Multidimensional Liquid Chromatography Coupled to Mass Spectrometry for the Analysis of Complex Mixtures of Proteins

Charles Robert Evans

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
2007

Approved by:

Professor James W. Jorgenson

Professor Gary L. Glish

Professor Matthew R. Redinbo

Professor Mark H. Schoenfisch

Professor Gary J. Pielak

ABSTRACT

Charles Robert Evans: Multidimensional Liquid Chromatography Coupled to Mass
Spectrometry for the Analysis of Complex Mixtures of Proteins
(Under the direction of James W. Jorgenson)

Proteomics – which is the analysis of the full complement of proteins produced by an

organism – plays a crucial part in a variety of fields of research, including basic biological

studies, pharmaceutical development, and clinical diagnostics. Separations are a key element

of proteomic analyses. The traditional means of separating protein mixtures is two-

dimensional gel electrophoresis, which offers very high resolution. Over the past several

decades, alternative methods for protein separations based on liquid chromatography have

been developed. These methods have advantages over gel-based analyses, including reduced

bias against certain classes of proteins, straightforward automation, and easier coupling to

mass spectrometry. However, in order to effectively manage the complexity of a proteome,

the separation technique must be able to resolve a very large number of components.

Multidimensional liquid chromatography (MDLC) is well-suited to this task. Chapter 1 of

this dissertation introduces the theoretical framework behind MDLC. Previous work

involving protein separations using comprehensive two-dimensional liquid chromatography

(LC x LC), which is a form of MDLC, is also discussed.

Chapters 2 and 3 of the dissertation focus on the development of methods for intact

protein separations using liquid chromatography. Several separation modes were evaluated,

including size exclusion, ion exchange, and reversed phase. Two-dimensional separations of

E. coli proteins were performed which use anion exchange in the first dimension and ultra-high pressure reversed-phase LC in the second dimension. Although high peak capacities were demonstrated, the technique was limited in that proteins could not be identified solely based on the intact protein molecular weight data which was obtained.

The research presented in Chapters 4 and 5 use the same intact-protein LC x LC separations, but also incorporates enzymatic digestion of proteins followed by LC-MS analysis of the resulting peptides. The resulting technique is a hybrid of "top-down" and "bottom-up" proteomics methods. It allows proteins to be identified on the basis of tandem mass spectra of peptides, but retains the information gained from intact protein MS. In Chapter 6, this technique was applied to study differential protein expression in yeast cultures grown under different conditions.

ACKNOWLEDGMENTS

There are many people who I would like to thank for their support during the completion of my dissertation. First is my research advisor, Dr. James Jorgenson. His guidance, advice and support were always first-rate, and his enthusiasm for research was an inspiration. Also crucial were fellow graduate students Jason Link, John Eschelbach and Will Thompson, who helped me learn the techniques I needed to carry out my research. Likewise, I sincerely appreciate the help of Brenna McJury, who worked with me in acquiring much of the data presented in the latter chapters of this dissertation, and will be carrying on the project. Also, I'd like to thank our collaborators at Waters Corporation, Keith Fadgen and Scott Berger, who provided me with resources and insights which helped make my dissertation possible.

Of equal importance are the people who supported me on a personal level over the course of my career as a graduate student. Thank you to my good friends Patty Dennis and Alicia Douglas, with whom I shared enjoyable camping trips, mountain bike rides and countless conversations. Special thanks to Patty, for helping me keep things in perspective over the years. Also, I appreciate the support and friendship of all the members of the Jorgenson research group. Finally, thank you to my family – my mother and father, sister and brother, and my extended family as well – for their constant support and encouragement, throughout my time in graduate school and at every step along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1D	one-dimensional
2D	two-dimensional
2D-LC	two-dimensional liquid chromatography
2D-PAGE	two-dimensional polyacrylamide gel electrophoresis
3D	three-dimensional
Å	Angstroms
ACN	acetonitrile
AutoME	automated maximum entropy de-convolution
AXC	anion exchange chromatography
BEH	bridged-ethyl hybrid particle
BPI	base peak intensity
BSA	bovine serum albumin
CapLC	Waters capillary LC system
CID	collision induced dissociation
cm	centimeter
CXC	cation exchange chromatography
Da	Dalton
DDA	data directed analysis
ESI	electrospray ionization
FTICR	Fourier transform ion cyclotron resonance
g	grams
GPTMS	3-glycidoxypropyltrimethoxysilane

HFBA	heptafluorobutyric acid
HIC	hydrophobic interaction chromatography
HILIC	hydrophilic interaction chromatography
HPLC	high performance liquid chromatography
Hz	hertz
ID	inner diameter
IEC	ion exchange chromatography
IEF	isoelectric focusing
IgG	immunoglobulin G
kDa	kilodaltons
krpm	1000 revolutions per minute
kV	kilovolts
L	liter
LB	Luria-Bertani
LC	liquid chromatography
LC x LC	comprehensive two-dimensional liquid chromatography
M	molar
m	meter
m/z	mass to charge ratio
MALDI	matrix assisted laser desorption ionization
MaxEnt	maximum entropy spectral de-convolution
MDLC	multidimensional liquid chromatography
mg	milligram

min	minute
mL	milliliter
mm	millimeter
mM	millimolar
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MudPIT	multidimensional protein identification technology
MW	molecular weight
N	normal
ng	nanogram
nL	nanoliter
nm	nanometer
nM	nanomolar
NMR	nuclear magnetic resonance
NPC	normal phase chromatography
OD	outer diameter
PDA	photodiode array
PEEK	polyetheretherketone
PLGS2	ProteinLynx Global Server 2 (Waters)
PMF	peptide mass fingerprinting
ppm	parts per million
psi	pounds per square inch
PTM	post-translational modification

Q	quadrupole
Q-TOF	quadrupole-time of flight
RPLC	reversed-phase liquid chromatography
RT	retention time
SCX	strong cation exchange
SDS	sodium dodecyl sulfate
sec	second
SEC	size exclusion chromatography
TFA	trifluoroacetic acid
TIC	total ion current
TOF	time of flight
UHP	ultra-high pressure
UHPLC	ultra-high pressure liquid chromatography
UPLC	ultra performance liquid chromatography
UV	ultraviolet
V	volts
v/v	volume to volume ratio
VI	LabVIEW virtual instrument
μg	micrograms
μL	microliter
μm	micrometer
μM	micromolar

LIST OF SYMBOLS

A	eddy diffusion term from van Deemter equation
B	longitudinal diffusion term from van Deemter equation
C	resistance to mass transfer term from van Deemter equation
d_p	particle diameter
H	plate height (height equivalent to a theoretical plate)
L	length
m	ion mass
M	molecular mass
n_c	peak capacity
t	time
u	mobile phase linear velocity
z	charge

CHAPTER 1: Multidimensional separations and proteomics: theory, background and applications

1.1   Proteomics

The study of the expression, structure and function of proteins has always been at the center of efforts to understand the processes which enable life and which are responsible for the occurrence of disease. Although the study of proteins is by no means a new endeavor, recent advances in analytical technology have enabled substantial changes in the means by which proteins can be studied. In a conventional approach, a single protein of interest is selected, isolated, purified, and then studied using any of a variety of techniques, which might include amino acid analysis, Edman degradation, x-ray crystallography, NMR, or mass spectrometry. This strategy of single-protein analysis is still widely employed, and is indeed often the only way by which detailed information regarding protein structure and function may be obtained. Proteomics, however, describes a more global approach to the study of proteins. Its aim is to characterize the entire complement of proteins – termed the proteome – produced by an organism. It offers a broader picture, albeit at the expense of some detail, and is often able to reveal significant information regarding patterns of protein expression which might not be detected by more targeted analyses of individual proteins. A major focus of this dissertation will be the development of analytical methods – particularly chemical separations – which are readily applied to proteomics.

1.1.1 Challenges associated with proteomics

The field of proteomics poses numerous analytical challenges. First is the complexity of proteomes in terms of the sheer number of components. Genetic data predicts that the prokaryotic bacterium *E. coli*, a relatively simple unicellular organism, is capable of producing in excess of 4,000 distinct proteins.[1] The human proteome is substantially more complex. Human genes encode an estimated complement of 20,000-30,000 proteins, while the actual number of distinct proteins is substantially higher due to the possibility of multiple proteins originating from a single gene and due to post-translational modifications (PTMs).[2] These modifications, such as phosphorylation, glycosylation, and ubiquitination, are known to have important effects on proteins such as regulating their activity, specifying their intended location within a cell, or targeting them for destruction.[3] Therefore, PTMs must be considered in any thorough attempt to characterize the proteome of an organism. Furthermore, in any given organism, tissue or cell, different proteins are likely to be present at concentrations spanning six or more orders of magnitude.[4] In multicellular organisms, the picture is also complicated by the heterogeneity of protein distribution – proteins expressed in one type of cell or tissue may differ dramatically from those expressed in others. Finally, the proteome of an organism is not a static entity. The proteins being produced and their relative abundance are likely to change in response to environmental conditions or other factors. All of these factors make proteome analysis a formidable task.

While no single analytical technique yet developed has all of the capabilities necessary to fully characterize the proteome of an organism, certain criteria are particularly important in selecting a workable strategy for proteome analysis. Some of the major steps typically involved in proteomics include separation of a complex mixture of proteins into individual components or more manageable subsets, identification and study of the proteins

which are present, and quantification to assess their absolute or relative abundance. For the separation component of the analysis, high resolution is a requirement. If the separation is not able to adequately resolve the components of the mixture, which is likely to contain hundreds or thousands of proteins or peptides, identification and quantification will be difficult or impossible. After the separation, some technique must be used to detect and identify the proteins, and to study features of interest, such as post-translational modifications. For identification, a criterion of great importance is sensitivity. Enhanced sensitivity will improve both the number and the accuracy of the protein identifications which can be attained. While numerous techniques have been investigated, the most versatile method for identifying proteins is mass spectrometry, which has, over the past twenty years, become an indispensable part of proteomics research. Finally, to quantify proteins, a technique with a large dynamic range is required in order to ensure that a wide range of concentrations can be detected simultaneously. Traditionally, some form of protein staining or radioisotope labeling followed by visualization of spots on a gel has been used, but in recent years mass spectrometry has gained some acceptance for relative quantification in proteomics.[5, 6]

Clearly, proteomics cannot be performed by relying exclusively on any one analytical aspect of the task – whether separation, identification, or quantification. However, improvements in any one of these areas can enhance the capabilities of the method as a whole. The focus of this dissertation will be on the development of high-resolution separation techniques, with a view toward their application to the field of proteomics.

1.1.2 Choice of strategies: bottom up versus top down

Two major strategies exist for proteomic studies, which have been termed "bottom-up" and "top-down".[7] In the bottom-up approach, the sample containing a mixture of

proteins is first subject to digestion using a proteolytic enzyme – most commonly trypsin, though others are also frequently used – which breaks down the proteins into peptides. The mixture of peptides is then separated and these peptides are analyzed using a mass spectrometer, such that, in an ideal case, a mass spectrum or tandem mass spectra are obtained of each peptide. From these data, the peptides can be sequenced manually or, more typically, sequenced and identified using probability-based database searching. From the search results, a list of proteins likely to have been present in the original sample is also produced.

In top down proteomics, no enzymatic digestion of the proteins is performed. Instead, the intact proteins are separated and then sent to the mass spectrometer. If any information beyond the intact protein molecular weight is desired – which in itself is usually insufficient to identify the protein – then the proteins are subjected to gas-phase fragmentation inside the mass spectrometer.[8] Patterns of fragmentation are then studied in order to identify the proteins.

Both approaches have distinct advantages and disadvantages. From a separation standpoint, the bottom-up approach is typically regarded as more straightforward, since reversed-phase liquid chromatography (RPLC), which is a high resolution separation technique and is easily coupled to mass spectrometry, is well suited to the analysis of peptide mixtures. Intact protein separations, particularly using reversed-phase liquid chromatography, are more prone to difficulties such as poor peak shape and sample carryover[9-12]. Therefore, fewer proteins can be resolved in a single separation than is possible with peptides. However, the proteolytic cleavage performed as a part of a bottom-up analysis increases the difficulty of the separation, because when a mixture containing

several hundred proteins is digested, several thousand or even tens of thousands of peptides will be produced. In intact protein separations, the complexity of the mixture is not increased by enzymatic digestion, so the separation does not need to be able to resolve as many components.

In terms of data collection and analysis, the bottom-up approach involving tandem mass spectra (MS/MS) of peptides is well established, and numerous database searching programs exist which are routinely used to accurately identify proteins from these spectra.[13] However, since mass spectra are typically not obtained for every peptide predicted from the protein sequence, some low abundance proteins may be missed entirely, and certain information such as the location of post-translational modifications may not be obtained for even high abundance proteins. An advantage of the top-down approach is that whole proteins are introduced into the mass spectrometer. This allows the possibility of performing more thorough studies in terms of protein structure and analysis of post-translational modifications.[14] Since gas-phase fragmentation of proteins produces very complex mass spectra, however, only high-resolution mass spectrometers such as Fourier transform – ion cyclotron resonance (FTICR) instruments are generally regarded as being well suited to the task. Recently, there has been some interest in analyses which combine aspects of both bottom-up and top-down methods.[15-17] To some extent, this allows the best characteristics of each technique to be shared in a single analysis. Ultimately, the choice between a bottom-up and top-down approach is dictated by the type of information which is desired from the study and the instrumentation which is available to carry it out.

1.1.3 Conventional separation methods for proteomics

Regardless of whether a top-down or bottom-up approach is used, separations are an important part of nearly all proteomic studies. Historically, most separations of proteins have

been carried out using gel electrophoresis, a technique in which proteins are resolved by their different rates of migration through a polymeric gel in the presence of an applied electric field. The most powerful gel electrophoresis technique for proteomics is known as 2D polyacrylamide gel electrophoresis, or 2D-PAGE. First reported in 1975 by O'Farrell,[18] 2D-PAGE resolves proteins into spots on a flat gel using two different separation mechanisms. First, proteins are separated according to their isoelectric point via isoelectric focusing (IEF). This is typically performed using a gel strip with an immobilized pH gradient.[19] When a protein is introduced into the gel, it migrates to the location at which its isoelectric point matches the pH of the surrounding gel medium. At this location the net charge of the protein is zero, so it is no longer influenced by the electric field, and its migration stops. Once all proteins have been allowed to focus into bands according to their isoelectric point, the applied electric field is turned off. To carry out the second separation, the IEF gel is then placed in contact with one side of a flat slab of polyacrylamide gel containing the surfactant molecule sodium dodecyl sulfate (SDS). SDS is composed of a hydrophobic twelve-carbon alkane group, which tends to complex with proteins, and a hydrophilic sulfate group, which imparts all proteins with a negative charge. An electric field is applied perpendicular to the orientation of the IEF gel, which causes the negatively charged SDS-coated proteins to enter the polyacrylamide gel and migrate toward the positive electrode. Larger proteins migrate more slowly, due to greater frictional hindrance in moving through the gel matrix. Thus the proteins are separated according to their size. Once good separation has been achieved, the electric field is turned off, leaving proteins distributed at various spots throughout the gel. The protein spots are then made visible by exposing the gel to a visible or fluorescent stain

which binds to the proteins, or by exposure of the gel to photographic film if the proteins were labeled with radioactive isotopes.

2D-PAGE is capable of producing very high resolution separations. For a typical mixture, where proteins are randomly distributed throughout the 2D gel, it is estimated that over 2000 components can be resolved.[19] Differential analyses are also possible with 2D-PAGE, in which changes in levels of protein expression between two similar samples can be detected. This is achieved by running two samples under identical conditions, then visually comparing the images to identify spots on the gels which differ substantially in terms of intensity from one run to the next. More recently, computer software has become available to automate the process of comparing gel spot intensities.

Although 2D-PAGE is a very powerful technique for protein separations, it has certain fundamental limitations. It is a labor-intensive analysis that usually requires at least two days to complete. It cannot be coupled to mass spectrometry in an on-line fashion. In order to obtain mass spectra of peaks of interest, protein spots must be excised from the gel manually or using robotics systems, after which the proteins must be extracted from the gel matrix and individually analyzed using MS. 2D-PAGE does not yield good separations of certain types of proteins, especially proteins that are extremely large or hydrophobic, which may not enter the gel, and proteins which are very acidic or basic, which tend to be poorly resolved. Also, 2D-PAGE is only semi-quantitative, and proteins present at very low abundance may fall be below the detection limit of the staining techniques used to visualize the proteins.[19] Given these shortcomings, there is ample reason to attempt to develop a technique which would offer some of the same capabilities of 2D-PAGE, with fewer of its limitations.

1.1.4 Application of liquid chromatography to proteomics

Many of the limitations of 2D-PAGE are not shared by separation methods based on liquid chromatography. When appropriate conditions are used, LC can separate a broader range of proteins than 2D-PAGE, including those that are large, hydrophobic, acidic or basic.[20] Depending on the column and detection technique used, LC can have greater sensitivity to proteins present in small amounts than 2D-PAGE. Perhaps most importantly, separation methods such as liquid chromatography can be fully automated and can be coupled with a variety of detection methods such as UV absorbance, laser-induced fluorescence or mass spectrometry. Given these advantages, it is not surprising that researchers have looked to liquid chromatography and related techniques to provide an alternative to 2D-PAGE. The ongoing challenge for separation scientists has been to devise LC-based separation methods and instrumentation that can equal 2D-PAGE in resolving power.

In order to develop LC methods which give separating power equivalent to 2D-PAGE, several approaches are available. One strategy is to use long columns and to increase the gradient length. While this approach is straightforward and has been shown to be successful,[21] to attain a desired increase in resolution using this method requires an exponential increase in run time. Thus this approach quickly reaches a practical limit in terms of the time available to run the separation. Another approach is to use LC columns packed with smaller particles. Over the past ten years, it has been demonstrated that using columns packed with particles with diameters between 1-2 μm, as opposed to the 3-5 μm range which were the previous standard for HPLC, enables higher resolution separations to be achieved while simultaneously reducing run time.[22-26] The tradeoff for these improvements is that much higher pressures are required to force mobile phase through the

packed bed of small particles. Pump technology has been developed to meet these requirements; the resulting technique has been termed ultra-high pressure liquid chromatography (UHPLC), or more recently, ultra-performance liquid chromatography (UPLC). Although UHPLC has been used to produce separations with peak capacities over one thousand[27] and has been shown to be applicable to proteomics[28], no single HPLC or UHPLC separation has yet demonstrated peak capacities as high as 2D-PAGE, especially for intact protein separations. To achieve such truly high peak capacities, often the only option is to perform a multidimensional separation.

1.2   Multidimensional separations

A multidimensional separation involves the coupling of two or more separation mechanisms in order to carry out a single analysis.[29, 30] Although simple in concept, multidimensional separations can be exceedingly powerful in terms of resolution. The need for multidimensional separations arises from the inability of one-dimensional separation methods to adequately resolve highly complex samples, such as those encountered in proteomics, as has already been discussed.

1.2.1   Background and theory

In a multidimensional separation, a sample is first subjected to separation via one method, and then the separated components are further separated by at least one additional method. Although there is no inherent limitation to the number of independent separation methods that can be coupled, practical constraints have limited the vast majority of the multidimensional separations reported to date to two dimensions. Two dimensional separations provide enhanced resolution and peak capacity as compared to 1D separations, where peak capacity is defined as the maximum number of peaks that can fit into the accessible separation space side by side with a resolution of 1.0 from each neighboring

9

peak.[31, 32] In an ideal case, the peak capacity of a 2D separation is given by Equation 1-1:

$$n_c = n_1 \times n_2$$ (Equation 1-1)

where n_c is the total peak capacity and n_1 and n_2 are the peak capacity of the first and second dimensions, respectively.[30] Hypothetically, if two techniques with a peak capacity of 100 each were used in a 2D separation, the resulting system would have a maximum peak capacity of 10,000.   Although this theoretically means that a maximum of 10,000 different components could be fully separated using this method, due to the usually random distribution of peaks and the statistical probability of peak overlap, the actual number of detectable components that can be expected to be fully resolved is substantially smaller than the peak capacity.[33, 34] Thus, achieving very high peak capacities using 2D separations is clearly desirable, though it may not be a trivial task due to the challenges associated with selecting and coupling two appropriate separation methods.

Due to the large number of 2D separations reported in the literature, it is important to make clear distinctions between different classes of 2D separations. One important concept is the difference between separations that are 2D-in-space versus 2D-in-time. A separation which is 2D-in-space results in a separation of analytes over a physical two-dimensional (planar) surface. 2D-PAGE, which was described earlier in this chapter, is a prime example of a 2D-in-space separation. Techniques that are 2D-in-time are carried out by first performing a 1D separation – for example, using a LC column – and then sequentially subjecting individual fractions from this first separation to a second 1D separation. 2D-LC, also known as LC x LC, is an example of a 2D-in-time method. Among techniques that are 2D-in-time, another distinction must be made between off-line and on-line multidimensional separations. Off-line methods involve collecting fractions from the first dimension in tubes

or other containers, then later subjecting the contents of these fractions to separation on the second dimension. On-line techniques employ switching valves or other instrumentation that allow the fractions from the first dimension to be transferred directly to and be separated by the second dimension, usually while the first dimensional separation continues simultaneously.

Certain theoretical considerations apply to all multidimensional separations, regardless of whether a technique is 2D in space or time, off-line or on-line. Giddings established two such fundamental requirements for ideal multidimensional separations.[29, 30] First, the separation mechanisms must be orthogonal. This means that there should be no correlation between the elution order of analytes in one dimension with their elution order in any subsequent dimension. For a 2D separation, this yields a chromatogram in which the separated components are distributed randomly over the separation space, which maximizes peak capacity. Hypothetical examples of orthogonal and non-orthogonal 2D separations are shown in Figure 1-1. Often, orthogonal separations are achieved by ensuring that the chemical or physical property by which analytes are separated is different for each dimension. For example, reversed phase liquid chromatography (RPLC), which separates based on hydrophobicity, is orthogonal to ion exchange chromatography (IEC), which separates according to coulombic interactions. Interestingly, however, it has been shown that two separations can be orthogonal even when their separation modes are identical, if some chemical condition is changed which alters the selectivity of one of the dimensions. Gilar *et al*. demonstrated a 2D reversed-phase / reversed phase separation of peptides which was orthogonal and generated high peak capacity when the two separations were carried out at substantially different pH values.[35]

Giddings' second criterion states that no resolution gained in the first dimension of the separation may be lost in any subsequent dimension. For on-line, 2D-in-time separations, this requirement is satisfied by designing instrumentation that collects fractions or transfers them from the first dimension to the second at a frequency sufficient to ensure the resolution of the first dimension is not substantially diminished. Therefore, in an on-line 2D separation the second dimension must be capable of completing separations very rapidly, in order to allow many second-dimension runs to occur within the time it takes for a single first-dimension separation to finish. The resulting need for high speed analyses in the second dimension represents an experimental challenge in carrying out 2D separations. Exactly how frequently the first dimension must be sampled depends on the width of the first-dimension peaks. Murphy et al. proposed that, in order to achieve optimal resolution, each peak in the first dimension should be sampled at least three times by the second dimension.[36] More recently, Horie et al. reported that sampling the first dimension column 1 to 2 times over the width of a first dimension peak is typically preferable, because the reduced sampling rate prevents excessively short run times on the second dimension.[37] This improves the total peak capacity of the 2D separation, even though the resolution contributed by the first dimension may be somewhat reduced. In spite the challenges associated with two-dimensional separations, methods that satisfy Giddings' criteria and produce good separations of complex mixtures have been demonstrated.

1.2.2 Previous work in multidimensional liquid chromatography

One multidimensional method that has been used commonly is an on-line, 2D-in-time technique called heart-cutting two-dimensional liquid chromatography. In this approach, the first-dimension separation is carried out like a standard 1D analysis on a chromatography column. A single desired segment of the first-column effluent is then transferred, typically

via a switching valve, to a second column for further separation via a different method. Heart-cutting is useful when extra resolution is needed to examine a small segment of peaks in a complex chromatogram, but it requires foreknowledge of the sample's composition. Since heart-cutting does not subject the entire sample to two-dimensional separation, it is of little use when every component of a sample is an analyte of interest, as in proteomics. The alternative to heart-cutting is known as "comprehensive" multidimensional separations, because all sample components are subjected to displacement on both dimensions.[38] Comprehensive multidimensional separations are more powerful than heart-cutting separations because they provide greater peak capacity.

A wide variety of combinations of liquid chromatographic modes have been used in comprehensive LC x LC separations, including ion exchange-reversed phase (IEC x RPLC),[39-45] size exclusion-reversed phase (SEC x RPLC),[46-49] reversed phase-size exclusion (RPLC x SEC),[36], ion exchange-size exclusion (IEC x SEC),[38] and normal phase-reversed phase (NPC x RPLC).[50] Although many different instrumental setups for LC x LC have been designed, most of the components used are similar, and include an injector, two isocratic or gradient LC pumps, a single column for the first dimension, one or more columns for the second dimension, one or more computer-controlled switching valves, and an appropriate detection system.

The first true comprehensive system for two-dimensional liquid chromatographic separation was reported in 1990 by Bushey and Jorgenson.[38] This system was used to analyze a mixture of protein standards and a sample of human serum. A cation exchange column operated in gradient elution mode for the first dimension is coupled to a size exclusion column used for the second dimension. An eight-port switching valve is used to

interface the two columns and allows on-line fraction transfer from the first column to the

second. A diagram of the instrumental setup of this LC x LC system is provided in Figure

1-2. The effluent of the cation exchange column is directed to a storage loop. When the

valve is switched, the isocratic LC pump forces the contents of the loop onto the size

exclusion column where they are further separated. Meanwhile, the effluent of the first

dimension is directed to the second storage loop. The flow rates on the two columns are

selected such that the analysis time of the second dimension exactly matches the amount of

time it takes for the first dimension to fill the storage loop. This allows all the effluent of the

first column to be transferred to the second dimension. A computer is used to control the

switching of the valve and the acquisition of data from a UV absorbance detector at the end

of the second column. In this manner the SEC column can analyze a large number of

fractions from the ion exchange column during the course of a run lasting a total of 2.5 to 6

hours. The chromatographic data are presented as a three-dimensional view of a 2D

chromatogram as shown in Figure 1-3. The system was estimated to have a total peak

capacity of approximately 130, which is the product of the peak capacities visually estimated

for each dimension. Other examples of column-switching LC x LC separations using dual

storage loops have since been reported in the literature, some with greater peak capacity.[36, 39,

40, 51] A 2D anion exchange-reversed phase system that was used to analyze a tryptic digest of

reduced porcine thyroglobulin gave a peak capacity over 2,000.[43]

A different approach, in which the effluent from the first column is not captured by

storage loops but instead is transferred directly to the head of one of two parallel second-

dimension columns, was reported in 1997 by Opiteck, Jorgenson and Anderegg.[47] This

system was used to analyze the fragments produced by tryptic digests of the proteins

ovalbumin and bovine serum albumin. The first dimension of their system consisted of six

size-exclusion columns connected in series. These were coupled via a pair of four-port

valves to two parallel reversed phase LC columns, which serve as the second dimension.

This study was also one of the first reports of on-line mass spectrometric detection used to

analyze a two dimensional separation. A diagram of the entire setup is shown in Figure 1-4.

As sample elutes from the series of SEC columns, it is routed to RPLC column alpha. Since

the aqueous buffers used for SEC are weak eluents for RPLC, the sample material is

concentrated in a narrow zone at the head of the column. When the two four-port valves are

switched simultaneously, a second LC pump starts a gradient to elute the sample from RPLC

column alpha, while the effluent from the first dimension is loaded onto RPLC column beta.

The effluent from the RPLC column being eluted is directed to a UV absorbance detector

followed by a 10:1 flow splitter and an electrospray mass spectrometer. The total peak

capacity of this system was estimated to be about 500. Other two-dimensional LC x LC

techniques using parallel columns in the second dimension have been reported, many of

which replace the two four-port switching valves with a single ten-port valve that

accomplishes the same purpose.[41, 42, 45, 52, 53]

1.2.3   Recent applications of multidimensional liquid chromatography to proteomics

The most prominent recent technique in which multidimensional liquid

chromatography has been applied to the field of proteomics is a bottom-up technique known

as Multidimensional Protein Identification Technology (MudPIT), developed by Yates and

colleagues.[54-60] As opposed to comprehensive LC x LC systems, which use separate columns

for each dimension coupled through a column switching valve, MudPIT uses a single

capillary column packed with two different stationary phases, with a pulled tip at the outlet

end for direct coupling to a mass spectrometer via electrospray ionization. The end of the

capillary nearest the pulled tip is packed with several centimeters of reversed phase material, followed by several centimeters of strong cation exchange (SCX) material. The column is connected to a polyetheretherketone (PEEK) microcross which splits the flow from a conventional quaternary LC pump, reducing it to 0.15-0.25 μL/min, and also provides a connection where the electrospray voltage is applied, as shown in Figure 1-5. The 2D separation is carried out using a multi-step gradient elution profile, an example of which is shown in Figure 1-6. To elute fractions of the sample from the ion-exchange portion of the column onto the reverse phase material, a step-gradient of buffer with increasing ionic strength is used. In between each of these incremental steps, a linear gradient of increasing organic content is used to separate the eluted fraction on the reversed phase segment of the column. The outlet of the column is interfaced with an electrospray mass spectrometer. Typically the mass spectrometers used with MudPIT have tandem MS capabilities to enable the identification of proteins via "shotgun" analysis. In this approach, the tandem mass spectra which are obtained are compared with a database containing known protein sequences using a probability-based searching method. The results of this analysis are used to identify the proteins present in the sample.

The capabilities of MudPIT as a separation method for the purpose of proteomics have been assessed by several studies. The peak capacity of the two-dimensional separation carried out in one 15-fraction MudPIT analysis was estimated to be approximately 3,200.[56] When the inherent peak capacity of the mass spectrometer was included, the estimated overall peak capacity increased to 23,000, which compares favorably with 2D-PAGE techniques. In practice, MudPIT is capable of identifying over 1,000 proteins in a single sample. Washburn and colleagues applied MudPIT to the analysis of the proteome of the

yeast *S. cerevisiae* and identified 5440 peptides using mass spectrometry, which were assigned via database searching to originate from 1484 unique proteins.[55] Among these were a substantial number of low-abundance and trans-membrane proteins, which suggests that the technique is largely unbiased and gives a representative sampling of the yeast proteome. Lack of bias represents a significant advantage over 2D-PAGE, which has limited dynamic range and typically performs poorly with trans-membrane proteins since they are hydrophobic and do not easily enter into the gel.

Another report indicates that the number of protein identifications can be further improved by using a three-phase MudPIT column in which an additional section of RP material is added at the inlet of the column to provide on-line removal of salts from the sample.[58] MudPIT has also been used for quantitative studies of proteomic samples via the use of ^{15}N isotope labeling.[57] An assessment of quantitative MudPIT suggests that the technique can provide useful data about relative levels of protein expression, but that experimental reproducibility could be improved by enhancing resolution in order to increase the number of peptides detected per protein.[60]

One of the greatest advantages of MudPIT is the relative simplicity of the instrumentation needed to carry out 2D separations. Only a single quaternary gradient LC pump, an appropriately packed capillary column, and minimal interfacing components are required. However, the simplicity of this arrangement also limits its flexibility. The two separation modes coupled in tandem must both utilize gradient elution. Further, the gradient in the first dimension must always be run in a step-wise fashion in order to elute concise fractions of sample onto the second column. The resolution contributed by the first dimension is limited by the number of fractions transferred, which is equal to the number of

steps used in the first dimension gradient. Each step requires substantial time, typically approximately 100 minutes, which includes periods for re-equilibration and washing of salts off the column. A maximum of fifteen steps in the first-dimension salt gradient have been reported,[55] which may suggest that the first dimension is under sampled in spite of long run times. Notably, the only combination of separation modes that has been used thus far in MudPIT is strong cation exchange – reversed phase, perhaps because this is one of few arrangements where suitable gradients can be applied alternately to two different stationary phases in a single column in order to bring about an effective, orthogonal 2D separation.

1.3   Scope of dissertation

Given the importance of the field of proteomics and the increasing interest in using chromatography as an alternative to 2D-PAGE, there is ample motivation for ongoing research into advancing the capabilities of LC for protein separations. Thus, although the research described in this dissertation involves many aspects of proteomics – including mass spectrometry and database searching for protein identification – the focus of the research will be on multidimensional liquid chromatography. First, an evaluation of various 1D chromatographic methods and research directed toward improving their capabilities for protein separations will be described (Chapter 2). Next, an off-line LC x LC separation using UHPLC for separation of intact proteins will be discussed (Chapter 3). Then, two hybrid top down / bottom up approaches to analysis of the *E. coli* proteome using off-line (Chapter 4) and on-line (Chapter 5) LC x LC-MS will be presented. Finally, the same hybrid top down / bottom up on-line LC x LC-MS technique will be applied to several samples of the proteome of the yeast *S. cerevisiae* grown under differing conditions, in order to detect differences in protein expression (Chapter 6).

1.4   Acknowledgment

1.5 References

(1) Regnier, F. E.; Amini, A.; Chakraborty, A.; Geng, M.; Ji, J.; Riggs, L.; Sioma, C.; Wang, S.; Zhang, X. *LC-GC North America* **2001**, *19*, 200-213.

(2) Southan, C. *Proteomics* **2004**, *4*, 1712-1726.

(3) Han, K. K.; Martinage, A. *International Journal of Biochemistry* **1992**, *24*, 19-28.

(4) Wu, L.; Han, D. K. *Expert Review of Proteomics* **2006**, *3*, 611-619.

(5) Paoletti, A. C.; Washburn, M. P. *Biotechnology & Genetic Engineering Reviews* **2006**, *22*, 1-19.

(6) MacCoss, M. J.; Matthews, D. E. *Analytical Chemistry* **2005**, *77*, 294A-302A.

(7) Wehr, T. *LC-GC North America* **2006**, *24*, 1006-1008.

(8) Reid, G. E.; McLuckey, S. A. *Journal of Mass Spectrometry* **2002**, *37*, 663-675.

(9) Kastner, M. *Protein Liquid Chromatography*; Elsevier: Amsterdam, Netherlands, 2000.

(10) Cohen, K. A.; Schellenberg, K.; Benedek, K.; Karger, B. L.; Grego, B.; Hearn, M. T. W. *Analytical Biochemistry* **1984**, *140*, 223-235.

(11) Cohen, S. A. *Analytical Chemistry* **1984**, *56*, 217-221.

(12) Cohen, S. A.; Benedek, K.; Tapuhi, Y.; Ford, J. C.; Karger, B. L. *Analytical Biochemistry* **1985**, *144*, 275-284.

(13) Ahrens, C.; Jespersen, H.; Schandorff, S. *Methods of Biochemical Analysis* **2005**, *45*, 249-272.

(14) Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196A-203A.

(15) Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *Journal of Proteome Research* **2006**, *5*, 135-146.

(16) Berger, S. J.; Millea, K. M.; Krull, I. S.; Cohen, S. A. In *Separation Methods in Proteomics*; Smejkal, G. B., Lazareu, A., Eds.; CRC Press LLC: Boca Raton, Florida, 2006, pp 387-417.

(17) Borchers, C. H.; Thapar, R.; Petrochenko, E. V.; Torres, M. P.; Speir, J. P.; Easterling, M.; Dominski, Z.; Marzluff, W. F. *Proceedings of the National Academy of the Sciences of the United States of America* **2006**, *2006*, 9.

(18) O'Farrell, P. H. *Journal of Biological Chemistry* **1975**, *250*, 4007-4021.

(19)   Wehr, T. *LC-GC North America* **2001**, *19*, 702-711.

(20)   Wehr, T. *LC-GC North America* **2002**, *20*, 954-957.

(21)   Wang, X.; Barber, W. E.; Carr, P. W. *Journal of Chromatography A* **2006**, *1107*, 139-151.

(22)   MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W. *Analytical Chemistry* **1997**, *69*, 983-989.

(23)   MacNair, J. E.; Patel, K. D.; Jorgenson, J. W. *Analytical Chemistry* **1999**, *71*, 700-708.

(24)   Jerkovich, A. D.; Mellors, J. S.; Jorgenson, J. W. *LC-GC North America* **2003**, *21*, 600-610.

(25)   Patel, K. D.; Jerkovich, A. D.; Link, J. C.; Jorgenson, J. W. *Analytical Chemistry* **2004**, *76*, 5777-5786.

(26)   Mellors, J. S.; Jorgenson, J. W. *Analytical Chemistry* **2004**, *76*, 5441-5450.

(27)   Plumb, R. S.; Rainville, P.; Smith, B. W.; Johnson, K. A.; Castro-Perez, J.; Wilson, I. D.; Nicholson, J. K. *Analytical Chemistry* **2006**, *2006*, 7278-7283.

(28)   Shen, Y.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 3090-3100.

(29)   Giddings, J. C. *Analytical Chemistry* **1984**, *56*, 1258A-1270A.

(30)   Giddings, J. C. *Journal of High Resolution Chromatography & Chromatography Communications* **1987**, *10*, 319-323.

(31)   Giddings, J. C. *Analytical Chemistry* **1967**, *39*, 1027-1028.

(32)   Grushka, E. *Analytical Chemistry* **1970**, *42*, 1142-1147.

(33)   Davis, J. M.; Giddings, J. C. *Analytical Chemistry* **1985**, *57*, 2168-2177.

(34)   Davis, J. M.; Giddings, J. C. *Analytical Chemistry* **1985**, *57*, 2178-2182.

(35)   Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Journal of Separation Science* **2005**, *28*, 1694-1703.

(36)   Murphy, R. E.; Schure, M. R.; Foley, J. P. *Analytical Chemistry* **1998**, *70*, 1585-1594.

(37)   Horie, K.; Kimura, H.; Ikegami, T.; Iwatsuka, A.; Saad, N.; Fiehn, O.; Tanaka, N. *Analytical Chemistry* **2007**, *79*, 3764-3770.

(38)   Bushey, M. M.; Jorgenson, J. W. *Analytical Chemistry* **1990**, *62*, 161-167.

(39)   Holland, L. A.; Jorgenson, J. W. *Analytical Chemistry* **1995**, *67*, 3275-3283.

(40)   Opiteck, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. *Analytical Chemistry* **1997**, *69*, 1518-1524.

(41)   Wagner, K.; Racaityte, K.; Unger, K. K.; Miliotis, T.; Edholm, L. E.; Bischoff, R.; Marko-Varga, G. *Journal of Chromatography A* **2000**, *893*, 293-305.

(42)   Unger, K. K.; Racaityte, K.; Wagner, K.; Miliotis, T.; Edholm, L. E.; Bischoff, R.; Marko-Varga, G. *Journal of High Resolution Chromatography* **2000**, *23*, 259-265.

(43)   Holland, L. A.; Jorgenson, J. W. *Journal of Microcolumn Separations* **2000**, *12*, 371-377.

(44)   Wagner, K.; Miliotis, T.; Marko-Varga, G.; Bischoff, R.; Unger, K. K. *Analytical Chemistry* **2002**, *74*, 809-820.

(45)   Haefliger, O. P. *Analytical Chemistry* **2003**, *75*, 371-378.

(46)   Erni, F.; Frei, R. W. *Journal of Chromatography* **1978**, *149*, 561-569.

(47)   Opiteck, G. J.; Jorgenson, J. W.; Anderegg, R. J. *Analytical Chemistry* **1997**, *69*, 2283-2291.

(48)   Opiteck, G. J.; Ramirez, S. M.; Jorgenson, J. W.; Moseley, M. A., III *Analytical Biochemistry* **1998**, *258*, 349-361.

(49)   Stroink, T.; Wiese, G.; Lingeman, H.; Bult, A.; Underberg, W. J. M. *Analytica Chimica Acta* **2001**, *444*, 193-203.

(50)   Murphy, R. E.; Schure, M. R.; Foley, J. P. *Analytical Chemistry* **1998**, *70*, 4353-4360.

(51)   Liu, Z.; Lee, M. L. *Journal of Microcolumn separations* **2000**, *12*, 241-254.

(52)   Opiteck, G. J.; Jorgenson, J. W.; Moseley, M. A., III; Anderegg, R. J. *Journal of Microcolumn Separations* **1998**, *10*, 365-375.

(53)   Murahashi, T. *Analyst* **2003**, *128*, 611-615.

(54)   Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates, J. R., III *Nature Biotechnology* **1999**, *17*, 676-682.

(55)   Washburn, M. P.; Wolters, D. A.; Yates, J. R., III *Nature Biotechnology* **2001**, *19*, 242-247.

(56)   Wolters, D. A.; Washburn, M. P.; Yates, J. R., III *Analytical Chemistry* **2001**, *73*, 5683-5690.

(57)   Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., III *Analytical Chemistry* **2002**, *74*, 1650-1657.

(58)   McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R., III *International Journal of Mass Spectrometry* **2002**, *219*, 245-251.

(59)   Wu, C. C.; MacCoss, M. J.; Howell, K. E.; Yates, J. R., III *Nature Biotechnology* **2003**, *21*, 532-538.

(60)   Washburn, M. P.; Ulaszek, R. R.; Yates, J. R., III *Analytical Chemistry* **2003**, *75*, 5054-5061.

1.6   Figures
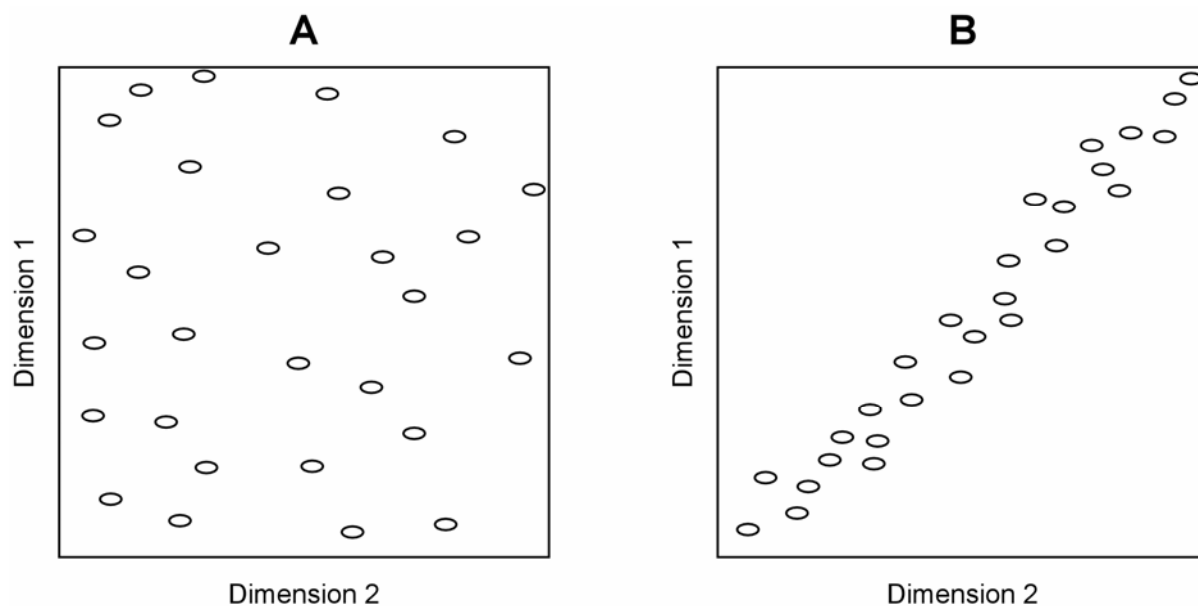
Figure 1-1: Hypothetical orthogonal (A) and non-orthogonal (B) two-dimensional separations. Note that, in the orthogonal separation, sample components are randomly distributed over the entire separation space, which maximizes peak capacity. The non-orthogonal separation shows a strong correlation between retention on the two dimensions, which results in a substantially diminished peak capacity.
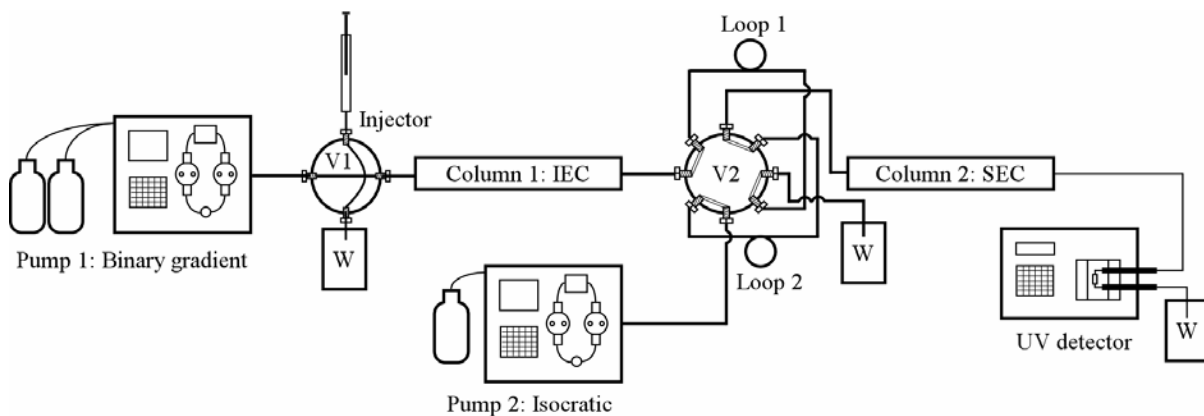
Figure 1-2: Schematic diagram of a 2D IEC-SEC instrument. Effluent from the ion exchange column is directed onto one of the two sample loops via the 8-port switching valve (V2). To separate the contents of the loop, valve 2 is switched, which sends the contents of the loop onto the size exclusion column. Meanwhile, the other storage loop fills with effluent from the ion exchange separation. Once the sample is fully separated on the size exclusion volume, V2 is switched and the process begins again. (Adapted from Ref. [38])
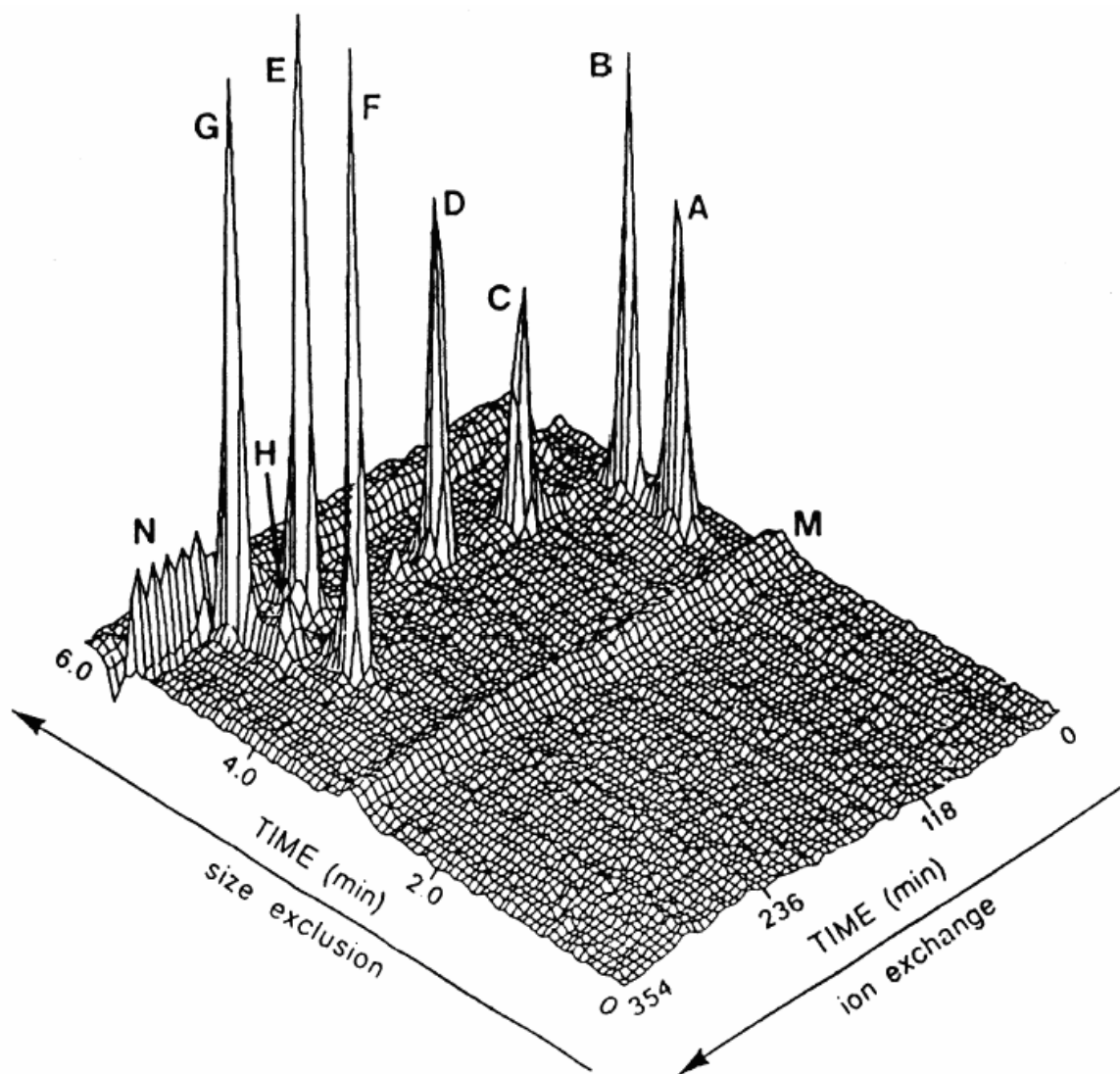
Figure 1-3: 2D ion exchange-size exclusion chromatogram of protein sample. Peak identities: (A) glucose oxidase, (B) ovalbumin, (C) β-lactoglobulin, (D) trypsinogen, (E) α-lactoglobulin, (F) conalbumin, (G) ribonuclease A, (H) hemoglobin; (M) exclusion volume "pressure" ridge, (N) inclusion volume "salt" ridge. Valve actuated every 6 minutes. UV detection at 215 nm (Reproduced with permission from Ref. [38])
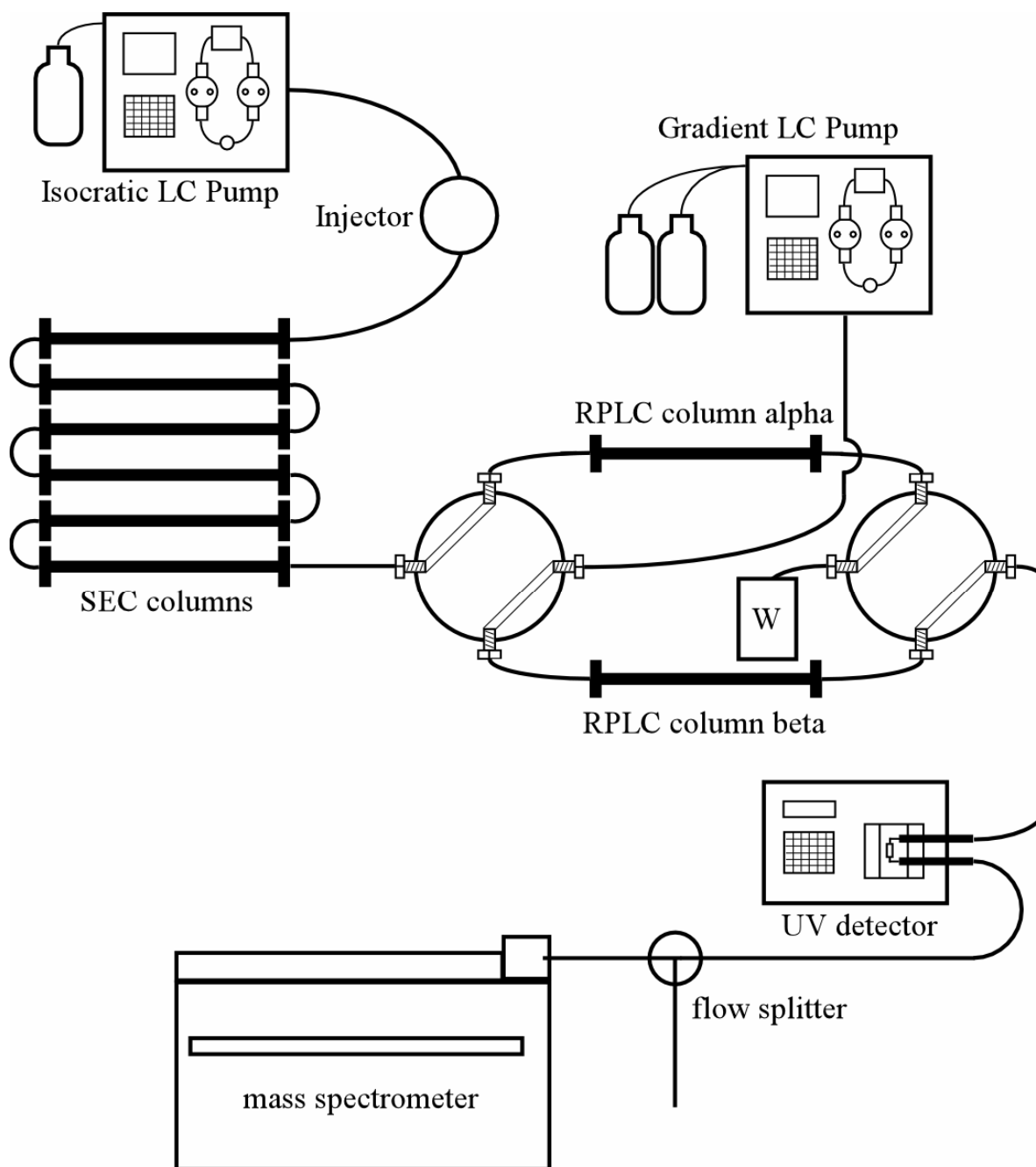
Figure 1-4: Schematic diagram of a 2D SEC-RPLC instrument with on-line mass spectrometric detection. The two four-port valves are switched simultaneously, directing the effluent of the size exclusion column to one of two reversed-phase columns while the other column is eluted using a gradient LC pump. After elution is complete, the valves switch again and the process repeats. (Adapted from Ref. [47])
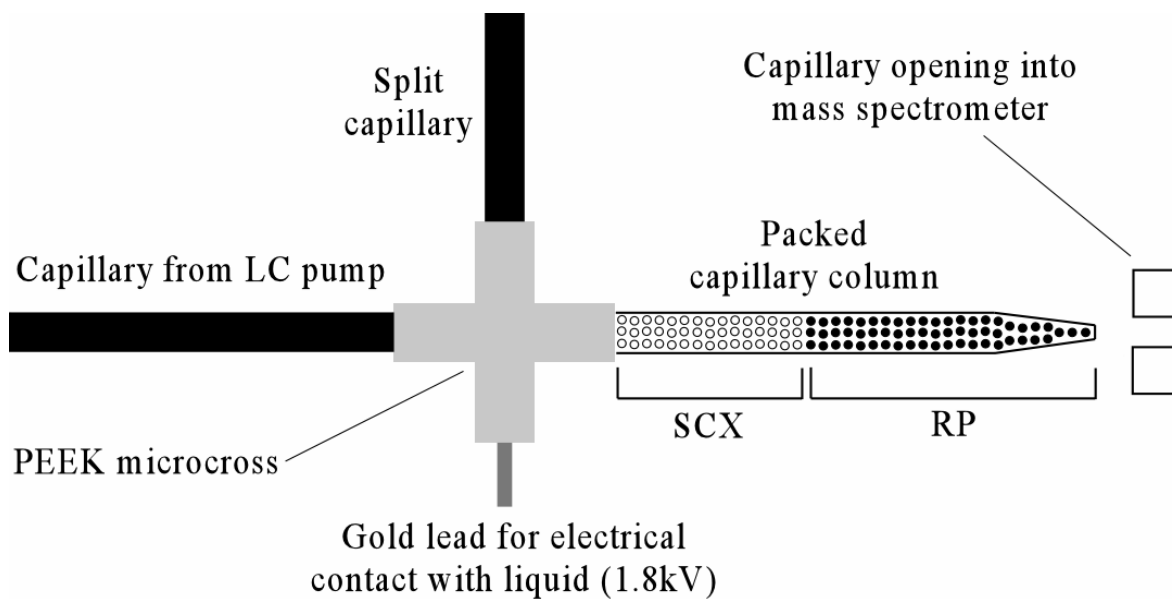
Figure 1-5: Diagram of a MudPIT column and electrospray interface. The micro-cross serves as a flow splitter in conjunction with a split capillary, and as the location at which the electrospray voltage is described. Further explanation can be found in the text.   (Adapted from Ref. [56]).
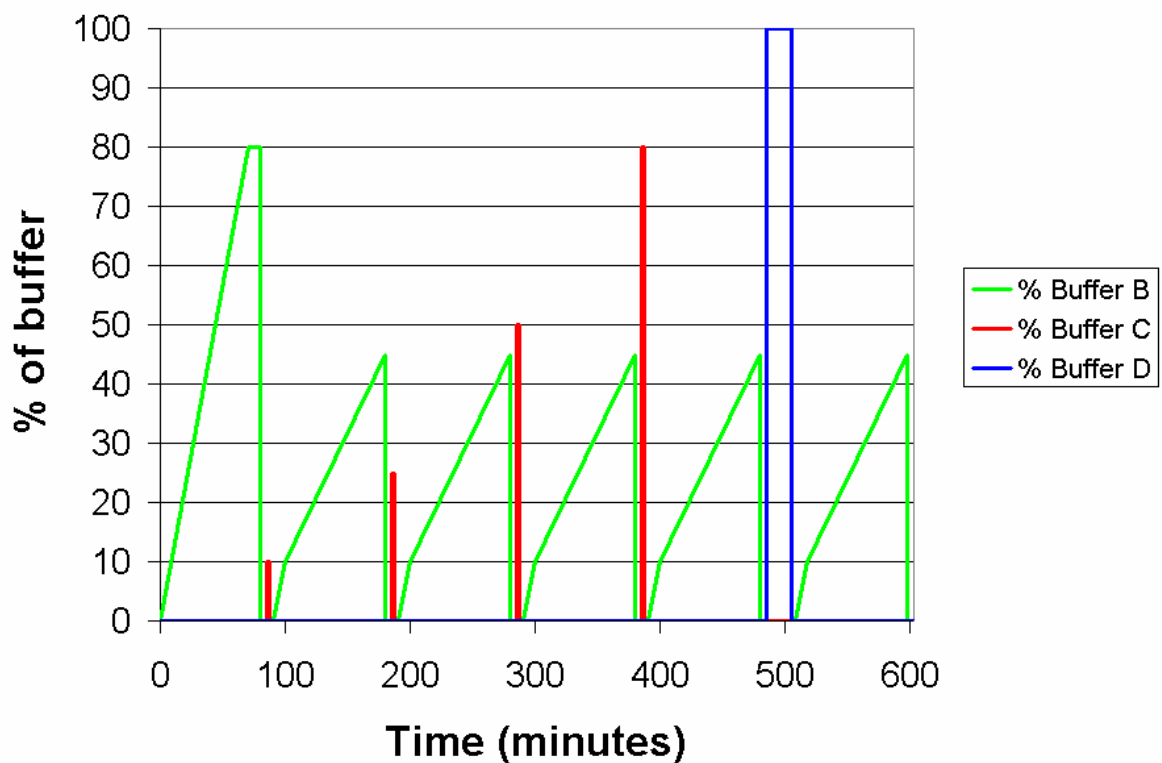
Figure 1-6: An example 5-step MudPIT gradient profile. The balance of the mobile phase at any time is made up of buffer A. Buffer A is 5% acetonitrile (ACN), 95% H_2O and 0.02% heptafluorobutyric acid (HFBA). Buffer B is 80% ACN, 20% H_2O and 0.02% HFBA. Buffer C is 250 mM ammonium acetate/5% ACN, 95% H_2O and 0.02% HFBA. Buffer D is 400 mM ammonium acetate/5% ACN, 95% H_2O and 0.02% HFBA. A further description of the parameters used to perform a MudPIT experiment can be found in the text (Adapted from Ref. [55]

CHAPTER 2: Size exclusion and ion exchange chromatography of intact proteins

2.1   Introduction

The ability to separate mixtures containing proteins is crucial to a wide variety of

chemical and biological research, including the study of cellular pathways, biomarker

discovery, and development of biopharmaceutical agents. Virtually the entire complement of

available separation techniques is routinely applied to proteins, including dialysis,

ultracentrifugation, isoelectric focusing, electrophoresis, and liquid chromatography. Each of

these methods has distinct advantages and disadvantages. For instance, while dialysis offers

an effective and inexpensive means of removing low or high molecular weight impurities

from a solution containing a protein of interest, this technique cannot separate many proteins

in a highly complex mixture. Likewise, capillary electrophoresis can offer fast, high-

resolution separations, but would be impractical for preparative-scale protein purification.

Among the strategies available for protein separations, liquid chromatography is

relatively unique in that it can be used to separate a sample based on many different chemical

properties, such as size, charge or hydrophobicity. Thus, a wide range of selectivity can be

achieved using the same fundamental technique: a sample is injected onto a column, and

separation occurs as analyte molecules differentially partition between a liquid mobile phase

and a stationary phase. The fact that the method remains consistent regardless of the

separation mechanism is important because it enables the user to choose a separation strategy

tailored to the task at hand simply by switching from one column to another, without having

to invest in entirely new instrumentation when the task changes. Additionally, the relative similarity of most LC separations makes it possible to couple one LC technique to another to perform a multidimensional separation. Although multidimensional LC will be the primary focus of later chapters in this dissertation, proper selection and optimization of each individual dimension is essential in order to make a multidimensional separation successful. Thus, this chapter describes research performed in order to select and optimize 1D-LC separation methods for intact proteins, such that they may be incorporated into a 2D separation.

2.1.1 Overview of LC methods available for intact protein separations

A wide variety of LC separation modes can be used to separate mixtures of proteins, including size-exclusion chromatography, ion exchange chromatography, hydrophobic interaction chromatography, hydrophilic interaction chromatography, and reversed-phase chromatography.[1] Specialized methods such as affinity chromatography are also available for isolation of a particular protein or class of proteins, though these techniques are not applicable to a general separation of a complex mixture containing unknown components. Given the range of options which are available, the chromatographer must carefully consider the nature of the sample to be analyzed and the characteristics of the available techniques when selecting a separation mode.

The easiest technique to implement for protein separations is size exclusion chromatography (SEC). This method separates proteins based on their size, using a mechanism relatively unique compared to most forms of chromatography. A size exclusion column consists of a packed bed of porous particles. The diameter of the pores is chosen such that it is of the same order as the hydrodynamic radius of the proteins being separated (10-100 nm is typical). Proteins or other molecules much smaller than the pore diameter are

31

easily able to enter and occupy the space within the pores. Proteins with diameters larger than the pores are completely excluded from entering, whereas proteins close to the size of the pores are partially excluded. To perform a chromatographic run, a sample is injected and analyte molecules are carried through the column by flow of a liquid mobile phase. Proteins too large to enter the pores flow directly through the column and thus elute first, whereas smaller molecules spend a portion of their time within the stagnant mobile phase inside the pores, and thus elute later. The result is a chromatogram with peaks arranged in order of size. SEC is carried out under isocratic conditions – the mobile phase composition remains the same throughout the run – which simplifies optimization of the separation but also limits its flexibility. While effective, SEC is generally regarded as providing low resolution. A well-optimized separation might have a peak capacity of 14 or less.[2]

Ion exchange chromatography (IEC) separates proteins based on their charge. The stationary phase consists of charged groups immobilized on the surface of a support particle. The immobilized charged groups are always associated with ions of opposite charge, which are not covalently bonded but remain attached due to electrostatic attraction. When a sample containing ions is introduced into the column, the sample will displace the electrostatically bound ions already present and will bind to the stationary phase. If the ionic strength of the mobile phase is low, the sample will remain bound since few ions are present in solution to displace it. To carry out a separation, a gradient of increasing salt concentration is applied. This increases the concentration of counterions competing for the charged surface sites, which causes analyte molecules to be displaced by ions from the mobile phase.[1] Proteins with greater charge density tend to be retained longer on the column. IEC offers good flexibility in that a range of stationary phases are available. Both positively charged (anion

exchange) and negatively charged (cation exchange) stationary phases can be used, and the separation can be adjusted by changing the pH of the mobile phase. In general, ion exchange offers higher peak capacities than size exclusion chromatography.[2]

Hydrophobic interaction chromatography (HIC) and hydrophilic interaction chromatography (HILIC) separate proteins based on their hydrophobicity and polarity, respectively. HIC employs a gradient from high to low salt concentration. The high salt concentration used at the beginning of the run causes proteins to bind to a mildly hydrophobic stationary phase – effectively, they are "salted out" of the mobile phase. As the salt concentration in the mobile phase is gradually decreased, proteins begin to enter the mobile phase and separation occurs. HILIC is analogous to normal phase liquid chromatography in that it uses a gradient of increasing solvent polarity, and more polar molecules are retained longer by the stationary phase. The only significant difference from normal-phase LC is that HILIC uses some water in the mobile phase as opposed to entirely organic solvents, which makes it feasible for protein separations.[3]

Reversed-phase liquid chromatography (RPLC) separates proteins based on their hydrophobicity. It uses a gradient of increasing organic solvent concentration as opposed to a salt gradient. At the start of a run, proteins are retained by interaction with a hydrophobic stationary phase, which might consist of C18, C8, C4 or phenyl groups immobilized on a bed of packed particles. As the concentration of organic solvent in the mobile phase increases, it becomes more favorable for the protein to partition into the mobile phase. Separation is achieved because more hydrophobic proteins are retained longer than less hydrophobic ones. Unfortunately, RPLC does have some disadvantages in terms of reproducibility and sample carryover from one run to the next when used for intact protein separations.[4-7] Nevertheless,

if these difficulties can be managed, RPLC offers high peak capacities and relatively short run times.

2.1.2 Considerations in selection of separation modes to be used for MDLC

When a multidimensional separation is to be performed, additional criteria become important in selecting and optimizing each LC dimension. First, one must ensure that the two dimensions produce orthogonal separations. For example, coupling of two reversed-phase separations carried out under identical conditions would offer no benefit over a 1D-LC separation, as discussed previously (section 1.2.1). Coupling of ion-exchange with reversed-phase LC would be a better choice due to the difference in the selectivity of the separations. Alternatively, the selectivity of one of the dimensions could be changed, for example, by altering the pH of the mobile phase in one reversed-phase separation. If the two dimensions are to be coupled in an off-line manner, by collecting fractions after the first dimension and then analyzing those fractions on the second dimension, then this may be the extent of the criteria that must be considered. Each separation method can be run under whatever conditions give the best performance within the desired run-time, and steps such as solvent evaporation or desalting can be carried out on the fractions that are collected if desired.

If an LC x LC separation is to be performed using on-line coupling between the dimensions, the criteria for selecting and optimizing the individual separations are more complex. First, the mobile phase from the first column must be miscible with that used for the second column, and any buffers or salts to be used must be soluble in both mobile phases. It is also desirable for the first dimension mobile phase to be a weak eluent on the second dimension. This is because it is preferable for all components in a fraction to concentrate at the head of the second column, to reduce the effects of band broadening and improve resolution. Finally, since it is necessary to sample the effluent from the first separation many

times using the second dimension, the second dimension separation must be run very rapidly compared to the first. This ensures that resolution gained on the first dimension is maintained in the second dimension, as stipulated by Giddings' second criterion (see section 1.2.1).

Certain combinations of separation modes satisfy these criteria better than others for online LC x LC of proteins. Solvent miscibility is seldom a problem, since mobile phases used for protein separations generally contain water. Buffer and salt solubility can be a concern, however, if high-salt techniques like HIC or ion exchange are coupled to techniques where substantial amounts of organic solvent are used in the mobile phase, like RPLC or HILIC. Eluent strength may also be an issue – if the effluent from an ion exchange column, which uses a low-to-high salt gradient, were directed onto a HIC column, which uses a high-to-low salt gradient, sample components would not be retained initially on the second column, and resolution would be diminished. The most significant challenge, however, is the difference in speed required between the two dimensions. The first dimension must typically be configured such that a single run lasts several hours. The second dimension runs must be much more rapid, typically lasting only for a period of a few minutes. These criteria dictate differences in the columns that should be used for each dimension. In the first dimension, resolution is optimized by using long columns, which may be packed with relatively large (5 μm or greater) support particles. The second dimension columns must be short to minimize dead time, and may benefit from the use of smaller (3 μm or less) support particles in order to allow high resolution separations at fast flow rates. The choice of separation modes is also important – reversed-phase LC is typically well-suited for use as the second dimension, since it can be run quickly and still provides relatively high peak capacity. Size exclusion and ion

exchange are more ideal for the first dimension, since commercially available columns using these separation modes suffer too much loss of resolution when used for short runs.

In the research presented in this chapter, size exclusion and ion exchange LC modes were evaluated for separation of intact proteins. Both commercially available and custom-packed columns were assessed in order to optimize their performance and to determine if and how they should be used as a first dimension of a multidimensional LC separation. As pertaining to the second separation dimension, the choice between separation modes was significantly more straightforward. Few other techniques have the advantages of reversed phase liquid chromatography for use as a second dimension in a LC x LC separation, such as its capability of fast run times and its ability to be directly coupled to electrospray mass spectrometry. These advantages have lead to its widespread use in multidimensional liquid chromatography – of the 13 publications on original LC x LC research cited in Chapter 1, only 2 used a separation mode other than RPLC in the second dimension. RPLC is not evaluated for 1D protein separations in this chapter; its use is described in the context of multidimensional separations presented later in this dissertation.

2.2   Size exclusion chromatography

As discussed previously, size exclusion chromatography is a good technique for protein separations since it is simple to implement and works well with most samples. It is also a good candidate for use as a first dimension of LC x LC. It can be run slowly without loss of resolution, and it uses aqueous mobile phases which are weak eluents for reversed-phase columns. Unfortunately, it is generally considered to be a low-resolution technique.[8] This preconception may not be due not to fundamental limitations of the method, but rather to the lackluster performance of standard columns. To assess this assumption, the performance of a commercially available size exclusion column was tested for the separation

of protein mixtures. Custom size exclusion columns were also prepared by packing fused silica capillaries with both commercial and custom-bonded porous particles. The performance of the commercial and custom size exclusion columns was compared in order to determine the best option for use in a 2D separation.

2.2.1 Experimental

2.2.1.1 Mobile phases and chemicals

Three different mobile phases were used for size exclusion separations. The first, a denaturing mobile phase, consists of 50 mM dibasic sodium phosphate (Fisher Scientific, Fair Lawn, NJ) and 3M urea (Fisher) in deionized water, adjusted to pH 7 with phosphoric acid (Fisher). The second, a non-denaturing mobile phase, consists of 50 mM dibasic sodium phosphate (Fisher) and 100 mM sodium sulfate (Sigma-Aldrich, St. Louis, MO) in deionized water, adjusted to pH 7 with phosphoric acid. The third, a non-denaturing mobile phase with a volatile buffer, consists of 400 mM ammonium acetate (Sigma-Aldrich) in deionized water, adjusted to pH 7 using ammonium hydroxide (Fisher) or acetic acid (Fisher). Deionized water was purified using a Barnstead Nanopure water system (Boston, MA). The reagents used for preparation of custom-bonded size exclusion particles were hydrochloric acid (Fisher), 3-glycidoxypropyltrimethoxysilane (Gelest, Morrisville, PA), acetone (Fisher), and diethyl ether (Fisher). All chemicals were ACS grade or higher and were used as received.

2.2.1.2 Protein standards and *E. coli* protein extract preparation

Samples used for size exclusion separations included mixtures of standard proteins and proteins extracted from the bacterium *E. coli*. Standard proteins used were thyroglobulin (MW 660 kDa), immunoglobulin G (MW 160 kDa), bovine serum albumin (MW 66 kDa), ovalbumin (MW 43 kDa) and myoglobin (MW 17 kDa). Typical concentrations injected were 0.1-0.2 mg/mL per protein, in a solvent of the mobile phase being used to carry out the

separation. The peptides leucine enkephalin (MW 556 Da) and tetraglycine (MW 228) were used as low MW standards at a typical concentration of 0.05 mg/mL. All proteins and peptides were purchased from Sigma-Aldrich and were used as received.

To obtain a more complex sample representative of the proteome of an organism, soluble proteins were extracted from a culture of the bacterium *E. coli* by Eric Hamlett in the Giddings research group in the department of Microbiology and Immunology at the University of North Carolina – Chapel Hill. The procedure he used is described below.

A single colony-forming unit of *Escherichia coli* K12 (strain MG1655 wild type) was isolated from a Luria-Bertani (LB) broth agar plate culture. The colony was aseptically transferred into 50% glycerol as a stock culture and stored at -80ºC until needed. An inoculation loop from the glycerol stock was transferred to 3mL of autoclaved LB broth and grown for 18 hours at 37ºC at a rotation of 250 rpm. 100 μL of this primary growth was transferred to 1.0 L of autoclaved LB broth and grown for 18 hours at 37ºC at a rotation of 250 rpm, after which point the growth had reached stationary phase – the point at which the cells had saturated the growth medium. The LB broth media was centrifuged at 5000 rpm for 20 minutes in order to pellet the cells, after which point the supernatant was poured off. To cleanse the cells of LB broth, the cell pellet was re-suspended in 20 mL of 0.1M phosphate buffer-saline (PBS) and then centrifuged at 5000 rpm for 20 minutes; this wash step was then repeated. The cleansed cell pellet was then re-suspended in 20 mL of lysis buffer. The lysis cell suspension was sonicated using a Fisher Dismembrenator model 300, using several bursts at 30% power. The suspension slurry was then centrifuged at 10,000 rpm for 20 minutes to re-pellet the cell debris. The golden-colored supernatant was transferred to a sterile ultracentrifuge tube. Non-solubilized proteins were pelleted (an oily droplet) by

centrifuging at 35,000 rpm for 45 minutes. Supernatant was transferred to a sterile tube. Benzonase, a nuclease, was added to the supernatant to digest chromosomal DNA. The digestion was performed on ice for a period of 20 minutes using 8 units of benzonase per mL of supernatant. The nuclease-treated extract was then concentrated and low MW components were partially removed using UltraFree centrifuge filters with a 5000 Da MW cutoff. 20 mL of supernatant was concentrated to 1.2 mL by centrifuging through the filters at 2000 rpm. This solution was then ready for analysis using size-exclusion LC.

2.2.1.3 Particle bonding

To prepare custom size exclusion columns, silica particles were chemically bonded with an inert surface coating containing diol functional groups to prevent chemical interactions with proteins. The starting material was 2 μm porous silica particles with 145 Å pores, provided by Waters Corporation (Milford, MA). To perform the chemical modification, the procedure set forth by Regnier et al.,[9] adapted by the recommendations of Porsch,[10] was followed. First, the particles were re-hydroxylated by refluxing 3 g of particles in 200 mL of 1.5 N hydrochloric acid in a round-bottomed flask for several hours. The slurry was allowed to cool, and the particles were washed with deionized water until the pH of the slurry was neutral. The particles were then boiled in deionized water for 30 minutes, after which they were again washed with deionized water. They were then dried for several hours in a vacuum oven at room temperature.

After drying, the re-hydroxylated particles were ready to be chemically modified using 3-glycidoxypropyltrimethoxysilane (GPTMS). A simplified diagram of the chemical bonding procedure is shown in Figure 2-1. 45 mL of a 10% (v/v) solution of GPTMS in water was prepared by adding the silane drop-wise to water. The pH of the solution was monitored using a pH meter while the silane was added, and was maintained at 8.5 by adding

drops of 10-3 to 10-1M solutions of aqueous potassium hydroxide as needed. The GPTMS solution was added to a round bottomed flask containing the re-hydroxylated particles, and the flask was sealed and swirled to mix its contents. The particles were left in the sealed flask for 24 hours at room temperature to allow the bonding reaction to occur. After the one-day reaction period, the particles were collected over a 0.2 μm filter, and were washed with deionized water. To open the epoxide ring of the oxirane now bonded to the particle surface, the particles were transferred to a flask and 50 mL of 10-3M hydrochloric acid were added. The flask was swirled periodically to mix the slurry. After one hour, the particles were collected by filtration and washed with deionized water, acetone, and diethyl ether. The particles were dried overnight on a watchglass at room temperature, then in a convection oven at 85ºC for 3 hours. The particles were then ready to be packed into capillary columns and used for size exclusion LC.

2.2.1.4 Column selection and column packing

To assess the capabilities of a commercially available size exclusion column for protein separations, a column was purchased from Tosoh Bioscience LLC (Montgomeryville, PA). The column, model number G2000SW$_{XL}$, has an inner diameter of 7.8 mm, a length of 30 cm, and is packed with 5 μm diameter silica particles with 125 Å diameter pores. Like the custom-bonded particles, these particles are derivatized with ligands containing diol functional groups. The manufacturer specifies that the column is useful for separations of globular proteins ranging in size from 5 kDa – 150 kDa. As an intermediate option between the commercial size exclusion column and custom-bonded particles, size exclusion packing material was also purchased from Tosoh Bioscience. The particles obtained are identical to those in the G2000SW$_{XL}$ column, except their pore diameter is 250 Å instead of 125 Å. This changes the useful size range for protein separations to 10-500 kDa.

Packing of fused-silica capillaries (Polymicro Technologies, Phoenix, AZ) with SEC particles was carried out using a high pressure slurry method, which has been described in detail elsewhere.[11] In short, the capillary to be packed is cut to the desired final column length plus approximately 40 cm. An outlet frit is formed by tapping one end of the capillary in a vial containing non-porous silica particles, then placing that end of the column into an electrical arc in order to sinter the particles and fix them in place. The open end of the capillary is placed into a high-pressure fitting, which is tightened into a packing reservoir containing the packing slurry. The slurry is prepared by adding approximately 20 mg of particles to 4 mL of deionized water, then placing the slurry in an ultrasonic bath for 20 minutes to fully suspend the particles. Packing is initiated by applying approximately 1000 psi of liquid pressure to the reservoir using a pneumatic amplifier pump (Haskel International, Burbank, CA). As the bed of packed particles in the capillary lengthens, the pressure is increased in order to keep the rate of packing approximately constant. For 5 μm particles, typically no more than 6000 psi of liquid pressure is required to pack a column to a length of approximately 50 cm within about one hour. Somewhat longer packing times and higher pressures were needed to pack the 2 μm particles.

Once the column has packed to the desired length, the pressure is turned off and residual pressure is released gradually through a valve. The column is cut to remove any excess capillary from the inlet end, and is hung vertically, inlet end up, to allow the bed to dry overnight. After drying, an inlet frit is added in the same manner as the outlet frit. To create a transparent "window" in the capillary to enable on-column UV absorbance detection, a small section of the polyimide is burned away in the area a few centimeters from the outlet end of the capillary. This can be accomplished before packing the column by passing the

41

desired region of the capillary through an electrical arc, or after packing by dripping a small amount of fuming sulfuric acid over the desired area.

2.2.1.5 Instrumentation and run conditions

Runs were performed using a Waters 600E Multisolvent Delivery System, which is a low-pressure-mixing gradient HPLC pump capable of flow rates from 0.01 – 20 mL/min at a maximum backpressure of 6000 psi. The pump outlet is connected to a 6-port computer-controlled injector valve (VICI Valco Instruments, Houston, TX). For runs using the G2000SW$_{XL}$ column, the outlet of the injector valve is connected directly to the column, and the column outlet is connected to an Applied Biosystems Model 785A programmable absorbance detector (Foster City, CA). For most size exclusion runs on the G2000SW$_{XL}$ column, a flow rate of 0.5 mL/min and sample injection volumes of 5-50 μL were used. All runs were performed under isocratic conditions.

For runs using capillary columns, a pre-column dynamic split is used, as depicted in Figure 2-2. Typically, the LC pump is operated at a flow rate of 0.2 mL/min. The majority of the flow from the pump is diverted to waste through a splitter capillary, while a small portion enters the packed column. The length and diameter of the splitter capillary determines the backpressure which is generated, which in turn determines the flow rate through the packed column. If a splitter capillary were configured to generate 500 psi of backpressure, the flow rate through a typical SEC column with an inner diameter of 75 μm and a length of 50 cm packed with 5 μm particles would be approximately 50 nL/min. Thus the split ratio would be approximately 4000:1. Typical on-column injection volumes are 1-10 nL. On-column UV absorbance detection was performed through the packed bed using a Linear 200 UVIS variable wavelength detector (Linear Instruments, Reno, NV) equipped with a capillary flow cell.

2.2.2   Results and discussion

The motivation for performing 1D SEC separations was to develop a method which could be used as a first dimension in a 2D separation of intact proteins. To achieve this goal, several parameters needed to be assessed. One major question was whether better results could be obtained using a commercially available column or with a custom-packed column. Additionally, appropriate mobile phase and run conditions needed to be selected. Finally, the selected technique needed to be tested with a complex sample representative of the proteome of an organism.

The first step in evaluating SEC for protein separations was to compare the performance of a standard, commercially available size exclusion column to a column packed in-lab. At first, certain difficulties were experienced in packing size exclusion columns. For the 5 μm SEC particles from Tosoh Bioscience, it was noted that if the column was allowed to pack gradually, the rate of packing would decrease dramatically – even to the point of coming to a complete halt – over the course of a few hours. It was suspected that this phenomenon was due to clogging of the column with small particulate matter, which impedes flow and prevents packing from proceeding. This hypothesis was verified by observing SEM images of particles before and after being placed in the packing reservoir, shown in Figure 2-3. These images demonstrate that newly-prepared slurry contains only whole, spherical particles, whereas slurry removed from the packing reservoir after several hours has many small, irregular-shaped particle fragments. The explanation for this degradation is relatively straightforward: the packing reservoir contains a magnetic stir-bar, which constantly stirs the slurry to prevent particles from settling out of suspension. Evidently, the highly porous size exclusion particles are vulnerable to fragmentation from physical impact with the stir-bar. The difficulty is reduced by packing columns as quickly as

possible, typically in less than 1 hour, before substantial fragmentation has time to occur. Unfortunately, this limited the useful length of the columns which could be packed to approximately 50 cm. Interestingly, the problem of particle fragmentation was not as significant with the 2 μm custom-bonded particles, which seem to be mechanically stronger than the 5 μm particles.

Once well-packed size exclusion columns were obtained, their performance was compared to that of the commercial size exclusion column. Figure 2-4 shows chromatograms from two SEC separations of four protein standards plus a low MW peptide standard. The first separation was performed on the G2000SW$_{XL}$ column containing 5 μm particles with 125 Å pores. The second chromatogram was generated using a capillary column packed with the 5 μm SEC particles with 250 Å pores. Note that while elution order is identical, resolution of most components is slightly higher on the capillary column. This can be attributed to the fact that it is nearly twice the length of the commercial column (55 cm as opposed to 30 cm). Also, the larger pore size of the particles in the capillary column would be expected to more effectively separate larger proteins; this explains the greater resolution between thyroglobulin and BSA (and the BSA dimer peak) on the capillary column. Peak capacities are estimated as 7 for the commercial column and 10 for the capillary column. Thus, both columns offer relatively mediocre separating power, although within a typical range for size exclusion chromatography.[8]

Based on standard chromatographic theory, substantial gains in separation efficiency can be achieved by using particles of smaller diameter, which results in higher peak capacity for a given column length.[1] To assess whether such improvements could be attained for size exclusion separations, capillary columns were packed with 2 μm porous silica particles

modified in-house with GPTMS. Figure 2-5 shows separations performed on this type of

column compared with a column packed with Tosoh 5 μm SEC particles. The column

packed with 2 μm particles was slightly longer (55 cm as opposed to 49 cm) and the mobile

phase flow rate was slightly higher; otherwise, run conditions were identical. Visual

comparison of the chromatograms suggests that the 2 μm particles give somewhat sharper

peaks. This would suggest that improved resolution can be obtained by using smaller

diameter particles for size exclusion separations.

Although SEC columns packed with 2 μm particles did produce somewhat higher

resolution separations for proteins under 50 kDa than columns with 5 μm particles, they did

not perform well with larger proteins, such as IgG or thyroglobulin. When such proteins

were injected on the column, typically either no peak was seen or the resulting peak was very

broad and tailed. This suggests that the bonding procedure used to deactivate the surface of

the silica particles is inadequate to prevent undesirable interactions with large proteins. This

limitation ruled out the use of the custom-bonded size exclusion particles for intact protein

separations. Capillary columns packed with the 5 μm SEC particles from Tosoh Bioscience

gave better results; however, the low flow rates used in capillary LC would make coupling of

this separation method to another LC separation difficult. Therefore, we returned to the

conventional-diameter commercial size exclusion column as the most practical option for

intact protein separations.

The G2000SW$_{XL}$ column was evaluated for separation of an *E. coli* soluble protein

extract. Figure 2-6 shows a chromatogram for a separation of 1mg of total *E. coli* protein

which was generated with the SEC column using 400 mM ammonium acetate (pH 7) as the

mobile phase. The chromatogram suggests the presence of many co-eluting peaks, which

appear within a relatively narrow window of retention time. Clearly, a technique to be used as the first dimension of a 2D separation would not be expected to completely resolve a sample with this many components. However, the relatively broad peaks and limited window of time in which peaks were observed suggests that SEC may not provide adequate peak capacity to effectively separate the sample into as many distinct fractions as desired. One option to address this issue would be to link many commercial SEC columns in series to increase resolution, as has been demonstrated successfully in the past.[12, 13] However, doubling column length is expected to increase peak capacity by only 40%,[14] so substantial expense would be involved in order to obtain significant improvements in performance. Therefore, it seemed more practical to investigate other separation modes as possible first dimensions for a 2D-LC separation of proteins.

2.3   Ion exchange chromatography

Ion exchange is another logical candidate for use as a first dimension separation of intact proteins. Like SEC, it uses entirely aqueous eluents, so it is compatible with on-line coupling to RPLC as a second dimension. Unlike SEC, ion exchange separations have many parameters which must be selected and optimized. First, there is the choice between anion exchange and cation exchange, which dictates the charge of the proteins which can be retained and separated by the column. Next, one must select between a "weak" or "strong" ion exchanger. Strong ion exchangers retain their charge regardless of pH, whereas for weak exchangers the extent of the surface charge is affected by the pH of the mobile phase. Changes in pH also affect the charge of the proteins being analyzed, which influences their retention on the column. Thus, the choice of pH and selection of an appropriate buffer are crucial to optimizing ion exchange separations. Finally, since ion exchange of proteins is carried out using gradient elution – typically by increasing the concentration of salt in the

mobile phase over the course of a run – the gradient must be optimized in order to obtain the best separation.

2.3.1 Experimental

2.3.1.1 Mobile phases and sample preparation

Both anion exchange and cation exchange separations were performed using binary gradients. The gradient profiles and composition of the mobile phases were changed as needed between runs to obtain the best separations possible. Therefore, precise mobile phase compositions and gradient parameters are specified along with the relevant chromatogram or description of a run in the text of the chapter; a typical mobile phase composition is specified here. For anion exchange separations, mobile phase A consisted of 50 mM ammonium acetate (Sigma-Aldrich) in deionized water, adjusted to pH 9 with ammonium hydroxide (Fisher). Mobile phase B was 1M ammonium acetate, adjusted to pH 9 with ammonium hydroxide. For cation exchange separations, mobile phase A was 50 mM ammonium formate (Sigma-Aldrich), adjusted to pH 3.5 with formic acid (Sigma-Aldrich). Mobile phase B was 1M ammonium formate, adjusted to pH 3.5 with formic acid. The *E. coli* protein extract sample used to evaluate the anion and cation exchange columns was the same as that used for evaluation of size exclusion columns; its preparation is described in section 2.2.1.2.

2.3.1.2 Column selection

Due to the wide range of columns available for ion exchange chromatography, selection was performed based on several criteria. Since it was unknown whether anion exchange or cation exchange would produce better separations of the *E. coli* proteome, both modes were investigated in order to compare their performance. Strong ion exchangers were selected, since it was deemed preferable to be able to perform separations over a wide range

of pH values without altering the charge of the stationary phase. Since working above pH 7 is typical of many ion exchange separations, the support particles needed to be able to withstand high pH. This ruled out most columns packed with silica particles. Fortunately, many commercially available ion exchangers are based on polymeric support particles and can be used even at the extremes of the pH scale. A final criterion was that the column needed to be usable as the first dimension in a 2D separation. Since off-line coupling of dimensions was to be employed, this stipulated a column with an inner diameter which would allow operation at a flow rate between 0.1-1 mL/min so that fractions of a reasonable volume could be collected.

Two columns which satisfied the above criteria – one anion exchange and one cation exchange – were obtained from Waters Corporation. The anion exchange column is the Biosuite Q 10 μm AXC. It is a stainless steel column 7.5 cm in length with an inner diameter of 7.5 mm. It is packed with 10 μm methacrylate-based polymeric particles with 1000 Å pores. The particles are bonded with quaternary amine functional groups, which give the surface a positive charge. The cation exchange column is the Waters Biosuite SP 10μm CXC. Its dimensions and support particles are the same as the AXC column, except the stationary phase is coated with sulfopropyl functional groups, which carry a negative charge.

2.3.1.3 Column packing

In addition to the commercial columns described above, a custom-packed anion exchange column was also prepared. The particles to be packed were identical to those in the Biosuite Q 10 μm AXC column, except they were 13 μm in diameter as opposed to 10 μm. These particles were obtained by emptying the particles from a semi-preparative scale column, the Waters Biosuite Q 13 μm AXC (15 cm x 21.5 mm ID). This was performed by removing one endfitting from the column, then pumping deionized water into the other end

of the column while it was held with its open end facing down above a clean glass beaker. Once all particles had been extruded from the column, they were transferred to a glass vial, which was sealed and stored at 4ºC until needed.

A slurry reservoir was designed and constructed to contain the particles and allow them to be packed into analytical LC columns with a conventional inner diameter (2-10 mm). This reservoir is pictured in Figure 2-7. The slurry reservoir was machined from clear polycarbonate. Its internal volume is approximately 150 mL. The seal between the lid and base is made with a Viton o-ring, which fits into a groove machined into the lid. The bottom portion of the internal chamber is tapered, which prevents particles from settling on a horizontal surface, and directs the slurry toward the outlet fitting at the base of the reservoir. The reservoir is pressurized using an HPLC pump through a connection for 1/16" tubing located in the lid. Although the maximum pressure of the reservoir has not been determined, it is typically not used above 200 psi due to the large internal volume and limitations in the strength of the polycarbonate. Also, due to limited chemical resistance, most organic solvents cannot be used with the reservoir.

Empty glass columns purchased from Omnifit (Cambridge, UK) were packed with the anion exchange particles. The columns have an inner diameter of 6.6 mm and are 40 cm long. One end has an adjustable endfitting, which allows the column to be used with a bed length from 32 - 40 cm. The maximum operational pressure is 900 psi. For packing, the outlet endfitting was fitted with a 10 μm polyethylene frit and the inlet frit was removed to allow the particles to enter the column. 20% (v/v) slurry of the anion exchange particles was prepared and added to the packing reservoir. The slurry solvent was 1M ammonium acetate (Sigma-Aldrich), adjusted to pH 8.5 with ammonium hydroxide (Fisher). It is important to

use a relatively high salt concentration when packing polymeric ion exchange particles, since the particles swell when the ionic strength of the buffer is low, which can cause gaps to form in the packing when the column is used.[1]

The reservoir was connected to the inlet of the column using 1/16" ID Teflon tubing. Packing was initiated by applying flow to the top of the reservoir at a rate of 5 mL/min using a Waters 600 HPLC pump. The backpressure was monitored as packing proceeded and flow was reduced as needed to keep the pressure below 200 psi (a safe limit for the pressure reservoir). The column packed to a length of 37 cm within a period of about 1 hour. Once packing was complete, flow was turned off and the column was allowed to depressurize. A frit was then added to the inlet endfitting, and its position was adjusted to remove any gap between the endfitting and the packed bed. A total of three Omnifit glass columns were packed using this procedure, each to a bed length of 37 cm. At this point, each column was tested chromatographically with standard proteins to ensure that it gave peaks which were symmetrical and as sharp as expected. Once all columns had been tested, they were connected in series using 0.005" ID PEEK tubing, which was positioned directly next to the frits to minimize void volume. A picture of the three packed anion exchange columns connected in series is shown in Figure 2-8. The connected columns are used as a single AXC column, with an effective bed length of 111 cm.

2.3.1.4 Instrumentation and run conditions

Runs were performed with the same HPLC pump, injector and detector as was used for the separations on the conventional-diameter SEC column, which are described in section 2.2.1.4. The flow rate used for the Biosuite columns was 0.5 mL/min. The custom-packed anion exchange column was operated at 0.39 mL/min in order to maintain the same linear mobile phase velocity, since its inner diameter is smaller (6.6 mm as opposed to 7.5 mm).

Injection volumes varied from 5-50 μL. All runs were carried out using binary gradient elution; specific gradient programs are specified along with the data in the results section.

2.3.1.5 Bradford protein quantification assay

The Bradford protein quantification assay[15] was used to measure the amount of protein in fractions collected from the ion exchange columns. The Bradford reagent was prepared as follows: 10 mg of Coomassie Brilliant Blue G-250 (Sigma) were dissolved in 5 mL of ethanol (Aaper, Shelbyville, KY) in a 100 mL volumetric flask; a deep blue solution is produced. 10 mL of 85% phosphoric acid (Fisher) were added to the flask and the solution was swirled to mix; the solution turned a red-brown color. The solution was diluted to mark with deionized water; the solution became a brown-green color. The solution was filtered using a 1 μm syringe filter into a dry flask. The filtrate, dark brown in color, is the ready-to-use Bradford reagent.

To use the Bradford reagent, the protocol for the microprotein assay[15] was followed. 1.5 mL of Bradford reagent was added to a centrifuge tube containing 0.1 mL of the buffer solution to be assayed for protein. The tube was mixed by vortexing. After an incubation period of 5 minutes, the absorbance of the solution was ready to be measured using a spectrophotometer. A Spectronic 401 (Milton Roy, Rochester, NY) was set to a wavelength of 595 nm and was zeroed using a blank solution prepared by adding 1.5mL of Bradford reagent to 0.1mL of buffer solution with no protein. The absorbance of the solution to be assayed was then measured. Standard curves were generated using the protein bovine immunoglobulin G over a concentration range of 5-250 μg/mL; separate curves were generated for the buffers used in anion exchange chromatography (50 mM ammonium acetate, pH adjusted to 9.0 using ammonium hydroxide) and cation exchange chromatography (50 mM ammonium formate, pH adjusted to 3.5 using formic acid).

The Bradford assay was used to determine the quantity of protein found in ion exchange separations of the *E. coli* protein extract. Anion and cation exchange separations of the *E. coli* protein extract were carried out as described in section 2.3.1.4. The effluent from the column was collected in fractions using microcentrifuge tubes, which were switched every 2 minutes. The fractions were then lyophilized in order to remove the mobile phase and allow the proteins to be concentrated. To perform lyophilization, after all fractions were collected, the tubes were capped and frozen by immersion in liquid nitrogen for approximately 1 minute. The tube caps were opened, and all tubes were placed in a vacuum centrifuge (SpeedVac Concentrator, Savant, Farmingdale, NY). The centrifuge was turned on and vacuum was applied. The tubes were left under vacuum until all mobile phase had been removed; typically, this took about 4-8 hours. After lyophilizing, the contents of the fractions were re-suspended in 0.1 mL of 50 mM buffer solution, which represents 10-fold concentration of the proteins as compared to before lyophilization. The fractions were then analyzed for protein using the Bradford reagent using the protocol described above.

2.3.2 Results and discussion

The motivation for switching from size exclusion to ion exchange chromatography was to obtain better resolution of complex protein mixtures, such as the *E. coli* protein extract sample, for use in a multidimensional separation. Therefore, experimental efforts were directed at optimizing separations of these samples, rather than first performing separations of standard proteins. The first attempt at separating the *E. coli* protein extract on the anion exchange column was performed using standard mobile phases recommended by the column manufacturer. A 20-minute gradient from 0 to 0.5M sodium chloride was used, with a buffer of 50 mM Tris-HCl (pH 8.5) present in the mobile phase throughout the run. Detection was performed using UV absorbance at a wavelength of 215 nm. A chromatogram

from this separation is shown in Figure 2-9. Peaks are observed to be spread over an 18-minute window of retention time; the estimated peak capacity of the separation was 20. Thus, even using a relatively short gradient and without extensive optimization, this approach demonstrated better separation capabilities than size exclusion chromatography.

Although the sodium chloride gradient produced a reasonable anion exchange separation, the use of a non-volatile salt posed difficulties in terms of the plan to collect and lyophilize fractions in order to remove the mobile phase prior to analysis by a second chromatographic dimension. The simplest solution was to use a mobile phase consisting of only components which are volatile under vacuum. Therefore, an anion exchange separation of the *E. coli* protein extract was performed using a gradient from 50 to 500 mM ammonium acetate, pH adjusted to 9.0 using ammonium hydroxide. Ammonium acetate, a volatile salt, serves as both the pH buffer and as the elution salt in this gradient. Likewise, a cation exchange separation was performed using ammonium formate, with the pH adjusted to 3.5 using formic acid. It was found that the background absorbance from ammonium acetate or formate was too strong to permit detection at 215 nm, so the wavelength was changed to 280 nm. At this wavelength, detection of proteins is based on absorbance of the side chains of the aromatic amino acids, as opposed to the peptide backbone.[16] Therefore, detection is not as universal nor is quantitatively equivalent for all proteins, although the traces still provide an idea of when proteins elute from the column.

Chromatograms of the ion exchange separations of the *E. coli* protein extract are shown in Figure 2-10. The anion exchange chromatogram appears to show good separation of the sample, with peaks spread relatively evenly over 30 minutes of separation space. Peaks in the chromatogram from the cation exchange separation, however, are much sparser.

A very large peak is observed to elute near the dead time of the column; further into the gradient only three, substantially smaller peaks are observed. Thus it seems that cation exchange chromatography, under the conditions used in this run, does not separate the *E. coli* protein extract effectively.

In order to attempt to improve the cation exchange separation of the *E. coli* proteins, four different buffer compositions with pH values ranging from 2.5 to 7.0 were investigated. No separations above neutral pH were attempted, since alkaline pH favors the conversion of ionizable molecules such as proteins to anions rather than cations. Chromatograms from these runs are shown in Figure 2-11. Although small peaks appear throughout the chromatograms, the majority of the sample seems to elute near the dead time of the column (approximately 4 minutes) in all of the runs. Decreasing the pH of the buffer did cause increased retention of some components in the sample. This is particularly notable with pH 2.5 citrate buffer, and to a lesser extent at pH 3.5 with the ammonium formate buffer. In these cases, there is some resolution between the components eluting near the dead time. Another substantial reduction in the pH might result in a better cation exchange separation, but this was not possible since the manufacturer of the column recommends against using it at a pH below 2.0.

The detection method which was used for these separations – UV absorbance at 280 nm – is not selective to proteins alone. Thus, a chromatogram which appears to have numerous peaks spread over a reasonable range of retention time is not a certain indication of a successful protein separation. Therefore, a quantitative assay of protein concentration in fractions collected from the ion exchange columns was performed. The assay was carried out using the Bradford method, which is based on the dye molecule Coomassie Brilliant Blue

G-250.[15] At acidic pH, the dye binds with proteins and undergoes a characteristic absorbance wavelength shift from 465 nm to 595 nm. The magnitude of this shift is proportional to the concentration of protein in solution, and can be measured using a visible-light spectrophotometer.

Standard curves for the Bradford protein assay were generated using bovine IgG and are shown in Figure 2-12. Separate curves were generated using protein solutions in buffers at different pH values, to match the buffers used in the ion exchange and cation exchange separations. The curves demonstrate that the response of the Bradford assay is generally linear over the range of protein concentrations assayed (5-250 µg/mL). The response was not notably different in one buffer compared to the other.

Results of assays performed on fractions collected from anion and cation exchange separations of the *E. coli* protein extract are shown in Figure 2-13. Bradford assay data is shown in terms of the raw absorbance value at 595 nm; these values are directly proportional to protein concentration. The "LB" and "HB" data shown in the figures are the results from control experiments performed to ensure that solutions containing no protein give little or no reaction with the Bradford assay. The LB (low salt blank) tube contained 100% mobile phase A, whereas HB (high salt blank) contained 100% mobile phase B. Both of these tubes were treated identically to the actual ion exchange fractions, including all lyophilization and reconstitution steps.   Although the LB and HB solutions showed some absorbance, the levels were low compared to fractions containing actual protein. This confirms that the Bradford assay does not experience substantial interference from residual salts from the ion exchange mobile phase.

The Bradford assay data are shown on the same graph as the UV absorbance chromatogram in order to assess the correlation between peaks that are observed and the presence of protein in the fractions. For the anion exchange separation, fractions 1-10 appear to contain significant quantities of protein. Interestingly, the UV chromatogram shows peaks in fractions 11-16 as well, while the Bradford assay detects no significant amount of protein in these fractions. This suggests that the later peaks in the UV chromatogram are non-protein components, perhaps small molecules with absorbance at 280 nm. Nonetheless, the assay does show that proteins are spread over a reasonable number of fractions, which validates the method as a means of separating mixtures of intact proteins. The results from the Bradford assay of the cation exchange fractions are less encouraging. No substantial absorbance at 595 nm was detected in any fraction, even though the same amount of protein was injected onto the column. It is suspected that some part of the cation exchange separation, fraction collection and lyophilization procedure interferes with the accuracy of the Bradford assay. Regardless, the assay did not provide convincing evidence that cation exchange chromatography separates the protein mixture effectively. Therefore it was determined that anion exchange chromatography was the best option for use as a first dimension in a multidimensional separation of intact proteins.

Once anion exchange was determined to be the preferred method, the separation of the *E. coli* protein extract was optimized further by lengthening the gradient. Figure 2-14A shows a separation of the *E. coli* protein extract performed using the Biosuite Q 10um AXC column. The 60 minute gradient described in the figure caption gives a peak capacity of approximately 35. This column was then used as the first dimension in 2D separations of intact proteins; the details of these 2D separations will be described in later chapters.

Although the Biosuite column was effective for 2D protein separations, ultimately, higher resolution separations than it could provide were desired. Thus, a custom-packed anion exchange column with 111 cm of total bed length was prepared and tested as described in section 2.3.1.3. The new column was used to perform separations of the same *E. coli* protein extract; an example of such a separation is shown in Figure 2-14B. Due to the increased length of the column, the gradient needed to be lengthened. A general rule is that, when attempting to improve resolution or peak capacity in an already-optimized separation by increasing column length, the quantity given by the expression:

$$\frac{(\text{gradient time} \times \text{flow rate})}{\text{column volume}} \qquad \text{(Equation 2-1)}$$

should be held constant.[17] In even simpler terms, if the linear mobile phase velocity is kept constant, then the gradient length should be increased directly proportional to the increase in column length. Since the linear flow rate was indeed the same on the short and long ion exchange columns, and the length of the long anion exchange column is 14.8 times greater than that of the short column, a 360 minute gradient on the long column is roughly equivalent to a 24 minute gradient on the short column. This is not to say that the resolution would be equivalent in these two hypothetical runs – clearly the long column will produce much higher resolution separations. Instead, it indicates that the long column can benefit from much longer run times than the short column. Just as increasing the gradient length on the short column from 30 minutes to 60 minutes improved peak capacity, lengthening the gradient on the long column from 360 to 720 minutes should also do the same, although this was not experimentally verified.

As expected, the result of the increased column and gradient length was substantially higher resolution separations. The peak capacity of the separation shown was estimated to be

85. As with the shorter column, the long anion exchange column was used as the first

dimension in several multidimensional protein separations, which are described in

subsequent chapters of this dissertation.

2.4   Conclusions

Size exclusion and ion exchange chromatography were evaluated as potential first

dimensions for a LC x LC separation of intact proteins. Although size exclusion

chromatography successfully separated mixtures of protein standards, its peak capacity was

not as high as desired and fractionation of a complex sample such as the *E. coli* protein digest

was poor. Capillary columns packed with commercial and custom-bonded size exclusion

columns did offer better resolution, but the magnitude of the improvement was not

substantial. Anion exchange chromatography produced significantly higher resolution

separations of the *E. coli* sample. Cation exchange chromatography, however, did not appear

to separate most of the proteins. The Bradford quantitative protein assay confirmed the

presence of proteins in many fractions from an anion exchange separation of the *E. coli*

sample, while proteins were not found in cation exchange fractions. Further improvement in

anion exchange peak capacity was attained by packing a longer column and increasing the

gradient time. Of the techniques investigated, it was concluded that anion exchange

chromatography is the best option for use as the first dimension in an LC x LC separation.

2.5 References

(1)   Neue, E. D. *HPLC Columns: Theory, Technology, and Practice*; Wiley-VCH: New York, 1997.

(2)   Gilar, M.; Olivova, P.; Daly, A. E.; Gebler, J. C. *Analytical Chemistry* **2005**, *77*, 6426-6434.

(3)   Linder, H.; Sarg, B.; Meraner, C.; Helliger, W. *Journal of Chromatography A* **1998**, *743*, 137-144.

(4)   Kastner, M. *Protein Liquid Chromatography*; Elsevier: Amsterdam, Netherlands, 2000.

(5)   Cohen, K. A.; Schellenberg, K.; Benedek, K.; Karger, B. L.; Grego, B.; Hearn, M. T. W. *Analytical Biochemistry* **1984**, *140*, 223-235.

(6)   Cohen, S. A. *Analytical Chemistry* **1984**, *56*, 217-221.

(7)   Cohen, S. A.; Benedek, K.; Tapuhi, Y.; Ford, J. C.; Karger, B. L. *Analytical Biochemistry* **1985**, *144*, 275-284.

(8)   Lecchi, P.; Gupte, A. R.; Perez, R. E.; Stockert, L. V.; Abramson, F. P. *Journal of Biochemical and Biophysical Methods* **2003**, *56*, 141-152.

(9)   Regnier, F. E.; Noel, R. *Journal of Chromatographic Science* **1976**, *14*, 316-320.

(10)   Porsch, B. *Journal of Chromatography A* **1993**, *653*, 1-7.

(11)   MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W. *Analytical Chemistry* **1997**, *69*, 983-989.

(12)   Opiteck, G. J.; Jorgenson, J. W.; Anderegg, R. J. *Analytical Chemistry* **1997**, *69*, 2283-2291.

(13)   Opiteck, G. J.; Ramirez, S. M.; Jorgenson, J. W.; Moseley, M. A., III *Analytical Biochemistry* **1998**, *258*, 349-361.

(14)   Gilar, M.; Daly, A. E.; Kele, M.; Neue, E. D.; Gebler, J. C. *Journal of Chromatography A* **2004**, *1061*, 183-192.

(15)   Bradford, M. M. *Analytical Biochemistry* **1976**, *72*, 248-254.

(16)   Moffatt, F.; Senkans, P.; Ricketts, D. *Journal of Chromatography A* **2000**, *891*, 235-242.

(17)   Snyder, L. R.; Dolan, L. J. *Journal of Chromatography A* **1998**, *799*, 21-34.

2.6   Figures

3-glycidoxypropyltrimethoxysilane

pH 8.5

10^{-3} M HCl
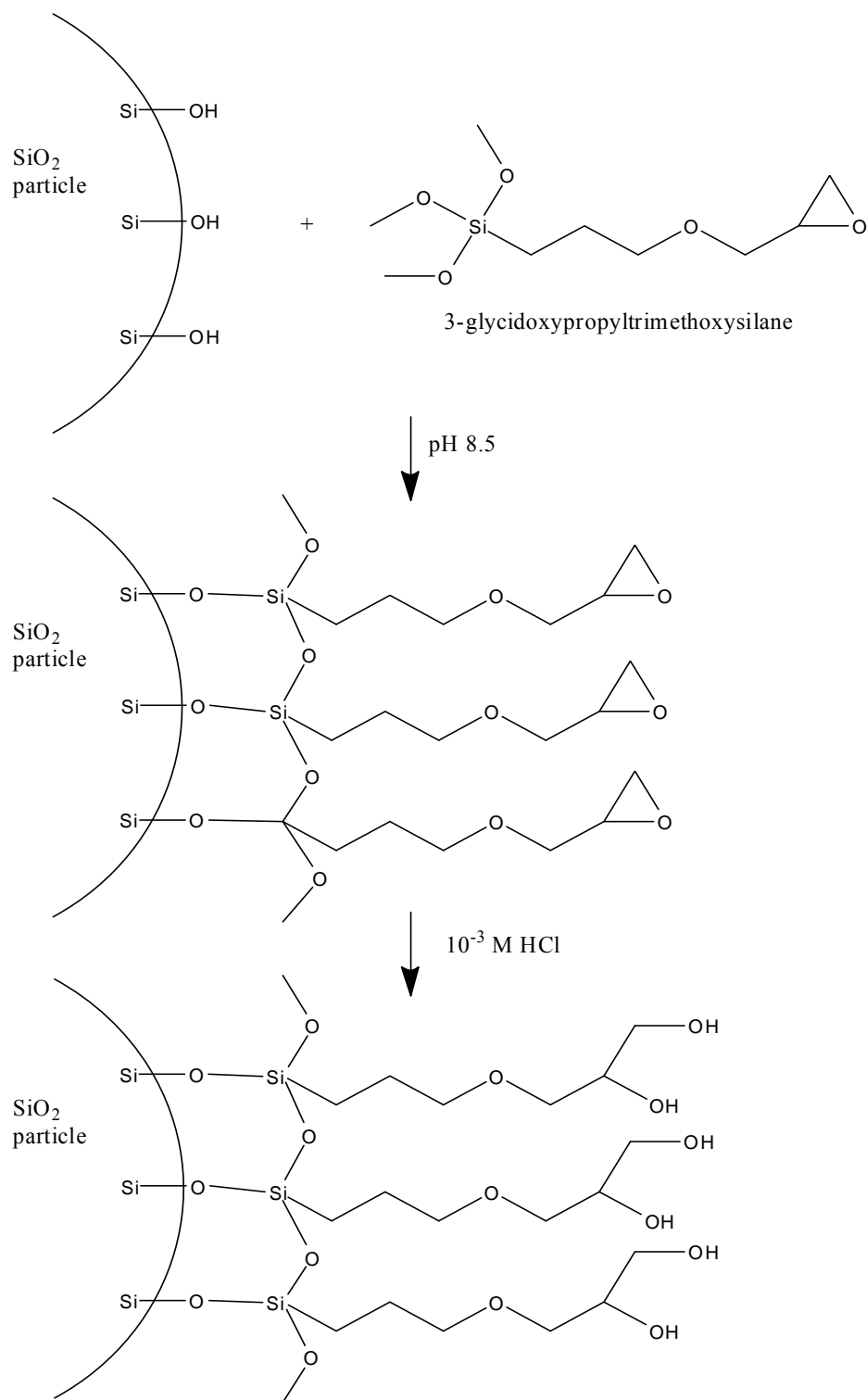
Figure 2-1: Scheme for surface modification of a silica particle with a diol phase using 3-glycidoxypropyltrimethoxysilane (GPTMS)
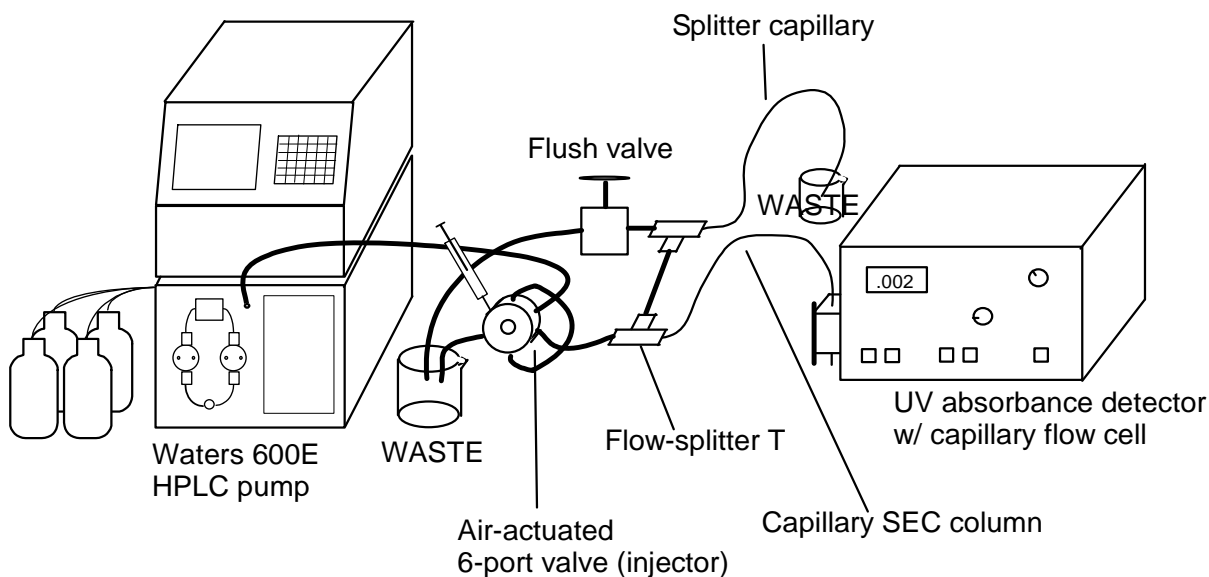
Figure 2-2: Diagram of instrumentation used to perform capillary LC separations. The capillary SEC column passes entirely through the flow-splitter tee and is sleeved by the PEEK tubing connected to the outlet of the 6-port injector, such that its inlet end is located as near as possible to (but not touching) the valve rotor. This ensures minimum delay volume between the HPLC pump / injector and the capillary column. A second tee connects the splitter capillary and a flush valve. When runs are being performed, the flush valve is closed so that flow is directed through the splitter capillary, which determines the backpressure at the column head. To flush the pump at high flow rates, as when changing mobile phase, the flush valve can be opened.
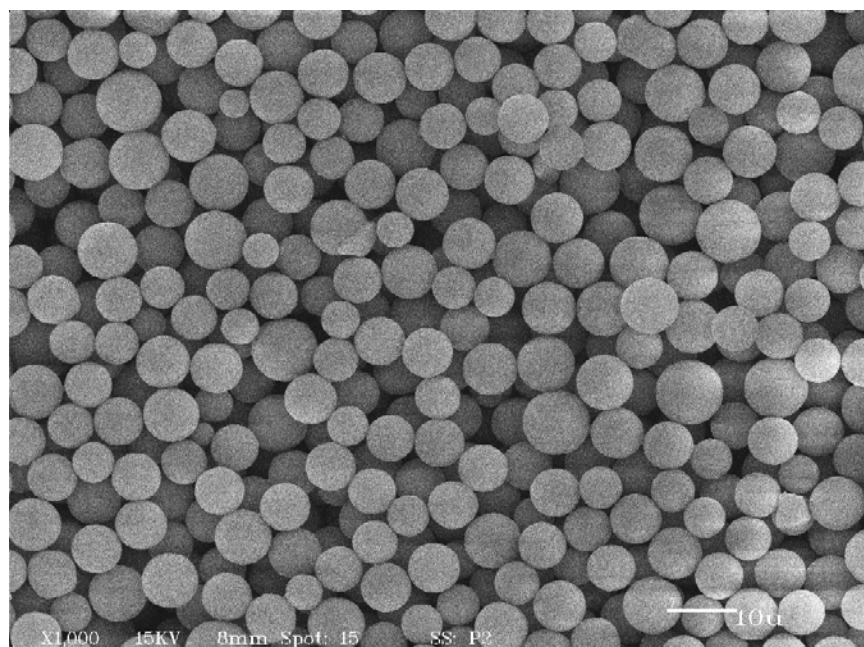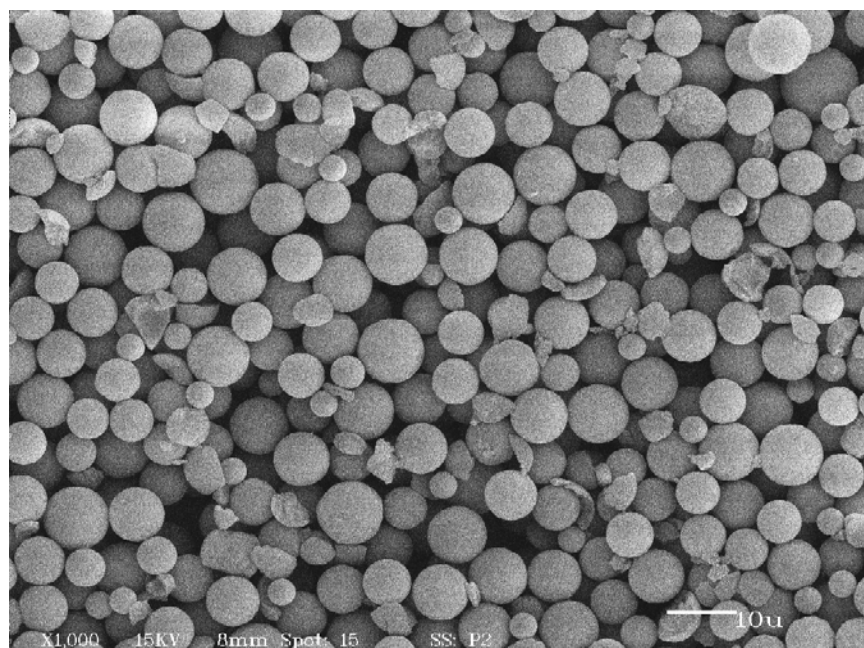
Figure 2-3: SEM images of 5 μm size exclusion particles from Tosoh Bioseparations, LLC. The top image shows particles as received; the bottom image shows particles retrieved from slurry which had been inside the stirred packing reservoir for several hours.
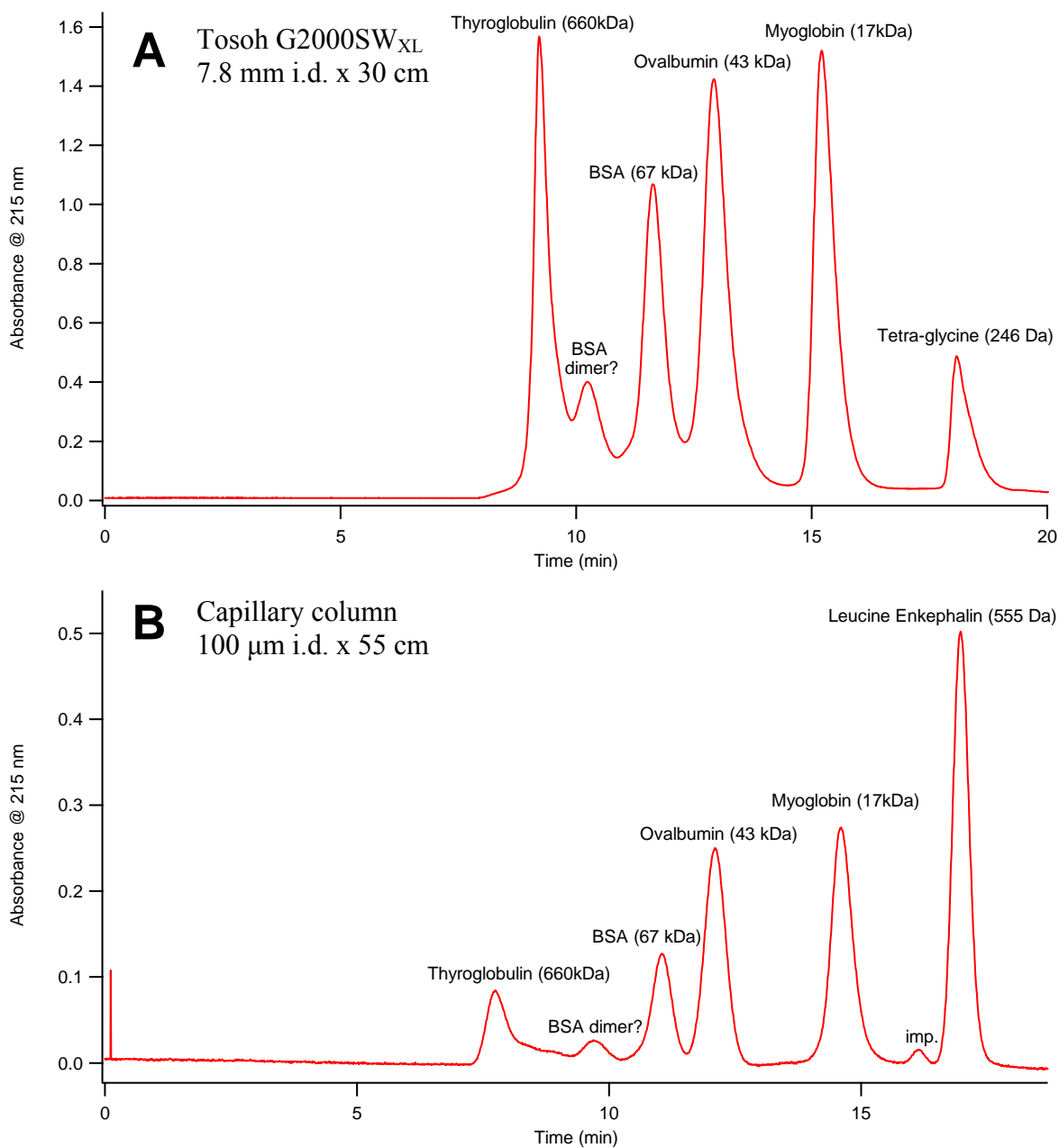
Figure 2-4: Size exclusion separations of protein standards. Chromatogram A was generated using a Tosoh Bioscience G2000SW$_{XL}$ column containing 5 μm particles with 125 Å pores. Chromatogram B was obtained using a capillary column packed with 5 μm particles with 250 Å pores. Note that while elution order is identical, resolution of most components is slightly higher on the capillary column.
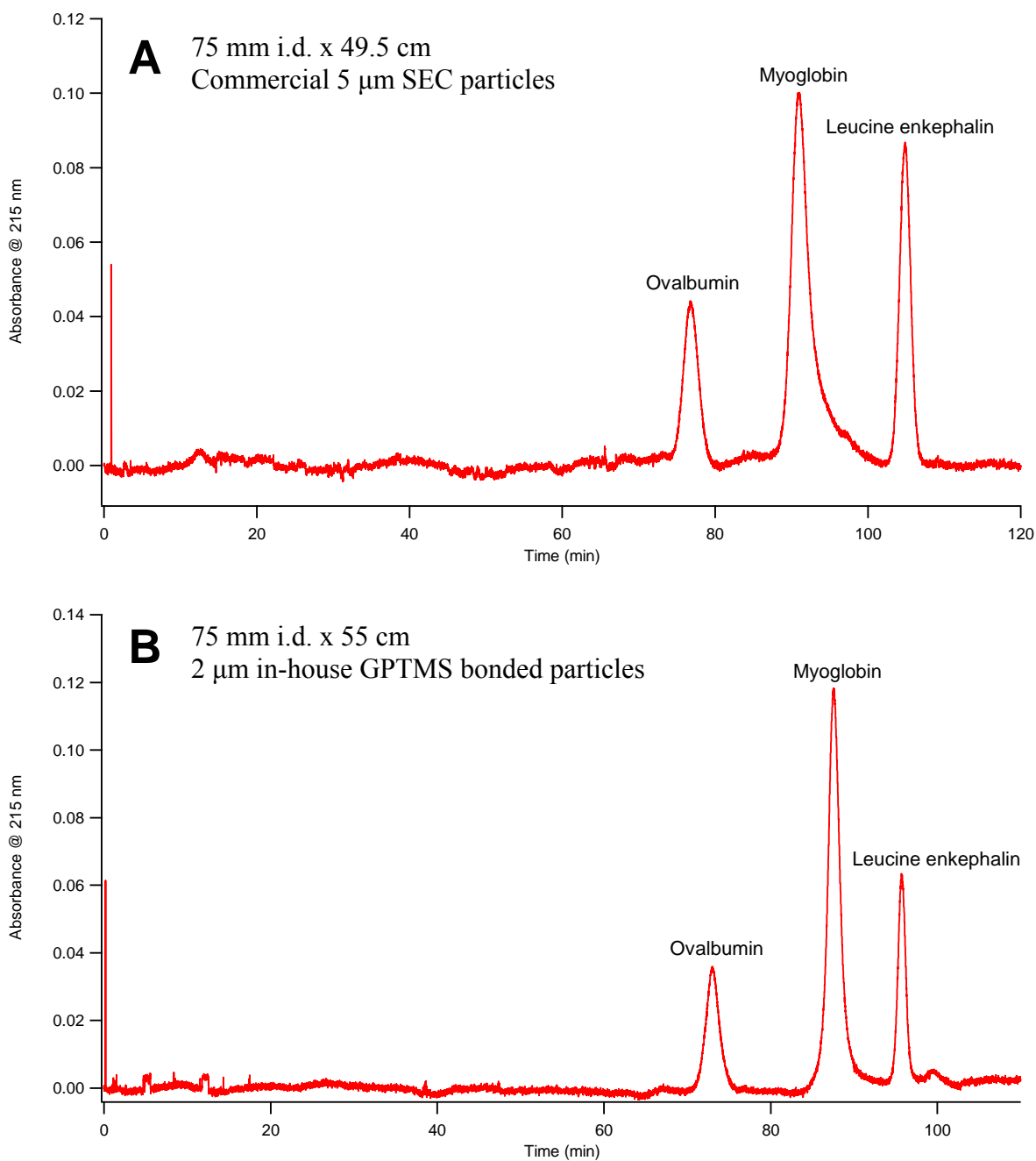
Figure 2-5: Comparison of performance of capillary columns packed with (A) commercial 5 μm SEC particles, and (B) 2 μm silica particles bonded in-house with GPTMS. Both separations were carried out using denaturing mobile phase (50 mM phosphate + 3M urea, pH 7)
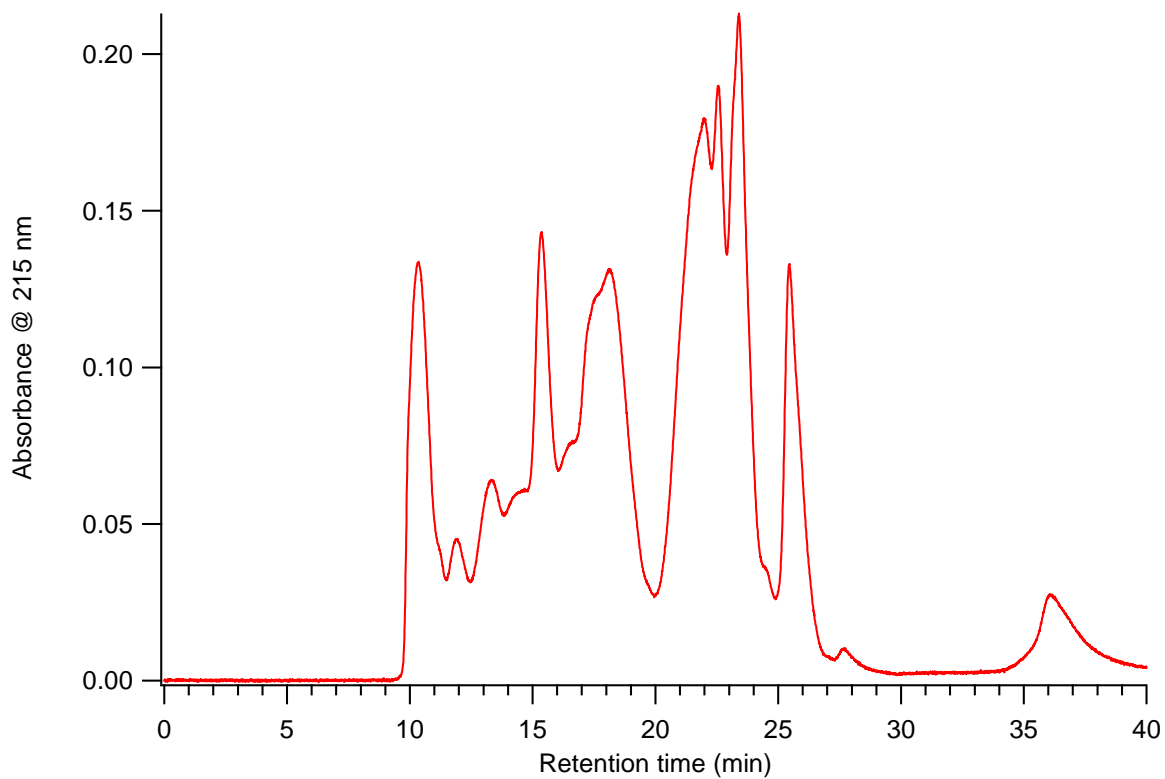
Figure 2-6: Size exclusion separation of the *E. coli* protein extract on the G2000SW$_{XL}$ column.

Figure 2-7: Polycarbonate slurry reservoir used to pack anion exchange columns. The green PEEK tubing at the top of the reservoir connects to a HPLC pump, which generates the flow used to pack the column. The metal valve at the bottom controls the flow of slurry into the column to be packed, which is connected using the tan PEEK fitting at the bottom of the image.
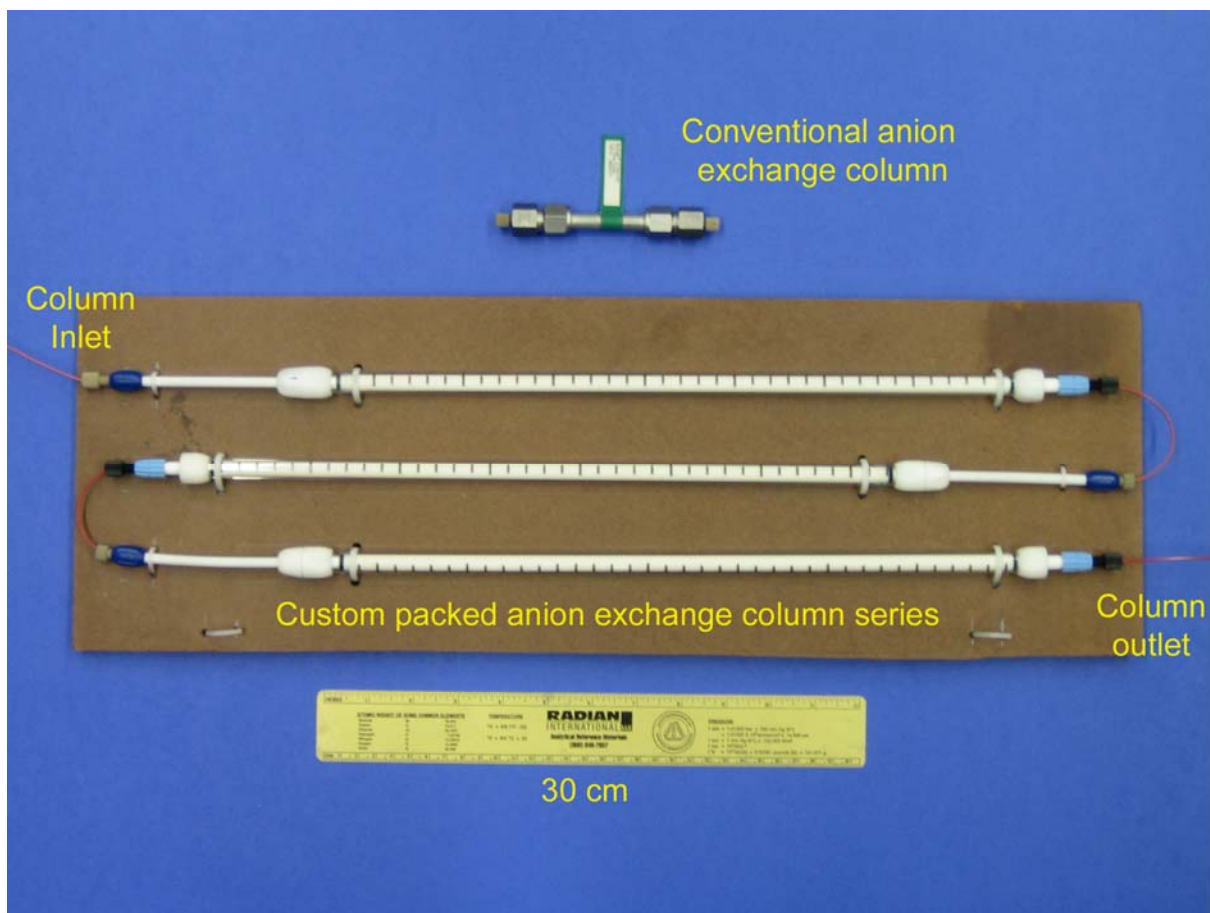
Figure 2-8: Photograph comparing conventional anion exchange column (7.5 cm long x 7.5 mm ID) and the custom-packed anion exchange column series (111 cm total length x 6.6 mm ID).
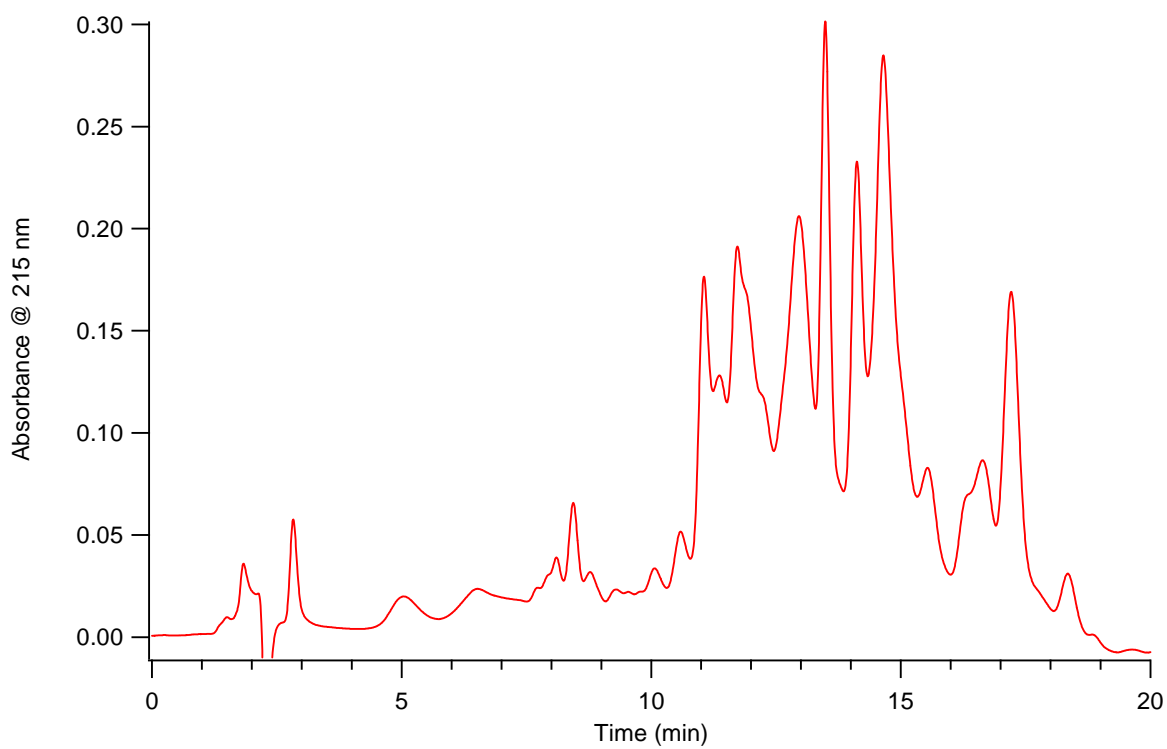
Figure 2-9: Anion exchange separation of the *E. coli* protein extract. Mobile phase A contained 50 mM Tris-HCl (pH 8.5); mobile phase B contained 50 mM Tris-HCl (pH 8.5) plus 500mM sodium chloride. A 20 minute linear gradient from 0 to 100%B was used.
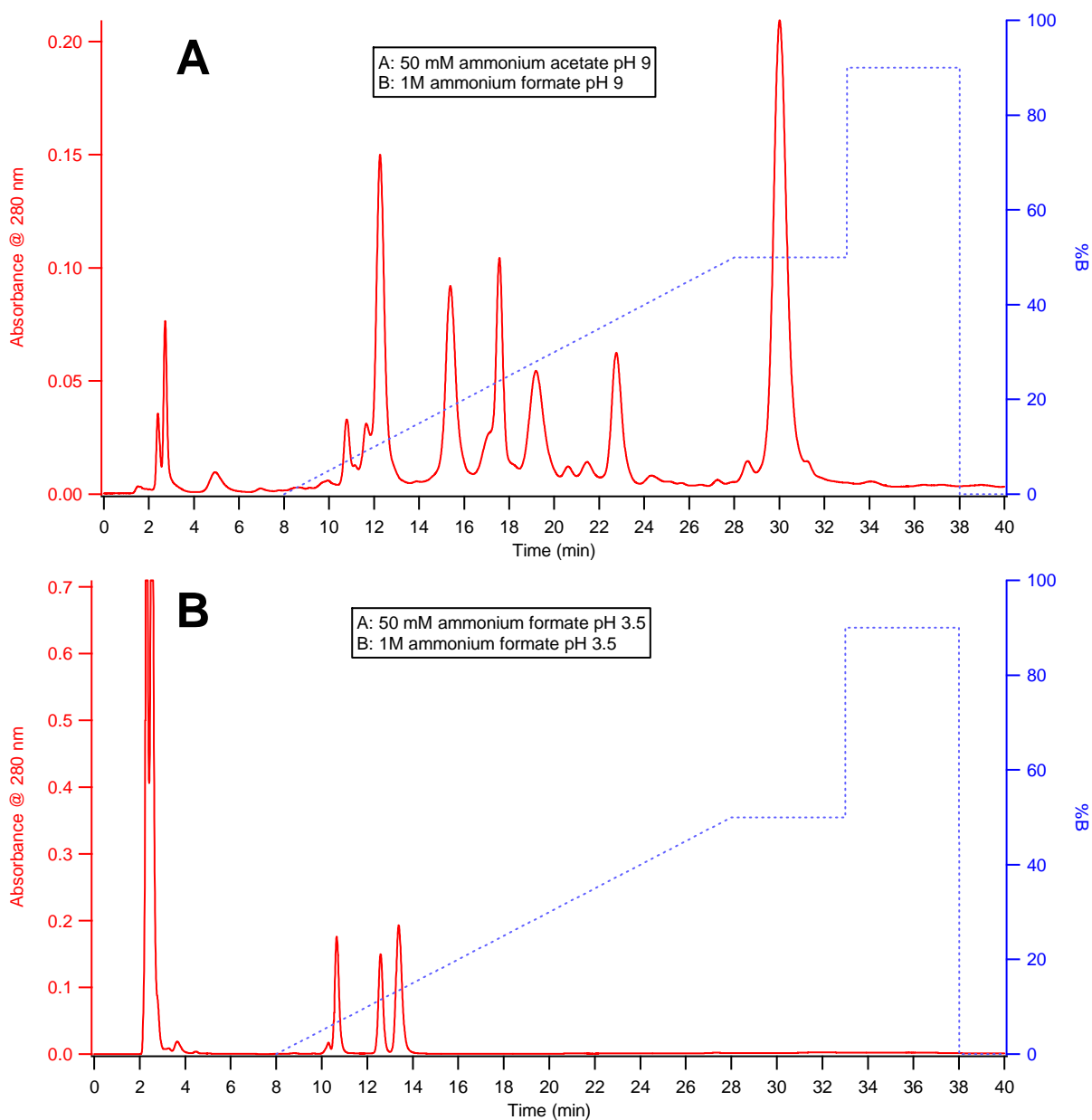
Figure 2-10: (A) anion exchange and (B) cation exchange separations of *E. coli* protein extract. The red trace is the UV absorbance chromatogram. The blue line represents the profile of the gradient in terms of percent of mobile phase B; the balance of the mobile phase was made up of mobile phase A.
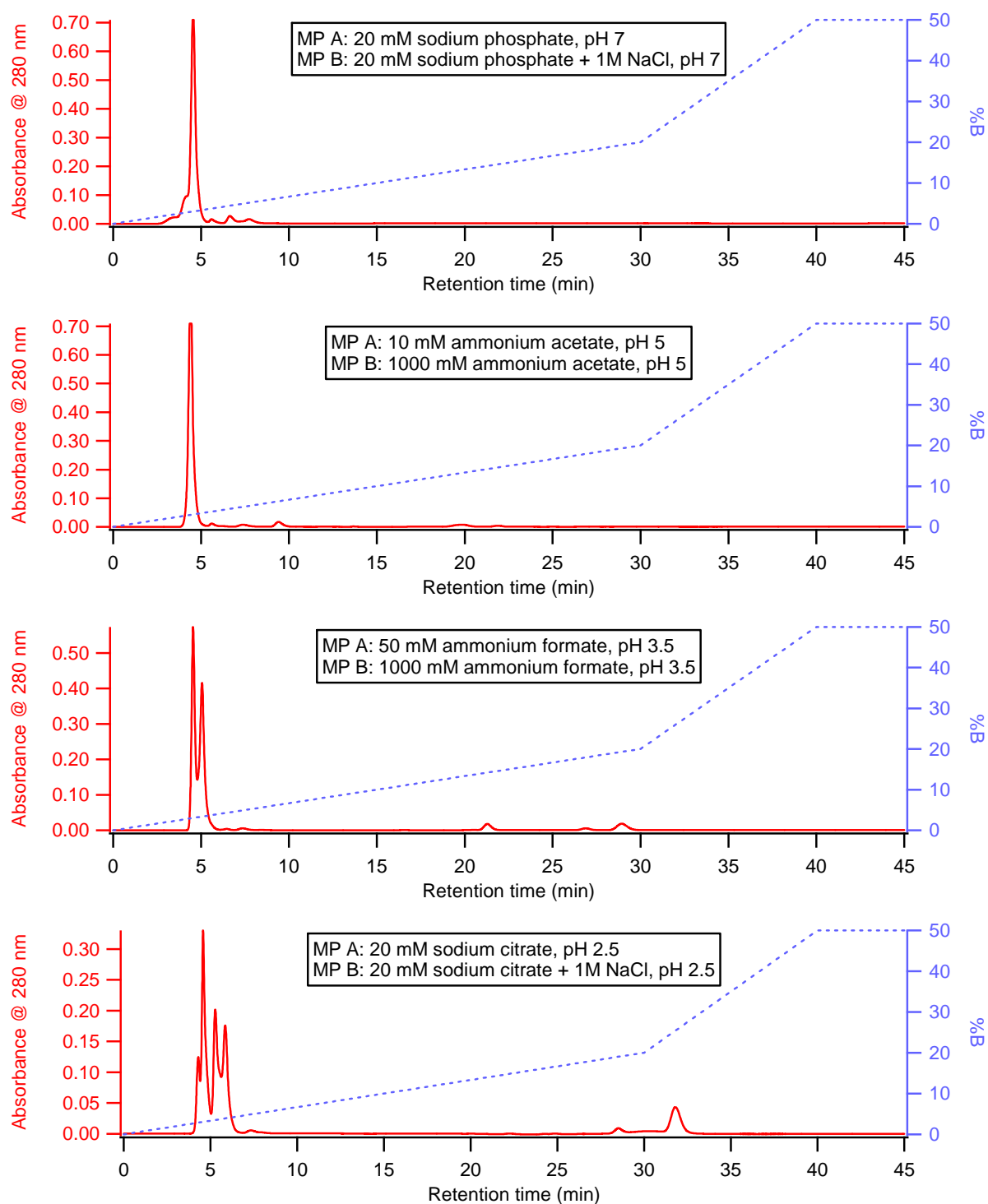
Figure 2-11: Cation exchange separations of the *E. coli* protein extract performed using four different buffer and pH conditions.
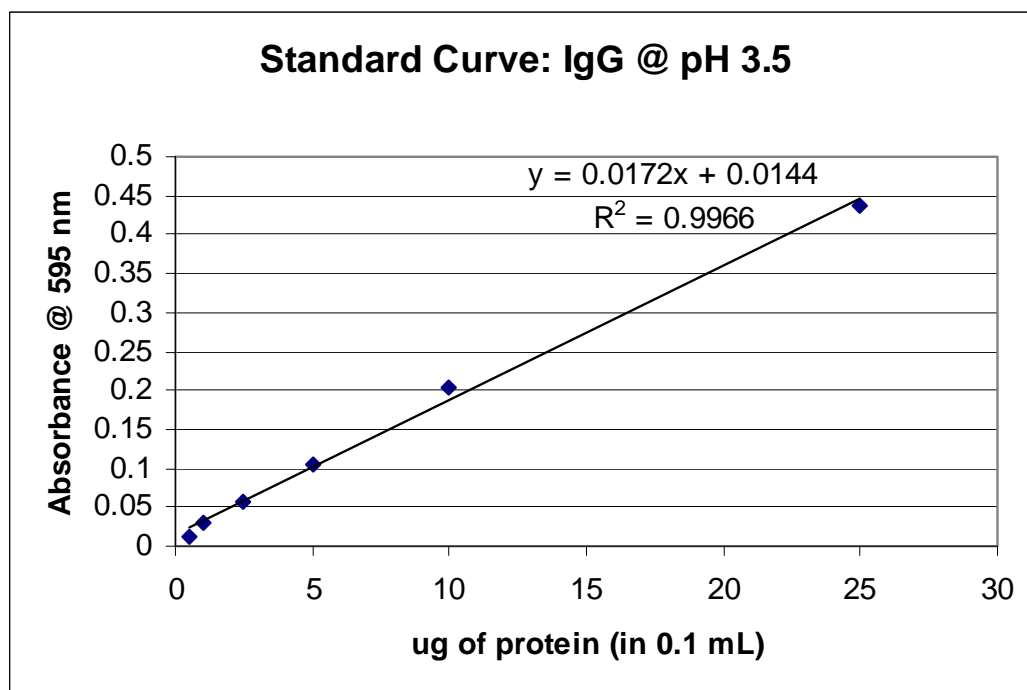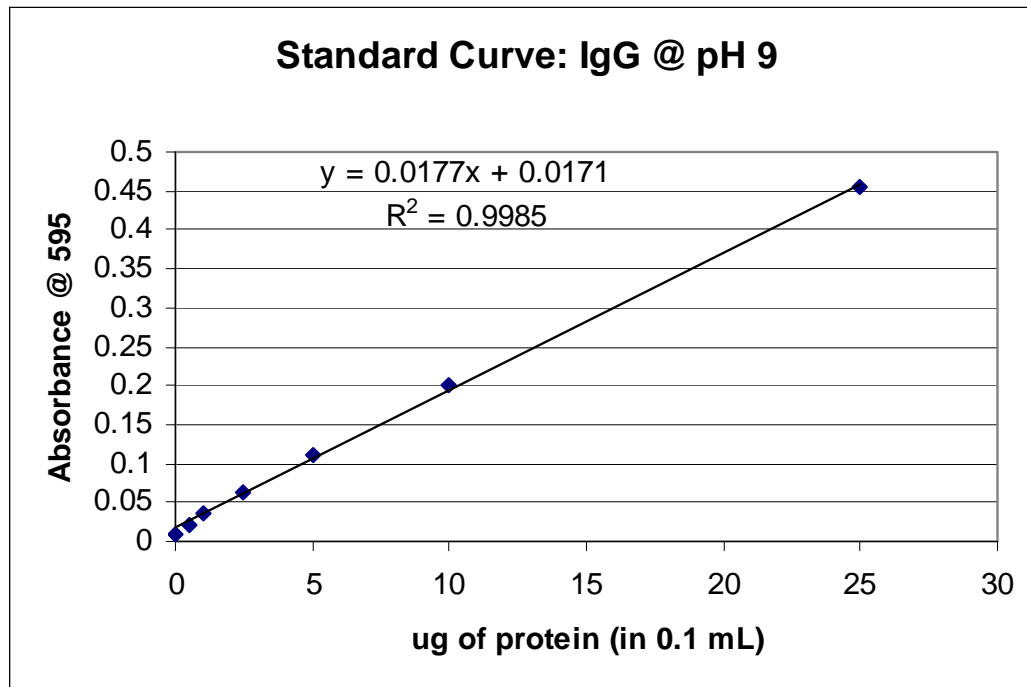
Figure 2-12: Standard curves for the Bradford assay. For the top curve, IgG was dissolved in 50 mM ammonium acetate, pH 9.0. In the bottom curve, IgG was dissolved in 50 mM ammonium formate, pH 3.5.
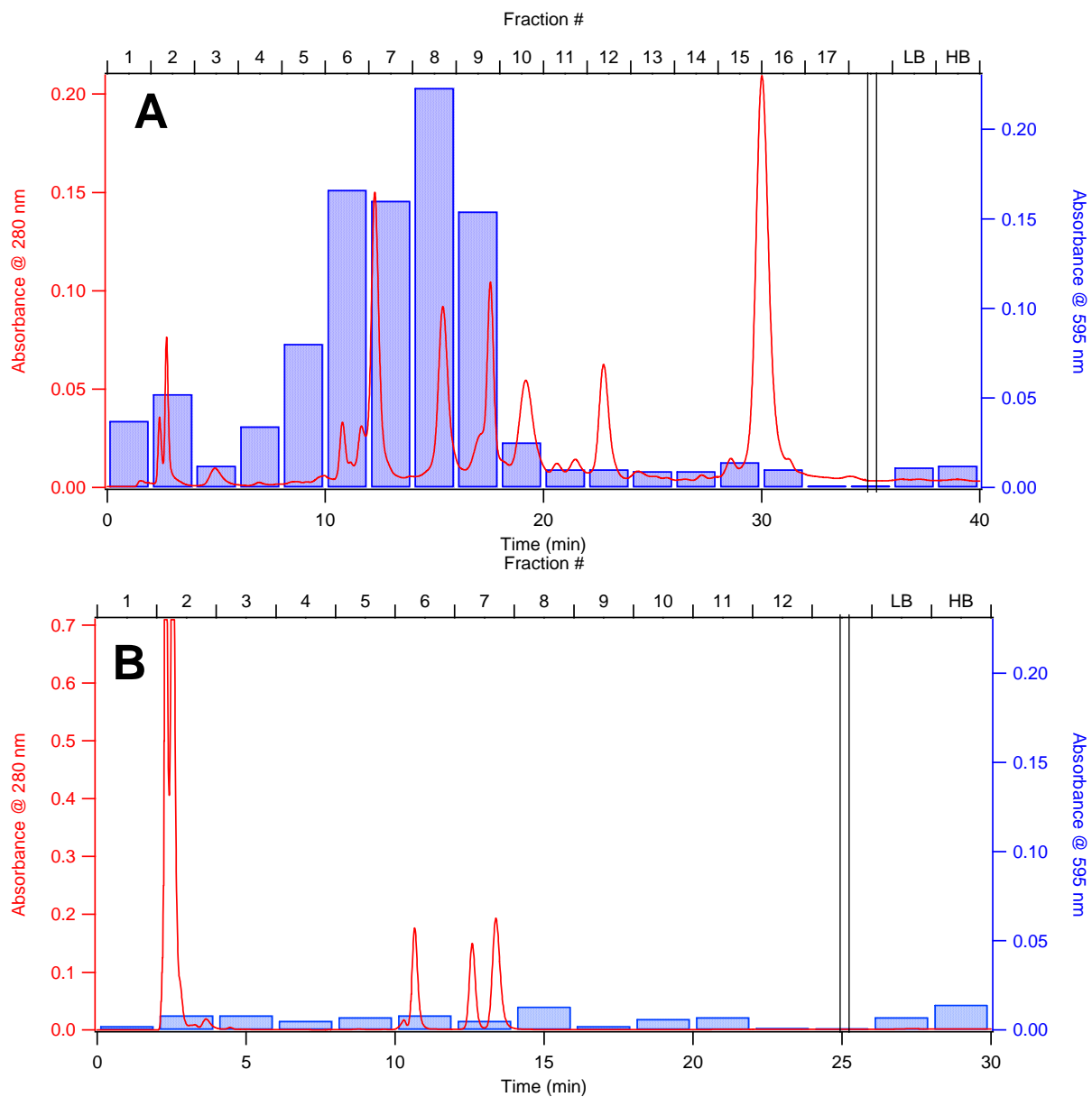
Figure 2-13: Results of the Bradford protein assay used to assess the quantity of proteins present in fractions collected from (A) anion exchange and (B) cation exchange separations of *E. coli* protein extract. The blue bars represent the measured absorbance of each fraction at 595 nm after mixing with the Bradford reagent, which is proportional to protein concentration. The red trace is the corresponding UV absorbance chromatogram.
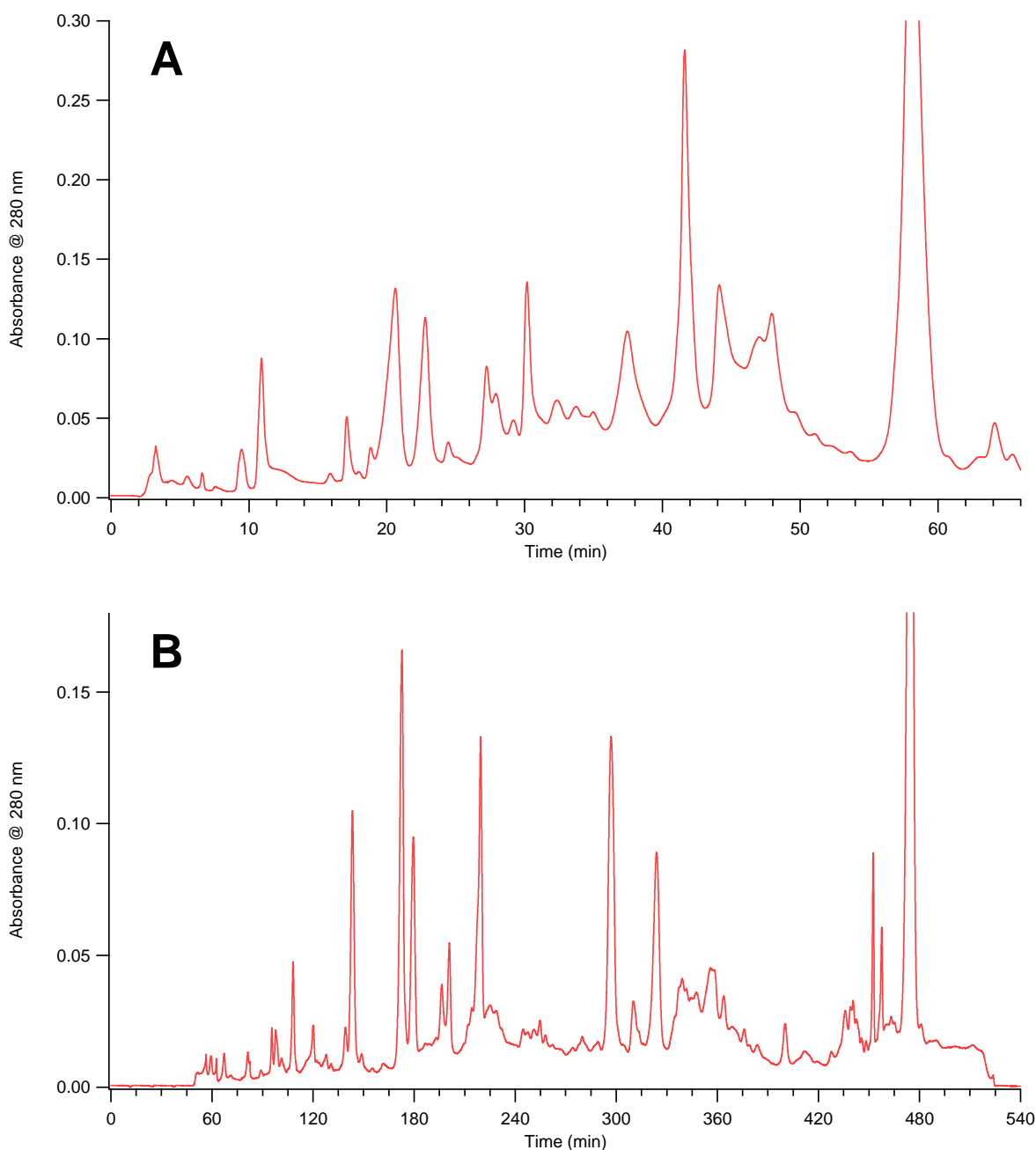
Figure 2-14: Comparison of separations of the *E. coli* protein extract on two anion exchange columns. Chromatogram (A) was obtained with the 7.5 cm Biosuite Q 10um AXC column using a 60 minute gradient from 25-500 mM ammonium acetate (pH 8.5). Chromatogram (B) was obtained using the custom-packed 111 cm anion exchange column using a 360 minute gradient from 25-500mM ammonium acetate (pH 8.5). In both chromatograms the final peak is shown off-scale to allow the detail of the remainder of the peaks to be seen.

CHAPTER 3: Off-line LC x UHPLC-MS separations of *E. coli* proteins

3.1   Introduction

The research described in this chapter falls within the field of top-down proteomics. In general terms, the goal of this method is to separate a complex mixture of intact proteins and to analyze them using mass spectrometry. Although still in development, top-down proteomics shows substantial promise for enabling identification of proteins and study of post-translational modifications without the need for enzymatic digestion.[1] The emphasis of most research in top-down proteomics to date has been within the field of mass spectrometry, especially development of gas-phase fragmentation techniques and methods for interpreting the resulting complex spectra. Separations are also important, however, since no mass spectrometer is able to simultaneously analyze all the proteins produced by an organism. Multidimensional liquid chromatography is a logical candidate, due to the high resolution it provides and the ease with which it can be coupled to mass spectrometry via electrospray ionization.

The focus of this chapter is the development of an LC x LC separation method for intact proteins. Anion exchange and reversed phase were chosen as the first and second dimensions for the separation, respectively, as presented in Chapter 2. Beyond the selection of the separation modes, however, numerous steps were necessary to develop a functional LC x LC method. First, the strategy for coupling the two dimensions was determined. Second, the parameters of the reversed-phase separation were selected and optimized. Third, since

MS is an essential part of any proteomic analysis, the separation method was coupled to a mass spectrometer. Finally, a procedure for analyzing the chromatographic and mass spectral data was developed. Once all aspects of the 2D separation were in place, the method was applied to the study of the *E. coli* proteome.

Before describing the experimental methods used to perform the research, several aspects of the techniques reported in this chapter require further introduction. One is the selection of a method of coupling one dimension to the other in LC x LC. Second is ultra-high pressure liquid chromatography, the technique to be used for the second dimension separation. Finally, since mass spectrometry plays an important role in the method, electrospray ionization MS of intact proteins is briefly introduced.

3.1.1   Off-line vs. on-line coupling in multidimensional LC

Although apparently simple, the selection of a method used to couple dimensions in a 2D separation merits consideration from a conceptual standpoint. The majority of previous research in the Jorgenson group involving LC x LC has employed on-line coupling of dimensions, in which fractions from the first column are loaded directly onto to the second column using automated switching valves. An alternative is off-line coupling, where the effluent of the first column is physically collected as fractions that are later injected individually onto the second column. Although off-line coupling may seem to be a less sophisticated approach, both methods have noteworthy advantages and disadvantages which must be weighed when designing a 2D separation.

The primary advantages of the on-line approach are automation and speed. An entire 2D separation can be carried out with no user intervention beyond the initial injection of the sample. This is very advantageous for the purpose of routine, high-throughput analyses. The entire run can be finished in essentially the time that it takes for the first dimension

separation to complete; therefore, the technique can be made quite rapid, especially compared to 2D gel electrophoresis. However, these advantages impose certain limitations. Most significantly, the second dimension separation must be configured such that it is very rapid compared to the first. This is necessary so that many second-dimension runs can be carried out during the time it takes to complete a single run on the first dimension in order to adequately sample the column. This usually means that the resolution contributed by the second dimension must be compromised to improve the analysis speed. Secondly, the on-line approach requires relatively complex instrumentation. Two LC pumps must be operated simultaneously, and operation of the switching valves used to connect the two dimensions must be timed precisely so that all of the effluent from the first column is transferred to and analyzed by the second column.

Off-line LC x LC is a comparatively simple approach. When minimally configured, it requires only a single LC pump, two separation columns of different types, and suitable containers for collecting fractions. A greater degree of automation can be achieved if a fraction collector is used to collect the fractions from the first column and an autosampler is used to inject the fractions onto the second column. There is no limit to the number of fractions that can be collected, so off-line coupling can be just as thorough in sampling the first dimension as on-line coupling. Fraction collection also permits sample manipulation between separation dimensions – for example, fractions could be concentrated via lyophilization, and reconstituted in a solvent more appropriate for analysis on the second dimension. Another major advantage of the off-line approach is that there is no need for the second dimension separation to be run faster than the first, because fractions can be stored and run when time permits. Therefore both separations can be optimized to provide the

highest possible resolution, which could allow the second dimension to contribute a much higher peak capacity to the overall 2D method. Offline LC x LC also allows flexibility in the amount of each fraction that is transferred to the second dimension. This amount could be increased to improve sensitivity, decreased to prevent second-dimension column overload, or a fraction could be skipped altogether if the detector monitoring the first-dimension column effluent indicates that no analytes are present. Finally, an off-line 2D separation allows the flexibility of analyzing any fraction multiple times on the second dimension without repeating the entire 2D separation, should the sample prove of particular interest or if an instrument failure occurs.

Of course, off-line sampling also has numerous disadvantages. Foremost is speed – off-line sampling is certain to take longer than an on-line approach because the fractions are usually run after the first dimension separation is complete, rather than as it is in progress. An additional factor is that the operator of the instrument may need to spend more time performing manual tasks such as setting up for fraction collection, collecting the fractions, and manipulating the fractions such that they are ready to be analyzed on the second dimension. Finally, since the liquid fractions come in contact with more tubes and surfaces when fraction collection is performed than when on-line coupling is used, there is a greater possibility of sample losses with off-line coupling.

In the case of the 2D separation to be described in this chapter, it was decided to use off-line coupling of dimensions, for two main reasons. First, the goal of the separation was to generate as high of a peak capacity as possible in order to analyze the *E. coli* proteome. Since the speed of the separation was not a primary concern, off-line coupling could be used to allow both dimensions to be optimized without concern as to run time. Second, it was

desired to use ultra-high pressure liquid chromatography (UHPLC) for the second dimension of the separation. UHPLC is an enhancement of HPLC which offers improved chromatographic performance without long run times.[2-6] A description of UHPLC and an explanation as to why it could not be used with on-line coupling are provided in the next section.

3.1.2   UHPLC in multidimensional separations: benefits and challenges

In two dimensional separations, as in any other form of chromatography, it is desirable to generate very high peak capacity in as short a time as possible. In reality, some compromise between speed and resolution must be made, the specifics of which depend on the nature of the sample to be analyzed. UHPLC has the potential to substantially enhance either the speed or the peak capacity of LC x LC, or even both simultaneously to some extent. The basis for this potential improvement is briefly described below, in terms of standard one-dimensional chromatographic theory.

3.1.2.1 UHPLC theory

The efficiency of a chromatographic separation can be described by the height equivalent of a theoretical plate (H), where lower values of H correspond to more efficient separations. The Van Deemter equation describes the relationship between H and mobile phase flow velocity (u) as the sum of three major terms, A, B and C, each of which represent a different contribution to band-broadening in a chromatographic column.

$$H = A + \frac{B}{u} + Cu \qquad \text{(Equation 3-1)}$$

The A-term corresponds to Eddy diffusion, the B-term to longitudinal diffusion, and the C-term to resistance to mass transfer. For more efficient separations, it is advantageous to minimize the value of all of these terms. The magnitude of two of the terms is known to be

related to the diameter of the particles (d_p) with which the column is packed: A is proportional to d_p and C is proportional to $d_p{}^2$. Therefore, it is desirable to use the smallest possible particles in order to achieve the highest efficiency. An added benefit of smaller particles is that the optimum linear velocity – that is, the value of u that gives the most efficient separation – increases as particle diameter decreases. Therefore by decreasing particle diameter, it is possible to perform both faster and more efficient separations.

The drawback to small particles is that they have higher flow resistance than conventional-sized particles when packed in a column, and therefore generate greater backpressures. Conventional LC pumps are limited to a maximum operating pressure around 400 bar (6,000 psi), which can quickly be exceeded with particles of 1-2 μm in diameter. One option is to decrease the column length to a few centimeters or less; this results in faster separations.[7] However, due to the loss of column length no gain in separation efficiency is achieved. An alternative is to use specialized pumps capable of producing substantially higher pressures in order to overcome the increased flow resistance. The resulting technology is termed ultra-high pressure liquid chromatography (UHPLC) or ultra performance liquid chromatography (UPLC). Runs have been successfully performed at pressures up to 6900 bar (100,000 psi) using packed capillary columns.[6] UHPLC has demonstrated plate counts over 350,000 for isocratic reversed-phase separations[4] and peak capacities over 1,000 using gradient elution.[8]

3.1.2.2 UHPLC for LC x LC: high speed vs. high peak capacity

If the main goal for a 2D separation is a short total analysis time, UHPLC is theoretically quite attractive. If on-line coupling of dimensions can be used, the limiting factor for the speed of a 2D separation is the time it takes to complete each second-dimension run. If this time is decreased, the first column can be sampled at shorter intervals and

therefore can be run faster, thus reducing the total analysis time. The high pressure capabilities of UHPLC pumps can be devoted to generating very fast runs in relatively short columns. With the small particles used in UHPLC, fast runs do not result in excessive compromise in terms of chromatographic efficiency, so relatively high performance can be maintained. Therefore fast UHPLC is potentially well-suited to use as the second dimension of an on-line 2D separation. At the time when the research presented in this chapter was performed, however, a practical difficulty prevented UHPLC from being used in this manner. Automated switching valves with low internal volumes – a necessary component for on-line coupling – were not compatible with use at ultra-high pressures. Very recently, some manufacturers have introduced valves which support operation at up to 1,400 bar (20,000 psi) and even some specialized valves which can be used up to 2800 bar (40,000 psi).[9] As the use of UHPLC becomes more widespread, a greater variety of such valves will become available. It is therefore reasonable to anticipate that on-line LC x LC separations using UHPLC will be feasible in the near future.

Another important benefit of UHPLC in the context of LC x LC is its potential to generate very high resolution separations. As previously discussed, the peak capacity of a multidimensional separation is equal to the product of the peak capacities of each dimension as long as all separations are orthogonal and no resolution gained using one dimension is lost in any subsequent dimension.[10] Since reversed-phase UHPLC can give peak capacities well into the hundreds as a 1D technique, it is easy to conceive that extremely powerful comprehensive 2D separations could be performed if UHPLC is appropriately combined with a second, orthogonal separation method. As already discussed, off-line coupling is a good option for LC x LC separations when the main goal is to maximize resolution. For this

reason, the research reported in this chapter uses off-line coupling to interface the first dimension to the second.

3.1.3   Electrospray ionization mass spectrometry of proteins

Within the past two decades, mass spectrometry has revolutionized the field of proteomics by enabling rapid, accurate identification and characterization of proteins. One crucial development which made these advances possible was the introduction of electrospray[11] and MALDI[12] ionization techniques. These methods allow large molecules such as proteins to be ionized and converted to the gas phase without undergoing degradation or fragmentation. Electrospray ionization is of particular interest in the present study because of the ease with which it can be coupled to liquid chromatography.

In electrospray ionization, a solution containing the sample to be analyzed is passed through a capillary with a narrow-diameter tip. In the case of LC-MS, this solution can be the effluent of a separation column. A high electric potential (typically several kV) is applied either to the capillary or to the solution, which causes a very fine spray of charged droplets to form at the capillary tip due to electrostatic forces. As the droplets pass through the air, solvent molecules evaporate, decreasing the size of the droplet and leaving it with a higher charge density. Once the charge density becomes too great – termed the Rayleigh limit – the droplet fissions to produce multiple droplets in a process called Coulombic explosion.[13] This process repeats, yielding progressively smaller droplets. Ultimately, the analyte molecules become completely free of solvent, while retaining the charge originally carried by the droplets. The precise mechanism by which the analyte molecules transition to the gas phase is still under debate, and may depend on the properties of the analyte.[13] Proteins tend to form multiply-charged protonated molecular ions, $[M+nH]^{n+}$, during electrospray. Regardless of the mechanism, the gas-phase ions enter the mass spectrometer through an orifice in a

sampling cone. Inside the mass spectrometer, the ions are placed under vacuum and are accelerated using an electric field. They are then separated based on their mass-to-charge ratio (m/z) using any of a variety of mass analyzers, and subsequently detected.

In this experiment, electrospray-time of flight mass spectrometry (ESI-TOF-MS) was used to detect proteins eluting from a reversed-phase LC column. Much of the work in top-down proteomics to date has used higher resolution mass analyzers, such as Fourier transform - ion cyclotron resonance (FTICR) instruments, in order to enable gas phase fragmentation of proteins.[1] In general, however, TOF analyzers provide adequate resolution to measure intact protein molecular weight within a few Daltons, as was the goal for this experiment. In addition to the information gained by measuring protein MW, mass spectrometry has the advantage of being a very sensitive detection method. When low flow rates are used, as in the capillary LC separations described in this chapter, its sensitivity easily surpasses that of optical detection methods such as UV absorbance.

3.2   Experimental

3.2.1   Overview

A general diagram of the instrumentation and procedure used for LC x UHPLC is shown in Figure 3-1. Conventional-pressure anion exchange chromatography is used to separate proteins in the sample according to their charge in the first dimension. Interfacing between the two dimensions is accomplished by fraction collection after dimension 1, followed by lyophilization of the volatile mobile phase and reconstitution of the fractions in order to concentrate the proteins. All fractions are then analyzed on the second dimension, ultra-high pressure reversed-phase liquid chromatography, which separates based on hydrophobicity. The outlet of the reversed-phase capillary column is directly interfaced with

a mass spectrometer to carry out on-line electrospray time-of-flight MS, which provides intact molecular weight information for all detectable proteins.

3.2.2   Samples and reagents

Ammonium acetate and formic acid (ACS reagent grade) were purchased from Sigma-Aldrich (St. Louis, MO). Ammonium hydroxide (ACS reagent grade) and acetonitrile (HPLC grade) were purchased from Fisher Scientific (Fair Lawn, NJ). Deionized water was purified using a Barnstead Nanopure System (Boston, MA).

The samples analyzed in these experiments were an extract of the soluble proteins of the bacterium *Escherichia coli.* Two different samples were used. The first is from *E. coli.* K12 strain MG1655 wild type, and was prepared by Eric Hamlet in the Giddings lab in the Department of Microbiology and Immunology at the University of North Carolina at Chapel Hill, using the procedure described in Section 2.2.1.2. The second extract was prepared from *E. coli* strain DH5α (Q-BIOgene, Irvine, CA) by researchers at Waters Corporation (Milford, MA), using the procedure set forth by Millea et al.[14]

3.2.3   Dimension 1: anion exchange chromatography

3.2.3.1 Columns and instrumentation

The first dimension of the 2D separation was performed using an anion exchange column. Initially, the column used was the Waters Biosuite Q 10 μm AXC (7.5 cm x 7.5 mm ID), referred to hereafter as the short AXC column. In order to improve resolution, a custom-packed anion exchange column (111 cm x 6.6 mm ID) was also used; it is referred to hereafter as the long AXC column. For a more detailed description of the preparation of this column, see section 2.3.1.3. The anion exchange gradient was generated by a Waters 600E quaternary gradient LC pump. The pump is connected to a Valco 6-port valve (VICI, Houston, TX) with a 100 μL sample loop, which is used to inject sample onto the column.

Detection was performed using an Applied Biosystems 785A UV absorbance detector
(Foster City, CA) set at 280 nm.

3.2.3.2 Chromatographic conditions

For anion exchange separations, mobile phase A was 25 mM ammonium acetate,
adjusted to pH 8.5 with ammonium hydroxide. Mobile phase B was 1M ammonium acetate,
adjusted to pH 8.5 with ammonium hydroxide. The gradient program is as follows: after
sample injection, a linear gradient from 0% to 50% B was run over 30, 60 or 120 minutes
(for the short AXC column) or 360 minutes (for the long AXC column). The mobile phase
was then ramped to 75% B over 5 minutes and was held at this composition for 15 minutes
(short AXC column) or 45 minutes (long AXC column) to ensure complete elution of all
proteins. The mobile phase was then returned to 100% A.   The flow rate for the short AXC
column was 0.5 mL/min; for the long AXC column it was 0.39 mL/min.

3.2.4   Fraction collection and lyophilization

After passing through the UV detector, the effluent from the anion exchange column
was sent to a Waters Fraction Collector II, which automated collection of fractions in 2 mL
polypropylene centrifuge tubes. For the runs with the short AXC column, fractions were
collected every 3 minutes, beginning as soon as the sample was injected, for a volume of 1.5
mL per fraction. For the long column, fractions were collected every 5 minutes, beginning
50 minutes after the sample was injected (the dead time of the column); each fraction had a
volume of 1.95 mL. The total number of fractions varied depending on the length of the
gradient: 15 were collected for the 30 minute gradient, 25 for the 60 minute gradient, 40 for
the 120 minute gradient, and 80 for the 360 minute gradient. After collection, fractions were
flash-frozen using liquid nitrogen and were then placed in a SpeedVac Concentrator
(Thermo-Electron, Bellefonte, PA), which was pumped down to pressures between 10^{-2} to

10^{-3} Torr using an Edwards double-stage rotary vacuum pump (Wilmington, MA). Once the fractions had been lyophilized to dryness, they were reconstituted in 100 μL of a solution of 10% acetonitrile and 90% water (v/v). The fractions were then ready for analysis on the second dimension.

3.2.5 Dimension 2: ultra-high pressure reversed phase liquid chromatography

3.2.5.1 Columns and instrumentation

The second dimension RPLC separations were performed using capillary columns prepared in lab. The columns are packed with 1.5 μm diameter bridged ethyl hybrid (BEH) particles with a C18 stationary phase. A typical column used for the separations reported in this chapter was 45 cm long and had an ID of 50 μm. A detailed characterization of the BEH particles, including a description of the procedure used to pack them in capillary columns, has been reported previously.[5] These particles have also been successfully used for intact protein separations at ultra-high pressures.[15, 16]

A custom-built LC system was used to perform gradient reversed phase separations at ultra-high pressures.[15] The system was designed and constructed by Geoff Gearhardt and Keith Fadgen at Waters Corporation. A photograph of the instrument is shown in Figure 3-2; a schematic diagram is shown in Figure 3-3. The instrument consists of two separate pumps. The first is a Waters CapLC, which is a conventional gradient pump with a 350 bar (5,000 psi) pressure limit. It also has an integrated autosampler. This pump is responsible for injecting the sample and generating the reversed phase gradient to be run on the column. The second pump is actually a hybrid of a Waters 1525 HPLC which has been modified to pump automotive brake fluid, and a custom-designed hydraulic-amplifier which amplifies the fluidic pressure produced by the 1525 to generate the ultra-high pressures used to carry out the separation. It is capable of generating pressures up to 3100 bar (45,000 psi). The outlet

of the CapLC and hydraulic amplifier pumps are connected to a 4-port stainless steel high-pressure union, as are the capillary column and an open-tubular split flow capillary (1.3 m long, 10 μm ID). The hydraulic amplifier pump connects to the 4-port union through the gradient storage tubing, which is a 5 meter long length of 0.020" ID stainless steel tubing used to store the gradient loaded by the CapLC pump. After loading, the gradient is sent onto the column by the hydraulic amplifier pump. Also in the fluidic flow-path are two air-actuated on/off pin valves, model ASFVO, from Valco. One valve is located between the CapLC and the 4-port union; the second is located at the outlet of a pressure release vent on the hydraulic amplifier pump. Although the valves are only rated for use up to pressures of 700 bar (10,000 psi), we have found that they give many months of reliable service at pressures up to 1700 bar (25,000 psi), and can tolerate occasional use at pressures up to 2800 bar (40,000 psi). All of the instrumentation is controlled via a PC, using both the software package MassLynx 4.0 (Waters) and a custom-written software program for the hydraulic amplifier pump.

A schematic diagram of the instrument which illustrates the process of performing a run is shown in Figure 3-3. A vial containing the sample to be injected is placed in the autosampler tray of the CapLC. With valves 1 and 2 in the open position, the CapLC generates an acetonitrile/water gradient, which travels through the 4-port union and onto the gradient storage tubing. Since the gradient will later be forced onto the column by the ultra-high pressure pump, which is located on the opposite end of the storage tubing, the gradient must be loaded in reverse – that is, beginning with the highest desired acetonitrile content and ending with the lowest. The CapLC is typically set to load the gradient at a flow rate of 40 μL/min, whereas the ultra-high pressure pump operates at 2 μL/min when a run is in

progress. Therefore, a gradient that will run for 60 minutes only takes 3 minutes to load onto the gradient storage tubing. Once loading of the gradient is complete, the CapLC loads a plug of the sample, typically 0.1-10 μL in volume, onto the storage tubing in the same manner. Although the column and splitter are not blocked while the gradient and sample are being loaded, they both have a very high flow resistance compared to the gradient storage tubing, so only a miniscule amount of flow enters the column before the run begins.

Once the sample has been loaded, the ultra-high pressure run is ready to begin, so valve 1 is closed in order to isolate the CapLC pump from ultra-high pressures. Valve 2 is also closed, and then the ultra-high pressure pump is activated in order to bring the system to the desired run pressure – 1600 bar (23,000 psi) is typical. Once this pressure is reached, the pump operates at a constant flow rate of 2 μL/min. The splitter capillary diverts most of the flow from the gradient storage tubing and maintains appropriate pressure at the head of the column. The flow rate through the separation capillary depends on this pressure, as well as the length and diameter of the column and the size of the packing material. For a 50 μm ID capillary column, a flow rate of about 100 nL/min is typical, which means the split ratio is approximately 20:1. The sample in the gradient storage tubing is the first liquid which enters the column. Next, the column experiences the stored gradient, which is transferred to the column in the normal manner of low to high acetonitrile concentration.

3.2.5.2 Chromatographic conditions

For reversed phase separations, mobile phase A was deionized water with 0.2% (v/v) formic acid (ACS reagent grade, Sigma-Aldrich). Mobile phase B was acetonitrile (HPLC grade, Fisher) with 0.2% (v/v) formic acid. Runs were performed using a linear gradient from 10% to 60% MP B over 60 minutes, followed by a 10 minute hold at 60%B, and then a return to initial conditions.

3.2.6 Electrospray – time of flight MS detection

The outlet of the separation column was coupled using a Teflon sleeve to a fused silica PicoTip nano-electrospray emitter purchased from New Objective (Woburn, MA). The emitter, which is model #FS360-20-5-CE, has an inner diameter of 20 μm which tapers to 5 μm at the tip. It has a conductive coating, which allows electrospray voltage to be applied by placing a light-weight wire loop around the emitter, then connecting the other end of the wire to the capillary voltage power supply on the mass spectrometer. The tip of the emitter was placed a few millimeters away from the sample cone of the mass spectrometer and was oriented perpendicular to its orifice. The emitter-to-cone distance was manually optimized to give the best signal.

On-line positive ion mode electrospray - time of flight MS was performed using a Micromass Q-TOF Micro instrument (Waters). A capillary voltage of 2000 V, a sample cone voltage of 40 V, and an extraction cone voltage of 3 V were used for runs in this experiment. The source temperature was set at 100 ºC; no desolvation gas was necessary for stable electrospray at the flow rates used. Mass spectra were acquired at a frequency of 2 Hz over a mass range from 450-1600 Da for the duration of the run. Although the instrument is capable of tandem mass spectrometry, it was operated in TOF-only mode for the experiments reported in this chapter.   All mass spectra were acquired using the software package MassLynx 4.0.

3.2.7 Data analysis

2D chromatograms were prepared by loading total ion current data from each reversed-phase chromatogram into the data analysis software program Igor Pro (Wavemetrics, Lake Oswego, Oregon). In Igor Pro, the data from all fractions in an anion exchange run were combined into a two-dimensional data set. The data was plotted as an

image to generate a 2D chromatogram. The upper limit of the intensity scale was usually set by the most intense peak in the chromatogram, although in some cases the most intense peaks were intentionally set off-scale in order to allow lower intensity components to be seen.

In the case of the run performed using the long AXC column, an additional method is used to generate an enhanced 2D chromatogram. All of the LC-MS data from the second-dimension runs were processed using AutoME, a software program written by Ignatius Kass at Waters Corporation. Figure 3-4 illustrates the steps AutoME uses. First, it divides a chromatogram up into time segments, typically 6 seconds wide, as indicated by the dashed vertical lines on the chromatogram in Figure 3-4A. All mass spectra acquired within each time segment are combined into a single spectrum, an example of which is illustrated in Figure 3-4B. An iterative mathematical process called maximum entropy (MaxEnt) is used to de-convolute the experimental data, which contains ions in multiple charge states, to give a model spectrum in which all components have zero charge.[17] An example of a mass spectrum before and after MaxEnt de-convolution is shown in Figure 3-4B and C. Various parameters can be specified which affect the manner in which de-convolution is performed. For this experiment, these parameters were set such that the resolution of the output spectrum is 1 Da, the time segment width is 6 seconds, the range of masses in the output spectrum is 3,000-60,000 Da, and the maximum number of iterations used is 50. Once all data from a chromatogram is processed using AutoME, a custom-written Igor Pro function is used to determine the most intense peak (termed the "base peak") in each time segment. The intensity of this peak is then plotted against the original retention time of the time segment; the result is called a base peak intensity (BPI) chromatogram. An example is shown in Figure 3-4D. When all of the base peak chromatograms for the RPLC runs in a 2D

separation are combined using Igor Pro, a 2D base peak intensity chromatogram can be produced.

To determine the molecular weight of the proteins in the *E. coli* sample, two different methods are possible. For those chromatograms processed using AutoME, a list of the de-convoluted protein masses is automatically generated as a part of data processing. However, AutoME processing could not be used on the runs performed using the short AXC column because the MS data were acquired with an intensity threshold to reduce noise, which precludes MaxEnt processing. Therefore, a manual de-convolution approach was employed. All chromatograms were thoroughly surveyed for peaks that gave a detectable charge envelope. When such a peak was found, all MS scans under the peak were summed, and the resulting mass spectrum was background subtracted, smoothed and centered according to MassLynx default parameters to convert the spectrum from continuum data to a line spectrum. Then an adjacent pair of ions from the same charge envelope in the mass spectrum was identified visually. Using the MassLynx "Find Manual" dialog, the ion pair's m/z values were used to discover all remaining ions in the same series and calculate the actual molecular weight of the intact protein. If more than one charge envelope was present in the same mass spectrum, all other detectable protein masses were also measured. This procedure was repeated in the same manner for all peaks in all of the chromatograms, and protein mass data were recorded in tabular form.

Identification of the proteins based on their molecular weight was not attempted because mass accuracy was not sufficient to distinguish between the thousands of proteins known to be produced by *E. coli*. Instead, the data were used to determine the number of probable unique protein masses detected in each fraction and in each 2D run. To generate

this list, two steps were required. First, multiple forms of the same protein must be disregarded. Within a mass spectrum of a chromatographic peak, often one main component is present along with several other less intense components with a higher or lower mass, typically within a few hundred Da of the most intense component. Potential explanations for these additional components include different post-translational modifications of the same protein, or the formation of non-covalent adducts between proteins and ions other than H^+ (e.g., Na^+, K^+, or other metals) during electrospray ionization. Regardless of the origin of these extra masses, since their retention time is identical to the main component, they are interpreted as being different forms of the same protein. Therefore only the most abundant mass in the spectrum was counted as a unique protein; all additional masses within a set range of this mass (typically +/-500 Da) were disregarded. Second, duplicate masses between fractions needed to be removed. If the same protein mass (within an error of +/- 5 Da) appeared in more than one fraction in a 2D run, it was counted as being found only in the fraction where its intensity was the greatest. Once duplicates and multiple forms of the same protein were eliminated from the list, it was considered complete and the total number of proteins detected was tallied.

3.3   Results and discussion

3.3.1   Anion exchange separations

Two different *E. coli* protein extracts, described in section 3.2.2, were analyzed in the experiments reported in this chapter. Although both samples are derived from the same species, differences between the strains and the procedures used to produce the extracts are likely to result in differences in the proteins which are present in each of the samples. To perform a simple comparison of the two protein extracts, anion exchange separations of both samples were run under identical conditions, using a 30 minute gradient from 25 to 500 mM

ammonium acetate (pH 8.5). UV absorbance chromatograms from both separations are shown in Figure 3-5. It is immediately apparent that there are substantial differences between the samples. Peaks appear to be spread over the same range of retention time, but prominent peaks present in one sample are often absent in the other. It appears that the strain DH5α *E. coli* sample prepared by Waters Corporation produces more total peaks. It is impossible to determine from these chromatograms how the samples differ in terms of protein identities and quantities. Nevertheless, it is clear that the two samples cannot be considered equivalent for the purpose of comparing runs performed under different chromatographic conditions. For the remaining runs reported in this chapter, runs performed using the short AXC column analyzed the protein extract from the MG1655 strain, whereas the long column was used to analyze the DH5α strain.

Figure 3-6 shows the UV absorbance chromatograms from three anion exchange separations of the MG1655 strain *E. coli* protein extract. The goal of these separations was to study the effect of changing the anion exchange gradient length on peak capacity and total number of proteins detected in the 2D separations. The sample is too complicated to be fully resolved by any of the anion exchange separations, as indicated by the presence of overlapping peaks throughout all of the chromatograms. Nevertheless, by extending the gradient from 30 minutes to 60 minutes or 120 minutes, some additional resolution is gained. Although it is difficult to assign a peak capacity to these separations without fully-resolved peaks, rough estimates for the peak capacity are 20, 30, and 42, for the 30, 60, and 120 minute gradients, respectively.

A chromatogram from a 360-minute gradient separation of the *E. coli* DH5α strain protein extract using the long AXC column is shown in Figure 3-7. As explained previously,

a direct comparison between the peaks observed in this separation and those from the short AXC columns is not possible due to the differences in the samples. However, it is clear that the peak capacity is substantially higher in this separation than in any of the runs performed on the short AXC column – 80 is a reasonable estimate.

3.3.2 UHP – RPLC separations

After being lyophilized and reconstituted, fractions from the anion exchange separations were analyzed on the second dimension, UHP-RPLC. For each anion exchange separation, a series of reversed-phase chromatograms are produced. A representative chromatogram of a reversed-phase separation of one anion exchange fraction is shown in Figure 3-8. The chromatogram is a plot of the total ion current (TIC) measured by the mass spectrometer as a function of retention time. Peaks typically do not appear until after 15-30 minutes in most of the reversed-phase chromatograms, mainly because of the delay associated with transferring the sample from the gradient storage tubing onto the reversed phase column. Further inspection of the chromatogram reveals that there is a great deal of variability in peak shape. Some peaks are sharp and symmetrical, having base widths as small as 10 seconds. Some proteins co-elute, causing overlapping peaks and distorted peak shape. Some non-overlapping peaks, particularly those appearing late in the chromatogram, are also broad and asymmetrical, having notable tailing and widths as large as several minutes. Assuming an average peak width of 20 seconds and an elution window of 33 minutes, an estimated peak capacity for this separation would be 100, which is a relatively high number for a 1D intact protein separation.

Although the observed inconsistency in peak shape is clearly not desirable, it is not atypical of reversed-phase protein separations due to non-ideal interactions between proteins and the stationary phase.[18-21] Another factor which contributes to the large width of some

peaks is column overloading. In order to be able to detect minor components of the sample, a relatively large volume of sample must be injected onto the column. This high loading saturates the stationary phase with the most abundant proteins and causes them to elute as intense, broad peaks. One option to remedy this problem is to reduce column loading and use a more sensitive mass spectrometer with a wider dynamic range. Due to the expense of cutting-edge mass spectrometers, this is not often feasible, so another means to combat the dynamic range issue is with higher chromatographic peak capacity. Although increasing peak capacity does not diminish column overloading, the chances of a broad peak overwhelming smaller adjacent peaks are reduced when the peaks are spread over a larger field of separation space. The high total resolving power of the UHP-RPLC separation is beneficial in this regard. This is also a significant argument in favor of 2D separations. The multiplicative peak capacity provided by 2D separations greatly increases separation space, which reduces the probability of peak overlap even when sample loading is high.

3.3.3   2D Chromatograms

Two-dimensional chromatograms can be generated by displaying a series of the 1D RPLC chromatograms side by side. Figure 3-9A is an example of such a plot, created using chromatograms from the RPLC separations of fractions from the 60 minute gradient AXC separation of the *E. coli* protein extract. The chromatograms are offset from one another both horizontally and vertically, so that the baselines of all of the traces can be seen. A different color is also used for each chromatogram to make them easier to distinguish. The result is a rudimentary "3D" visualization of the 2D separation. The figure does reveal the complexity of the sample and the distribution of its components over the both separation dimensions. One disadvantage of this type of plot, however, is that high intensity components which appear in chromatograms toward the front obscure the peaks that appear

94

further back. An alternate means of visualizing the data is to plot it on a planar 2D surface as shown in Figure 3-9B. First dimension retention time is plotted on the Y-axis, and second-dimension retention time is shown along the X-axis. Intensity is indicated by color rather than by peak height. Although the color scale is arbitrary, it allows simple visual comparison of peak intensity within a chromatogram and between chromatograms. In general, this method of visualization is more useful than a 3D plot, since all of the data is clearly visible in a single figure.

In Figure 3-10, 2D chromatograms are shown for separations of the *E. coli* protein extract performed using 30, 60 and 120 minute gradients on the short AXC column. The data are scaled such that the maximum peak intensity is represented by the color at the top of the scale (black). It is possible to estimate a peak capacity for the 2D separations by multiplying the peak capacity of each dimension. Although the peak capacities of the AXC separations were already determined from their UV chromatogram, these estimates are not valid for the 2D separation, since the maximum effective peak capacity from the first dimension cannot be greater than the total number of fractions collected. Since some, but fewer than half, of the peaks appear to be split between two fractions, it is reasonable to use 2/3 of the total number of fractions that contained protein peaks as a conservative estimate for the first-dimension peak capacity. If the second dimension peak capacity is taken to be 100 for all of the RPLC separations, the resulting total peak capacities of the 2D separations are approximately 900 for the 30 minute separation, 1200 for the 60 minute separation, and 1900 for the 120 minute separation. If the anion exchange column had been sampled by a greater number of fractions, these peak capacities would have been substantially higher

In comparing the three chromatograms, it is apparent that the general pattern of the elution profile is similar for all three runs. Numerous peaks appear within the first 12 minutes of anion-exchange retention time in all three 2D chromatograms. These correspond to the proteins that were un-retained or very lightly retained on the anion exchange column, which implies that they are the most basic proteins in the sample. There is then a gap of several minutes where few peaks appear; the gap is longer for the shallower anion exchange gradient and shorter for the steep one. The time after this gap at which peaks reappear corresponds to the point on the gradient when the salt concentration is high enough to begin eluting proteins that were retained on the anion exchange column. Peaks are spread over a fairly wide portion of the available separation space, although certain areas of the 2D plot clearly contain greater concentrations of protein than others. Additionally, there seems to be a weak correlation between retention on the two dimensions – proteins that elute early from the anion exchange column (the most basic proteins) also tend to elute earlier in the reversed-phase runs. This correlation is expected, however, since many of *E. coli*'s most basic proteins are ribosomal proteins, which also tend to be fairly hydrophilic and thus are expected to be lightly retained on the reversed-phase column. In general, however, the two separation methods appear to be relatively orthogonal in that peaks are spread over a relatively wide range of the available 2D separation space.

As the anion exchange gradient is lengthened, more detail becomes visible in the 2D chromatograms. Regions that appear laden with many overlapping, high intensity bands in the chromatogram from the steepest anion exchange gradient begin to separate into resolved peaks with the longer gradients. Another trend is that the intensity of all peaks seems to diminish as the anion exchange gradient is lengthened. This is most apparent in Figure

3-10C, which is the 2D chromatogram generated from the 120 minute anion exchange gradient. Not only is the maximum peak intensity lower, but also the peaks which elute late in the gradients (those in the upper right of the chromatogram) seem to be lower in intensity relative to other peaks. The notable decrease in sensitivity as the gradient is made shallower is consistent with previous observations from gradient reversed-phase and ion exchange separations of proteins.[22] Additionally, it has been shown that, for a given column length, the peak capacity of a separation will reach a maximum at some gradient length, and that making the gradient shallower beyond this point will result in no further improvement of peak capacity.[23, 24] Both of these factors emphasize that there is a point of diminishing returns when attempting to improve resolution by lengthening the gradient.

If the column length is increased, however, the potential exists to take advantage of much longer gradients, as discussed in section 2.3.2. A 2D chromatogram from a separation of the DH5α strain *E. coli* protein extract performed using the long AXC column is found in Figure 3-11. A total of 80 fractions were transferred to the second dimension, although only about the first 56 fractions contained detectable protein peaks, so only these fractions are displayed in the chromatogram. The data are scaled such that a few of the most intense peaks are off scale, in order to allow peaks of low intensity to be seen more clearly. In general, it is apparent that this separation has substantially higher peak capacity than any of the separations performed using the short AXC column. Using 2/3 of the number of fractions containing proteins as the AXC peak capacity, and 100 as that of the second dimension gives a total peak capacity of 3700. Another notable feature in the chromatogram is the "stripe" of peaks running throughout all fractions at a reversed-phase retention time of 35 minutes. Based on its mass spectrum, it was determined to be a non-protein component with an m/z of

679. Although its origin was initially unknown, it served as a convenient "alignment" marker which allowed retention times to be adjusted to compensate for retention time drift, which is a problem with the stored gradient UHP-RPLC pump. It was later determined that the source of this peak was a contaminant originating from nylon membrane filters used to prepare the LC-MS mobile phase;[25] eliminating this step removed the contaminant peak.

A second chromatogram was produced for the same 2D separation using the data de-convoluted by AutoME as described in section 3.2.7. The chromatogram, shown in Figure 3-12, is a plot of base peak intensity – that is, the intensity of the single most intense component at each point in the chromatogram. Also, only components within the mass range specified for de-convolution (3,000-60,000 Da) are considered, so non-protein components do not appear in the chromatogram. In general, the BPI chromatogram is substantially "cleaner" in appearance than the 2D TIC chromatogram. The contaminant peak seen in all fractions is gone, many areas in the TIC chromatogram which were filled with low intensity noise are completely empty, and some peaks which appeared to be tailed are more symmetrical. These improvements make it easier to distinguish actual protein peaks, and would facilitate comparisons of different chromatograms.

3.3.4 Protein mass spectrometry data

Many of the trends that can be visually observed by examining the 2D chromatograms are also supported by the mass spectrometry data. Figure 3-13 presents graphs of the number of probable unique protein masses found in each fraction of the three 2D runs performed using the short AXC column. As was noted in the chromatograms, in all of the runs there is a spike in the number of proteins detected in the second fraction, which corresponds to proteins not retained by the anion exchange column. This is followed by a

gap of several fractions where few proteins are detected. After the gap, the majority of the proteins elute over a range of the next 10 to 25 fractions, depending on the gradient length.

Several differences between the three runs are also apparent in the MS data. For one, lengthening the anion exchange gradient does spread the proteins over a greater number of fractions, as anticipated. For the 30 minute gradient all proteins elute within 15 fractions, whereas they are spread over 31 fractions for the 120 minute gradient. Also notable is the fact that the number of proteins detected in each fraction is substantially less for the 120 minute anion exchange gradient run than for the 30 or 60 minute gradients. As the gradient is lengthened, many of the proteins that elute in only one fraction in the shorter gradients are spread over two or more fractions in the longest gradient, and thus are more dilute. This reduces their MS signal intensity and causes some proteins to fall beneath the detection limit, which prevents them from being counted. Just as is the case for the trend observed from visual inspection of the 2D chromatograms, the decrease in the number of proteins detected is consistent with expected drop-off in sensitivity as gradient length is increased.[22]

A plot of the proteins found by fraction for the separation performed using the long anion exchange column is shown in Figure 3-14A. It is immediately apparent that the single largest concentration of proteins was found in the first fraction; this is probably because a substantial portion of the proteins were not significantly retained by the anion exchange column. The proteins which are retained, however, are spread over a relatively large number of anion exchange fractions. Since a different sample was analyzed, a specific comparison of the number of proteins found using the two AXC columns would not be valid.

Another means of examining the data is by looking at the molecular weight of the proteins found in the *E. coli* samples. Figure 3-14B and Figure 3-15 show MW distribution

histograms for the proteins detected using the long AXC column and the short AXC column, respectively. Proteins were observed over a mass range from 3 to 78 kDa in all of the three runs on the short AXC column; on the long column no proteins were observed above 60 kDa because this was the upper limit set for the AutoME de-convolution program used to determine protein MW for this run. Clearly, the majority of the proteins detected fall toward the low end of the mass range; the bin with the largest number of proteins is 9-12 kDa in all separations except the one performed on the long column. For the long column, the most proteins were found in the 3-6 kDa range, which contrasts with the short column runs in which relatively few proteins were detected in this range. This is explained by the fact that the sample analyzed on the short column – *E. coli* strain MG1655 – was depleted of proteins below 5 kDa using a MW cutoff filter. The DH5α strain protein extract was not prepared in this manner, so the smaller proteins remain. Otherwise, all four protein distributions for all of the samples look relatively similar, with substantial differences only in the number of proteins detected in each size range.

It is also possible to compare the observed MW distribution data to that which would be expected based on known *E. coli* proteins. Using the SwissProt database, accessible at http://ca.expasy.org/srs5/, the average MW of an *E. coli* K12 strain MG1655 protein was determined to be 35.1 kDa, and the median was 30.6 kDa. In comparison, for the data from the 2D separation using a 60 minute gradient on the short AXC column, the average MW was 21.9 kDa and the median was 18.1 kDa. This indicates that the methods used in this study were biased significantly toward detection of smaller proteins. This is not surprising, since electrospray ionization is known to give better results with smaller proteins than with larger ones. Additionally, the theoretical average and median protein MW do not take

abundance into account. It is likely that many of the most abundant *E. coli* proteins have a relatively low MW; these proteins are more likely to be detected than the less abundant proteins, many of which may have a higher MW.

The data from all four 2D runs are summarized in Table 3-1. The total number of unique protein masses found in all fractions increased from 209 to 247 for the 30- and 60-minute anion exchange gradients, respectively. This suggests that the increased peak capacity contributed by lengthening the anion exchange gradient allows more proteins to be resolved and detected. This trend did not continue when the gradient was lengthened further, however, as the number of proteins detected decreased to 176 for the 120 minute anion exchange gradient. This drop-off probably results from the fact that the proteins become too dilute when spread over several fractions. The resulting diminished signal intensity offsets any gain in resolution achieved by lengthening the anion exchange gradient beyond a certain point. Therefore, of the separations performed on the short AXC column, the 60 minute gradient gave the best results. A total of 422 probable unique proteins were detected in the 2D run performed using the long AXC column. Although direct comparisons with the short column are not valid because a different strain of *E. coli* was used, the improved peak capacity generated by the 2D separation using the long column does seem to enable the detection of a greater number of proteins.

3.4   Conclusions

The peak capacity achieved using the long AXC column as part of a 2D separation, estimated as 3700, is one of the highest reported to date for an LC x LC separation of intact proteins. This peak capacity could be increased substantially if the first dimension were sampled more frequently, so that no resolution was lost between dimensions. Also, much potential exists to further enhance the capabilities of the system through optimization of both

the anion exchange and UHP-RPLC separations. Although the peak capacity of this system still falls substantially short of the capabilities of 2D gel electrophoresis, it is free from many of the cumbersome problems that plague gel-based separations. As further refinements are made, it is expected that multidimensional LC using ultra-high pressures will become a practical technique for separation of complex samples.

From a proteomic standpoint, the results of the study were somewhat less encouraging. The total number of probable unique proteins detected – 422 using the long anion exchange column – is only approximately 10% of the roughly 4400 proteins predicted by the *E. coli* genome. It is worth noting that other more exhaustive studies of the *E. coli* proteome have successfully characterized only about 1600 *E. coli* proteins.[26, 27] Therefore, it is possible that some proteins predicted by the genome are not transcribed or are produced in such small quantities as to render them undetectable to presently available techniques. Still, the present study only detected about 25% of those proteins that have been characterized previously. There are several factors that may contribute to why this percentage is low. First, the technique by which proteins were extracted from *E. coli* was intended to recover only proteins in the cytosol; thus all membrane-bound proteins would be excluded from the sample. Secondly, electrospray MS favors detection of small proteins; proteins above 30 kDa may be under-represented in the proteins detected in this study. Perhaps the most significant challenge, however, is the fact that the concentration of different proteins in the sample spans many orders of magnitude. To see the less abundant proteins requires injecting a large amount of sample, but doing so causes column overloading with the most abundant proteins. To remedy this situation would require a mass spectrometer with improved

sensitivity and dynamic range, a column with higher loading capacity, or a 2D separation

with a peak capacity high enough to compensate for the effects of overloading.

It is also important to note that none of the proteins detected in this study were

identified. For true proteomics, this is a critical limitation. Various techniques could be

applied to attempt to identify the proteins. Top-down proteomics involving gas-phase

fragmentation would be possible; however, it generally requires the use of a mass

spectrometer with higher resolution than the TOF analyzer used in these studies. An

alternative is to incorporate a step involving enzymatic digestion followed by MS/MS and

database searching into or after the intact protein LC x LC separation. This concept of a

hybrid top down / bottom up approach to proteomics will be the focus of the following

chapter.

3.5   References

(1)   Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196A-203A.

(2)   MacNair, J. E.; Lewis, K. C.; Jorgenson, J. W. *Analytical Chemistry* **1997**, *69*, 983-989.

(3)   MacNair, J. E.; Patel, K. D.; Jorgenson, J. W. *Analytical Chemistry* **1999**, *71*, 700-708.

(4)   Jerkovich, A. D.; Mellors, J. S.; Jorgenson, J. W. *LC-GC North America* **2003**, *21*, 600-610.

(5)   Mellors, J. S.; Jorgenson, J. W. *Analytical Chemistry* **2004**, *76*, 5441-5450.

(6)   Patel, K. D.; Jerkovich, A. D.; Link, J. C.; Jorgenson, J. W. *Analytical Chemistry* **2004**, *76*, 5777-5786.

(7)   Majors, R. E. *LC-GC Europe* **2003**, *16*, 8-13.

(8)   Shen, Y.; Zhang, R.; Moore, R. J.; Kim, J.; Metz, T. O.; Hixson, K. K.; Zhao, R.; Livesay, E. A.; Udseth, H. R.; Smith, R. D. *Analytical Chemistry* **2005**, *77*, 3090-3100.

(9)   Anspach, J. A.; Maloney, T. D.; Brice, R. W.; Colon, L. A. *Analytical Chemistry* **2005**, *77*, 7489-7494.

(10)   Giddings, J. C. *Journal of High Resolution Chromatography & Chromatography Communications* **1987**, *10*, 319-323.

(11)   Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64-71.

(12)   Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T. *Analytical Chemistry* **1991**, *63*, 1193A-1203A.

(13)   Gaskell, S. J. *Journal of Mass Spectrometry* **1997**, *32*, 677-688.

(14)   Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *Journal of Proteome Research* **2006**, *5*, 135-146.

(15)   Link, J. C., University of North Carolina at Chapel Hill, Chapel Hill, 2004.

(16)   Eschelbach, J. W.; Jorgenson, J. W. *Analytical Chemistry* **2006**, *78*, 1697-1706.

(17)   Ferrige, A. G.; Seddon, M. J.; Green, B. N.; Jarvis, S. A.; Skilling, J. *Rapid Communications in Mass Spectrometry* **1992**, *6*, 707-711.

(18)   Kastner, M. *Protein Liquid Chromatography*; Elsevier: Amsterdam, Netherlands, 2000.

(19)   Cohen, K. A.; Schellenberg, K.; Benedek, K.; Karger, B. L.; Grego, B.; Hearn, M. T. W. *Analytical Biochemistry* **1984**, *140*, 223-235.

(20)   Cohen, S. A. *Analytical Chemistry* **1984**, *56*, 217-221.

(21)   Cohen, S. A.; Benedek, K.; Tapuhi, Y.; Ford, J. C.; Karger, B. L. *Analytical Biochemistry* **1985**, *144*, 275-284.

(22)   Maltsev, V. G.; Nasledov, D. G.; Trushin, S. A.; Tennikova, T. B.; Vinogradova, L. V.; Volokitina, I. N.; Zgonnik, V. N.; Belenkii, B. G. *Journal of High Resolution Chromatography* **1990**, *13*, 185-192.

(23)   Stout, R. W.; Sivakoff, S. I.; Ricker, R. D.; Snyder, L. R. *Journal of Chromatography* **1986**, *353*, 439-463.

(24)   Gilar, M.; Daly, A. E.; Kele, M.; Neue, E. D.; Gebler, J. C. *Journal of Chromatography A* **2004**, *1061*, 183-192.

(25)   Tran, J. C.; Doucette, A. A. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 652-656.

(26)   Han, M.-J.; Lee, S. Y. *Microbiology and Molecular Biology Reviews* **2006**, *70*, 362-439.

(27)   Corbin, R. W.; Paliy, O.; Yang, F.; Shabanowitz, J.; Platt, M.; Lyons, C. E., Jr.; Root, K.; McAuliffe, J.; Jordan, M. I.; Kustu, S.; Soupene, E.; Hunt, D. F. *Proceedings of the National Academy of the Sciences of the United States of America* **2003**, *100*, 9232-9237.

3.6   Tables

Anion exchange column	*E. coli* strain analyzed	Anion exchange gradient length	# of probable unique proteins found	# of fractions with probable unique proteins
Short	MG1655	30 min	209	15
Short	MG1655	60 min	247	21
Short	MG1655	120 min	176	31
Long	DH5α	360 min	422	47

Table 3-1: Summary of data from four LC x UHPLC separations of an *E. coli* protein extract

3.7   Figures
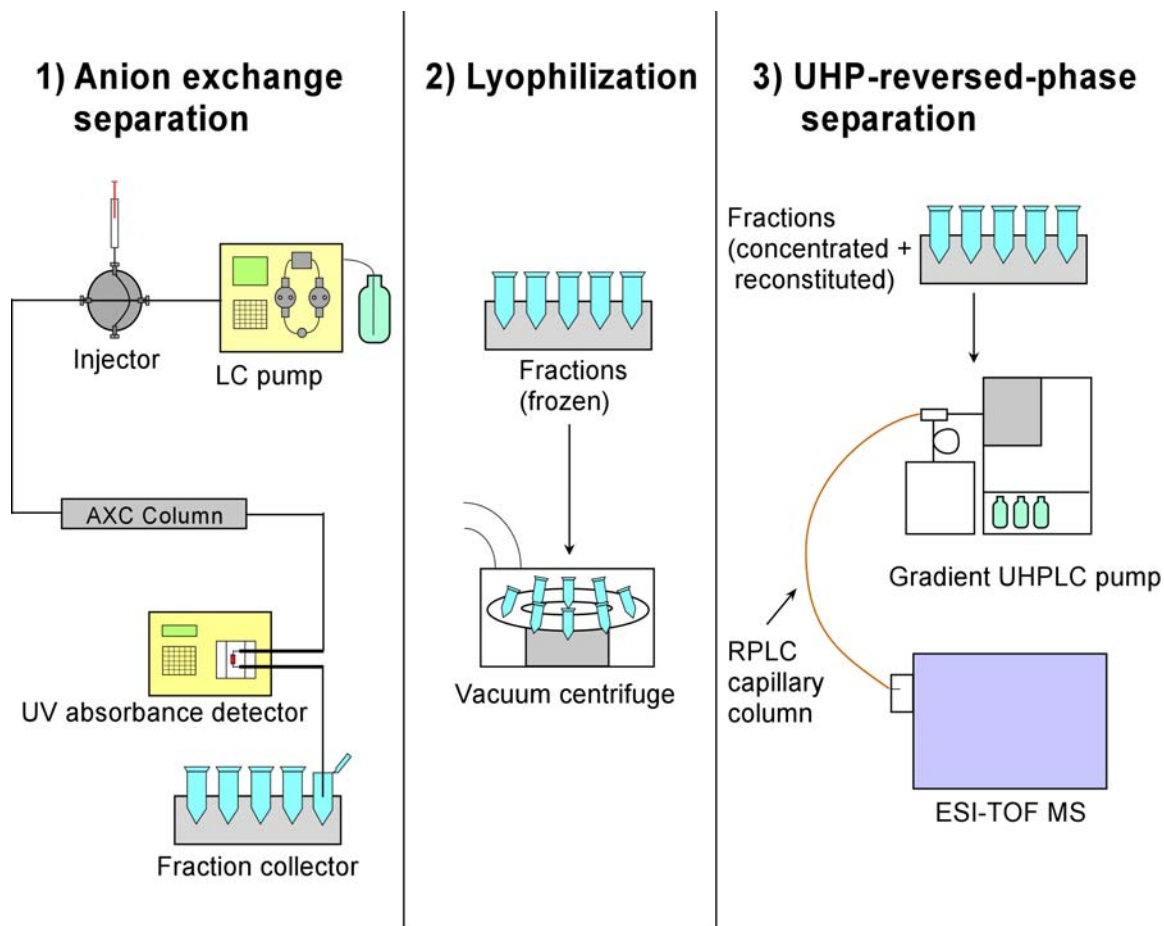
Figure 3-1: Overview of instrumentation and procedure used for off-line LC x UHPLC-MS separations
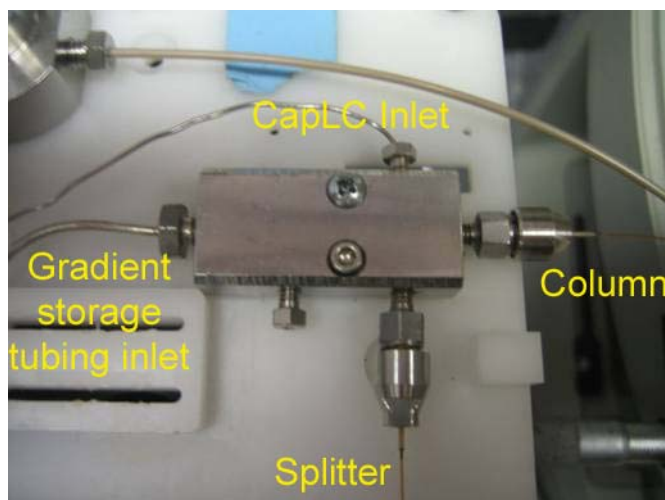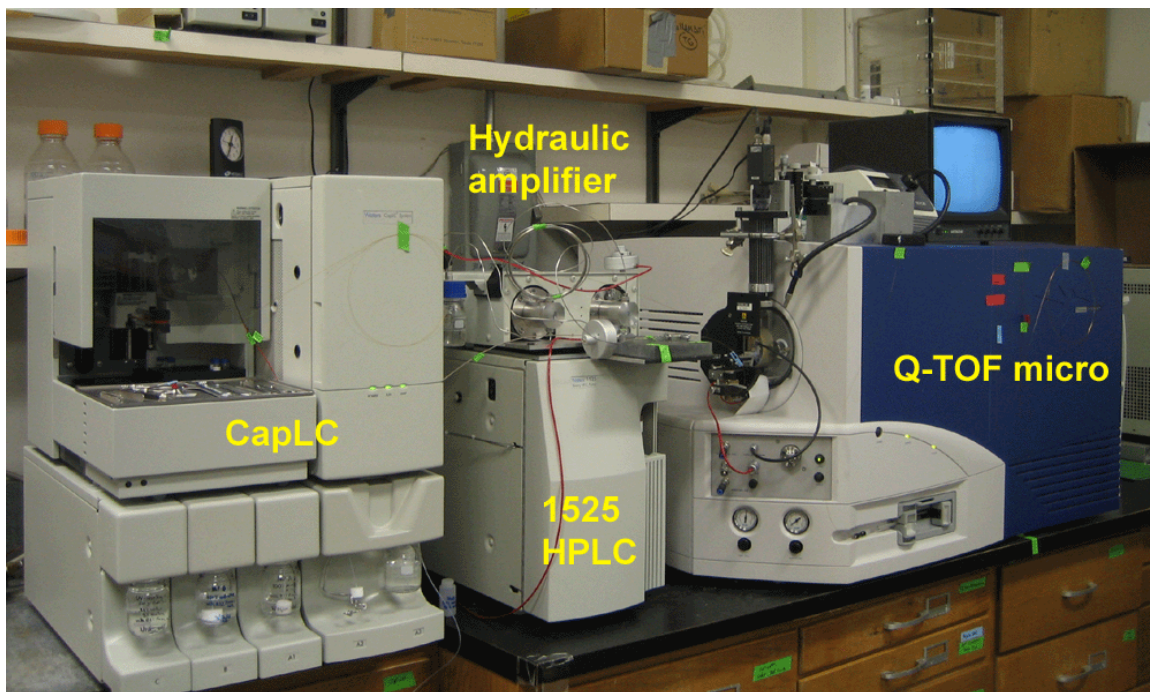
Figure 3-2: Photographs of the pre-loaded gradient UHPLC instrument. (A) shows the entire instrument; (B) shows a close-up view of the 4-port high pressure union
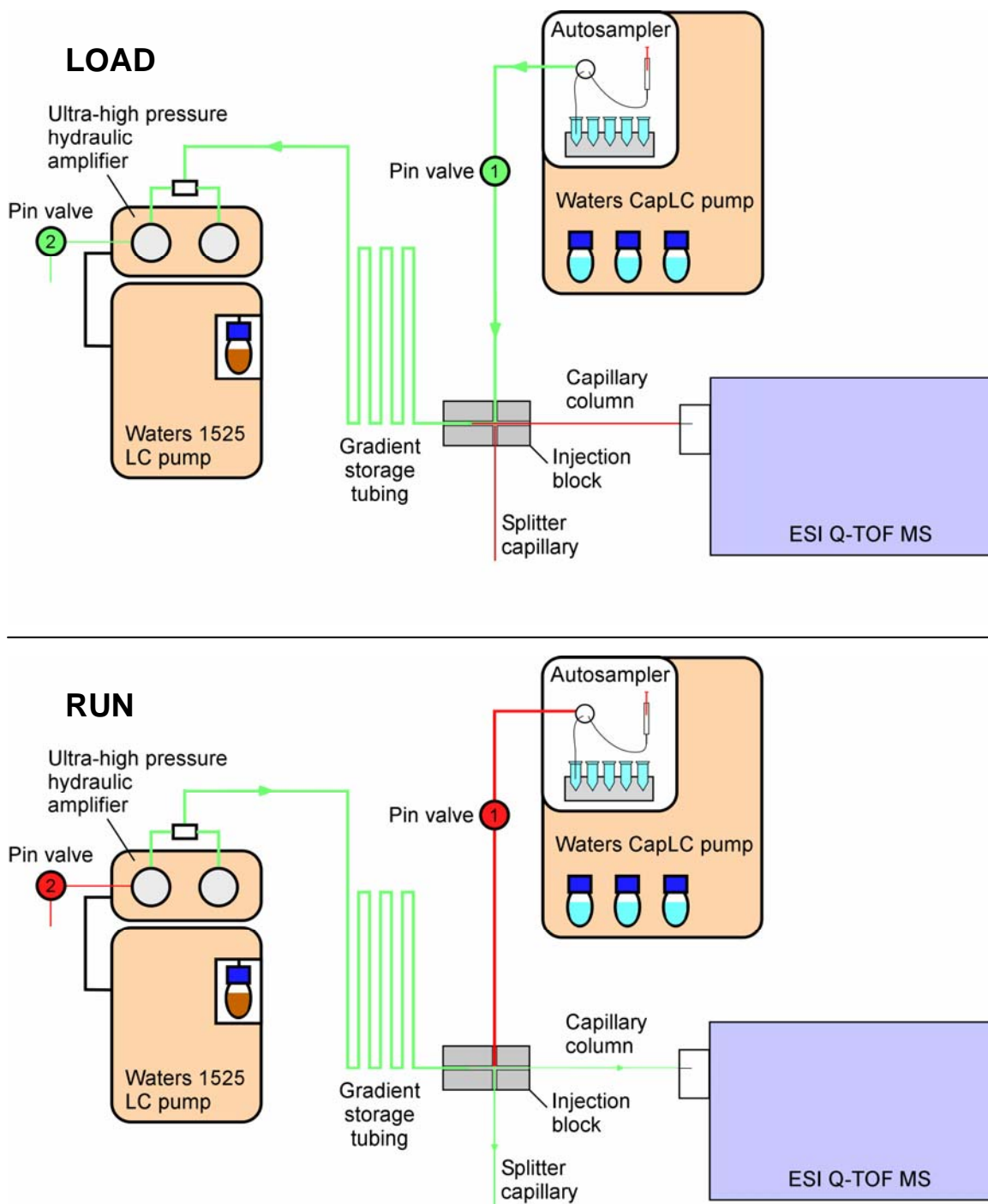
Figure 3-3: Diagram illustrating the operation of the preloaded gradient UHPLC system. In "LOAD" mode, both pin valves are open. The CapLC generates a gradient and injects the sample, both of which pass through the 4-port union (labeled "injection block") and are loaded into the gradient storage tubing. In "RUN" mode, both pin valves are closed and the hydraulic amplifier pump pressurizes the gradient storage tubing, forcing the pre-loaded sample and gradient onto the column. More details regarding the operation of the instrument are found in the text.
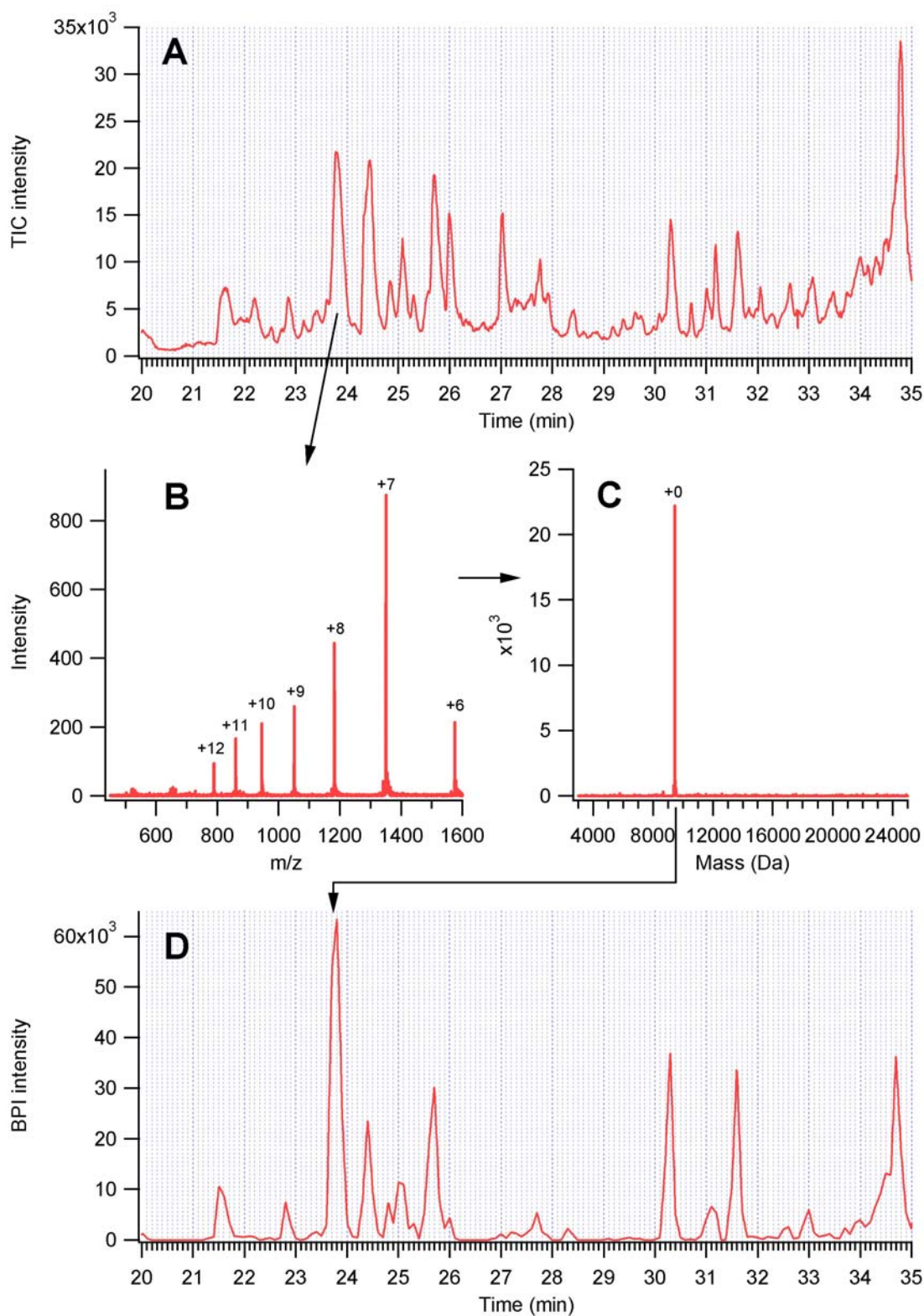
Figure 3-4: Process of generating a BPI chromatogram for intact proteins using maximum entropy de-convolution. Full details are explained in Section 3.2.7 of the text.
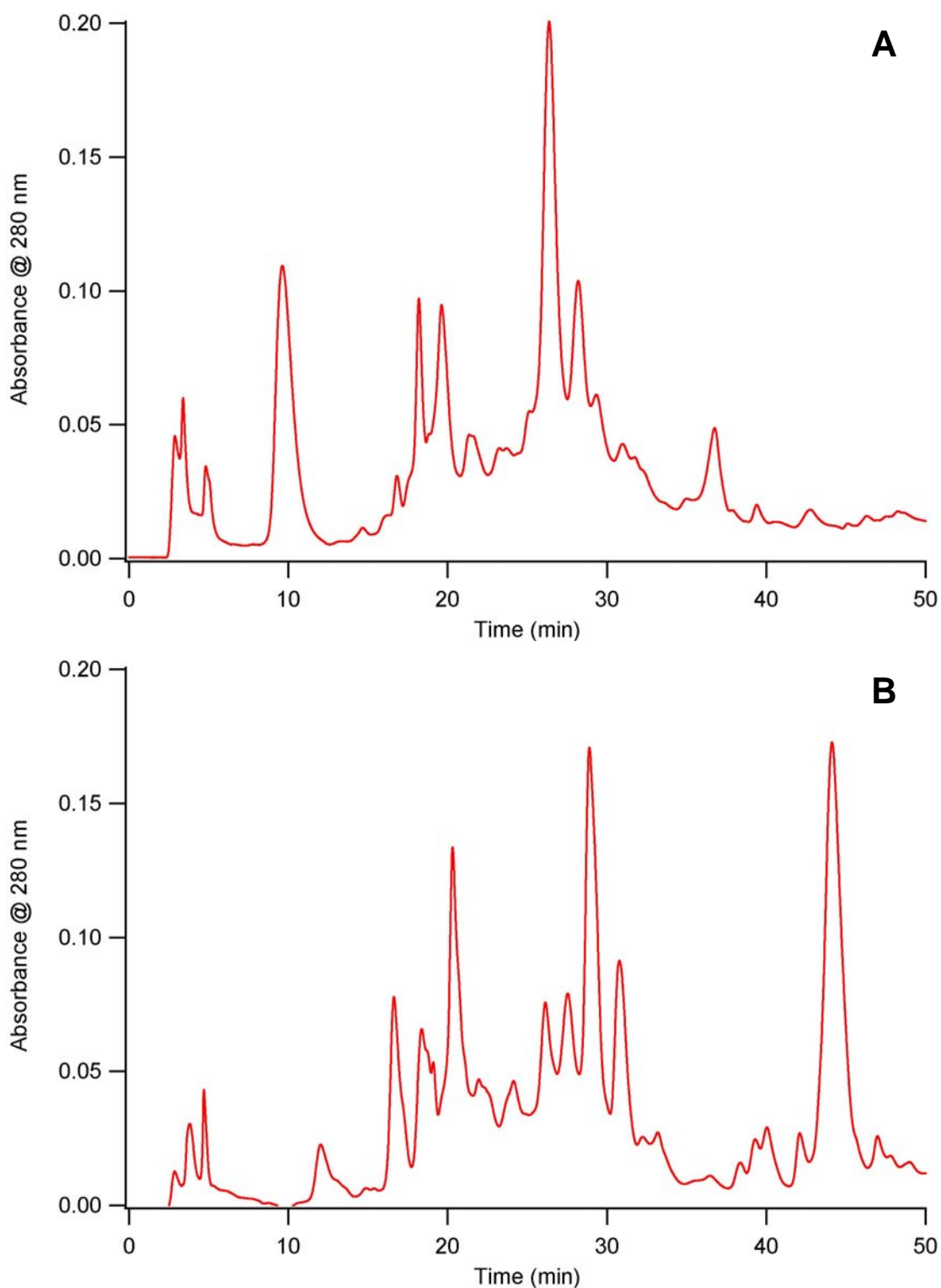
Figure 3-5: Comparison of anion exchange separations of *E. coli* protein extracts prepared from strain MG1655 wild type (A) and strain DH5α (B). Both samples were separated on the short AXC column using a 30 minute gradient from 25 to 500 mM ammonium acetate.
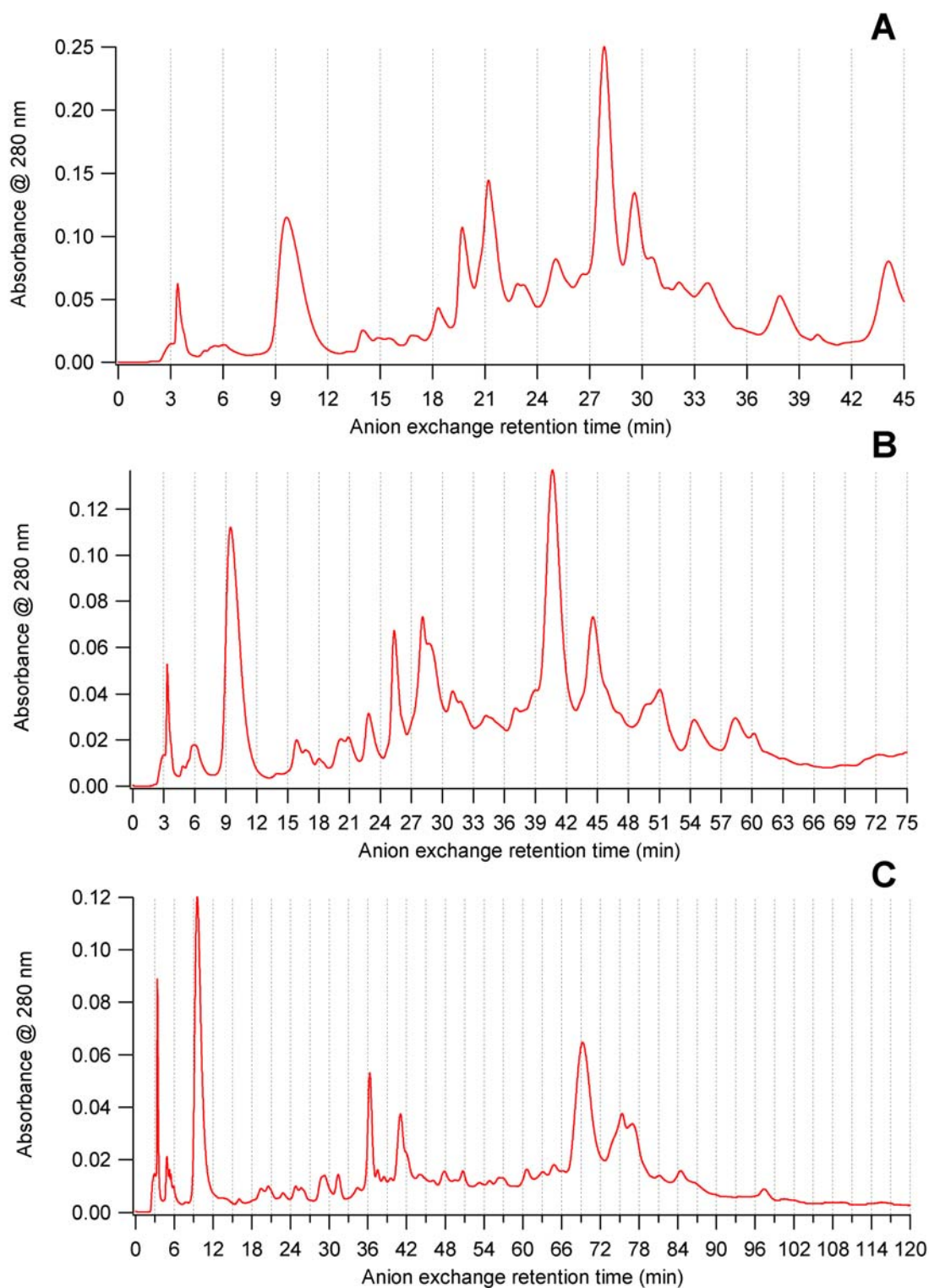
Figure 3-6: Anion exchange separations of the *E. coli* protein extract on the short AXC column. The separations were carried out using a gradient from 25 to 500 mM ammonium acetate over 30 min (A), 60 min (B), and 120 min (C). Fractions were switched at intervals indicated by the dashed vertical lines on the chromatograms.
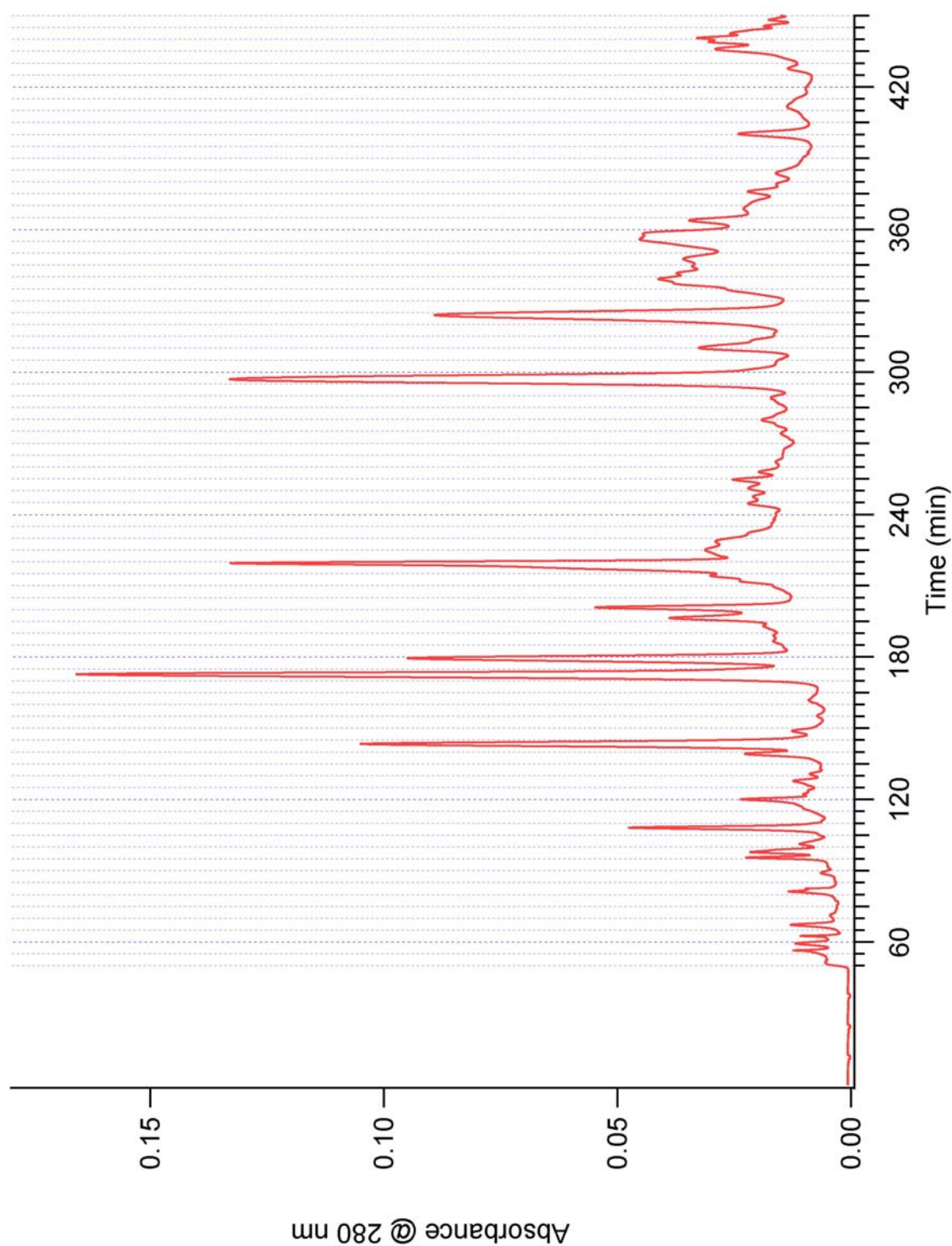
Figure 3-7: Anion exchange separation of the *E. coli* protein extract on the long AXC column. The separation was carried out using a gradient from 25 to 500 mM ammonium acetate over 360 min.
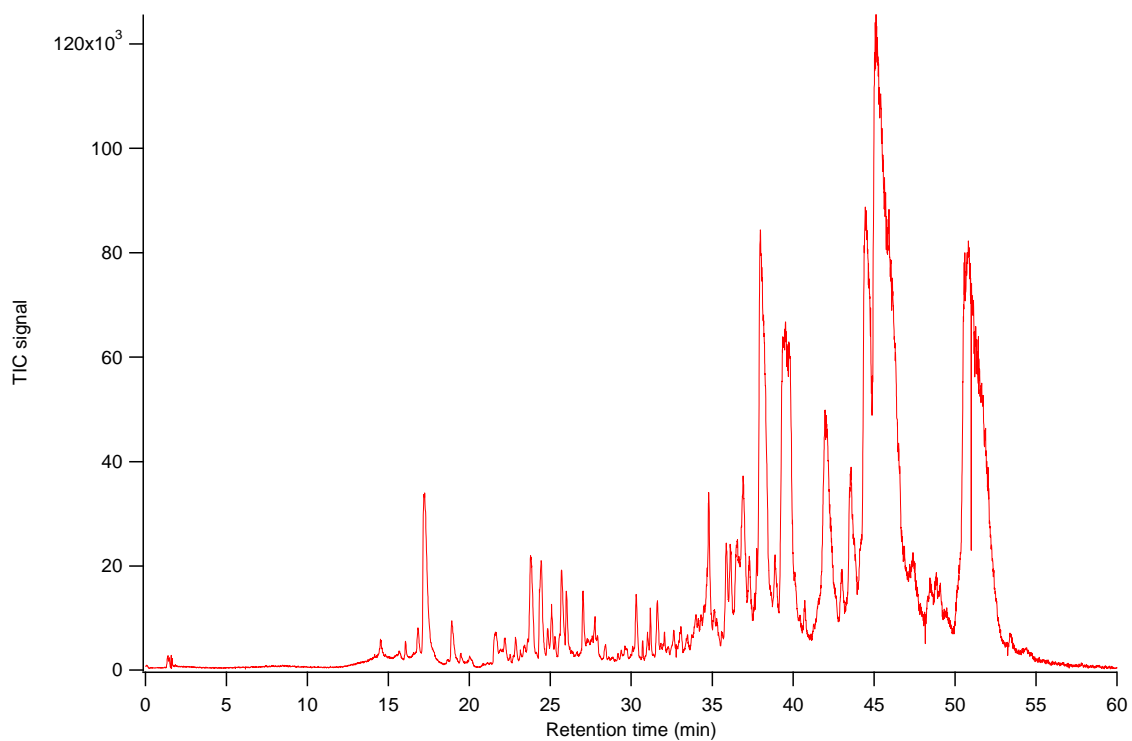
Figure 3-8: UHP-RPLC separation of one anion exchange fraction of an *E. coli* protein extract on a capillary LC column using nano-ESI-MS detection. The gradient used was a linear ramp from 10 to 65% acetonitrile (with 0.2% formic acid throughout) over 60 minutes.
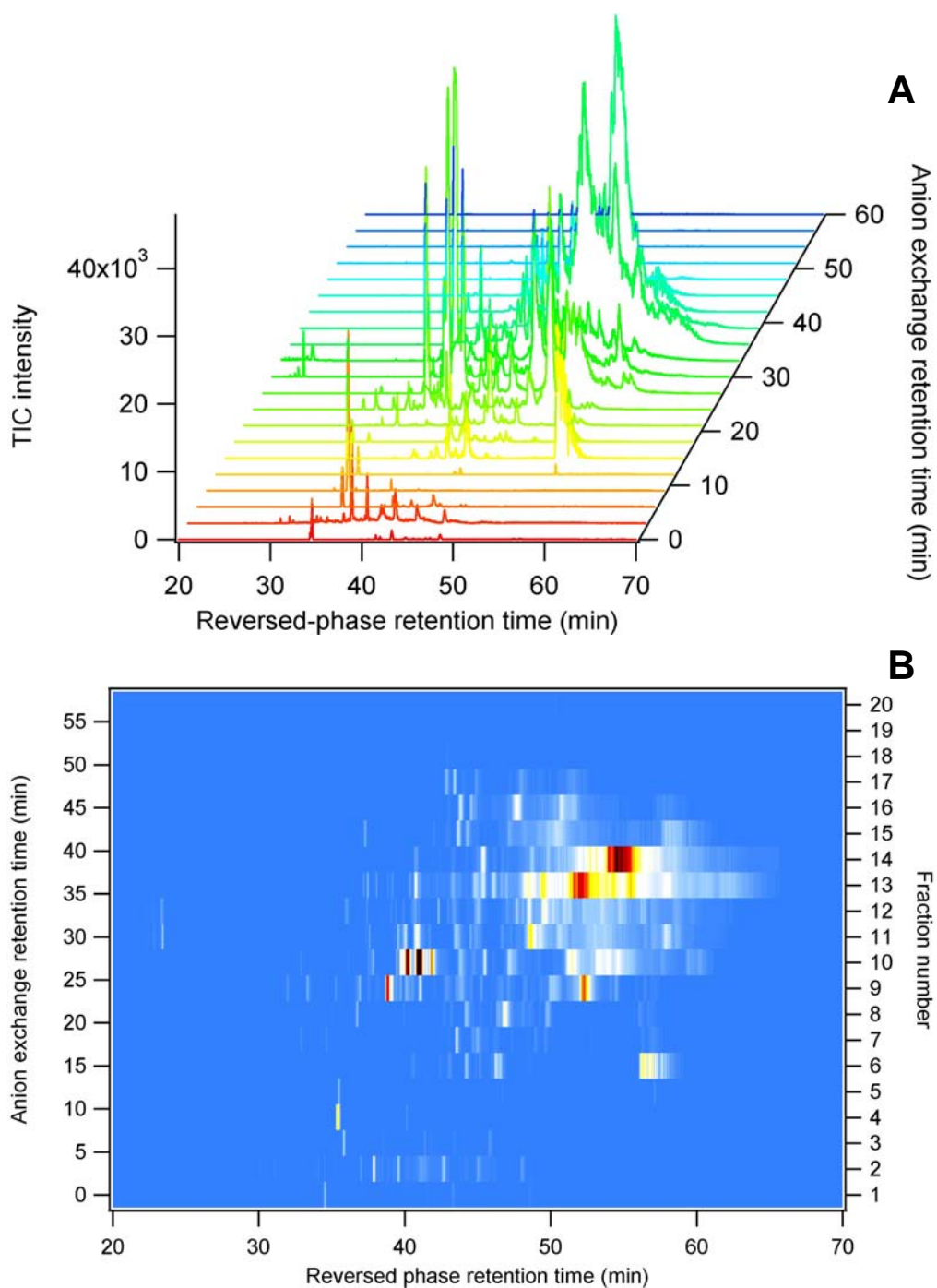
Figure 3-9: LC x LC data displayed in different formats. In (A), 1D chromatograms are displayed side-by-side with a horizontal and vertical offset, which gives a rudimentary 3D visualization of the separation. In (B), the same data is displayed on a planar 2D surface, and intensity is represented by a color scale as opposed to peak height.
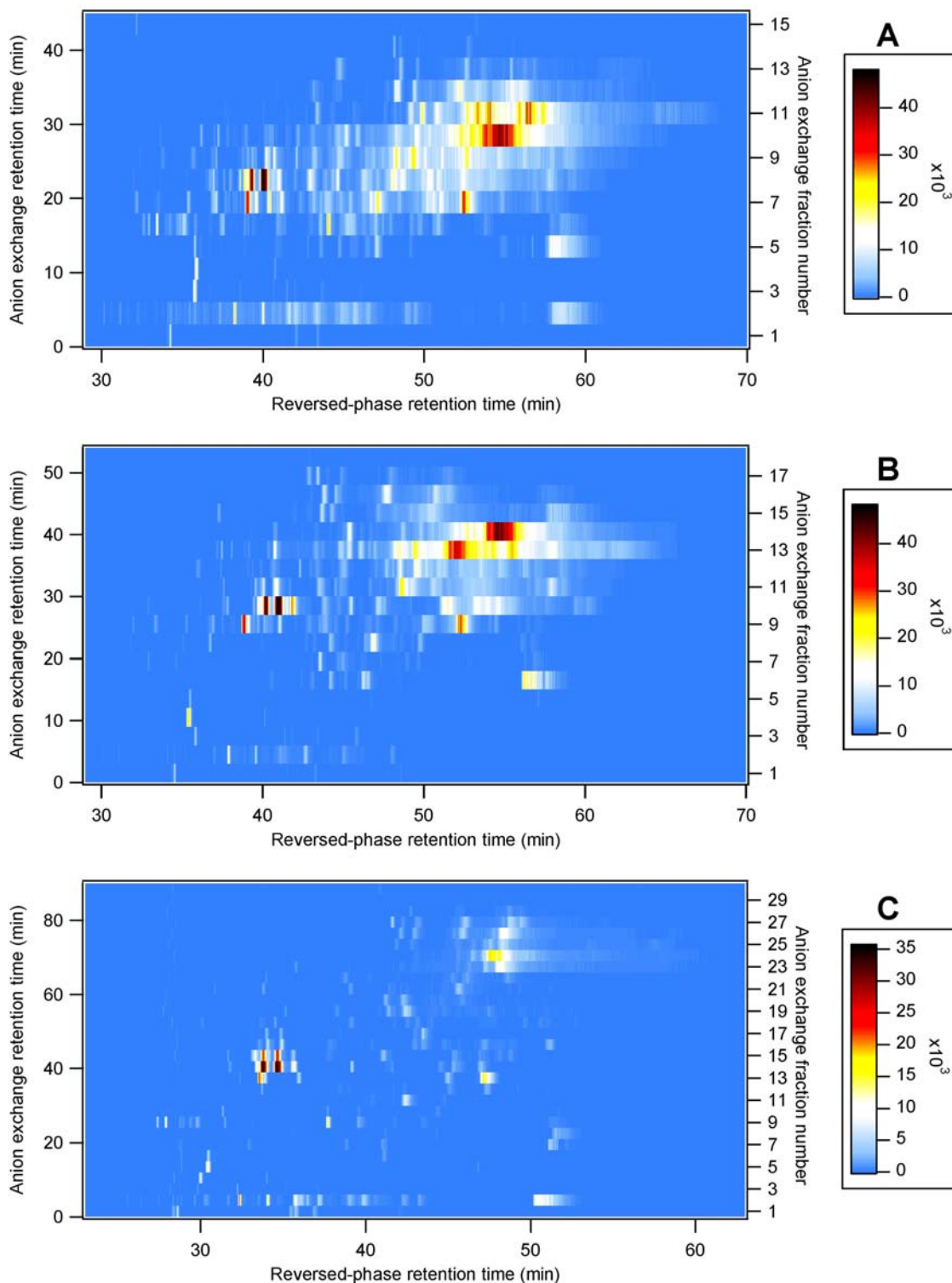
Figure 3-10: 2D chromatograms of the *E. coli* extract generated using the short AXC column with three gradients. All separations used a linear gradient from 25-500 mM ammonium acetate. The gradient length was 30 min for (A), 60 min for (B) and 120 min for (C). Peak intensity is based on TIC signal.
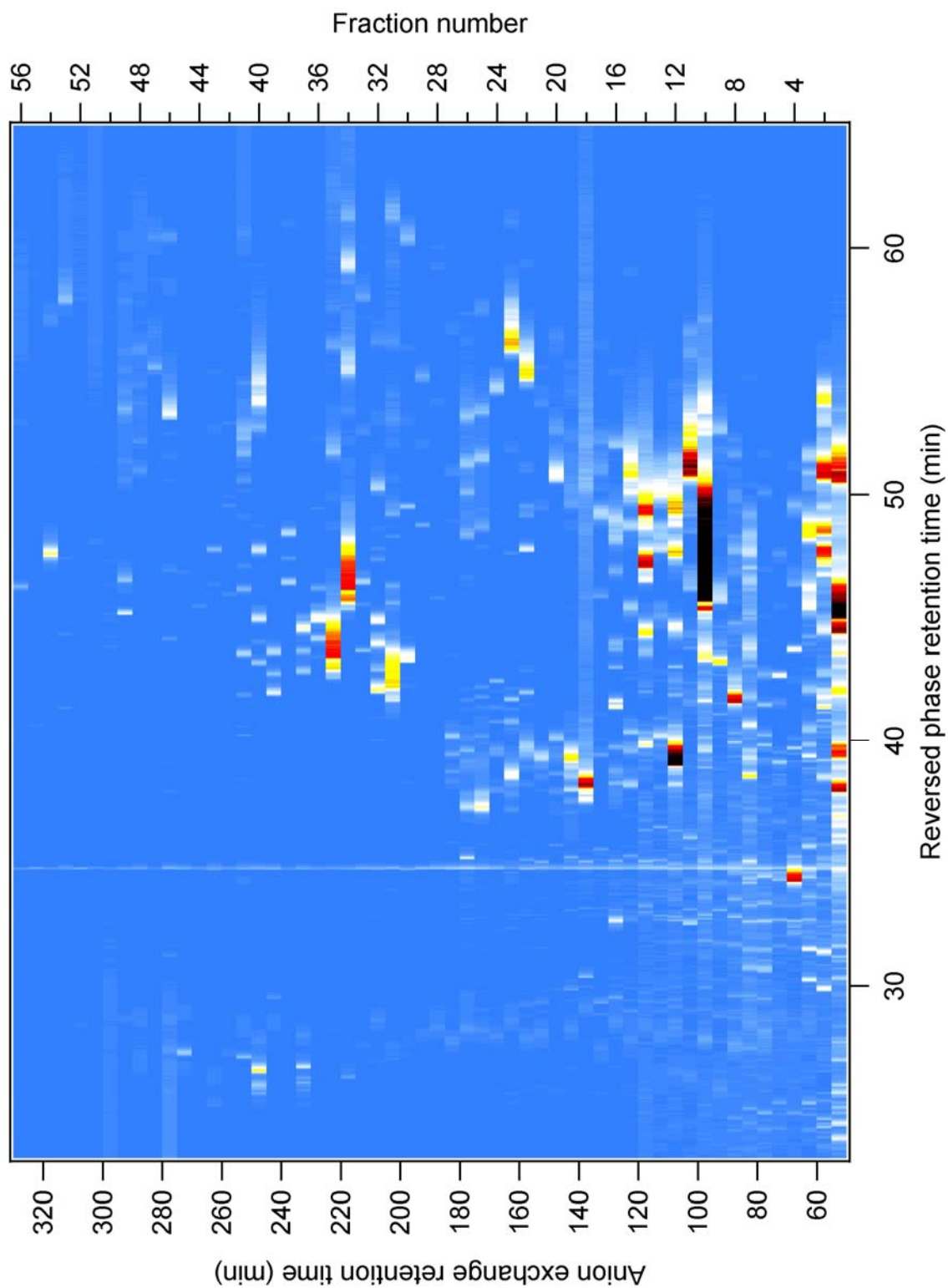
Figure 3-11: Total ion current (TIC) 2D chromatogram of the *E. coli* extract separated using the long AXC column. The gradient was a linear ramp from 25-500 mM ammonium acetate over 360 minutes.
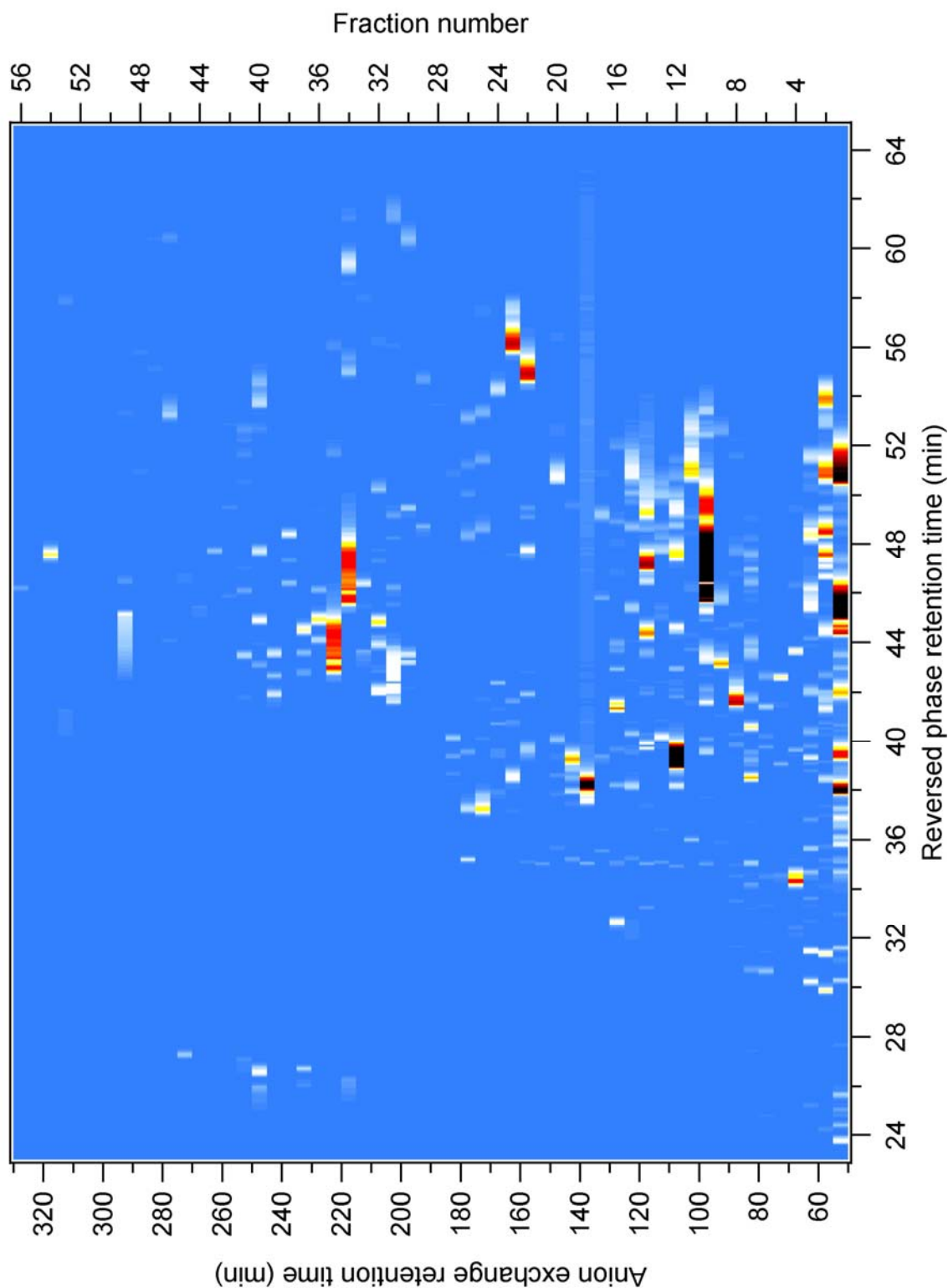
Figure 3-12: Base peak intensity (BPI) 2D chromatogram of the *E. coli* extract separated using the long AXC column. The chromatogram is from the same separation as Figure 3-11, except data were de-convoluted using AutoME, and the base peak intensity of components within the mass range of 3,000-60,000 Da were plotted.

118
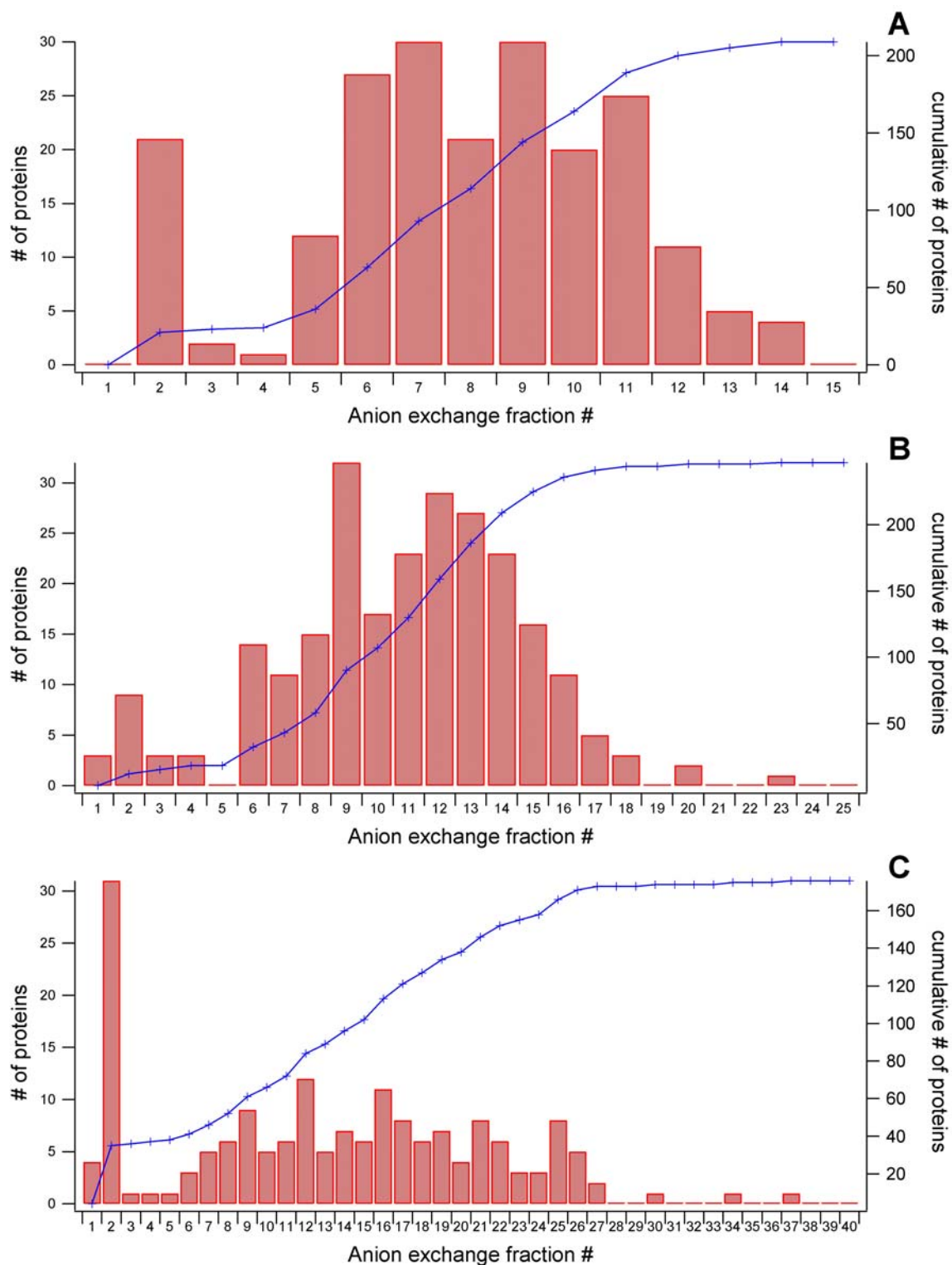
Figure 3-13: Number of proteins found in each anion exchange fraction for 2D separations of the *E. coli* protein extract using the short AXC column. Plots (A), (B), (C) represent the 30, 60, and 120 minute gradients, respectively.
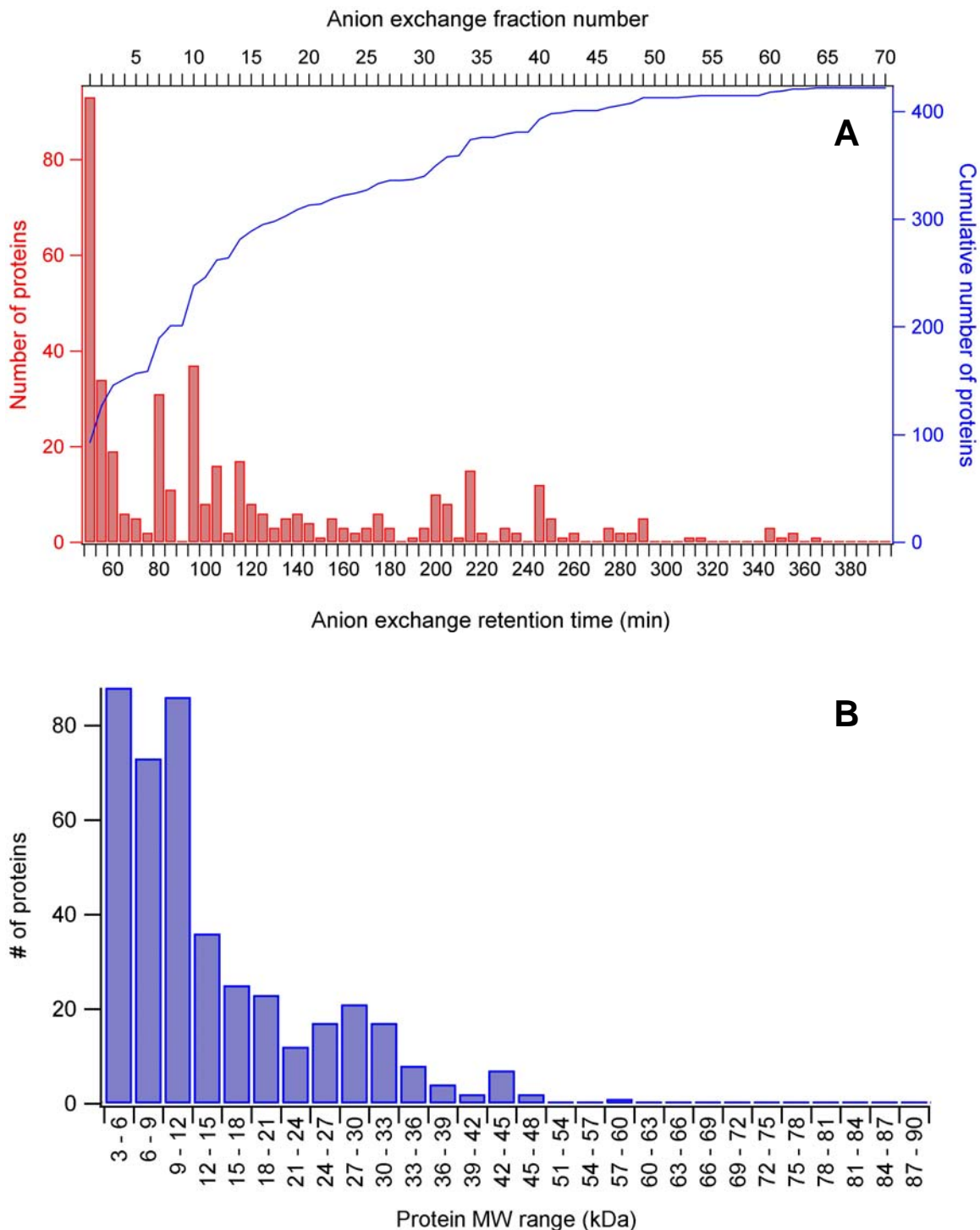
Figure 3-14: (A) Number of proteins found in each anion exchange fraction for 2D separation of the *E. coli* protein extract using the long AXC column; (B) MW distribution histogram for proteins found for the 2D separation using the long AXC column

120
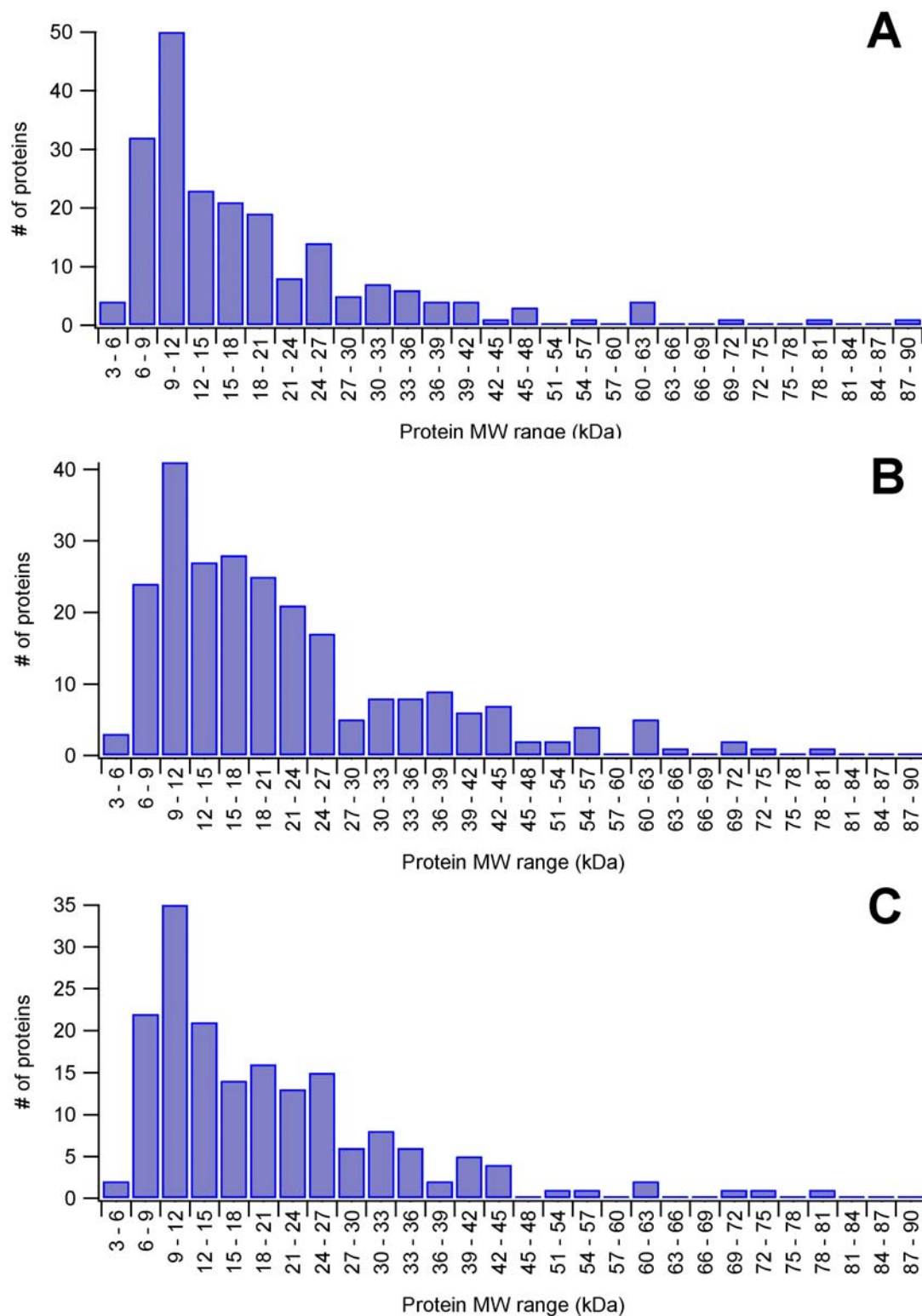
Figure 3-15: MW distribution histograms for proteins found in the *E. coli* protein extract for the 2D separation using the short AXC column. Plots (A), (B), (C) represent the 30, 60, and 120 minute gradients, respectively.

CHAPTER 4: Analysis of *E. coli* proteins and protein digests using off-line LC x UHPLC-MS: a hybrid top-down/bottom-up approach to proteomics

4.1   Introduction

In Chapter 3, a top-down technique for protein analysis using LC x UHPLC-MS was

reported. Although the separation generated high peak capacity, a major disadvantage was

that it was not possible to identify the proteins which were detected because only their intact

molecular weight was measured using MS. In this chapter, a method is presented which uses

top-down intact-protein analysis, but also allows proteins to be identified by incorporating

elements of a bottom-up approach. This hybrid strategy is then applied to analysis of the *E.*

coli proteome.

4.1.1   Bottom-up proteomics: overview

Before presenting research related to combined top-down / bottom-up proteomics, it

is important to have a well-established theoretical understanding of both methods as

independent techniques. Up to this point in the dissertation, substantial emphasis has already

been placed on top-down proteomics, since all of the separations and mass spectra presented

have pertained to intact proteins. Bottom-up proteomics[1, 2] – which is also commonly termed

shotgun proteomics – has only been explained in general terms. Since it plays a critical part

in the research presented in this chapter, it is important to introduce some of its theoretical

and practical aspects.

Like a top-down experiment, bottom up proteomics begins with a sample containing a

mixture of proteins, which is subjected to one or more stages of separation and is analyzed

using mass spectrometry. The distinguishing feature which all bottom-up analyses have in common is that, at some point before the sample is analyzed by MS, the proteins are enzymatically digested to produce peptides. Once peptide MS data are acquired, they are analyzed and used to identify the proteins in the original sample.

There are two main methods used to identify proteins from peptide MS data: peptide mass fingerprinting (PMF) and tandem mass spectrometry (MS/MS).[2] A diagram showing an overview of a PMF experiment is shown in Figure 4-1. In this example, the sample consists of a single protein which was digested to peptides using trypsin. The sample is separated using RPLC coupled to ESI-TOF-MS, which produces the chromatogram shown in Figure 4-1A. The mass spectra of individual peaks are then examined (Figure 4-1B), and the m/z and charge state of all detected components are recorded in a peak list table (Figure 4-1C). The peak list is input into a database searching program (Figure 4-1D), which compares the peptide masses from the MS data with calculated peptide masses which would be expected if known proteins stored in the database were digested using trypsin. Based on the comparison, the program predicts the identity of the protein(s) which were present in the sample. Some limitations of PMF include the fact that many peptides must be detected from each protein to ensure accurate identification, and that it works best with samples containing only a single protein or a relatively simple mixture.

The alternative to PMF is tandem MS (MS/MS). An overview of this method is shown in Figure 4-2. As in PMF, a sample containing enzymatically digested protein(s) is separated using LC (Figure 4-2A) and mass spectra of the peptides are obtained (Figure 4-2B). Unlike PMF, multiple stages of mass spectrometry are also performed during the course of the analysis. Individual peptide ions are selected using a mass analyzer and are

then subjected to fragmentation via collision induced dissociation (CID) or some other method. Fragmentation typically occurs predominantly along the peptide backbone, although the mechanism can be complex and has been reviewed in detail elsewhere.[3] Mass spectra of the fragment ions (Figure 4-2C) allow can then be compared with a database containing known protein sequences using a probability-based searching method in order to identify the proteins in the sample (Figure 4-2D). MS/MS has the advantage of being a more selective process than PMF; thus proteins can be identified based on mass spectra of just a few – or even one – peptide. It also is generally more successful than PMF with mixtures of greater complexity. A disadvantage of MS/MS is that the duty cycle is low since the instrument is constantly being cycled between MS survey and MS/MS modes.

4.1.2   Top-down and bottom-up proteomics: complementary techniques

Although it is a well-established method of identifying proteins, there are substantial disadvantages inherent to all forms of bottom-up proteomics. For one, the complexity of the sample is increased when an already complex mixture of proteins is digested to peptides. This makes the separation more difficult. Also, sequence coverage of any given protein is often low. Thus, any unknown post-translational modifications on portions of the protein which were not detected will not be discovered.

Top-down proteomics offers an advantage over bottom-up in that it allows the intact molecular weight of all proteins to be measured, which is useful for study of protein processing or detection of post-translational modifications.[4] Its major problem, however, is that identification of proteins is not often straightforward. Identification has been demonstrated by using MS/MS of intact proteins to search for patterns of fragmentation which can uniquely identify a protein. However, high-cost mass analyzers such as an FTICR[5] or Orbitrap[6] are required to attain the needed high resolution and mass accuracy. An

alternative is to use a charge-state reduction method, in which ion-ion reactions are used to convert multiply charged ions to singly charged ions.[7, 8] This allows intact protein MS/MS to be performed using lower resolution mass analyzers, such as a quadrupole ion trap. Although this method has gained some acceptance, further development will be required to make charge state reduction a practical mainstream technique. Thus, protein identification using the top-down approach remains difficult.

Clearly, there are distinct limitations associated with both top-down and bottom-up proteomic methods. Fortunately, the weaknesses of one technique tend to be strengths of the other. Simpler separations and better ability to detect post-translational modifications are advantages of the top-down approach, whereas well-established methods of identifying proteins favor bottom-up. Therefore, it is logical to consider using both methods in order to gain a more complete picture of the proteome being studied. Performing two entirely separate complex analyses on a single sample is inefficient, however. Thus, development of a method in which both top-down and bottom-up proteomics are integrated into a single analysis is a worthwhile goal.

4.1.3   Prior work in combining top-down and bottom-up workflows

In the past, selecting between top-down and bottom-up was treated as an "either-or" choice. As the field of proteomics has developed, however, there has been increased recognition of the complementary nature of the two techniques. One of the first steps toward combining elements of both techniques was separation of the sample into fractions at the intact protein level preceding digestion. A wide variety of techniques have been used to fractionate intact proteins before bottom-up analysis, including gel electrophoresis,[9] off-gel electrophoresis,[10] and many different modes of chromatography.[11-13] Multi-dimensional fractionation, including LC x LC, has also been used.[14] Fractionation reduces the complexity

of the protein mixture to be digested, which improves the outcome of the peptide separation and enhances the likelihood of successful protein identification using MS or MS/MS. Although protein fractionation incorporates the separation aspect of the top-down approach, intact proteins are not introduced into a mass spectrometer. Since all information about the identity and composition of the proteins is obtained from mass spectra of peptides, techniques which use protein fractionation before digestion are still bottom-up analyses.

Several examples of true hybrid top-down / bottom-up proteome analyses have been reported in the literature.[15-20] Unlike bottom-up methods that use protein fractionation, these involve acquiring mass spectra of both intact proteins and protein digests. One of the first studies demonstrating the use of this strategy for the characterization of the complete proteome of an organism was published in 2002 by VerBerkmoes et al.[15] Their research pertained to proteins extracted from *Shewanella oneidensis*, a metal-reducing bacterium of interest in the field of bioremediation. Various stages of centrifugation were used to separate membrane and cytosolic proteins. The proteins were then fractionated using anion exchange chromatography. A portion of each fraction was analyzed directly using FTICR-MS, while a second portion was digested using trypsin and then analyzed using LC-MS/MS. A total of 868 proteins were identified using the bottom-up approach, though only 70 proteins were detected using top-down, and only 20 of these were identified. The relatively low number identified by the top-down approach was attributed to matrix ionization effects in performing ESI-MS of the fractions, which each contain numerous proteins.

Berger et al. reported a further refined hybrid top down / bottom up technique which they termed "middle-out" proteomics.[18, 19] In their approach, intact proteins were fractionated using on-line LC x LC, with anion exchange as the first dimension and reversed-

phase as the second. Using a flow split, a portion of the second dimension effluent was directed to an ESI-TOF mass spectrometer to measure intact protein MW, while the majority was diverted to a fraction collector. The fractions were digested using trypsin and analyzed using MALDI-Q-TOF-MS/MS. The technique was applied to both yeast ribosome proteins[19] and to *E. coli* cytosol.[18] In the yeast ribosome protein study, approximately 85% of the known 80 unique ribosomal subunit proteins were detected by either the top down or bottom up approach. In the study of *E. coli* cytosol – a much more complex sample – a total of 46 proteins were detected and identified by top-down and bottom-up approaches, 55 were identified using bottom-up only, and 57 were detected using top-down alone. In both studies, numerous post-translational modifications were identified on many of the proteins found, including N-terminal methionine removal or addition, N-terminal acetylation, and phosphorylation.

Most recently, Sharma et al. reported a proteomic profiling method for *S. oneidensis* using AXC fractionation.[20] One portion of the fractions was analyzed using RPLC-FTICR-MS, while another portion was digested with trypsin and analyzed using RPLC-ion trap MS/MS. A total of 715 intact proteins were detected using the top-down approach, and 447 proteins were identified using the bottom-up peptide analysis. Approximately 12% of the intact proteins were identified by correlating the intact protein MW with the predicted MW of proteins identified from the bottom up approach, after considering mass shifts due to PTMs.

4.2 Experimental methods

4.2.1 Overview

Figure 4-3 gives an overview of the instrumentation and procedure used to perform a hybrid top-down / bottom up analysis of the *E. coli* cytosol proteome. The sample, which

contains a complex mixture of proteins, is separated into fractions using anion exchange chromatography. The fractions are then lyophilized and reconstituted in a smaller volume of buffer. The reconstituted fractions are divided into two equal parts. One part of each fraction is separated and analyzed "as is" using gradient ultra-high pressure RPLC-MS. This provides intact protein molecular weight data. The second part of each fraction is digested using trypsin to produce a mixture of peptides. The peptides are then separated and analyzed, also using UHP-RPLC-MS. Proteins are identified by searching the acquired peptide LC-MS data against the SwissProt database.

4.2.2 Samples and reagents

Reagents used for LC mobile phase were ammonium acetate and formic acid (ACS reagent grade), which were purchased from Sigma-Aldrich (St. Louis, MO), and ammonium hydroxide (ACS reagent grade) and acetonitrile (HPLC grade), which were purchased from Fisher Scientific (Fair Lawn, NJ). Deionized water was purified using a Barnstead Nanopure System (Boston, MA). Reagents used in the trypsin digest procedure were ammonium bicarbonate (Sigma-Aldrich, St. Louis, MO), RapiGest SF acid-labile surfactant (Waters Corporation, Milford, MA), dithiothreitol (Research Products International, Mt. Prospect, IL), iodoacetamide (Sigma-Aldrich), TPCK-modified trypsin (Pierce, Rockford, IL), and trifluoroacetic acid (Sigma-Aldrich).

The sample analyzed in this experiment was an extract of the soluble proteins of the bacterium *Escherichia coli.* It was prepared from *E. coli* strain DH5α (Q-BIOgene, Irvine, CA) by researchers at Waters Corporation using the procedure set forth by Millea et al.[18] The total protein concentration in the sample was measured as approximately 100 mg/mL using a modified Coomassie dye-binding assay (BioRad, Hercules, CA). Prior to injection,

the sample was diluted to 20 mg/mL in 10 mM ammonium acetate (pH 8.5). Approximately

4 mg of total protein was analyzed in the complete 2D separation run.

4.2.3   Anion exchange fractionation

The first dimension of the 2D separation was performed using a Waters Biosuite Q 10

μm anion exchange column, which is 7.5 cm long and has an ID of 7.5 mm. A Waters 600E

quaternary gradient LC pump was used to produce the anion exchange gradient. The pump is

connected to a Valco 6-port valve (VICI, Houston, TX) with a 200 μL sample loop, which is

used to perform injections onto the column. Detection was performed using an Applied

Biosystems 785A UV absorbance detector (Foster City, CA) set at 280 nm.

For anion exchange separations, mobile phase A was 10 mM ammonium acetate,

adjusted to pH 8.5 with ammonium hydroxide. Mobile phase B was 1M ammonium acetate,

adjusted to pH 8.5 with ammonium hydroxide. The gradient program is as follows: after

sample injection, a linear gradient from 0% to 50% B was run over 60 minutes, followed by a

linear ramp to 75% B over 5 minutes. The mobile phase composition was held at 75% B for

20 minutes, and then returned to 100% A.   The flow rate through the column was 0.5

mL/min.

After passing through the UV detector, the effluent from the anion exchange column

was sent to a Waters Fraction Collector II, which allowed automated collection of fractions

in polypropylene centrifuge tubes. Fractions were collected every 3 minutes, beginning as

soon as the sample was injected, for a volume of 1.5 mL per fraction. A total of 20 fractions

were collected. After collection, fractions were flash-frozen using liquid nitrogen and were

placed in a SpeedVac Concentrator (Thermo-Electron, Bellefonte, PA), which was pumped

down to pressures between 10^{-2} to 10^{-3} Torr using an Edwards double-stage rotary vacuum

pump (Wilmington, MA). Once the fractions had been lyophilized to dryness, they were

reconstituted in 50 μL of a solution of 50 mM ammonium bicarbonate in deionized water and vortexed. After all lyophilized protein had dissolved, 25 μL was transferred from each fraction to a corresponding Waters CapLC sample vial to be used for intact protein LC-MS analysis. 50 μL of additional ammonium bicarbonate buffer was added to these sample vials, which brought their total volume up to 75 μL. The 25 μL of the reconstituted anion exchange fractions remaining in the microcentrifuge tubes was left in place to be digested using trypsin.

4.2.4 Trypsin digest of fractions

The procedure for trypsin digestion of fractions was based upon the recommended in-solution digest protocol provided with the RapiGest SF acid-labile surfactant. The advantage of using RapiGest over standard protein denaturants, such sodium dodecyl sulfate or urea, is that RapiGest does not modify peptides or suppress endoprotease activity, and it can be easily removed after use by hydrolysis under acidic conditions.[21] Slight modifications were made to the standard procedure to account for the expected quantity of protein in each fraction. The exact procedure used is given below.

RapiGest SF powder was re-constituted by adding 150 μL of 50 mM ammonium bicarbonate to 1 mg of powder in a glass vial, then vortexing to mix. 3 μL of the RapiGest solution was added to each of the centrifuge tubes containing the re-constituted anion exchange fractions, and all tubes were vortexed. The fractions were incubated in a convection oven at 80º C for 15 minutes, vortexing periodically. The fractions were then centrifuged for 4 minutes at 12 krpm. 1 μL of 100 mM dithiothreitol was added to each fraction, and they were again vortexed to mix. They were incubated at 60º C for 30 minutes, and then centrifuged for 4 minutes at 12 krpm. 1 μL of 200 mM iodoacetamide was added to each fraction and vortexed. They were incubated in the dark at room temperature for 30

minutes, and were then centrifuged for 4 minutes at 12krpm. 10 μL of a 0.1 μg/μL solution

of TPCK-modified trypsin in 50 mM ammonium bicarbonate was added to each fraction and

vortexed. They were then incubated overnight at 37º C. The next morning, the fractions

were removed from the convection oven and were centrifuged at 12 krpm for 4 min. 40 μL

of deionized water and 4 μL of 10% trifluoroacetic acid were added to each fraction to

quench the enzymatic reaction and to hydrolyze the RapiGest. The fractions were vortexed

and then incubated at 37º C for 45 minutes. They were then centrifuged for 10 minutes at 13

krpm. The supernatant of each fraction, which totaled just over 80 μL, was transferred to

CapLC sample vials for LC-MS/MS analysis.

4.2.5   Intact protein separation using gradient UHP-RPLC-MS

4.2.5.1 LC instrumentation and run conditions

The second-dimension separations of intact proteins were performed using gradient

ultra-high pressure reversed phase liquid chromatography (UHP-RPLC). The instrument

used to perform the separation was the same as the pre-loaded gradient pump described in

section 3.2.5.1, with one exception. The high pressure 4-port union, where the column,

splitter capillary, CapLC pump and hydraulic amplifier pump are interfaced, was replaced

with a lower dead-volume version. A comparison of the old and new 4-port unions is shown

in Figure 4-4. The new union has a much smaller internal volume, which reduces the delay

between when a run is started and when the first peaks begin to elute from the column. The

parameters used for performing runs were identical to those specified previously, with the

exception of the UHPLC pump flow rate, which was increased to 4 μL/min. The column

used for the separations was a 50 μm ID capillary, 15 cm in length, packed with 1.8 μm

bridged-ethyl hybrid silica particles modified with a C18 stationary phase. The particles

have 300 Å pores, which are suited for use with large molecules such as intact proteins. The

column was operated at a pressure of approximately 1,600 bar (23,000psi). The flow rate through the column was approximately 400 nL/min, giving a split ratio of approximately 10:1.

For reversed phase separations of intact proteins, mobile phase A was deionized water with 0.2% (v/v) formic acid. Mobile phase B was acetonitrile with 0.2% (v/v) formic acid. Runs were performed using a linear gradient from 10% to 60% MP B over 30 minutes, followed by a 5 minute hold at 60%B, and then a return to initial conditions. The CapLC autosampler loaded 2.5 μL of sample onto the gradient storage loop for each run. Based on the split ratio, the volume of sample injected onto the column was approximately 250 nL.

4.2.5.2 MS instrumentation and run conditions

The outlet of the reversed-phase capillary column was coupled to a fused silica nano-electrospray emitter using a Waters Universal Nanoflow Sprayer. The sprayer has a zero dead volume union which allows the column to be coupled end-to-end with the emitter. Since the union is made of stainless steel, it also serves as an electrical contact to the fluid inside the emitter. Electrospray can be produced by applying an elevated electrical potential (the "capillary voltage" from the MS) anywhere on the sprayer. The nano-electrospray emitter, marketed under the name NanoEase Emitter, was obtained from Waters Corporation. It is a 20 μm ID / 90 μm OD fused silica capillary, sheathed by and bonded to a 100 μm ID / 360 μm OD fused silica capillary. Approximately 2 cm of un-sheathed 90 μm OD capillary protrudes from the larger capillary at the outlet end of the emitter. The outlet end of the capillary is not tapered; the 90 μm diameter of the capillary is small enough that a stable Taylor cone forms at the end of the emitter when operating at flow rates of 200 nL/min or higher. A gentle sheath flow of room-temperature nitrogen gas can be applied around the emitter tip in order to improve the stability of the Taylor cone. The tip of the emitter was

placed a few millimeters away from the sample cone of the mass spectrometer and was oriented perpendicular to its orifice. The emitter-to-cone distance was manually optimized to give the best signal.

On-line positive ion mode electrospray - time of flight MS was performed using a Micromass LCT instrument (Waters Corp.). A capillary voltage of +2900 V, a sample cone voltage of +35 V, and an extraction cone voltage of +2 V were used for runs in this experiment. The source temperature was set at 100 ºC. Mass spectra were acquired at a frequency of 2 Hz over a mass range from 550-1600 Da for the duration of the runs. All mass spectra were acquired using the software package MassLynx 4.0 (Waters Corp.).

4.2.6   Protein digest separation using gradient UHP-RPLC-MS

4.2.6.1 LC instrumentation and run conditions

Like the intact protein separations, the separations of protein digests were also performed using gradient ultra-high pressure reversed phase liquid chromatography (UHP-RPLC). A second preloaded gradient UHPLC pump, identical to the one described previously, was used to inject the sample and generate the gradient. The column differed slightly from the one used for intact protein separations. It was a 50 μm ID capillary, 30 cm in length, packed with 1.5 μm bridged-ethyl hybrid silica particles modified with a C18 stationary phase. The particles have 145 Å pores, which are suited for use with smaller molecules such as peptides. The column was operated at a pressure of approximately 1,600 bar (23,000psi); the flow rate through the column was around 150 nL/min. Since the UHPLC pump was operated at 4 μL/min, the split ratio was approximately 26:1.

For reversed phase separations of protein digests, mobile phase A was deionized water with 0.2% (v/v) formic acid. Mobile phase B was acetonitrile with 0.2% (v/v) formic acid. Runs were performed using a linear gradient from 1% to 50% MP B over 30 minutes,

followed by a 5 minute hold at 50%B, and then returned to initial conditions in 1 minute. The CapLC autosampler loaded 1 μL of sample onto the gradient storage loop for each run. Based on the split ratio, the volume of sample injected onto the column was approximately 40 nL.

4.2.6.2 MS instrumentation and run conditions

The outlet of the reversed-phase capillary column was coupled to a fused silica nano-electrospray emitter using the same Waters Universal Nanoflow Sprayer described in section 4.2.5.2. Due to the reduced flow rate as compared to the column used for protein separations, it was necessary to use a pulled-tip emitter as opposed to the blunt-ended NanoEase emitter. Thus, a fused silica Pico-Tip emitter was purchased from New Objective (Woburn, MA). This nano-electrospray emitter has an inner diameter of 20 μm, which tapers to 10 μm at the tip. It does not have an electrically conductive coating, since the voltage is still applied through the metal junction of the sprayer.

On-line positive ion mode electrospray - time of flight MS and MS/MS were performed using a Waters Q-TOF micro. All mass spectra were acquired using a capillary voltage of +2000 V, a sample cone voltage of +30 V, and an extraction cone voltage of +2 V, with the source temperature set at 100 ºC. For the first set of runs of the protein digest fractions, peptide mass fingerprinting (PMF) was used. Therefore, the instrument was operated in TOF-only mode, with the quadrupole set to allow all ions to pass to the TOF mass analyzer. Mass spectra were acquired at a frequency of 1.67 Hz over a mass range from 450-1600 Da for the duration of the runs.

The protein digest fractions were separated and analyzed a second time, using a data-directed analysis (DDA) method obtain MS/MS spectra of the peptides as they eluted from the column. The workflow of the DDA experiment is as follows. Initially, the mass

spectrometer runs a survey scan over the mass range from 450-1600 Da, in which the quadrupole is set to pass all ions to the TOF analyzer and the collision energy is set at its default low-energy value (7 V) so that ions are not fragmented. If no peaks above a threshold of 25 counts/second are detected, the instrument continues running survey scans at a frequency of 0.91 Hz. When the intensity of a peak exceeds a set threshold, the instrument switches to MS/MS mode. This sets the quadrupole to allow only ions with a m/z that matches the above-threshold peak to pass to the TOF analyzer. The collision energy is increased in order to induce fragmentation, and the resulting fragment ions between 200-1800 Da are detected. In order to obtain good quality MS/MS spectra, the collision energy is stepped through four different voltages, each with a scan time lasting 1.1 seconds, including a 0.1 second inter-scan delay time. The precise values of the collision energy depend on the m/z of the precursor ion, and are set using a collision energy profile table (shown in Table 4-1). After all MS/MS scans are complete (4.4 seconds after initiating the first scan), the instrument returns to performing survey scans, and continues to do so until another peak exceeds the threshold. The instrument is set to be able to trigger MS/MS on two precursor ions simultaneously. This allows data to be obtained on two co-eluting peptides. In order to avoid acquiring redundant data, the instrument is programmed not to switch to MS/MS mode on any peaks already analyzed within the last 60 seconds. To prevent MS/MS from being triggered on isotope peaks of the same component, a mass range of +/- 2.3 Da from the main peak is also excluded for the same period of time.

4.2.7   Data analysis

Data analysis for this experiment consisted of two steps. First, data from the top down and bottom up portions of the experiment were worked up independently, in order to produce a list of the intact protein masses detected and of the proteins identified using

peptide database searching. Then, it was possible to consider the data from the experiment as a whole, by attempting to match intact protein masses with protein identities.

4.2.7.1 Intact protein data workup

Data workup for the intact protein separations was performed using the methods described in Section 3.2.7. Maximum entropy de-convolution was performed on all mass spectra within the LC-MS chromatograms using AutoME. Parameters for the de-convolution were set as follows: the time segment width was 6 seconds, the output spectrum resolution was 1Da, the range of masses in the output spectrum was 5,000-80,000 Da, and the maximum number of iterations was 50. 2D chromatograms and protein peak lists were both generated from the AutoME-processed data.

4.2.7.2 Protein digest data workup

LC-MS data from the runs of the trypsin-digested fractions were collected and analyzed using two different methods: peptide mass fingerprinting (PMF) and data directed analysis (DDA) MS/MS. A different data workup method is necessary for each of these methods. In order to identify proteins from the PMF data, peak lists were generated which contain the mass and charge of all components in each chromatogram above a user-defined threshold. Although this can be performed manually, typically an add-on package to MassLynx 4.0 called ProteinLynx was used to generate the peaks lists. ProteinLynx uses an algorithm which detects peaks, recognizes their charge state via a de-convolution routine known as MaxEnt3, and records the information in a peak list file. Once the lists were generated, the web-based database search program Aldente (http://www.expasy.org/tools/aldente/) was used to match the masses with those of peptides that would be expected from tryptic digests of known *E. coli* proteins listed in the Swiss-Prot protein knowledgebase (http://us.expasy.org/sprot/). Numerous parameters can be specified

to define the criteria which Aldente uses to perform the search. A detailed list of the search parameters used for this experiment is provided in Table 4-2. Based on the output of the search, a list of proteins possibly present in the sample was produced. Aldente assigns each protein a score which indicates the probability that the identification is correct.

The runs performed using DDA MS/MS were analyzed up using ProteinLynx Global Server 2 (PLGS2, Waters Corp.), which is a software package that automates processing of LC-MS data for proteomics. Unlike Aldente, which requires the user to generate a peak list, PLGS2 allows LC-MS/MS data acquired in MassLynx to be loaded directly into the program. Generally, data analysis using PLGS2 is performed with two components: the data preparation tool and workflow designer. The data preparation tool automates noise reduction, de-isotoping, and centroiding of the mass spectra acquired during the runs. In essence, this generates data equivalent to the peak list used by Aldente. The workflow designer allows automation of databank search queries using the processed mass spectra. A variety of parameters must be specified; they are shown in Table 4-3. As with Aldente, data were searched against the Swiss-Prot *E. coli* database, and the end result was a list of proteins which were identified as being present in the sample. Each protein is assigned a probability score, along with other statistics.

4.2.7.3 Comparison of intact protein and peptide data

The intact protein and peptide data analyses generated two lists: one with intact protein masses, and the other with the names of proteins identified and a predicted mass for each protein. It was then necessary to correlate one list with the other and thereby associate intact protein masses with a protein identity. In some cases, this was as simple as comparing the predicted mass of an identified protein with the intact protein mass list from the same anion exchange fraction and finding a near-exact (within 5 Da) match. When no exact match

137

was apparent, but the difference between the observed mass of a protein and a peptide-predicted mass was within several hundred Daltons, the mass difference was checked against the Delta Mass database of common post-translational modifications (http://www.abrf.org/index.cfm/dm.home). If a logical match was found, it was recorded as a possible PTM of the identified protein.

4.3   Results and discussion

From the standpoint of chromatography, this experiment can be thought of as two distinct 2D separations: one in which the sample is intact proteins throughout, and the other in which the sample begins as proteins which are digested to peptides between the first and second dimensions. Thus, the chromatographic results from these two separations are reported and discussed in separate sections. From the viewpoint of proteomics, however, the mass spectral data from the top-down and bottom-up analyses are complementary. Thus, they are presented here as a single entity; where possible, data from the two techniques were correlated in order to reveal information not evident from either technique alone.

4.3.1   Anion exchange fractionation: chromatographic results

The first step in the hybrid top-down/bottom-up analysis was to fractionate the *E. coli* protein extract using anion exchange chromatography. The chromatogram for a 60-minute gradient separation of the sample is shown in Figure 4-5. It reveals a complex pattern of incompletely resolved peaks, which is expected due to the complexity of the sample. A total of 20 fractions were collected. Although the largest peak appears in the last fraction, it is unlikely that it and some of the other late-eluting components are actually proteins, based on observations from previous 2D-LC runs of this sample (Chapter 3). This is plausible because UV absorbance detection at 280 nm is not selective to proteins; other smaller molecules may exhibit substantial absorbance at this wavelength.

The Waters Biosuite Q 10 μm AXC column, 7.5 mm ID x 7.5 cm long, was used to perform the separation, as opposed to the custom-packed long anion exchange column. The long column would have generated greater resolution and thus allowed a larger number of distinct fractions to be collected, which would improve the total peak capacity of the 2D separation. The tradeoff for increased resolution is longer total analysis time. Given the fact that each anion exchange fraction is divided into two parts, each of which must be run using RPLC-MS, the increase in total analysis time would have been substantial. Thus, the separation provided by the short AXC column was judged to be adequate for this initial evaluation of the top down / bottom up approach to proteomics.

4.3.2 Intact protein separation: chromatographic results

Following the anion exchange separation, the fractions were lyophilized, re-constituted in a smaller volume of solvent, and then divided into two equal parts. For the portion of the fractions used for intact protein analysis, no further modification of the proteins was performed. Ultra-high pressure RPLC was used to separate the proteins, with the outlet of the column coupled to on-line ESI-TOF-MS. The chromatograms which resulted from the RPLC separations were combined to generate a 2D chromatogram, two versions of which are shown in Figure 4-6. The chromatogram in Figure 4-6A was produced using the total ion current data from the LC-MS runs, whereas Figure 4-6B is a base peak intensity chromatogram generated using the same LC-MS data after de-convolution by AutoME. As noted in Chapter 3, AutoME de-convolution eliminates much of the background noise from non-protein components in the sample. Particularly notable is its removal of the non-protein contaminant peak which appears at a reversed-phase retention time of approximately 14 minutes in all fractions (see section 3.3.3 for explanation).

139

In terms of chromatographic performance, the proteins seem to be spread over a relatively wide range of the 2D separation space. Peaks are observed in all 20 of the anion exchange fractions, though they are less abundant in the earliest and latest fractions. The second-dimension peak capacity seems to be diminished somewhat as compared to the runs performed in Chapter 3. This is due to the fact that the gradient length was decreased from 60 minutes to 30 minutes in order to reduce the total analysis time. The first dimension peak capacity is estimated to be 2/3 of the total number of fractions containing peaks, to account for some peaks being split between fractions as explained in Section 3.3.3. This gives a value of 13. The second-dimension peak capacity is estimated as being 75. Thus the total peak capacity of the 2D separation is approximately 975. As demonstrated in Chapter 3, it would easily be possible to improve this peak capacity by using a longer anion exchange column and gradient in the first dimension and by increasing the gradient length in the second dimension.

The AutoME processed data were used to generate a list of proteins detected in the *E. coli* protein extract sample. The number of proteins detected in each fraction is illustrated as a histogram in Figure 4-7. The chart shows that a relatively small number of proteins eluted from the anion exchange column in the first several fractions. The majority of the proteins, however, were relatively evenly distributed between fractions 6-19. This suggests that most proteins are well-retained by the anion exchange column, which contrasts favorably with many of the runs performed in Chapter 3, where the first few anion exchange fractions often appeared to contain more proteins than any other fractions. The improvement in retention is probably due to the change in initial buffer concentration from 25 mM to 10 mM. Although not a dramatic reduction, the decrease in ionic strength is evidently sufficient to allow a

substantial number of proteins to be retained by the anion exchange column which would otherwise have eluted near the column void time.

4.3.3   Protein digest separation: chromatographic results

The proteins in the second portion of the anion exchange fractions were first denatured using an acid-labile surfactant to expose the entire peptide backbone. They were then chemically reduced with dithiothreitol in order to cleave disulfide bonds, and alkylated using iodoacetamide to prevent disulfide bonds from re-forming. They were then digested using trypsin, an enzyme which cleaves the peptide backbone at lysine and arginine residues. Once the digestion was complete, the surfactant was degraded by acid hydrolysis, and the fractions were ready for analysis. They were separated using gradient UHP-RPLC; the outlet of the column was coupled to an ESI-Q-TOF mass spectrometer.

Two runs of each trypsin-digested fraction were performed, which were identical in terms of chromatographic conditions but differed in the mode in which the mass spectrometer was operated.   In the first set of runs, the MS was operated in TOF-only mode, in order to obtain the MW of the peptides to enable protein identification using peptide mass fingerprinting. For the second set of runs, data-directed MS/MS analysis was used to switch between a survey MS scan and MS/MS mode for fragmentation of detected ions. These modes result in LC-MS chromatograms that are substantially different in appearance, as illustrated Figure 4-8. For runs of the same anion exchange fraction, the PMF mode (Figure 4-8A) seems to produce a much higher resolution chromatogram than the MS/MS mode (figure 4-6B1). This is not due to superior chromatographic performance, since both runs were run on the same column under identical conditions. Instead, it is an issue of sampling frequency. The PMF experiment uses a constant sampling rate of 1.67 Hz; thus, a typical

peak with a base width of 8 seconds will be sampled 13 times by the mass spectrometer, which yields a relatively smooth chromatographic trace.

In MS/MS, the sampling frequency for the survey scan varies throughout the run due to the data-directed workflow. If no MS/MS scans are being performed, the survey scan sampling rate is one point every 1.1 seconds (0.91 Hz). When one MS/MS scan is active, the survey scan sampling frequency drops to one point every 4.4 seconds (0.23 Hz). With two MS/MS scans active simultaneously, the sampling frequency decreases to one point every 8.8 seconds (0.11 Hz). Thus, assuming a peak width around 8 seconds, many peaks will only be sampled once or twice in MS survey mode. This is inadequate to generate a chromatogram where peaks can be clearly discerned, which is the case in Figure 4-8B1. MS/MS also produces two additional "channels" of data which contain the spectra in which selected ions were fragmented to discern the peptide sequence. These channels are shown in Figure 4-8B2 and B3, which plot the total intensity of fragment ions as a function of time. The data appear in 4-scan "bursts", interspersed with areas of no signal. This reflects the fact that MS/MS scans are discontinuous – one ion is selected and analyzed using MS/MS for 4 scans, after which the instrument returns to MS survey mode, selects a different ion, and the process continues. The intensity-vs.-time representation of MS/MS data has limited practical use, however, since only by observing the full mass spectra can peptide sequence information be obtained.

Since the PMF runs produced visually superior chromatograms, they were combined to produce a 2D base-peak chromatogram of the peptide separation, which is shown in Figure 4-9. Unlike the intact protein data, no special de-convolution is necessary to produce the BPI chromatogram. This is because peptides are smaller molecules than proteins, and when

142

ionized by electrospray they generally exist predominantly in only one charge state. Therefore, MassLynx can generate a BPI chromatogram by simply plotting the intensity of the most abundant ion as a function of time.

One feature of the 2D chromatogram which is immediately evident is that the reversed-phase separation of peptides produces sharper peaks (or "bands", as they appear in the 2D chromatogram) than a separation of intact proteins. This is because peptides are generally more uniform in terms of chemical properties than proteins, and are also less likely to exhibit non-ideal interactions with the stationary phase, which often cause proteins to elute as broadened or asymmetrical peaks. The peak capacity of the second dimension of the 2D peptide separation is estimated to be 180; since the anion exchange peak capacity is unchanged at 13, the total peak capacity is about 2,300. Although the peak capacity is about 2.5 times higher than that of the intact protein separation, the complexity of the mixture has increased to an even greater extent. It has been reported that the average number of tryptic peptides expected to be produced per *E. coli* protein is around 13.[22] Since the increase in peak capacity is 5 times less than this value, the intact protein separation would be expected to offer better resolution of the components in the sample than the peptide separation. This conclusion is supported by a visual comparison of the experimental data. In the most crowded region of the 2D peptide chromatogram (Figure 4-9), more peptides appear to co-elute with or are poorly resolved from their nearest neighbors than is the case for the 2D protein separation (Figure 4-6). Therefore, from a chromatographic standpoint, there is an advantage to working with intact proteins rather than protein digests.

4.3.4 Top down / bottom up *E. coli* proteome analysis: mass spectral results

A sensitive and accurate means of detecting and identifying proteins is as important to the success of a proteomic analysis as the separation method used to resolve the sample

into its components. Mass spectrometry and the interpretation of mass spectral data is the technique used in this experiment which enables identification and characterization of proteins. Thus, it is important to consider the results of the *E. coli* analysis presented in this chapter from the perspective of mass spectrometry, as well as from a chromatographic one.

4.3.4.1 Proteins detected using intact protein, PMF, and MS/MS methods

The simplest means of assessing the results of the top down / bottom up *E. coli* proteome analysis is to examine the total number of proteins detected by each of the methods used to analyze the sample. Intact protein LC-MS data was de-convoluted using AutoME and a list of intact protein masses detected in the sample was generated. All duplicates and adducts were removed, which resulted in a total of 233 proteins. For the peptide LC-MS data, two different data analysis methods were used – PMF and MS/MS – both of which employed a database searching approach to identify tryptic peptides originating from known *E. coli* proteins. After all duplicate proteins were removed from the final lists, PMF identified 256 proteins and MS/MS identified 159. Based on these data alone, it would seem that PMF was the most successful technique. However, this simple comparison of numbers of proteins detected does not take into account the accuracy of the identifications, or to what extent the techniques complement one another.

In order to determine how many of the detected proteins were found by multiple methods, all three protein lists were combined and sorted in order of protein MW. For the intact protein data, the MW data came from the observed protein masses; for the peptide data, the MWs predicted by the database search program were used. The lists were then surveyed to find masses which were within 5 Da of one another (a conservative estimate for the mass accuracy of the intact protein data). Matching masses were recorded as proteins found using multiple techniques. A Venn diagram which displays the overlap of the number

of proteins found by all three methods is shown in Figure 4-10A. Of the possible total of 510 distinct proteins found, only 25 (4%) were detected by all three techniques. The overlap between PMF and intact protein data was 31/233, or 13%. The overlap between MS/MS and intact protein data was about twice as high, at 64/233, or 27%. The overlap between MS/MS and PMF data was approximately the same, at 68/256, or 27%.

Since the goal of the experiment was to correlate top-down and bottom-up techniques, it was necessary to decide which bottom-up data set to use for further comparisons. Although PMF produced a higher total number of protein identifications than MS/MS, fewer of its proteins matched with those found using the top-down method. This reduced percentage of overlap implies that PMF may have produced a higher number of false protein identifications than MS/MS. This is not difficult to believe, since PMF relies on peptide MW alone to identify peptides. The accuracy of PMF is very sensitive to the mass accuracy of the instrument used to acquire the data, which in the case of this experiment was prone to drift of up to 200 ppm (0.2 Da for a 1 kDa peptide) due to reliance on external calibration. MS/MS, on the other hand, employs the study of fragmentation patterns of selected ions, which is an inherently more discriminating process and is less prone to false identifications, in spite of limited mass accuracy. It was therefore determined that the MS/MS dataset would be used for further correlations between top-down and bottom up analyses. A simplified version of the Venn diagram showing only the overlap between these two methods is shown in Figure 4-10B. Complete tables were produced which list all proteins detected by both methods (Table 4-4), as well as those found only in intact protein (Table 4-5) and MS/MS (Table 4-6) datasets. Descriptive names of the identified proteins are provided in the appendix of this dissertation.

4.3.4.2 Assessment of overlap between top-down and bottom-up methods

One discouraging feature of the results presented in this chapter is the relatively low overlap (approximately 27%) between the number of intact proteins detected and proteins identified based on bottom-up peptide analysis. Although low, this value is comparable to those reported by similar studies found in the literature.[17, 18, 23] Still, it is worthwhile to consider the factors which may contribute to the observed lack of overlap. The simplest explanation is that many proteins may not have been detected by both the top-down and the bottom-up approach. This is particularly likely for low-abundance proteins, where the protein is near the detection limit. In the MS/MS runs, some proteins may also have been missed due to a low scan frequency. If a peptide happened to elute from the column just after the instrument began to perform MS/MS on other ions, the new peptide might have finished eluting by the time the instrument returned to survey scan mode. The probability of missing components in MS/MS analyses could be reduced by increasing the peak capacity of the LC x LC separation, by using a mass spectrometer with a higher scan rate and duty cycle, or by performing replicate runs of each sample.

Another factor that may contribute to the low overlap is the fact that there is an inherent difference in selectivity for top-down and bottom-up approaches. This is made clear by examining the MW distribution histogram shown in Figure 4-11. From the graph, it is immediately apparent that a larger number of intact proteins were detected than were found using MS/MS. However, the more significant aspect of the data is the fact that the distribution of the proteins identified by MS/MS is shifted toward higher masses than the intact protein data. No masses above 60 kDa were found in the intact data, whereas proteins up to 111 kDa were found using MS/MS. Numerically, the average MW for the intact protein data is 21.0 kDa and the median is 18.6 kDa, whereas for the MS/MS data the

average is 33.1 kDa and the median is 29.2 kDa. Interestingly, the values from the MS/MS data are much closer to – although still lower than – the theoretically predicted 35.1 kDa average and 30.6 kDa median calculated from all known *E. coli* proteins (see Section 3.3.4).

The differences in the MW distributions can be explained by considering the nature of the two techniques. Generally speaking, small intact proteins (MW < 30 kDa) are ionized and detected more readily using electrospray-MS than larger proteins. Conversely, MS/MS may favor detection of high MW proteins. This is because, for a given concentration of protein, more peptides are produced when a large protein is enzymatically digested than a small one. The probability of detecting one or more of these peptides in a bottom-up analysis is higher than finding one of the comparatively few peptides produced by a digest of a smaller protein. Thus, top down analysis may be biased toward small proteins, whereas bottom-up may be biased toward larger proteins.

4.3.4.3 Detection of post-translational modifications

A final possible reason for the low degree of overlap – and also an important feature of the data in its own right – is the presence of post-translational modifications (PTMs). In this experiment, when an exact match was found between an observed intact protein MW and the predicted MW of a protein identified by the bottom-up method, it was viewed as confirmation that the identification was correct. Some predicted protein masses already incorporated certain PTMs, such as removal of a signal sequence of approximately 20 amino acids from the N-terminus of the protein. However, many of the proteins in this sample are likely to have PTMs which are not recorded in the databases used for bottom-up protein identification. In some cases, consideration of other PTMs improved the correlation between top-down and bottom up data. Table 4-7 shows nine proteins which were found using both intact protein and peptide MS/MS analyses when certain common modifications were

considered, such as acetylation, phosphorylation, and removal (or non-removal) of the N-terminal methionine. The PTMs listed in the table cannot be assigned with absolute confidence, since the only evidence for their presence is the difference in mass between the observed and predicted protein MW. Nevertheless, the ability to detect possible PTMs demonstrates that the hybrid top down / bottom up technique can give information about the proteins present in the sample which neither method could provide on its own.

Although common post-translational modifications could be assigned based on differences between observed and predicted protein MW, it was not possible to identify the entire range of PTMs that may occur or to discern multiple PTMs on the same protein. If these modifications could be detected, however, the actual overlap between the top-down and bottom-up data might be substantially higher than was determined in this experiment. To improve chances of unambiguously determining PTMs additional improvements would be required – such as significantly higher resolution and mass accuracy, and/or MS/MS of intact proteins. Another daunting challenge is determining the location of PTMs on a protein. Various MS/MS techniques have been used to locate specific PTMs within peptides.[24, 25] Such methods are relatively early in development for use at the intact protein level, however.[26] Thus, characterization of post-translational modifications remains a difficult, yet significant, aspect of the development of top-down / bottom-up proteomics.

4.3.5   Visualization of LC-MS data using 2D chromatospectra

In addition to the chromatographic and proteomic analyses already described, chromatospectra were used as a supplementary means of examining the data generated in this experiment. In essence, a chromatospectrum is a plot analogous to a 2D chromatogram, except one of the axes contains MS data. In this experiment, chromatospectra were applied to visualization of AutoME-de-convoluted intact protein data. Since it was desirable to

display the results from an entire LC x LC separation in one plot, it was necessary to combine all of the data from one of the dimensions into a single data set. Because the peak capacity of the reversed-phase separation is higher than for the anion exchange separation, the data from all anion exchange fractions were combined and placed in a table containing reversed-phase retention time, protein MW, and intensity.

Once the data was properly formatted, a chromatospectrum was generated using an Igor Pro function written by John Eschelbach. The details of this custom-programmed function have been described elsewhere.[27] When applied to the data gathered in this experiment, the resulting 2D plot is shown in Figure 4-12. The chromatospectrum shows reversed-phase retention time on the X-axis and protein MW on the Y-axis, with component intensity represented by color, with blue being the lowest intensity and black the highest. The apparent "peak capacity" given by the MS dimension is truly spectacular – for this analysis, a range from 5,000 to 60,000 Da is displayed, and the data were de-convoluted to a resolution of 1 Da. This yields an effective peak capacity of 55,000, which, when multiplied by the reversed-phase peak capacity of 75, gives a total peak capacity of 4.1 million. It is important to note that, on its own, a mass spectrometer would not be able to resolve anywhere near 55,000 proteins in a single mass spectrum, at least with current technology. In fact, due to the occurrence of ionization suppression and the fact that spectra are more complex preceding de-convolution, the practical peak capacity for ESI-TOF-MS of proteins is probably closer to 10. Nonetheless, when intact protein mass spectrometry is used as a part of a LC x LC analysis, it allows a very detailed visualization of the data to be produced.

A chromatospectrum with a peak capacity over 4 million could not be adequately displayed on a standard sheet of paper, because the width of each data point would be too

small to be visually discerned. Thus, Figure 4-12 was generated using 100 Da bins, which reduces the visually observable peak capacity to 41,000. Although a large reduction from 4.1 million, this is still high compared to a typical LC x LC chromatogram. To access the higher resolution data, a specific region of the chromatospectrum can be displayed using a smaller bin size. This is illustrated using the area inside the dashed rectangle, which is expanded in Figure 4-13. The bin size in this chromatospectrum was 1 Da. The expanded view reveals much more detail, which was not evident in the lower-resolution chromatospectrum. With the increased resolution, peaks separated by only a few tens of Daltons can be clearly discerned.

It is important to consider what practical uses chromatospectra might serve in a proteomic analysis. One logical application for "low-resolution", full mass range chromatospectra would be in making comparisons between samples. Since protein mass measurement is not subject to nearly as much drift as chromatographic retention times, a chromatospectrum would have an advantage over a 2D chromatogram in terms of reproducibility in one of its dimensions. Higher resolution chromatospectra could also prove useful in detecting multiple forms of a protein. As seen in Figure 4-13, proteins with multiple forms appear as parallel bands separated by tens or hundreds of Daltons. These could be due to a number of post-translational modifications, which are of great interest in proteomic studies. The visual nature of the chromatospectrum makes it easy to detect such patterns within the data.

4.4   Conclusions

In this chapter, a hybrid top-down / bottom-up approach was shown to be a successful means of studying the proteome of *E. coli*. LC x LC coupled to mass spectrometry was the fundamental technique used to carry out the analysis. The LC dimensions were coupled off-

line, which allowed fractions from the first dimension, anion exchange, to be divided in two parts. One set of fractions were analyzed as intact proteins using UHP-RPLC-MS; this served as the top-down portion of the experiment. The second set was digested using trypsin and analyzed using UHP-RPLC-MS/MS; this represented the bottom-up portion of the analysis. Data from the two workflows were analyzed and compared in order to match intact proteins detected with proteins identified using database searching of peptide MS/MS spectra. In so doing, numerous post-translational modifications were revealed which were not predicted by the peptide databank entries.

Although numerous *E. coli* proteins were successfully characterized using the method described in this chapter, the technique does suffer from certain shortcomings. First, the total number of proteins detected – approximately 200 – is well under $1/10^{th}$ of the total proteins expected to be produced by *E. coli*. Clearly, no strategy used to date has been able to successfully characterize all of the proteins produced by any organism, though greater numbers of identifications have been reported elsewhere.[22] As discussed, the number of proteins characterized could be improved by performing a higher resolution anion exchange separation, collecting more fractions, and lengthening the reversed-phase gradients. Making these improvements would, however, substantially increase the analysis time. Unfortunately, the existing method is already relatively slow, due to the time required to analyze 20 fractions off-line using both top-down and bottom-up methods. In order to increase the number of fractions analyzed without suffering an unreasonable increase in analysis time, a substantial change in approach would be necessary. It was this fact which motivated a switch to on-line LC x LC, which will be described in the next chapter.

4.5   References

(1)   Swanson, S. K.; Washburn, M. P. *Drug Discovery Today* **2005**, *10*, 719-725.

(2)   Wehr, T. *LC-GC North America* **2006**, *24*, 1006-1008.

(3)   Paizs, B.; Suhai, S. *Mass Spectrometry Reviews* **2004**, *24*, 508-548.

(4)   Kelleher, N. L. *Analytical Chemistry* **2004**, *76*, 196A-203A.

(5)   Bogdanov, B.; Smith, R. D. *Mass Spectrometry Reviews* **2005**, *24*, 168-200.

(6)   Scigelova, M.; Makarov, A. *Practical Proteomics* **2006**, *1*, 16-21.

(7)   Reid, G. E.; McLuckey, S. A. *Journal of Mass Spectrometry* **2002**, *37*, 663-675.

(8)   Good, D.; Coon, J. *BioTechniques* **2006**, *40*, 783-789.

(9)   Schirle, M.; Heurtier, M. A.; Kuster, B. *Molecular and Cellular Proteomics* **2003**, *2*, 1297-1305.

(10)   Tan, A.; Pashkova, A.; Zang, L.; Foret, F.; Karger, B. L. *Electrophoresis* **2002**, *23*, 3599-3607.

(11)   Baddock, V.; Steinhusen, U.; Bommert, K.; Otto, A. *Electrophoresis* **2001**, *22*, 2856-2864.

(12)   Wang, H.; Hanash, S. *Mass Spectrometry Reviews* **2005**, *24*, 413-426.

(13)   Butt, A.; Davison, M. D.; Smith, G. J.; Young, J. A.; Gaskell, S. J.; Oliver, S. G.; Beynon, R. J. *Proteomics* **2001**, *1*, 42-53.

(14)   Liu, H.; Berger, S. J.; Chakraborty, A.; Plumb, R. S.; Cohen, S. A. *Journal of Chromatography B* **2002**, *782*, 267-289.

(15)   VerBerkmoes, N. C.; Bundy, J. L.; Hauser, L.; Asano, K. G.; Razumovskaya, J.; Larimer, F.; Hettich, R. L.; Stephenson, J. L., Jr. *Journal of Proteome Research* **2002**, *1*, 239-252.

(16)   Hayter, J. R.; Robertson, D. H. L.; Gaskell, S. J.; Beynon, R. J. *Molecular and Cellular Proteomics* **2003**, *2*, 85-95.

(17)   Hamler, R. L.; Zhu, K.; Buchanan, N. S.; Kreunin, P.; Kachman, M. T.; Miller, F. R.; Lubman, D. M. *Proteomics* **2004**, *4*, 562-577.

(18)   Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *Journal of Proteome Research* **2006**, *5*, 135-146.

(19)   Berger, S. J.; Millea, K. M.; Krull, I. S.; Cohen, S. A. In *Separation Methods in Proteomics*; Smejkal, G. B., Lazareu, A., Eds.; CRC Press LLC: Boca Raton, Florida, 2006, pp 387-417.

(20)   Sharma, S.; Simpson, D. C.; Tolic, N.; Jaitly, N.; Mayampurath, A. M.; Smith, R. D.; Pasa-Tolic, L. *Journal of Proteome Research* **2007**, *6*, 602-610.

(21)   Yu, Y.-Q.; Gilar, M.; Lee, P. J.; Boubier, E. S. P.; Gebler, J. C. *Analytical Chemistry* **2003**, *75*, 6023-6028.

(22)   Corbin, R. W.; Paily, O.; Yang, F.; Shabanowitz, J.; Platt, M.; Lyons, C. E., Jr.; Root, K.; McAuliffe, J.; Jordan, M. I.; Kustu, S.; Soupene, E.; Hunt, D. F. *Proceedings of the National Academy of the Sciences of the United States of America* **2003**, *100*, 9232-9237.

(23)   Kreunin, P.; Urquidi, V.; Lubman, D. M.; Goodison, S. *Proteomics* **2004**, *4*, 2754-2765.

(24)   Sweet, S. M. M.; Creese, A. J.; Cooper, H. J. *Analytical Chemistry* **2006**, *78*, 7563-7569.

(25)   Harvey, D. J. *Expert Review of Proteomics* **2005**, *2*, 87-101.

(26)   Scherperel, G.; Reid, G. E. *Analyst* **2007**, *132*, 500-506.

(27)   Eschelbach, J. W. Doctoral Thesis. University of North Carolina - Chapel Hill, Chapel Hill, 2006.

4.6   Tables

m/z range	Collision energy 1 (V)	Collision energy 2 (V)	Collision energy 3 (V)	Collision energy 4 (V)
400.00-500.00	20	22	24	26
500.01-600.00	22	24	26	28
600.01-700.00	24	26	28	30
700.01-800.00	26	28	30	32
800.01-900.00	28	30	32	34
900.01-up	30	32	34	36

Table 4-1: Collision energy profile used for MS/MS experiments

Parameter	Value
Database	UniProt KB / Swiss-Prot
Taxon	Escherichia coli
Protein MW range	0-285,000
Protein pI range:	0.00-14.00
Include varsplice	yes
Include fragment	yes
Enzyme	trypsin
Missed cleavage	1
Resolution	monoisotopic
Ion mode	[M]
PTM	observed / by similarity
Max PTM per peptide	2
Modifications	CAM C – C_2H_4ON – fixed
	MSO M – O – variable
Shift max (Da)	0.00
Slope max (ppm)	200
Internal error (ppm)	25
pValue max	1×10^{-3}
Random try max	100,000
Random fine min	3,000
Peptide: Scoring missed cl	0.50
Peptide: Scoring intensity	1.00
Peptide: Scoring C-term R	1.00
Peptide: Scoring C-term K	1.00
Peptide: Scoring C-term L	1.00
Protein: Scoring Coverage	1.00
Protein: Scoring MW	1.00
Protein: Scoring pI	1.00

Table 4-2: Parameters used to perform peptide mass fingerprinting database searching using Aldente (http://www.expasy.org/tools/aldente/)

Data processing parameter	Value
Mass Accuracy	No lockspray calibration
Noise reduction (electrospray survey and MS/MS)	Adaptive
Perform de-isotoping	Yes
Deisotoping type	Slow (MaxEnt 3)
Maximum iterations	50
Workflow template parameter	**Value**
Search engine type	PLGS
Databanks	SWISSPROT_NCBI-1.0
Species	*E. coli*
Peptide tolerance	50 ppm
Fragment tolerance	0.3 Da
Estimated calibration error	0.005 Da
Molecular weight range	0-100,000 Da
pI range	0-14
Minimum peptides to match	1
Maximum hits	200
Primary digest reagent	Trypsin
Secondary digest reagent	None
Missed cleavages	2
Fixed modifications	Carbamidomethyl C
Variable modifications	Oxidation M
Validate	Yes
Filter	None

Table 4-3: Parameters used to perform LC-MS/MS data workup using ProteinLynx Global Server 2.0 (Waters) Data processing parameters control how the MS spectra were processed prior to database searching; workflow template parameters control how the MS/MS database search is performed

Ref #	Protein	Peptides matched	Coverage %	Predicted MW	Observed MW	AXC Fraction	RT
1	CSPC_ECOLI	2	43	7271	7272	3	18.18
2	CSPE_ECOLI	1	18	7332	7333	2	19.05
3	YJBJ_ECOLI	5	59	8325	8326	6	17.43
4	HDEB_ECOLI	1	16	9065	9062	10	17.50
5	PTHP_ECOLI	1	35	9119	9119	8	20.64
6	RS15_ECOLI	2	42	10138	10142	6	20.53
7	CH10_ECOLI	2	25	10387	10387	14	19.69
8	FETP_ECOLI	1	26	10821	10822	10	19.00
9	YGIW_ECOLI	2	27	11976	11977	10	19.40
10	YEGP_ECOLI	1	17	12024	12026	8	16.14
11	RL7_ECOLI	1	10	12164	12162	16	28.25
12	GRCA_ECOLI	3	35	14284	14285	10	19.60
13	YCCU_ECOLI	2	27	14701	14701	8	26.74
14	OSMC_ECOLI	3	41	14957	14955	9	21.51
15	RS6_ECOLI	1	8	15316	15316	16	18.64
16	NDK_ECOLI	2	28	15332	15333	13	24.18
17	HNS_ECOLI	2	15	15408	15410	12	21.22
18	SODC_ECOLI	1	5	15739	15737	9	16.31
19	USPA_ECOLI	2	36	15935	15934	18	25.46
20	YJGK_ECOLI	2	11	16865	16867	12	20.92
21	TPX_ECOLI	3	35	17704	17704	12	26.72
22	PTGA_ECOLI	4	30	18120	18121	14	23.59
23	OSMY_ECOLI	11	60	18161	18162	8	16.74
24	BFR_ECOLI	2	16	18495	18494	19	24.94
25	FABA_ECOLI	3	20	18838	18842	17	20.28
26	FTNA_ECOLI	1	8	19424	19425	12	21.62
27	YFBU_ECOLI	1	5	19536	19537	13	24.28
28	AHPC_ECOLI	5	35	20630	20628	14	24.29
29	WRBA_ECOLI	6	46	20714	20712	17	23.68
30	SODF_ECOLI	2	15	21135	21136	10	24.20
31	RPIA_ECOLI	1	4	22860	22863	12	24.22
32	KAD_ECOLI	6	35	23586	23587	9	22.21
33	RPE_ECOLI	2	8	24554	24553	19	25.34
34	ARTI_ECOLI	5	27	25042	25042	7	21.01
35	FABG_ECOLI	1	5	25560	25561	8	24.84
36	ARGT_ECOLI	1	6	25785	25785	7	20.71
37	DEOD_ECOLI	3	20	25819	25818	15	25.87
38	FLIY_ECOLI	2	10	26068	26069	8	21.24
39	HISJ_ECOLI	1	6	26233	26233	7	20.41
40	TPIS_ECOLI	1	5	26972	26972	13	22.98
41	UDP_ECOLI	2	14	27028	27029	13	25.58
42	KDUD_ECOLI	7	28	27070	27067	16	27.75
43	FABI_ECOLI	1	6	27733	27735	10	25.40
44	GPMA_ECOLI	7	31	28425	28427	10	23.70
45	RBSB_ECOLI	18	75	28474	28476	6	21.33
46	EFTS_ECOLI	4	20	30292	30293	14	22.59

Table 4-4: *E. coli* proteins identified by both intact protein and peptide MS/MS analyses. (Table is continued and term definitions are included on the following page).

Ref #	Protein	Peptides matched	Coverage (%)	Predicted MW (Da)	Observed MW (Da)	AXC Fraction	RT (min)
47	HCHA_ECOLI	3	13	31059	31062	12	23.62
48	YTFG_ECOLI	3	9	32126	32127	6	22.43
49	MDH_ECOLI	10	35	32337	32338	8	25.74
50	DGAL_ECOLI	12	51	33368	33371	7	21.91
51	PROX_ECOLI	1	7	33727	33727	7	20.81
52	CYSK_ECOLI	4	21	34358	34359	13	23.88
53	ASPG2_ECOLI	1	4	34593	34593	9	21.41
54	TALB_ECOLI	7	21	35088	35090	11	25.71
55	TALA_ECOLI	4	18	35659	35660	10	23.30
56	6PGL_ECOLI	2	9	36308	36304	18	20.36
57	GLPQ_ECOLI	5	19	38200	38201	14	20.79
58	ALF_ECOLI	3	9	39016	39014	15	23.37
59	MALE_ECOLI	4	12	40707	40709	9	22.31
60	EFTU_ECOLI	11	35	43182	43188	16	22.84
61	GLYA_ECOLI	8	24	45317	45320	12	23.22
62	UGPB_ECOLI	2	7	46124	46126	7	21.51
63	TNAA_ECOLI	16	45	52773	52775	11	24.21
64	DPPA_ECOLI	1	5	57407	57406	7	23.71

Table 4-4: (Table is continued from the previous page) *E. coli* proteins identified by both intact protein and peptide MS/MS analyses. Heading terms are defined as follows: **Ref #:** Internal reference number; see below. **Protein**: SwissProt database entry name; **Peptides matched**: number of tryptic peptides identified using ESI-q-TOF MS/MS analysis; **Coverage %**: percent coverage of predicted protein sequence based on peptides matched; **Predicted MW**: average protein mass calculated from predicted protein sequence in SwissProt database entry; **Observed MW**: de-convoluted molecular weight of protein determined by ESI-TOF MS; **AXC Fraction**: anion exchange fraction number; **RT**: intact protein reversed-phase retention time.

A descriptive name of all proteins in this table is provided in the appendix of this dissertation, listed by Ref #.

MW	Frac.	RT	MW	Frac.	RT	MW	Frac.	RT	MW	Frac.	RT
5064	10	17.40	11691	12	25.02	19490	11	19.11	30979	9	24.01
5458	2	15.05	11745	12	22.72	19572	15	21.67	31094	12	23.32
5474	2	13.95	11896	2	16.75	19685	14	21.69	31234	2	19.95
5591	3	15.08	11911	6	16.33	19729	11	24.61	31658	19	26.94
5760	17	21.78	12205	16	28.95	19892	12	14.12	32262	10	26.20
6318	6	19.23	12220	16	28.65	19951	14	22.49	32899	20	26.14
6856	10	17.40	12294	16	28.45	19963	14	24.09	33271	9	23.61
6867	2	16.95	12424	16	23.14	20112	17	23.18	33827	17	25.18
6945	18	24.36	12692	12	25.92	20324	10	20.70	34375	13	23.68
7136	16	21.14	13003	13	17.78	20453	14	23.29	34486	17	23.08
7151	7	13.81	13111	14	21.99	20472	7	22.71	36489	17	24.18
7349	14	17.99	13520	8	22.14	20603	15	25.87	38704	10	23.20
8157	10	24.80	13737	10	20.60	20888	6	20.43	40990	12	28.13
8275	11	10.31	13757	8	23.44	20945	6	20.33	41258	14	25.29
8342	6	17.33	14085	12	21.72	20994	19	23.54	42150	17	22.88
8393	18	22.96	14102	12	21.52	21151	10	23.80	42852	10	26.30
8536	6	19.23	14151	19	29.15	21222	19	26.64	43543	11	21.21
8670	12	21.92	14297	10	20.40	21574	16	26.34	43555	11	21.01
8749	16	20.44	14364	8	17.84	21666	19	27.85	45010	6	21.03
9005	7	17.51	14442	6	19.23	21738	2	22.15	49317	6	20.73
9078	10	17.10	14528	20	27.14	21760	13	23.78	50264	11	23.41
9094	10	17.30	14607	10	20.10	22082	16	25.74	51349	14	27.19
9161	10	18.20	14662	17	29.59	22297	9	26.91	52670	15	26.67
9283	4	17.57	14939	12	20.72	22973	4	22.17	57689	17	26.88
9386	11	20.21	15291	15	23.27	23457	10	19.70	58727	12	23.42
9461	7	16.71	15709	11	23.41	23531	6	20.63			
9477	7	16.51	15946	18	26.16	23783	19	22.64			
9577	8	17.74	16099	7	19.51	23858	11	23.41			
9705	7	16.61	16155	17	28.29	23906	13	25.08			
9737	11	18.01	16284	16	24.44	24175	13	25.88			
9755	11	17.61	16659	18	22.76	24532	10	21.80			
9802	6	20.13	16958	9	25.21	24572	12	24.82			
9827	7	19.41	17036	14	22.09	24971	4	21.47			
9838	13	19.78	17057	6	19.43	24985	2	21.25			
9903	12	20.12	17317	15	23.47	25863	8	20.84			
9967	12	19.82	17523	18	20.76	26313	13	26.58			
10314	6	19.43	17647	16	24.34	26959	8	24.44			
10602	13	20.38	17791	4	22.17	27029	13	25.58			
10615	6	20.83	17980	12	26.52	28549	7	21.61			
10651	2	19.95	18082	2	19.85	28578	7	21.71			
10779	11	20.11	18178	8	16.54	28679	6	21.93			
10919	16	24.04	18214	6	21.83	28702	8	21.74			
11035	19	22.44	18248	8	21.64	28864	6	22.53			
11157	14	22.99	18389	9	17.11	29296	10	24.60			
11224	9	21.11	18795	17	27.18	29364	12	23.22			
11400	9	17.11	18963	10	25.50	29389	6	23.63			
11496	15	21.67	19199	12	23.02	30206	6	22.13			
11674	12	26.72	19408	12	21.72	30417	6	23.33			

Table 4-5: *E. coli* proteins detected by intact protein analysis only. Heading terms are defined as follows: **MW**: de-convoluted molecular weight of protein determined by ESI-TOF MS; **Frac.**: anion exchange fraction number; **RT**: reversed-phase retention time (in minutes)

Ref #	Protein	Prd MW	Pep mtc	% Cov	Ref #	Protein	Prd MW	Pep mtc	% Cov
65	RL29_ECOLI	7273	1	22	113	DHAS_ECOLI	40018	1	5
66	YCCJ_ECOLI	8524	1	16	114	PGK_ECOLI	40987	14	55
67	RL28_ECOLI	8875	1	13	115	SUCC_ECOLI	41393	6	24
68	DBHB_ECOLI	9226	2	33	116	AGP_ECOLI	43560	1	3
69	DBHA_ECOLI	9535	4	58	117	AAT_ECOLI	43573	2	6
70	RL25_ECOLI	10693	2	10	118	DEOB_ECOLI	44370	2	6
71	RL24_ECOLI	11185	1	17	119	SURA_ECOLI	45078	1	3
72	IHFA_ECOLI	11354	1	10	120	ENO_ECOLI	45524	7	28
73	RL21_ECOLI	11564	1	12	121	IDH_ECOLI	45757	2	5
74	RS10_ECOLI	11736	2	23	122	ACEA_ECOLI	47522	2	5
75	RL18_ECOLI	12770	1	8	123	CISY_ECOLI	48015	3	10
76	GLRX4_ECOLI	12879	1	11	124	TIG_ECOLI	48193	3	9
77	YJGF_ECOLI	13480	2	48	125	SYS_ECOLI	48414	2	7
78	RL11_ECOLI	14744	2	17	126	ATPB_ECOLI	50194	1	3
79	RL15_ECOLI	14980	2	18	127	DLDH_ECOLI	50557	1	2
80	SKP_ECOLI	15692	2	20	128	KPYK2_ECOLI	51226	2	8
81	RL9_ECOLI	15769	5	40	129	6PGD_ECOLI	51481	2	9
82	RL13_ECOLI	16019	1	14	130	ALDA_ECOLI	52142	9	22
83	ASNC_ECOLI	16888	1	7	131	ASPA_ECOLI	52356	2	5
84	RL10_ECOLI	17580	1	7	132	DCEA_ECOLI	52685	6	16
85	YBAY_ECOLI	17682	1	9	133	YHJJ_ECOLI	53050	1	3
86	PPIB_ECOLI	18153	2	20	134	ATPA_ECOLI	55222	5	11
87	RL6_ECOLI	18773	3	24	135	OPGG_ECOLI	55365	2	8
88	RS7_ECOLI	19888	1	10	136	GLPK_ECOLI	56100	20	50
89	NUSG_ECOLI	20400	1	8	137	ALDB_ECOLI	56306	2	5
90	CLPP_ECOLI	21563	1	10	138	CH60_ECOLI	57198	11	28
91	YRBC_ECOLI	21733	1	7	139	PUR9_ECOLI	57329	2	6
92	RL3_ECOLI	22244	1	10	140	PPCK_ECOLI	59643	8	27
93	ALKH_ECOLI	22284	4	25	141	RS1_ECOLI	61158	4	13
94	SODM_ECOLI	22966	1	4	142	G6PI_ECOLI	61530	3	7
95	CRP_ECOLI	23640	2	12	143	SYD_ECOLI	65913	1	2
96	GLRX2_ECOLI	24350	1	6	144	GLMS_ECOLI	66763	2	4
97	RL1_ECOLI	24598	3	16	145	DNAK_ECOLI	68984	7	15
98	GLNH_ECOLI	24963	6	33	146	HTPG_ECOLI	71423	2	4
99	RS3_ECOLI	25852	1	5	147	TKT1_ECOLI	72212	2	5
100	YBGI_ECOLI	26892	1	8	148	PNP_ECOLI	77101	3	5
101	UDP_ECOLI	27028	2	14	149	EFG_ECOLI	77450	18	39
102	DEOC_ECOLI	27734	2	11	150	MASZ_ECOLI	80357	1	2
103	FARR_ECOLI	28273	1	3	151	MAO2_ECOLI	82417	4	9
104	SUCD_ECOLI	29646	1	33	152	CATE_ECOLI	84163	6	9
105	DAPD_ECOLI	29892	1	3	153	PFLB_ECOLI	85226	2	2
106	KDUI_ECOLI	31076	3	16	154	SYFB_ECOLI	87378	2	4
107	GLTI_ECOLI	31229	2	11	155	ACON2_ECOLI	93498	3	4
108	KDGK_ECOLI	33962	3	15	156	ADHE_ECOLI	95996	2	3
109	G3P1_ECOLI	35401	13	48	157	SYV_ECOLI	108192	1	1
110	USPE_ECOLI	35576	1	6	158	GABT_ECOLI	45775	1	3
111	YGHZ_ECOLI	38832	1	8	159	IF3_ECOLI	20564	2	13
112	SERC_ECOLI	39652	1	3					

Table 4-6: *E. coli* proteins detected by peptide MS/MS analysis only. Heading term definitions are the same as in Table 4-4 (abbreviated versions used). A descriptive name of all proteins in this table is provided in the appendix of this dissertation, listed by Ref #.

Ref #	Protein	Pept mtch	Cov %	AXC Frac.	RT	Pred. MW	Obs. MW	ΔMW	Assigned modification
66	YCCJ_ECOLI	1	16	18	22.96	8524	8393	-131	Met loss
67	RL28_ECOLI	1	13	7	17.51	8875	9005	130	No Met loss
68	DBHB_ECOLI	2	33	10	17.30	9226	9094	-131	Met loss
69	DBHA_ECOLI	4	58	8	17.74	9535	9577	42	Acetylation
70	RL25_ECOLI	2	10	2	19.95	10693	10651	-43	No acetylation
72	IHFA_ECOLI	1	10	9	21.11	11354	11224	-130	Met loss
79	RL15_ECOLI	2	18	12	20.72	14980	14939	-42	No acetylation
82	RL13_ECOLI	1	14	7	19.51	16019	16099	80	Phosphorylation
129	6PGD_ECOLI	2	9	14	27.19	51481	51349	-132	Met loss

Table 4-7: Additional *E. coli* proteins possibly identified by both intact protein and peptide MS/MS analyses when certain common post-translational modifications were considered. Heading term definitions are the same as in Table 4-4 (in abbreviated form), with the addition of the following: **ΔMW**: Difference in MW calculated by subtracting Predicted MW from Observed MW; **Assigned modification**: post-translational modification of the protein assigned based on delta-mass information. A descriptive name of all proteins in this table is provided in the appendix of this dissertation, listed by Ref #.

4.7 Figures
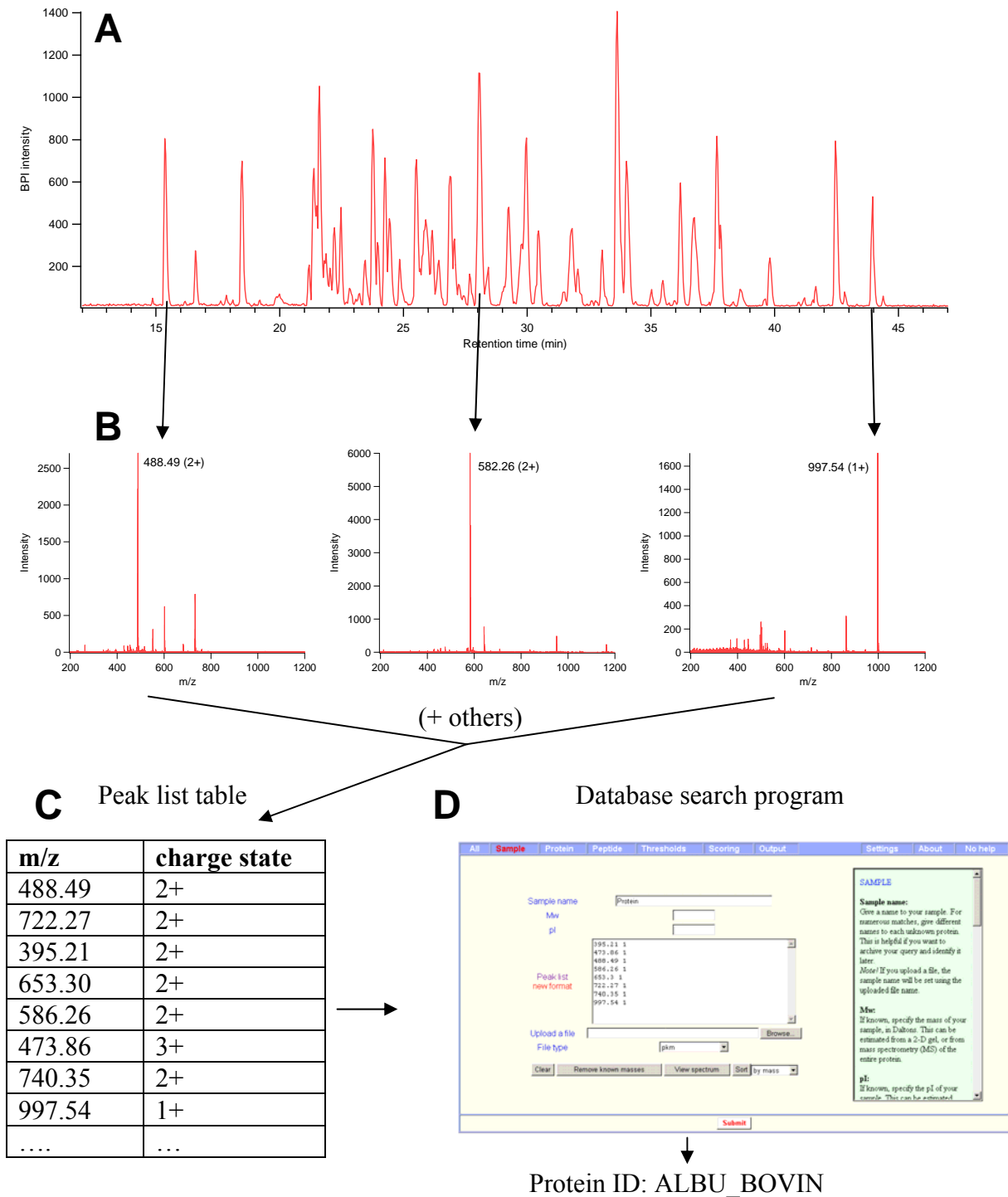
Figure 4-1: Overview of peptide mass fingerprinting (PMF) method. An LC-MS separation of a peptide mixture is performed (A); mass spectra for all major components are generated (B) and their m/z and charge state are recorded in a peak list (C). The peak list is searched against a database (D), which produces an identification of the protein(s)
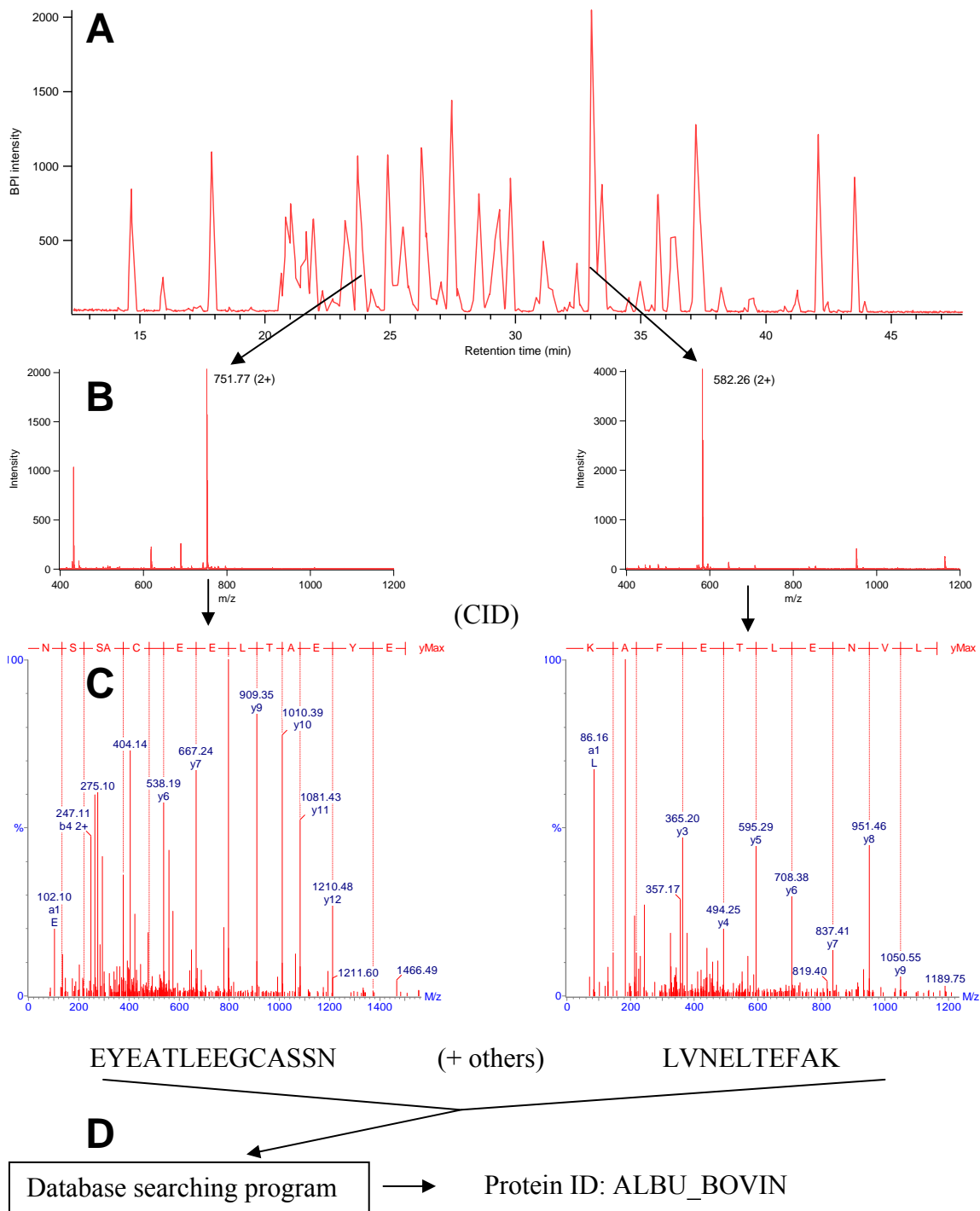
Figure 4-2: Overview of bottom-up protein identification using DDA MS/MS of peptides. An LC-MS separation of a peptide mixture is performed (A); MS survey mode is used to detect major components (B). These components are then selected and fragmented by CID; the resulting MS/MS spectra can be used to sequence the peptides. The peptide sequences are searched against a database (D), which identifies the protein(s).
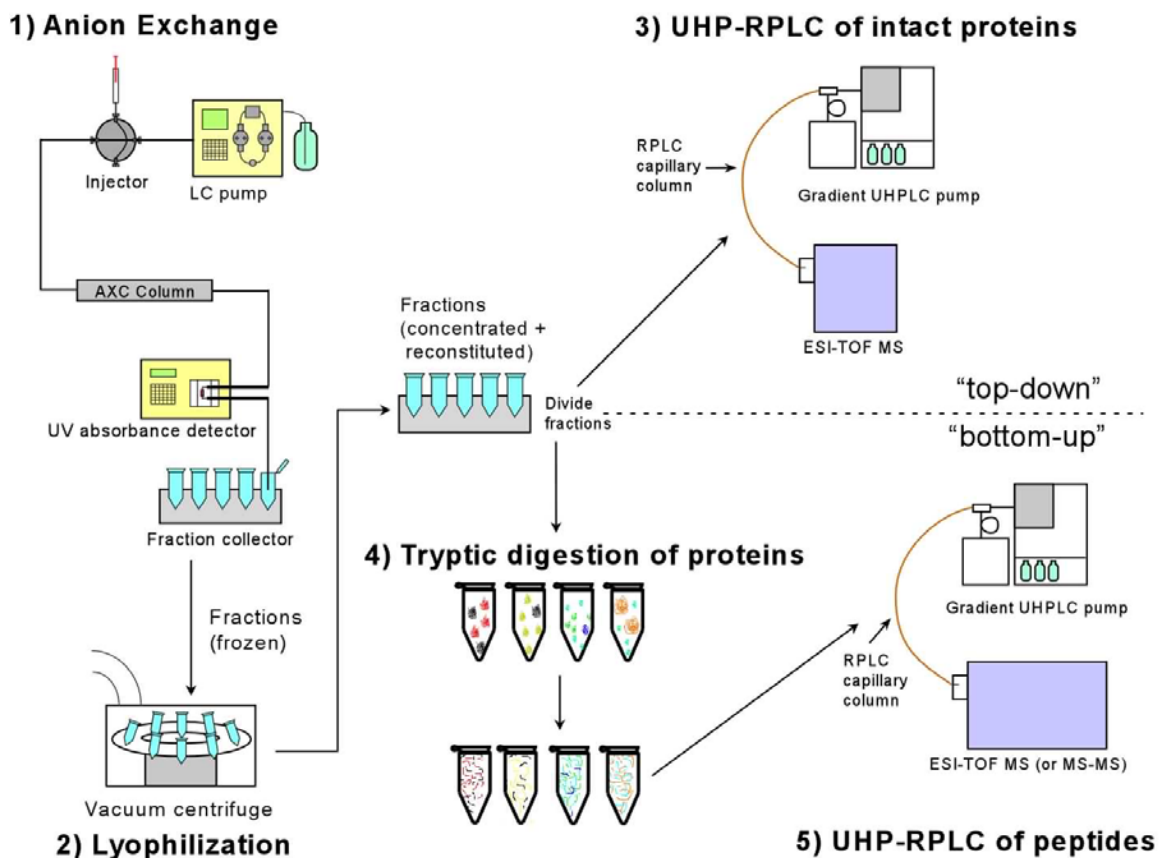
Figure 4-3: Overview of instrumentation and procedure used for hybrid top-down / bottom-up LC x UHPLC-MS analysis of complex protein mixtures
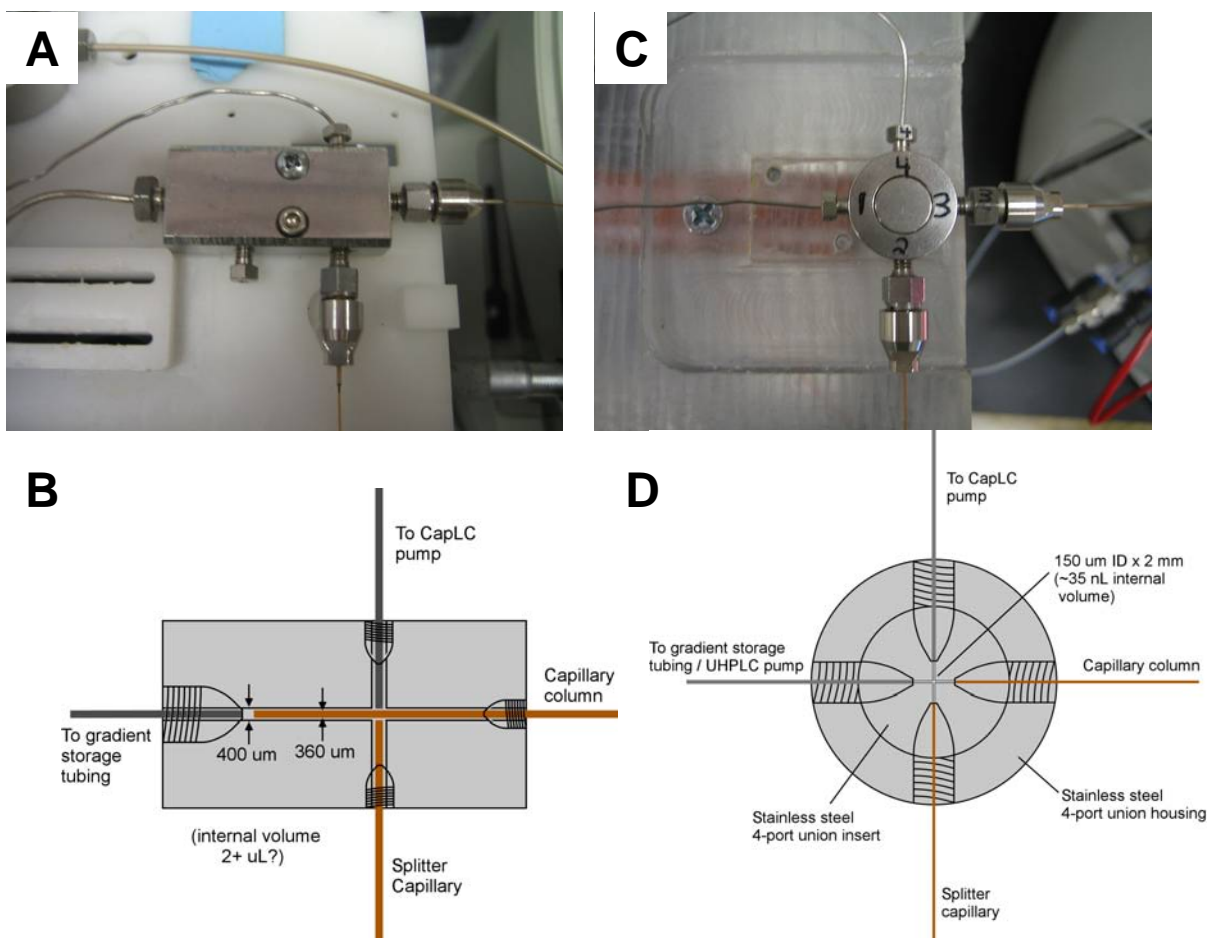
Figure 4-4: Comparison of the old high pressure 4-port union (A & B) with the new one (C & D). The new union has a much lower internal dead volume, which reduces the delay between the start of the run and when the first peaks appear. Both unions use the same 1/32" fittings to hold fused silica capillary. A complete diagram of the pump can be found in Figure 3-4.
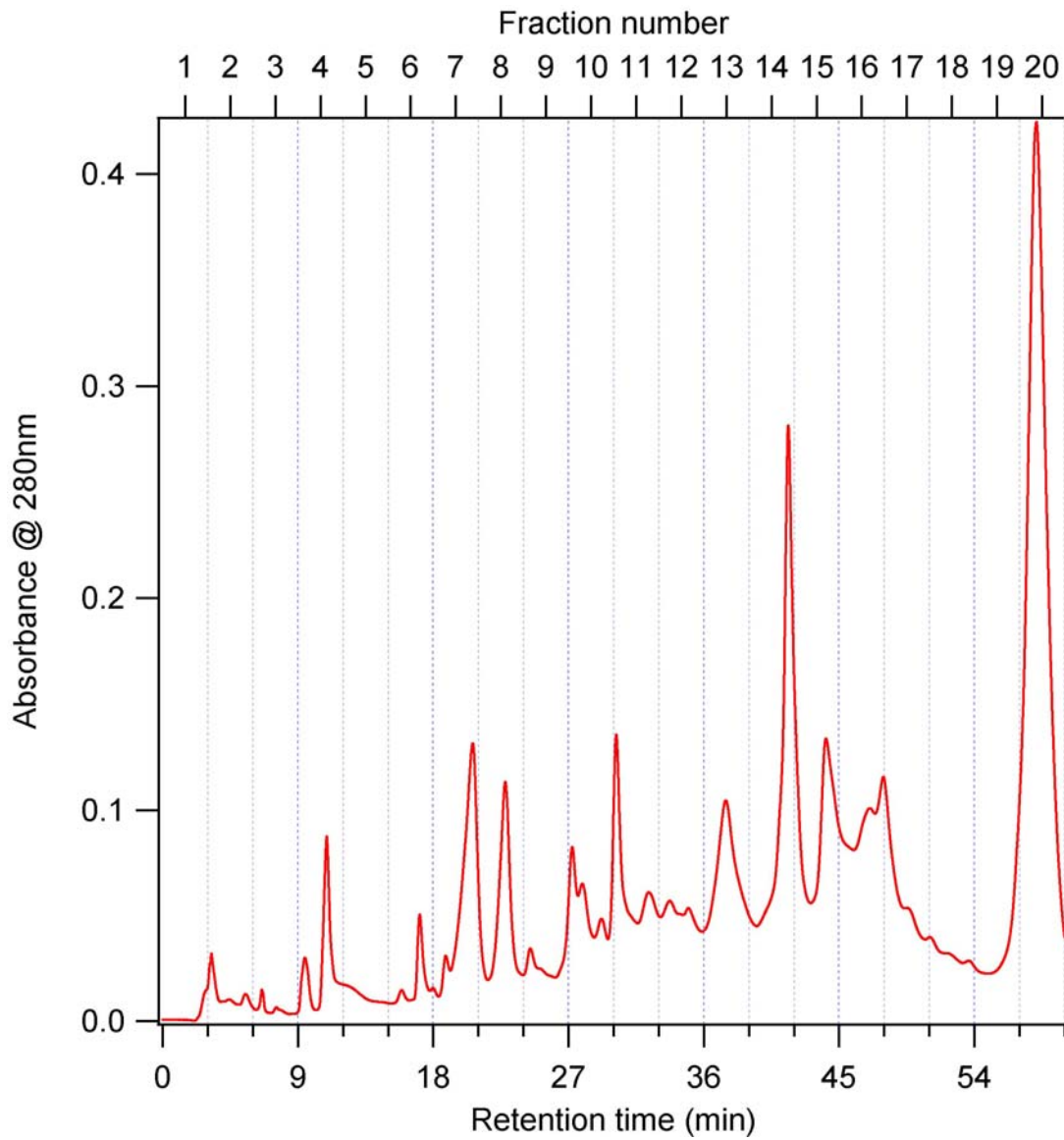
Figure 4-5: First dimension anion exchange separation of the *E. coli* protein extract. The run was performed using a 60 minute linear gradient from 10 to 500 mM ammonium acetate (pH 8.5). Detection was UV absorbance at 280 nm. A total of 20 fractions were collected.
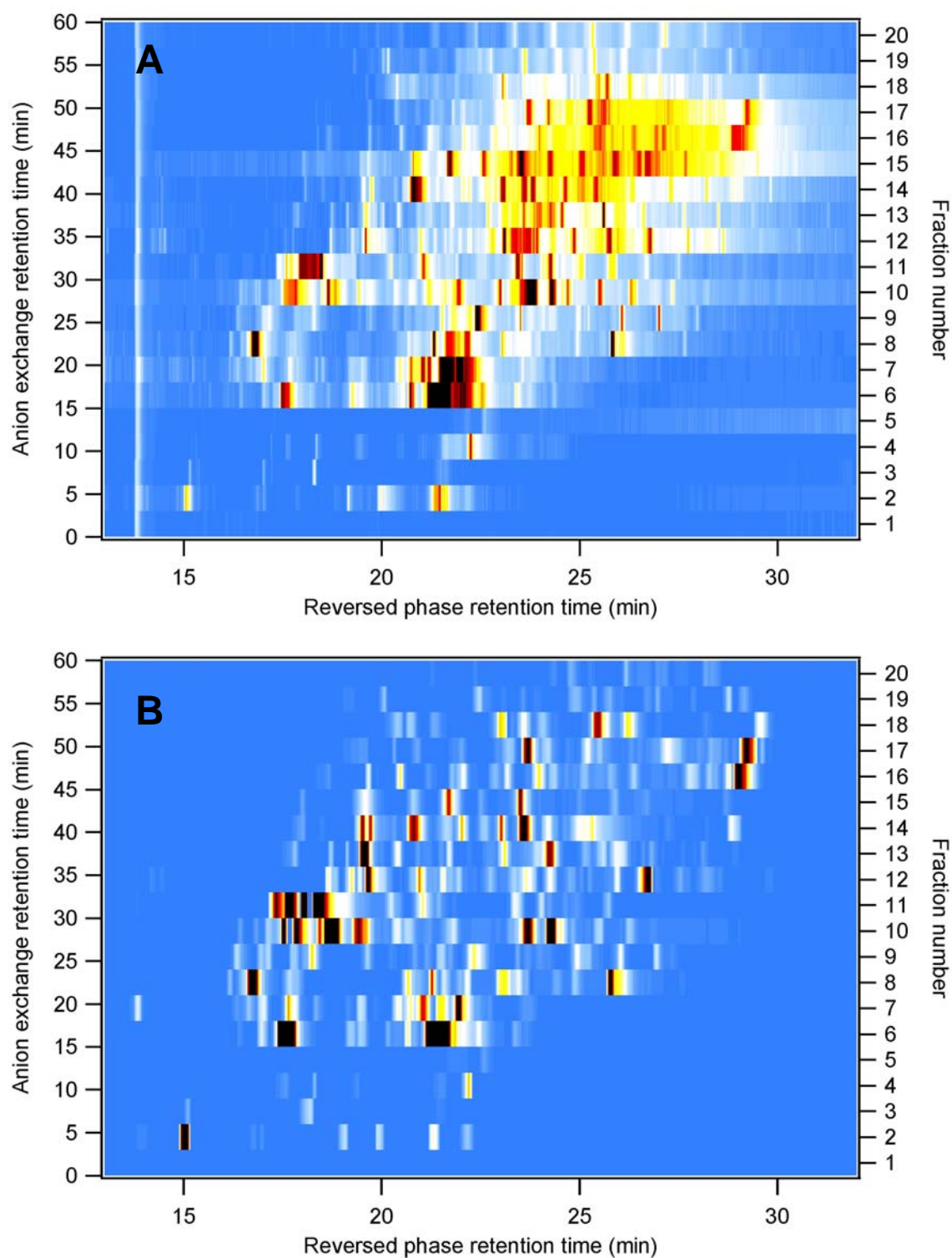
Figure 4-6: 2D chromatograms from an off-line AXC x RPLC separation of intact *E. coli* proteins. (A) was generated using the TIC data; (B) is a base peak chromatogram for the same run, generated using AutoME de-convoluted data. Run conditions are defined in the text.

167
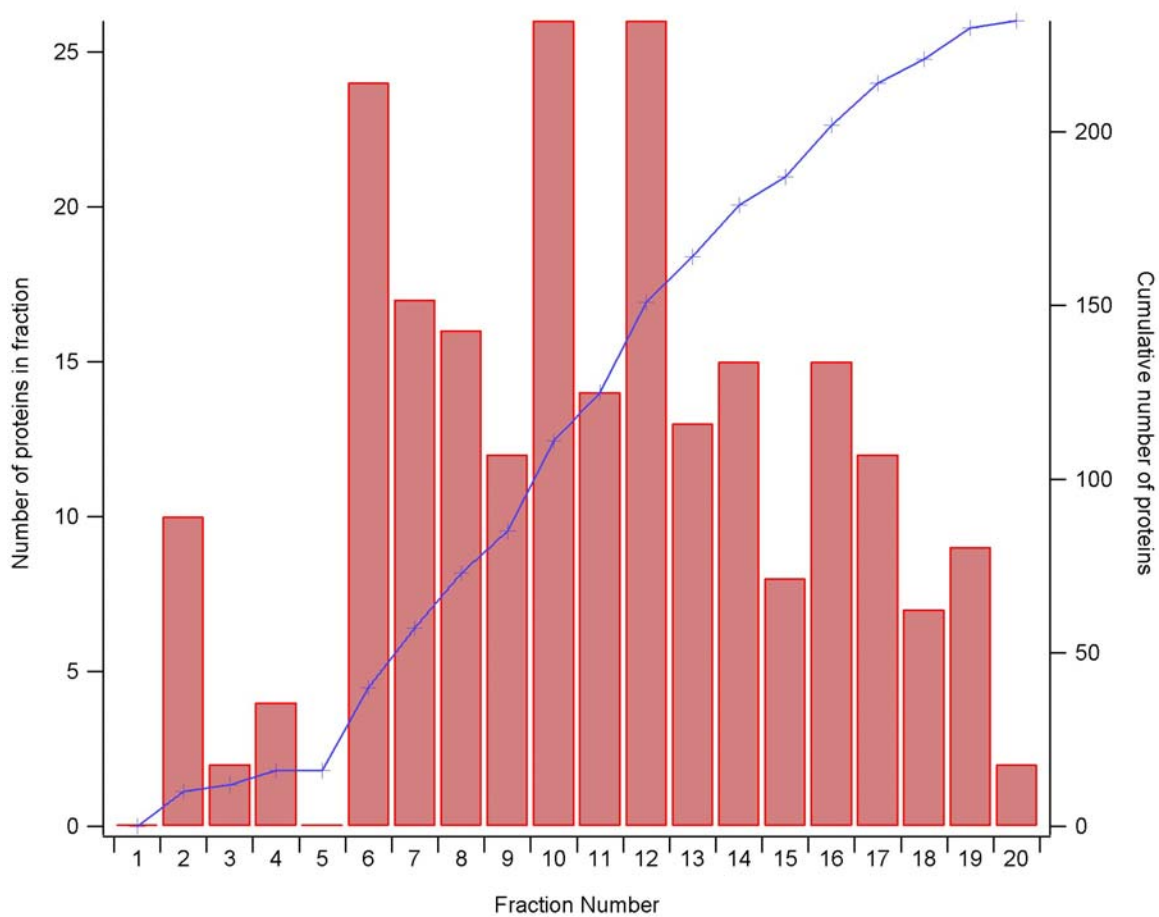
Figure 4-7: Number of proteins found in each anion exchange fraction for the 2D separation of the *E. coli* protein extract.
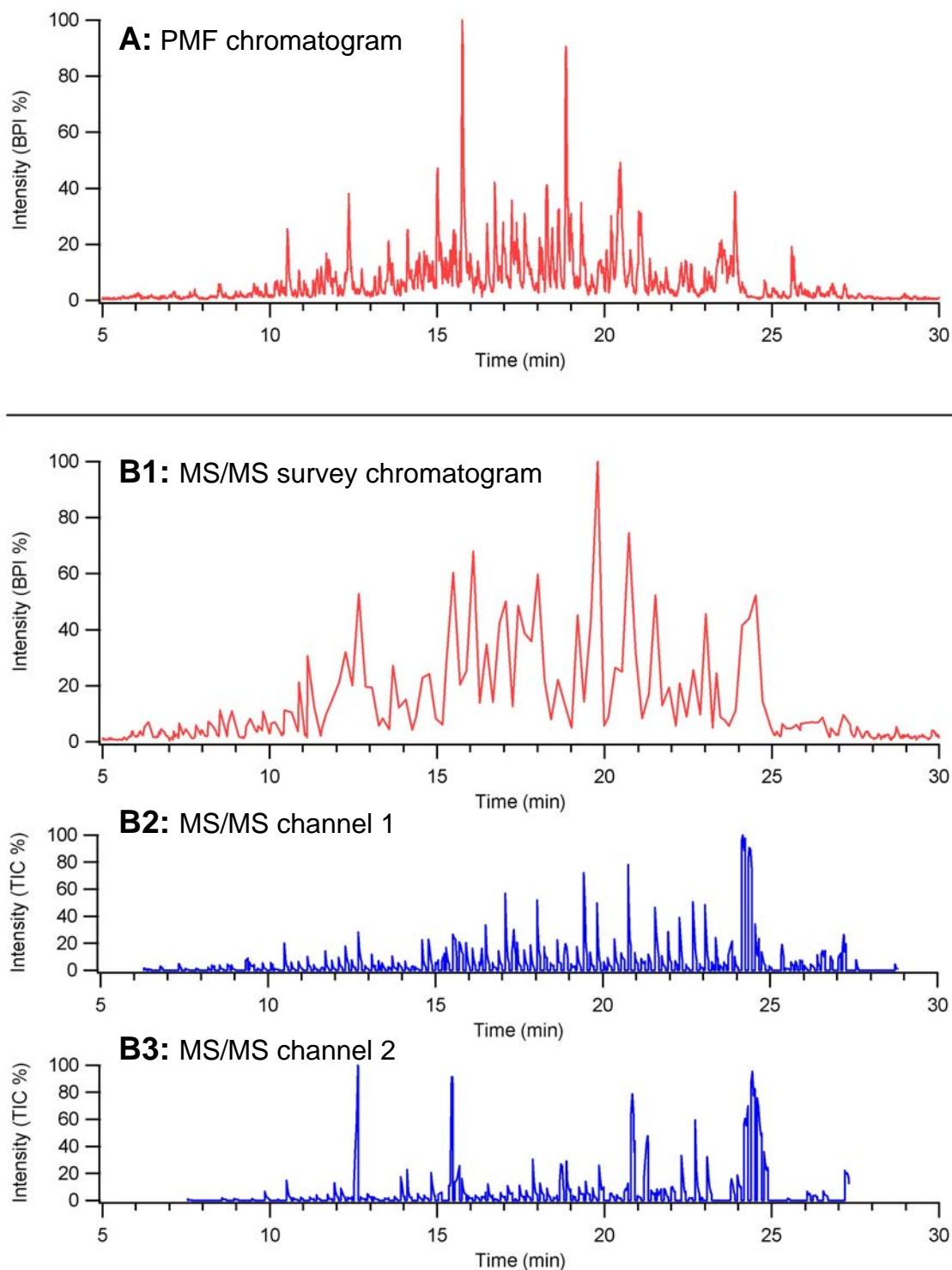
Figure 4-8: Comparison of PMF and MS/MS chromatograms of the same *E. coli* protein digest sample. (A) Is the chromatogram from the PMF experiment; (B1) is the survey chromatogram from the MS/MS experiment. (B2) and (B3) are chromatographic representations of the two MS/MS channels from the same run as (B1); they indicate when MS/MS was performed and the total ion intensity of the component being scanned. Run conditions are defined in the text.
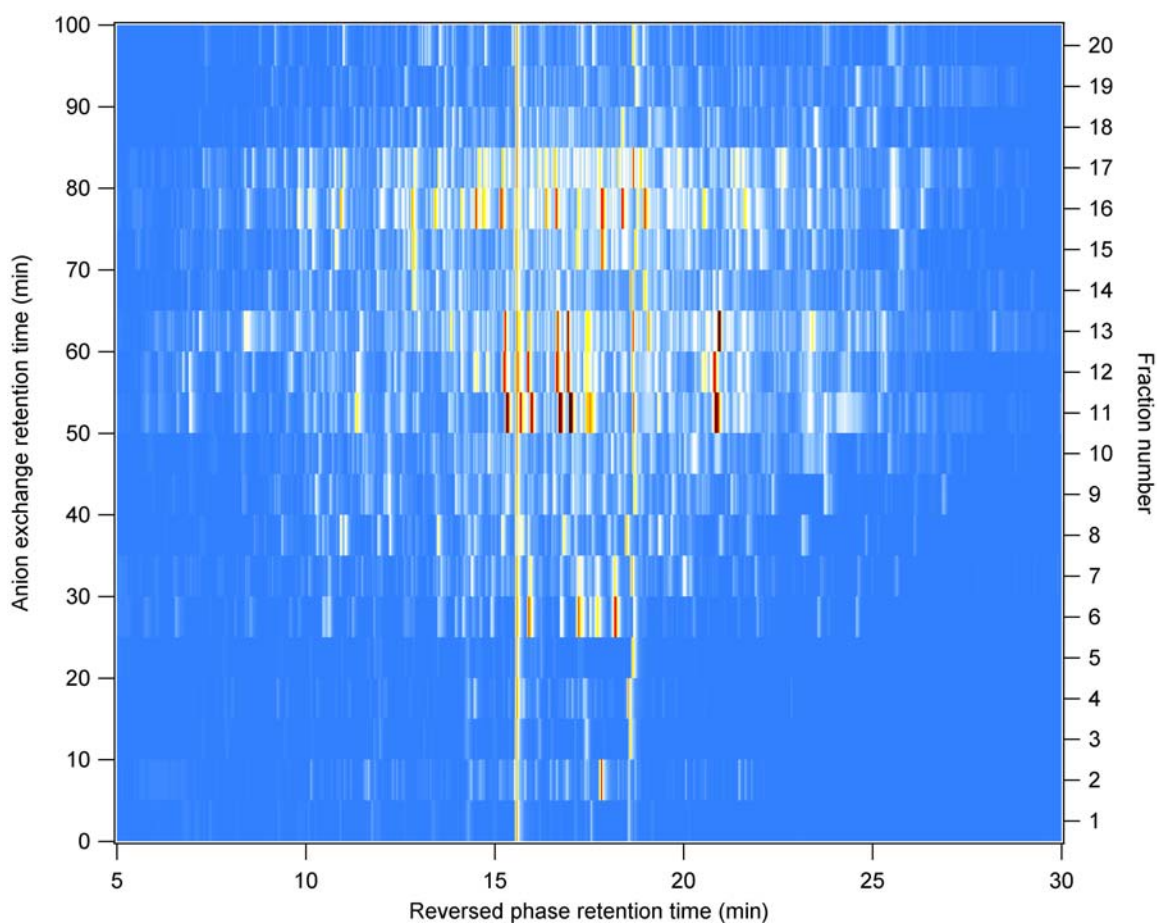
Figure 4-9: 2D chromatogram of an off-line AXC x RPLC separation of peptides from trypsin-digested *E. coli* proteins. The chromatogram was generated using the UHP-RPLC runs from the PMF experiment. Run conditions are defined in the text.
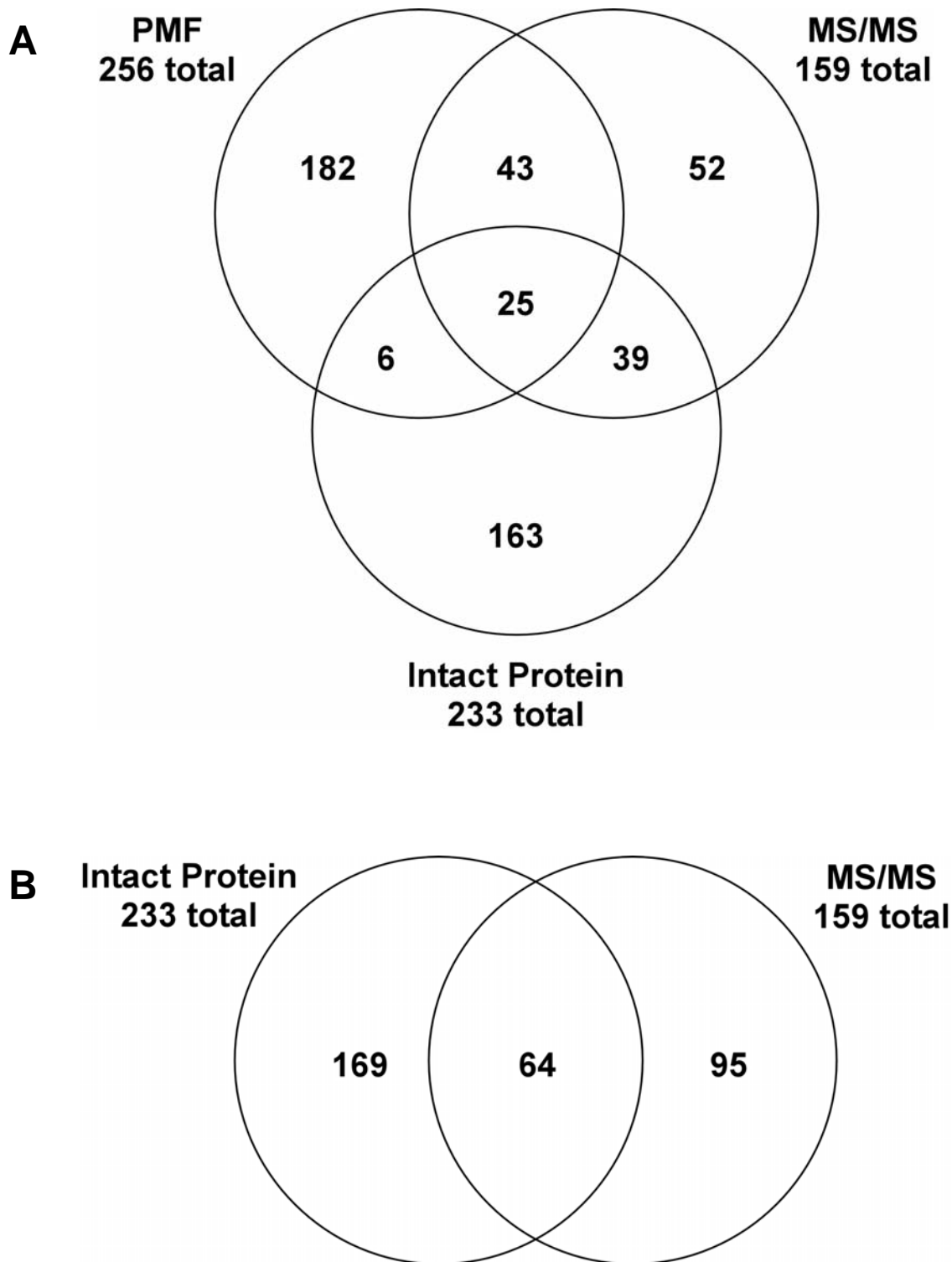
Figure 4-10: Venn diagrams illustrating overlap of *E. coli* proteins detected using bottom up (PMF and MS/MS) and top-down (intact protein) analyses. (A) Compares all three of the methods; (B) is a simplified diagram comparing only intact protein and MS/MS strategies
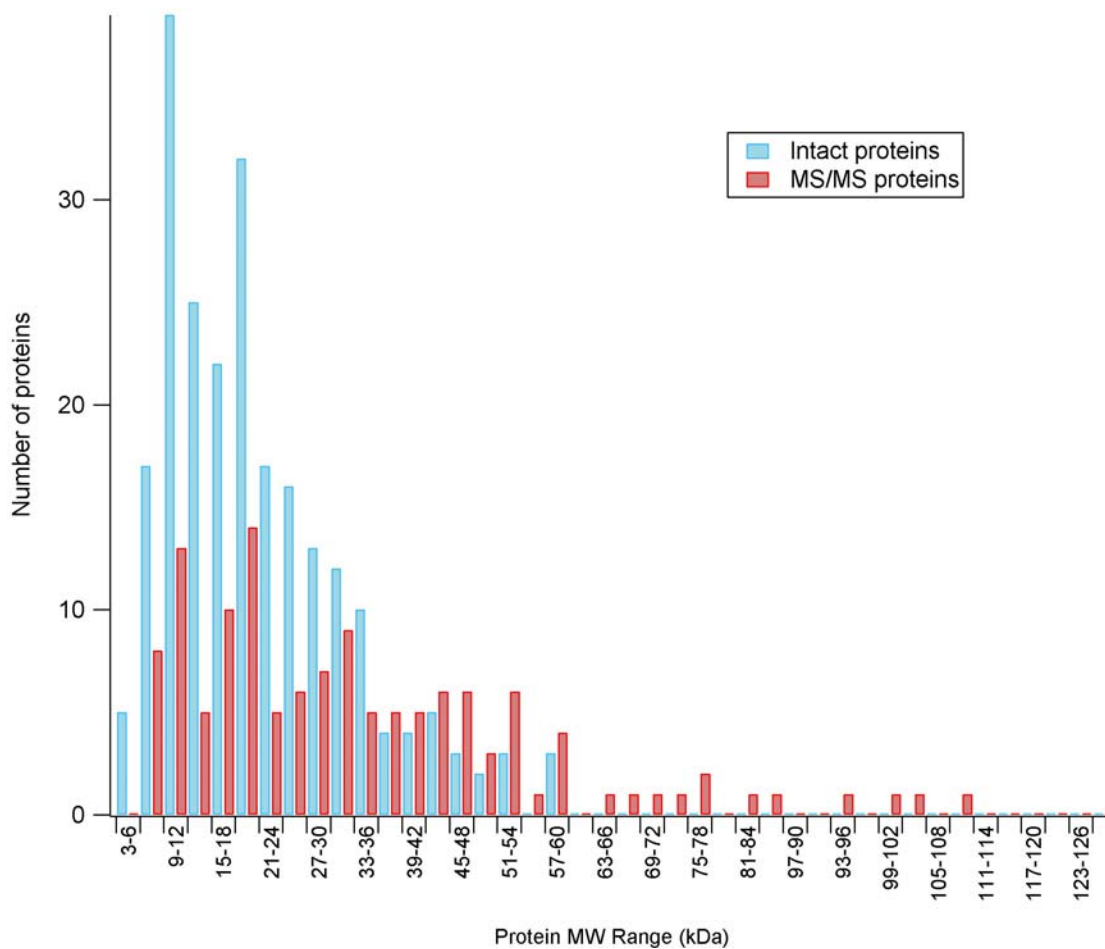
Figure 4-11: Molecular weight distribution of *E. coli* proteins detected from the intact protein and peptide MS/MS data
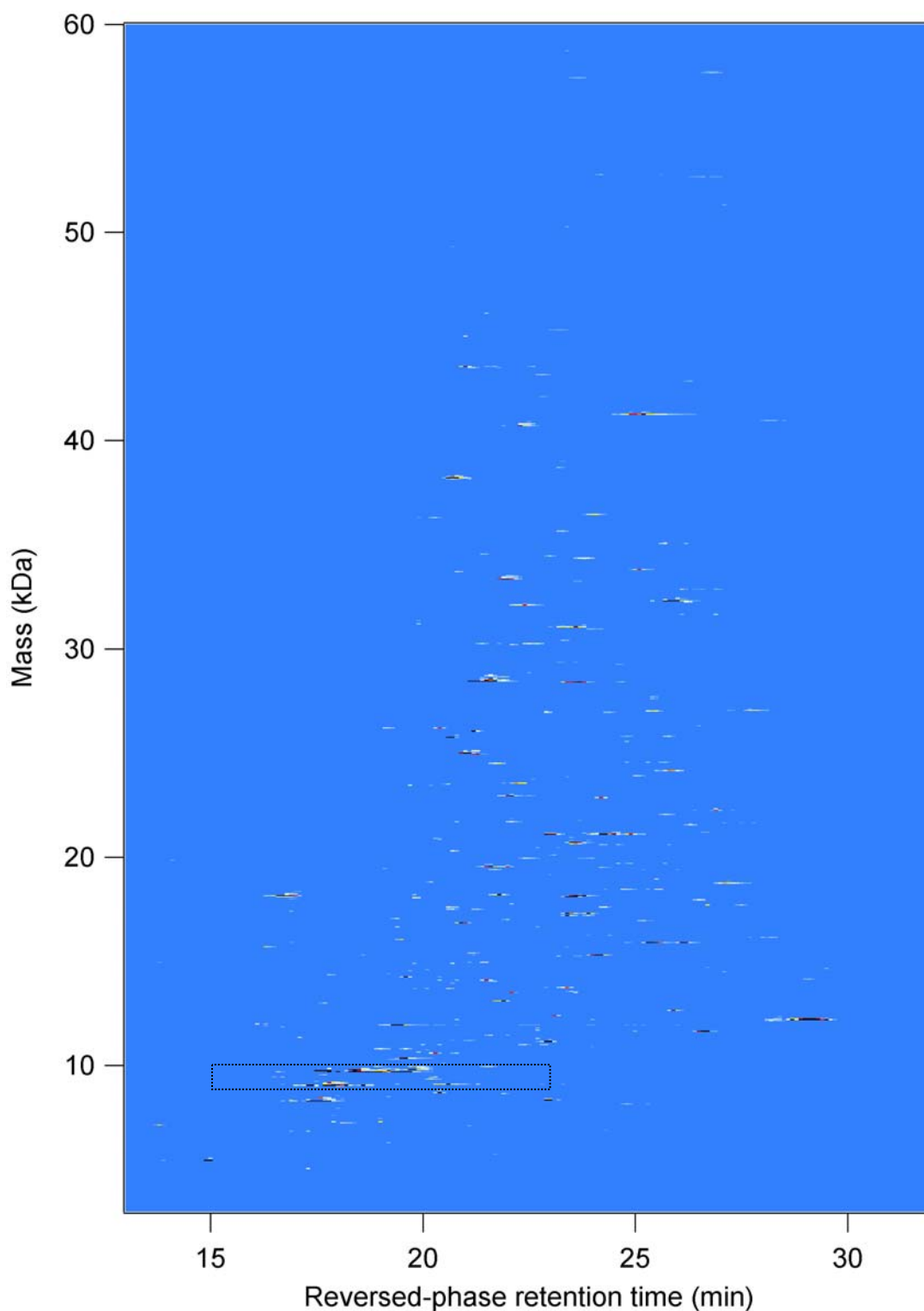
Figure 4-12: "Low-resolution" chromatospectrum of the full dataset from a 2D AXC x RPLC separation of intact *E. coli* proteins. Data from all anion exchange fractions were summed. MW Range: 5-60 kDa, RT range: 13-32 min, bin size: 100 Da. Total data points: 104,550. The area within the box is expanded in Figure 4-13.
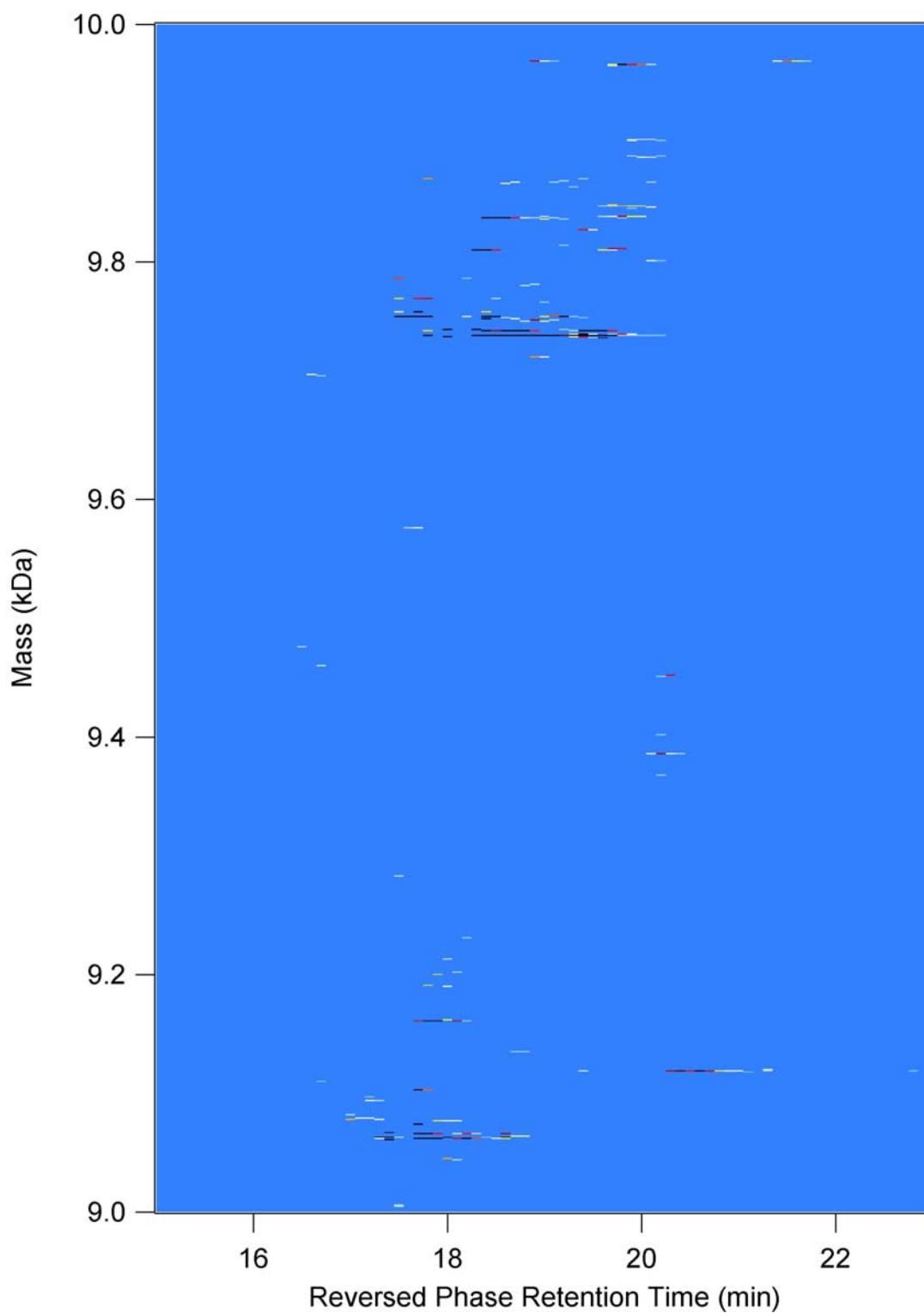
Figure 4-13: "High-resolution" chromatospectrum of a selected time/mass window (see Figure 4-12) from a 2D AXC x RPLC separation of intact *E. coli* proteins. MW Range: 9-10 kDa, RT range: 15-23 min, bin size: 1 Da. Total data points: 80,000.

CHAPTER 5: On-line LC x LC-MS of intact proteins, followed by peptide LC-MS/MS: an extension of the top down / bottom up method

5.1   Introduction

The work described in Chapter 4 demonstrated that a hybrid of top down and bottom up proteomics methods can provide useful information when analyzing a complex sample. However, the data obtained suggested that the technique needed higher chromatographic resolution of intact proteins prior to digestion in order to reduce the complexity of the fractions and improve the success of peptide identification using MS/MS. To do so without an unreasonable increase in analysis time necessitates a substantial change in approach. On-line LC x LC is a logical option, because it can substantially reduce analysis time while maintaining or improving separation performance. A switch to on-line coupling of dimensions is not without tradeoffs, however. It requires substantially more careful optimization of the timing of each dimension than off-line methods. Also, as discussed previously, it is not yet practical to use ultra-high pressure separations in on-line LC x LC. In spite of the drawbacks, it was judged that there were sufficient advantages to merit moving to an on-line approach.

In this chapter, a method is reported which uses on-line LC x LC for intact protein separations, followed by fraction collection and bottom-up protein analysis. This technique differs from the research presented in Chapter 4 in two main ways: first, the intact protein separation is carried out using on-line coupling of dimensions as opposed to off-line, and second, fraction collection and trypsin digestion occur after two dimensions of LC separation

as opposed to just one. Bottom-up analysis of peptides using LC-MS/MS remains a part of the technique; however, the coupling between the top-down and bottom-up portion of the analysis is still off-line via fraction collection. The methods which were developed were applied to analysis of the *E. coli* proteome. Results from these studies are presented and compared with the findings from the off-line LC x LC method reported in Chapter 4.

5.2   Experimental methods

5.2.1 Overview

An overview of the instrumentation and procedure used to perform an on-line 2D separation of intact proteins and bottom-up analysis of protein digests is shown in Figure 5-1. The first dimension of the on-line LC x LC separation is performed using gradient anion exchange chromatography. The effluent from the anion exchange column enters a 10-port switching valve. The valve directs the AXC column effluent to the front of one of two reversed-phase columns, which serve as the second dimension of the 2D separation. Any proteins eluting from the AXC column are trapped at the front of the RPLC column until the valve is switched, at which point they are eluted using a reversed-phase gradient. The valve is switched multiple times over the course of a 2D run, which allows the second dimension to sample the first multiple times. For some experiments, a small portion of the effluent from the reversed-phase column was diverted to an ESI-quadrupole mass spectrometer to obtain intact protein MW information. For all runs, the majority of the effluent was sent to a UV absorbance detector, then to a fraction collector. For the bottom-up portion of the experiment, the fractions are lyophilized in a vacuum centrifuge, and then digested using trypsin to produce peptides. They are then analyzed using capillary RPLC-MS/MS and the resulting spectra are searched against a database in order to identify the proteins present in the sample.

5.2.2 Chemicals and samples

Chemicals used for LC mobile phase were ammonium acetate, formic acid and trifluoroacetic acid (ACS reagent grade), which were purchased from Sigma-Aldrich (St. Louis, MO), and ammonium hydroxide (ACS reagent grade) and acetonitrile (HPLC grade), which were purchased from Fisher Scientific (Fair Lawn, NJ). Deionized water was purified using a Barnstead Nanopure System (Boston, MA). Chemicals used in the trypsin digest procedure were: ammonium bicarbonate (Sigma-Aldrich, St. Louis, MO), RapiGest SF acid-labile surfactant (Waters Corporation, Milford, MA), dithiothreitol (Research Products International, Mt. Prospect, IL), iodoacetamide (Sigma-Aldrich), TPCK-modified trypsin (Pierce, Rockford, IL), and trifluoroacetic acid (Sigma-Aldrich).

The samples which were analyzed consisted of both standard proteins and cellular protein extracts. The standard proteins – lysozyme, ribonuclease A, cytochrome C, myoglobin, β-lactoglobulin, bovine serum albumin, and ovalbumin – were purchased from Sigma-Aldrich and were used as received. The *Escherichia coli* protein extract analyzed in this experiment was the same sample described in Chapter 4. Prior to injection, the sample was diluted to 20 mg/mL in 10 mM ammonium acetate (pH 8.5).

5.2.3 Evaluation of commercial columns for RPLC of intact proteins

In previous chapters, all reversed-phase separations of proteins were performed using capillary columns operated at ultra-high pressures. It was not possible to use these columns in an on-line LC x LC system because automated switching valves which could be operated at pressures above 1400 bar (20 kpsi) were not available. Also, it was desired to collect fractions from the effluent of the second dimension columns. Therefore, even if suitable switching valves were available, the capillary columns used in UHPLC would not be practical because they operate at flow rates which are too low for fraction collection.

177

Consequently, it was necessary to use conventional-diameter, conventional-pressure reversed phase HPLC columns. Therefore, two different commercial RPLC columns were evaluated under controlled conditions before selecting a column to use for on-line 2D separations of complex samples. The goal of these tests was to determine if satisfactory separations could be attained, and which type of column offered the best performance.

Reversed-phase columns marketed as appropriate for protein separations were obtained from Waters Corporation. The first column was the Waters Biosuite pPhenyl 10 μm, which is 7.5 cm long x 4.6 mm ID. It is packed with 10 μm polymeric particles with 1000 Å pores, which are the same base particle as those found in the Biosuite AXC column. The particles are bonded with phenyl functional groups to give the stationary phase a hydrophobic surface. The second column was a Waters Symmetry300 C4, which is 5 cm long x 4.6 mm ID and is packed with 5 μm silica particles with 300 Å pores. The bonded stationary phase contains n-butyl groups. A Waters 600E quaternary gradient LC pump was used to produce the reversed-phase gradients. The pump is connected to a Valco 6-port valve (VICI, Houston, TX) with a sample loop, which is used to perform injections onto the column. Detection was performed using an Applied Biosystems 785A UV absorbance detector (Foster City, CA) set at 280 nm.

Columns were tested using a mixture of 7 standard proteins. Each was dissolved at a concentration of 0.1 mg/mL in starting mobile phase; the sample injection amount was typically 20 μL. In addition to the standard protein sample, the *E. coli* protein extract was also used to evaluate the performance of the columns with a more complex mixture. The mobile phase composition and gradient program for the separations were varied in order to

determine the method which gave the best performance; specific run conditions are described with the results presented in Section 5.3.

5.2.4 On-line LC x LC: instrumentation

Once an RPLC column was selected, the on-line LC x LC system was constructed. A diagram illustrating the components of the system and how they are operated is shown in Figure 5-2. All fluidic connections are made using 0.005" (127 μm) ID PEEK tubing. The first dimension of the 2D separation is performed using anion exchange chromatography. Initially, the Waters Biosuite Q 10 μm AXC column (7.5 cm x 7.5 mm ID) was used. In order to improve resolution, it was later replaced with the custom-packed long anion exchange column (111 cm x 6.6 mm ID). Sample is injected onto the first dimension column using a Valco 6-port valve (VICI, Houston, TX) with a sample loop. The anion exchange gradient is generated by a Waters 600E quaternary gradient LC pump.

The effluent from the AXC column enters a two-position, 10-port valve (Cheminert C2H-1000EH, VICI). From there, it is directed onto the head of the first of two identical reversed-phase columns (labeled column 1 in Figure 5-2A). Since the mobile phase from the AXC column is 100% aqueous, all proteins are retained at the front of the reversed-phase column. The residual anion exchange mobile phase passes through the column and back into the 10-port valve, where it is diverted to waste. At the same time as sample is being loaded onto column 1, the inlet of RPLC column 2 is connected to a second LC pump. This pump, which is a Hewlett Packard model 1050 HPLC (Agilent, Santa Clara, CA), is used to generate gradients to elute proteins from the RPLC columns. After a set period of time – typically 30 minutes – the 10-port valve is switched, which directs the effluent of the AXC column to RPLC column 2. Meanwhile, the proteins which had previously been loaded onto RPLC column 1 are eluted using a gradient generated by the second LC pump. In order to

perform a full 2D run, the process of loading one column while eluting the other is repeated

multiple times until all components have eluted from the AXC column and have been

separated on one of the RPLC columns.

In the first run performed on the 2D system, all of the effluent from the RPLC column

being eluted was directed to a UV absorbance detector set at 215 nm (Applied Biosystems

785A, Foster City, CA) and then to a fraction collector (Waters Fraction Collector II). In

later analyses, approximately 1/20th of the RPLC column effluent was diverted to a Waters

ZQ mass spectrometer for on-line ESI-MS analysis. The mass spectrometer was operated

using a standard z-spray source in the positive ion mode, with a capillary voltage of +3000V

and a sample cone voltage of +35V. Data were acquired at a rate of 1 sample/sec over a

range of 450-1600 m/z. The ZQ uses a single quadrupole mass analyzer, which gives lower

resolution than a TOF instrument, although it is adequate for measuring intact protein MW.

The instrument was calibrated regularly using a solution of sodium formate.

5.2.5   On-line LC x LC: software control and timing

In order for the LC x LC system to work properly, the timing of many components

must be precisely controlled. This was accomplished using a computerized control program

called a virtual instrument (VI) written in-house using LabVIEW 6.1 (National Instruments,

Austin, TX). This program coordinated the timing of the sample injection, valve switch

events, gradient start signals, data acquisition, and fraction collection. Although the

LabVIEW program controls the timing of the entire LC x LC analysis, two other software

programs were used for data acquisition. A modified version of Stripchart Recorder XP

(another custom LabVIEW VI, written by Matthew Monroe) was used to record the signal

from the UV detector, and MassLynx 4.1 (Waters Corp.) was used to acquire LC-MS data.

5.2.6   On-line LC x LC: run conditions

For the anion exchange separation, mobile phase A was 10 mM ammonium acetate, adjusted to pH 8.5 with ammonium hydroxide. Mobile phase B was 1M ammonium acetate, adjusted to pH 8.5 with ammonium hydroxide. For the first 2D separation, which used UV detection only, reversed-phase mobile phase A was deionized water with 0.1% trifluoroacetic acid, and B was acetonitrile with 0.1% trifluoroacetic acid. When MS detection was used, the ion pairing agent was changed to 0.2% formic acid, since TFA is known to cause ion suppression in ESI-MS.[1]

Selection of run lengths and gradient profiles is a more complex process for on-line LC x LC than for off-line. This is because the second dimension separations must be much faster than the first so that the effluent of the first dimension can be sampled multiple times over the course of the 2D run.[2] Since the second dimension separation speed is the limiting factor for the length of the entire analysis, it was desired to keep the RPLC run time as low as possible. Evaluations of the RPLC columns had determined that a good compromise between speed and resolution was reached when they were operated at a flow rate of 0.5 mL/min and with a turn-around time of 30 minutes between runs. In order to allow for a post-gradient column wash and full re-equilibration to initial conditions within this time period, a reversed-phase gradient length of 20 minutes was used. The length of the anion exchange gradient was set so that an adequate number of fractions could be transferred from the first dimension to the second. For the short anion exchange column, it was desired to transfer a total of 10-12 fractions, so a 300 minute anion exchange gradient was used. The flow rate was set at 0.1 mL/min to keep the gradient volume near the same as previous runs using this column. For the long anion exchange column, it was decided to analyze a total of 30 fractions, so the gradient length was increased to 900 minutes. The flow rate used for this

column was also 0.1 mL/min. Several different anion exchange and reversed-phase gradient profiles were used for the LC x LC runs reported in this chapter. AXC gradient profiles are displayed in Table 5-1 and Table 5-2; reversed-phase gradient profiles are displayed in Table 5-3 and Table 5-4. The specific profile used for a given 2D run is reported with the results in Section 5.3.

5.2.7 Fraction collection, lyophilization, and trypsin digest

Some of the 2D runs reported in this chapter were performed solely to investigate the capabilities of the online LC x LC system for intact protein separations. In these runs, no fractions were collected, and bottom-up proteome analysis was not performed. For the one run in which fraction collection was performed, the goal was to spread the proteins out over as many fractions as possible, and to avoid collecting fractions in empty regions of the chromatogram. In trial runs of reversed-phase separations of the *E. coli* protein extract, no protein peaks were typically observed before 7 minutes into the gradient or after 23 minutes. Therefore, the fraction collector was programmed to delay starting the collection of fractions until 7 minutes after the start of every reversed-phase gradient, and to change tubes every 1.5 minutes until 23.5 min. This yielded 11 fractions per reversed-phase run, with a volume of 0.75 mL each. A total of 11 reversed phase runs were performed over the course of the 2D separation, so the total number of fractions collected was 121. Once all fractions were collected, they were flash-frozen using liquid nitrogen and then lyophilization was performed using a vacuum centrifuge as described in Section 4.2.4. Once completely dry, the fractions were reconstituted using 25 μL of 50 mM ammonium bicarbonate. The proteins in the fractions were then reduced, alkylated, and digested using trypsin according to the procedure explained in Section 4.2.4.

5.2.8 LC-MS/MS of protein digests

Capillary LC was performed using a conventional-pressure CapLC pump (Waters Corp.). A diagram of the instrument is shown in Figure 5-3. Its operation is quite straightforward compared to the stored-gradient UHPLC pump described previously. The autosampler injects the desired amount of sample onto the column, and then a real-time gradient is generated to elute the sample components. The column used was a Waters Symmetry C18, which is 15 cm long and has an ID of 320 μm. It is packed with 5 μm silica particles with 100 Å pores which are bonded with a C18 stationary phase. Although referred to as a capillary column, its inner diameter is several times larger than the custom-packed capillary columns described previously in this dissertation. Typical flow rates used with this column were 5-10 μL/min; therefore, fluidic connections between system components could be made with 0.0025" (67 μm) ID PEEK tubing without causing excessive band broadening.

For the RPLC peptide separations, mobile phase A was deionized water with 0.2% formic acid and mobile phase B was HPLC-grade acetonitrile with 0.2% formic acid. A 30 minute gradient from 1-30% B was used to separate the peptides, followed by a 5 minute column wash at 75% B, after which the mobile phase was returned to 1% B. Typically 10 μL of sample were injected onto the column. This is a large volume of sample relative to the column diameter; however, it was necessary in order to obtain adequate MS/MS signal intensity for sample components over a wide range of concentrations.

For experiments in which UV detection was performed, the outlet of the column was coupled to a CapLC photodiode array (PDA) detector, which is an integrated part of the CapLC instrument. This detector allows UV absorbance to be monitored over a range of wavelengths from 190 - 400 nm. The detector flow cell has an internal volume of 250 nL, which is small enough to prevent significant band broadening from harming the resolution of

peaks as they pass through the cell. The effluent from the detector was coupled to the inlet of the Q-TOF Micro mass spectrometer (Waters Corp.). For those experiments in which UV detection was not performed, the outlet of the column was connected directly to the MS. The MS was operated using a standard z-spray source in the positive ion mode, with a capillary voltage of +3000V and a sample cone voltage of +40V. The MS source temperature was 100º C and the desolvation gas flow rate and temperature were set at 250 L/hr and 120º C, respectively. MS/MS analysis of the components eluting from the column were performed using a data-directed analysis (DDA) method. The workflow of the DDA experiment was the same as described in Section 4.2.6.2.

5.2.9   Data analysis

For the LC x LC intact protein separation in which no MS data were acquired, a 2D chromatogram was generated using UV absorbance data. For the other LC x LC runs, 2D chromatograms were generated from the LC-MS data which was de-convoluted using AutoME as described in Section 3.2.7. AutoME processing parameters were set as follows: the time segment width was 6 seconds, the output spectrum resolution was 1Da, the range of masses in the output spectrum was 5,000-80,000 Da, and the maximum number of iterations was 50. The de-convoluted data were also used to generate a list of the molecular weights of intact proteins found in the sample. ProteinLynx Global Server 2.0 (Waters Corp.) was used to process LC-MS/MS data from runs of the trypsin-digested fractions and to search the MS/MS spectra against the SwissProt database. A more detailed account of this process can be found in Section 4.2.7.2.

5.3   Results and discussion

5.3.1   Evaluation of commercial columns for RPLC of intact proteins

Reversed-phase separations of intact proteins are sometimes associated with negative attributes such as asymmetrical peak shape, poor sample recovery, and sample carryover between runs.[3-7] These occurrences may be due to undesired interactions between the proteins and chemical moieties in the stationary phase, such as residual silanol groups on silica particles or metal contaminants.[8] Since the chemistry of the stationary phase can have a substantial impact on separation quality, two different types of reversed-phase columns were tested to assess their performance for protein separations. The first column – the Waters Biosuite pPhenyl – contained methacrylate-based polymeric particles bonded with a phenyl stationary phase. The second – a Waters Symmetry300 – contained silica particles with a C4 stationary phase.

The columns were first evaluated using a mixture of seven standard proteins, which are listed in Section 5.2.2. Separations were performed using 15 and 30 minute gradients from 10-70% acetonitrile with 0.1% trifluoroacetic acid. Chromatograms from these runs are shown in Figure 5-4. Peak capacities were computed for all of the separations and are displayed along with the chromatograms in the figure. For both columns, peak capacity increased by approximately 45% when the gradient was lengthened from 15 to 30 minutes. Of the two columns, the Symmetry offers better chromatographic performance – it gives peak capacities approximately 50% higher than the Biosuite column for equivalent gradient lengths. This is not surprising, since the Symmetry column contains 5 μm particles while the Biosuite column uses 10 μm particles. The columns were also evaluated by performing separations of an *E. coli* protein extract using a 15 minute gradient, as shown in Figure 5-5. Due to the complexity of the sample, neither column is able to resolve the components into

185

individual peaks. However, the chromatogram produced by the Symmetry column does show more detail, due to a smaller average peak width.

In general, these data suggest that the Symmetry column produces higher resolution separations of proteins than the Biosuite column. However, closer examination of the chromatograms in Figure 5-4 reveals a concern related to the performance of the Symmetry column. The Biosuite column produces seven distinct, relatively symmetrical peaks, plus one minor peak, which is most likely a contaminant from one of the protein standards. Thus, the number of major peaks matches the number of components in the sample. The chromatograms from the Symmetry column show four sharp peaks, two somewhat broader peaks, and a variety of other peaks too small to be one of the main sample components. Thus, one of the seven proteins in the sample appears to be missing from the chromatogram. Also, some of the peaks are somewhat asymmetrical. These observations indicate that non-ideal interactions with some of the proteins may be occurring with the silica-based Symmetry column. Therefore, the polymer-based Biosuite column was selected to be used for the second dimension of the on-line 2D system, because of the likelihood that it will produce more consistent separations of a wide range of proteins.

5.3.2   On-line LC x LC of intact proteins: chromatographic results

Once the reversed-phase columns had been selected and the on-line LC x LC instrument was prepared, a 2D separation of the *E. coli* protein extract was performed. Anion exchange run conditions for this separation are specified in Table 5-1; reversed-phase run conditions are in Table 5-3. For this run, no MS detector was used; only UV absorbance data were obtained for the separation. The UV data were used to generate a 2D chromatogram, which is shown in Figure 5-6. Peaks in the chromatogram appear to be spread over a relatively wide range of the separation space, although the last two anion

186

exchange fractions contained relatively few proteins. The peak capacity of this separation

was estimated to be approximately 300, which is low compared to LC x LC separations

reported in previous chapters. The loss in peak capacity can be attributed to the low number

of anion exchange fractions and the reduced peak capacity of the reversed-phase separation

as compared to runs performed using ultra-high pressure LC. From a proteomic standpoint,

however, better use is made of the available peak capacity, since 121 fractions were collected

during the run, all of which were subsequently digested and analyzed using a bottom-up

method. This is 6 times more fractions than were collected and analyzed in the off-line LC x

LC experiment reported in Chapter 4. Another advantage of this method is the substantial

increase in speed. The entire 2D separation was performed in approximately 6 hours; an

equivalent separation would have taken several days to complete using the off-line method.

A second LC x LC separation of the *E. coli* protein extract was performed using

electrospray mass spectrometry as the detection method. Anion exchange run conditions for

this separation are specified in Table 5-1; reversed-phase run conditions are in Table 5-4. In

order to assess the reproducibility of the online 2D separation, this run was performed in

duplicate. Chromatograms for both of the runs are shown in Figure 5-7. Protein peaks

appear to be more well-defined in these chromatograms than in the UV chromatogram

(Figure 5-6) due to the greater sensitivity and selectivity of electrospray mass spectrometry

toward intact proteins, particularly when maximum entropy de-convolution is used to process

the data as described in Section 3.2.7. The peak capacity is essentially unchanged from the

previous LC x LC run, which is expected since the gradient lengths were the same.

Reversed-phase retention times were shifted lower by several minutes because the gradient

was altered slightly and the acidic modifier was changed from trifluoroacetic acid to formic

acid. Based on a visual comparison of the two chromatograms, reproducibility of the 2D

separation appears to be relatively good. Most peaks appear in the same position and have

approximately equal intensity. There are some instances, however, where a peak appears to

have shifted to an adjacent anion exchange fraction, or to be split between two fractions in

one chromatogram while appearing in only one fraction in the other. Although it is difficult

to ascertain the cause of these shifts with certainty, it is possible that a slight change in the

pH of the anion exchange mobile phase between 2D runs is responsible for the retention time

drift. No fraction collection was performed with these 2D separations.

A third on-line LC x LC separation was carried out using the long AXC column as

the first dimension. As before, the run was performed in duplicate. Anion exchange run

conditions for this separation are specified in Table 5-2; reversed-phase run conditions are in

Table 5-4. The improved resolution contributed by the longer anion exchange column

increased the peak capacity of the 2D separation to approximately 900. As before, the

general appearance of the two replicate chromatograms is similar, indicating that

reproducibility is fairly good. Closer inspection does reveal that a substantial number of

components shifted to adjacent anion exchange fractions, lending more evidence to the

hypothesis that a drift in pH of the anion exchange mobile phase is causing a shift in

retention of the proteins. Again, no fraction collection was performed following these 2D

separations. If it had been, the total number of fractions collected could have been well over

300, which illustrates the potential of the technique to divide the complex sample into fine

segments before bottom-up protein analysis.

5.3.3 On-line LC x LC of intact proteins: mass slice chromatograms

In the process of comparing chromatograms from replicate 2D runs, a useful way of

visually representing LC x LC-MS data was developed. The data can be divided into

segments based on mass ranges and then displayed as a series of 2D chromatograms. For example, one chromatogram could be generated using only components with a mass between 5 and 15 kDa, a second between 15 and 25 kDa, and so on. The data from the 2D separation of the *E. coli* protein extract shown in Figure 5-8A was processed using custom-written functions in Igor Pro (Wavemetrics, Lake Oswego, OR) to divide the data into 11 kDa mass ranges. The resulting data were displayed as a series of 2D plots shown in Figure 5-9, which are termed "mass slice chromatograms." The plots are analogous to a selected ion chromatogram (SIC) for LC-MS data, except they are presented in 2D format and contain a range of masses as opposed to a single mass. The first chromatogram on the page, Figure 5-9A, shows the full mass range data (3000-80,000 kDa). The remaining seven chromatograms, Figure 5-9B-H, display the same data divided into mass slices. The result is that the number of components which appear in any individual 2D chromatogram is reduced. Some of the mass slice chromatograms contain more peaks than others; this is because a large portion of the *E. coli* proteins detected have a mass between 10-35 kDa. Note that all peaks present in the full mass range chromatogram (A) can be found in one of the mass slice chromatograms (B-H).

The main use for the mass-slice chromatograms is anticipated to be in comparing sets of LC x LC data. When attempting to compare two peaks in a crowded chromatogram, it is often difficult to be certain whether the same components have been selected, especially considering the fact that retention times may have shifted slightly. When the chromatogram is divided into mass slices and only peaks within the same mass range are compared, the number of interfering peaks is reduced, which increases the reliability of the comparison. Also, mass-slice chromatograms may reveal peaks which are not visible in full mass range

189

chromatograms. This would be the case if two peaks with different masses co-eluted in terms of retention times on both the first and second LC dimensions. While such peaks could not be distinguished in a full-mass-range chromatogram, they would appear as separate components in the mass slice chromatograms. Finally, it should be noted that mass slices are not limited to ~10 kDa segments. The data could be divided into much finer bins (for example, 1kDa or 100Da) in order to further simplify comparison between chromatograms.

5.3.4 RPLC-MS/MS of peptides: chromatographic results

After the LC x LC separation with fraction collection was performed, fractions were digested using trypsin to produce peptides. The peptides in each fraction were separated and analyzed using LC-MS/MS. Due to the large number of fractions to be run – 121 in total – it was necessary to use instrumentation capable of automated, relatively fast runs. It was therefore decided to use a conventional-pressure (350 bar / 5000 psi) capillary LC system, which offered reduced separation performance compared to the stored gradient UHPLC system, but substantially faster sample throughput due to more reliable automation.

Two chromatograms from a RPLC separation of one of the trypsin-digested LC x LC fractions are shown in Figure 5-10. The first chromatogram, Figure 5-10A, displays UV absorbance at 193 nm.   This was found to be the most sensitive wavelength for detection of peptides. The peak capacity for this separation was estimated to be 55, which is 3 times less than the peak capacity obtained using gradient UHPLC reported in Chapter 4. The loss of peak capacity is not surprising, given that this column uses 5 μm particles, as opposed to the 1.5 μm particles used for UHPLC. The complexity of the sample somewhat exceeds the resolving capabilities of the column, as is indicated by numerous overlapping peaks distributed throughout the chromatogram and the absence of regions of flat baseline.

The second chromatogram from the same run, shown in Figure 5-10B, displays the base peak intensity from the MS survey scan. Due to the low sampling rate associated with data-directed MS/MS, the resolution of the chromatogram is poor. This makes it nearly impossible to correlate specific peaks in the UV chromatogram with peaks in the MS chromatogram. Protein identification was still possible, since it is based on the quality of the MS/MS spectra rather than chromatographic resolution. Nevertheless, better coverage of the proteins in the sample might have been obtained if a higher resolution separation were performed, because it would increase the number of different peptides that could be sampled using MS/MS.

5.3.5 Top-down / bottom-up *E. coli* proteome analysis: mass spectral results

To assess the effectiveness of the on-line LC x LC separation when applied to the field of proteomics, the MS data were analyzed using automated maximum entropy de-convolution (AutoME, Waters Corp.) to generate a list of the intact protein masses found in the sample. Likewise, the MS/MS data from the trypsin-digested fractions were searched against the SwissProt database of known *E. coli* proteins using ProteinLynx Global Server 2.0 (Waters Corp.) in order to identify the proteins which were detected. The results were compared with previous analyses, to determine if the on-line LC x LC approach offered any benefit in terms of increased number of proteins detected and identified. The overlap between the proteins detected using top-down (intact protein MS) and bottom-up (peptide MS/MS) analyses was also assessed, to determine if the increased fractionation offered by LC x LC helped improve the correlation between the methods.

5.3.5.1 Comparison with previous experiments

The simplest comparison to make involving proteomic datasets is between proteins that have been identified by database searching. This is because each protein is assigned a

191

unique name and identifying code; a comparison between two lists can be made simply by checking to see if a protein present in one list is also present in the other. Thus, the proteins identified using MS/MS data in this experiment were compared in this manner with those identified from the off-line LC x LC experiment reported in Chapter 4. A Venn diagram illustrating the correlation between the proteins identified in each experiment is shown in Figure 5-11A. Encouragingly, the overlap is relatively high – 72% of the proteins identified using MS/MS in Chapter 4 were also found using the methods reported in this chapter. Moreover, the total number of proteins identified increased from 159 to 241 in moving from the off-line to the on-line LC x LC separation. This is evidence that the larger number of fractions collected allowed better MS/MS coverage of the proteins in the sample. Table 5-5 lists the 127 "new" proteins that were not found in Chapter 4 but were identified by peptide MS/MS database searching using the method reported in this chapter.

The fact that the overlap between the two methods was not 100%, in spite of an increased total number of proteins detected, indicates that protein identifications via MS/MS can be somewhat hit-or-miss. For example, if a peptide starts to elute from the RPLC column when other ions are being analyzed using MS/MS, but finishes eluting before the MS/MS scans are complete, it may never be subjected to fragmentation and therefore cannot be used to identify the protein from which it originated. A means of improving the correlation further would be to return to UHPLC for peptide separations. This would improve resolution of peaks, which would allow a greater portion of the peptides to be analyzed using MS/MS.

5.3.5.2 Overlap between intact protein masses and identified proteins

Unlike comparing lists of proteins which have been uniquely identified, there is some ambiguity associated with matching proteins in two different lists based on their molecular

weights. One reason for this uncertainty is that an arbitrary error margin must be assigned to account for mass error. Also, when comparing experimentally measured intact protein masses to predicted protein MW values, there is no way of knowing in advance if the predicted mass accounts for all post-translational modifications. Nonetheless, this is the only method by which the correlation between top-down and bottom-up data collected in this experiment could be assessed. A direct comparison of the proteins found in the same fractions using the top-down and bottom-up methods was not possible, since intact protein MS data was not collected for the 2D separation in which fractions were analyzed using the bottom-up approach (Figure 5-6). However, intact protein MS data was collected in a later LC x LC separation of the same sample under nearly identical conditions (Figure 5-7). This allowed a reasonable assessment of the overlap between top-down and bottom-up analyses to be made, which is illustrated using a Venn diagram in Figure 5-11B. The overlap between intact proteins and proteins identified by MS/MS was determined to be 40%. This is a moderate increase from the overlap of 27% obtained in Chapter 4. Thus, some improvement in the correlation between top-down and bottom-up data was gained due to the extra resolution provided by LC x LC separation prior to fractionation. As discussed in Chapter 4, the remaining 60% of non-overlapping proteins may be due in part to post-translational modifications not predicted in the protein databases, or different selectivity of electrospray-MS for intact proteins as compared to peptides.

A final comparison was made between the total number of intact proteins detected using on-line LC x LC with the short AXC column (Figure 5-7) versus the long AXC column (Figure 5-8). By increasing the column length, the total number of intact proteins detected increased from 233 to 299. Thus, the longer anion exchange column provided a relatively

modest 28% increase in total number of intact proteins detected. It is likely that this increase would have been greater if the long anion exchange column were sampled more frequently by the second dimension, in order to take advantage of more of the resolution which it is capable of generating.

5.4   Conclusions

The data generated in this experiment demonstrated that on-line LC x LC can be successfully used in the context of a hybrid top-down / bottom-up proteomic experiment. The number of intact proteins detected and proteins identified using MS/MS data both increased modestly as compared to the off-line LC x LC approach. The relatively small gains in proteins detected illustrates the fact that "digging deeper" into the proteome of an organism becomes exponentially more difficult the farther down one attempts to probe, due to the large range of protein concentrations that are present. Ultimately, the number of proteins which can be detected is largely dependent on the sensitivity and dynamic range of the mass spectrometer. MS instruments which are substantially more sensitive than the ones used in this study are commercially available at the present time, and future improvements in this area will doubtless result in further advances in the field of proteomics.

A more significant improvement demonstrated in this chapter was the increase in analysis speed achieved by using on-line LC x LC for separations of intact proteins. An analysis which previously required several days to complete could be performed in just a few hours, and was completely automated once the sample was injected. Moreover, the sensitivity and the resolution of the technique was maintained to an adequate degree, such that the number of proteins detected and identified remained the same or improved as compared to off-line separations. The bottom-up portion of the analysis, which was still performed off-line using fraction collection, remained relatively time-consuming.

Nevertheless, an improvement in speed of one portion of the method represents a significant

step toward making the technique more practical for day-to-day proteomics research.

5.5   References

(1)   Garcia, M. C.; Hogenboom, A. C.; Zappey, H.; Irth, H. *Journal of Chromatography A* **2002**, *957*, 187-199.

(2)   Giddings, J. C. *Journal of High Resolution Chromatography & Chromatography Communications* **1987**, *10*, 319-323.

(3)   Neue, E. D. *HPLC Columns: Theory, Technology, and Practice*; Wiley-VCH: New York, 1997.

(4)   Kastner, M. *Protein Liquid Chromatography*; Elsevier: Amsterdam, Netherlands, 2000.

(5)   Cohen, K. A.; Schellenberg, K.; Benedek, K.; Karger, B. L.; Grego, B.; Hearn, M. T. W. *Analytical Biochemistry* **1984**, *140*, 223-235.

(6)   Cohen, S. A. *Analytical Chemistry* **1984**, *56*, 217-221.

(7)   Cohen, S. A.; Benedek, K.; Tapuhi, Y.; Ford, J. C.; Karger, B. L. *Analytical Biochemistry* **1985**, *144*, 275-284.

(8)   Sadek, P. C.; Carr, P. W.; Bowers, L. W. *Journal of Liquid Chromatography* **1985**, *8*, 2369-2386.

5.6   Tables

Time (min)	% B
0	0
10	5
300	50
350	50
355	0

Table 5-1: AXC gradient profile used with the short anion exchange column. MP A was 0.01M ammonium acetate, pH 8.5. MP B was 1M ammonium acetate, pH 8.5.

Time (min)	% mobile phase B
0	0
30	5
900	50
1050	50
1071	0

Table 5-2: AXC gradient profile used with the long anion exchange column. Mobile phase A was 0.01M ammonium acetate, pH 8.5. Mobile phase B was 1M ammonium acetate, pH 8.5

Time (min)	% mobile phase B
0	10
1	20
6.25	35
19	50
20	70
21	70
22	10

Table 5-3: RPLC gradient profile for LC x LC separations with UV detection only. Mobile phase A was 0.1% (v/v) trifluoroacetic acid in deionized water. Mobile phase B was 0.1% (v/v) trifluoroacetic acid in acetonitrile.

Time (min)	% mobile phase B
0	5
2	20
20	50
21	70
22	5
23	70
24	5

Table 5-4: RPLC gradient profile for LC x LC separations with UV and MS detection. Mobile phase A was 0.2% (v/v) formic acid in deionized water. Mobile phase B was 0.2% (v/v) formic acid in acetonitrile.

Ref #	Protein	Prd MW	Pep mtc	% Cov	Ref #	Protein	Prd MW	Pep mtc	% Cov
160	RL30_ECOLI	6411	1	24	208	YCAC_ECOLI	23100	1	5
161	YDCH_ECOLI	6470	1	18	209	ENGB_ECOLI	23561	1	5
162	RL31_ECOLI	7871	1	20	210	YLIJ_ECOLI	23713	2	17
163	IF1_ECOLI	8118	1	25	211	TRAT1_ECOLI	23809	1	2
164	YODD_ECOLI	8579	2	57	212	NARL_ECOLI	23927	2	12
165	GLRX3_ECOLI	9006	1	30	213	YCEH_ECOLI	24177	1	6
166	YCIN_ECOLI	9386	1	24	214	PYRH_ECOLI	25839	1	5
167	RS20_ECOLI	9553	1	13	215	CPXR_ECOLI	26312	1	6
168	RS17_ECOLI	9573	1	23	216	RS2_ECOLI	26612	1	6
169	YIIU_ECOLI	9635	4	58	217	YGGE_ECOLI	26635	2	7
170	HDEA_ECOLI	9741	2	59	218	HDHA_ECOLI	26779	1	5
171	YHCO_ECOLI	10796	1	22	219	LPXA_ECOLI	28080	1	8
172	YGIN_ECOLI	11532	1	9	220	PANB_ECOLI	28237	2	9
173	THIO_ECOLI	11675	1	11	221	TRPA_ECOLI	28724	1	7
174	BOLA_ECOLI	11994	1	15	222	YEHZ_ECOLI	30204	1	6
175	YDEI_ECOLI	12025	1	12	223	ALSB_ECOLI	30385	2	10
176	PHNA_ECOLI	12345	1	15	224	NADE_ECOLI	30637	1	5
177	GLNB_ECOLI	12425	1	10	225	THTM_ECOLI	30681	2	14
178	RL19_ECOLI	13002	1	13	226	DKGA_ECOLI	31110	2	11
179	YCFF_ECOLI	13241	1	14	227	AGAY_ECOLI	31294	1	2
180	YJBR_ECOLI	13519	1	8	228	YJJW_ECOLI	31491	1	2
181	RL16_ECOLI	14281	1	12	229	PANC_ECOLI	31598	1	4
182	YHHA_ECOLI	14738	1	14	230	YBBN_ECOLI	31791	1	4
183	YAEH_ECOLI	15096	1	9	231	YTFQ_ECOLI	32126	3	13
184	USPG_ECOLI	15935	1	18	232	GLSA1_ECOLI	32903	3	16
185	FUR_ECOLI	16795	3	27	233	RSEB_ECOLI	33294	1	4
186	YHBC_ECOLI	16821	1	7	234	ACCD_ECOLI	33322	1	7
187	DKSA_ECOLI	17528	1	8	235	THIB_ECOLI	34207	1	4
188	BCP_ECOLI	17634	1	13	236	PTNAB_ECOLI	34916	1	4
189	GREA_ECOLI	17641	1	7	237	ACCA_ECOLI	35110	3	12
190	PPIA_ECOLI	18077	1	14	238	QOR_ECOLI	35172	1	4
191	SSPB_ECOLI	18262	1	7	239	RPOA_ECOLI	36512	2	9
192	MOAB_ECOLI	18534	1	7	240	TDH_ECOLI	37239	1	1
193	LRP_ECOLI	18756	1	7	241	ALF1_ECOLI	37978	2	7
194	SSB_ECOLI	18844	1	8	242	GALM_ECOLI	38190	1	7
195	HSLV_ECOLI	18962	1	6	243	GCST_ECOLI	40016	4	15
196	YAJQ_ECOLI	19047	2	17	244	IADA_ECOLI	41084	1	3
197	ATPD_ECOLI	19332	2	22	245	YQHD_ECOLI	42097	1	4
198	APT_ECOLI	19859	2	23	246	FABB_ECOLI	42613	1	2
199	WCAF_ECOLI	19962	1	3	247	RIR2_ECOLI	43386	2	9
200	YDJA_ECOLI	20059	1	9	248	TOLB_ECOLI	43602	1	2
201	RL5_ECOLI	20170	1	12	249	ASTC_ECOLI	43665	1	5
202	EFP_ECOLI	20460	1	6	250	ODO2_ECOLI	43880	5	15
203	GMHA_ECOLI	20815	1	8	251	ISCS_ECOLI	45090	1	3
204	YDJY_ECOLI	21740	1	2	252	DEGP_ECOLI	46829	2	6
205	GRPE_ECOLI	21798	1	7	253	PURA_ECOLI	47214	1	5
206	FKBB_ECOLI	22085	2	11	254	GLMM_ECOLI	47412	2	6
207	RL4_ECOLI	22087	1	12	255	DHE4_ECOLI	48581	1	6

Table 5-5: *E. coli* proteins identified by peptide MS/MS analyses, which were not detected in Chapter 4. (This table is continued and the heading terms are defined on the following page).

Ref #	Protein	Prd MW	Pep mtc	% Cov	Ref #	Protein	Prd MW	Pep mtc	% Cov
256	ACCC_ECOLI	49321	1	3	272	SYQ_ECOLI	63347	1	2
257	KPYK1_ECOLI	50729	2	7	273	DHSA_ECOLI	64422	1	2
258	GABD_ECOLI	51720	2	10	274	ODP2_ECOLI	65965	1	2
259	GLNA_ECOLI	51773	4	15	275	YEJA_ECOLI	67521	1	1
260	IMDH_ECOLI	52022	3	13	276	TKT2_ECOLI	73043	3	8
261	SYN_ECOLI	52439	1	2	277	OPDA_ECOLI	77167	2	5
262	PEPD_ECOLI	52784	1	2	278	CATA_ECOLI	80024	3	6
263	SYE_ECOLI	53816	4	11	279	TRAD1_ECOLI	81489	1	1
264	NORR_ECOLI	55236	1	1	280	CLPB_ECOLI	95585	6	7
265	AHPF_ECOLI	56177	1	2	281	SYA_ECOLI	96032	3	7
266	PUR1_ECOLI	56357	3	6	282	AMPN_ECOLI	98788	4	5
267	SYK2_ECOLI	57695	2	5	283	CAPP_ECOLI	99063	1	2
268	USHA_ECOLI	58209	1	2	284	ODP1_ECOLI	99537	1	1
269	FUMA_ECOLI	60167	3	7	285	SYI_ECOLI	104166	5	10
270	YTFM_ECOLI	62527	1	1	286	ODO1_ECOLI	105062	3	5
271	MAO1_ECOLI	63197	1	2					

Table 5-5: (Table is continued from previous page) *E. coli* proteins identified by peptide MS/MS analyses, which were not detected in Chapter 4. Heading terms are defined as follows: **Ref #**: Internal protein reference number; see below. **Protein**: SwissProt database entry name; **Prd MW**: average protein mass calculated from predicted protein sequence in SwissProt database entry**; Pep mtc**: number of tryptic peptides identified using ESI-q-TOF MS/MS analysis; **% Cov**: Percent coverage of predicted protein sequence based on peptides matched.

A descriptive name of all proteins in this table is provided in the appendix of this dissertation, listed by Ref #.

5.7   Figures
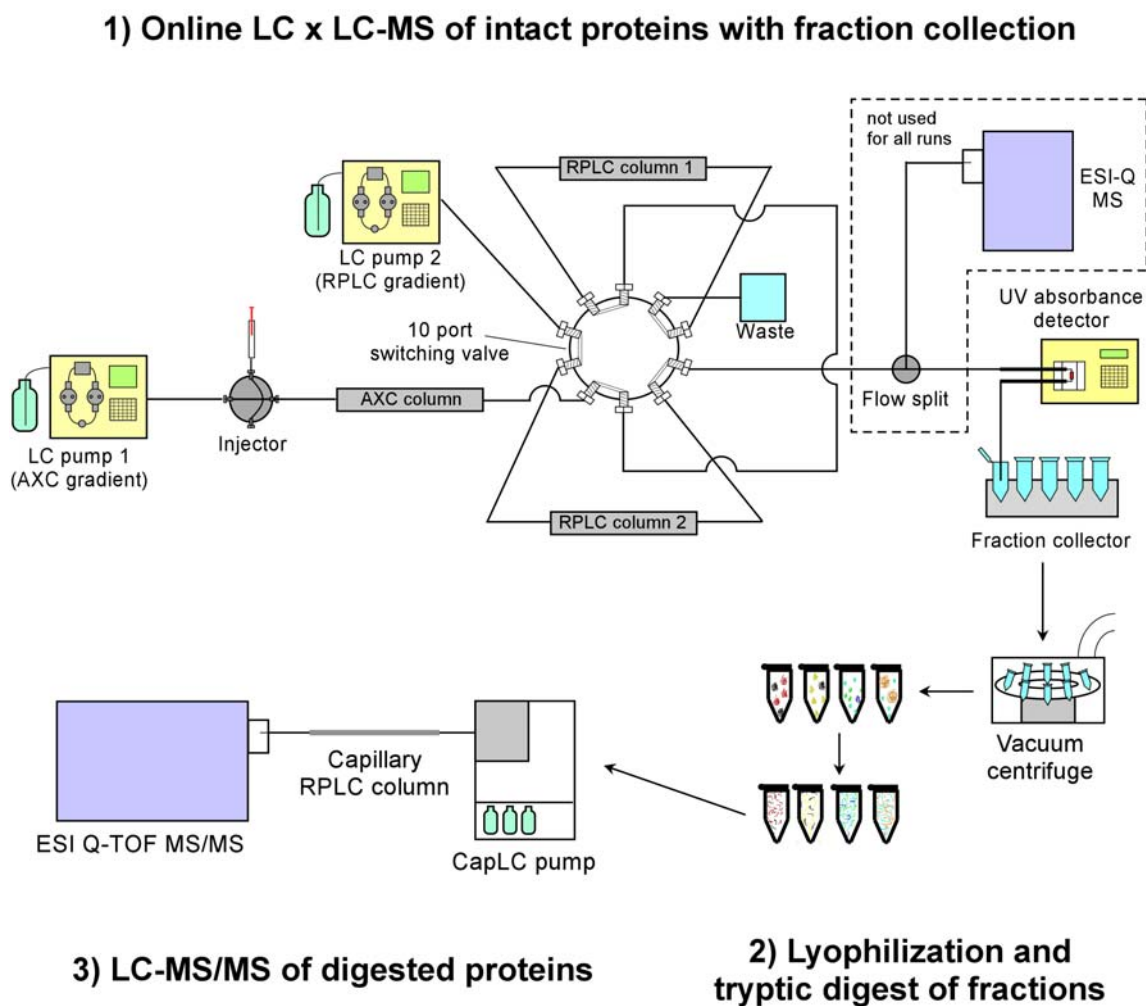
Figure 5-1: Overview of instrumentation and procedure used for top-down LC x LC-MS and bottom-up LC-MS/MS analysis of complex protein mixtures

Position A: load column 1, elute column 2
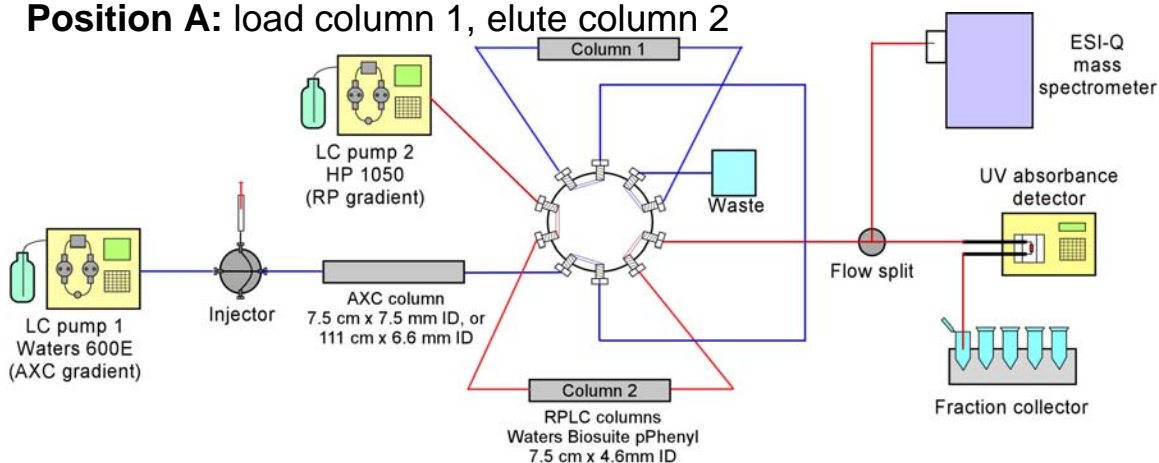
Position B: load column 2, elute column 1
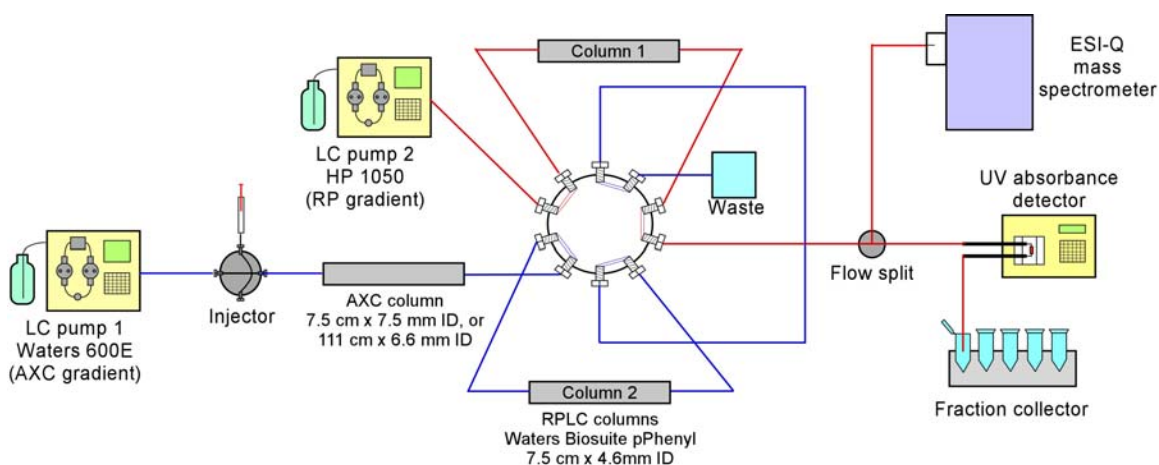
Figure 5-2: Diagram illustrating operation of the on-line LC x LC system. Fluidic connection lines are colored to illustrate flow paths through the system. The 10-port switching valve can be in one of two positions, A or B. In position A, the effluent from the AXC column is directed to RPLC column 1, and sample components are trapped at the head of the column. Meanwhile, RPLC column 2 is eluted by a reversed-phase gradient produced by LC pump 2. When the gradient is finished, the valve is switched to position B, which causes the effluent of the AXC column to be directed to column 2, and allows the sample components on column 1 to be eluted.
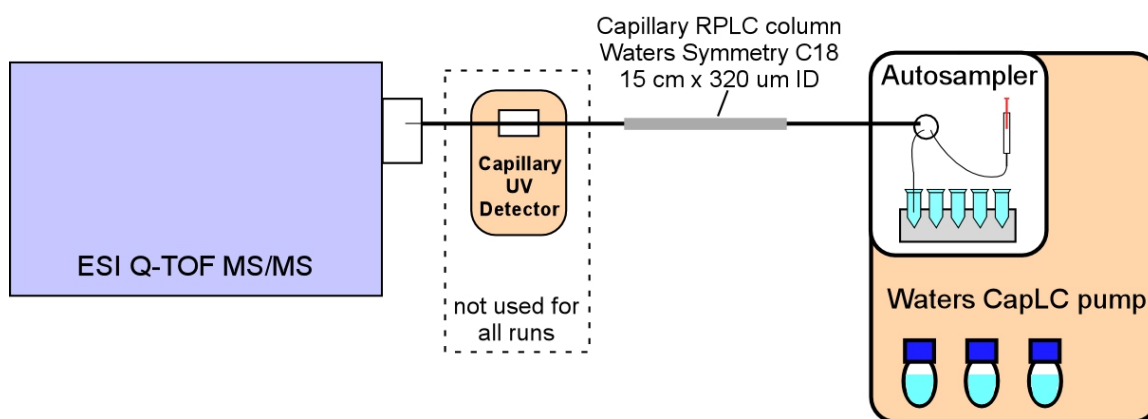
Figure 5-3: Diagram of the conventional pressure CapLC-MS system used to perform bottom-up proteomic analyses. The UV detector was a photodiode array, which is built into the CapLC system. It is shown as a separate component in this diagram for clarity.
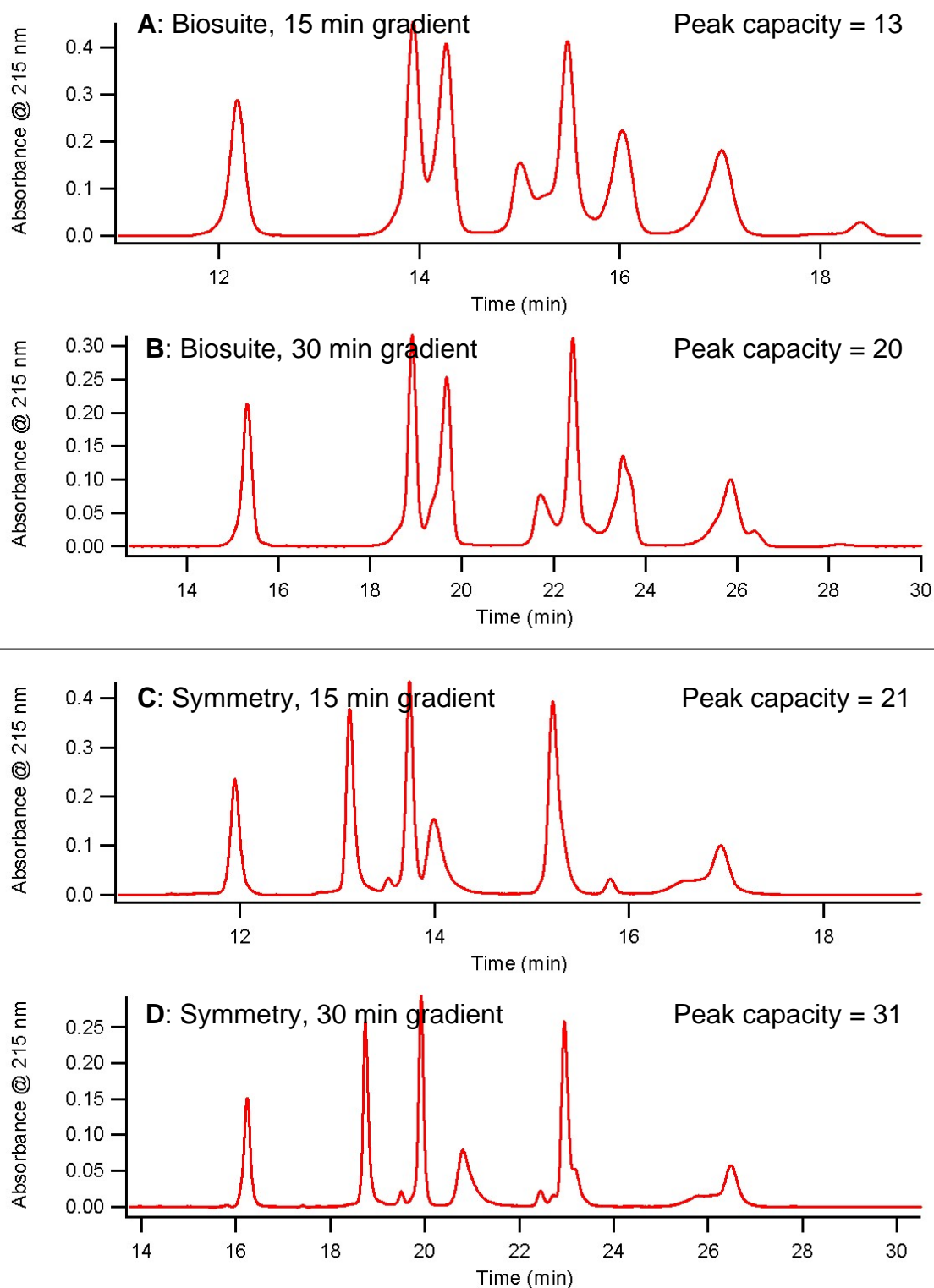
Figure 5-4: Comparison of separations of seven standard proteins at 0.1 mg/mL using polymeric (A, B) and silica-based (C, D) RPLC columns. The gradient went from 10 to 70% ACN with 0.1% TFA over the specified length of time (15 or 30 minutes).
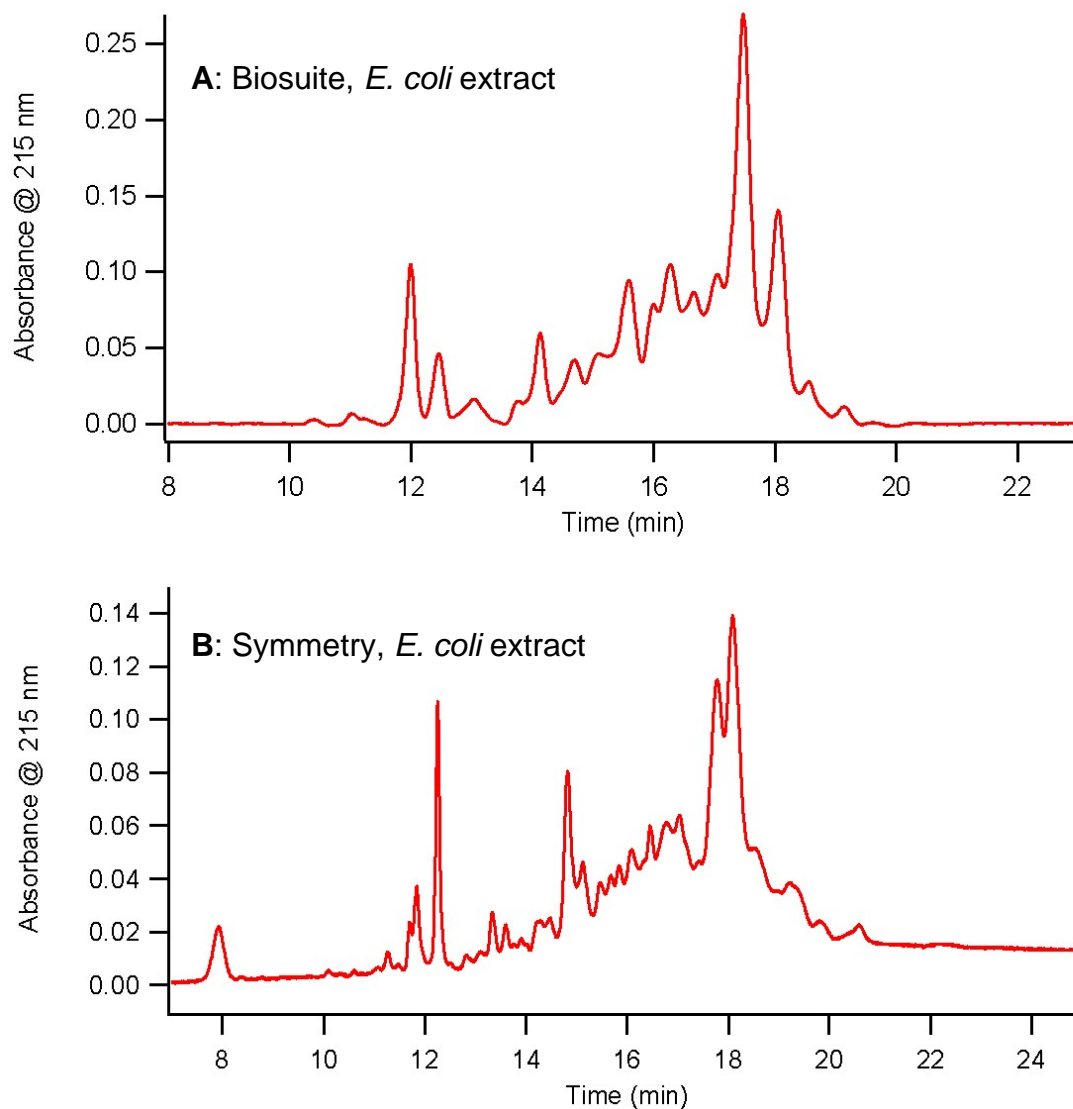
Figure 5-5: Comparison of separations of an *E. coli* protein extract using polymeric (A) and silica-based (B) RPLC columns. The gradients used were: (A): 10-70% ACN over 7.5 min; (B): 10-70% ACN over 15 min; (C): 20-60%ACN over 15 min; (D): 20-80% ACN over 15 min. 0.1% TFA was in all mobile phases throughout the runs.
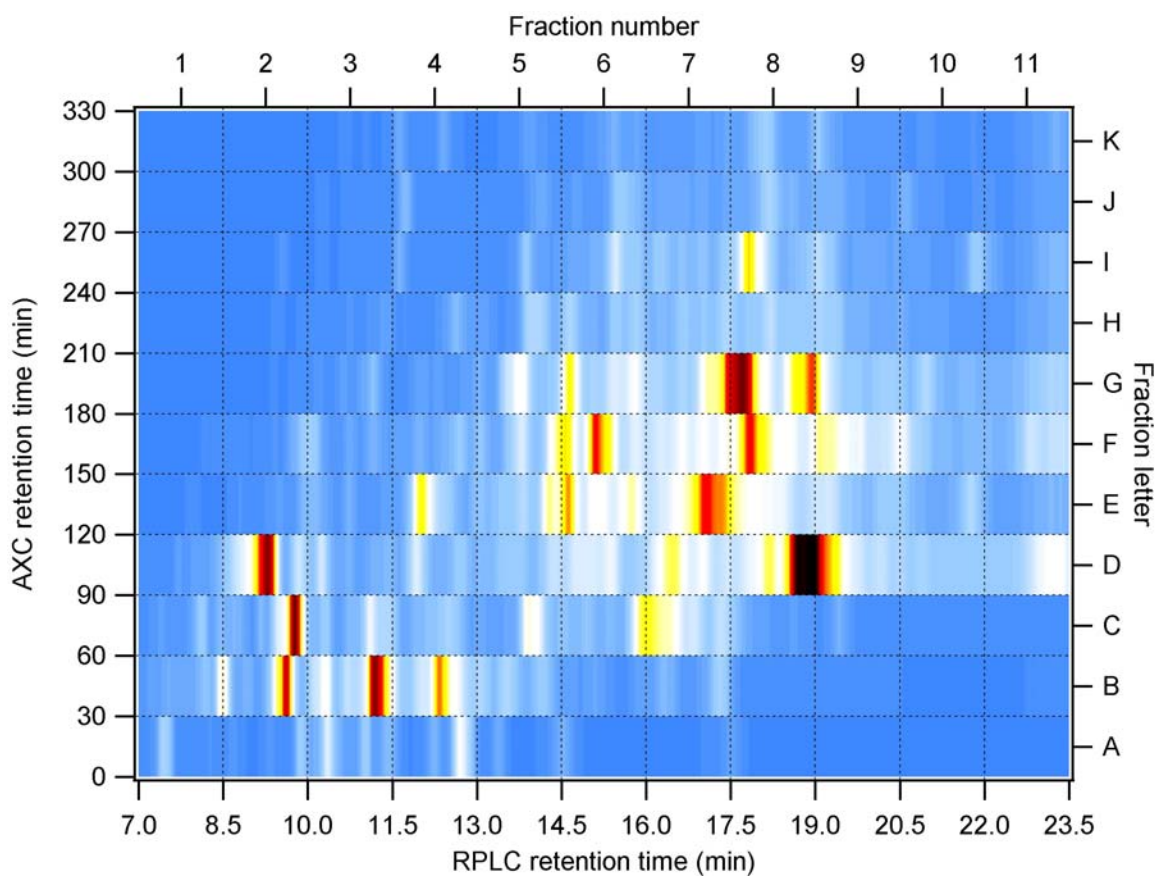
Figure 5-6: UV absorbance chromatogram from an online LC x LC separation of the *E. coli* protein extract. Detection was performed at a wavelength of 215 nm. A total of 121 fractions were collected. Anion exchange run conditions are specified in Table 5-1; reversed-phase run conditions are specified in Table 5-3.
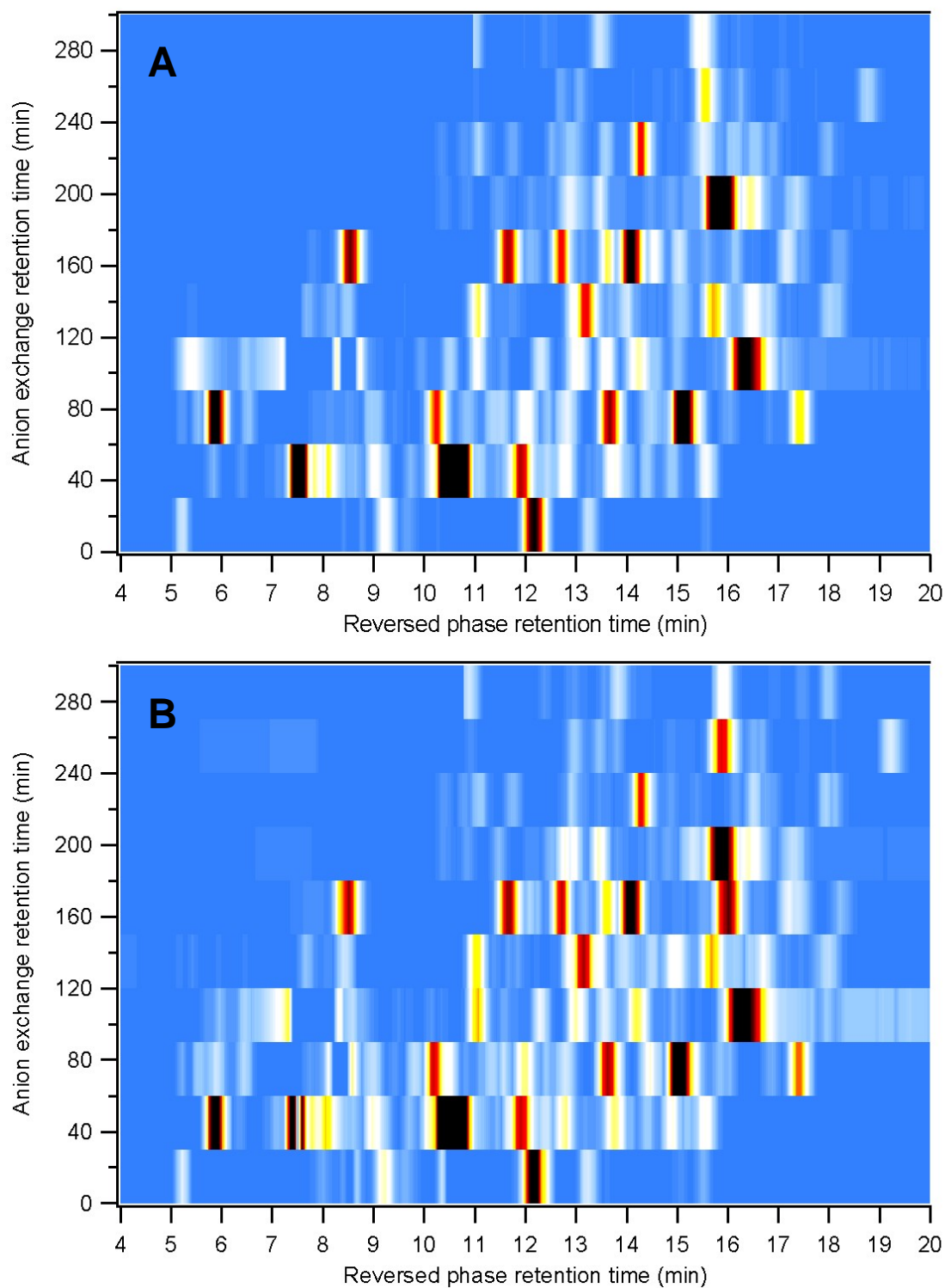
Figure 5-7: 2D chromatograms from replicate on-line LC x LC separations of the *E. coli* protein extract using the short AXC column. The chromatograms were generated using AutoME-de-convoluted data and are displayed using an identical color-scale range. Run conditions are specified in the text.
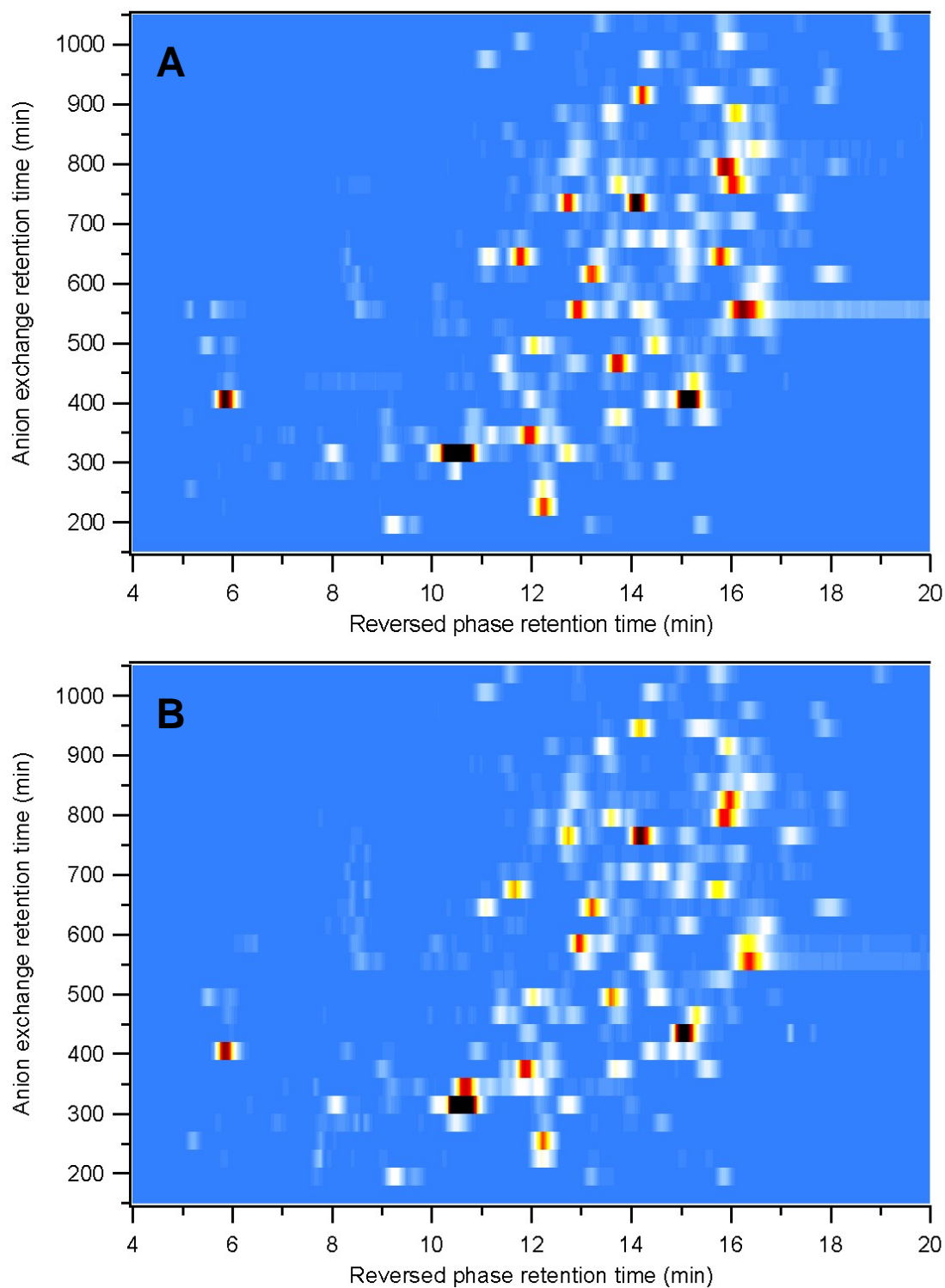
Figure 5-8: 2D chromatograms from replicate on-line LC x LC separations of the *E. coli* protein extract using the long AXC column. The chromatograms were generated using AutoME-de-convoluted data and are displayed using an identical color-scale range. Run conditions are specified in the text.
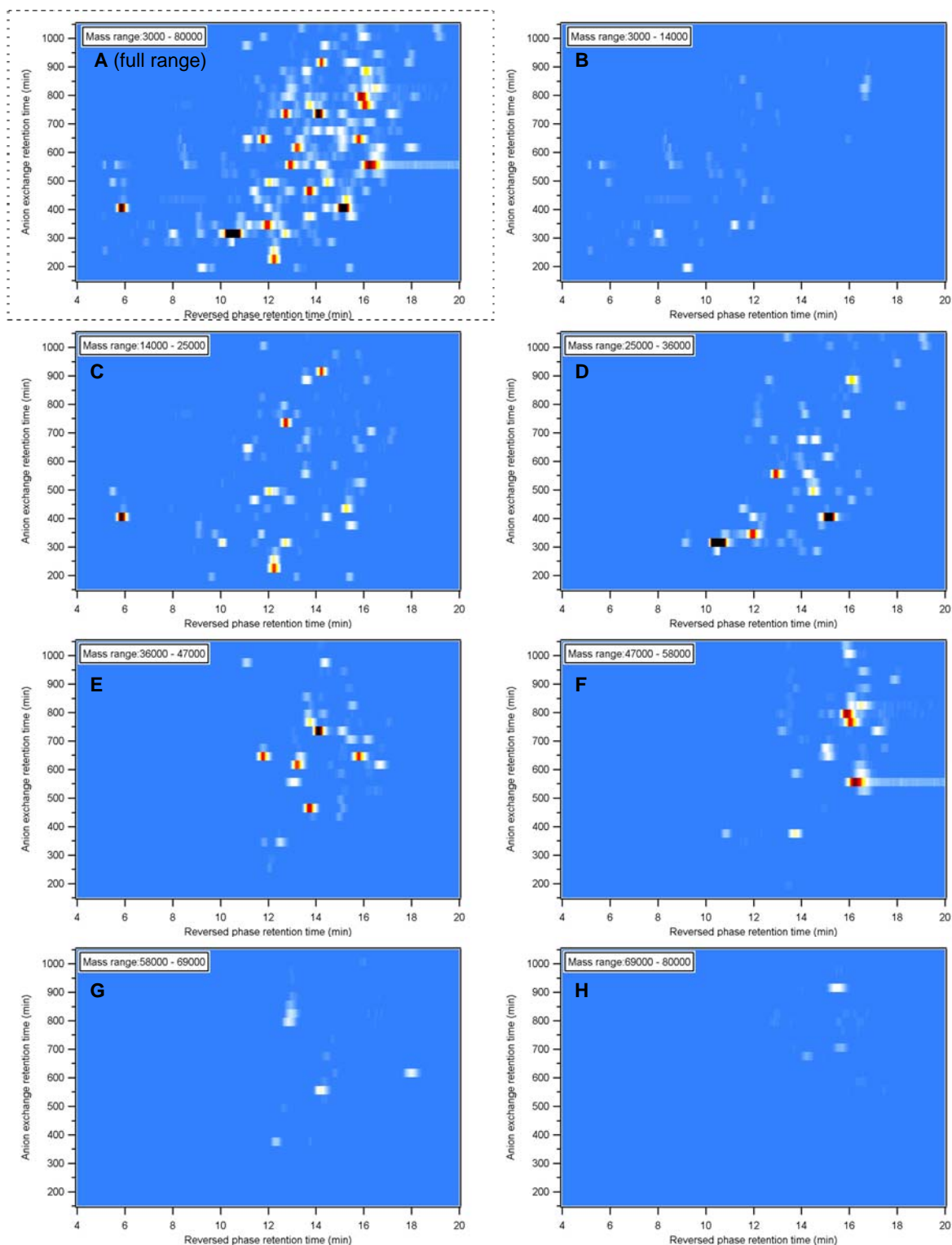
Figure 5-9: Mass slice 2D chromatograms of the *E. coli* protein extract. Chromatogram (A) shows the full-mass range data; the other seven chromatograms (B-H) display 11 kDa segment "mass slices" of the same data.
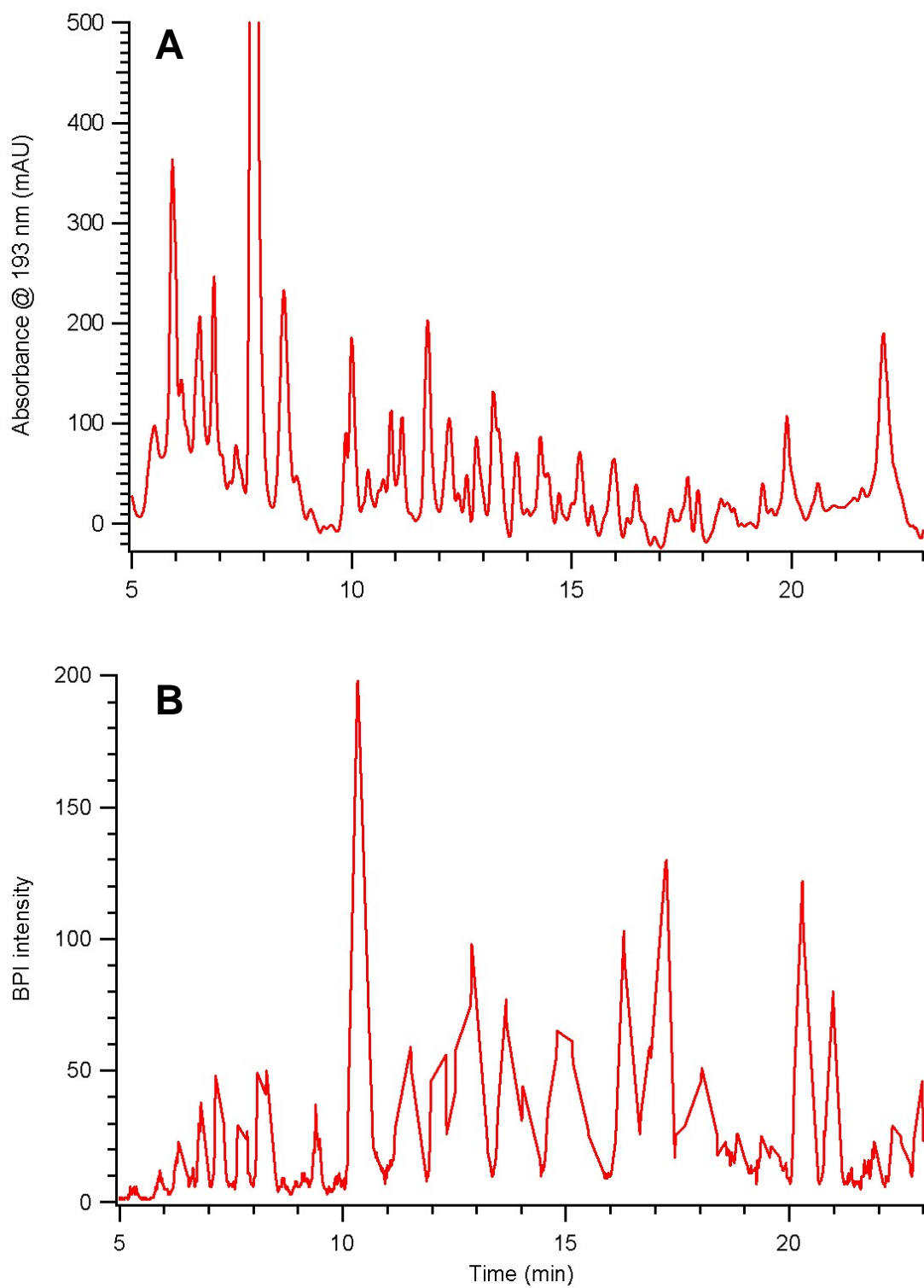
Figure 5-10: Chromatograms displaying UV absorbance (A) and MS base peak intensity (B) for the RPLC separation of peptides from a trypsin digest of LC x LC fraction F8 (see Figure 5-6).
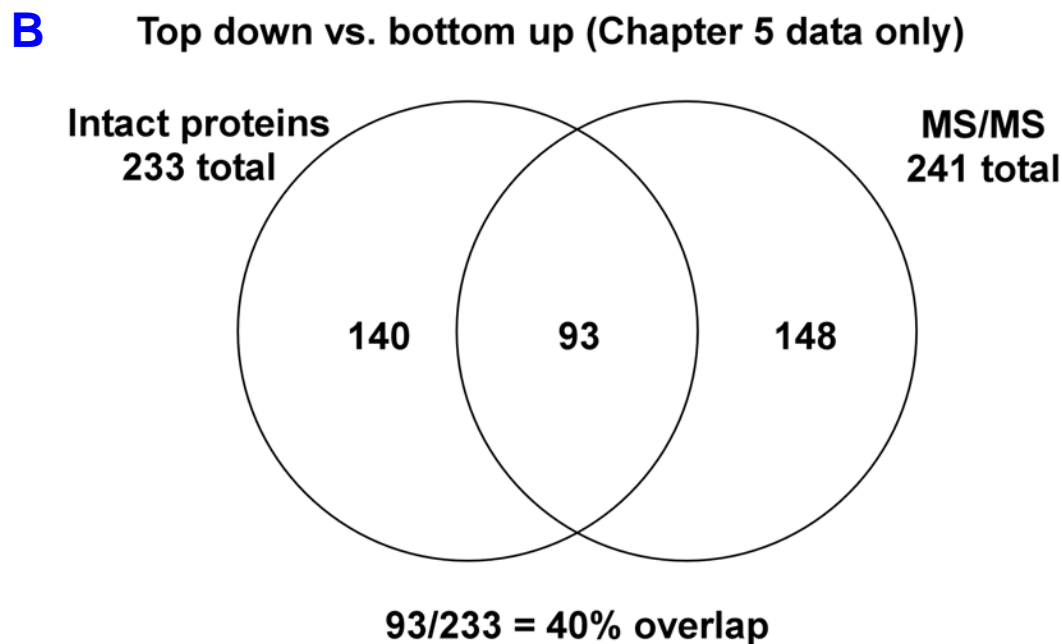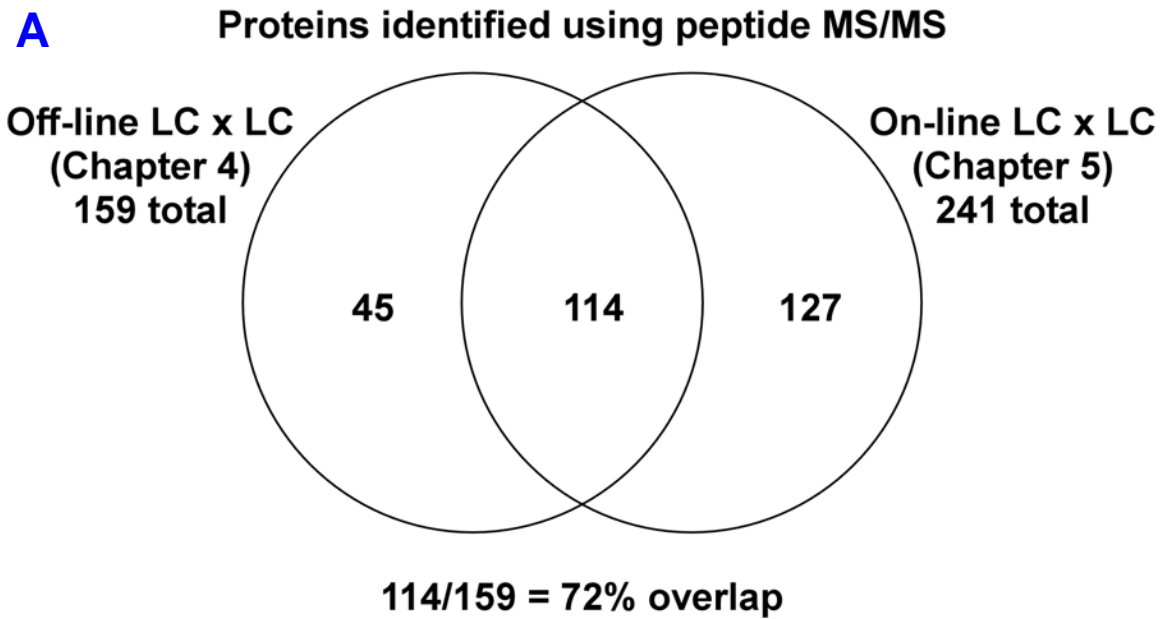
Figure 5-11: Venn diagrams illustrating overlap between proteins found using different analysis methods. (A) compares the number of proteins identified using MS/MS from off-line LC x LC (Chapter 4) and on-line LC x LC (Chapter 5). (B) illustrates the overlap between proteins detected using intact protein LC x LC-MS versus those identified using LC-MS/MS of the peptide fractions.

CHAPTER 6: LC x LC-MS for differential analysis of the proteome of yeast grown under different conditions

6.1   Introduction

The objective of proteomics is to determine meaningful biological information by analyzing the entire complement of proteins expressed by an organism. Qualitative proteomics – which involves determining the identities of the proteins in a sample – has been the focus of the methods reported in this dissertation thus far. While this is a useful means of gaining understanding of many fundamental biological processes, in many cases it is also important to know how two or more samples differ in terms of the quantity of the various proteins they contain. For example, the rapidly developing field of biomarker discovery relies in part on the detection of patterns of change in protein expression in order to allow faster, more accurate diagnosis of disease.[1-4] It is therefore often necessary to compare the amounts of various proteins produced by a healthy organism or tissue to those from an organism with a specific disease. Methods designed to address this type of challenge fall within the field of differential proteomics.[5, 6]

Most differential proteomics performed to date has been carried out using gel electrophoresis.[6] A typical experiment is performed by separating the samples to be compared on 2D gels under identical conditions. The proteins are stained, and images of the gels are captured and converted to digital form. Image manipulation and analysis is performed to overlay matching spots and to detect spots which differ substantially in intensity. These regions of the gel are then excised, and in-gel digestion is performed on the

proteins. The peptides which are released are then analyzed using bottom-up methods such as MALDI-TOF-MS or LC-ESI-TOF-MS followed by database searching to identify the proteins which differed in quantity between samples. While effective, this approach is subject to all of the shortcomings of gel electrophoresis, such as poor resolution of some proteins, the lack of dynamic range associated with staining methods, and the requirement of substantial manual labor. More recently, alternative approaches for differential proteomics have been demonstrated which employ strategies based on LC-MS or multidimensional LC-MS.[1, 6] The majority of these techniques are based on bottom-up proteomics. Frequently, they involve some form of isotopic labeling in order to differentiate one sample from another using mass spectrometry,[7, 8] although some label-free methods have been reported.[9]

In the research presented in this chapter, multidimensional liquid chromatography coupled to mass spectrometry was used to perform a differential analysis of several samples of soluble proteins extracted from *Saccharomyces cerevisiae* (bakers yeast), a unicellular eukaryote which has been extensively used as a model organism in the fields of cellular biology and biochemistry.[10, 11] Differences in protein expression were brought about by culturing yeast on different growth media and by harvesting protein at different stages in its growth cycle. LC x LC-MS analyses of intact proteins from the yeast cultures were then used to detect variances in protein expression between the samples. Proteins which were found to be up- or down-regulated in one or more of the samples were digested using trypsin and were identified via bottom-up LC-MS/MS. Most of the analytical techniques used are similar to those described in Chapter 5. Thus, the focus of this chapter is on the application of the hybrid top down / bottom up method to differential proteomics.

6.2   Experimental

6.2.1   Overview

An overview of the method used to perform differential proteomic analysis of *S. cerevisiae* is shown in Figure 6-1. Protein extracts from yeast cultures grown under different conditions were separated using on-line LC x LC. The effluent from the second dimension was split such that a small portion was used to perform on-line electrospray-TOF MS, while the majority was collected in fractions. The intact protein LC x LC-MS data were used to generate 2D mass-slice chromatograms. The mass slice chromatograms from the different samples were compared, and proteins which increased or decreased significantly in intensity were selected for further analysis. The proteins in these fractions were digested using trypsin, and then separated and detected using capillary RPLC-MS/MS. The resulting MS/MS spectra were searched against a database of known yeast proteins in order to identify the proteins present in the fractions. The specific proteins which were detected to be up- or down-regulated were identified by matching their observed MW from the intact protein MS data with the predicted MWs from the peptide data.

6.2.2   Chemicals and samples

The chemicals used in this experiment were the same as those described in Chapter 5. The yeast protein extract samples were prepared from cultures of *S. cerevisiae* by researchers at Waters Corporation (Milford, MA). To induce differences in protein expression, the yeast was cultured using growth media with different carbon sources and the proteins were harvested at different phases of growth. One culture was grown on glycerol medium and harvested at log-growth phase. A second was grown on glucose medium and was harvested at log-growth phase. The third was grown on glucose medium and was harvested at stationary phase. Soluble proteins were extracted from each of the yeast cultures according

to a procedure which has been described previously.[12] The final extract samples contained

cytosolic yeast proteins in 25 mM ammonium bicarbonate, along with several protease

inhibitors. The concentration of protein in the extracts was measured using a Coomassie Plus

Bradford assay kit (Pierce, Rockford, IL) using bovine serum albumin as a standard. Based

on this assay, the concentration of protein in the glycerol/log phase sample was 13.5 mg/mL;

in the glucose/log phase sample it was 11.0 mg/mL, and in the glucose/ stationary phase

sample it was 10.5 mg/mL. All solutions were diluted to a concentration of 10.0 mg/mL with

anion exchange mobile phase A and were spiked with bovine heart cytochrome C at a

concentration of 10 ng/mL to serve as an internal standard.

6.2.3   Intact protein LC x LC-MS

The instrumentation used for the on-line LC x LC-MS separations of intact proteins

was identical to that which is described in Section 5.2.6. Most of the chromatographic

conditions were also similar, though slight modifications were made to the anion exchange

and reversed phase gradients in order to improve the distribution of the yeast proteins over

the available separation space. Mobile phase compositions and gradient profiles are provided

in Table 6-1 and Table 5-4. A total of 30 reversed-phase separations were performed in

every LC x LC run. Fraction collection was initiated 2 minutes after the start of each

reversed phase gradient, and fractions were switched every 45 seconds using an automated

fraction collector until 16 fractions had been collected, at which point collection was

suspended until 2 minutes after the start of the next gradient. For each LC x LC run, a total

of 480 fractions, 0.75 mL each, were collected in 1 mL 96-well microtiter plates. After

collection, the fractions were stored at -20º C until needed.

A total of six LC x LC separations were performed. The first three runs were carried

out in order to analyze differences in protein expression between the three yeast samples.

The log-phase glycerol-medium sample was analyzed first, followed by the log-phase glucose-medium sample and the stationary-phase glucose-medium sample. After these runs were complete, the glycerol log-phase sample was analyzed three additional times using the same LC x LC-MS method. These replicate runs served to assess the run-to-run reproducibility, and thereby to verify the technique's applicability to differential analysis of complex mixtures of proteins.

6.2.4   Intact protein data processing

The focus of this experiment was to selectively identify proteins which were up or down-regulated in one of the yeast samples, rather than to detect and identify as many proteins as possible. Therefore, the intact protein LC x LC-MS data were examined to detect proteins that changed in intensity between different yeast samples. The yeast culture grown on glycerol medium was compared to the sample grown on glucose medium (both samples harvested at logarithmic phase), and the yeast culture harvested at logarithmic growth phase was compared to the sample harvested at stationary phase (both samples were grown on glucose media).

In order to compare the abundance of proteins in the samples, LC x LC-MS data were first processed using automated maximum entropy de-convolution (AutoME) according to the parameters specified in Chapter 5. The resolution for maximum entropy de-convolution was set at 2 Da as opposed to 1 Da to reduce processing time. The de-convoluted data were used to generate 2D chromatograms using custom functions written for the data analysis software package Igor Pro (WaveMetrics, Lake Oswego, Oregon). The bovine heart cytochrome C peak, an internal standard added to each of the yeast samples, was located in each chromatogram and its area was measured. These peak areas were used to standardize the intensities of all other peaks for all three yeast samples. The standardized chromatograms

216

were displayed using the same z-axis range (color scale) in order to allow accurate comparisons between samples. The maximum intensity was selected such that some of the tallest peaks are off-scale; this made lower intensity peaks more visible. In cases where the intensity of the largest peaks needed to be compared, the chromatograms were re-scaled appropriately. 2D mass-slice chromatograms spanning a range of 1 kDa each were generated as described in Section 5.3.3. The mass slice chromatograms from two different yeast samples were then visually compared. Any protein which was observed to increase or decrease significantly in intensity was noted by recording the mass slice in which it was found and the LC x LC fraction in which it was most intense.

The de-convoluted intact protein data were also used for quantitative analysis of the samples. Custom-written Igor Pro functions were used to identify components in the LC x LC-MS data and to determine their intensity, 2D retention time, and mass. The intensity of a component was calculated by summing its intensity (as reported by maximum entropy de-convolution) for all consecutive time segments in which it appeared. This included summing components which were spread over multiple anion exchange fractions. The 2D retention time and mass of each component were set as the values determined from the time segment at the apex of the peak. To compare the intensity of components in different samples, an additional Igor Pro function was used to tabulate components with matching mass (+/- 150 ppm), anion exchange retention time (+/- 2 fractions) and reversed-phase retention time (+/- 2 min).

6.2.5   Protein digestion and peptide LC-MS/MS

After reviewing the intact protein data, fractions that contained proteins which were up- or down-regulated between samples were selected for further analysis. These fractions were thawed, digested using trypsin as described in Chapter 5, and then analyzed using

capillary RPLC-MS/MS. The chromatographic run conditions and mass spectrometer

settings were the same as described in Chapter 5, except that a custom-packed 8 cm x 150

μm ID capillary column with 1.5 μm bridged-ethyl hybrid C18-modified particles (Waters)

was used to perform the RPLC separations. The LC flow was split prior to entering the

column. Approximately 1 μL/min flowed through the column and 4 μL/min flowed to waste

via a splitter capillary. The outlet of the RPLC capillary was directly connected to an

uncoated fused silica nanospray emitter with a 20 μm ID and a 10 μm tip (New Objective,

Woburn, MA). Spray voltage was applied through a zero dead volume union incorporated

into the nanoflow sprayer (Waters Corporation). MS/MS scans were performed in data

directed analysis mode, using the same parameters as described in Chapter 5.

6.2.6   Peptide data processing

The peptide LC-MS/MS data were processed using ProteinLynx Global Server 2.3

(Waters Corporation). MS and MS/MS data were background subtracted using adaptive

mode, de-isotoped using the MaxEnt 3 algorithm, and centroided. The MS/MS spectra were

then searched against a database of known yeast proteins from the Swiss-Prot protein

knowledgebase release 54.2 (http://us.expasy.org/sprot). A fixed modification of

carboxyamidomethylcysteine and a variable modification of methionine oxidation were used

in the search parameters.

After the database search was complete, the proteins identified at the peptide level

were compared with the intact proteins detected in the corresponding fractions.   A match

was considered to have been found when the MW of a protein identified from the peptide

MS/MS data, as given in its Swiss-Prot database entry, corresponded with the MW of an

intact protein found in the same fraction. Although in many cases the MW correspondence

was precise, a tolerance of +/- 500 Daltons was allowed to account for the possible presence

of post-translational modifications not incorporated into the MW reported in the database. This wide mass range is justified by the fact that not only the mass but also the 2D retention time needed to be the same in order for a match to be considered valid.

6.3 Results

6.3.1 Reproducibility of intact protein LC x LC-MS analysis

Three replicate LC x LC-MS runs of the yeast protein extract grown on glycerol medium were used to assess the reproducibility of the method. Full-mass-range 2D chromatograms from the separations of these samples are shown in Figure 6-2. As expected, the general pattern of peaks is quite similar in all three chromatograms. Closer inspection reveals that some retention time drift did occur, both in the anion exchange and reversed phase dimensions. This may be due to slight changes in the composition of the anion exchange and reversed phase mobile phases due to evaporation of volatile buffer components, since runs were performed over the course of three days. Since this drift affects all peaks approximately equally, it does not significantly hamper visual comparison of the chromatograms.

In terms of peak intensity, most peaks appear to be relatively consistent from run to run, although a few peaks vary noticeably. Since reproducible peak intensity is important when performing relative quantification, the extent of the variation was investigated further by generating a log-log intensity scatter plot, which is shown in Figure 6-3. In this plot, proteins are represented as points on an x-y plane, where the x-coordinate is the intensity of a given protein measured in one 2D run and the y coordinate is its intensity in another run. Since three replicate runs were performed, but only two intensities can be plotted as a single point, a total of three different sets of coordinates were plotted on the same graph: run 1 vs. run 2, run 2 vs. run 3, and run 2 vs. run 3. In order to better visualize the full range of the

219

data, a logarithmic scale is used on both axes. For replicate runs of a sample, an ideal

distribution would be a plot in which all points fall along the line y = x, displaying no run-to-

run variation in intensity. The data shown in Figure 6-3 follow this trend relatively closely

(±10-20%) for proteins with high intensity. For lower intensity proteins, however, the

variability is significantly higher – over 10-fold for proteins with an intensity of 500, which

was the limit of detection for this method. The shape of the distribution is relatively close to

that of a simple linear equation, $y = \pm (mx + b)$, plotted on a log-log scale. This equation

takes into consideration a constant noise component, b, and a response factor uncertainty

component, m. Several different forms of this equation are plotted in Figure 3, which

incorporate a greater or lesser proportion of the experimentally-measured protein intensity

data points depending on the m and b values which are selected. The equation $y = 1.2x +$

$10^{4.4}$ was determined to contain 96.5% of the experimental data points from the replicate runs

of the glycerol-medium yeast protein extract; thus it was judged to be an acceptable threshold

for determining whether a difference in protein intensity was significant for runs of different

samples.

**6.3.2   Detection and relative quantification of differentially expressed proteins using
intact protein LC x LC-MS data**

Full-mass-range 2D chromatograms from the separation of the three different yeast

protein extracts are shown in Figure 6-4. As with the replicate runs, there are substantial

similarities between the patterns of peaks in the chromatograms. This is expected, since all

of the samples originate from the same organism, and therefore largely carry the same

complement of proteins. There is some retention time drift between chromatograms,

particularly in the anion exchange dimension, but it does not substantially hamper visual

comparison of the peaks. It is also apparent from the data that the intensities of many peaks

are substantially different from one chromatogram to the next. Furthermore, some peaks appear to be present in one or two of the chromatograms, but are completely absent from others. Although it is possible to observe some intensity changes in the full-mass-range chromatograms, many may be missed due to the presence of overlapping peaks.

To facilitate visual comparison of peak intensity from one sample to the next, mass slice chromatograms were used. Figure 6-5 shows three representative mass slice chromatograms with a 1kDa mass range which were used to compare glucose versus glycerol yeast protein samples. Figure 6-6 depicts mass slice chromatograms for the comparison between logarithmic growth phase and stationary growth phase samples. Some of the mass slice chromatograms contained no detectable peaks, or no peaks which were noted to change intensity. However, many chromatograms did have components which changed intensity between samples, or components which were present in one chromatogram but completely absent from another. When this was the case, the LC x LC fraction in which the component was most abundant was recorded. Once all mass slice chromatograms had been analyzed, a total of 164 fractions (out of a total of 1440) had been selected for further analysis in order to attempt to identify the proteins using bottom-up LC-MS/MS.

Although visual comparison of the mass slice chromatograms allowed facile selection of fractions containing proteins of interest, it did not provide a quantitative estimate of change between samples. Additionally, it did not address whether an observed difference in peak intensity was significant. Therefore, log-log intensity scatter plots were generated using the de-convoluted peak intensities for the two relevant comparisons: 1) proteins from yeast grown on glucose and glycerol media, and 2) proteins from yeast harvested at log and stationary growth phase. Only those peaks which had been selected for further analysis

based on the mass-slice chromatograms and were also successfully identified based on the peptide LC-MS/MS data (Section 6.2.6) were included in the scatter plots. Both of these plots are shown on the same axes in Figure 6-7. Also included in the plot is the data from the replicate runs of the glycerol sample, and the curve associated with a 96.5% confidence interval. Points from the glucose/glycerol and log/stationary comparisons which fall outside of the curve are considered significantly different, which means the protein was present at different abundances in the two samples.

Although it was straightforward to determine whether the change in abundance of a protein was significant for components which were detected in both samples, a substantial number of components were found in only one of two samples. Since an intensity value of zero cannot be displayed on a log-log plot, these components could not be assessed for significance of differential expression. To work around this issue, the missing proteins were assigned an intensity of 500 – which is the approximate limit of detection for the method – for the sample in which they were not found. Thus, for a protein detected in only one sample, its intensity must meet the threshold for statistical significance as if it were present in the second sample at an abundance equivalent to the limit of detection. Over the entire intensity range, a total of 52 proteins satisfied the criteria for statistical significance for change in abundance at the 96.5% threshold. Of these, 30 proteins were from the glucose/glycerol comparison and 22 from the log/stationary comparison.

6.3.3   Identification of differentially expressed proteins using peptide LC-MS/MS data

Those fractions which had been selected for "bottom-up" analysis in order to identify the proteins they contained were lyophilized, digested using trypsin, and analyzed using capillary RPLC-MS/MS. The MS/MS data from these runs were searched against the Swiss-Prot database of known yeast proteins using ProteinLynx Global Server 2.3. This generated

a list of proteins present in the fractions. A total of 230 unique yeast proteins were identified in all of the fractions which were analyzed; however, many of these proteins were not those which were detected as being up- or down-regulated from the LC x LC-MS data. After a fraction-by-fraction comparison of the list of identified proteins with the list of proteins which changed intensity, a total of 76 matches were found. Thus, of the total 164 differentially expressed proteins which were "searched for" by LC-MS/MS analysis, approximately 46% were successfully identified. As stated previously, 52 of these proteins met the criteria for statistical significance of differential expression. These proteins are listed in Table 6-3 (glucose vs. glycerol comparison) and Table 6-4 (logarithmic vs. stationary comparison). Also included in the tables are the proteins' Swiss-Prot accession number (http://us.expasy.org/sprot), ordered locus gene name (http://www.yeastgenome.org), MW, intensity, and the extent to which they were up- or down-regulated, expressed as a "fold-change", which was calculated by dividing the intensity of the more abundant protein by the intensity of the less abundant protein.

6.4 Discussion

The fact that a substantial portion of the intact proteins which were detected as being differentially expressed were not successfully identified is clearly a limitation of this method. Nonetheless, the observed portion of successful identifications (46%) is in line with or better than the typical success rate of correlating top-down and bottom-up proteomic data, as reported in the literature[2, 12, 13] and in Chapters 4 and 5 of this dissertation. Many of the same factors which prevent higher overlap from being achieved in other studies are likely also at work in this experiment. These include difficulty detecting low abundance proteins, differences in selectivity of electrospray ionization towards intact proteins and peptides, and the presence of post translational modifications not predicted in protein database entries. The

last of these reasons can be viewed as both an obstacle to be overcome and as a potential source of useful information. Once a correlation between intact protein MW and an identified protein is established, the mass shift provides a means of probing the modified state of the protein to determine which PTMs may be present. As enhanced top-down proteomics methods are developed, it is anticipated that correlating intact proteins detected with proteins identified by bottom-up methods will become more routinely feasible, which will allow more constructive data regarding PTMs to be obtained.

Of the differentially expressed yeast proteins which were successfully identified, many are involved in metabolic pathways. It is logical to expect that such proteins would be well-represented in this study, since two of the samples were grown on different media, which would be anticipated to cause differences in metabolic processes within the yeast cells.[14] Likewise, the transition from logarithmic to stationary growth phase is associated with a saturation of the growth medium and a depletion of the readily available carbon source, which would also be expected to induce changes in metabolism.[15]

Figure 6-8 and Figure 6-9 visually illustrate the major metabolic pathways of yeast, and highlight the metabolic proteins which were differentially expressed in the glucose/glycerol (Figure 6-8) and log/stationary (Figure 6-9) comparisons. In Figure 6-8, it is particularly notable that numerous enzymes which are involved in the citric acid cycle are up-regulated in the yeast grown on glycerol media. Yeast is unable to derive energy from glycerol via fermentation to ethanol, as is possible with glucose, and therefore must use the citric acid cycle. It is therefore logical that enzymes involved in this cycle would be present in larger quantities in the glycerol sample than in glucose. Conversely, some proteins are up-regulated in the glycerol sample. For example, the protein phosphoglycerate mutase 1

(PMG1) is known to be unnecessary for growth on glycerol, which explains the fact that it was detected as being more abundant in the glucose sample.[16]

Aside from proteins directly involved in metabolic pathways of yeast, several other proteins which were detected as being differentially expressed can be associated with biochemical changes caused by growing yeast under different conditions. For instance, mitochondrial matrix factor 1 (MMF1) and its homologue HMF1, were detected as being up-regulated in the glycerol sample. These proteins are known to be essential for the growth of yeast on non-fermentable carbon sources, such as glycerol.[17] Likewise, the protein alanine:glyoxylate aminotransferase 1 (AGX1), which is involved in the biosynthesis of glycine, was detected as being more abundant in the glycerol sample. This observation correlates with the fact that expression of this protein is known to be repressed when yeast is grown on a glucose-containing medium.[18] A final example is NADPH dependent methylglyoxal reductase (GRE2), which catalyzes the reduction of methyl glyoxal, a cytotoxic compound. This protein was found to be up-regulated in the sample harvested at stationary phase, which is logical given that accumulation of waste products is likely to occur when the growth medium reaches saturation with yeast cells.[19]

Some of the proteins which were detected as being differentially expressed in the samples have no obvious link to changes which would be expected to occur due to differences in the growth medium or a shift from logarithmic growth to stationary phase. The expression of these proteins could be related to metabolic changes in an indirect manner. Alternatively, it is possible that some instances in which a protein was detected as being marginally up- or down-regulated were due to random variation between samples rather than biochemical changes. Inclusion of biological replicates in future studies would improve

certainty in interpretation of the data. Nevertheless, the fact that numerous well-documented changes in protein expression were detected using this method demonstrates its ability to reveal useful information when used for differential analysis of samples from a living organism.

6.5   Conclusions

Multidimensional liquid chromatography coupled to mass spectrometry shows significant potential as a method for differential proteomics studies. The hybrid top-down / bottom-up workflow allows differentially expressed proteins to be detected at the intact protein level, which reduces sample complexity as compared to purely bottom-up analyses. The use of 2D mass-slice chromatograms facilitates visual comparison of the intact protein data, and reduces the potential for confusion of neighboring peaks. Only fractions containing proteins observed to change in intensity between samples are subjected to further analysis by trypsin digest and LC-MS, which reduces the total workload. While the method is not able to discern subtle changes in protein expression, changes of a reasonable magnitude are easily detected. One area which requires further development is a reliable means of correlating data from the top-down and bottom-up portions of the experiment. Nevertheless, this label-free, gel-free approach to differential proteomics holds substantial advantages in terms of convenience and automation over more traditional methods. Therefore, variations of this method – and multidimensional LC in general – may see increasing application in real-world proteomic analyses.

6.6   Acknowledgment

I would like to acknowledge the contributions of Brenna McJury, a fellow graduate student, to the data presented in this chapter. In addition to performing some of the

experiments, Brenna analyzed a significant portion of the data and provided several figures, which she graciously agreed to allow me to include in this work.

6.7 References

(1)   Qian, W.-J.; Jacobs, J. M.; Liu, T.; Camp, D. G., II; Smith, R. D. *Molecular and Cellular Proteomics* **2006**, *5*, 1727-1744.

(2)   Hamler, R. L.; Zhu, K.; Buchanan, N. S.; Kreunin, P.; Kachman, M. T.; Miller, F. R.; Lubman, D. M. *Proteomics* **2004**, *4*, 562-577.

(3)   Hoffman, S. A.; Joo, W.-A.; Echan, L. A.; Speicher, D. W. *Journal of Chromatography B* **2007**, *849*, 43-52.

(4)   Roy, S. M.; Anderle, M.; Becker, C. H. *International Journal of Mass Spectrometry* **2004**, *238*, 163-171.

(5)   Pandey, A.; Mann, M. *Nature Biotechnology* **2000**, *405*, 837-846.

(6)   Monteoliva, L.; Albar, J. P. *Briefings in Functional Genomics and Proteomics* **2004**, *3*, 220-239.

(7)   Qian, W.-J.; Monroe, M. E.; Liu, T.; Jacobs, J. M.; Anderson, G. M.; Shen, Y.; Moore, R. J.; Anderson, D. J.; Zhang, R.; Calvano, S. E.; Lowry, S. F.; Xiao, W.; Modawer, L. L.; Davis, R. W.; Tompkins, R. G.; Camp, D. G., II; Smith, R. D. *Molecular and Cellular Proteomics* **2005**, *4*, 700-709.

(8)   Ong, S. E.; Blagoev, B.; Kractchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M. *Molecular and Cellular Proteomics* **2002**, *1*, 376-386.

(9)   Wang, W.; Zhou, H.; Lin, H.; Roy, S.; Shaler, T. A.; Hill, L. R.; Norton, S.; Kumar, P.; Anderle, M.; Beker, C. H. *Analytical Chemistry* **2003**, *75*, 4818-4826.

(10)   Kolkman, A.; Daran-Lapujade, P.; Fullaondo, A.; Olsthoorn, M. M.; Pronk, J. T.; Slijper, M.; Heck, A. J. *Molecular Systems Biology* **2006**, *2*, 1-16.

(11)   Ghaemmaghami, S.; Huh, W.-K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. *Nature Biotechnology* **2003**, *425*, 737-741.

(12)   Millea, K. M.; Krull, I. S.; Cohen, S. A.; Gebler, J. C.; Berger, S. J. *Journal of Proteome Research* **2006**, *5*, 135-146.

(13)   Kreunin, P.; Urquidi, V.; Lubman, D. M.; Goodison, S. *Proteomics* **2004**, *4*, 2754-2765.

(14)   Fraenkel, D. G. In *The Molecular Biology of the Yeast Saccharomyces*; Strathern, J., Jones, E. W., Broach, J. R., Eds.; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1982, pp 1-37.

(15)   Van Doorn, J.; Valkenburg, J. A. C.; Scholte, M. E.; Oehlen, L. J. W. M.; Van Driel, R.; Postma, P. W.; Nanninga, N.; Van Dam, K. *Journal of Bacteriology* **1988**, *170*, 4808-4815.

(16)   Rodicio, R.; Heinisch, J. J.; Hollenberg, C. P. *Gene* **1993**, *125*, 125-133.

(17)   Kim, J.-M.; Yoshikawa, H.; Shirahige, K. *Genes to Cells* **2001**, *6*, 507-517.

(18)   Schlosser, T.; Gatgens, C.; Weber, U.; Stahmann, K.-P. *Yeast* **2003**, *21*, 63-73.

(19)   Garay-Arroyo, A.; Covarrubias, A. A. *Yeast* **1999**, *15*, 879-892.

6.8   Tables

Time (min)	% mobile phase B
0	0
20	7
600	67
675	67
680	0

Table 6-1: AXC gradient profile used for the first dimension of the on-line LC x LC separation performed using the long AXC column. Mobile phase A was 0.01 M ammonium acetate, pH 9.0. Mobile phase B was 1 M ammonium acetate, pH 9.0. The flow rate was 0.2 mL/min. The total volume of sample injected was 225 µL per run; the mass of protein injected was 2.25 mg.

Time (min)	% mobile phase B
0	5
2	20
5	30
15	45
15.5	90
16	90
16.5	5

Table 6-2: RPLC gradient profile used for the second dimension of the on-line LC x LC separation. Valve switch events in the 2D separation were separated by 20 minutes in order to allow for full re-equilibration with the starting mobile phase before the next sample was loaded. Mobile phase A was 0.2% (v/v) formic acid in deionized water. Mobile phase B was 0.2% (v/v) formic acid in acetonitrile. The flow rate was 1.0 mL/min.

Swiss-Prot name	Ordered locus name	Description	Pep. hits	Intact mass	Up-reg. in	Fold change
ACBP	YGR037C	Acyl-CoA binding protein	1	9930.17	Glucose	1.5
TRX2	YGR209C	Thioredoxin 2	2	11202.1	Glucose	-
GLRX1	YCL035C	Glutaredoxin 1	4	12245.7	Glucose	21.6
SODC	YJR104C	Superoxide dismutase Cu Zn	1	15722.9	Glucose	1.9
YC026	YCL026C-B	Uncharacterized protein YCL026C B	2	20906.2	Glucose	9.3
TPIS	YDR050C	Triosephosphate isomerase	4	26666.3	Glucose	1.7
MRS1	YIR021W	Mitochondrial RNA splicing protein MRS1	1	41078.3	Glucose	13.5
CYS3	YAL012W	Cystathionine gamma lyase	5	42414.4	Glucose	-
OYE2	YHR179W	NADPH dehydrogenase 2	7	44884.2	Glucose	9.6
G6PI	YBR196C	Glucose 6 phosphate isomerase	24	61216.5	Glucose	1.3
IPB2	YNL015W	Protease B inhibitors 2 and 1	2	8459.56	Glycerol	8.6
PROF	YOR122C	Profilin	1	13589.5	Glycerol	11.8
HMF1	YER057C	Protein HMF1	1	13775.5	Glycerol	4.7
PMG1	YKL152C	Phosphoglycerate mutase 1	8	27482.4	Glycerol	2.0
YM71	YMR226C	Uncharacterized oxioreductase	6	29071.6	Glycerol	5.3
BMH1	YER177W	Protein BMH1	10	30004.2	Glycerol	1.8
BMH2	YDR099W	Protein BMH2	10	30974.2	Glycerol	1.6
SUCA	YOR142W	Succinyl CoA ligase, ADP forming subunit alpha	4	33222.7	Glycerol	15.5
MDHM	YKL085W	Malate dehydrogenase, mitochondrial	5	33836.2	Glycerol	6.3
ADH1	YOL086C	Alcohol dehydrogenase	1	36646.2	Glycerol	-
TAL1	YLR354C	Transaldolase	6	36951.3	Glycerol	10.7
IDH2	YOR136W	Isocitrate dehydrogenase NAD subunit 2	3	37802.1	Glycerol	-
RIR4	YGR180C	Ribonucleoside diphosphate reductase small chain 2	4	40130.7	Glycerol	5.2
MDHC	YOL126C	Malate dehydrogenase, cytoplasmic	2	40604.2	Glycerol	-
AGX1	YFL030W	Alanine:glyoxylate aminotransferase 1	6	41778.4	Glycerol	-
AATC	YLR027C	Aspartate aminotransferase, cytoplasmic	5	45893.4	Glycerol	7.4
CISY1	YNR001C	Citrate synthase	7	49221.9	Glycerol	35.9
GLYM	YBR263W	Serine hydroxymethyltransferase	3	51605.7	Glycerol	-
ALDH4	YOR374W	K+ activated aldehyde dehydrogenase	15	53979.1	Glycerol	-
TKT1	YPR074C	Transketolase	8	73677.7	Glycerol	-

Table 6-3: Proteins identified as being significantly differentially expressed in yeast grown on glucose medium versus glycerol medium (both samples were harvested during log-growth phase). Heading terms are defined as follows: **Swiss-Prot name:** Swiss-Prot database entry name (all proteins have _YEAST suffix); **Ordered locus name:** Systematic gene name **MW:** Measured intact protein molecular weight; Pep. Hits: number of peptide MS/MS hits used to identify protein; **Up-reg in.**: Identity of the sample in which the protein was present at a higher abundance; **Fold change**: ratio of abundance in samples (more abundant / less abundant). '-' indicates that the protein was not detected in one of the samples, thus a fold change could not be calculated.

Swiss-Prot name	Ordered locus name	Description	Pep. hits	Intact mass	Up-reg in	Fold change
ACBP	YGR037C	Acyl-CoA binding protein	1	9930.17	Log	2.5
CH10	YOR020C	10kDa Heat shock protein, mitochondrial	3	11283.9	Log	29.5
HSP12	YFL014W	12kDa Heat shock protein	5	11604.6	Log	373.1
YL364	YLR364W	Glutaredoxin-like protein	2	12069	Log	5.1
TBCA	YOR265W	Tubulin specific chaperone A	3	12248.4	Log	-
PNC1	YGL037C	Nicotinamidase	2	24993.9	Log	17.9
MRP8	YKL142W	Uncharacterized protein MRP8	6	25004.8	Log	-
BMH1	YER177W	Protein BMH1	10	29998.4	Log	-
BMH2	YDR099W	Protein BMH2	10	30968.7	Log	-
G3P3	YGR192C	Glyceraldehyde 3 phosphate dehydrogenase 3	11	35617.8	Log	122.6
CYS3	YAL012W	Cystathionine gamma lyase	5	42414.4	Log	18.5
PGK	YCR012W	Phosphoglycerate kinase	16	44652.7	Log	-
SODC	YJR104C	Superoxide dismutase Cu Zn	1	15719.6	Stat	1.5
COFI	YLL050C	Cofilin	8	15809.9	Stat	1.9
KGUA	YDR454C	Guanylate kinase	1	20901.8	Stat	7.1
SODM	YHR008C	Superoxide dismutase Mn mitochondrial	4	23080.5	Stat	3.7
PMG1	YKL152C	Phosphoglycerate mutase 1	8	27477.9	Stat	6.8
G3P1	YJL052W	Glyceraldehyde 3 phosphate dehydrogenase 1	11	35742.1	Stat	-
TAL1	YLR354C	Transaldolase	6	36942.8	Stat	7.4
DHOM	YJR139C	Homoserine dehydrogenase	11	38404.9	Stat	2.2
MRS1	YIR021W	Mitochondrial RNA splicing protein MRS1	1	41070.4	Stat	1.4
G6PI	YBR196C	Glucose 6 phosphate isomerase	24	61202.9	Stat	1.7

Table 6-4: Proteins identified as being significantly differentially expressed in yeast harvested during log-growth phase versus stationary phase (both samples were grown on glucose medium). Heading terms are defined as follows: **Swiss-Prot name:** Swiss-Prot database entry name (all proteins have _YEAST suffix); **Ordered locus name:** Systematic gene name; **MW:** Measured intact protein molecular weight; **Pep. Hits**: number of peptide MS/MS database hits used to identify protein; **Up-reg in.**: Identity of the sample in which the protein was more abundant; **Fold change**: ratio of abundance in samples (more abundant / less abundant). '-' indicates that the protein was not detected in one of the samples, thus a fold change could not be calculated.

6.9   Figures

Sample: 3 yeast
protein extracts

1) On-line LC x LC-MS
to separate
intact proteins

2) Create & compare 2D
mass slice chromatograms

3) Select fractions with
differentially expressed
proteins & perform digest

4) Capilary
RPLC-MS/MS to
analyze peptides

5) Database search to
identify differentially
expressed proteins
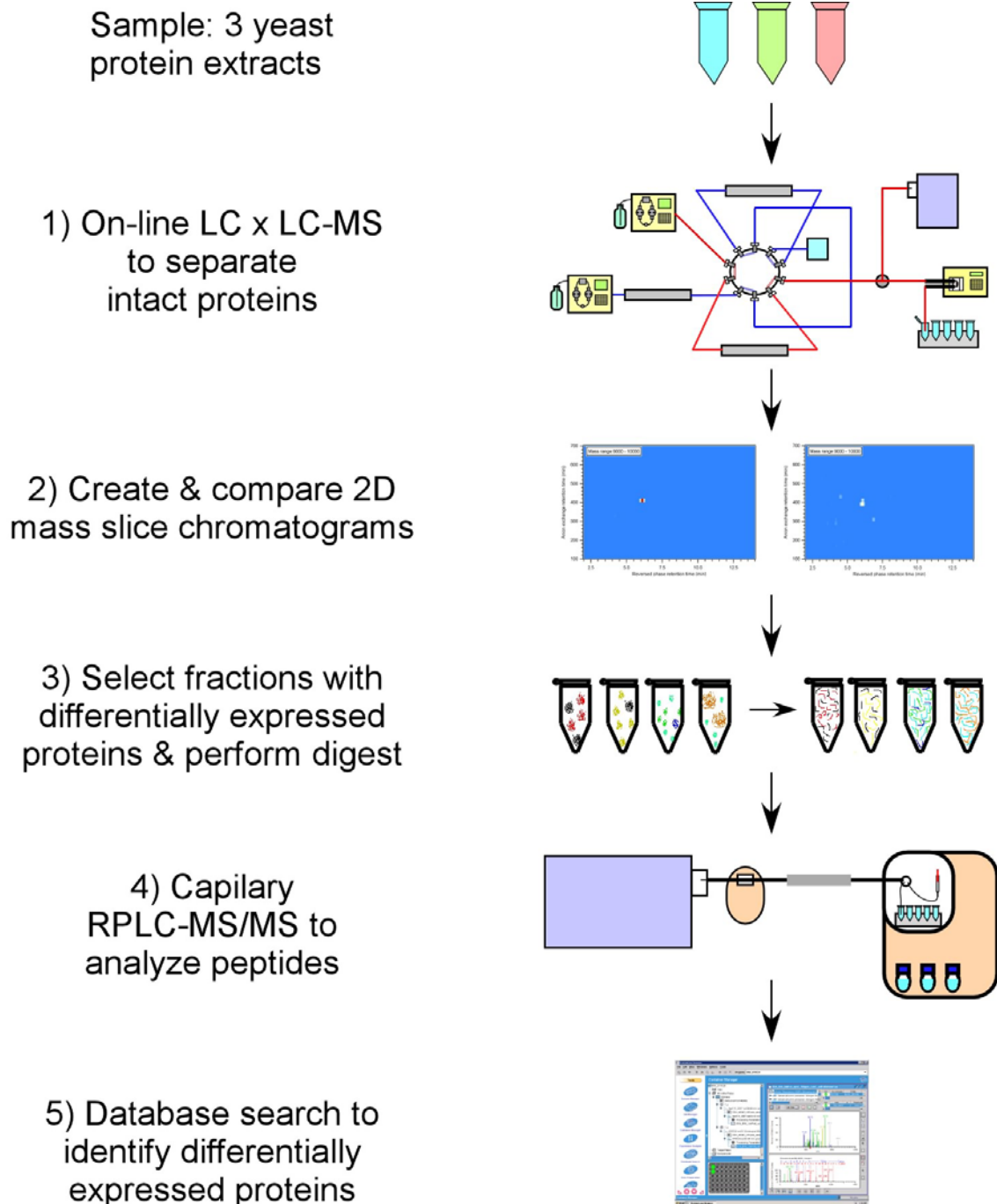
Figure 6-1: Overview of differential proteomics approach
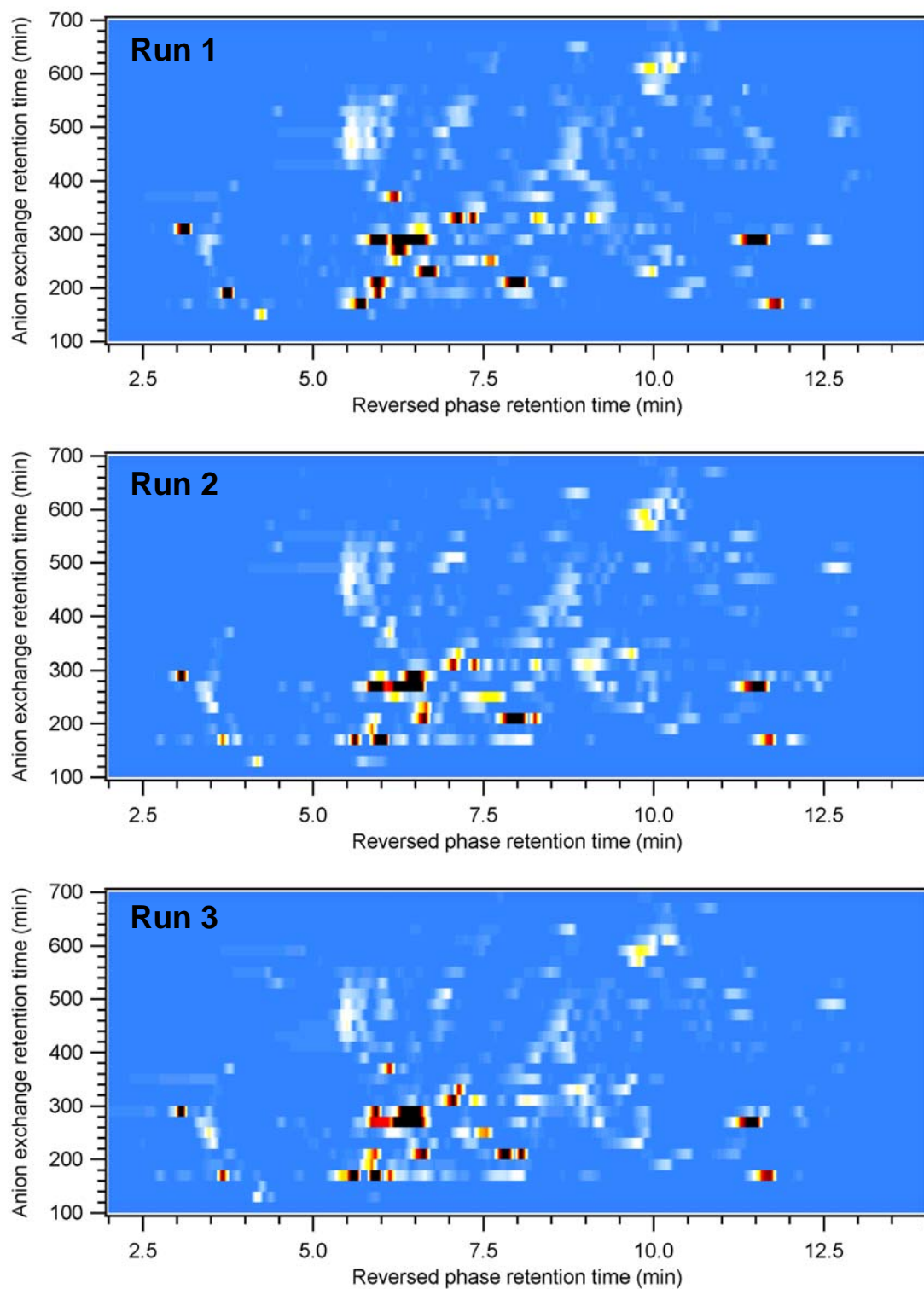
Figure 6-2: 2D chromatograms from replicate online LC x LC separations of a protein extract from a yeast culture grown on glycerol medium, harvested during log-growth phase.
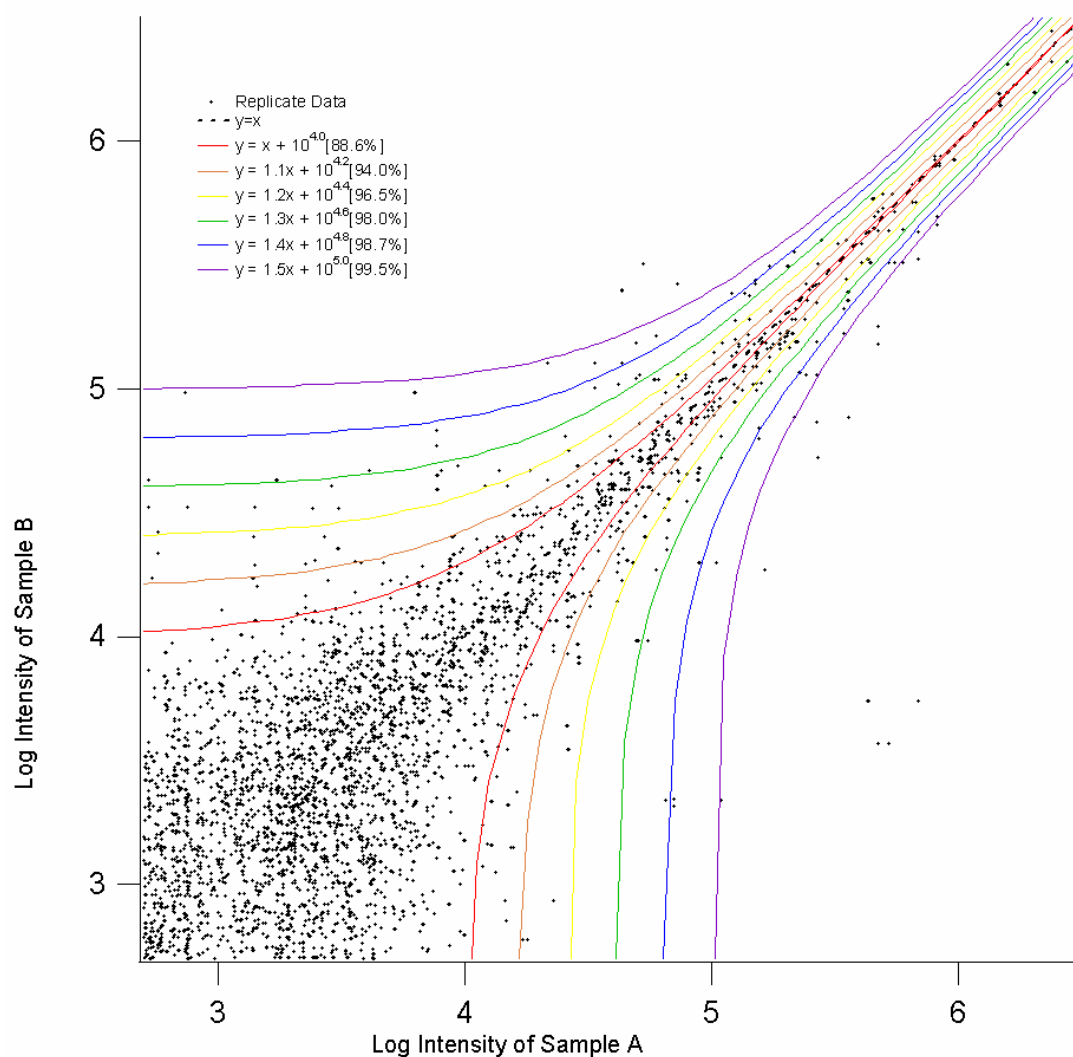
Figure 6-3: Log-log plot comparing intensities of protein peaks in three replicate runs of a yeast protein extract. Each black dot represents the intensity of a protein found in two different samples. The colored curves are linear equations which were determined to contain different portions of the points in the data set. They are useful for assigning bounds for typical run-to-run variability as a function of peak intensity.
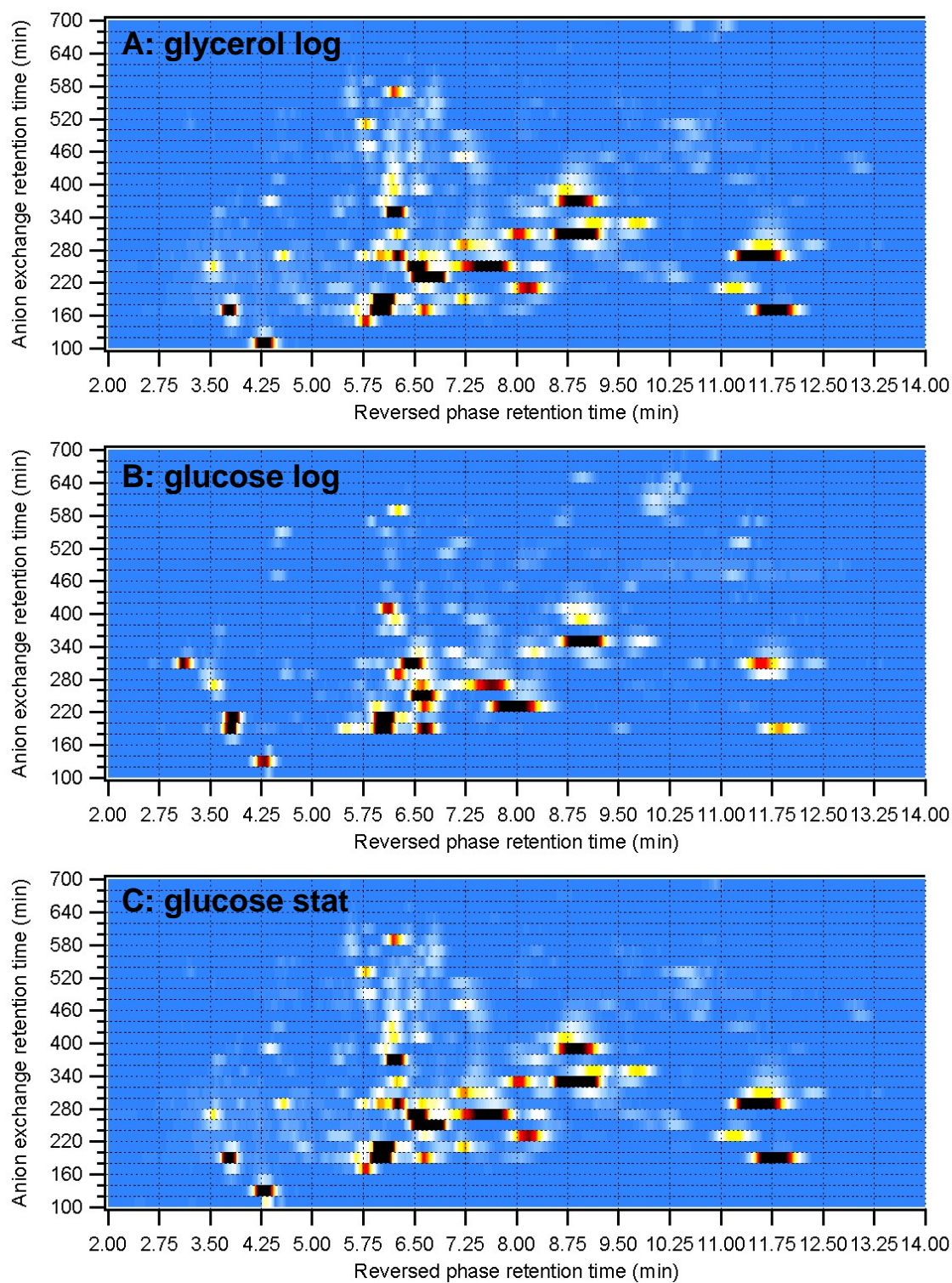
Figure 6-4: 2D chromatograms from online LC x LC separations of three yeast protein extracts. The grid represents the time when fractions were switched. Run conditions are defined in Table 6-3 and Table 6-4.

Figure 6-5: Mass-slice chromatograms for comparison between yeast cultures grown on glucose medium and glycerol medium

Glucose medium, log phase **Glucose medium, stationary phase**
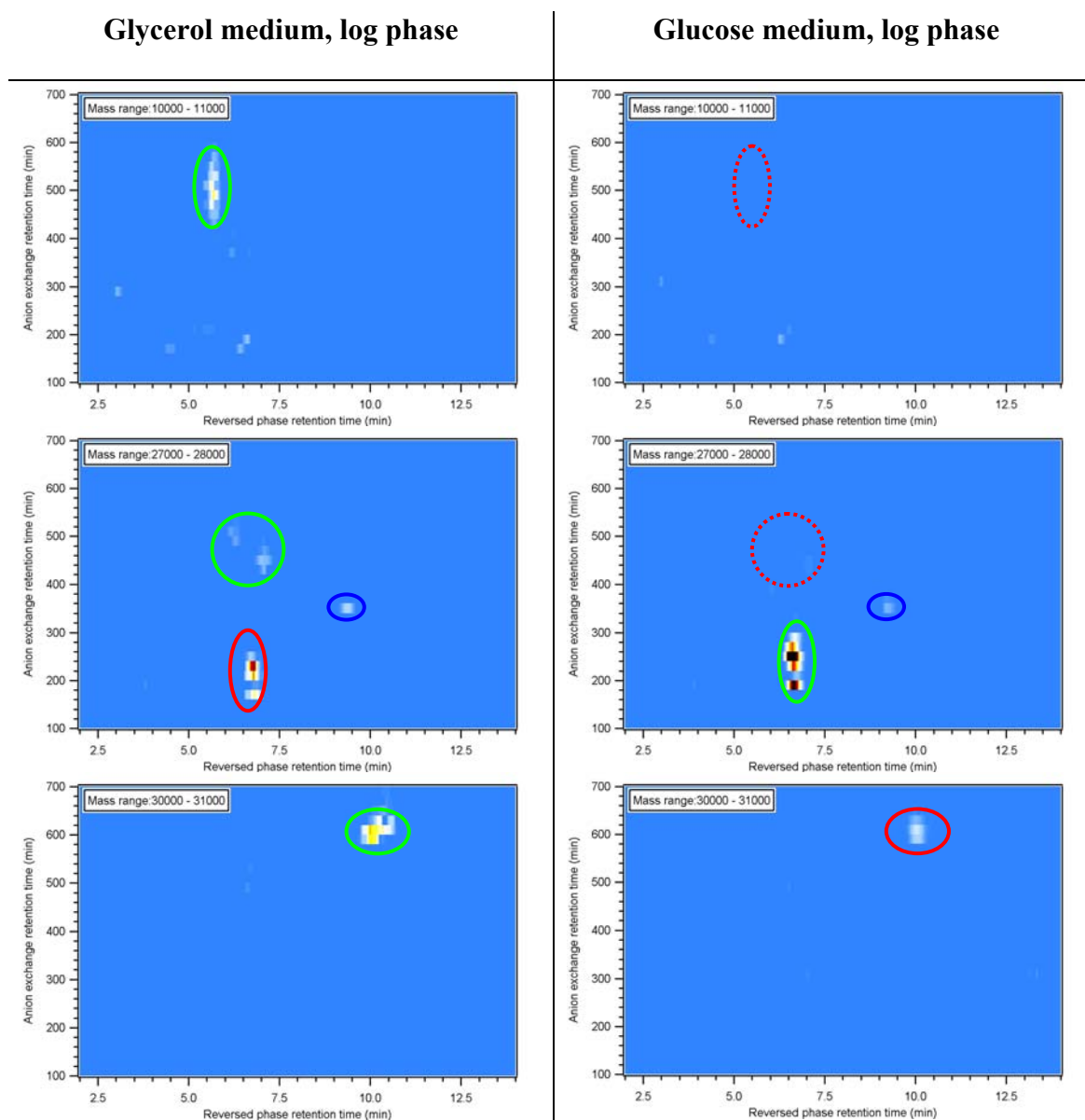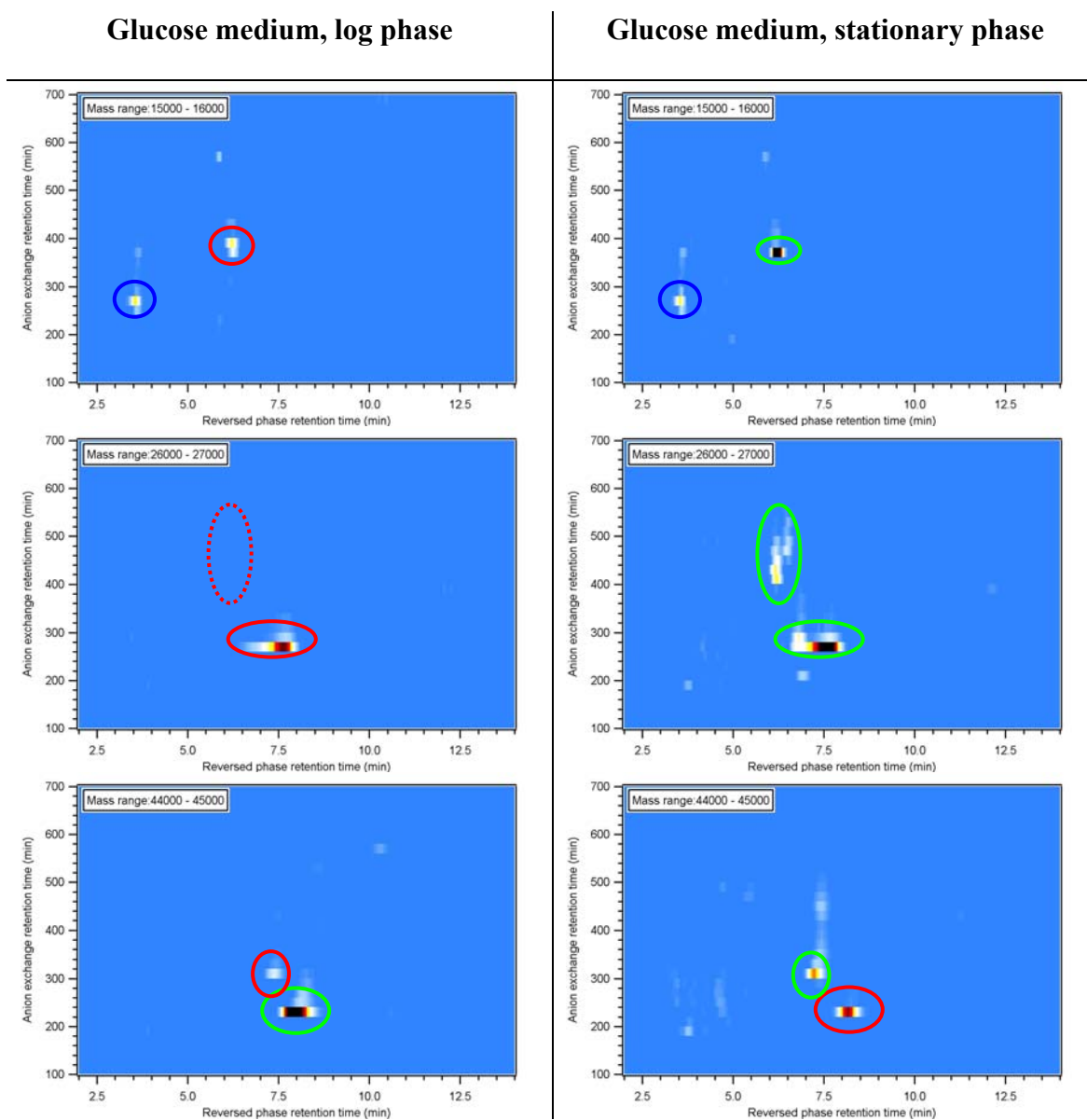
Key:

Protein intensity essentially unchanged

Protein up-regulated in sample

Protein down-regulated in sample

Protein entirely absent from sample

Figure 6-6: Mass-slice chromatograms for comparison between yeast cultures harvested at log-growth phase and stationary phase
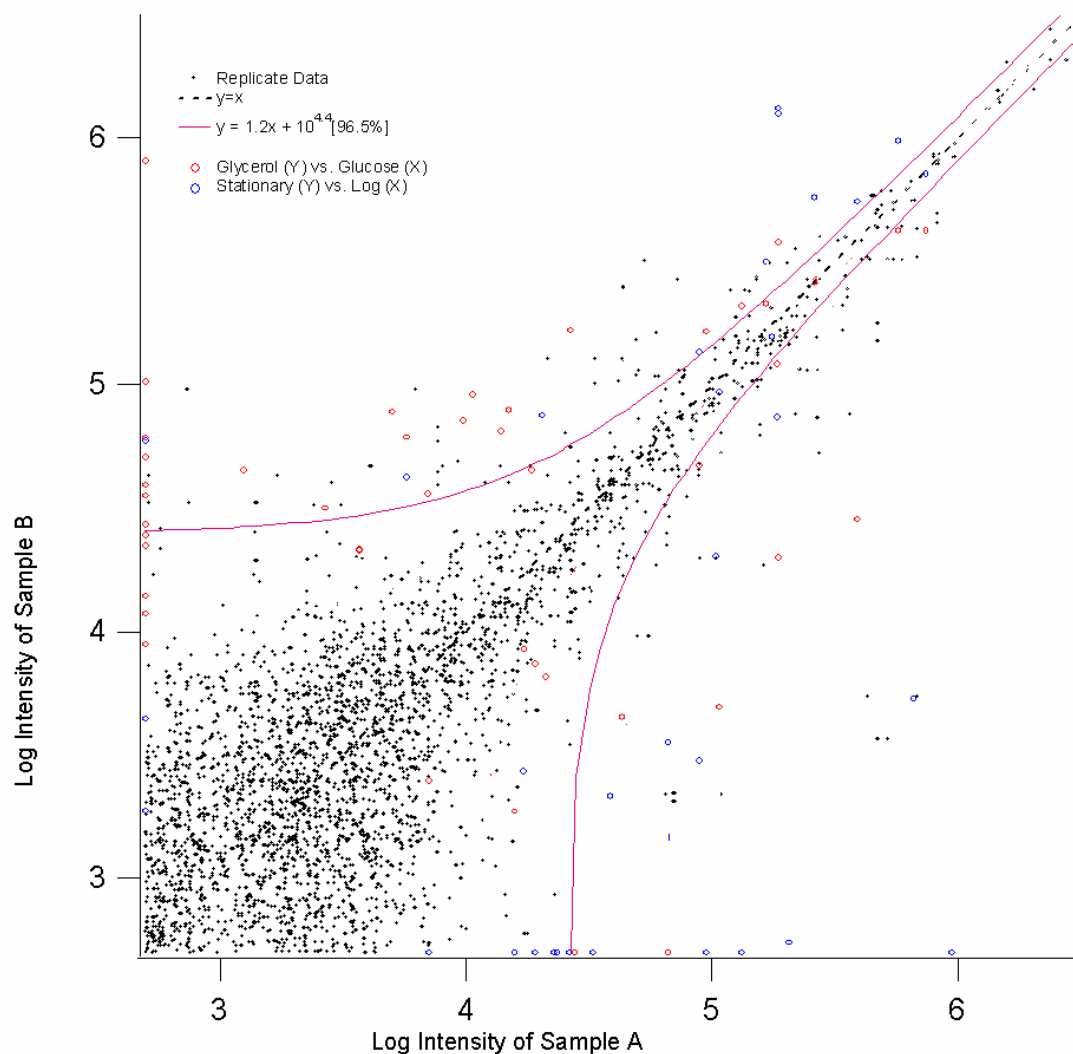
Figure 6-7: Log-log plot displaying differential peak intensity for both glucose vs. glycerol and log vs. stationary yeast protein extract comparisons. The comparison data are overlaid upon the data from replicate runs of the glycerol sample. The curve shown represents a "confidence interval" of 96.5%. Points in the glucose/glycerol and log/stationary phase comparisons which fall outside the curve are considered to be significantly different, indicating a change in protein expression between the samples.
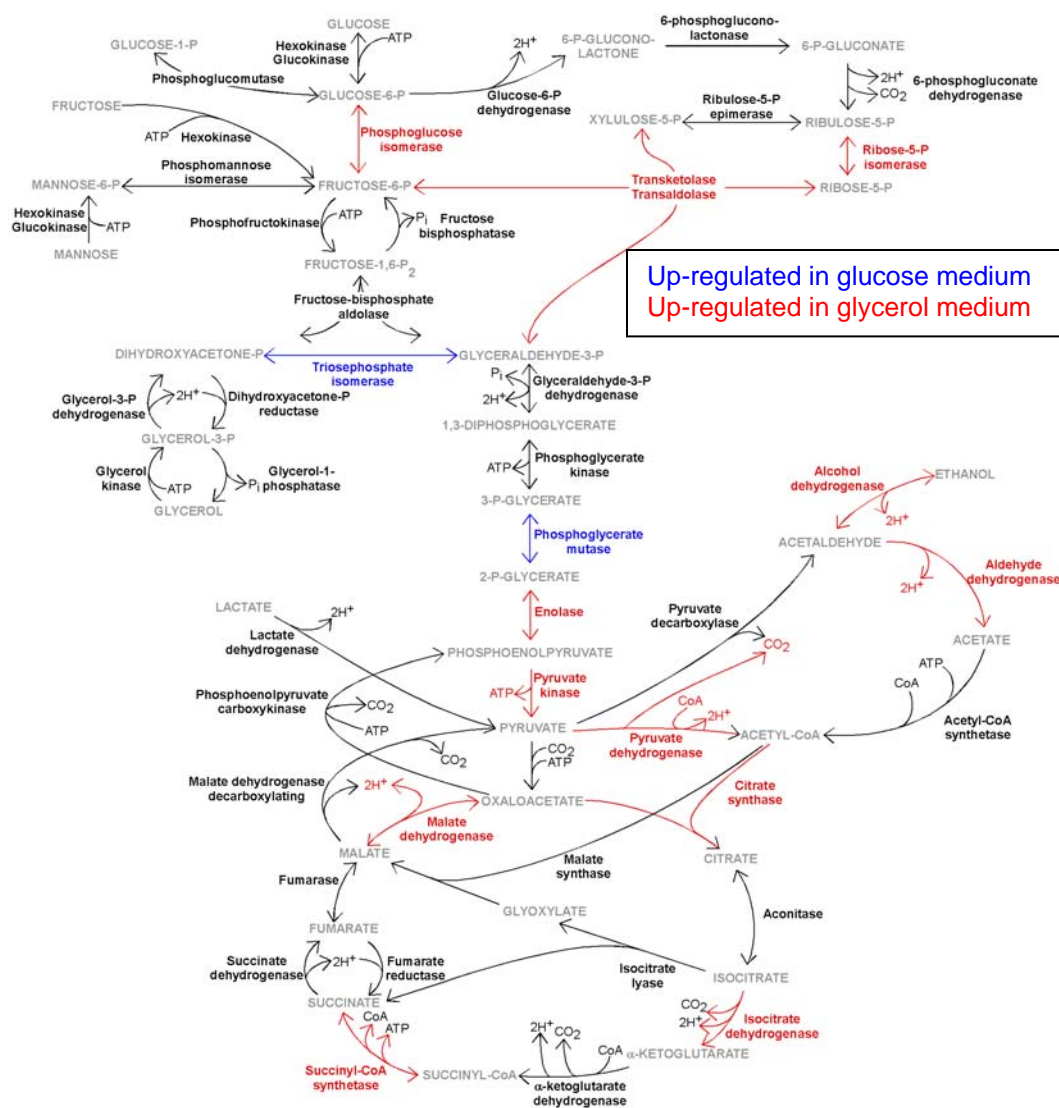
Figure 6-8: Yeast metabolic proteins differentially expressed in the samples cultured on glucose and glycerol media. Proteins more abundant in the glucose sample are shown in blue, proteins more abundant in the glycerol sample are shown in red, and proteins not significantly different in abundance are shown in black.
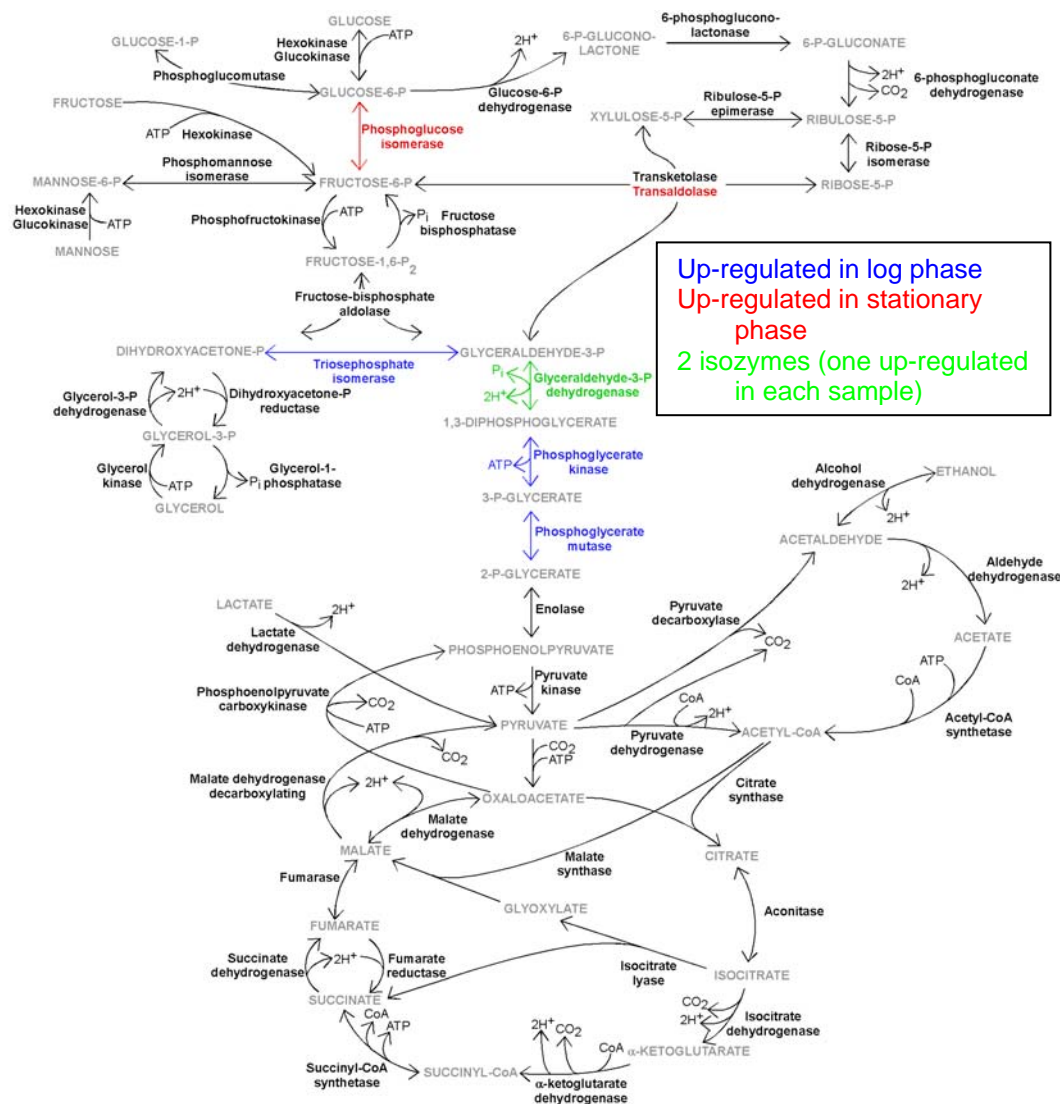
Figure 6-9: Yeast metabolic proteins differentially expressed during logarithmic growth phase and stationary phase. Proteins more abundant during log growth phase are shown in blue, proteins more abundant in stationary phase are shown in red, and proteins not significantly different in abundance are shown in black. The protein shown in green has two isoforms; one isoform was more abundant in each sample.

APPENDIX

A complete table of the *E. coli* proteins identified in Chapters 4 and 5 is provided

below. Terms are as follows: **Ref #**: number assigned in Table 4-4, Table 4-6, or Table 5-5;

Protein: SwissProt database entry name; **Description**: SwissProt protein description

| Ref #	Protein	Description
1	CSPC_ECOLI	Cold shock-like protein cspC (CSP-C).
2	CSPE_ECOLI	Cold shock-like protein cspE (CSP-E).
3	YJBJ_ECOLI	UPF0337 protein yjbJ.
4	HDEB_ECOLI	Protein hdeB precursor (10K-L protein).
5	PTHP_ECOLI	Phosphocarrier protein HPr (Histidine-containing protein).
6	RS15_ECOLI	30S ribosomal protein S15.
7	CH10_ECOLI	10 kDa chaperonin (Protein Cpn10) (groES protein).
8	FETP_ECOLI	Probable Fe(2+)-trafficking protein.
9	YGIW_ECOLI	Protein ygiW precursor.
10	YEGP_ECOLI	UPF0339 protein yegP.
11	RL7_ECOLI	50S ribosomal protein L7/L12 (L8).
12	GRCA_ECOLI	Autonomous glycyl radical cofactor.
13	YCCU_ECOLI	Uncharacterized protein yccU.
14	OSMC_ECOLI	Peroxiredoxin osmC (EC 1.11.1.15) (Osmotically-inducible protein C).
15	RS6_ECOLI	30S ribosomal protein S6 [Contains: 30S ribosomal protein S6, fully modified isoform; 30S ribosomal protein S6, non-modified isoform].
16	NDK_ECOLI	Nucleoside diphosphate kinase (EC 2.7.4.6) (NDK) (NDP kinase)
17	HNS_ECOLI	DNA-binding protein H-NS (Histone-like protein HLP-II) (Protein H1)
18	SODC_ECOLI	Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
19	USPA_ECOLI	Universal stress protein A.
20	YJGK_ECOLI	Uncharacterized protein yjgK.
21	TPX_ECOLI	Thiol peroxidase (EC 1.11.1.-) (Scavengase P20).
22	PTGA_ECOLI	Glucose-specific phosphotransferase enzyme IIA component (EC 2.7.1.-)
23	OSMY_ECOLI	Osmotically-inducible protein Y precursor.
24	BFR_ECOLI	Bacterioferritin (BFR) (Cytochrome b-1) (Cytochrome b-557).
25	FABA_ECOLI	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase (EC 4.2.1.60)
26	FTNA_ECOLI	Ferritin-1 (EC 1.16.3.1).
27	YFBU_ECOLI	UPF0304 protein yfbU.
28	AHPC_ECOLI	Alkyl hydroperoxide reductase subunit C (EC 1.11.1.15) (Peroxiredoxin)
29	WRBA_ECOLI	Flavoprotein wrbA (Trp repressor-binding protein).
30	SODF_ECOLI	Superoxide dismutase [Fe] (EC 1.15.1.1).
31	RPIA_ECOLI	Ribose-5-phosphate isomerase A (EC 5.3.1.6) (Phosphoriboisomerase A)
32	KAD_ECOLI	Adenylate kinase (EC 2.7.4.3) (ATP-AMP transphosphorylase) (AK).
33	RPE_ECOLI	Ribulose-phosphate 3-epimerase (EC 5.1.3.1) (Pentose-5-phosphate 3-epimerase) (PPE) (R5P3E).
34	ARTI_ECOLI	Arginine-binding periplasmic protein 1 precursor.
35	FABG_ECOLI	3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) (3-ketoacyl-acyl carrier protein reductase).
36	ARGT_ECOLI	Lysine-arginine-ornithine-binding periplasmic protein precursor (LAO-binding protein).

Ref #	Protein	Description
37	DEOD_ECOLI	Purine nucleoside phosphorylase deoD-type (EC 2.4.2.1) (PNP) (Inosine phosphorylase).
38	FLIY_ECOLI	Cystine-binding periplasmic protein precursor (CBP) (Protein fliY)
39	HISJ_ECOLI	Histidine-binding periplasmic protein precursor (HBP).
40	TPIS_ECOLI	Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Triose-phosphate isomerase).
41	UDP_ECOLI	Uridine phosphorylase (EC 2.4.2.3) (UrdPase) (UPase).
42	KDUD_ECOLI	2-deoxy-D-gluconate 3-dehydrogenase (EC 1.1.1.125) (2-keto-3-deoxygluconate oxidoreductase).
43	FABI_ECOLI	Enoyl-[acyl-carrier-protein] reductase [NADH] (EC 1.3.1.9) (NADH-dependent enoyl-ACP reductase).
44	GPMA_ECOLI	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (EC 5.4.2.1)
45	RBSB_ECOLI	D-ribose-binding periplasmic protein precursor.
46	EFTS_ECOLI	Elongation factor Ts (EF-Ts).
47	HCHA_ECOLI	Chaperone protein hchA (Hsp31) (EcHsp31).
48	YTFG_ECOLI	Uncharacterized oxidoreductase ytfG (EC 1.-.-.-).
49	MDH_ECOLI	Malate dehydrogenase (EC 1.1.1.37).
50	DGAL_ECOLI	D-galactose-binding periplasmic protein precursor (GBP) (D-galactose/D-glucose-binding protein) (GGBP).
51	PROX_ECOLI	Glycine betaine-binding periplasmic protein precursor.
52	CYSK_ECOLI	Cysteine synthase A (EC 2.5.1.47) (O-acetylserine sulfhydrylase A) (O-acetylserine (Thiol)-lyase A) (CSase A) (Sulfate starvation-induced protein 5) (SSI5).
53	ASPG2_ECOLI	L-asparaginase 2 precursor (EC 3.5.1.1) (L-asparaginase II) (L-
54	TALB_ECOLI	Transaldolase B (EC 2.2.1.2).
55	TALA_ECOLI	Transaldolase A (EC 2.2.1.2).
56	6PGL_ECOLI	6-phosphogluconolactonase (EC 3.1.1.31) (6-P-gluconolactonase) (Pgl).
57	GLPQ_ECOLI	Glycerophosphoryl diester phosphodiesterase precursor (EC 3.1.4.46)
58	ALF_ECOLI	Fructose-bisphosphate aldolase class 2 (EC 4.1.2.13)
59	MALE_ECOLI	Maltose-binding periplasmic protein precursor
60	EFTU_ECOLI	Elongation factor Tu (EF-Tu) (P-43).
61	GLYA_ECOLI	Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase)
62	UGPB_ECOLI	sn-glycerol-3-phosphate-binding periplasmic protein ugpB precursor.
63	TNAA_ECOLI	Tryptophanase (EC 4.1.99.1) (L-tryptophan indole-lyase) (TNase).
64	DPPA_ECOLI	Periplasmic dipeptide transport protein precursor
65	RL29_ECOLI	50S ribosomal protein L29.
66	YCCJ_ECOLI	Uncharacterized protein yccJ.
67	RL28_ECOLI	50S ribosomal protein L28.
68	DBHB_ECOLI	DNA-binding protein HU-beta (NS1) (HU-1).
69	DBHA_ECOLI	DNA-binding protein HU-alpha (NS2) (HU-2).
70	RL25_ECOLI	50S ribosomal protein L25.
71	RL24_ECOLI	50S ribosomal protein L24.
72	IHFA_ECOLI	Integration host factor subunit alpha (IHF-alpha).
73	RL21_ECOLI	50S ribosomal protein L21.
74	RS10_ECOLI	30S ribosomal protein S10.
75	RL18_ECOLI	50S ribosomal protein L18.
76	GLRX4_ECOLI	Glutaredoxin-4 (Grx4) (Monothiol glutaredoxin).
77	YJGF_ECOLI	UPF0076 protein yjgF.
78	RL11_ECOLI	50S ribosomal protein L11.
79	RL15_ECOLI	50S ribosomal protein L15.

Ref #	Protein	Description
80	SKP_ECOLI	Chaperone protein skp precursor (Seventeen kilodalton protein)
81	RL9_ECOLI	50S ribosomal protein L9.
82	RL13_ECOLI	50S ribosomal protein L13.
83	ASNC_ECOLI	Regulatory protein asnC.
84	RL10_ECOLI	50S ribosomal protein L10 (50S ribosomal protein L8).
85	YBAY_ECOLI	Uncharacterized lipoprotein ybaY precursor.
86	PPIB_ECOLI	Peptidyl-prolyl cis-trans isomerase B (EC 5.2.1.8) (PPIase B)
87	RL6_ECOLI	50S ribosomal protein L6.
88	RS7_ECOLI	30S ribosomal protein S7.
89	NUSG_ECOLI	Transcription antitermination protein nusG.
90	CLPP_ECOLI	ATP-dependent Clp protease proteolytic subunit precursor
91	YRBC_ECOLI	Protein yrbC precursor.
92	RL3_ECOLI	50S ribosomal protein L3.
93	ALKH_ECOLI	KHG/KDPG aldolase [Includes: 4-hydroxy-2-oxoglutarate aldolase
94	SODM_ECOLI	Superoxide dismutase [Mn] (EC 1.15.1.1) (MnSOD).
95	CRP_ECOLI	Catabolite gene activator (cAMP receptor protein)
96	GLRX2_ECOLI	Glutaredoxin-2 (Grx2).
97	RL1_ECOLI	50S ribosomal protein L1.
98	GLNH_ECOLI	Glutamine-binding periplasmic protein precursor (GlnBP).
99	RS3_ECOLI	30S ribosomal protein S3.
100	YBGI_ECOLI	UPF0135 protein ybgI.
101	UDP_ECOLI	Uridine phosphorylase (EC 2.4.2.3) (UrdPase) (UPase).
102	DEOC_ECOLI	Deoxyribose-phosphate aldolase (EC 4.1.2.4) (Phosphodeoxyriboaldolase)
103	FARR_ECOLI	Fatty acyl-responsive regulator (P30 protein).
104	SUCD_ECOLI	Succinyl-CoA ligase [ADP-forming] subunit alpha (EC 6.2.1.5)
105	DAPD_ECOLI	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase
106	KDUI_ECOLI	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (EC 5.3.1.17)
107	GLTI_ECOLI	Glutamate/aspartate periplasmic-binding protein precursor.
108	KDGK_ECOLI	2-dehydro-3-deoxygluconokinase (EC 2.7.1.45)
109	G3P1_ECOLI	Glyceraldehyde-3-phosphate dehydrogenase A (EC 1.2.1.12) (GAPDH-A).
110	USPE_ECOLI	Universal stress protein E.
111	YGHZ_ECOLI	Uncharacterized protein yghZ.
112	SERC_ECOLI	Phosphoserine aminotransferase (EC 2.6.1.52)
113	DHAS_ECOLI	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11) (ASA dehydrogenase)
114	PGK_ECOLI	Phosphoglycerate kinase (EC 2.7.2.3).
115	SUCC_ECOLI	Succinyl-CoA synthetase beta chain (EC 6.2.1.5) (SCS-beta).
116	AGP_ECOLI	Glucose-1-phosphatase precursor (EC 3.1.3.10) (G1Pase).
117	AAT_ECOLI	Aspartate aminotransferase (EC 2.6.1.1) (Transaminase A) (ASPAT).
118	DEOB_ECOLI	Phosphopentomutase (EC 5.4.2.7) (Phosphodeoxyribomutase).
119	SURA_ECOLI	Chaperone surA precursor (Peptidyl-prolyl cis-trans isomerase surA)
120	ENO_ECOLI	Enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase)
121	IDH_ECOLI	Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)
122	ACEA_ECOLI	Isocitrate lyase (EC 4.1.3.1) (Isocitrase) (Isocitratase) (ICL).
123	CISY_ECOLI	Citrate synthase (EC 2.3.3.1).
124	TIG_ECOLI	Trigger factor (TF).
125	SYS_ECOLI	Seryl-tRNA synthetase (EC 6.1.1.11) (Seryl-tRNA(Ser/Sec) synthetase)

Ref #	Protein	Description
126	ATPB_ECOLI	ATP synthase subunit beta (EC 3.6.3.14) (ATPase subunit beta)
127	DLDH_ECOLI	Dihydrolipoyl dehydrogenase (EC 1.8.1.4)
128	KPYK2_ECOLI	Pyruvate kinase II (EC 2.7.1.40) (PK-2).
129	6PGD_ECOLI	6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44).
130	ALDA_ECOLI	Aldehyde dehydrogenase A (EC 1.2.1.22) (Lactaldehyde dehydrogenase)
131	ASPA_ECOLI	Aspartate ammonia-lyase (EC 4.3.1.1) (Aspartase).
132	DCEA_ECOLI	Glutamate decarboxylase alpha (EC 4.1.1.15) (GAD-alpha).
133	YHJJ_ECOLI	Protein yhjJ precursor.
134	ATPA_ECOLI	ATP synthase subunit alpha (EC 3.6.3.14) (ATPase subunit alpha)
135	OPGG_ECOLI	Glucans biosynthesis protein G precursor.
136	GLPK_ECOLI	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase)
137	ALDB_ECOLI	Aldehyde dehydrogenase B (EC 1.2.1.-).
138	CH60_ECOLI	60 kDa chaperonin (Protein Cpn60) (groEL protein).
139	PUR9_ECOLI	Bifunctional purine biosynthesis protein purH
140	PPCK_ECOLI	Phosphoenolpyruvate carboxykinase [ATP] (EC 4.1.1.49)
141	RS1_ECOLI	30S ribosomal protein S1.
142	G6PI_ECOLI	Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI)
143	SYD_ECOLI	Aspartyl-tRNA synthetase (EC 6.1.1.12) (Aspartate--tRNA ligase)
144	GLMS_ECOLI	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]
145	DNAK_ECOLI	Chaperone protein dnaK (Heat shock protein 70)
146	HTPG_ECOLI	Chaperone protein htpG (Heat shock protein htpG)
147	TKT1_ECOLI	Transketolase 1 (EC 2.2.1.1) (TK 1).
148	PNP_ECOLI	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8)
149	EFG_ECOLI	Elongation factor G (EF-G).
150	MASZ_ECOLI	Malate synthase G (EC 2.3.3.9) (MSG).
151	MAO2_ECOLI	NADP-dependent malic enzyme (EC 1.1.1.40) (NADP-ME).
152	CATE_ECOLI	Catalase HPII (EC 1.11.1.6) (Hydroxyperoxidase II).
153	PFLB_ECOLI	Formate acetyltransferase 1 (EC 2.3.1.54) (Pyruvate formate-lyase 1).
154	SYFB_ECOLI	Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
155	ACON2_ECOLI	Aconitate hydratase 2 (EC 4.2.1.3) (Citrate hydro-lyase 2)
156	ADHE_ECOLI	Aldehyde-alcohol dehydrogenase
157	SYV_ECOLI	Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).
158	GABT_ECOLI	4-aminobutyrate aminotransferase (EC 2.6.1.19)
159	IF3_ECOLI	Translation initiation factor IF-3.
160	RL30_ECOLI	50S ribosomal protein L30.
161	YDCH_ECOLI	Uncharacterized protein ydcH.
162	RL31_ECOLI	50S ribosomal protein L31.
163	IF1_ECOLI	Translation initiation factor IF-1.
164	YODD_ECOLI	Uncharacterized protein yodD.
165	GLRX3_ECOLI	Glutaredoxin-3 (Grx3).
166	YCIN_ECOLI	Protein yciN.
167	RS20_ECOLI	30S ribosomal protein S20.
168	RS17_ECOLI	30S ribosomal protein S17.
169	YIIU_ECOLI	Uncharacterized protein yiiU.
170	HDEA_ECOLI	Protein hdeA precursor (10K-S protein).
171	YHCO_ECOLI	Uncharacterized protein yhcO.
172	YGIN_ECOLI	Protein ygiN.
173	THIO_ECOLI	Thioredoxin-1 (Trx-1) (Trx).

Ref #	Protein	Description
174	BOLA_ECOLI	Protein bolA.
175	YDEI_ECOLI	Uncharacterized protein ydeI precursor.
176	PHNA_ECOLI	Protein phnA.
177	GLNB_ECOLI	Nitrogen regulatory protein P-II 1.
178	RL19_ECOLI	50S ribosomal protein L19.
179	YCFF_ECOLI	HIT-like protein ycfF.
180	YJBR_ECOLI	Uncharacterized protein yjbR.
181	RL16_ECOLI	50S ribosomal protein L16.
182	YHHA_ECOLI	Uncharacterized protein yhhA precursor (ORFQ).
183	YAEH_ECOLI	UPF0325 protein yaeH.
184	USPG_ECOLI	Universal stress protein G.
185	FUR_ECOLI	Ferric uptake regulation protein (Ferric uptake regulator).
186	YHBC_ECOLI	UPF0090 protein yhbC.
187	DKSA_ECOLI	DnaK suppressor protein.
188	BCP_ECOLI	Putative peroxiredoxin bcp (EC 1.11.1.15) (Thioredoxin reductase)
189	GREA_ECOLI	Transcription elongation factor greA
190	PPIA_ECOLI	Peptidyl-prolyl cis-trans isomerase A precursor (EC 5.2.1.8)
191	SSPB_ECOLI	Stringent starvation protein B.
192	MOAB_ECOLI	Molybdenum cofactor biosynthesis protein B.
193	LRP_ECOLI	Leucine-responsive regulatory protein.
194	SSB_ECOLI	Single-stranded DNA-binding protein (SSB)
195	HSLV_ECOLI	ATP-dependent protease hslV (EC 3.4.25.-) (Heat shock protein hslV).
196	YAJQ_ECOLI	UPF0234 protein yajQ.
197	ATPD_ECOLI	ATP synthase delta chain (EC 3.6.3.14).
198	APT_ECOLI	Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).
199	WCAF_ECOLI	Putative colanic acid biosynthesis acetyltransferase wcaF
200	YDJA_ECOLI	Protein ydjA.
201	RL5_ECOLI	50S ribosomal protein L5.
202	EFP_ECOLI	Elongation factor P (EF-P).
203	GMHA_ECOLI	Phosphoheptose isomerase (EC 5.3.1.-)
204	YDJY_ECOLI	Uncharacterized protein ydjY precursor.
205	GRPE_ECOLI	Protein grpE (HSP-70 cofactor) (Heat shock protein B25.3) (HSP24).
206	FKBB_ECOLI	FKBP-type 22 kDa peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)
207	RL4_ECOLI	50S ribosomal protein L4.
208	YCAC_ECOLI	Uncharacterized protein ycaC.
209	ENGB_ECOLI	Probable GTP-binding protein engB.
210	YLIJ_ECOLI	Uncharacterized GST-like protein yliJ.
211	TRAT1_ECOLI	TraT complement resistance protein precursor.
212	NARL_ECOLI	Nitrate/nitrite response regulator protein narL.
213	YCEH_ECOLI	Uncharacterized protein yceH (G20.3).
214	PYRH_ECOLI	Uridylate kinase (EC 2.7.4.-) (UK) (Uridine monophosphate kinase)
215	CPXR_ECOLI	Transcriptional regulatory protein cpxR.
216	RS2_ECOLI	30S ribosomal protein S2.
217	YGGE_ECOLI	Uncharacterized protein yggE.
218	HDHA_ECOLI	7-alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) (7-alpha-HSDH).
219	LPXA_ECOLI	Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase
220	PANB_ECOLI	3-methyl-2-oxobutanoate hydroxymethyltransferase (EC 2.1.2.11)
221	TRPA_ECOLI	Tryptophan synthase alpha chain (EC 4.2.1.20).

Ref #	Protein	Description
222	YEHZ_ECOLI	Uncharacterized protein yehZ precursor.
223	ALSB_ECOLI	D-allose-binding periplasmic protein precursor (ALBP).
224	NADE_ECOLI	NH(3)-dependent NAD(+) synthetase (EC 6.3.1.5)
225	THTM_ECOLI	3-mercaptopyruvate sulfurtransferase (EC 2.8.1.2)
226	DKGA_ECOLI	2,5-diketo-D-gluconic acid reductase A (EC 1.1.1.274) (2,5-DKG
227	AGAY_ECOLI	Tagatose-1,6-bisphosphate aldolase agaY (EC 4.1.2.-) (TBPA).
228	YJJW_ECOLI	Uncharacterized protein yjjW.
229	PANC_ECOLI	Pantoate--beta-alanine ligase (EC 6.3.2.1) (Pantothenate synthetase)
230	YBBN_ECOLI	Uncharacterized protein ybbN.
231	YTFQ_ECOLI	ABC transporter periplasmic-binding protein ytfQ precursor.
232	GLSA1_ECOLI	Glutaminase 1 (EC 3.5.1.2).
233	RSEB_ECOLI	Sigma-E factor regulatory protein rseB precursor.
234	ACCD_ECOLI	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
235	THIB_ECOLI	Thiamine-binding periplasmic protein precursor.
236	PTNAB_ECOLI	PTS system mannose-specific EIIAB component (EIIAB-Man)
237	ACCA_ECOLI	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha
238	QOR_ECOLI	Quinone oxidoreductase (EC 1.6.5.5) (NADPH:quinone reductase)
239	RPOA_ECOLI	DNA-directed RNA polymerase alpha chain (EC 2.7.7.6)
240	TDH_ECOLI	L-threonine 3-dehydrogenase (EC 1.1.1.103).
241	ALF1_ECOLI	Fructose-bisphosphate aldolase class 1 (EC 4.1.2.13)
242	GALM_ECOLI	Aldose 1-epimerase (EC 5.1.3.3) (Mutarotase).
243	GCST_ECOLI	Aminomethyltransferase (EC 2.1.2.10)
244	IADA_ECOLI	Isoaspartyl dipeptidase (EC 3.4.19.-).
245	YQHD_ECOLI	Alcohol dehydrogenase yqhD (EC 1.1.1.-).
246	FABB_ECOLI	3-oxoacyl-[acyl-carrier-protein] synthase 1 (EC 2.3.1.41)
247	RIR2_ECOLI	Ribonucleoside-diphosphate reductase 1 subunit beta (EC 1.17.4.1)
248	TOLB_ECOLI	Protein tolB precursor.
249	ASTC_ECOLI	Succinylornithine transaminase (EC 2.6.1.81)
250	ODO2_ECOLI	Dihydrolipoyllysine-residue succinyltransferase component
251	ISCS_ECOLI	Cysteine desulfurase (EC 2.8.1.7) (ThiI transpersulfidase)
252	DEGP_ECOLI	Protease do precursor (EC 3.4.21.-).
253	PURA_ECOLI	Adenylosuccinate synthetase (EC 6.3.4.4) (IMP--aspartate ligase)
254	GLMM_ECOLI	Phosphoglucosamine mutase (EC 5.4.2.10).
255	DHE4_ECOLI	NADP-specific glutamate dehydrogenase (EC 1.4.1.4) (NADP-GDH).
256	ACCC_ECOLI	Biotin carboxylase (EC 6.3.4.14) (Acetyl-CoA carboxylase subunit A)
257	KPYK1_ECOLI	Pyruvate kinase I (EC 2.7.1.40) (PK-1).
258	GABD_ECOLI	Succinate-semialdehyde dehydrogenase [NADP+] (EC 1.2.1.16) (SSDH).
259	GLNA_ECOLI	Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase).
260	IMDH_ECOLI	Inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205)
261	SYN_ECOLI	Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)
262	PEPD_ECOLI	Aminoacyl-histidine dipeptidase (EC 3.4.13.3) (Xaa-His dipeptidase)
263	SYE_ECOLI	Glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
264	NORR_ECOLI	Anaerobic nitric oxide reductase transcription regulator norR.
265	AHPF_ECOLI	Alkyl hydroperoxide reductase subunit F (EC 1.6.4.-)
266	PUR1_ECOLI	Amidophosphoribosyltransferase (EC 2.4.2.14)
267	SYK2_ECOLI	Lysyl-tRNA synthetase, heat inducible (EC 6.1.1.6)

Ref #	Protein	Description
268	USHA_ECOLI	Protein ushA precursor [Includes: UDP-sugar hydrolase (EC 3.6.1.45)
269	FUMA_ECOLI	Fumarate hydratase class I, aerobic (EC 4.2.1.2) (Fumarase).
270	YTFM_ECOLI	Uncharacterized protein ytfM precursor.
271	MAO1_ECOLI	NAD-dependent malic enzyme (EC 1.1.1.38) (NAD-ME).
272	SYQ_ECOLI	Glutaminyl-tRNA synthetase (EC 6.1.1.18) (Glutamine--tRNA ligase)
273	DHSA_ECOLI	Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1).
274	ODP2_ECOLI	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate
275	YEJA_ECOLI	Uncharacterized protein yejA precursor.
276	TKT2_ECOLI	Transketolase 2 (EC 2.2.1.1) (TK 2).
277	OPDA_ECOLI	Oligopeptidase A (EC 3.4.24.70).
278	CATA_ECOLI	Peroxidase/catalase HPI (EC 1.11.1.6) (Catalase-peroxidase)
279	TRAD1_ECOLI	Protein traD.
280	CLPB_ECOLI	Chaperone protein clpB (Heat-shock protein F84.1).
281	SYA_ECOLI	Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (AlaRS).
282	AMPN_ECOLI	Aminopeptidase N (EC 3.4.11.2) (Alpha-aminoacylpeptide hydrolase).
283	CAPP_ECOLI	Phosphoenolpyruvate carboxylase (EC 4.1.1.31) (PEPCase) (PEPC).
284	ODP1_ECOLI	Pyruvate dehydrogenase E1 component (EC 1.2.4.1).
285	SYI_ECOLI	Isoleucyl-tRNA synthetase (EC 6.1.1.5) (Isoleucine--tRNA ligase)
286	ODO1_ECOLI	2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)