Strategies for Differential Proteomic Analysis by Liquid Chromatography-Mass Spectrometry

Jordan Thomas Stobaugh

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill
2012

Approved by:

James W. Jorgenson

Gary L. Glish

R. Mark Wightman

Dhiren R. Thakker

Gary J. Pielak

ABSTRACT

JORDAN THOMAS STOBAUGH: Strategies for Differential Proteomic Analysis by Liquid Chromatography-Mass Spectrometry
(Under the direction of James W. Jorgenson)

As the field of proteomics has evolved, the analytical techniques used in this field of research has also evolved, transitioning from primarily gel-based techniques to separations based on liquid chromatography. Enzymatic digestion is often a first step for comprehensive analysis, known as bottom-up proteomics, but this comes at a cost of greatly increasing the sample complexity. Analysis of intact proteins, or top-down proteomics, is not without difficulties either as both the chromatography and mass spectrometry are more challenging. Various single and multidimensional techniques have been used to increase the peak capacity for bottom-up approaches, as, for the time being, they appear better suited for comprehensive proteomic analyses.

The focus of Chapter 2 is determining what is possible given commercial instrumentation and common analytical methods for a bottom-up analysis. A single- and multidimensional separation were compared in the analysis of differentially grown yeast samples. These same samples were used in Chapter 3 in a pre-fractionation approach in which the intact protein samples were fractionated, then digested. The digested fractions were then analyzed by LC/MS to determine protein identity and relative abundance. The work presented in Chapter 5 relies on the methods developed in Chapter 3 in the differential proteomic analysis of a beta-arrestin 1,2 double-knockout.

Lastly, in an effort to increase the number of protein identifications without substantially increasing analysis time, new instrumentation was designed to facilitate chromatography at pressures not attainable with commercial instrumentation. This was the focus of Chapter 4 where a nanoAcquity UPLC was modified for XUPLC operation, allowing for an increase in backpressure from 10,000 psi to 40,000 psi. With the increase in back pressure, columns with more efficiency either through smaller particle diameters or increased length could be used in the analysis of peptide samples.

ACKNOWLEDGEMENTS

None of the work presented in this dissertation would have been possible without contributions on many levels from many individuals. It would only seem appropriate to begin with Dr. Jorgenson. Through his approach to research and guidance of students, I was able to grow tremendously as a researcher with significant latitude to try new things, but lean on his expertise for advice when needed. I cannot thank him enough for the opportunities provided to me by studying under his direction. Brenna Richardson taught me many of the techniques as a new graduate student that were needed for this research, and for that I am grateful. I also must thank Kaitlin Fague, who worked alongside me for much of the work and will be carrying the projects forward. Keith Fadgen and Martin Gilar from Waters Corporation deserve many thanks as well. They each provided much support by aiding in troubleshooting and for supplying the necessary parts to keep the systems operational. No data would have been collected without their continued support. Keith has also been instrumental in providing critiques for the XUPLC system. With regard to the motivation for the XUPLC system, Ed Franklin was instrumental in showing what could be accomplished through use of long capillary columns.

There are many people who helped me on a personal level during my time as a graduate student. For helping me stay sane by providing me an outlet I need to thank the students and mentors of FIRST Robotics Team 900, whom I mentored for three years. In particular, I would like to thank mentors Marshall Massengill and Austin Page for their help

in mentoring and the fun we had while doing so. Nicholas Dobes also deserves a thank you for the amusing opportunities he provided for me with car repair and other such tasks involving my garage. Laura Blue also provided much support and stress relief through conversations at our desks when everything seemed to be going wrong. Lastly, I must thank my family for their unwavering support – my mother for helping to keep things in perspective and my father for his guidance and valid opinions on my research. Megan, my wife, probably deserves the most gratitude of all. Thank you for everything. Thank you for moving far from home, leaving family, friends, and a career behind to allow me to pursue my dreams. Thank you for Abigail, our daughter, and the joy she has brought us. None of this would have been possible without you. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2D-UPLC	two-dimensional ultra performance liquid chromatography
2DE	two-dimensional gel electrophoresis
AUC	area under the curve
AXC	anion-exchange chromatography
AXC20	anion-exchange chromatography with 20 fractions
AXC40	anion-exchange chromatography with 40 fractions
BEH	bridged-ethyl hybrid
BPI	base peak index
CID	collision-induced dissociation
ECD	electron-capture dissociation
ESI	electro-spray ionization
ETD	electron-transfer dissociation
FT-ICR	Fourier transform ion cyclotron resonance
FT-MS	Fourier transform mass spectrometry
GELFrEE	gel-eluted liquid fraction entrapment electrophoresis
IEC	ion-exchange chromatography
IMS	ion mobility separations
kDa	kilo-dalton
LC	liquid chromatography
LC/MS	liquid chromatography mass spectrometry
LC/MS/MS	liquid chromatography tandem mass spectrometry
LOD	limit of detection

MALDI	matrix-assisted laser desorption ionization
mAU	milli-absorbance units
MDLC	multidimensional liquid chromatography
MS/MS	tandem mass spectrometry
MS^E	Waters Corporation data-independent acquisition mode
MudPIT	multidimensional peptide identification technology
MW	molecular weight
nL	nanoliter
NPLC	normal-phase liquid chromatography
PTM	post-translational modification
QTOF	quadrupole time-of-flight
RP	reversed-phase
RP18	reversed-phase chromatography with 18 fractions
RP20	reversed-phase chromatography with 20 fractions
RPLC	reversed-phase liquid chromatography
SCX	strong-cation exchange chromatography
SEC	size-exclusion chromatography
SILAC	stable isotope labeling by amino acids in cell culture
TFA	trifluoroacetic acid
TIC	total ion current
UHPLC	ultra-high pressure liquid chromatography
UPLC	ultra performance liquid chromatography
WT	wild-type

XUPLC	extreme ultra performance liquid chromatography
µL	microliter
µm	micrometer
dKO	double-knockout
UV	ultraviolet
PLGS	ProteinLynx Global Server

CHAPTER 1: INTRODUCTION AND BACKGROUND FOR PROTEOMICS BY MULTIDIMENSIONAL LIQUID CHROMATOGRAPHY MASS SPECTROMETRY

1.1 Differential Proteomics

1.1.1 Definition and Challenges

Qualitative information was typical of proteomic analyses in early proteomics work with much effort being focused solely on the total number of identifications from a given sample. Significant improvements in mass spectrometry with much improved data analysis tools have contributed greatly to towards the expansion of quantitative approaches and the ability to use proteomics for biomarker discovery. The focus of differential proteomics is to discern differences in protein expression between samples in an effort to gain additional information into the processes that are occurring that would be otherwise difficult to ascertain by examining a single sample. The challenges associated with differential proteomics are similar to those of any proteomic analysis. The number of proteins possible is significant. For example, there are over 5,500 proteins encoded by the yeast genome.[1] For mammalian cells there are more.[2] Expression levels can range from thousands of copies per cell down to a few copies with the most interesting proteins often found at the lower end of this spectrum.[3,4] Cytosolic proteins are typically the easiest proteins to identify based on their solubility. Unfortunately, many proteins associated with cell signaling are found to be associated with the membrane. Preventing these protein from precipitating and having a sample amenable to analysis is a significant challenge.

1.1.2 Approaches: Top-Down vs. Bottom-Up

Analysis of intact proteins, top-down proteomics, is a challenging problem both in terms of the chromatography and mass spectrometry and is therefore not as common as methods involving the analysis of peptides. Identification of proteins through mass alone is not a reliable approach for making the identification due to post-translational modifications that can alter the molecular weight of the target protein. Therefore, MS/MS of intact proteins is required of top-down proteomics. While collision-induced dissociation (CID) is a common method of fragmentation, for top-down methods the preferred methods are electron-capture dissociation (ECD) or electron-transfer dissociation (ETD).[5,6,7] Both of these methods provide for a more complete fragmentation of the peptide backbone and tend to retain labile post-translational modifications (PTMs). Fourier transform-ion cyclotron resonance (FT-ICR) mass spectrometers are commonly used for top-down experiments as are a new class of mass spectrometers, Orbitraps.[8] With resolution in excess of 100,000, both of these instruments have the requisite resolution needed to resolve the isotopic charge envelopes of co-eluting proteins.[5,9,10] Advantages of top-down proteomics include the ability to discern isoforms, better characterize PTMs, and better sequence coverage.[11] Quantitation of the intact protein has also been reported as an advantage of top-down experiments as the protein abundance is measured directly and not inferred through peptide intensity.[12,13] Despite these advantages, there are a number of limitations of this technique that have restricted its widespread adoption. The chromatographic challenge of separating intact proteins is significant and conditions used for effective protein chromatography, such as the inclusion of trifluoroacetic acid (TFA) as an ion-paring agent, hinder electrospray ionization (ESI). Peptides, on the other hand, are readily separated without TFA and the use of formic acid to

improve ESI does not significantly hinder their separation. One of the key disadvantages of top-down proteomics is the need for high-resolution instrumentation and the associated expense of such mass spectrometers. Larger amounts of protein are also frequently needed and the samples analyzed trend towards simpler mixtures or even purified proteins.[8] Some recent work has analyzed more complex mixtures by top-down methodology, but fewer identifications are possible with current methodologies via top-down proteomics when compared to bottom-up approaches.[13,14]

As a result of obtaining a greater number of identifications, in addition to the lower instrumentation cost, and general experimental simplicity, bottom-up approaches are substantially more prevalent for proteomic analyses.[8] In a true bottom-up analysis, the intact proteins are not subject to an initial separation, but are digested, often by trypsin, and the resultant peptide mixture analyzed by LC/MS/MS.[8] While sample complexity is greatly increased due to the digestion, the chromatography of peptides is less problematic than intact proteins. Gradients of water and acetonitrile with formic acid as a modifier are typical of peptide separations and can be done on a C18 stationary phase without much in the way of special considerations. Though the chromatography is less challenging when attempting to resolve peptides with respect to proteins, because there are many more components in the mixture, the peak capacity of even the most rigorous separation methods are overwhelmed. That said, mass spectrometers offer another dimension of separation and increase the effective peak capacity of the analysis.[15] This approach holds an advantage over top-down methods with regard to proteome coverage, which is an important factor when analyzing complex samples.[16] Additionally, there is often an increase in sensitivity for bottom-up methods over top-down approaches. Much of this advantage is due to the improvement in

bioinformatics with significant work being done to improve probabilistic scoring schemes as well as incorporation of additional criteria during the searching.[17-19] Some of the work has been to create libraries of proteomics data to aid in the searching and these efforts are onoing.[20] Bottom-up approaches, however, suffer from a limited dynamic range and are not capable of quantifying across the expression range of proteins in cells.

There are additional limitations to what can be accomplished through bottom-up strategies. Ultimately, peak capacity remains an issue – especially for purely "shotgun" samples, where no prior separation has occurred to simplify the sample. While the mass spectrometer does provide another dimension of separation, even the most advanced mass spectrometers currently available can be overwhelmed by exceedingly complex samples. Further improvements to both mass spectrometry and liquid chromatography are needed to more deeply examine proteomic samples. Although bottom-up approaches clearly require greater peak capacity than presently available in order to better resolve sample components, there are additional challenges that require attention. Incomplete sequence coverage, the loss of labile PTMs, and an inability to determine the origin of certain peptides as a result of redundant sequences are all areas of potential improvement for bottom-up proteomic approaches.[8]

1.1.4 Quantitative proteomics

Much progress has been made with regard to quantitative proteomics in both relative and absolute terms. Various labeling schemes have been created to aid in the quantitation. Label-free methods are also common. Each approach has certain advantages and limitations that make them more suitable for particular proteomic challenges. A common labeling strategy that been cited frequently is stable-isotope labeling by amino acids in cell culture

(SILAC).[21,22] Through the incorporation of heavy and light isotopes in growth media, two proteomes can be distinguished from one another based on the mass of the identified peptides. Typically, the labeled amino acids used in this approach are lysine and arginine. For the most basic of SILAC analyses, one cell culture is grown with a light amino acid while the other is grown in heavy amino acid media. By mixing the cells in a 1:1 ratio, any proteins that are observed in a 1:1 ratio are known to be equally expressed. If the ratio is in favor of the light amino acid, meaning that a greater intensity of peptides was found from the light amino acid culture, then there was a greater expression of that particular protein for the light amino acid culture than the heavy. There are a few criteria that must be met in order to be sure of a differential identification. The first is that the incorporation of heavy or light amino acids must exceed 95%. Another criterion is that the peptides used in the identification, and therefore the determination of any associated fold-changes in expression, must have good abundance and a good signal-to-noise ratio. Ratios on the order of 1.3-2.0 have been used as the minimum cut-off for differential determination.[21]A key limitation of SILAC is the narrow range of samples that can be analyzed by this method. Typically, bacterial or yeast cell cultures are used with this method. A significant amount of effort has been placed though on creating cell lines with stable-isotope labeling. This has expanded the use of the technique, but working beyond controlled cell lines has proven challenging. One group has been able to incorporate this labeling method into mice. Initially, this was accomplished by feeding a lysine-free diet and providing lysine through alternative means.[22] Over a number of weeks, the isotope content for a range of cell types analyzed from the mice increased to approximately 80%. However, this level of labeling was not enough and extended feeding did not help due to recycling of amino acids. Only by breeding mice for several generations

and providing a lysine deficient diet for each generation was the incorporation of the label able to reach 93%. Other common labeling techniques involve the use of isobaric tags to modify peptides allowing for relative and absolute quantitation during tandem mass spectrometry.[23-25] Protein samples are digested and then tagged with different isobaric tags. The two digested protein samples are then combined in equal amounts. During the MS^1 scan, the tagged precursor ions will not be differentiable, however, the MS^2 scan will reveal the reporter ions from the tag in addition to the expected fragments. The intensity of these reporters can then be compared for quantitative purposes.[25] Overall, both chemical and metabolic labeling strategies provide for more accurate quantitation of proteins, but the expense of the isotopic labels and the limited-scope of available samples discourages significant applications.

Much work has been done to improve the quantitation of protein abundances based on the MS signal for label-free methods. The motivation for this type of work is clear. By not requiring a label, a much wider range of samples can be analyzed. Integrating peak areas for peptides, termed area under the curve (AUC), quantifies proteins based on the intensity of the precursor ion. One such method is data-independent acquisition, known as MS^E from Waters Corporation. All ions are passed to the collision cell and subjected to a high energy scan hopefully fragmenting all peptides allowing for sequencing of a greater number of peptides than what is possible through data-directed approaches. The MS/MS data is then assigned to the precursor ions based on retention time and intensity of the precursor. For quantitation, this method relies on the comparison of intensities of the most abundant peptides from a given protein to those from a standard of known quantity. Another method for quantifying label-free data is through spectral counting. Spectra from peptides assigned to a protein are

summed to determine abundance and normalized against a number of factors including the length of primary sequence. Studies have indicated that each of these methods produce similar findings with regard to protein identifications.[26] The data-independent acquisition, termed MS^E by Waters Corporation, did produce slightly higher sequence coverage and a greater number of peptides assigned to proteins on average. The greatest challenge associated with label-free techniques, regardless of the quantitation method, lies in the analysis of large data sets. Ultimately, as software and MS technology has improved, so has the confidence in using label-free techniques for quantitative purposes, and can be considered as a reliable alternative to labeling strategies. Detection limits are often reported in the 10fmol to 100pmol range with good linearity.[27] Already these strategies offer an improvement over classical gel approaches with regard to dynamic range, which will only continue to improve with advancement in software and instrumentation. This will allow for greater utility in the area of biomarker discovery as more is learned regarding entire protein networks involved in biological processes.

1.2. Multidimensional separations for differential proteomics

1.2.1 Background and theory

It has long been recognized that the peak capacity of single-dimension separations is insufficient to completely resolve the components of even modestly complex samples. This peak capacity problem is further compounded by the fact that for a random distribution of analytes, should the number of components exceed 37% of the peak capacity, significant overlapping of peaks will occur.[28] Proteomics samples often contain more than a thousand proteins or tens of thousands of peptides and even the best single-dimension separations fall well short in terms of peak capacity. A multidimensional separation has the ability to

dramatically increase the peak capacity and Giddings specified two key criteria for this type of separation.[29] The first is that, for a separation to be multidimensional, the modes of separation must be orthogonal. Many different modes of chromatography have been paired for multidimensional separations. Examples include ion-exchange followed by reversed phase,[30,31] size-exclusion followed by reversed-phase,[32,33] and normal-phase followed by reversed-phase,[34] in addition to others. Altering the sample is another way to achieve orthogonality by adjusting the pH of the mobile phase in multidimensional separations. In 2005, Gilar *et al* demonstrated the validity of this approach by multidimensional separation using reversed-phase chromatography in both dimensions for peptide analysis. In this work, a high-pH first dimension was followed by a low-pH second dimension.[35] Whichever the mechanism, the important point is that the two separation steps must be independent. Assuming orthogonal modes between the first and second dimension, the theoretical peak capacity of the entire system is simply the peak capacity of the first dimension multiplied by the peak capacity of the second dimension. In practice, the peak capacity often falls short of the theoretical maximum, but still provides for a significant improvement over most single-dimension approaches.[36]

The second criterion, according to Giddings,[28] regarding the operation of multidimensional, separations is that the peak capacity of the first dimension must be transferred to the second dimension in order for the peak capacities to be multiplicative. Sampling the first dimension at a rate of three times the peak capacity is frequently mentioned as an appropriate minimum sampling rate needed to preserve the resolution from the initial separation.[37,38] Often this proves to be impractical, as second dimension run times

would need to be exceedingly fast. For more infrequent sampling, the peak capacity of the first dimension of separation is reduced to the number of fractions collected.

The impracticality of sampling the first dimension as frequently as required is most notably associated with online methodologies, where the two modes are coupled, although there are a number of advantages associated with online approaches.[39] Automation is the most frequently cited as there is an inherent increase in sample throughput as there is less time required on the part of the investigator manipulating the sample and online analyses tend to require less time. Also, as there is complete transfer of sample from one dimension to the next, sample loss and contamination is significantly reduced. There are significant challenges to overcome; however, when implementing an online multidimensional separation. Custom instrumentation is often needed, which presents control issues, but more importantly are concerns related to the coupling of the LC methods. Mobile phase immiscibility and compatibilty is a concern as is the precipitation of salts from buffers. Previous work in the Jorgenson lab with online separations of intact proteins addressed these issues by utilizing volatile ammonium acetate buffers for anion-exchange chromatography in the first dimension and gradients of water and acetonitrile with formic acid for the second.[40] These were not the ideal mobile phase conditions for either dimension of separation. At a pH of 9.0, the anion-exchange mobile phase was poorly buffered and the pH would change over the course of the analysis. The inclusion of formic acid into the reversed-phase separation was needed to improve ionization for the intact mass determination by time-of-flight mass spectrometry, whereas TFA would have been preferred for reversed-phase protein separations.

Offline separations provide an alternative to the more frequently cited online multidimensional separations. As for the online approach, there are disadvantages.[39] These

include, a more labor- and time-intensive approach and an associated difficulty with automation as fractions must be collected from the first dimension. There is a greater potential for sample contamination so care must be taken to avoid such challenges and recovery can be problematic depending on the nature of the analyte. However, despite these potential pitfalls, there are significant advantages to an offline separation. Instrumentation is very simple. Often all that is needed is an LC pump and a fraction collector. The incompatibility of solvents can largely be mitigated as the two modes are no longer directly coupled. Most important of all is that the two methods can be independently optimized without concern for sampling rates allowing for more flexibility in method development.

1.2.2 Two-dimensional gel electrophoresis

Classically, 2D gel electrophoresis (2DE) was used for differential protein expression analysis.[41] 2DE works by first separating proteins by their isoelectric point utilizing isoelectric focusing and then by molecular weight with electrophoresis with both separations occurring on a polyacrylamide gel.[42] Spots are then excised and the proteins, which are digested in gel, are analyzed by traditional chemical sequencing techniques or more commonly by MALDI-MS. The resulting accurate mass data can then be compared to a database containing enzymatically digested protein sequence information in order to discern the identity of the protein, as is done with protein mass fingerprinting. An advantage for this method is that it is considered to be among the most rapid methods for identifying protein expression differences. However, there are numerous drawbacks rendering it a less than ideal method for comprehensive proteomic analyses. A significant drawback is the small dynamic range. As a result of the visual procedure needed for identifying differential proteins, only proteins of large abundance are observed. Proteins of large abundance are typically not the

most interesting for further study. Proteins involved in cell signaling, though, are of great interest for a wide variety of research topics and these proteins are often among the least expressed.[27] In order to analyze these less abundant proteins, further work to enrich the sample through fractionation is often needed, resulting in increased analysis time. Increased fractionation cannot overcome the inability for hydrophobic proteins, which are frequently membrane proteins, to enter the gel.[42] This again has implications for the ability to study cell-signaling pathways as many of the receptors are membrane-associated proteins. Other challenges associated with 2DE include difficulty in spot picking and poor resolution of very large, small, and very basic proteins. Additionally, interfacing 2DE with mass spectrometry is more difficult than methods utilizing liquid chromatography as it requires more sample handling and is more labor-intensive than strategies based on liquid chromatography. Despite these limitations, proteomics through 2DE has demonstrated promising results with the detection of proteins associated with tumors in liver carcinoma, lung adenocarcinoma, fibrosarcoma, lymphoma, breast cancer, and prostate cancer.[42] Clearly, proteomics, and the use of multidimensional separations as part of the workflow, has a profound implication on the future of biomarker discovery and should aid in the development of future therapies.

1.2.3 Gel-Eluted Liquid-Fraction Entrapment Electrophoresis

A technique that has become more prevalent recently is gel-eluted liquid-fraction entrapment electrophoresis (GELFrEE),[43] which uses tube gels to separate proteins. Fractions are collected from the outlet of the gels, permitting greater flexibility for downstream analysis. The primary advantage for this method of protein fractionation is that the proteins do not have to be excised from the gels. As a result of them eluting in solution, there are many more possible avenues for analysis available for these samples. After a detergent

removal step, intact proteins can be analyzed by FT-MS. The Kelleher group has successfully demonstrated this workflow in the analysis of HeLa cells.[44] Following the GELFrEE fractionation, nano-RPLC-MS was used to identify proteins in the low mass range, up to 12 kDa.

While comprehensive proteomics by intact protein analysis with mass spectrometry has been somewhat limited owing to challenges associated with ionization of large macromolecules, it is not an inherent limitation of the GELFrEE separation. Doucette has reported the separation of intact proteins from 10 kDa up to 150 kDa in just over three hours.[43] Additionally, they observed narrow bands and excellent reproducibility so long as the gels were not overloaded. While most of the implementation of this technique has been for top-down proteomics, the ability to separate the proteins into narrow, reproducible bands is attractive for use in a pre-fractionation strategy. By first separating a sample by MW, a single fraction is likely to contain most if not all of the peptides for a given protein. The reduction in sample complexity as a result of the fractionation affords bottom-up separations a better chance of identifying and quantifying more proteins.

1.2.4 Multidimensional peptide identification technology

To overcome the lack of peak capacity associated with bottom-up proteomics, numerous strategies involving multidimensional separations have been employed. One of the most notable is the multidimensional protein identification technology (MudPIT) approach pioneered by Yates for the analysis of peptides.[45,46] This method utilizes a biphasic capillary column with strong-cation exchange (SCX) as the first dimension and reversed-phase as the second dimension of separation directly coupled to a mass spectrometer. These columns are prepared by first packing reversed-phase particles into a length of fused-silica capillary

followed by packing of the SCX particles. Peptides are eluted step-wise from the SCX portion of the column and are subsequently trapped on the reversed-phase half of the column. Between each step-wise elution, a revered-phase gradient was employed to elute peptides from the column and were analyzed by an LCQ ion-trap mass spectrometer. In order to completely analyze their sample of *S. cerevisiae,* the sample was divided into three parts, a soluble fraction and two insoluble fractions. The results of this analysis were the most impressive to date and resulted in the identification of 1,484 proteins.[45] Additionally, Yates determined that the identified proteins spanned all subcellular portions of the cell and that they represented a wide variety of proteins in terms of the pI and MW range, abundance, and hydrophobicity. For bottom-up proteomics, this method has been successfully employed in a wide-range of proteomic challenges.

1.2.5 Pre-fractionation as alternative MDLC approach

An alternative approach to either top-down or bottom-up approaches has been reported in last few years, but has yet to garner much attention as of yet. The method separates intact proteins in the first dimension with fraction collection. These fractions are then digested before being subjected to an LC/MS/MS analysis for identity determination.[47,48] This same approach for proteomic analysis was independently arrived at through work in the Jorgenson lab and reached many of the same conclusions offered by Martosella *et al* and Dowell *et al*.

In many respects, this type of analysis is analogous to a tandem mass spectrometric analysis. In the mass spectrometer, sample is introduced, selected (by mass), fragmented by collision-induced dissociation, and the fragments detected and analyzed in order to elucidate the identity of the species in question. The liquid chromatographic pre-fractionation approach

is quite similar. Sample is injected onto the first column where it is selected based on the separation mechanism chosen for this dimension. The fractions containing separated proteins are then subjected to a reaction that creates many fragments through enzymatic digestion. These fragments are then analyzed in order to determine the identity of the unknown species, in this case, proteins. While the MS/MS analysis is clearly an online format and the proposed pre-fractionation approach is not, offline multidimensional separations have long been considered just that. Many of the earliest examples of multidimensional separations demonstrated offline approaches with the first comprehensive two-dimensional technique not reported until 1990 by Bushey and Jorgenson.[49] One could argue that the pre-fractionation approach is also capable of satisfying Giddings criteria for a multidimensional separation. The resulting 2D chromatograms from such analyses show peptides scattered widely across the available separation space, indicative of orthogonality. The sampling rate needed to preserve the peak capacity from the first dimension in order to properly transfer it to the second dimension is set by the fraction collection parameters and can be adjusted accordingly in order to satisfy the second criterion.

To a degree, the precedent for labeling a multidimensional separation for which a chemical reaction has modified the sample between chromatographic dimensions was already set by Gilar in 2005.[35] The previously discussed high/low pH multidimensional peptide separations protonate the sample between dimensions. Therefore, between similarities to MS/MS analysis and the high/low pH separations, one could safely say that a pre-fractionation of intact proteins with subsequent digestion and peptide analysis is indeed a multidimensional chromatographic technique. Nevertheless, as will be demonstrated in

Chapters 3 and 5, the approach serves to increase the ability for investigators to identify and quantify proteins, the ultimate goal of any proteomic approach.

1.3 Scope of Dissertation

In general, a greater emphasis is being placed on improving the introduction of the proteomic sample to the mass spectrometer. The scope of work presented here will compare current methodology to new methods with the overall goal of improving differential proteomics by refining multidimensional separation strategies for both intact proteins and peptides. In all cases, spectral counting was used to generate the relative quantitation data with the goal of the experimentation to improve the identification of differential proteins. Information regarding isoforms and many post-translation modifications will inevitably be lost; however, for comprehensive studies presented herein, the goal is to provide basic insight into biological processes that could potentially serve as a starting point for a more thorough, targeted analysis.

1.4 References

1. Kellis, M.; Patterson, N.; Endrizzi, M.; Birren, B.; Lander, E. S., Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 2003, *423* (6937), 241-254.

2. Southan, C. Has the yo-yo stopped? An assessment of human protein-coding gene number. *Proteomics* **2004**, 4, 1712-1726.

3. Ghaemmaghami, S.; Huh, W.; Bower, K.; Howson, R.; Belle, A.; Dephoure, N.; O'Shea, E.; Weissman, J., Global analysis of protein expression in yeaset. *Nature* 2003, *425* (6959), 737-741.

4. Wu, L.; Han, D.K. Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Review of Proteomics* **2006**, 3, 611-619.

5. Zubarev, R.A.; Kelleher, N.L.; McLafferty, F.W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J. Am. Chem. Soc.* 1998, *120*, 3265-3266.

6. Coon, J.; Ueberheide, B.; Syka, J.; Dryhurst, D.; Ausio, J. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl. Acad. Sci.* 2005, *101*, 9528-9533.

7. Bunger, M.; Cargile, B.; Ngunjiri, A.; Bundy, J.; Stephenson, J. Automated proteomics of *E. coli* via top-down electron-transfer dissociation mass spectrometry. *Anal. Chem*. 2008, *80*, 1459-1467.

8. Yates, J.; Ruse, C.; Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* 2009, *11*, 49-79.

9. Macek, B.; Waanders, L.; Olsen, J.; Mann, M. Top-down protein sequencing and MS3 on a hybrid linear quadrupole ion trap-orbitrap mass spectrometer. *Mol. Cell Proteomics.* 2006, *5*, 949-958.

10. McLafferty, F.; Horn, D.; Breuker, K., Ge, Y.; Lewis, M. Electronc capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. *J. Am. Soc. Mass Spectrom.* 2001, *12*, 245-249.

11. Hawkridge, A.; Heublein, D.; Bergen, H.; Cataliotti, A.; Burnet, J. Muddiman, D. Quantitative mass spectral evidence for the absence of circulating brain natriuretic peptide (BNP-32) in severe human heart failure. *Proc. Natl. Acad. Sci.* 2005, *102*, 17442-17447.

12. Waanders, L.; Hanke, S.; Mann, M. Top-down quantitation and characterization of SILAC-labeled proteins. *J. Am. Soc. Mass Spectrom.* 2007, *18*, 2058-2064.

13. Du, Y.; Parks, B.; Sohn, S; Kwast, K.; Kelleher, N. Top-down approaches for measuring expression ratios of intact yeast proteins using fourier transform mass spectrometry. *Anal. Chem.*, 2006, *78*, 686-694.

14. Sharma, S.; Simpson, D.C.; Tolic, N.; Jaitly, N.; Mayampurath, A. Proteomic profiling of intact proteins using WAX-RPLC 2-D separations and FTICR mass spectrometry. *J. Proteome Res.* 2007, *6*, 602-610.

15. Evans, C.R.; Jorgenson, J.W. Multidimensional LC-LC and LC-CE for high-resolution separations of biological molecules. *Anal. Bioanal. Chem.* 2004, *378,* 1952-1961.

16. Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science*. 2006, *312*, 212-217.

17. 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*, 21-26

18. Keller, A.; Eng, J.; Zhang, N.; Li, X.J.; Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* 2005, *1*, 2005-2017.

19. Moore, R.E.; Young, M.K., Lee, T.D. Method for screening peptide fragment ion mass spectra prior to database searching. *J. Am. Soc. Mass Spectrom.* 2000, *11*, 422-426.

20. Frewen, B.; MacCoss, M.J. Using BiblioSpec for creating and searching tandem MS peptide libraries. *Curr. Protoc. Bioinformatics*. 2007, *20*, 13.7.1-13.7.12.

21. Mann, M. Functional and quantitative proteomics using SILAC. *Nature.* 2006, *7*, 952-959.

22. Krüger, M.; Moser, M; Siegfried, U.; Thievessen, I.; Luber, C.A.; Forner, F.;Schmidt, S.; Zanivan, S.; Fässler, R.; Mann, M. SILAC Mouse for Quantitative Proteomics Uncovers Kindlin-3 as an Essential Factor for Red Blood Cell Function. *Cell*. 2008, 134, 353-364.

23. Köcher, T.; Pichler, P.; Schutzbier, M.; Stingl, C.; Kaul, A.; Teucher, N.; Hasenfuss, G.; Penninger, J. M.; Mechtler, K. High Precision Quantitative Proteomics Using iTRAQ on an LTQ Orbitrap: A New Mass Spectrometric Method Combining the Benefits of All. *Journal of Proteome Research.* 2009, *8* (10), 4743-4752.

24. Thompson, A.; Schäfer, J.; Kuhn, K. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, 75, 1895-1904.

25. Ross, P.L.; Huang, Y.N.; Marchese, J.N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D.J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, 3, 1154-1169.

26. Geromanos, S.J.; Vissers, J.P.; Silva, J.C.; Dorschel, C.A. The detection, correlation, and comparison of peptide precurosor and product ions from data independent LC-MS with data dependent LC-MS/MS. *Proteomics.* 2009, *9*, 1683-1695.

27. Neilson, K.A.; Ali, N.A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S.C.; Haynes, P.A. Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* 2011, *11*, 535-553.

28. Davis, J.M.; Giddings, J.C. Statistical theory of component overlap in multicomponent chromatograms. *Anal Chem.* 1983, *55*, 418-424

29. Giddings, J.C. Concepts and Comparisons in Multidimensional Separation. *J. High Resol. Chromatog.* 1987, *10*, 319-323.

30. Holland, L.A.; Jorgenson, J.W. Separation of Nanoliter Samples of Biological Amines by a Comprehensive Two Dimensional Microcolumn Liquid Chromatography System. *Anal. Chem.* **1995**, 67, 3275-3283.

31. Wagner, K.; Miliotis, T.; Marko-Varga, G.; Bischoff, R.; Unger, K.K. An Automated On-Line Multidimensional HPLC System for Protein and Peptide Mapping with Integrated Sample Preparation. *Anal. Chem.* **2002**, 74, 809-820.

32. Opiteck, G.J.; Jorgenson, J.W.; Anderegg, R.J. Two-Dimensional SEC/RPLC Coupled to Mass Spectrometry for the Analysis of Peptides. *Anal. Chem.* **1997**, 69, 2283-2291.

33. Opiteck, G.J.; Ramirez, S.M.; Jorgenson, J.W.; Moseley, M.A., III. Comprehensive Two-Dimensional High-Performance Liquid Chromatography for the Isolation of Overexpressed Proteins and Proteome Mapping. *Anal. Biochem.* **1998**, 258, 349-361.

34. Murphy, R.E.; Schure, M.R.; Foley, J.P. One- and Two-Dimensional Chromatographic Analysis of Alcohol Ethoxylates. *Anal. Chem.* **1998**, 70, 4353-4360.

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

36. Malerod, H.; Lundanes, E.; Greibrokk, T. Recent advances in on-line multidimensional liquid chromatography. *Anal. Methods*. **2009**, *2,* 110-122.

37. Davis, J.M.; Giddings, J.C. Statistical method for estimation of number of components from single complex chromatograms: theory, computer-based testing, and analysis of errors. *Anal. Chem.* **1985**, 57, 2168-2177.

38. Davis, J.M.; Giddings, J.C. Statistical method for estimation of number of components from single complex chromatograms: application to experimental chromatograms. *Anal. Chem.* **1985**, 57, 2178-2182.

39. Dugo, P.; Kumm, T.; Cacciola, F.; Dugo, G.; Mondello, L. Multidimensional Liquid Chromatographic Separations Applied to the Analysis of Food Samples. *Journal of Liquid Chromatography & Related Technologies*. 2008, *31,* 1758-1807.

40. Evans, C. R. *Multidimensional Liquid Chromatography Coupled to Mass Spectrometry for the Analysis of Complex Mixtures of Proteins*. University of North Carolina at Chapel Hill, Chapel Hill, 2007.

41. O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, 250, 4007-4021.

42. Monteoliva, L.; Albar, J.P. Differential proteomics: An overview of gel and non-gel based approaches. *Briefings in Functional Genomics & Proteomics*. 2004, *3*, 220-239.

43. Tran, J.C.; Doucette, A.A.; Multiplexed Size Separation of Intact Proteins in Solution Phase for Mass Spectrometry. *Anal. Chem.* 2009, *81*, 6201-6209.

44. Lee, J.E.; Kellie, J.F.; Tran, J.C.; Tipton, J.D.; Catherman, A.D.; Thomas, H.M.; Ahlf, D.R.; Durbin, K.R.; Vellaichamy, A.; Ntai, I.; Marshall, A.G.; Kelleher, N.L. A Robust Two-Dimensional Separation for Top-Down Tandem Mass Spectrometry of the Low-Mass Proteome. *J. Am. Soc. Mass Spectrom.* **2009**, 20, 2183-2191.

45. Wolters, D.A.; Washburn, M.P.; Yates, J.R. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Anal. Chem.* 2001, *73*, 5683-5690.

46. Washburn, M.P.; Ulaszek, R.; Deciu, C.; Schieltz, D.M.; Yates, J.R. Analysis of Quantitative Proteomic Data Generated via Multidimensional Protein Identification Technology. *Anal. Chem.* 2002, *74*, 1650-1657.

47. Martosella, J.; Zolotarjova, N.; Liu, H.; Nicol, G.; Boyes, B.E., Reversed-Phase High-Performance Liquid Chromatographic Prefractionation of Immunodepleted Human Serum Proteins to Enhance Mass Spectrometry Identification of Lower-Abundant Proteins. *J. Prot. Res.* **2005**, 4, 1522-1537.

48. Dowell, J.A.; Frost, D.C.; Zhang, J.; Li, L. Comparison of Two-Dimensional Fractionation Techniques for Shotgun Proteomics. *Anal. Chem.* **2008**, 80, 6715-6723.

49. Bushey, M.M.; Jorgenson, J.W. Automated instrumentation for comprehensive two-dimensional high-performance liquid chromatography of protein. *Anal. Chem.* **1990**, 62, 161-167.

CHAPTER 2: BOTTOM-UP ANALYSIS OF *S. CEREVISIAE* BY UPLC & 2D-UPLC

2.1 Introduction

2.1.1 Bottom-up Proteomics

The proliferation of bottom-up proteomics, often termed "shotgun" proteomics, is due to the simplicity of the experiment, facilitating improvements in throughput over previous methods.[1] The simplicity lies in the relatively few number of steps required to prepare the sample and perform the analysis. Intact proteins are enzymatically digested, often by trypsin, and then analyzed by LC/MS. Alternative approaches involve the use of either 1D or 2D gel electrophoresis. With either of these electrophoretic methods, there are significant number drawbacks, which were discussed in greater detail in Chapter 1. The two largest reasons for moving away from gel electrophoresis is the increased time required on the part of the investigator to prepare the sample as well as the associated increase in contamination due to the additional steps required. The improved throughput of bottom-up proteomics, allowing a greater number of samples to be analyzed in a day, owes itself to the straightforward sample preparation procedure as well as the ability to perform separations with peak capacities in the low hundreds within a reasonable amount of time.

Peak capacities are always insufficient to resolve the components from even modest protein digests. Typical peak capacities for the methods allowing the greatest throughput, encompassing no more than two hours for an LC/MS analysis, are approximately 300.[2] For whole-cell lysates, tens of thousands of peptides are present. Many investigators utilizing

bottom-up techniques will frequently employ other measures to reduce sample complexity. Analysis of a sub-proteome by only digesting proteins from certain organelles is a common method for reducing sample complexity.[3] Other approaches focus on the removal of the most abundant proteins. In human serum, for example, a handful of proteins are responsible for more than 99% of the protein present.[1] There is little hope of identifying and quantifying biomarkers when there is that much protein interference present and depletion strategies are particularly well-suited for this type of sample. Dynamic range for bottom-up techniques has been reported to be approximately 10^3.[4-6] In complex samples, such as human plasma samples, there is an expected range in protein concentration spanning seven orders of magnitude, demonstrating a need for further improvements to dynamic range.[7] In order to increase peak capacity and ideally improve the dynamic range of bottom-up methods, multidimensional separations are frequently employed. A number of reviews, in fact, state that they are necessary and pay little attention to single-dimension separations of average performance.[1,3,8,9] While multidimensional separations afford an increase in peak capacity, it comes at the cost of analysis time. The previously discussed MuDPIT approach is a favorite among proteomics investigators, but does require many hours of analysis.

2.1.2 2D nanoAcquity by High/Low pH

While the MudPIT technique pioneered by Yates more than 10 years ago has received much attention, other multidimensional platforms exist to improve upon the peak capacity of single-dimension LC analyses. One method, commercialized by Waters Corporation, utilizes a reversed-phase separation for both the first and second dimension. Orthogonality is achieved by changing the pH significantly enough between the two dimensions to allow for a change in selectivity of the peptides.[10] Figure 2-1 shows a diagram of this system for the flow

paths required during the course of the first and second dimension separations. The first dimension is the high pH dimension. In this setup, sample is injected onto the first dimension column at a low percentage of organic modifier, typically acetonitrile. Peptides are then eluted in stepwise fashion from the first dimension at specific concentrations of acetonitrile. The number of fractions is determined by the extent of fractionation desired by the investigator. Peptides eluting from the first dimension column must then be diluted with the low pH mobile phase with a dilution factor of the aqueous fraction substantial enough to cause the peptides to be retained on the trap column. Once an elution step is complete, the valve configuration is changed and a gradient of water and acetonitrile at a pH less than 3 is run to elute the peptides from the analytical column.

The advantage of using such a system is the increase in the peak capacity enabling better detection and analysis of the peptides with the convenience of a commercialized platform and associated support that accompanies commercial instrumentation. There are, however, a number of disadvantages. First, a second UPLC pump must be purchased and with additional pumps comes additional maintenance requirements. A second disadvantage of this platform is that great care must be taken during method development to properly set up the methods. In order to properly determine the elution steps for the first dimension of separation, many fractions must be taken in order to determine how many peptides, determined via the total ion current, are eluting. This data is then used in conjunction with software to determine the "cuts." Ideally, the second-dimension gradient should be optimized for each fraction to achieve the greatest use of the available separation space. Lastly, a multidimensional separation requires significantly more analysis time and decreases the number of samples that can be analyzed in a given period of time.

2.1.3 Ion mobility in proteomic applications

While one way to increase peak capacity of the overall analysis is to incorporate a second dimension of chromatographic analysis, another is to increase the peak capacity prior to the mass spectrometer. Ion mobility separations (IMS) are gas phase separations that result from the difference in drift times due to the difference in velocities of the ions that result from their collision with a buffer gas as they are pulled through a drift cell by an applied electric field.[11] The difference in velocities is due to the variation in the cross-sectional area of the molecules. Separation based on shape is somewhat orthogonal to the separation based on mass, resulting in an additional separation prior to mass spectrometric analysis. By combining IMS-MS with liquid chromatography, three or even four-dimensional separations are possible enabling huge peak capacities with peak capacities of 2×10^5 being reported for an HPLC-IMS-MS technique.[12] With UHPLC instead of HPLC, the peak capacities would improve an additional 5-fold owing to the improve resolution of UHPLC over HPLC.[2] The coupling of this technique with chromatography is efficient as even narrow peaks from UHPLC separations are slow with respect to IMS separations. Peaks in UHPLC are often 10 seconds wide and frequently wider. IMS separations require 10-60 ms while the time required to obtain a spectrum from TOF-MS is on the order of 100 μs.[12] As a result of the difference in timings required for each dimension of analysis, there are more than enough data points that can be acquired to properly sample analytes as they elute from the LC column.

Using liquid chromatography as a mechanism for introducing the greatly simplified samples to the mass spectrometer has been shown to decrease detection limits in addition to increasing the peak capacity of the analysis. For one study, IMS-MS was compared to LC-

IMS-MS for peptide analysis. Without the use of chromatography to introduce the sample, the limit of detection (LOD) was found to be 40 attomoles. When used in conjunction with LC, an LOD of approximately 1 attomole was the result.[13] Ion mobility technology is commercially available. The type of ion mobility used on one such platform, the Synapt G2 HDMS platform by Waters Corporation, is traveling-wave ion mobility. The distinct advantage of a traveling-wave setup is that the duty cycle of the ion mobility separation does not adversely affect the sensitivity of the mass spectrometer.[14] An additional benefit further improving the sensitivity of this method is the reduction in background ions as a direct result of the IMS separations.[13]

2.2 Experimental

2.2.1 Overview of experimental methods

The work presented in this chapter focuses on the ability of commercialized instrumentation with commonly used workflows to analyze a complicated sample – whole-cell lysates from Baker's yeast. A one-dimensional UPLC separation was conducted with Baker's yeast lysates that were grown differentially – one on dextrose-based media, the other on glycerol-based media. The proteins that were identified, and their associated fold-changes, were then compared to 2D UPLC methods available using a 2D-nanoAcquity platform. Lastly, to assess the effect of improved mass spectrometric analysis, the same one-dimensional method used to analyze the samples in this chapter with a previous generation of mass spectrometer were analyzed by one with an improved detector design and ion mobility.

2.2.2 Preparation of S. cerevisiae protein extracts

YAPD and YAPG growth media was prepared by combining 6.0 g of yeast extract, 12.0 g of Peptone, 12.0 g of glucose or 5 mL of glycerol, 60 mg adenine hemisulfate, and

600 mL of water and an additional 10 g bacto-agar for media for the plates. *S. cerevisiae* (BY4741) was the cell line used for analysis. Plates of each growth media were streaked with yeast and incubated until sizeable colonies were obtained. This process required two days for dextrose and four days for glycerol-based media. A single colony was then used to inoculate a 150 mL small-scale culture. These cultures were grown to an O.D. greater than two before being used to inoculate a 2 L (in a 4 L flask) prep scale batch. Both the glycerol and dextrose-based media samples were harvested when the O.D. was 2.0. For the glyercol-based media, significantly more time was required to reach this endpoint. Cells were centrifuged at 7000g in a Sorvall GS-3 rotor for 30 minutes until pelleted. Cells were then stored at -80°C until processed.

Cells were resuspended by pipet in 2 volumes of 50 mM ammonium bicarbonate with protease inhibitors present (Pierce protease inhibitor tablets, 88661) prepared to manufacturer's recommendations. A homogenate was prepared by 8 passes through a chilled french press cell dropwise at 20,000 psi. The homogenate was centrifuged (Beckman JA20 rotor, 30,000 g, 20 min, 4 °C) and a cytosolic fraction was prepared from the cleared lysate by ultracentrifugation at 120,000 g for 90 min, 4 °C. Cytosolic fractions were determined to be between 10-13 mg/ml of total protein using the Bio-Rad (Hercules, CA) protein assay with BSA standard. Immediately prior to analysis, each fraction was diluted with 25 mM ammobium bicarbonate to provide an equivalent protein concentration in all samples.

2.2.3 Trypsin digestion of fractions

Fractions were transferred to microcentrifuge tubes, then lyophilized and reconstituted in 25 µL of 50 mM ammonium bicarbonate. Three microliters of 6.67% (w/v) RapiGest™ SF in buffer were added (15 min, 80 ºC) to denature the proteins. The proteins

were reduced by adding 1uL of 100mM dithiothreitol (30 min, 60ºC), and then alkylated with 1uL of 200 mM iodoacetamide (30 min, room temperature, protected from light). The proteins were then digested by adding 10 µL of 320 ng/µL TPCK-modified trypsin in 50 mM ammonium bicarbonate (overnight, 37ºC). Once the digestion was complete, the RapiGest™ SF was degraded using 44 µL of 1% (v/v) trifluoroacetic acid (2 h, 60ºC). The fractions were centrifuged for 20 minutes at 14,000 x g to pellet the hydrolyzed surfactant, after which they were ready for analysis. The samples were transferred to sample vials and spiked with 4.21 µL of a 1 pmol/L internal standard bovine serum albumin digest.

2.2.5 Capillary UPLC-MS/MS of tryptic peptides

Each fraction was analyzed using capillary RPLC-MS/MS using a Waters nanoAcquity/QTOF Premier system. The separations were performed on a 250 mm x 75 µm ID capillary column packed with 1.7 µm silica bridged-ethyl particles with a C18 stationary phase (Waters). Mobile phase A was Optima-grade water with 0.1% formic acid (Fisher), and mobile phase B was Optima-grade acetonitrile with 0.1% formic acid (Fisher). At a flow rate of 300 nL/min, gradients varying in length of 60 to 180 minutes spanning 5-40% acetonitrile were used to separate the peptides. Each gradient was followed by a 5 minute column wash at 85% B, after which the mobile phase was returned to 5% B. Two µL of sample was injected per run for an average column load of approximately 1 µg of digest protein per run. The outlet of the UPLC column was directly connected to an uncoated fused silica nanospray emitter with a 20 µm ID and a 10 µm tip (New Objective, Woburn, MA) operated at 2.7 kV. Data-independent acquisition, or MS^E, scans were performed and the instrument was set to perform scans from m/z 50-1990 over 0.6 sec with the collision energy alternating between low energy (5.0V), and a high energy ramp (15-40V).

2.2.6 Protein identification by PLGS 2.5 & Scaffold 3.3.

After the MS^E data was collected, the mass spectra were processed by ProteinLynx Global Server 2.5 (Waters Corporations, Milford, MA) for peptide assignments. Database searching and statistical analysis was then completed by importing the sequenced data into Scaffold 3.3 (Proteome Software, Portland, OR). To keep the protein false discovery rate to approximately 5%, a minimum of 4 peptides, each with a confidence level of 95% were used to make the identification of a protein. Two different thresholds were used for protein identifications. To assess the chromatography, if a protein was identified with a confidence greater than 95% in at least one of three replicate runs, then it was included in the final protein count for that set of experiments. A more stringent criterion, however, was applied to those proteins used for the differential analysis. In order for a protein to be used in the differential comparison, it must have been found in at least two of the three replicate runs at a confidence level greater than 95% for a given sample. This was done to allow for Scaffold 3.3 to perform a Fischer's exact test on the data and to increase the confidence in the differential comparison.

2.3 Results & Discussion

2.3.1 Differential comparison of Baker's yeast by 1D UPLC.

The goal of this study was to determine the effectiveness of commercially available instrumentation and software tools for the purpose of making a differential proteomic analysis. Baker's yeast was chosen as it could readily be grown on different carbon sources, which one would expect to result in the differential regulation of various metabolic pathways. Baker's yeast is also a moderately complex proteome with approximately 5,500 proteins. While one might not consider the methods discussed to be "high throughput," they do allow

for a greater number of samples to be analyzed in a day than more comprehensive techniques, which will be discussed in Chapter 3.

The shortest single-dimension technique employed utilized a 60-minute gradient for an injection-to-injection time of 90 minutes. A total of 58 proteins were found using the criterion that a protein must be found in at least two of three replicate runs to be counted as an identification for the differential comparison. Figure 2-2 is a Venn diagram of the proteins identified by this approach. A total of 30 proteins were identified from the sample containing yeast grown on dextrose and 49 from the sample originating from the glycerol media. Of these, nine were unique to dextrose, 28 to glycerol, with 21 found in common. These numbers, however, do not take into account any statistical analysis regarding whether or not they were expressed differentially. Only 31 of the 58 proteins identified were expressed differentially as determined by the Fisher's exact test using a P value <0.05. This was further reduced by two to 29 as a result of two of the identifications having a fold-change of less than 1.5, the minimum accepted threshold. Table 2-1 lists the differential proteins, their associated fold-change, and which sample they were found to be up-regulated in. What can immediately be noticed from this is that the majority of the identifications, 23 of 29, were considered differential as the identified protein was only found in one of the two samples.

Metabolic pathways of interest are shown in Figure 2-3 with proteins identified, but not differentially shown in black, those found to be up-regulated in the dextrose sample in blue, and those up-regulated in the glycerol sample in red. Proteins not identified, but involved in these pathways are shown in bold-face grey type. What can be noticed immediately is that very few proteins that were searched for were identified. In fact, only 20 of 53 proteins were found. Cleary this is the result of the few number of identifications found

overall, which is an inherent weakness for this type of analysis. While the digestion is quick and easy to do and the separation does not require much time, a significant burden is placed on the mass spectrometer and associated software. A total of 1 μg of digested protein was loaded onto the trap column for the analysis. Loading greater amounts than this has deleterious effects on the chromatography, as peak shapes suffer because of overloading. With potentially more than a thousand proteins in the sample spanning several orders of magnitude in expression levels, the analysis poses a significant challenge for the mass spectrometer, as the peak capacity of the UPLC separation is insufficient to resolve the components. In addition to the overall low number of protein identifications, the number of peptides used to make the identifications was explored. These numbers varied considerably. On average, 8.4 peptides were used in an identification; however, the relative standard deviation was 114%. The median value of peptides used to make the identification was found to be 7.0 peptides with a high of 27 and a low of 4. Given the large range of expression levels possible, this is not an unexpected result. A key, limiting factor in the number of protein identifications possible via this approach is the lower limit of detection of the peptides needed to make those identifications.

For the 60-minute method, glucokinase 1 (HXKG), alcohol dehydrogenase 2 (ADH2), aldehyde dehydrogenase 4 (ALDH4), phosphoenolpyruvate carboxykinase (PCKA), malate dehydrogenase (MDHM), citrate synthase 1 (CISY1), and succinyl-CoA ligase (SUCA) were found to be up-regulated in the glycerol sample. ADH2 and ALDH4 expression is dextrose repressed, therefore their apparent up-regulation in the glycerol sample matches expectations as there was no dextrose present.[15,16] HXKG is involved in the first irreversible step of the metabolism of dextrose and is also found in the cytosol. It is one of three enzymes that *S.*

cerevisiae uses to catalyze the reaction, but it is not responsible for the main role in this reaction. It has been demonstrated though, that when shifted to a non-fermentable carbon source, such as glycerol, expression of this enzyme increases.[17] Malate dehydrogenase is involved in the TCA pathway, which during fermentation leads to the formation of organic acids.[18] This is likely due to the cell's response to being in glycerol and needing more mitochondrial enzymes in order to generate energy.

During gluconeogenesis, PCKA works to catalyze the reaction of oxaloacetate to phosphoenolpyruvate. The process of gluconeogenesis is what allows organisms, such as yeast, to grown on carbon sources other than dextrose, such as glycerol. PCKA expression is repressed when dextrose is present and is expected to be up-regulated on glycerol, which fits with the results from this analysis.[19] One enzyme was found to be up-regulated in the dextrose sample. This was hexokinase 2 (HXKB). According to the work of Lobo et al, this is an expected result as the gene encoding for HXKB is repressed and those encoding for HXKA and HXKG are up-regulated when cells are moved to a non-fermentable carbon source.[20]

While it was encouraging that the findings from the proteins which were determined to be differentially expressed generally matched predictions given the drastic change in carbon source, the overall number of protein identifications and the sparse coverage of the few metabolic pathways examined demonstrated the limitations of this particular experiment. In order to explore the effect of gradient length on the number of identifications and coverage of the metabolic pathways, gradients of 90, 120, 150, and 180 minutes were also run in addition to the 60-minute gradient. By extending the gradient time, but keeping the flow rate constant at 300 nL/min, and the starting and ending points of the gradient (5-40%

acetonitrile) consistent, the percent change in the organic modifier across the analytical column was reduced for increasing gradient times. It was expected that peak capacity would increase to a point and then level off as the maximum separation capability of the column is being used, relying less on the gradient at the long times. With increased peak capacity, it was believe that there would also be a greater number of identifications. For the 90-minute gradient, this was correct, as there was a slight increase in identifications meeting the threshold used for differential identifications from 58 to 80. For gradients longer than 90 minutes, there was little to no improvement observed. The number of identifications for even longer methods were 78, 83, and 86 for the 120-, 150-, and 180-minute gradients, respectively. For each of the four longer methods, approximately the same number of proteins were identified in the dextrose-based sample, the glycerol-based sample, or both (Figure 2-4). The number of proteins determined to be actual differential hits based on the Fisher's exact test within Scaffold 3.3 was found to increase from 35 to 45 identifications from the 60- to 90-minute gradient lengths. This number remained relatively consistent with 42, 43, and 47 identifications for the three longest methods. The consistent number of identifications is indicative of the maximum peak capacity being achieved with a gradient length of approximately 90 minutes. This will be discussed in greater detail in section 2.3.2.

With the greater number of identifications made with longer gradients, ideally there would be greater coverage of the metabolic pathways of interest to this study. Figures 2-5 through 2-8 show the pathways and up-regulation of metabolic enzymes for the 90-, 120-, 150-, and 180- minute gradient lengths. In general, the results from the longer gradients were the same as what was previously observed when using the shorter gradient length. HXKA was an additional identifications in all four of the longer gradients and was found to be up-

regulated in glycerol. This data is in agreement with what was previously discussed regarding the enzymes involved in this particular process. At the two longest gradient times, there was a change in the apparent regulation of HXKB. It was no longer determined to be up-regulated in the dextrose sample, and was found to be expressed roughly equally in both samples. The scores for HXKB in the glycerol sample were consistent for all the methods. In the dextrose sample, it was either not found at all in the 60- and 120- minute gradients. For the 90-minute gradient, the counts were low and it was found to be up-regulated in glycerol with a fold-change of 1.9. A greater score was found for HXKB in the two longer gradients, causing it to not be a differential protein as determined by Scaffold 3.3. Another protein of interest that was found by increasing the gradient length beyond 90 minutes was glycerol-3-phosphatase (GPP1). This enzyme is responsible for the production of glycerol needed for cellular processes.[21] It was only found in the dextrose sample, which given its role, is not surprising. Two other proteins were determined to be up-regulated in the dextrose sample at longer run times, phosphoglucose isomerase (G6PI) and phosphofructokinase, subunit alpha (K6PF1). According to literature, both of these enzymes are involved in glycolysis, which is the process of splitting one molecule of glucose (dextrose) into two molecules of pyruvate with a net gain of 2 ATP.[22] G6PI converts glucose-6-phosphate to fructose-6-phosphate. K6PF1 then converts fructose-6-phosphate to fructose 1,6-bisphosphate.

2.3.2 Analysis of gradient length and available peak capacity.

From the previous section regarding the differential comparison, it is evident that the low number of identifications made the task of looking for trends in the up-regulation of metabolic enzymes difficult. While the differential proteins that were identified did fall in line with expectations from literature, the metabolic pathways of interest were not well

covered and there were relatively few additional proteins identified outside of these. Extending the gradient from 60 to 90 minutes did result in a nearly 30% increase in identifications. Lengthening them beyond this point did little to improve the number of identified proteins, although a few additional proteins were able to be identified that were of direct interest as they were involved in the targeted processes. As mentioned earlier, this would seem to indicate that there were no additional increases in peak capacity when the gradients were extended beyond 90 minutes. To examine this possibility further, representative chromatograms of the separations are shown in Figures 2-9 and 2-10 for the five gradients. Facilitating this analysis, PLGS 2.5 utilizes the Apex3D algorithm, which approximates peak shape and determines retention times as part of the data-independent acquisition. By examining this data, one can readily find the full-width, half-maximum of the peaks fit to a Gaussian. Using the half-height to determine peak width, it was observed that peak widths increased from 0.375 minutes to 0.508 minutes with an increase in gradient length from 60 to 90 minutes (Figure 2-11). While the peaks were wider in the 90-minute analysis, there was an increase in peak capacity from 160 to 180. By extending the gradient a further 30 minutes, the peak capacity increased only slightly to 190 and the same peak capacity was observed for the 150-minute gradient. At 180 minutes, the peak capacity was not able to be calculated from the PLGS data as the peaks had broadened and were tailing sufficiently enough that the program was now fitting multiple peaks to a single eluting peptide. This caused the apparent peak width to decrease from 0.802 minutes with the 150-minute gradient to 0.505 minutes, comparable to what was seen with a 90-minute gradient. Another complicating factor that was found was the inability for the data acquisition software, Masslynx 4.1, to collect data for the entire specified time. For the 180-minute gradient, 210

minutes were needed to see the column return to initial conditions. Each of the runs is cut off before this point. Upon examination of the data files themselves it was noticed that they were all of identical size. The problem is not necessarily related to any of the experimental parameters, but is more of a limitation of the software. As Masslynx operates in a 32-bit environment, it is unable to collect more than 4 GB of data. Fortunately, with this system there is little need to go to these lengths as there was no improvement to the analysis by going to analysis times this long.

2.3.3 2D nanoAcquity

As gradient times in excess of 90-minutes did not produce an increase in protein identifications and left the coverage of the metabolic pathways of interest with many unidentified proteins, a different approach was utilized to try to improve upon these results. A multidimensional analysis was implemented to further increase the peak capacity of the system beyond the 180-190 peaks possible with a 1D-UPLC system. The high/low pH RP-RP mode of operation was chosen as it had been shown to have greater orthogonality and therefore greater peak capacity than other multidimensional separation formats using the nanoAcquity platform. A total of 5 fractions were taken from the first dimension and each was subjected to a 30-minute gradient. Figure 2-12 shows the results from this analysis in terms of identified proteins. A total of 137 proteins were identified at least once with a confidence of 95%. Of those, 63 were found only in the dextrose sample, 27 only in the glycerol sample, and 47 found in common between the two samples. A more interesting number to look at though, is the number of differential proteins identified after the statistical analysis in Scaffold 3.3 was completed. A 1D UPLC separation of varying gradient lengths was producing an average of 42 differential identifications, with 45 being identified in the

90-minute gradient, which was considered to give the best productivity, identifying 22.5 proteins per hour. The 2D UPLC separation identified 83 differential proteins, which is clearly a significant improvement. In order to get twice the identifications, however, the analysis time was increased 3-fold as this analysis requires 6 hours per injection. In terms of productivity, this resulted in 13.8 identifications per hour. With the additional identifications, one would hope to see an increase in the coverage of the metabolic pathways. Unfortunately, as shown by comparing Figure 2-13 to Figure 2-3, this was not the end result. The proteins and their relative regulation for the 6-hour long 2D UPLC method was found to be identical to the results from the 1D 180-minute method, which was very similar to the other gradient lengths.

There are a few reasons why this 2D analysis did not result in a significant increase in the coverage of the metabolic pathways. Identifying twice the number of proteins sounds impressive, but when coupled with the fact that in order to achieve this result the analysis time was tripled and the total numbers remain low, this is less impressive. Speaking first to the instrumentation, while 6 hours are required of the analysis, much of this time is spent equilibrating the analytical column to initial conditions. Time is also spent eluting peptides off of the first dimension and onto the second dimension trap column so that they can be analyzed by the second dimension column. As a result, a 30-minute long gradient was the longest that could be used for an individual step. With five fractions, the total gradient time was 150 minutes, an increase of 67%. Given that perspective, identifying twice the proteins was an impressive outcome, but unfortunately in order to achieve those results, a substantial amount of time had to spent on “housekeeping” tasks. One of the reasons for using the high/low pH mode of 2D operation was the improved orthogonality, as demonstrated by

Gilar et al, over other 2D formats such as those utilizing strong cation exchange chromatography in the first dimension and reversed phase chromatography in the second (SCX/RP). Figure 2-14 shows the base peak index (BPI) chromatogram and Figure 2-15 shows the total ion current (TIC) chromatogram for a representative injection using the high/low pH operation. Good intensity for individual peptides was observed as shown from the BPI chromatogram, but a general trend of peptides eluting early in the first dimension also eluting early in the second dimension can be seen. By examining the TIC chromatogram, however, this effect is clearly seen. The first fraction, where 12% acetonitrile was used to elute peptides from the first dimension, shows the majority of peptides eluting from approximately 10 to 23 minutes. By the time the last fraction is taken, this elution envelope shifted out to later retention time and peptides eluted from 20 to 41 minutes. The effect this has on the analysis is that the experimental peak capacity falls well short of the theoretical maximum, as peptides were not eluting across the entire separation space. In fact, the elution envelope from the first dimension represents only 43% of the total gradient time.

Subsequent to having the instrumentation configured for a 2D analysis with the nanoAcquity UPLC, factors that could have lead to improved performance were determined. The choice of which percentage acetonitrile to use for the first dimension fractions were based on a visual inspection of the chromatograms after several different possibilities were tried. An improved approach would be to run a very detailed fractionation scheme over the course of more than a day, eluting peptides at one of two percent intervals. The summation of the ions could then be used to create a curve. By fitting an equation to this curve and integrating it, even cuts with respect to the total ion count could be achieved. Another improvement could be made to the gradients themselves. The same gradient was chosen for

each of the five fractions as this had been standard practice for all samples run by UPLC. With more method development, the gradients could have been more specifically tailored to the peptides eluting in a give fraction. Ultimately, these improvements could result in more identifications, but it is doubtful that it would lead to a substantial increase when compared to the number of entries in the yeast genome. Even with improvements in gradients and fractionation of the peptides, there was still a substantial amount of time spent regenerating columns and not collecting data. Chapter 3 discusses improvement to fractionation in multidimensional analyses and the work presented in Chapter 4 aims to improve productivity by spending less time regenerating columns during an analysis. The single largest improvement, which is by far the most expensive and practically difficult to achieve, is to use a better mass spectrometer.

2.3.4 Data from Synapt G2

Mass spectrometers, like most instrumentation, are constantly being improved upon. Occasionally there are significant improvements in performance through the inclusion of new technology. The system used to acquire the data presented to this point was a nanoAcquity UPLC with a QTOF Premier quadrupole time-of-flight mass spectrometer. While a good mass spectrometer with a resolution of 10,000 and an expected sensitivity of 200 counts with an 50 pg/uL solution of leucine-enkephalin (Leu-Enk), it is technology from 2004 and is prior to a significant advancement in QTOF instrumentation available from Waters Corporation. For comparison, the same bottom-up samples were run on a similar system under identical LC conditions with the only difference being the use of a Synapt G2 mass spectrometer. The expected sensitivity of this instrument is about 10-fold greater than the

QTOF-Premier. Additionally, ion mobility is a feature allowing for an even greater peak capacity as peptides eluting at the same time can be further differentiated.

The gradient chosen for the comparison was the 90 minute gradient as it performed the best among the 1D methods explored. A total of 80 proteins were identified with the QTOF Premier. When used with the SynaptG2, the number rose to 485 representing a 6-fold increase in the number of identifications. Additionally, 295 of these proteins were found to be differentially expressed by Scaffold 3.3. Figure 2-16 shows the metabolic pathways that were expected to be affected by the different growth media. With better detection, the pathways were covered in greater detail. All but three proteins were identified and these were not found by any of trials involving different gradient lengths for data collected on the Premier. The regulation of proteins found in common between the two techniques were in agreement with the exception of HXKB, which was previously mentioned to appear to be up-regulated in the dextrose sample in shorter gradient lengths, but found to not be differentially expressed at longer gradient times. For the data collected on the Synapt G2, it was determined to not be differentially expressed.

The quality of data was found to be much better from the SynaptG2. Figure 2-17 shows a histogram of the correct and incorrect assignments for peptides as it relates to the peptide score. The higher the peptide score, the more correct assignments one would expect to see. Looking at the distribution between the incorrect, shown in blue, and the correct, shown in red, one can see that there is a great deal of overlap between the two distributions. Effectively, this limits the number of peptides available for assignment as peptides with scores below the cutoff are ignored. The false discovery rate (FDR), or the rate at which peptides are assigned to the random database appended to the end of the actual database

result in false positives, is significantly affected by the peptide score and peptide confidence. With the data from the Premier, the peptide confidence had to be set at 95% and a minimum of 4 peptides had to be used in order to keep the FDR at approximately 5%. Lowering the peptide confidence to allow peptides that were only 90% or reducing the number of peptides used for an assignment to three, resulted in a massive increase of the FDR to greater than 20%. In contrast, data from the SynaptG2, with wider spacing in peptides scores between correct and incorrect assignments, allowed these thresholds to be set much lower. The peptide confidence could be set as low as 20% and the number of peptides used for an identification could be reduced to 2, resulting in an FDR of only 8%.

2.4 Summary and Conclusions

The purpose of this work was to serve as a baseline for comparison for later studies involving similar samples. The results from the one-dimensional analysis of the total bottom-up samples on the nanoAcquity/QTOF-Premier system was not surprising given the lack of peak capacity. While throughput is a primary concern for some analytical problems where vast numbers of samples need to be analyzed, often looking at a particular analyte, others are more concerned with the broad scope of what is in the sample. The protein digests from whole-cell lysates fall into the latter category and spending more time on the analysis to obtain improvements in the results is worthwhile to the investigator. Increasing the peak capacity did serve to improve the number of differential identifications substantially, nearly doubling them when a multidimensional separation of the peptides was employed. Ultimately the greatest increase in differential identifications was due to the improved results afforded by the more advanced mass spectrometric detection. With a 6-fold increase in identifications, the metabolic pathways of interest were completely covered. One could argue that if the data

collected from the multidimensional peptide separation had been obtained on a SynaptG2, that over 700 proteins could have been identified, 500 of them differentially. Later chapters will focus on different way to improve the peak capacity of the chromatographic separation using the QTOF-Premier for detection. These results should be compared to those obtained on the QTOF-Premier with the knowledge that the improved mass spectrometry already available commercially could allow for an even deeper look at the proteome.

2.5 References

1. Han, D. K.; Wu, L., Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Review of Proteomics* **2006**, 3, 611.

2. McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H., Ion mobility-mass spectrometry: a new paradigm for proteomics. *International Journal of Mass Spectrometry* **2005**, 240, 301-315.

3. Swanson, S. K.; Washburn, M. P., The continuing evolution of shotgun proteomics. *Drug Discovery Today* **2005**, 10, 719-725.

4. Ghaemmaghami, S.; Huh, W. K.; Bower, K., Global analysis of protein expression in yeast. *Nature* **2003**, 425, 737-741.

5. Washburn, M.P.; Wolters, D.; Yates, J. R., Large scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, **2001**, 19, 242-247.

6. de Godoy, L. M.; Olsen, J. V.; de Souza, G. A.; Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biology* **2005**, 7, 50.

7. Anderson, N. L.; Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Molecular & Cellular Proteomics* **2002**, 1, 845-867.

8. Han, X.; Aslanian, A; Yates, J. R., Mass spectrometry for proteomics. *Current Opinion in Chemical Biology* **2008**, 12, 483-490.

9. Lin, D.; Tabb, D. L.; Yates, J. R., Large-scale protein identification using mass spectrometry. *Biochimica et Biophysica Acta* **2003**, 1646, 1-10.

10. Gilar, M.; Olovova, P.; Daly, A. E.; Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second dimensions. *Journal of Separation Science* **2005**, 28, 1694-1703.

11. Kanu, A. B.; Dwivedi, P.; Tam, M.; Matz, L.; Hill, H. H., Ion mobility-mass spectrometry. *Journal of Mass Spectrometry* **2008**, 43, 1-22.

12. Valentine, S.J.; Kulchania, C.A.; Barnes, S.; Clemmer, D. E., Multidimensional separations of complex peptide mixtures: a combined high-performance liquid chromatography/ion mobility/time-of-flight mass spectrometry approach. *International Journal of Mass Spectrometry* **2001**, 212, 97-109

13. Angel, T. E.; Aryal, U. K.; Hengel, S. M.; Baker, E. S.; Kelly, R. T.; Robinson, E. W.; Smith, R. D., Mass spectrometry-based proteomics: existing capabilities and future directions. *Chemical Society Reviews* **2012**, 41, 3912-3928.

14. Pringle ,S. D.; Giles, K.; Wildgoose, J. L.; Williams, J. P.; Slade, S. E.; Thalassinos, K.; Bateman, R. H.; Bowers, M. T.; Scrivens, J. H., An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupoles/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* **2007**, 261, 1-12.

15. Wills, C., Amino acid substitutions in two functional mutants of yeast alcohol dehydrogenase. *Nature* **1979**, 279, 734-736.

16. Larochelle M, et al. Oxidative stress-activated zinc cluster protein Stb5 has dual activator/repressor functions required for pentose phosphate pathway regulation and NADPH production. *Molecular & Cellular Biology* **2006**, 26, 6690-6701.

17. Clifton, D., Functional studies of yeast glucokinase. *Journal of Bacteriology* **1993**, 175, 3289-3294.

18. Grandier-Vazeille, X., et al. Yeast mitochondrial dehydrogenases are associated in a supramolecular complex. *Biochemistry* **2001**, 40, 9758-9769.

19. Haarrasilta, S.; Oura, E., On the activity and regulation of anaplerotic and gluconeogenetic enzymes during the growth process of baker's yeast. The biphasic growth. *Eur J Biochem* **1975**, 52, 1-7.

20. Lobo, Z.; Maitra, P. K., Physiological role of glucose-phosphorylating enzymes in Saccharomyces cerevisiae. *Archives of Biochemistry & Biophysics* **1977**, 182, 639-645.

21. Pahlman, A. K.; Granath, K.; Ansell, R.; Hohmann, S.; Adler, L., The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress. *Journal of Biological Chemistry* **2001**, 2, 3555-3563.

22. Aguilera, A.; Zimmermann, F. K., Isolation and molecular analysis of the phosphoglucose isomerase structural gene of Saccharomyces cerevisiae. *Molecular and General Genetics* **1986**, 202, 83-89.

2.6 Tables

1D UPLC – 60 minute Gradient		
Accession Number	Fold Change	Up Reg
ALDH4_YEAST	G Only	G
KPYK1_YEAST	1.4	Change<1.5
ADH2_YEAST	6.9	G
HSP72_YEAST	1.8	D
ATPB_YEAST	G Only	G
HXKG_YEAST	G Only	G
MDHM_YEAST	G Only	G
HSP75_YEAST	2.2	D
ATPA_YEAST	G Only	G
OAT_YEAST	G Only	G
HSP26_YEAST	G Only	G
ALF_YEAST	1.4	Change<1.5
PCKA_YEAST	G Only	G
CISY1_YEAST	G Only	G
QCR2_YEAST	G Only	G
IF4A_YEAST	2.4	D
BMH2_YEAST	G Only	G
CDC48_YEAST	G Only	G
RS7A_YEAST	G Only	G
SUCA_YEAST	G Only	G
EF2_YEAST	3.5	D
RIR2_YEAST	G Only	G
AHP1_YEAST	3.5	D
HSP76_YEAST	D Only	D
HXKB_YEAST	D Only	D
EF1G2_YEAST	D Only	D
ACBP_YEAST	D Only	D
SYDC_YEAST	D Only	D
SNU13_YEAST	D Only	D
METE_YEAST	D Only	D
OYE2_YEAST	D Only	D

Table 2-1. Proteins identified as being differentially expressed between the dextrose-based and glycerol-based baker's yeast samples are shown here. These proteins were found to have a P-value of less than 0.05 when conducting a Fisher's exact test.

2.7 Figures

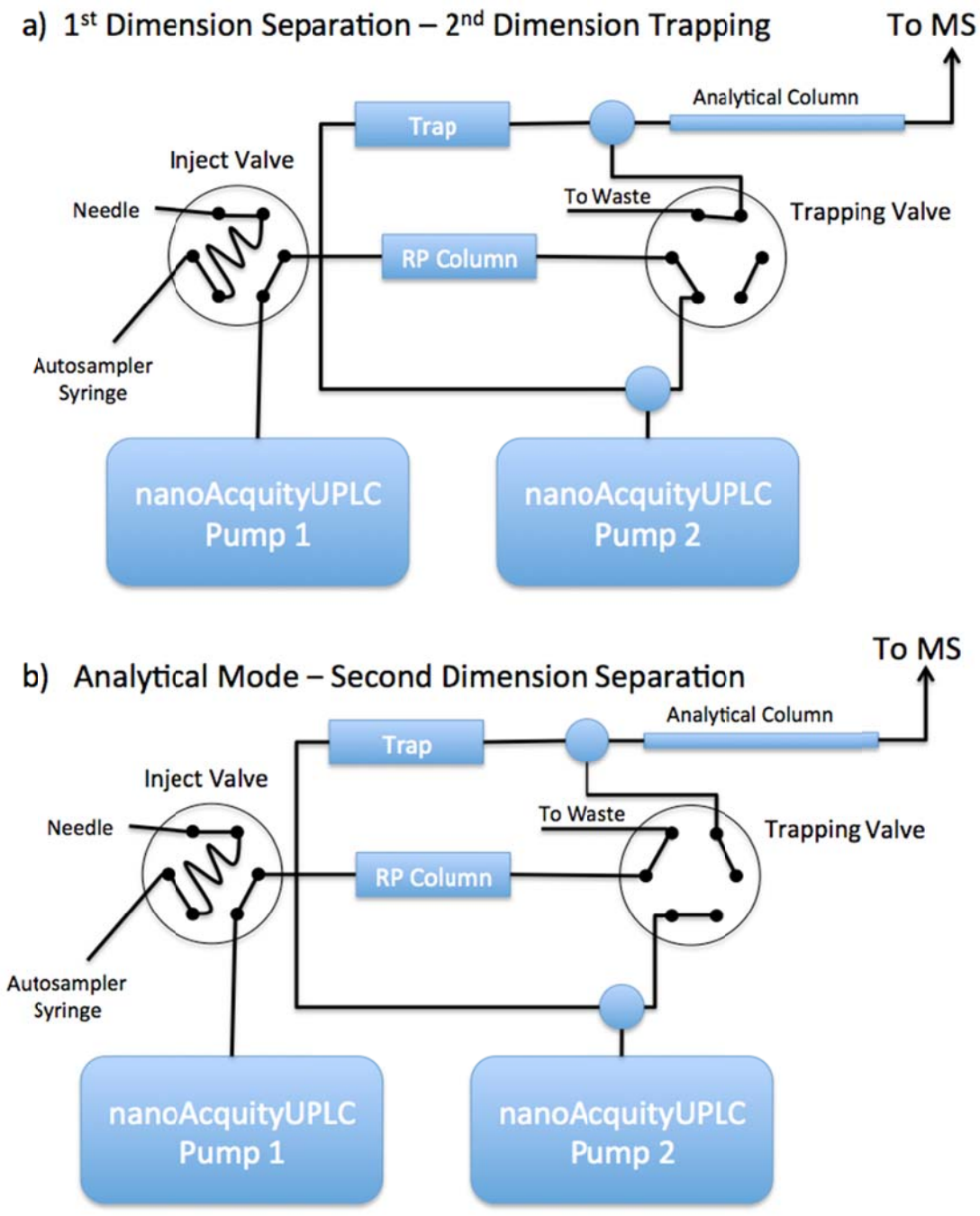


Figure 2-1. a) During the first dimension of separation, flow from the first pump at a pH of 10 flows through the first-dimension column. Effluent from this column is diluted with low pH mobile phase from the second pump before being trapped on the trap column. During the second dimension of analysis (b), flow from the first pump is diverted after the first-dimension column to waste and a gradient is pumped from the second pump through the analytical column with detection by mass spectrometry.

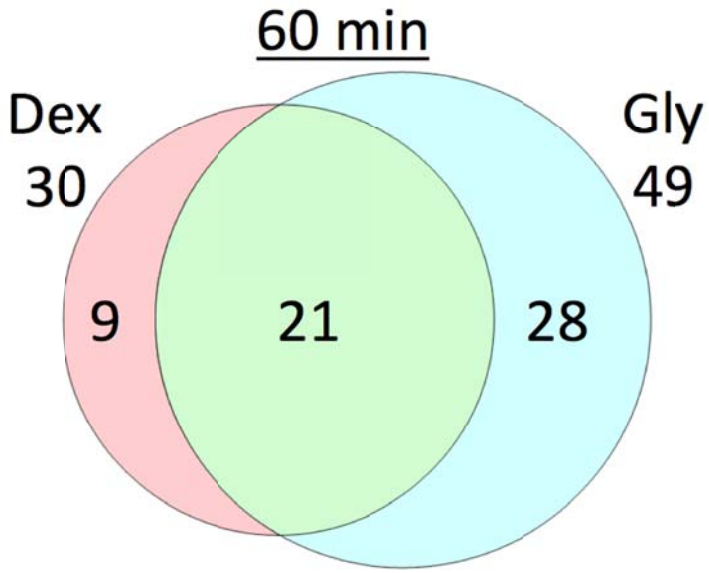


Figure 2-2. The separation of peptides from the dextrose-based sample and glycerol-based sample was by reversed-phase UPLC with a 60 minute long gradient from 5-40% acetonitrile. For the purposes of a differential comparison, proteins were only counted if they were identified with a confidence greater than 95% by Scaffold 3.3 in two of three replicates. This figure indicates the number of proteins identified in each sample and the number of proteins identified in common. For the differential comparison, proteins were required to have a p-value less than 0.05 after a Fisher's exact test was applied to the data.

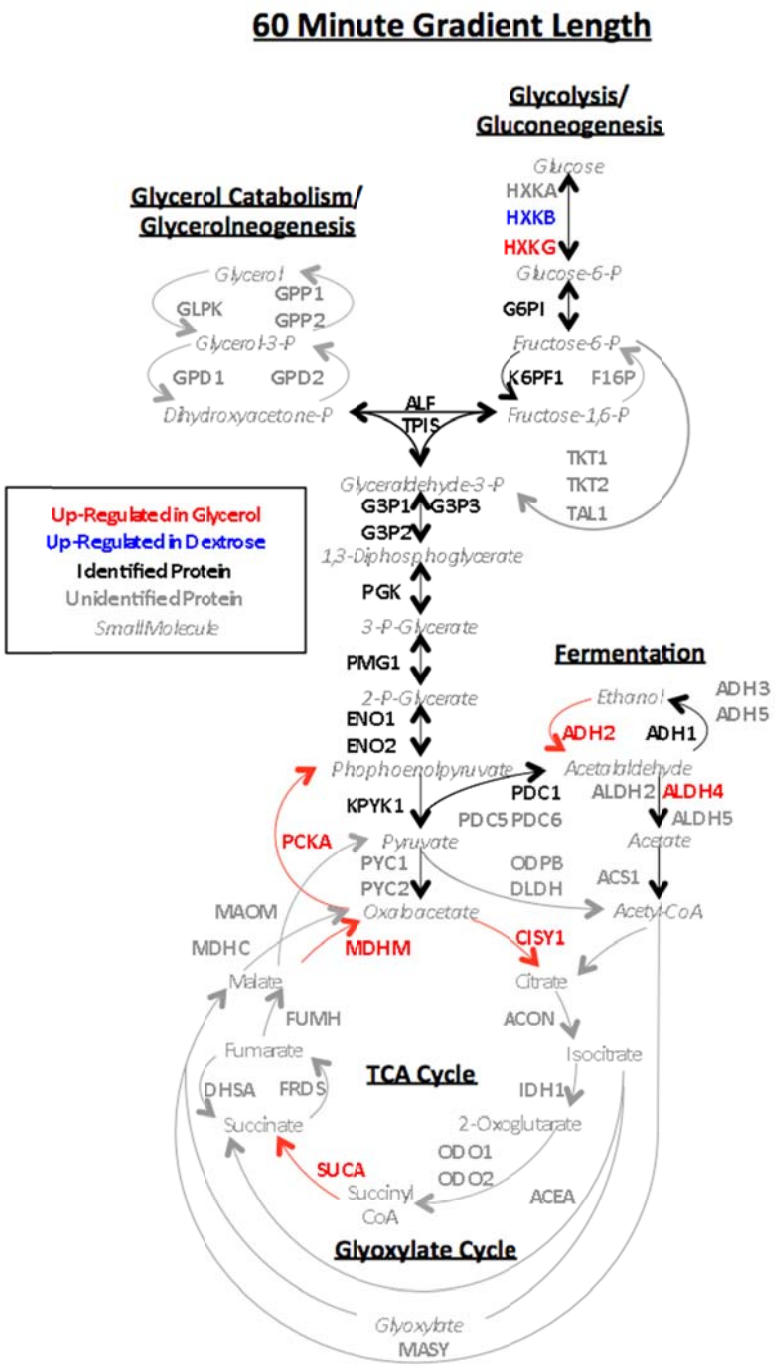


Figure 2-3. Given the change in carbon source from dextrose to glycerol, it was expected that there would be differences in protein expression for the metabolic pathways. Simplified pathways and their associated proteins are shown here. Proteins found to be up-regulated in glycerol are shown in red. Those up-regulated in the dextrose sample are shown in blue. Proteins identified, but not determined to be differentially expressed are shown in boldface black with unidentified proteins shown in gray. The gradient length used in the peptide separation was 60 minutes in length from 5-40% acetonitrile.

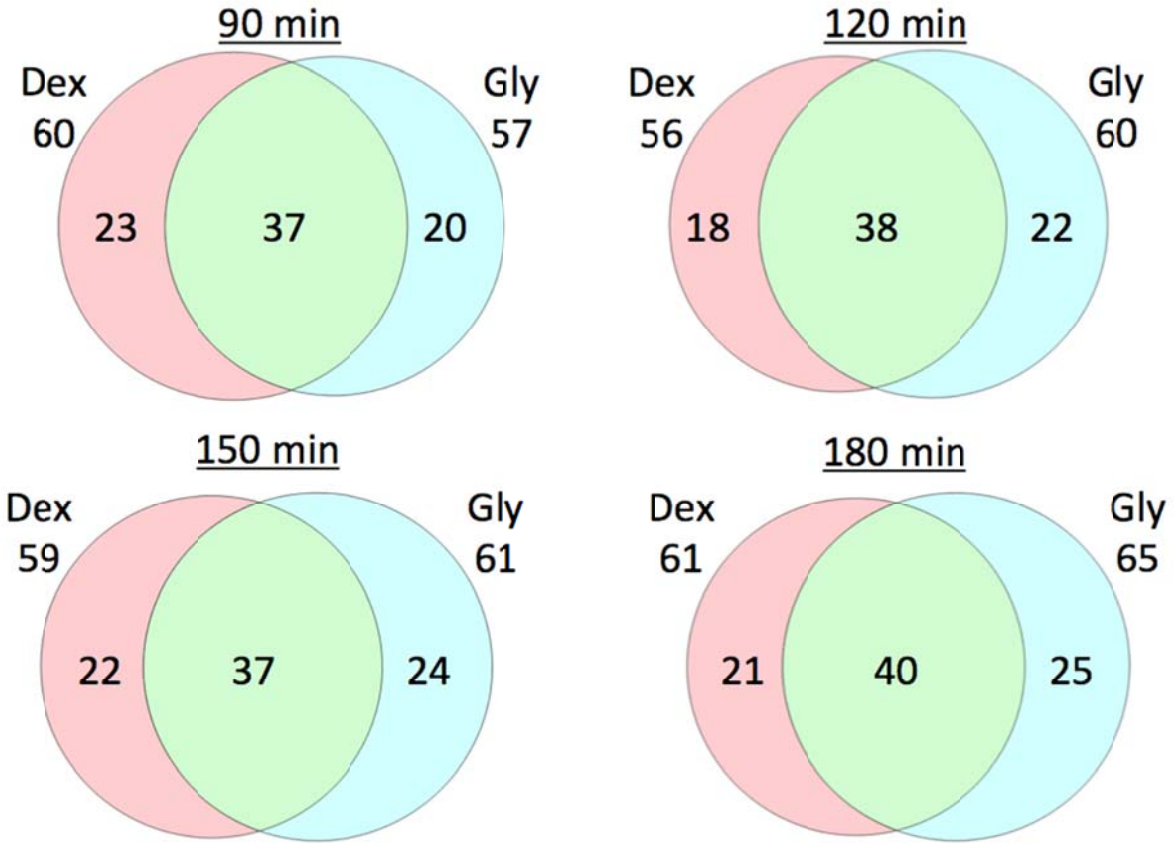


Figure 2-4. Additional gradient lengths were attempted. The gradient lengths used were (a) 90-, (b) 120-, (c) 150-, and (d) 180-minutes. For the purposes of a differential comparison, proteins were only counted if they were identified with a confidence greater than 95% by Scaffold 3.3 in two of three replicates. This figure indicates the number of proteins identified in each sample and the number of proteins identified in common. For the differential comparison, proteins were required to have a p-value less than 0.05 after a Fisher's exact test was applied to the data.

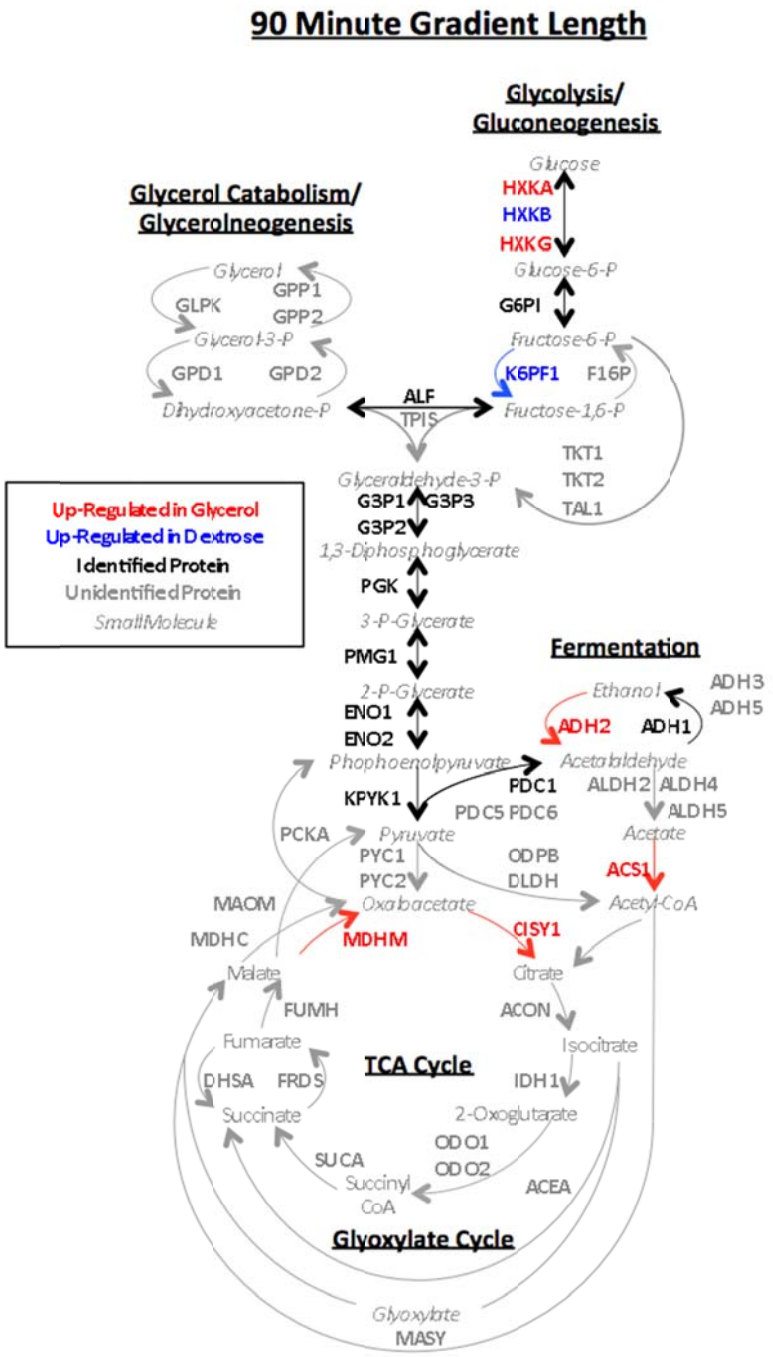


Figure 2-5. Given the change in carbon source from dextrose to glycerol, it was expected that there would be differences in protein expression for the metabolic pathways. Simplified pathways and their associated proteins are shown here. Proteins found to be up-regulated in glycerol are shown in red. Those up-regulated in the dextrose sample are shown in blue. Proteins identified, but not determined to be differentially expressed are shown in boldface black with unidentified proteins shown in gray. The gradient length used in the peptide separation was 90 minutes in length from 5-40% acetonitrile.

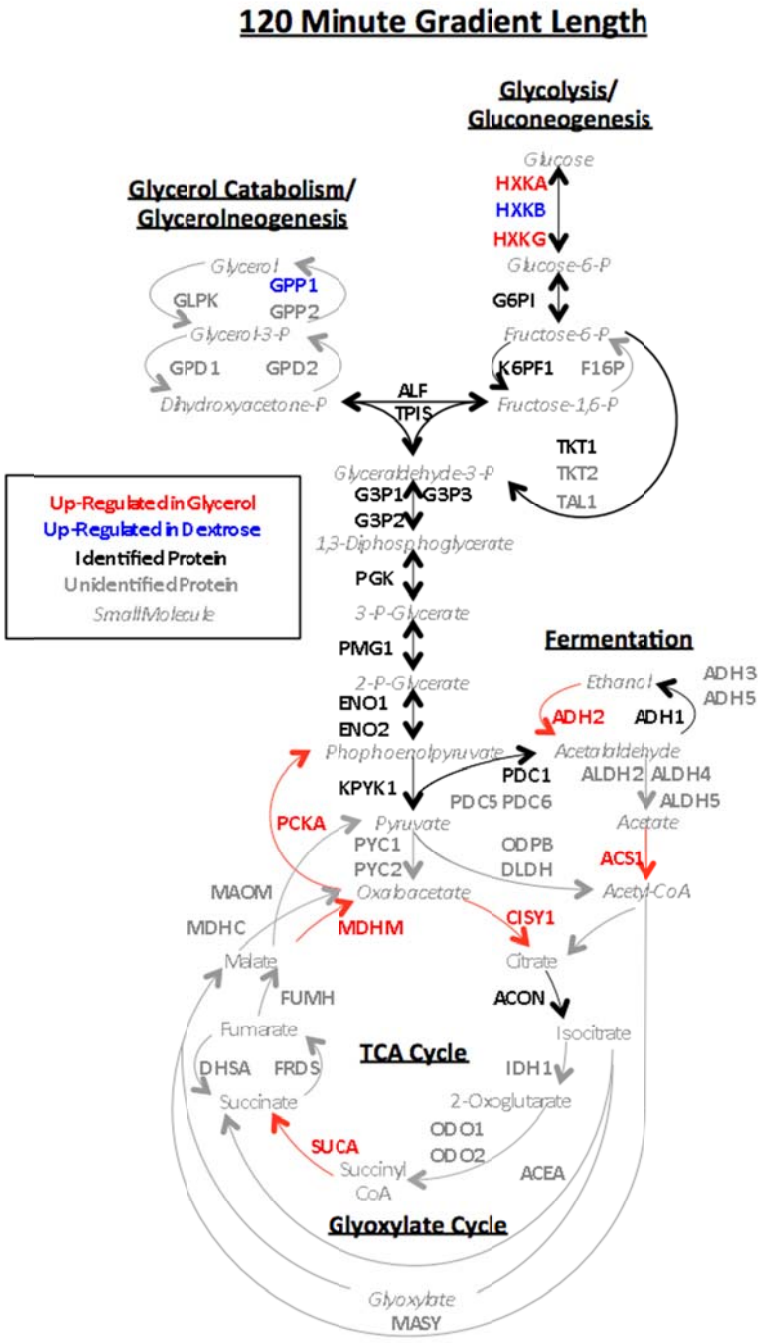


Figure 2-6. Given the change in carbon source from dextrose to glycerol, it was expected that there would be differences in protein expression for the metabolic pathways. Simplified pathways and their associated proteins are shown here. Proteins found to be up-regulated in glycerol are shown in red. Those up-regulated in the dextrose sample are shown in blue. Proteins identified, but not determined to be differentially expressed are shown in boldface black with unidentified proteins shown in gray. The gradient length used in the peptide separation was 120 minutes in length from 5-40% acetonitrile.

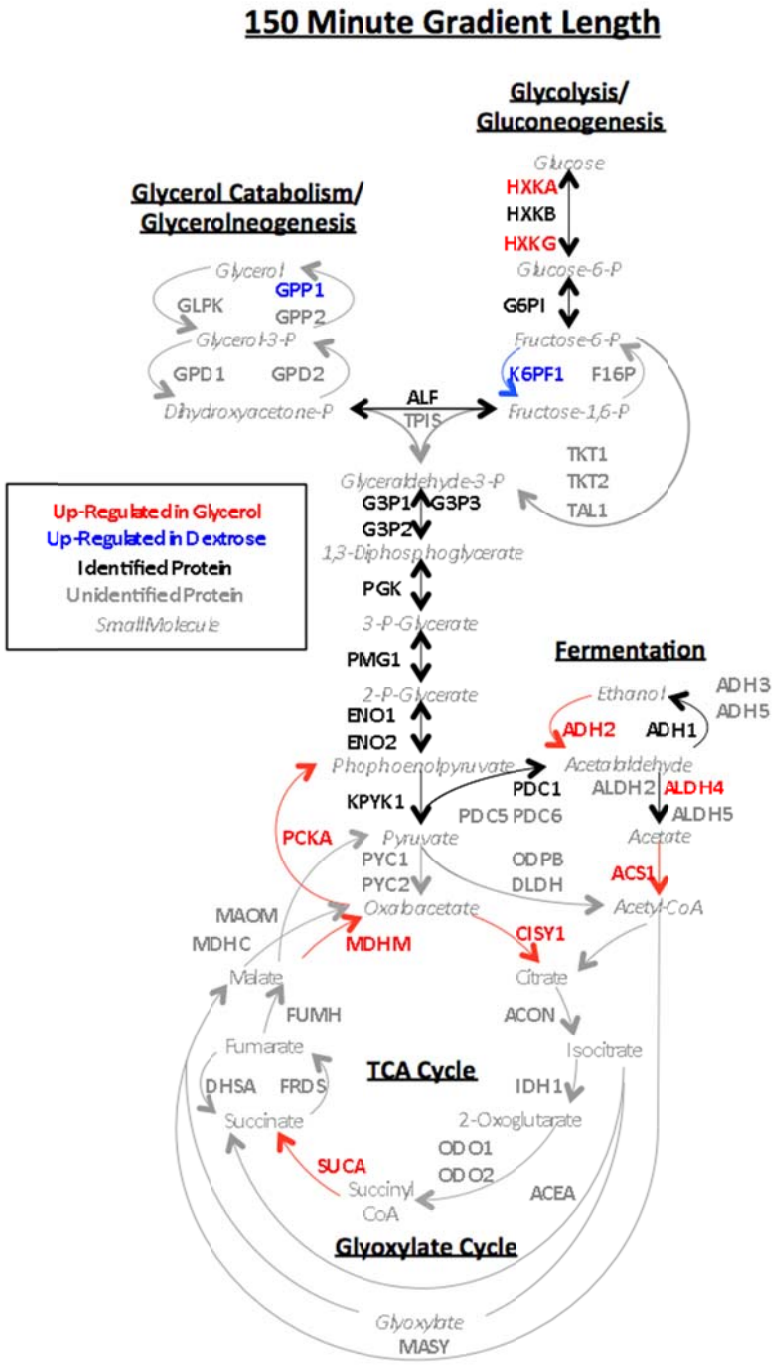


Figure 2-7. Given the change in carbon source from dextrose to glycerol, it was expected that there would be differences in protein expression for the metabolic pathways. Simplified pathways and their associated proteins are shown here. Proteins found to be up-regulated in glycerol are shown in red. Those up-regulated in the dextrose sample are shown in blue. Proteins identified, but not determined to be differentially expressed are shown in boldface black with unidentified proteins shown in gray. The gradient length used in the peptide separation was 90 minutes in length from 5-40% acetonitrile.

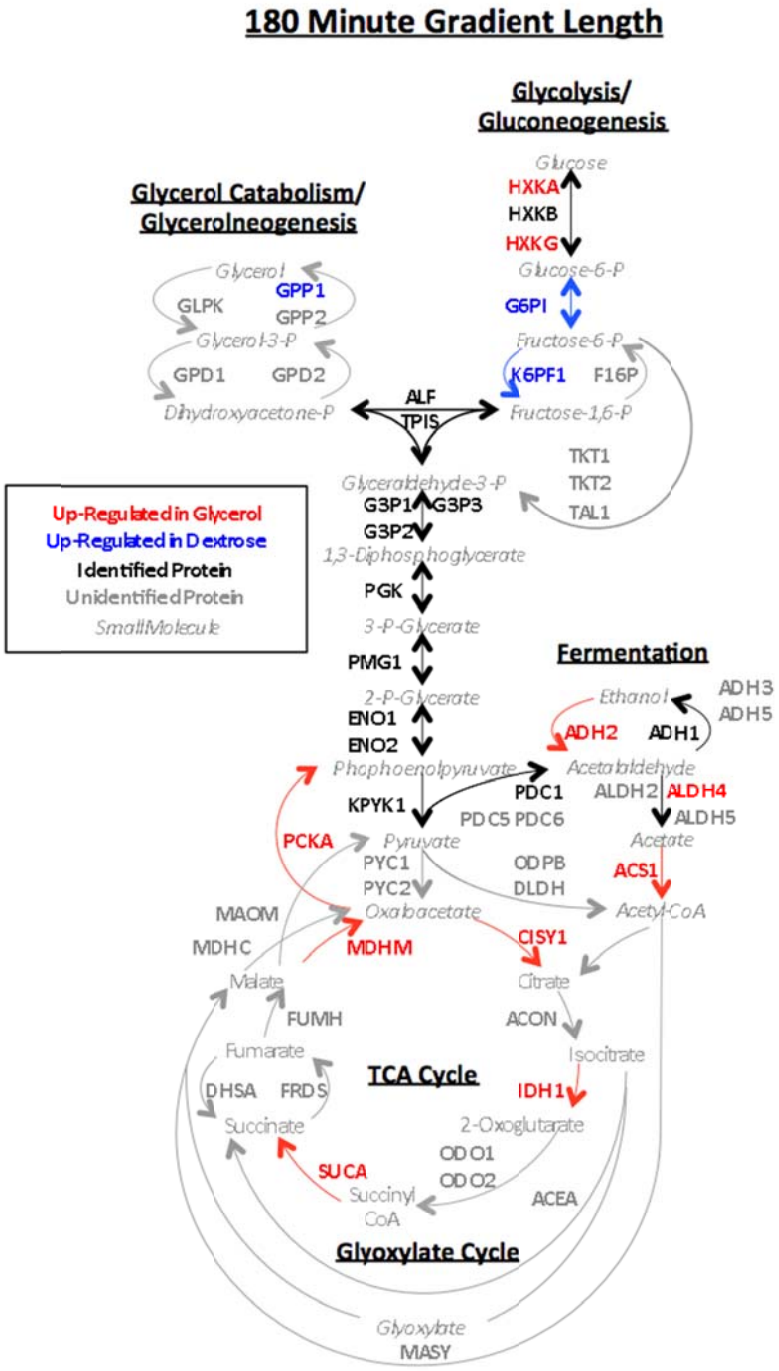


Figure 2-8. Given the change in carbon source from dextrose to glycerol, it was expected that there would be differences in protein expression for the metabolic pathways. Simplified pathways and their associated proteins are shown here. Proteins found to be up-regulated in glycerol are shown in red. Those up-regulated in the dextrose sample are shown in blue. Proteins identified, but not determined to be differentially expressed are shown in boldface black with unidentified proteins shown in gray. The gradient length used in the peptide separation was 90 minutes in length from 5-40% acetonitrile.

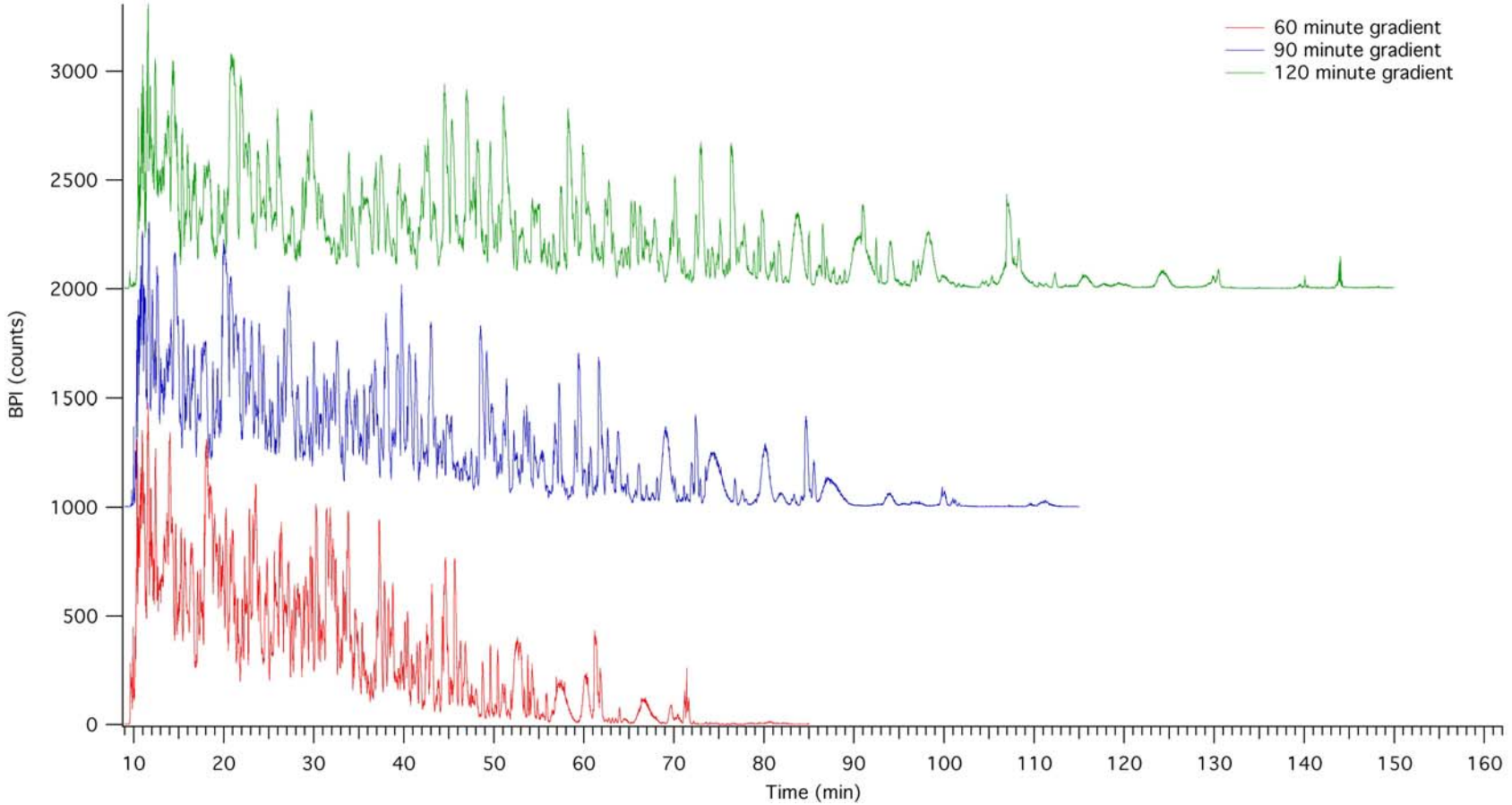


Figure 2-9. Representative chromatograms of the glycerol-based, yeast cell lysate digest for gradient lengths of 60-, 90-, and 120-minute long gradients. The 60-minute method shows a chromatogram that is crowded with poorly resolved peaks. This improves slightly with the 90-minute gradient. A more thorough analysis is needed than what is possible through visual inspection to determine improvements to peak capacity.

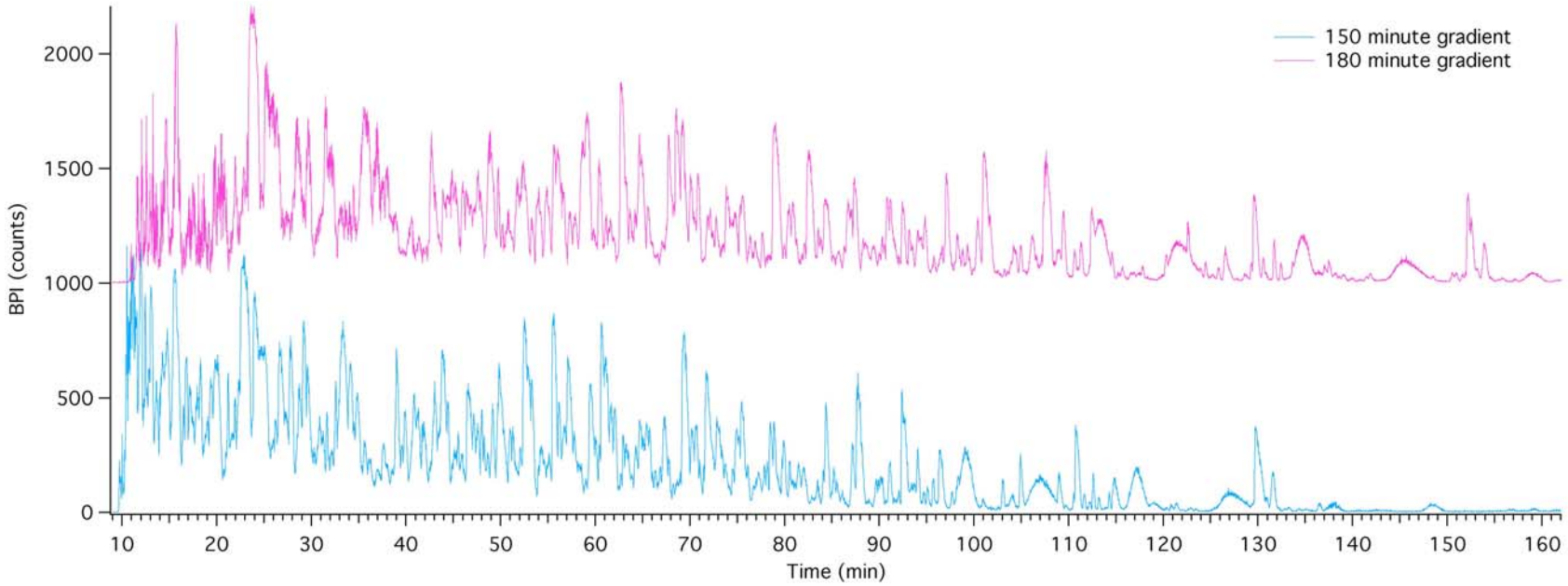


Figure 2-10 Representative chromatograms of the glycerol-based, yeast cell lysate digest for gradient lengths of 150- and 180-minute long gradients. For both of these gradients, it does not appear as though the resolution is improving. Peak widths are broader at the longer gradient times. A more thorough analysis is needed than what is possible through visual inspection to determine improvements to peak capacity.

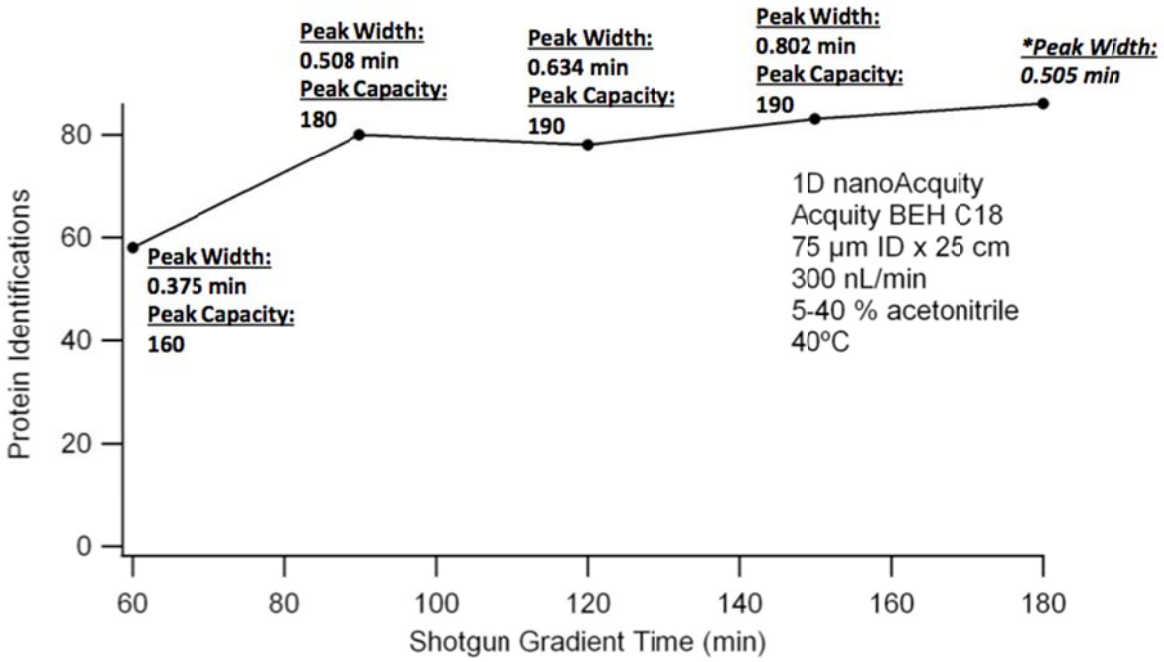


Figure 2-11. From the chromatograms shown above, the data shown here is the result of the PLGS 2.5 data processing. The Apex3D algorithm used to process the raw mass spectral data attempts to fit Gaussian peaks to the eluting peptides. From this data, the full-width, half-maximum (FWHM) of the peaks are readily available. Peak capacities here are calculated from the FWHM. Gradients longer than 90 minutes were not shown to give more protein identifications. There was a slight increase in peak capacity in lengthening the gradient from 90- to 120-minutes. At 180 minutes, the gradient becomes shallow enough and the peaks tail significantly enough that the Apex3D algorithm begins fitting multiple peaks per actual peak.

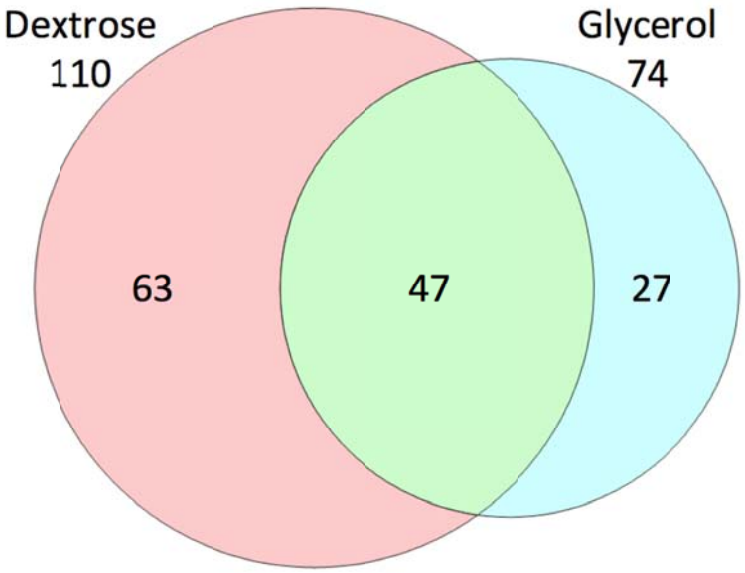


Figure 2-12. Proteins that were identified in one of three analyses with a confidence greater than 95% by Scaffold 3.3 are counted here. The multidimensional UPLC system utilizing a high pH first dimension and a low pH second dimension resulted in a greater number of identifications. The amount of time required for the multidimensional approach, however, was significantly more than required of the single-dimension analyses.

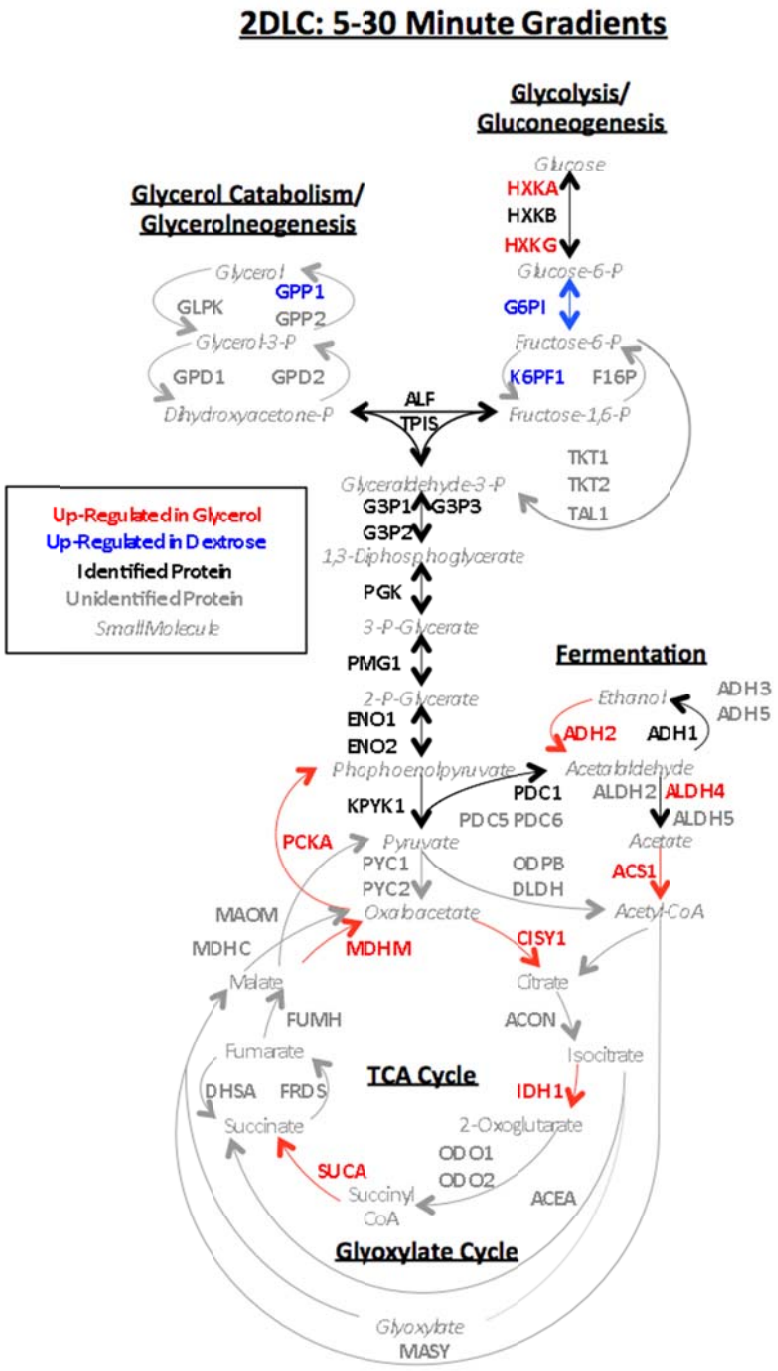


Figure 2-13. For the multidimensional peptide separation, the coverage of the metabolic pathways of interest was about the same as for the longer, single-dimension separations. The coverage was found to be 26 of the 53 proteins searched for.

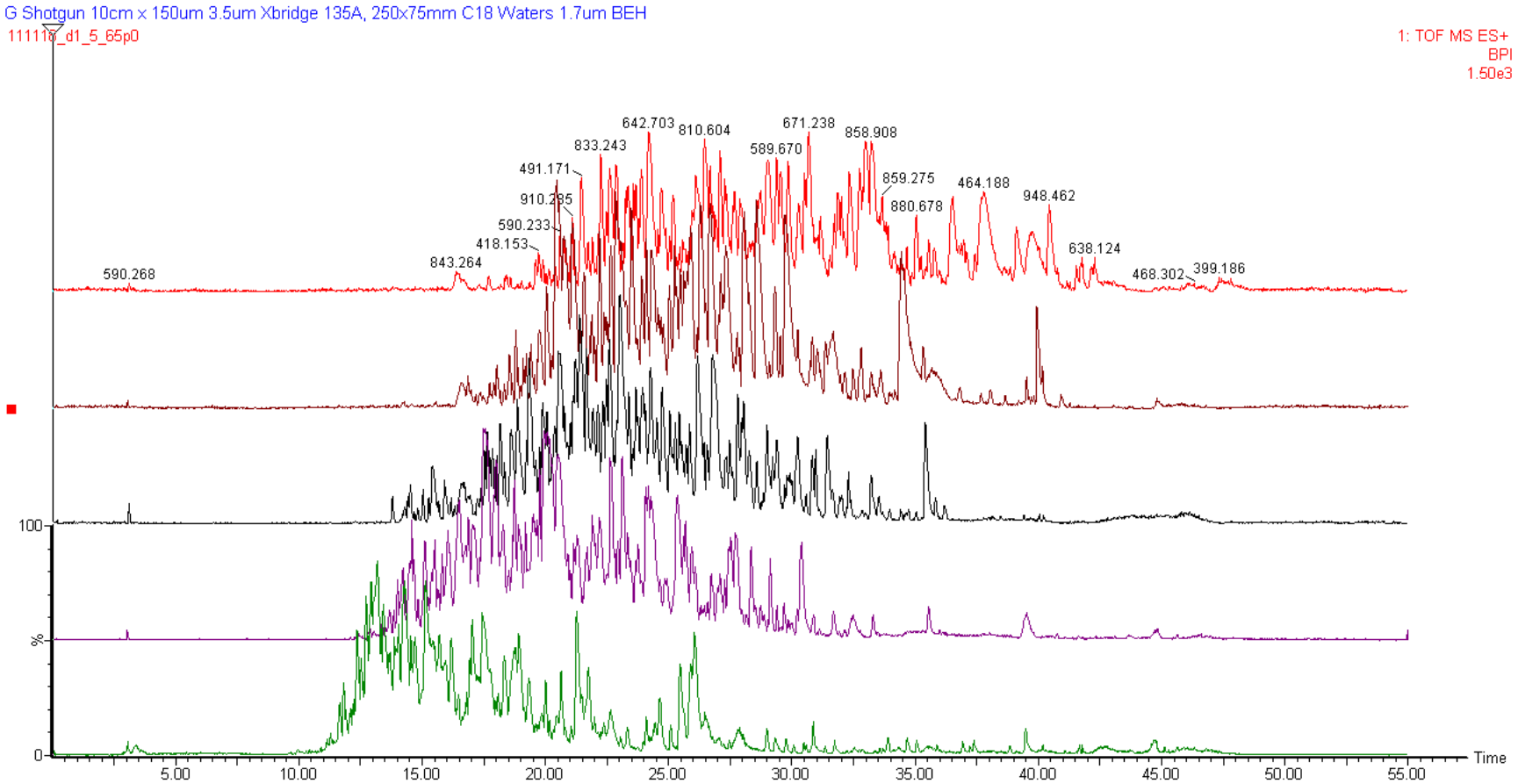


Figure 2-14. These chromatograms are from the multidimensional bottom-up analysis of the dextrose-based yeast sample. The chromatograms here show the most intense ion eluting at any given point. Overall intensities were found to be good with 1,500 counts being a reasonable value. There is a noticeable correlation between peptides that elute early in the first dimension also eluting early in the second dimension.

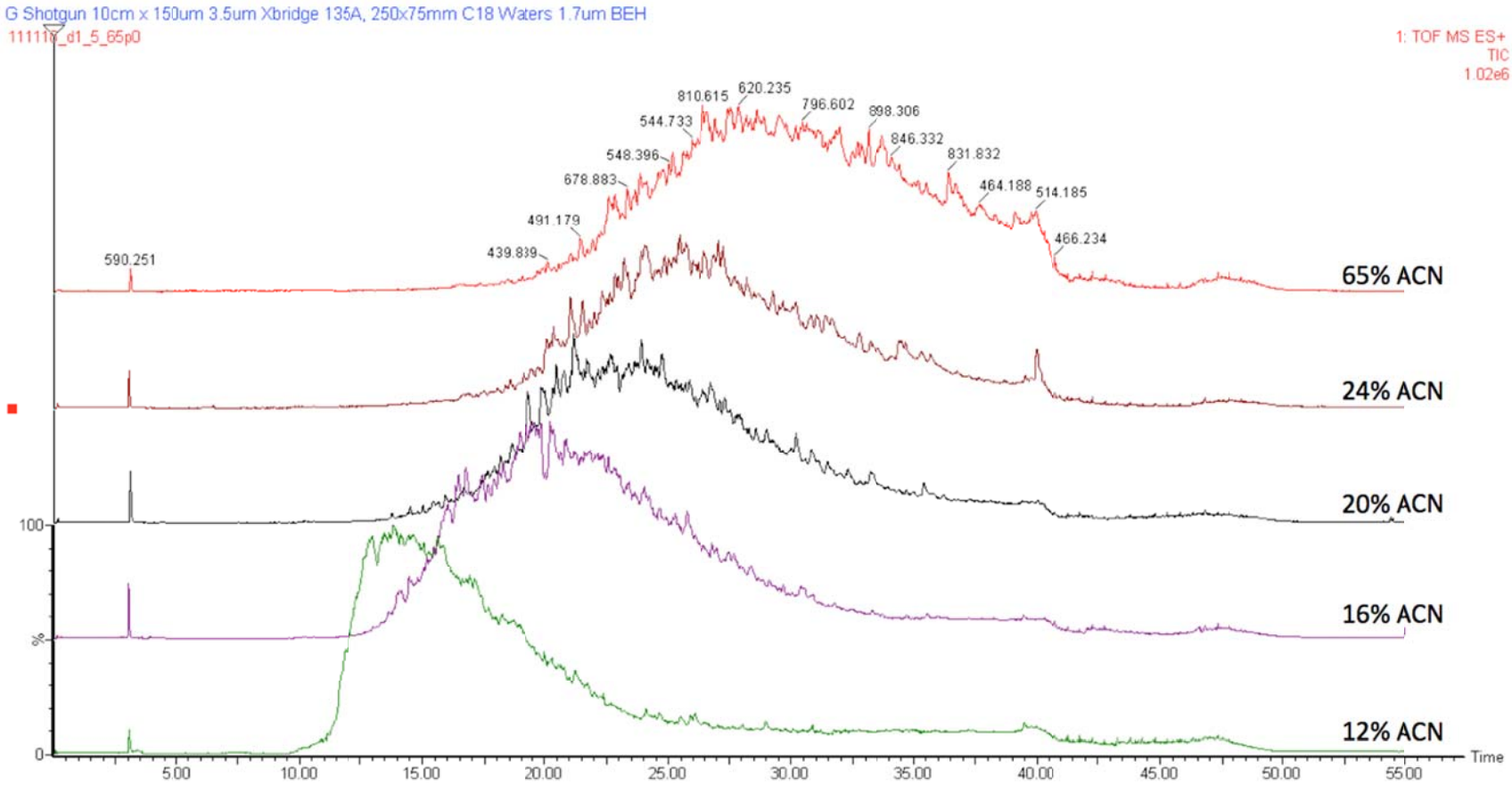


Figure 2-15. The chromatograms here are from the same analysis as in Figure 2-14. Here, the total ion current has been plotted, providing a clearer picture of the elution window and the correlation of peptides eluting early in the first dimension also eluting early in the second dimension, as shown in green in the bottom trace. The percentage acetonitrile used to elute peptides from the first dimension are shown for each fraction.

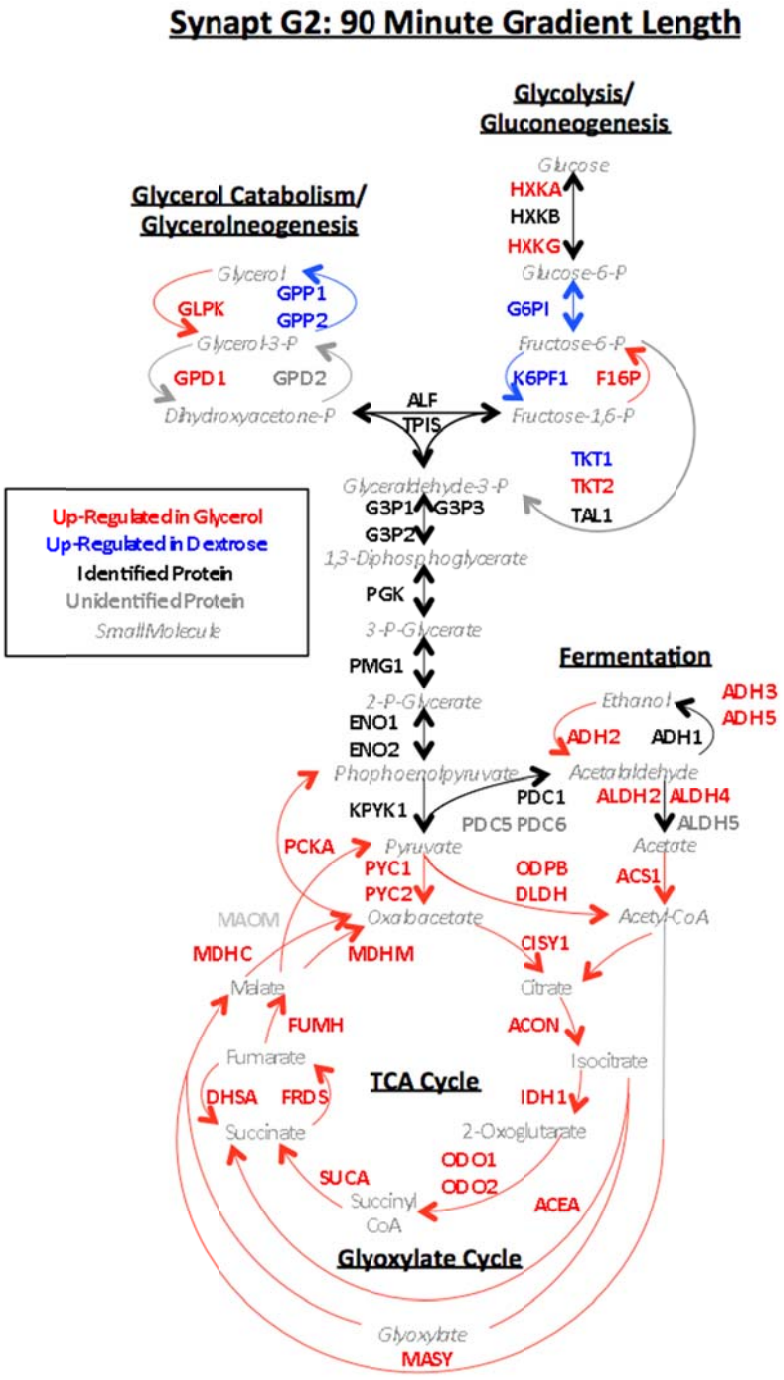


Figure 2-16. All of the other metabolic pathway figures were comprised of data collected on a Waters QTOF Premier mass spectrometer. For this analysis, the same conditions used for the 90-minute gradient conducted in the Jorgenson lab were used by the Duke Proteomics Core facility to obtain data on their Waters SynaptG2 HDMS mass spectrometer. The newer and more advanced mass spectrometer, which also utilized an ion mobility separation, was able to identify all but two of the proteins searched for.

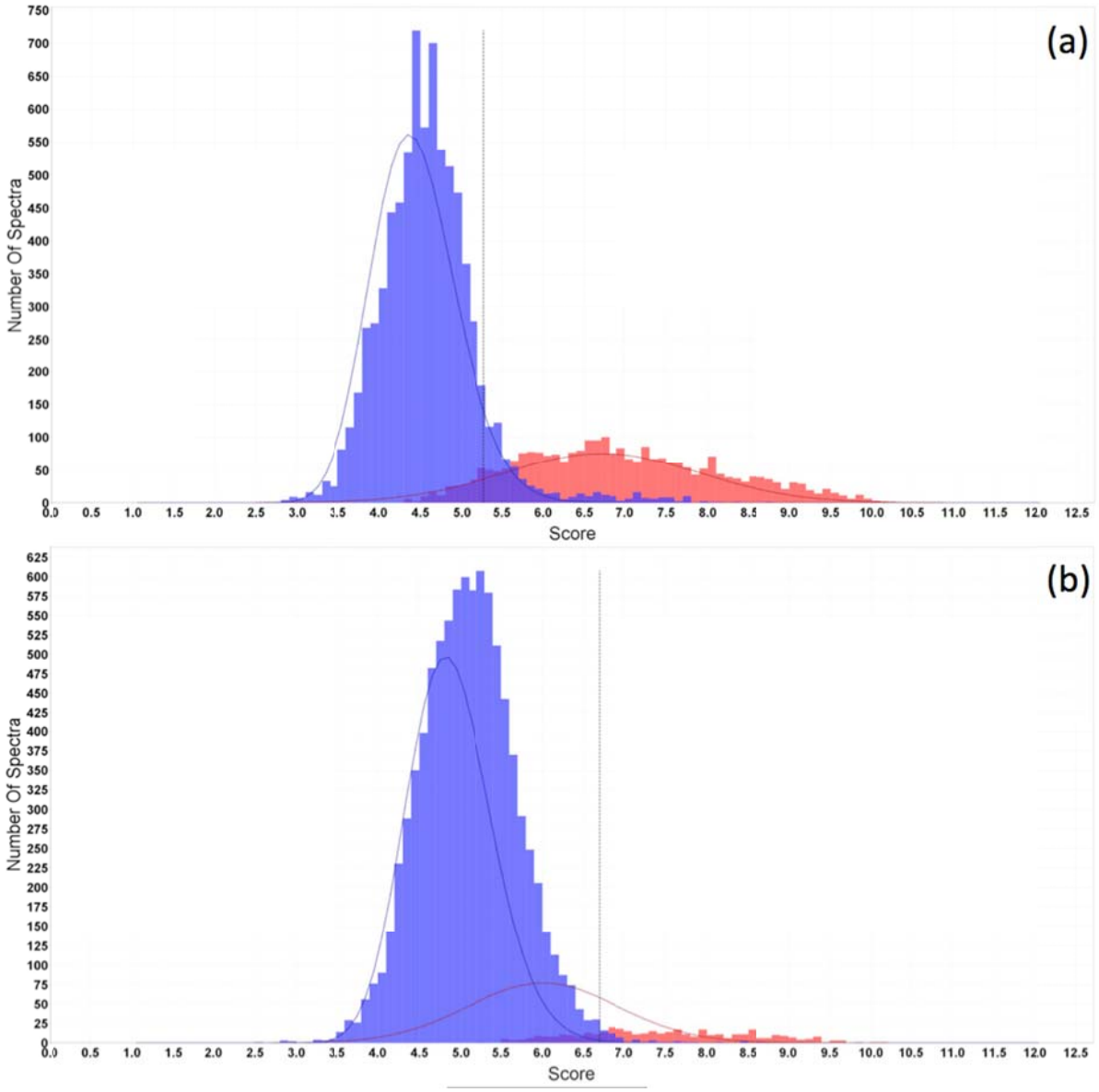


Figure 2-17. Scaffold 3.3 attempts to assign all peptides. Using the peptides score as part of this function, the software attempts to eliminate incorrectly assigned peptides from being used to identify proteins. Data was collected on two different mass spectrometers, (a) a SynaptG2 and a (b) QTOF Premier. There were a greater number of correctly assigned peptides from the SynaptG2 data and the differences in scores between the incorrect and correct populations were also greater. The improved data facilitated the additional identifications observed with the analysis completed by the more advanced mass spectrometer.

CHAPTER 3: PRE-FRACTIONATION OF DIFFERENTIALLY GROWN *S. CEREVISIAE* CELL LYSATES

3.1 Introduction

Complete bottom-up methods for proteomic analysis result in samples that are exceedingly complex with regard to the number of components to be analyzed. Further complicating the analysis of these samples is the wide-range of expression levels proteins are found at within the cell. In order to identify proteins of low abundance, larger injection volumes are needed in order to load enough material onto the column so that identifications are possible. However, one cannot simply load more sample in many situations without deleterious effects. Peptides in large abundance will be overloaded in the case of large injections, creating a situation where these peaks begin to saturate the column and detector and have pronounced tailing effects resulting in poor resolution of these components. Ultimately, the quality of the data from the mass spectrometer suffers resulting in less identifications.

Many different types of multidimensional separations have been employed to increase the peak capacity of proteomic samples in an effort to improve the number of identifications. The work in this chapter focuses on two pre-fractionation strategies where intact proteins are separated and the resulting fractions digested and analyzed in a bottom-up approach. Pre-fractionation allows for a more comprehensive analysis. Certainly there is an increase in analysis time over traditional bottom-up methods, but the increase in peak capacity is needed to more thoroughly analyze complicated proteomic samples. Initially, pre-fractionation was

completed by anion-exchange chromatography. This method was chosen as it had previously been used with some success in both online and offline methodology. The results of this pre-fractionation scheme were compared to a reversed-phase pre-fractionation approach to evaluate the effect of the chromatographic method as it applies to intact protein separations. In addition to the comparison between anion-exchange and reversed-phase pre-fractionation, the effect of the number of fractions collected during the intact protein fractionation was briefly explored as was applying a multidimensional separation at the peptide level resulting in three stages of chromatography for the overall analysis.

3.2. Experimental

3.2.1 Overview of Experimental Methods

A workflow for the experimental methods described here can be found in Figure 3-1. Intact protein samples of the cytosolic portion of a yeast cell lysate grown differentially were separated by gradient elution by anion-exchange and reversed-phase chromatography. For each method, 40 fractions were collected, lyophilized, and digested. All 40 fractions from the anion-exchange pre-fractionation approach were initially analyzed by UPLC-MS^E. They were later combined to form 20 fractions to allow for a comparison with the reversed-phase pre-fractionation method where only 20 fractions were analyzed due to analysis time considerations. By having an equal number of fractions, a more direct comparison between these methods was facilitated. Fractions from the reversed-phase approach were also subject to a 2D-UPLC-MS^E separation in addition the single-dimension of peptide analysis. All of the peptide data was processed by ProteinLynx Global Server 2.5 (PLGS 2.5). Data was then exported into Scaffold 3.3 for statistical analysis.

3.2.2 Chemicals and sample preparation

The chemicals used for the anion-exchange pre-fractionation mobile phases were ammonium acetate and ammonium hydroxide purchased from Fisher Scientific (Fair Lawn, NJ). Deionized water was purified using a Barnstead Nanopure System (Boston, MA). Mobile phases for the reversed-phase pre-fractionation contained Optima™ grade acetonitrile and isopropanol, and trifluoroacetic acid, also from Fisher Scientific. Chemicals used for the peptide analysis were Optima LC/MS™ grade water and acetonitrile, formic acid, and ammonium hydroxide. Chemicals used in the trypsin digestion were: ammonium bicarbonate, iodoacetamide and trifluoroacetic acid, purchased from Sigma (St. Louis, MO), RapiGest SF, an acid-labile surfactant provided by Waters Corporation (Milford, MA), dithiothreitol (Research Products International, Mt. Prospect, IL), and TPCK-modified trypsin (Pierce, Rockford, IL). The preparation of the samples used in this analysis was previously described in section 2.2.2.

3.2.3 Intact Protein Pre-fractionation

The intact proteins were separated by strong anion-exchange and reversed-phase chromatography. The first pre-fractionation separation was performed using anion exchange chromatography with a Waters BioSuite Q SAX, 7.5 mm ID x 75mm with 10μm particles and 1000 Å pores operated at 50°C. The mobile phase compositions and gradient profile is shown in Table 3-1. A linear gradient from 25 mM to 750 mM ammonium acetate, pH 9.0, was run from 0-60 minutes and held at 750 mM until 90 minutes. Two-minute wide fractions were collected from 2 to 82 minutes, and every other fraction pooled, yielding 40 fractions.

Reversed-phase method development utilized two columns. The first column tested was a Waters BEH C4 silica-based column containing 1.7 μm particles with 300 Å pores and

dimensions of 2.1 mm x 50 mm. This column was operated at 50 °C. A gradient from 15-50% acetonitrile with 0.2% TFA was employed for the separation. The second reversed-phase column tested and which was used to a much greater extent was a 4.6 mm ID x 250 mm PLRP-S column with 5 μm particles (Agilent, Santa Clara, CA). After an initial separation, the gradient was changed for improved resolution. Separation conditions for the final PLRP-S method can be found in Table 3-2. During fraction collection, one-minute wide fractions were collected from 2 to 42 minutes and every other fraction pooled, yielding 20, two-minute wide fractions. Fractions were stored at -80º C until lyophilization.

3.2.4 Protein digestion and Peptide LC-MS/MS

Fractions were transferred to microcentrifuge tubes, then lyophilized and reconstituted in 25 μL of 50 mM ammonium bicarbonate. Three microliters of 6.67% (w/v) RapiGest™ SF in buffer were added (15 min, 80 ºC) to denature the proteins. The proteins were reduced by adding 1uL of 100mM dithiothreitol (30 min, 60ºC), and then alkylated with 1uL of 200 mM iodoacetamide (30 min, room temperature, protected from light). The proteins were then digested by adding 10 μL of 320 ng / μL TPCK-modified trypsin in 50 mM ammonium bicarbonate (overnight, 37ºC). Once the digestion was complete, the RapiGest™ SF was degraded using 44 μL of 1% (v/v) trifluoroacetic acid (2 h, 60ºC). The fractions were centrifuged for 20 minutes at 14,000 x g to pellet the hydrolyzed surfactant, after which they were ready for analysis. The samples were transferred to LC vials and spiked with 4.21 μL of a 1 pmol/L internal standard Bovine serum albumin digest (Waters Corporation).

Each fraction was analyzed using capillary RPLC-MS/MS using a Waters nanoAcquity/QTOF Premier system. The separations were performed on a 250 mm x 75 μm

ID capillary column packed with 1.7 μm silica bridged-ethyl particles with a C18 stationary phase (Waters). Mobile phase A was Optima Grade water with 0.1% formic acid (Fisher), and mobile phase B was Optima-grade acetonitrile with 0.1% formic acid (Fisher). At a flow rate of 300 nL/min, a 90 minute gradient from 5-40% B was used to separate the peptides, followed by a 5 minute column wash at 85% B, after which the mobile phase was returned to 5% B. Two μL of sample was injected per run for an average column load of approximately 1 μg of digest protein per run . The outlet of the RPLC column was directly connected to an uncoated fused silica nanospray emitter with a 20 μm ID and a 10 μm tip (New Objective, Woburn, MA) operated at 2.7 kV. Data-independent acquisition or MS^E scans were performed and the instrument was set to perform parent ion scans from m/z 50-1990 over 0.6 sec at 5.0 V. The collision energy was then ramped from 15-40 V over the subsequent 0.6 sec scan.

3.2.5 Multidimensional LC/MS of peptides

Each of the fractions from the reversed-phase pre-fractionation study were also investigated by a high/low pH RP-RP-MS^E analysis. The first-dimension column was a 150 μm ID x 150 mm XBridge C18 column with 3.5 μm particles with 130 Å pores. Mobile phase A was 25 mM ammonium formate, pH 10.0. Mobile phase B was acetonitrile with no ion pairing agent present. The second-dimension column and analytical conditions were the same as described in the previous section. Sample loading was increased to 3 μL. The 2D with dilution method was operated with a 13:1 dilution factor being applied after the first dimension before the trap column of the second dimension using 99.5% water with 0.2% formic acid. This ensured that peptides eluting from the first dimension would be trapped for second dimension analysis. Peptides were eluted from the first-dimension in 5 fractions by

increasing the acetonitrile percentage for each step. Peptides were eluted in steps of 12%, 16%, 20%, 24%, and 65% acetonitrile.

3.2.6 Peptide data processing

The peptide LC-MS/MS data were processed using ProteinLynx Global Server 2.5 (Waters). The MS^E spectra were searched against a database of known yeast proteins from the Uni-Prot protein knowledgebase (www.uniprot.org) with a 1X randomized sequence appended to the end. After the database search was complete, the results were imported into Scaffold 3.3. (Proteome Software, Portland, OR). A Fisher's exact test was performed on the triplicate samples. Proteins with a p-value less than 0.050 between the two yeast samples and a fold change greater than 1.5 were considered to be differentially expressed.

3.3 Results & Discussion

Previous efforts in our group focused on a hybrid approach of top-down and bottom-up methodologies involving an online two-dimensional separation of intact proteins by strong anion-exchange chromatography followed by reversed-phase chromatography in an online format. While those efforts are not the focus of this work, it did serve as the starting point for the methods demonstrated here. Initially, pre-fractionation of intact proteins for bottom-up analysis was completed by strong-anion exchange chromatography using the same packing material as had previously been used for online analyses. A total of 40, 2-minute fractions were collected, digested, and analyzed for the anion exchange fractionation utilizing a shallow linear gradient. Initially, all 40 fractions were analyzed requiring more than 240 hours of analysis time per intact protein sample, as the digested fractions were run in triplicate. Subsequent to this analysis, it was determined that for the reversed-phase fractionation, every other fraction would be pooled after digestion in order to reduce the

analysis time in half, to a more manageable amount. Additionally, each intact protein separation was done three times to have replicates from the intact protein separation and not just for the peptide separations to more thoroughly evaluate the method. Each of the resulting 20 reversed-phase peptide samples were then analyzed by UPLC-MS^E in two ways. First, they were analyzed by a one-dimensional separation and then by a more comprehensive and time intensive two-dimensional approach. Lastly, each of the 40 anion-exchange fractions were pooled as in the case of the reversed-phase pre-fractionation approach in order to make a more straightforward comparison on the basis of gradient fractionation.

3.3.1 Chromatograms of Intact Protein Separations

Qualitative results from the UV data (Figure 3-2) of the first dimension provided good initial insight without the substantial investment in time required by the digestion and bottom-up analysis of each fraction. For both the dextrose- and glycerol-based yeast samples, the anion exchange separation exhibited a substantial absorption signal in the UV data at the dead time followed by a lull in elution until the ionic strength of the mobile phase had increased enough to begin to elute proteins from the column. The large signal occurring at the dead time was assumed to be proteins not retained by anion exchange chromatography. Fraction collection continued until all proteins were assumed to have been eluted owing to a lack of peaks in the UV data and a return to baseline conditions. The observed elution profile was not completely unexpected as some proteins will not be bound by the anion-exchange resin whereas the majority will and only elute once the gradient increases sufficiently in ionic strength to cause a change in the retention. Attempts were made to lower the ammonium acetate concentration in order to prevent proteins from eluting in the dead time, however, precipitation of the proteins prevented a further reduction in ionic strength and a balance had

to be maintained between starting conditions and sample solubility. An initial concentration of 25 mM ammonium acetate, pH 9.0, was determined to be an appropriate concentration given the constraints. While it is possible that a non-linear gradient could result in a more even elution profile, it was believed a further flattening of the gradient during the elution window would have only resulted in broader peaks without an increase in resolution. It is also possible that conditions could have been ramped more quickly to the point where proteins began to elute, however downstream analysis time would not have been significantly reduced. The reduction in fractions would have only been by one or two with the additional method development requiring a greater investment in time than the bottom-up analysis of these few additional fractions.

Reversed-phase chromatography was explored as an alternative to the anion exchange method. Greater resolution for the reversed-phase method was expected as reversed-phase chromatography typically yields higher resolution separations than ion exchange for proteins owing to the nature of the interactions. After exploring a few different column chemistries, a PLRP-S column, comprised of polystyrene-divinyl benzene particles, was selected for use for pre-fractionation of intact proteins by reversed-phase. It was chosen over silica-based materials due to the rugged nature of the material, allowing for temperatures up to a rated maximum of 200°C, which was advantageous when working with protein samples prone to precipitation. Columns with silica-based particles were considered for this separation and initially, a 2.1 x 50 mm ID XBridge column from Waters with 1.7 um C4 bridged-ethyl hybrid particles and 300 Å pores was tried as it is specifically marketed for protein separations because of the larger porse size and short alkyl chain length. A gradient of water and acetonitrile containing 0.2% TFA was used with a column temperature of 50 °C.

Problems were quickly encountered with use of this column as the sample precipitated on contact with the inlet frit, clogging this column. In contrast to the previous work with elevated temperature RPLC of intact proteins, water/acetonitrile gradients were found to be insufficient to prevent protein precipitation.[1] Isopropanol as an organic modifier has previously been reported to work well with more chromatographically challenging proteins because of its greater eluotropic strength and was therefore used in combination with the PLRP-S column.[2] The new mobile phase B contained 50% isopropanol and 50% acetonitrile with 0.2% TFA and the temperature was also increased to 90 °C as elevated temperatures have also been reported to improve the resolution of intact protein separations. After the sample was injected there was a slight increase in back pressure, however flow was not stopped and most proteins were observed eluting in a short window between 30 & 40 minutes (Figure 3-3). This served as the starting point for further method development, in order to better separate the proteins. The gradient used for fraction collection contained an initial hold to trap the proteins on the head of the column with a quick ramp to 25%, slightly below the gradient strength needed to elute the proteins. The gradient steepness was then reduced to allow for the protein separation and ramped at the end to remove any remaining proteins. This method provided for separations with almost no ghosting and reproducible results as observed by the intact protein separations of both the dextrose and glycerol-based yeast separations run in triplicate. The overlay of the replicates for both samples on the PLRP column are shown in Figure 3-4. A comparison of the samples by overlaying the UV data shows a similar elution profile with regard to intensity and overall chromatographic behavior, but definite differences are observed Figure 3-5.

Similarly to the anion-exchange method, a substantial signal was observed at the dead time in the UV data; however, this did not transition into a long delay and instead peaks were observed more consistently throughout the gradient. Furthermore, the resolution of the peaks was significantly improved in the reversed-phase separation with many more peaks observed. Using a 50:50 mixture of acetonitrile and isopropanol with 0.2% trifluoroacetic acid along with elevated temperatures resulted in no ghosting and a consistent back pressure between sample and blank runs. Lastly, fraction collection was continued well beyond the last observed peak and until the baseline stabilized in an effort to collect as many proteins as possible.

3.3.2 Proteins per fraction

The most straightforward comparison to make initially is between the two methods where the gradient was fractionated evenly for both intact protein pre-fractionation methods. It is therefore necessary to start the discussion by comparing the results of fractionating the anion-exchange and reversed-phase methods into 20 fractions. A more quantitative look at the pre-fractionation strategies was possible once all of the fractions had been digested and analyzed by UPLC-MS^E. In Figure 3-6, non-unique protein identifications per fraction were represented by the lighter shading in a bar graph for both separations. A non-unique protein was defined as any protein found within a given fraction, thus, if a protein were to be found in multiple fractions it would be counted in each fraction. The number of proteins found per fraction from the anion exchange pre-fractionation mimics the general trends observed from the UV data, with proteins found initially followed by a lull and then more identifications, as the gradient progressed. In fractions 1 and 2, 34 non-unique proteins were observed whereas the following three fractions contained a total of 25 non-unique protein hits. For the reversed-

phase separation there was again a large UV absorption signal initially, which did not drop off as quickly as for the anion-exchange separation, however the results from the protein identifications were much different with only 9 non-unique identifications in the first three fractions in total. Additional discrepancies between the UV and MS data can be observed, demonstrating that UV data is convenient for determining that the separation occurred without significant problems, but is not useful for any other conclusions. Another area of interest as it relates to the UV data and overall protein identifications occurs towards the end of fraction collection. In both pre-fractionation strategies, the UV trace decreased significantly over the last five fractions collected.

If only non-unique protein identifications were taken into consideration, shown in the more lightly shaded region of Figure 3-6, it would appear as though fraction collection was terminated prematurely in both cases and especially so for the reversed-phase separation, which is not suggested by the UV data. However, by also counting a protein only in the fraction containing its most intense spectral counts, as determined through Scaffold, it was shown that only a small number of proteins were uniquely identified in the last five fractions. Had fraction collection been stopped even earlier, reducing the number of bottom-up samples and therefore reducing overall analysis time, the overall productivity of the analysis as measured in protein identifications per hour, would have been improved. Specifically for the reversed-phase separation, fraction collection was terminated after 44 minutes as the baseline had stabilized and no peaks were observed after 32 minutes (fraction 15). Fractions 16-20 only contributed 33 more unique identifications for the reversed-phase separation and only 16 additional for the anion-exchange pre-fractionation at a cost of 30 hours analysis time per method including replicate analysis. In total, there were 784 and 345 unique protein

identifications from the reversed-phase and anion-exchange pre-fractionation methods respectively, which comprises the data observed in the unique identifications in Figure 3-6. Roughly 25% of these proteins were only identified in one of three replicate runs and were therefore excluded from any potential differential analysis. For the purpose of a differential comparison, a protein must have been identified with a confidence of 95% by Scaffold 3.3 in at least two of three replicate runs. With this stricter criterion, the total identifications are reduced to 546 and 262 for the reversed-phase and anion-exchange pre-fractionation methods, respectively. Of the 262 proteins identified by the anion-exchange method, 220 were found to be in common with the reversed-phase method and only 42 found only by anion-exchange pre-fractionation, while 326 were found solely by reversed-phase (Figure 3-7). Protein hits were counted as a unique identification if found with a confidence greater than 95% by Scaffold with a minimum of three peptides, each with a confidence greater than 95%.

By examining the data from the initial anion-exchange fractionation approach, where the 40 collected fractions were digested and each run individually, we can begin to explore the effect of gradient fractionation. Increased fractionation has the potential to reduce fraction complexity assuming the resolution of the separation mechanism being used, is sufficient enough to avoid splitting a given protein across many fractions. However, if fractionation is too frequent, then one could potentially expect a decrease in the number of identifications as the concentration of each protein would be significantly impacted during the digestion step and the peptides needed for identification would be also be spread at lower levels across many fractions. The trends observed from the 20 fraction anion exchange analysis are also observed for this 40 fraction separation with a substantial number of proteins being identified initially and followed by a lull in elution before the rest of the

proteins eluted (Figure 3-8). Overall, there were more unique protein identifications by this method. By doubling the number of fractions, 370 unique proteins were identified, an increase of 41%, with the added cost of doubling the analysis time from 120 to 240 hours, for replicate analysis. A Venn diagram demonstrating the overlap in identifications between the two anion-exchange methods can be observed in Figure 3-9, with 231 proteins being found in common resulting in only 31 proteins being found solely by the AXC20 (anion-exchange fractionation with 20 fractions) method. Finding additional proteins unique to the approach with fewer fractions is not unexpected. Because the peak capacity of the peptide separation is still inadequate to properly resolve all of the components despite pre-fractionation of the intact proteins, multiple peptides are eluting at any given time. This is true for both the 20 and 40 fraction approaches. As a result, occasionally a peptide will be better sequenced, seemingly at random, and will result in an identification. This result can also be observed by running the same method multiple times. While the majority of the proteins will be identified in all replicates, there will be some that are only identified in one or two replicates. A second consideration for the additional identifications coming from a method with fewer intact protein fractions is in the splitting of the protein peaks as they elute. A protein split between two fractions may result in peptides that fall below the limit of detection. By analyzing fewer fractions, there are some proteins which will no longer be split and therefore the peptides may now be closer to the limit of detection for the mass spectrometer. The AXC40 (anion-exchange fractionation with 40 fractions) approach with greater identifications overall, identified an additional 139 proteins not observed by the less comprehensive approach, albeit with increased analysis time. Overall, for these three methods compared, the reversed-phase pre-fractionation method provided for the greatest number of protein identifications

identifying a large majority of the proteins identified by either anion exchange fractionation scheme (Figure 3-10).

Another option for changing the fractionation scheme is to subject the peptides to a multidimensional separation, as was done in Chapter 2 for the differential shotgun analysis of yeast. Using the fractions collected from the reversed-phase pre-fractionation strategy, each of the 20 fractions from the glycerol pre-fractionation strategy was subjected to a five-fraction multidimensional analysis. The gradient used during the low pH second dimension for this method was 30 minutes in length, increasing the total gradient time per intact protein fraction from 90 minutes to 150 minutes. However, the total analysis time per intact protein fraction increased from just over two hours to nearly 6 hours as a result of the steps needed to re-equilibrate columns. Only the glycerol sample was analyzed by this method and was only run once in order to evaluate the merits of this approach. Triplicate analysis of the RP20 (reversed phase fractionation with 20 fractions) method required 120 hours identifying 415 proteins in they glycerol sample. This is an equivalent amount of time needed for a single analysis by this multidimensional peptide approach. Triplicate analysis for the 2DLC method required 360 hours of analysis time. Only 557 proteins were identified by fractionating the peptides five times from the first dimension separation in the 2D-RPLC-RPLC method – an increase of only 142 new identifications (25%) -- despite a substantial increase in the amount of time required (300%) (Figure 3-11). The time during the analysis during which a gradient was running and therefore there was potential for peptides to elute for a single 2D analysis was 150 minutes as opposed to the 90 minutes of gradient time allowed for the 1D method. This results in an increase of gradient time of only 67% as the majority of the time, 210 minutes, was spent re-equilibrating the system and collecting fractions from the first

dimension. Given the modest increase in gradient time, the results are not unexpected. Further examination of the chromatograms also shows that the effective use of the available gradient space was poor, calling into question the degree of orthogonality of the method (Figure 3-12). Additional method development to yield better separations for each step is possible; however, flattening the gradient significantly enough to elute the peptides from the first step-wise elution across the entire 30 minute gradient window would not likely result in a significant improvement in resolution, but broader peaks.

3.3.4 Fractions per protein

In order to further evaluate the efficiency of the fractionation, we looked at the number of fractions a unique protein was found in and plotted the results for all unique proteins from both the anion-exchange and reversed-phase 20 fraction pre-fractionation approaches. Looking at raw numbers, there were more unique proteins found in all cases by the reversed-phase method. Given that there were more than twice as many proteins identified by this method, this is not surprising. When normalized, however, the fractions per protein results are consistent across both methods as shown in Figure 3-13. Since the majority of proteins, 60% an 61% for reversed phase and anion exchange, respectively, are being found in only one fraction, this indicates that the width of each fraction was more than sufficient to capture the entire peak as it eluted. Inevitably, proteins will elute during a fraction change resulting in a significant number, 20% & 18%, being found in two fractions. Somewhat surprising was that 20% of the identified proteins for each pre-fractionation method were found spread across three or more fractions, indicating that these proteins were eluting over a larger portion of the gradient demonstrating the difficult nature of intact protein separations. One possible explanation for this is differences in post-translational

modifications, which would change the retentivity of the proteins. This would result in peptides used in the identification of a protein being spread across many fractions and would require a detailed look at the peptides and associated modifications, if they were present on the peptides used in the identification, to determine any differences.

The results differ only slightly when looking at the other two methods of analysis – the AXC40 and multidimensional peptide approaches. With finer fractionation, it was expected that proteins would be split more frequently among the fractions and this was case with a reduction in the percentage of proteins identified in only one fraction from 61% to 41%. This was associated with an increase in the number of proteins identified in two fractions from 18% to 28%. While the additional fractionation of the anion-exchange gradient caused more proteins to be split across fractions, an increase in peak splitting was also expected to be detected by the multidimensional peptide separation. This approach has previously been reported to yield detection of proteins at lower levels, increasing the probability that there would be an increase in the number of fractions a protein was identified in. However, when comparing the two peptide analysis methods, one- and two-dimensional UPLC, there was only a slight shift towards a greater number of fractions per protein (Figure 3-14) with the multidimensional peptide method.

3.3.5 Molecular weight distribution of identified proteins

The distribution of proteins identified based on molecular weight was another area of interest for comparison between the two modes of pre-fractionation. When plotted as a histogram with 10 kDa bins from 0-230 kDa, the typical distribution of proteins with the majority of the proteins identified between 10 kDa and 40 kDa was observed (Figure 3-15). No significant bias was observed as the median molecular weight of proteins identified by

the reversed-phase and anion-exchange pre-fractionation methods was determined to be 38.0 kDa and 37.6 kDa, respectively. Another way to represent this data is to plot the mass of a protein against the fraction in which it eluted for all of the proteins identified in order to better visualize the results. This provides for a qualitative inspection lending insight into how well the gradient space of the first-dimension pre-fraction method is being utilized as well as the molecular weight range of identified proteins. The substantial increase in proteins identified by the reversed-phase pre-fractionation approach is most evident between10 kDa and 40 kDa and can be observed in Figure 3-16. One can also see the increase in the number of proteins found by reversed-phase pre-fractionation around the 100 kDa mark. With regard to the chromatography, the same general elution patterns can be discerned as observed in the UV data, although a drift in molecular weight towards larger proteins was observed between fractions 4 and 15 for the reversed-phase separation. Median values for the mass of proteins identified by the AXC40 and two-dimensional peptide analysis of the RP20 fractions was in agreement with the other two methods with 37.7 kDa and 37.2 kDa for the AXC40 and RP20-2D (reversed phase fractionation with 2D peptide analysis), respectively.

3.3.6 Differential comparison for RP-20 & AXC-20 fractionation strategies

Three of the methods -- AXC20, RP20, and AXC40 -- analyzed both samples in triplicate to allow for a differential comparison of the number of proteins identified per sample and the overlap between samples demonstrated in the Venn diagrams. The RP20 method exhibited better overlap between the dextrose and glycerol yeast samples than either of the anion exchange methods with 22% and 34% of the proteins being unique to the dextrose and glycerol yeast samples, respectively (Figure 3-17). The AXC20 method, which will be the other data set used for the differential comparison below, was found to have a

substantially increased number of proteins that could potentially be considered as differential proteins (Figure 3-18). Nearly half of the proteins identified in each sample were uniquely identified. A similar trend was observed for the AXC40 method (Figure 3-19). Upon closer examination of the differential proteins, it was observed that for many of these identifications, the intensity of these proteins was low, with only a few spectral counts. For proteins considered to be differentially expressed, one must question the extent of the differential expression for proteins with low spectral counts being identified solely in the glycerol- or dextrose-based samples. This is because at the low intensity levels reported, it is possible that for the sample where a given protein was not identified, the limit of detection for the analysis may play a greater role in the differential classification than what is the result of differences in the biological processes. Impacting the limit of detection for the anion-exchange pre-fractionation strategies is the problem of protein recovery. With proteins likely precipitating on the inlet frit of the anion-exchange column, a smaller amount of the sample is making it through the system. With less protein in the fractions there would be fewer peptides and therefore the limit of detection for the entire analysis would be adversely affected.

Of the proteins involved in the metabolic pathways of interest, found to be differentially expressed, clear trends emerged with regard to the effectiveness of intact protein pre-fractionation by the reversed-phase method in comparison to the anion-exchange separation. Table 3-3 lists many of these proteins with the average number of peptides used in the identification as well as the average coverage percentage as determined by Scaffold for both the dextrose and glycerol samples. For the dextrose sample, 21 peptides were found per protein on average for the differentially expressed proteins fractionated by reversed phase and only 9 for the anion-exchange pre-fractionation strategy with the coverage being 49%

and 24%, respectively. A similar trend was observed for the glycerol sample with 27 peptides and 59% coverage for the reversed-phase method while only 13 peptides and 36% coverage on average from the anion exchange method for the differential identifications.

Literature shows that growing yeast on different carbon sources effects expression of metabolic enzymes. The differential protein intensities measured by the methods described here follows the trends in protein expression predicted by the literature.[3] A stronger correlation exists between the literature and the results from the reversed-phase pre-fractionation method than with the anion-exchange pre-fractionation method due to the greater number of protein identifications as well as greater intensity as determined by spectral counts with the reversed-phase pre-fractionation method. Several metabolic pathways are altered when yeast is grown on glycerol versus its preferred carbon source, dextrose. Glycolysis, gluconeogenesis, glycerol catabolism, glycerolneogenesis, fermentation, tricarboxylic acid cycle (TCA), glyoxylate cycle, and pentose phosphate pathway will be discussed in more detail (Figure 3-20).

Glycolysis is the digestion of glucose to pyruvate, which can then be converted into energy through the TCA cycle, glyoxylate cycle, or fermentation. When glucose is not present, the cell undergoes gluconeogenesis, which is the formation of glucose from pyruvate. Many of the same proteins are involved in glycolysis and gluconeogenesis. As expected, there is no general trend in the differential expression of these proteins as they are active in yeast grown on both media sources. The proteins involved in both these processes were identified by both methods. The one exception is F16P which transforms Fructose-1,6-P to Fructose-6-P during gluconeogenesis while K6PF1 is involved in the reverse reaction during

glycolysis. F16P protein was found to be up-regulated in the reversed-phase fractions of the glycerol sample, but this protein was not detected in the anion-exchange fractions.

Analogous to glycolysis and gluconeogenesis are glycerol metabolism, which converts glycerol into pyruvate, and glycerolneogenesis, which converts pyruvate into glycerol. For the yeast grown on glycerol, it is predicted that the proteins used in glycerol catabolism such as GLPK and GPD1 would be up-regulated[4,5] while the proteins used in glycerolneogenesis, such as GPP1 and GPP2, would be down-regulated.[6] This phenomenon is observed in the reversed-phase fractions, but these proteins were not detected in the anion exchange fractions.

After its biogenesis, pyruvate is fermented into ethanol if there is an excess amount of glucose present. A protein complex is formed by PDC1, PDC5 and PDC6.[7] This complex is involved in the conversion of pyruvate to acetaldehyde during fermentation. These three subunits were detected in the reversed-phase fractions with PDC5 and PDC6 being up-regulated in the dextrose sample. Only PDC1 was identified by anion exchange, and it was similarly determined to be up-regulated in the dextrose sample. In the absence of glucose, pyruvate enters the TCA or glyoxylate cycle.[8] Additionally, any ethanol that may be present is converted into acetate to be used in the TCA or glyoxylate cycle.[9,5, 10,11,12]

In these pathways, 49 of 52 proteins were identified in the reversed-phase fractions with 28 being up-regulated in the glycerol sample, 6 up-regulated in the dextrose sample, and 15 identified in both, but not differentially. In the anion-exchange fractions, only 37 of 52 proteins were identified with 14 being up-regulated in the glycerol sample, 3 in the dextrose sample, and 20 identified in both, but not differentially.. Fewer differential proteins were observed in the anion-exchange fractions because many of the proteins were lower in

intensity and the run-to-run variation was great enough to obscure classification as a differential protein. The metabolic pathways for yeast grown on glycerol occur in the mitochondria while fermentation occurs in the cytoplasm. When looking at the data from the reversed-phase pre-fractionation method, our results show that out of 100 mitochondrial proteins identified, 37 were up-regulated in the glycerol sample while only 16 were up-regulated in the dextrose sample.

The pentose phosphate pathway takes glucose and transforms it into ribulose-5-phosphate which is used in the formation of ribosomes and subsequently DNA synthesis. Cells grown in a glucose deficient environment will break down the ribulose-5-phosphate to be used in the TCA cycle. Therefore, an abundance of ribosomal proteins should exist in the yeast grown on dextrose. Again, using the data from the reversed-phase pre-fractionation, the results indicate 32 out of the 55 ribosomal proteins identified were up-regulated in the dextrose sample while only 3 were up-regulated in the glycerol sample.

3.4 Conclusions

Martosella et al, explored recovery from reversed-phase columns as a function of temperature and determined that elevated temperatures, 80°C, were necessary for good recovery. While the reversed-phase method utilized a PLRP-S column operated at 80°C, the anion-exchange method utilized a BioSuite Q SAX column with an upper temperature limit of 50°C, which was employed for the fractionation. With our particular cell lysate, it was found that the inlet frit of the anion exchange columns would clog after repeated injections, creating a substantial increase in back pressure requiring routine cleaning to maintain them and it is likely that a closer look at the anion exchange pre-fractionation would confirm recovery as being an issue with this method of pre-fractionation. Clogging of inlet frits was

not found to be problematic for the PLRP-S column with no increase in back pressure observed after repeated injections. Taking all of this into consideration, protein recovery was most likely critically important to the improved performance of the reversed-phase method over anion-exchange. An interesting note regarding the inlet frits was determined after the pre-fractionation of the intact protein samples had occurred and the peptide workflow was well on its way. The Waters BEH300 C4 column was re-examined briefly, and it was found that by diluting the sample by at least 50% in 80% formic acid, the sample would remain soluble long enough to proceed past the inlet frit of this column. This sample preparation step also allowed for subsequent blank runs to remain largely clear of ghosting proteins when utilizing this column.

By fractionating intact proteins, more of the resultant peptides for a given protein are contained in fewer fractions as opposed to multidimensional separations at the peptide level, providing greater confidence in identifications. Fractionation of intact proteins by either anion-exchange or reversed-phase chromatography allowed for greater peak capacity in an environment where there is a need as there are approximately 5,500 proteins encoded by the yeast genome.[13] The reversed-phase method was determined to be particularly powerful as temperatures in excess of 80°C were utilized to improve recovery and peak shape and greater recovery allowed for greater proteome coverage, which enabled the differential comparison. Many of the proteins involved in the metabolic pathways investigated within this work were only found by the reversed-phase method. These methods also allowed for milligram quantities of protein to be analyzed as opposed to μg quantities common to gel electrophoresis-based methods without overloading the first dimension of separation. Greater sample loading allowed for greater flexibility with peptide analysis as there was a greater

volume of sample available for analysis as well as a greater amount of peptides from less abundant proteins, allowing their identification. More importantly, however, is that the ability to load a greater amount of protein allows for the possibility of finding proteins of lower abundance.

With over 120 hours needed to complete the analysis for a single sample and twice that in order to make a differential comparison, this method does require a significant investment in time. A parameter to be explored in the future and one that is of interest is the extent of fractionation required in order to generate useful data. The effect of fractionation has been studied previously by many groups and will continue to be studied as there is always a balance to be maintained between analysis time and proteome coverage.[14] If proteins of interest could be identified and quantified by only splitting the intact protein separation into 10 fractions, then this would be of great benefit to the investigator. Alternatively, owing to the improved resolution of the reversed-phase method, one could arguably benefit from increased fractionation, depending on the goals of the investigator. What is likely a more prudent method to explore than applying the same fractionation scheme to the entire gradient, is to base the fractionation off of the UV data. Areas where there is significant signal intensity in the UV data could be more finely fractionated and the end of the gradient could be relegated to one large fraction, as there were few new protein identifications made during this time, yet a significant amount of analysis time was required when evenly fractionating the gradient.

3.5 References

1. Martosella, J.; Zolotarjova, N.; Liu, H.; Nicol, G.; Boyes, B. E., Reversed-Phase High-Performance Liquid Chromatographic Prefractionation of Immunodepleted Human Serum Proteins to Enhance Mass Spectrometry Identification of Lower-Abundant Proteins. *Journal of Proteome Research* **2005**, *4* (5), 1522-1537.

2. Dillon, T. M.; Bondarenko, P. V.; Rehder, D. S.; Pipes, G. D.; Kleemann, G. R.; Speed, R. M., Optimization of a reversed-phase high-performance liquid chromatography/mass spectrometry method for characterizing recombinant antibody heterogeneity and stability. *J. Chromatogr., A* **2006**, *1120*, 112-120.

3. Fraenkel, D. G., Carbohydrate metabolism [in yeast]. *Cold Spring Harbor Monogr. Ser.* **1982**, *11B*, 1-37.

4. Sprague, G. F.; Cronan, J. E., Isolation and characterization of Saccharomyces cerevisiae mutants defective in glycerol catabolism. *Journal of Bacteriology* **1977**, *129* (3), 1335-1342.

5. Grauslund, M.; Lopes, J. M.; Rønnow, B., Expression of GUT1, which encodes glycerol kinase in Saccharomyces cerevisiae, is controlled by the positive regulators Adr1p, Ino2p and Ino4p and the negative regulator Opi1p in a carbon source-dependent fashion. *Nucleic Acids Research* **1999**, *27* (22), 4391-4398.

6. Cronwright, G. R.; Rohwer, J. M.; Prior, B. A., Metabolic Control Analysis of Glycerol Synthesis in Saccharomyces cerevisiae. *Applied and Environmental Microbiology* **2002**, *68* (9), 4448-4456.

7. ter Schure, E. G.; Flikweert, M. T.; van Dijken, J. P.; Pronk, J. T.; Verrips, C. T., Pyruvate Decarboxylase Catalyzes Decarboxylation of Branched-Chain 2-Oxo Acids but Is Not Essential for Fusel Alcohol Production by Saccharomyces cerevisiae. *Applied and Environmental Microbiology* **1998**, *64* (4), 1303-1307.

8. Brewster, N. K.; Val, D. L.; Walker, M. E.; Wallace, J. C., Regulation of Pyruvate Carboxylase Isozyme (PYC1, PYC2) Gene Expression in Saccharomyces cerevisiae during Fermentative and Nonfermentative Growth. *Archives of Biochemistry and Biophysics* **1994**, *311* (1), 62-71.

9. Pavlik, P.; Simon, M.; Schuster, T.; Ruis, H., The glycerol kinase (GUT1) gene of Saccharomyces cerevisiae : cloning and characterization. *Current Genetics* **1993**, *24* (1), 21-25.

10. Blumberg, H.; Hartshorne, T. A.; Young, E. T., Regulation of expression and activity of the yeast transcription factor ADR1. *Molecular and Cellular Biology* **1988**, *8* (5), 1868-1876.

11. Kratzer, S.; Schüller, H.-J., Transcriptional control of the yeast acetyl-CoA synthetase gene, ACS1, by the positive regulators CAT8 and ADR1 and the pleiotropic repressor UME6. *Molecular Microbiology* **1997**, *26* (4), 631-641.

12. Simon, M.; Binder, M.; Adam, G.; Hartig, A.; Ruis, H., Control of peroxisome proliferation in Saccharomyces cerevisiae by ADR1, SNF1 (CAT1, CCR1) and SNF4 (CAT3). *Yeast* **1992**, *8* (4), 303-309.

13. Kellis, M.; Patterson, N.; Endrizzi, M.; Birren, B.; Lander, E. S., Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **2003**, *423* (6937), 241-254.

14. Gu, H.; Huang, Y.; Carr, P. W., Peak capacity optimization in comprehensive two dimensional liquid chromatography: A practical approach. *Journal of Chromatography A* **2011**, *1218* (1), 64-73.

3.6 Tables

Time (min)	25mM Ammonium Acetate, pH 9.0 (%A)	1M Ammonium Acetate, pH 9.0 (%B)
0	100	0
60	0	100
90	0	100
90.1	100	0

Table 3-1. Gradient program used for the intact protein pre-fractionation of the yeast cell lysates by anion-exchange chromatography. The column used was a Waters BioSuite Q SAX, 7.5 mm ID x 7.5 cm with 10 µm particles and 1000 Å pores. The flow rate was 0.5 mL/min and the column was held at a temperature of 50 °C, the maximum per manufacturer recommendations. Forty fractions were collected from 2 to 82 minutes.

Time (min)	90:5:5 H_2O:ACN:IPA + 0.2% TFA (%A)	50:50 ACN:IPA + 0.2% TFA (%B)
0	100	0
2	100	0
25	95	5
40	50	50
45	35	65
45.1	0	100
50	0	100
50.1	100	0

Table 3-2. Gradient program used for the intact protein pre-fractionation of the yeast cell lysates by reversed-phase chromatography. An Agilent PLRP-S, 4.6 mm x 25 cm column with 5 µm pores particles and 300 Å pores. The flow rate was 1.0 mL/min. A column temperature of 80 °C was used – well below the rated maximum of 200 °C. Fractions were collected every minute from 2 to 42 minutes for a total of 40 fractions collected. These fractions were later pooled to form the 20 fractions analyzed.

	Dextrose				Glycerol			
	Peptides		% Coverage		Peptides		% Coverage	
Entry	AXC	RP	AXC	RP	AXC	RP	AXC	RP
K6PF1	1.0	26.3	1.6	37.3	0.7	26.7	1.3	34.3
6PGD1	1.0	32.7	2.9	66.0	7.0	38.0	20.0	77.7
ACS1	7.3	-	12.3	-	13.3	38.0	24.0	52.0
ACON	-	8.3	-	13.7	5.7	34.3	11.9	48.7
ADH1	16.7	25.3	54.0	66.0	17.7	22.3	53.7	65.0
ADH2	13.7	19.0	30.0	40.7	28.3	38.3	63.0	71.3
ADH3	-	2.7	1.7	7.2	1.3	4.0	5.7	13.8
ADH6	1.3	2.7	6.9	14.6	1.7	1.0	8.6	6.1
ALDH5	-	-	-	-	0.7	3.7	3.9	0.5
CISY1	-	-	-	-	16.7	26.7	41.0	54.0
DLDH	2.0	7.3	6.0	23.5	2.0	37.3	7.4	77.3
ENO1	13.3	20.7	42.0	78.7	15.0	24.7	67.3	78.0
ENO2	35.7	54.0	57.0	88.3	44.0	51.3	70.3	85.3
ALF	3.3	30.7	6.9	76.7	9.7	32.7	21.3	71.0
FUMH	-	1.3	-	3.7	10.3	18.0	34.0	48.0
FRDS	3.7	10.7	13.0	33.0	10.0	7.7	28.3	22.3
HXKG	0.7	2.7	1.9	7.8	-	39.0	1.0	71.3
G6PI	7.0	36.3	19.3	69.0	30.0	37.3	56.3	69.3
G3P1	10.0	29.0	23.7	52.3	25.3	36.0	63.0	81.0
G3P2	11.7	24.0	44.3	84.0	13.3	22.3	58.7	85.7
G3P3	8.0	10.7	71.3	88.0	9.7	10.3	81.7	91.7
HXKA	4.0	19.0	11.7	54.0	4.7	35.3	13.0	79.3
HXKB	9.0	33.3	25.7	79.3	10.0	29.3	34.7	81.0
ACEA	1.7	-	4.0	-	4.0	11.7	8.7	30.3
ALDH4	12.3	1.7	24.3	4.0	36.3	54.7	66.0	80.7
MDHC	3.3	-	11.3	2.3	10.7	16.3	35.3	62.7
MDHM	7.7	6.0	38.0	31.0	21.3	24.0	70.3	71.3
PCKA	1.0	-	2.5	-	1.7	36.7	3.7	60.7
PGK	40.0	53.3	77.0	87.0	38.3	49.3	79.3	82.7
PMG1	13.0	27.0	46.7	80.0	17.3	22.3	66.7	75.3
PYC1	4.7	6.0	5.7	9.6	6.3	15.0	8.7	19.3
PDC1	12.3	53.0	19.3	80.0	12.0	47.3	24.3	71.0
ODPB	1.0	2.7	4.1	8.1	4.7	4.3	14.3	16.3
KPYK1	32.0	62.0	76.0	87.7	32.3	57.3	75.7	84.0
SUCA	2.7	13.7	15.1	52.7	15.0	21.7	61.7	82.0
SUCB	-	-	-	-	2.0	23.7	6.9	46.3
TKT1	3.0	22.7	7.4	42.7	3.7	24.3	8.9	42.7
TKT2	-	-	-	-	1.7	5.0	3.1	12.0
TPIS	11.7	21.3	52.0	85.0	19.7	19.7	86.3	84.7
Average:	9.2	21.5	24.7	48.6	13.3	26.9	35.6	58.6

Table 3-3. The number of peptides used in an identification as well as the coverage percentage of the proteome is shown for both anion-exchange and reversed-phase pre-fractionation of proteins determined to be involved in the metabolic pathways of interest and differentially expressed. Overall, there were more peptides found and a larger coverage percentage for the reversed-phase pre-fractionation.

3.7 Figures

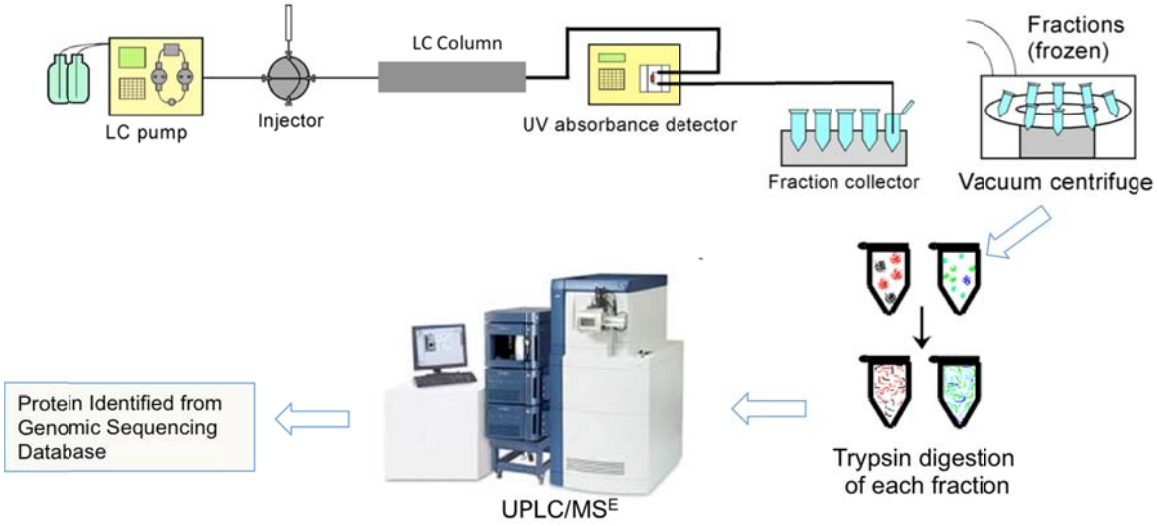


Figure 3-1. Experimental workflow for the intact protein pre-fractionation and bottom-up analysis of each digested fraction. An LC pump generated a gradient, which began the specified gradient program once the injection occurred. A Waters 2487 UV detector at 280 nm recorded data for the intact protein separations. Fractions were then lyophilized and digested with trypsin. Fractions were then analyzed by either 1D or 2D-UPLC/MS^E using a Waters nanoAcquity UPLC system with mass spectrometric detection by a Waters Q-TOF Premier mass spectrometer. The data was then processed through ProteinLynx Global Server 2.5 and Scaffold 3.3.

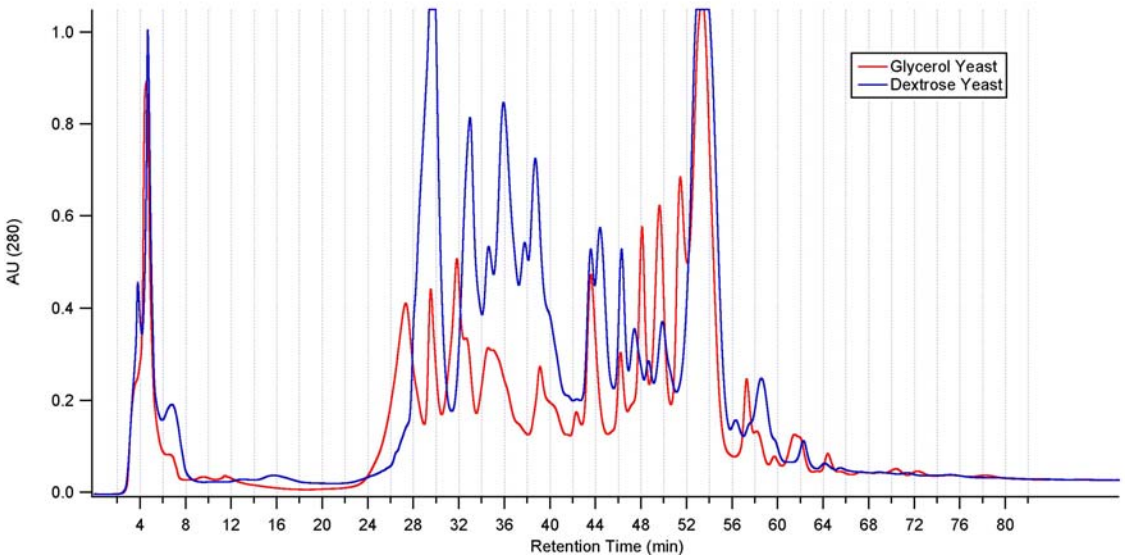


Figure 3-2. Anion-exchange pre-fractionation of yeast cell lysates differentially grown on dextrose and glycerol-based media. Approximately 4 mg of each sample was inject for analysis. The column used was a Waters BioSuite Q SAX, 7.5 mm ID x 7.5 cm with 10 µm particles and 1000 Å pores. The flow rate was 0.5 mL/min and the column was held at a temperature of 50 °C, the maximum per manufacturer recommendations. A linear gradient from 25 mm ammonium bicarbonate, pH 9.0 to 750 mM ammonium bicarbonate, pH 9.0, was run from 0-60 minutes with an additional 30 minute hold at 750 mM ammonium bicarbonate. Forty fractions were collected from 2 to 82 minutes. For the anion-exchange pre-fractionation, the intact protein separation was only done once.

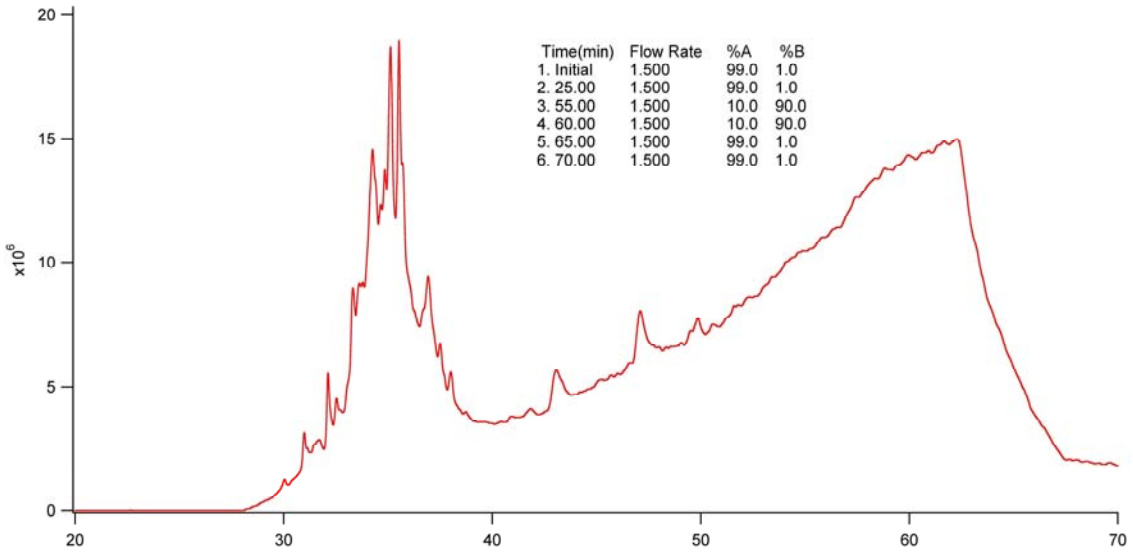


Figure 3-3. The initial separation of the dextrose yeast cell lysate on the PLRP-S column, an Agilent PLRP-S, 4.6 mm x 25 cm column with 5 µm particles and 300 Å pores. Mobile phase A was comprised of 80% water, 10% acetonitrile and 10% isopropanol plus 0.2% TFA. Mobile phase A was a 50/50 mixture of acetonitrile and isopropanol with 0.2% TFA added. The temperature was held at 90 °C. This gradient was substantially steeper than the final gradient conditions used for the pre-fractionation, but demonstrated the utility of this column and chromatographic conditions.

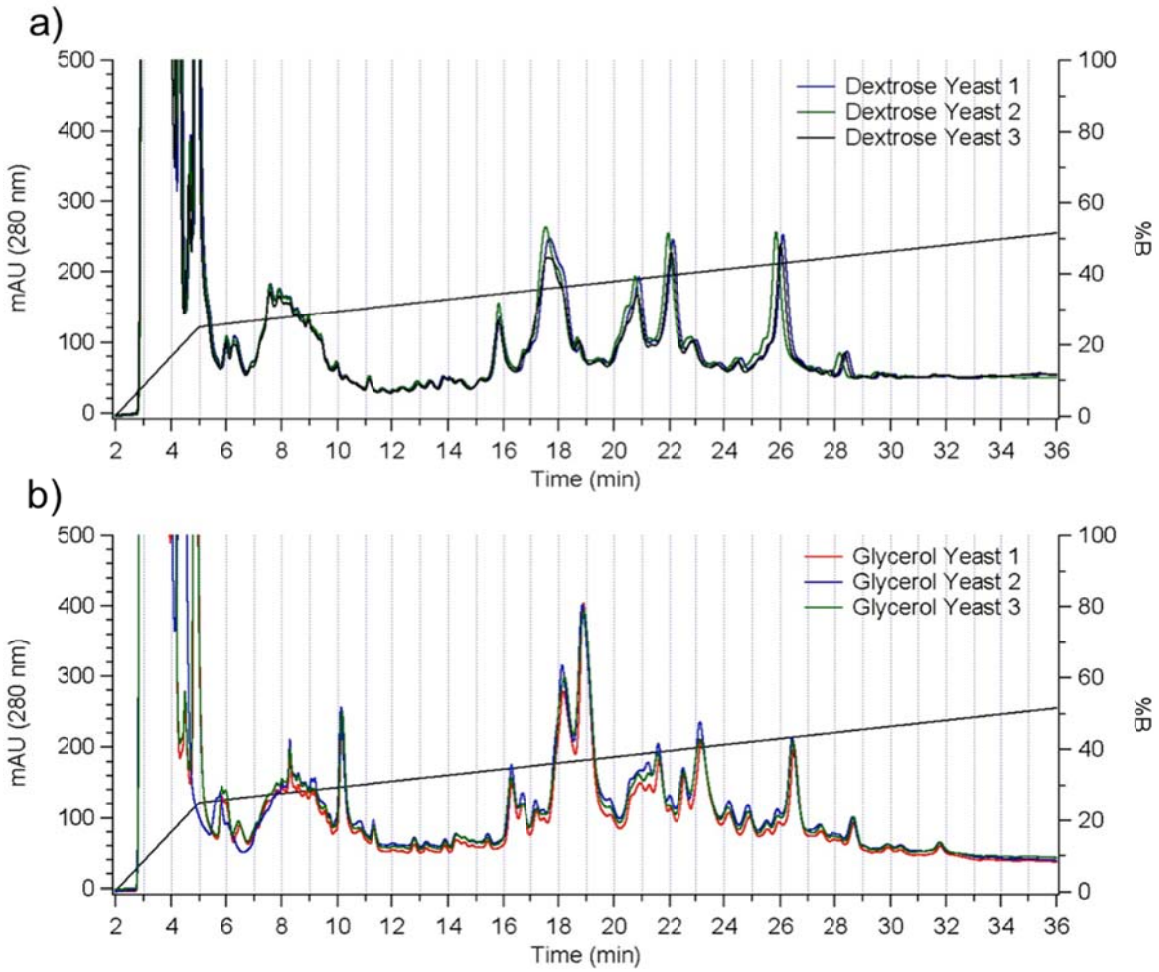


Figure 3-4. For the reversed-phase pre-fractionation methods, a more rigorous replicate analysis was employed to better evaluate the reproducibility of the entire method. As such, each intact protein sample injected three times with fractions collected and digested for each of the 6 runs. An Agilent PLRP-S, 4.6 mm x 25 cm column with 5 μm pores particles and 300 Å pores was used for the separations. The flow rate was 1.0 mL/min. A column temperature of 80 °C was used – well below the rated maximum of 200 °C. Fractions were collected every minute from 2 to 42 minutes for a total of 40 fractions collected. These fractions were later pooled to form the 20 fractions analyzed. For each sample, the UV traces overlay quite well demonstrating good method reproducibility. The gradient conditions are overlayed in black and correspond to the right axis.

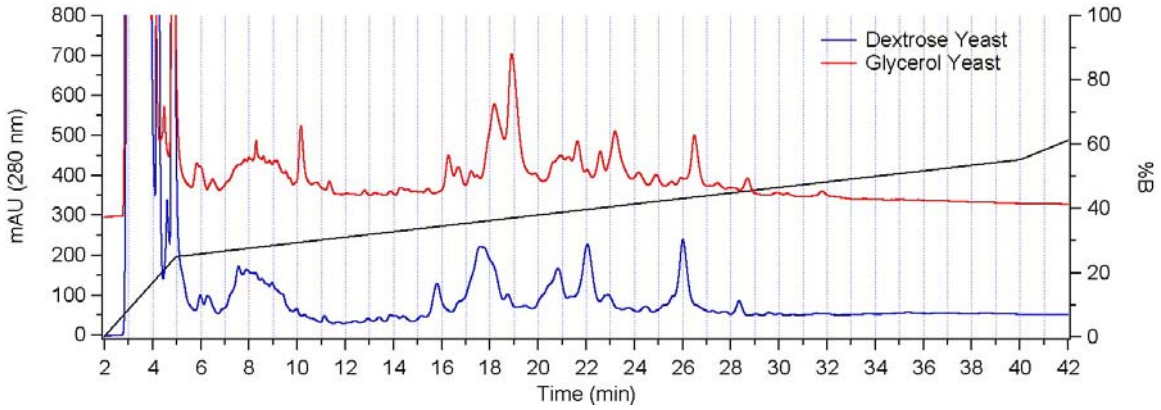


Figure 3-5. While the samples are very similar and show similar elution trends, there are also differences prevalent between the two sample types. The glycerol yeast data, offset by 300 mAU, show an intense peak at 10 minutes which is missing altogether in the dextrose yeast, as an example. The gradient conditions are overlayed in black and correspond to the right axis.

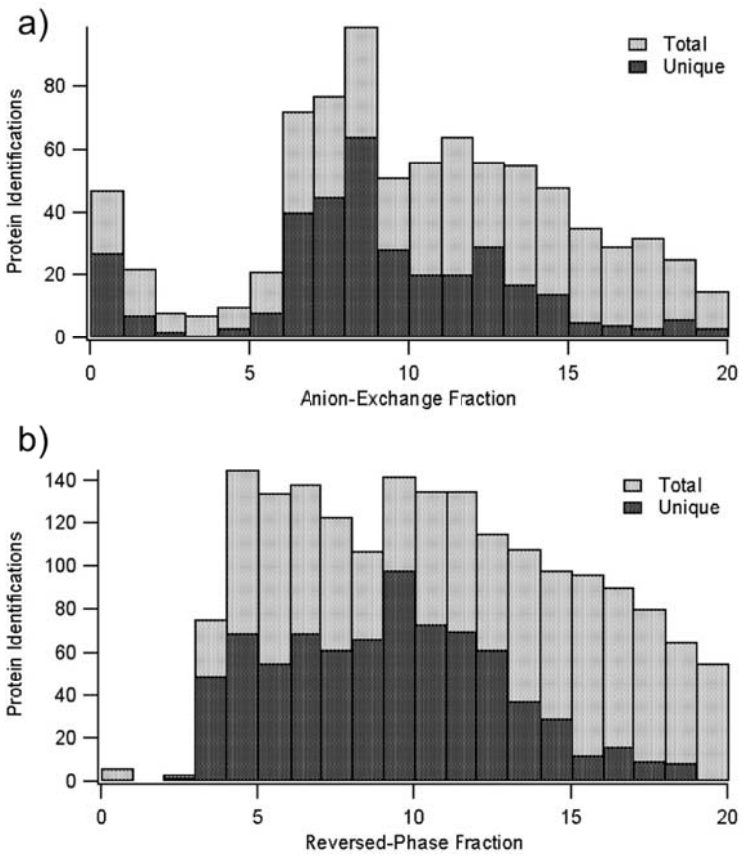


Figure 3-6. The total number of protein identifications for each fraction are shown in light grey. Plotting the data in this way is useful as it gives an idea as to what is eluting off of the LC column in each fraction and the complexity of the sample reaching the mass spectrometer. However, for the purpose of determining overall protein identifications, it is more convenient to represent the data by only plotting proteins in the fraction of greatest intensity, as determined through Scaffold 3.3. Fewer identifications were made by the anion-exchange method (a) than by the reversed-phase method (b). For both intact protein separations, few new unique proteins were identified in the final five fractions.

Figure 3-7. The reversed-phase pre-fractionation method identified 546 proteins in 2/3 replicate runs whereas there were only 262 proteins identified in 2/3 replicate runs of the peptide fractions. Only 16% of the proteins identified by anion exchange were not identified by the reversed-phase approach. Twenty fractions were analyzed by each method to allow for a uniform comparison based on gradient fractionation.

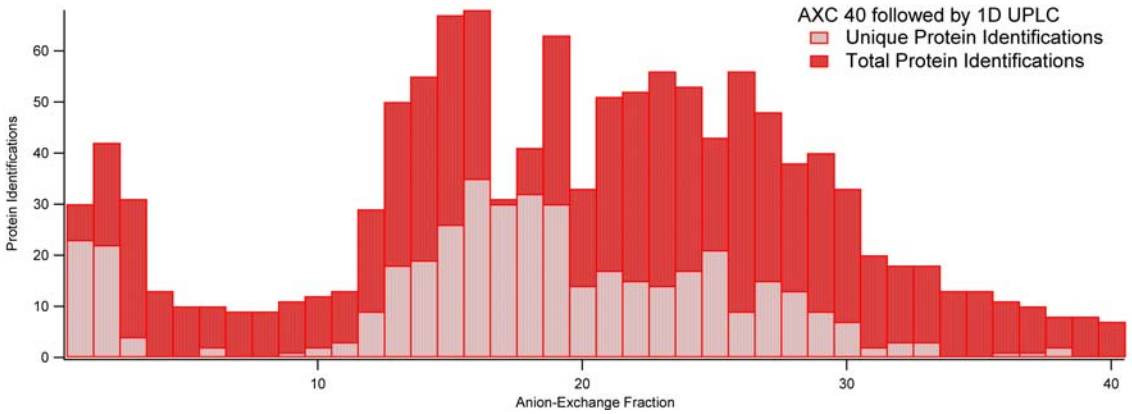


Figure 3-8. After running all 40 fractions from the anion-exchange pre-fractionation method, similar trends are observed as seen previously in Figure 6. The total number of protein identifications for each fraction are shown in dark red. Plotting the data in this way is useful as it gives an idea as to what is eluting off of the LC column in each fraction and the complexity of the sample reaching the mass spectrometer. However, for the purpose of determining overall protein identifications, it is more convenient to represent the data by only plotting proteins in the fraction of greatest intensity, as determined through Scaffold 3.3, and is shown in light red. Finer fractionation where there are large numbers of proteins eluting and courser fractionation over portions of the method where there are few unique proteins being identified could increase throughput

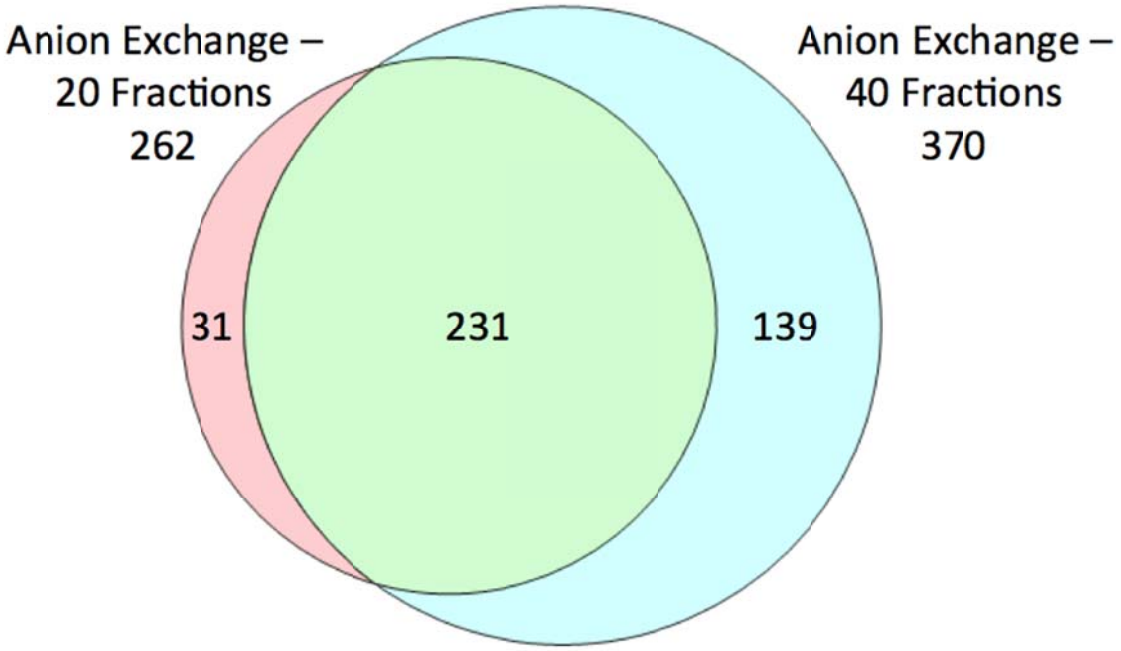


Figure 3-9. For the anion-exchange approach, all 40 fractions were analyzed in addition to being pooled to form 20 fractions. Finer fractionation allowed for a small increase in the number of proteins identified, but required an increase in analysis time from 120 to 240 hours.

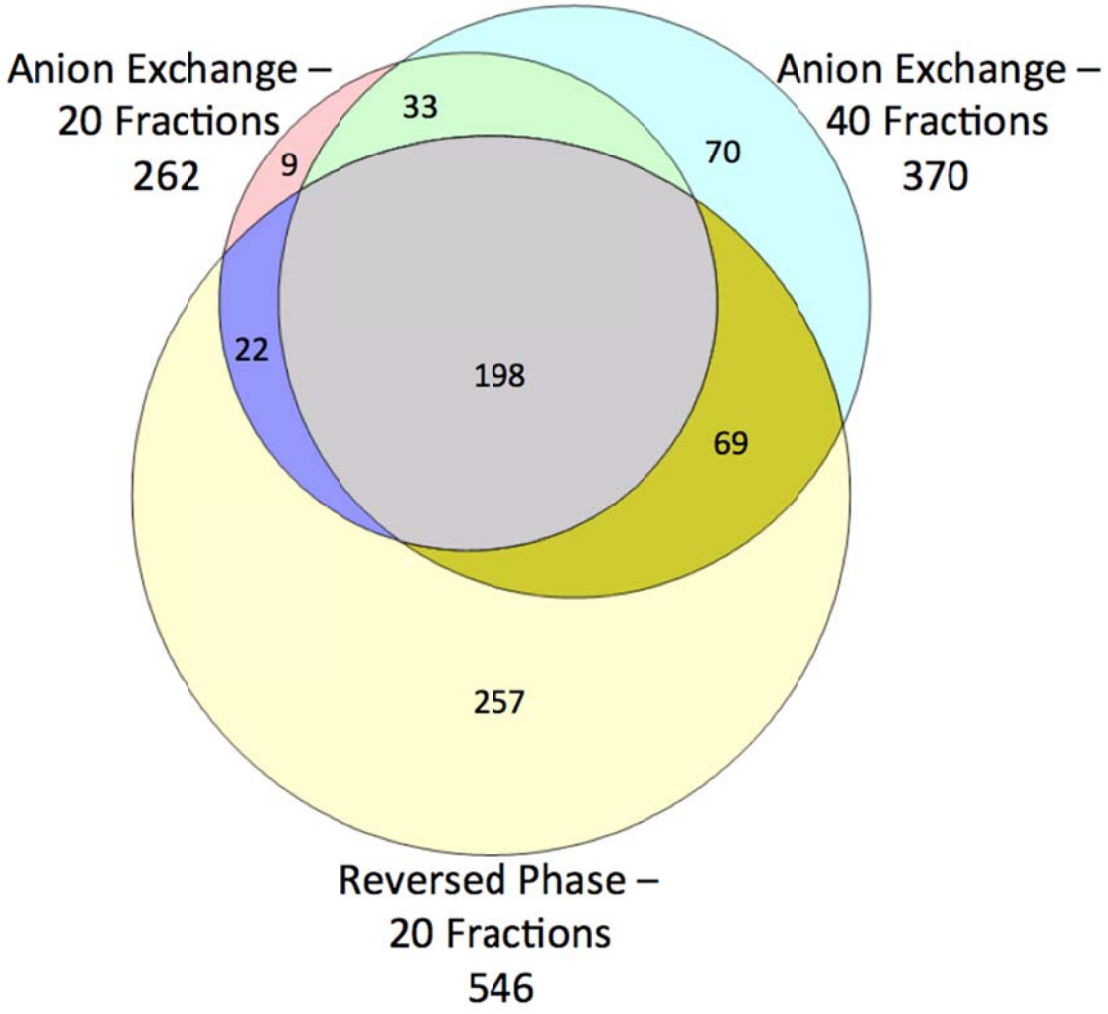


Figure 3-10. Reversed-phase pre-fractionation allowed for a greater number of protein identifications than either anion-exchange method. A total of 658 unique proteins were identified in at least 2/3 replicate runs for these fractionation strategies. Only 17% (112) were not found by reversed-phase pre-fractionation.

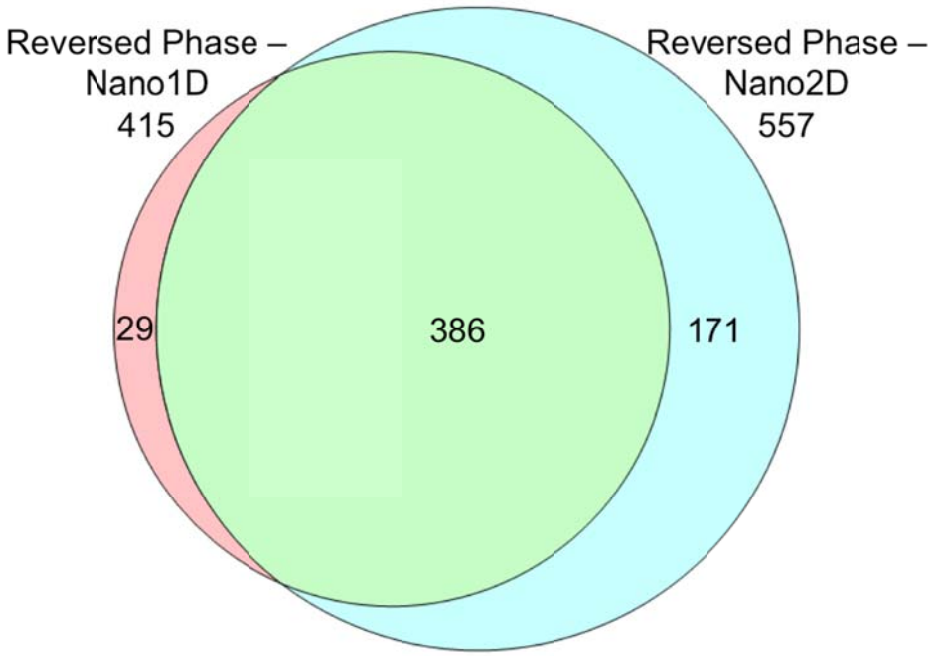


Figure 3-11. An alternative approach to increasing the fractionation of the intact protein separation was to increase the peak capacity of the peptide separation by incorporating a multidimensional separation at the peptide level. This comparison was made by analyzing only the glycerol yeast sample. While a greater number of protein identifications were made with the multidimensional separation, it is questionable whether the gains in additional identifications are worth the additional analysis time.

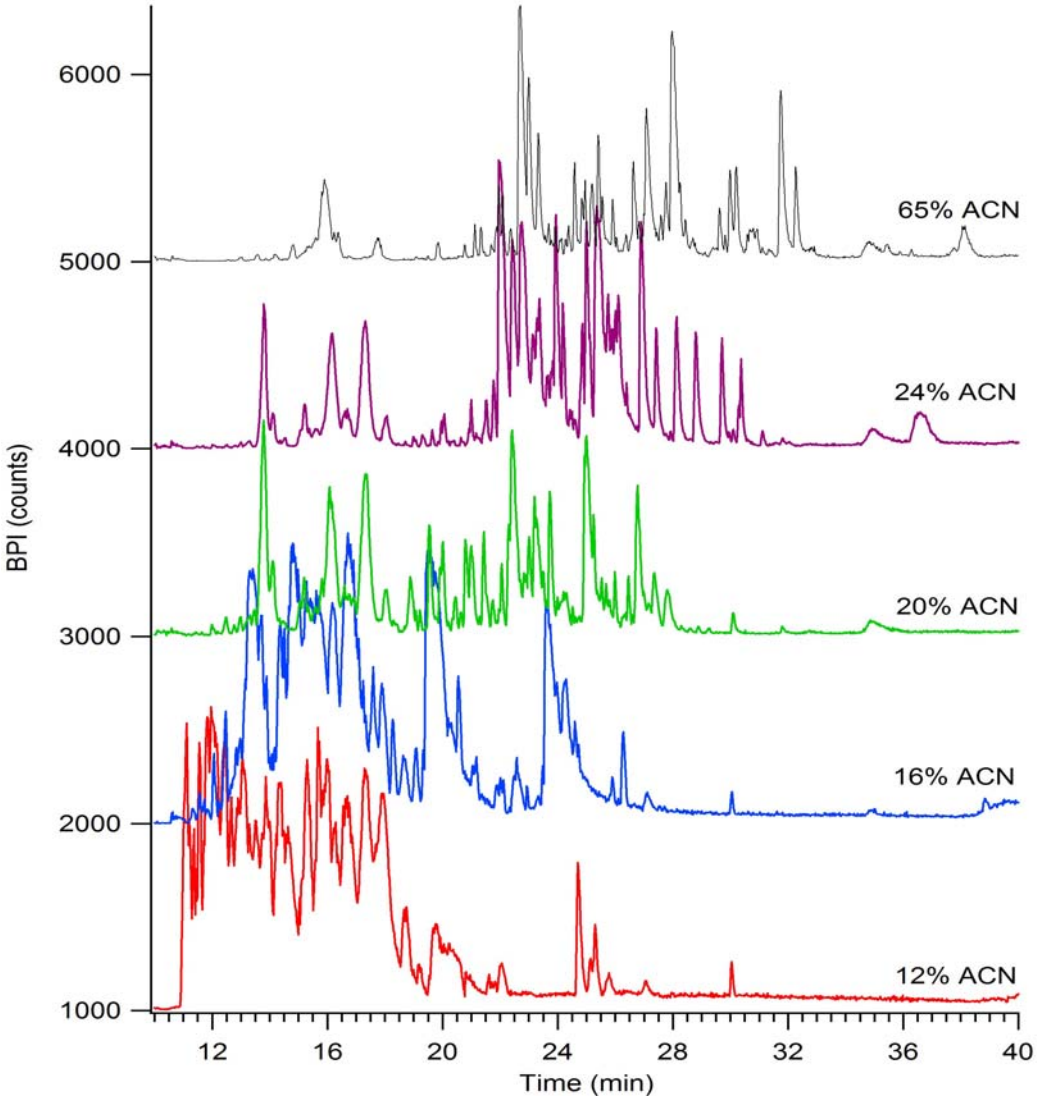


Figure 3-12. A representative set of chromatograms from the reversed-phase fractionation of yeast grown on glycerol. Reversed-phase chromatography was used in both dimensions of the peptide separation by operating the first dimension at a pH 10 and the second dimension at pH 3. The first-dimension column was a custom-packed 150 µm ID x 10 cm XBridge C18 with 3.5µ particles and 130 Å pores. The second-dimension column was a Waters Acquity BEH C18 column with 1.7 µm particles and 130 Å pores. Five fractions were taken from the first dimension and peptides were eluted from the first column in five isocratic steps of 12, 16, 20, 24, and 65% acetonitrile. There is a general trend of peptides eluting early in the first-dimension also eluting early in the second-dimension suggesting a lack some level of correlation between the two dimensions and limiting the peak capacity of the overall system.

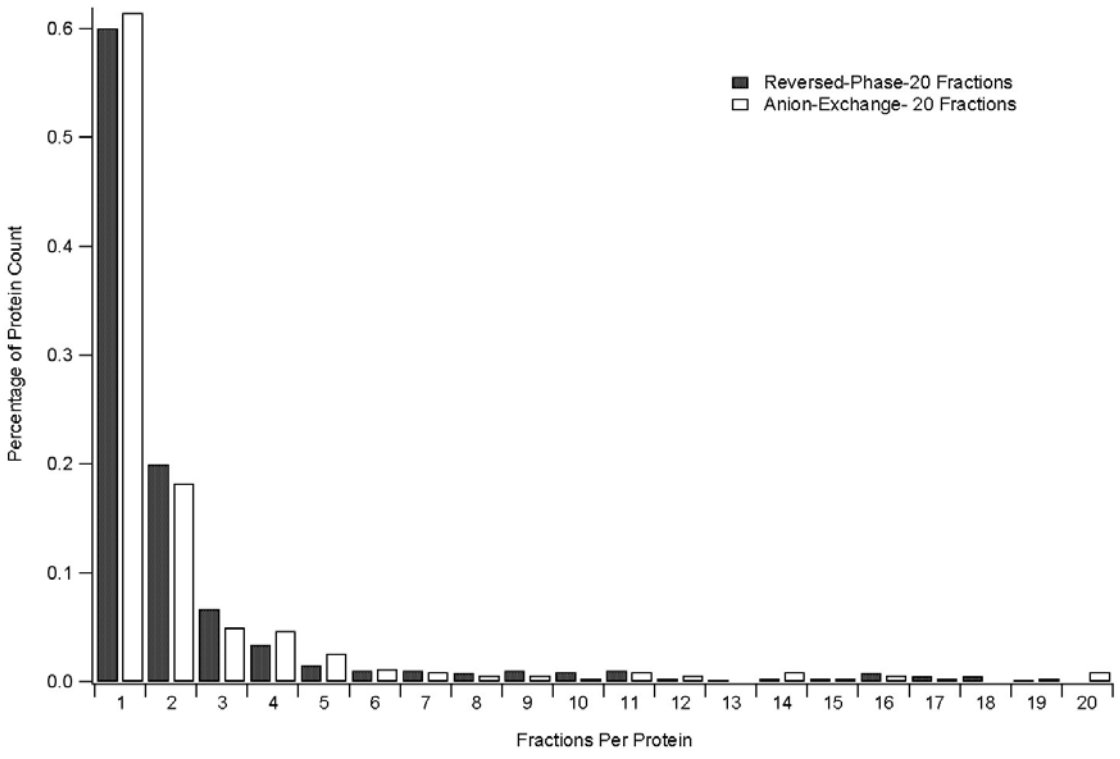


Figure 3-13. The percentage of unique proteins is plotted as a function of the number of fractions a unique protein is identified in. Both anion-exchange and reversed-phase pre-fractionation strategies exhibit similar behavior when compared on the basis of gradient fractionation. The majority of proteins (60%) elute in only one fractions indicating that the fractions were approximately equal to the width of a peak.

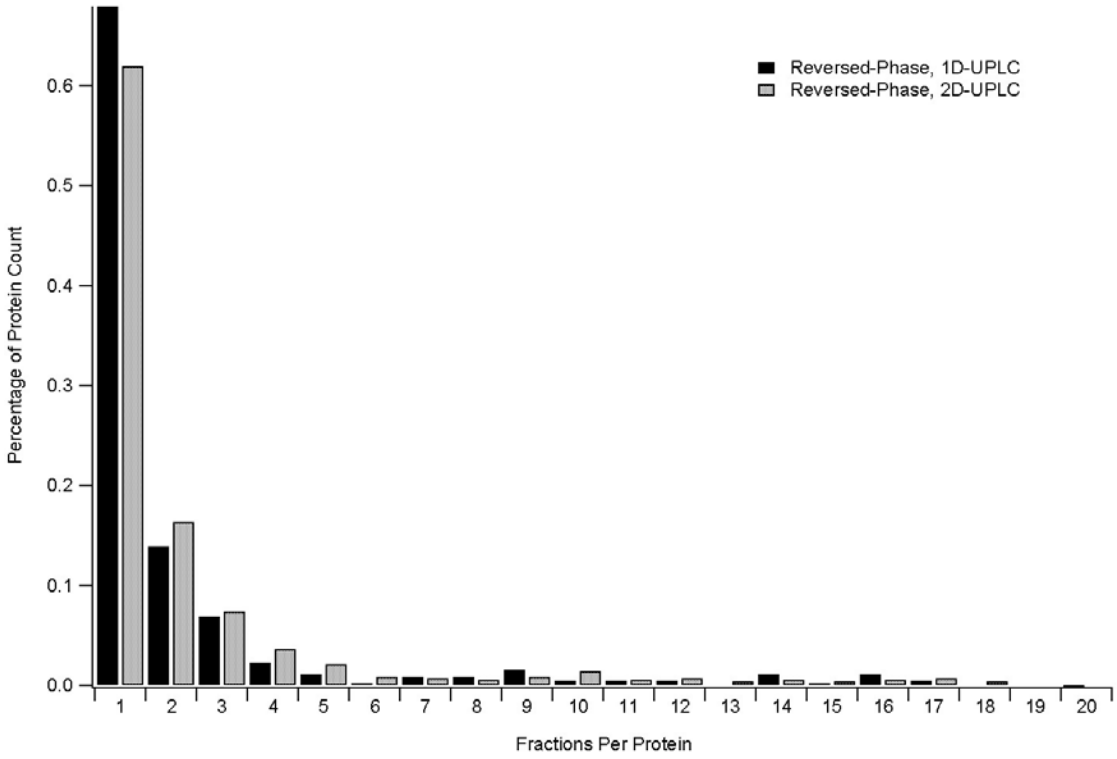


Figure 3-14. Only a slight increase in the number of fractions a unique protein was identified in was associated with the more comprehensive peptide analysis. It has been reported that greater peak capacity afforded by the multidimensional separation allows for the detection of peptides at lower concentrations and therefore the ability to detect proteins at low concentration.

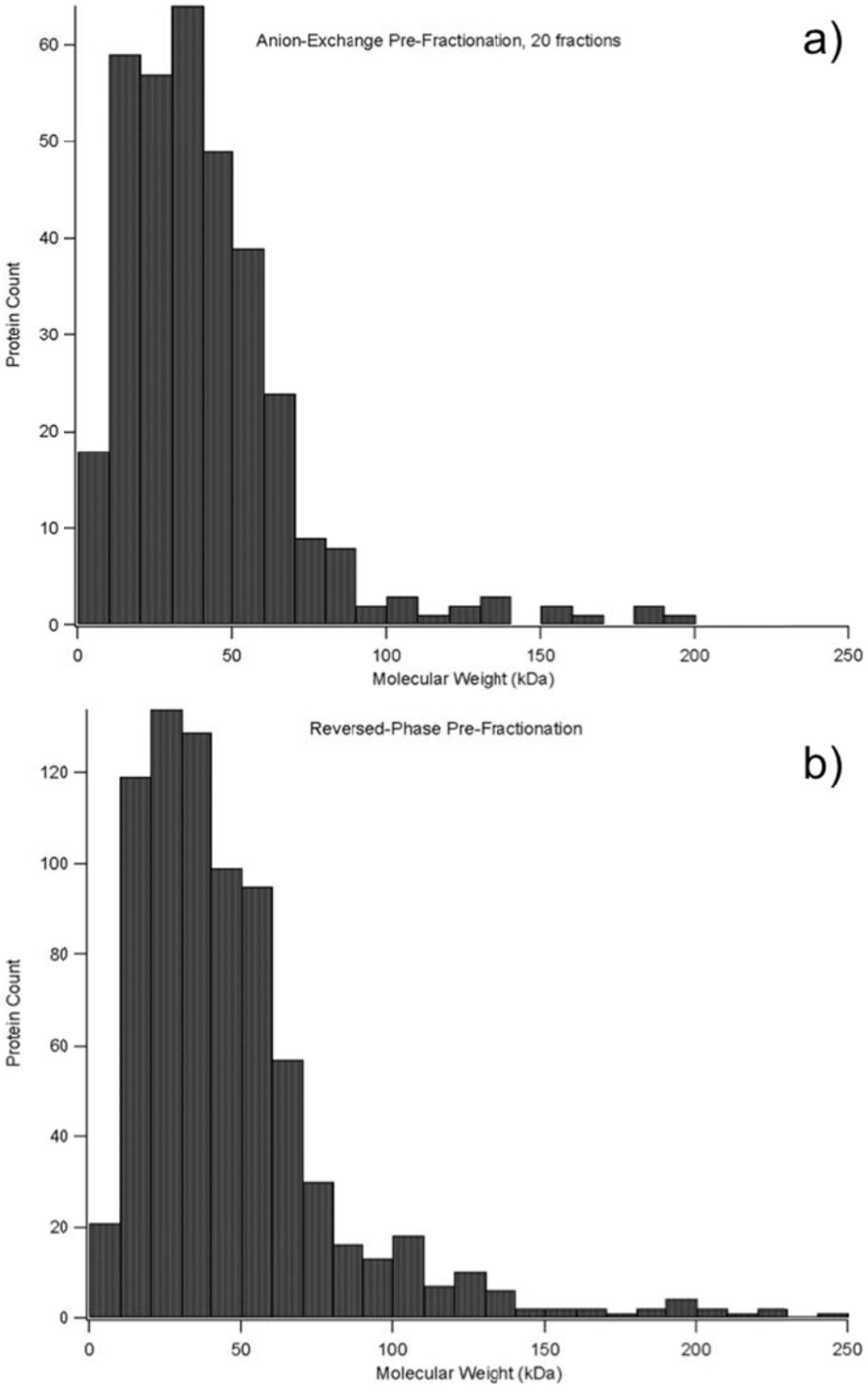


Figure 3-15. There was no bias associated with the molecular weight of the intact proteins identified by either pre-fractionation method. The median mass of proteins identified was 37.6 kDa for anion exchange (a) and 38.0 kDa for reversed phase (b).

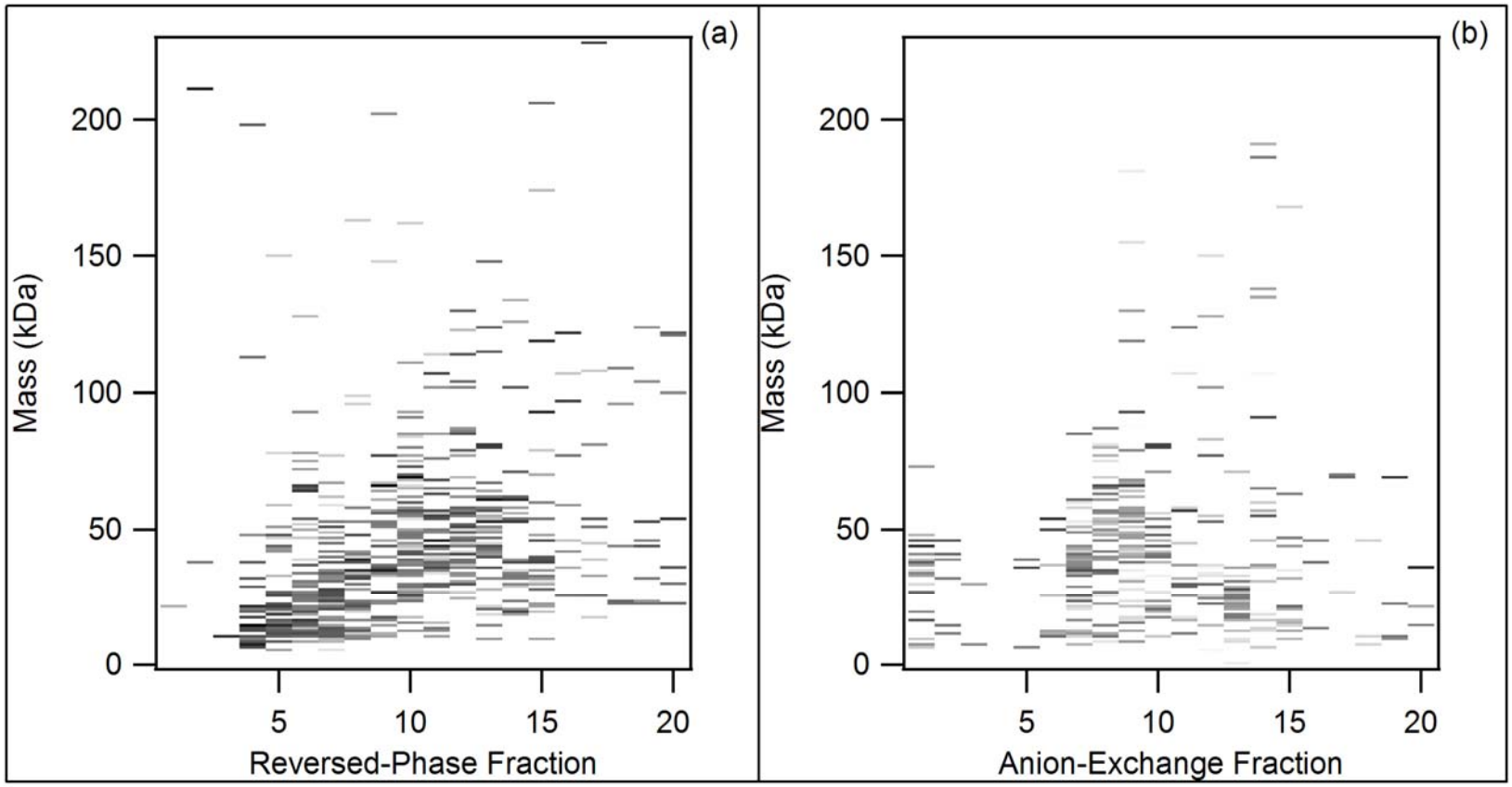


Figure 3-16. The mass of proteins for the reversed-phase (a) and anion-exchange (b) pre-fractionation methods are shown as a function of the fraction each protein eluted in greatest intensity. More abundant proteins are represented by darker bars.

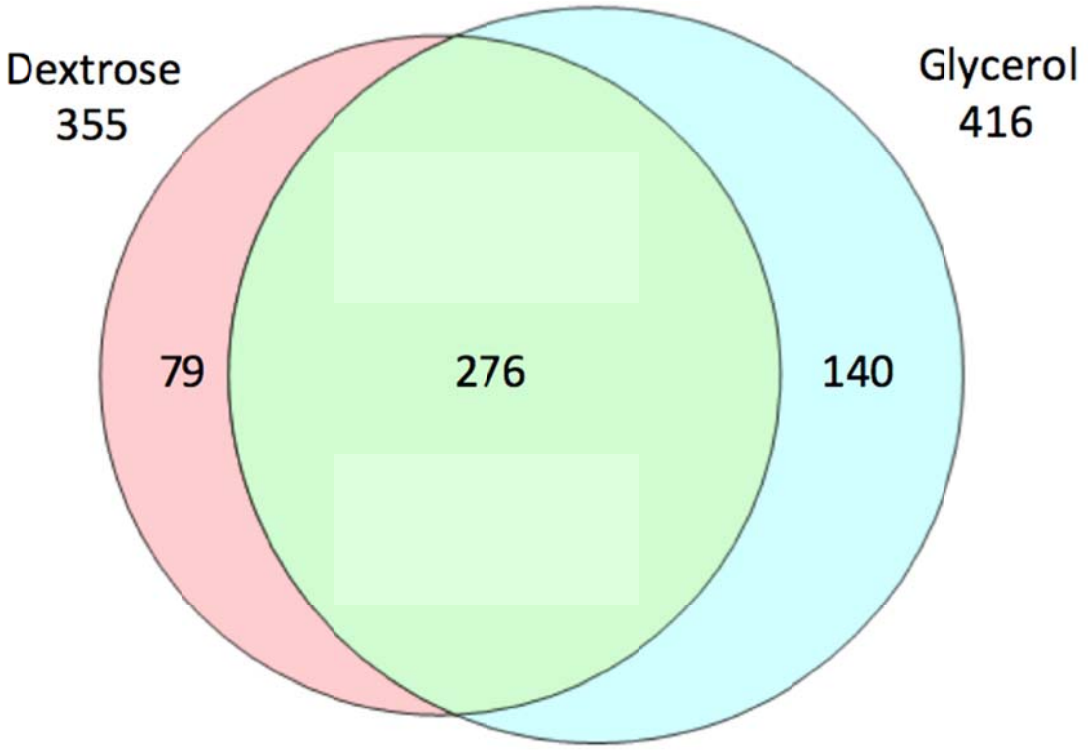


Figure 3-17. The reversed-phase pre-fractionation approach showed better overlap between the two samples than either anion exchange fractionation approach and significantly more proteins overall. While the samples were grown on different carbon sources, a significant amount of agreement is expected between the two.

Figure 3-18. Greater variability in the identified proteins was observed between the two samples when pre-fractionated by anion-exchange chromatography. In general, proteins identified by the anion exchange methods were less intense than by the reversed phase method suggesting that the decreased overlap between the two samples may be due in part to poor recovery.

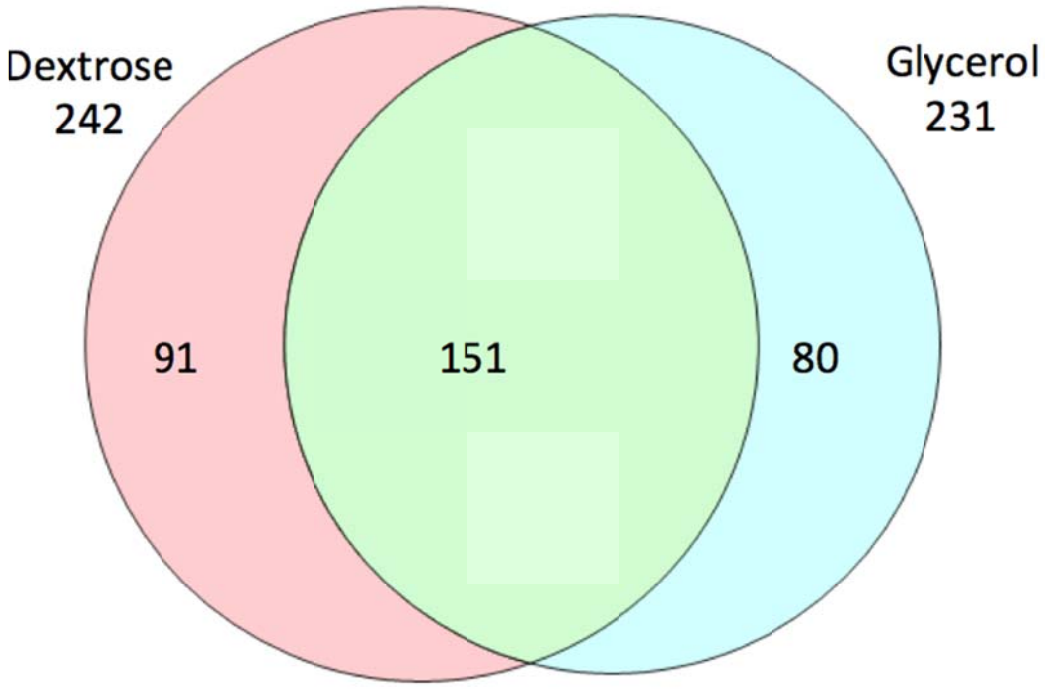


Figure 3-19. By increasing the fractionation of the anion exchange method from 20 to 40 fractions, not only was there an increase in the number of identifications, but better agreement between the samples with respect to the proteins identified.

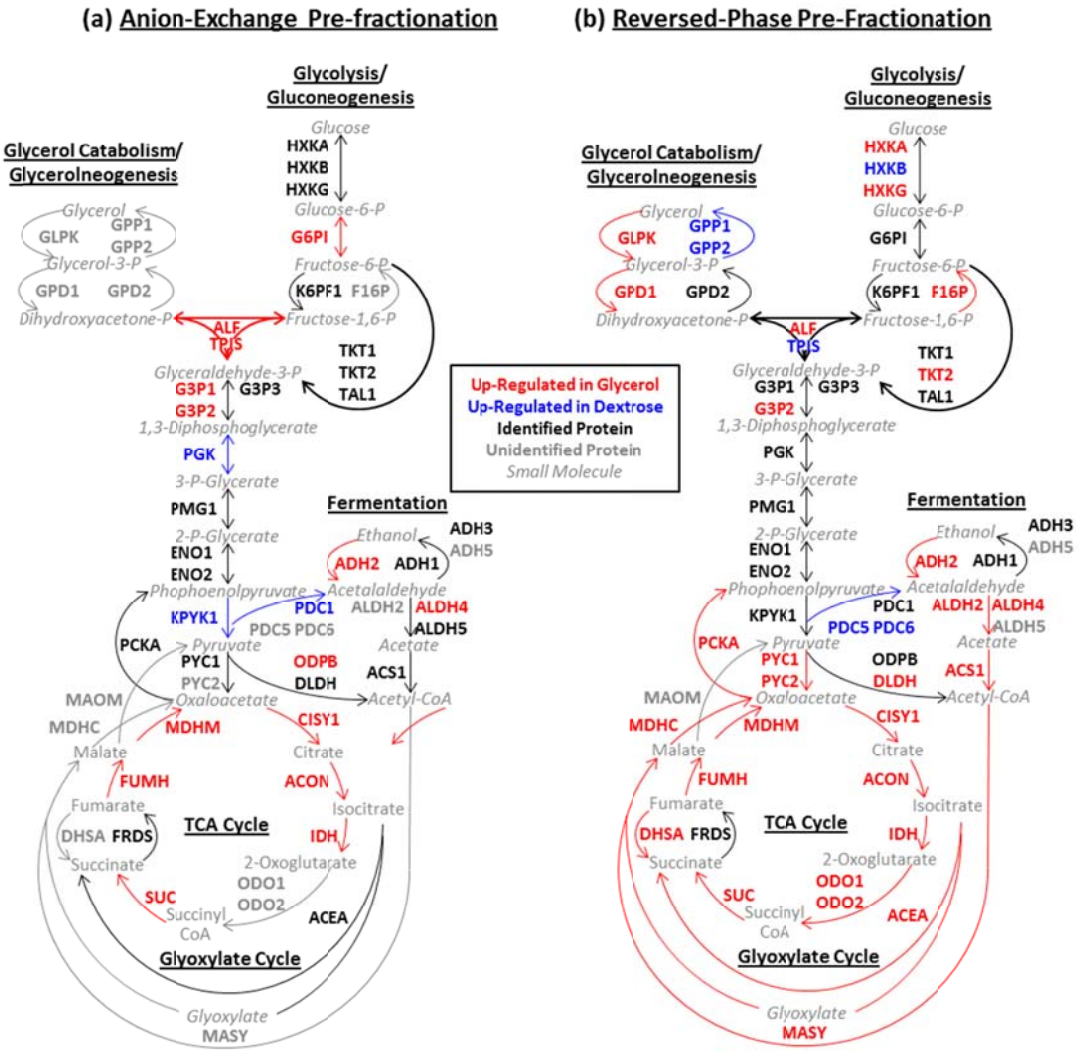


Figure 3-20. The up-regulation of the TCA cycle and the glyoxylate cycle are much more evident from the reversed-phase pre-fractionation approach. Many more proteins were identified by the reversed-phase method and these additional identifications were needed in order to form conclusions. The up-regulation of proteins in these pathways meets expectations given the nature of the differential sample.

CHAPTER 4: XUPLC MODIFICATION OF A NANOACQUITY UPLC

4.1. Introduction

4.1.1 Motivation for ultra-high pressure separations

Many applications in various analytical environments require fast analysis times and moderate peak capacities. For these situations there are seemingly endless combinations of columns and instrumentation available. Other applications and analytical challenges, such as those frequently encountered in metabolomics and proteomics, are somewhat less concerned with throughput as with overall peak capacity. Capillary chromatography is well suited for high peak capacity analyses in proteomic applications as these can be more readily packed to lengths longer than conventional chromatography columns, allowing for greater peak capacities.[1] Additionally, the small inner diameters of capillary columns consume less sample without decreasing the concentration of analytes as they elute from the column, making them an ideal choice for samples of limited quantity.[2] The nanoAcquity from Waters is an instrument commonly used for proteomics applications owing to its ability to provide flow down to 100 nL/min allowing for the use of capillary columns of 75 μm in ID. In Chapter 2, the performance ceiling in terms of peak capacity for the nanoAcquity system was found with an observed experimental peak capacity of approximately 200 and, as shown in Figure 4-1, lengthening the gradient beyond 90 minutes did not result in an appreciable increase in peak capacity and no increase in protein identifications. Ultimately, more peak capacity was desired as the result of the studies in Chapter 2 and not even the

multidimensional bottom-up analysis had sufficient resolving power for the total bottom-up sample analyzed. While the results from Chapter 3 with intact protein pre-fractionation did allow for a greater number of proteins identified than by the total bottom-up methods, a substantial amount of time was required for this type of analysis. Achieving larger peak capacities has been the focus of much work of both single and multidimensional approaches for a long time. For single-dimension analyses, the use of smaller particle diameters, longer columns, or a combination thereof provides for an increase in peak capacity as a result of the increase in theoretical plates.

Packing such columns on the capillary scale has also been done routinely. MacNair and Jorgenson demonstrated the peak capacity possible through one-dimensional capillary chromatography in 1997 by packing a 46 cm long column with 1-micron non-porous C18 particles.[1] What allowed this separation was the high pressure available to drive the flow and fittings to create low volume connections. By 2004, UHPLC at 100,000 psi was reported.[3] Currently, typical columns used for proteomics applications on the nanoAcquity are 75 microns in ID with a length of 25 cm packed with 1.9 μm bridged-ethyl hybrid (BEH) n-octadecylsilyl (C18) particles. Flow rates for these columns are on the order of 300 nL/min resulting in backpressures of approximately 8,000 psi, with 10,000 psi as the limit for the nanoAcquity. Unfortunately, columns with greater performance either through increased length or smaller particle diameters – such as those shown by earlier work in the Jorgenson lab – exceed the performance specifications of currently available commercialized instrumentation.[1,3,4] In order to circumvent the back pressure problem associated with smaller particle diameters, some researchers have opted for using more traditional HPLC type packing materials, utilizing 5 micron particles packed into columns exceeding a meter in

length.[5,6] A recent publication demonstrated peak capacities in excess of 1,000 with analysis times on the order of 12 hours. Peak widths were reported to be less than a minute wide for many of the peptides.[7]

Multidimensional separations have been used extensively to increase the peak capacity of an analysis. In concept, the technique is a good fit for proteomic applications, as the resulting peak capacities of two orthogonal methods can be quite large owing to the multiplicative effect. In practice, multidimensional separations create many challenges. Much of the work presented in Chapter 3 relied on the concept of multidimensional separations in order to achieve improvements in peak capacity to allow for a more thorough proteomic analysis. These techniques focused on offline approaches, considered to be among the most time consuming of multidimensional techniques, but with greater flexibility in designing the experiment.[8,9] For the analysis of peptides, many online techniques have been used for proteomic studies with commercialized platforms, such as the high/low pH nanoAcquity UPLC described in Chapter 2. These instruments do provide for an increase in peak capacity, however, their productivity is hindered by column equilibration after each second dimension analysis. Experience using such approaches in this laboratory have found that nearly half of the analysis time in a run spanning 6 hours is devoted to such housekeeping tasks with no additional information being gleaned. As a result, a high-efficiency single dimension separation becomes increasingly attractive as the equilibration time is significantly less.

4.1.2 Previous prototype instrumentation

Prototype instrumentation developed out of collaboration between the Jorgenson lab and Waters Corporation facilitated gradient elution for capillary UHPLC.[10,11,12] The hydraulic amplifier system (Figure 4-2) was routinely operated at 25,000 psi with capabilities

to go higher and approach 45,000.[10] This system relied on a Waters CapLC system to form the gradient and make the sample injection. During the "Load" mode of operation, a gradient was formed in reverse and pushed onto the gradient storage loop. The sample was then injected and placed at the head of this loop. For "Run" operation, two pneumatically-actuated on/off valves from Valco, Inc. were then closed. At this point, the custom instrumentation provided the UHPLC conditions. A modified Waters 1525 binary HPLC pump was reconfigured to pump DOT3 brake fluid to hydraulic cylinders. Attached to the end of the cylinders was a much smaller sapphire plunger, which allowed for the pressure amplification at a ratio of approximately 9:1. Therefore, by providing modest HPLC-like pressures from the 1525 pump, UHPLC pressures were attainable. The hydraulic amplifiers forced the gradient and sample out of the gradient storage loop and onto the analytical column.

This prototype instrumentation allowed for capillary LC/MS under conditions not previously possible and in a completely automated fashion. However, despite the capabilities afforded by this approach, the system had several deficiencies. The pump design utilized open-loop control for the "constant flow" operation. Constant flow was attempted by controlling the flow on the low pressure side of the hydraulic amplifier. The lack of feedback often resulted in a measurable difference in flow rates as gradients progressed through the column, creating an increase in back pressure as the viscosity of the water/acetonitrile mixtures changed over the course of the gradient. The pump heads themselves required custom check valves and therefore increased maintenance. Custom software was required making use somewhat cumbersome and the system relied on an autosampler not known for its efficiency with samples. Lastly, the system, as implemented, used a split flow/split injection system without trapping.[10] The split flow was required as the flow rates required for

the columns utilized was far below the minimum constant flow rate possible. The split injection was a result of this need to split flow. Trapping was not originally designed into the system and therefore was never utilized with this instrumentation. The system did however, provide enough in the way of preliminary results to show what was possible with higher pressure pumping capabilities and allowed for testing of many components at these pressures.

4.1.3 Purpose of nanoAcquity modification

There were a number of motivating factors for creating a new system for UHPLC operation. Lessons learned from the prototype system were used in the development of the XUPLC modification. Design simplicity was a must as the previous system relied on many custom parts, which were difficult to replicate and created maintenance dilemmas. An emphasis was placed on creating a solution using as many commercially available parts as possible. Between simplicity in design and commercial parts, it was hoped that the system robustness would be improved. Another motivation behind using off the shelf items was to allow for other investigators to more easily adopt this technology, enabling an expansion of the technique, whereas the custom instrumentation used before limited this possibility.

The autosampler of the nanoAcquity is also more efficient than that of the CapLC system. A 1 μL injection on the CapLC would actually consume several microliters of sample. With the nanoAcquity, a 1 μL injection consumed 1 μL. The nanoAcquity sample manager also was completely enclosed allowing for cooler temperatures. This stands in contrast to the open design of the CapLC where the minimum achievable temperature in a warm laboratory was no better than 14°C. With the nanoAcquity sample manager, 6° C is achievable.

4.2 System design & experimental parameters

The Waters nanoAcquity UPLC is commonly housed on a cart (Figure 4-3). As part of normal operation, the column heater module swings out in front of the mass spectrometer source. It was desired to maintain the approximate footprint of the instrumentation as well as facilitate the use of the nanoAcquity in its standard configuration once the modification was complete. Figure 4-4 shows the completed setup, illustrating that only a modest increase in the amount of space required was needed. Additionally, the column heater module of the nanoAcquity can still be positioned to allow for normal operation. System design for the XUPLC modification was modeled after the trapping operation of the nanoAcquity. Since the nanoAcquity is a proven platform, the goal was to attempt to scale up operation with as few changes to the operational methodology as possible. There were significant changes required as it pertains to routing of transfer lines, additional valves, and software methods. Nevertheless, the idea of trapping the sample on a short column, less retentive than the analytical column, and then eluting peptides with a gradient and no flow splitting remained at the heart of the system.

4.2.1 Pneumatics

Compressed gases, either air or N_2, facilitate the operation of high pressure valves as well as provide for the driving force for 30,000+ psi operation. A general overview schematic of the pneumatics can be seen in Figure 4-5 and Figure 4-6 is a picture showing additional details such as the location of ball valves and mounting locations. A convenient way to operate the system was through use of a nitrogen gas cylinder. Pressure needed to be regulated at the tank to a pressure not exceeding the limitations of the downstream regulators – 100 psi was an appropriate amount. Gas flow was then split to provide pressure to each of

the downstream regulators, one providing gas to the pneumatic amplifier and the other to the pin valves. A ball valve was placed after the tee and prior to the regulator to allow the user to quickly shut off gas flow to the high pressure pneumatic amplifier. This was not deemed necessary for the pin valve air supply and was not included in the system. As an additional safety precaution, a manual vent in the form of another ball valve was placed after the regulator for the pneumatic amplifier to allow for a manual system vent.

For the pneumatic amplifier, a high accuracy gauge (0.25% accuracy) was obtained from McMaster-Carr. The large dial size is also helpful for the user when specifying the pressure for operation as the pressure used during XUPLC operation is manually controlled via this regulator and gauge. A direct-acting, 3-way, 2-position solenoid valve (Humphrey model 320) with a Cv rating of 1.0 was used to control the gas supply to the pneumatic amplifier. A higher Cv rating allows a greater flow rate of gas. It was initially desired to have a greater flow rate than provided through this valve, but it was necessary to use this valve as larger valves did not meet the necessary pressure range specifications. Ultimately, the concern over flow rates was found to not be a problem and the valve used was found sufficient for operation of the pneumatic amplifier.

The on/off valves from Valco, Inc. are used to manipulate flow paths and isolate sensitive components from the high operating pressures. These are pneumatically actuated valves, requiring approximately 40 psi to operate. However, they do not require a significant amount of flow, therefore exceptionally small pneumatic solenoid valves could be used. Valves with a Cv rating of 0.01 (Humphrey model A101x) were used. The manifold configuration further simplified plumbing as a single supply line from the regulator could be used to provide gas to all the valves.

The pneumatic component that makes the XUPLC modification possible is the pneumatic amplifier. A Haskel DSXHF-903 (-903) triple-stage pump obtained from Haskel, Burbank, CA, was chosen for operation which has a maximum pressure output of 75,000 psi. A convenient consequence of using the -903 is the amplification ratio. At approximately 1000:1 (actually 1038:1), setting the gas supply pressure becomes easy as one can look at the gauge for the regulator and effectively think of the numbers as representing thousands of psi on the liquid side, so 30 psi for the inlet pressure corresponds to ~30,000 psi output. Additionally, the -903 has a substantial displacement at 2.3 mL. With this amount of displacement, the pump should never need to cycle in order to refill during operation, providing an even pressure for operation. If the pump were cycling during the course of an analysis, it would be expected that retention times would shift as the pressure would fluctuate.

4.2.2 Fluidic connections

An overall fluidic schematic can be seen in Figure 4-7. The modification begins at the inject valve of the nanoAcquity, bypassing its built-in second, 6-port valve often referred to as the "trapping" valve. Before going into detail regarding the flow paths, it is first important to identify many of the key components required for operation. The gradient storage loop allows for the nanoAcquity, with its closed-loop pump control, to form a precise gradient and store it for use during operation above the nanoAcquity's limit of 10,000 psi. This gradient storage loop is approximately 500 μL in internal volume and is 130' of 0.005" ID x 1/32" OD 316 stainless steel tubing, obtained from Valco Precision Sampling, Baton Rouge, LA (TSS.505). Stainless steel ferrules are compression set on the tubing and then brazed with silver solder, enabling leak-free connections above 40,000 psi.

Another piece of hardware from Valco enabling XUPLC operation are the UHPLC on/off valves (ASFVO40K1). Mentioned earlier in the pneumatics section, these valves work by being spring-loaded in the off position. The spring drives a coned piston into a similarly coned orifice, closing off the flow path. The valves are opened by the application of 40 psi of air, which pushes on a disc and opposes the action of the spring. For operation, four of the UHPLC on/off valves are required. Three of these remain in the closed position, while one must operate in the open position during the high-pressure operation in order to create the correct flow path for elution. Initial operation of the system is limited by the open valve as the seal surrounding the piston must hold at 40,000 psi and it is not possible to maintain leak-free operation above this pressure. Through use of valves with the 10,000 psi limit from Valco, it was experimentally determined that if the applied pressure is on the same axis as the piston, the valves can form a seal in the closed position and hold at pressures far in excess of the rated 10,000. Pressures as high as 30,000 psi were routinely applied to these valves when operated in the closed position. Lifetime was likely shortened as a result of operating at these pressures and these valves were eventually switched out in place of the 40,000 psi rated equipment. A final key piece of hardware from Valco were the microvolume tees (MT.5XCS6). Used in four locations, these tees allowed for low volume connections to be made under XUPLC conditions. While only rated to 10,000 psi, these two-piece stainless steel tees showed no signs of failure after months of use at 30,000 psi. Their internal bore was 150 microns. Ideally this bore size would be reduced to reduce effects of extra-column broadening and the potential for gradient mixing as both the fused-silica capillary and stainless steel tubing were of smaller ID.

An additional component necessary for operation was the column heater, where the trap and analytical columns were housed. The temperature control module (TCMII) and column heater module (CHM) were provided courtesy of Waters Corporation. While more typically used in HPLC as a column heater and mobile phase pre-heater, the size of the heater module along with the computer temperature controller was a convenient device for providing a constant, elevated temperature for the trap and analytical columns through the course of the XUPLC analyses. Lastly, prototype capillary fittings capable of holding at the requisite pressures were provided courtesy of Waters Corporation.

4.2.3 Flow paths

For operation, there are three distinct modes, each encompassing a different valve state configuration and therefore flow path. These are “gradient loading”, “sample trapping”, and “running”. Figure 4-8 shows the flow path for gradient loading with the desired flow path shown in green. The valco on/off valves are represented by either red or green circles with green showing an open valve and red showing a valve in the closed position. The analytical columns used for XUPLC analyses provide a significant amount of flow resistance and therefore the flow through the column at the flow rates and pressures observed during gradient formation and sample trapping is negligible. During gradient loading, a gradient was loaded at a flow rate of 5 μL/min. This was to keep mobile phase A and B under closed-loop control as the flow sensors do not provide linear feedback at flow rates in excess of 5 μL/min. Gradients are formed in reverse, placing the end of the gradient farther down the gradient storage loop as it is a “first on, last off” type of operation. Figure 4-9 demonstrates the flow configuration during the second-phase of operation, sample trapping. During sample trapping, the gradient storage loop was isolated by closing both the vent and gradient storage isolation

valves to prohibit any undesired mixing or additional flow due to mobile phase compressibility. The trap valve was then opened to allow flow through the trap column, but not through the analytical column. Sample trapping allows for a large volume of low concentration sample to be injected onto the system.

Once the gradient is loaded and the sample trapped on the trap column, the valve states are reconfigured one last time to permit XUPLC operation (Figure 4-10). The vent valve, nanoAcquity valve, and trap valve are all closed and the gradient storage isolation valve is opened to permit flow from the -903 to the mass spectrometer. Additionally, a nanoAcquity vent valve is opened behind the nanoAcquity valve. This additional valve serves two purposes. It allows for the nanoAcquity to gradually decrease the flow rate once the analysis is commenced and also provides a safety should the nanoAcquity valve fail during operation. This would prevent the nanoAcquity from ever experiencing pressures in excess of the rated maximum. There is one final safety feature of this system to prevent damage to components. The ID of the fused-silica capillary that connects the outlet of the -903 to the Valco tee at the end of the gradient storage loop is 15 microns in ID. Combined with being approximately a meter in length, this capillary line would only permit a flow rate of less than 100 μL/min if a catastrophic leak occurred beyond it at 30,000 psi. Additionally, during operation, typical flow rates of ~300 nL/min only cause a pressure drop of 100 psi, which is 1/3 of 1% of the applied pressure.

4.2.4 Software interface

The previous prototype instrumentation required not only the use of Masslynx for LC/MS control, but a significant amount of custom electronics and a custom software interface as well. As these custom components began to fail, it created significant problems

in the operation of the system. At one point there were three operational hydraulic amplifier prototypes, which has since been reduced to one, and the functionality of that system is only possible because of a workaround. Thus, when designing the present system, it was desired to use the built-in functionality of the control software provided by Waters Corporation with a nanoAcquity UPLC/MS system. Little was changed with regard to normal operation within Masslynx 4.1, the software that is primarily responsible for the operation of the mass spectrometer. The largest change was to how one creates a sample list.

A screen capture of the sample list from Masslynx can been seen in Figure 4-11. Normally, each row corresponds to a single analysis. Because of the need to store a gradient prior to trapping the sample, multiple rows are needed for a single analysis. Methods pertaining to each step were created to facilitate ease of use. The first line needed is a gradient loading method. At the end of the gradient loading, valves states are reconfigured to allow for trapping and the mobile phase conditions are changed to be amenable to this mode of operation. The second line required for operation is for sample injection and trapping. There are a significant number of timing events associated with this method. Initial settings entered into the events tab must be set to allow trapping operation. This should be a continuation of the valve states used at the end of the gradient loading method with mobile phase conditions amenable to sample trapping (0.5% acetonitrile). Trapping continues for an amount of time defined by the user and typically a volume three times the sample loop volume is used. Once this is complete, valve states are changed to allow for high pressure operation. When this occurs, an additional signal is sent to the mass spectrometer to begin data collection and the -903 pump is activated by applying pressure via the pneumatic solenoid valve. For short analysis times, the end of this method could be configured to turn

off the -903 pump, depressurize the system, and return the valve states to the "gradient loading" configuration to allow for the next run to proceed. Typically, more data analysis time is required and therefore "linker" methods are employed.

The needs for the "linker" methods results from a limitation of 32-bit operating systems as mentioned in Chapter 2. With these high-resolution separations, data collection rates were at times exceeding 2GB/hr per function (one for low energy collisions, one for high energy). Thus, in less than two hours, each function would reach the maximum 4GB limit imposed by the 32-bit system. While the mass spectrometer continues to perform scans and the XUPLC system continues to elute analytes, no data is saved. In order to overcome this limitation, multiple runs were linked together. There are a few disadvantages to this type of operation, one of which will be discussed in more detail in section 4.3.3. The most obvious disadvantage is that approximately 15 seconds of data are lost for each additional linked run. This is certainly not a large amount of data, but it would be ideal to not lose any data. Unfortunately upgrading to a 64-bit operating system is not sufficient as Masslynx was written in a 32-bit environment and significant software changes would be required to facilitate the use of the 64-bit operating system.

While Masslynx is the program that is used to control the mass spectrometer and provides access to modify both LC and MS conditions, the nanoAcquity console and nanoAcquity inlet method editor are the two programs that are actually running the nanoAcquity. Masslynx provides a convenient link to these programs and streamlines the user experience by requiring only one application to be physically opened by the user. The nanoAcquity console is critically important to the operation of this system. There has been much discussion regarding valve state changes with little mention of how this actually occurs.

During setup, column flushing, and other non-routine activities, the valve states can be directly controlled by accessing the "rear panel" from the troubleshooting pull-down menu within the nanoAcquity console. This is highlighted with a red oval in Figure 4-12. In order to have enough valve states to properly control all aspects of the system, switches on both the sample manager module as well as the nanoAcquity binary solvent manager were needed. These can be toggled by clicking on the icon for each of the swtiches changing between a red, open switch, and a green, closed switch. Conveniently enough for operation, "red" corresponds to a closed Valco on/off valve, while "green" indicates that the pneumatic solenoid is energized to allow gas flow to the on/off valve, opening it (Figure 4-13). Fortunately, changing these switches can be automated. This is accomplished by accessing the inlet method editor. Once in this part of the program, one can edit the timing events associated with both the pump and the autosampler by navigating to the "Events" tab and inputting the appropriate values (Figure 4-14). In order to assist in the creation of methods, a spreadsheet was created where inputting the gradient loading time will output all the values needed across all the tables for all windows.

4.2.5 Constant Pressure & Elevated Temperature

Constant flow is the standard for HPLC and UHPLC instrumentation. For flow rates in excess of 5 μL/min, instrumentation from Waters Corporation operates under open-loop control, relying on the stepper motors powering the pistons to deliver the desired flow. For flow rates in the nanoliter/min range, sensors measuring the flow rate were added to the Acquity UPLC to create the nanoAcquity UPLC. The pressure limitation of the Acquity is 15,000 psi, however the nanoAcquity is restricted to operation below 10,000 psi as a result of the flow sensors. Developing a sensor that operates at 30- or 40,000 psi, as desired for this

work, and can regulate flow down to a few nanoliters per minute is a significant engineering challenge. Ultimately, repeatability is the desired outcome. Using the -903 pump provided a simple solution, although this solution comes with the requirement that constant pressure operation is acceptable. There are a few drawbacks to constant pressure systems. If there is a leak, there will be no corresponding pressure readout indicating a loss of pressure. If there is a clog, there will be no system warning that the pump has gone overpressure. Lastly, the flow rate will change during the course of gradient operation. The change in flow rate is the result of a change in the viscosity of the mobile phase due to the water/acetonitrile gradient. Prior to assembly of the system, the change in viscosity over the course of the gradient was modeled at four temperatures – 25, 45, 65, and 85°C – from 0% to 100% acetonitrile, as observed in Figure 4-15. At first glance, there appears to be a significant change in viscosity among the temperatures and this is true when comparing the different temperatures. However, what is more important to consider is the change in viscosity over the course of the gradient, which is narrower than plotted, typically in the range of 1-40% acetonitrile. Normalized data for the four temperatures was plotted (Figure 4-16) for a more appropriate gradient range. At 25°C, there is a relative difference of less than 15% through the course of the gradient. By increasing the operational temperature to 65°C, this relative difference is reduced to approximately 5%. While the relative viscosity change over the gradient was reduced at elevated temperature, the viscosity overall was reduced. The consequence was that for a given pressure, the flow rate would be greater and therefore an analysis that might require 4 hours at room temperature, would require 2 hours at elevated temperature, resulting in shorter analysis times with equivalent performance in terms of peak capacity.

4.2.6 Columns & Running conditions

Three capillary columns were used in the evaluation of the XUPLC system. The first was 200 cm x 75-micron ID packed with 1.9 μmμ diameter Acquity BEH-C18 particles. The second column used to evaluate the system was 117 cm x 75-micron ID packed with 1.5 micron diameter BEH-C18 particles. The final column used for analysis was a 25 cm x 75-micron ID column packed with 1-micron BEH-C18 particles. For trapping operation, a trap column was packed with 5-micron Symmetry-C18 particles. Traps were 150 micron in ID x 6.5 cm in length. All particles were supplied by Waters Corporation and the columns were custom-packed in house.

4.3 Results & discussion

As with any new system, there were a number of unforeseen problems encountered. Proper formation of the gradient and determining valve timings was a significant challenge. This was made difficult by the length of the columns used during setup and countless leaks at the many ultra-high pressure capillary connections. However, much was learned regarding system setup and operation as a result of these issues. Likely the single most challenging problem ahead is in the capillary fittings themselves. While the prototype fittings available for this work do provide a solution to facilitate the advances in instrumentation shown here, they are undeniably challenging to work with and require a skilled user. This is an area where additional engineering resources should be invested in order to facilitate the adoption of XUPLC. Despite, these difficulties, preliminary chromatograms from the three columns tested are promising. This data, along with discussions regarding system setup, operation, and troubleshooting will be discussed in the remainder of this section.

4.3.1 Gradient Clipping

As a result of using a gradient storage loop, there is a significant delay volume between the pump outlet and the entrance to the storage loop itself. As a result, when the end of the gradient is formed by the pump, an additional amount of flow equivalent to the volume of connecting tubing is necessary to properly place the gradient. If too much time elapses and a greater amount of volume is delivered than required, peak elution will be delayed and the analysis unnecessarily prolonged. If not enough volume is delivered, then the initial portion of the gradient is clipped. This gradient clipping is demonstrated in Figure 4-17. It can be seen at around 18 minutes, there is an abrupt rise in the baseline and elution begins. This is a result of the gradient being clipped and a substantial number of peptides are being eluted simultaneously. Since the gradient is loaded in reverse, the last portion to be loaded is the low organic part of the gradient. When setting up methods, a systematic approach is needed to avoid this problem. The most straightforward is to run several consecutive runs using a standard and to introduce longer and longer delay times for gradient placement. When the TIC chromatogram (not shown) begins to show a gradual rise and fall over the course of the analysis, the correct gradient placement delay has been found.

4.3.2 Leaks

By far, the most challenging problem with the XUPLC modification is leaks. Large leaks are often easy to identify, as a pool will rapidly form around the suspect connection. Small leaks, on the other hand, tend to evaporate before pooling as the system is operated at elevated temperature. Frequently, these problems will be diagnosed through observation of the chromatogram. Figure 4-18 shows an example of what was supposed to be a gradient requiring in excess of 4 hours to complete. It would appear as though a leak developed not

long after 60 minutes as peak shapes suddenly become much more closely spaced together and elution is completed by 90 minutes. Identifying which connection is leaking becomes the challenge. If a leak occurs after the trap column, then there will likely be few or no peaks observed, as the majority of the sample will not reach the analytical column. This was not the source of the leak for Figure 4-18. For this particular chromatogram, the leak was prior to the trap column. The result is that the sample is eluted from the analytical column, but in a compressed elution window, as the majority of the gradient volume is lost through the leak. These problems can largely be avoided by systematic and careful installation of the capillary fittings. When setting up the system, it has also been useful to pressurize it beyond the desired operating pressure. Thus, for 30,000 psi operation, the system should be tested at 33 or 34,000 psi. If a fitting holds for 15 minutes at the elevated pressure, then it will likely not be problematic at the operating pressure.

4.3.3 Data Analysis of linked files

The data analysis of the files generated is a seemingly mundane detail, but warrants discussion as there are issues that users must be aware of. The ProteinLynx peptide scoring algorithm assigns peptides scores and then these scores are used to determine the confidence in a protein identification. One of the criteria used is expected retention time. Ordinarily, this causes no issues as chromatograms are collected in one file and contain all the peptides from the beginning through the end of the gradient. The long gradient times with the longer columns employed with XUPLC create the aforementioned file size problems, which must be circumvented by essentially tricking the mass spectrometer control software. What happens when this occurs is that peptides eluting at the beginning of subsequent linked runs will have an unexpected retention time. If the files are individually processed, this turns out to be okay,

as the order of the peptides still makes sense. If the files are merged, an option within ProteinLynx, then a number of issues present themselves. Peaks are now out of an expected order. When the mass spectral data is merged, peaks can also become overlapped that should not be. The short-term workaround, until a more permanent fix within ProteinLynx is found, is to process the files individually. At times a protein will be identified in more than one fraction, but with manual manipulation of the data, these duplicates could be removed. Importing the data into Scaffold at this point would also likely simplify the problem as the data could be grouped together as its own "sample category".

4.3.4 Chromatographic performance of XUPLC

The column initially used to test the XUPLC modification was the 200 cm column packed with 1.9-micron BEH-C18 particles. An example chromatogram is shown in Figure 4-19. For this analysis, a 35 μL gradient was employed for the separation. The dead time of the column was observed to be 55 minutes and peaks eluted until 210 minutes. This was a very aggressive gradient with a roughly 7.2% change in acetonitrile composition per column volume. Typically, it is recommended by Waters to employ a gradient change of 1% per column volume. Nevertheless, even with this steep approach, full-width, half-maximum (FWHM) peak widths were found to be 0.25 minutes resulting in an experimental peak capacity of 660 for this separation. The sample analyzed during this run was the total bottom-up dextrose yeast sample previously analyzed in Chapter 2. In the four hours required for this analysis, 153 proteins were identified. Unfortunately, a flattening of the gradient, with a run time extending beyond 600 minutes, did not appreciably increase the peak capacity of the system and, strangely, resulted in fewer protein identifications. Given that the change in

acetonitrile per column volume was much closer to the desired 1% mark, this was both surprising and troubling. A similar phenomenon was noticed with the 117-cm column as well.

Figure 4-20 shows the results of the use of 1.5-micron particles in a 117-cm column. A 45-uL gradient was employed, resulting in a roughly 3.4% gradient change across the column – still three times the recommended gradient slope. The resulting average FWHM peak widths were nearly the same with a width of 0.26 minutes, an increase of half a second. With an elution window of 145 minutes, this resulted in a peak capacity of approximately 570 and from this analysis, 128 protein identifications were made. The third column used to analyze the system performance was much shorter at 25 cm, but with the smallest particle diameter of the three at 1 micron in diameter. Using this shorter column resulted in lower peak capacities, but analysis times comparable with what is more typically of normal nanoAcquity operation. Three different gradients with volumes of 15-, 25-, and 35 μL were used resulting in a gradient change across the column of 2.0%, 01.2%, and 0.9%, respectively (Figures 4-21 – 4-23). The peak capacity of the short gradient was determined to be 160, on par with what was seen from the shortest nanoAcquity method utilized, albeit in two thirds of the time. Increasing the gradient length to 25 μL resulted in an analysis time of 90 minutes and a peak capacity of 250, a 56% increase over standard nanoAcquity operation. Further flattening of the gradient below the recommended level resulted in broader peaks, with an average peak width of 0.37 minutes. For the 15- and 25 μL gradients peak widths were 0.26 minutes, on average. Interestingly, there was a decrease in the number of proteins identified, despite the increase in peak capacity. This would seem to indicate that the mass spectrometer was having difficulty characterizing the peptides with peak widths less than 15 seconds wide. This is not surprising considering the relative standard deviation for peak widths ranged

between 20 and 25% meaning that some peaks were quite narrow. Narrow peaks, while the goal of the XUPLC separations would have been difficult for the Apex3D algorithm to properly characterize given the frequency of the scans and the need for a minimum of 10 points per peak for proper characterization.

When compared to standard nanoAcquity operation with a QTOF Premier, the XUPLC system outperformed the nanoAcquity in terms of total unique protein identifications and peak capacity. This data is summarized in Table 4-1. Also included in this table is analysis time, peak capacity per hour, and proteins identified per hour. In addition to the data from the XUPLC operation and standard nanoAcquity operation, results from the SynaptG2 discussed in Chapter 2 are also included for comparison. Excluding the SynaptG2 results, the most proteins were identified by the method using the 200-cm long column packed with 1.9 μm particles, with proteins being identified at a rate of 0.64/minute. While the 117-cm long column with 1.5 μm particles had four more identifications than the shorter column with 1 μm particles, the "productivity" was much better for the shorter column with 0.99 identifications/minute as opposed to 0.68 identifications/minute for the longer column. It is unclear why the shorter column produced a 46% increase in proteins identified/minute over the 117-cm long column. Peak widths for the shorter column were wider at 0.37 minutes as a result of the shallow gradient in comparison to 0.26 minutes for the 117 cm column. This would lead one to believe that the mass spectrometer was having difficulty handling the narrow peaks eluting from the longer column. A steeper gradient on the short column had productivity values that remained high, 0.94 identifications/minute and peak widths of 0.26 minutes. There is likely a balance to be maintained between overall peak capacity and peak widths. Clearly this is an area that needs to be investigated further to determine the

appropriate conditions for proteomic analyses with higher pressure operation. Lastly, as in Chapter 2, the data from the SynaptG2 was the best in terms of proteins identified per hour at 4.0. If one were to look at the difference between the QTOF Premier and the SynaptG2, there was a 6-fold increase in protein identifications. Extrapolating, one could expect upwards of 900 protein identifications using the 200-cm long column where 153 proteins were identified with the QTOF Premier.

The examples shown thus far employ gradients that are quite steep in comparison to what is typically used. Using a longer gradient on the 117 cm long column, fewer proteins were identified despite twice the analysis time. The resulting gradient used in the longer analysis had a 1.8% change in acetonitrile concentration across the column as opposed to 3.4% for the faster gradient. Qualitatively, the chromatogram looks good at first glace. However, when looking closely at a smaller portion of the elution window, as in Figure 4-24, one can immediately notice asymmetric peaks. The peak at 178 minutes is a good example. This peak would appear to be overloaded. Looking at more of the peaks in this window, they all look overloaded. However, the intensity of these peptides are lower than many of the peaks used in the standard tryptic enolase digest used for system performance monitoring. Lowering the injection volume did not help this problem and, in hindsight, given the fact that all of the peaks exhibit this tailing, it should not have helped. The problem likely lies ultimately in the system construction. For typical trapping operation on a nanoAcquity under normal operating parameters, a trap column is only 2 cm long. Due to the high pressure fittings required to "float" the trap column in excess of 30,000 psi, the minimum length possible was roughly 6.5 cm (Figure 4-25). It is suspected that the extra length required by the high pressure fittings is causing a misshaped peak shape to elute from the trap column.

With a steep gradient, there is enough gradient focusing to mask this problem. However, under more shallow gradients, there the analytes do not properly focus on the head of the analytical column and the poor peak shape is maintained throughout the analytical column. If this is the case, then peak shapes could be substantially improved and therefore peak capacities would increase substantially as well. There is some indication of how sharp a peak could be by looking at the front of the peaks. If the peak shape more closely approximated a Gaussian, one could expect peak capacities to potentially double given the extent of the tailing observed. An additional possibility is the release from the trap column in relation to the retention of the analytical column. Sorbent choices will significantly impact the performance of a trapping setup. As gradients become shallower, the need for less retentive trap columns relative to the analytical column becomes increasingly important. Trapping operation allows for a greater sample load to be placed on the column, as the ID of the trap column is frequently larger than the ID of the analytical column. With only a small portion of the sample re-concentrating at the head of the column, it is easier to avoid situations of mass overload. However, the re-concentrating or re-focusing effect is critical to the success of trapping operation. Once the sample releases from the trap column, the peak shape will be poorer than that of what is achievable via the analytical column. To prevent the peak shape of analytes eluting from the analytical column being dictated by the performance characteristic of the trap column, the sample must momentarily trap and re-concentrate at the head of the analytical column before the gradient increase in strength sufficiently enough to cause elution from the analytical column.

4.4 Future directions

While the system has performed well, the peak shape problem likely caused by the trap column is an area that warrants further attention. There are a few ways to potentially solve this problem. The most obvious, given the difference in trap column lengths between the nanoAcquity and XUPLC operation, is to shorten this column. Unfortunately, doing so would require a substantial redesign of fused-silica capillary fittings to create such a solution. Practically, this is not possible in the short-term. Another approach would be to change the packing material from the 5 μm Symmetry C18 particles to something less retentive. A 5 μm BEH-C8 would be a good alternative to test. The last option, which is likely to produce to most significant change in results and improve the chromatography the most, is to change the flow paths. Figure 4-26 is a schematic for the new design, which came about after discussion with Keith Fadgen from Waters Corporation. The former design utilized a forward-trapping, forward-flush methodology, where analytes are trapped at the head of the trap column and then must be flushed through the rest of the trap column before reaching the analytical column. With the redesign, a reversed-trapping, forward-flush approach is possible. This places the analytes at the outlet of the trap column during trapping, enabling them to release and then re-focus at the head of the analytical column without traveling through the trap column. A "cleaner" release is expected, which should result in improved peak shapes and reduced tailing. Operational modes are shown in Figures 4-27, 4-28, and 4-29.

Currently the system has been operated at 30,000 psi. With the flow path configuration used for the collection of the data shown in this chapter, the maximum operating pressure was 40,000 psi due to the gradient storage isolation valve that was required to operate in the open position. The proposed changes to the flow paths would

remove this valve and, given previous experience with this type of valve, one could safely operate the system in excess of 50,000 psi. The likely point of failure would most definitely be the capillary fittings when operating at 50,000 psi. They already are the source of many of the problems experienced during setup at 30,000 psi. Continued work is needed to develop better fittings that are more user-friendly and capable of handling the added pressure.

4.5 References

1. MacNair, J.E.; Patel, K.D.; Jorgenson, J. W., Ultrahigh-Pressure Reversed-Phase Capillary Liquid Chromatography: Isocratic and Gradient Elution Using Columns Packed with 1.0-μm Particles. *Analytical Chemistry* **1999**, 71, 700-708.

2. Neue, U. D. *HPLC Columns: Theory, Technology, and Practice*, Wiley-VCH: New York, 1997.

3. Patel, K. D.; Jerkovich, A. D.; Link, J. C.; Jorgenson, J. W., In-Depth Characterization of Slurry Packed Capillary Columns with 1.0-μm Nonporous Particles Using Reversed-Phase Isocratic Ultrahigh-Pressure Liquid Chromatography. *Analytical Chemistry*, **2004**, 76, 5777-5786.

4. Mellors, J. S.; Jorgenson, J. W., Use of 1.5-μm Porous Ethyl-Bridged Hybrid Particles as a Stationary-Phase Support for Reversed-Phase Ultrahigh-Pressure Liquid Chromatography. *Analytical Chemistry*, **2004**, 76 5441-5450.

5. Wang, X. L.; Barber, W. E.; Carr, P. W., A practical approach to maximizing peak capacity by using long columns packed with pellicular stationary phases for proteomic research. *Journal of Chromatography A,* **2006**, 1107, 139-151.

6. Wang, X. L.; Stoll, D. R.; Schellinger, A. P.; Carr, P. W., Peak capacity optimization of peptide separations in reversed-phase gradient elution chromatography: fixed column format. *Analytical Chemistry*, **2006**, 78, 3406-3416

7. Zhou, F.; Lu, Y.; Ficarro, S. B.; Webber, J. T.; Marto, J. A., Nanoflow Low Pressure High Peak Capacity Single Dimension LC-MS/MS Platform for High-Throughput, In-Depth Analysis of Mammalian Proteomes. *Analytical Chemistry*, **2012**, 84, 5133-5139.

8. Martosella, J.; Zolotarjova, N.; Liu, H.; Nicol, G.; Boyes, B. E., Reversed-Phase High-Performance Liquid Chromatographic Prefractionation of Immunodepleted Human Serum Proteins to Enhance Mass Spectrometry Identification of Lower-Abundant Proteins. *Journal of Proteome Research* **2005**, *4* (5), 1522-1537.

9. Dowell, J. A.; Frost, D. C.; Zhang, J.; Li, L., Comparison of Two-Dimensional Fractionation Techniques for Shotgun Proteomics. *Analytical Chemistry* **2008**, *80* (17), 6715-6723.

10. Link, J. C., *Development and Application of Gradient Ultrahigh Pressure Liquid Chromatography for Separations of Complex Biological Mixtures*. PhD dissertation, University of North Carolina at Chapel Hill. Chapell Hill, **2004**.

11. Evans, C. E. *Multidimensional Liquid Chromatography Coupled to Mass Spectrometry for the Analysis of Complex Mixtures of Proteins*. PhD dissertation, University of North Carolina at Chapel Hill, Chapel Hill, **2007.**

12. Richardson, B. *Multidimensional Separation of Intact Proteins for Differential Proteomics Employing Topdown and Bottom-up Proteomic Strategies.* PhD dissertation, University of North Carolina at Chapel Hill. Chapel Hill, **2010**.

13. Thomson, J. D.; Carr, P., High-Speed Liquid Chromatography by Simultaneous Liquid Optimization of Temperature and Eluent Composition. *Analytical Chemistry* **2002**, 74-4150-4159.

14. Chen, H.; Horvath, C. J. J., High Speed High-Performance Liquid Chromatography of Peptides and Proteins. *Journal of Chromatography A*, **1995**, 705, 3-20.

4.6 Tables

Column	Analysis Time (min)	Acetonitrile Change	Peak Capacity	Peak Capacity / Min	Ave Peak Width, FWHM (min)	Proteins ID'd	Proteins ID'd / Min
200 cm, 1.9 µm	240	7.2%	660	2.75	0.25	153	0.64
117 cm, 1.5 µm	180	3.4%	570	3.17	0.26	128	0.71
25 cm, 1 µm	60	2.0%	160	2.67	0.26	58	0.97
	90	1.2%	250	2.78	0.26	85	0.94
	120	0.9%	250	2.08	0.37	119	0.99
nanoAcquity QTOF Premier	120	1.1%	180	1.50	0.51	80	0.67
nanoAcquity SynaptG2	120	1.1%	180	1.50	0.51	485	4.04

Table 4-1. Three columns were evaluated on the XUPLC system. These include a 200-cm column packed with 1.9 µm particles, a 117-cm column packed with 1.5 µm particles, and a 25-cm column packed with 1 µm particles. The inner diameter for each column was 75 µm. The pressure supplied by the Haskel -903 pump was approximately 30,000 psi and the temperature was held at 65°C. Gradients of water and acetonitrile were used to elute the peptides and started between 1% & 3% acetonitrile with a final acetonitrile concentration of 40% at the end of the gradient. Peak capacities were calculated by dividing the elution window, which was defined as the period of time between the first eluting peak and last eluting peak, by the FWHM peak width. The analysis time corresponds to the amount of time required from injection-to-injection. Detection was by a Waters QTOF Premier mass spectromter. Proteins were identified with PLGS 2.5. The last two lines of the table show data from a standard nanoAcquity setup utilizing a 90-minute gradient previously determined as the preferred gradient length with one data set collected with a QTOF Premier and the other with a Waters Synapt G2 and are included for comparison with the XUPLC separations.

4.7 Figures

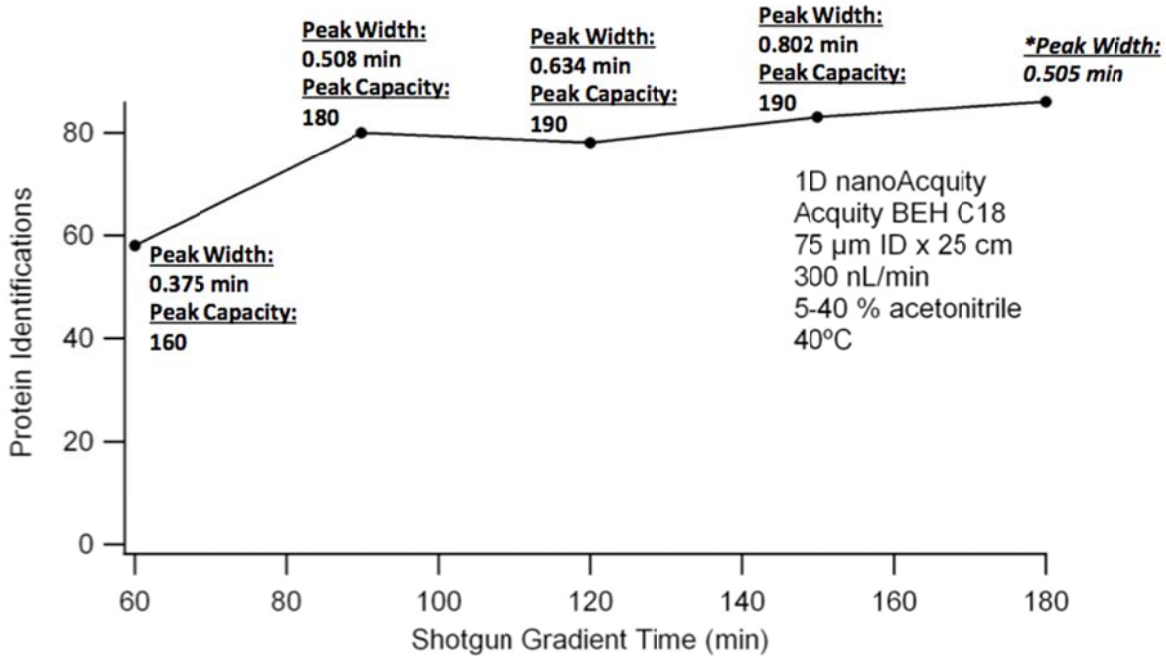


Figure 4-1. The Apex3D algorithm used to process the raw mass spectral data attempts to fit Gaussian peaks to the eluting peptides. From this data, the full-width, half-maximum (FWHM) of the peaks are readily available. Peak capacities here are calculated from the FWHM. Gradients longer than 90 minutes were not shown to give more protein identifications. There was a slight increase in peak capacity in lengthening the gradient from 90- to 120-minutes. At 180 minutes, the gradient becomes shallow enough and the peaks tail significantly enough that the Apex3D algorithm begins fitting multiple peaks per actual peak.

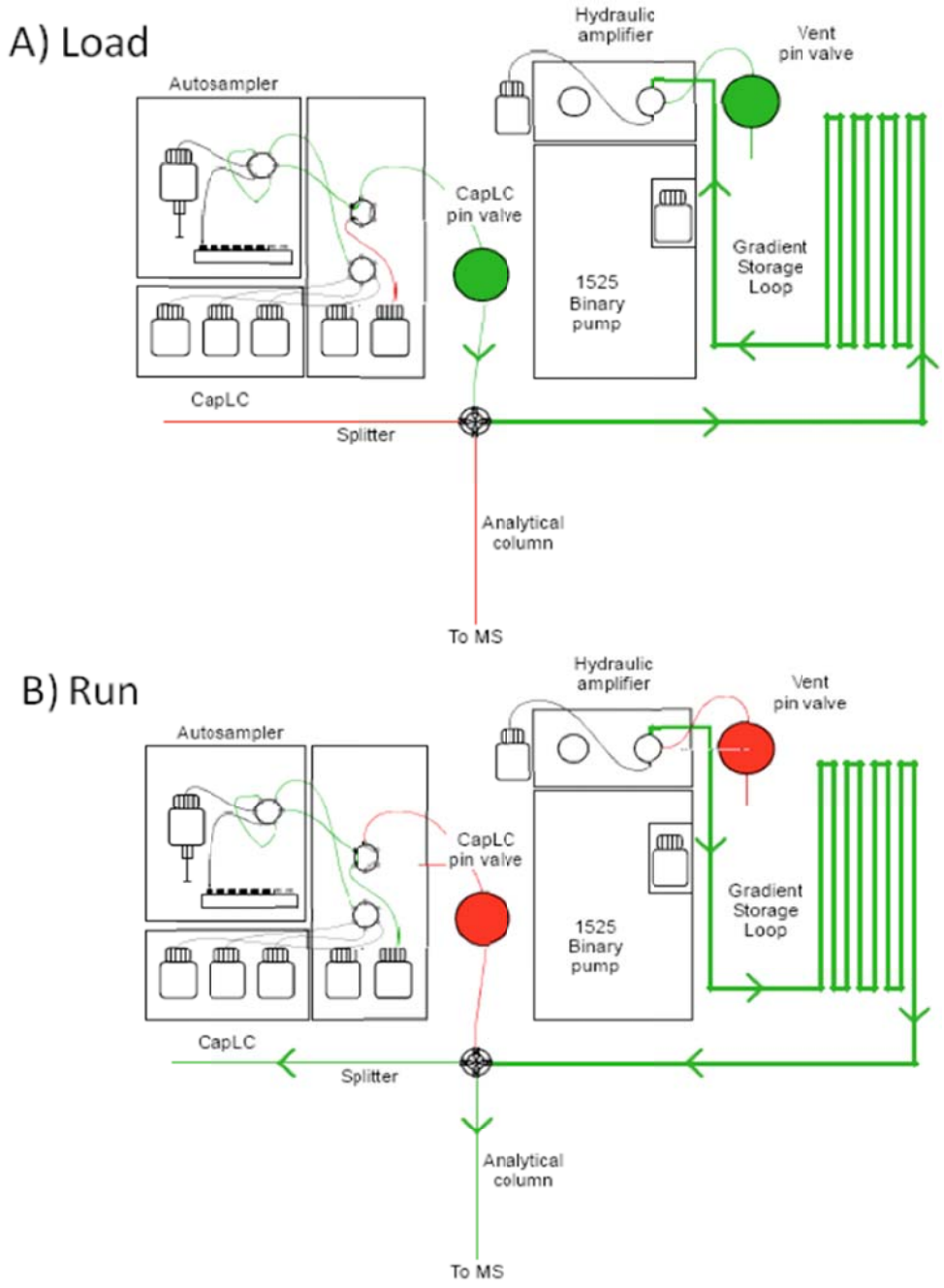


Figure 4-2. The prototype hydraulic amplifier system previously used for automated UHPLC gradient analyses. During sample loading and gradient formation (A), flow would originate from the CapLC system and fill the gradient storage loop. After loading, the two "pin valves" or Valco on/off valves would be closed and the hydraulic amplifier activated, forcing flow out of the storage loop and through the column and splitter.

Figure 4-3. A Waters nanoAcquity UPLC with detection by a Waters QTOF-Premier mass spectrometer. The pressure limit of the nanoAcquity is 10,000 psi.

Figure 4-4. The nanoAcquity UPLC with the XUPLC modification installed. Pneumatic components are primarily locatd on the left side of the nanoAcquity stack with the exception of the Valco on/off valves. The column heater is mounted to position the outlet of the capillary column near the mass spectrometer source. A gradient storage loop is wound around a spool and mounted below the temperature control module on the lower right of the nanoAcquity.

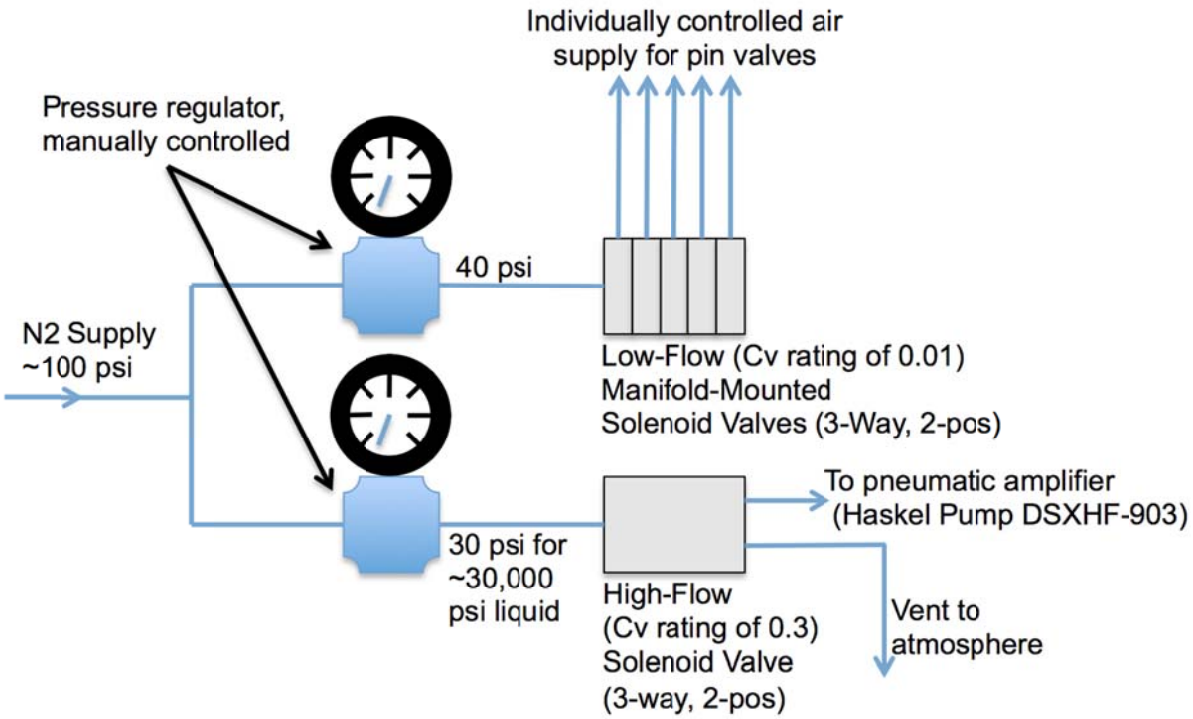


Figure 4-5. A general schematic for the pneumatics sub-system. Gas supply (air or N_2) needs to be provided at approximately 100 psi. Two regulators are need, one to regulate down the pressure to 40 psi for the solenoid valves responsible for switching the Valco on/off valves and the other to supply pressure to the Haskel -903 pneumatic amplifying liquid pump. The outlet pressure of the -903 is roughly equal to 1,000 psi liquid for 1 psi gas.

Figure 4-6. A photograph of the pneumatics sub-system showing mounting locations. Note the ball valves for manual shut-off and release of gas pressure to the -903. These are installed as a safety mechanism in case rapid decompression is required without the use of the computer controls.

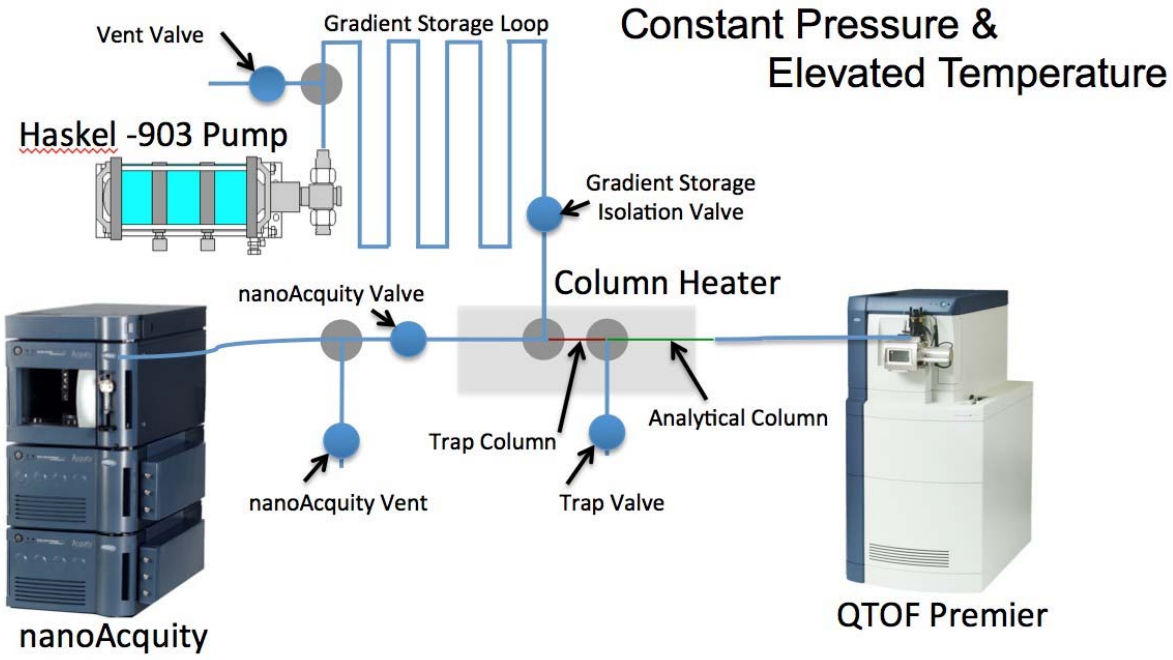
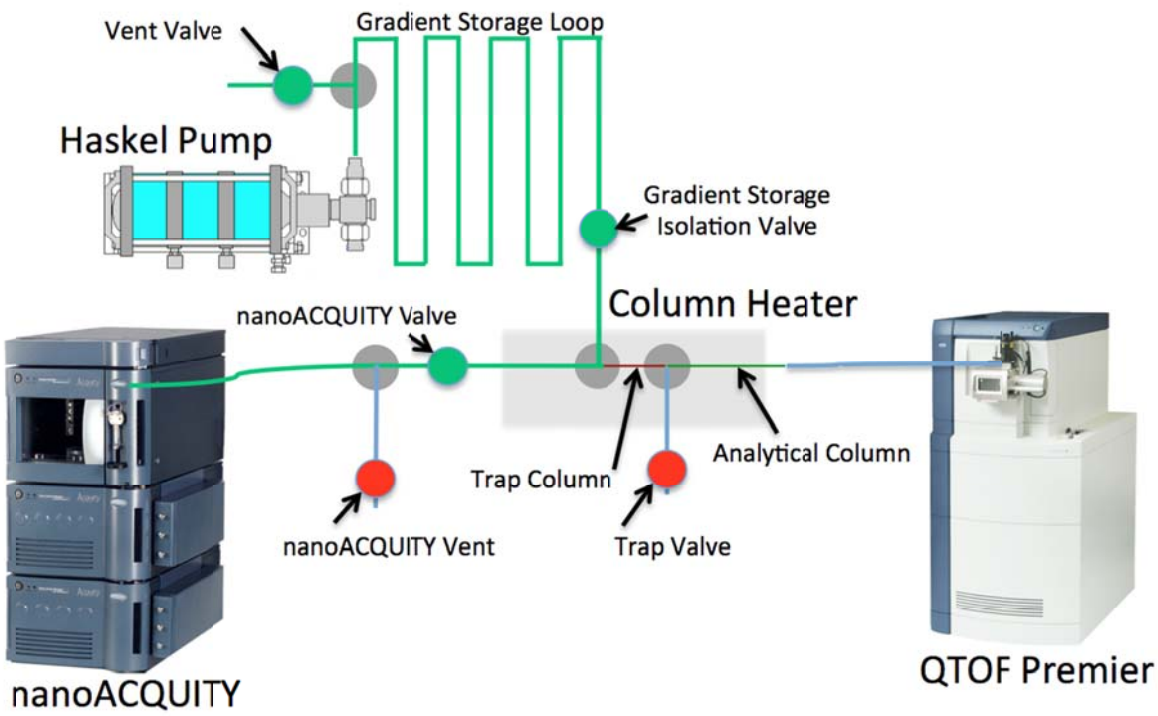
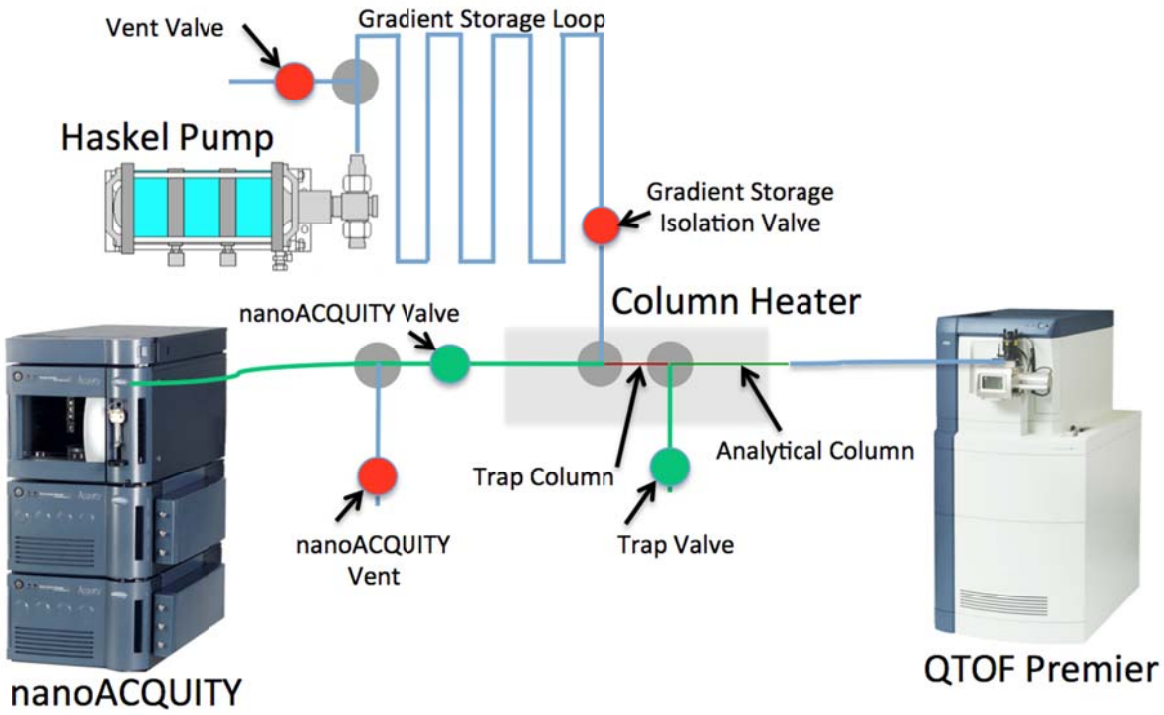


Figure 4-7. Fluidic schematic overview for the XUPLC system. The blue circles represent Valco on/off valves. Grey circles represent tees from Valco. The trap column, analytical column, and associated tees are housed inside the Waters TCMII for elevated temperature operation.



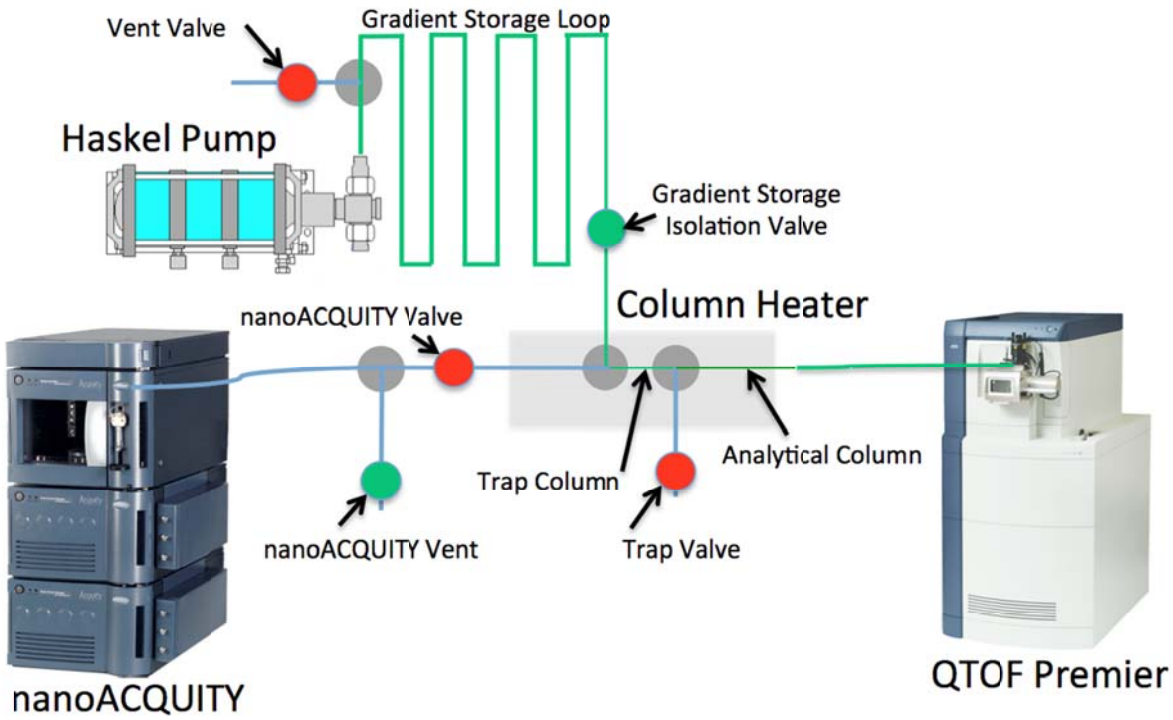
	nanoLC Valve	Trap Valve	Vent Valve	nanoLC Vent	Isolation Valve
Gradient Loading	**Open**	**Closed**	**Open**	**Closed**	**Open**
Sample Trapping	Open	Open	Closed	Closed	Closed
Running	Closed	Closed	Closed	Open	Open

Figure 4-8. During gradient loading, the nanoAcquity vent valve and trap valve are closed. Flow from the nanoAcquity fills the gradient storage loop. The resistance to flow from the analytical column is sufficient enough to prevent the gradient from being split between the storage loop and the column.



	nanoLC Valve	Trap Valve	Vent Valve	nanoLC Vent	Isolation Valve
Gradient Loading	Open	Closed	Open	Closed	Open
Sample Trapping	**Open**	**Open**	**Closed**	**Closed**	**Closed**
Running	Closed	Closed	Closed	Open	Open

Figure 4-9. During sample loading and trapping the valves are configured to allow flow from the nanoAcquity through the trap column and then to waste. Analytes are trapped on the column. Trapping conditions are typically 0.5% acetonitrile and a minimum volume of three times the sample loop volume is used to ensure all of the sample has been flushed onto the column. During this operation, the gradient storage loop is isolated to prevent any changes to the gradient.



	nanoLC Valve	Trap Valve	Vent Valve	nanoLC Vent	Isolation Valve
Gradient Loading	Open	Closed	Open	Closed	Open
Sample Trapping	Open	Open	Closed	Closed	Closed
Running	**Closed**	**Closed**	**Closed**	**Open**	**Open**

Figure 4-10. During running operation the vent, nanoAcquity and trap valves are closed and the gradient storage isolation valve is opened. This forces flow driven by the -903 pump through the storage loop, past the trap column where analytes are eluted, and onto the analytical column where analytes are briefly re-focused prior to elution. The nanoAcquity vent valve is open during operation to prevent damage to the nanoAcquity should the nanoAcquity valve fail.

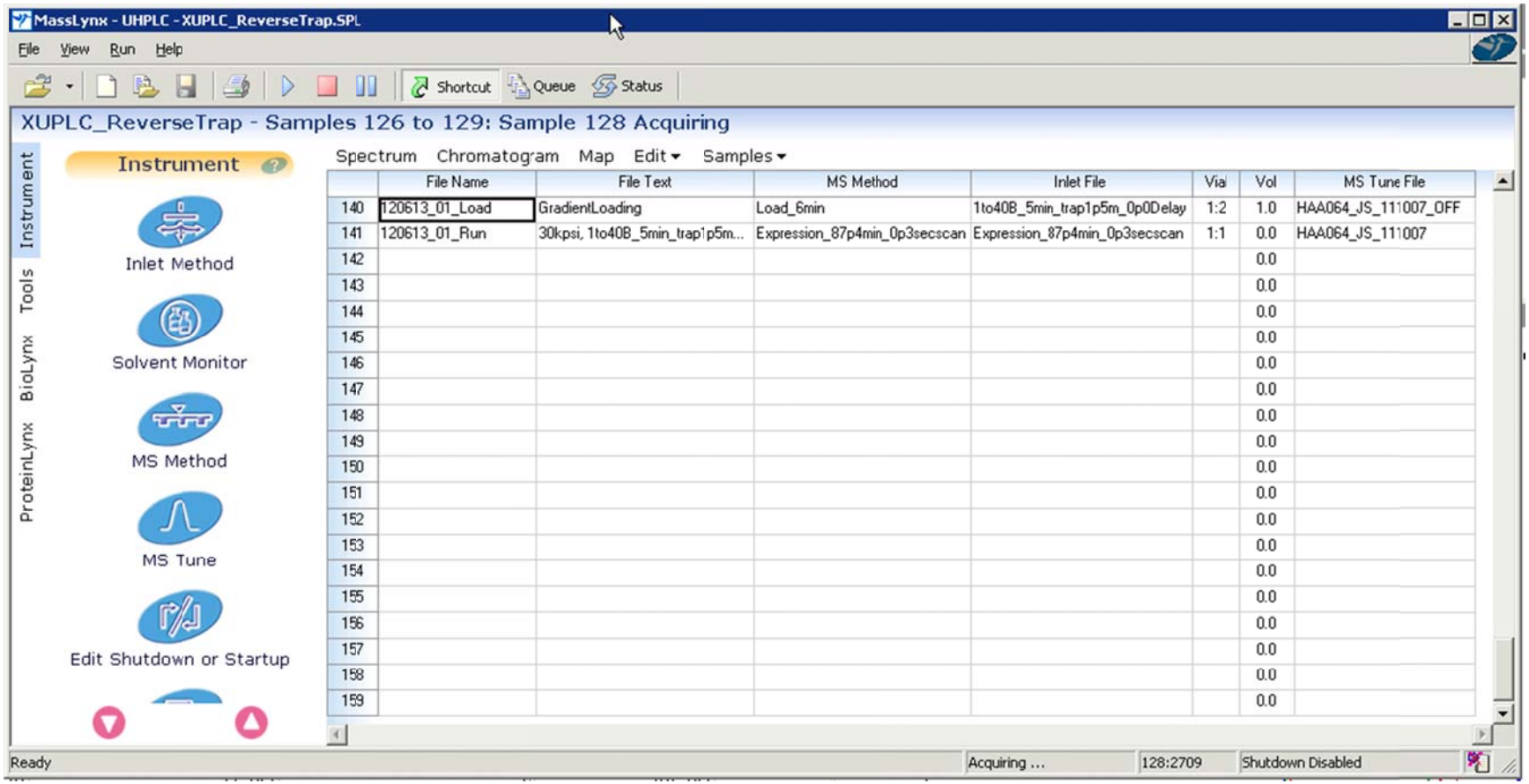


Figure 4-11. Typically, one line of the sample list corresponds to an LC/MS analysis. For XUPLC operation, a minimum of two lines and frequently more are required for a given sample injection. The first row is responsible for loading the gradient. The second line allows for the sample injection and starts the mass spectrometer. Often a third row is used to stop the analysis and return the system to initial conditions for the next run, although two can be used for shorter analyses. If more lines are needed, they can be added in an intermediate position in order to maintain valve states for running operation.

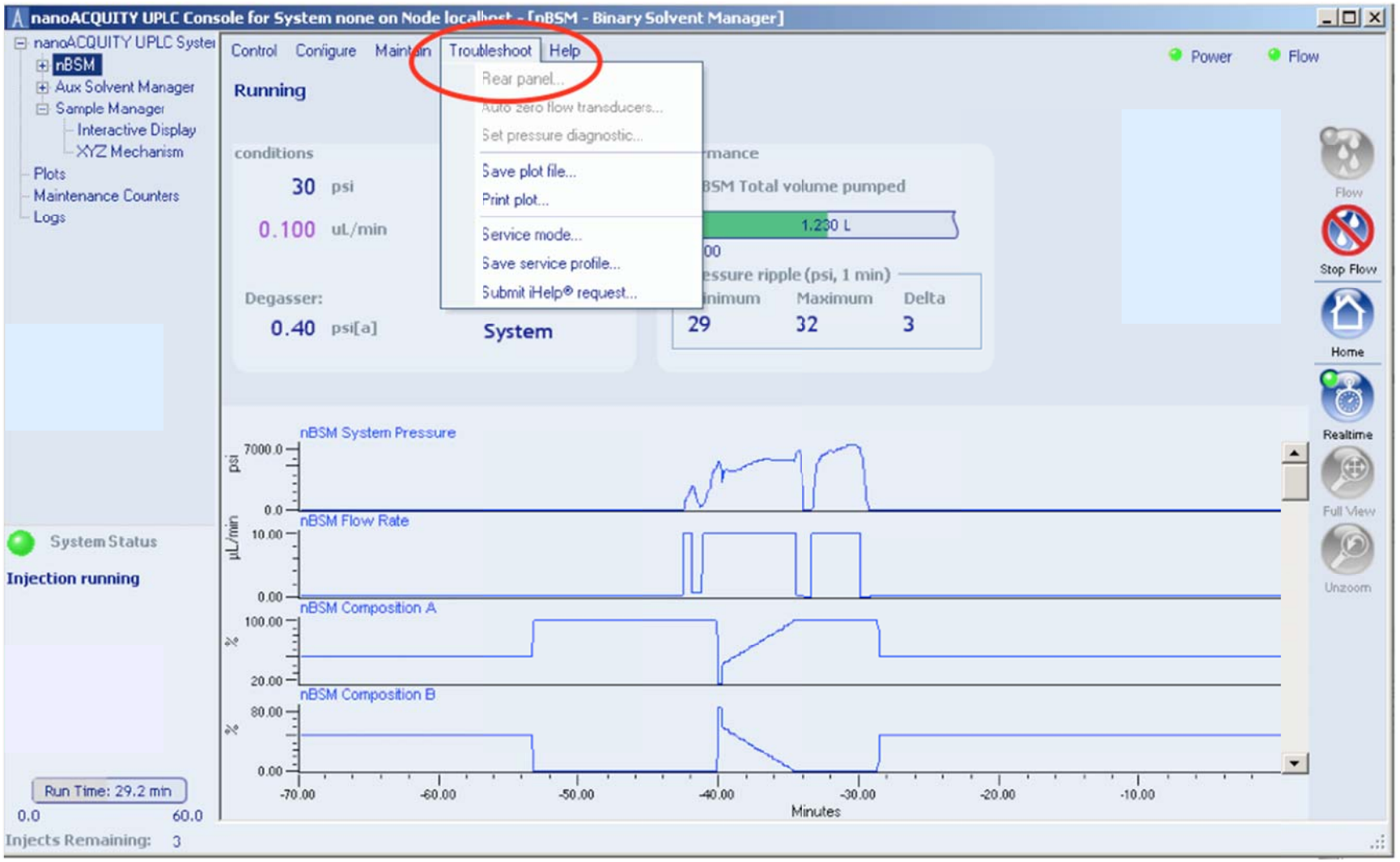


Figure 4-12. On the rear of the nanoAcquity are several switches required for XUPLC operation. These control the Valco on/off valves responsible for manipulating the flow path. In order to access these during manual operation, one must navigate to the "Rear panel…" menu under "Troubleshoot" for both the pump and autosampler.

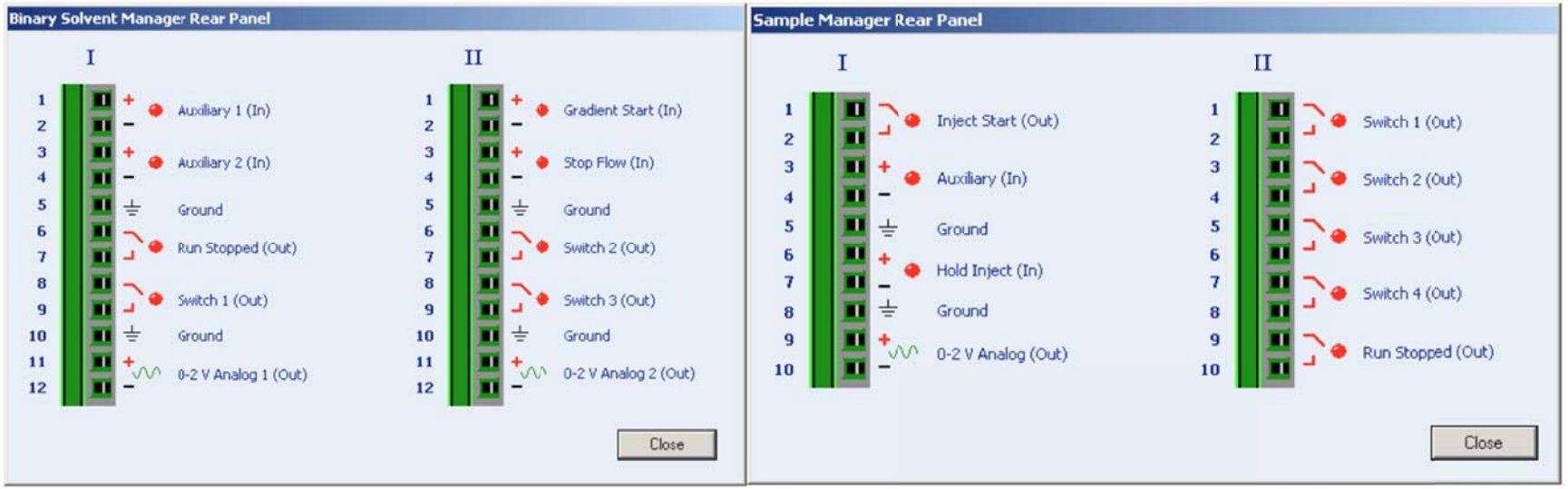


Figure 4-13. Within the rear panel are many switches. The ones needed for operation are switches 1, 2, and 3 on the binary solvent manager rear panel and switches 1, 2, 3, and 4 on the sample manager rear panel.

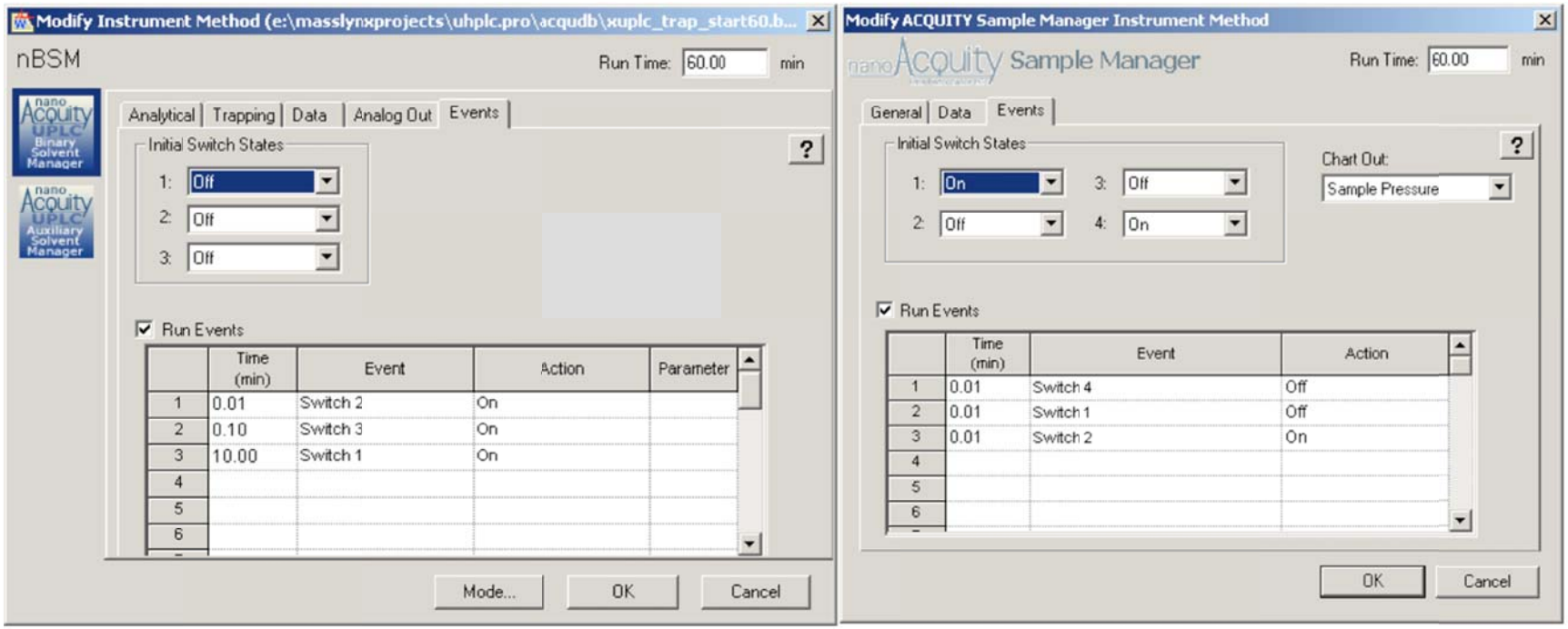


Figure 4-14. For automated operation, timing events are included in inlet file, which is the method file used by the nanoAcquity during operation. For both the pump and autosampler, there is an events tab where switches and their associated on/off states can be defined. To aid in method development, a spreadsheet was developed to allow the user to more easily program the methods.

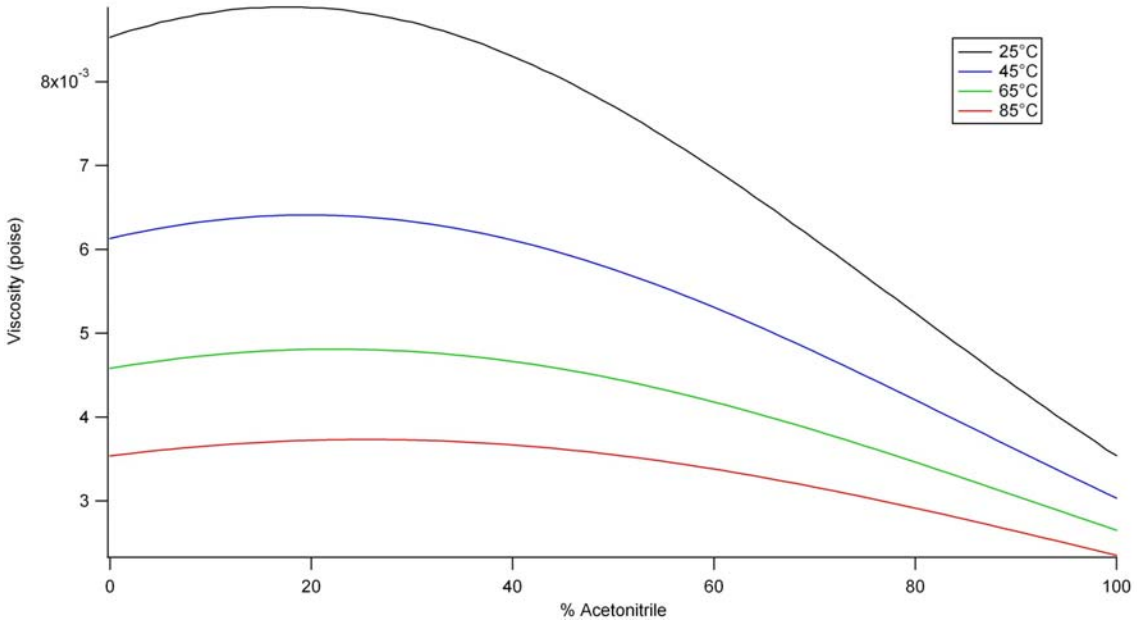


Figure 4-15. Curves showing the change in viscosity as a function of the acetonitrile percentage and the temperature are shown here at 25, 45, 65, and 85°C spanning 0% to 100% acetonitrile. Most gradients range from 1-40% acetonitrile.[13,14]

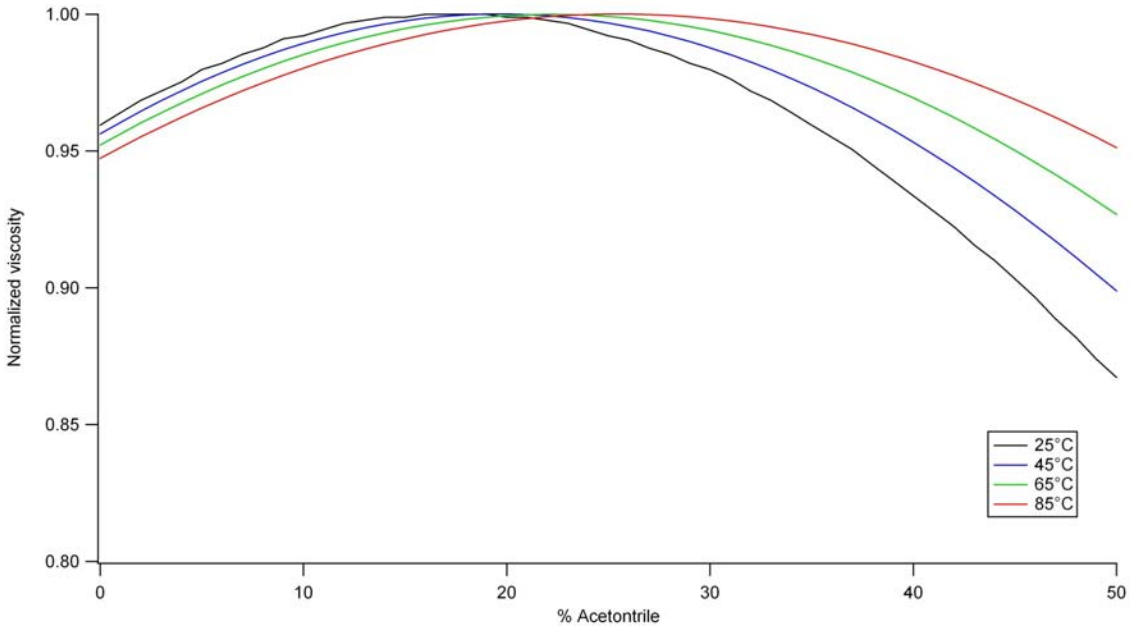


Figure 4-16. Data shown in Figure 4-15 was normalized in order to demonstrate the % change in viscosity. The figure was also modified to show the difference in viscosity across a more typical LC gradient, spanning 0-50% acetonitrile.

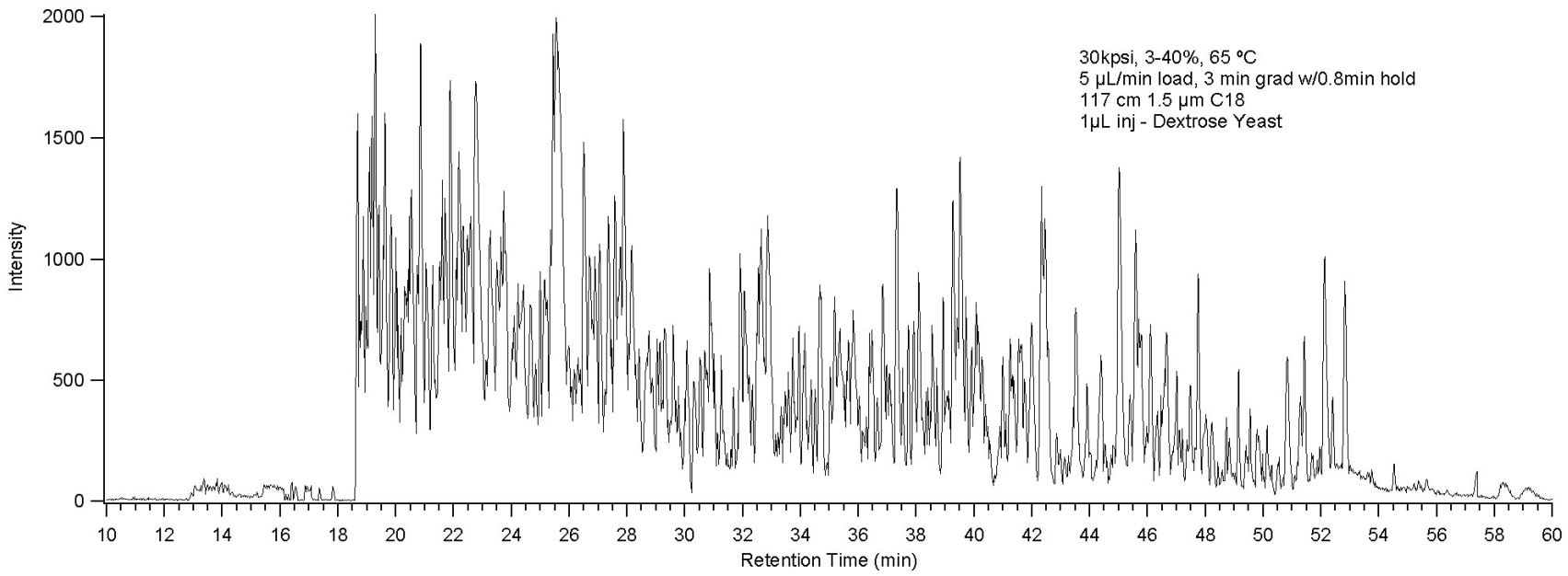


Figure 4-17. Method development for the XUPLC system is more difficult than for standard nanoAcquity UPLC operation. Clipping of the front-end of the gradient can occur if an incorrect delay time is used. The delay is necessary to account for the system volume between the pump and the gradient storage loop. This can be observed in the chromatogram shown as at 18.5 minutes there is a sudden change from baseline conditions to full-intensity peptide elution. A more appropriate delay volume would result in a gradual rise from the baseline as peptides begin to elute.

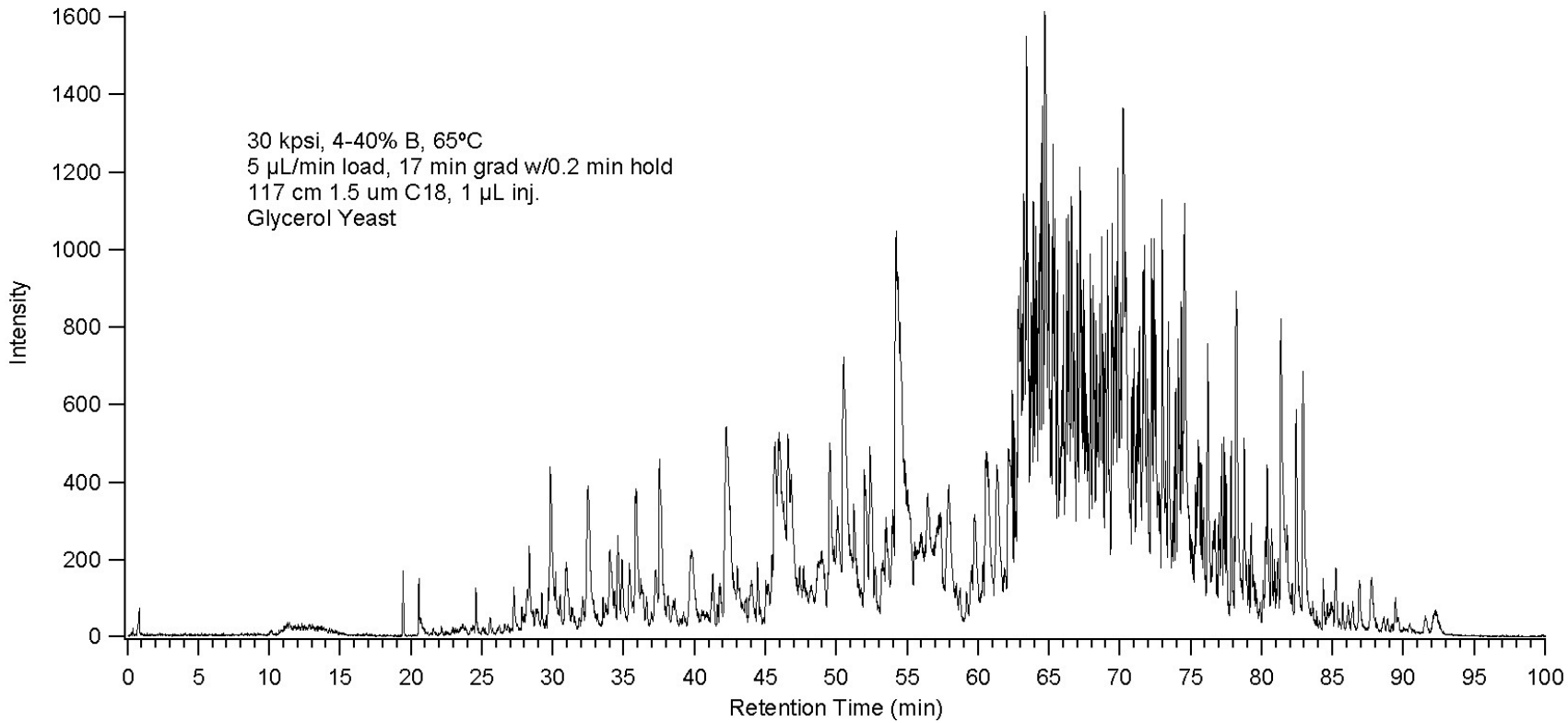


Figure 4-18. As a result of the elevated pressures (30,000 psi), leaks can develop. These are routinely found on the fused-silica capillary fittings. The volume of gradient loaded should have resulted in a nearly 300-minute long analysis time. From this chromatogram, it would appear as though a leak developed at approximately 63 minutes. A leak is suspected as the gradient is unexpectedly progressing much more rapidly with a significant number of peaks being detected in a compressed timescale. The location of the leak is likely one of the three connections on the tee prior to the trapping column as there are still peptides being detected. A leak at the second tee results in a significant loss of sample and no peaks are seen as a result.

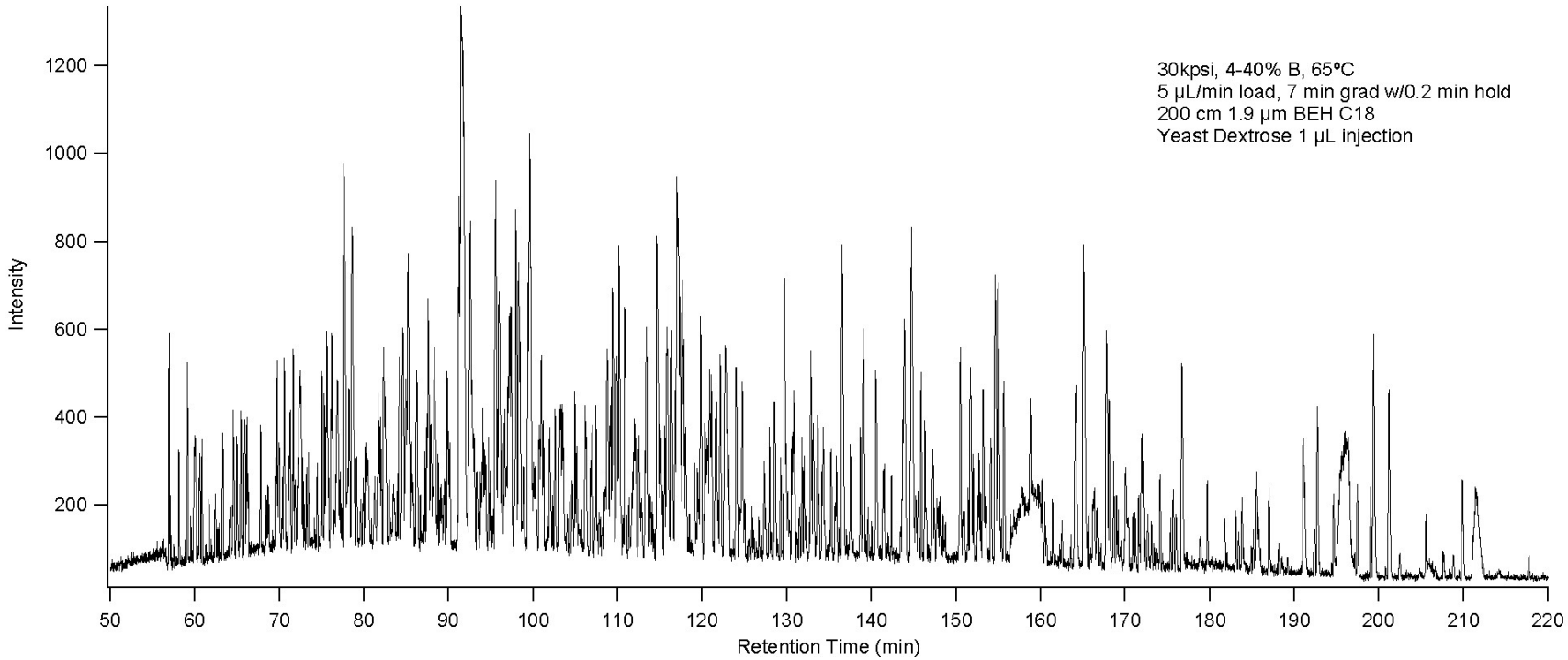


Figure 4-19. A 200-cm x 75 µm ID capillary column packed with 1.9 µm BEH C18 particles was operated at 30,000 psi and 65°C. The gradient was 35 µL of 4-40% acetonitrile. A peak capacity from the full-width, half-maximum data outputted from ProteinLynx yields a peak capacity of 660 for this separation. A total of 153 proteins were identified from this separation.

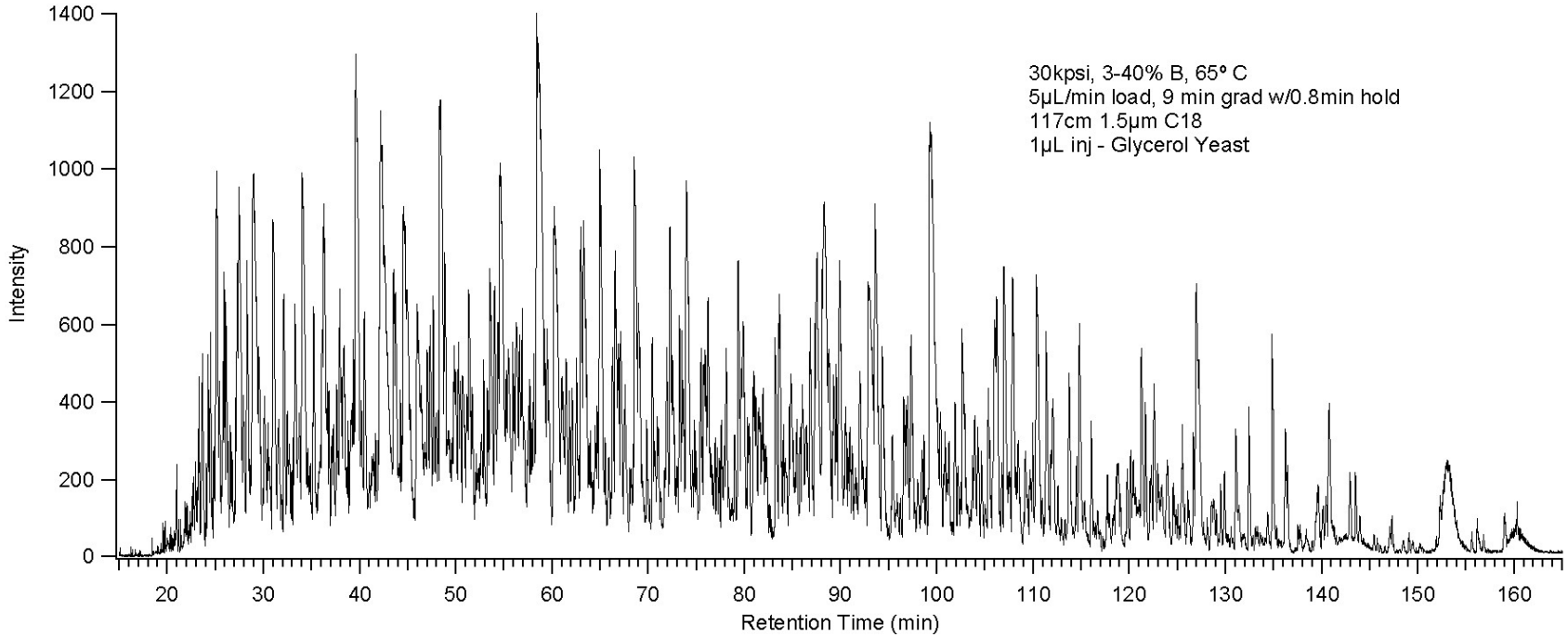


Figure 4-20. A 117-cm x 75 µm ID capillary column packed with 1.5 µm BEH C18 particles was operated at 30,000 psi and 65°C. The gradient was 45 µL of 3-40% acetonitrile. A peak capacity from the full-width, half-maximum data outputted from ProteinLynx yields a peak capacity of 570 for this separation. A total of 128 proteins were identified in this separation.

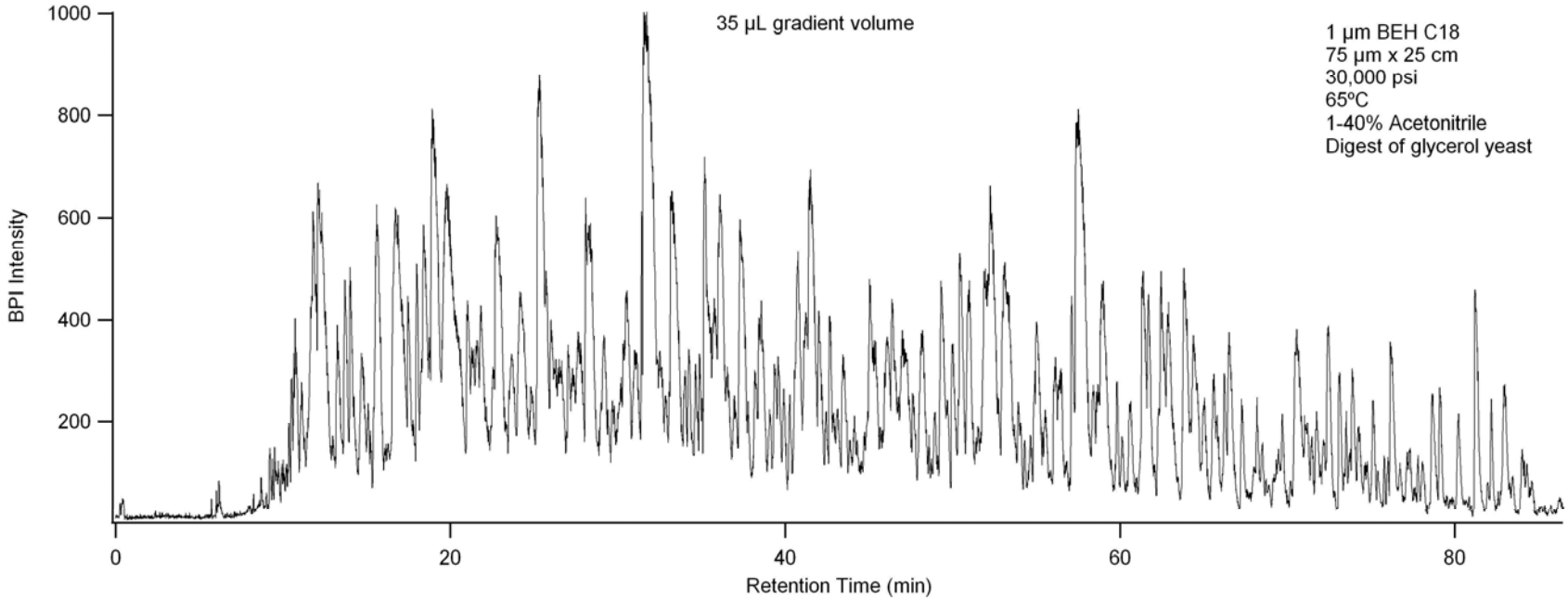


Figure 4-21. The elevated pressure permits the use of longer columns with modestly sized particles (1.5 and 1.9 µm). Also possible is the use of 1 µm particles. A 25 cm x 75 µm ID capillary column was packed with 1 µm BEH C18 particles and operated at 65°C. The gradient had a volume of 35 µL spanning 1-40% acetonitrile. Run times more consistent with what is typically of the nanoAcquity UPLC are possible with this shorter column. In this analysis, 119 proteins were identified in a total analysis time of 120 minutes with a peak capacity of 250. For standard nanoAcquity operation, a 90-minute gradient resulting in a total run time of 120 minutes resulted in 80 protein identifications and a peak capacity of 190.

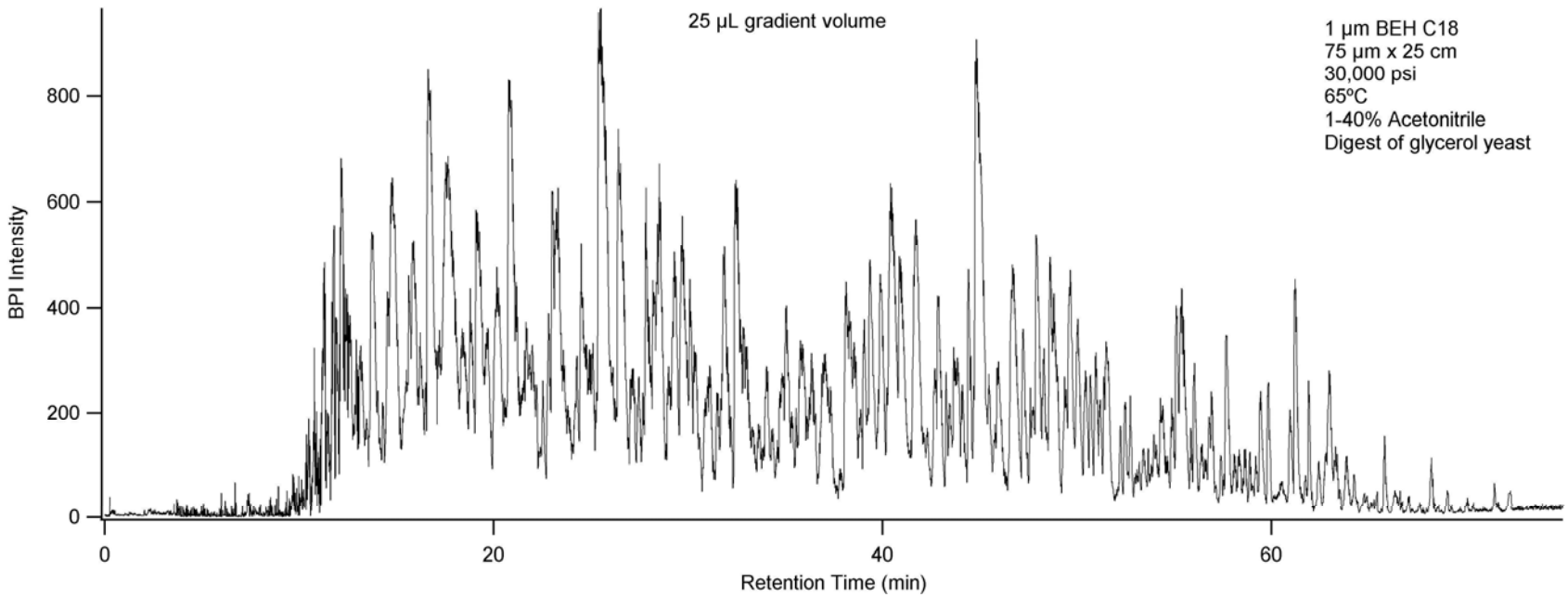


Figure 4-22. A 25 cm x 75 µm ID capillary column was packed with 1 µm BEH C18 particles and operated at 65°C. The gradient had a volume of 25 µL spanning 1-40% acetonitrile. Run times more consistent with what is typically of the nanoAcquity UPLC are possible with this shorter column. In this analysis, 85 proteins were identified in a total analysis time of 120 minutes with a peak capacity of 250. The same peak capacity was the result of narrower peaks with this length of gradient as compared to the gradient used in for the chromatogram in Figure 4-21. The reduction in identified peaks can likely be attributed to the narrow peaks not being characterized as well by the Apex3D algorithm within ProteinLynx.

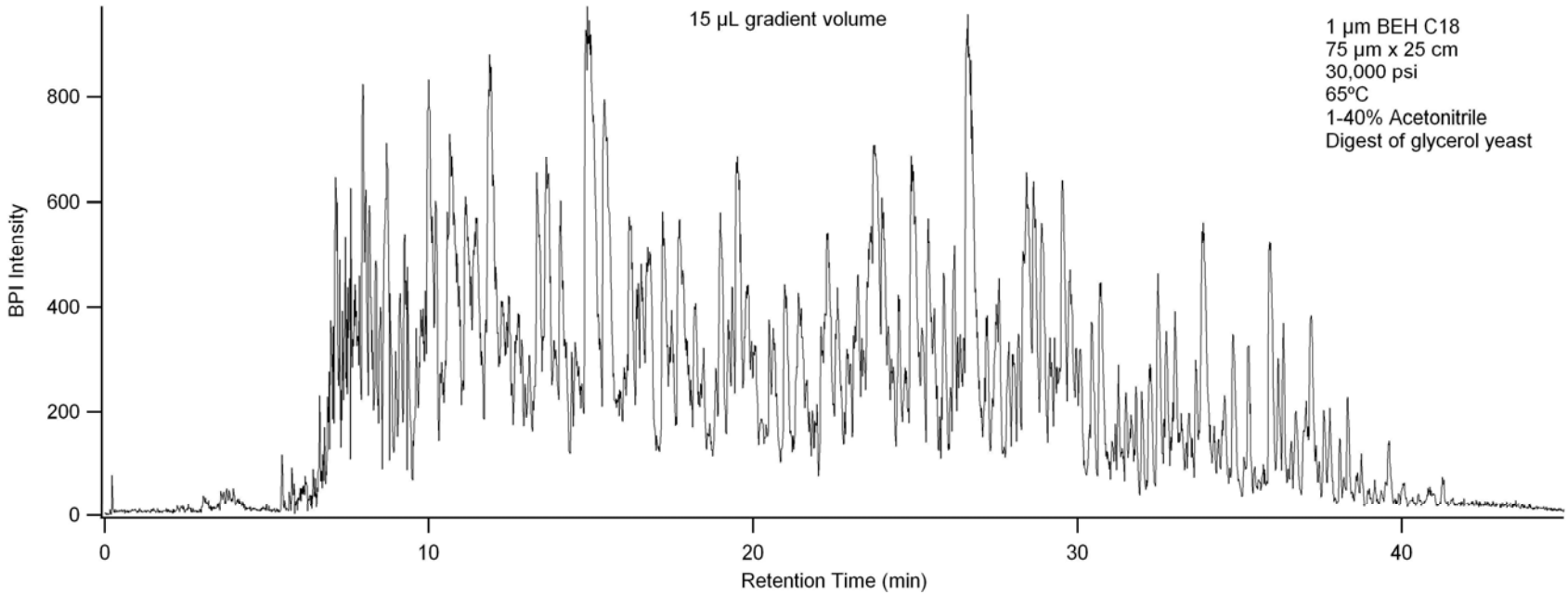


Figure 4-23. A 25 cm x 75 µm ID capillary column was packed with 1 µm BEH C18 particles and operated at 65°C. The gradient had a volume of 15 µL spanning 1-40% acetonitrile. In this analysis, 59 proteins were identified in a total analysis time of 120 minutes with a peak capacity of 160. This method requires 60 minutes from injection-to-injection and has a comparable peak capacity to the nanoAcquity in 2/3 the analysis time.

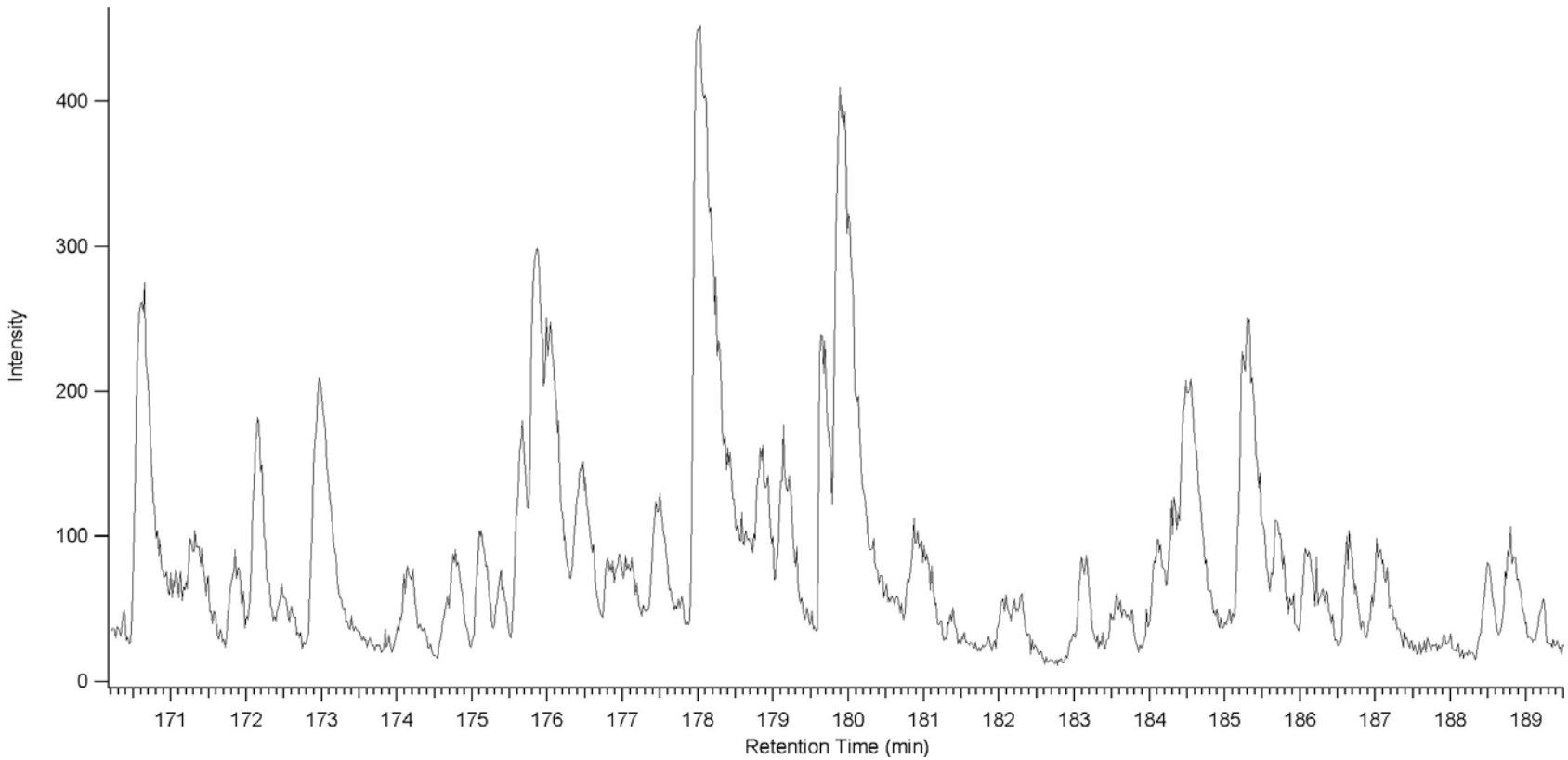


Figure 4-24. The same 117-cm column with 1.5 μm particles as used in the analysis shown in Figure 4-20 was used here with a gradient approximately twice the length. There was only a slight improvement in peak capacity and fewer proteins were identified. Close examination of the peak shape reveals substantial tailing. Mass overload is not suspected as this is exhibited by low intensity peaks as well. Reducing the mass of sample injected did not result in peak shape improvements either. The forward-trapping, forward-flush operation is suspected to be contributing to this problem. The trap column is much longer (6.5 cm) than what is typically used in this type of operation on the nanoAcquity (2 cm). This additional length may be broadening the peaks sufficiently such that they cannot properly focus at the head of the analytical column.

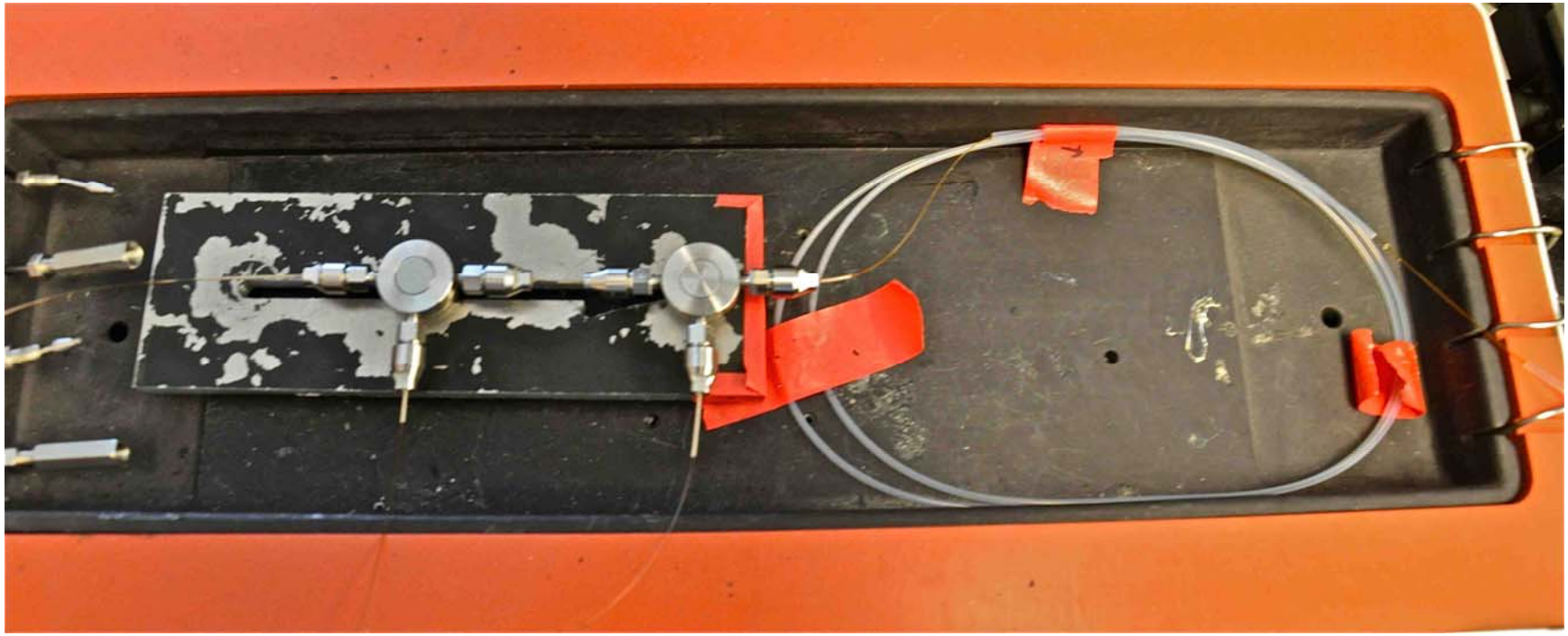


Figure 4-25. Inside the column heater module are the trap column, analytical, and hardware required for operation. The two tees and either end of the trap column are mounted to a slotted piece of ¼" aluminum plate to facilitate assembly. The tees can be mounted with a screw from the back. The analytical column is coiled and taped to the surface of the heater to ensure that it is at the proper temperature. A Teflon junction is used to connect the outlet of the column, which remains inside the heater, to a piece of capillary serving as a transfer line between the column and the spray tip.

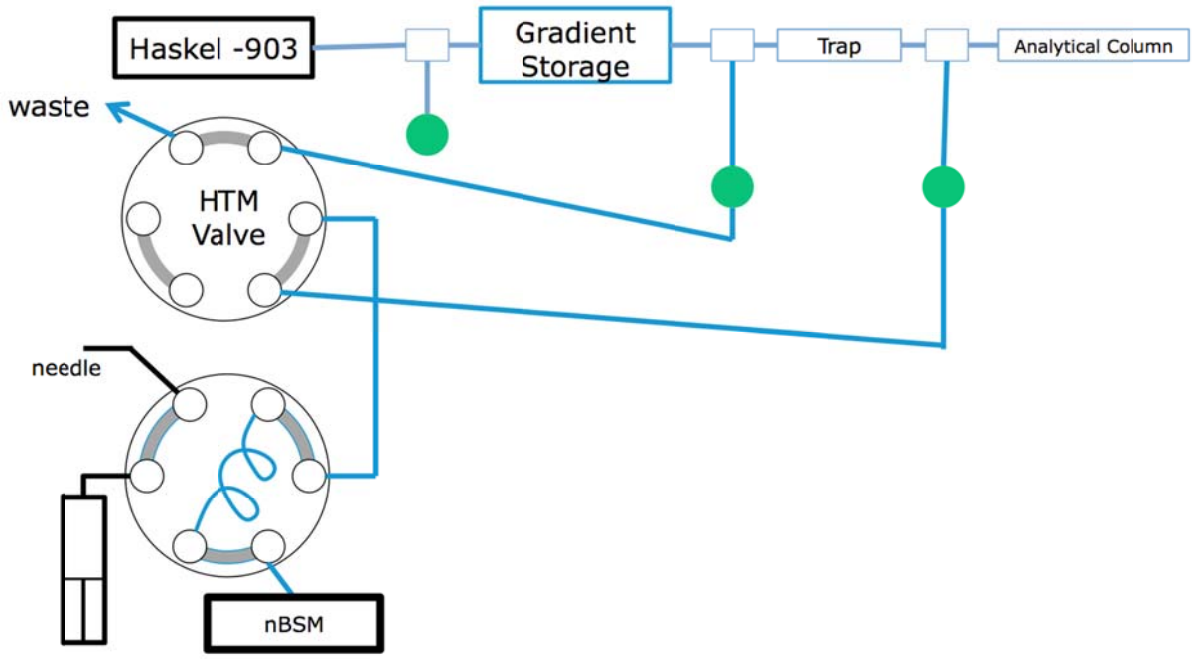


Figure 4-26. A fluid schematic for reversed-trapping with a forward flush. This method utilizes both the inject valve and the trapping (HTM) valve that are part of the nanoAcquity. This proposed flow paths also remove the gradient storage loop isolation on/off valve, removing the 40,000 psi upper limit from the system. The reversed-trapping operation should provide for improved peak shape as the sample does not have to travel through the length of the trap column after release.

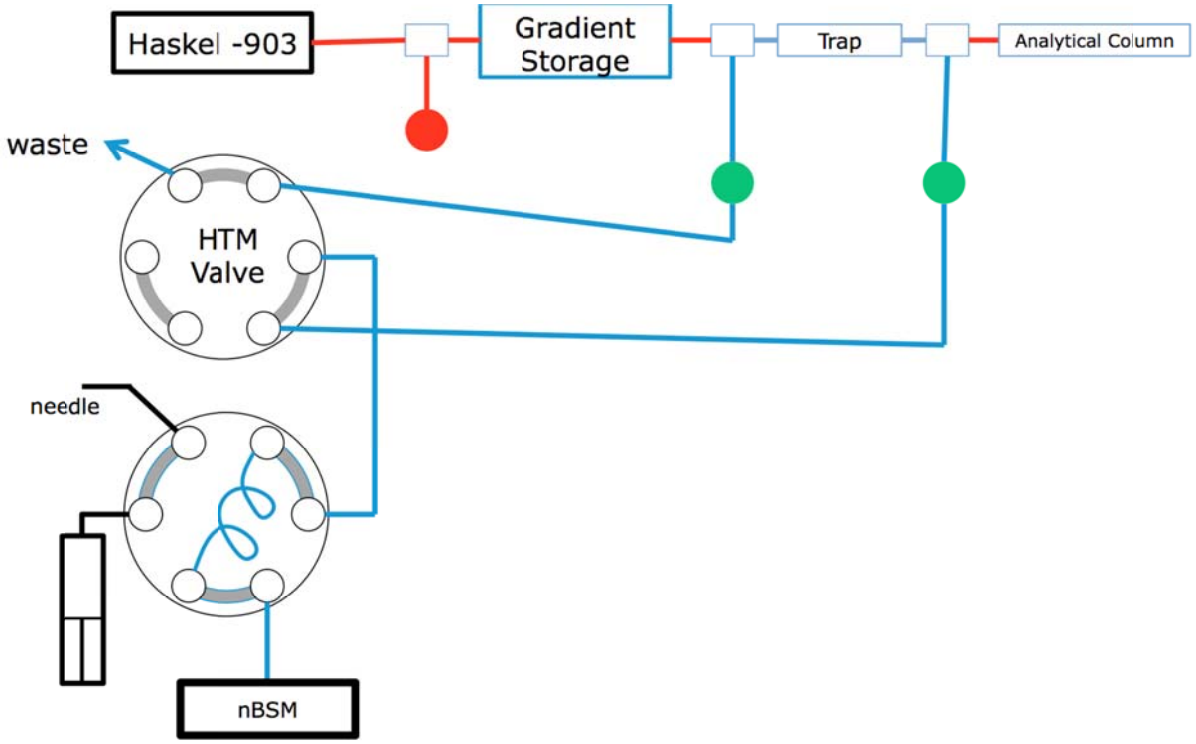


Figure 4-27. Valve configuration for sample loading and trapping. In this configuration, sample is placed onto the end of the trap column.

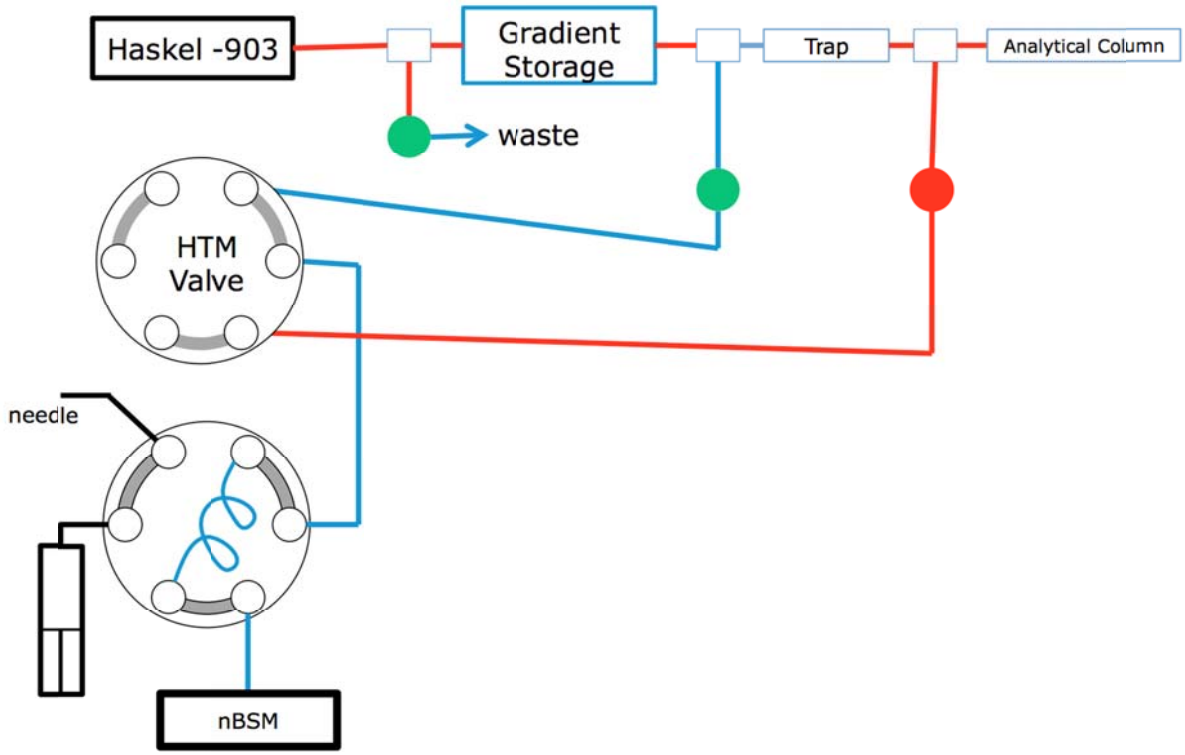


Figure 4-28. Valve configuration for gradient loading. Using the second valve (HTM Valve) on the nanoAcquity facilitates reversed-trapping as after the sample is trapped, by switching the HTM Valve and associated on/off valves, the gradient can be loaded.

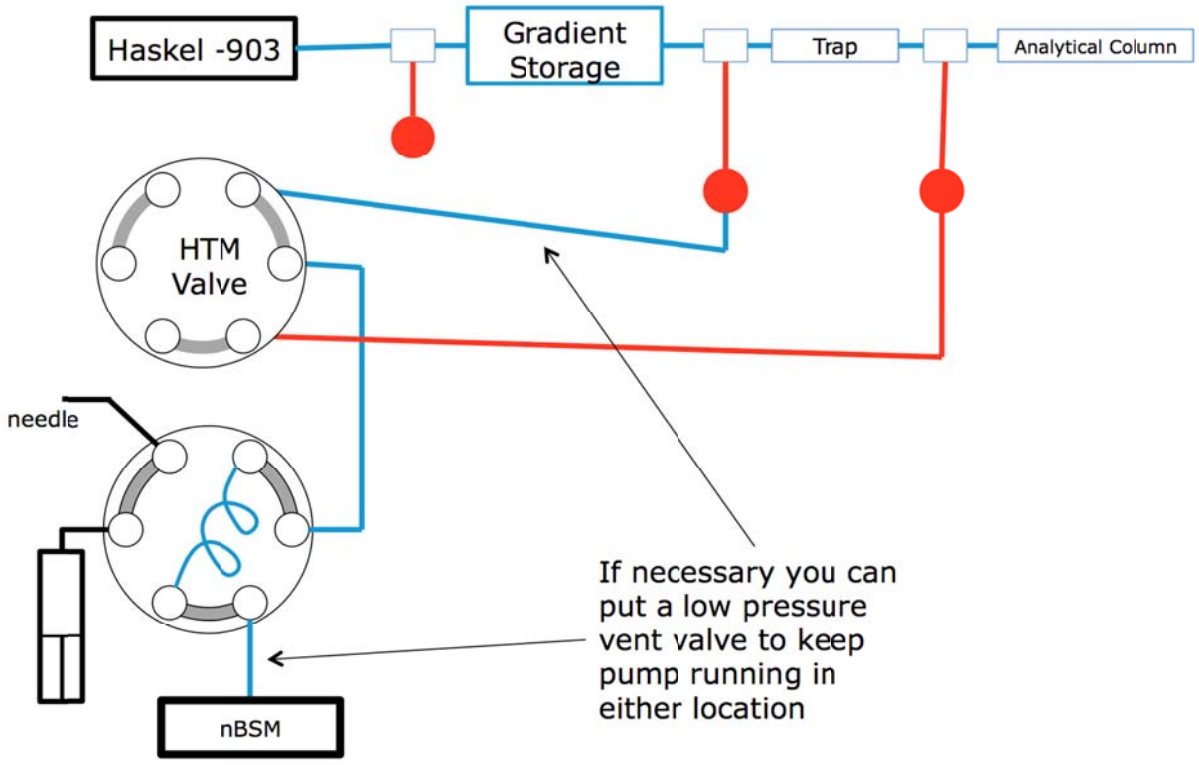


Figure 4-29. Valve configuration for running under XUPLC conditions. Adding the low pressure valve, as indicated in the figure, would allow for a gradual reduction in flow from the nanoAcquity after the gradient is loaded. It would also allow the pump to continue pumping at a low rate during the analytical portion of the method and is likely a worthwhile investment.

CHAPTER 5: APPLICATION OF PRE-FRACTIONATION METHODOLOGY FOR A DIFFERENTIAL ANALYSIS OF A MOUSE EMBRYONIC FIBROBLAST BETA-ARRESTIN 1,2 DOUBLE-KNOCKOUT

5.1 Introduction

5.1.1 Beta-arrestin background

G-protein coupled receptors (GPCRs), which are seven-transmembrane domain proteins, comprise the largest and most diverse class of cell surface receptors.[1] It has been estimated that nearly half of all drugs available for clinical use target these proteins whether to block a signaling pathway altogether or to stimulate a pathway by mimicking a GPCR ligand. Pathways GPCRs are involved in include those responsible for neurotransmission, physiologic homeostasis and metabolism as well as controlling cell growth, proliferation, differentiation, and apoptosis across many cell types.[2] Given their prominent role in the regulation of cellular processes, proteins that are associated with these receptors are of interest to researchers for other possible avenues of therapeutic advancement. Over the past decade, arrestin proteins, specifically β-arrestin proteins (arrestins 2 and 3), have been shown to have a role in cell signaling beyond their classically known function as GPCR desensitizing proteins.[2-4] Under the classical understanding of β-arrestin function, once a GPCR is bound it can no longer activate G-proteins and ceases the cell signaling cascade. The new information that has been reported shows that while β-arrestins effectively desensitize the GPCRs, they also participate in a signaling pathway on their own right.[2,5,6]

G-protein signaling is initiated by the binding of a ligand, such as catecholamine, by a GPCR after a systemic release of the ligand. A heart attack would be one example of such an event. Upon binding of the ligand, the receptor acts to induce a conformation change in the heterotrimeric G-protein complex thereby promoting the exchange of GDP for GTP (Figure 5-1).[7,8] Once this occurs, The G-protein complex dissociates into G_{α} and $G_{\beta\gamma}$ subunits. Each subunit is then capable of activating a different second messenger pathway as part of the signaling process. The GPCR, once the G-protein complex has dissociated, is free to induce changes and initiate signaling cascades in additional G-protein complexes as the ligand remains bound until phosphorylation by GRK (G-protein coupled receptor kinase). By phosphorylating the receptor, a change in GPCR conformation is induced promoting the binding of β-arrestins upon which the GPCR can no longer signal additional G-proteins and marks the beginning of signal pathway down-regulation. In the classical understanding of β-arrestin function, after the binding of the β-arrestin, additional proteins are recruited forming a clathrin-coated pit as part of endocytosis. After the endosome is formed, the receptor can either be degraded or recycled.

The role of the β-arrestin, however, is not confined to receptor desensitization as previously thought, but has additional roles in initiating additional second messenger pathways. One such pathway involves the transactivation of the epidermal growth factor receptor (EGFR) (Figure 5-2). This transactivation process is initiated after the β-arrestin promotes an Src kinase to phosphorylate matrix-metalloproteinase (MMP). MMP, in turn, causes the splitting of the Heparin-Binding epidermal growth factor (HB-EGF) outside the cell, which directly causes the EGFR transactivation by acting as a ligand.[8] Additional signaling occurs within the cell, but ultimately this pathway is determined to have beneficial

effects not seen by the classically known G-protein pathway. In fact, through a mouse model, it has been demonstrated that chronic stimulation of the G-protein mediated pathway has many adverse cardiac effects.[4,9,10] These include a decrease in the strength of contraction resulting in the enlargement of the heart as well as thickening of the muscle tissue in the heart. Conversely, signaling that occurs as the result of the β-arrestin pathways have been found to have many cardioprotective effects including myocyte proliferation and anti-apoptotic effects.[8]

Unlike the signaling onset for G-protein mediated pathways, the onset of the beta-arrestin dependent pathway is slower and increases gradually. This signaling pathway also is sustained for much longer and does not experience the rapid decrease as is associated with the G-protein mediated pathway once bound by a beta-arrestin.[4] The prolonged signaling effect combined with the cardioprotective effects makes for an appealing target for further research. The β-arrestin pathway is of interest and could lead to a new class of drugs to treat heart disease, the largest killer of both men and women in the United States. This result is a long way off, however. Most work to this point has focused on the associations between β-arrestins and other proteins. Yates and Lefkowitz studied this by applying the MudPIT technique to a mouse cell model.[11] From this work they were able to find 173 proteins that were associated with β-arrestin 1, 266 proteins associated with β-arrestin 2, and 102 proteins that were found to be associated with both forms for a total of 337 unique proteins. These associations were also broken down in to classes, demonstrating that proteins were found to be involved in a wide variety of functions.

5.1.2 Previous studies of MEF lysates by online 2DLC

The differential analysis of the mouse embryonic fibroblast cell lysate with a β-arrestin 1,2 double-knockout had previously been studied in the Jorgenson lab by Brenna Richardson utilizing a combined top-down/bottom-up approach with a multidimensional separation of intact proteins followed by a flow split to allow for simultaneous fraction collection and intact mass determination via TOF-MS.[12] The online multidimensional separation utilized anion-exchange chromatography in the first dimension followed by reversed-phase in the second dimension. Once the intact data was collected, fold-changes were determined and fractions of interest then further analyzed by bottom-up methods to hopefully identify proteins of similar mass, allowing for identifications. This technique identified a total of 68 differential masses from the intact data with about 50% of these masses matching proteins identified through bottom-up measures.

Ultimately, it was deemed necessary to revisit this differential comparison with different methods to try and improve upon the results found by Richardson. A more comprehensive analysis relying less on mass correlations was believed to have a better chance of not only identifying proteins of interest, but also allowing for more confident identifications as sequencing data would be available as a result of the workflow previously discussed in Chapter 3.

5.2 Experimental

5.2.1 Overview of experimental method

Much of the analysis of the mouse embryonic fibroblast (MEF) cell lysate samples is the same as was done previously in Chapter 3 for the fractionated analysis of yeast. A different fractionation scheme was used for the reversed-phase analysis as an attempt to

improve the number of protein identifications through more complete fraction collection. The most significant difference between the two methods is in the sample preparation. Whereas for the yeast, samples could be prepared ahead of time and thawed prior to use, this was found to have detrimental consequences for the analysis of the MEF lysates. In order to maintain sample solubility without the addition of salts and surfactants incompatible with liquid chromatography, the cells were thawed, lysed, and analyzed within the same day. While kept on ice, the lysate samples were never frozen. This resulted in an improvement in the protein content of the samples and allowed for a greater amount of protein to be used for the analysis. A second key difference was in the sample injection itself. With the yeast samples, the inclusion of formic acid with the sample injection was not found to be necessary to maintain sample solubility beyond the inlet frits of the reversed-phase column. For the MEF samples, a dilution with an equal volume of 88% formic acid was necessary to allow the proteins to move beyond the inlet frit and concentrate at the head of the column. Backpressure was significantly reduced during the reversed-phase separation and subsequent blanks were found to be clean as a result of the inclusion of formic acid in the sample injection.

5.2.2 Preparation of MEF protein extracts

The mouse embryonic fibroblasts cells were provided by Dr. Kevin Xiao from the Lefkowitz lab at Duke University School of Medicine. β-arrestin 1,2 double knockout mouse embryos were prepared as previously described.[13] Mouse embryonic fibroblast cells (MEF) were prepared and cultured according to the 3T3 protocol of Todar and Green.[14] Cell harvesting and lysis was performed by first harvesting the cells in a phosphate buffered saline solution, centrifuging, and removing the supernatant to form a cell paste. The solution for re-

suspension was composed of 25 mM ammonium bicarbonate with the following phosphatase inhibitors: 1 mM potassium fluoride, 1 mM sodium pervanadate, 1 μM microcystin, and 10 nM calyculin A. Protease inhibitors from Roche were added as well at 1 tablet/10 mL. Cells were re-suspended in 0.5 mL of the buffer described above and subjected to three freeze/thaw cycles. An additional 0.5 mL of buffer was added and samples were dounced for 20 strokes. Sonication was performed for a total of 50 seconds by sonicating for 10 seconds followed by a cooling time of 20 seconds at 30% power (Branson 250 sonifier, Danbury, CT). Centrifugation at 15,000 x g for 30 minutes at 4°C was performed in order to remove cell debris and insoluble protein. Filtration was performed with a 0.2 μm filter to further clean up the sample. A Bradford protein quantitation assay was then performed as done for the yeast samples in Chapter 2. The concentration of protein in each sample was found to be 4.0 mg/mL for the double-knockout sample and 5.5 mg/mL for the wild-type sample with relative standard deviations of 4.6 and 7.8%, respectively.

5.2.3 LC/MS conditions

For both the anion-exchange and reversed-phase intact protein pre-fractionation strategies, 2.5 mg of protein was injected onto the column for each sample type. For the reversed-phase separation, it was necessary to dilute the sample to a final volume of 1.24 mL with formic acid to maintain solubility past the inlet frit. The anion exchange column used for the analysis was a Waters BioSuite Q SAX, 7.5 mm x 7.5 cm with 10 μm particles and 1000 Å pores and was operated at the manufacturer's maximum recommended temperature of 50°C. Mobile phase A for the anion exchange separation was 25 mm ammonium acetate at pH 9.0 with mobile phase B being 1M ammonium acetate, also at pH 9.0. Gradient conditions for the anion exchange separation can be found in Table 5-1. The reversed phase

column was an Agilent PLRP-S, 4.6 mm x 30 cm with 5 μm particles and 300 Å pores and was operated at 80°C. Mobile phase A was 90% water, 5% acetonitrile, and 5% isopropanol with 0.2% TFA. Mobile phase B was 50% acetonitrile and 50% isopropanol with 0.2% TFA. Gradient conditions for the reversed-phase separation can be found in Table 5-2. The peptide separation was by RP-UPLC-MS^E and was previously described in section 3.2. Mass spectrometric data processing was also previously described and again utilized PLGS 2.5 for peptide assignments and Scaffold 3.3 for database searching and statistical analysis.

5.3 Results & Discussion

5.3.1 Comparison of AXC40 to RP18

The anion exchange fractionation was completed using the same liquid chromatographic conditions as was done for the anion exchange separation of Baker's yeast in Chapter 3. As was seen with the yeast sample, there was a significant elution of proteins in the dead volume followed by a lull in the elution (Figure 5-3). Peaks of similar width were observed and the separation was otherwise unremarkable with differences being observed between the two sample types. The reversed-phase fractionation scheme was modified to collect fractions over more of the gradient in response to the elution pattern observed from the analysis of yeast, where a significant number of proteins were tailing into later fractions as discussed in Chapter 3. This was accomplished in two ways. More fractions were collected over a longer period of time. In all, 54, 1-minute wide fractions were collected. These were combined to form 18, 3-minute wide fractions, which then underwent further analysis by LC/MS. Fractions were pooled in this manner out of consideration for computational resources and overall analysis time. Increasing the number of fractions creates increasingly large data sets, which tax the software and systems used to analyze the data. Further

improvements to software would facilitate finer fractionation and allow for more comprehensive analyses. The time aspect also is a concern, as running 18 fractions in triplicate for two intact protein samples requires upwards of 200 hours. The gradient was also steepened to reach a higher percent of the 50/50 acetonitrile/isopropanol mobile phase more quickly to ensure all proteins had eluted. By increasing the gradient steepness, the elution window was compressed (Figure 5-4) with respect to what had previously been seen for the yeast samples.

It was assumed that the number of unique proteins would decrease slightly as a result of this additional grouping and steeper gradient. However, it was believed that overall there would be a greater number of protein identifications made as a result of the reversed-phase pre-fractionation than what was found by anion-exchange methods. This expectation was based on the results shown in Chapter 3 where with even fractionation of the gradients, reversed-phase outperformed anion-exchange by a 2:1 margin. When fractionation was doubled for the anion-exchange pre-fractionation, there was still a significant advantage in choosing reversed-phase for the pre-fractionation. In practice, these assumptions proved to be incorrect for the mouse samples. Ultimately, the anion-exchange pre-fractionation with 40 fractions was able to identify 355 unique proteins from the mouse embryonic fibroblast cells whereas only 215 identifications were made from the reversed-phase fractionation (Figure 5-5) when using the criterion that proteins must be identified in two of three replicate analyses with a confidence greater than 95%. Somewhat surprising, however, was the lack of overlap between the two methods, with only 82 proteins found in common and could be indicative of a more complex sample. The reversed-phase method resulted in the identification of 61% of the total number of unique identifications made by the anion-exchange method. Also of

interest is the fact that 133 additional proteins unique to the reversed-phase approach, representing an increase in protein identifications of 37%, were observed. The discrepancy in identified proteins and the small amount of overlap between the data sets can likely be attributed to the more complex sample composition of the mouse cell line with the potential for many more proteins to be present and likely with a greater variety of post-translational modifications. This can present issues with the chromatography of the intact proteins and proteins, which are well behaved during the anion-exchange separation, may not chromatograph well for reversed-phase method, and vice-versa. An example of a problematic class of proteins are those which are not retained, such as positively charged proteins in the anion exchange separation. Also troublesome are proteins that elute across a substantial number of fractions. More hydrophobic proteins are those likely to precipitate on the inlet frits, and this effect was observed in particular when fractionating the MEF samples on the reversed-phase column (5-6) Another possible explanation for the lack of overlap may be due to different frit materials or mesh sizes used in the frits by the manufacturers. Differences in protein adsorption at the inlet frits would not be unexpected between the two columns. Therefore, the proteins observed, or not, in one method with regard to another, may not be the result of the chromatographic mechanism used to separate the proteins, but an undesirable consequence of practical column construction considerations. Inclusion of formic acid to solubilize proteins, especially membrane proteins, has been reported and was necessary to produce reproducible injections for the mouse embryonic fibroblast cell lysates (Figure 5-6).[15] Without formic acid injections, there was a substantial rise in the back pressure from the column and less signal from the UV detector as observed in Figure 5-7. By diluting the sample into a 50% formic acid solution, no rise in the backpressure over blank runs was

observed and a more substantial UV signal resulted, indicating that more of the proteins in the sample were remaining soluble beyond the frits.

For the remainder of the results and this discussion, less stringent requirements than those used for the differential comparison were used. Where previously a protein must have been identified in two of three replicates with a confidence greater than 95%, a protein only was required to be found in one of three replicates with a confidence greater than 95% in order to be counted. This was done to facilitate a better evaluation of the chromatography where a differential comparison was not of primary concern. Previously mentioned was the difference in fractionation schemes applied to this mouse embryonic fibroblast cell line than what had previously been done with yeast samples. Not surprisingly, this had a profound effect on the number of fractions a protein eluted across and ultimately the number of protein identifications. The overwhelming majority of proteins, 78%, fractionated by reversed-phase chromatography were identified in only one fraction. While this initially appears to be a beneficial consequence of the coarse fractionation scheme, the broad fractionation approach may have resulted in a peptide sample too complex to be properly analyzed by the second dimension workflow. Figure 5-8 and Figure 5-9 show the difference between the reversed-phase and anion-exchange fractionation schemes with respect to the number of fractions per protein identification. The increased and more even fractionation of the anion-exchange method does split more peaks with only 31% being found in one fraction. The percent of proteins being found in two fractions was determined to be 11% for reversed-phase pre-fractionation and 22% for anion-exchange pre-fractionation and those found in three or more fractions was 11% for reversed-phase pre-fractionation and 47% for anion exchange pre-fractionation.

To further evaluate the fractionation schemes, proteins counts were plotted as a function of the fraction in which they eluted. Thus, in areas of the chromatogram where there were few proteins eluting off of the first dimension separation, few proteins were identified. For both fractionation schemes, reversed-phase (Figure 5-10) and anion-exchange (Figure 5-11), there are regions of especially dense protein elution as well as areas in the separation where there were few additional proteins identified. The trends generally follow what was observed for the analysis of yeast in Chapter 3. For the reversed-phase separation, essentially no proteins were identified during the first 6 minutes of fraction collection and relatively few new proteins were identified during the last half of the fraction collection. Given the somewhat relaxed requirements for identification purposes, 623 unique proteins were identified, but not quantified when viewing the data in this way. At times, upwards of 120 proteins were identified from a 3-minute wide fraction with the bulk of the protein elution occurring between fractions three and nine. For the anion exchange separation, after the initial elution of proteins in the dead time, the expected lull in elution followed, which then transitioned into a period of significant elution. This too tapered off, although it was more consistent than the reversed-phase elution of proteins with only the last fourth contributing little in the way of new identifications. Using the less stringent requirements for identification, 930 unique proteins were identified. When comparing the thresholds for identifications versus identifications used for fold-change identification we do see a marginal difference between the two sample types. For the reversed-phase pre-fractionation, only 71% of the total number of proteins identified once at a confidence level of 95% were also identified a second time allowing for use in the differential comparison. This is in comparison to the anion-exchange method where 81% of the proteins found once were also

found twice or more. This 10% difference could be due to random variation, but it may be more indicative of the inefficiency of the reversed-phase pre-fractionation combining too many proteins into one fraction. It is possible that with improved fractionation of the reversed-phase separation, fewer proteins and therefore fewer peptides would be contained within a given fraction. The result is that the mass spectrometer would have a better opportunity to fragment these peptides without raising the background noise significantly and could present the data analysis programs with cleaner data in order to better sequence the peptides. With better data, more proteins could be identified and the gap between the two methods could likely be narrowed. Ultimately, the data shown comparing the number of fractions per protein and conversely the number of fractions per protein indicate that there is a balance to be maintained with fractionation. The most efficient route is likely that which abandons fractionation based on even time intervals in favor of a more targeted approach using the UV data as a guide.

5.3.2 Comparison of Pre-fractionation to Online 2DLC results

This cell line was previously studied as part of the research undertaken by Brenna Richardson in the last chapter of her dissertation. The approach utilized by Richardson involved a mixed top-down, bottom-up methodology where the intact mass was determined by mass spectrometry and later compared to bottom-up data from fractions collected during the top-down analysis. A number of proteins were found to be differentially expressed between the wild type and double-knockout samples with fold changes that would be considered to be significant (greater than 1.5).

Unfortunately, none of these proteins were found to be of interest, as they are not believed to be involved in any of the pathways of interest for β-arrestin proteins. There are a

number of reasons why this may be the case. A key limitation of this technique is the need to be able to detect intact proteins. Few proteins were ever observed beyond 40 kDa, so immediately there was a significant reduction in the number of proteins one could expect to observe. The other, more significant contributor to the difficult nature of this analysis is in regard to the abundance of signaling proteins. Typically proteins involved in signaling pathways are not expressed in large quantities making for a more difficult analytical challenge to observe these proteins. Additionally, proteins involved in signaling are often found to be associated with the cell membrane. Only cytosolic proteins were analyzed by this method owing to solubility requirements for the two-dimensional separation of intact proteins used in this analysis. Other challenges limiting the number of protein identifications are due to the limited number of intact masses that were able to be determined. This is due to a competition for ionization. Also, as a result of the difficulty associated with the ionization of large proteins, few proteins were identified above 40 kDa. Lastly, an additional challenge, which was never overcome despite work to create data filtering programs and comparison tools, was the problem of associating a deconvoluted mass from the intact protein separation, with the mass reported from the bottom-up analysis. Without knowing which post-translational modifications were present, if any, results from this type of comparison would always be suspect, with few ways to confirm identifications. So, when one takes into account all of these factors, it is of little surprise that the differentially expressed proteins identified were not found to be of interest to the collaborators at Duke University.

It was hoped that a more comprehensive analysis, where peptides were being analyzed by mass spectrometry as opposed to intact proteins, would sufficiently lower the limits of detection to enable detection and relative quantitation of proteins associated with β-

arresting signaling pathways. Overall, there was a significant increase in the number of differential proteins identified by both reversed-phase and anion-exchange approaches. 32 differential proteins were identified by the online method utilized by Richardson in her analysis of these samples. With the offline pre-fractionation approach, 141 were found by the reversed-phase method and 295 by the anion-exchange method, again with a fold-change greater than 1.5 as the minimum threshold. When comparing the proteins identified by Richardson to the list of differential proteins generated by the anion exchange pre-fractionation approach, approximately half of the 32 masses identified in the earlier work were identified. Unfortunately, these overlapping proteins were not always found to be up-regulated in the same sample, and actually about half of this subset was found to indicate the opposite expression level. Fold changes for both analysis techniques were typically less than 4. When looking specifically at the spectral counts, as determined through Scaffold, for the proteins found in common between the two techniques, the numbers were low. Therefore, it is not necessarily unexpected that the results are contradictory given the relatively small fold changes at low intensity levels. For the reversed-phase pre-fractionation there was even less overlap between the two techniques given the small number of proteins identified overall by this fractionation approach.

5.3.3 Comparison of Pre-fractionation to Literature

Likely the best research to date on the beta-arrestin interactome comes from the collaboration between Yates and Lefkowitz employing the MudPIT approach as well as gel-based approaches for a comprehensive analysis. In this work, interacting proteins from HEK293 cells overexpressing both beta-arrestin 1 and 2 were immunoprecipitated and analyzed by mass spectrometry.[11] Many replicates (6-8) were conducted in an effort to

increase the reliability of the identifications. From this work a total of 337 unique protein identifications were made spanning a wide-range of functions and locations within the cell. The majority (53%) were found in the cytoplasm, but there were also significant contributions from proteins found in the nucleus, nucleolus, and membrane. Functions included signal transduction, nucleic acid binding, cellular organization, metabolic functions, and a substantial portion of unknown functions at almost 42% of the total.[11] This type of analysis was much different from that which was conducted in the research presented in this chapter, however. For the research completed by Yates and Lefkowitz, the purpose of the study was to induce as many associations between beta-arrestins and other interacting proteins. The overexpression of beta-arrestins as well as the immunoprecipitation would be expected to help with the limit of detection immensely as their approach was exceptionally targeted with respect to the analysis of cell lysates with few modifications.

Analysis of the cell model used in the studies presented here with the more subtle changes than the samples used in the research of Yates and Lefkowitz unfortunately shared a similar outcome with the methods attempted by Richardson in her studies. Despite a substantial increase in the number of differential proteins, an increase of approximately 10-fold, only a few proteins were identified that were also identified as interacting with beta-arrestins by the Yates study. Interestingly enough, there was one additional match from the reversed-phase data than from the anion-exchange data, which was quite a surprise given the fewer number of overall identifications. In total, there were four classes of proteins from the anion exchange method and five classes of proteins from the reversed phase method that matched the reported results. These results are summarized in Tables 5-3 & 5-4 and described in the following text.

The 14-3-3 (1433_mouse) scaffolding proteins were identified as being down-regulated in the double-knockout sample and were unique to the anion exchange data set. Six isoforms were identified in total with three present only in the wild-type sample and three with an average 5-fold change in expression level. Little is known regarding the interactions between these two classes of proteins. Both beta-arrestins and the 14-3-3 proteins are known to function as scaffolding proteins, but are assumed to serve as scaffolds for different types of proteins as the interactomes of beta-arrestin and the 14-3-3 proteins was only determined to be 8%. Also found from the anion exchange pre-fractionation was annexin, a protein known to participate in signaling as well as two myosin proteins (MYH9 and Q3UH59). The annexin proteins were found in the reversed-phase data in addition to both myosin proteins. Also found was another signaling protein, catenin-beta (CTNB1). Several proteins involved with cellular organization were identified. These included several forms of tubulin (TBB3, TBB6, TBB2C, and TBB5), cofilin 1 (COF1), and caldesmon (CALD1). For the reversed-phase method, the annexin and catenin proteins were only identified in the double-knockout sample. The annexin result for the reversed-phase method was the same as observed for the anion-exchange method. The remainder of the proteins identified in the reversed-phase method were identified as being either only found in the wild-type sample, suggesting down-regulation in the double-knockout sample, or were clearly determined to be down-regulated in the double-knockout sample with an associated value for the fold change in regulation. COF1, TBB3, and TBB6 were found to be only in the wild-type sample with TBB2C, CALD1, and TBB5 being down-regulated in the double-knockout sample with fold changes of 2.3, 2.1, and 1.6, respectively.

5.3.4 Comparison of MEF to Yeast

The comparison between the cell lysates from mouse embryonic fibroblast cells and yeast reveals striking differences. With the yeast analysis in Chapter 3, it was clear that the reversed-phase fractionation with only 20 fractions was a significant improvement over the anion exchange methods, even when twice the number of fractions were analyzed. For this mammalian sample, the increased fractionation appears to be of greater importance as there was a significant increase in protein identifications from the anion exchange method with greater fractionation than the reversed-phase method. A possible explanation for this is that there are simply more proteins to be identified in the mammalian samples creating a more complex sample. As increased fractionation would help alleviate sample complexity, the increased fractionation of the anion-exchange method utilized in this comparison would be expected to be of great utility and allow for the improvements in protein identifications over methods with fewer fractions. Clearly, each sample must be treated individually and methods must be tailored for each analysis. While it would be convenient to apply methods without any modifications to any potential sample, these results indicate the sample complexity of mammalian cells in comparison to yeast require more experimental planning on the part of the investigator. However, simply increasing the fractionation is not the answer as with 18 fractions, the analysis time required is already a significant investment that is not likely to be practical for routine or even semi-routine analysis.

To compare specifics of the two samples, when analyzing Baker's yeast, the AXC40 method provided 370 identifications in comparison to 546 from the RP20 method, or 68%. Of the 370 identifications, though, 70 were unique to the AXC40 method, resulting in an additional 13% increase in protein identifications when the results of two methods are

combined. When analyzing the MEF samples, the situation was much different. As already mentioned, the reversed-phase method found fewer proteins than the anion exchange method, with reversed-phase identifying only 61% as many proteins. The lack of overlap was striking and the fact the additional contributions from the method identifying fewer overall proteins rose from 13% to 37% is surprising.

5.4 Conclusions

The ability to determine proteins that are up or down-regulated as the result of the beta-arrestin double-knockout ultimately may prove to be too subtle a challenge for the type of analysis being conducted. As with the approach by Yates and Lefkowitz, a more targeted methodology will likely be needed to learn anything beyond what has previously been reported. The fact that very few proteins involved in cell signaling that were previously reported to interact with beta-arrestin proteins were identified by the pre-fractionation approaches is somewhat discouraging. Coming back to the idea that signaling proteins are often among the least expressed and often associated with the membranes, future work should focus on inclusion of membrane proteins in the analysis as well as the use of better mass spectrometry through newer instrumentation with lower limits of detection. That said, there are a couple alternatives that can be tried in the interim that may allow for more identifications in less time via targeted fractionation, the use of the XUPLC system described previously, and the incorporation of both of these strategies along with immunoprecipitation studies.

5.5 Future Studies

5.5.1 Improved fractionation

Comparisons of the UV traces to the number of proteins identified per fraction for both the yeast and mouse samples revealed that the UV data was a good predictor for where unique proteins would be identified. This is somewhat contradictory to what has recently been reported in the literature,[16] but given our results across both sample types for multiple fractionation schemes involving two different intact protein separation strategies, the UV data appears to be sufficient for guiding a targeted fractionation scheme. With targeted fractionation, one could lump together fractions at the beginning and end of the intact protein separation method where little is occurring and few unique protein identifications are expected. This allows for a reduction in analysis time that could be used in one of two ways. If one desired to maintain a substantial number of fractions, say 20, then by combining fractions at the beginning and end would allow for finer fractionation across the portion of the UV trace where there were many peaks eluting. This sort of strategy is particularly appealing for the MEF samples, as the reversed-phase method clearly did not fractionate this sample enough. The same approach could be applied to the anion-exchange pre-fractionation and one could likely cut the number of fractions significantly without deleterious effects as fine fractionation could be maintained where needed and fractions where little new information is being gleaned could be combined. The other option would be to simply reduce the number of fractions even below the 18 analyzed for the reversed-phase method without further reducing the number of protein identifications. The needs of the investigator ultimately would decide how finely to fractionate the intact proteins with the amount of analysis time available for a given sample being the ultimate factor in deciding how many

fractions to collect and analyze. Regardless of the outcome, since the UV data is collected during fraction collection one would be well suited to finely fractionate the intact protein separation and later combine fractions in a manner consistent with level of activity observed across the chromatogram.

5.5.2 XUPLC of peptides

Improved fractionation is expected to improve productivity, which is defined as the number of protein identifications per hour. Chapter 2 looked at analyzing fractions by a commercially available multidimensional approach with results that were less than satisfactory given the amount of time required for such an analysis. In Chapter 4, the use of XUPLC showed significant promise. While fractionation approaches already require a significant investment in time, the use of meter-long columns and hours-long analysis times begins to lose some appeal. However, the use of shorter columns packed with smaller diameter packing materials resulting in run times comparable to what is currently conducted on a commercially available nanoAcquity UPLC system is intriguing. As previously reported, twice the number of proteins were identified given a 33% increase in analysis time, which is certainly a large improvement in productivity. Combining the targeted fractionation scheme with XUPLC would likely provide for significant improvements in protein identifications, and therefore an improved differential analysis.

5.5.3 Immunoprecipitation of beta-arrestin interacting proteins

The study of the cell lysates has yielded little in the way of useful information regarding beta-arrestin interactions despite multiple analysis techniques being applied. A more targeted approach, which was undertaken by Yates and Lefkowitz produced far more useful information. Given the relative success of this method, it would seem prudent to

incorporate some of the techniques used in the preparation of the sample to be analyzed by LC/MS before attempting to further improve the LC/MS methodology. Preparing a better sample is likely to yield more fruitful results than spending greater and greater amounts of time fractionating intact proteins and analyzing the peptides with the myriad of techniques discussed here and in previous chapters. However, with a better sample, the peak capacity provided by intact protein pre-fractionation in combination with large peak capacity separations in reasonable analysis times as the result of using XUPLC with smaller particle diameters could certainly be a worthwhile use of time.

5.6 Tables

Time (min)	25mM Ammonium Acetate, pH 9.0 (%A)	1M Ammonium Acetate, pH 9.0 (%B)
0	100	0
60	0	100
90	0	100
90.1	100	0

Table 5-1. Gradient program used for the intact protein pre-fractionation of the yeast cell lysates by anion-exchange chromatography. The column used was a Waters BioSuite Q SAX, 7.5 mm ID x 7.5 cm with 10 µm particles and 1000 Å pores. The flow rate was 0.5 mL/min and the column was held at a temperature of 50 °C, the maximum per manufacturer recommendations. Forty fractions were collected from 2 to 82 minutes.

Time (min)	90:5:5 H_2O:ACN:IPA + 0.2% TFA (%A)	50:50 ACN:IPA + 0.2% TFA (%B)
0	100	0
2	100	0
5	75	25
45	30	70
45.5	0	100
50	0	100
50.5	100	0
60	100	0

Table 5-2. Gradient program used for the intact protein pre-fractionation of the yeast cell lysates by reversed-phase chromatography. An Agilent PLRP-S, 4.6 mm x 25 cm column with 5 µm pores particles and 300 Å pores. The flow rate was 1.0 mL/min. A column temperature of 80 °C was used – well below the rated maximum of 200 °C. Fractions were collected every minute from 2 to 42 minutes for a total of 40 fractions collected. These fractions were later pooled to form the 20 fractions analyzed.

Anion-Exchange Pre-Fractionation

Differential Protein	Regulation Relative to WT	Fold Change
1433B	Down Regulated	WT Only
1433E	Down Regulated	5.6
1433F	Down Regulated	WT Only
1433G	Down Regulated	5.1
1433T	Down Regulated	WT Only
1433Z	Down Regulated	4.4
ANXA1	Up Regulated	dKO Only
ANXA6	Up Regulated	dKO Only
ANXA8	Up Regulated	dKO Only
MYH9	Down Regulated	3.0
Q3UH59	Down Regulated	WT Only
CTNB1	No Difference	-
TBB3	No Difference	-
TBB6	No Difference	-
TBB2C	No Difference	-
TBB5	No Difference	-
COF1	No Difference	-
CALD1	No Difference	-

Table 5-3. Proteins found by the anion-exchange pre-fractionation method that are known to associate with beta-arrestin proteins are shown here. Only 5 of the 18 proteins in this list were differentially expressed and found in both samples allowing a fold change to be calculated. Nearly half, 7 of 18, were found to be expressed within a fold change of 1.5 in each sample.

Reversed-Phase Pre-Fractionation

Differential Protein	Regulation Relative to WT	Fold Change
ANXA1	Up Regulated	dKO Only
ANXA6	Up Regulated	dKO Only
ANXA8	Up Regulated	dKO Only
MYH9	Down Regulated	3.1
Q3UH59	Down Regulated	WT Only
CTNB1	No Difference	-
TBB3	Down Regulated	WT Only
TBB6	Down Regulated	WT Only
TBB2C	Down Regulated	2.3
TBB5	Down Regulated	1.6
COF1	Down Regulated	WT Only
CALD1	Down Regulated	2.1

Table 5-4. For the reversed-phase pre-fractionation, there were fewer proteins identified that are known to interact with the beta-arrestin proteins. There were a few differences between this data and that from the anion-exchange pre-fractionation data as fold changes were calculated for some of the tubulin proteins (TBB“X”), which was not previously observed.

5.7 Figures

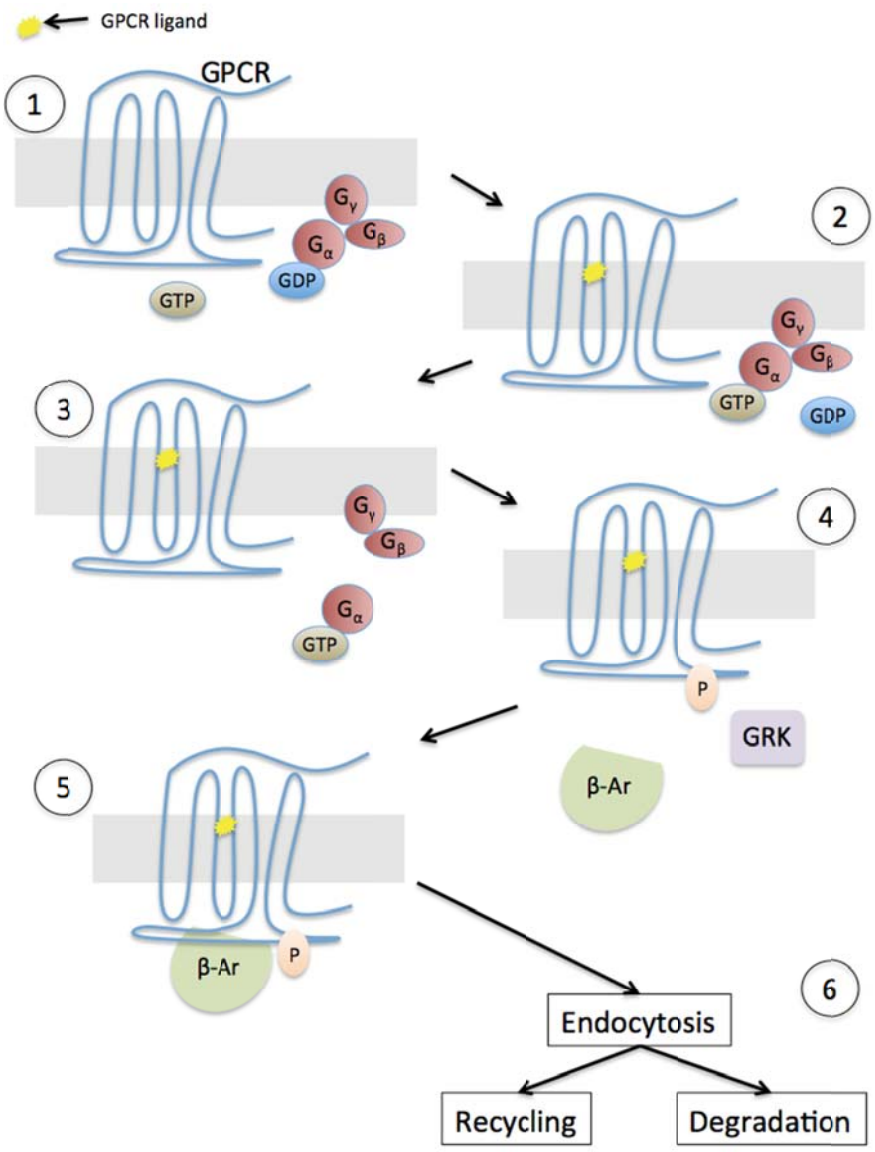


Figure 5-1. When a GPCR ligand (such as a catecholamine) is bound, the GPCR induces a change on the G-protein complex causing it to preferentially bind GTP over GDP. Once GTP is bound by the G_α sub-unit, it dissociates from the complex and both the Gα and $G_{\beta\gamma}$ sub-units are capable of promoting independent second messenger pathways. In order to regulate signaling, GRK phosphorylates the GPCR, inducing another structural change allowing for the binding of a β-arrestin. After β-arrestin binding, the GPCR can no longer initiate signaling in G-proteins and the receptor is either recycled or degraded after endocytosis.

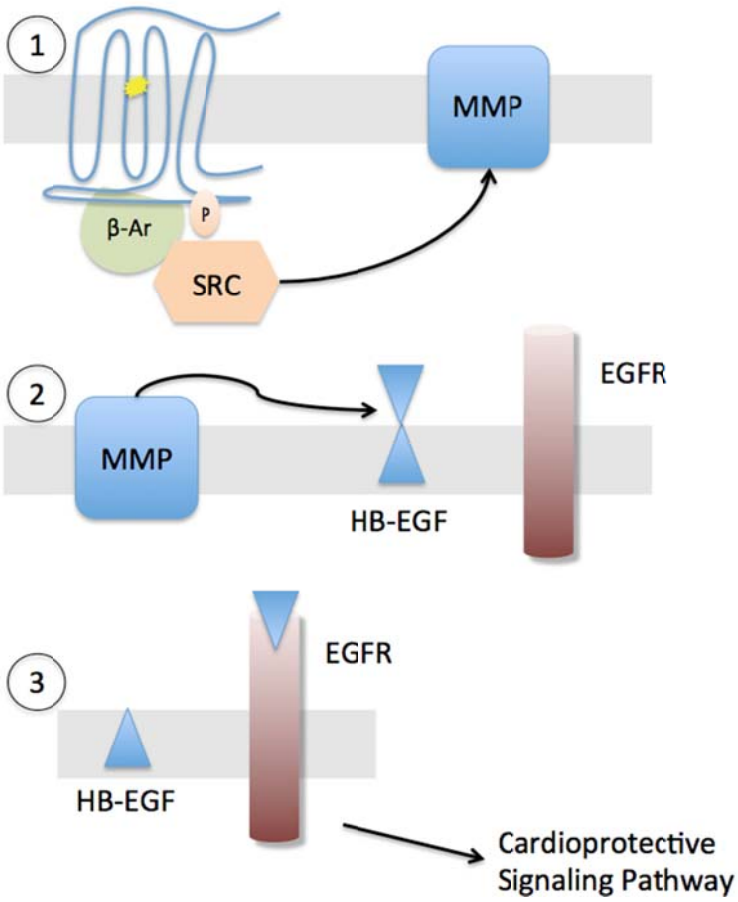


Figure 5-2. The classically understood role of β-arrestins is in regulation of GPCR signaling. More recently, pathways demonstrating the role of β-arrestins as signaling initiators have been found. One such pathway believed to have cardioprotective effects is shown here. The β-arrestin recruits an SRC kinase, which phosphorylates matrix-metalloproteinase (MMP). The pathway then transits the cell membrane as MMP then cleaves Heparin-Binding epidermal growth factor (HP-EGF). The cleaved portion of HP-EGF then activates the epidermal growth factor receptor (EGFR). The EGFR continues to signaling cascade in this cardioprotective pathway. More research is needed regarding this and other potential β-arrestin signaling pathways as they are currently not well understood.

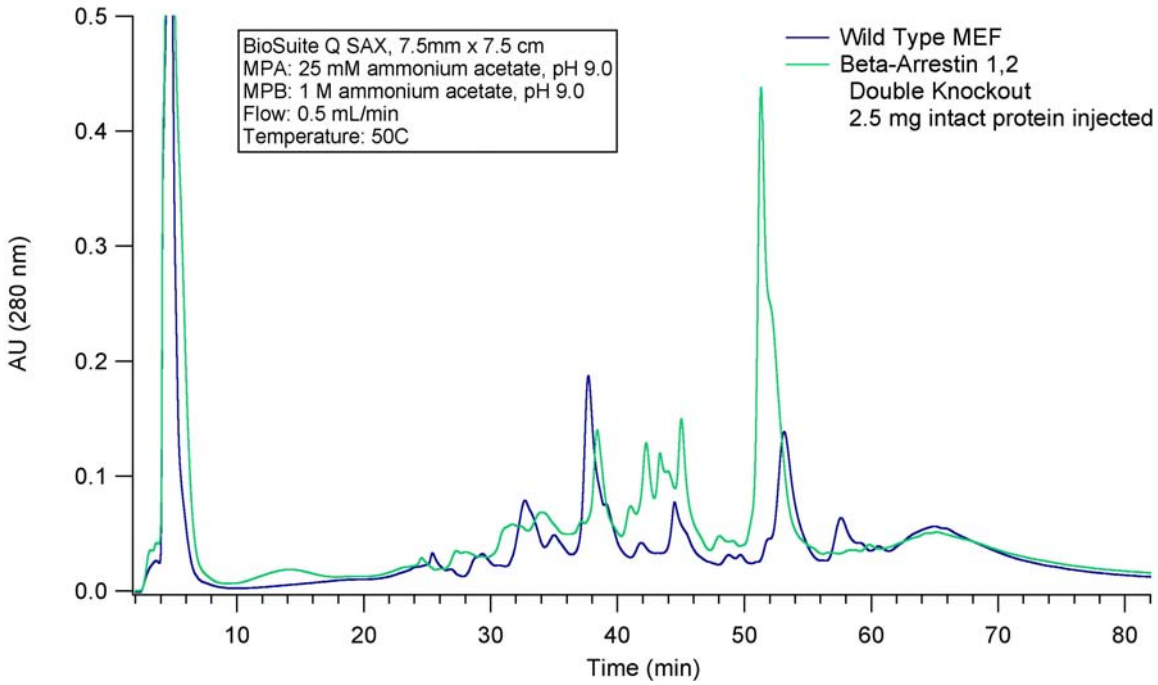


Figure 5-3. The intact protein pre-fractionation of a MEF cell lysate by anion exchange chromatography with UV detection. A BioSuite Q SAX column was used for the separation. For both the wild type and double-knockout samples, 2.5 mg of protein were injected on column. A gradient of 25-1000 mM ammonium acetate at pH 9.0 over 60 minutes was used for elution of the proteins. The column was operated at the maximum manufacturer recommended temperature. Forty fractions were collected from four to 84 minutes. Each of the 40 fractions was subject to a trypsin digestion and ran by UPLC-MS^E.

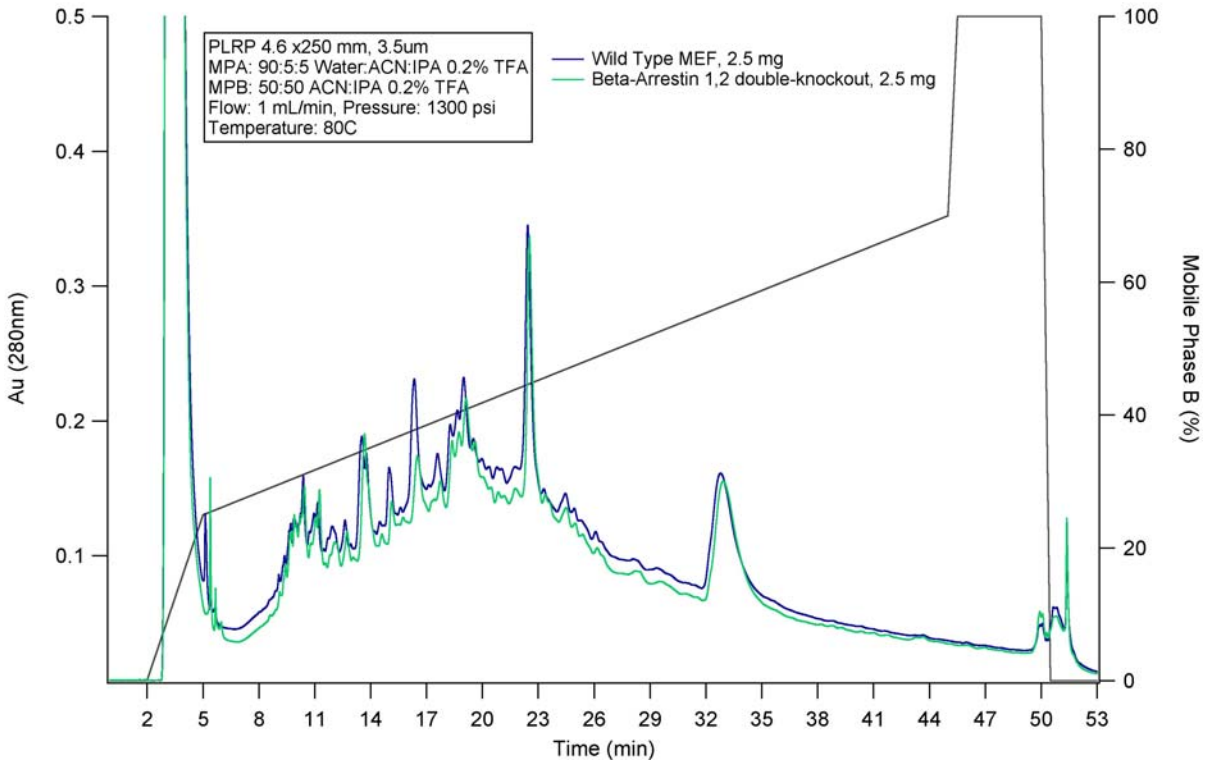


Figure 5-4. The intact protein pre-fractionation by reversed-phase chromatography with UV detection. An Agilent PLRP-S 4.6mm x 30 cm was used for the separation. Mobile phase B was comprised of 50% acetonitrile and 50% isopropanol with 0.2% TFA. Mobile phase A was 90% water and 10% mobile phase B with 0.2% TFA. The maximum temperature of a PLRP-S column is 200°C, however 80°C was determined sufficient for the elution of proteins. The appearance of the chromatograms for the two samples is more similar than in the anion-exchange pre-fractionation. A total of 54, minute-wide fractions were taken and combined to form 18, three-minute wide fractions, which were subject to a trypsin digestion and analyzed by UPLC-MS^{E}.

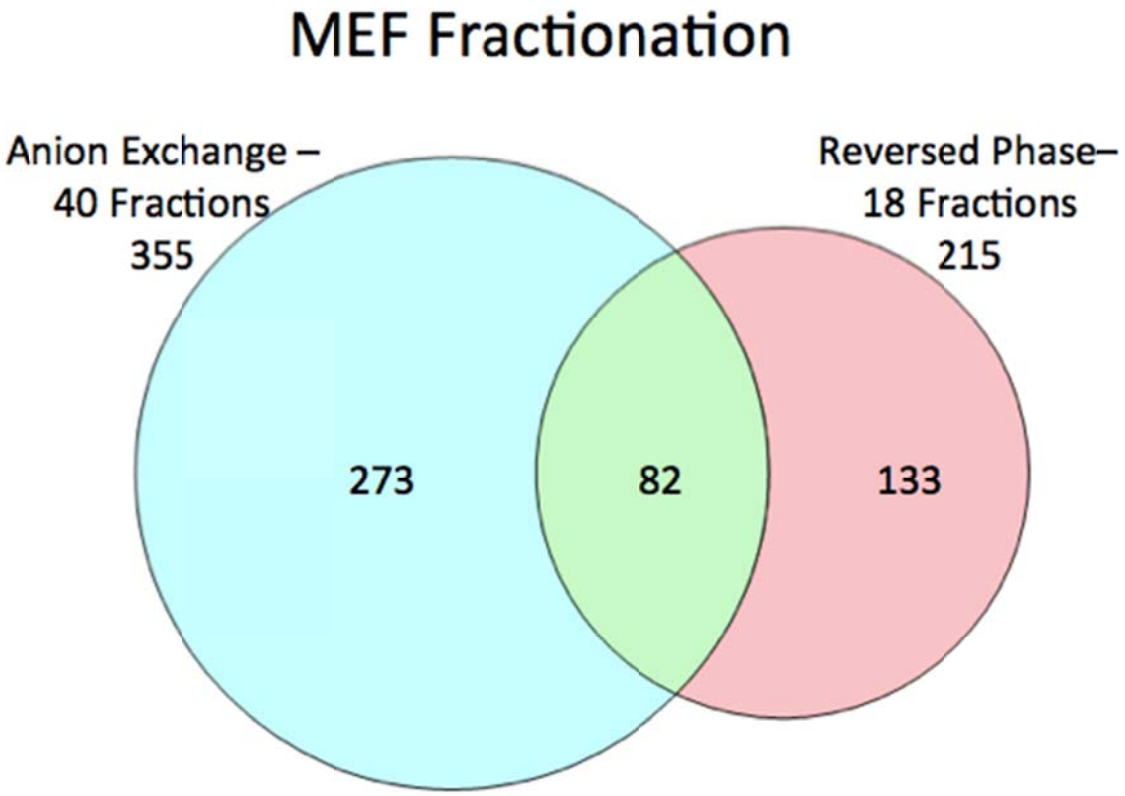


Figure 5-5. For the differential comparison, a protein must have been found in at least two of three replicate runs with a confidence greater than 95%. The proteins identified and counted from both the wild-type and double-knockout samples are represented in this diagram indicating a surprisingly low overlap between the two pre-fractionation methods than what was previously found when analyzing the Baker's yeast samples.

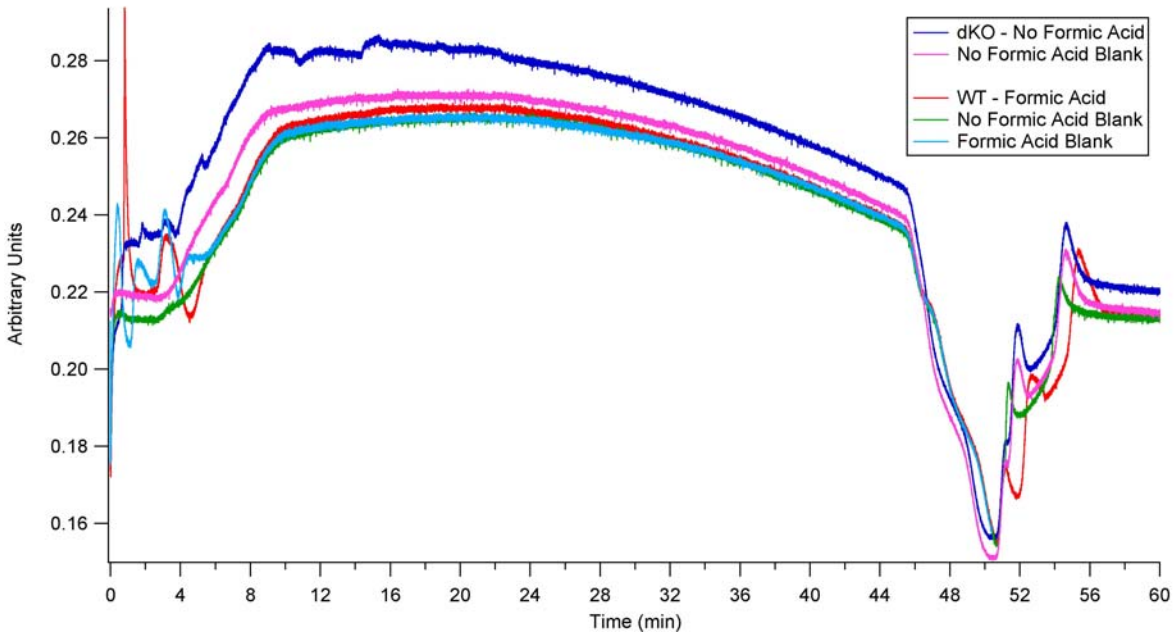


Figure 5-6. A pressure trace using arbitrary units was obtained during the intact protein pre-fractionation utilizing the reversed-phase method. The use of formic acid was necessary to reduce the backpressure not only during the run, but also for subsequent runs. Without formic acid, the baseline pressure would continue to rise with each subsequent sample injection. With formic acid, after an initial spike in the pressure, the sample runs and blank runs were nearly identical.

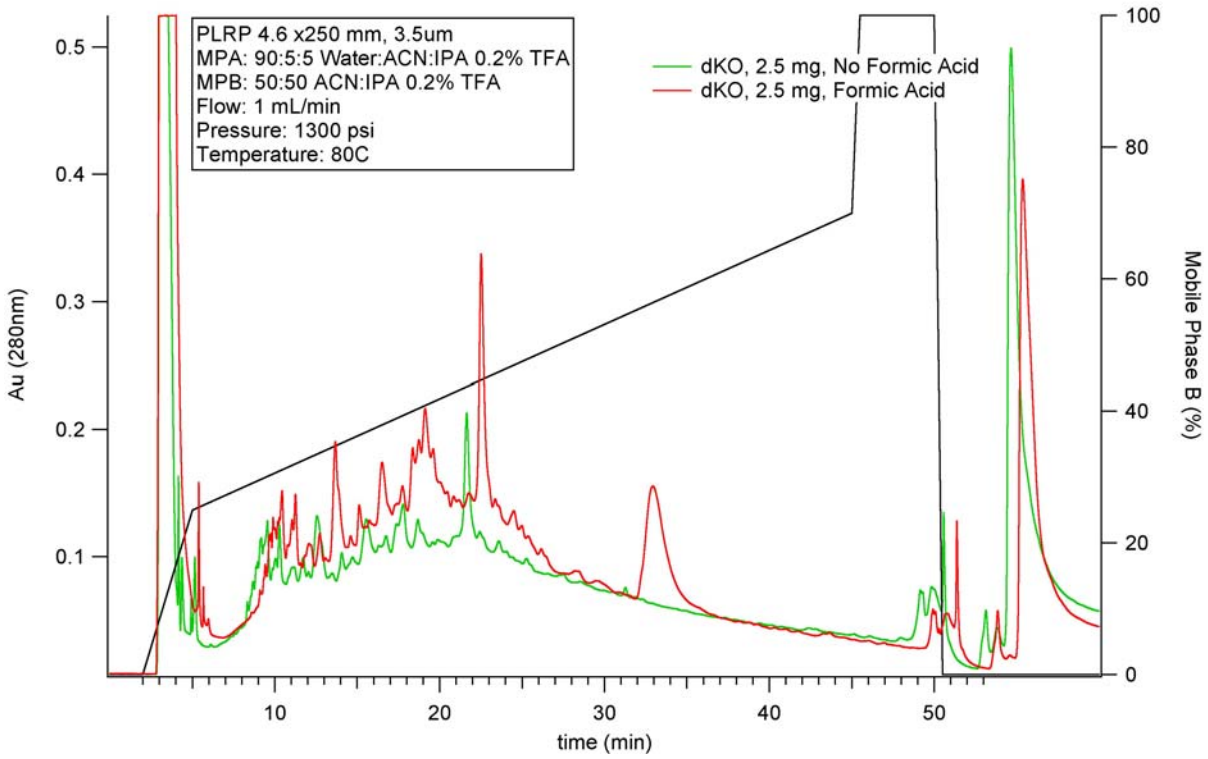


Figure 5-7. The use of formic acid in the injection also had an effect on the data observed in the UV trace. The peak at 33 minutes is noticeably absent from the injection without formic acid. Additionally, greater signal intensity was observed when formic acid was used.

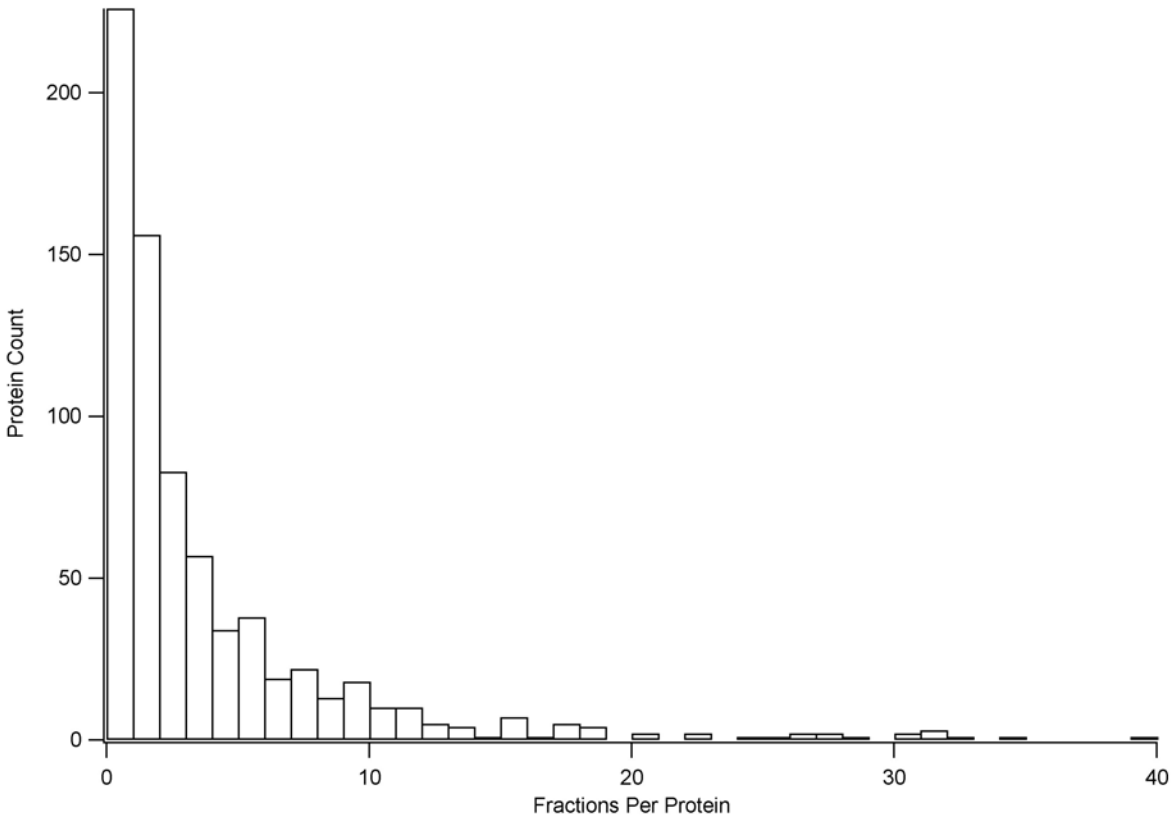


Figure 5-8. A histogram showing the number of fractions a given protein was identified in is shown for the anion-exchange pre-fractionation method. By fractionating the gradient 40 times, there were a substantial number of proteins split across two or more fractions. Data for both the wild-type and double-knockout samples are included in this figure.

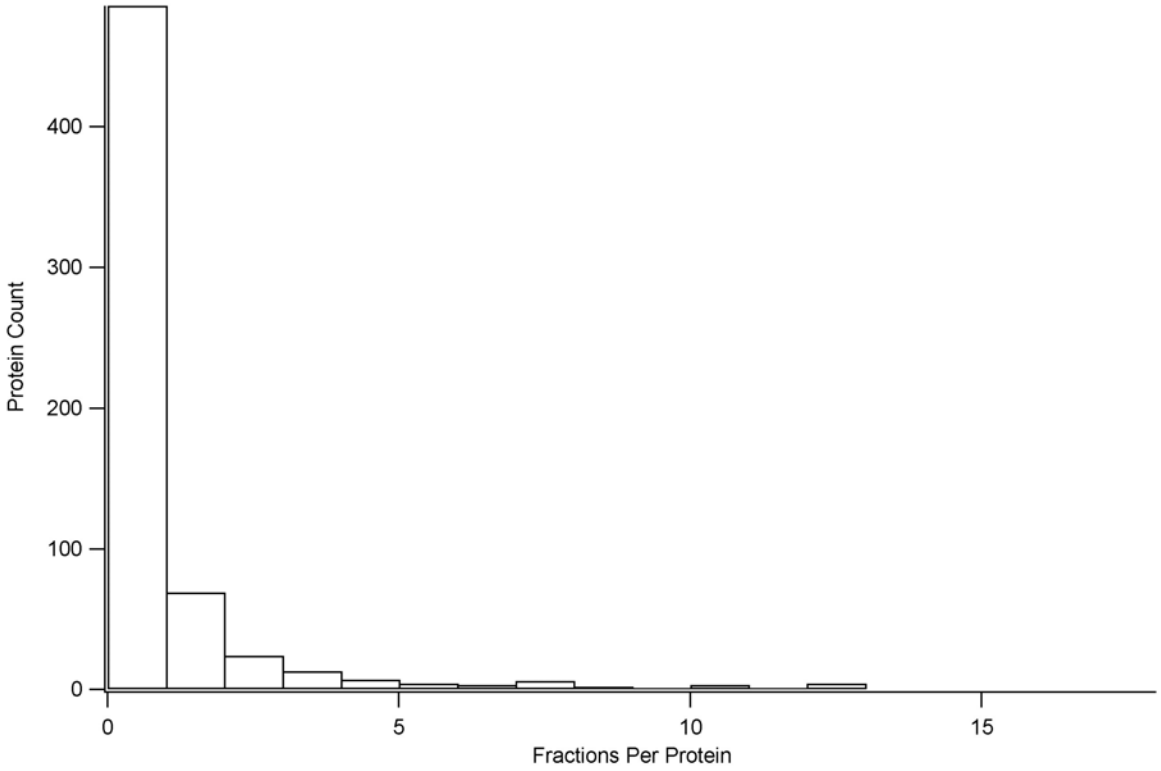


Figure 5-9. A histogram showing the number of fractions a given protein is identified in is shown for the reversed-phase pre-fractionation method. By only fractionating the gradient 18 times, an overwhelming majority of the proteins identified by this method were found in only one fraction. Data for both the wild-type and double-knockout samples are included in this figure.

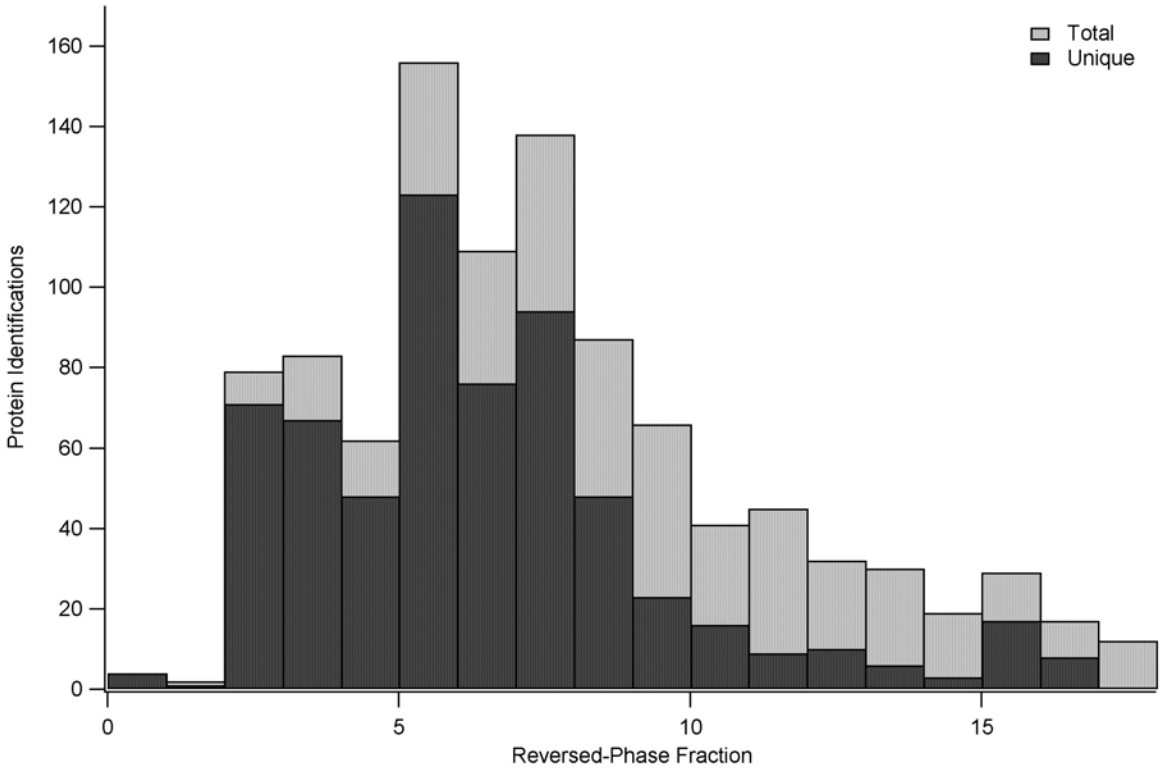


Figure 5-10. The number of total proteins identified in a single fraction as well as the number of unique proteins per fraction is shown for the reversed-phase method. For the unique protein data, a protein was only counted in the fraction in which it eluted with the greatest number of spectral counts as determined by Scaffold 3.3. As with the anion-exchange pre-fractionation, little information in the form of new identifications was made in the later fractions as compared to earlier fractions. For the reversed-phase method, proteins were generally better retained and there were not a substantial number of proteins eluting in the dead volume. Data for both the wild-type and double-knockout samples are included in this figure.

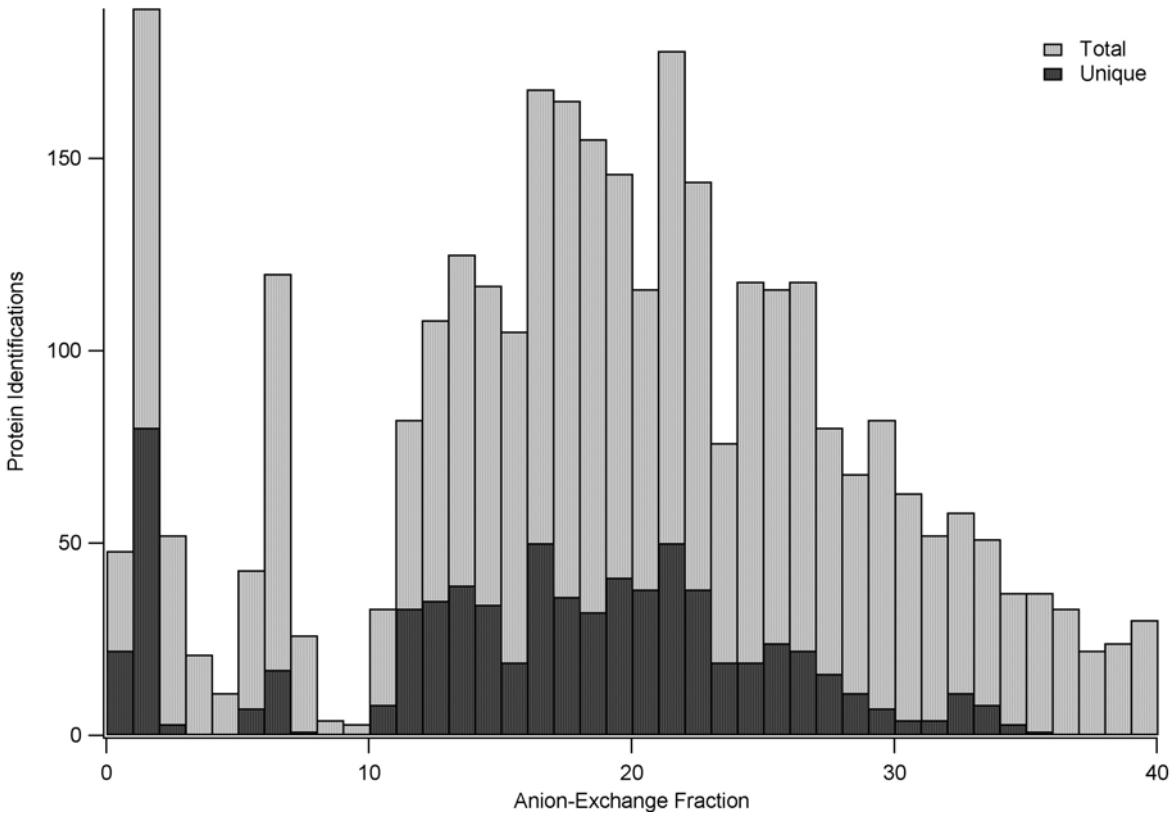


Figure 5-11. The number of total proteins identified in a single fraction as well as the number of unique proteins per fraction is shown for the anion-exchange method. For the unique protein data, a protein was only counted in the fraction in which it eluted with the greatest number of spectral counts as determined by Scaffold 3.3. While many proteins were being identified in the last several fractions, little new information was gained as little new identification were made in these fractions. There are also a significant number of protein identifications at the dead volume indicating there are many proteins that are unretained by anion-exchange chromatography. Data for both the wild-type and double-knockout samples are included in this figure.

5.8 References

1. Luttrell, L. M.; Gesty-Palmer, D., Beyond Desensitization: Physiological Relevance of Arrestin-Dependent Signaling. *Pharmacological Reviews* **2010**, 62, 305-330.

2. Lymperopoulos, A., GRK2 and β-arrestins in cardiovascular disease: Something old, something new. *Am J Cardiovasc Dis.* ***2011****, 1, 126-137.*

3. Lefkowitz, R.J.; Shenoy, S.K. Transduction of Receptor Signals by Beta-Arrestins. *Science* **2005**, *308*, 512-517.

4. Lefkowitz, R.J.; Rajagopal, K.; Whalen, E.J. New Roles for β-Arrestins in Cell Signaling: Not Just for Seven-Transmembrane Receptors. *Molecular Cell* **2006**, *24*, 643-652.

5. Noma, T.; Lemaire, A.; Prasad, S. V. N.; Barki-Harrington, L.; Tilley, D.G.; Chen, J.; Le Corvoisier, P.; Violin, J. D.; Wei, H.; Lefkowitz, R.J.; Rockman, H. A. Beta-arrestin-mediated beta1-adrenergic receptor transactivation of the EGFR confers cardioprotection. *Journal of Clinical Investigation* **2007**, *117*, 2445-2458.

6. Zhai, P.; Myamamoto, M.; Galeotti, J.; Liu, J.; Masurekar, M.; Thaisz, J.; Irie, K.; Holle, E.; Yu, X.; Kupershmidt, S.; Roden, D.; Wagner, T.; Yatani, A.; Vatner, D. E.; Vatner, S. F.; Sadoshima, J. Cardiac-specific overexpression of AT1 receptor mutant lacking Gαq/Gαi coupling causes hypertrophy and bradycardia in transgenic mice. *Journal of Clinical Investigation* **2005**, *115*, 3045-3056.

7. Moore, C. A. C.; Milano, S. K.; Benovic, J. L., Regulation of Receptor Trafficking by GRKs and Arrestins. *Annual Review of Physiology* **2007**, 69, 451-482.

8. Patel, P. A.; Tilley, D. G.; Rockman, H. A. Biased Ligands for Better Cardiovascular Drugs: Dissecting G-Protein-Couple Receptor Pharmacology. *Journal of the American Heart Association* **2008**, *72*, 1725-1729.

9. Tilley, D. G.; Kim, I. M.; Patel, P. A.; Violin, J. D.; Rockman, H. A. β-Arrestin Mediates β1-Adrenergic Receptor-Epidermal Growth Factor Receptor Interaction and Downstream Signaling. *Journal of Biological Chemistry* **2009**, *284*, 20375-20386.

10. Mangmool, S.; Shukla, A. K.; Rockman, H. A. β-Arrestin–dependent activation of Ca^{2+}/calmodulin kinase II after β_1–adrenergic receptor stimulation. *Journal of Cell Biology* **2010**, *189*, 573-587.

11. Xiao, K,; McClatchy, D. B.; Shukla, A. K.; Zhao, Y.; Chen, M.; Shenoy, S. K.; Yates, J. R.; Lefkowitz, R. J. Functional specialization of β-arrestin interactions revealed by proteomic analysis. *Proceedings of the National Academy of the Sciences of the United States of America* **2007**, *104*, 12011-12016.

12. Richardson, B. *Multidimensional Separation of Intact Proteins for Differential Proteomics Employing Topdown and Bottom-up Proteomic Strategies.* PhD dissertation, University of North Carolina at Chapel Hill. Chapel Hill, **2010**.

13. Kohout, T. A.; Lin, F. T.; Perry, S. J. ; Conner, D. A.; Lefkowitz, R. J., β-Arrestin 1 and 2 differentially regulate heptahelical receptor signaling and trafficking. *Proceedings of the National Academy of the Sciences of the United States of America* **2001**, *98*, 1601-1606.

14. Todaro, G. J.; Green, H., Quantitative Studies of the Growth of Mouse Embryo Cells in Culture and their Development into Established Lines. *Journal of Cell Biology* **1963**, *17*, 299-313.

15. Washburn, M. P.; Wolters, D.; Yates, J. R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature* **2001**, 19, 242-247.

16. Lasaosa, M.; Jakoby, T.; Delmotte, N.; Huber, C. G.; Heinzle, E.; Tholey, A. Rapid selection of peptide containing fractions in off-line 2-D HPLC in shotgun proteome analysis by screening with MALDI TOF MS. *Proteomics*. **2009**, *9*, 4577-4581.