EXAMINING THE VARIABLES WHICH INFLUENCE STRONG CATION EXCHANGE OF PEPTIDES, AND THEIR IMPLICATIONS ON THE FIRST DIMENSION OF SEPARATION IN LIQUID CHROMATOGRAPHY TANDEM MASS SPECTROMETRY

by

Kirsten M. Jones

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
August 2018

Dedication

For my Grandparents Jim and Jean Tremayne for their belief that education is the greatest gift you can bestow upon a child -and their generosity which accompanied that belief. And especially for James Douglas Tremayne (1926 -2012), who taught me the value of patience, kindness, honesty, optimism, and humour in the face of adversity.

If you can dream—and not make dreams your master;
If you can think—and not make thoughts your aim;
If you can meet with Triumph and Disaster
And treat those two impostors just the same;
If you can bear to hear the truth you've spoken
Twisted by knaves to make a trap for fools,
Or watch the things you gave your life to, broken,
And stoop and build 'em up with wornout tools:

If you can make one heap of all your winnings
And risk it on one turn of pitch-and-toss,
And lose, and start again at your beginnings
And never breathe a word about your loss;
If you can force your heart and nerve and sinew
To serve your turn long after they are gone,
And so hold on when there is nothing in you
Except the Will which says to them: "Hold on;"

-Rudyard Kipling

Table of Contents

List of Tables

List of Figures

Abstract

Orthogonal separations are often employed to improve the capacity of mass spectrometry for proteome analysis. The gold standard of orthogonal peptide-level separation utilises strong cation exchange (SCX) coupled to reversed phase chromatography in an online format. However, the gradient conditions used in SCX fail to account for maximal sample recovery, nor for optimising the degree of fractionation; most SCX separations primarily rely on salt gradients, rather than employing pH ramps to recover peptides. Additionally, online separations must compromise the types of solvents used to reduce matrix effects experienced in electrospray ionisation mass spectrometry. This thesis endeavours to understand the conditions governing retention and recovery of peptides from SCX, primarily pH and ionic strength, using two novel offline approaches.

The ProTrap XG presented in Chapter 2 is a centrifugal column chromatography device which allows for multiple simultaneous separations, with an almost unlimited choice in separation buffers. Using this system, it was determined that altering the pH of the eluting buffer was the primary variable impacting peptide retention on the SCX column, while ionic strength played a minor role in the separations. A pH gradient in the absence of salt was optimised, loading peptides at pH 2.1 and eluting between pH of 3 and 11, with an average recovery of 80 ± 11% and 2600 ± 800 *E. coli* peptides identified per replicate. Attempts at optimising the *E. coli* peptide separation through combining pH and salt resulted in 90% recovery. However, the device had challenges in reproducibility due to variable flow rates and spin times, which arose due to variations in column consistency. This is addressed in chapter 3 with a novel automated HPLC-based separation platform capable of multiple solvent deliveries for optimised SCX separation.

The Strong Cation Exchange Clean-up Liquid Chromatography (S-CLC) system separates yeast peptides and desalts the fractions and measures recovery by LC-UV. Two key components distinguish this system from traditional multidimensional chromatography systems, namely two distinct pumps are employed (one to load/ elute on SCX and the other to load/ elute from reversed phase). The second is to use an autosampler to inject the solvents used to recover peptides from SCX, which like the ProTrap XG, allows tailored selection of buffers to optimise separation in automated fashion. Separation by S-CLC confirmed that pH is the primary variable in altering the retention of peptides in SCX, while salt gradients provide counterions that compete for interaction with the column. Salt and pH were combined to elicit the removal of peptides from the SCX column and compared to the gold standard method. While both methods identified a similar number of peptides the MuDPIT method had higher levels of overlap between fractions, while the combined salt and pH gradient had decreasing overlap.

Based on these results, an optimised separation on SCX should be developed using univariate separations with pH in the absence of salt, and a salt gradient under constant pH conditions. Once this is complete, salt and pH gradients can be combined until obtaining even peptide partitioning. On a high-resolution mass spectrometer, it is expected that there will be a significant increase in the number of identified proteins when compared to traditional proteomic separation methods.

List of Abbreviations and Symbols Used

$(NH_4)_3C_6H_5O_7$	Ammonium Citrate
$\times g$	Times the Force of Gravity
°C	Degrees Celsius
µg	Micrograms
µL	microLiters
2D-LC	Two-dimensional Liquid Chromatography
Å	Angstrom
ACN	Acetonitrile
Arg	Arginine
Asp	Asparagine
BCA	Bicinchoninic Assay
BSA	Bovine Serum Albumin
BSA-CRM	Bovine Serum Albumin-Certified Reference Material
CA-TRIS	Citric acid and TRIS Gradient
CRM	Certified Reference Material
Da	Dalton
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
ESI	Electrospray Ionisation
ESI-MS	Electrospray Ionisation Mass Spectrometry
FTICR	Fourier-Transform Ion Cyclotron Resonance
GELFrEE	Gel Elution Liquid Fraction Entrapment Electrophoresis
Gln	Glutamine
Glu	Glutamic Acid
HGP	Human Genome Project
HPLC	High Performance Liquid Chromatography
HPLC-MS	High Performance Liquid Chromatography-Mass Spectrometry

HPLC-UV	High Performance Liquid Chromatography Ultraviolet Spectroscopy
IAA	Iodoacetamide
ICAT	Isotope-Coded Affinity Tagging
ID	Identification
ID/min	Identifications per Minute
IEF	Isoelectric Focussing
kDa	Kilodalton
kV	Kilovolt
LB	Luria Bertani
LC	Liquid Chromatography
LC-MS	Liquid Chromatography- Mass Spectrometry
LC-MS/MS	Liquid Chromatography- Tandem Mass Spectrometry
LC-UV	Liquid Chromatography- Ultraviolet spectroscopy
LOD	Limit of Detection
LOQ	Limit of Quantitation
LTQ	Linear Ion Trap Quadrupole
Lys	Lysine
m/z	Mass to Charge
MALDI	Matrix-assisted Laser Desorption/Ionisation
mAU	Mass Absorbance Units
MDLC	Multidimensional Liquid Chromatography
min	Minutes
mM	Millimolar
MS Excel	Microsoft Excel
MuDPIT	Mulitdimensional Protein Identification Technology
nm	Nanometer
PBS	Phosphate Buffered Saline
PEHA	Pentylethylenehexamine
pH	Potential Hydrogen

pKa	-log (Acid Dissociation Constant)
qTOF	Quadrupole Time of Flight
RP	Reversed Phase
rpm	Rotations Per Minute
S/N	Signal-to-Noise
S-CLC	Strong Cation Exchange Clean-Up Liquid Chromatography
SCX	Strong Cation Exchange
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SPE	Solid Phase Extraction
SPG	Salt and pH Gradient Method
TFA	Trifluoroacetic Acid
TRIS	Tris(hydroxymethyl)aminomethane
UV	Ultraviolet Spectroscopy
UV-VIS	Ultraviolet and Visible Spectroscopy
VIS	Visible Spectroscopy
λ	Wavelength

Acknowledgements

First and foremost, I want to thank my parents, and my grandparents without their financial backing, and guidance I would not have been able to complete the either of my degrees.

I want to thank **Dalhousie University** for the amazing opportunities that this institution has helped to provide me.

In particular:

- My supervisor, **Dr. Alan Doucette**, and my committee members, **Dr. Robert White** and **Dr. Devenand Pinto** for their support, advice and patience with me.
- **Dr. Heather Andreas**, and **Dr. Roderick Chisholm** for their advice on both life and school.
- Current and past group members and floor mates for sharing this learning experience with me. Especially **Andrew Crowell** for his help over the last 5+ years. **Nicole Unterlander** for helping to read and provide comments on Chapter 2, and **Chelsi Wicks** and **Steven Driscoll**, for putting up with my 'dumb' stats moments.
- Without **Mike Boutilier** and **Ryan MacKinnon** a vast majority of the equipment in our research group would not be functioning today. I am incredibly indebted to them for their assistance.
- Dalhousie Rowing in particular: **Erica Gagnon** for being one of the best coaches I've ever had on the water, her amazing positivity towards coaching us in athletics, and in life was inspiring. My winter crew! **Ryker Been**. and **Myshel Enderlein**, for waking up at 5:30 every morning this winter to train with me right through exam season, and for preventing boredom in the Killam while I was writin. **Eleri McEachern,** for amazing hugs, cheesecake and deep chats about the meaning of life and belonging.
- **#14.**, **#10**. and **#31** for reminding me about: why I fell love with chemistry, what I really want to do with my life and, incredible grit, and determination when the cards are admittedly at times stacked against you. The three of you are amazing humans and I think I may have learnt more from you than you did from me. Go Tigers.

The writing process, and the research process was a difficult one for me, so there are some additional people outside of Dal that I also need to thank:

Without **Leanne Huck**, I probably wouldn't be able to walk, let alone been able to do the research that I did, played the sports that I play, or have been able to do in my day to day. Her work has done wonders in terms of my quality of life during this entire process.

Everyone in grad school needs to surround themselves with patient people that understand that time is fluid, and that I will probably be late to basically everything, likely a little grumpy, or overly sensitive.
So: **Katie MacVicar, Fiona Warde, Emily Baker, Jaime Jones, Bart Soroka, Taylor Burns, Eric Bergmann, Arielle Detraz, Adam Norton, Lauren Bilinsky** and **Tomasz Hines** thank you for being so patient.

Beverley Gillis for the donation of her time, and her incredible patience in helping to edit this beast cover to cover. I'm sorry if any subsequent additions I made to this thing ruined your handy-work.

Alex Patterson, Emily Baker and Thomas Hines for reading through my very rough
Adam Norton: Who was supposed to check my thesis for formatting, and instead read and edited this cover to cover.
Mum, **Cecilly,** and **David** for triple checking my formatting and my references.

Chapter 1 Introduction

1.1 Preamble

The discovery of the structure of DNA by Watson and Crick[1] , Wilkins[2], as well as Franklin[3] in 1953 and methods to sequence DNA by Sanger[4] in 1977 are considered watershed moments in biological innovation. These discoveries led to the development of the Human Genome Project (HGP) in 1990, one of the largest collaborative research projects in modern history.[5] The HGP project endeavoured to map the genetic blueprint and help unlock the origins of many diseased states.[5–7] In most respects the HGP was successful resulting in the sequencing of approximately 21,000 human genes along with the genomes for many other organisms.[6,7] However, the process was also humbling to the scientific community, as an organism's DNA is only a miniscule component contributing to our knowledge of a disease. As an example, a genetic predisposition does not always result in the corresponding illness.[8] Since the completion of the HGP, genomic testing, though useful, remains an imprecise disease detection tool. One of the reasons for this is that not all genes are expressed at the same time. This often makes it impossible to predict through sequencing the exact moment when genes associated with a disease will begin to express - if at all.[6] Consequently, a number of -omics fields such as transcriptomics, metabolomics, and proteomics (the latter being the focus of this thesis) have arisen to better understand the products of our genetic code. The term proteome was first coined in 1994 by Wilkins at a conference as a description of the comprehensive study of all protein expressed by a genome.[9–11] Thus, proteomics is the systematic

analysis of the structure, function and composition of all protein sequences expressed by an organism at any given time.

In detail, proteins are biopolymers made up of approximately twenty different naturally occurring α-amino acid monomers. Polymers containing less than 50 amino acids are generally known as peptides. Each protein contains a C-terminal carboxyl group, and an N-terminal amine group. Each amino acid is distinguished by its differing side chains ('R' groups) attached to the (α) carbon (Figure 1-1).[12,13] A polypeptide chain is formed through a condensation reaction between an N-terminal amine group of one amino acid, and the carboxyl terminus of another amino acid to form a peptide bond. These polypeptide chains readily adjust their conformation as a direct result of the chemical properties of each amino acid, resulting in many differing biological effects.[12]

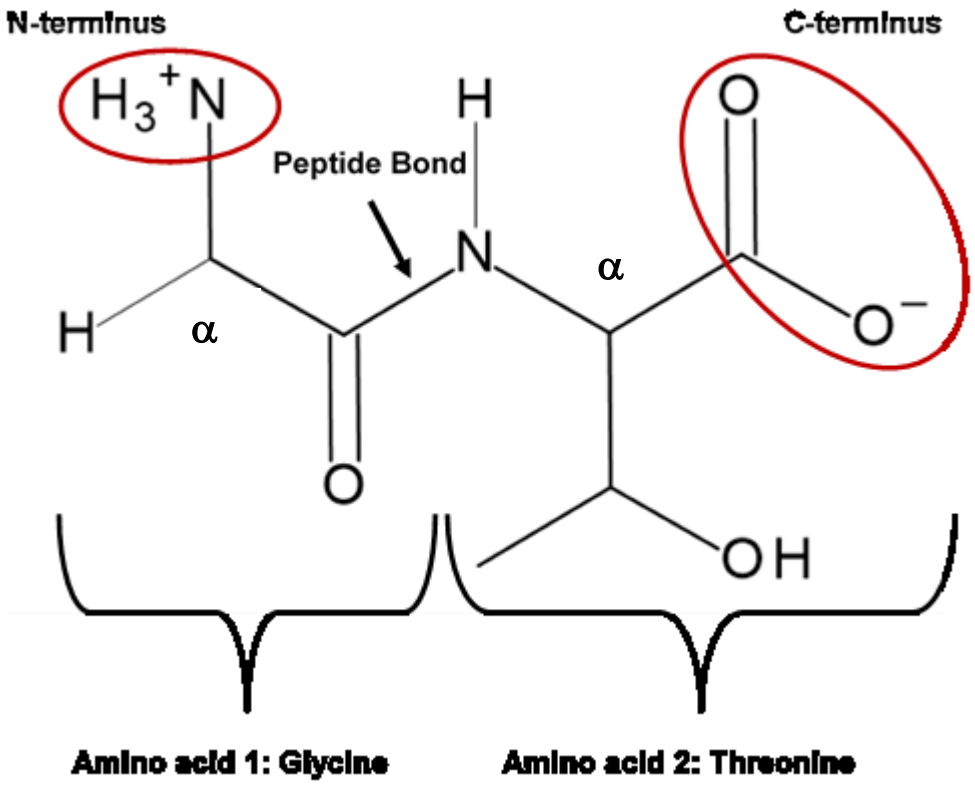


Figure 1-1 An illustration of a peptide formed by two amino acids, glycine and threonine. Note that this figure is shown at a neutral pH of 7, so the amine group (N-terminus) of glycine is protonated, and the carboxylic acid (C-terminus) of threonine is deprotonated.

There are two key components to the analysis of a proteome sample. The first is the utilisation of separation techniques to isolate proteins and the second is the actual detection and identification of the protein sequences. Consequently, it can be argued that the precursor of proteomics occurred in 1951 with Dr. Fredrich Sanger.[14] Sanger's work combined the core principles of separation and identification. His progress on fully sequencing the protein insulin spans a decade's worth of publications beginning with the isolation of alpha and gamma protein chains utilising gel electrophoresis; then, identifying the amino acids through sequencing both the individual chains and insights into protein conformation.[15–18] Sanger's techniques, together with those of Edman[19], who also developed a method of sequence determination, define a research era; in the mid

twentieth century research was primarily focussed on isolating individual proteins through various separatory techniques, and then analysing the structure and composition of protein isolates.

It was not until the onset of the digital age that modern proteomic analysis came to the fore. The advancements in the design of detection platforms such as mass spectrometry (MS) and separations devices now allow for the analysis of proteins and peptides at a significantly faster rate than the methods developed by Sanger and Edman.[20–22] This development was in part possible due to the exponential developments of computing. The development of this processing power allows the researcher to collect and store the large amounts of mass spectrometric data, and allows for the verification of proteomic sequences using sequencing algorithms such as SEQUEST.[20–23] The SEQUEST algorithm matches mass spectra to libraries of protein sequences which are often derived from the genetic blueprint of the target organism.[23] As a direct product of the genetic code, the proteins associated with disease states known as biomarkers, now have the potential to be monitored on a large scale. This allows for insight into the modifications, activities, and relative abundances of proteins and provides the opportunity for improved diagnostics, and monitoring of disease.[24] However, the applications of proteomic techniques extend beyond pathology as they also provide an understanding of cellular biology, and allow for pharmaceutical companies to screen for drug targets, and model responses to drug therapies.[24,25]

As of April 2018 proteomic research has fully shifted into the age of big data with approximately 75,000 proteins catalogued and detected from *homo sapiens* on the UNIProt Database alone, exceeding the number of genes found within our DNA.[26,27] The

number of identified proteins exceeds the number of genes due to modifications which occur during and after the transcription process; consequently, the term proteoform is utilised to describe a sub family of proteins encoded by the same gene.[28] While 75,000 proteins may sound extensive, the database is incomplete with some estimates such as those by Ponomarenko *et al.*,[28] who suggest that the human genome has the potential to encode several million proteins. While this seems quite large, the complexity of the proteome sample continues to make total characterisation of any proteome elusive. Despite the continued advancement in mass spectrometry as a detection method, the preparation of samples for analysis, and an understanding of how these methods work remains as much of a priority now as it was for Edman and Sanger.

1.2 Identification of Proteins

Though a variety of techniques exist to characterize proteins, the most robust and utilised instrument for modern protein analysis is the mass spectrometer. In general, proteomics utilises tandem mass spectrometry where a parent ion is first isolated and then fragmented to determine the order of amino acids making up the protein or peptide, typically by comparison to predetermined protein databases.[20,21] This section will describe the principles and approaches to mass spectrometric analysis as well as a brief discussion on how key components of the instrumentation work for the analysis of proteins.

1.2.1 Bottom-up vs. Top-Down Proteomics.

In general, all MS detection methods in proteomics fall under one of two categories; either top-down or bottom-up MS analysis. Top-down proteomics is the analysis of intact proteins by mass spectrometry and can be used to gain insight on both

protein modifications and the sequences of smaller proteins.[25] However, intact analysis of proteins often experiences challenges with solubility of large or hydrophobic proteins prior to their introduction to the mass spectrometer.[29] Additionally, this method often experiences challenges with dynamic range (an instrument's ability to discern low abundance proteins amongst several high abundance proteins).[30,31] While the shortcomings in the instrumental dynamic range are being addressed by the manufacturers of mass spectrometers, the best way of tackling challenges with dynamic range in the lab is to improve the separation of proteins prior to mass spectrometric analysis in order to reduce the overlap between proteins allowing for low abundance proteins to be discerned.[21,30,32,33]

As an alternative to top-down proteomics, bottom-up analysis is the analysis of proteins which have undergone proteolytic cleavage (digestion) to produce smaller amino acid chains (approximately 50 amino acids or less) or peptides.[12] For the most part, peptides are more soluble than their intact 'parent' proteins, allowing for a variety of separation methods to be performed. However, digestion of sample reduces the information on some proteins composed through alternative splicing, and post-translational modifications.[18,19,] Despite these shortcomings, bottom-up techniques are most popular due to the increased number of identifications of complex proteomes, compared to top-down proteomics.[22,24,29,34] To ensure the best results from bottom-up analysis, the separations resolution, which is the ability to identify and discern peaks in an instrument, must be optimized. [22]

1.2.2 Theoretical Aspects of Mass Spectrometry

Since the creation of mass spectrometry by J.J Thomson,[35,36] this method has developed into one of the most popular methods for the identification of unknown proteomic analytes.[24,37] Current MS instruments include an ionisation source which ionizes analytes, and a mass analyser which separates gas phase ions either through the use of scanning magnetic fields, or electrostatic interactions according to their mass to charge ratio (*m/z*), prior to detection to form a mass spectrum.[30,38] The overall performance of a mass spectrometer is evaluated according to three key components: sensitivity, limit of detection and mass resolution, and their corresponding uncertainties: signal-to-noise, dynamic range and mass accuracy. [27]

Sensitivity describes the relationship between signal intensity strength and the concentration of an analyte. This differs from the limit of detection (LOD) which describes the smallest concentration of analyte required to be confidently distinguished from background noise. Consequently, improving the sensitivity will lower the limit of detection. Both sensitivity and limit of detection are metrics of signal intensity and noise, while mass resolution is measured according to the full width of a spectral curve measured at half the maximum amplitude.

The uncertainty of signal intensity within a mass spectrometer is evaluated through dynamic range, signal-to-noise ratio and signal to background, where dynamic range is the ratio between the most intense signal and the least intense signal which are both simultaneously detected.[30] This is important to proteomics, as the complex samples often contain a mixture of high-mass and low-mass analytes. Another form of uncertainty, signal-to-noise (S/N) is a measure of error in the signal intensity of an

instrument caused by the electronics of the instrument.[30] A high signal-to-noise ratio increases the precision associated with quantitation and lowers the limit of detection. However, it is important to note that signal-to-noise is not a measure of uncertainty caused by matrix effects. Uncertainty in the matrix, called signal-to-background, is often introduced through contaminants, and in some cases exasperated by the ionisation source such as electrospray ionisation (Section 1.3.1)[30] This is particularly important in proteomics where the signal-to-background noise can be a significant source of error due to the number of competitor ions and adducts such as salts, surfactants, and buffers which are used as part of sample preparative steps. In some mass spectrometers such as orbitrap[39] and fourier-transform ion cyclotron resonance[40] (FT ICR), the uncertainty in signal intensity can be improved through the reduction of scan speed within the instrument to improve resolution.

In detail, resolution in terms of mass spectrometry describes both the mass accuracy, which relates the ability of an instrument to deliver a true molecular formula from a target analyte (Equation 1-1) and the overall resolving power of an instrument.[30] The ability to discern neighbouring peaks from one another (resolving power) is particularly important for the identification of peptides with multiple charges. However, adjustments to the resolving power to scanning mass spectrometers such as quadrupoles and ion traps should be limited as it reduces the percent transmittance, thus decreasing the overall signal intensity of the instrument. Consequently, adjustments to scan speed and resolving power are often a compromise between optimal resolution and optimal sensitivity. Given that both signal intensity and resolution are essential for the proper identification and the quantification of proteomic analytes, the development of

instruments which are both sensitive and have high resolution is particularly important to proteomics.

$$\textbf{Mass Accuracy} = m/z_{\text{Theoretical}} - m/z_{\text{Experimental}} \qquad \textbf{Equation 1-1}$$

The type and quality of spectra obtained from a mass spectrometer is dependent on the type of ion source and the mass analyser. For example, the development of instruments such as the Quadrupole Time of Flight[41] (qTOF), Orbitrap[39], and Fourier Transform Ion Cyclotron Resonance[40] all have a high resolution and are highly sensitive. This allows for fast scan speeds with improved signal-to-noise ratio and limit of detection compared to early mass spectrometers, such as the linear ion trap quadrupole (LTQ), the instrument used in this thesis.[42] This is particularly advantageous, as it allows for smaller amounts of sample to be injected into the mass spectrometer, and with a larger number of distinguishable peptides.

1.2.2.1 Ionisation sources

As mass spectrometry has developed, a variety of ionisation techniques have emerged to allow for a wider array of applications. Ionisation techniques can be subdivided into hard ionisation sources which require analytes to already be in the gas phase, and soft condensed phase ionisation sources. Hard gas phase ionisation sources include the early sources such as electron ionisation and though still in use today, are generally incompatible with proteomics. Due to the challenge associated with getting large biological molecules into the gas phase prior to ionisation, and the fragmentation which occurs on long chains during the ionisation process, a hard ionisation process is less than desirable. Prior to 1980, the lack of a soft condensed phase ionisation interface to facilitate the transition between liquid and gas phase was one of the key obstacles

preventing biological science from capitalizing on the benefits of mass spectrometry – this changed with the advent of electrospray ionisation (ESI)[43,44] and matrix-assisted laser desorption/ionisation (MALDI).[45,46]

The development of ESI by John Fenn allowed large thermally labile ions to transition from a liquid to a gaseous state for analysis by mass spectrometry.[38,43,44] ESI in particular has become the most popular ionisation source utilised in peptide and protein analysis today as it allows for the direct analysis of proteomic separations through liquid chromatography by facilitating the transition of large biological molecules from the liquid to gas phase.[30] As an added benefit for analysis of large biomolecules, the soft ionisation technique provided by ESI allows proteins and peptides to acquire multiple charges to transition into the gaseous phase without fragmentation prior to entry into the mass analyser.[38] However, this ionisation process can result in a decreased signal to background, reducing the instrumental sensitivity, and increasing the limit of detection when samples are not prepared appropriately. Consequently, it is not enough to simply have a good mass spectrometer and an interface to analyse complex proteomic samples; there must be a deeper understanding of how sample preparation can impact the quality of analysis.

1.2.2.2 Electrospray Ionisation and Matrix Effects

Electrospray ionisation begins with the injection of an analyte diluted into a volatile solvent continuously flowing into a spray capillary. This capillary has a potential difference of 1 to 5 kV relative to a counter electrode (the mass spectrometer) and establishes an electric field.[30,47] This electric field deforms the solvent meniscus at the end of the spray capillary to form a Taylor cone.[30,45] For a positive applied voltage at the

spray capillary, the tip of the cone has a high positive charge density which results in the formation of a small jet (containing an excess of similarly charged ions) which eventually overcomes surface tension to form a series of droplets that disperse through columbic forces.[30,45]

When we consider an electrospray system running in positive mode, negatively charged functional groups such as carboxylates are protonated, leaving a series of positively charged droplets travelling through the electric field towards the inlet of the mass spectrometer. The release of ionised proteins and peptides into the gas phase is best described through the charge residue model.[48] As charged droplets approach the inlet, the volatile organic solvent evaporates and the charge density increases. Eventually these droplets reach the Rayleigh limit, where electrostatic repulsion overcomes the surface tension, initiating a cascade into a series of ever smaller charged droplets until solitary ions are ejected into the gas phase.[30,38] Although this ionisation method is highly effective, the final ion cascade is often the source of ion suppression caused by matrix effects which can be exasperated by direct analysis methods due to improper solvent choice or sample clean-up.

In general, proteomic samples often experience problems with decreased signal to background ratio and MS resolution due to three key mechanisms of adduct formation which are known to decrease signal-to-noise in electrospray ionisation, and reduce resolution.[48] In positive mode the adduct formation processes are known as cationisation, salt adduction and non-specific ligand binding.[48] In cationisation, cations such as sodium or potassium bind to carboxylates instead of protons, spreading the ion spectral count from a single intense peak, to a wide range of adduct peaks, otherwise known as mass

heterogeneity (Figure 1-2A).[48–50] This causes a decrease in signal to background ratio and reduces the overall resolution of the instrument. Cationisation differs from interference due to salt adduction which is caused by the interaction of counterions within the charged droplets.[33] In particular, counter ions (such as Cl^-) impact the surface tension of the droplets by reducing charge repulsion.[33,51] This decreases the rate of droplet shrinkage by non-specific ion-pairing which is favoured as the ion concentration increases due to solvent evaporation. Ion-pairings between cations such as K^+, and counter ions such as Cl^-, can also occur to form salts leading to the formation of clustered adduct interactions with proteins, which once again leads to mass heterogeneity and increased measurement uncertainty (Figure 1-2B).[48] Finally, non-specific ligand binding occurs when proteomic analytes form adducts with residual compounds in solution. Some examples of these ligands include surfactants or buffers utilized during previous steps of sample preparation.[48] Without proper methods of sample preparation, these ligands will form adducts. Consequently, desalting (also known as cleaning) proteomic samples should be considered essential prior to mass spectrometric analysis.[30,38]

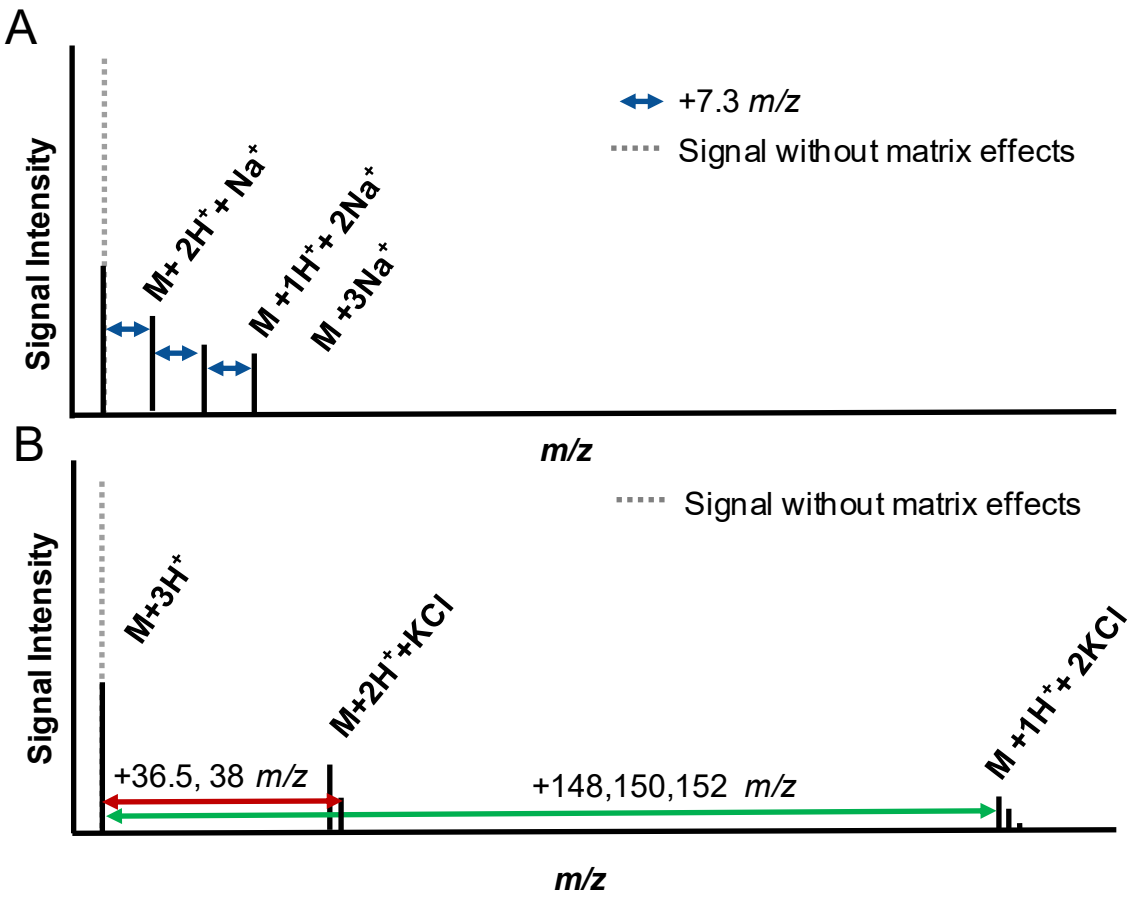


Figure 1-2 The impact of adduct formation on the signal-to-background for $M+3H^+$ ions. In particular, cationisation (**A**) is shown using Na^+, salt adduction (**B**) is shown using KCl. Note the shifts caused by the KCl has multiple mass shifts for each change in net charge resulting from Cl isotopes.

1.3 Separation of Proteins.

While solvent and sample compatibility with mass spectrometry can make sample preparatory steps like separations sound less desirable, they are an integral part of proteomics analysis.[20] Without some form of separation, tens of thousands of proteins or peptides of varying concentration would co-elute into the mass spectrometer for detection at once.[52] In addition to the need to separate a complex proteome into fractions and clean protein samples ahead of analysis, it is also important to conserve the sample. Each preparative step in the process of analysing a proteomic sample is subject to sample loss which compounds and reduces the ability to identify low abundance analytes in the sample.[53,54] Consequently, the separations process in proteomic mixtures must be studied in order to effectively produce preparative workflows that not only conserve the integrity of proteomic samples, but are also efficient and effective at providing good separations that are compatible with mass spectrometry.[21,55]

There are two types of separation methods in the context of mass spectrometry: offline separations which are not directly coupled to the MS, and online separations which feed compounds into the MS as they are separated in real time.[21] There are many advantages to both methods, and many would argue that online separation methods are preferential to reduce loss in sample due to poor protein recovery in preparative steps.[52,56,57] However, online methods often require compromising the quality of the separation to ensure that a method's solvents are compatible with ionisation interfaces. This often results in poor separation which can reduce the number of identifications, calling into question the purpose of performing such a separation in the first place. Additionally, the length of time and the labour involved in the offline steps of a

proteomic workflow make the analysis process as a whole less accessible, and less user friendly.

1.3.1 Electrophoresis

The first major development in terms of protein separation came with the development of electrophoresis by Arne Tiselius in 1926. Tiselius later applied his work to the separation of blood serum proteins for which he earned the first Nobel prize associated with proteomic analysis in 1948.[58–60] Since then, many different forms of this technique have been developed by the scientific community. The four types of electrophoresis of particular importance to modern proteomics are isoelectric focusing (IEF)[61] in gels, sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE),[62] gel elution liquid fraction entrapment electrophoresis (GELFrEE)[33] and two-dimensional electrophoretic processes[63] and are depicted in Figure 1-3. However, the underlying principles governing how these electrophoresis techniques work remains the same. Charged analytes are suspended in a fluid or gel medium and are exposed to an electric field which induces the movement of charged analytes through the media. The rate of migration of each analyte through the media is determined by its molecular size and charge, resulting in separation. In modern methods of gel-electrophoresis, separated proteins can be visualised through staining, and targeted protein bands can be excised from the gel.[29] The proteins are often extracted from the gel following enzymatic digestion, most commonly through the use of trypsin which cleaves at the C-terminal side of lysine (Lys) and arginine (Arg) residues.[13]

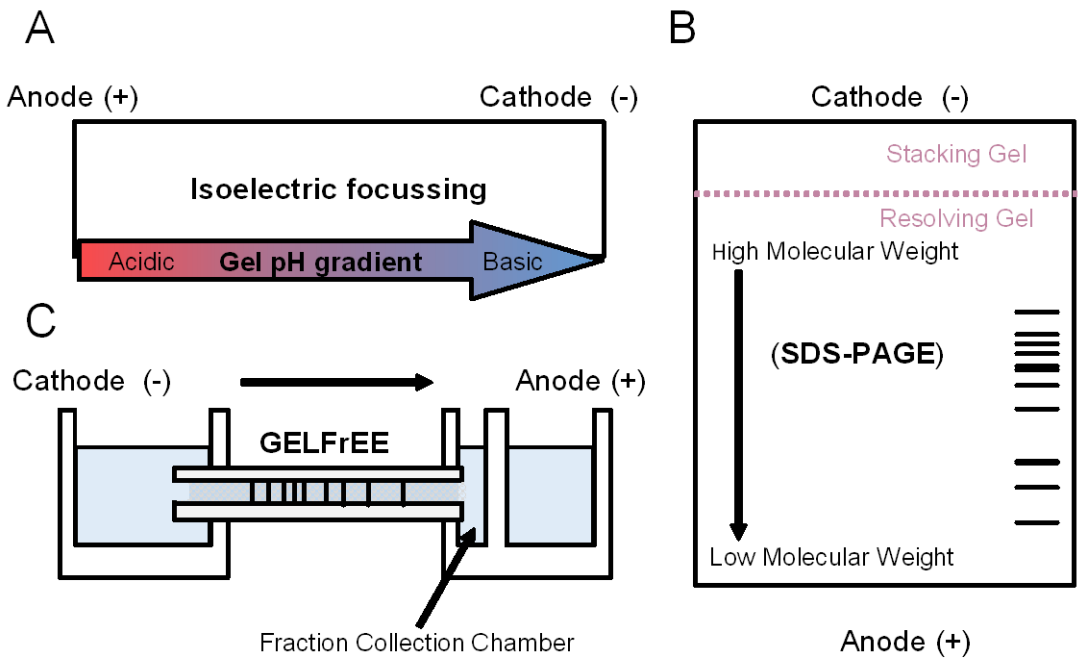


Figure 1-3 A diagram summarising three key methods of electrophoresis. (**A**) Isoelectric focussing which relies on a pH gradient embedded in the gel for separation of proteins based on isoelectric point. (**B**) SDS-PAGE separates proteins according to their molecular weight with smaller proteins travelling through the gel matrix faster than larger proteins. (**C**) GELFrEE uses an SDS-PAGE gel matrix in a tube which unloads into a fraction collection chamber allowing the researcher to collect intact proteins at various time points.

1.3.1.1 Isoelectric Focusing

Isoelectric focusing is a form of zone electrophoresis which uses a pH gradient embedded within a gel media with a more acidic region being located at the anode (Figure 1-2). [64,65] The separation exploits the polyprotic nature of a protein to obtain separation. In detail, all peptides contain at least two ionizable components - the C-terminus (approx. pKa 3-4), and the N-terminus (approx. pKa 9). Additionally, seven of the twenty main amino acids (cysteine, tyrosine, aspartic acid (Asp), glutamic acid (Glu), Lys, Arg, and histidine (His)) also have side chains which contain weakly acidic or basic functional groups.[66]

Thus, it can be expected that the net charge of a protein will change according to the pH of the surrounding environment.[64,65]

In isoelectric focusing, the application of an electric field causes proteins to migrate through the gel according to its electric charge (cations towards the cathode, and anions towards the anode), according to the pH gradient within the gel. For example, a protein starting in the acidic region will have a positive net charge and will begin to migrate towards the cathode.[64,65,67] As the migration occurs, the environments become progressively more basic resulting in the deprotonation of some amino acid side chains, and the C-terminus until the isoelectric point (the point at which the net charge of the protein is zero) is reached.[64,65,67] Once the protein has no net charge, the migration of the protein stops. Diffusion in either direction will result in the reformation of a charge which will force movement in the opposite direction, hence the name isoelectric focussing.[64,65] In general, the average resolution of an IEF gel is defined by the slope of the pH gradient, thus longer gels will have a shallower slope and a significantly higher resolution.[68] However, long gels are prone to the migration of defined pH bands within the gel due to the longer run times and requires a high electric field which can affect the sample due to joule heating.[63,67,68] Consequently, isoelectric focusing is often utilised as a form of prefractionation of complex proteomes rather than the primary mode of separation.[64,65]

1.3.1.2 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis is the most widely utilised electrophoresis technique with approximately 4.76 new publications uploaded to PubMed *per diem* in 2017 alone.[69] The technique utilises sodium dodecyl sulfate, an amphiphilic surfactant which is composed of a strongly anionic polar head (sulfate) and a

long hydrophobic tail (twelve carbon chain). These detergents increase a protein's overall solubility that aids the denaturisation process of the hydrophobic regions of proteins, and confers a net negative charge onto them. This ensures that all proteins will migrate towards the anode through the polyacrylamide gel upon the application of an electric field. The overall separation of proteins arises from their varying molecular weight where smaller proteins migrate through the gel matrix faster than those which are larger.[67] This method when run alongside a molecular weight ladder, allows the researcher to estimate the molecular weight of the proteins within the proteome, but proper identification still requires mass spectrometry.[67] Samples are recovered through isolating specific spots in the gel, or through where a lane is divided into fractions and are digested, to account for challenges with resolving a proteome which exceeds 10,000 proteins. Additionally the recovery of peptides, especially those which are hydrophobic from gels, is variable.[70] This impacts the recovery and analysis of low abundance proteins using mass spectrometry. Despite this, SDS-PAGE remains a popular separation tool for proteome analysis by mass spectrometry.

1.3.1.3 Two-Dimensional Gel Electrophoresis.

As is the case with almost all proteomic separations, one mode of electrophoretic separation is insufficient in providing complete resolution of the components of proteomic samples. While complete resolution is not essential, complex proteomes have a lot of overlap which can be reduced by additional separation. This resulted in the development of two-dimensional gel electrophoresis (2D-GE) by O'Farrell and Klose in 1975.[63] Two-dimensional separation requires orthogonal modes of separation as is seen through coupling of IEF with SDS-PAGE. This process has been known to separate

between 1500 and 10,000 proteins per gel; however, a resolution of 2000 proteins is considered the average.[31,68] While this is high, it is insufficient for the separation of complex proteomes which can exceed 30,000 proteins. Additional challenges to this separation method include: the gel-to-gel reproducibility, the limited dynamic range in terms of molecular size and isoelectric point which impacts the staining of the gel, the amount of time required for this separation process and the technical skill required to perform this separation.[47,67,50,52,53]

1.3.1.4 GELFrEE

Gel Elution Liquid Fraction Entrapment Electrophoresis, GELFrEE, is a technique developed in the Doucette group to address the poor sample recovery of intact proteins from SDS-PAGE.[73] The device consists of two chambers filled with running buffer solution and connected by a glass cylinder filled with the same discontinuous polyacrylamide gel matrix as found in SDS-PAGE (Figure 1-2B). Unlike SDS-PAGE where sample is separated within the gel, stained and excised, proteins in GELFrEE are allowed to migrate out of the gel into a small chamber where the proteins are collected as time-based fractions. [73]

While GELFrEE does not generally result in the collection of a single protein within any given fraction, excised proteomic separations in SDS-PAGE are also rarely discrete. GELFrEE removes the in-gel digestion process which allows the user to separate intact proteins, increasing the efficiency of separation and the overall protein recovery. However, the proteins within these fractions must undergo SDS depletion methods prior to analysis in mass spectrometry to prevent the formation of adducts.[49]

1.3.1.5 Summary of Electrophoresis and its Role in Proteomics.

While electrophoresis remains a key method for separation in proteomics, there are some features of the technique which make it a less-than-desirable approach to quantify and identify proteins. In particular, many gel-based electrophoresis techniques rely on the ability to visualize proteins in order to excise key areas of interest from the gel. This may not be possible with low abundance proteins. Excised portions of gel require digestion in order to extract the protein, a process which has challenges recovering large peptides.[64] This decreases the overall recovery of the sample which is particularly deleterious for samples obtained from IEF, or SDS-PAGE. However, the main disadvantage with electrophoresis is the challenges creating an automated, online user-friendly interface, with large loading capacities. This differs drastically from methods such as high-performance liquid chromatography which can perform multiple separations and collect fractions in an automated fashion with a few clicks of a button.

1.3.2 High Performance Liquid-Chromatography (HPLC)

It was not until the development of ESI which couples liquid chromatography (LC) to mass spectrometry, that LC began to compete with electrophoresis as a preferred separation method. Regardless of its use as an online or offline separation system, HPLC relies on the separation of compounds between two phases: a stationary phase held in place inside a column, and a mobile phase which carries target analytes as they separate through the column. The support matrix of the stationary phase should have minimal interaction with the target analyte. Consequently, the support is often charge neutral with a low specific binding capacity. The support matrices utilised within this thesis are rigid

solid silica particles (Jupiter), and polymer based (POROS) resins. Both support resins can withstand high pressure associated with HPLC. However, the size, surface area and permeability of these particles are of importance to the capacity and resolution of the separation. As described in Equation 1-2, where the resolution of a column is the product of it's efficiency, selectivity and capacitance. The efficiency (width of each peak) of the column is proportional to the length of the column (*L)* and inversely proportional to the diameter of the particle (d_p). The selectivity of the column is determined by the relative retention of the peaks (α) and the capacitance of the column, and the retention is proportional to the capacity factor for a component held on the column (k'_x) and inversely proportional to the average capacitance of the analytes on the column (k'_{av}) .

$$\textbf{Resolution} = \underbrace{\left(\frac{\sqrt{3500\left(\frac{L(cm)}{d_p(\mu m)}\right)}}{4}\right)}_{\textbf{Efficiency}} \underbrace{\left(\frac{\alpha+1}{\alpha}\right)}_{\textbf{Selectivity}} \underbrace{\left(\frac{k'_x}{1+k'_{av}}\right)}_{\textbf{Capacitance}}$$

Equation 1-2

The POROS resins are composed of a polystyrene divinyl benzene backbone with short diffuse pockets that diverge from large channel pores which bisect the particle. The porosity of the overall resin particle can lead to irreversible adsorption or poor column capacitance. The polymer which forms the support matrix of the POROS R2 resin is responsible for defining the mode of separation. In contrast, bonded phase packing materials, rely on a bonding phase which attaches to the support matrix. This determines the functionality of the resin and is what defines each mode of separation. For example, normal phase resins have polar functional groups affixed to the stationary phase while

reversed phase resin contains non-polar groups attached to the support matrix. The two chromatographic modes of separation which are relevant to this thesis are reversed phase (RP) and strong cation exchange chromatography (SCX).

1.3.2.1 Reversed Phase Chromatography

Reversed phase chromatography is one of the most popular chromatographic methods for proteomic separations because of the high resolving power of the mode and its compatibility with ESI-MS.[31] Non-polar polypeptides adsorb readily to the hydrophobic stationary phase under aqueous polar mobile phase conditions, and desorb from the resin as the mobile phase increases in hydrophobicity through the incorporation of organic modifier (most commonly acetonitrile).[31,34,45,74] This method is extremely versatile and can be utilised for sample clean-up through utilizing displacement methods for chromatography where protein or peptide samples are injected under low organic solvent composition. Proteins and peptides, which are more hydrophobic than the mobile phase, are retained while the original solvent matrix readily passes through the column.

The majority of RP-HPLC utilises resins with a pore size of approximately 100 Å; however, the majority of the resins in proteomics have larger diameters. For example, the Jupiter resin and the POROS R2 resin utilised in this thesis have pore sizes of 300 Å and 4000 Å, respectively, which is more amenable for large biomolecules. This in part is due to the nature of the resin where smaller diameter particles in the Jupiter resin composed of hard silica surface will result in sharper peaks and higher resolution. However, the diffuse microchannels and large pores in the POROS R2 act to decrease the back pressure and increase the surface area, increasing the overall binding capacity of the resin.

The mobile phase of reversed phase separation can be optimised through the choice in organic modifier and the addition of ion-pairing agents. The most common organic modifier is acetonitrile because it does not adsorb in the UV region, and reduces column back pressure due to the decreased viscosity of the solvent. Ion-pairing agents are utilized to improve the solubility and improve the peak shape of separated samples. In particular, trifluoroacetic acid (TFA) is often utilised in offline systems due to its volatility. However, formic acid is utilised in online systems because TFA forms adducts with positive charged analytes causing signal suppression in ESI.

1.3.2.2 Ion Exchange Chromatography

Ion exchange chromatography is one of the most versatile separation modes available to proteomics.[31] There are two types of ion exchange resins: anion exchange resins which have positively charged basic functional groups bound to the stationary phase, and cation exchange resins which contain negatively charged conjugate bases of acidic functional groups.[13,31,45] These two types of ion exchange are subdivided further into strong ion exchangers and weak ion exchangers. Strong ion exchanger have functional groups which maintain their charge over a wide pH range.[13] Weak ion exchangers contain weakly acidic or basic functional groups and can become protonated or deprotonated depending on the column pH conditions.[13,31,45] In proteomics the most popular form of ion exchange is strong cation exchange, which is also the focus of this thesis.[13] The concepts influencing ion exchange are introduced in this chapter and discussed further in Chapter 2.

The ion exchange process relies on the exploitation of coulombic interactions between the analyte and the stationary phase as seen in Equation 1-2.

$$F = \frac{Z_a Z_S}{Dr^2}$$ **Equation 1-3**

The force of the interaction (F) between an analyte and the stationary phase is proportional to the charge of both the analyte (Z_a) and the stationary phase (Z_s) and is inversely proportional to the dielectric constant (D), and the distance between each of the charges (r). As proteins are polyprotic, the acidity of the mobile phase can be utilised to alter the affinity for the column by altering the charge state of a protein. The retention between the stationary phase and protein analytes is strongest under acidic conditions and weakest after the pH has increased past the isoelectric point of the polypeptide chain.[31]

The choice of solvent is an essential component of proper separation in ion exchange. Buffer choice has a strong influence on retention. For example, buffers which have an affinity to the media are a poor choice, while buffers that are readily soluble and have a high buffering capacity are good candidates for the mobile phase.[31] In addition to buffers, non-buffering salts can be added to reduce the force of the coulombic interactions between the stationary phase and analytes.[75] These competitive interactions can be beneficial when they are used to remove analytes from the column but are a hindrance when they are remaining in the sample due to poor sample preparation. For example, the presence of ionic detergents that are left over from sample preparation will disrupt the coulombic interactions between peptides and the resin. This disruption will reduce the retention of the analytes.[31,76]

1.3.2.3 Multidimensional-Dimensional Chromatography

Unfortunately, like in the case of one dimensional electrophoresis, the resolving power of one dimension of HPLC is often insufficient in terms of resolution. Consequently, different techniques which pair other separation methods with HPLC such

as GeLC (it pairs gel electrophoresis with liquid chromatography), and multidimensional liquid chromatography (MDLC) solutions are often utilised in proteomics. When utilised correctly, MDLC can have positive impacts in terms of identification of proteins. In particular, the MDLC peak capacity is multiplicative.[77,78] In this thesis for example, if a strong cation exchange separation method collected in ten even fractions (a peak capacity of ten), and then further separated it using a second column with a peak capacity of 600 (as is in the case of nano revered phase capillary columns)[79], the system would be capable of resolving a total of 6000 proteins.[31]

In general, there are two types of MDLC: online and offline chromatography. Online MDLC is the direct pairing of two chromatographic methods one after the other. This can be utilised for sample clean-up or for separations. Online MDLC separation methods have the potential for very high resolution. However, this assumes that the two modes of separation are orthogonal, and do not have any challenges with peak broadening during the separation. In terms of mass spectrometry, one form of MDLC, Multidimensional Protein Identification Technology (MuDPIT), is considered the gold standard of separation methods in proteomics.[80] This separation method utilises a biphasic column where the first phase is SCX, and the second phase is RP chromatography. This method relies on the complete retention of the digested proteome in the first dimension. Elution from the first dimension is discrete and stepwise with each step of separation in the first dimension followed by a gradient to separate the fraction on reversed phase. As with any direct analysis separations method in mass spectrometry, this method must minimize adduct formation which can reduce the quality of separation in the first dimension.

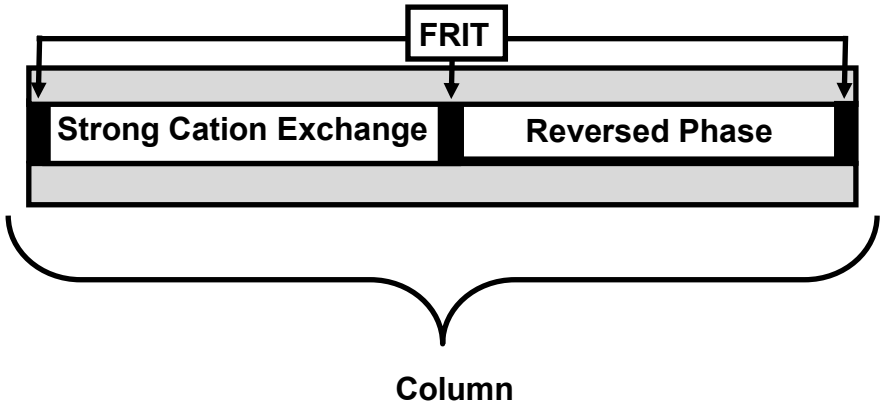


Figure 1- 4 A diagram outlining the contents of a biphasic MuDPIT column (grey), which contains two 'orthogonal' modes of separation, SCX, and RP (white) held in place through frits (black).

In contrast to online two-dimensional separation, offline two-dimensional separation allows for both modes of separation to be run independently. Samples are collected from the first separation in fractions before they are further separated and detected in the second dimension. This allows for sample processing after the first separation. One popular sample-processing step involves desalting and drying of proteomic fractions prior to reconstituting samples in an appropriate solvent for additional separations. This increases the compatibility with mass spectrometry. While binning fractions result in a reduction of peak resolution, the optimisation of separation in the first dimension can allow for the isolation of key proteins into a specific fraction and reduce the total run time for the analysis of the proteome.

1.4 Tying Mass Spectrometry and Separation together.

Development in both separations and mass spectrometry plays a key role in proteomics. During 1999, Gygi *et al.*[81] performed a sequence identification of proteins by using isotope encoded affinity tagging (ICAT) where half of a proteomic sample was exposed to an isotopically enriched reagent, chemically linked to biotin while the other

was exposed to the light biotin reagent. The biotin reagent interacts with the sulfhydryl groups of cysteine found in an amino acid chain. The samples were mixed and then peptides were passed through a column which bound to all peptides tagged with the biotin. This allowed Gygi *et al.* to reduce the sample complexity from 344,855 to 30,619 peptides, with a total of 800 peptides detected in total on an ion trap mass spectrometer.[81] Three years later the first MuDPIT method was published with Washburn *et al.* identifying 1,347 peptides using an ion trap.[80] When the same MuDPIT method was repeated by Webb[82] 12 years later, using an orbitrap with a better mass accuracy and MS/MS acquisition speed, a total of 19,173 peptides identified using the old method. This increase in the total number of protein identifications by an order of magnitude.[82] This increase is due to the improvements in mass accuracy and MS/MS acquisition speed. Given these increases, it is almost understandable why in proteomics some believe that improvements to mass spectrometry negate the need for separation. However, when plotted as a ratio of the number of identifications per minute (ID/min) in Figure 1-5, a different story takes place. Newer instruments with a faster MS/MS acquisition speed and better mass accuracy should have a higher average rate of identification (number of total peptide identifications (ID) divided by the length of analysis time) than older instruments. However, when you compare the rate of identification for ICAT (13.3 ID/minute) using an older MS, to that of standard MuDPIT (13.3 ID/min) performed by Webb,[82] the rate of identification is equivalent. This demonstrates that the standard MuDPIT as a separation method does not separate peptides in a way that allows the high-resolution instrument to reach its full potential. This emphasizes why chromatographic resolution is still important for mass spectrometry.

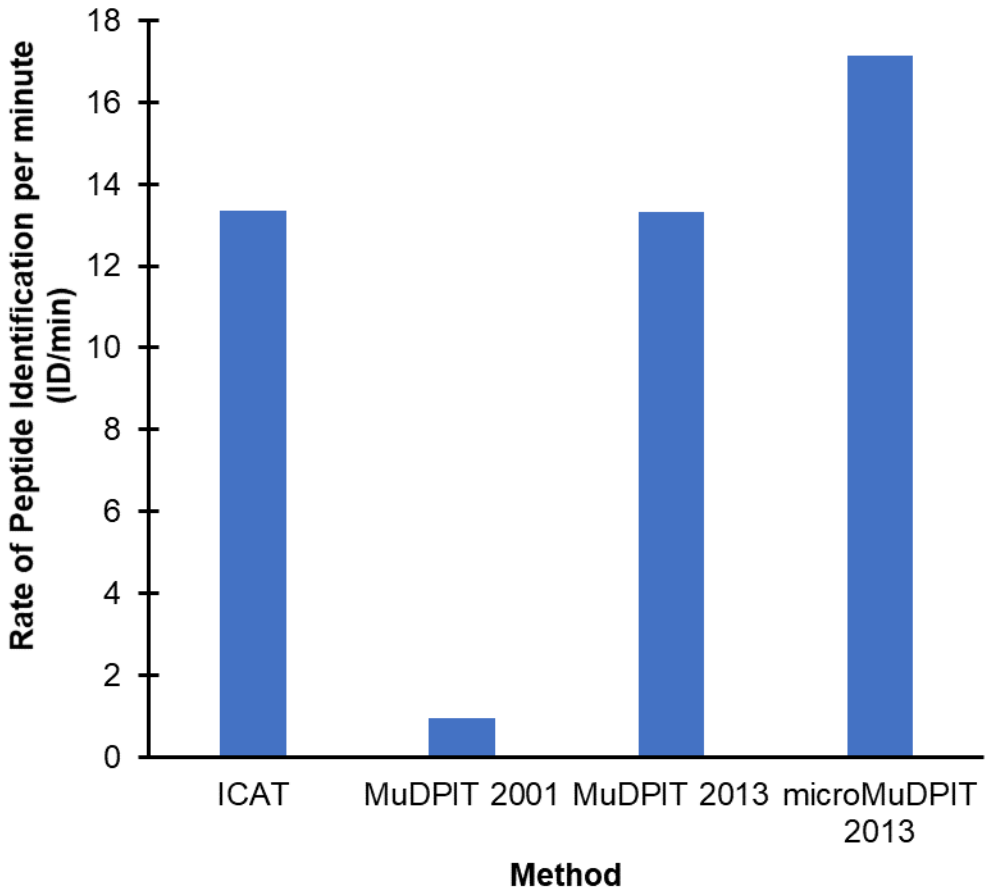


Figure 1-5 The average rate of identification for four different mass spectrometric experiments. Despite the fact that ICAT[81] was performed on a lower resolution mass spectrometer, the peptide identification rate (13 ID/min) is similar to that of the standard MuDPIT[80,82] method performed on a high-resolution mass spectrometer in 2013.

The improvements in terms of scan speed and resolution have and continue to make a huge impact in proteomics. Given that modern instruments are capable of collecting and processing 20 spectra per second, a 75-minute separation holds the potential for sequencing up to 90,000 peptides. However, when considering a more complex protein digest with 500,000 peptides, the same 75-minute run would need to identify approximately 111 peptides a second. Alternatively, with proper two dimensional separation, there is a reduction in peptide overlap, while increases the number of identifications on the mass spectrometer.

1.5 Summary and Research Objectives

Given that proteomic samples are inherently complex, the identification of a proteome is reliant both on the quality of the mass spectrometer, and the separation processes which reduces the total overlap between analytes during each run and enhances resolution. Without a high-resolution mass spectrometer, a good signal-to-noise ratio, and large dynamic range, the number of identifications and the accuracy of those samples would be reduced. Matrix effects can be mitigated through the appropriate use of separation which not only eliminate competing adducts, but also reduce sample complexity - a fact that is often ignored in modern proteomics where mass spectrometers can appear to be the focus, and separation an afterthought. This is apparent in many approaches to multidimensional separation where the key variables which govern separation of peptides on SCX resins are often ignored. Chromatographic theory should not be overlooked as an essential component of the proteomic researcher's toolbox.

SCX as a mode of separation is an ideal pairing for proteomic samples because it is designed to separate weakly basic compounds, has a high loading capacity, and is orthogonal to RPLC.[31,83] There are many buffer systems to choose from which allow the researcher to tailor the mobile phase of the separation. Despite this, many researchers are publishing results which do not take advantage of the major mode of separation (pH), and instead focus on separations that utilise a salt gradient.[84] This may indicate a lack of understanding on how SCX separates compounds, given that there still is not a formal model which can fully describe how separations with complex molecules such as proteins occur. However, performing full pH gradients in traditional HPLC platforms is also challenging. Binary pumps are limited to mixing two solvents, and quaternary pumps

limit the researcher to mixing four solvent systems. As the buffering capacity is limited to a standard deviation of one pH unit away from the pKa, it is almost impossible to create a gradient which extends the full pH range using two to four solvents, in the traditional HPLC system.

This thesis will discuss the development of two systems which allow the user to easily perform SCX separations of proteomes. Using these systems, an analysis of the major variables that influence the retention of peptides on SCX resins will be performed. Finally, a standard approach to optimising these separations will be developed using the knowledge of how these variables influence separation.

Chapter 2 discusses the development of a centrifugal liquid chromatography device called the ProTrap XG as an offline approach to SCX. This offline system allows the researcher to separate multiple samples simultaneously. This chapter begins by examining the impact of organic solvent, salt and pH on separations using *Escherichia coli* before transitioning to an optimised approach to separations. This is done through creating a multivariate separation which combines both salt and pH to influence retention on the SCX column.

Chapter 3 modifies the traditional HPLC set-up through establishing a separation gradient through an autosampler rather than mixing one using a pump system. This method provides an almost infinite approach to gradient design for SCX in an automated format while simultaneously preparing samples for analysis by mass spectrometry. This SCX separation uses proteins extracted from *Saccharomyces cerevisiae* (Baker's Yeast) to examine the impact of salt, and pH on separation. Once the impact of salt and pH on separation is determined, these variables are combined and a standardized approach for

optimising a separations gradient is proposed, tested and compared to the results from methods used to separate peptides on SCX by the gold standard (MuDPIT) method.

Chapter 2 SCX Separation of Peptides Using the ProTrap XG

2.1 Introduction

Strong cation exchange (SCX) has found great use in the field of proteomics, particularly as the first component in orthogonal separation methods.[85] The coupling of SCX with RP has been popularized with the development of what is now considered a gold standard in online proteomic separation and analysis, the MuDPIT approach.[57] The mechanistic principles of SCX are well suited to the separation of peptides due to their polyprotic nature[13], and both components of the separation method can be performed independently, as is the case in this thesis.

All SCX resins contain negatively charged functional groups which are not influenced by pH. These groups are commonly sulfonate and are attached to a support surface. In the case of this thesis, the support is a polymeric resin backbone consisting of porous crosslinked polystyrene divinylbenzene. The negatively charged functional groups on the surface of the resin interact electrostatically with cations in solution, called counter ions, and become attracted to the stationary phase. Likewise, anionic species in solution undergo electrostatic repulsion with the SCX surface. In general, separation with SCX resin takes place by altering the dynamic equilibria which exists between the stationary support and counter ions. This is done by adjusting one or more components of the mobile phase: the salt concentration, the solution pH or the organic content of the mobile phase solvent as discussed in detail below.

2.1.1 The Impact of Salt in Ion Exchange

Of the three components that influence interactions among analytes and the SCX resin, adjustments to the salt concentration of the mobile phase tend to be the most popular. Separation using salt gradients is described by the 'relative selectivity coefficient' (k)[75,86] which in the case of strong cation exchange quantifies the relative affinity of one counter ion towards the ionic exchange resin over another. Experimentally, the selectivity coefficient can be calculated using the ratio of mole fractions in both the mobile and stationary phase as shown in equation 2.1.[75,86]

$$aA_{stationary} + bB_{mobile} \xrightarrow{k_{a/b}} aA_{mobile} + bB_{stationary}$$

$$k_{a/b} = \frac{[A]^a_{stat}\,[B]^b_{mobile}\,\gamma^b_B}{[A]^a_{mobile}\,\gamma^a_a\,[B]^b_{stat}}$$

Equation 2-4

In this case the equilibria of two counter ions A and B between the stationary and the mobile phase are described by the relative selectivity coefficient ($k_{a/b}$). Where the ionic strength of each counter ion in the mobile phase is described by its activity (in the case of A: $[A]^a_{mobile}\gamma^a_{\mathrm{A}}$).[45] If in this case we assume that A had a greater affinity than B for the stationary phase, increasing the concentration of A in the mobile phase would result in an increased concentration of B in the eluent as equilibrium is restored to the column. Or, in the case of a salt gradient in proteomics, any peptide bound to the column will maintain its interaction with the stationary phase until the affinity of the counter ions in the mobile phase exceed that of the peptide for the resin.[75,86]

For optimal separation, using a salt gradient, peptides should be loaded onto the column in the absence of salt. After the peptides are loaded onto the column, the concentration of counter ions in the mobile phase is gradually increased to elute peptides

in order of their relative affinity for the stationary phase (lowest affinity to highest affinity).[45] This rate of elution and the quality of the separation can be moderated through the choice of cation relative to H^+ through the relative selectivity coefficicient.[76,86,87] Cations with increased relative selectivity compared to H^+ result in a steeper gradient, where the majority of the peptides are eluted quickly, while cations with a lower relative selectivity result in a shallower gradient on a molar by molar basis.

Determining the model for correctly predicting the relative selectivity an ion exchange resin has for one counter ion over a proteomic system continues to a challenge for academia, and lags behind other separation techniques.[88] However, five key trends have been established regarding ion affinity.[45] First, there is a greater affinity for polyvalent ions at low concentrations. Second, the affinity of ions which have the same valency increases with atomic number.[75] These trends can be seen when comparing the selectivity coefficients for the following ions: $Al^{3+} >>> K^+ > NH_4^+ > Na^+ > Li^+$, where K^+ is observed with a greater affinity for the resin than Na^+ due to its larger atomic mass.[45] The listed affinities also show a third trend, the effect of charge density which is the relationship between the charge and the overall ionic radius. Thus, Aluminium (III) has a stronger affinity for ionic exchange resin than the Group I metals.[45] Consequently, the amount of salt required to release a given peptide from the SCX resin is a function of the type of salt used as well as the concentration of that salt as discussed below.[13,45,76] Fourth, larger organic ions tend to have greater affinity than smaller metal ions. The order of increasing ion affinity for a resin remains constant; however, the magnitude of the ions selectivity, and therefore the strength of relative ion selectivity can be altered.

The final way to alter the strength of ion selectivity while maintaining the conditions on the column is through adjusting the pore size of the resin through crosslinking. Polymeric resins with larger amounts of crosslinking have smaller pore sizes.[13] This results in sieve affects where smaller ions with higher charge states (+2, +3…) are selected over lower charge density. The reverse is observed with decreased cross-linking and larger pore sizes where the relative selectivity coefficient decreases with a larger pore size.[76] This is important because pore size for proteomic separations is increased, to allow for the movement of large molecules through the pores, and to allow for additional binding sites within the resin.[13] While this is beneficial in terms of analyte binding, it reduces the deviation between the relative selectivity coefficient of the atomic counter ions, this may be important when using perfusion resins to separate samples.

2.1.2 The Impact of pH on Ion Exchange

In a digested proteome, several thousands of peptides will be produced with various physical and chemical properties. As discussed in chapter 1, the electrochemical charge of a peptide is dictated by the pH of a solution. This allows the researcher to adjust the overall affinity of the peptides to a cationic stationary phase. This has led strong cation exchange to be touted as "the most useful mode of high performance ion-exchange chromatography for peptide/protein separations".[89,90] The diversity in the zwitterionic properties among peptides is provided through its amino acid building blocks. For example, there are roughly 3.2 million possible amino acid combinations for a five-chain peptide which results in the differing protic moieties and isoelectric points associated with peptides. Unlike elemental counter ions whose affinity is fixed, adjusting

the pH changes the relative selectivity of the peptides on the ion exchange column through altering their charge state.

In general, below pH 3.0 most tryptic peptides possess a net positive charge and therefore have a high affinity for the column. At this pH, all basic amine groups (N-terminus, lysine, and arginine) are fully protonated (+1 charge) while the acidic residues (carboxyl terminus plus the acidic side chains) are also protonated, conferring a neutral charge.[13] An acidic pH represents the point where the peptides have the highest affinity for the SCX media, and therefore represents the optimal conditions for loading peptides onto the column.

As the pH is increased in solution, the net charge of the peptides will be reduced which also decreases the affinity of peptides for the stationary phase. However, it is important to note that the dispersion of charge across peptides can be asymmetric depending on the composition of the peptide, allowing for continued electrostatic interactions with the column despite a net neutral charge (Figure 2-6). As the pH continues to increase past the isoelectric point, peptides exhibit a net negative charge which shifts the equilibrium to the mobile phase, allowing most of the peptides to elute from the column entirely.

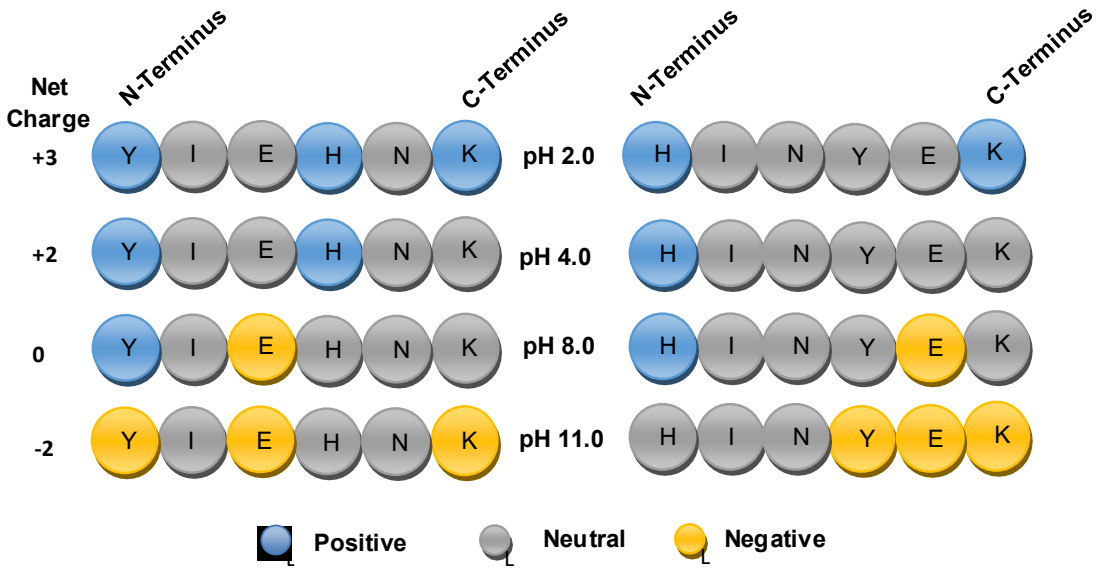


Figure 2-6 Two different peptides with the same amino acids and charge states, as an example of charge distribution.

2.1.2.1 Donnan Effects

If using a pH gradient in SCX, known as chromatofocusing, to separate peptides, it is essential to consider the impact of the stationary phase on the local pH environment experienced by peptides retained on the column through Donnan effects.[91] Donnan effects are the product of the selective uptake of ions across porous ionic surfaces, which result in the non-uniform distribution of ions throughout a system.[91] Given that the negative charge is fixed on the stationary phase, co-ions such as OH^- are expelled while cations such as H^+ are actively absorbed into the porous microenvironment. In the context of SCX, Donnan effects result is a shift in pH equilibrium where the pH of the buffered mobile phase flowing through the column is often one unit greater than that in the surrounding macroenvironment.[13] To minimize the impact of Donnan equilibria and properly maintain the pH in the mobile phase, the eluent must be set within the buffering range.[13,91]

2.1.3 The Impact of Organic Solvent on SCX Separation

It is important to remember that SCX resin is comprised of more than negatively charged functional groups which have electrostatic interactions with peptides. The backbone of the resin can interact also with the peptides, either through hydrophobic or hydrophilic interactions. This is often ignored when considering peptide separation, but has significant impact on resolution and recovery for separation of complex proteome mixtures especially when performing strong cation exchange.[90] Alpert *et.al*[92] has found that the addition of aprotic solvents such as acetonitrile reduces the retention of hyrophobic peptides by the resin backbone of SCX columns.[89,93] In particular, their analysis performed using a silica-based SCX resin, polySULFOETHYL A, found 25% acetonitrile was enough to reduce the retention of hydrophobic peptides on the stationary phase. This was advantageous as hydrophobic peptides are often challenging to recover from traditional SCX columns due to the highly charged and polar mobile phase. The same separation performed under 50% acetonitrile resulted in the removal of more hydrophobic peptides, while hydrophilic peptides are retained on the column longer based on their electrostatic interactions.[89,92] When increasing the organic solvent composition of the mobile phase past 70%, Alpert *et al.*[92] observed that the column ceased to behave like SCX separation, where peptides were being separated by hydrophilicity is the dominant mode of separation with hydrophobic peptides eluting first and hydrophilic peptides eluting second.[89,92,93]This can be problematic for the quantification of separations in an offline mode and for separations in online two-dimensional liquid chromatography (2D-LC) systems.

In an offline mode, samples are separated prior to analysis in mass spectrometry. This can be either performed in one dimension or two. In the first dimension, samples are separated using strong cation exchange. This separation often requires a buffer containing a large amount of salt which means that quantifying peptides through LC-UV must be performed after peptide clean-up. This clean-up uses a reversed phase column to remove both matrix effects, which impact quantitation, and to prepare peptides for subsequent analysis via mass spectrometry. The addition of 70% acetonitrile utilised in the Alpert method would require sample dilution prior to clean-up.[89] In contrast with 2D-LC, the presence of higher concentrations of organic solvent in the 1st dimension can negatively impact the subsequent separation of peptides by reversed phase chromatography. For example, if separations were performed in organic solvents with greater than 70%, the total amount of peptide eluted from the first phase of the column (SCX) would elute hydrophobic peptides followed by hydrophilic peptides which is the exact opposite elution order of reversed phase chromatography. Thus, the separation methods are no longer orthogonal. Increases in ACN following a salt gradient which are intended to separate peptides that are only in the second dimension of the column, are also separating peptides on the SCX resin. In summary, while the organic solvent is not a major mode of separation, it should be considered as a method to reduce the stationary phases overall affinity for peptide analytes. When you consider biphasic separation methods, running high concentrations of acetonitrile through an SCX column prior to RP separation may result in a reduction of orthogonal separation.

2.1.3.1 Critical Analysis of SCX Separation Methods in Proteomics.

Knowledge of the three components of the mobile phase which contribute to SCX separation can allow for a critical analysis of conventional SCX separation methods utilized in proteomics today. As mentioned in Chapter 1, the Multidimensional Protein Identification Technology (MuDPIT) method produced by Washburn *et al.* is considered the gold standard proteomic analysis.[57] Their group showed effectively that online two-dimensional separation increases the overall protein identifications in mass spectrometry.[82,94] However, it is also important to consider the overall design of directly coupling two-dimensional separation to mass spectrometry. For instance, while MuDPIT can identify more peptides, most of these identifications occur within the first few hours of separation on the column. This method, depending on the variation, can take between 12 and 24 hours to perform in its entirety. These lengthy runs reduce the number of replicates and the number of samples which can be examined at any given time. In addition to these challenges, two-dimensional chromatography in some ways can handcuff the experimentalist. In particular, the online method requires volatile solvent systems which limits the choice in buffers, and salt ions which can be used in the gradient system.

The Webb *et al.* method[82] consists of three buffer conditions. Buffer A and B are composed of a pH 2.6 solution that contains water/ acetonitrile/ formic acid in concentrations of 95%/ 5%/ 0.1%, and 20/ 80/ 0.1% respectively. Buffer C contains 500 mM aqueous ammonium acetate buffer at pH 6.8.[82] First the sample is loaded at pH 2.6, where positively charged peptides are bound to the SCX column and peptides which are not retained are (hopefully) bound to the reversed phase column. Then gradients are alternated between gradual salt and pH bumps of Buffer C interspersed with an ACN gradient, as seen in Figure 2-3. The final salt bumps contain increased amounts of ACN from Buffer B. Overall the shortened modified method runs for 15 hours and utilises 39 ammonium acetate salt bumps. The solvent gradients for the SCX separation using the biphasic column utilise all three key components identified as essential for separation, the use of a salt and pH gradient, and increased ACN concentration in the final few runs. The increased ACN composition in the final runs are an attempt to reduce the hydrophobic interactions of peptides within the stationary phase of the column.[82] However, it is important to note that while ammonium acetate is considered a volatile additive, it still causes residue build-up in MS that must be cleaned frequently.

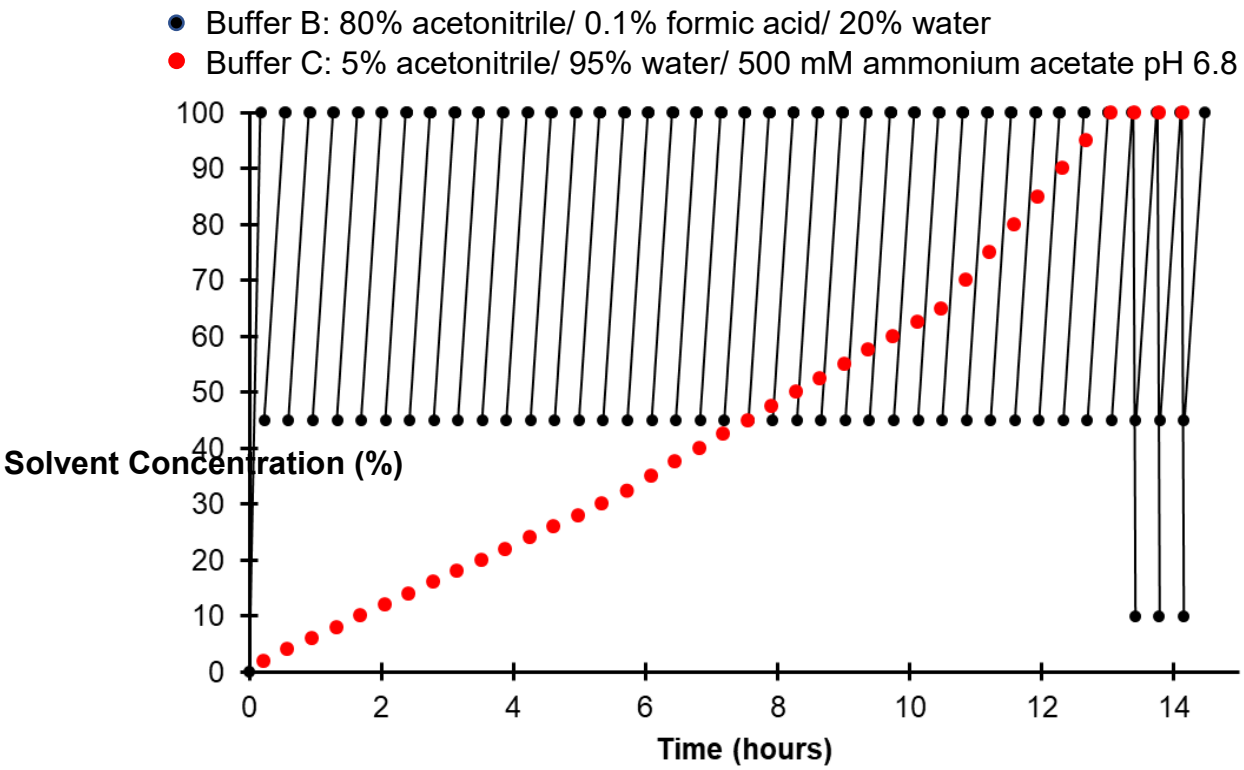


Figure 2-7 The solvent gradient used in Webb's *et al.*[82] microMuDPIT method. Buffer A, not depicted, contains 0.1% formic acid/water at pH 2.6. Buffer B is composed of 20% water, 80% acetonitrile and 0.1% formic acid and buffer C is an aqueous 500 mM ammonium acetate solution. Buffer A mixes with both B and C solvents to form the solvent gradients. Buffer B is responsible for removing peptides bound to the RP portion of the SCX column. Buffer C is responsible for the elution of peptides from the SCX column.

While there is no disputing the efficacy of the MuDPIT technique in terms of total number of peptide (and protein) identifications, the quality of the separation is compromised through the nature of conventional online HPLC. The pH is below the buffering range of ammonium acetate (pH 4-6) during the early portion of the run and is above the buffering range at the end of the run which is not an ideal environment to counteract Donnan effects. Additionally, the nature of online separation means the reversed phase gradient which the run begins with may cause peptides to prematurely elute from the SCX column. This reduces the orthogonality of the overall separation, and impacts chromatography, hindering the overall identification of peptides by mass spectrometry.

2.1.4 Offline Methods of Separation

2.1.4.1 An Introduction to the ProTrap XG

Offline separations offer the greatest opportunity to utilize SCX to its greatest potential. However, offline separation methods have their challenges: they are often considered slower than their online counterparts; they are more time consuming; they require greater sample preparation and they have challenges in terms of sample recovery.[53] Here, the challenges of traditional offline methods are addressed through the introduction of a novel device intended to integrate all aspects of the proteomics workflow. This device, the ProTrap XG, was designed at Dalhousie by Mark Wall and Alan Doucette (chemistry), and molded by Robert Warner (engineering). The device was originally created as a centrifugal sample preparative device for automated acetone precipitation and digestion before being developed here as a separation device Figure 2-8.[95] The device is produced from polypropylene and consists of a filtration cartridge which connects either to a plug or interchangeable solid phase extraction (SPE) column housings (Figure 2-8BII). The top filtration cartridge contains a membrane for protein precipitation allowing for the depletion of salts as well as SDS, prior to subsequent resolubilization and protein digestion within the upper filtration cartridge. This method ensures the sufficient depletion of surfactants for mass spectrometry and prevents sample loss, improving overall peptide recovery and detection.[33,96,]

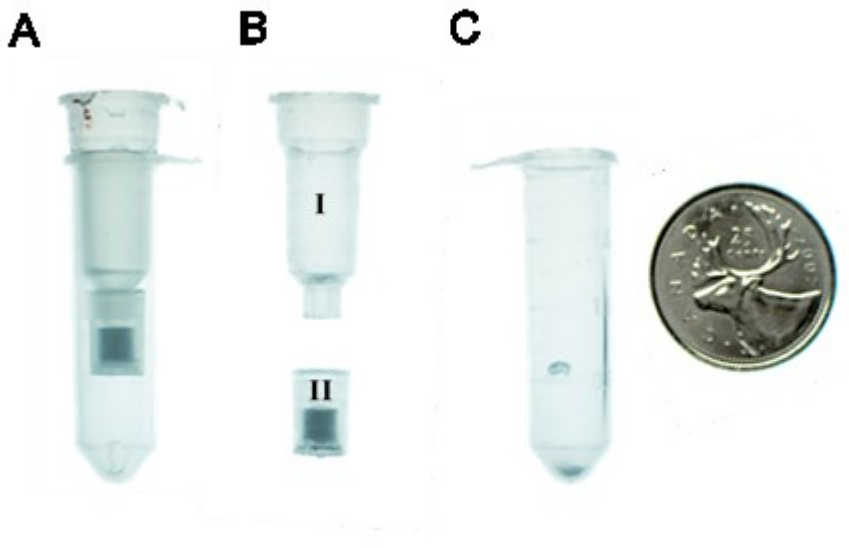


Figure 2-8 The components of the ProTrap XG (**A**), using a quarter for scale. The main components of the device when disassembled (**B**) include a filtration cartridge (**I**), and the solid phase extraction column (**II**) which are housed in a 2 mL microcentrifuge tube when fully assembled (**C)**.

By adapting the ProTrap XG to utilize SCX resin rather than reversed phase material, we can transition away from an online HPLC system. Unlike HPLC which is for the most part limited to separating a single sample at a time, centrifugal separation allows multiple replicates to be run concurrently. This allows for more efficient separations with the ProTrap XG. In addition to offering a means of parallel separation, there is increased flexibility in the choice of buffer system, when compared to traditional binary HPLC pumps which can only mix two solvents. This allows for the preparation of multiple buffers with differing pH values close to their individual pKa's, ensuring the buffer can maintain pH as it travels throughout the column. Performing offline analysis will allow researchers to quantify specific fractions of interest rather than being required to process all fractions sequentially. This can further reduce run time regarding sample clean-up and quantitation of peptide prior to mass spectrometry. This is particularly important in

medical applications where specific biomarkers can be used to assist the diagnosis of diseased states and help develop treatment plans. This would not be possible with traditional 2DLC-MS due to its run time of 12 hours per sample. This chapter will focus on examining the properties of separation in centrifugal column chromatography to ensure consistency with reported properties of SCX exchange. These principles will then be used to develop an approach towards producing an optimised method of peptide separation.

2.2 Materials and Methods.

2.2.1 Materials

2.2.1.1 Biological samples, Growth Media and Assays

Proteins were either obtained as individual proteins from Millipore Sigma (Oakville, ON) as is the case of Bovine Serum Albumin (BSA, cat# C3934) or harvested from *Escherichia coli* (*E. coli* -K12) donated by Dr. Andrew Roger in the Department of Biochemistry at Dalhousie. The *E. coli* was cultured and harvested using pre-mixed luria betani broth agar powder (LB, cat # L3022), and phosphate buffered saline tablets (PBS, cat # P4417), both purchased from Millipore Sigma, while the petri dishes (cat# CA62407) were obtained from ThermoFisher Scientific (Rockford USA). The proteins were assayed using a bicinchoninic acid (BCA) protein assay (cat #23225) and bovine serum albumin certified reference material (BSA-CRM) (cat #23209) also purchased from Thermo Scientific.

The proteins were digested using iodoacetamide (IAA, cat# 163-2109) and dithiothreitol (DTT, cat# 161-0611) obtained from Bio-Rad (Hercules, USA); tris(hydroxymethyl)aminomethane (Tris, cat# 161-0719), trypsin (cat# T6763) and trifluoroacetic acid (TFA, cat# T6508) were obtained from Millipore Sigma; and acetone (cat# BP2403) from Fisher Scientific (Ottawa, Canada).

2.2.1.2 Buffering Salts and Solvents

Potassium chloride (cat# P217500) sodium chloride (cat# S-2830) were obtained from ACP (Montreal, QC), while magnesium chloride (cat# 7786), aluminium chloride

(cat# 7446), lithium chloride (L4408), ammonium chloride (A9434) and calcium chloride hexahydrate (cat#442909) were acquired from Millipore Sigma.

Buffers were created using citric acid (cat# C0759), potassium carbonate (cat# 60109-F), pentylethylenehexamine, PEHA (cat# 292753) obtained from Millipore Sigma while the ammonium acetate (cat# A-637) and the acetic acid (cat# 351271-212) were purchased from Fisher Scientific.

HPLC-UV and HPLC-MS solvents were prepared using acetonitrile (ACN, cat# A955), methanol (cat# A452), formic acid (FA, cat# 94318, Millipore Sigma), TFA (See Biological Samples Growth Media and Assays) and Milli-Q grade 18.2 MΩcm^{-1} water.

2.2.1.3 Instrumentation

Milli-Q grade water was purified to 18.2 MΩ cm using a Sartorius Atrium mini water purifier (Sartorius, Germany). Centrifugation was performed using a Legend Micro 21R centrifuge (Thermo Scientific), an Accuspin Micro centrifuge (Fisher Scientific), a Savant SpeedVac Plus (Thermo Scientific), or an IEC clinical centrifuge (Thermo Scientific). Protein assays were measured using either an Agilent 8453 Spectrophotometer (Santa Clara, USA), or an Agilent G1315B Diode array detector.

HPLC-UV separations were performed on a hybridized Agilent 1100/1200 system which was a Gilson203 fraction collector (Middleton, USA). Once ready for mass spectrometry samples were separated using a hybridized Agilent 1050/1100/1200 system and analysed using a Thermo Scientific LTQ linear ion trap mass spectrometer.

2.2.2 Sample Preparation

2.2.2.1 Escherichia coli growth

E. coli K12 stocks were maintained on 1.5% LB plate, stored at 4 °C, and harvested according to the conventions outlined in the Qiagen Guide to Good Microbiological Practice until necessary for study.[97] An *E. coli* subculture was created by adding a single colony of *E. coli* obtained from plated stocks to 50 mL of LB broth (1 g/L LB) and incubated at room temperature overnight on a gyratory shaker (Model G2, New Brunswick Scientific Co INC, New Brunswick, NJ). The primed media was added to 200 mL of LB media, and then incubated at 37 °C in a thermal shaker (Sanyo, Waterford, UK) until an optical density of 0.8 at 600 nm was obtained. The media was placed on ice, aliquoted into 20 fractions, and the cells were isolated by centrifugation using an IEC clinical centrifuge (5,000 × g, 15 min). The pelleted cells were washed with PBS buffer and subdivided into 2 mL aliquots. These were rinsed twice with water (1 mL per wash, 5,000 ×g, 15 min) to remove traces of LB media, and PBS. The cleaned cell pellets were then stored at -20 °C.

2.2.2.2 Extraction of Protein from E. coli

An aliquot containing an *E. coli* cell pellet was resuspended in 200 µL of 50 mM TRIS (pH 8) then flash frozen using liquid nitrogen. The cells were then ground utilizing a pellet pestle. The protein from the resulting cell lysate was clarified through centrifugation (16000 × g, 30 min at 4°C) using a Legend Micro 21R centrifuge. The protein content of the retained supernatant determined using the BCA Assay ranged from 2.4 - 4 mg/mL (Section 2.2.4.1).

2.2.2.3 Preparation of Bovine Serum Albumin (BSA) Stocks

Bovine serum albumin stocks were made up as standards for liquid chromatography. Approximately 25 mg of BSA was dissolved in 10 mL water. The absorbance of the 1 mL BSA stock was measured at 279 nm. The concentration was determined using the extinction coefficient ($\varepsilon_{\lambda = 280}$ = 43,824 M^{-1} cm^{-1}) and then diluted to a final concentration of 2.0 g/L. The stocks were divided into 500 µL aliquots, and frozen at -20 °C.

2.2.2.4 Digestion

The frozen 2.0 g/L BSA stocks were thawed, 100 µL of 500 mM TRIS was added prior to the addition of 50 µL of 200 mM DTT. Once DTT was added the sample was incubated in a water bath (57 °C for 20 min) before the addition of 100 µL of 200 mM IAA and incubated for 20 minutes at room temperature in the dark. Finally, trypsin was added to the protein in a 1:20 ratio (mass/mass) and incubated at 37 °C for 8 hours. The reaction was quenched through the addition of 10 µL of 10% TFA in water. Additional water was added as needed to obtain a final concentration of 50 mM TRIS, 10 mM DTT,

20 mM IAA, 0.1% TFA, and 1 g/L of digested protein in a total volume of 1 mL. This procedure is the same for the digestion of extracted *E. coli* proteins Section (2.2.2.2).

2.2.2.5 Peptide Clean-up

Following digestion, *E. coli* peptides were cleaned using a Mega Bond Elute C18 EWP (Part # 12256130 Agilent) solid phase extraction (SPE) column. The column was primed using 2 mL of methanol, then equilibrated with 1 mL of 10% acetonitrile/ 0.1% TFA/ water. The *E. coli* peptides from 1 mL of digested *E. coli* sample was loaded into the SPE column in the same solvent conditions as the equilibration. The SPE column was then washed with 1 mL of 10% ACN/0.1% TFA/water and eluted using 1 mL of 80% acetonitrile/ 0.1% TFA/ water. Ten microliters of the eluted *E. coli* peptides were retained and quantified using LC-UV (Section 2.3.5.2) while the remainder was dried under vacuum and frozen at -20 °C in 1 mg peptide aliquots (as determined by LC-UV) until use.

2.2.3 SCX ProTrap

2.2.3.1 Construction

The original ProTrap XG device was modified by removing the membrane within the upper filtration cartridge and packing the SPE housing with SCX resin for the purpose of ion exchange separation of peptides. The SPE housing was packed with POROS XS SCX resin (50 µm beads, XS-041, Applied Biosystems, Bedford, USA) between two PTFE filters (Sterlitech, 0.45 µm thick) with diameters of 5/32" and 3/16" for the top and bottom filters respectively. A detailed schematic is provided in Figure 2-8. The resin and filters in the SPE column were held firmly in place with an endcap Figure 2-8. The assembled device (cartridge + column) fits within a 2 mL microcentrifuge tube.

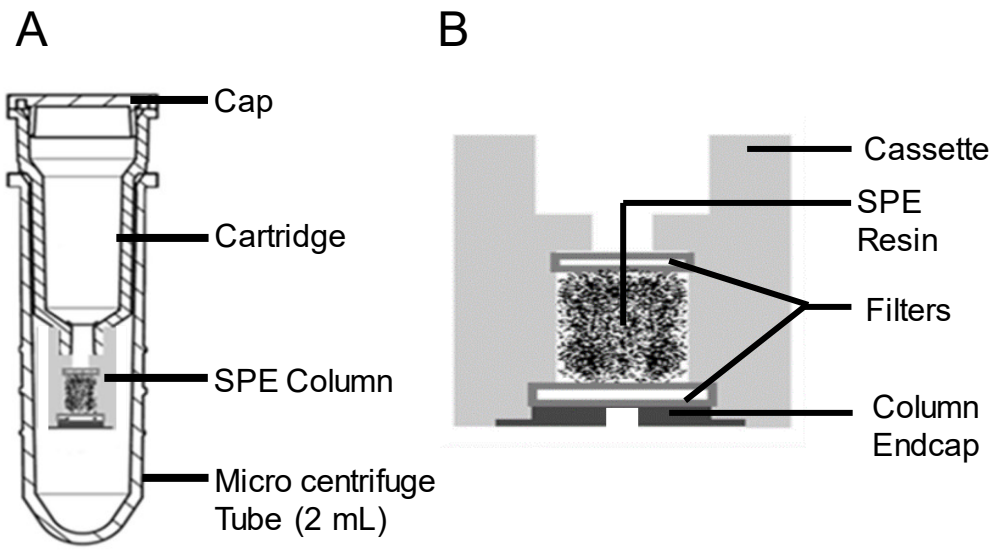


Figure 2-9 Detailed schematic of the ProTrap XG. The assembled device (**A**) contains the cap, cartridge, SPE column, and micro centrifuge tube. The SPE column (**B)** is composed of a Cassette which holds the SPE resin sandwiched between two filters and held in place by an endcap.

2.2.3.2 Separation of Peptides using the ProTrap XG.

The steps utilised in the separation and analysis of peptides using the SCX ProTrap XG is summarized in Figure 2-9. The SPE housing is primed by flowing 400 µL of methanol through the column at 2400 rpm (600 ×g) for 1 minute in an Accuspin Microcentrifuge. The cartridge is then equilibrated using two passes of 400 µL of the loading solvent, most commonly 15% ACN/0.1% TFA/water at 2400 rpm (for 1 minute. Once the column is primed and equilibrated, the filtration cartridges were capped to prevent the resin from drying out. When cartridges were ready for use, they were de-capped, to prevent the formation of a vacuum during centrifugation. *E. coli* peptides were dissolved in loading buffer prior to being passed through the ProTrap XG at 2400 rpm for 2 minutes. To ensure maximal binding. the collected eluent was passed through the column a second time and the flow through solvent was retained for analysis. Next,

the retained peptides were eluted in stepwise fashion from the SCX ProTrap XG through a single pass of 400 µL of the appropriate eluting buffer (see Table 2-2). The retained fractions were capped and chilled at 4 °C prior to clean up and quantitation by RP-LC-UV (Section 2.2.5.2).

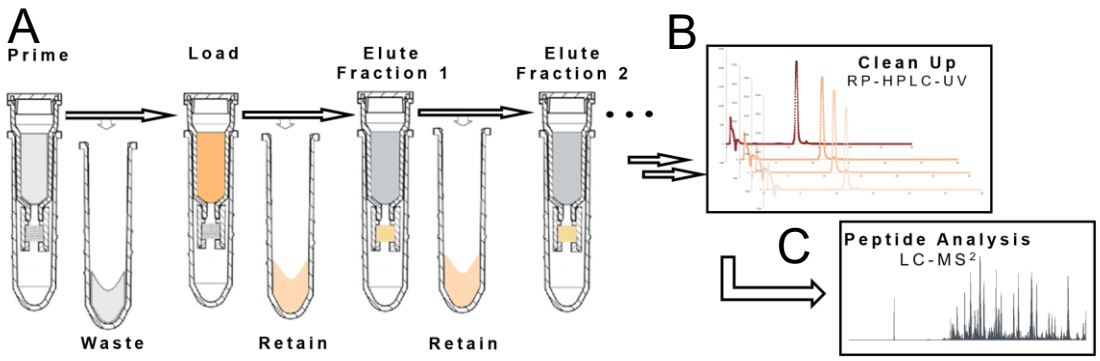


Figure 2-10: Summary of the sample separation and analysis process for the SCX fractions (**A**) The process of peptide separation into fractions using the ProTrap XG. (**B**) The clean-up and quantification of each fraction using reversed phase high performance liquid chromatography-UV (RP-HPLC-UV) while measuring the absorbance at a wavelength of 214 nm. (**C**) The analysis of peptides using reversed phase high performance liquid chromatography tandem mass spectrometry (RP-HPLC-MS2).

To determine the optimal loading conditions for the separation of peptides, a series of experiments were performed and are summarized in Table 2-1. The first experiment examined the loading capacity of the column through loading a total of 1000 µg of BSA peptidesonto the ProTrap XG. To determine the appropriate mass of *E. coli* peptides to be exposed to the column, the loading capacity was assessed using the ProTrap XG with a loading buffer containing 0.1% formic acid/ water/ 5% acetonitrile.

This was followed by an experiment which examined the retention of *E. coli* peptides under increasing concentrations of acetonitrile. One hundred micrograms of *E. coli* peptides were loaded onto the column under three conditions containing increased amounts of acetonitrile: 30, 40 and 70%. Once the impact of ACN on retention and on the

column capacity was determined, a series of experiments were designed to examine the best conditions for recovering and separating *E. coli* bound to the column. These experiments are divided into two key variables of SCX separation salt and pH.

Table 2-1 The buffer composition used to examine the loading conditions for the separation of peptides on the ProTrap XG

Experiment	Loading Buffer
Loading Capacity Section 2.3.1	5% ACN/ 0.1% Formic Acid/ water Injection 1. 50 µg of BSA Injection 2. 50 µg of BSA Injection 3. 100 µg of BSA Injection 4. 200 µg of BSA Injection 5. 400 µg of BSA Injection 6. 200 µg of BSA Total mass of BSA exposed to the column 1000 µg
Organic Solvent and Retention Section 2.3.2	Condition 1. 30% ACN, 0.1% TFA Condition 2. 40% ACN, 0.1% TFA Condition 3. 70% ACN, 0.1% TFA

The first series of experiments examining the impact of salt separation are summarized in Table 2-2 and Table 2-3. The first experiment examines the impact of ionic strength on sample recovery. Two hundred µg of *E. coli* peptide was loaded into the column in 15% ACN/ 0.1% TFA/ water and eluted using seven different cationic species: Al^{3+}, Ca^{2+}, Mg^{2+}, K^{+}, NH_4^{+}, Na^{+} and Li^{+} obtained at a concentration of 0 mM, 250 mM and 500 mM (Table 2-2). Once this experiment was completed, an additional experiment was performed which examines the impact of a potassium gradient under four different fixed pH conditions (pH 2.0, 4.5, 6.5 and 10) on the separation of *E. coli* peptides (Table 2-3).

The second series of experiments examine the impact of pH on the separation of *E. coli* peptides was examined using a universal buffer pentaethylene hexamine and the

second method examines the use of an alternative pH gradient using two buffers in the absence of salt (Table 2-4).

Table 2-2 *The* buffer composition for understanding the impact of ionic strength on the separation of *E. coli peptides.*

Impact of ionic strength on the retention of peptides Section 2.3.3	Load Conditions: 15% ACN/ 0.1%TFA/ water Cationic species were obtained in from $AlCl_3$, $MgCl_2$, $CaCl_2$, KCl, NaCl, and LiCl Fraction 1.[Cation] = 0 M; 15% ACN/ 0.1%TFA/ water Fraction 2. [Cation] = 0.250 M; 15% ACN/ 0.1% TFA/ water Fraction 3. [Cation] = 0.500 mM; 15% ACN/ 0.1% TFA/ water

Table 2-3 The buffer composition for the separation of *E. coli peptides* using a 500 mM K^+ Gradient

Salt Gradients with a fixed pH 2.1 Section 2.3.3	Load:15% ACN/ 0.1%TFA/ water 100 mM KCl/ 15% ACN/ 0.1%TFA/ water 200 mM KCl/ 15% ACN/ 0.1%TFA/ water 300 mM KCl/ 15% ACN/ 0.1%TFA/ water 400 mM KCl/ 15% ACN/ 0.1%TFA/ water 500 mM KCl/ 15% ACN/ 0.1%TFA/ water
Salt Gradient at pH 4.5 Section 2.3.3	Load:15% ACN/ 0.1%TFA/ water 100 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 200 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 300 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 400 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 500 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water
Salt Gradient at pH 6.5 Section 2.3.3	Load:15% ACN/ 0.1%TFA/ water 100 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 200 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 300 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 400 mM KCl/ 15% ACN/ 10 mM Citric Acid/ water 500 mM KCl/15% ACN/ 10 mM Citric Acid/ water
Salt Gradient at pH 10.0 Section 2.3.3	Load:15% ACN/ 0.1%TFA/ water 100 mM KCl/ 15% ACN/ 10 mM TRIS/ water 200 mM KCl/ 15% ACN/ 10 mM TRIS/ water 300 mM KCl/ 15% ACN/ 10 mM TRIS/ water 400 mM KCl/ 15% ACN/ 10 mM TRIS/ water 500 mM KCl/ 15% ACN/ 10 mM TRIS/ water

Table 2-4 Elution gradients utilised for separation of *E.coli* peptides within the ProTrap XG

Pentaethylene Hexamine Section 2.3.4	Load: 15% ACN/ 0.1%TFA/ water Each fraction contains 5 mM PEHA/16% ACN/ nano Water buffered to the following pH: Fraction 1. pH 3.02 Fraction 2. pH 4.01 Fraction 3. pH 4.93 Fraction 4. pH 6.02 Fraction 5. pH 7.02 Fraction 6. pH 8.06 Fraction 7. pH 8.95 Fraction 8. pH 9.90 Fraction 9. pH 11.12
Optimised pH Gradient without supplemented salt Section 2.3.4	Each Fraction Contains 450 µL pH 4.0 10 mM Citric Acid / 15% ACN/ water pH 4.25 10 mM Citric Acid/ 15% ACN/ water pH 5.0 10 mM Citric Acid/ 15% ACN/ water pH 6.0 10 mM Citric Acid/ 15% ACN/ water pH 6.50 10 mM Citric Acid/ 15% ACN/ water pH 9.0 10 mM TRIS/ 15% ACN/ water pH 11.0 10 mM TRIS/ 15% ACN/ water
Select Salt and pH gradient Section 2.3.5	Spin speed: 3.5 rpm for 2 minutes Load: 15% ACN/ 0.1%TFA/ water Fraction 1: 20% ACN/ 0.1%TFA/ water/50 mM KCl pH 2.1 Fraction 2: 20% ACN/ 0.1%TFA/ water 100 mM KCl pH 2.1 Fraction 3: 20% ACN/ 0.1%TFA/ water 150 mM KCl pH 2.05 Fraction 4: 20% ACN/ 50 mM $C_2H_3O_2K_2$/ water 0 mM KCl pH 5.0 Fraction 5: 20% ACN/ 50 mM $C_2H_3O_2K_2$/ water 50 mM KCl pH 5.0 Fraction 6: 20% ACN/ 50 mM TRIS/ water 0 mM KCl pH 7.5 Fraction 7: 20% ACN/ 50 mM TRIS / water 50 mM KCl pH 7.5 Fraction 8: 20% ACN/ 50 mM TRIS / water 100 mM KCl pH 7.5 Fraction 9: 20% ACN/ 50 mM TRIS / water 150 mM KCl pH 7.5

2.3.4 Mass Spectrometry

A Thermo Scientific LTQ linear ion trap mass spectrometer was used for the analysis of peptide sequences through the use of a dual nano-electrospray ionisation interface.[98] An Agilent 1050 Isocratic load pump at 5% ACN/ 0.1% FA/ water loads and equilibrates the column that is not spraying. Cleaned peptides were reconstituted in 5% ACN/ 0.1% FA/ water at a concentration of 0.1 µg/µL and 10 µL were injected into the system through a 1100 Agilent autosampler. Peptide samples were directed onto the column which is not being analysed through a voltage-controlled switch valve. Each of

the 30 cm × 75 µm capillary columns from New Objective (Woburn, USA) were self-packed with C12 Jupiter beads (4 µm, 90 Å pore size) from Phenomenex (Torrance, USA). An Agilent 1200 capillary pump flowing at 280 µL/min with a split flow resulted in a final flow entering the column at approximately 250 nL/min. The pump also operated a gradient on the column separating peptides for analysis via mass spectrometry. The gradient utilised a binary (A/B) solvent system, where A contained 0.1% FA in water, and B contained 0.1% FA in ACN. The timetable gradient is summarized in Table 2-5. The LTQ was operated under a data-dependant mode where MS/MS is performed on the top five peaks. Two replicate injections of each sample were run, the first on column 1 and the second on column 2.

Table 2-5 The timetable for the separation gradient of peptides for LC-MS/MS

Time (min)	%B
0	5
0.1	7.5
90	20
115	25
120	35
121	80
125	80
125.1	5

2.2.5 Assays

2.2.5.1 BCA Assay

E. coli protein concentration was determined using calibration curves created in 50 mM TRIS pH 8 with concentrations ranging from 0 to 250 µg/mL using a BSA-CRM. To determine the concentration of extracted protein samples, a serial dilutions (1/4, 1/8, 1/16, 1/32, 1/64 fold) of the extracted *E. coli* proteins were performed. The assay was

performed through the addition of 300 µL of BCA working reagent to 15 µL of each calibration standard and to the unknown sample in duplicate. The calibration standards and unknowns were incubated simultaneously for 30 minutes in a water bath (60 °C). The absorbance of the samples and standards were then measured using a 500 µL glass cuvette (λ_{abs} = 562 nm) with an Agilent 8453 Spectrometer (Santa Clara, USA). The unknown concentration of protein samples were obtained through a linear fit of the BSA-CRM calibration curve.

2.2.5.2 Quantifying Peptide

Peptide samples retained from the ProTrap XG were desalted and quantified using an Agilent 1100/1200 hybridized HPLC system containing an autosampler with a 100 µL sample loop, a diode array detector (λ_{abs} = 214 nm) and a fraction collector. The 1 mm × 50 mm reversed phase column was packed in house using 5 µm Waters Spherisorb S5 OD52 C18 beads (Milford, MA, USA). Sample and solvent blanks (98 µL) were loaded in 5% ACN/ 0.1% TFA/ water using the autosampler and the column was eluted using a stepped gradient at 8 minutes to 85% ACN/0.1% TFA water. The peptides eluted as a single peak and fractions were collected between 10 and 12 minutes. The peptide peak was quantified through manual integration in Microsoft Excel (Microsoft, Calgary, AB) Figure 2-11. The peptide yield for sample was determined through comparison to a calibration curve created from 0.5, 1, 2, 4, 6, 8, 10, 15, 20 µg of digested BSA. The cleaned sample fractions were fully dried under vacuum using a Speed-Vac before storage at -20 °C for later analysis using LC-MS.

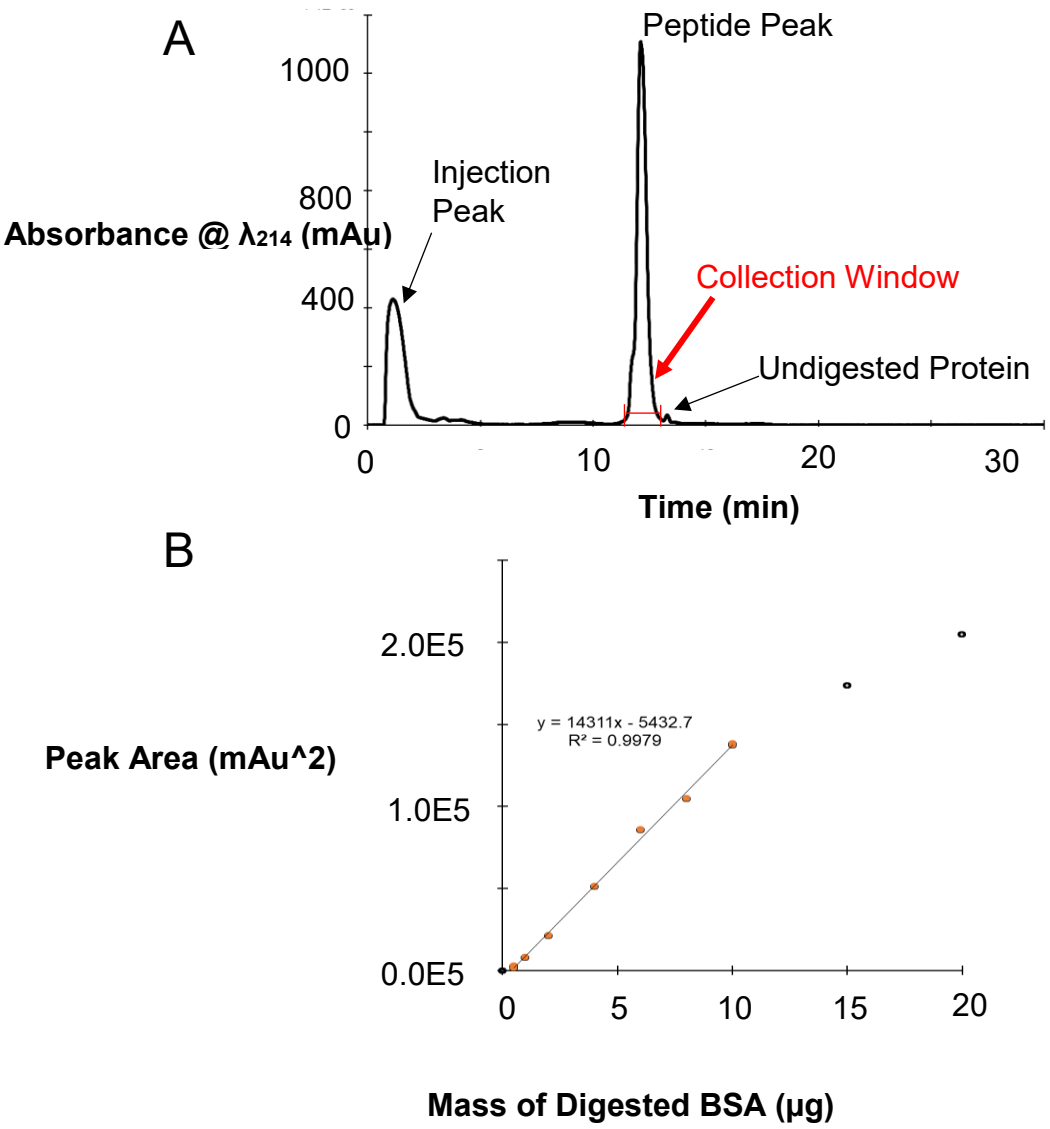


Figure 2-11 A sample LC-UV spectra and calibration curve LC-UV used to describe peptide quantification. The components of a blank subtracted chromatogram (**A**) obtained from the clean-up method on a C18 column. Of note, the collection window where samples are obtained using a fraction collector. A calibration curve (**B**) produced through analysis of the peak area of a series of digested bovine serum albumin standards. The linear range (orange) with a limit of detection of 0.5 µg and a Limit of Linearity of approximately 10 µg.

2.2.6 Data Analysis

2.2.6.1 Peak Integration

All blank LC-UV spectra from runs were overlaid and plotted in Microsoft Excel. Blank runs showing evidence of residual protein on the column were omitted as outliers.

The remaining blank spectra were averaged and the intensity readout was pasted into Column A of an Excel spreadsheet to perform data analysis (See Appendix A) that can process up to five spectra at a time. Referring to Figure 2A, the measurement window was established by examining the spectra for initial and final elution time points (t_1 and t_2). The area of the elution peak was calculated through summing the absorbance readouts over the peak elution window (t_1 and t_2). These areas from manual integration are proportional to the mass of peptide and are compared to digested BSA calibration standards to determine an unknown peptide concentration. The calibration curve has a linear range between 2 and 10 µg with a LOD of 0.5 µg, as shown in Figure 2-11B.

2.2.6.2 Peptide Database Searching

Thermo Proteome Discoverer software (v. 1.4) which employs the SEQUEST[23] searching algorithm, for the identification of peptide sequences using tandem mass spectrometry. The MS spectra were searched against organism specific FASTA sequences (*E. coli*, 4313 proteins, downloaded 2018) obtained from UniProt, with a fragment tolerance of 1.0 Da, and a medium peptide confidence interval.[27] Other search parameters included: the modifications to cysteine [static carbamidomethylation (+57.0215 Da), and methionine (dynamic oxidation (+15.9949 Da)], and up to two missed trypsin cleavages. The false positive rate was 5%, with a medium peptide confidence and a minimum of 2 peptides per protein.

2.2.6.3 Calculating the Net Charge of the Peptides Identified Using MS

The net charge of each of the individual peptides was calculated using the spreadsheet found in Appendix (B). The pKa values for the amino acid side chains and the C and N-termini were used to estimate the charge abundance at a specified pH. This approximation was calculated by multiplying the prevalence of each amino acid side chain within the peptide by their respective charge abundance and then adding the charge of the C and N termini.

2.3 Results

2.3.1 Loading Capacity

Strong Cation Exchange resin is known to have a higher binding capacity than traditional reversed phase resin.[13,87] To determine the appropriate mass of *E. coli* peptides to be exposed to the column, the loading capacity was assessed using the sequential loading of BSA peptides onto the ProTrap XG. The loading buffer was selected to match the MuDPIT method (0.1% formic acid/ water, 5% ACN). The eluent was analysed using HPLC to determine the amount of unretained peptide on the column (Figure 2-12). Referring to Figure 2-12, exponential behaviour was observed for the successive loading of *E. coli* peptide onto the column. Initially there is no statistical difference in terms of the mass lost in each of the successive loadings up to 200 µg of peptide. As expected, the addition of more peptide onto the column results in increasing amount of peptide loss, where a lack of retention appears to increase rapidly for the loading of 400, 800 and 1000 µg onto the column with losses of 13 ± 6 µg, 72 ± 5 µg of and 93 ± 10 µg respectively.

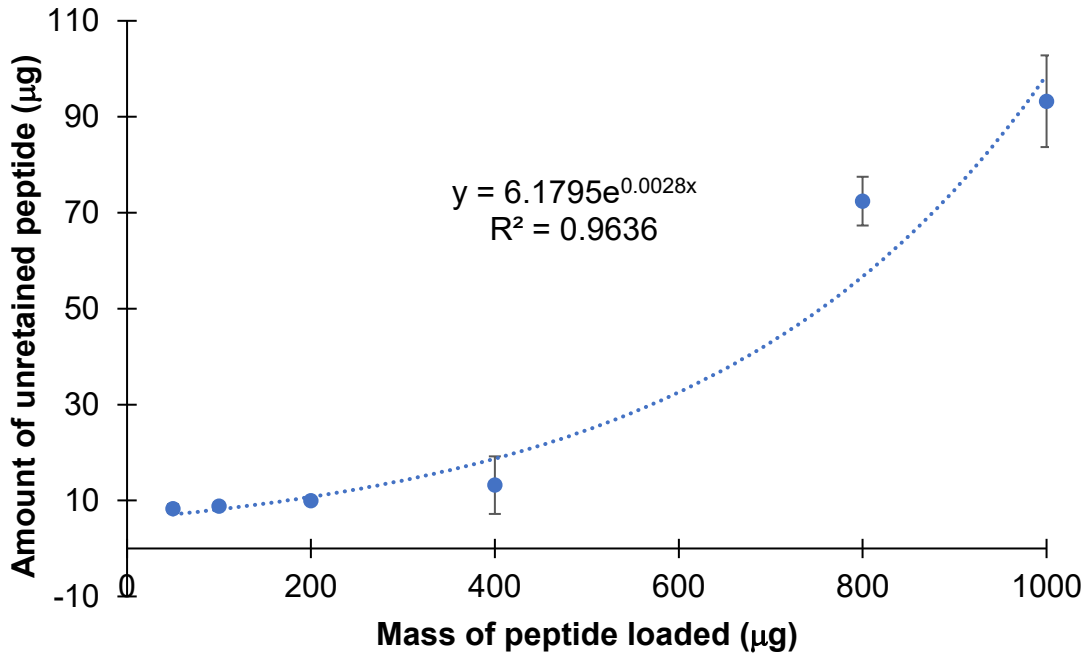


Figure 2-12 The amount of unretained peptide (µg) as a function of the total mass of peptide loaded onto the column. Successive injections of BSA peptides to the column shows an exponential relationship between total peptide loss and the recovery of peptides (N = 2).

While the amount of peptide lost in the individual fraction appears to be quite high, it is important to consider the cumulative loss of peptide in context with the total amount of peptide exposed to the column. The mass of peptide loss should also be examined with context using the cumulative loss of peptide expressed as a percentage of the total amount of peptide loaded onto the column. The cumulative mass of unretained peptide relative to total amount of peptide for 50 µg (17 ± 2%) shows no significant difference at the 95% confidence interval from the cumulative loss of 1 mg of peptide expressed as a percentage (21 ± 1%). While the average amount of peptide retained on the column could potentially be reduced by a small increase of ACN in the mobile phase, even under current conditions 1 mg of *E. coli* peptide remains within the capacity of the ProTrap SCX column (79 ± 1%).

2.3.2 Impact of ACN on Separation Using SCX

Acetonitrile is commonly utilized to decrease the backpressure on a column in chromatography. This allows the user the opportunity to run sample through the column using faster spin speeds. This is important given that one of the objectives is the collection of samples within 2 minutes of spin time. However, it is also important to be cautious about the use of acetonitrile within these runs as high concentrations of ACN alter the mode of separation within the resin.[89] A quick analysis was performed to assess the impact of acetonitrile on retention on the column and are summarized in Figure 2-13. As expected, as the concentration of acetonitrile in the buffer is increased, there is a decrease in retention where the total amount of peptide retained on the column was inversely proportional to the amount of organic solvent in the load solution. This is consistent with literature which states that concentrations of ACN that are in excess of 25% will result in the elution of peptide from the column. [89,92] The remaining experiments were performed using 15% acetonitrile which does not significantly impact recovery, but does reduce back pressure.

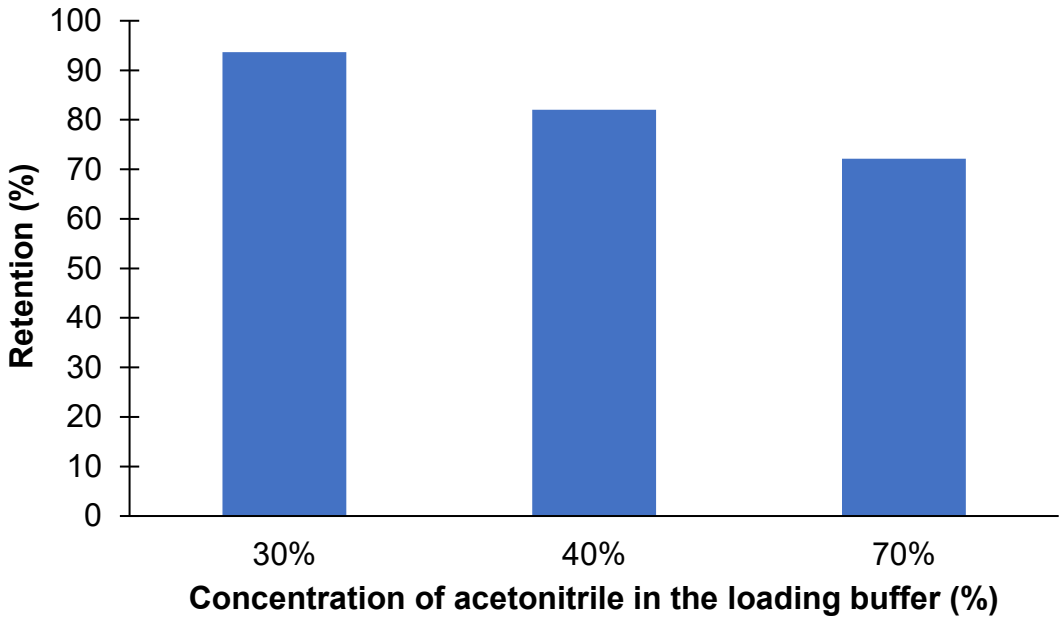


Figure 2-13 The relationship between the retention of peptide as a function of acetonitrile concentration in the load buffer (N = 1).

2.3.3 The Impact of Cations on the Retention of Peptides on SCX

Three factors associated with salt gradients are imperative to consider when examining peptide separation: ionic selectivity, salt concentration, and resin properties. As the resin is consistent throughout this research, the remaining ionic selectivity is examined in this section while the impact of salt concentration is examined in the next. Previous work done in the Doucette group utilized K^+ ions over Na^+ due to challenges with the effective removal of BSA from the ProTrap XG column.[99] Consequently, the impact of charge density on the total amount of recovery as a function of concentration was examined to determine the ideal salt ion for separation (Figure 2-14). Seven different cationic species Al^{3+}, Ca^{2+}, Mg^{2+}, K^+, NH_4^+, Na^+ and Li^+ were obtained from chloride salts. This was done to eliminate the impact of differing counter ions. Ionic species with a higher ionic charge density such as Al^{3+} are expected to elute a larger number of peptides than ions, such as Li^+ with a small ionic radius due to a stronger selectivity coefficient.

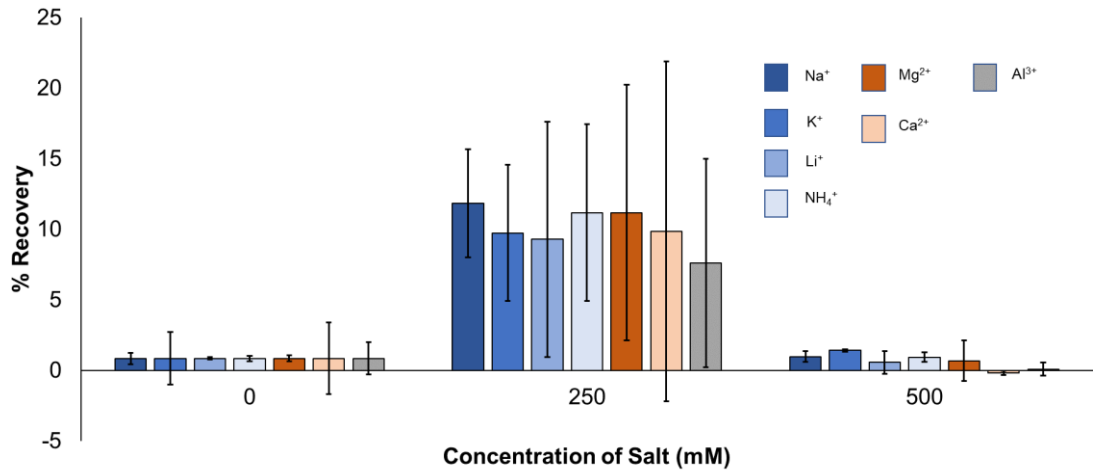


Figure 2-14: The recovery of *E. coli* peptides using seven different cationic salts as listed (Na^+, K^+, Li^+, Mg^{2+}, NH_4^+, Ca^{2+}, Al^{3+}) (N = 4). Data are organised according to net charge, singly charged cations Na^+, K^+, Li^+, NH_4^+(shades of blue), doubly charged cations Mg^{2+} and Ca^{2+} (shades of orange), and one triply charged cation Al^{3+} (grey).

Increasing the concentration of salt results in the elution of peptides with little to no quantifiable recoveries in the absence of salt, and recovery above the limit of detection in the presence of 250 mM cation concentration. However, the experimental results show no difference in either the elution profile, or the total recovery of *E. coli* peptides among the different types of cations. Most peptides were eluted in the first salt fraction with a cationic concentration of 250 mM, above which, there was little to no detectable amount of peptide removed from the column. Comparison of the peptide recovery based off cationic strength shows no statistical advantage in using any particular salt ion due to the large error observed in the recovery for all of the differing cations, for example: the recovery of Mg^{2+} was 16 ± 6%, Al^{3+} was 14 ± 2%, and Li^+ 15 ± 5% which could be due to discrepancies with packing in the prototype system, causing non-uniform flow profiles through the ProTrap XG during centrifugation. Error among the four replicates might be a result of column packing. However, the low overall recovery and the lack of detectable

difference in recovery, despite the notable differences in charge density among the four replicates indicate a reduction in the magnitude of selective affinity for cations.

The relationship between the affinity for a resin and each cation is quantified through the use of the ion selectivity coefficient which is dependent on particle size, pore size and level of crosslinking within the polymer. The use of larger particles and pore sizes due to reduced crosslinking in ion exchange resin is often required for proteomic samples for the optimal retention of amino acid chains, but this also reduces the ion selectivity coefficient. The POROS XS resin utilized in the ProTrap XG has a relatively large particle size of 50 μm, to reduce the back pressure for chromatographic separation within the centrifuge. As a perfusion chromatography resin, large through channels travel through the SCX resin, rather than have small diffusive pores. This allows for an increased binding capacity of protein and peptides, but reduces the magnitude between the relativity selectivity of individual ions, resulting in very little difference among the selectivity coefficients in terms of recovery. Consequently, using this resin, any cation could be used for separation on the column.

Given that the affinity for various cationic salts was low with a total recovery averaging less than 30% in Figure 2-14, it is important to ascertain if there is a viable method for recovering peptides from the column using a salt gradient. One clear method is by reducing the overall affinity for the columns is best by considering the polyprotic nature of peptides, the effect of pH on the binding affinity of peptides and the rate of exchange among the cations and the resin. The impact of salt gradients on overall recovery was examined using potassium chloride salt gradients with a maximum concentration of 500 mM (Figure 2-14). All samples were loaded under acidic conditions

before K^+ salt gradients were performed under four constant pH conditions using 10 mM buffer at pH 2.1, 4.5, 6.5, 10. It is expected that peptides will have a reduced affinity for the SCX column at higher pH values, allowing for the salt gradient to elute a greater portion of the peptides.

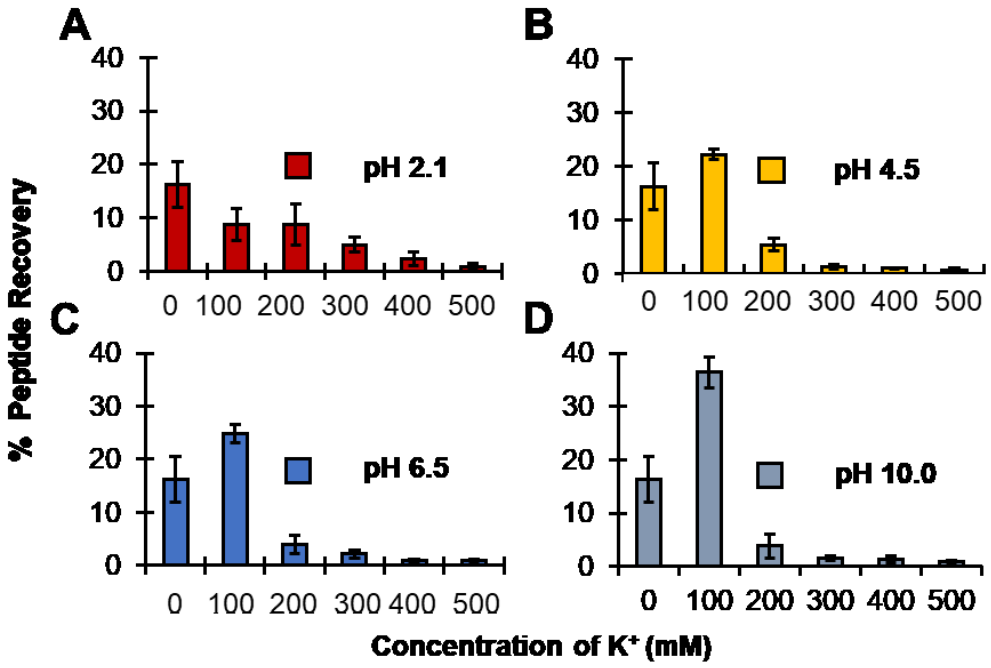


Figure 2-15: The recovery of *E. coli* peptides using a 500 mM K+ salt gradient at four fixed pH values from the ProTrap XG. (**A**) The K^+ separation gradient in a 0.1% TFA 15% ACN buffer at pH 2.1. (**B**) The K^+ separation gradient in a 15% ACN, 10 mM citric acid buffer at pH 4.5. (**C**) The K^+ separation gradient in a 15% ACN, 10 mM citric acid buffer at pH 6.5. (**D**) The K^+ separation gradient in a 15% ACN, 10 mM TRIS buffer at pH 10.

Referring to Figure 2-15, as expected, the total recovery of the peptides increases under strongly basic conditions, with the highest total recovery at pH 10 (60 ± 6%) and the lowest total recoveries under acidic and near neutral conditions pH 2.1 (43 ± 7%), pH 4.5 (49 ± 4%), and pH 6.5 (47 ± 5%). The difference in the overall recovery of peptides under the two pH conditions which are greater than 2 pH units away from their

theoretical lowest and highest pKa values there is a significant difference in recovery ($P < 0.05$). The impact of increasing the pH on the overall recovery on the column is not significant for salt gradients performed under the strongly acidic conditions and pH 4.5 and 6.5 ($P < 0.05$). The highest individual recoveries of peptide are in the first salt fraction of 100 mM K^+.

The first fraction recovery for each run is 9 ± 3%, 24 ± 1%, 22 ± 2% and 36 ± 3% for pH 2.1, 4.5, 6.4, 11 respectively. This demonstrates an increase in pH the peptide affinity does decrease on the column.

However, if ionic salt concentration was the greatest contributor to separation, the release of peptide would increase with each increasing salt fraction. Regardless of pH conditions in Figure 2-15, 95 ± 2% of all of the peptides recovered from the columns occurs below 300 mM of salt, this is consistent with a trend observed in Figure 2-14 which also shows peptide recovery under increasing salt concentration; however, there is little to no peptide recovery in the third salt fraction of 500 mM regardless of type of salt ion. This indicates there is a limit to the impact of salt concentration on overall recovery of peptides from the column and suggests that pH may be a stronger contributor to separation over salt concentration and should be examined to confirm.

2.3.4 Examining the impact of pH on SCX Separation

Given that pH impacts the retention affinity of peptides to the column the separation, a pH gradient was performed to examine these effects on separation. Ideally, a single buffer system consisting of a polyprotic species with pKa's between pH 2 and 10 would allow for the creation of several fractions. Consequently, a hexaprotic compound, pentylethylenehexamine (PEHA) is an ideal buffer candidate to examine the relationship

between peptide affinity on the column and overall separation of peptide from the column. The impact of a pH gradient on the separation and recovery of peptides was examined using a PEHA gradient from pH 2-8 and is summarized in Figure 2-16.

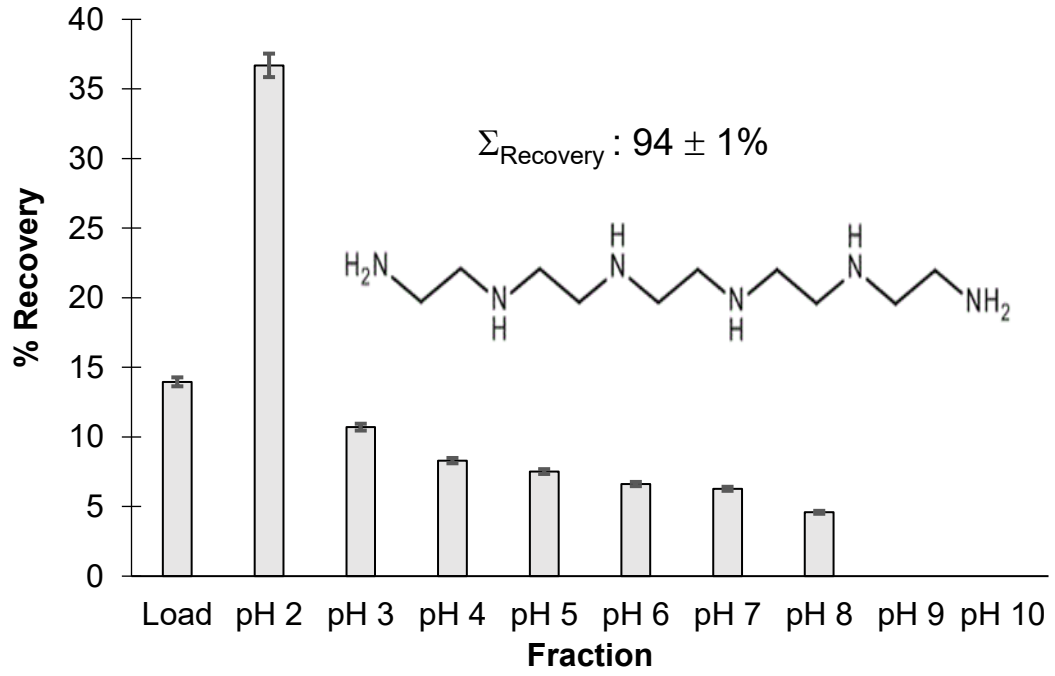


Figure 2-16 The recovery of 200 µg *E. coli* peptide in the absence of salt gradient conditions using 5 mM PEHA (structure top right) gradient by individual fraction (N = 3).

The overall recovery from the column is significantly higher than that of the overall salt column with 95 ± 1% recovery using PEHA compared to the 60 ± 6% recovery from the salt gradient performed at pH 10. However, the quality of fractionation is poor with 37 ± 1% eluting in the first fraction to the PEHA fraction although it shares the similar pH conditions. Furthermore, the elution profile of the PEHA gradient is similar to the salt gradients held at a constant pH where the recovery gradually tapers until there is no detectable amount of peptide eluted from the column. Additionally, there are no spikes of peptide recovery near the pKa values which is inconsistent with what one

would expect given the polyprotic nature of peptide chains. In detail, the third fraction at pH 3.0 has 11 ± 0.3% while the final detectable amount of peptide recovered at pH 8 contains 4.6 ± 0.1% of peptide.

The large amount of peptide recovered in the second fraction of Figure 2-16 could be attributed to the ionic strength of the PEHA. Each of the six amine groups when protonated carry a positive charge giving the buffer an overall net charge of +6 under strongly acidic conditions. which is evenly dispersed throughout the structure, increasing the binding affinity for the column. Consequently, the peptides likely have a stronger affinity for the SCX resin than peptides which may only have a few charged species that are unevenly dispersed throughout the amino acid chain. This indicates that PEHA as an elution buffer may be acting as both a cationic salt, and a pH modifier.

To determine if the charges found on PEHA can be attributed to the large amounts of peptide eluted from the column, a pH gradient must be performed with a neutrally charged buffer species under acidic conditions. Unfortunately, polyprotic species that are neutral in charge when fully protonated and carry at least four pKa's are rare. As a result, this experiment requires two different buffers for the fractionation of 200 μg of *E. coli* peptides across the entire pH range (Figure 2-17).

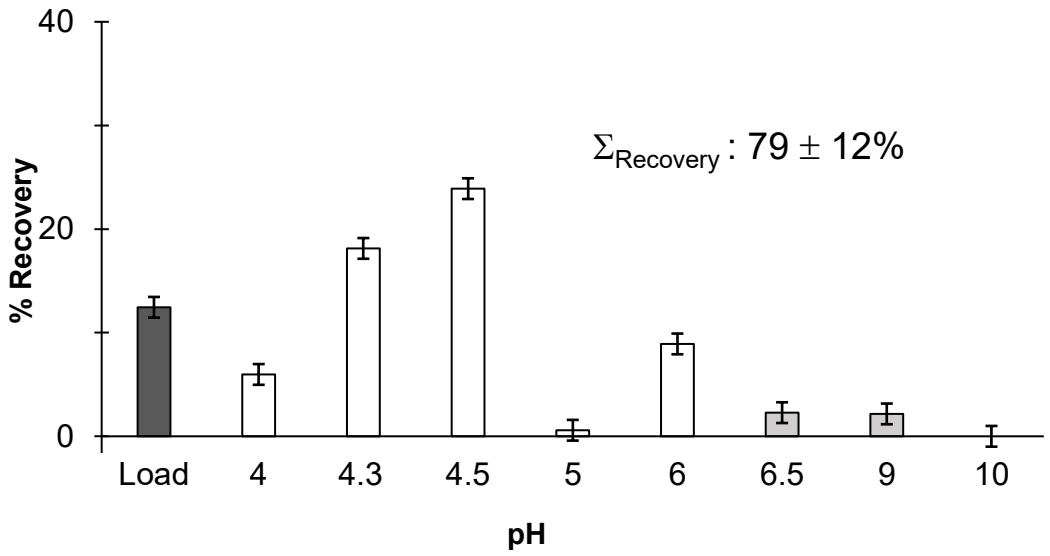


Figure 2-17 The recovery of 200 µg *E.coli* peptide across a broad pH range using four solvent conditions (N = 5). Samples were loaded in 0.1% TFA 15% ACN pH 2.0 (dark grey) and eluted using ammonium citrate buffer pH 2.0-6.0 (white), followed by TRIS pH 6.5-10.0 (light grey). The total recovery of peptides from the column was 79 ± 12% in the absence of salt.

In reference to Figure 2-17, there is no significant difference ($p < 0.05$) in the recovery of peptide from the load fractions of PEHA and those of the citric acid - TRIS gradient (CA-TRIS). However, there is a significant difference in the overall elution profile of the peptides for the remaining fractions utilizing the CA-TRIS gradient. In particular, the CA-TRIS gradient follows pKa trends where there are spikes in recovery at or near isoelectric points for peptides. For example, after loading the peptides, the citric acid-gradient steps to a pH of 4.0 with a recovery of 6.0 ± 1.2%. Once the pH approaches the pKa of Glu's side chain (4.3) the recovery increases to 23 ± 10% and 24 ± 4% recovery at a pH 4.3 and 4.5 respectively. Immediately after, at a pH of 5, there is no detectable amount of peptide recovered from the column. However, unlike in salt gradients when increasing the concentration of cations does not result in the increase of recovery of peptide after we reach the limit of the detection, an increase in pH to 6.0

results in 9 ± 2% recovery of peptide, followed by 2 ± 1% and 2 ± 2% recovery for pH 6.5 and 9 respectfully. There is no detectable amount of peptide at a pH of 10.0 and additional pH washes were not successful in removing peptide from the column and yielded an overall recovery from the column of 79 ± 11%.

The pattern of elution recovery for PEHA closely followed that of a salt gradient, while a pH gradient composed of CA-TRIS followed the expected elution trend where there are spikes in the amount of peptide eluted at or near known pKa's or the average isoelectric points for a peptide. Consequently, it is likely that the ionic strength of PEHA allows the buffer to behave as both a competitor for the ion exchange resin and as a buffer. This results in the removal of large percentages of the peptide from the column in the first fraction which is not ideal for an optimised gradient. However, given that overall recovery of the peptide is highest using PEHA, a gradient that combines both pH and salt gradients is expected to have an increased overall recovery of samples from the SCX column.

Using the information that both salt and pH can be utilised to influence the recovery of peptides, several experiments were performed attempting to produce an optimised separation gradient, to varying degrees of success. The following experiment was selected as it demonstrated both the potential of the system and the key challenges faced by the ProTrap XG as a separation system. Three ProTrap XG columns were loaded with 400 µg of *E. coli* in 0.1% TFA, 15% ACN at pH 2.1, and separated using a stepped salt gradient at pH 2.1, 5.0 and 7.5 with the results of the experiment shown in Figure 2-18.

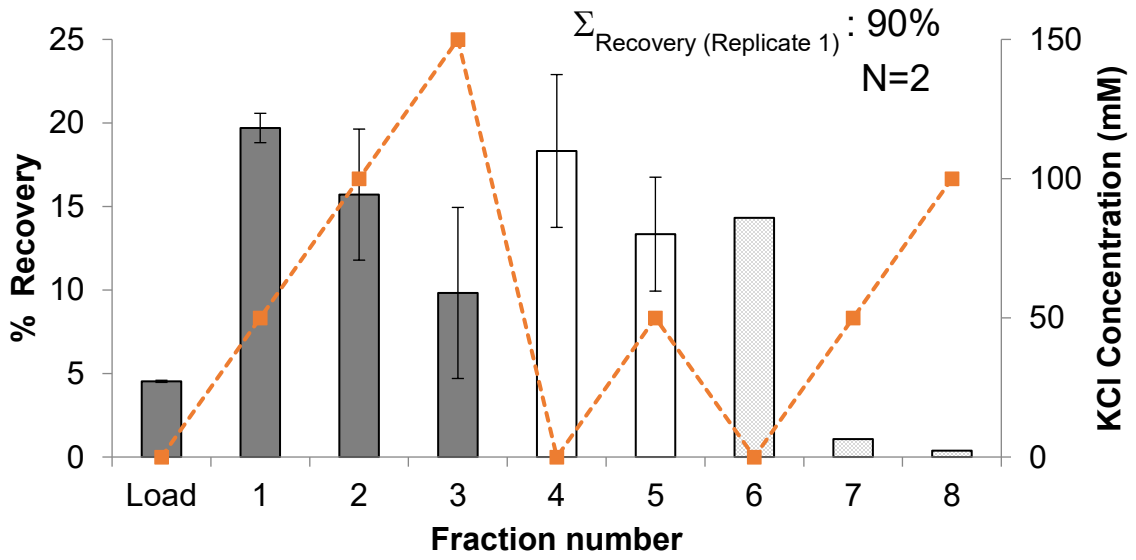


Figure 2-18 The impact of using an elution gradient which uses both pH and stepped salt gradient to separate the samples. The salt gradient (orange) is plotted on the secondary axis, while the buffer conditions are shown using the colour of each bar pH 2.1 (dark grey), pH 5.0 (white) pH 7.5 (light grey). Collection of data ceased with replicate 2 after fraction 6, with a total% recovery of 75%.

Refering to Figure 2-18, the recovery obtained in one of the replicates from the combined salt-pH gradient had one of the highest observed recoveries, 90% of the combined gradients. The high recoveries seen in this combined salt and pH method, along with the optimised pH gradient in the absence of salt, and in the PEHA gradient, all demonstrate the potential of the ProTrap XG system to run separations with high levels of recovery. However, this separation using a combined gradient also demonstrates the challenges when working with a prototype system. The two columns which were loaded without issue at spin speed of 3500 rpm (2600 ×*g*) for a duration of 2 minutes began to experience challenges with column spin time by the second fraction. By the time the separation had reached the fourth fraction, the amount of time required for both samples to fully elute at 2500 rpm (600 ×*g*) had increased to four minutes. The flow rate for the first replicate by the sixth fraction was seven minutes. Separation of peptide in the second replicate was suspended for fractions six to eight because it was unable to flow all 450 µL of eluting buffer in under ten minutes. The challenges with consistent flow rate coincide with the increased variability among replicates, something which is observed in every attempt at separation.

2.3.5 Mass Spectrometric Analysis of pH Gradient Fractions

The quality of the optimised pH gradient in the absence of salt was evaluated using mass spectrometry through two 1 µg injections of each of the fractions of three of the five replicates (Replicates 1, 4, 5) with a cumulative average number of identifications of 2600 ± 800 peptides per replicate. The quality of the separation was tested through comparing the overlap in peptide identifications among fractions shown in

Figure 2-19. If separation exists among fractions, the amount of overlap should decrease with fractions that are further apart from one another, defined as the degree of separation.

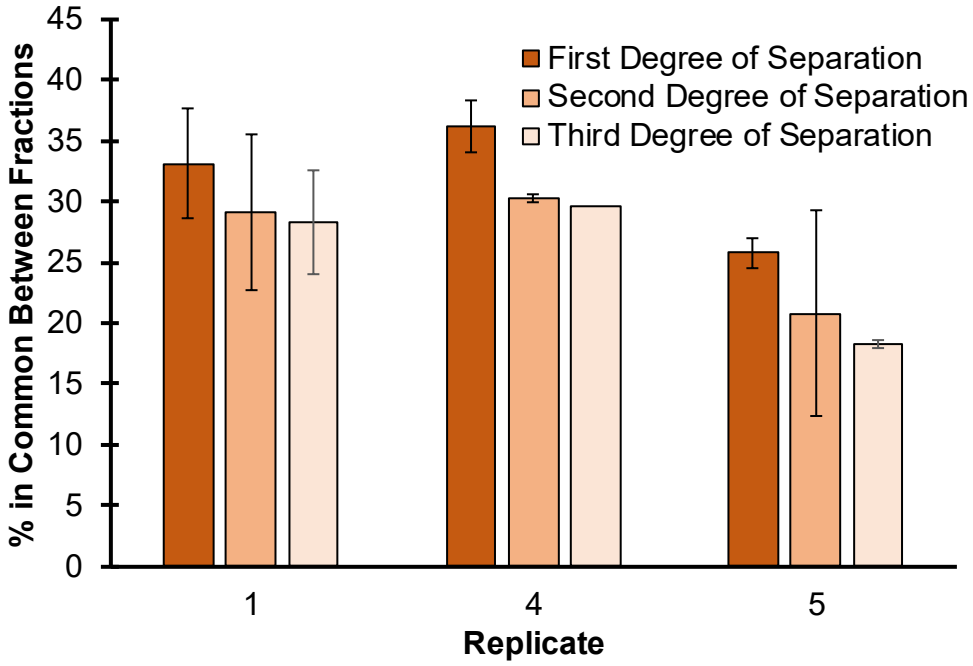


Figure 2-19 The average amount of overlap in peptide identifications for each degree of separation for each replicate. First degree of separation N = 4, second degree of separation N = 3, third degree of separation N = 2.

Referring to Figure 2-19, for each of the three replicates tested using mass spectrometry, there is a decrease in the number of identified peptides which are shared among fractions that are further apart from each other in the collection sequence rather than close together. The decrease is significant in the case of Replicate 4 ($p < 0.05$) but is not statistically significant when considering the trend across all three replicates. This could be for two reasons, the first is that the separations columns contribute to the variance in peptide recovery, which translates into variance in terms of sequence identification. Additionally, the use of a low-resolution mass spectrometer such as the LTQ have a limited dynamic range and a slow scan speed. This reduces the total number of identifications, which could make difference between fractions appear smaller than

they are. Despite this the overall pattern of decreasing overlap is expected because increasing pH should cause the peptides on the column to elute gradually. In detail, peptides are expected to elute as their charge state shifts from positive through neutral to negative decreasing the overall affinity of a peptide in a column. This demonstrates that the pH gradient could be separating the peptides rather than simply spreading the sample out across the fractions.

2.4 Conclusions

An SCX centrifugal chromatography platform called the ProTrap XG was developed to allow for the rapid separation of protein samples in an offline format. This system has many advantages over conventional LC methods, allowing for the selection of multiple buffers to maintain pH across a wide gradient (pH 2 to 11), and for replicate separations to be performed simultaneously. The flexible approaches to buffer selection provide the opportunity to tailor separation to specific proteomes. This allows for the analysis into how organic solvent, salt and pH impact the separation of peptides using SCX in the hopes of producing an optimised separation gradient. High concentrations of acetonitrile were found to reduce the retention of peptides on the column, which is consistent with the work of Alpert *et al.*[92] who found that concentrations of acetonitrile in excess of 25% reduced the retention of peptides on SCX columns.[89,92] Increasing the concentration of salt does result in the elution of peptides from the column. However, the efficacy of salt in the removal of peptides from the column appears to be linked to the charge state of the peptides which is controlled by the pH of the mobile phase. Consequently, pH is the major contributor in separations using SCX. The development of an optimised pH gradient using two buffers in the absence of salt resulted in the removal

of 80 ± 12% (N = 5) of *E. coli* peptides from the column and identified an average of 2600 ± 800 peptides per replicate. This demonstrates that the ProTrap XG has the potential to separate peptides through centrifugal SCX with high levels of recovery.

Interestingly, the use of PEHA as a buffer, which has multiple charge states varying from +5 to neutral had poor separation but high levels of recovery. This shows that increasing the cationic strength of the mobile phase increases the rate of peptide elution from the column. This is expected as separation using a salt gradient is governed by the relative affinity of one counter ion over another. However, attempts to examine the impact of relative affinity on separation using ammonium and various elemental cations were not quantifiable. This could partially be attributed to the increased porosity of the resin which reduces the ionic selectivity of the resin. Reductions in selectivity reduces the overall difference in peptide recovery between each of the ions which when paired with the variability in ProTrap XG columns are below the device's current limit of detection.

Optimised SCX separation systems should utilise salt in tandem with a variety of buffers which have a neutral charge state in strongly acidic conditions and are titrated to a pH within their buffering capacity. The ProTrap XG is a promising separation system, allowing for many of the results to be completed. However, the results among replicates are highly variable. This hinders the ability to perform accurate comparisons among differing methods and in the production of an optimised separation gradient. The system often has challenges with reproducibility either through ensuring that replicates separations go until completion, or through the gradual reduction in flow rate which appears to occur at random among fractionation steps. These issues must be addressed prior to commercialisation for the platform to be successful.

Chapter 3 SCX Separation of Peptides Using the S-CLC

3.1 Introduction

With proper application of a salt and pH gradient, SCX has the potential for providing good separation of peptides to improve detection ahead of mass spectrometry. However, there is no standard approach or method to determine an appropriate SCX gradient for a proteomic system. This is because in part the mechanism of SCX separation in proteomics has yet to be fully modelled. Salt, pH, and organic solvent are three main variables known to influence peptide separation on the column.[13] However, separation gradients in HPLC platforms are limited by the solvent pumps which can combine and deliver, at most, four independent solvents. Considering an online SCX approach, many buffers or high concentrations of salt can cause matrix effects which reduce signal-to-noise in mass spectrometry. Consequently, there is a need to develop a chromatographic platform that allows simple delivery of a complex salt, pH, and organic solvent system to maximise the resolution of SCX separation.

In Chapter 2, an offline separation system called the ProTrap XG was modified to perform SCX chromatography in a centrifugal system. This system allowed for multiple separations to be performed simultaneously and it allowed the user to tailor the separation gradient for a specific proteome. The findings from this novel platform demonstrated that adjustments to the pH of a system were the major driving force for SCX separation. However, the system faced several challenges that impacted the ability to properly examine how these separations take place. There is considerable variability in flow rate from column to column which impacts the recovery and reproducibility of the data.

The work presented in this chapter re-examines the influence of these three main variables to develop effective strategies for optimized SCX separation. This is done in a more conventional column-based SCX platform operated by HPLC. Rather than employ the typical binary or even quaternary solvent delivery system which provides limited flexibility in delivering a separation gradient, the current system establishes a means of eluting peptides retained on the SCX column using an autosampler. The development of this automated solvent delivery system allows a closer examination of the factors that influence separation and has the potential to guide the user towards development of an optimised SCX gradient using a much more diverse array of solvents. This would be applicable to high-throughput multi-dimensional HPLC system where fractions are separated first through SCX, are then desalted, and condensed on a second reverse phase column.

An approach to optimisation was performed through examination of four different gradient systems employed in SCX. The first two univariate separations are examined utilizing salt gradients, each applied at two different, but constant pH conditions, pH 2.1 and pH 6.5. A third univariate separation employs a stepped-pH-gradient delivered in the absence of salt. Finally, the microMUDPIT SCX buffer gradient was performed using a combination of increasing salt and pH. The quality of these separations was then examined based on the total peptide recovery and degree of peptide separation across the fractions, as determined through LC-MS/MS analysis of the SCX-fractionated proteome. These findings again enforce the importance of pH in controlling peptide separation by SCX and point out deficiencies in favoured strategies for multidimensional separations.

3.2 Materials and Methods

3.2.1 Materials

Most of the materials utilised in Chapter 3 are listed in Chapter 2 (See Section 2.3.0). The remaining materials are as follows: *Saccharomyces cerevisiae* (Baker's yeast), was obtained from a local grocery store as a dry pellet (active dry yeast, No Name brand) and used within 1 month of purchase. HPLC-UV separations were performed used a modified pump system containing both an Agilent 1100 binary pump and Agilent 1100 quaternary pump.

3.2.2 Sample Preparation

3.2.2.1 Protein Extraction

The lyophilized yeast was prepared as a 12.5% (weight/volume) aqueous suspension with 250 mM Tris (pH 8) under gentle agitation for 30 minutes. The sample was centrifuged at 5000 x g for 2 minutes resulting in 5 mL of yeast sediment for every 2.5 grams of yeast. The wet yeast pellet was washed twice in a 1:1 volume ratio with 50 mM TRIS buffer (pH 8), 5000 x g for 10 minutes, then reconstituted in a 1:1 volume ratio with 50 mM Tris buffer. The suspended yeast was lysed through grinding under liquid nitrogen with a pellet pestle for five minutes. Approximately 0.2 grams of ground yeast cells were weighed into pre-chilled 1.5 mL microcentrifuge tubes and were kept at -20 °C until just prior to use. Sample vials were thawed, and 50 mM Tris (pH 8) was added at a ratio of 1.2 mL per gram of the ground yeast. The sample was then spun at 14,800 x g for 30 minutes at 4°C, and the supernatant was assayed by BCA for protein content (approx. 6-7 g/L on average).

3.2.2.2 Protein Precipitation and Digestion

The protein supernatant was reduced and alkylated using DTT and IAA according to the procedures described in Section 2.2.2.4. Upon completion, the protein sample was precipitated using acetone. Cold acetone (-20 °C) was added in a 4:1 volume ratio to the protein sample and was chilled for one hour in the freezer (-20 °C) and spun at 14800 x g for five minutes. The acetone supernatant was removed and replaced with an equal volume of cold acetone and centrifuged again 14800 x g for five minutes. The bulk of the supernatant was removed, and the samples were placed in a Speed Vac to fully evaporate the solvent. The protein pellet was reconstituted in 50 mM Tris (pH 8.0) to an assumed concentration of 2.0 g/L protein (assuming 100% recovery through precipitation). This was vortexed for 1 minute, mixed with 0.1 g/L trypsin and digested overnight at 37 °C in a water bath. The digestion was quenched through the addition of 10% TFA, added to a final concentration of 0.1% TFA in the sample and a final peptide concentration of 1.0 g/L.

Unless otherwise stated, samples were prepared for HPLC injection by diluting to a final concentration of 1.1 µg/µL in 10% acetonitrile/ 1 mM HCl/ water to ensure that 100 µg of peptide was loaded into the column from a 95 µL injection.

3.2.3 Instrumental Set-up

3.2.3.1 Automated Strong Cation Exchange Fractionation and Clean-Up

The first dimension of separation, SCX, is the source of fractionation while the reversed phase column serves only to concentrate the fractionated peptides and remove any salts and buffers from the sample. These salts are incompatible with mass

spectrometry hence the name strong cation exchange/clean-up liquid chromatography or, S-CLC (Figure 3-20).

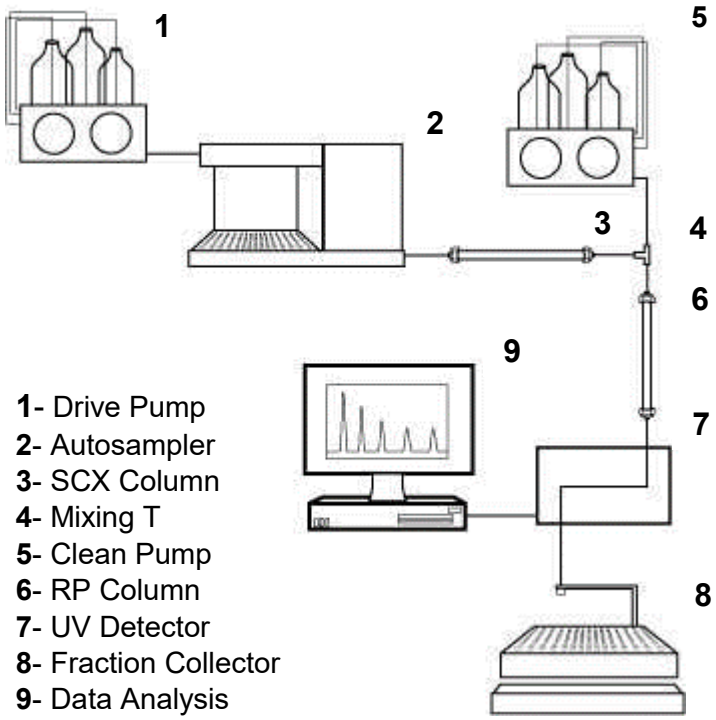


Figure 3-20 Schematic diagram of the SCX peptide fractionation and clean-up system. (**1**) Drive pump continuously flowing 10% ACN/ 90% water/ 1 mM HCl. (**2**) The autosampler, used to both load peptide samples and to inject the separation gradient. (**3**) The first column, which was packed with POROS XS SCX resin. (**4**) The mixing T, which combines the mobile phase eluting from the SCX column with the solvent flowing from (**5**) the Clean-up pump. (**6**) The second column contains reverse phase resin which focusses and cleans the eluted peptide fractions before eluting to (**7**) the UV-VIS detector and collection using (**8**) a fraction collector. After the experiment was complete samples were analysed (**9**) using Chemstation and Microsoft Excel. *Figure courtesy of Jessica Nickerson (Dept. Chemistry, Dalhousie University).*

In reference to Figure 3-20, the S-CLC system consists of two HPLC pumps: the first pump, the 'drive' pump (**1**), and the second pump known as the 'clean-up pump' (**5**). For this thesis, the drive pump was an Agilent 1100 Quaternary pump which flows a continuous supply of 1 mM HCl/ 10% acetonitrile/ water at 50 µL/min through to an Agilent 1100 autosampler (**2**). The solvent lines lead from the autosampler to an SCX column (1 x 50 mm) packed in house using POROS XS Resin (**3**). The autosampler was both the source of the sample, and the eluting solvents, establishing the mixed solvent

gradient profile (Figure 3-20). Samples were injected at a volume of 95 µL for 8 minutes to allow samples to load onto the SCX column. Likewise, when performing separations, each elution buffer was injected as an independent run with a volume of 50 µL for 8 minutes to allow for complete interaction with the SCX column.

After a peptide was eluted from the SCX column, it entered a mixing T (**4**) where it mixes with the mobile phase from the 'clean-up pump' delivered at 200 µl/min. The clean-up pump in this work, a binary Agilent 1100, initially flows a mixture of 80% solvent A (0.1% TFA, water) and 20% solvent B (0.1% TFA, acetonitrile). The mixed solvents from the drive and clean-up pumps were directed to the RP clean-up column (**6**) packed in house using POROS R2 resin 25 µm (cat# 1-1159-05, Thermo Fisher) on a 1 x 50 mm column. This column retains peptides while allowing buffer contaminants to pass through. Once the peptides were bound onto the clean-up column, the clean-up pump switched to 100% Solvent B (0.1%TFA, acetonitrile) ten minutes into the run. The absorbance of the elution peak was measured at 218 and at 280 nm using an Agilent 1100 UV-Vis absorbance detector (**7**) before the sample was collected by a Gilson 203 fraction collector.

3.2.4 Methods for Quantitative Analysis.

3.2.4.1 Analysis of Peptide Recovery from SCX LC System

Quantitative analysis of the peptide recovered from the column was performed utilising Microsoft Excel (See section 2.2.5.1). Two calibration curves were created measuring the absorbance at $\lambda = 214$ nm and $\lambda = 280$ nm using the digested yeast extract through injecting 1, 2, 4, 6, 8, 10, 15, 20 µg into the system (Figure 3-21). This was

performed using the instrumental set up as discussed in section (3.2.2) by replacing the SCX column with a union (Figure 3-21). The instrumental LOD is three times the standard deviation of the lowest replicate. The LOD for each of the calibration curves was 1.13 (λ_{214}) and 1.75 µg (λ_{280}) respectively. The limit of quantitation (ten times the LOD) was 11.3 and 17.5 µg for 214 and 280 nm respectively.

3.2.4.2 Peptide Identification

Mass spectrometry was performed using the same conditions as listed in Section (2.2.6.3) In brief, samples were collected using a Gilson fraction collector, dried under vacuum, and stored at -20 °C. Samples were reconstituted in 5% ACN/ 0.1% formic acid/ water to a final concentration of 0.1 µg/µL. Samples containing less than three µg of peptide were pooled with their neighbouring fraction to ensure enough sample was available for two replicate injections of ten µL with a third available in case of system failure. In the case of the salt and pH gradient (SPG), duplicate injections of eight µL were performed. This adjustment to the procedure was due to a power outage impacting the first attempt at peptide analysis. Once samples were recorded, they were database searched according to the standard methods listed in Section (2.2.6.3).

3.2.4.3 Calculation of Net Charge, Isoelectric Point, and Identification Overlap

The net charge of each of the individual peptides were calculated using the spreadsheet found in Appendix (B). The pKa values for the amino acid side chains and the C and N-termini were used to estimate the charge abundance at a specified pH. This approximation was calculated by multiplying the prevalence of each amino acid side chain within the peptide by their respective charge abundance and then adding the charge of the C and N termini.

The isoelectric point was approximated using the IPC calculator created by L.P. Kozlowski. The peptide sequences were uploaded into the program and the average pI was calculated using ten different algorithms per sequence.[100]

The overlap in peptide identifications among replicates, fractions, and methods were obtained through the use of Venny.[101]

3.2.5 Experimental Procedures

3.2.5.1 Determining the Appropriate Loading Conditions

Optimal loading conditions were tested through repeated addition of digested bovine serum albumin (BSA) in two key experiments. The details regarding buffer conditions for both experiments are found in Table 3-6. The first experiment done was to determine the lowest pH to fully retain peptides on the column and was performed through successively loading 3 x 100 µg of digested BSA in triplicate under five loading conditions. The first four loading conditions were in the absence of salt with pH conditions at 2.1, 3.2, 4.5, 5.5, and the second at a pH of 2.1 with 100 mM NaCl. Once an optimal pH was determined, the second experiment was performed to determine the loading capacity of the SCX column through successively loading 12 x 48 µg of BSA peptide until a total of 570 µg of BSA were exposed to the column.

Table 3-6 The buffer conditions for the experiments found in Section 3.3.2.1.

Experiment	Conditions
Solvent loading conditions Section 3.3.2.1	1. 5% ACN/ 0.1% TFA/ water 2. pH 3.2/ 5% ACN 50mM $(NH_4)_3C_6H_5O_7$ / water 3. pH 4.5/ 5% ACN/ 50 mM $(NH_4)_3C_6H_5O_7$ / water 4. pH 5.5/ 5% ACN/ 0.1% TFA/ water 5. pH 2.1/ 5% ACN/ 100 mM NaCl/ water
Loading capacity Section 3.3.2.1	10% ACN/ 0.1%TFA/ Water

3.2.5.2 Univariate Separation on the S-CLC

To examine the impact of salt and pH on the separation of peptides by SCX, 100 µg of digested yeast protein was injected onto the column. The peptides were then separated either using a salt gradient under constant pH conditions, or by a pH gradient in the absence of additional salt (Table 3-7). Buffer stocks were created using 1 M citric acid, and then titrated using ammonia to produce 500 mM ammonium citrate buffers at the appropriate pH.

Table 3-7 A summary of the solvent conditions required for examining the key variables which influence the separation of peptides via SCX

Experiment	SCX separation conditions
Salt gradient at pH 2.1 Section 3.3.3.1	(Load): 0 mM NaCl/ 10% ACN/ 0.1% TFA/ water pH 2.1 1: 50 mM NaCl/ 10% ACN/ 0.1% TFA/ water 2: 100 mM NaCl/ 10% ACN/ 0.1% TFA/ water 3: 200 mM NaCl/ 10% ACN/ 0.1% TFA/ water 4: 400 mM NaCl/10% ACN/ 0.1% TFA/ water Column clean: 100 mM NaCl/ 10% ACN/ 50 mM K_2CO_4 pH 11
Salt gradient at pH 6.50 Section 3.3.3.1	(Load): 0 mM NaCl/ 10% ACN/ 0.1% TFA/ water pH 2.1 1. 50 mM NaCl/ 50 mM $(NH_4)_3C_6H_5O_7$ 10% ACN/ 90% water 2. 100 mM NaCl/ 50 mM $(NH_4)_3C_6H_5O_7$ / 10% ACN/ 90% water 3. 200 mM NaCl/ 50 mM $(NH_4)_3C_6H_5O_7$ / 10% ACN/ 90% water 4. 400 mM NaCl/ 50 mM $(NH_4)_3C_6H_5O_7$ / 10% ACN/ 90% water Column clean: pH 11 100 mM NaCl/ 50 mM K_2CO_4 / 10% ACN/ 90% water
pH gradient Section 3.3.3.2	(Load): 0 mM NaCl/ 10% ACN/ 0.1% TFA/ water pH 2.1 1. Buffer pH 3.50 mM $(NH_4)C_6H_7O_7$ / 10% ACN/ 90% water 2. Buffer pH 4.5: 50 mM $(NH_4)C_6H_7O_7$ / 10% ACN/ 90% water 3. Buffer pH 5.5: 50 mM $(NH_4)_2C_6H_6O_7$ / 10% ACN/ 90% water 4. Buffer pH 6.5: 50 mM $(NH_4)_3C_6H_5O_7$ / 10% ACN/ 90% water 5. Buffer pH 9.0: 50 mM $(NH_4)_3C_6H_5O_7$ / 10% ACN/ 90% water Column clean: Buffer pH 11.0/ 50 mM K_2CO_4 / 400 mM NaCl/ 10% ACN/ 90% water

3.2.5.3 Modifying the microMUDPIT Method for Separation on the S-CLC System

To compare the results of an optimised method to a gold standard, the MuDPIT gradient shown in Figure 2-3, Section 2.1.4 was modified to remove the organic phase washes as they are not necessary for a one-dimensional separation and would interfere with the peptide clean-up on the RP column. The number of salt steps was reduced from 40 to 8; however, the ratio of A (10% acetonitrile/ 0.1% formic acid/ water pH 2.6) to B (10% acetonitrile 0.1% formic acid/ 500 mM ammonium acetate ($NH_4C_2H_3O_2$) / water pH 6.8)[56] was maintained through calculating the slope at each part of the gradient to ensure the rate of the gradient remained similar despite fewer fractions (Table 3-8. Seventy-five percent of the fractions in the modified MuDPIT method occurred during the first portion of the gradient. This corresponds to six fractions in the condensed method which maintained the slope of the MuDPIT gradient which ranged from 0 to 50% B. The remaining fractions were divided evenly across the slope of the second portion of the gradient. The final modification to the buffer system was the change from 0.1% formic acid to 0.1% TFA in both A and B. The volatility of TFA allows it to be removed from fractions to prevent contamination of the samples in later runs.

Table 3-8 A summary of the solvent conditions required for the MuDPIT Method modified for separation using the S-CLC.

Experiment	SCX separation conditions
Modified microMUDPIT Section 3.3.3.3	Load 10% ACN/ 0.1%TFA/ water A: 10% ACN/ 0.1% TFA/ Water pH 2.1 B: 10% ACN/ 0.1% TFA/ Water 500 mM $NH_4C_2H_3O_2$ pH 6.49 1. 10% B (50mM $NH_4C_2H_3O_2$ pH 5.26) 2. 20% B (100 mM $NH_4C_2H_3O_2$ pH 5.64) 3. 30% B (150 mM $NH_4C_2H_3O_2$ pH 5.84) 4. 40% B (200 mM $NH_4C_2H_3O_2$ pH 5.97) 5. 50% B (250 mM $NH_4C_2H_3O_2$ pH 6.08) 6. 70% B (350 mM $NH_4C_2H_3O_2$ pH 6.24) 7. 90% B (450 mM $NH_4C_2H_3O_2$ pH 6.35) 8. 100% B (500 mM $NH_4C_2H_3O_2$ pH 6.24) Column clean: Buffer pH 11.0: 50 mM K_2CO_4 / 400 mM NaCl/ 10% ACN/ 90% water

3.2.5.4 Combining the pH Gradient with Salt for Optimised Separation

As an example of how to expand into a multivariate separation of a specific proteome, the results from the univariate separations by salt and pH gradients alone were combined to establish a mixed mode gradient. One hundred µg of digested yeast peptides were injected onto the column and separated using fractions which contained a NaCl gradient from 10 to 100 mM and a 50 mM ammonium citrate ($(NH_4)_3C_6H_5O_7$) buffer gradient ranging in pH from 2.1 to 11. Full details on the mixed mode gradient is described in Table 3-9.

Table 3-9 The solvent conditions required for the initial stages of optimising a solvent gradient for the separation of yeast peptides.

Experiment	SCX separation conditions
Gradient combining salt and pH Section 3.3.3.4	Load: 10% ACN/ 0.1% TFA/ water 1. Buffer pH 3: 50 mM $(NH_4)_3C_6H_5O_7$ / 10 mM NaCl/ 10% ACN/ 90% water 2. Buffer pH 4.5: 50 mM $(NH_4)_3C_6H_5O_7$ / 10 mM NaCl/ 10% ACN/ 90% water 3. Buffer pH 5.5: 50 mM $(NH_4)_3C_6H_5O_7$ / 50 mM NaCl/ 10% ACN/ 90% water 4. Buffer pH 6.5: 50 mM $(NH_4)_3C_6H_5O_7$ / 100 mM NaCl/ 10% ACN/ water 5. Buffer pH 9.0: 50 mM $(NH_4)_3C_6H_5O_7$ / 100 mM NaCl/ 10% ACN/ 90% water Clean: Buffer pH 11.0: 50 mM K_2CO_4 / 400 mM NaCl/ 10% ACN/ 90% water

3.3 Results and Discussion

3.3.1 System Design

3.3.1.1 What is the S-CLC?

The strong cation exchange clean-up liquid chromatography (S-CLC) system works largely in part due to three key features. Firstly, the buffer in the drive pump was carefully selected to contain 1 mM HCl. This maintains the acidity of the column between injections but lacks the ionic strength to interfere with the pH of the buffers injected through the S-CLC, if each buffer has a sufficient buffering capacity. The second key feature arises from the discrepancy in flow rate between the two pumps when they reach the mixing T. Samples elute from the SCX column at a flow rate of 50 µL/min and mix with the clean-up pump flowing at 200 µL/min. The mixing resulted in a five-fold dilution of sample flowing from the SCX column, more specifically, a separation buffer with a 50 mM concentration was diluted to 10 mM in the mixing T. The concentration of the acetonitrile from the clean-up pump was also diluted by solvent from the drive pump. For instance, under loading conditions, 20% of acetonitrile delivered by the clean-up pump became approximately 16% as it entered the R2 column (Figure 3-21).

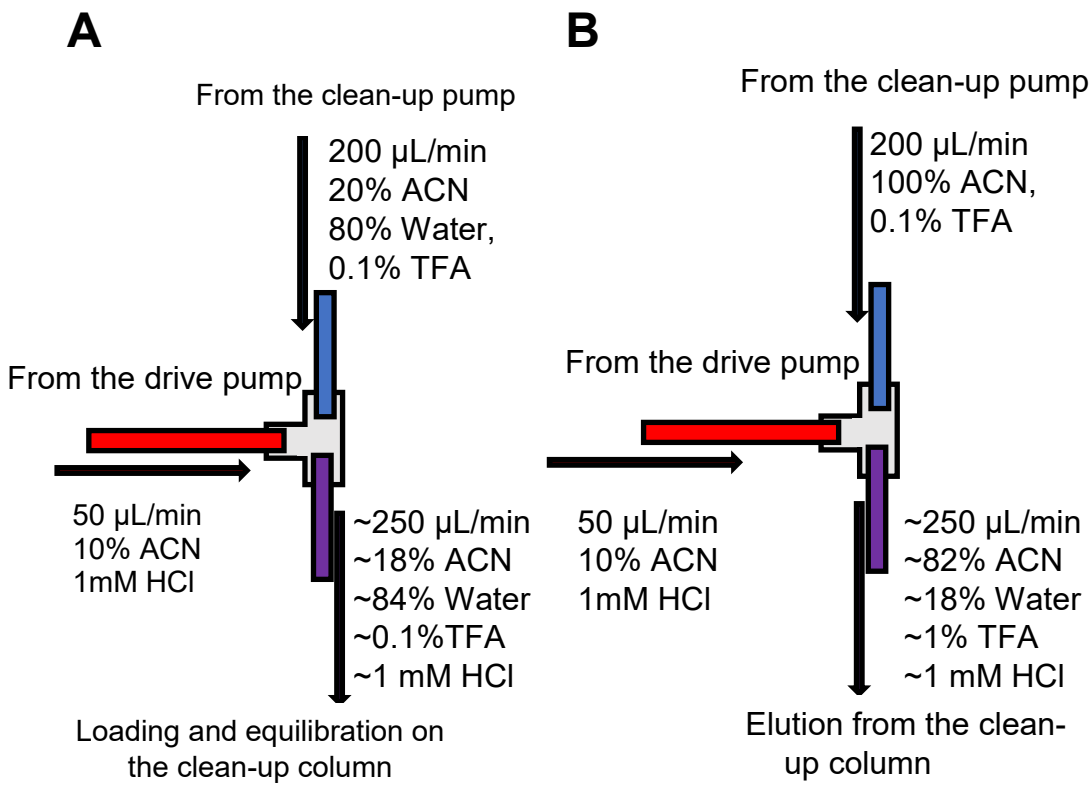


Figure 3-21 A diagram of the solvent conditions before and after entering the T-mixer in the S-CLC. (**A)** displays the solvent conditions while loading peptides and equilibrating the R2 clean-up column while (**B)** displays the solvent conditions during the removal of peptides from the R2 column.

3.3.1.1 Quantifying Data Obtained From the S-CLC

To ensure that peptides eluted from the SCX column were retained by the RP column, the SCX column was removed and replaced with a union. A calibration curve was created by measuring the absorbance at 214 and 280 nm for two injections ranging from 2 to 20 µg with yeast peptides (Figure 3-22). The calibration curves for both separations were linear with a LOD of 1.1 µg at λ_{214} and 1.8 at λ_{280}. The LOD for this system at $\lambda_{214\ was}$ twice the size of the traditional RP HPLC system (0.5 µg). This is expected given the two separate drive pumps are running simultaneously which causes fluctuations in flow rate and increases the signal-to-noise.

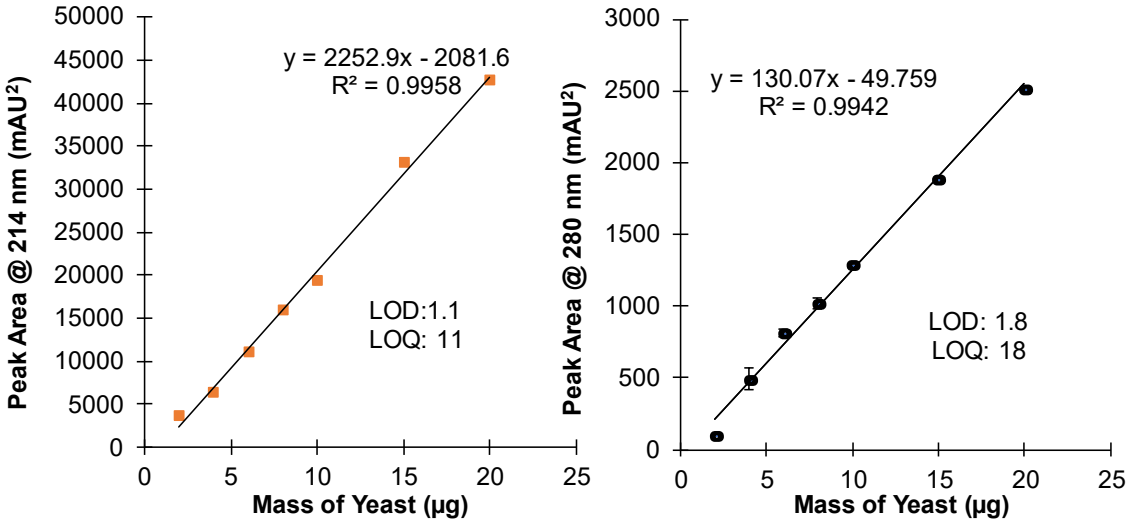


Figure 3-22 Sample calibration curves for λ_{214} and λ_{280} used to determine the amount of yeast recovered in each fraction from the column.

3.3.2 Examining Peptide Retention on the S-CLC

Given the operation of a new system, it was pertinent to determine the loading capacity of the SCX resin utilised in this chromatographic system. This was tested through two experiments discussed below, using digested BSA to help determine appropriate loading conditions for the system. When sample was loaded onto the SCX column, any unretained peptides from the SCX column were caught on the reverse phased column, cleaned, and quantified using LC-UV. Given, that pH, and salt were found to reduce the SCX column affinity for peptides, the loading conditions of the column were first examined (Section 3.3.2.1) prior to examining the overall column capacitance (Section 3.3.2.2).

3.3.2.1 Determining the Appropriate Buffer Conditions for Sample Loading

Given that the work in Chapter 2 determined that pH and salt have a significant impact on the retention of peptides, it was pertinent to examine the impact of pH and salt on the loading capacity of the column. The results of repeated injections of 100 µg can be

found in Figure 3-23. It was expected that increasing the pH or salt concentration of a buffer used to load sample onto an SCX column will result in a decrease in the retention of sample.

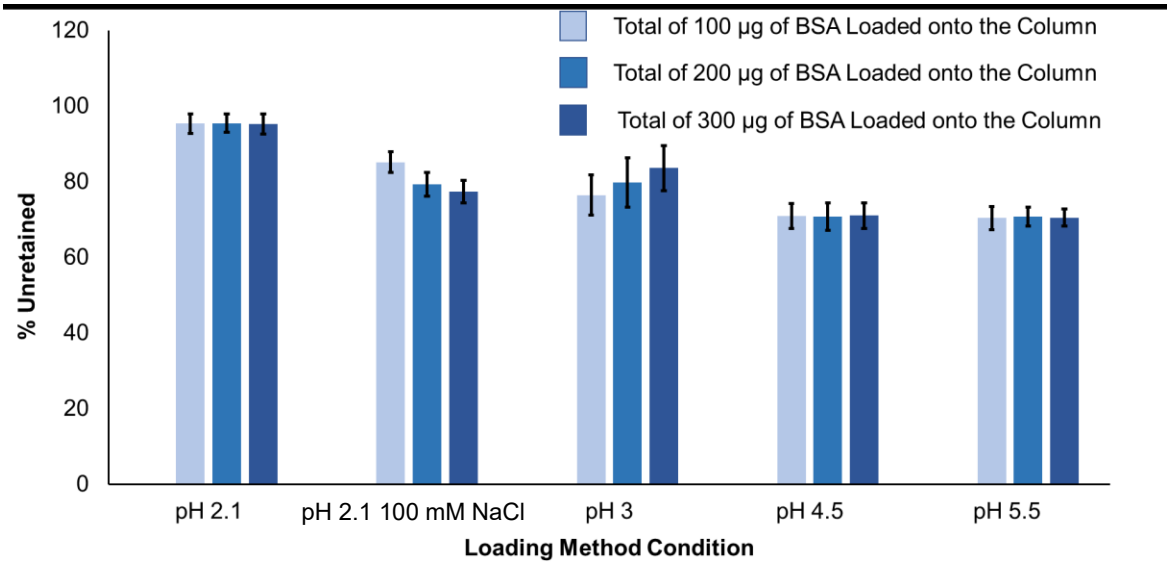


Figure 3-23 The retention of three successive injections of 100 µg of BSA peptides on the S-CLC under a variety of loading conditions. (N = 3)

Referring to Figure 3-23 under the lowest pH conditions (2.1), there is the largest amount of peptide retention by the column. Beginning with the 100 µg injection, over 95% of the peptides were retained on the column. Interestingly, the second and third 100 µg injections added to the column experienced the same 5% of sample loss. This indicates that it is not at the capacity of the column since successive injections should show a decrease in retention. At this pH, the peptides with multiple positive charges have a strong affinity for the column.

As expected, increasing the pH of the loading conditions decreased the retention of peptides on the column. From the outset increases to a pH of 3.2 resulted in a drop in recovery to 80%. This was also seen at pH 4.5 and 5.5 which had recoveries of 29 ± 3%

and 30 ± 3% respectively. These results were expected as the pH of the loading buffer approached and surpassed the pKa of Asp's (3.9) and Glu's (4.3) side chains. This decreased the column's affinity for peptides and resulted in an increase in the total number of unretained peptides caught by the clean-up column.

The impact of salt on the retention of peptides loaded onto the column shows a noticeable decrease in the retention of peptides between the first fraction (85 ± 3%) and the second (69 ± 3%). Comparing the total peptide recovery at pH 2.1 (95 ± 3%) in the absence of salt, with the recovery in the presence of salt (42 ± 4%) shows that salt decreased the overall loading capacity of the column. However, the retention of the peptides on the column in the first injection, under low pH conditions containing NaCl, was greater than the retention of peptide in any of the three increased pH conditions in the absence of salt. This shows that while salt decreased the overall capacity of the column, the samples have some salt tolerance. This indicated that samples should be loaded under conditions pH 2.1 or lower, and that the presence of cations in the buffer solution must be avoided as much as possible to ensure the maximal retention of peptide on the SCX column.

3.3.2.1 Loading Capacity of the SCX Column

The experiment summarised in Figure 3-23 determined that under low pH conditions the capacity of the column exceeds 300 µg. The capacity of the SCX column was further tested using a total of 570 µg of digested BSA loaded in 48 µg increments. In addition, the amount of acetonitrile in the load pump was doubled to 10% to reduce column back pressure. The amount of protein retained in each injection is shown in Figure 3-24.

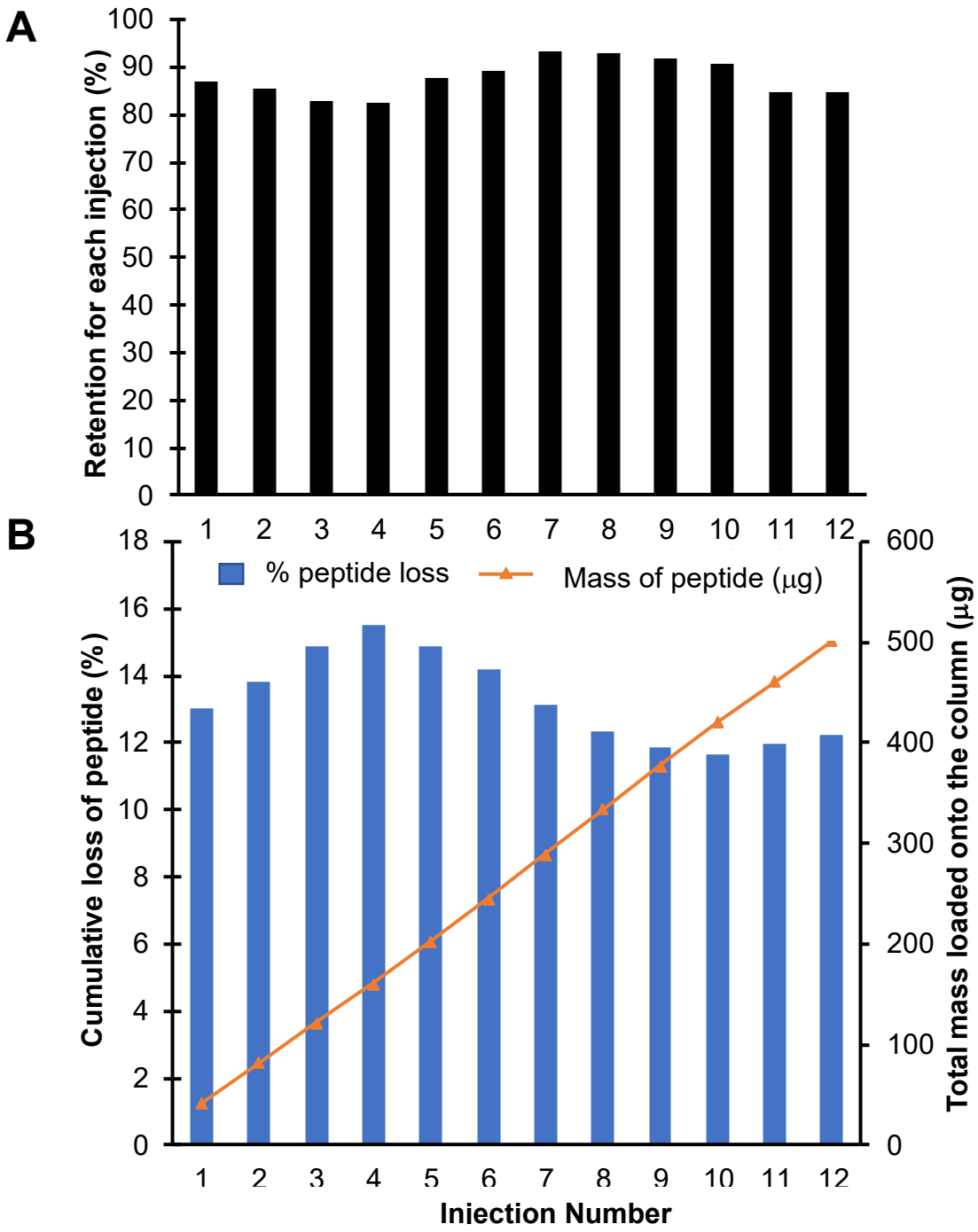


Figure 3-24 The retention of peptide on the S-CLC SCX column. (A)The amount of peptide retained in each of the twelve 48 µg injections, which has an average retention of 94± 2%. (B) The cumulative loss of peptide expressed as a percentage (primary axis: blue) when compared to the total mass of peptide loaded onto the column (percent loss of peptide retained and unretained on the column as a percentage of the total amount loaded with each injection of 48 µg). This shows that the percent retention shows no trend in sample loss.

Referring to Figure 3-24A, there is an average of 94 ± 2% recovery for each 48 µg injections. While Figure 3-24B, shows that with each successive injection that the cumulative amount of peptide that was unretained by the column as a percentage is conserved. As an additional comment, it was observed experimentally that overall percent retention for sample in loading can drop due to the incomplete removal of peptides from previous runs. This is known as column fouling.

3.3.3 Comparing Separation Gradients Using S-CLC

3.3.4.1 The Impact of Salt Gradients on the Separation of Peptides on SCX

From Chapter 2 it was determined that salt had a minor influence on SCX separations. A salt gradient was performed under low pH conditions, pH 2.1 (10% ACN/0.1% TFA/ water) and was compared to a salt gradient in weak acidic conditions, pH 6.5 (10% ACN/ 50 mM Citric Acid/ water). Given that low pH conditions are optimal for binding, attempting to perform a salt gradient under these conditions may yield poor sample recoveries, as the fully protonated peptides will have a strong affinity for the SCX resin. The comparison of the salt gradient performed in highly acidic (pH 2.1) conditions to the gradient in weakly acidic conditions (pH 6.5) is shown in Figure 3-25.

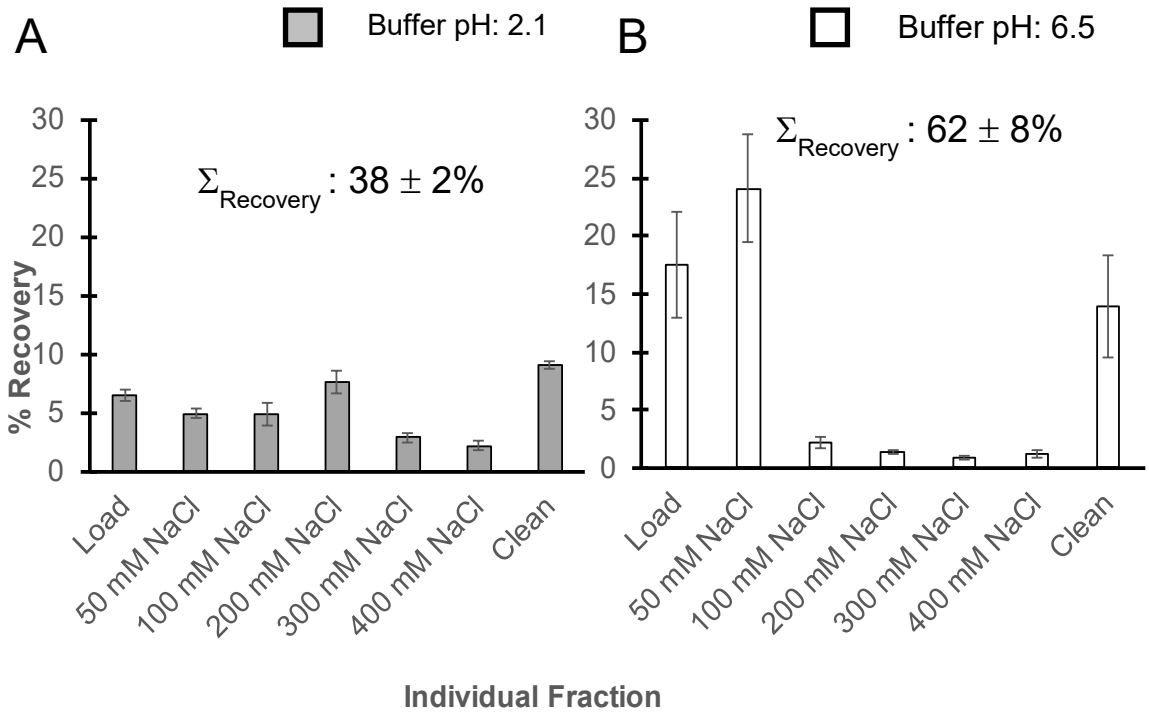


Figure 3-25 The recovery of peptide from the S-CLC using a salt gradient under fixed pH conditions and under strongly acidic conditions at a pH pf 2.1 (**A**), and under weakly acidic conditions with a pH of 6.5 (**B)** N = 2.

While the strongly acidic conditions represent optimal binding conditions, as expected, this pH does not translate to optimal conditions to recover peptides using a salt gradient. As seen in Figure 3-25A, only 38 ± 2% of yeast peptides loaded onto the column were recovered with the largest fraction of peptides eluted in the 200 mM NaCl fraction containing 8 ± 1 µg. This is below the 14 µg target recovery for each fraction (100 µg evenly spread across 7 fractions). The clean-up fraction, which exposed the column to a buffer at pH 11 with 100 mM NaCl, should theoretically remove all the peptides still bound to the column; however, only 5 ± 0.2% of peptides were recovered in this fraction. There are two key theories which can help explain this phenomenon. The first postulates that over time permeant bonds on the SCX may occur due to a change in peptide conformation.[31] However this would require the peptides to spend an extended

period of time on the column, and the total run time for one replicate in this experiment is between four and five hours. It is far more likely that the mobile phase is unable to interact with all the peptides within the perfusion resin. Attempts to improve this include larger or multiple injections of the clean-up.

In contrast to separation at pH 2.1, separating under weak acidic conditions yields significantly greater total recoveries when using a salt gradient. In this case, the recovery was 62 ± 8% which corresponds to over a 60% improvement when compared to the salt gradient at lower pH. At pH 6.5, the peptides eluted in the first salt fraction (50 mM NaCl) indicates that these peptides are weakly retained at a high pH. These trends in peptide recovery are consistent with the results observed in Chapter 2. This also demonstrates that the buffering capacity of the eluting solvent is sufficient to overcome the influx of 1 mM HCl into the SCX column from the drive pump.

Interestingly, the clean fraction (pH 11 with 100 mM NaCl), was more successful at pH 6.5 (14 ± 4μg) than at pH 2.1 (4.6 ± 0.2 μg). This may be because loading and separating at a low pH for an extended period of time may result in irreversible binding to the column.[31] This shows that it is essential to switch off the low pH buffer quickly. Here it is reinforced that the pH of the elution buffer plays a significant role in the recovery of peptides from SCX. Although 60% total recovery was an improvement over the low pH salt gradient run, subsequent experiments use a combination of salt and pH to further improve recovery.

3.3.4.2 The Impact of pH on SCX Separation of Peptides.

One hundred micrograms of yeast digest were injected into the SCX column and the samples were eluted using a stepped-pH-gradient from a pH of 3 to 11 using ammonium citrate and sodium carbonate. It is expected that increasing the pH will elicit the recovery of peptide from the SCX column. The results of the experiment are shown in Figure 3-26.

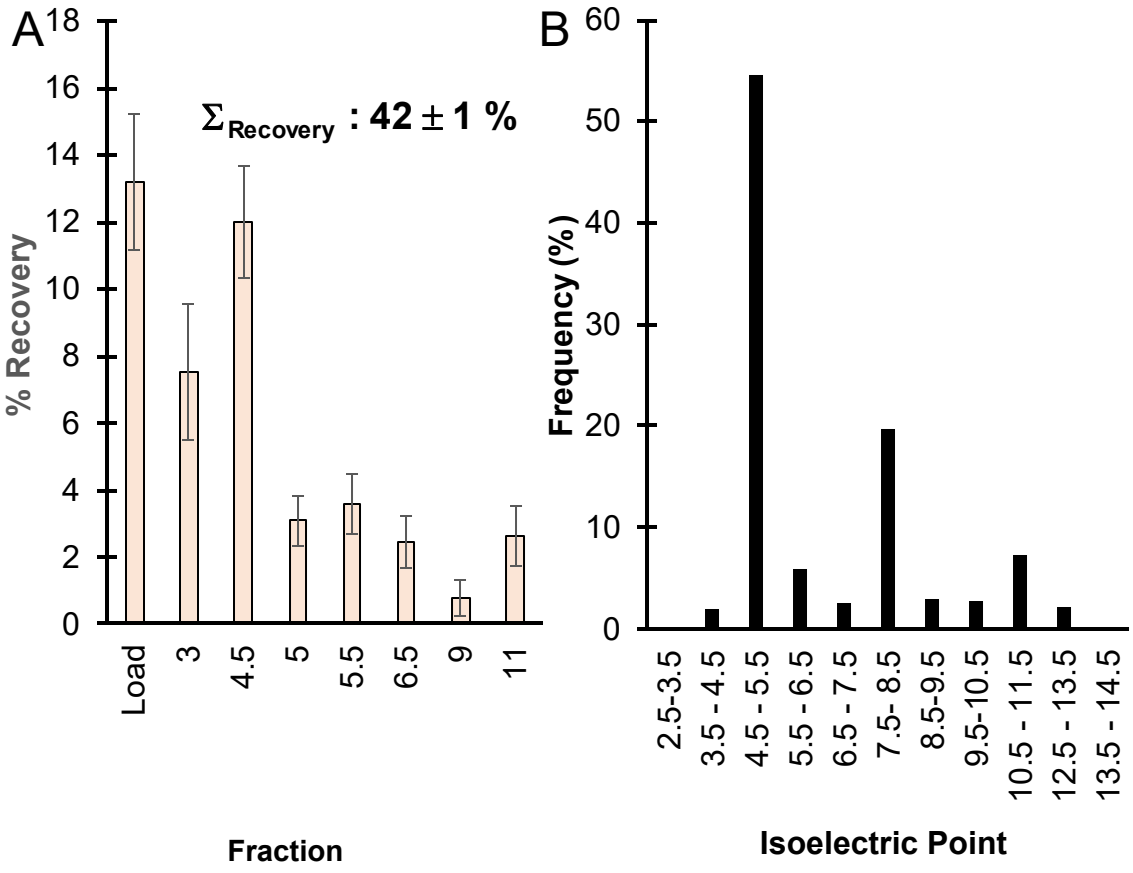


Figure 3-26 (**A**) The recovery of peptide using a pH gradient in the absence of salt (N = 3). (**B**) A histogram creating using the calculated isoelectric point of the 3398 peptides identified by mass spectrometric analysis of replicate 4 from the pH gradient in chapter two.

Referring to Figure 3-25, the overall trend of peptide recovery is like those observed in Chapter 2 under changing pH conditions. Each successively higher pH buffer recovered an additional portion of the total sample bound to the SCX column. The total recovery under changing pH conditions was 42 ± 1%, being comparable to the lower pH salt gradient.

Unlike the salt gradient, where peptide recovery was progressively smaller at the high salt fractions, there are noticeable spikes in recovery for peptides after the injection of buffer at a pH of 4.5 (12 ± 2%), at pH 5.5 (3.6 ± 0.9%), and at pH 11 (2.6 ± 0.9%). This correlates with the isoelectric point distribution of the peptides. The pI distribution plotted above in Figure 3-25B was created using peptides fractionated by the ProTrap XG and identified by LC-MS from the pH gradient in the absence of salt (Chapter 2). The pI's of the peptides have a wide distribution ranging from 3 to 13. Fifty-five percent of the identified peptides have a calculated pI between a pH of 4.5 and 5.5. As the pH of the eluting buffer transitions above the pI of the peptide, the charge state converts from positive, retained on SCX, through neutral to negative, unretained. The highest pH buffer, clean fraction, accounts for 3% of the peptides bound to the column; at this pH most peptides should elute from the column. The distribution of charge around the peptide chain plays a role in retaining the peptides which may explain why certain proteins with high pI above 11 were identified. This may also explain why we do not observe ~95% recovery or higher from the pH gradient separation in S-CLC.

3.3.4.3 Separation of Yeast Using the microMUDPIT SCX Gradient

Given that salt and pH alone were found to be insufficient in obtaining high recovery (>90%) for the peptides on the SCX column, a multivariate gradient was

employed. This combined a pH and salt gradient in an attempt to an increase the overall peptide recovery. The MuDPIT approach is widely considered the gold standard of two-dimensional proteome separation. Here, Webb's[82] MuDPIT method was adapted into a 10-fraction S-CLC separation (Section 3.2.5.3) and the recovery of the peptides from each of the fractions is summarized in Figure 3-27. The objective is to evenly divide the total mass of peptide injected on SCX across the 10 S-CLC fractions. This maximizes the potential to detect the peptides by LC-MS by reducing complexity of the proteomic mixture. The experiments were performed using two replicates of 100 μg of digested yeast peptides, with each fraction of the separation further examined by mass spectrometry.

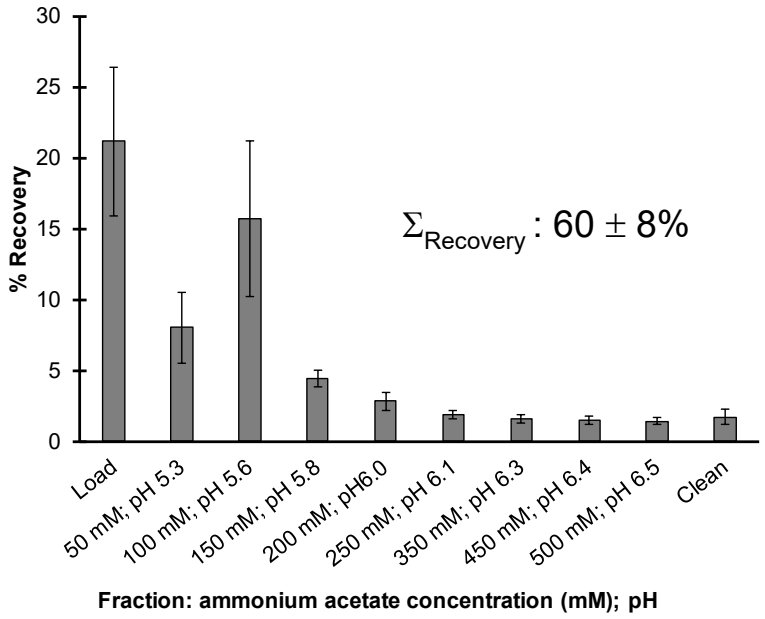


Figure 3-27 The recovery of 100 μg of yeast peptides from the column for each fraction using the modified form of the microMUDPIT method (N = 2).

The total recovery using the MuDPIT gradient was comparable to the high pH salt gradient wherein 60 ± 8% recovery was observed. Unfortunately, the mass distribution of

peptides across the peptides did not follow the theoretical ideal scenario of 10% per fraction. Instead, the bulk of the recovered peptide (~75%) was obtained in the first three fractions. As the separation proceeds to higher salt concentration and pH, there are no spikes in recovery for the remaining seven fractions; in fact, the recovery becomes progressively lower. This can be explained by considering that the pH range employed from the MuDPIT gradient is minimal with the pH of the first fraction at 5.3 and the final fraction at pH 6.5.

While it would be desirable to compare the total recovery obtained in this work to Webb's publication, this information is not provided in that work. However, Elschembroich *et al.*[102] used the twelve-hour MuDPIT method to examine mouse skeletal muscle and found that of the 6000 peptides identified, approximately 73% of peptides were first identified in the first four fractions which follow the same recovery trends as those observed in the present study. This logically concludes that MuDPIT is not an optimised gradient. Ideally, the gradient should transition through a pH of 2, 3 and 4, prior to transitioning to a higher pH for a greater number of identifications. However, the point of this work was not to improve MuDPIT, but to examine how this 'so called' gold standard of separation performs and compares to our multivariate method.

3.3.5 Towards an Optimal Multivariate Gradient

A separation was performed using both pH and salt to increase the recovery and improve the separation of the sample. In the microMuDPIT method, an increase in counterion concentration must always be associated with an increase in pH.[54] With S-CLC, the pH and salt gradient no longer need to be linked, allowing the researcher to prepare each elution buffer at any desired pH and salt concentration. This offers the greatest opportunity to tailor a separation

method according to the needs of the experiment and is a relatively facile way of approaching optimisation. This is seen through a single replicate experiment using the digested yeast sample and the pH gradient shown in Figure 3-26, with the addition of salt (Section 3.2.5.4). The recovery data for the combined gradient is shown below in Figure 3-28.

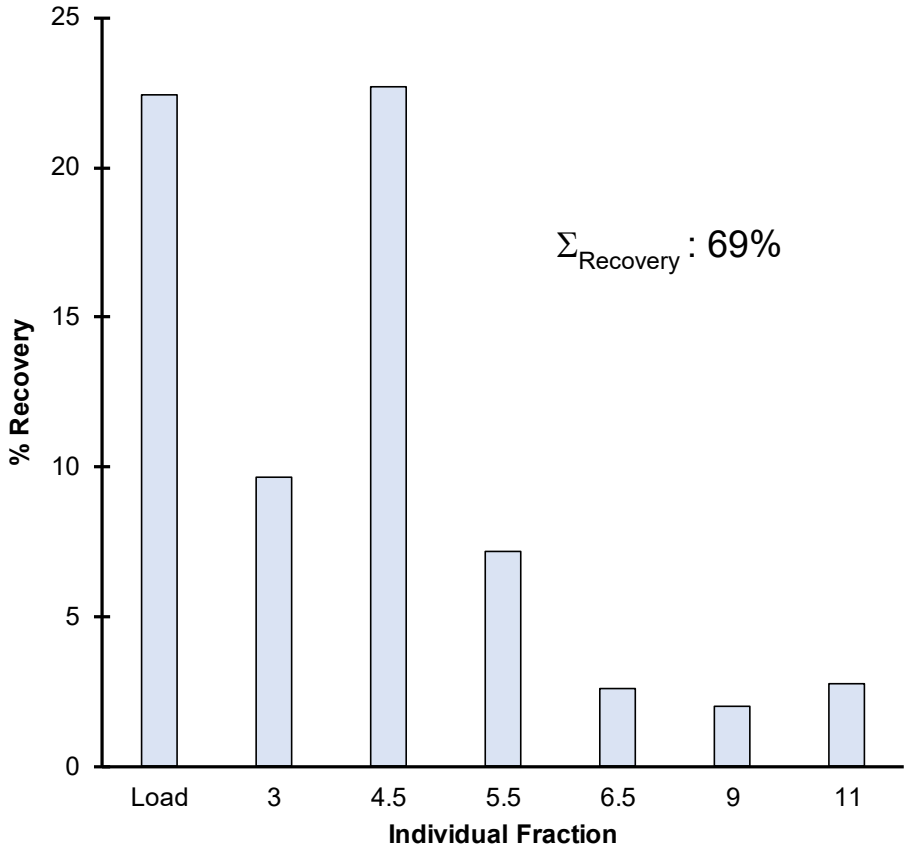


Figure 3-28 The recovery of 100 μg of yeast peptides using a gradient that combines a salt and pH gradient (N = 1).

The recovery obtained with the combined salt – pH gradient was the highest of all the recoveries observed with 69% total yield. This is a statistically significant improvement on the recovery with salt along at low pH as well as the recovery with a pH gradient at low salt ($P < 0.05$). One clear advantage of using a stepped gradient in S-CLC is that each individual buffer can be prepared as desired by the user. This approach can induce a greater spread of sample across each fraction. For example, at a pH of 3 with 10

mM NaCl, ~10% of peptide is eluted from the column. This is not statistically different in terms of recovery from the univariate pH gradient at pH 3. However, there is an observable difference in the next fraction at pH 4.5 where the mixed gradient fraction recovered 23% of the peptide and the univariate pH gradient recovered 12 ± 2%. This can be explained when considering the properties which influence retention onto the column. Increasing the pH of the mobile phase decreases a peptide's affinity for the strong cation exchange resin by deprotonating their amino acid residues and lowering their net positive charge. A pH of 4.5 is above the pKa of glutamic acid (pKa = 4.3) and aspartic acid (pKa = 3.9). This means that a significant fraction of peptides at any given moment will now have segments of their chain with a negative charge. This negative charge decreases the overall affinity of the peptide for the column. Depending on the peptides charge dispersion, many of these peptides may still be retained on the column. However, the reduced affinity on the column resulting from a change in pH from 3 to 4.5 is now sufficient for a low concentration of salt to effectively compete for interaction with the column. This results in larger amounts of peptide recovered in the mixed fraction than using pH or salt alone.

In a system whose predominant mode of separation is based on pH rather than the salt concentration, a researcher should understand the influence of pH before adjusting individual salt steps in the gradient. This allows researchers to compare the impact of salt on retention when adding salt at any given pH. From the data obtained from S-CLC and the ProTrap XG, increasing the salt concentration under constant pH will result in the increased recovery of peptides. Spikes in recovery, such as those observed at pH 4.5 in the presence of 10 mM NaCl should result in method adaptation. The researcher could

either insert an additional fraction between pH 3 and 4.5, at a pH around 4, or they could remove salt from the fraction containing 4.5 and resume the salt gradient in subsequent fractions.

3.3.5 Analysis of Separations Using Mass Spectrometry

Separated fractions of peptides were analysed using LC-MS/MS according to the methods outlined in Section 3.2.4.3 with individual peptides identified using proteome discoverer. As a comparative summary of all SCX separation methods employed, the total number of identified peptides are shown in Figure 3-29A, relative to the total recovery of the same separation platforms Figure 3-29B. This is also compared to the total number of peptides identified using an unfractionated-yeast-peptide sample obtained from four injections onto the mass spectrometer (Figure 3-29).

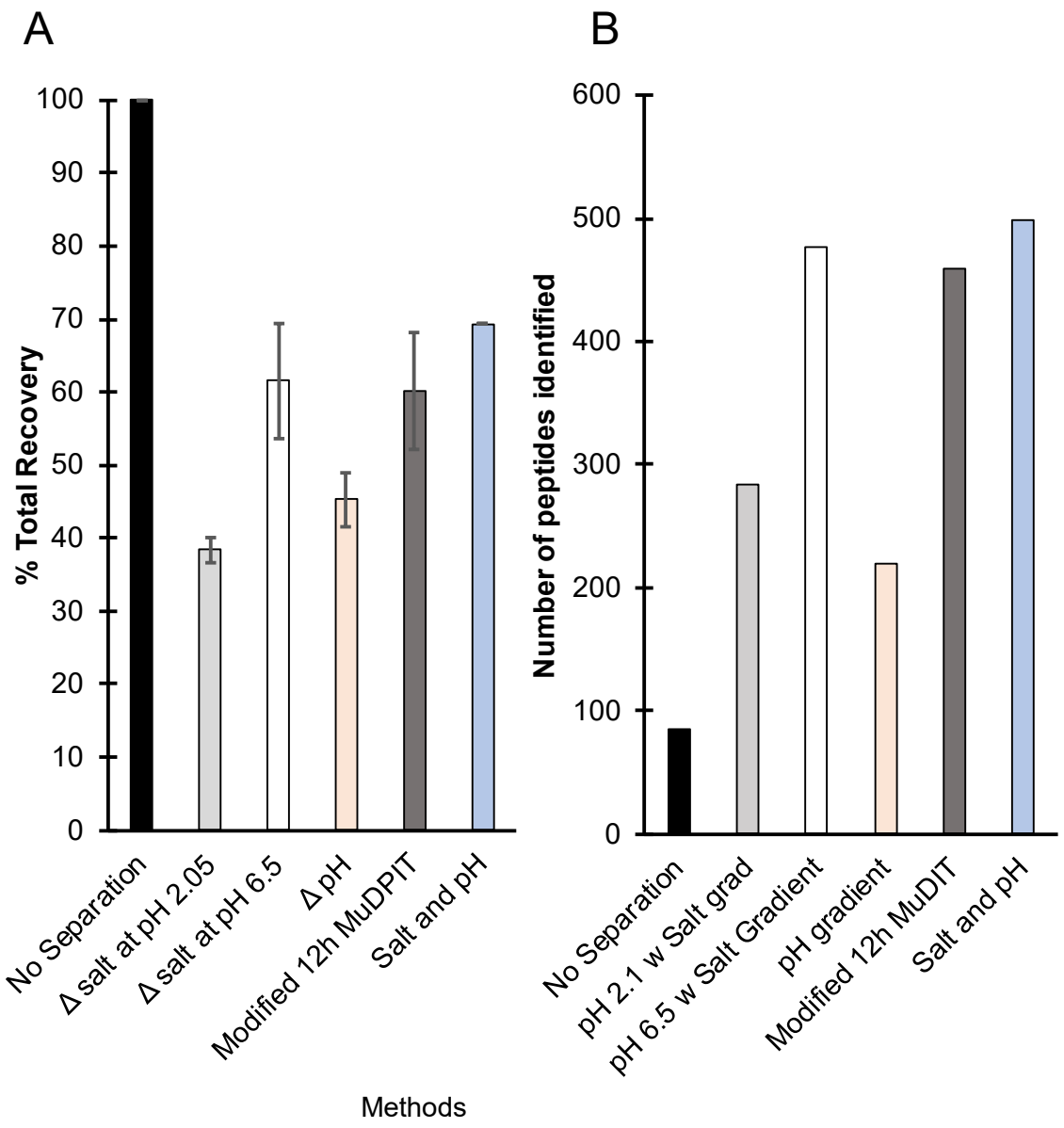


Figure 3-29 (**A**) The total amount of peptides recovered from each of the separation approaches using the S-CLC. (**B**) The total number of peptides identified from the fractions obtained from one replicate of each separation method.

Figure 3-29 shows that high recovery does not always translate to higher peptide count. This is most apparent when examining the unfractionated yeast peptides where 100% of the total number of peptides yielded only 85 unique identifications. In contrast, the salt gradient at pH 2.1 only recovered an average of 38 ± 2% of the peptide from the column but identified 284 peptides. However, the impact of adding salt to the pH gradient is clear when considering the 24% difference in recoveries between the pH gradient in the absence of salt, and the pH gradient in the presence of salt. This also doubled the total number of peptides identified. High recovery translates into an increased number of identifications when combined with separation. This is a result of the multiplicative peak capacity described by Giddings[78] where the best separation on S-CLC will result in the highest possible combined peak capacity with the RP separation, and should yield more identifications on the mass spectrometer.

The peptides identified from the fractions obtained from the salt gradient at pH 6.5, the modified MUDPIT method, and the combined salt and pH gradients on S-CLC were compared to each other in Figure 3-30. There is little difference in the percentage overlap when comparing the peptides identified in each of the three methods (Figure 3-30A vs 30B, 30B vs 30C, 30A vs 30C). Unfortunately, this demonstrates the limitations in terms of resolution of the mass spectrometer where increased peptide identifications should allow for a better understanding of the relationship between peptide identifications

and the resolution of the separation.

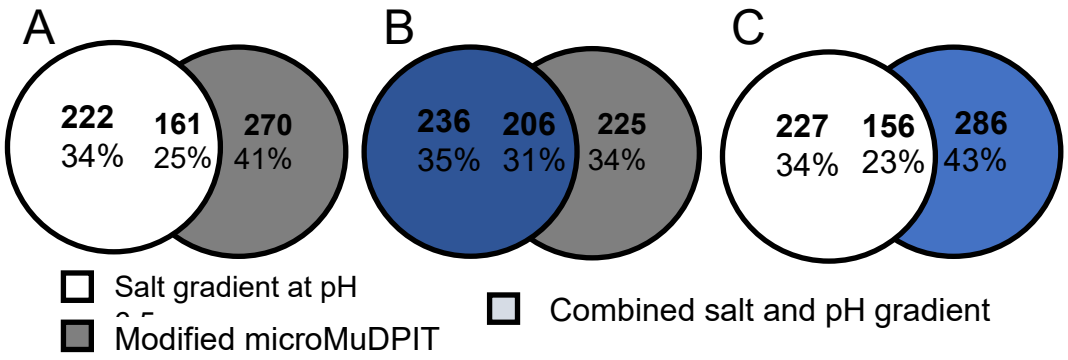


Figure 3-30 The number of identified peptides which are conserved between each separation method. (**A**) The salt gradient at pH 6.5 compared with microMuDPIT, (**B**) MicroMUDPIT compared to the SPG, and (**C**) the salt gradient at pH 6.5 compared with the SPG method.

3.3.7.1 Analysing the Quality of Separation for the Single Variate Separation Methods.

The quality of separation of peptide fractions was assessed through comparing the amount of overlap in peptides identified between each fraction for the various univariate separation gradients (Figure 3-31). Higher quality separation should result in fewer shared identifications. The degree of overlap should drop as one compares fractions that are further apart from each other in the collection sequence i.e. the degree of separation between fractions. For example, fractions 1 and 2 only have one degree of separation and should have more peptides in common than a comparison of fractions 1 and 4 which have three degrees of separation.

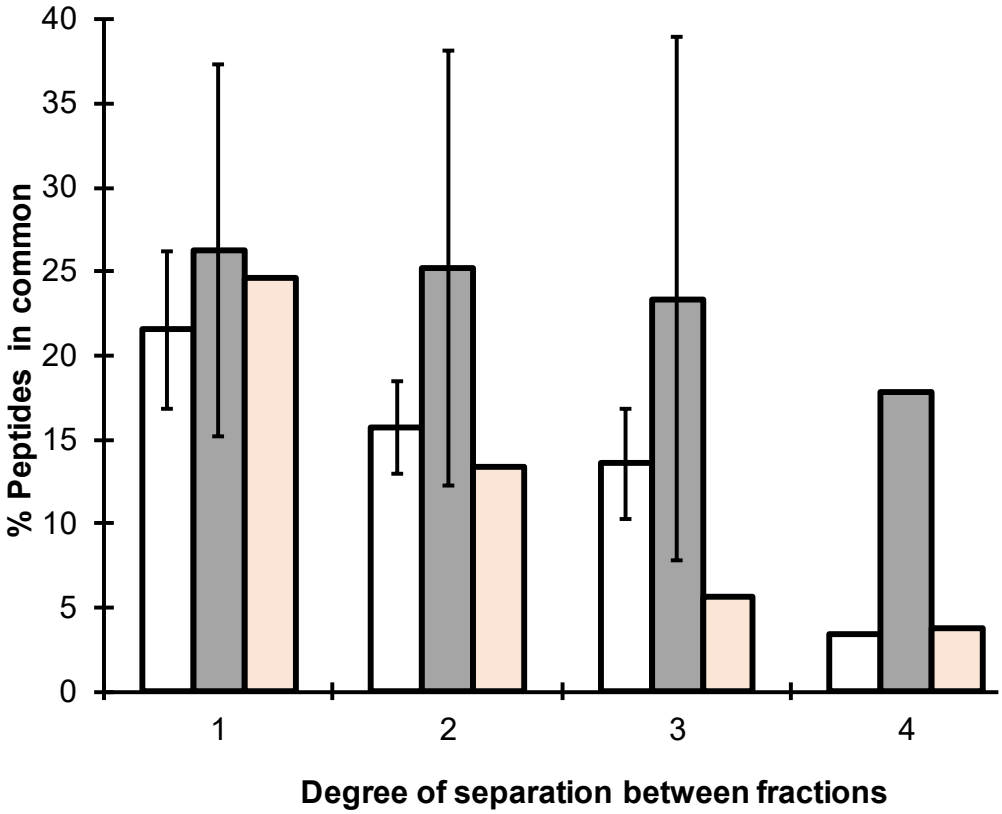


Figure 3-31 Assessment of the quality of the peptide separation through comparing the average overlap between fractions for (A) the salt gradients performed at pH 2.1 (grey), and 6.50 (white), and the pH gradient in the absence of salt (orange). The standard deviation of replicates for each degree of separation decreases N = 4 for the first degree of separation, N =3 for the second, N = 2 for the third.

In reference to Figure 3-31, there is a clear reduction in shared peptides between fractions using a pH gradient in the absence of salt, and those using the salt gradient at fixed pH. The pH gradient allows the peptides on the column to gradually reduce in affinity. This demonstrates that although the overall recovery of the peptides is poor, gradually changing the pH conditions is an effective way of separating peptides and is consistent with the results found in Figure 2-15. Likewise, a salt gradient under strong acidic conditions also results in the reduction of overlap between fractions which are not collected one after the other. An increase in salt results in the elution of peptide by

adjusting the charge affinity equilibria, as the net charge of the peptides remain unchanged by the buffer fraction to fraction. This differs for the separation of peptides using a salt gradient at pH 6.5. Unlike in the case of a pH gradient, the salt gradient at pH 6.5 has a rapid jump in pH in the transition from the loading condition (pH 2.1) to the separation buffer (pH 6.5). This increase would cause multiple peptides to alter their charge state as the pH rises above the pKa of several amino acid side chains. The impact of the ion selectivity coefficient to re-establish equilibria would be a minor factor in comparison to the rapid change in pH which would cause the elution of peptides with a wide variation in charge states.

It is hypothesized under fixed pH conditions, increasing the concentration of salt should elute peptides with successively higher net charge in the later fractions. To test this, the charge states were calculated for all identified peptides and their distribution per fraction was compared for the two salt gradients, with results plotted in Figure 3-32. Overall, there is less variance in charge within each fraction for peptides recovered at a pH of 2.1 (Figure 3-32A), than the fractions obtained using the same salt gradient at pH 6.50 (Figure 3-32B). The variance is so small at pH 2.1 that the boxes indicating the upper and lower quartiles are not visible for the first three fractions. Overall the median charge for separation appears unchanged in Figure 3-32A. However, the distribution of the charge appears to shift upwards due to the appearance of the boxes for the upper quartile. This indicates that increasing concentrations of salt is beginning to recover peptides which have a larger charge state.

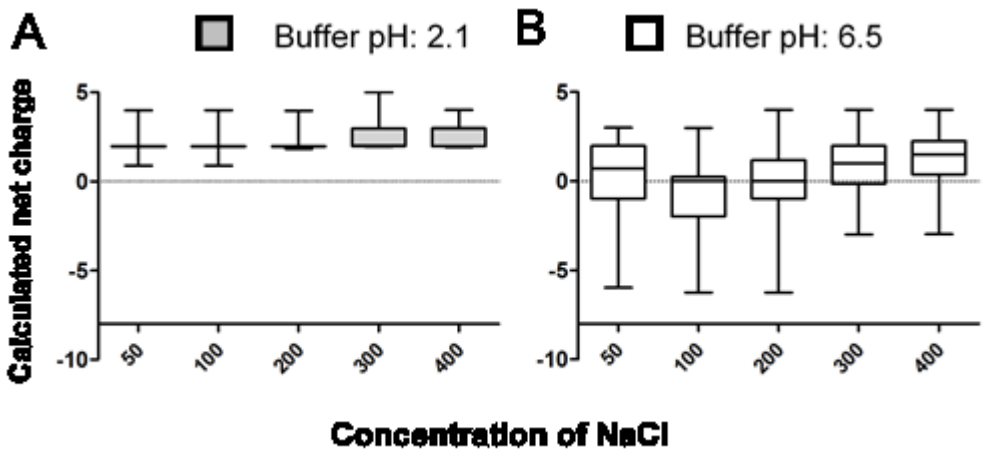


Figure 3-32 A comparison of the variability of charge state calculated for all of the peptides identified for each salt gradient fraction at pH 2.1 (A) and pH 6.5 (B).

In contrast to Figure 3-32A, the first three fractions in Figure 3-32B have a wide distribution of charge with both the upper lower quartile visible and the overall distribution of charge in the first three fractions ranging from -6 to +3, -6.5 to +3, and -6.5 to +4 for the 50, 100 and 200 mM NaCl fractions respectively. While the final two fractions show a decrease in overall variance, the impact of salt on the separation does not have any visible trends. This is consistent with the high amount of overlap detected between each fraction. The separation may simply be due to a limit in the availability of cations rather than the charged state and overall affinity of the samples themselves. This is apparent from the elution profile detected through LC-UV where most of the peptides are eluted by the salt gradient in the first few fractions. Fewer fractions result in more overlap, therefore a lower amount of overall separation.

3.3.7.2 Analysing the Quality of Separation from Multivariate Separation Conditions

Both the microMuDPIT and the multivariate gradient separation use a combination of pH and salt to separate yeast peptides. The degree of peptide overlap for

these separation platforms is plotted in Figure 3-33. There is a high degree of variance in all the points due in part to the differing sample sizes. Because there are more peptides identified in certain fractions than others, this skews the percent overlap. Despite this, in the case of separating peptides using the SPG method, the average overlap in identified peptides between each fraction is decreasing Figure 3-33 (B) while the overlap between fractions in the microMUDPIT method Figure 3-33 (A) remains essentially constant. This indicates that overall there is better separation on the SCX column using our method as compared to the microMuDPIT.

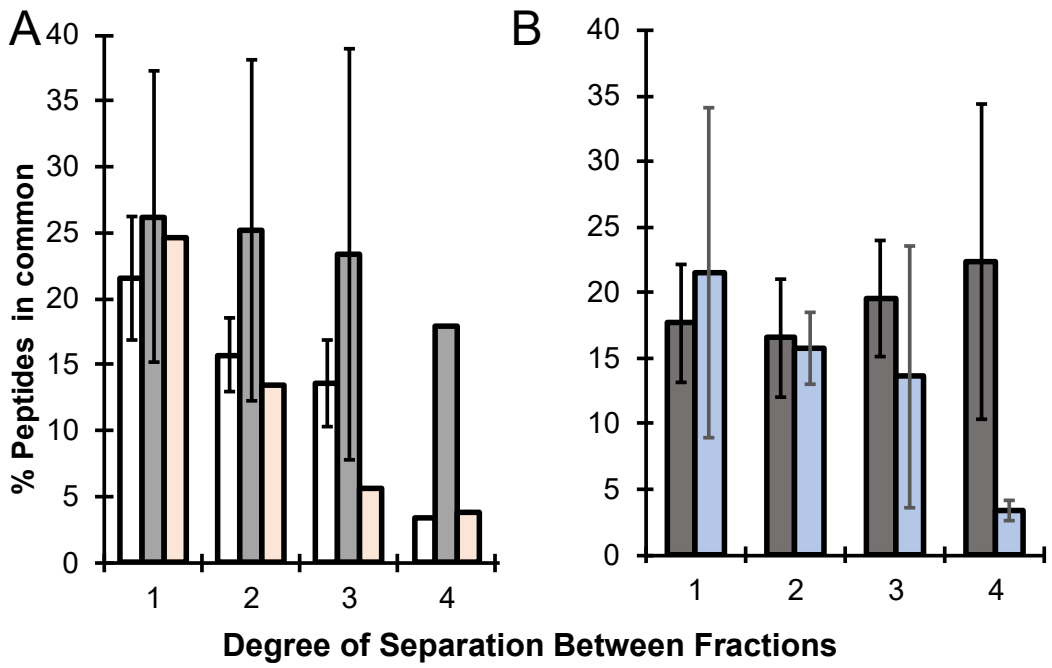


Figure 3-33 *An assessment of the quality of separation for (**A**)univariate separation (salt gradients performed at pH 2.1 (grey), and 6.50 (white), and the pH gradient in the absence of salt (orange)) and (**B**) Multivariate separations the Modified microMUDPIT (Dark Grey) and the SPG method (blue). The standard deviation of replicates for each degree of separation decreases N = 4 for the first degree of separation, N = 3 for the second, N = 2 for the third.*

The poor separation using the MuDPIT method likely arises due to the binary A: B mixing gradient which is unable to provide a wide pH gradient (pH 5.3 - 6.6) unlike our method which employs any pH range that is desired (in this case, pH 3 - 11). This

means that in the case of the MuDPIT method, there is a large jump in pH between loading the peptides at pH 2.1 and the first fraction (pH 5.3). Given that pH plays a dominant role in separation by altering the affinity of the peptides to the column, the jump in pH likely causes the same problem as observed with the salt gradient at pH 6.5. This is easily addressed through the adaptation of a method which allows for a buffer gradient that spans from a low pH to a high pH.

3.4 Conclusions

An LC approach was developed to allow for automated offline SCX separation using an autosampler with a reversed phase clean-up. The ability to select a specific pH, and a specific salt concentration for each fraction allows the user to quickly evaluate and alter various gradients to achieve an optimised separation for a specific proteome. This is not possible in a system where salt and pH are directly linked to each other like in a traditional binary or quaternary gradient system. Additionally, a standard approach of creating an optimised method was also developed where first the user performs a pH gradient in the absence of salt, and a salt gradient held at a fixed pH. An analysis of these two methods can be combined to separate the proteome as evenly as possible. The SPG method in terms of the number of peptide identifications and overall recovery is similar to the traditional gold standard method. However, the degree of separation in the SPG gradient showed a decreasing degree of overlap which indicates a better quality of separation. This could be further explored with a high-resolution mass spectrometer. This method also offers the researcher the opportunity to develop methods with the intention to recover peptides with specific properties in a select fraction for further analysis.

Chapter 4 Conclusion

4.1 Conclusion

Strong cation exchange has a long history of separation spanning many research fields.[88] In theory, strong cation exchange is an ideal mode of separation for peptides.[13] However, in recent years, many proteomic researchers have shifted away from SCX separation in favour of other separation methods that are orthogonal to RPLC (e.g. IEF, high pH reversed phase, capillary electrophoresis) due to challenges with sample recovery and a reduced number of peptide identifications. However, many online separation systems using SCX are likely to have poor resolution due to application of suboptimal gradient systems. This is brought on in part by limitations in the ability for conventional pump systems to deliver an appropriate solvent gradient for separation, and the overall incompatibility of those separation solvents with mass spectrometry. Consequently, the application of offline SCX separations could alleviate these concerns. But first, it is essential to fully examine the variables which control the separation of peptides on SCX. Once the variables which govern the separation of peptides in SCX are understood, they can be utilised to develop a method to optimise the separation of proteomes on SCX. For the purposes of this thesis, an optimised separation is one that not only minimises the degree of overlap between fractions, but also evenly partitions the peptides across the bins (collected fractions), doing so with high yield.

This thesis discussed the development of two devices which allow for offline separation of peptides on SCX. The second chapter discusses a centrifugal separation device called the ProTrap XG, and the third chapter discusses the development of a

multidimensional liquid chromatography system called S-CLC. One of the main advantages of both systems is the potentially limitless choice in solvents available to elute peptides from SCX. These include buffers with ranges of pH, salts at any desired concentration, and inclusion of organic modifiers to alter the solvent composition.

The loading capacity for both the ProTrap XG and the S-CLC were examined, and each was determined to exceed 300 µg of recovery. This is above the desired mass of peptide to be loaded onto the column (200 µg in the case of the ProTrap XG and 100 µg in the case of the S-CLC). Working with the ProTrap XG determined that samples must be loaded in solvents which have a less than 30% organic solvent content. In addition, work using S-CLC found that samples must be loaded onto the column under extremely low pH conditions, below a pH of 2.1, to ensure that most of the peptides are fully protonated and thus bind to the SCX column with maximal efficiency. This is supported by the initial work with pH conditions in chapter 2. Though critical to improving recovery in SCX separation of peptides, neither of these conditions are currently being met in favoured proteomics platforms including the 'gold standard' MuDPIT approach. Consequently, one can expect protein recovery to suffer with the MuDPIT platform.

In addition to determining the conditions for optimal peptide retention on an SCX column during the loading phase, this thesis examines the variables which influence the recovery of peptide from the column. The key variables which impact retention are salt concentration and pH. These variables are examined independently as well as in tandem, comparing the impact of a range of solvent gradients for separation in SCX. It was found using the ProTrap XG and confirmed with S-CLC that pH is the primary variable which governs the elution of peptides from the SCX column using two independent proteomes,

yeast, and *E. coli*. Increasing the pH of the mobile phase reduces both the charge state of the peptides and the affinity of the peptide for the column. In chapter 2, an optimised pH gradient in the absence of salt resulted in a separation with minimal peptide overlap across the fractions. These results were reproduced independently using the S-CLC separation in Chapter 3.

Work with each separation platform also found that the use of cationic salts alone is insufficient for the total recovery of peptides from the column. Under acidic conditions at pH 2.1, it was observed using both the ProTrap XG (Chapter 2), and S-CLC (Chapter 3), that cationic salts alone were insufficient for the total recovery of peptides from the column. This is expected when considering the affinity equilibria, where multiple cations are required to simultaneously displace multiple positive point charges along peptides. This is supported by work in Chapter 2 where PEHA which has a charge state of +6 at a pH of 2.0 recovered ~35% percent of the overall *E. coli* sample. This is further supported through work using the S-CLC where the charge state of the peptides identified by MS using a salt gradient at pH 2.1 steadily increased proportionally with salt concentration. This demonstrates that increasing salt is a useful secondary means of recovering peptides from the column through assisting with the displacement of peptides which have a weak affinity to the column. Combining salt and pH in a gradient was therefore found to be the most efficient approach to the separation of peptide.

All the gradient methods were tested against a modified microMUDPIT gradient. The SPG method was more effective than using a single variable alone. Furthermore, the SPG method, which is admittedly still unoptimised, has the largest recovery and peptide identification rate out of all the methods. The method developed by Webb *et al.*[82] has a

limited pH range (pH 5 – pH 6.5) in comparison to the wide pH range utilised in SPG (pH 3-11). The limited range in pH used in the MuDPIT method prevents good quality separation of peptides.

This information about the factors controlling separation in SCX can be used to develop an optimal approach to proteome separation. It is recommended that the user employ a pH gradient in the absence of salt, and then perform a salt gradient at a fixed pH under moderately acidic conditions (pH 4-5) to examine the ionic strength of the cation being used. This allows the researcher to see the impact that salt has when added to the pH gradient and provides information to the user on the best way to perform separations in their systems.

4.2 Future Work

Although there has been a lot of work performed to understand SCX, there remains a lot to do with the development of both systems to produce fully optimised separations. In the case of the ProTrap XG, the challenges with column reproducibility which arose from using prototype-stage columns constructed by hand, should be alleviated before any further attempts at creating or analysing 'optimised' separations. The large variability between replicates in terms of flow rate and run time should be solved through ensuring that a precise amount of resin is added to each ProTrap XG column, and that the resin is packed evenly.

In the case of the S-CLC, there is a need to find a method of recovering all the peptide from the column during regeneration, thus reducing column fouling and ensuring reproducibility from run to run. This is not a concern with the ProTrap XG as the

columns are designed to be disposable. One way to potentially regenerate the S-CLC column is to switch the drive pump to a PEHA gradient between loading new samples. Work using the ProTrap XG found that under acidic conditions, 5 mM PEHA was successful for the removal of peptides from the SCX column and did not damage the reverse phase resin during sample clean-up. The high ionic strength and large molecular weight of this cationic buffer have a strong affinity for the column. Flushing through a third buffer under basic conditions should, in theory, remove the PEHA from the SCX allowing for further separation of peptides. However, there are concerns that such a strong base may not be easily removed from the column.

Given that the work with both systems was hindered in terms of analysis by the current mass spectrometer, work using a higher resolution instrument should be considered essential to the success of any future analyses and publications which may result from them.

It would be beneficial to explore the development of a theoretical model that disseminates the distribution of peptides produced in a hypothetical digestion according to their calculated isoelectric points, aiding in dictating an appropriate pH gradient. This is possible through the use of the IPC calculator and the MS digest programme provided by the proteome prospector a programme hosted by the UCSF web page to both digest and calculate the pI of the samples.[100,103] Using the theoretical peptide sequences, the excel charge calculator could be modified into a programme to allow for a faster analysis of the charge states of the individual peptides at any desired pH. Once the theoretical analysis is completed, experiments using the suggested approach should be attempted in the lab followed by an analysis of the peptides, which are obtained in fractions. These

could then undergo a clustered analysis of the charges on the peptides to properly determine how salt interacts with peptides on the column; However, this would require multiple separation replicates and the use of a high-resolution mass spectrometer to increase the overall number of peptide identifications to ensure accuracy from run to run.

References

(1) Watson J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature.* **1953**, *171* (4356), 737–738.

(2) Wilkins M. H.F.; Stokes, A. R.; Wilson, H.R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature.* **1953**, *171* (4356), 738–740.

(3) Franklin R. E.; Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature.* **1953**, *171* (4356), 740–741.

(4) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5463–5467.

(5) A Brief History of the Human Genome Project - National Human Genome Research Institute (NHGRI) https://www.genome.gov/12011239/a-brief-history-of-the-human-genome-project/ (accessed May 14, 2018).

(6) What were the goals of the Human Genome Project? - *Genetics Home Reference* https://ghr.nlm.nih.gov/primer/hgp/goals (accessed May 14, 2018).

(7) The ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature.* **2012**, *489* (7414), 57–74.

(8) Cooper, D. N.; Krawczak, M.; Polychronakos, C.; Tyler-Smith, C.; Kehrer-Sawatzki, H. Where Genotype Is Not Predictive of Phenotype: Towards an Understanding of the Molecular Basis of Reduced Penetrance in Human Inherited Disease. *Hum. Genet.* **2013**, 132 (10), 1077–1130.

(9) Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I. Progress with Gene-Product Mapping of the Mollicutes: Mycoplasma Genitalium. *Electrophoresis.* **1995**, *16* (1), 1090–1094.

(10) Wilkins, M. R.; Sanchez, J.-C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Williams, K. L. Progress with Proteome Projects: Why All Proteins Expressed by a Genome Should Be Identified and How to Do It. *Biotechnol.* Genet. *Eng. Rev.* **1996**, *13* (1), 19–50.

(11) Dwivedi, R. C.; Spicer, V.; Harder, M.; Antonovici, M.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Krokhin, O. V. Practical Implementation of 2D HPLC Scheme with Accurate Peptide Retention Prediction in Both Dimensions for High-Throughput Bottom-Up Proteomics. *Anal. Chem.* **2008**, *80* (18), 7036–7042.

(12) Patthy, L. *Protein Evolution,* 2nd ed.; Blackwell Publishing LTD: Singapore, 2008.

(13) Cirkovic Velickovic, T.; Ognjenovic, J.; Mihajlovic, L. Separation of Amino Acids, Peptides, and Proteins by Ion Exchange Chromatography. *In Ion-exchange technology II*; Inamuddin, Luqman, M., Eds.; Springer, Dordrect: London, **2012**, 1–34.

(14) Sanger, F.; Tuppy, H. The Amino-Acid Sequence in the Phenylalanyl Chain of Insulin. I. The Identification of Lower Peptides from Partial Hydrolysates. *Biochem. J.* **1951**, *49* (4), 463–481.

(15) Sanger, F.; Thompson, E. O.; Kitai, R. The Amide Groups of Insulin. *Biochem. J.* **1955**, *59* (3), 509–518.

(16) Sanger, F.; Smith, L. F.; Kitai, R. The Disulphide Bridges of Insulin. *Biochem. J.* **1954**, *58* (330th Meeting), vi–vii.

(17) Sanger, F.; Tuppy, H. The Amino-Acid Sequence in the Phenylalanyl Chain of Insulin. 2. The Investigation of Peptides from Enzymic Hydrolysates. *Biochem. J.* **1951**, *49* (4), 481–490.
(18) Sanger, F. Chemistry of Insulin; Determination of the Structure of Insulin Opens the Way to Greater Understanding of Life Processes. *Science* **1959**, *129* (3359), 1340–1344.
(19) Pehr, Edman. Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chem. Scand.* **1950,** *4*, 283–293.
(20) Westermeier, R.; Naven, T. *Proteomics in Practice: A Laboratory Manual of Proteome Analysis;* Wiley-VCH: Weinheim, 2002.
(21) Westermeier, R.; Naven, T.; Höpker, H.-R. *Proteomics in Practice: A Guide to Successful Experimental Design;* Wiley-VCH, Weinheimn, 2008.
(22) Bakalarski, C. E.; Elias, J. E.; Villén, J.; Haas, W.; Gerber, S. A.; Everley, P. A.; Gygi, S. P. The Impact of Peptide Abundance and Dynamic Range on Stable-Isotope-Based Quantitative Proteomic Analyses. *J. Proteome Res*. **2008**, *7* (11), 4756–4765.
(23) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994,** *5* (11), 976–989.
(24) Veenstra, T. D.; Yates, J. R. *Proteomics for Biological Discovery;* Wiley-Liss: Hoboken, N.J., 2006.
(25) Hamacher, M. *Proteomics in Drug Research; Methods and principles in medicinal chemistry;* v. 28; Wiley-VCH: Weinheim, 2006.
(26) Breuza, L.; Poux, S.; Estreicher, A.; Famiglietti, M. L.; Magrane, M.; Tognolli, M.; Bridge, A.; Baratin, D.; Redaschi, N.; UniProt Consortium, T. U. The UniProtKB Guide to the Human Proteome. *Database* (Oxford) **2016**, *2016*.
(27) Boutet, E.; Lieberherr, D.; Tognolli, M.; Schneider, M.; Bairoch, A. *UniProtKB/Swiss-Prot. In Plant Bioinformatics;* Humana Press: Totowa, NJ, 2007: pp 89–112.
(28) Ponomarenko, E. A.; Poverennaya, E. V; Ilgisonis, E. V; Pyatnitskiy, M. A.; Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V; Archakov, A. I. The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016**, *2016***,** ID 7436849.
(29) Westermeier, R.; Naven, T.; Hpker, H.-R. *Proteomics in Practice;* Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2008.
(30) Gross, J. *Mass Spectrometry*, 3rd ed.; Springer-Verlag: Heidlberg, 2017.
(31) Simpson, R. J. *Purifying Proteins for Proteomics: A Laboratory Manual;* C. S. H. Laboratory Press, 2004.
(32) Twyman, R. M.; Twyman, R. M. *Principles of Proteomics; Advanced Text*; BIOS Scientific Publishers: New York, 2004.
(33) Crowell, A. M. J.; MacLellan, D. L.; Doucette, A. A. A Two-Stage Spin Cartridge for Integrated Protein Precipitation, Digestion and SDS Removal in a Comparative Bottom-up Proteomics Workflow. *J. Proteomics.* **2015**, *118*, 140–150.

(34) Mant, C. T., Hodges, R. S.; High-Performance Liquid Chromatography of Peptides and Proteins: Separation, Analysis and Conformation; *CRC Press*: Boca Raton, **1991.**
(35) Thomson, J. J. XL. Cathode Rays. London, Edinburgh, Dublin Philos. *Mag. J. Sci.* **1897**, *44* (269), 293–316.
(36) Thomson, J. J. XLVII. On Rays of Positive Electricity. London, Edinburgh, Dublin Philos. *Mag. J. Sci.* **1907,** *13* (77), 561–575.
(37) Münzenberg, G. Development of Mass Spectrometers from Thomson and Aston to Present. *Int. J. Mass Spectrom.* **2013,** *349–350* (1), 9–18.
(38) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, *246* (4926), 64–71.
(39) Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. Rapid '*de Novo*' Peptide Sequencing by a Combination of Nanoelectrospray, Isotopic Labeling and a Quadrupole/Time-of-Flight Mass Spectrometer. Rapid Commun. *Mass Spectrom.* **1997**, *11* (9), 1015–1024.
(40) Makarov. Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. *Anal. Chem.* **2000**, *72* (6), 1156–1162.
(41) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev*. **1998**, *17* (1), 1–35.
(42) Douglas, D. J.; Frank, A. J.; Mao, D. Linear Ion Traps in Mass Spectrometry. *Mass Spectrom. Rev*. **2005,** *24* (1), 1–29.
(43) Searcy, J. Q.; Fenn, J. B. Clustering of Water on Hydrated Protons in a Supersonic Free Jet Expansion. *J. Chem. Phys.* **1974**, *61* (12), 5282–5288.
(44) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459.
(45) Harris, D. C.; Lucy, C. A. *Quantitative Chemical Analysis;* W.H. Freeman: New York, 2016.
(46) Karas, M.; Bachmann, D.; Hillenkamp, F. Influence of the Wavelength in High-Irradiance Ultraviolet Laser Desorption Mass Spectrometry of Organic Molecules. *Anal. Chem.* **1985**, *57* (14), 2935–2939.
(47) Cole, R. B. *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation, and Applications;* Wiley: New York, 1997.
(48) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), p 2-9.
(49) Cech, N. B.; Enke, C. G. Practical Implications of Some Recent Studies in Electrospray Ionization Fundamentals. *Mass Spectrom. Rev.* **2001**, *20* (6), 362–387.
(50) Annesley, T. M. Ion Suppression in Mass Spectrometry. *Clin. Chem.* **2003**, *49* (7), 1041–1044.
(51) Arrua, R. D.; Talebi, M.; Causon, T. J.; Hilder, E. F. Review of Recent Advances in the Preparation of Organic Polymer Monoliths for Liquid Chromatography of Large Molecules. Anal. Chim. Acta **2012**, *738*, 1–12.

(52) Van Midwoud, P. M.; Rieux, L.; Bischoff, R.; Verpoorte, E.; Niederländer, H. A. G. Improvement of Recovery and Repeatability in Liquid Chromatography−Mass Spectrometry Analysis of Peptides. *J. Proteome Res.* **2007**, *6* (2), 781–791.
(53) Magdeldin, S.; Moresco, J. J.; Yamamoto, T.; Yates, J. R. Off-Line Multidimensional Liquid Chromatography and Auto Sampling Result in Sample Loss in LC/LC–MS/MS. *J. Proteome Res.* **2014**, *13* (8), 3826–3836.
(54) Pezeshki, A.; Vergote, V.; Van Dorpe, S.; Baert, B.; Burvenich, C.; Popkov, A.; De Spiegeleer, B. Adsorption of Peptides at the Sample Drying Step: Influence of Solvent Evaporation Technique, Vial Material and Solution Additive. *J. Pharm. Biomed. Anal.* **2009**, *49* (3), 607–612.
(55) Tabb, D. L.; McDonald, W. H.; Yates, J. R. DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics. *J. Proteome Res.* **2002**, *1* (1), 21–26.
(56) Fonslow, B. R.; Stein, B. D.; Webb, K. J.; Xu, T.; Choi, J.; Park, S. K.; Yates, J. R. Digestion and Depletion of Abundant Proteins Improves Proteomic Coverage. *Nat. Methods* **2013**, *10* (1), 54–56.
(57) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.
(58) Tiselius, A. Electrophoresis of Serum Globulin: Electrophoretic Analysis of Normal and Immune Sera. *Biochem. J.* **1937**, *31* (9), 1464–1477.
(59) Svedberg, T.; Tiselius, A. A New Method for Determination of the Mobility of Proteins. *J. Am. Chem. Soc.* **1926,** *48* (9), 2272–2278.
(60) Tiselius, A. Electrophoresis and Adsorption Analysis as Aids in Investigations of Large Molecular Weight Substances and Their Breakdown Products, *Nobel Lect.* **1948**.
(61) Svensson, H. Isoelectric Fractionation, Analysis, and Characterization of Ampholytes in Natural PH Gradients. I. The Differential Equation of Solute Concentrations at a Steady State and Its Solution for Simple Cases. *Acta Chemica Scandinavica.* **1961,** 325–341.
(62) Cleveland, D. W.; Fischer, S. G.; Kirschner, M. W.; Laemmli, U. K. Peptide Mapping by Limited Proteolysis in Sodium Dodecyl Sulfate and Analysis by Gel Electrophoresis. *J. Biol. Chem*. **1971,** *252* (3), 1102–1106.
(63) O'Farrell, P. H. High Resolution Two-Dimensional Electrophoresis of Proteins. *J. Biol. Chem.* **1975**, *250* (10), 4007–4021.
(64) Rickwood, D.; Hames, B. D. *Gel Electrophoresis of Proteins: A Practical Approach;* IRL Press: London, 1981.
(65) Allen, R. C.; Budowle, B. *Gel Electrophoresis of Proteins and Nucleic Acids: Selected Techniques;* W. de Gruyter: Berlin; New York, 1994.
(66) Garrett, R. H.; Grisham, C. M. *Biochemistry;* Brooks/Cole, Cengage Learning: Belmont, CA, 2010.
(67) Westermeier, R.; Fichmann, J. *Electrophoresis in Practice: A Guide to Methods and Applications of DNA and Protein Separations;* VCH, 1997.
(68) Poland, J.; Cahill, M. A.; Sinha, P. Isoelectric Focusing in Long Immobilized PH Gradient Gels to Improve Protein Separation in Proteomic Analysis. *Electrophoresis* **2003**, *24* (7–8), 1271–1275.

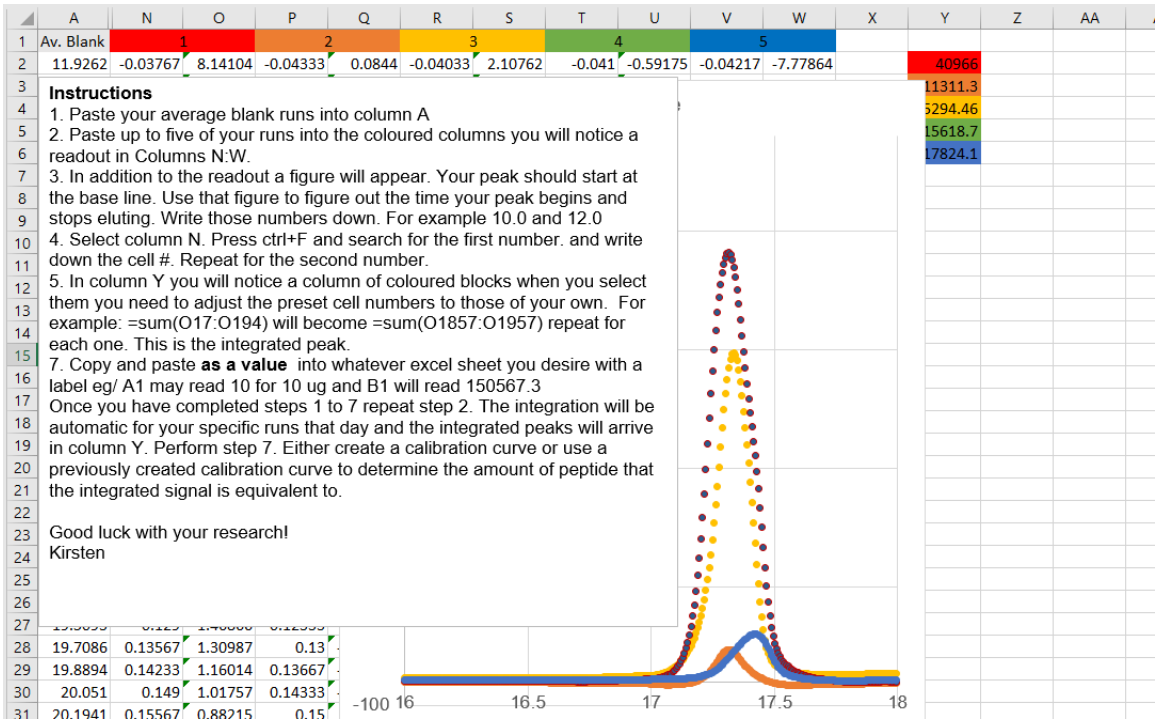
(69) Alexandru Dan Corlan. Medline trend: automated yearly statistics of PubMed results for any query http://dan.corlan.net/medline-trend.html (accessed Jul 10, 2018).
(70) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **1996**, *68* (5), 850–858.
(71) Celis, J. E.; Bravo, R. . *Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications;* Academic Press: New York, 1984.
(72) Smejkal, G. B.; Lazarev, A. Separation Methods in Proteomics; CRC Taylor & Francis: Boca Raton, FL, 2006.
(73) Tran. J. C.; Doucette, A. A. Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Anal. Chem.* 2008, *80* (5), 1568-1573.
(74) Chan, K. C.; Issaq, H. J. *Fractionation of Peptides by Strong Cation-Exchange Liquid Chromatography*; Humana Press, Totowa, NJ, 2013; pp 311–315.
(75) Dorfner, K. Ion Exchangers: Properties and Applications, 2nd ed.; Coers, A. F., Ed.; Ann Arbour Science Publishers Inc: Ann Arbor, 1972.
(76) De Dardel, F.; Arden, T. V. *Ion Exchangers. Ullmann's Encyclopedia of Industrial Chemistry;* Wiley-VCH, 2000; pp 264–322.
(77) Martin, M. On the Potential of Two- and Multi-Dimensional Separation Systems. Fresenius. *J. Anal. Chem.* **1995**, *352* (7–8), 625–632.
(78) Giddings, J. C. Two-Dimensional Separations: Concept and Promise. *Anal. Chem.* **1984**, *56* (12), 1258A–1270A.
(79) Daly, A. E.; Gilar, M.; Kele, M.; Neue, U. D.; Dourdeville, T.; Gebler, J. C. The Chromatographers Look on the Multi-Dimensional Chromatography of Complex Peptide Mixtures; Technical Report for Waters Corp: Milford 2004.
(80) Wolters, D. A.; Washburn, M. P.; Yates, J. R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* **2001**, *73* (23), 5683–5690.
(81) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999,** *17* (10), 994–999.
(82) Webb, K. J.; Xu, T.; Park, S. K.; Yates, J. R. Modified MuDPIT Separation Identified 4488 Proteins in a System-Wide Analysis of Quiescence in Yeast. *J. Proteome Res.* **2013**, *12* (5), 2177–2184.
(83) Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvic, B. M.; Yates, J. R. I. Direct Analysis of Protein Complexes Using Mass Spectrometry. *Nat. Biotechnol.* **1999**, *17* (7), 676–682.
(84) Mostovenko, E.; Hassan, C.; Rattke, J.; Deelder, A. M.; van Veelen, P. A.; Palmblad, M. Comparison of Peptide and Protein Fractionation Methods in Proteomics. *EuPA Open Proteomics* **2013**, *1*, 30–37.
(85) Schieltz, D. M.; Washburn, M. P. Analysis of Complex Protein Mixtures Using Multidimensional Protein Identification Technology (MuDPIT). *C.S.H. Protoc.* **2006**, *2006* (5), 1-6.
(86) Libereti, L.; Passino, R. Ion Exchange Kinetics in Selective Systems. In Ion Exchange and Solvent Extraction; Marinsky, J. A., Marcus, Y., Eds.; Marcel Dekker INC: New York, **1985**, 175–207.

(87) Lang, K. M. H.; Kittelmann, J.; Pilgram, F.; Osberghaus, A.; Hubbuch, J. Custom-Tailored Adsorbers: A Molecular Dynamics Study on Optimal Design of Ion Exchange Chromatography Material. *J. Chromatogr.* A **2015**, *1413*, 60–67.
(88) Gussakovsky, D.; Neustaeter, H.; Spicer, V.; Krokhin, O. V. Sequence-Specific Model for Peptide Retention Time Prediction in Strong Cation Exchange Chromatography. *Anal. Chem.* **2017**, *89(21)*, 11795-11802.
(89) Crimmins, D. L. Strong Cation-Exchange High Performance Liquid Chromatography as a Versatile Tool for the Characterization and Purification of Peptides. *Anal. Chem. Acta*. **1997,** *352*, 21–30.
(90) Lorne Burke, T. W.; Mant, C. T.; Black, J. A.; Hodges, R. S. Strong Cation-Exchange High-Performance Liquid Chromatography of Peptides: Effect of Non-Specific Hydrophobic Interactions and Linearization of Peptide Retention Behaviour. *J. Chromatogr. A*. **1989**, *476*, 377–389.
(91) Safronov, A. P.; Shklyar, T. F.; Borodin, V.; Smirnova, Y. A.; Sokolav, S. Y.; Pollack, G. H.; Blyakhman, F. A. Donnan Potential in Hydrogels of Poly(Methacrylic Acid) and it's Potassium Salt. In Water and the Cell; Pollack, G. H., Camaron, I. L., Wheatley, D. N., Eds.; Springer, Dordrect: Dordecht, **2006**, 273–284.
(92) Alpert, A. J.; Petritis, K.; Kangas, L.; Smith, R. D.; Mechtler, K.; Mitulović, G.; Mohammed, S.; Heck, A. J. R. Peptide Orientation Affects Selectivity in Ion-Exchange Chromatography. *Anal. Chem.* **2010**, *82* (12), 5253–5259.
(93) Alpert, A. J. Hydrophilic-Interaction Chromatography for the Separation of Peptides, Nucleic Acids and Other Polar Compounds. *J. Chromatogr. A* **1990**, *499*, 177–196.
(94) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. Comparison of Three Directly Coupled HPLC MS/MS Strategies for Identification of Proteins from Complex Mixtures: Single-Dimension LC-MS/MS, 2-Phase MudPIT, and 3-Phase MudPIT. *Int. J. Mass Spectrom.* **2002**, *219* (1), 245–251.
(95) Doucette, A. A.; Wall, M. J. Filtration and Extraction Assembly. U.S. Patent 9592500, **2013.**
(96) Crowell, A. *Methods to Isolate Proteins from Detergent-Containing Solutions for Proteome Analysis,* MSc Thesis, Dalhousie University, **2014**.
(97) Practice,M.; The Qiagen Guide to Good Microbiological Practice Part III — Growth of *E. coli Cultures.* **1999**, *1*, 17-19.
(98) Orton, D. J.; Wall, M. J.; Doucette, A. A. Dual LC-MS Platform for High-Throughput Proteome Analysis. *J. Proteome Res*. **2013**, *12* (12), 5963–5970.
(99) Jones, K. Ion Exchange Chromatography for Proteome Analysis Using Mass Spectrometry, Undergraduate Thesis, Dalhousie University, **2014**.
(100) Kozlowski, L. P. IPC – Isoelectric Point Calculator. *Biol. Direct* **2016**, *11* (1), 55.
(101) Oliveros, J. C. Venny. An interactive tool for comparing lists with Venn's diagrams http://bioinfogp.cnb.csic.es/tools/venny/index.html (accessed Jul 18, 2018).
(102) Elschenbroich, S.; Ignatchenko, V.; Sharma, P.; Schmitt-Ulms, G.; Gramolini, A. O.; Kislinger, T. Peptide Separations by On-Line MudPIT Compared to Isoelectric Focusing in an Off-Gel Format: Application to a Membrane-Enriched Fraction

from C2C12 Mouse Skeletal Muscle Cells, *J. Proteome Res.* **2009**, *8* (10), 4860–4869.
(103) MS-Digest http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msdigest (accessed Jul 18, 2018).

Appendix A: MS Excel File: Peak Area Calculator

This is a screen capture of the excel file used to determine the peak area of LC Data. This will work for S-CLC data or standard peptide clean -ups.
Instructions on how this calculator are found in the file so that anyone might use it in the future.



Appendix B: MS Excel File: Charge Calculator.

This is a screen capture of the excel file.
Instructions on how this calculator are found in the file so that anyone might use it in the future.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1																						pH:		pH=pKa+lo
2	PEPTIDE	A	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	V	W	Y	Total Charge	pKa
3	EPVSDWTDDVEAR	1	0	3	2	0	0	0	0	0	0	0	0	0	1	0	1	1	1	2	1	0	1.9990211	Protonate
4	NAPAAVLYEDGLK	3	0	1	1	0	1	0	0	1	2	0	1	0	1	0	0	0	0	1	0	1	0.999824	Deprotona
5	SPALVFEDADLDK	2	0	3	1	1	0	0	0	1	2	0	0	0	1	0	0	1	0	1	0	0	0.9995723	R
6	ELQDIANPIMSK	1	0	1	1	0	0	0	2	1	1	1	1	0	1	1	0	1	0	0	0	0	0.999824	alpha
7	AI																						0.9995979	percent pr
8	IH																						1.9997484	percent de
9	KI																						2.9997739	
10	TA																						0.9997482	
11	G																						1.9990176	
12	Q																						2.9997494	
13	H'																						1.999647	
14	H																						1.9996971	
15	RI																						1.9997482	
16	H																						1.9995969	
17	KI																						1.9997739	
18	KI																						2.999823	
19	YI																						2.9998497	
20	AI																						1.9997739	
21	LI																						1.9995478	
22	M																						1.9998497	
23	NI																						1.9999499	
24	VI																						0.9998741	
25	AI																						0.9995979	

How this works.
Using the pKa for the amino acid sequences and the pH the fractional abundance of protonated to deprotonated side chains at that specified pH can be calculated. This number is multiplied by the charge +1 and 0, or -1 and 0 depending on if the residue is basic or acidic.
If you want to see the math, it's hidden in columns Y to AG
The total number of each amino acid residue in your sequence are tabulated in Columns B through V.
If the amino acid has a charge, for example if Histadine (H) appears three times in your peptide chain, the charge fractional abundance is multiplied by 3. This is done for every amino acid and is then summed.

Instructions.
1. Input pH of the buffer that the sample was fractionated in (Yellow box)
2. Paste the peptide sequences into A3 and ensure that the calculator output extends low enough to process all of your ID'd peptides.
3. Copy your selected text and paste.

Good Luck with your research,
-Kirsten

Appendix C: List of Equations