2018

Development of methods for the analysis of human protamine via 2D LC-MS/MS

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

DEVELOPMENT OF METHODS FOR THE ANALYSIS OF HUMAN PROTAMINE VIA 2D LC-MS/MS

by

JACOB MATTHEW SAMUEL

B.A., Vanderbilt University, 2016

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2018

Approved by

First Reader

Robin W. Cotton, Ph.D.
Associate Professor, Program in Biomedical Forensic Sciences
Department of Anatomy & Neurobiology

Second Reader

Claude Mallet, Ph.D.
Senior Scientist
Waters Corporation

ACKNOWLEDGMENTS

Special acknowledgements go to Dr. Cotton, for coming up with the idea and giving me the opportunity to explore and do my thesis research in a field I was not yet very familiar with. Acknowledgements also go to Dr. Mallet for a large portion of the resources used for this research in addition to being a strong guiding and teaching force throughout.

DEVELOPMENT OF METHODS FOR THE ANALYSIS OF HUMAN PROTAMINE VIA 2D LC-MS/MS

JACOB MATTHEW SAMUEL

ABSTRACT

Protamine, a set of small basic proteins (P1 and P2), play a key role in compacting and protecting the DNA in sperm. As such, the structure of how P1 and P2 bind to DNA and potentially themselves and each other is of interest to several fields including forensics. In forensic DNA analysis, protamine binding of DNA is taken advantage of in the "differential extraction" procedure in which a sample that contains sperm and non-sperm cells can have DNA from the two different cell types separated and extracted at different points thus preventing a mixture of DNA. A key component of this greater structure and what makes the differential extraction functional are the disulfide bonds formed by protamine. So as a first step to elucidating the protamine-DNA complex, methods to analyze human protamine via 2D-Liquid Chromatography-Mass Spectrometry (2D-LC-MS) were developed in the hopes they could be used for disulfide bond mapping. Methods and multiple strategies for digestion, 2D-LC-MS were investigated and developed using chum salmon protamine. Digestion strategies were developed for Chymotrypsin and Lys-C, Trypsin or Arg-C with incubation times and substrate:enzyme mass ratios optimized. Various "trap and elute" 2D-chromatography configurations were tested for analysis intact and digested protein. Using H_2O with 2% NH_4OH as the loading mobile phase and H_2O and Acetonitrile both with 0.5% formic acid as the eluting mobile phases with the first dimension column being an HLB

(Hydrophilic Lipophilic Balance) 2.1 x 30 mm column and the second dimension being an C18 2.1 x 100 mm was found to produce the highest signal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D	Two-dimensional
ACD	At-Column Dilution
ACN	Acetonitrile
BEH	Ethylene Bridged Hybrid
CID	Collision Induced Dissociation
DTT	Dithiothreitol
ESI	Electrospray Ionization
FA	Formic Acid
g	Gram
HCl	Hydrochloric Acid
HLB	Hydrophilic Lipophilic Balance
IPA	Isopropanol
LC	Liquid Chromatography
LOD	Limit of Detection
m/z	Mass to Charge ratio
MeOH	Methanol
mL	Milliliter
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometer
ng	Nanogram
NH_4OH	Ammonium Hydroxide
P(1-4)	Protamine 1-4 (context dependent)
Q-TOF	Quadrupole-Time of Flight tandem Mass Spectrometer
TIC	Total Ion Chromatogram
µg	Microgram

1. INTRODUCTION

1.1 Significance of Protamine in Forensics

Deoxyribonucleic Acid (DNA) has been a famed identifier in the forensic world and beyond, making DNA analysis a key staple of forensic laboratories. While there are a variety of ways DNA evidence can be left behind, DNA evidence is generally left behind, not as isolated DNA strands but packaged in cells such epithelial cells, white blood cells or sperm cells. A common initial step in DNA analysis is isolation of the DNA from the cells via extraction by breaking down the components of the cell and nucleus, and separating some or all of those components from the DNA so that the DNA is ready for analysis. As mentioned, DNA can be left behind in a variety of different cell types and, as such, this can lead to some different obstacles in isolating the DNA. Differences in cell structure for the most part are not a problem except when it comes to sperm cells.

Cells commonly have their DNA as part of the macromolecular complex known as chromatin which consists of a repeating unit known as a nucleosome. The nucleosome consists of eight histones (2 of each H2A, H2B, H3 and H4) with 147 base pairs (bp) of DNA wrapped around the eight histones and another, smaller, histone (H1) between each nucleosome.(1, 2) Sperm cells differ here from other cells typically encountered in DNA analysis in that during sperm maturation, the large majority of histones are replaced with protamine, a low molecular weight, basic protein with 2 major variants (P1 and P2) that greatly condenses the chromatin and results in the cell becoming transcriptionally silent.(3, 4) One major difference between histones and protamines is

the presence of disulfide bonds. These S-S bonds plays a key role in DNA extraction from sperm cells as a reducing agent is required to free the DNA. This also allows for differential extraction, a technique in which a sample that contains epithelial cells and sperm cells (possible in sexual assault cases) can have the DNA from each cell type extracted at different times thus allowing for the deconvolution of a potential mixture.(5) Understanding the structure protamines and how they bind DNA could have further implications on this extraction method and larger implications on related fields. Damaged protamine and even improper ratios of P1 to P2 have been linked to infertility.(6-10) Additionally, the protamine ratio has been implicated in proper sperm head shape formation which is key for hydrodynamically efficient sperm in regards to competition with other sperm.(11) Protamine's primary role in condensing the DNA has further implications in embryonic development, unpacking of the sperm DNA and cellular programming once merged with the ovum. (12, 13)

Unfortunately, there are clear gaps in the current literature on fully elucidating the structure of human protamines as they are bound to DNA in mature sperm. There is, however, a larger body of information on protamine in other species, which, depending on the species, could have various implications on what might be expected to be seen in the human protamine DNA complex. While a lot is known, a definitive understanding of protamines structure *in vivo*, in its larger conformation, how it binds to DNA and potentially how the two protamines bind to each other is still unknown. As a first step to get a more complete understanding the higher structure of protamine, this thesis aims to map the disulfides of human protamines 1 and 2.

1.2 Background on Protamine

As mentioned earlier, there are 2 protamines, P1 and P2, synthesized from the PRM1 and PRM2 genes with P2 having two variants, P2a and P2b (sometimes seen in the literature as P3). PRM2 is believed to have come about as a duplication of PRM1 due to extremely similar sequence and PRM1 being seen throughout mammals while PRM2 is limited to mostly rodents and primates resulting in PRM1 being relatively more conserved while PRM2 shows much greater divergence. Additionally, P2b, has, so far, only been seen in human, ape and Old World Monkey sperm cells. (3, 14) Due to P2b having the same sequence as P2a except for being a threonine, arginine and histidine shorter on the N-terminus, they will be further referred to collectively as just P2.

Sequence of P1:

10 20 30 40 50
ARYRC CRSQS RSRYY RQRQR SRRRR RRSCQ TRRRA MRCCR PRYRP RCRRH

Sequence of P2a:

10 20 30 40 50
RTHGQ SHYRR RHCSR RRLHR IHRRQ HRSCR RRKRR SCRHR RRHRR GCRTR KRTCR RH

Sequence of P2b:

10 20 30 40 50
GQ SHYRR RHCSR RRLHR IHRRQ HRSCR RRKRR SCRHR RRHRR GCRTR KRTCR RH

Sequence of Proprotamine:

10 20 30 40 50
MVRPY RVRSLS ERSHE VYRQQ LHGQE QGHHG QEEQG LSPEH VEVYE RTHGQ SHYRR

60 70 80 90 100
RHCSR RRLHR IHRRQ HRSCR RRKRR SCRHR RRHRR GCRTR KRTCR RH

Figure 1. Sequences of P1, P2a, P2b and proprotamine. Note that P2b is the last 54 amino acids of P2b and that P2a is the final 57 amino acids of proprotamine.

Despite its sparse presence amongst species, P2 and its dysfunction has been implicated in many of the issues mentioned above, such as fertility, indicating a

necessary and disparate role.(8, 10, 15) A major difference between the two is that P1 is synthesized as a mature protein from PRM1 while P2 is synthesized from a larger precursor coded for by PRM2 called proprotamine. Proprotamine is converted to P2 during DNA condensation so the proprotamine binds to DNA and is then subject to proteolytic cleavage resulting in P2. The portion of proprotamine that is cleaved off is also believed to play an unknown role in the sperm cell.(14) While proprotamine is 102 amino acids long, the mature forms of P1, P2a and P2b are 50, 57 and 54 amino acids respectively (Figure 1)(16-18)

1.2.1 Composition of Protamine

The protamines are largely made up of arginine, existing at ~ 50% with the other basic amino acids, histidine and lysine showing up in considerably smaller amounts. P2 is the more basic of the two with a greater histidine content. The other amino acids present include: tyrosine, serine, glutamine, threonine, leucine, isoleucine, proline and cysteine. While not extremely varied, the amino acids present, and further studies investigating their presumptive roles, do provide some context clues for what the secondary structure of protamine bounds to DNA could be and how it functions. It also provides key context for how the protein could possibly be manipulated and analyzed. The most common amino acid, arginine, is thought to play a critical role in binding DNA as the positive charge of its guanidium group aids binding to negatively charge phosphate groups of DNA.(19) Arginine is generally seen in short stretches in the center of the sequence making this portion of the protein the DNA binding domain demarcating the

terminal ends for other purposes.(3, 20) Additionally, there are several serine and threonine seen in sequence which have been proven to be sites for phosphorylation(12, 21). Functionally, the protein is phosphorylated post-translation and once again upon entering the egg as a means of neutralizing the positive charge on protamine, thus diminishing the electrostatic interaction between protamine and the DNA. This allows the chromatin to expand and frees the sperm DNA to play its key role in fertilization(3, 13).

The presence of multiple cysteines, as shown in many of the experiments involving protamine, indicate the presence of disulfide bonds which play a key structural role not only in the protein conformation but the stability of the sperm DNA.(8) In bull protamine, four of seven cysteines are terminal cysteines involved in intramolecular disulfide bonds which are believed to fold back the terminal domains over the central, arginine rich regions. The remaining cysteines are believed to be in intermolecular disulfide bonds which are key in creating a larger network of protamines and DNA which greatly increases the stability of the chromatin. These intermolecular bonds are thought to play a key role in shielding the arginine from competition in its electrostatic interactions with DNA by salts. This makes protamines with these intramolecular disulfide bonds remarkably stable at high salt concentrations as shown by Transmission Electron Microscopy (TEM) studies comparing bull protamine (has intramolecular disulfides) and chum salmon protamine (no intramolecular disulfide bonds). (22)

In addition to disulfide bonds, several cysteine's, along with histidines, are believed to make a Cys_2 His_2 zinc finger motif in P2 which causes secondary structure transitions (23) which is believed to create further stability in binding to DNA (15, 24)

These salt bridges are believed to be key in preventing an excess of disulfide bonds from forming and preventing the DNA from being over compressed. Thus there may be a fine balance in which DNA is protected in the sperm cell but is free enough for an easy and timely release of the genetic material in the ooplasm. (13)

1.2.2 Transition from Histones to Protamines

One other key source of information regarding how protamine and DNA interact, is the specific transition of histones to protamine which speaks to the environment and specific circumstances under which this protein-DNA complex forms. While the DNA is initially bound to histones, it is bound to slightly different variants of histones as compared to a somatic cell, such as the H1T, which is seen as being bound less tightly and therefore is conducive to removal. Some of these are present initially but some are expressed later, playing more active roles in the transition to protamines. (2) The transition begins following meiosis where an increase in acetylation and ubiquitination of the histones occurs weakening the bonds to the DNA. This leads way to the transition proteins further reducing the affinity of the DNA to the histones. There are 4 human transition proteins (TP1-4) and with similar characteristics to histones and protamines: basic, small and having abundant positively charged amino acids. TP1 is the most abundant and plays a key role in relaxing the nucleosomal core and removing the histones. During this time, the DNA, in its chromatin complex, starts to change in secondary structure and condense. More due to the function of TP2, the DNA becomes stabilized and transcription becomes silent although the transition proteins exact

contribution to these changes are not fully known.(5, 8, 24, 25) The next phase is protamine binding, DNA condensation and further DNA stabilization. This emphasizes the importance of studying protamine from mature sperm cells due to all of the elaborate steps it takes, which this brief description just scratched the surface of, to get protamine in this specific configuration. (26)

1.2.3 Models in the Literature for Protamine-DNA Binding

As with most proteins, the structure informs its function as the above structures and the environment in which it occurs start to indicate how protamine fulfills its role of binding and condensing the sperm DNA. DNA bound to protamine is compact at least 6 times as much as it is in somatic cells resulting in a nucleus ~1/20th the volume.(3, 5) The complex is thought to take on a toroidal doughnut shape each containing 50 kb of DNA with up to ~ 50,000 toroidal unit within a sperm nuclei(8, 25). The exact mechanism of DNA binding and the structure of the resulting complex has, unfortunately, has not yet been completely elucidated but there are several models which attempt to explain the binding mechanism and, in turn, the higher order structures of the protamines.

One model, designed for mammals with only P1 and not accounting for P2, has protamine in an extended conformation with the central arginine stretches half in the minor groove of DNA and partially out and interacting with phosphate groups in the major groove of on an opposite DNA strand through electrostatic interactions and hydrogen bonding. The spacing works out that one protamine segment traverses one turn of DNA with the central arginine neutralizing the majority of the phosphodiester

backbone with some assistance from amino acids in the N-terminal. This slightly distorts DNA from its B structure to a modified B structure seen in protamine complexes. As mentioned above, the terminal ends, being phosphorylated, are repulsed from the DNA strand and play a larger role in created intra and intermolecular disulfide bonds, hydrogen bonds and hydrophobic interactions with other protamine. These intermolecular disulfide bonds connect to protamines occupying the minor grooves up and downstream the original protamine making a larger protamine complex locked around the strand of DNA. The orientation of these protamine and the specifics of intermolecular disulfide bonds are a point of ambiguity in this model with two possibilities given. The first is that the protamine are oriented in an alternating fashion resulting in intermolecular disulfide bonds between the C-termini of adjacent protamines and between the N-termini of adjacent protamines. The second is that intermolecular disulfide bonds occur between protamines bound to different DNA strands which results in protamine that could align in either an alternating or in the same orientation along the DNA strand. Given the first possibility, protamine are then interacting with the major groove of an adjacent strand of DNA, which results in linear arrays structure for the whole DNA-protamine complex.(27) Given the second possibility, this interaction with the major groove still exists but the additionally intermolecular disulfide bonds give the additional framework for more interconnectivity between protamines and DNA strands resulting in a structure potentially more complicated than a linear array. (4) A group studying bull protamine expanded on the previous model, giving more specific position to the disulfide bonds. They had two intramolecular disulfide bonds, one near each terminus, an intermolecular disulfide bond

near the C terminus with protamines on the same strand of DNA and two possible intermolecular disulfide bonds near the N terminus with protamines on different strands of DNA. A major difference in this model is that protamine are largely bound in the major groove, and the intermolecular disulfide bonds connect protamine-bound DNA strands in a hexagonal tightly packed lattice. (20)

Yet another model, based on literature review in 1999 that took into account a wide variety of papers, suggested that DNA was packed in a hexagonal shape (the hexagon being in the x-y plane and DNA helices heading into the z plane). Adjacent DNA helices would be rotated at different angles and in between them would be protamine in a globular structure thus allowing different parts of protamine to interact with differently rotated DNA helices. In this model, the tertiary structure is formed due to hydrophobic interactions and hydrogen bonds with the DNA binding to protamine again coming from electrostatic interaction but with the major groove. (26)

A model proposed in 2006 that accounts for P2 has the 2 protamines in a dimeric β sheet structure with the cysteines of the respective protamines aligned. The interaction between arginines and the phosphate groups of DNA is maintained in this model but the structure of the bound DNA is vastly different as a "straight ladder" structure that more easily accommodates the β sheets.(19)

While there are a fair number of models with consistent mechanics about how P1 binds DNA, the same cannot be said of P2 without serious, unsupported changes to the structure of the DNA. From the sequences, despite similarities there are clear differences which may thwart applying the DNA binding patterns ascribed to P1. For example,

serines and threonines in P2 are seen throughout the sequence whereas P1 has these potential phosphorylation sites on the N and C-termini.(21) Another clear difference is P2's unique role in the binding and coordination of zinc. P2 clearly plays a vital. role as its degradation/knockdown has grave implications on fertility but how it fits into the P1-DNA binding complex is not clear. As such, mapping the disulfide bonds of human protamine, particularly the intermolecular disulfide bonds, should be a good first step in determining how P1 and P2 bind to each other and how protamine binds to DNA.

1.3 Instrumentation Theory

1.3.1 Mass Spectrometry

One of the most common tools to analyze proteins, in a variety of applications such as structural elucidation, quantitation, proteome coverage, differential expression etc. currently, is mass spectrometry. Mass spectrometry has been used in countless experiments including being a foremost tool in mapping the human proteome.(28) While there is a wide diversity of mass spectrometers, they are all divided into the same general parts: the ionization source, the mass analyzer, the ion detector, the computer and the vacuum pumps.

In Electrospray Ionization (ESI), the ionization source used here, the sample enters in solution and passes through a nebulizer, whose function is to turn the solution into an aerosol mist, with a high potential difference applied to it. This results in the spraying of charged droplets; the polarity of the charge is dependent on charge on the needle allowing for a positive and negative ionization mode. Additionally, most ions are

already formed in the solution resulting in the solution composition being vital to ion formation and if/to what extent species are multiply charged.(29) This is particularly key for protein analysis, as proteins, due to their size, can and often will have multiple charge states (Figure 2) allowing for the analysis of proteins with molecular weights outside of the given mass spectrometers mass range. The region in the spray of nebulizer is heated, resulting in evaporative loss of the solvent resulting in shrinkage of these charged droplets. The droplets eventually reach a point in which the surface tension of the droplet cannot sustain the charge it bears (known as the Rayleigh limit) which results in "Coulombic Explosion", in which the droplet explodes into smaller droplets. These smaller droplets undergo the same process until gas-phase analyte ions are all that is left which can then enter the mass analyzer. (30-32)

In the quadrupole type mass analyzer, 4 cylindrical electrodes in a diamond arrangement with electrodes across from each other having the same potential. Radio Frequency (RF) and DC voltage is applied to the electrodes in an alternating fashion resulting in a periodic change in polarity. As an ion enters, its flight pattern is affected by this alternating voltage and its mass to charge ratio (m/z) resulting in only ions of a specific m/z passing through to the next section of the mass spectrometer. (30, 33) A range of m/z values can be scanned, thus showing all the ions from the sample present in that range, or focusing on just one m/z and letting only ions of that m/z through for more targeted analysis.

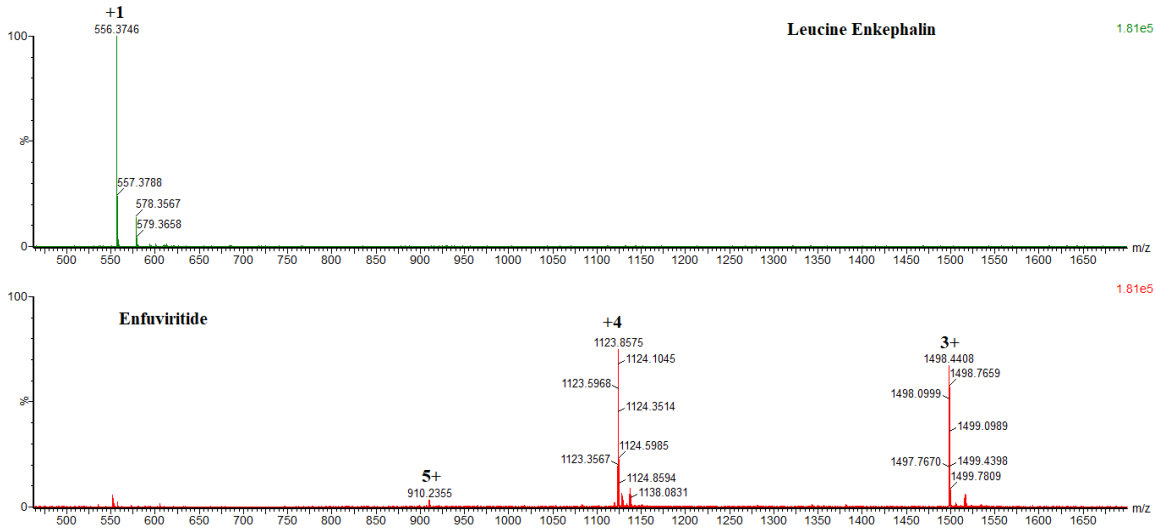

Figure 2. Comparison of spectra of a singly charged vs a multiply charged species. (Top) Mass spectrum of leucine enkephalin (molecular weight = 556.3) (Bottom) Mass spectrum of enfuviritide (molecular weight = 4,492.1) showing three of its charge states. Charges states are noted in bolded text above the peaks.

In a different type of mass analyzer, known as a Time of Flight (TOF), ions are accelerated down a flight tube by an electric field resulting in all ions having the same kinetic energy thus the velocity of each ion is dependent on its m/z. This allows m/z of an ion to be determined based on the time it takes to reach the detector at the end of the flight tube which allows for high mass accuracy and resolution (Figure 2) which can be useful in analysis of proteins as one protein can results in a variety of peaks in a mass spectrum and additional mass accuracy and resolution (Figure 3) can increase confidence in identifying key peaks.(33, 34)

After the mass analyzer, ions head toward a detector, one type of which is an electron multiplier. The sample ions collide with a dynode resulting in secondary emission in which electrons are released. Instead of a single dynode, a series of dynodes is used, causing repeated secondary emissions amplifying the number of electrons exponentially which results in a detectable current. The signal is eventually processed

and presented through a computer as a mass spectrum, plotting m/z against intensity or relative abundance. While not a specific area in the sample path, vacuum pumps are a vital component as the system is under vacuum to prevent any unintended collisions.

The instruments used here are tandem Mass Spectrometers (MS/MS) which means the sample ions enter a mass analyzer but then exit to a collision cell in which fragmentation of the ions can occur by collision-induced dissociation (CID). In CID, the sample ions are accelerated towards a neutral gas such as argon, nitrogen or helium. The charge used to accelerate the ions and the mass of the neutral gas can affect how much the ion is fragmented and can be manipulated to produce more desirable fragmentation. (30, 35) This allows for another level of identification in analyzing the fragmentation pattern as fragmentation happens in a predictable fashion based on structure.(30, 35) After fragmentation, the ions exit the collision cell and enter a second mass analyzer where the fragment ions are sorted by m/z and then exit to a detector as described above. The instruments used in these experiments are triple-quadrupole, meaning the first and second mass analyzers are both quadrupole type, and Quadrupole-Time of Flight (Q-TOF) in which the first mass analyzer is quadrupole type and the second is a TOF.

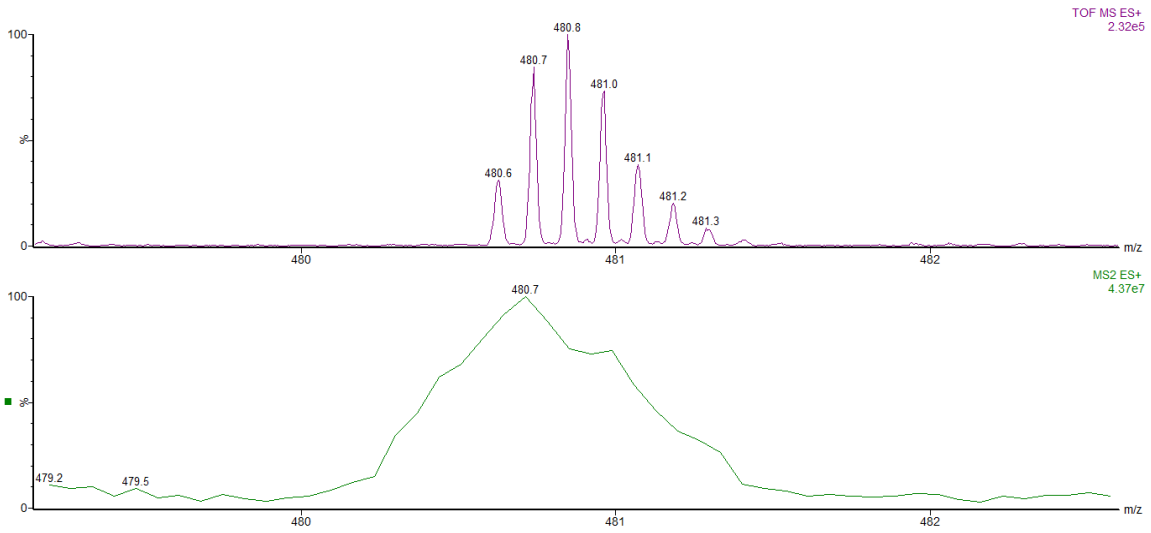

Figure 3. Comparison of resolution between triple-quadrupole and Q-TOF. (Top) scan of one of the chum salmon protamine achieved with a Q-TOF instrument. (Bottom) scan of the same species achieved with a triple quadrupole MS. Each peak of the isotopic distribution now visible due to the enhanced resolution.

1.3.2 Liquid Chromatography

Liquid Chromatography is the most widely used tool to separate out the components of a sample and it is often used in conjunction with mass spectrometry as to prevent complex samples, with a multitude of analytes, from being introduced to the mass spectrometer all at once. Liquid chromatography works on the basis of the sample being carried in the liquid mobile phase being pumped through a column at high pressure. The column contains the stationary phase which consists of silica particles with specific ligands attached; analytes in the mobile phase are attracted to either the stationary phase or mobile phase depending on key characteristics of both. As such, different analytes, will have different affinities to each phase changing the time it takes for each analyte to elute off of the column (retention time) thus separating them. The most common characteristic taken advantage of for this purpose is polarity in which the most common

configuration is “reverse-phase” in which the stationary phase is non-polar, such as with C_{18} or C_8 ligands, and the mobile phase is polar.(35)

Liquid chromatography can also be categorized based on certain parameters of the system such as particle size of the stationary phase and pressures used. High Performance Liquid Chromatography (HPLC) uses 3.5- 5 μm particle size whereas Ultra Performance Liquid Chromatography (UPLC) uses a sub-2 μm particle size resulting in higher pressures used but also shorter analysis times and increased resolution.(36) What was described above is considered a single dimension chromatography setup but with the addition of a second, third or more columns, multi-dimensional chromatography setups can be achieved. A 2D system was used for this project, in a configuration known as “trap and elute” with the first column used as a trapping column. Instead of the sample moving directly into the analytical column, it is first trapped onto a smaller column with a typically larger particle size (≥10 μm). That results in a lower resolution (Rs) and separating power but the purpose of this first column is not to separate, but to trap and focus the analyte at the column head in a narrow tight band. This is achieved by maximizing the affinity of the analyte to the column packing which is done by optimizing the flow rates, chemistry and additives present during loading. The analyte is then pulled off this column with back-flush elution and is then brought to the analytical column for separation as described above.(37)

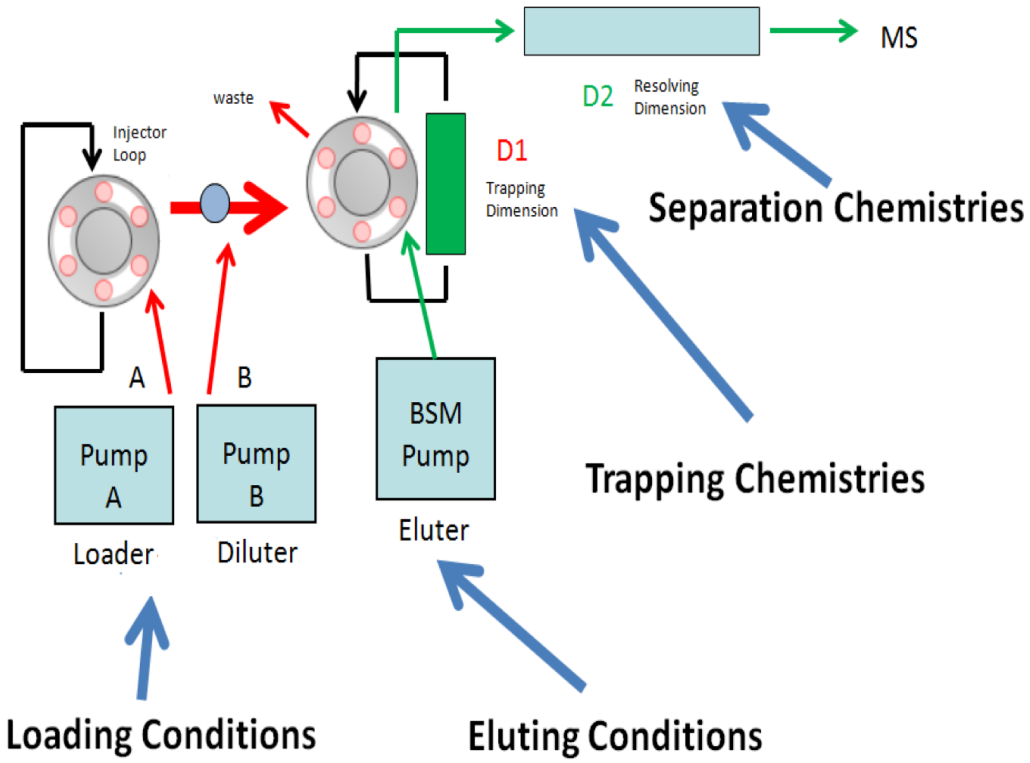

Figure 4. 2D-LC Configuration: Trap and Elute A schematic showing the relationship between the various pumps and columns in the 2d LC configuration used for all the experiments described.

A variation on a typical chromatography setup, made possible by the trapping column and used in these experiments, is At-Column Dilution (ACD) which occurs before the sample enters either column. In ACD, the loading stream carrying the sample, flows through a 3-way mixer where it merges with a diluting stream that is being pumped at a higher flow rate. The flow rate of the loading stream and the diluting stream are set so that the volume injected is diluted based on the ratio of the flow rates. Typically, injecting a sample that is dissolved in 100% organic solvent results in little to no interaction with the column as the analyte has a much greater affinity for the organic solvent it is in. This results in the analyte passing right through the column and is known as break through. The main benefit of this technique is that it allows samples in any

percentage of organic to be injected without breakthrough as the sample can be diluted and the percentage of organic solvent decreased through the ACD, to as much as 5% as seen in past experiments, before trapping on the first dimension column. (37) Being able to inject samples in 100% organic solvent results in less time needed for solvent exchange to prevent breakthrough and results in better conditions for injection and analysis for samples that are not as stable or soluble in aqueous solvents.

1.4 Research Objective

The overall goal of this research is to map the disulfide bonds of human protamine 1 and 2 via LC-MS/MS. The research described here goes about developing and optimizing protocols for the mass spectrometry, liquid chromatography and digestion aspects of this experiment by testing on related proteins (chum salmon(*Oncorhynchus keta)* protamines (of which there are 4) with experiments being verified with the analysis of more well characterized proteins and peptides (Leucine Enkephalin and Enfuviritide). Testing began on chum salmon protamines and enfuviritide due to cost and availability and as a means of benchmarking the developed protocols. Protamine from Chum Salmon will be referred to as P1, P2, P3 and P4, generally referring to their 8+ charge state as a consistent point of reference. (Figure 5)

Oncorhynchus keta (Chum Salmon)

P1 (m/z for the 8+ charge state = 509.1125):

10 20 30
PRRRR ASRRI RRRRR PRVSR RRRRG GRRRR

P2 (m/z for the 8+ charge state = 530.6250):

10 20 30
PRRRR RSSSR PIRRR RRPRA SRRRR RGGRR RR

P3 (m/z for the 8+ charge state = 532.3750):

10 20 30
PRRRR SSSRP VRRRR RPRVS RRRRR RGGRR RR

P4(m/z for the 8+ charge state = 541.0125):

10 20 30
PRRRR SSRRP VRRRR RPRVS RRRRR RGGRR RR

Figure 5. Chum Salmon Protamine Sequence. The sequences, names (as referred to in this paper) and 8+ charge state m/z values for Chum Salmon protamines.

2. MATERIALS AND METHODS

2.1 Standards and Reagents

Acetonitrile (ACN), Methanol (MeOH), Isopropanol (IPA) and Acetone were optima grade and purchased from Fisher Scientific (Waltham, MA, U.S.A.) along with Formic Acid and Ammonium Hydroxide (NH_4OH). The MilliQ grade water was obtained from EMD Millipore Sigma (Darmstadt, Germany). Hydrochloric acid (37%) (HCl), Tris(hyrdoxymethyl)aminomethane, Calcium Chloride ($CaCl_2$), Ammonium Bicarbonate, 1,4-Dithiothreitol (DTT), USP testing grade Protamine Sulfate, Enfuviritide acetate salt, Leucine Enkephalin acetate salt hydrate, Type I-S α Chymotrypsin from bovine pancreas, sequencing grade Endoproteinase Lys-C from *Lysobacter enzymogenes* and TPCK treated Trypsin from bovine pancreas were obtained from Sigma-Aldrich (Steinheim, Germany). Sequencing grade endoproteinase Arg-C from *Clostridium histolyticum* was obtained from Worthington Biochemical Corporation (Lakewood, NJ, U.S.A.).

2.2 Instrumentation and Software

An ACQUITY UPLC ® (Waters Corporation, Milford, MA, U.S.A.) was used in a 2D "trap and elute" configuration with ACD. This was achieved by using a Binary Solvent Manager (BSM) pump for loading samples (the loader) and another BSM pump for creating the dilution stream (the diluter) both of which flow water with the major variation being in pH. The diluter and the loader both are connected to a 50 µL mixer where the ACD took place which then lead to the 1st dimension trapping column. A final BSM pump (the elution pump) was used for pulling compounds off of the trap column to

the second dimension column and for creating the gradient for subsequent elution off of the second dimension column. This change in flow paths is controlled by the column manager. In a given experiment, three minutes is given for loading the sample onto the trap column in which the flow path of the sample only involves the loader pump, the diluter pump and the trap column. After three minutes, a valve switch occurs and the flow of the elution pump now includes the trap and pulls the trapped compounds off with backflush elution towards the analytical column. The second dimension column elution flows then to a mass spectrometer for detection. A Xevo TQS and a Synapt G2-S Q-TOF (Waters Corporation, Milford, MA, U.S.A.) with ESI in positive ion mode were used for detection.

MassLynx © version 4.1 (Waters Corporation, Milford, MA, U.S.A.) software was used to analyze all chromatograms and spectra along with controlling the entire LC-MS system. Biolynx, Peptide Editor and Protein Sequencer, were used in analysis of specific chromatograms and spectra, specifically in predicting or matching digest fragments or peptide MS/MS fragment ions and predicting sequences based off of those ions.

2.3 Liquid Chromatography Columns

All columns used were from Waters Corporation, Milford, MA, USA. The analytical columns assessed include: C18 2.1 x 100 mm, 1.7 µm, 300 Å, C18 2.1 x 150 mm, 1.7 µm, 300 Å and C4 2.1 x 100 mm, 1.7 µm, 300 Å. The trap columns tested include: C18 2.1 x 5, 1.7 µm, 300 Å, C4 2.1 x 5, 1.7 µm, 300 Å, Hydrophilic Lipophilic

Balanced (referred to as Polymer), 2.1 x 30 mm, 20 µm, C18 2.1 x 30, 10 µm and C8 2.1 x 30 , 10 µm. These columns will be referred to by their stationary phase, internal diameter (i.d.) and length.

2.4 Methods

2.4.1 Preparation of Stock Solutions

Protamine Sulfate and Leucine Enkephalin were each dissolved in water and Enfuviritide was dissolved in a 50/50 water/methanol solution. All protein stock solutions were made at 1 mg/mL and stored at 5 C. Chymotrypsin and trypsin were dissolved in 1 mM HCl, Lys-C was dissolved in water and Arg-C was dissolved in 50 mM tris HCl (100 mM) $CaCl_2$. All proteases were made at a final concentration of 10 ug/mL and stored at -20 C. These stock solutions were used to create the lower concentration amounts seen throughout the following experiments.

2.4.2 Infusion

Solutions of the four proteins/peptides, leucine enkephalin, enfuviritide and chum salmon protamines were prepared at 5 ug/mL, 1 ug/mL, 100 ng/mL and 10 ng/mL in water. At 5 ug/mL, 1 ug/mL and 100 ng/mL, additional solutions were prepared with the addition of 0.2% formic acid or 0.2 % NH_4OH. These solutions were then directly infused into the Synapt G2-S Q-TOF in scan mode. The resulting signal was compared between sensitivity and resolution mode and combined (combining a flow of 0.2 mL/min (95% H_2O .5% FA, 5% ACN .5% FA) from one the elution pump) vs no combined flow rate. The intensity of the signal was noted along with the charge states seen if applicable

along with charge states and masses seen for chum salmon protamine being compared to references. (38, 39)

2.4.3 2D-LC Optimization

Initial evaluations started with using no column on the second dimension, the elution pump having mobile phase A be water and mobile phase B be acetonitrile and the Xevo TQS as a detector being the static parameters. The LC program, which also remained the same for this portion, started at 5% B for the first 3 minutes (loading the compounds onto the trap), then ramped up to 5% B over 1 minute, stayed at 95% B for 1 minute and then immediately dropped back down to 5% B and remained there for 1 minutes resulting in a final run time of 6 minutes. The parameters considered in this evaluation included the pH of the flow of the loader and diluter (water with 2% formic acid vs water no addition vs water with 2% NH_4OH), the trap columns (C18 2.1 x 5 mm vs C4 2.1 x 5 mm), the pH of both mobile phase A and B of the elution pump (addition of 0.5% formic acid vs 0.5% NH_4OH) and with ACD vs without ACD. ACD is programmed by having the diluter flowing at 2 mL/min while the loader flows at .1 mL/min resulting in a 20 x dilution factor. To program ACD as off, the diluter was set to .1 mL/min while the loader was set to 2 mL/min. All possible combinations of these parameters were tested on chum salmon protamine and leucine enkephalin which was used as a benchmark to assess that the system was functioning properly. Whether the four chum salmon protamines were present, their signal and present charge states were noted. After the initial assessment of all possible combinations, the combinations that worked were rerun with the Synapt G2-S

Q-TOF as the detector with the same criteria noted.

Further parameters for the LC portion were optimized while keeping the trap column, loader and diluter pH, elution solvent pH and ACD constant as determined by previous evaluations. All experiments analyzed protamine sulfate at 5 µg/mL, 1 µg/mL and 100 ng/mL along with leucine enkephalin and enfuviritide at the same concentrations as benchmarkers to assess how the system was functioning. This initial optimization was performed on intact protein and would have to be later verified on digested protein fragments. All experiments used C18 2.1 x 5 mm for the trapping column, water with 2% formic acid for the loader and diluter flows, ACD on, and 0.5% formic acid to the elution pump mobile phase. The peak shape, signal and separation of the four chum salmon protamines were noted in the chromatograms and spectra. Columns tested included the C18 2.1 x 100 mm, the C4 2.1 x 100 mm and the C18 2.1 x 150 mm. Mobile phases tested included ACN, MeOH, IPA and Acetone. Flowrates on the elution pump tested included 0.5 mL/min - 0.1 mL/min in increments of 0.1. Parameters were assessed individually. The final gradient chosen was 5% B for the first three minutes (loading), then remaining at 5% for another minute, ramping up to 50 % B over two minutes, ramping up to 60% B in one minute, holding at 60% B for one minute and then dropping down to 5% B in 0.5 minutes and staying there for another 0.5 minutes resulting in a total run time of 9 minutes. The final LC method which consisted of pH 3 water (2% addition formic acid) for the loader and diluter streams flowing at .1 mL/min and 2 mL/min respectively for a 5% ACD and water and IPA with .5% formic acid for mobile phases A and B for the elution pump flowing at .2 mL/min. The trapping column used was the

C18 2.1 x 5 mm while the second dimension column was the C18 2.1 x 100 mm with the Synapt G2-S Q-TOF as the detector.

2.4.4 Enzymatic Digest Optimization

Digests were optimized for a combination of Lys-C and chymotrypsin as tested on enfuviritide, trypsin as tested on chum salmon protamine and Arg-C as tested on chum salmon protamine. Due to the similarity cut sites and resulting fragments, Trypsin and Arg-C digests were evaluated contemporaneously. Both were evaluated as their cut sites provide slightly different fragments, there have been experiments using the data from multiple digests for enhanced coverage (40, 41) and they have different optimal pH ranges which could play a role in sample preparation. These specific proteases were chosen due to the ability of the resulting digest fragments, when applied to human protamine P1 and P2, to isolate the majority of cysteines (Figure 6). Enfuviritide is used here as a benchmarker as enfuviritide digested with chymotrypsin has been previously characterized (42); enfuviritide was also used as neither chymotrypsin nor lys-C have cut sites on any of the 4 chum salmon protamines. To initially establish that the digest is working at all, initial digests are set up with parameters as influenced by what is seen in the literature and then, once confirmed and fragments identified, performed again with a range of parameters that are systematically tested.

Initial digests were done at a 1:50 mass ratio (enzyme:substrate), incubated at 36 C for 18 hours, quenched with 2% formic acid and analyzed with the same LC method as described in the 2nd phase of LC validation. Chymotrypsin digest of enfuviritide was

initially done in 100 mM Tris HCl/ 2 mM $CaCl_2$ buffer (pH 8.6) and 50 mM ammonium bicarbonate buffer with a starting concentration of 10 ug/mL in 1 mL which was diluted with water to 1 ug/mL with 250 µL being injected. As Lys-C digestion was to occur with chymotrypsin digestion, the initial parameters for Lys-C were determined based on chymotrypsins optimized conditions. Trypsin and Arg-C digests were initially done in 50 mM ammonium bicarbonate and 50 mM tris-HCl, 5 mM $CaCl_2$ (pH 7.7) with 50 mM Dithiothreitol (DTT) respectively at 50 µg/mL in 200 µL of buffer and diluted to 5 µg/mL with 250 µL injected.

Each digest was also run alongside a negative or blank in which the same reaction was made but without the protein to be digested which could be used as a baseline to identify peaks unique to the digested protein. Additionally, an "enzyme negative", in which the same reaction was prepared but without the protease, was also prepared with each digest to have a reference for the signal of the undigested protein being treated to these conditions.

Lys-C and Chymotrypsin Digests (Chymotrypsin primary cuts: FYW, secondary cuts: MLH; Lys-C primary cuts: K)

Enfuviritide:

Y | TSL | IH | SL | IEESQNQQEK | NEQEL | L | EL | DK | W | ASL | W | NW | F

Human P1:

ARY | RCCRSQSRSRY | Y | RQRQRSRRRRRRSCQTRRRAM | RCCRPRY | RPRCRRH

Human P2:

RTH | GQSH | Y | RRRH | CSRRRL | H | RIH | RRQH | RSCRRRK | RRSCRH | RRRH | RRGCRTRK | RTCRRH

Trypsin and Arg-C Digests (Trypsin primary cuts: RK; Arg-C primary cuts: R)

Chum Salmon P1

Arg-C: PR|R|R|R|ASR|R|IR|R|R|R|R|PR|VSR|R|R|R|R|GGR|R|R|R
Trypsin: PR | RRR | ASR | R | IR | RRRRPR | VSR | RRRR | GGR | RRR

Chum Salmon P2

Arg-C: PR|R|R|R|R|SSSR|PIR|R|R|R|R|PR|ASR|R|R|R|R|GGR|R|R|R
Trypsin: PR | RRRR | SSSRPIR | RRRRPR | ASR | RRRR | GGR | RRR

Chum Salmon P3

Arg-C: PR|R|R|R|SSSR|PVR|R|R|R|R|PR|VSR|R|R|R|R|R|GGR|R|R|R
Trypsin: PR | RRR | SSSRPVR | RRRRPRV | SR | RRRRR | GGR | RRR

Chum Salmon P4

Arg-C: PR|R|R|R|SSR|R|PVR|R|R|R|R|PR|VSR|R|R|R|R|R|GGR|R|R|R
Trypsin: PR | RRR | SSR | RPVR | RRRRPR | VSR | RRRRR | GGR | RRR

Human P1

Arg-C: AR|YR|CCR|SQSR|SR|YYR|QR|QR|SR|R|R|R|R|R|SCQTR|R|R|AMR|CCR|PR|YR|PR|CR|R|H
Trypsin: AR | YR | CCR | SQSR | SR | YYR | QR | QR | SR | RRRRR | SCQTR | RR | AMR | CCRPR | YRPR | CR | RH

Human P2

Arg-C:
R|THGQSHYR|R|R|HCSR|R|R|LHR|IHR|R|QHR|SCR|R|R|KR|R|SCR|HR|R|R|HR|R|GCR|TR|KR|TCR|R|H
Trypsin: R | THGQSHYR | RRHCSR | RR | LHR | IHR | R | QHR | SCR | RR | K | R | R | SCR | HR | RRHR | R | GCR | TR | K | R | TCR | RH

Figure 6. Sequences and predicted cut sites of all digests. Predicted cut sites for all digests performed and for human P1 and P2 as generated by Expasy Peptide Cutter. Cysteines are highlighted in yellow. While basic rules for the cut sites of each protease are given, exceptions are not stated (but reflected in cut site choices). Not all exceptions are absolute, some are drastic decreases in occurrence.

Establishing that a digest worked involved comparing the digest chromatogram to the enzyme negative to show that the intact protein was no longer present, and then

identification of peaks seen in the digest chromatogram not seen in the negative. The spectra of these peaks were then examined and the masses correlated to a reasonable sequence and charge state. Lists of possible fragments in multiple charge states were generated, allowing for up to 4 missed cut sites and using charge filter (checks to see if the charge state is reasonable for the fragment) The peptide editor was also used to match fragments in spectra with each identification and spectra being reviewed manually. Select, high intensity peaks of varying length and various cuts/missed cuts were then chosen and used as markers for assessing the optimization of the digest with varying parameters.

Upon establishing a digest worked, the digest was then run varying the parameters of ratio of enzyme:substrate (1:50, 1:100, 1:200) and time of incubation (based off literature for each enzyme). For these, the enfuviritide digests were conducted at 50 µg/mL and diluted down to 5 µg/mL for a 250 µL injection while the chum salmon protamine digests were conducted at 100 µg/mL and diluted down to 10 µg/mL for a 250 µL injection. The digests were assessed on the intensity of the marker peaks, any other peaks present and comparison of intensity of fragments that require missed cut sites or partial digestion to fragments that require complete digestion.

2.4 Optimization of Liquid Chromatography Parameters for Digest Fragments

Due to low signal intensity on protamine sulfate digests, the smaller mass of the digest fragments allowing for a wider variety of trap columns to be used and the chemical similarity of the chum salmon protamine digest fragments and the human protamine

digest fragments, further optimization of the LC method was performed on Arg-C and trypsin fragments of chum salmon protamines prepared as described above except at a 1:100, enzyme:substrate ratio and diluted to a final concentration of 10 µg/mL for a 50 µL injection. This involved using the previously optimized LC method, with both ACN and IPA low pH elution, but varying the trap columns and the pH of the loading and diluting streams. Trap columns tested included: Polymer 2.1 x 30, C8 2.1 x 30, and C18 2.1 x 30. The loading streams were either water no addition, water with 2% formic acid or water with 2% NH_4OH. Chromatograms and mass spectra were evaluated as previously by looking for signal intensity, peak shape of previously identified fragments and the diversity in fragments seen. Upon finding a suitable method that produced higher signal, digest optimization was performed as described above with trypsin and Arg-C (no DTT) spanning across incubation times of 4, 18 and 24 hours and enzyme:substrate ratios of 1:50, 1:100 and 1:200. The digests were made in 200 µL of the appropriate buffer at an initial concentration of 100 µg/mL, quenched with 2% formic acid and diluted to 10 µg/mL for a 50 µL injection to be analyzed by LC-MS/MS. LC analysis was performed using water with 2% NH_4OH as the loading and diluting mobile phase flowing at 0.1 and 2 mL/min respectively, water and acetonitrile with 0.5% formic acid as the elution mobile phase flowing at 0.2 mL/min, the Polymer 2.1 x 30 mm as the trapping column, and the C18 2.1 x 100 mm as the second dimension column.

3. RESULTS/DISCUSSION

3.1 Infusion

Infusion of chum salmon protamines produced wildly varying results across the three pH's (Table 1). At pH 3, chum salmon protamine produces a mass spectra consistent with previous publications with charge states seen between 4+ and 12+. (Figure 7) Additionally, ratios of protamines were roughly similar to what was previously reported with P1, P2 and P4 being in a similar amount with P4 the highest, and P3 notably less.

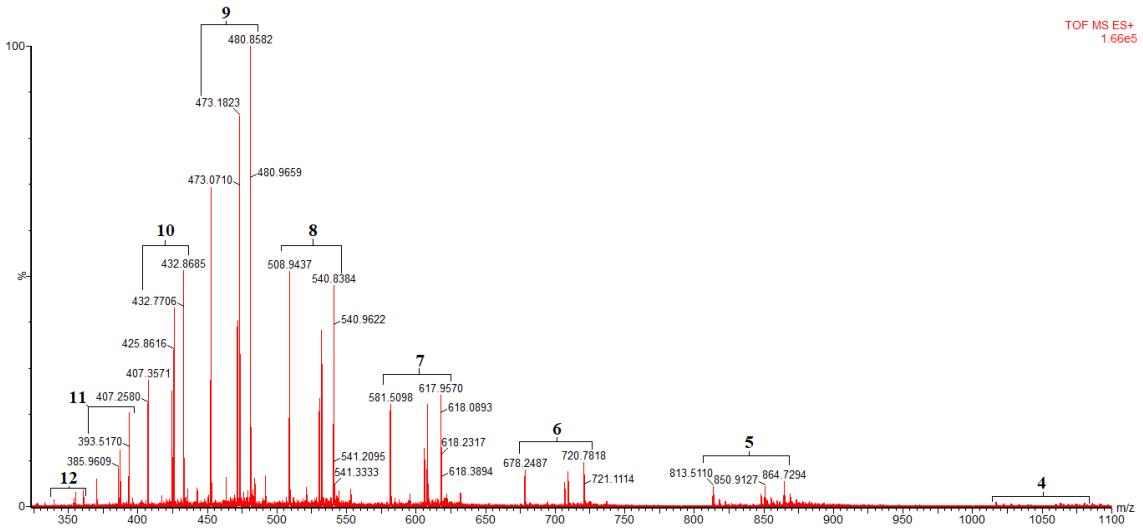

Figure 7. Scans of infusion of Chum Salmon Protamine at pH 3. A scan of an infusion of chum salmon protamine at 5 µg/mL with an addition of 0.2% formic acid (pH 3). The charge states seen (4-12) are labeled.

Additionally, at pH 3, peaks corresponding to chum salmon protamine were seen as low as 10 ng/mL in concentration.

At pH 7, the expected peaks for the various charge states of chum salmon protamine are not seen, however a multitude of other peaks are seen in the same mass range which are most likely sulfate adducts. At pH 10, no signal that could be linked to

chum salmon protamine was seen. Analysis with a combined flow rate for pH 3 showed nothing novel except for a decreased signal.

Considering all the infusion results indicates analysis of chum salmon protamine should be conducted under low pH for the highest signal and the most easily interpretable spectrum.

3.2 2D-LC Optimization

3.2.1 2D LC Optimization of Trapping Conditions

Resulting chromatograms were evaluated assessing peak shape, retention time and signal. Spectra were evaluated by checking for the 8+ charge state of chum salmon protamine. If not present, other charge states were checked and other high intensity peaks in the chromatogram investigated. The chromatographic signal for P1 and the signal intensity of the related mass spectra were used to compare the various methods.

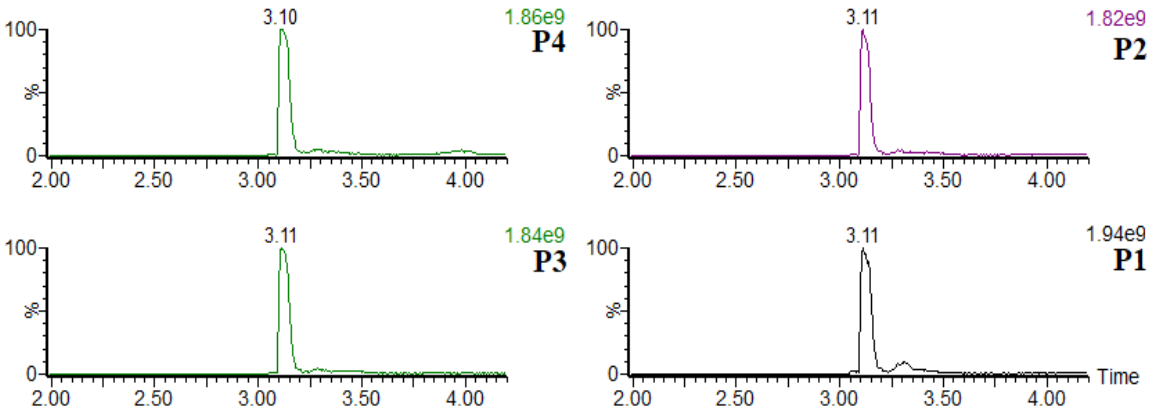

Figure 8. Chromatograms of P1-4 for ACD on, low pH elution, pH 3 load on C18 trap.
Chromatograms of the four chum salmon protamine. The number in the top right of each chromatogram is the signal intensity and which protamine species is shown.

Upon evaluation with ACD off, the majority of methods that produced signal for chum salmon protamine used low pH elution with either low pH or neutral loading with the C18 2.1 x 5 mm trap producing higher signal than the C4 2.1 x 5 mm trap. Across the board, high pH loading resulted in loss of the compounds and high pH elution causes loss of the compounds in all except low pH loading with the C18 2.1 x 5 mm trap. In this 2D setup, this could be caused by the compounds, not trapping, resulting in the compounds flowing straight to waste or the elution stream not pulling the compound off efficiently, resulting in no notable signal.

The highest intensity signal was produced via C18 2.1 x 5 mm trap, low pH elution and loading. Additionally, ACD on, consistently produced higher intensity signal for all methods (Figure 10) it worked for but it did result in 2 less methods working (pH 3 loading, pH 3 elution, C4 2.1 x 5 mm trap and pH 3 loading, pH 10 elution, C18 2.1x 5 mm trap). This trend of ACD producing higher signal was also seen with benchmarker leucine enkephalin. As such, experiments conducted after this experiment largely used low pH loading and elution, with the C18 2.1 x 5 trap column and ACD on.

Additionally, the use of the trap columns conferred a desalting effect to the analysis. In trapping compounds on the first dimension, ions from the salts, in this case most notably sulfate from the salt preparation of protamine, flow directly to waste in the first three-minute loading period of the LC program.

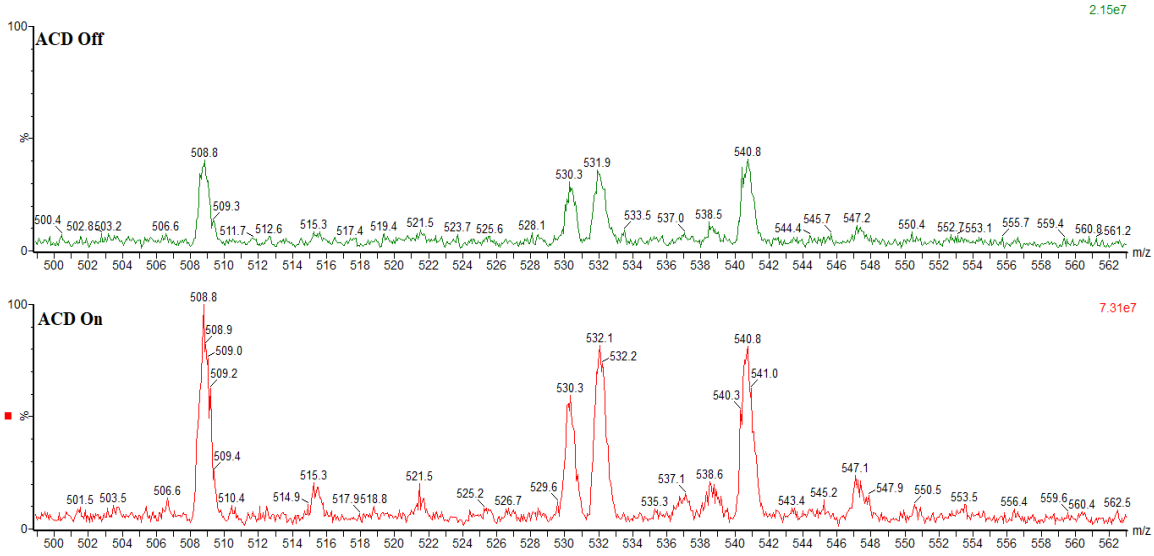

Figure 9. Comparison of ACD on vs ACD off. (Top) Spectra of chum salmon protamine (P1-4) from pH 7 loading, C18 trap and pH 3 elution with ACD off. (Bottom) Spectra of same species and conditions but with ACD on and the vertical axis on the same scale as the top spectra. Signal intensity is noted in top right hand corner.

Comparison of the mass spectrum resulting from the infusion of the chum salmon protamine on a Q-TOF mass spectrometer as compared to a mass spectrum from this evaluation shows the loss of sulfate adduct peaks (Figure 11), which in the 8+ charge state, are peaks with an m/z 12 greater than that of the chum salmon protamine peaks.

This confirmed desalting will have important implications in creating the final workflow in purifying human protamine from sperm and making an extract that is mass spectrometry amenable while still useful to the application.

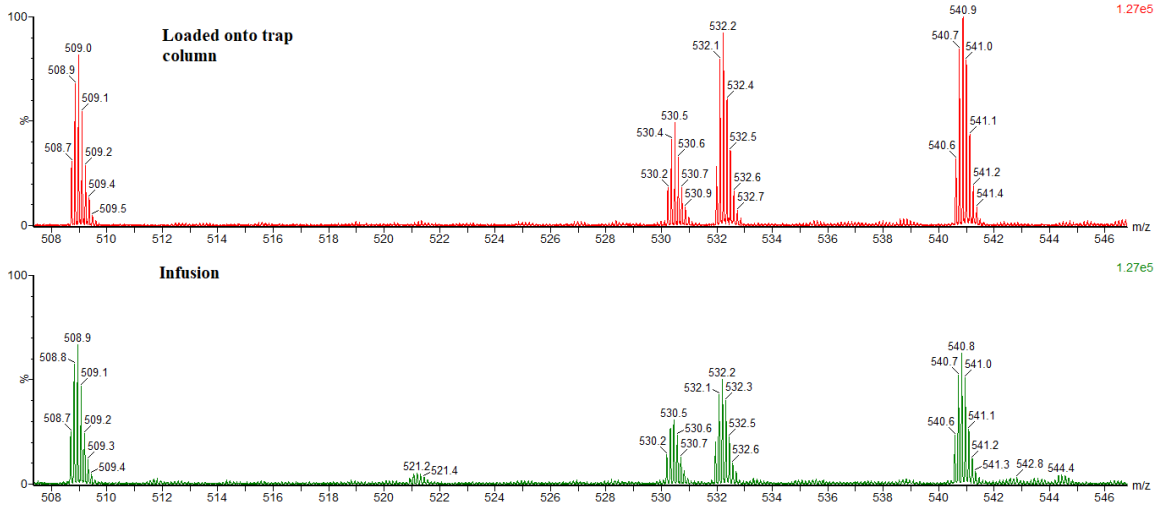

Figure 10. Comparison of spectra from ACD on, low pH elution, pH3 load on C18 trap vs infusion at low pH. (Top) Spectra of chum salmon protamine loaded onto trap column and then analyzed by MS. (Bottom): Infusion of chum salmon protamine. Note the sulfate adduct peaks (521.2 (12 + P1), 542.8 (12 + P2) and 544.4 (12 + P3)) disappear when using the trap column.

3.2.2 2D LC Optimization of Separation Condition

Each of these parameters were evaluated by analyzing the resulting chromatograms and mass spectra for presence of the four chum salmon protamines, with focus given to the 8+ charge state for consistency. For each of the four chum salmon protamines, chromatographic signal, mass spectrum signal (for the highest intensity charge state), separation and peak shape were taken into account when comparing across the various parameters. Chromatograms and spectra of enfuviritide and leucine enkephalin were also analyzed as standards to ensure the validity of the experiment and assess the health of the system. The "starting point" for these parameters was using H2O and Acetonitrile with 0.5% formic acid for the elution mobile phase, flowing at 0.5 mL/min using the C18 2.1 x 100 mm as the 2nd dimension column.

For evaluation of the analytical column, different columns were tested as

mentioned above in the methods section while using H_2O and ACN, both with 0.5% formic acid, as the elution mobile phase.

Across the three columns, variations were minimal, especially with regards to signal. The biggest difference was in retention time (Figure 11), with the C4 column having the shortest, and separation of the four protamines, again with C4 providing the most separation. That being said, the protamines are not at all resolved by any column and the "increased separation" provided by the C4 column is in the range of hundredths of a minute. Given the lack of notable variation and the extremely short retention time, chum salmon protamine seems to have minimal interaction with either stationary phase, C4 or C18. Given no significant benefit, further experiments were run on the C18 2.1x100 as that has been the starting and comparison point for most other experiments before this one so most prior data had been generated with this column.

The various elution mobile phases were also tested as mentioned above except that, due to the high backpressures created by MeOH and IPA, those elution mobile phases were tested with a flow rate of 0.3 mL/min. Overall, varying the elution mobile phase had a much more notable effect with regards to both signal intensity and peak shape.

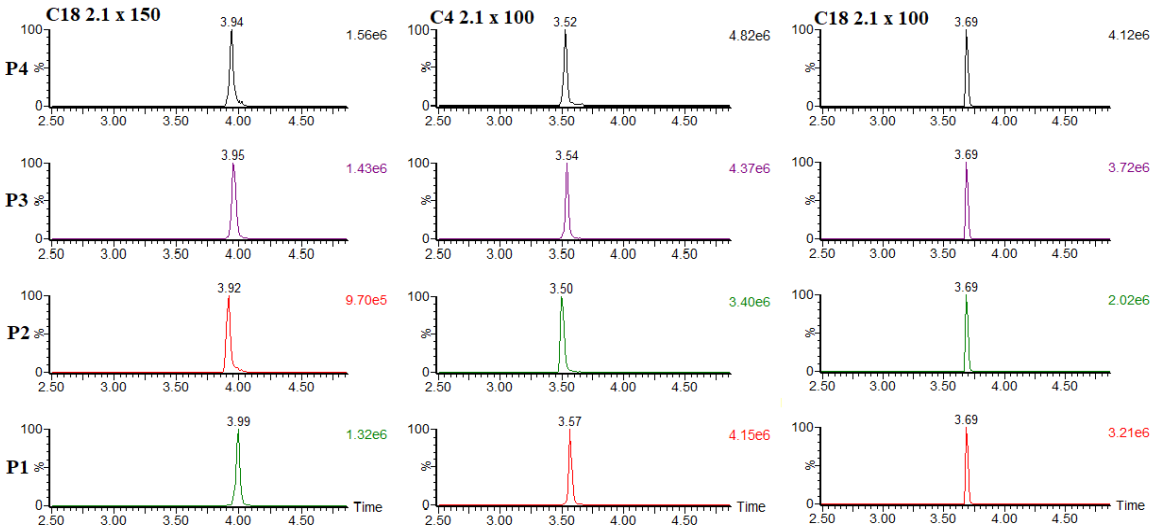

Figure 11. Chromatograms of Chum Salmon Protamine produced by various analytical columns. Bold text at top of each column indicates which column used. Bold text at beginning of each row indicates which protamine is shown.

In terms of signal intensity, at 5 µg/mL all four solvents produce similar signal; at 100 ng/mL, however, MeOH produced the lowest signal intensity, resulting in no significant signal being seen unlike the other three solvents. (Figure 13) Acetone produced the highest chromatographic signal, however it resulted in a slightly distorted peak shape with a low left-hand shoulder.

In terms of separation of the four chum salmon protamines, results, again, are minimal. MeOH gives the most separation with the greatest between species being 0.04 minutes so these differences are negligible. Given these results, IPA and ACN produced the highest signals while maintain Gaussian peak shape. However, IPA evaluations were conducted at a lower flow rate. This result was taken into consideration in addition to the flow rate testing evaluations to produce a more even comparison.

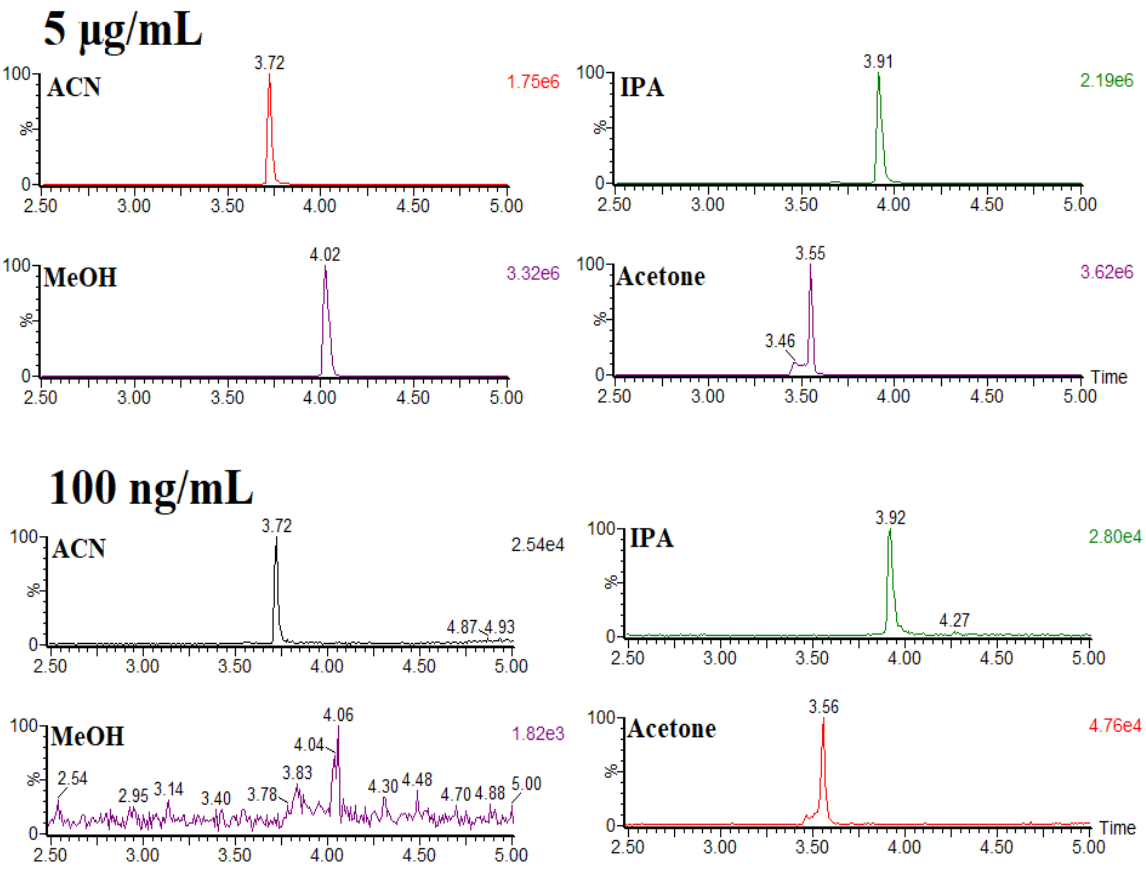

Figure 12. Comparison of chromatograms of P1 produced by different elution solvents. Top four chromatograms are at a concentration of 5 µg/mL and bottom four at 100 ng/mL. For each chromatogram, the elution solvent is written in the top left corner and the signal intensity in the top right.

All flow rate evaluations were run using acetonitrile and H2O with 0.5 % formic acid as the elution solvents. Changing the flow rate produced a notable change in both signal and separation of chum salmon protamines.

There is a notable trend in decreasing the flow rate resulting in a higher signal (Figure 13) which holds true for P1-P4. This comes at the price of wider peaks and also eventually poor peak shape as the solvent at a low flow rate does a less efficient job of pulling the compound off of the trap efficiently in a tight band. As such, this trend stops at 0.2 mL/min, with 0.1 mL/min being too broad resulting in a lower chromatographic

signal. This notable broadening and degradation of peak shape is seen, more so at 100 ng/mL with 0.1 mL/min; the extent of degradation and broadening varies across P1-P4 with P3 faring the worst showing distorted peak shape across several lower flow rates.

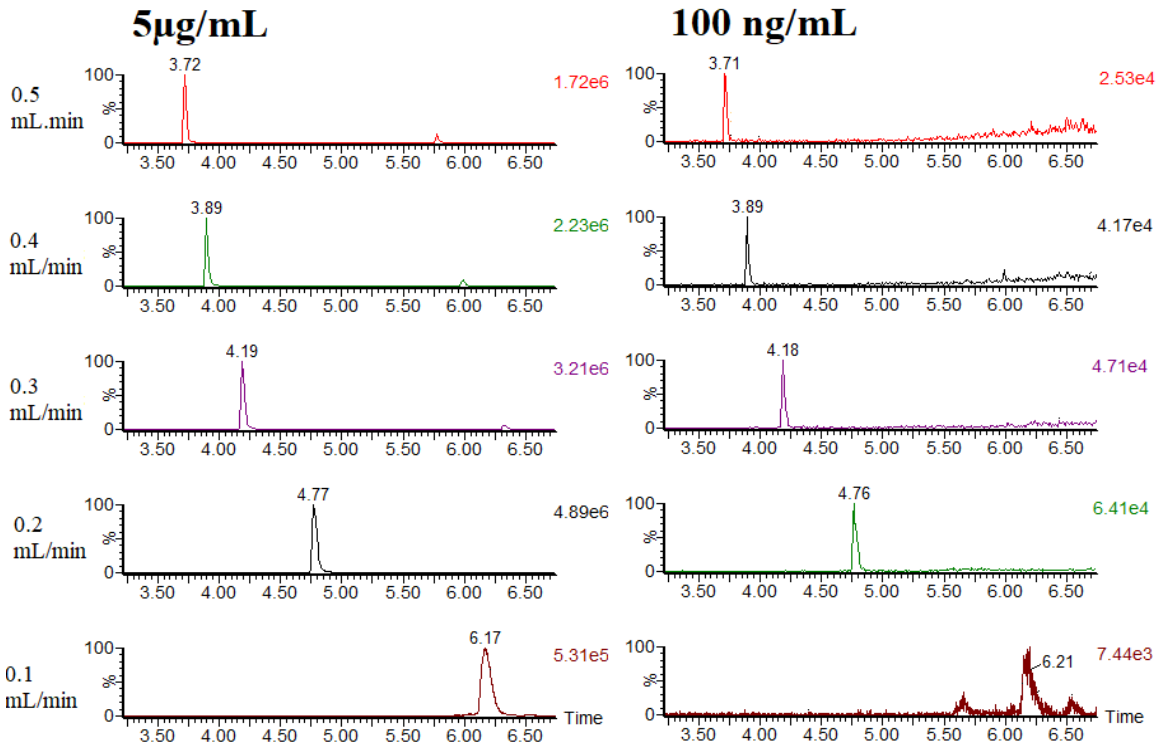

Figure 13. Comparison of chromatograms of P3 across flow rates. The bolded text on top indicates concentration, text on the left hand side indicates flow rate and text on the top right of each chromatogram is for signal intensity.

In terms of separation, flow rate, in comparison to the other two variables tested, has the greatest effect. Notable separation is achieved but it is only seen starting at 0.2 mL/min with 0.3 mL/min resulting in a max separation of 0.08 minutes. From there, however, there is a vast improvement in separation for each increment of 0.1 mL/min to the flow rate with 0.2 mL/min having a max separation of 0.23 minutes and 0.1 mL/min having a max separation of 0.51 minutes. However, due to peak broadening, and different

amounts of separation between the four protamine species, as flow rate is lowered, this does not result in resolving all four species of protamine.

P2 shows the most separation from the other species; as the flowrate is lowered, P2 can be seen becoming more and more resolved in the total ion chromatogram (TIC), which shows all chum salmon protamines at once, (Figure 14), with P2 almost achieving a resolution (Rs) of 1.5 (returning to baseline) at 0.2 mL/min with the other protamines.

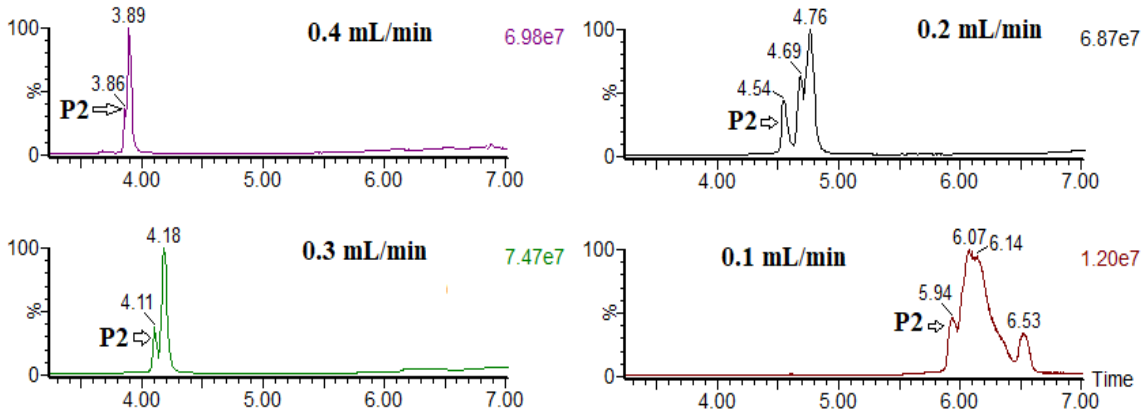

Figure 14. Comparison of flowrates effect on separation of P2 from other chum salmon protamine species. Each chromatogram has the flow rate noted, and P2 pointed out with an arrow.

Given the resolution achieved, and intensity of signal, 0.2 mL/min was chosen for future experiments. While there was a trend for species specific broadening and distorted peak shape at lower concentration, it was not expected future work for this application would require analysis at lower concentrations, the final method will be intended for digest fragments and not the intact protein and peak shape at 5 µg/mL remained Gaussian for all four species.

In light of the previously described evaluations, the column chosen to move forward with was the C18 2.1 x 100 mm. However, the results from mobile phase and

flow rate were not as clear cut. Given the data, ACN with 0.2 mL/min flow rate was chosen but given the high signal intensity achieved with IPA, this prompted investigation into using IPA as elution mobile phase with a lower flow rate.

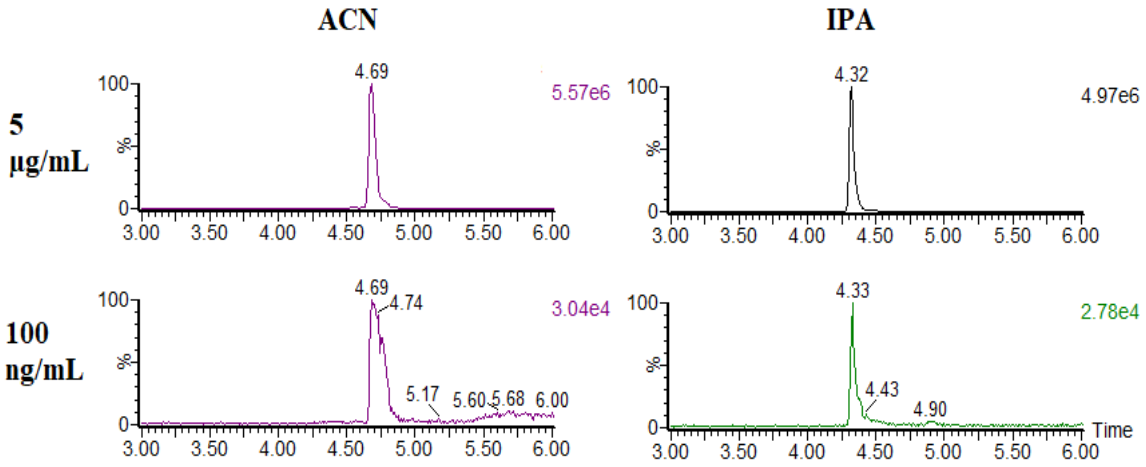

Figure 15. Comparison of chromatograms of P4 from ACN and IPA elution at 0.2 mL/min. Text at top indicates elution mobile phase, text on the left side indicates concentration and text in right hand corner of each chromatogram is for signal.

In this comparison, they produced similar signals with ACN producing a slightly higher signal (Figure 15) and producing notably more separation across protamine species. However, at 100 ng/mL, IPA elution maintains a more Gaussian peak shape for all 4 species whereas ACN has peak shape distortion for some protamine species at low concentrations. Given these results, both elution mobile phases were considered for further testing on digest fragments which would then hopefully provide a clearer answer on which would be most suited for this application.

3.4 Digest Optimization

3.4.1 Chymotrypsin + Lys-C digest of Enfuviritide

Initial tests to optimize simultaneous chymotrypsin and Lys-C digestion began with optimizing a digest protocol only using chymotrypsin. These tests assessed enzyme: substrate ratio at 1:50, 1:100 and 1:200 by weight, incubation time at 36 C for 1,2 18 and 24 hours and digest buffer being 50 mM ammonium bicarbonate or 50 mM tris-HCl, 5 mM $CaCl_2$ (pH 7.7).

Digest fragments were identified based on previous literature results, predicted partial and complete digest fragments and investigation of prominent unique peaks in the sample chromatogram. Fragments were identified based on their mass and charge state (Figure 16). This resulted in a list of fragments (Table 1) found across the digests. Note that in calculation of some of these mass fragments, the N-term of enfuviritide is acetylated and the C-term is amidated. A group of those fragments were used as a point of comparison to compare the efficacy of the different ratios for digestion. A more complete digestion would be indicated by a high signal intensity seen in peaks A, C, D, E and G and a lower signal intensity in the remaining peaks as the first group of peaks result from more cuts and the remaining peaks result from more missed cuts.

These assessments showed several notable trends across the various parameters. Using ammonium bicarbonate or a low time of incubation (1-2 hours) produced lower signal for certain low missed cut site fragments such as F or G (Figure 17) while showing a higher intensity signal for certain high missed cut site fragments indicating a more partial digest.

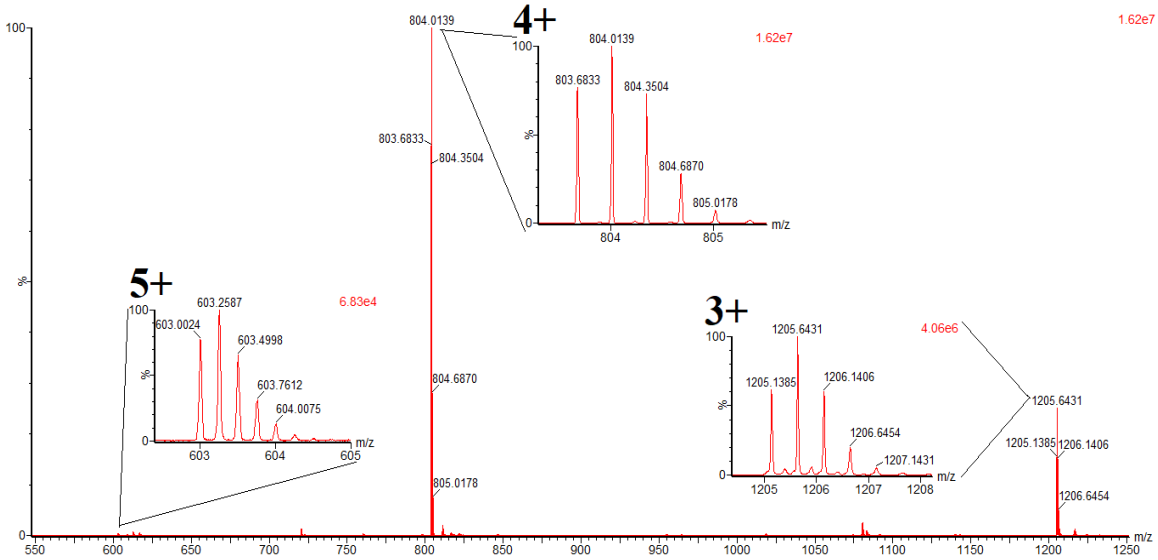

Figure 16. Spectra of Chymotrypysin digest fragment F. Text at top left indicates signal intensity. Each charge state is labeled and has the spectra zoomed in next to it.

Table 1. Fragments identified from Chymotrypsin digest of Enfuviritide. "Ac-" refers to N-term acetylation. "-Am" refers to C-term amidation. X(Y') indicates that fragment X is a part of fragment Y. Z (X+Y) indicates that fragment Z is a combination of fragments X and Y.

Name	Sequence	MW (Theoretical)	MW (seen)	Charge States seen	Missed Cut Sites	Retention Time (min)
Intact	Ac-YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF-Am	4490.1707	1123.2941 1497.4082 899.0463	3, 4, 5	x	6.48
A	Ac-YTSL	524.2482	525.2292	1	1	5.55
B	IHSLIEESQNQQEKNEQELLELDKW	3079.5098	1541.1372 1027.4857 770.8635 616.8701	2,3,4,5	5	5.68
C	ASLW	475.2431	476.2285	1	1	5.64
D	NWF	464.2012	465.1823	1	1	5.62
F (B')	IHSLIEESQNQQEKNEQELL	2408.1892	1205.1385 803.6833 603.0024	2,3,4	3	5.50
G (B')	ELDKW	689.3384	690.2927 345.6404	1,2	1	5.32
H (B')	SLIEESQNQQEKNEQELLELDKW	2829.3668	1415.9023 944.0806 708.3010	2,3,4	4	5.74

Name	Sequence	MW (Theoretical)	MW (seen)	Charge States seen	Missed Cut Sites	Retention Time (min)
I (A + F)	Ac-YTSL IHSLIEESQNQQEKN EQELL	2914.4196	1458.4290 972.4355 729.5707	2,3,4	5	6.06
J (A+B)	Ac-YTSL IHSLIEESQNQQEKN EQELLELDKW	3585.7475	1196.2876 897.2939	3,4	7	5.82
K(B+C)	IHSLIEESQNQQEKN EQELLELDKWASL W	3536.7423	1179.9489 885.1406 708.3010	3,4,5	8	5.69
L (A+B+C)	Ac-YTSLIHSLIEESQNQ QEKNEQELLELDK WASLW	4042.9800	1348.8082 1011.7202	3,4	10	6.35

Across the remaining combinations (tris buffer, 18 or 24-hour incubation and 1:50 or 1:100 ratio), signal across the various digest fragments was higher for low missed cut site fragments and showed little variation. Given that, both 18 and 24 hour digests and 1:50 and 1:100 ratios using the tris-HCl buffer were used as a starting point for optimizing the simultaneous Lys-C, chymotrypsin digest.

In testing a digest using two proteases at once, there were two main strategies seen in the literature(40): begin digesting the protein with both enzymes at the same time or to take a staggered approach and start one enzyme earlier. As such, both staggered and simultaneous digestion were assessed at 18 hours with a 1:50 ratio and 1:100 ratio; simultaneous digestion was also assessed at 24 hours with a 1:50 ratio. Staggering was done by adding Lys-C first and incubating for 4 hours at 36 C and then adding Chymotrypsin.to react for the remaining 14 hours.

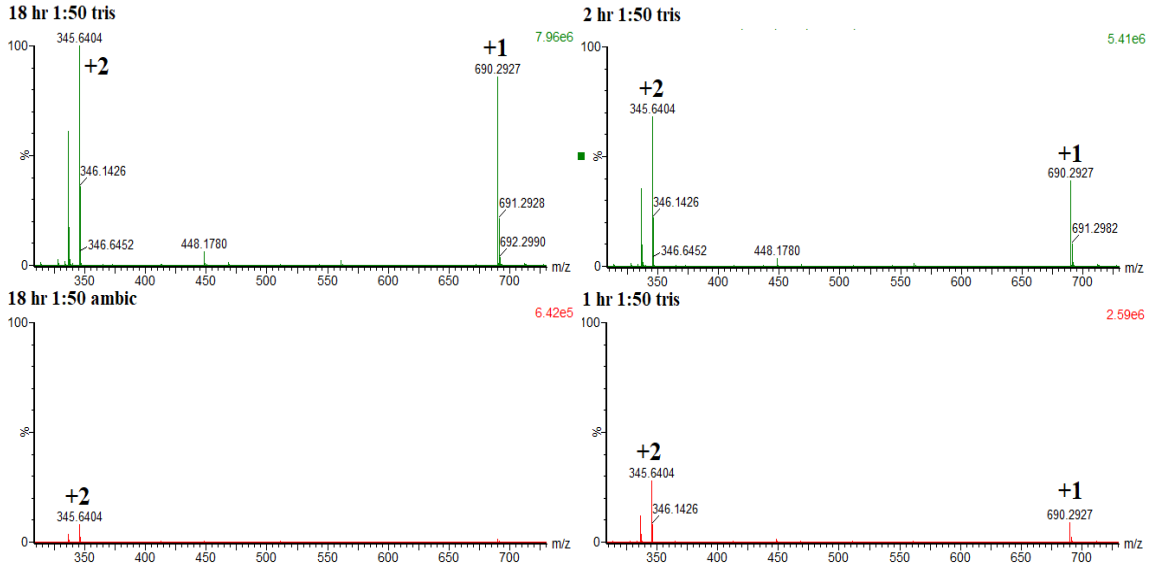

Figure 17. Comparison of spectra of fragment G for different incubation times and buffers for Chymotrypsin digest. Text at top left of each chromatogram indicates the digest conditions and top right indicates signal (vertical axes are linked across spectra). The peaks of interest are 345.6404 (2+ charge state) and 690.2927 (1+ charge state).

Success of the digest was checked as previously described. Fragments were again identified (Table 2) and it was found that all of the higher mass fragments found in the chymotrypsin digest (B, I, J, K and L) were no longer seen and 8 new fragments seen along with the remaining smaller chymotrypsin digest fragments were seen.

Table 2. Fragments identified from Chymotrypsin and Lys-C digest of Enfuviritide. * indicates fragment seen in previous literature. "Ac-" refers to N-term acetylation. "-Am" refers to C-term amidation. X(Y') indicates that fragment X is a part of fragment Y. Fragment names are carried over from the chymotrypsin digest fragments (Table 12). Z (X+Y) indicates that fragment Z is a combination of fragments X and Y.

Name	Sequence	MW (Expected)	MW (seen)	Charge States Seen	Missed Cut Sites	Retention Time (min)
Whole	Ac-YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF-Am	4490.1707	1123.2941 1497.4082 899.0463	3, 4, 5	x	6.48
A	Ac-YTSL	524.2482	525.2292	1	1	5.55
C	ASLW	475.2431	476.2285	1	1	5.64
D	NWF	464.2012	465.1823	1	1	5.62

Name	Sequence	MW (Expected)	MW (seen)	Charge States Seen	Missed Cut Sites	Retention Time (min)
F (B')	IHSLIEESQNQQEKNEQELL	2408.1892	1205.1385 803.6833 603.0024	2,3,4	3	5.50
G (B')	ELDKW	689.3384	690.2927 345.6404	1,2	1	5.32
M(B')	SLIEESQNQQEK	1431.6842	1432.6859 716.8495 478.2372	1,2,3	1	5.16
N(B')	IHSLIEESQNQQEK	1681.8271	1682.8383 841.9239 561.6177	1,2,3	2	5.22
O(B')	NEQELLELDK	1229.6139	1230.6317 615.8182	1,2	2	5.51
P (E+N)	TSLIHSLIEESQNQQEK	1982.9909	992.5089 662.0035	2,3	3	5.57
Q(B')	NEQELLELDKW	1415.6933	1416.7167 708.8566	1,2	3	5.83
R	WASLW	661.3224	662.3297	1	2	5.99
S(A+E)	Ac-YTSLIHSLIEESQNQQEK	2188.0648	1095.0427 730.3644	2,3	4	6.00
T (G')	ELDK	503.2591	504.2636	1	1	5.51

In terms of results, the staggered digest produced notably lower intensity signal for 5 digest fragments (C, F, G, M and Q) with the lowest being 100x weaker than the simultaneous digest. (Figure 18). For the remaining fragments, signal intensity was similar to the simultaneous digest. Fragment S defies this generalization though, having a signal 9-18 x higher than the signal seen in the non-staggered digests. Given that fragment S, comes from a missed chymotrypsin cut site and that the only fragments (C, F and G) that require two chymotrypsin cuts are among the fragments that showed low intensity for the staggered digest, it seems that the staggered digest had a weaker, more partial chymotrypsin digest.

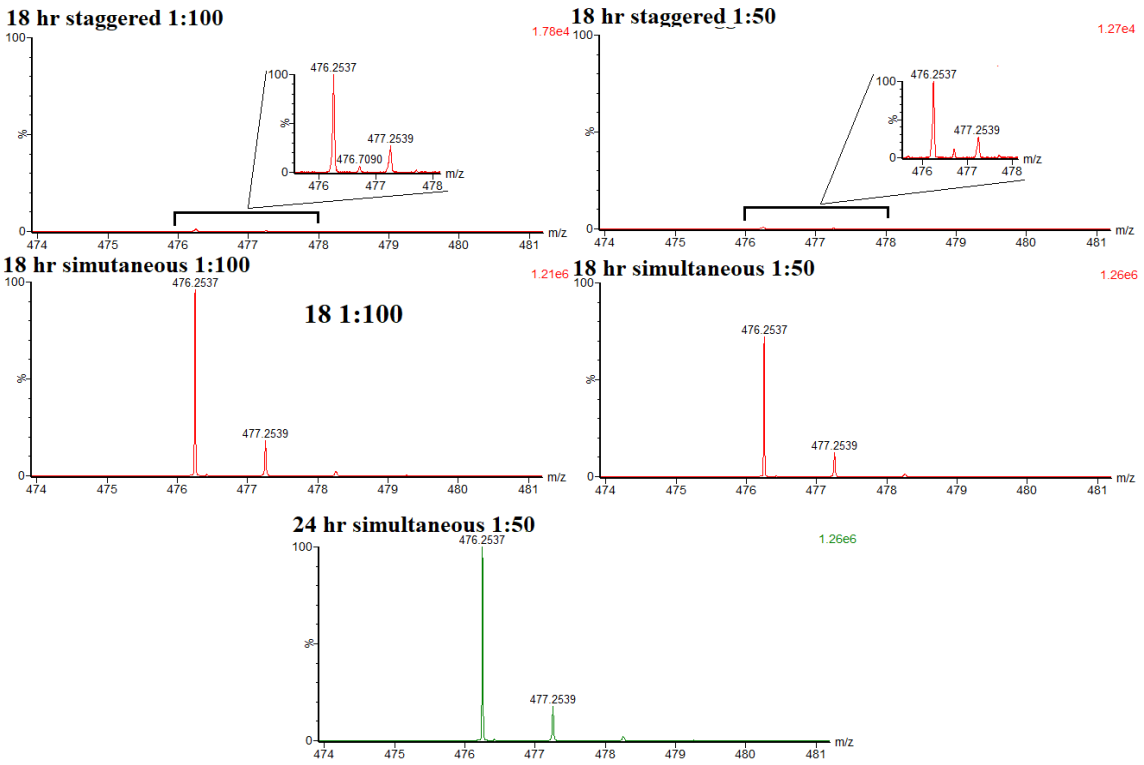

Figure 18. Comparison of spectra of fragment C for various Chymotrypsin and Lys-C digests. Text on top of spectra indicates digest conditions while text on top right indicates signal. Vertical axes are linked for all spectra. Spectra with low signal intensity have their peaks magnified.

The remaining digests, again, have negligible differences in intensity across peaks except for fragment S which is higher for the 18 hour, 1:100 digest. Despite that, it does not seem that the 18 hour, 1:100 digest had a relatively more incomplete chymotrypsin digest as there are no other indications of that in the signal intensity of the other fragments. Given that, either the 18 hour or 24 hours, 1:50 or 1:100 non-staggered digest could be used on human protamine P1 and P2.

3.4.2 Trypsin and Arg-C (separate) digest of Chum Salmon Protamine

3.4.2.1 2D-LC Optimization for Analysis of Protamine Digest Fragments

Initial analysis of digests of chum salmon protamine with Trypsin and Arg-C produced very few fragments compared to what was expected and at very low intensity. This prompted further optimization of the previously described 2D-LC method for analysis of protamine digest fragments. The initial trap column used was the C18 2.1 x 5 mm, initially used for its 300 Å pore size as the method was optimized on the intact chum salmon protamine (~ 4000 Da). Since Arg-c and Trypsin produce smaller fragments (<1000 Da) upon digesting protamine sulfate, this pore size confers no advantage and could potentially lead to lack of retention on the first dimension. As such, three other trap columns, Polymer 2.1 x 30 mm, C18 2.1 x 30 mm and C8 2.1 x 30 mm and their loading and eluting pH's (pH 3, 7 and 10) were assessed, keeping all other parameters from the optimized 2D-LC method the same with both ACN and IPA being tested as elution solvents.

This larger evaluation ultimately proved successful in showing key trends in what parameters produced high signal. Each chromatogram and mass spectra was evaluated based on looking for three predicted digest fragments (Table 3) consistently seen across methods.

Any method that used an elution mobile phase with 0.5% NH_4OH as an additive produced poor peak shape and either no or much lower signal.

Table 3. Fragments used to assess 2D-LC optimization for digest fragments.

Name	m/z seen	Theoretical m/z (+1 charge)	Retention Time	Sequence	Charge
A2	264.1712 527.3422	527.3412	5.43	RPVR	1,2
B2	361.2173	361.2194	5.40	VSR	1
C2	263.4807 394.7206	787.4300	5.43	SSSRPVR	2,3

Conversely, the best conditions were found to be ACN as an elution mobile phase with 0.5% formic acid as an additive, with the loading and diluting streams using H_2O at pH 10 across all three trap columns and elution mobile phases (ACN) (Figure 19)). IPA elution, for these methods produced a relatively lower signal and resulted in peak shape distortion and tailing. Each of these methods with ACN elution produced the strongest signal across the four peaks while maintain Gaussian peak shape and producing a variety of digest fragments.

These conditions may be best for these fragments due to the basic nature of the protamines, and its resulting fragments, due to the high percentage of arginine. At high pH, arginine is neutral which could increase the interaction that it and fragments containing it have with the non-polar stationary phase of the trapping column, resulting in more fragments being retained. At low pH, during elution, arginine becomes charged which decreases its interaction with the stationary phase, making pulling the fragments off efficiently in a tight band possible.

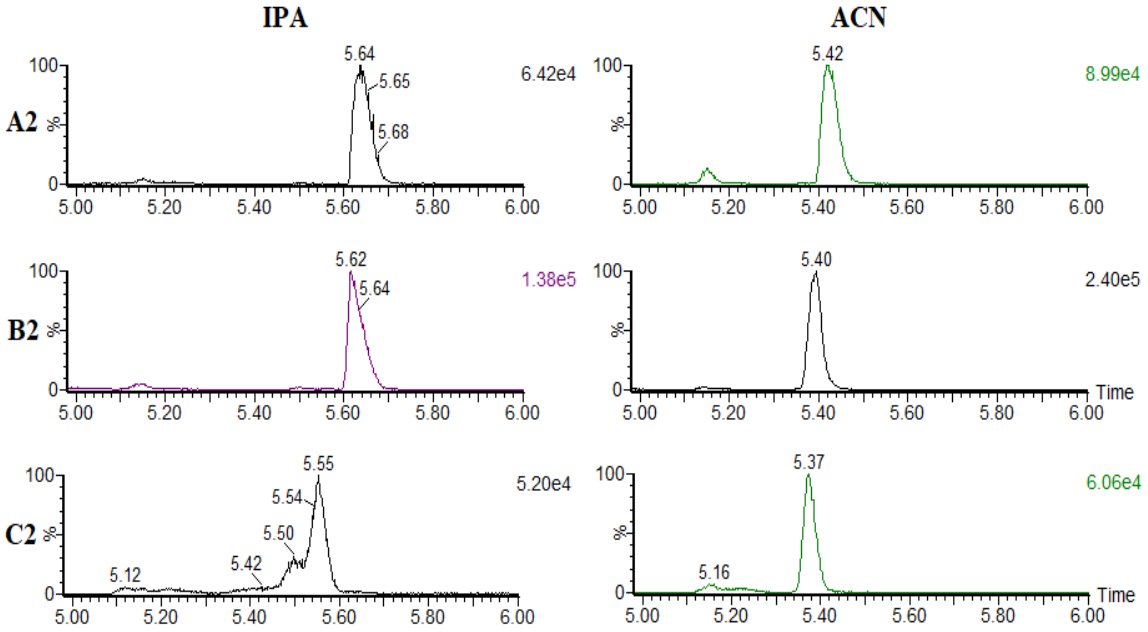

Figure 19. Comparison of chromatograms of Trypsin digest of protamine using low pH elution, high pH loading on a polymer trap. Text on top of spectra indicates elution solvent, text on the left hand side indicates which fragment is shown. Signal intensity is noted on the top right of each chromatogram.

In addition to a higher signal, an identified fragment list was able to be made for each digest which are shown in Tables 4 and 5. 26 fragments were identified for Arg-C and 31 for trypsin, only considering peaks with a signal > 1e5 signal. A threshold was used as not every peak of the digest needs to be identified for the sake of this application and there were a sufficient number of fragments to confirm and assess the digest using only the higher intensity peaks. A list of peaks that varied in intensity, charge state, number of missed cut sites and covering all four species were selected to assess a protamine sulfate digest with Arg-C or Trypsin (Table 6). This allows for testing protamine sulfate digest across a variety of parameters such as enzyme/substrate ratio and incubation time for optimization.

Table 4. Fragments identified from Arg-C digest of chum salmon protamine.

m/z seen	Theoretical m/z (+1 charge)	Retention Time	Sequence	Missed Cut Sites	Charge	Seen in
264.1712, 527.3422	527.3412	5.43	RPVR	0	1,2	P1
361.2173	361.2194	5.40	VSR	0	1	P1,P3, P4
315.5144, 472.7729	943.5120	5.43	RSSSRPVR or SSSRPVRR	1	2,3	P3
320.1868, 479.7808	957.5468	5.45	SSSRPIRR or RSSSRPIR	1	2,3	P2
263.4807, 394.7206	787.4300	5.43	SSSRPVR	0	2,3	P3
185.6990, 247.4931, 370.7389, 740.4824	740.4750	5.24	PRRRR or RRRPR	3 or 2	1,2,3,4	P1,P2,P3,P4
175.1160	175.1117	5.25	R	0	1	P1,P2,P3,P4
136.5856, 272.1696	272.1717	5.24	PR	0	1,2	P1,P2,P3,P4
166.1101, 331.2181	331.2211	5.15	RR	1	1,2	P1,P2,P3,P4
222.6523, 444.3035	444.3041	5.43	RIR or IRR	1	1,2	P1
244.1611, 487.3216	487.3212	5.16	RRR	2	1,2	P1,P2,P3,P4
223.1322, 445.2628	445.2630	5.16	GGRR or RGGR	1	1,2	P1,P2,P3,P4
201.1221	201.1262	5.16	GGRRR or RGGRR	2	3	P1,P2,P3,P4
253.1429	253.1457	5.16	RSSR	1	2	P4
301.1857, 601.3669	601.2641	5.16	RRGGR or RGGRR or GGRRR	2	1,2	P1,P2,P3,P4
323.1956	323.1988	5.16	RRASR	2	2	P1
214.6367, 428.2723	428.2728	5.24	PRR or RPR	1 or 0	1,2	P1,P2,P3,P4
505.2854	505.2841	5.16	RSSR	1	1	P4
195.4591, 292.6879, 584.3757	584.3739	5.29	PRRR or RRPR	2 or 1	1,2,3	P1,P2,P3,P4
259.1608, 517.323	517.3205	5.29	VSRR	1	1,2	P1,P3, P4

m/z seen	Theoretical m/z (+1 charge)	Retention Time	Sequence	Missed Cut Sites	Charge	Seen in
228.4819, 342.2221, 683.4445	683.4424	5.43	RPVRR	1	1,2,3	P4
288.2007	288.2030	5.43	IR	0	1	P1
367.5486, 550.8243, 944.5516	944.5384	5.43	RSSSRPVR or SSSRPVRR	2	1,2,3	P3
429.2553	429.2568	5.43	SSRRPVR	1	2	P4
557.8323	557.8312	5.43	RRSSSRPIR or RSSSRPIRR or SSSRPIRRR	3	2	P2
788.4424	788.4373	5.43	SSSRPVR	4	1	P3

Table 5. Fragments identified from Trypsin digest of chum salmon protamine

m/z seen	Theoretical m/z (+1 charge)	Retention Time	Sequence	Missed Cut Sites	Charge	Seen in
244.1611, 487.3216	487.3212	5.16	RRR	2	1,2	P1,P2,P3,P4
223.1321, 445.2617	445.2630	5.16	GGRR or RGGR	1	1,2	P1,P2,P3,P4
301.1824	301.1857	5.16	RRGGR or RGGRR or GGRRR	2	2	P1,P2,P3,P4
323.1959	323.1988	5.16	RRASR	2	2	P1
166.1099, 331.2169	331.2211	5.15	RR	1	1,2	P1,P2,P3,P4
379.2336	379.2362	5.16	RRRGGR or RRGGRR or RGGRRR or GGRRRR	3	2	P1,P2,P3,P4
214.6367, 428.2654	428.2728	5.16	PRR or RPR	1	1,2	P1,P2,P3,P4
194.4588, 292.6879, 584.3753	584.3739	5.29	PRRR or RRPR	2 or 1	1,2,3	P1,P2,P3,P4
259.1608, 517.3196	517.3205	5.29	VSRR	1	1,2	P1,P3, P4

m/z seen	Theoretical m/z (+1 charge)	Retention Time	Sequence	Missed Cut Sites	Charge	Seen in
247.4926, 370.7389, 740.4827	740.4750	5.24	PRRRR or RRRPR	3 or 2	1,2,3	P1,P2,P3,P4
136.5853, 272.1693	272.1717	5.24	PR	0	1,2	P1,P2,P3,P4
186.1196, 232.3954, 309.5265, 463.7907, 926.5812	926.5755	5.43	RRPRVSR or RPRVSRR	2	1,2,3,4,5	P1,P3,P4
200.8026, 300.7029	599.3980	5.43	RIRR or IRRR	2	2,3	P1
264.1709, 527.3416	527.3412	5.43	RPVR	0	1,2	P4
222.6523, 444.3028	444.3041	5.43	RIR or IRR	1	1,2	P1
228.4819, 342.2216	682.4351	5.43	RPVRR	1	2,3	P4
248.1488, 371.7224	741.4358	5.43	RPRASR	1	2,3	P2
271.4207, 361.5606	1081.6693	5.43	RRRPRVSR, RRPRVSRR, RPRVSRRR	3	3,4	P1,P3, P4
288.2005	288.2030	5.43	IR	0	1	P1
361.2166	361.2194	5.43	VSR	0	1	P1,P3, P4
367.5485, 550.8241, 944.5471	944.5384	5.43	RSSSRPVR or SSSRPVRR	2	1,2,3	P3
385.7385, 770.4778	770.4744	5.43	RPRVSR	1	1,2	P1,P3, P4
263.4800, 394.7199	787.4300	5.43	SSSRPVR	0	2,3	P3
449.7746	449.7757	5.43	RRPRASR or RPRASRR	2	2	P2
315.5140, 472.7724	943.5312	5.43	RSSSRPVR or SSSRPVRR	1	2,3	P3
320.1855, 479.7803	957.5468	5.45	SSSRPIRR or RSSSRPIR	1	2,3	P2
557.8312	557.8312	5.43	RRSSSRPIR or RSSSRPIRR	3	2	P2

m/z seen	Theoretical m/z (+1 charge)	Retention Time	Sequence	Missed Cut Sites	Charge	Seen in
619.8935	619.8925	5.43	RRRRPRVSR or RRRPRVSRR or RRPRVSRRR or RPRVSRRRR	4	2	P1,P3, P4
628.8759	628.8740	5.43	RRRSSSRPVR or RRSSSRPVRR or RSSSRPVRRR	3	2	P3
413.5942	413.5974	5.43	RRRPRVSRR, or RRPRVSRRR or RPRVSRRRR	4	3	P1, P3, P4
404.2503	404.2537	5.27	RRRRPRASR, RRRPRASRR, RRPRASRRR	4	3	P2

Table 6. Fragments used for Arg-C and Trypsin digest evaluation. Sequences highlighted with the same color indicate the same fragment but in multiple charge states.

m/z seen	Charge State	Missed Cuts	Sequence	Name
288.2007	1	0	IR	A3
361.2166	1	0	VSR	B3
264.1709	2	0	RPVR	C3
527.3422	1	0	RPVR	D3
185.6990	4	3 or 2	PRRRR or RRRPR	E3
247.4931	3	3 or 2	PRRRR or RRRPR	F3
370.7389	2	3 or 2	PRRRR or RRRPR	G3
740.4824	1	3 or 2	PRRRR or RRRPR	H3
244.1611,	2	2	RRR	I3
487.3216	1	2	RRR	J3
557.8312	2	3	RRSSSRPIR or RSSSRPIRR or SSSRPIRRR	K3

Based on signal intensity and diversity of peaks, the parameters chosen for this digest optimization are: Acetonitrile and H_2O with 0.5 % formic acid as the elution mobile phase flowing at 0.2 mL/min, H_2O with 2% NH_4OH (pH 10) as the loading and diluting mobile phase (flowing at 0.1 mL/min and 2 mL/min respectively), a Polymer, 2.1 x 30 mm as the 1st dimension trap column and a C18 2.1 x 100 mm as the 2nd dimension analytical column.

3.4.2.2 Digest Optimization

This evaluation varied incubation period with 4, 18 and 24 hours and enzyme/substrate mass ratio at 1:50, 1:100 and 1:200. Evaluations were carried out as described previously. Largely, varying these parameters showed no significant change in intensity of any of the peaks except for a few cases. Trypsin at a 1:200 ratio produced extremely low to non-existent signal for all incubation times. For Arg-C, the 1:50 ratio caused a notable decrease in signal only in the 4+ charge state of fragment E3 which is not a major concern for this application.

Due to the lack of variance, 18 hours, 1:200 was chosen for Arg-C to conserve the amount of Arg-C used. Trypsin is optimized at 1:100 18 hours, due to loss of signal seen in the 1:200 samples, lack of variance and also because it coincides with the other digests (no major difference seen between 1:100 18 hours and 1:100 24 hours).

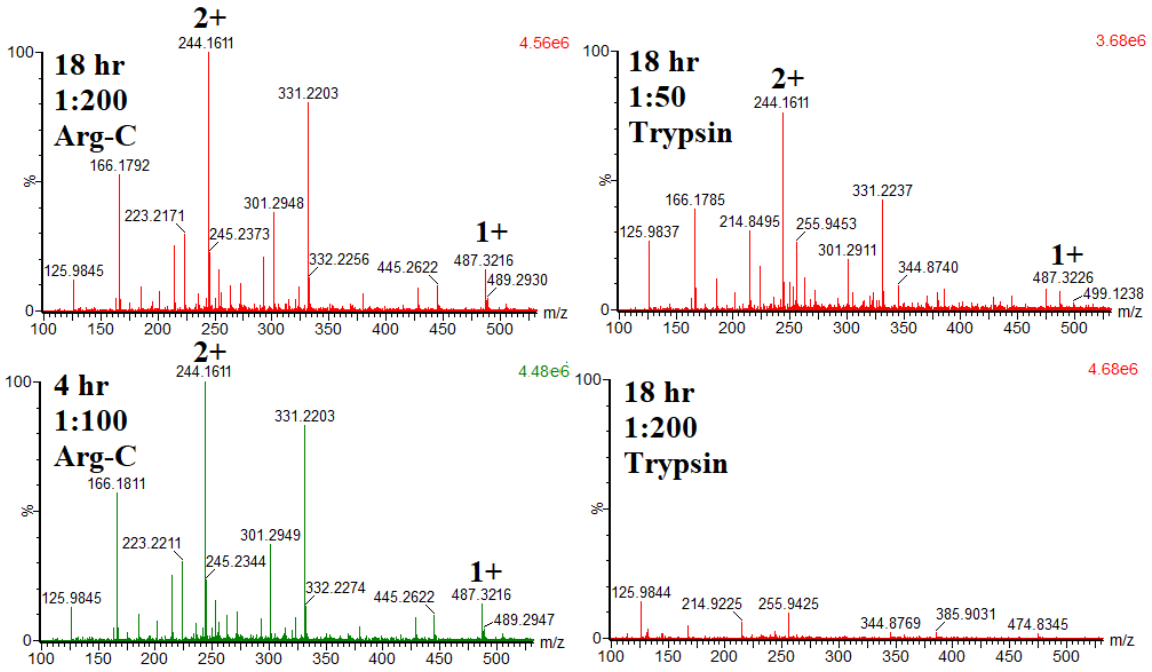

Figure 20. Comparison of Spectra of fragment I3 and J3 resulting from various digest conditions. Each chromatogram is labeled with the digest conditions in the top left, signal intensity in the top right (vertical axes are linked), and the charge states labeled (I3 = 2+, J3 = 1+)

4. CONCLUSION

This work produced methodologies for digesting a chum salmon protamine and an optimized method for analysis via 2D-LC-MS/MS for the identification of fragments and then sequencing of key fragments for the purposes of mapping of disulfide bonds. These were done in a manner that, hopefully, could be applied to a purified extract of human protamine from human sperm for the mapping of disulfide bonds in human P1 and P2. In addition to one optimized method proposed, a variety of other methods have been characterized that produce comparable signal and a different proportion of fragments that can act as an alternative if issues arise in applying the proposed method to human protamines.

It was determined that use of Acetonitrile at a flow rate of 0.2 mL/min as an elution mobile phase is optimal, however IPA is an acceptable alternative and analysis with both may prove useful as different fragments are seen with each mobile phase. Elution mobile phases must be acidic with addition of 0.5% formic acid and loading and diluting mobile phases should be prepared at pH 10 for optimal signal intensity and number of fragments present. In terms of column chemistry, all of the trap columns tested (Polymer, C18 or C8), produced comparable results, with polymer being producing slightly higher signal, using these mobile phases. The same results are observed with the 2^{nd} dimension columns (2.1x100 mm C18, 2.1 x 150 mm C18 and 2.1 x 100 C4) with no major variance seen among them.

It was determined that Arg-C digestion can be functional for this application and produce similar fragments at similar intensities across a variety of conditions. Trypsin

was found to be less robust, not working at a 1:200 ratio or particularly well at a 4-hour incubation time but all other combinations produced successful results. Chymotrypsin and Lys-C were found to work together in a digest simultaneously and produce comparable signal at an 18-hour incubation time with 1:50 and 1:100 ratios and a 24-hour incubation time with a 1:50 ratio.

5. FUTURE DIRECTIONS

Considering the differences between chum salmon protamine and human P1 and P2, these methods may require some adjustment to transfer over. Some issues seen with moving to the next step is that these digests are performed at neutral to slightly basic pH which are conditions that could allow for disulfide bond scrambling. It is recommended to work at low pH to prevent this thus some additional experimentation may need to be done if disulfide scrambling is seen with the methods reported above. Arg-C digestion may prove useful here as it has a lower pH range of 7.2-8 where trypsin has an optimal pH of 8 and Lys-C has an optimal pH of 8.6-8.8 (chymotrypsin has a wider pH range but it needs to be paired with Lys-C)(43). To test this, a known protein with well characterized disulfide linkages should be run alongside the human protamine as a positive control.

Another potential source of difficulty would be the actual digestion itself. Digestion of purified protein standards should occur with ease due to a minimum of interfering molecules the protein could bind to and potential loss of higher order structure. Many other digest protocols, designed for extracts from biological samples, include the use of Guanidine Hydrochloride (GuHCl) or Urea to denature proteins and DTT to cleave disulfide bonds as to open up the protein and allow more interaction with the protease. The proposed digest methods were able to produce fairly consistent results across a wide variety of parameters without any of these additives potentially due to clean nature of the standards being worked with. In working with a sample extract from sperm, depending on how purified the protein is, addition of these chemicals (excluding DTT)

may be necessary in addition to using longer incubation times or higher substrate:enzyme ratios. This may be exceptionally difficult considering what is known about the human protamines higher order structure in being bound to DNA and P1 and P2 potentially having intra- and intermolecular disulfide bonds.

Additionally, while methods have been clarified for sequencing of fragments, sequencing of disulfide linked fragments can prove more complex due to the numbers of ways the bonded peptides could fragment under CID in addition to all the ways a peptide could already fragment. Additional software tools, specifically for the analysis of disulfide linked peptides may need to be considered.

LIST OF JOURNAL ABBREVIATIONS

Anal Bioanal Chem	Analytical and Bioanalytical Chemistry
Anim Reprod Sci	Animal Reproduction Science
Asian J Androl	Asian Journal of Andrology
Biochem J	Biochemical Journal
Biophys J	Biophysical Journal
Biosci Rep	Bioscience Reports
BMC Evol Biol	BMC Evolutionary Biology
Brief Funct Genomic Proteomic	Briefings in Functional Genomics and Proteomics
Chem Res Toxicol	Chemical Research in Toxicology
Eur J Biochem	European Journal of Biochemistry
Exp Cell Res	Experimental Cell Research
Forensic Sci Int Genet	Forensic Science International: Genetics
Genome Biol	Genome Biology
Hum Reprod Update	Human Reproduction Update
J Am Soc Mass Spectrom	Journal of the American Society for Mass Spectrometry
J Androl	Journal of Andrology
J Biol Chem	The Journal of Biological Chemistry
J Proteome Res	Journal of Proteome Research
J Theor Biol	Journal of Theoretical Biology
Microsc Res Tech	Microscopy Research and Technique
Mol Hum Reprod	Molecular Human Reproduction
Nucleic Acids Res	Nucleic Acids Research

Proc Biol Sci	Proceedings of the Royal Society B: Biological Sciences
Protein Expr Purif	Protein Expression and Purification
Syst Biol Reprod Med	Systems Biology in Reproductive Medicine

BIBLIOGRAPHY

1. Contrepois K, Ezan E, Mann C, Fenaille F. Ultra-high performance liquid chromatography-mass spectrometry for the fast profiling of histone post-translational modifications. J Proteome Res. 2010;9(10):5501-5509.

2. Bao J, Bedford MT. Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. Reproduction. 2016;151(5):R55-70.

3. Balhorn R. The protamine family of sperm nuclear proteins. Genome Biol. 2007; 8(9):227.

4. Balhorn R, Weston S, Thomas C, Wyrobek AJ. DNA packaging in mouse spermatids. Synthesis of protamine variants and four transition proteins. Exp Cell Res. 1984;150(2):298-308.

5. Cotton RW, Fisher MB. Review: Properties of sperm and seminal fluid, informed by research on reproduction and contraception. Forensic Sci Int Genet. 2015;18:66-77.

6. Castillo J, Amaral A, Azpiazu R, Vavouri T, Estanyol JM, Ballescà JL, et al. Genomic and proteomic dissection and characterization of the human sperm chromatin. Mol Hum Reprod. 2014;20(11):1041-1053.

7. Castillo J, Estanyol JM, Ballescá JL, Oliva R. Human sperm chromatin epigenetic potential: genomics, proteomics, and male infertility. Asian J Androl. 2015;17(4):601-609.

8. D'Occhio MJ, Hengstberger KJ, Johnston SD. Biology of sperm chromatin structure and relationship to male fertility and embryonic survival. Anim Reprod Sci. 2007;101(1-2):1-17.

9. Manochantr S, Chiamchanya C, Sobhon P. Relationship between chromatin condensation, DNA integrity and quality of ejaculated spermatozoa from infertile men. Andrologia. 2012;44(3):187-199.

10. Castillo J, Simon L, de Mateo S, Lewis S, Oliva R. Protamine/DNA ratios and DNA damage in native and density gradient centrifuged sperm from infertile patients. J Androl. 2011;32(3):324-332.

11. Lüke L, Campbell P, Varea Sánchez M, Nachman MW, Roldan ER. Sexual selection on protamine and transition nuclear protein expression in mouse species. Proc Biol Sci. 2014;281(1783):20133359. Epub 2014/03/26.

12. Brunner AM, Nanni P, Mansuy IM. Epigenetic marking of sperm by post-translational modification of histones and protamines. Epigenetics & Chromatin. 2014; 7(1):2.

13. Björndahl L, Kvist U. A model for the importance of zinc in the dynamics of human sperm chromatin stabilization after ejaculation in relation to sperm DNA vulnerability. Syst Biol Reprod Med. 2011;57(1-2):86-92.

14. Lüke L, Tourmente M, Dopazo H, Serra F, Roldan ER. Selective constraints on protamine 2 in primates and rodents. BMC Evol Biol. 2016;16:21.

15. Quintanilla-Vega B, Hoover DJ, Bal W, Silbergeld EK, Waalkes MP, Anderson LD. Lead interaction with human protamine (HP2) as a mechanism of male reproductive toxicity. Chem Res Toxicol. 2000;13(7):594-600.

16. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2018;46(5):2699.

17. McKay DJ, Renaux BS, Dixon GH. The amino acid sequence of human sperm protamine P1. Biosci Rep. 1985;5(5):383-391.

18. McKay DJ, Renaux BS, Dixon GH. Human sperm protamines. Amino-acid sequences of two forms of protamine P2. Eur J Biochem. 1986;156(1):5-8.

19. Biegeleisen K. The probable structure of the protamine-DNA complex. J Theor Biol. 2006;241(3):533-540.

20. Vilfan ID, Conwell CC, Hud NV. Formation of native-like mammalian sperm cell chromatin with folded bull protamine. J Biol Chem. 2004;279(19):20088-20095.

21. Pirhonen A, Linnala-Kankkunen A, Mäenpää PH. Identification of Phosphoseryl Residues in Protamines from Mature Mammalian Spermatozoa1. Biology of Reproduction. 1994;50(5):981-986.

22. Hutchison JM, Rau DC, DeRouchey JE. Role of Disulfide Bonds on DNA Packaging Forces in Bull Sperm Chromatin. Biophys J. 2017;113(9):1925-1933.

23. Gatewood JM, Schroth GP, Schmid CW, Bradbury EM. Zinc-induced secondary structure transitions in human sperm protamines. J Biol Chem. 1990;265(33):20667-20672.

24. Dadoune JP. Expression of mammalian spermatozoal nucleoproteins. Microsc Res Tech. 2003;61(1):56-75.

25. Oliva R. Protamines and male infertility. Hum Reprod Update. 2006;12(4):417-435. Epub 2006/03/31.

26. Raukas E, Mikelsaar RH. Are there molecules of nucleoprotamine? Bioessays. 1999;21(5):440-448.

27. Balhorn R. A Model for the Structure of Chromatin in Mammalian Sperm. Journal of Cell Biology. 1982;93:298-305.

28. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, et al. Mass-spectrometry-based draft of the human proteome. Nature. 2014;509(7502):582-587.

29. Lomeli SH, Peng IX, Yin S, Loo RR, Loo JA. New reagents for increasing ESI multiple charging of proteins and protein complexes. J Am Soc Mass Spectrom. 2010; 21(1):127-131.

30. Gross J. Mass Spectrometry, A Textbook. Heidelberg, Germany: Springer International Publishing; 2017.

31. Griffiths WJ, Jonsson AP, Liu S, Rai DK, Wang Y. Electrospray and tandem mass spectrometry in biochemistry. Biochem J. 2001;355(Pt 3):545-561.

32. Gates P. Electrospray Ionisation (ESI): University of Bristol; 2014 [cited 2018 05/19]. Available from: http://www.chm.bris.ac.uk/ms/esi-ionisation.xhtml.

33. Wysocki VH, Resing KA, Zhang Q, Cheng G. Mass spectrometry of peptides and proteins. Methods. 2005;35(3):211-222.

34. Chernushevich Igor V, Loboda Alexander V, Thomson Bruce A. An introduction to quadrupole–time-of-flight mass spectrometry. Journal of Mass Spectrometry. 2001; 36(8):849-865.

35. Levine B. Principles of Forensic Toxicology. Levine B, editor. Washington, DC: American Association for Clinical Chemistry Press; 2016.

36. Xu Q. Ultra-High Performance Liquid Chromatography and Its Applications. Xu QA, editor. Hoboken, New Jersey: John Wileu & Sons, Inc; 2013.

37. Mallet C, Murphy B. Multi-Dimensional Chromatography Compendium: Trap and Elute vs. At-Column Dilution. Milford, Massachusetts: Waters Corporation; 2015. Available from http://www.waters.com/webassets/cms/library/docs/720005339en.pdf

38. Gucinski AC, Boyne MT, Keire DA. Modern analytics for naturally derived complex drug substances: NMR and MS tests for protamine sulfate from chum salmon. Anal Bioanal Chem. 2015;407(3):749-759.

39. Hoffmann JA, Chance RE, Johnson MG. Purification and analysis of the major components of chum salmon protamine contained in insulin formulations using high-performance liquid chromatography. Protein Expr Purif. 1990;1(2):127-133.

40. Biringer RG, Amato H, Harrington MG, Fonteh AN, Riggins JN, Hühmer AF. Enhanced sequence coverage of proteins in human cerebrospinal fluid using multiple enzymatic digestion and linear ion trap LC-MS/MS. Brief Funct Genomic Proteomic. 2006;5(2):144-153.

41. Choudhary G, Wu SL, Shieh P, Hancock WS. Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. J Proteome Res. 2003;2(1):59-67.

42. van den Broek I, Sparidans RW, Schellens JHM, Beijnen JH. Enzymatic digestion as a tool for the LC–MS/MS quantification of large peptides in biological matrices: Measurement of chymotryptic fragments from the HIV-1 fusion inhibitor enfuvirtide and its metabolite M-20 in human plasma. Journal of Chromatography B. 2007;854(1):245-259.

43. Switzar L, Giera M, Niessen WM. Protein digestion: an overview of the available techniques and recent developments. J Proteome Res. 2013;12(3):1067-1077.

CURRICULUM VITAE