1KG	Sigma
Hydrochloric acid	30721-1L-GL	Sigma-Aldrich
Hydroxylamine hydrochloride	55459-50G	Fluka
Iodoacetamide (IAA)	I6125-10G	Sigma
L-Tryptophan	T0254-5G	Sigma-Aldrich

MilliQ-water (H_2O)	-	Millipore (USA)
N-acetyl DL Tryptophan (Ac-Trp)	A6251-5G	Sigma
N-Hydroxysuccinimide (NHS)	130672-25G	Sigma
N,N-Dimethylethylenediamine (N,N-DMEDA)	D158003-100G	Sigma
NHS Mag Sepharose™ (NHS-beads)	28-9513-80	GE Healthcare
Pierce™ BCA Protein Assay Kit	23227	Thermo Scientific
Polymer-bound 4-Benzyloxy-2,6-dimethoxybenz aldehyde (ALD-beads)	538213-5G	Aldrich
Sodium cyanoborohydride	156159-10G	Aldrich
Sodium deoxycholate (SDC)	D6750-25G	Sigma-Aldrich
Triethylammonium bicarbonate (TEAB)	T7408-100ML	Sigma
Trifluoroacetic acid 100% (TFA)	61030 T6508-25ML	Riedel-de Hägen Honeywell
Trypsin	T6567-5X20UG	Sigma-Aldrich
Urea	33247-250G	Sigma-Aldrich

The Lys-C was a vial of stock solution from the -20˚C-freezer in the lab, that has since been misplaced and cannot be found. The information on production date, concentration and initials of the person who made it, is therefore unavailable.

3.2 Materials

Table 3-2: List of equipment and materials

Material/equipment	Description	Manufacturer/provider
Biofuge Fresco	Centrifuge	Heraeus Instruments
Bond Elute OMIX, 96 C18 100 uL	Desalting pipette tip	Agilent Technologies
Centrifugal Filter Modified PES 30K, 500 µL	Purification filter	VWR
CentriVap Concentrator	Evaporator	Labconco
Disposable Antistatic Microspatula	Spatula	VWR
EASY-nLC 1000	nLC-instrument	Thermo Scientific
Eppendorf Protein LoBind	Sample vials, 0.5 mL/1.5 mL	Eppendorf
FALCON tube 15mL	Sample vial Stock solution storage	Eppendorf
Heraeus Megafuge 16	Centrifuge	Thermo Scientific
MagNA Lyser Green Beads	vial and beads for cell lysis	Roche
MagNA Lyser instrument	cell lysis instrument	Roche
MagRack 6	Magnet rack	GE Healthcare
MaxQuant	Data analysis	Max Planck Institute of Biochemistry
Micropipette	0.5–2.5 µL 0.5–10 µL 10–100 µL 10–200 µL 100–1000 µL 1–5 mL	Eppendorf Research
Mini Star Silverline	Microcentrifuge	VWR
MS 1 Minishaker	Shaker/Vortexer	IKA-Werke
NanoDrop™1000 Spectrophotometer	Peptide concentration measurement	Thermo Scientific

Parafilm “M”	Clingfilm for sample vial covering.	Bemis
PEAKS Studios 8	Data analysis	Bioinformatics Solutions Inc.
Pierce Low Protein Binding Microcentrifuge Tubes, 2.0 mL	Sample vial, 2 mL	Thermo Scientific
Pierce ™ BCA Assay Kit	BSA standard and working reagent for determination of protein concentration.	Thermo Scientific
Pipette tip	10 µL 250 µL 1000 µL 5 mL	Thermo Scientific
Q Exactive™ hybrid quadrupole-Orbitrap™ mass spectrometer	MS-instrument	Thermo Scientific
SpectraMax 190 Microplate Reader	Protein concentration measurement	Molecular Devices
Termaks Inkubator	Incubator	Termaks AS
Tissue Culture plate, 96-well	Microwell plate	BD Falcon

4 Methods

4.1 Sample preparation

The steps described in 4.1 were used for preparation of all samples, regardless of whether they were to be N-terminal enriched or C-terminal enriched.

4.1.1 Cell lysis

1. Centrifuged and frozen *Escherichia coli* (ATCC® 25922™) culture aliquots were retrieved from a -80˚C freezer, and thawed at room temperature (RT).
2. To remove possible contaminants, the *E. coli* pellets were centrifuged at 16,000 × g, before removing the supernatant and washing with 200 µL H_2O. The process was then repeated.
3. After the second wash, the pellet was re-suspended in 500 µL lysis buffer, consisting of 100mM TEAB and 2% SDC.
4. The resuspended cell culture was transferred to MagNA Lyser Green Beads-vials, and lysed using the MagNA Lyser instrument at 7,000 RPM for 50 seconds.

4.1.2 Determination of protein concentration

Protein concentration was determined using a Pierce™ BCA Assay Kit from Thermo Scientific, with the microplate procedure as described by the user guide[27].

1. BCA Assay standards were prepared with six dilutions of Bovine Serum Albumin (BSA) in 5 mM TEAB, with BSA concentrations ranging from 0–250 µg/mL.
2. The protein sample was prepared in five dilutions in 5 mM TEAB, ranging from 1:20–1:1000 of original concentration.
3. 25µL of standards and sample dilutions were pipetted into a 96-well microplate in three parallels, and 200 µL of working reagent added were added to the microplate wells.
4. The microplate was incubated with gentle agitation for 30 minutes at 37 ˚C.
5. After incubation, the plate was cooled at RT for 10 minutes before absorbance was recorded on the microplate reader at 562 nm.

4.1.3 Reduction and alkylation of disulphide bridges

1. The desired amount of protein, calculated using the results obtained from the BCA Assay, was transferred to a 1.5mL Eppendorf Protein Lo-Bind-tube and diluted to the desired concentration using 100mM TEAB.

2. The disulphide bridges were reduced by adding 200 mM dithiothreitol (DTT) to the samples, to a concentration of 5 mM DTT. The sample was then incubated with gentle agitation in a Termaks Incubator for 30 minutes at 54 ˚C.
3. After incubation, the sample was cooled at RT for 10 minutes.
4. When the sample was cooled to RT, the cysteine sulfhydryls were carbamidomethylated by adding 600 mM iodoacetamide (IAA) to a final concentration of 15 mM and incubating the samples in the dark at RT for 30 minutes.
5. To remove excess IAA in the sample, DDT was added to a final concentration of 5 mM a second time, followed by incubation for 10 minutes at RT.

4.2 Methods for enrichment of N-terminal peptides

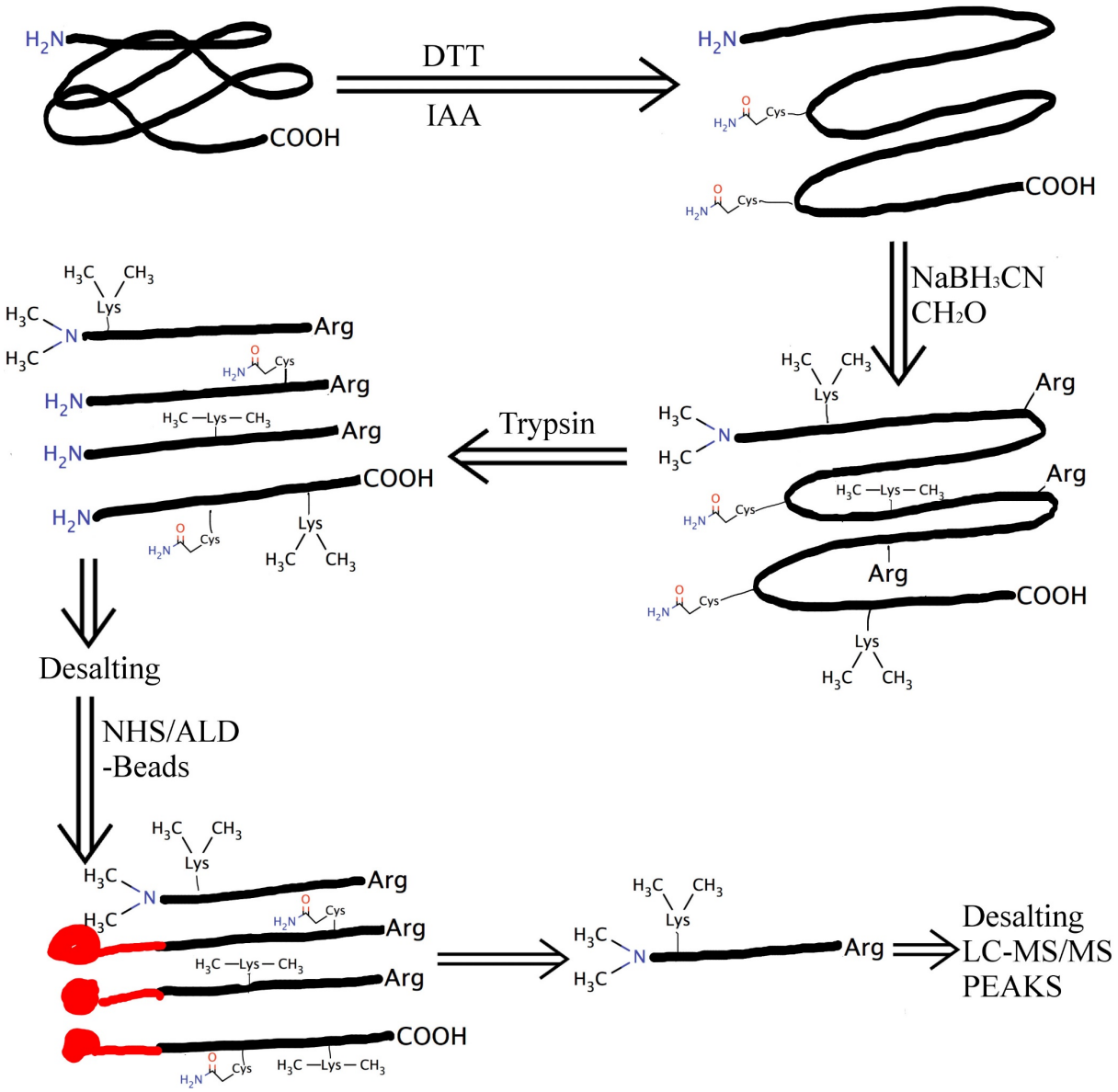


Figure 4-1: N-terminal enrichment process overview

4.2.1 Dimethyl labelling of N-termini and lysines

Dimethylation of protein N-termini and lysine ϵ-amines was accomplished through reductive methylation. Three different methods were tested, utilizing different reagents for quenching the reaction.

4.2.1.1 Dimethylation using glycine and hydroxylamine for quenching

1. The dimethyl labelling solution was prepared by weighing sodium cyanoborohydride ($NaBH_3CN$) in a 1.5 mL Eppendorf Safe-Lock, and then adding TEAB, CH_2O and H_2O to concentrations of 200 mM $NaBH_3CN$, 100 mM CH_2O, and 100 mM TEAB.
2. The labelling solution was added to the protein sample, to a final concentration of 40 mM $NaBH_3CN$ and 20 mM CH_2O, and the sample was vortexed for 10 seconds.
3. The sample was incubated for two hours at 37 ˚C with gentle agitation.
4. After two hours, an additional 50% of labelling buffer volume was added. The sample was vortexed for 10 seconds and left in the incubator with gentle agitation for additional 30 minutes.
5. After labelling, excess CH_2O was quenched by adding 1.2 M glycine hydrochloride to a final concentration of 60 mM. The sample was vortexed for 10 seconds and incubated for 10 minutes at RT.
6. 2.6 M hydroxylamine hydrochloride was added to a final concentration of 130 mM. The sample was vortexed for 10 seconds and incubated for 15 minutes at RT.

4.2.1.2 Dimethylation using ammonium acetate and hydroxylamine for quenching

5. Step 1–4 is performed as described in 4.2.1.1.
6. Excess CH_2O was quenched by adding 1.2 M ammonium acetate to a final concentration of 60 mM. The sample was vortexed for 10 seconds and incubated for 10 minutes at RT.
7. 2.6 M hydroxylamine hydrochloride was added to a final concentration of 130 mM. The sample was vortexed for 10 seconds and incubated for 15 minutes at RT.

4.2.1.3 Dimethylation using only hydroxylamine for quenching

5. Step 1–4 is performed as described in 4.2.1.1.
6. Excess CH_2O was quenched by adding 2.6 M hydroxylamine to a final concentration of 130 mM. The sample was vortexed for 10 seconds and incubated for 30 minutes at RT.

4.2.2 Sample clean-up

There were three different methods of sample clean-up tested for enrichment of N-terminal peptides.

4.2.2.1 Acid precipitation of detergent

Precipitation of SDC was the only sample clean-up method performed after protein digestion.

1. After digestion, trifluoroacetic acid (TFA) was added to the sample to a final concentration of 1%.
2. The sample was centrifuged for 20 minutes at 16,000 × g.
3. After centrifugation the supernatant was carefully pipetted out and transferred to a new sample tube and the remaining pellet was discarded.

4.2.2.2 Ethanol precipitation of proteins

1. The sample was transferred to a 15mL test tube, and 10x sample starting volume of ice cold EtOH was added.
2. The sample was chilled at -20 ˚C for at least 1 hour.
3. After cooling, the sample was centrifuged for 30 minutes at 4692 × g and 4 ˚C.
4. The supernatant was removed, and the pellet was washed with 500µL ice-cold acetone
5. After centrifugation for 5 minutes at 4692 × g, the supernatant was removed, and the sample was dried for 30 minutes at RT.

4.2.2.3 Filter aided sample preparation (FASP)

This clean up method, and subsequent digestion, is a modified version of a method published by Pasing et al. in 2016 for use with hydrophobic samples[28].

The amount of protein per filter was limited to 200–250 µg to prevent the filters from possibly getting clogged.

1. 100 µL washing buffer, consisting of 8.0 M urea and 100 mM TEAB, was added to the filter(s). The filter was then centrifuged for 20 minutes at 16,000 × g and 21 ˚C, and the liquid was discarded.
2. The sample was added to the filter(s), followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.
3. 200 µL washing buffer was added to the sample, followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.

   This step was repeated twice using 100 µL washing buffer.

4. 100 µL of 100 mM TEAB was added to the sample, followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.
   This step was repeated twice.
5. Digestion of the protein samples was performed using either trypsin of chymotrypsin as described in 4.2.3, using 200 µL of the appropriate digestion solution. After digestion the sample was centrifuged for 20 minutes at 16,000 × g and 21 ˚C.
6. 50 µL 100mM TEAB was added to the sample, followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C.
7. 50 µL H_2O was added to the sample, followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C.
8. The sample was desalted as described in 4.2.4.

4.2.3 Protein digestion

4.2.3.1 Trypsin

1. Calcium chloride ($CaCl_2$) was added to the sample, to a final concentration of 1 mM.
   If FASP had been used for purification, 200 µL of 100 mM TEAB and 1 mM $CaCl_2$ were added to the sample on the filter.
2. Trypsin was added to the sample at an enzyme:protein ratio of 1:20 (w/w).
   The sample was carefully aspirated and dispensed 10 times using a micropipette.
3. The sample was incubated for 18 hours at 37 ˚C, with gentle agitation.

4.2.3.2 Chymotrypsin

1. $CaCl_2$ was added to the sample, to a final concentration of 10 mM.
2. Chymotrypsin solution was added to the sample at an enzyme:protein ratio of 1:20 (w/w). The sample was carefully aspirated and dispensed 10 times using a micropipette.
3. The sample was then incubated for 18 hours at 25 ˚C, with gentle agitation.

4.2.3.3 Trypsin with Lys-C pre-digestion

1. Lys-C was added to the sample at an enzyme:protein ratio of 1:100 (w/w).
2. The sample was incubated for 6 hours at 37 ˚C, with gentle agitation.
3. Calcium chloride ($CaCl_2$) was added to the sample, to a final concentration of 1 mM.
4. Trypsin was added to the sample at an enzyme:protein ratio of 1:20 (w/w).
   The sample was carefully aspirated and dispensed 10 times using a micropipette.
5. The sample was incubated for 18 hours at 37 ˚C, with gentle agitation.

4.2.4 Desalting

The desalting was performed using the OMIX C18 100 µL-tip, following the manufacturers protocol with some changes made to better accommodate this experiment.

1. TFA was added to the sample, to a final concentration of 1%.
   If method 4.2.2.1 was used for sample clean-up, this step was skipped, as TFA had already been added.
2. Conditioning buffer (50% ACN in H_2O) was aspirated and then dispensed as waste. This was done twice.
3. Equilibration buffer (1% TFA in H_2O) was aspirated and then dispensed as waste. This was done twice.
4. The sample was slowly aspirated and dispensed 10 times.
   If the sample volume was >1000 µL, aspiration and dispensing was repeated as many times as needed to ensure that the total sample volume had passed through the column.
5. Rinsing buffer (0.1% TFA in H_2O) was aspirated and dispensed as waste. This was done twice.
6. Elution buffer (0.1% FA and 80% ACN in H_2O) was aspirated and then dispensed in a new sample vial.
   If the amount of protein in the sample was over 79 µg*, steps 3–6 were repeated as many times as needed to desalt the total amount of protein in the sample.
7. The elution buffer was evaporated using the CentriVap at 60 ˚C.
8. The sample was reconstituted with 100mM TEAB, to a concentration of 1–2 µg/µL.

4.2.5 Enrichment of N-terminal peptides

4.2.5.1 Enrichment using NHS Mag Sepharose™ (NHS-beads)

1. The NHS-beads were pipetted over to an Eppendorf Protein LoBind 1.5 mL sample vial, and the liquid was removed.
2. The beads were washed three times using a volume of 100 mM TEAB equal to the volume of NHS-beads, vortexing the beads for 10 seconds, and removing the liquid using the magnet rack and a micropipette.
3. The sample was added to the beads and incubated for 2 hours at 37 ˚C while shaken using a MS 1 Minishaker at 1000 RPM, to make sure the beads didn't sink to the bottom.

* Maximum loading capacity according to the manufacturer.

4. After incubation, the sample vial was placed in the magnet rack and the liquid was transferred to a new sample tube.
5. The NHS-beads were washed with ¼ × sample volume and vortexed for 10 seconds. The liquid was combined with the one from step 4.
   This was repeated once.
6. The liquids were combined, and the beads discarded.
7. The sample was desalted and dried following the procedure described in 4.2.4, skipping the final step, and then analysed.

4.2.5.2 Enrichment using polymer-bound 4-Benzyloxy-2,6-dimethoxybenzaldehyde (ALD-beads)

1. The ALD-beads were weighed out in a n Eppendorf Protein LoBind1.5 mL sample vial.
2. To remove possible contaminants, the beads were washed with 100 mM TEAB. The beads were spun down using a microcentrifuge, and the liquid was discarded. This process was repeated three times.
3. The sample solution was added to the beads, along with $NaBH_3CN$ to a final concentration of 40mM.
4. The sample was incubated for 2 hours at 37 ˚C, while shaken vigorously using a MS 1 Minishaker at 1000 RPM, to make sure the beads didn't sink to the bottom.
5. 1.2M glycine was added to a final concentration of 60 mM, before incubating for 10 minutes at RT,
6. NH_2OH was added to a final concentration of 130 mM and incubated for another 15 minutes at RT.
7. The beads were spun down using the microcentrifuge, and the liquid was transferred to a new sample vial.
8. The ALD-beads were washed with ¼ × sample volume and vortexed for 10 seconds. The liquid was combined with the one from step 7. This was repeated twice.
9. The liquids were combined, and the beads discarded.
10. The sample was desalted and dried following the procedure described in 4.2.4, skipping the final step, and then analysed.

4.3 Method for enrichment of C-terminal peptides

This method used 200 µg protein from a stock of previously lysed, reduced and alkylated proteins. The stock had a protein concentration of 1.85 µg/µL, making the starting volume of the sample 108.1 µL.

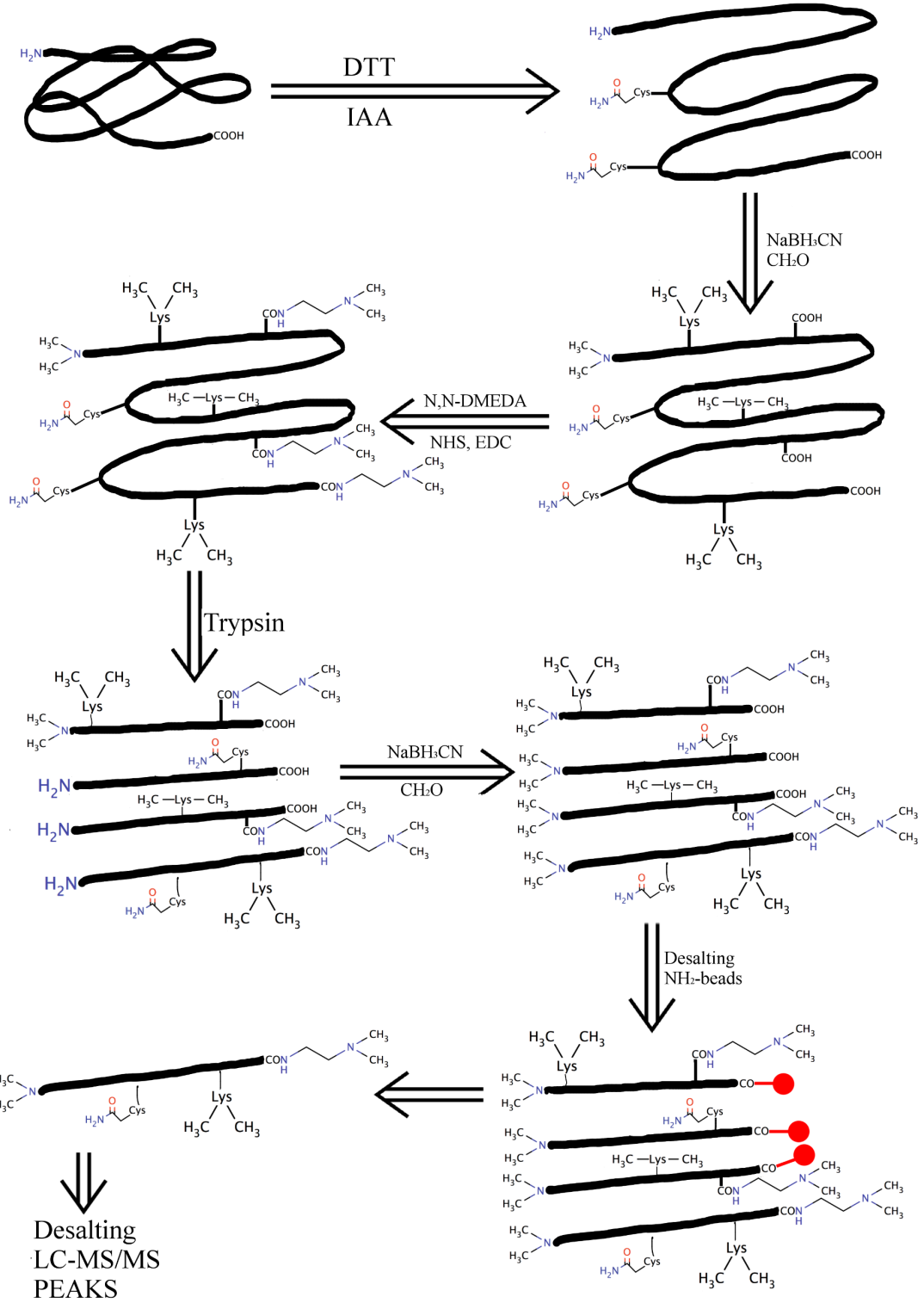


Figure 4-2: C-terminal enrichment process overview

4.3.1 Dimethylation of N-termini and lysine ϵ-amines

The method described in 4.2.1.1 was used to label primary amines.

1. 27 µL labelling buffer was added to the sample.
2. The sample was incubated for 2 hours at 37 ˚C, with gentle agitation.
3. 13.5 µL labelling buffer was added, and the sample was incubated for an additional 30 minutes at 37 ˚C, with gentle agitation
4. 7.82 µL of 1.2 M glycine hydrochloride was added, to a final concentration of 60 mM, and the sample was incubated for 10 minutes at RT, with gentle agitation.
5. 8.23 µL 2.6 M hydroxylamine solution was added, to a final concentration of 130 mM, and the sample was incubated for 15 minutes at RT, with gentle agitation.

4.3.2 FASP

The sample was processed using a modified version of the FASP method from 4.2.2.3.

The liquid was discarded after every step.

1. 100 µL washing buffer was added to the filter. The filter was then centrifuged for 20 minutes at 16,000 × g and 21 ˚C.
2. The sample was added to the filter(s), followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.
3. 200 µL washing buffer was added to the sample, followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.
   This step was repeated twice using 100 µL washing buffer.
4. 100 µL of 100 mM TEAB was added to the sample, followed by centrifugation for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.
   This step was repeated once.
5. The sample was washed with 100 µL H_2O and was centrifuged for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.
6. The sample was washed with 100 µL 200 mM MES/2.0 M guanidine and was centrifuged for 20 minutes at 16,000 × g and 21 ˚C. The liquid was discarded.

4.3.3 Labelling of C-termini and aspartate/glutamate carboxylates

1. The sample was reconstituted on the filter with 89.1 µL 500 mM MES, 89.1 µL 5M guanidine and 29.8 µL N,N-dimethylethylenediamine (N,N-DMEDA).
2. pH was adjusted to 4–5 using 32 µL 12M HCl.
3. 12.16 µL 200 mM NHS/400 mM EDC solution was added, to a final concentration of 10mM NHS and 20 mM EDC, and the sample was left to incubate at RT for 1 minute before measuring pH.
4. After checking that the pH ≈ 5, the sample was incubated for one hour at 25 ˚C.
5. 12.8 µL of NHS/EDC solution was added, and the pH was checked after incubating at RT for one minute; pH ≈ 5.
6. The sample was incubated at 25 ˚C for another hour, with gentle agitation.
7. 13.48 µL NHS/EDC solution was added, and the pH was checked after incubating at RT for one minute; pH ≈ 5.
8. The sample was incubated for 16 hours.
9. The sample was centrifuged for 20 minutes at 16,000 × g and 21 ˚C, using the Hearaeus Megafuge 16.

4.3.4 Washing

The liquid was discarded after every step.

1. The sample was washed with 200 µL H_2O and centrifuged at 16,000 × g and 21 ˚C.
2. The sample was washed with 100 µL H_2O and centrifuged at 16,000 × g and 21 ˚C.
3. The sample was washed twice with washing buffer. After every wash the sample was centrifuged at 16,000 × g and 21 ˚C.
4. The sample was washed twice with 100 µL 100 mM TEAB. After every wash the sample was centrifuged at 16,000 × g and 21 ˚C.

4.3.5 Digestion

The method in 4.2.3.1 was used for digestion, with a few steps added.

1. 200 µL 100 mM TEAB and 1 mM $CaCl_2$ was added to the sample.
2. 10 µL 1 µg/µL trypsin was added to the sample. To mix, the sample was carefully aspirated and dispensed 10 times with a micropipette.
3. The sample was incubated for 18 hours at 37 ˚C, with gentle agitation.
4. The sample was centrifuged for 20 minutes at 16,000 × g and 21 ˚C.

5. The sample was washed with 50 µL 100 mM TEAB and centrifuged for 20 minutes at 16,000 × g and 21 ˚C.
6. The sample was washed with 50 µL H_2O and centrifuged for 20 minutes at 16,000 × g and 21 ˚C.
7. The liquids were combined in a sample vial, and the filter was discarded.

4.3.6 Dimethylation of neo-N-termini

The method described in 4.2.1.3 was used to label neo-N-termini amines.

1. 77.5 µL labelling buffer was added.
2. The sample was incubated for 2 hours at 37 ˚C, with gentle agitation.
3. 38.8 µL labelling buffer was added.
4. The sample was incubated for 30 minutes at 37 ˚C, with gentle agitation
5. 22.4 µL 2.6 M NH_2OH solution was added, to a final concentration of 130 mM.
6. The sample was incubated for 30 minutes at RT, with gentle agitation.

4.3.7 Desalting

The desalting was performed using the OMIX C18 100 µL-tip, following the manufacturers protocol with some changes made to better accommodate this experiment.

1. HCl was added to the sample, to a final concentration of 75 mM.
2. Conditioning buffer (50% ACN in H_2O) was aspirated and then dispensed as waste. This was done twice.
3. Equilibration buffer (75 mM HCl) was aspirated and then dispensed as waste. This was done twice.
4. The sample was slowly aspirated and dispensed 10 times.
5. Rinsing buffer (20 mM HCl) was aspirated and dispensed as waste. This was done twice.
6. Elution buffer (0.1% FA and 80% ACN in H_2O) was aspirated and then dispensed in a new sample vial.
7. Steps 3–6 was repeated two more times, combining the eluate from the 6. step in the same sample vial.
8. The elution buffer was evaporated using the CentriVap at 60 ˚C.

4.3.8 C-terminal enrichment using amine terminated magnetic particles (NH_2-beads)

1. After desalting, the sample was reconstituted with 500 µL 200 mM MES/2 M guanidine.
2. 100µL NH_2-beads were pipetted over to a 1.5 mL sample vial on the MagRack6, and the liquid was removed.
3. To remove possible contaminants, 500 µL of 200 mM MES/2.0 M guanidine equal to the volume of NH_2-beads were added, and the beads were vortexed for 10 seconds. The vial was then placed on the MagRack6 to remove and discard the supernatant. This process was repeated three times.
4. The sample was added to the beads, and 55.6 µL 100 mM NHS/500 mM EDC solution was added, to a final concentration of 10 mM NHS and 50 mM EDC.
5. After 1 minute the pH was controlled; pH ≈ 4–5.
6. The sample was incubated for 3 hours at 25 ˚C while shaken vigorously, using a MS 1 Minishaker at 1000 RPM, to make sure the beads didn't sink to the bottom.
7. 61.7 µL NHS/EDC solution was, and after one minute the pH was controlled; pH ≈ 5.
8. The sample was incubated for 16 hours at 25 ˚C, while shaken vigorously.
9. The sample was placed in a MagRack6 and the liquid was pipetted to a new sample vial.
10. The NH_2-beads were washed with 200µL MES/guanidine buffer, and the beads were vortexed for 10 seconds. This was repeated once.
11. The liquids from step 10 were combined with the liquid from step 9, and the beads were discarded.
12. The samples was desalted and dried following the procedure described in 4.2.4, skipping the last step, and then analysed.

4.4 Sample analysis

The steps described in 4.4 were used for analysis of all samples, regardless of whether they were to be used in N-terminal enrichment or C-terminal enrichment.

4.4.1 Determining peptide concentration

Before the nLC-MS/MS analysis, the peptide concentration was determined with spectrophotometry using a NanoDrop™1000 spectrophotometer and the "Protein A280" setting. Absorbance was measured at $\lambda = 205$nm.

1. The sample was reconstituted in 10 µL 0.1% TFA.
2. The spectrophotometer was conditioned with 2 µL H_2O.
3. 2 µL 0.1% TFA was used to set a reference point.
4. A "blank" sample of 2 µL 0.1% TFA was ran.
5. Added 2 µL of the sample, and measured absorbance in 4 parallels.
   If more than one sample had the concentration measured in quick succession, step 3–5 was repeated for as many times as needed.

4.4.2 nLC-MS/MS

The nLC-MS/MS analysis was performed with a Thermo Scientific™ Q Exactive™ Hybrid Quadropole-Orbitrap™ Mass Spectrometer, connected to a Thermo Scientific™ EASY-nLC™ 1000 Liquid Chromatograph.

The nLC utilized a C18 column (2 µm, 100 Å, 75 µm, 50 cm), and two different gradients.

- A short gradient was 30 minutes long, with 4–40% ACN in 0.1% FA.
  The injection volume was 2 µL, for a total of 0.5 µg peptides in 0.1% TFA
- A long gradient was 60 minutes long, with 4–8% ACN in 0.1% FA for the first 10 minutes and 8–40% for the last 50 minutes.
  The injection volume was 3 µL, for a total of 0.75 µg peptides in 0.1% TFA.

Flow rate was 250 nL/min for both gradients.

MS settings:

Table 4-1: Q-Exactive settings

Method of Q Exactive	
OVERALL METHOD SETTINGS	
Global Settings	
Chrom. Peak width (FWHM)	9 s
Experiment	
FULL MS / DD-MS^2 (TOPN)	
General	
Polarity	Positive
Default charge state	2
Full MS	
Resolution	70,000
AGC target	1e6
Maximum IT	100 ms
Scan Range	400 to 2,000 m/z
Spectrum data type	Profile
dd-MS^2 / dd-SIM	
Resolution	17,500
AGC Target	1e5
Maximum IT	60 ms
Loop count	10
TopN	10
Isolation window	2.0 m/z
(N) CE / stepped (N) CE	nce: 28
Spectrum data type	Profile
dd Settings	
Intensity threshold	3.3e4
Charge exclusion	unassigned, 1, 5 – 8, >8

4.4.3 Data analysis

PEAKS Studio 8.0 was used for data analysis. The settings used can be seen in Table 4-2.

Table 4-2: PEAKS search settings.

Error Tolerance	Precursor mass	10.0 ppm
	Fragment ion:	0.05 Da
Enzyme	Specified by each sample	
	Allow non-specific cleavage at — — ends of the peptide	one
	Maximum allowed variable PTM per peptide	5
PTM	Fixed	Carbamidomethylation (all samples)
	Variable	Lysine dimethylation (all samples)
		N-term dimethylation (N-terminal samples)
		N-term acetylation (N-terminal samples)
		C-term N,N-DMEDA (C-terminal samples)
		Aspartate/glutamate N,N-DMEDA (C-terminal samples)
Database	Select database	E. coli
De Novo Tag Options	Available de novo tags:	DENOVO 2
General options	Estimate FDR with decoy-fusion	

Peptide FDR was set to 1% and peptide information was exported as a .csv-file.

The .csv-file was imported in Excel, and was processed using the following functions:

- Average peptide mass was calculated using =AVERAGE(Mass)
- The number of peptides was counted using =COUNTIF(*)-1 (-1 to exclude the header row)
- Number of N-terminal peptides was counted =COUNTIF("?(*"), and then removing "false positives" by removing K-leading peptides with only one +28.03 Da modification, as well as C-leading peptides without a dimethyl modification.
- % N-terminal peptides was calculated by =N-term/Peptides
- The amount of lysines were counted using the function =COUNTIF("*K*"), and the amount of modified lysines (termed K_{mod}) were counted using =COUNTIF("*K(*"). The ratio was then calculated by = K_{mod}/K.

4.5 Analysis of model systems

To test whether or not some of the steps used in the enrichment processes worked by principle, two simple systems were set up. The N-terminal model used benzylamine as a model substance, and the C-terminal model used tryptophan and N-acetyl tryptophan (Ac-Trp). These model systems were analysed in a simpler way than the proteome samples, using ultra performance LC (UPLC) and triple quadrupole MS (TQMS). Since these model systems were about proof of concept, the aim was fast results rather than reproducibility or high accuracy.

The model samples were separated using C18 column (2.5 µm, 100 Å, 1.0 x 50 mm) with a 5-minute gradient of 1–95% ACN in 0.1% FA and a flow rate of 0.38 mL/min. The injection volume was 10 µL.

After separation the samples were first analysed by the TQMS in full scan mode looking for masses (m/z) between 50–500 Da, to look for potential contaminants and the mass of the target ions. The samples were then analysed in selected ion recording (SIR) mode, with the MS selecting for the specific masses of the ion supposed to be present in the sample. See Table 4-3 for masses. These masses match the ions found in the complementing MS full scan, which may deviate from the calculated masses, as the MS-instrument had not been calibrated in a while.

Table 4-3: m/z for selected ions.

Ion	m/z	Calculated m/z
Benzylamine + H^+	107.8	108.15
Dimethylbenzylamine + H^+	135.8	136.21
Trp + H^+	205.00	205.23
Dimethyl-Trp + H^+	233.00	233.28

5 Results and Discussion

5.1 Validation of dimethyl labelling

A sample of 50 µg protein was processed using glycine and hydroxylamine for quenching the dimethylation, acid precipitation of SDC for clean up, and trypsin for digestion. Results from this sample showed that 97.6% of all lysines had been labelled after processing.
The lysines in control samples for trypsin digestion and chymotrypsin digestion, as well as in three randomly selected samples, were also counted. In these five samples an average of 99.0% of lysines were labelled.

One sample had been labelled using a labelling buffer that was a week old, and with this sample the importance of using freshly prepared labelling buffer shone through. The PEAKS-searched showed that only 5.3% of the lysines had been labelled, and enrichment of the sample yielded a mere 15 N-terminal peptides (0.3% of all peptides found).

5.2 Testing of enzyme digestion

In addition to chymotrypsin and trypsin, a digestion method using endoproteinase Lys-C in addition to trypsin was tested. A sample of 50 µg protein was processed using glycine and hydroxylamine for quenching the dimethylation, and trypsin for digestion. but before trypsination the sample was incubated for 6 hours at 37 ˚C with 0,5 µg Lys-C. The results, however, showed that the Lys-C pre-treated sample was only 1.5 percentage points higher in peptides ending with a lysine. It has been reported that monomethylation of lysines will inhibit Lys-C[29], so it would be natural that dimethylation has the same effect. Digestion using Lys-C in a pre-digestion step was therefore not investigated any further.
Digestion using trypsin resulted in fewer peptides in total compared to chymotrypsin, but as can be seen in Figure 5-1, the number of N-terminal peptides identified by PEAKS was more than double compared to samples that had been digested with chymotrypsin. A possible explanation for this is the specificity difference for the two enzymes.
As mentioned in 1.4.2, trypsin has a specificity for lysine and arginine due to the negatively charged aspartate in the 189-position. However, as with Lys-C, dimethylation of lysine greatly reduces the specificity of trypsin for lysine, resulting in cleavage almost exclusively at arginine leaving a very limited amount of cleavage sites[30] —and therefore longer peptides; while the specificity for several hydrophilic amino acids in chymotrypsin could give shorter peptides due increased amount of cleavage sites.

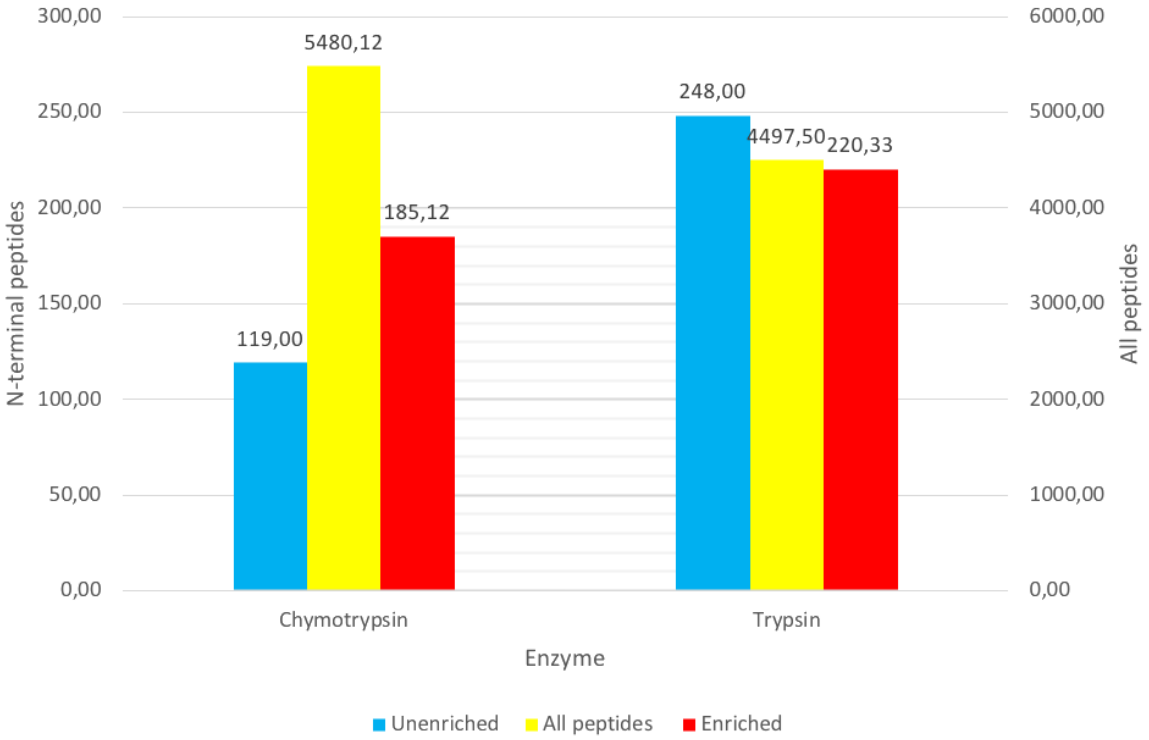


Figure 5-1: Average peptide yield per enzyme

5.3 N-terminal enrichment

5.3.1 NHS-beads vs. ALD-beads

17 different samples were set up in parallel to find which of the two bead types would give the best results. The samples were set up/prepared according to Table 5-1 and

Table 5-2, and were processed using glycine and hydroxylamine for quenching of the dimethyl labelling, acid precipitation of SDC as sample clean-up, and chymotrypsin for digestion. Eight of these samples were enriched using the NHS-bead method, and nine were enriched using the ALD-bead method.

Table 5-1: NHS-enriched parallels

Volume	Protein	Amount of NHS-beads		
1 mL	2 mg	500 µL (NHSA1)	250 µL (NHSA2)	125 µL (NHSA2)
1 mL	1 mg	500 µL (NHSB1)	250 µL (NHSB2)	125 µL (NHSB3)
0.5 mL	0.5 mg	500 µL (NHSC1)	250 µL (NHSC2)	

Table 5-2: ALD-enriched parallels

Volume	Protein	Amount of ALD-beads		
1 mL	2 mg	15 mg (ALDA1)	5 mg (ALDA2)	1 mg (ALDA3)
1 mL	1 mg	15 mg (ALDB1)	5 mg (ALDB2)	1 mg (ALDB3)
0.5 mL	0.5 mg	15 mg (ALDC1)	5 mg (ALDC2)	1 mg (ALDC3)

In general, enrichment using NHS-beads gave a higher number and ratio of N-terminal beads than enrichment using ALD-beads, but as can be seen in Figure 5-2 there is only a difference of about 10 peptides between the averages. It was therefore chosen to move forward using the ALD-beads due to the much lower price of these compared to the NHS-beads.

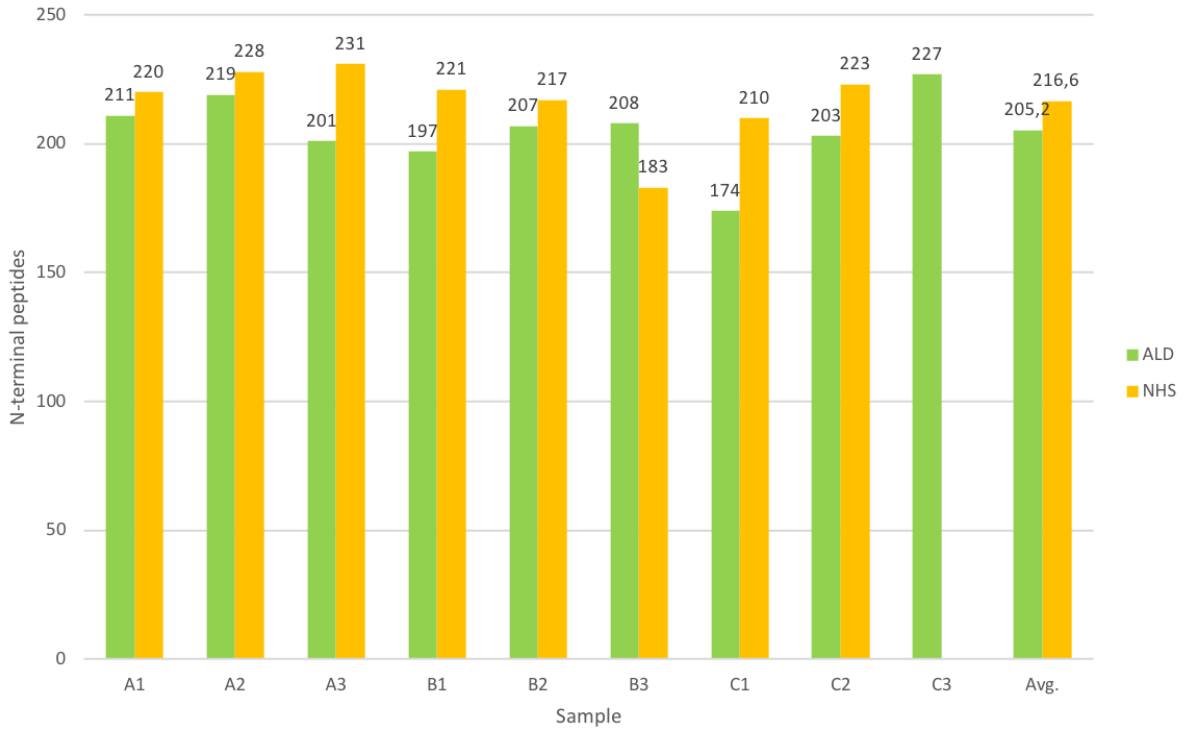


Figure 5-2: Comparison of results from the 17 different parallels

5.3.2 N-terminal enrichment-model using benzylamine

Due to trouble in achieving a good amount of N-terminal peptides during the first four months, a model was set up using benzylamine as a substitute for proteins/peptides.

0.5 ml of 666 µM benzylamine, in 100 mM TEAB, was processed using both the NHS and the ALD enrichment methods. The analysis of these two samples on UPLC with a Triple Quadrupole MS (UPLC-TQMS) showed that ≥98% of the benzylamine had been removed from the sample.

A new sample of 666 μM benzylamine was dimethylated, using glycine and hydroxylamine for quenching. The sample was then analysed on UPLC, and the results showed that the labelling had been successful. However, when the sample was desalted and analysed again, the results showed that glycine was still present after desalting. The presence of glycine in the sample is a great problem for the enrichment. With the free amine in glycine, and the rather high amount added, glycine can bind to the beads during the enrichment process and lead to loss of peptides. This discovery led to a change in the sample clean-up method from detergent precipitation to protein precipitation.

5.3.3 Detergent precipitation vs. protein precipitation vs. FASP

As explained in the results from the benzylamine model, glycine was still present in the sample after clean-up using the detergent precipitation and a new method for sample clean-up was needed. Ethanol precipitation of proteins was tested as the new method, with three different samples containing a different amount of proteins. The samples were labelled using glycine and hydroxylamine for quenching, and chymotrypsin for digestion. As can be seen in Table 5-3 the results were even worse than before. This is like due to the fact that glycine is insoluble in ethanol[31], and is precipitating along with the proteins. However, the fourth sample tested using ethanol precipitation of proteins, in which the labelling reaction quenched using ammonium acetate in place of glycine, still had a much lower number and ratio of N-terminal peptides than samples that were processed using the acid precipitation of the detergent.

The third and final clean-up method that were tested was the filter aided sample preparation. It was used on only three samples for the N-terminal enrichment. One of these samples, digested with trypsin and enriched with NHS-beads, had 221 N-terminal peptides out of a total of 4238 peptides. While not the highest total number of N-terminal peptides found in a sample, at 5.21% it is the highest ratio of N-terminal peptides seen in a processed sample during this project. However, this sample was run on a longer gradient than the other samples in the N-terminal experiments, and is therefore not necessarily comparable to them.

The two other samples that had been prepared using FASP were both enriched with ALD-beads but digested with different enzymes. The chymotrypsin sample had results that was lower than what had been seen with similar samples using acid precipitation of detergent as a clean-up method (155 N-terminal peptides/2.46% for the FASP-sample, and an average of 205,2 N-terminal peptides/3.61% for comparable samples.

The FASP-sample that had been digested using trypsin, and enriched using ALD-beads, had the third highest number of N-terminal peptides seen, but the ratio was somewhat lower compared to the sample with the best ratio (4.45% vs. 4.91%).

5.3.4 Summary

On average, trypsin yielded more N-terminal peptides than chymotrypsin (average 220.33 for trypsin and average 185.12 for chymotrypsin). However, the chymotrypsin number if affected by the poor performance of ethanol precipitation of proteins (average of 123.7 N-terminal peptides), which was only tested using chymotrypsin. If these results are excluded from the average, the Number of N-terminal peptide goes up to an average of 204.5.

In the 17 parallels set up, NHS-beads gave slightly more peptides than ALD-beads (216.6 for NHS, 205.2 for ALD).

Ethanol precipitation of proteins was only tested using chymotrypsin and ALD-beads, but given the low average (123.7 / 2.29% N-terminal peptides), it hard to imagine changing digestion or enrichment method would improve this.

The best results in terms of absolute number of N-terminal peptides identified, was a sample which underwent detergent precipitation, digested by chymotrypsin and enriched using NHS-beads. Out of 5805 peptides found by PEAKS in the sample, 231 peptides (3.98 %) had a dimethyl modification that implied it was an N-terminal peptide (+28.03 Da for non-lysine N-terminal amino acids, or +56.06 Da modifications for N-terminal lysines).

The sample with the best ratio, 5.21 %, of N-terminal peptides was in a sample purified with FASP, digested with trypsin and enriched with NHS-beads, which yielded 221 N-terminal peptides out of 4238 total peptides.

Table 5-3: Results from the N-terminal enrichment

Sample	*Amount of protein*	*Clean Up*	*Digestion Enzyme*	*Enrichment*	*Amount of beads*	*Peptide average mass, Da*	*Peptides*	*N-terminal peptides*	*Ratio of N-terminal peptides*
171018_Chymotrypsin _Enrichment	200 µg	Detergent precipitation	Chymo	NHS-beads	100 µL	1348,96	5362	151	2,82 %
171018_Trypsin _Enrichment	200 µg	Detergent precipitation	Trypsin	NHS-beads	100 µL	1620,42	4441	218	4,91 %
180106_ALDA1	2 mg	Detergent precipitation	Chymo	ALD-beads	14.92 mg	1554,56	5756	211	3,67 %
180106_ALDA2	2 mg	Detergent precipitation	Chymo	ALD-beads	5.07 mg	1610,50	5677	219	3,86 %
180106_ALDA3	2 mg	Detergent precipitation	Chymo	ALD-beads	0.99 mg	1548,30	5793	201	3,47 %
180106_ALDB1	1 mg	Detergent precipitation	Chymo	ALD-beads	15.00 mg	1494,56	5831	197	3,38 %
180106_ALDB2	1 mg	Detergent precipitation	Chymo	ALD-beads	4.99 mg	1523,62	5732	207	3,61 %
180106_ALDB3	1 mg	Detergent precipitation	Chymo	ALD-beads	1.02 mg	1473,60	5713	208	3,64 %
180106_ALDC1	0.5 mg	Detergent precipitation	Chymo	ALD-beads	15.01 mg	1475,80	5321	174	3,27 %
180106_ALDC2	0.5 mg	Detergent precipitation	Chymo	ALD-beads	4.95 mg	1569,02	5625	203	3,61 %
180106_ALDC3	0.5 mg	Detergent precipitation	Chymo	ALD-beads	1.00 mg	1545,78	5708	227	3,98 %
180106_NHSA1	2 mg	Detergent precipitation	Chymo	NHS-beads	500 µL	1597,37	5487	220	4,01 %
180106_NHSA2	2 mg	Detergent precipitation	Chymo	NHS-beads	250 µL	1557,32	5488	228	4,15 %
180106_NHSA3	2 mg	Detergent precipitation	Chymo	NHS-beads	125 µL	1556,41	5805	231	3,98 %
180106_NHSB1	1 mg	Detergent precipitation	Chymo	NHS-beads	500 µL	1585,98	5346	221	4,13 %
180106_NHSB2	1 mg	Detergent precipitation	Chymo	NHS-beads	250 µL	1580,70	5296	217	4,10 %

180106_NHSB3	1 mg	Detergent precipitation	Chymo	NHS-beads	125 µL	1530,53	5462	183	3,35 %
180106_NHSC1	0.5 mg	Detergent precipitation	Chymo	NHS-beads	500 µL	1577,98	5423	210	3,87 %
180106_NHSC2	0.5 mg	Detergent precipitation	Chymo	NHS-beads	250 µL	1644,90	5081	223	4,39 %
180208_NoGlyA	0.5 mg	Protein precipitation	Chymo	ALD-beads	2.1 mg	1343,47	5380	123	2,29 %
180208_NoGlyB	1 mg	Protein precipitation	Chymo	ALD-beads	2.4 mg	1348,31	5310	113	2,13 %
180208_NoGlyC	2 mg	Protein precipitation	Chymo	ALD-beads	2.2 mg	1347,71	5684	133	2,34 %
180215_NoGlyD	1 mg	Protein precipitation	Chymo	ALD-beads	1.9 mg	1443,48	5542	148	2,67 %
180226_ALD10	1.2 mg	Protein precipitation	Chymo	ALD-beads	11.4 mg	1327,35	5074	110	2,17 %
180226_ALD20	1.2 mg	Protein precipitation	Chymo	ALD-beads	22.1 mg	1320,94	5360	115	2,15 %
180227_FASP_test	1 mg	FASP	Chymo	ALD-beads	21.1 mg	1434,38	6301	155	2,46 %
180309_Trypsin_FASP	1 mg	FASP	Trypsin	ALD-beads	11.3 mg	1879,22	4989	222	4,45 %
180320_Chymo_test	50 µg	—	Chymo	—	—	1274,16	3926	119	3,03 %
180320_Trypsin_test	50 µg	—	Trypsin	—	—	1688,64	3417	248	7,26 %
Same sample as above on a long nLC-gradient						1797,42	4577	254	5,55 %
180403_FASP3	0.5 mg	FASP	Trypsin	NHS-beads	500 µL	1765,36	4238	221	5,21 %

5.4 C-terminal enrichment

5.4.1 C-terminal Enrichment-model using tryptophan and N-acetyl tryptophan

A model using tryptophan and N-acetyl tryptophan (Ac-Trp) was set up to test some of the steps in the C-terminal enrichment process, using a modified version of a protocol for C-terminal peptide enrichment published by Schilling et al. in *Protocol Exchange from Nature Protocol*s[15].

1.3 mg Ac-Trp was dissolved in 1000 µL buffer containing 200 mM MES, 2.0 M guanidine and 1M N,N-DMEDA (in the Schilling et al. protocol ethanolamine was used instead of N,N-DMEDA). The sample was then adjusted to pH 5 using 165 µL 100% TFA. NHS/EDC was added to 10/20 mM three times, incubating at 25 ˚C for one hour after the first two additions and for 16 hours after the third addition. The pH was controlled 1 minute after every addition of NHS/EDC. After the 16-hour incubation, the solution was analysed on UPLC-TQMS to confirm that the process had been successful.

Once the N,N-DMEDA and Ac-Trp coupling had been confirmed, a sample of 1 mg tryptophan was dimethylated —using N,N-DMEDA for quenching— and desalted. Analysis on UPLC-TQMS showed that quenching with N,N-DMEDA seemed to have worked, but that a large amount of DM-tryptophan appeared to have been removed during the desalting. Therefore, the enrichment process itself ended up not being tested using this model.

The labelling process was then tested using a trypsinated protein sample, which was analysed with the methods described in 4.4. The PEAKS-search showed very few peptides that had been labelled with N,N-DMEDA, but due to time restrictions the experiment was pushed forward to the next phase.

5.4.2 Enrichment of C-terminal peptides using amine terminated magnetic particles

When the data from the sample put through the processes as described in 4.3 was analysed, PEAKS found only 61 peptides in the sample. Only two of these peptides had an N,N-DMEDA label, and only one of these two peptides had the label C-terminally. This is a far cry from the 10s-100s of peptides anticipated in Schillinger et al.'s protocol[15].

Due to the limited amount of time for this thesis, any further testing to find out what could be the problem with the C-terminal was not possible. However, one of the key difficulties during this part was working with N,N-DMEDA. The NHS/EDC reaction requires a pH of 4–5, but the addition of N,N-DMEDA to 1.0 M brought the pH in the buffer up to 11–12. Lowering the pH to 4–5 required a large amount of HCl, which in turn lowered the concentration of both N,N-DMEDA and the MES/guanidine.

6 Conclusion

While none of the results have been validated or tested in more than one parallel, it would seem that none of the methods presented here would be usable to increase the amount of terminal peptides in a complex proteome sample. In theory, the methods should be possible, as similar approaches have been reported, with much higher numbers than achieved in this project.

On the N-terminal side, it appears that trypsin gives a higher number of N-terminal peptides — at least that could be detected with PEAKS —than chymotrypsin. Additionally, ethanol precipitation of proteins as a sample clean-up method seems to give a lower number of N-terminal peptides compared to acid precipitation of detergent and the FASP method. The FASP method and the acid precipitation of detergent gave similar results. Clearly there are one or more variables in our approach that are not under control, or that is treated the wrong way. If this variable is found, the approach should give good results based on the currently published results.

The method for enrichment of C-terminal peptides was a single test to investigate if it would be possible to enrich for C-terminal peptides. However, the results were not positive, presumably because of the same unknown variable that is present in the N-terminal enrichment process.

7 Future Aspects

The need for a robust protocol for both N-terminal and C-terminal is evident, however in the current work there are undisputedly some steps and variables not under control and in further need of investigation. Dimethylamine was chosen as an amine label to better mimic the charge distribution of tryptic peptides, but this may, for unknown reasons, have introduced problems during the enrichment process.

Since the depletion of benzylamine seemed to work, the overall method should be possible, but is in need of an in-depth investigation of the variables in sample preparation and enrichment methods.

Upon success of the N-terminal enrichment method, the C-terminal enrichment can be addressed. Also here should a model system be utilized to investigate the different variables such as labelling buffers, labelling reagents, N-terminal labelling, and other variables that might be of importance.

References

1. Nawin Mishra. *Introduction to Proteomics: Principles and Applications*. (John Wiley & Sons, Inc., 2010).
2. Crick, F. On Protein Synthesis. *Symp. Soc. Exp. Biol.* **12,** 138–163 (1958).
3. Crick, F. Central Dogma of Molecular Biology. *Nature* **227,** 561–563 (1970).
4. Branden, C. & Tooze, J. *Introduction to Protein Structure*. (Garland Publishing, 1991).
5. Tanaka, M. & Miyahima, A. Onconstatin M, a multifunctional cytokine. in *Reviews of Physiology, Biochemistry and Pharmacology* 39–52 (Springer Berlin Heidelberg, 2003). doi:10.1007/s10254-003-0013-1
6. Wahl Rouleau, A. F. *et al.* Regulation of inflammatory responses by oncostatin M. *J. Immunol.* **162,** 5547–5555 (1999).
7. Walker, E. C. *et al.* Oncostatin M promotes bone formation independently of resorption when signaling through leukemia inhibitory factor receptor in mice. *J. Clin. Invest.* **120,** 582–592 (2010).
8. Malik, N. *et al.* Molecular Cloning, Sequence Analysis, and Functional Expression of a Novel Growth Regulator, Oncostatin. M. *Mol. Cell. Biol.* **9,** 2847–2853 (1989).
9. Linsley, P. S., Kallestad, J., Ochs, V. & Neubauer, M. Cleavage of a Hydrophilic C-Terminal Domain Increases Growth- Inhibitory Activity of Oncostatin M. *Mol. Cell. Biol.* **10,** 1882–1890 (1990).
10. May, L. T., Shaw, J. E., Khanna, A. K., Zabriskie, J. B. & Sehgal, P. B. Marked cell-type-specific differences in glycosylation of human interleukin-6. *Cytokine* **3,** 204–211 (1991).
11. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
12. Yates, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* **11,** 49–79 (2009).
13. Gevaert, K. *et al.* Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* **21,** 566–569 (2003).
14. Van Damme, P. *et al.* Complementary positional proteomics for screening substrates of endo-and exoproteases. *Nat. Methods* (2010). doi:10.1038/nmeth.1469
15. Schilling, O., Barré, O., Huesgen, P. F. & Overall, C. M. Enrichment of natural and proteolytically generated protein C-termini on a proteome-wide scale. *Protocol Exchange* (2010). Available at: http://dx.doi.org/10.1038/nprot.2010.100.
16. Wedemeyer, W. J., Welker, E., Narayan, M. & Scheraga, H. A. Disulfide Bonds and Protein Folding. *Biochemistry* **39,** 4207–4216 (2000).
17. Lin, Y. *et al.* Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. *Anal. Biochem.* **377,** 259–266 (2008).
18. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry*. (W.H. Freeman and Company, 2008).

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

20. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **102,** 4501–4524 (2002).

21. McDonald, L. & Beynon, R. J. Positional proteomics: Preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nat. Protoc.* **1,** 1790–1798 (2006).

22. Timmer, J. C. *et al.* Profiling constitutive proteolytic events in vivo. *Biochem. J.* **407,** 41–8 (2007).

23. Rogers, L. D. & Overall, C. M. Proteolytic Post-translational Modification of Proteins: Proteomic Tools and Methodology*. *Mol Cell Proteomics* **12,** 3532–3542 (2013).

24. *Methods in Enzymology - Proteomics in Biology, Part A*. (Academic Press, 2017).

25. Demir, F., Niedermaier, S., Kizhakkedathu, J. N. & Huesgen, P. F. *Profiling of Protein N-Termini and Their Modifications in Complex Samples*. *Protein Terminal Profiling: Methods and Protocols* (Springer Nature, 2017). doi:10.1007/978-1-4939-6850-3

26. Marino, G., Eckhard, U. & Overall, C. M. Protein Termini and Their Modifications Revealed by Positional Proteomics. *ACS Chem. Biol.* **10,** 1754–1764 (2015).

27. Pierce Biotechnology. *User Guide: Pierce BCA Protein Assay Kit (MAN0011430 Rev. A)*.

28. Pasing, Y., Colnoe, S. & Hansen, T. Proteomics of hydrophobic samples: Fast, robust and low-cost workflows for clinical approaches. *Proteomics* **17,** 1–9 (2017).

29. Wigle, T. J. *et al.* Accessing protein methyltransferase and demethylase enzymology using microfluidic capillary electrophoresis. *Chem. Biol.* **17,** 695–704 (2010).

30. Golghalyani, V., Neupä, M., Wittig, I., Bahr, U. & Karas, M. ArgC-Like Digestion: Complementary or Alternative to Tryptic Digestion? *J. Proteome Res.* **16,** 978–987 (2017).

31. Lide, D. R. *CRC Handbook of Chemistry and Physics*. (CRC Press Inc., 1998).