Early Studies in Ultrahigh Pressure Liquid Chromatography of Intact Proteins

John W. Eschelbach

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
2006

Approved By

Advisor: Professor James W. Jorgenson

Reader: Professor Gary L. Glish

Reader: Professor R. Mark Wightman

ABSTRACT

John W. Eschelbach

EARLY STUDIES IN ULTRAHIGH PRESSURE LIQUID CHROMATOGRAPHY OF INTACT PROTEINS

(Under the direction of James W. Jorgenson)

The analysis of proteomic mixtures is quite complex due to the abundance of species to characterize. Most modern methods utilize a bottom-up approach, but as mass spectrometer methods improve, the need to high-resolution top-down methods have also been identified. The separation of intact proteins for top-down proteomics by reverse-phase liquid chromatography (RPLC) has received renewed interest due to the ease of coupling to ESI-MS. Unfortunately, modern RPLC methods do not have high enough resolving power to analyze complex proteomic mixtures with great success.

It has been well documented that the use of smaller diameter packing material in a chromatographic column can greatly increase the resolving power. These particles, which have a diameter of <2 μm, require a substantially higher backpressure to produce an equivalent flow. Ultrahigh pressure liquid chromatography (UHPLC) can produce pressures up to 100 kpsi and can be used with these smaller particles. Previous work has explored isocratic separations small organics and gradient separation of peptides. This work will investigate gradient UHPLC of intact proteins.

A custom gradient UHPLC system capable of up to 40 kpsi for use with proteomic samples is presented. This system was first used to investigate separations of an *E Coli* lysate on a column packed with 1.5 um porous particles. The results present some of the highest chromatographic peak capacities seen to date. Next, the carryover behavior of four standard proteins was explored. It is evident from the results that the use of pressures above 15 kpsi greatly improves the carryover and recovery of intact proteins, which is an unanticipated benefit of using ultrahigh pressures.

Reversed-phase methods show the greatest separation power, but also require the use of long gradient times and mobile phases that reduce the activity of the protein in its native form. Hydrodynamic chromatography (HDC) utilizes the parabolic flow profile in the interstitial spaces between particles in a packed column to drive separation of analytes based on their size in reverse order. Since there is no direct interaction with the stationery phase, non-porous silica (NPS) beads can be utilized, making this technique well-suited for ultrahigh pressures.

Based on the hydrodynamic radius of proteins in solution, particles in the size range of 0.2-0.6μm are required for proteins between 25 – 300 kDa. The preparation of sub-micron NPS packed capillary columns and performance of these columns applied to protein separations by HDC was also investigated as a potential alternative to RPLC methods.

Καλεπα τα Καλα

ACKNOWLEDGEMENTS

Many people have contributed to the success of this dissertation during my time at the University of North Carolina-Chapel Hill. I would like to thank Dr. James Jorgenson for the guidance and advice he provided during my time in his group. His ability to quickly grasp research ideas and provide unique insights proved invaluable. I will never forget the times that group members and I would debate ideas for hours, only to have "JJ" come to the correct (and only sometimes the same) conclusion during a short 15 minute conversation. Additional thanks to my readers, Prof. Gary Glish, and Prof. R. Mark Wightman and my committee members, Prof. Lee G. Pedersen and Prof. Kevin Weeks.

I would also like to thank several former and current group members. First, Dr. Keith Fadgen and his colleague, Dr. Geoff Gerhardt of Waters Corp., provided extensive support for the UHPLC project. Next, Dr. Jason Link did much of the initial UHPLC troubleshooting; the work that he and I completed formed the foundation for much of this thesis. Also, Dr. Matthew Monroe first mentored me when I joined the group, and I continuously used some of his data and programming skills. Dr. Will Thompson and Dr. Scott Mellors were also a great resource for discussion of limitless topics during our time in the group. Finally, thanks to current member Charles Evans for assistance with the UHPLC and MS project. Special thanks for Rachel Lieberman for assistance in collecting some of the final data for the HDC project.

In addition to the support provided by members of the lab, several others have also contributed to this works' success. Thanks to my parents for always encouraging me to pursue what I enjoyed and for providing me with the foundation and resources to do so. Finally, thank you to my beautiful wife, Katherine, for the love and never-ending support you provided during the ups-and-downs of dissertation research. You always believed in me, and that is something for which I will always be grateful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations

2D	two dimensional
Å	angstrom
ACN	acetonitrile
AU	absorbance units
AUFS	absorbance units full scale (detector gain factor)
avg	average
BEH	bridged-ethyl hybrid particle
BPI	base peak index chromatogram
BSA	bovine serum albumin
CapLC	Waters capillary LC
cm	centimeter
Da	Dalton
DAQ	data acquisition
DI	deionized (water)
ESI	electrospray ionizatioin
EtOH	ethanol
FTV	freeze-thaw valve
GST	gradient storage tubing
gUHPLC	gradient ultrahigh pressure liquid chromatography
HDC	hydrodynamic chromatography
HETP	height equivalent to a theoretical plate
HPLC	high performance liquid chromatography
Hz	hertz
i.d.	inner diameter
kDa	kiloDalton
L	liter
LC	liquid chromatography
M	molar
m/z	mass to charge ratio

mg	milligram
min	minute
mL	milliliter
mm	millimeter
mM	millimolar
MS	mass spectrometry/spectrum
MS/MS	tandem mass spectrometry
MW	molecular weight (mass)
Myo	Myoglobin
ng	nanogram
nL	nanoliter
nmol	nanomole
NPS	nonporous silica
o.d.	outer diameter
ODS	octadecylsilane (C18)
Ova	Ovalbumin
PBS	phosphate buffered saline
pg	picogram
pmol	picomole
PS	polystyrene
psi	pounds per square inch
RNaseA	Ribonuclease A
RPLC	reversed phase liquid chromatography
RT	retention time
s.p.	stationary phase
S/N	signal to noise ratio
SDS	sodium dodecyl sulfate
sec	second
SEC	size exclusion chromatography
SIC	selected ion chromatogram
TEOS	tetraethyl orthosilicate

THF	tetrahydrofuran
TIC	total ion chromatogram
TOF	time-of-flight
µL	microliter
µm	micrometer (micron)
µM	micromolar
UHMPWE	ultra-high molecular weight polyethylene
UHPLC	ultrahigh pressure liquid chromatography

LIST OF SYMBOLS

d_c	inner diameter of the column
C	rotational constant (HDC)
Δ	change
D_m	diffusion coefficient of the solute in the mobile phase
d_p	diameter of packing particles
F	volumetric flow rate
γ	interparticle tortuosity factor for longitudinal diffusion
h	height
H	plate height (height of an equivalent theoretical plate)
H_A	contribution to H from eddy diffusion
H_B	contribution to H from longitudinal diffusion
H_C	contribution to H from resistance to mass transfer
H_{det}	contribution to H from detector
H_{virt}	contribution to H from virtual column
k'	retention/capacity factor
L	length
M	ion mass
M	molecular mass
N	number of theoretical plates
n_c	peak capacity
P	pressure
Φ	flow resistance factor
r	radius
R_a	stokes radius
$R_{a,\,eff}$	effective stokes radius
R_c	effective capillary channel radius
R_{eff}	effective radius of polymer in solution
P	column aspect ratio
R_s	resolution

t	time
τ	retention ratio (HDC)
t_r	retention time (elution time)
u	linear velocity
u_{opt}	optimum linear velocity
V	potential
w_{avg}	average peak width (at 4σ)
w_{base}	peak width at the base of the peak (at 4σ)
ε	interparticle porosity
η	viscosity
λ	aspect ratio (HDC)
λ	scaling factor for eddy diffusion (van Deemter theory)
σ	standard deviation
σ^2	variance

CHAPTER 1
PROTEOMICS & ULTRAHIGH PRESSURE LIQUID CHROMATOGRAPHY

1.1 PROTEOMICS AND CHROMATOGRAPHY

The completion of the Human Genome Project in 2000 was a major scientific achievement that has led to an increased effort to characterize the human proteome. For this effort, commonly referred to as proteomics, the number and scale of problems to overcome are far greater than those of genomics. Solutions to these problems offer the promise of a more complete understanding of human biology and disease, which may lead to novel medicines and more rapid rapid drug development.

1.1.1 Proteomics Background

It is generally accepted that there are around 25,000 protein-coding genes identified by the human genome project. Estimates for the total number of unique proteins, however, vary greatly from 50,000 to 500,000 due to an incomplete understanding of how post-translational modifications play a role in protein expression.[1, 2] Even the most conservative of estimates poses a monumental challenge from a separation and identification perspective, especially due to the large sample sets often necessary for analysis. Further complicating the problem of sample complexity is the fact that the dynamic range required to detect proteins is ~10^6-10^9.[3] Thus, proteomics requires a technique that is high-resolution for protein separations, high-sensitivity for detection, and relatively high-speed for large sample sizes.

Two major approaches to proteomic analysis approaches have arisen. The more common approach, "bottom-up", first digests the intact proteins to their peptide fragments. The fragments are then separated and sequenced by mass spectrometry. A second, less common method is known as "top-down". With this method, the proteins are both separated and analyzed in the intact form.

Each approach has known advantages and disadvantages. Bottom-up has been historically preferred since analysis of peptide mixtures by mass spectrometry (MS) and tandem MS-MS are well-established, and numerous techniques exist to sequence and identify the proteins from peptide fragments. MS of intact proteins for top-down proteomics has gained interest, but is still more difficult than the MS of peptides. From a separations perspective, bottom-up is also easier to perform since chromatography of peptides is easier than the chromatography of intact protein.

The drawbacks to bottom-up, however, come out of these strengths. Since bottom-up methods require digestion, the sample will be more complex since there will always be more peptides than parent proteins. Additionally, the mixture information is "scrambled" and it can be difficult to connect a detected peptide to the parent protein. Alternatively, top-down approaches reduce the sample complexity since digestion is not needed and also maintain the original protein for analysis. This provides complementary information, such as post-translational modifications, that can be more difficult to observe by a pure bottom-up approach.[4-7] As mixture complexity increases, top-down approaches may become advantageous, since fewer components exist to separate and analyze.

1.1.2 Need for Improved Chromatographic Methods

The preferred method for proteomic separations has historically been slab gel-electrophoresis. While this method has long been shown to have extremely high resolving

power for proteins,[8] it is also tedious to perform as sample automation and interfacing with mass spectrometry are difficult. As an alternative, high resolution chromatographic methods for proteomics are especially desirable due to the ease of interface with electrospray ionization (ESI) methods for mass spectrometry.[9, 10] Additionally, chromatographic methods are generally fast and easily automated for high throughput.

While much of the recent work in proteomics has focused on improving identification of proteins by various mass spectrometry methods, chromatographic methods have played a critical role in protein and peptide analyses as well.[10, 11] Of these methods, reversed-phase liquid chromatography (RPLC) has been the most widely used[12] and continues to play an important role. RPLC utilizes a hydrophobic stationary phase coated on the packing material in a chromatographic column. As the proteins and peptides elute, the degree of interaction with the stationary phase controls the relative separation.

Although RPLC is widely used, it is rarely reported with a resolution high enough to work within the realm of unpurified complex biological samples.[13] Additionally, RPLC of intact proteins has been historically problematic due to problems with peak shape, protein carryover or "ghosting", and resolution.[14-19] In order to utilize RPLC for top-down proteomic methods, these limitations will need to be overcome, but RPLC offers tremendous potential.

1.2 ULTRAHIGH PRESSURE LIQUID CHROMATOGRAPHY

One method for improving the resolving power of a column has been the use of smaller particles operated at elevated pressures, so-called ultrahigh pressure. The use of ultrahigh pressure for liquid chromatographic separations was pioneered in our laboratory by MacNair during the late 1990's.[20, 21] His work showed the potential advantages for using columns

packed with 1-μm particles at ultrahigh pressures. Many in our lab have since followed exploring various aspects of UHPLC from fundamental[22-24], engineering[25, 26], and applied[27, 28] points of view. While UHPLC theory has been presented in depth, a brief review follows along with an extension to proteomics.

1.2.1 Motivations for Utilizing Ultrahigh Pressures

The advantages for UHPLC can be seen from basic chromatographic theory for column performance. The column performance is governed by the Height Equivalent to a Theoretical Plate (HETP), *H*, and can be written as:

$$H = H_A + H_B + H_C = A + \frac{B}{u} + Cu \qquad \textbf{(1.1)}$$

where *u* is the linear velocity of the column. *A*-, *B*-, and *C*-terms are constants used to describe contributions to *H* from multiple flow paths in the column (so called Eddie Diffusion), longitudinal diffusion and resistance to mass transfer, respectively.[29]

The *A*-term, can be further defined as function of particle diameter, d_p, by:

$$H_A = \lambda d_p \qquad \textbf{(1.2)}$$

The constant λ is a geometrical factor that scales the impact of particle diameter on the eddy diffusion term, and is approximated as 0.5.[29]

Next, the B-term can be written as:

$$H_B = 2\gamma \frac{D_m}{u} \qquad \textbf{(1.3)}$$

where D_m is the diffusion coefficient in of the analyte in the mobile phase. The constant γ, the interparticle tortuosity factor, describes the obstruction of the packing material to free diffusion of analytes. [29] It is generally estimated as 0.5 for packed capillary columns.

Finally, the C-term is defined by:

$$H_C = c\frac{d_p^2}{D_m}u \qquad \textbf{(1.4)}$$

where *c* represents the many factors that affect mass transfer in a chromatographic column, and can be estimated as 0.2 for packed capillary columns.

Combining the preceding equations, the van Deemter equation can be simplified to:

$$H = \frac{d_p}{2} + \frac{D_m}{u} + \frac{d_p^2}{5D_m}u \qquad \textbf{(1.5)}$$

An aspect of column performance as related to UHPLC is observed in Eq. 1.5. This shows that *H* is dependent on particle diameter for a given column and thus smaller particles lead to decreased plate heights, indicating higher column efficiency. This is further illustrated by Figure 1-1 which shows plots of Eq. 1.5 for decreasing values of d_p.

While *H* describes the performance of a given column, it does not indicate the overall column efficiency or separation power. Instead, the concept of theoretical plates, *N*, is used. This can be related to *H* by:

$$N = \frac{L}{H} \qquad \textbf{(1.6)}$$

where L is the column length. In general, the greater the number of theoretical plates a column produces, the higher the quality of separation. A combination of this Eq. with Eq. 1.5 indicates that *N* is inversely related to d_p. Therefore, the use of smaller diameter packing material leads to an increased number of theoretical plates and ultimately a high quality separation.

Unfortunately, the improvement in *N* does not come without a penalty, which is the increased backpressure of the column required to drive the separation. This can be seen from the Kozeny-Carman equation which describes the pressure drop, ΔP, across a packed bed.[29]

$$\Delta P = \frac{180\eta L u (1-\varepsilon)^2}{d_p^2 \varepsilon^2} \quad \textbf{(1.7)}$$

where η is solvent viscosity, and ε is interparticle porosity. This indicates that at a fixed u, the pressure required is inversely related to d_p^2, meaning that the pressure requirements increase at much faster rate than the improvement in separation. Further, the minimum of the van Deemter plot (H_{min}) in Figure 1-1 can be seen to not only decrease, but also shift to a higher u as d_p decreases. This point on the plot is commonly referred to as u_{opt} and can be related to pressure by:[20]

$$\Delta P_{opt} \propto \frac{1}{d_p^3} \quad \textbf{(1.8)}$$

indicating that pressure requirements for running small particles at H_{min} is even greater. Simply decreasing d_p 5-fold predicts a similar increase in N, but a 125-fold increase in the backpressure needed. Not surprisingly, the mechanical limitation of commercial LC equipment is quickly reached when using smaller d_p.

Commercial pumps have traditionally been limited to <6 kpsi. This allowed use of particles ~3-5 μm, and limited use of < 3 μm particles on shortened columns. In order to utilize smaller particles on sufficiently long columns to increase N, custom methods are needed to handle the ultrahigh pressure requirements. The primary reason for exploring ultrahigh pressure methods, therefore, has been to obtain improvements in separation quality by using particles < 2 μm.

1.2.2 Application to Proteomics

The complexity of the samples common to proteomic applications require high-resolution techniques. Traditional RPLC quickly reaches a point where the sample complexity overwhelms the separation and little can be done to improve it. UHPLC, however, is well-

suited for these complex separations because of the higher separation efficiency inherent to the method. Additionally, as outlined above, RPLC methods are easily interfaced to mass spectrometry for detection. Since UHPLC does not affect this interface, it is a natural fit to apply it to proteomic methods.

The theory presented in the previous section was developed for isocratic separation of small molecules. While the basic reasoning is unchanged, in that smaller particles lead to improved separations, the nature of proteins require a few additional considerations. First, proteins are much larger in size than small molecules and thus have a smaller diffusion coefficient, D_m. While a small organic molecule may exhibit a $D_m \approx 1\text{x}10^{-5}$ cm^2/sec, proteins can easily have a D_m an order of magnitude lower, $\sim 1\text{x}10^{-6}$. From Eq. 1.5, it can be seen that H is dependent on D_m as well as d_p. The result of this dependence is shown in Figure 1-2 for a fixed d_p. As D_m decreases, the contribution from the B-term is reduced, but the contribution from the C-term is increased due to poor mass transfer caused by the slow rate of diffusion. The net effect is that H_{min} is unchanged, but u_{opt} decreases by the decrease in D_m. Theoretically, this means that longer columns or smaller particles could be utilized because less pressure is needed to reach u_{opt}.

The second important consideration when applying RPLC to proteomics is the less ideal retention behavior of the protein. The retention time (t_r) for a given analyte can be defined as

$$t_r = \frac{L}{u}(1 + k') \qquad \textbf{(1.9)}$$

where k' is the retention factor, and indicates how strongly an analyte is retained in the stationary phase (s.p.). Higher values indicate greater preference for the s.p., while lower values indicate the opposite as the analyte spends more time in the mobile phase and thus migrates faster. In general, the k' of small organic molecules is relatively insensitive to small

changes in organic concentration. Proteins, however, have a much greater sensitivy and tend to exhibit more of an "on/off" retention behavior when changing the mobile phase.[29] As such, isocratic methods for protein separations are rather difficult and gradient methods are instead used.

The concept of theoretical plates, *N*, is only applicable to isocratic separations. For gradient separations, the concept of peak capacity, n_c, is instead used to evaluate the quality of separation. The peak capacity of a separation is defined by the theoretical maximum number of chromatographic peaks that could fit into a given time window and can be written as:

$$n_c = \frac{t_2 - t_1}{w_b} \qquad \textbf{(1.10)}$$

where t_2 and t_1 are the limits of the elution time window and w_b is the widths of the peak at base. This indicates that peak capacity is increased by either expanding the elution window, which is achieved by increasing the gradient time, or by decreasing the peak width. Since UHPLC methods result in reduced peak widths, the use of smaller particles should also increase n_c, making the method applicable to proteomic separations as well.

1.3 COLUMN OPTIMIZATION CONSIDERATIONS

Ultrahigh pressure LC methods allow for the use of smaller packing material which enhances overall separation performance. Examining the background theory presented in 1.2.1 shows that $N_{\max} \propto L/d_p$, indicating that plate count could grow infinitely by using ever longer columns and smaller particles. Of course, the pressure required would grow infinitely as well leading to a rather a rather mundane conclusion since pressure is limited, even in the case of UHPLC. Jorgenson & Guthrie developed a theory for a similar scenario with open

tubular capillaries.[30] By accepting that a pressure limit existed, it was shown that an optimum dimension could be calculated, and that beyond a certain point, smaller diameters decreased column performance. This approach can be expanded to packed capillaries and is useful for estimating optimal column parameters within UHP limits.

A thorough derivation of optimal *N* is presented in Appendix A. For the case of fixed pressures or fixed dead times, *N* can be written as:

$$N(P,d_p)=\frac{\alpha P d_p^2}{\beta P^{1/2} d_p^2+\chi+P d_p^4}\ \textbf{(a)} \qquad N(t,d_p)=\frac{\alpha' t d_p^2}{\beta' t^{1/2} d_p^2+\chi' t+\rho d_p^4}\ \textbf{(b)} \quad \textbf{(1.11)}$$

where N is now a function of d_p and either *P* (a) or *t* (b). The constants α, β, χ, ρ are the various column parameters, as defined in Appendix A. It is then possible to create a multidimensional plot from Eq. 1.11 to observe trends in *N* for varying parameters.

1.3.1 Fixed Dead Time

A plot of *N* vs. P and d_p is shown in Figure 1-3 for small organic molecules (A, D_m= $1x10^{-5}$) and proteins (B, D_m= $1x10^{-6}$) at a fixed dead time of 4 mins. Additionally, the corresponding column length (from Eq. A.4b) for a given *P* and d_p is also overlaid. The dead time along with the D_m, and η define the constants (α, β, χ) in Eq. 1.11a. From this plot, it is immediately evident that, analagous to observations with open tubular columns, an optimal particle diameter exists for any given pressure. This optimal d_p is shown to continually decrease as pressure continues to rise. While it is clear that higher pressures will always allow for better results, it is also apparent that smaller particles eventually have a negative effect if the pressure is fixed.

This somewhat counter-intuitive result can be explained by thinking of the plots as inverted van Deemter curves. When pressure is fixed, and d_p decreased, the column length must also be shortened in order to maintain the same dead time. If the optimum linear

velocity (u_{opt}) for the smaller d_p could be reached, then N would essentially be unchanged since the two decreases would offset each other. However, since backpressure has a greater dependence on d_p than L (Eq. 1.7), the linear velocity in the column must also be decreased since pressure and dead time are fixed. From Figure 1-1, it can be seen that, even for small particles, performance begins to decrease as the linear velocity becomes too slow and longitudinal diffusion (B-term) dominates.

A comparison of the effect of D_m for small organics and proteins on the optimal N is also shown by Figure 1-3a & b. Not surprisingly, a similar trend to that seen in Figure 1-2 is evident as the optimal d_p shifts to larger values for smaller molecules with higher D_m. This shift is again related to the B- and C-terms. The higher D_m increases the B-term effect, meaning that small decreases in u will reduce performance more so than with a slower D_m. The C-term, however, has been reduced shifting u_{opt} to a higher value.

Finally, it is also interesting to note that for plate counts below the optimum, multiple values of d_p, P and L give similar results. For instance, if only 40,000 plates were desired for small organics, Figure 1-3a shows that a 75-cm column packed with 5μm particles operated at 50 kpsi would give similar results to a 25-cm column packed with 3 μm particles run at 15 kpsi. Of course, numerous other combinations exist on the 40,000 plate "band". This potential versatility further illustrates the advantage of UHPLC as it is not always possible to obtain the exactly desired column dimensions. An UHP pump could allow 5-μm particles, which have been historically well characterized, on long columns to give similar performance to 1.5 μm particles, a rather new packing technology.

1.3.2 Fixed Pressure

The analysis of N_{opt} in the previous section assumed that dead time was fixed and pressure was variable. Since it is clear that higher pressures lead to better results, it is useful to

instead fix P and the maximum available pressure with current pump technology, and instead vary dead time since this is a variable that is easily modified. Eq. 1.11b can be used to plot N vs t and d_p at a fixed pressure. This is shown for small organic molecules (A, D_m= 1x10^{-5}) and proteins (B, D_m= 1x10^{-6}) at a fixed pressure of 30 kpsi in Figure 1-4. Similar to the results from a fixed pressure, an optimum d_p is again evident for each dead time and longer dead times lead to higher values of N. A difference, however, is that as dead time increases, N_{opt} shifts slightly to higher values of d_p. This is the result of slower linear velocities caused by longer dead times and columns reducing the C-term. Finally, the different values of D_m in Figure 1-4a & b produce similar results to the fixed dead time plots with the optimum for small molecules occurring at a higher d_p than that of proteins.

The main difference between a fixed pressure and dead time is that increases in pressure causes u to increase and the B-term to be reduced, while longer dead times decrease u and reduce the C-term. It should also be noted that the above theory was developed for isocratic separations and can not be directly applied to gradient elution. Since protein separations are only usefully achieved by gradient chromatography, the results presented can only be viewed as a general guide. Nonetheless, this comparison shows that smaller molecules are actually better suited for 2-3 μm particles at ultrahigh pressures while proteins and peptides will perform better on ~1 μm particles.

1.4 SUMMARY

The complexity of samples in proteomic work requires fast, high resolutions methods. Traditional proteomic methods have been shown to be high-resolution, but are typically slow and cumbersome. Most methods are based on bottom-up approaches, but top-down methods appear to offer advantages due to the reduction of overall sample complexity. Additionally,

the desire to analyze samples with information rich techniques (such as mass spectrometry) requires a convenient interface. The optimum particle diameter for molecules with high diffusion coefficients, such as proteins, also requires ultrahigh pressures. The advantages of UHPLC techniques are therefore a good fit and possible alternative to traditional methods. Utilizing smaller particles at ultrahigh pressures enhances the resolution, while LC methods are easily interfaced to MS for analysis.

While our lab has extensive experience applying UHPLC to small organic molecules and peptides, we have yet to fully explore UHPLC and intact proteins for top-down proteomic methods. Of particular interest is the potential improvement in separation efficiency over that of conventional particles. The work presented in this thesis will explore various aspects of UHPLC and proteins.

1.5 REFERENCES

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

(2) Hellmich, W.; Pelargus, C.; Leffhalm, K.; Ros, A.; Anselmetti, D. *Electrophoresis* **2005**, *26*, 3689-3696.

(3) Lion, N.; Rohner, T. C.; Dayon, L.; Arnaud, I. L.; Damoc, E.; Youhnovski, N.; Wu, Z.-y.; Roussel, C.; Josserand, J.; Jensen, H.; Rossier, J. S.; Przybylski, M.; Girault, H. H. *Electrophoresis* **2003**, *24*, 3533-3562.

(4) Kettman, J. R.; Frey, J. R.; Lefkovits, I. *Biomolecular Engineering* **2001**, *18*, 207-212.

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

(6) Nemeth-Cawley Jennifer, F.; Tangarone Bruce, S.; Rouse Jason, C. *Journal of proteome research* **2003**, *2*, 495-505.

(7) Cooper, J. W.; Wang, Y.; Lee, C. S. *Electrophoresis* **2004**, *25*, 3913-3926.

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

(9) Shi, Y.; Xiang, R.; Horvath, C.; Wilkins, J. A. *Journal of Chromatography, A* **2004**, *1053*, 27-36.

(10) Shen, Y.; Smith, R. D. *Electrophoresis* **2002**, *23*, 3106-3124.

(11) Neverova, I.; Van Eyk, J. E. *Journal of Chromatography, B: Analytical Technologies in the Biomedical and Life Sciences* **2005**, *815*, 51-63.

(12) Chen, H.; Horvath, C. *Journal of Chromatography, A* **1995**, *705*, 3-20.

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

(14) Kastner, M., Ed. *Protein liquid chromatography*; Elsevier: Amsterdam ; New York, 2000.

(15) Cohen, S. A.; Benedek, K. P.; Dong, S.; Tapuhi, Y.; Karger, B. L. *Analytical Chemistry* **1984**, *56*, 217-221.

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

(17) Nice, E. C.; Lloyd, C. J.; Burgess, A. W. *Journal of Chromatography* **1984**, *296*, 153-170.

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

(19) Burton, W. G.; Nugent, K. D.; Slattery, T. K.; Summers, B. R.; Snyder, L. R. *Journal of chromatography* **1988**, *443*, 363-379.

(20) MacNair, J. E. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 1998.

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

(22) Jerkovich, A. D., University of North Carolina, Chapel Hill, NC, 2003.

(23) Mellors, J. S. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2005.

(24) Thompson, J. W., University of North Carolina, Chapel Hill, NC, 2006.

(25) Link, J. C. Doctoral Dissertation, Univ. of North Carolina, Chapel Hill, NC, 2004.

(26) Patel, K. D. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2001.

(27) Monroe, M., University of North Carolina, Chapel Hill, NC, 2002.

(28) Sousa, J. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2004.

(29) Neue, U. D. *HPLC columns : theory, technology, and practice*; Wiley-VCH: New York, 1997.

(30) Jorgenson, J. W.; Guthrie, E. J. *Journal of Chromatography* **1983**, *255*, 335-348.

1.6 FIGURES

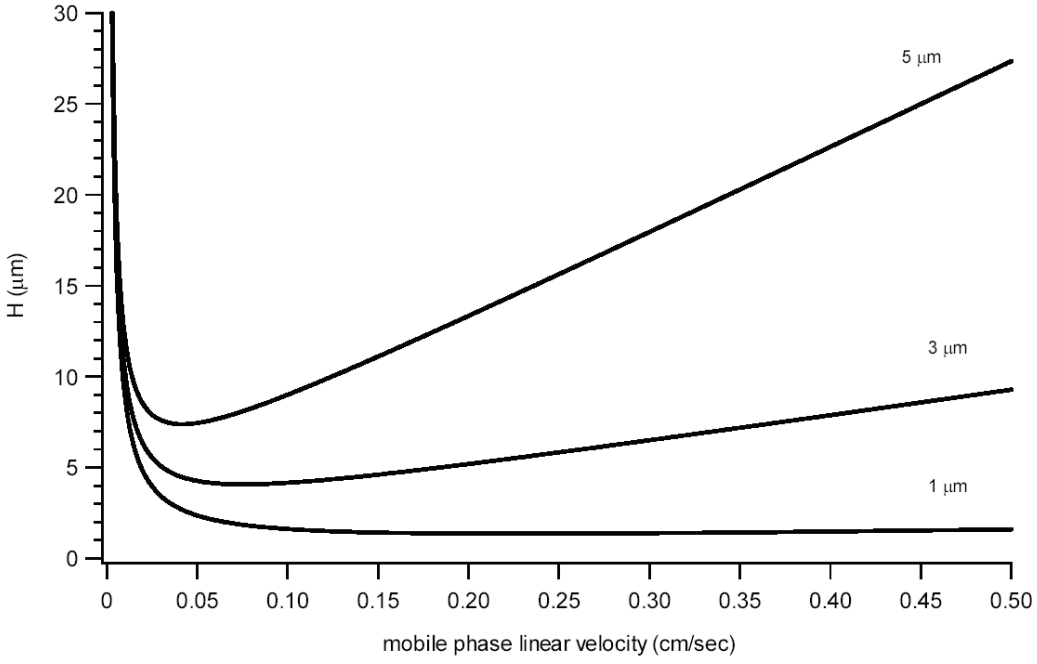


Figure 1-1: Theoretical van Deemter plots for 5-, 3- and 1-μm diameter particles used in column packing.

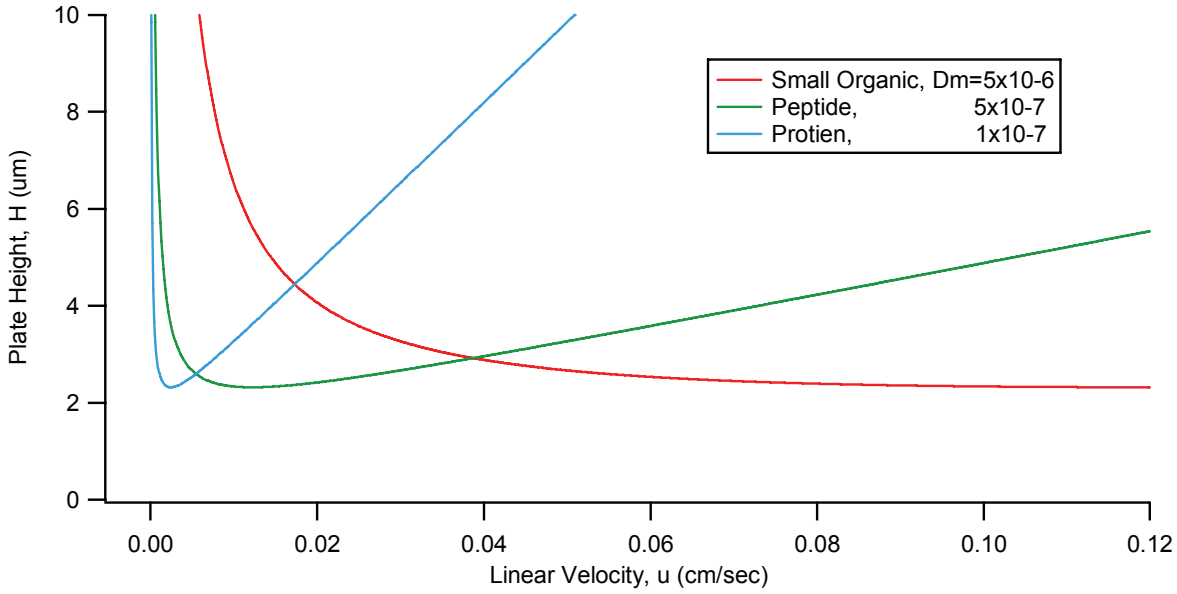


Figure 1-2: Effect of diffusion coefficient (D_m) on the van Deemter equation at a fixed d_p. H_{min} is unchanged, but u_{opt} is continually decreasing.

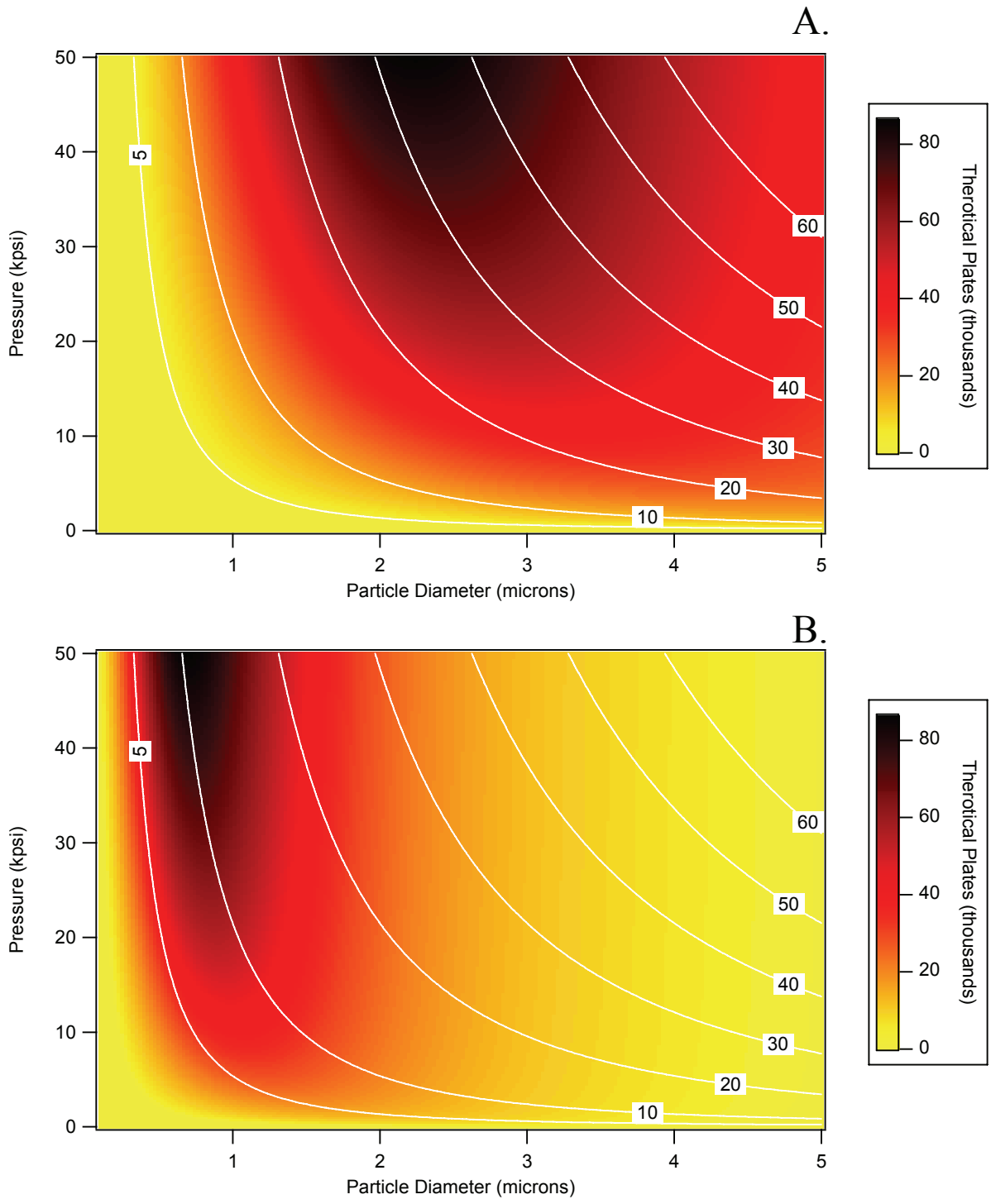


Figure 1-3: Plate optimization plots at fixed time for A) small organic (D_m=1e-5) and B) protein (D_m=1e-6). Overlays are isolength lines, in cm, from Eq A.4b. t=240 sec.

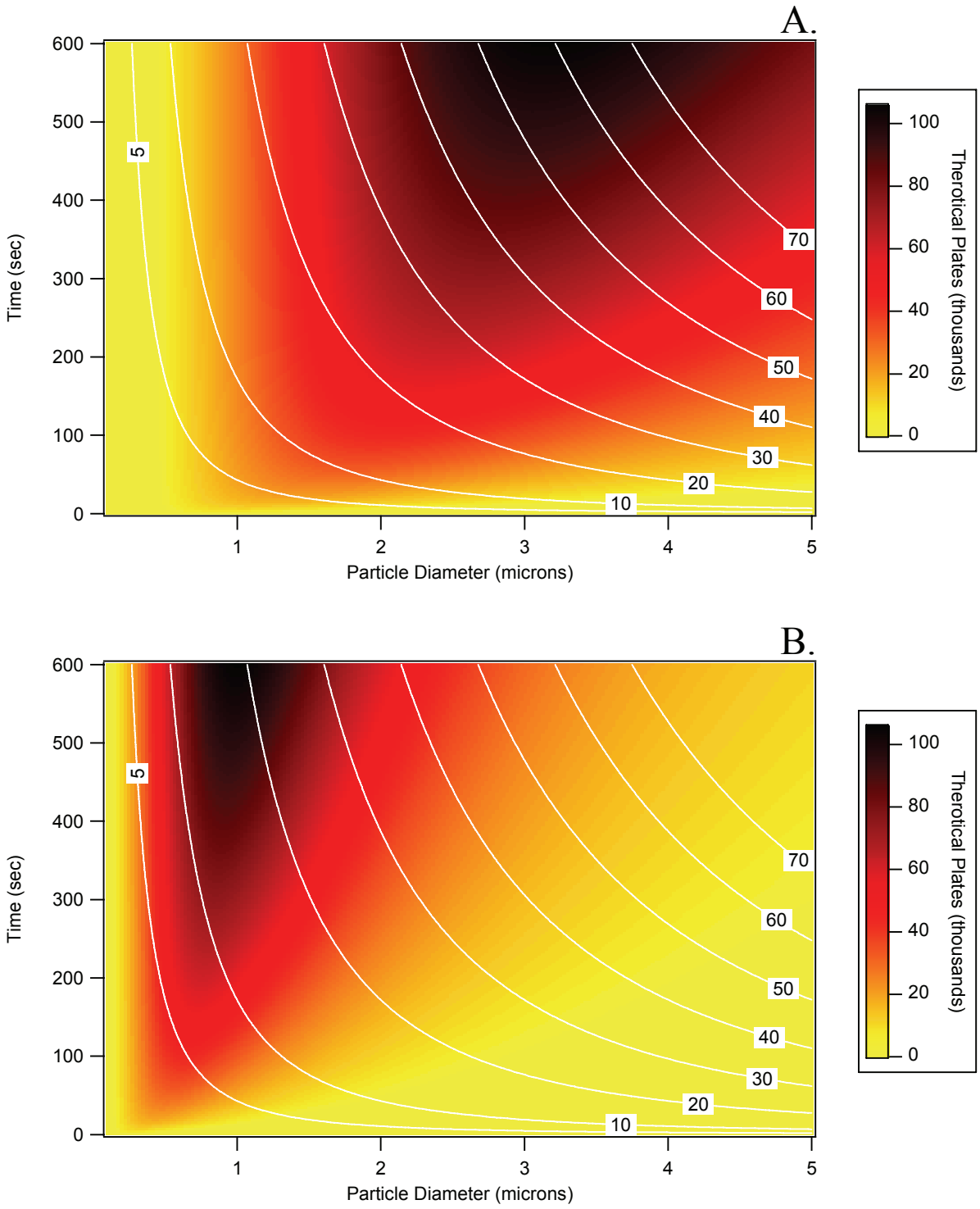


Figure 1-4: Plate optimization plots at fixed pressure for A) small organic (D_m=1e-5) and B) protein (D_m=1e-6). Overlays are isolength lines, in cm, from Eq A.4b. P=30 kpsi. Note change in Z-scale from Figure 1-3.

CHAPTER 2
GRADIENT ULTRAHIGH PRESSURE RPLC FOR THE SEPARATION OF INTACT PROTEINS

2.1 EVOLUTION OF GRADIENT UHPLC

In the previous chapter, the benefits of using small particles and gradient UHPLC for proteomics work were introduced. It was also noted that the pressures required for UHPLC far exceeds current commercial HPLC pump technology and that custom instrumentation is needed. In this chapter, the preliminary application of UHPLC to top-down proteomic separations is explored. A novel instrument for UHPLC using a hybrid of commercial and custom parts is also introduced.

Early Jorgenson Lab work in UHPLC focused on isocratic methods, however, gradient instrumentation has also been through several generations of development. The fundamental different in instrumentation is that isocratic UHPLC can utilize constant-pressure pumps while gradient UHPLC requires constant-flow pumps due to the changing mobile phase composition during the gradient. While ultrahigh pressure technology for constant-pressure pumps existed for other industrial applications and was adapted for UHPLC, constant-flow pumps were simply not available in any form and had to be designed from scratch. A brief review of the evolution from a single-pump exponential dilution gradient to the current commercial hybrid design follows.

2.1.1 Single-Syringe Exponential Gradient UHPLC

The first use of gradient UHPLC (gUHPLC) in our lab and the literature was demonstrated by MacNair & Patel.[1, 2] This system, being mostly a proof-of-concept for gUHPLC, consisted of only a single constant-flow syringe pump capable of generating ultrahigh pressures. Since one pump was initially built, only an exponential dilution gradient could be generated. By using only the first half of the gradient dilution time, a pseudo-linear gradient would result and peptide separations could be performed. This system was used with protein digests with LIF detection, and was capable of producing separations with peak capacities ~250-500 in under 2 hrs.

There were several disadvantages to this system noted by MacNair. First, the lack of a true linear gradient lead to greater band broadening towards the end of the exponential dilution. Next, there was no straightforward way to re-equilibrate the column after the gradient was complete since the dilution vessel had to be manually opened and filled. Finally, it was difficult to perform gradients over a wide mobile phase range due to the need to operate in the pseudo-linear region of the dilution gradient.

2.1.2 Dual Syringe Linear Gradient UHPLC

To overcome the problems encountered with the single-syringe configuration, Patel constructured an identical second syringe UHP pump.[3] This configuration had two independent UHP pumps that could be used to generate true linear gradients by controlling the flow rate of each pump relative to the other. Patel initially used this system with LIF detection for protein digest separations producing similar results to MacNair. Monroe later followed this work by integration of a time-of-flight mass spectrometer (TOF-MS) to gain advantages in sensitivity and informatics.[4] Finally, Link demonstrated the advantages of

porous particles to gUHPLC separation of peptides due to the increased ability load sample onto the column.[5]

While the system proved successful for generating linear gradients at ultrahigh pressures, several disadvantages were noted over the lifetime of the instrument. Since the instrument was completely lab-built, all maintenance had to be performed by the lab as well. This led to increased downtime of the instrument as things as straight forward as UHPLC seals were machined by hand and often were inconsistent in performance. The final instrument, shown for scale in Figure 2-1a, was also extremely large and cumbersome to work on. Therefore, a system composed of commercialized parts that could be replaced as needed was desired. A smaller footprint and ease of access to the pump was also needed.

2.1.3 Waters Hybrid UHPLC Prototype

Collaboration with Waters Corporation (Milford, MA) brought about the basics for the current gUHPLC. The initial design proposed by Waters Corp. was a more compact version of Patel's design that consisted of a dual-syringe setup capable of ultrahigh pressures. The design utilized commercially available seals and was significantly smaller or more modular, making maintenance less cumbersome. Link initially evaluated this system (designated UHPLC System 1) and, unfortunately, found it to be too unreliable at generating UHP gradients.[5] The initial design also did not incorporate an auto-sampler, which had been another drawback of earlier systems, since proteomic work often involved the need for high sample throughput.

As a solution, Link and our collaborators at Waters developed a preloaded gUHPLC system. This version used the syringe pump from the earlier Waters prototype, but incorporated an autosampler and second LC pump to preload the gradient and sample. This final design, shown in Figure 2-1b for comparison, represents the current state of our

gUHPLC and is utilized throughout this research. A detailed explanation of the system follows.

2.2 GRADIENT UHPLC INSTRUMENTATION FOR PROTEINS

The various advancements of gradient UHPLC have allowed for the move from a completely custom-built and maintained system to a hybrid of commercial equipment coupled to custom pieces. The final Waters preloaded gUHPLC system was the focus of a great deal of research by Link.[5] His work explored not only applications, but a thorough characterization of important figures of merit for the system. The reader is referred to his work for more specific characterization details. The system and a summary of its performance is discussed in this section as it is a central component to gUHPLC methods. Additionally, since Link was only able to perform initial studies on the system, several notes about usage of the system over a longer time frame are also provided.

2.2.1 Preloaded Gradient System Components

The gradient UHPLC is composed of both commercially available and custom components. An overall schematic is shown in Figure 2-2. The system can be broken down into three individual pieces: hydraulic amplifier, gradient capillary LC (CapLC), and high pressure union & valves. Each piece plays a specific role in the overall gUHPLC.

2.2.1.1 Hydraulic Amplifier

Commercially available pumps with pressure capability over 15 kpsi are currently not available, and any gUHPLC will require some form of customization. It is desirable to use as many commercial components as possible to minimize development of the pump itself. Therefore, a commercial Waters 1525 Binary Gradient pump (Waters Corp., Milford, MA) was modified to pump motor vehicle brake fluid (Castrol N.A., Wayne, NJ), which serves as

an hydraulic fluid in this application. The 1525 pump is connected to a custom-built syringe-driven hydraulic amplifier with a 30:1 amplification, allowing for the generation of ultrahigh pressures from the nominal pressures of the Waters 1525 pump. A diagram of the hydraulic amplifier is shown in Figure 2-3. It should be noted from this diagram that while the hydraulic fluid is doing the work, it is physically isolated from the mobile phase solvents. Early versions of the amplifier used differing solvents in pistons A & B for real-time gradients, however, this preloaded configuration uses only water in the piston heads.

Commercially available seals capable of holding ultrahigh pressures were obtained from Bal Seal Engineering (Foothills Ranch, CA) and were made out of ultra-high molecular weight polyethylene (UHMWPE). The seals were modified slightly with the addition of a #13 neoprene o-ring (McMaster-Carr, Atlanta GA) on the outside of the seal. This o-ring served as a static seal and increased the pressure capabilities to 40 kpsi. The final seal design was found to have a long lifetime, with the o-ring static seal being the main component to fail. Over three years of consistent use, each piston head was only rebuilt once and the problem was traced to catastrophic o-ring failure. This long lifespan is likely due to the fact that the seal only experiences DI water and is never exposed to a harsher organic solvent. O-ring failure is likely due to absorption of water which ultimately softens the seal. Lifetime of the o-ring could possibly be improved with the use of a different material, but this was not explored.

Several different check valve designs were initially explored by Link. A specially fabricated ball & seat cartridge design by Waters Corp. was eventually settled upon. Earlier designs suffered from poor reproducibility in low pressure sealing and were prone to clogs. This design was improved, but problems still arose with low-pressure sealing. Careful pump

operation normally alleviates the problem, but evaluation of newer designs is ongoing. Finally, all stainless tubing used in this system was brazed with silver solder before use since Waters fittings are only capable of 6 kpsi. As noted by Link, we have yet to see a brazed fitting fail at our operating pressures.

2.2.1.2 Capillary LC

The second major component to this system is a commercial Capillary LC (Waters Corp.). This component is unmodified and serves two main purposes. First, by using a commercial gradient system the gradient generation was found to be significantly more reproducible. Second, the CapLC incorporates an auto sampler, which was a significant limitation of earlier designs. Work with proteins typically involves numerous samples, and some form of run-to-run automation is required. Overall, few problems were encountered with the CapLC and autosampler.

2.2.1.3 High-Pressure Union & Valving

The hydraulic amplifier and CapLC are coupled via a high pressure 4-port union (custom made, Waters Corp.). This union contains a 400 μm through hole to which various capillary and stainless steel tubing connections can be made. The outlet of the hydraulic amplifier is connected to the inlet of the union via 6 m of 0.020" i.d. (~1.5ml) stainless steel gradient storage tubing (GST). This large dead volume of tubing serves to hold the preloaded gradient from the CapLC before ultrahigh pressures are applied.

An open-tubular 120-cm x 10 μm i.d. splitter capillary, packed chromatographic capillary and gradient inlet capillary from the CapLC are attached to the remaining ports on the union to create a closed system capable of ultrahigh pressures. A picture and internal layout of the 4-port union are shown by Figure 2-4a & b. The column was positioned ~17 mm in front of the splitter outlet to create a narrow injection plug. Pressure at the head of the column is

controlled by the length and inner diameter of the splitter capillary and the volumetric flow rate of the amplifier syringe pump, typically 4 µl/min.

A novel freeze-thaw valving design has been implemented to reduce the dead volumes that are present in the system. A freeze-thaw valve (FTV) utilizes liquid CO_2 to freeze a small volume of liquid inside a capillary column.[6] Once frozen, this plug is capable of withstanding UHPLC pressures in narrow i.d. (< 50 µm) capillary columns, essentially creating a "closed" valve. By heating the capillary, the plug can be thawed and the valve "opened". The valve, shown in Figure 2-5, consists of a capillary sandwiched between two copper plates to which CO_2 is applied. A resistive thermofoil heater is also integrated into the design in order to open the valve by thawing the frozen plug. Since the capillary is acting as the valve, only the width of the copper plates, typically 2 cm, introduce dead volume. For a 30 µm i.d. capillary, ~14 nl of dead volume would be introduced, which is considered minimal for µl flow rates.

This configuration has also proven to be quite robust. Few leaks are introduced by the various ports, with the typical source being the GST connections. The FTV are even more reliable as a single capillary has undergone hundreds of F-T cycles with no breakage. Additionally, since no mechanical wear takes place as in a typical valve, the FTV never develops leaks.

2.2.2 Complete System Operation

A standard HPLC generates a gradient in real time by mixing various proportions of the solvents to form a gradient. Link evaluated an early prototype with such a system, but found it to be too unreliable for routine work. The solution to the problem was to use a preloaded gradient from the CapLC. This is a non-traditional method for generating gradients, as it

adds to the overall run time and can lead to gradient broadening in larger volume connection tubing, but works well for UHPLC methods.

2.2.2.1 Procedure

The gradient for the system is generated in advance of the run and pre-loaded onto the gradient storage tubing of the hydraulic amplifier using the CapLC. Two FTV are used to isolate the low pressure components from the ultrahigh pressure side of the pump. Before the run, valves A & B (Figure 2-6a) are opened and the gradient is loaded in reverse onto the storage tubing. Gradients were loaded at 40 µl/min from the CapLC. Next, the autosampler is used to load the sample and push it onto the storage tubing. Once gradient and sample loading are complete, the two valves are closed and the amplifier can be used to push both the sample and gradient into the 4-port union (Figure 2-6b). The flow rate from the amplifier was 4 µl/min, or 10x slower than the loading flow rate. This provided a convenient way to calculate the gradient preloading as 1 min of load time was equivalent to 10 mins of analysis time. An illustration showing the relative positions of the amplifier pump, GST, gradient, sample and column is shown by Figure 2-7. A sample plug is, therefore, injected onto the head of the column, followed by a linear gradient run at the ultrahigh pressure of the hydraulic amplifier.

An injection is determined by split ratio of the system. This was generally ~4 %, indicating that only $1/20^{th}$ of the sample is actually injected onto the column. This was a disadvantage over earlier systems which were direct-injection, but also allowed for completely automated operation. The system could be setup for batch runs and left unattended, something not possible with earlier gUHPLC designs.

2.2.2.2 Problems Encountered

There were two main operational issues that were encountered after extended use of the system: improper gradient loading and flow variations from column-splitter dead time mismatch. Both were accounted for, but are still important to consider for future work with the system.

First, since the gradient is preloaded it is critical that the full gradient and injection plug are pushed past the column in the 4-port union. Ideally, this volume would be as small as possible. Any excess volume would result in a delayed analysis. Unfortunately, precise positioning of the injection plug is not practical because the gradient is being loaded 10x faster than it is actually being run. Additionally, Link saw that injection of acetone plugs were significantly broadened during loading. This indicated that the gradient loading needed to be extended to fully compensate for the broadened plug. The solution was careful calibration of the system volume from injector to GST. This volume typically was ~15 µl, indicating that at least an extra 0.5 mins of loading was required for each gradient.

Secondly, it was noticed during flow calibration that the system was not entirely constant-flow, as originally designed. To measure the volumetric flow, a 80 cm x 200 µm i.d. capillary was butt-connected to the outlet of the column. The linear velocity of a flow front was determined by recording the time for a front to move a fixed distance. Column linear velocity was measured over 2.44 mm using a calibrated microscope reticule to view the flow front at 10x magnification. Split velocity was measured in the same 200 µm i.d. capillary over 50 cm without magnification. Volumetric flow was then calculated from the i.d. of the calibration capillary and the measured linear velocity.

Figure 2-8 shows the results of simultaneous flow measurements taken from both the end of the column and the splitter during a 0-100% water/acetonitrile gradient. It is evident that

the flow decreases first in the splitter and then in the column before increasing again. This is primarily a result of mismatch in dead times of the column and splitter. A 30-cm x 75 μm i.d. packed capillary column with 1.5 μm particles has a dead time of ~1 min at 25 kpsi while a 120-cm x 10 μm open tube splitter has a dead time of only ~2 sec. Any changes in viscosity that result from the gradient will first affect the flow of the splitter and then the column since the dead times are not matched. Theoretically, a 15 m x 18 μm i.d. restrictor capillary would be required to match the dead time of the column, which is simply not practical since the capillary is so long. Instead, this effect was noted and ignored for most separations as long as a similar gradient program was utilized.

2.2.3 Detection

Mass Spectrometry pairs nicely with LC methods for proteomics because of the extra information that is obtained relating to protein MW. UHPLC-MS methods have been the focus of much of the research in our lab in the recent past.[4, 5, 7] These methods were used without modification and only basic details follow.

Detection was accomplished via electrospray time-of-flight mass spectrometry (ESI-TOF-MS). The ESI interface consisted of platinum coated 20-μm pulled to 5-μm i.d. fused-silica PicoTips™ (Model #FS360-20-5-CE, New Objective, Woburn, MA) which were butt-connected via a Teflon® sleeve to the outlet of the column. A LCT-TOF (Micromass, Ltd., Milford, MA) was used as the mass spectrometer. The column-tip assembly was positioned manually at 90° to the inlet cone of the MS, and a spray voltage of ~2 kV was typically used. Sample and extraction cone potentials were set to 40 V and 10 V, respectively, to favor the higher masses of intact proteins, but limit fragmentation. Other mass spectral parameters were not modified from their default system values.

2.2.4 System Advantages

The gUHPLC initially developed and characterized by Link has been used with great success over the past few years in the Jorgenson Lab. The system has proven robust enough that three units based on Link's original design are in current use. Several advantages to this have proven useful. First, the use of commercial parts and seals has significantly increased system usage and downtime is now limited. Second, the integration of an autosampler and ability to automate batches of runs has greatly increased sample throughput. Third, the system footprint has been greatly reduced. This is further illustrated in Figure 2-1a & b. The LCT-MS present in both pictures gives an indication of scale and how much larger the original gUHPLC was. The smaller footprint has allowed for more systems to be utilized with various instruments in lab. Finally, the system operates over a wide pressure range, making future use of smaller particles a possibility.

2.3 PROTEIN SEPARATIONS USING UHPLC

The gUHPLC developed in our lab allows for the use of smaller diameter packing material in proteomic separations. The proof-of-concept for the system was approached in two ways. An analysis was first completed on an *E. Coli* lysate as this has become a standard way to compare new methods to SDS-PAGE. Next, four standard proteins <100 kDa were chosen in order to characterize the system in a more fundamental manner. These two analyses provided initial information on the performance of the gUHPLC.

2.3.1 Column Preparation & Conditions

2.3.1.1 Column Packing

All columns were prepared by using fused-silica capillaries (Polymicro, Inc., Phoenix, AZ) with a 360 μm outer diameter (o.d.), 50 μm inner diameter (i.d.) and approximately 35

cm in length. Columns were slurry packed with 1.5-µm diameter spherical C18 modified bridged-ethyl hybrid (BEH) 150 Å porous particles obtained from Waters Corp. The particle diameter, calculated by the volume average ($d_{p,50\%vol}$), was 1.53 µm. Both the packing procedure[8] and characterization of the particles has been previously reported.[9, 10] Briefly, the particles were suspended at a concentration of 10 mg/ml in 100% acetone, followed by 15 min sonication to reduce particle aggregation. Outlet frits prepared for this work did not call for the extra gap needed with electrochemical detection, as described previously,[8] and were thus prepared flush with the end of the column using 3.5-µm silica glass beads. Once fully packed, columns were pressurized in aqueous mobile phase to a higher pressure than the intended run pressure and allowed to depressurize overnight to maintain bed integrity. Inlet frits were then made using a heated wire stripper at an inlet pressure of 15 kpsi before running the column on the UHPLC system. Each column was conditioned by running two gradients before acquiring data.

2.3.1.2 Mobile Phases

Mobile phase A consisted of 5% HPLC-grade acetonitrile (Fisher Scientific, Fair View, NJ), 95% water and 0.2% (v/v) formic acid (88%, Aldrich, St. Louis, MO). Mobile phase B was prepared with 90% acetonitrile, 10% water and 0.2% (v/v) formic acid. All mobile phases were filtered with a 0.1 µm nylon filter before use on the chromatographic system

Acetonitrile was used as received from the vendor, and water was purified using a NANOpure ultrapurified water system (Barnstead International, Boston, MA) to a minimum resistance of 18MΩ*cm.

2.3.1.3 Gradient

The run gradient consisted of a 3 min delay at 1% mobile phase B after the sample had been injected, before the gradient began. A linear gradient from 1%B to 70%B over 120

mins, 0.6%B per min, was used followed by a 5 min hold at 70%B. The reverse, reconditioning gradient of 70%B to 1%B over 5 mins, 14%B per min, was then used before the following run would be injected.

2.3.2 Analysis of *E. Coli* Lysate.

2.3.2.1 E. Coli. Lysate Preparation.

A sample containing the soluble proteins extracted from an *E. Coli* lysate was provided by the Giddings laboratory in the Department of Microbiology and Immunology at UNC. The procedure has been described in detail by Link.[5] In summary, the primary growth was created from stock *E. Coli*, incubated at 37°C for 18hrs. A 100 µl aliquot is then transferred to 1 L of phosphate buffered saline (PBS) before a second 18 hr growth period at 37°C. The secondary growth was concentrated into a pellet by centrifugation and supernatant discarded. The concentrated *E. Coli* pellet was then washed with PBS by suspension and reconcentration. This washed *E. Coli* sample was finally suspended in PBS with the protease inhibitors pepstatin A and leupeptin. Cells were finally lysed by sonication bursts to rupture the walls.

Next, the cell debris was removed by centrifugation and the supernatant was transferred to a fresh centrifugation tube. This faction contained both soluble and insoluble proteins. A final pellet step was performed to remove the insoluble proteins and the supernatant was again removed. Chromosomal DNA was digested with benzonase before the final 10:1 concentration. A 5 kDA MW cutoff filter was also used during the final concentration step. The final sample, therefore, contained only soluble proteins with MW >5 kDa.

Before injection, the final sample was diluted 10:1. A 1 µl plug was loaded onto the GST before the column split. With a 4% system split ratio, roughly 40 nl is injected onto the column for analysis.

2.3.2.2 Data Analysis

The gUHPLC data was analyzed for both the peak capacity of the separation and to identify as many proteins by MS as possible within S/N limits. Peak capacity, defined in Eq. 1.10 as $n_c=(t2-t1)/w_b$, is used to characterize the efficiency of a given separation. Ideally, n_c is calculated by the average peak width of every peak in a given time window. Previous work in our lab had automated this process for chromatographic data using Gaussian fits to each peak and calculating the width from the fit.[4] Unfortunately, the data generated by the MS acquisition software (MassLynx 4.1, Waters Corp.) is not readily imported into this earlier software. As a first time solution, peak width was instead calculated by manually measuring peak width at half-height ($w_{1/2}$) and converted to w_b for calculating the n_c. This is less desirable as it adds a potential human error component to the calculation, but still serves as acceptable method for proof-of-concept work.

A disadvantage of ESI-MS is that it generates an envelope of charge states, z, that create a range of observed m/z values in the mass spectra. An example of an infusion of Ribonuclease A is shown in Figure 2-9a. As can be seen, the actual MW of RNaseA (13.7 kDa) is not observed in the mass spectra. Software based methods have been implemented in MassLynx that allow for deconvolution of the ESI-envelope to an approximate MW of the species that generated the envelope. The MassLynx method, known as Transform, requires the user to identify two adjacent peaks from which the MW of the parent species is calculated. A Transform deconvoluted spectra is shown in Figure 2-9b. For evaluation of the *E. Coli*. sample, a Transform deconvolution was performed on each manually identified peak in order to determine an approximate MW of the protein present. Finally, more advanced methods for lysate data analysis and display have been developed and are discussed in Appendix B.

2.3.2.3 Results and Performance

Results. The gUHPLC separation of the Giddings *E. Coli* sample is shown in Figure 2-10. A total of 88 peaks were identified as proteins. A few peaks were found to have multiple proteins after the deconvolution, while other peaks had too little signal to perform the deconvolution routine. Often times, protein envelopes did appear in many of these baseline peaks, but had to be discarded because the error produced by the Transform routine was too large. Results of the manual deconvolution are shown in Table 2-1. Peaks noted by (*) were apparent proteins, but discarded from the MW calculation because of an inability to resolve a deconvoluted MW. Additionally, 15 MW are highlighted indicating that the MW was present in multiple locations in the chromatogram. Whether this was a true duplicate or two separate proteins with similar MW is difficult to determine, although true duplicates are likely. Accounting for duplicates and overlaps, 81 unique proteins are present in 68 peaks. A scatter plot of deconvoluted MW from Table 2-1 vs. retention time (t_r) is shown in Figure 2-11. The general trend appears that while larger MW proteins only elute later in the chromatogram, smaller MWs elute across a broad time and gradient window.

It is generally accepted that *E. coli*. has over 4000 proteins, spread over a lower molecular weight range.[11] This makes identification fairly tedious as the error introduced by Transform often resulted in multiple hits. Additionally, the relatively modest 2% coverage indicates much more work is needed in enhancing the sensitivity of our instrument since it is likely more proteins would be identified with improved S/N. Finally, the value of 4000 includes insoluble proteins and proteins with MW <5 kDa, both of which were removed during the sample preparation, indicating that our coverage is better than 2%. Even so, much work needs to be completed to approach the capabilities of SDS-PAGE.

Peak Capacity. While the separation underperformed in regards to protein coverage, the analysis of peak capacity was significantly better. The median peak width at base calculated from Table 2-1 is 18 sec, which over the 120 min elution window, gives a peak capacity of 383. Typical peaks capacities for RPLC of proteins are rarely reported above 150.[12] This is also similar to the performance seen by MacNair, Patel and Monroe on previous gUHPLC systems when separating peptides. When initially complete, the work presented here was some of the highest peak capacities reported to date for proteins. Smith and co-workers have since reported n_c of over 1000 using a UHPLC method, although time was greatly sacrificed as the runs are generally >8hrs, 4-fold longer than our already long analysis time.[13] While we have not explored analysis times on this order, it is likely similar results could be achieved simply by the use of a much slower gradient.

Figure 2-11 also shows how the peak width varies over the course of the gradient. In general, peak width increases as the gradient progresses. This increase is likely an artifact from the gradient mismatch problem discussed earlier which causes the column flow to drop during the gradient. Additionally, it is of note that larger MW proteins tend to produce wider peaks. This may indicate that the interaction kinetics between the protein and stationary phase may be a limiting factor to higher n_c in samples with large MW proteins.

2.3.3 Analysis of Standard Proteins

The analysis of the *E. Coli* sample provided good evidence of the applicability of gUHPLC to complex sample mixtures. A drawback, however, was that the exact proteins being injected were not specifically known and are not individually characterized. Even after determination of the MW, it was difficult to uniquely identify the exact proteins, making the elution trends seen difficult to interpret based on the nature of the protein. As an alternative,

analysis of standard proteins is advantageous since their structure and properties are well-characterized.

2.3.3.1 Methods

Proteins and Sample Preparation. Four model proteins, whose chromatographic behavior and physical properties are historically well-characterized, were chosen for this study. Each was obtained in powdered form from Sigma Chemical Co. (St. Louis, MO) and used without further purification. The four proteins, listed with MW and purity reported by the vendor, were Ribonuclease A (RNaseA, 13.5 kDa, min. 90%), Myoglobin (Myo, 17 kDa, 95%), Ovalbumin (Ova, 43 kDa, 99%) and Bovine Serum Albumin (BSA, 67 kDa, 99%). .

Stock solutions for each protein were prepared separately at a concentration of ~5 mg/ml in mobile phase A and stored at 0°C until use. Samples for protein carryover studies were prepared by combing each of the four proteins into a sample vial and diluting to the target concentration, typically 300 ng/µl- 35 ng/µl, with mobile phase A before analysis. Each sample was prepared fresh daily and stored at 10°C until analysis.

Gradients. Mobile phases used were the same as above. Two main gradients were used. The normal run gradient was a linear gradient from 1%B to 90%B over 55 mins, 1.6%B per min, was used followed by a 5 min hold at 90%B. The reverse, reconditioning gradient of 90%B to 1%B over 5 mins, 18%B per min, was then used before the following run would be injected. Additionally, a slow gradient was used to improve the resolution and peak capacity. It consisted of a rapid change from the starting condition of 1%B up to 40%B over 4 mins. A 6 min hold at 40% was then used to insure proper equilibration. The gradient was then ramped from 40%B to 90%B over the next 80 mins, 0.6%B/min, before finally reconditioning the column back to the starting conditions.

MaxEnt Data Analysis. In addition to the Transform procedure described above, MassLynx offers a less user-biased, but more time consuming method, known as MaxEnt. This method is an iterative statistical method that attempts to fit simulated spectra to the acquired one using a Maxium Entropy algorithm, and has been shown to be useful for deconvolution of ESI data.[14-16] The main advantages are automation and reducing the human-bias introduced by manual selection of peaks. This algorithm was used in deconvolution of BSA spectra collected here. The software parameters are as follows: MW range was 60 kda-70 kda, resolution of 1 da, and $w_{1/2}$=0.5 da. The MIR parameter was set to 50% for both left and right. Finally, the algorithm was allowed to go to convergence, and typically converged after <20 iterations.

Reduction of Proteins. Previous work has noted proteins that have not been reduced in solution may perform worse chromatographically. A simple protocol for carboxymethylation of the cysteine residues was used in order to determine what effect, if any, reduction has on these proteins.[17] Briefly, the proteins were first dissolved in a denaturing buffer consisting of 6M guanidinium HCl in 0.6M Tris-HCl at pH 8.6. β-Mercaptoethanol was added as a denaturant along with 500 mM iodoacetamide for the methylation. The vial was then purged with N_2, and wrapped in foil before incubating for 30 mins at 37°C. Each of the four proteins was prepared individually and later diluted and combined before analysis. Unused sample was stored at 0°C before analysis.

2.3.3.2 Separations

Standard Proteins. Four standard proteins were chosen over a wide MW range and separated using gUHPLC. An example separation is shown in Figure 2-12. From this result, it is clear that the separation of standard proteins is not nearly as clean as the separation of proteins from *E. Coli.* While RNaseA is relatively narrow in width, the other three proteins

are extremely broad with Ova eluting over more than a 10 min window. The elution order is fairly similar to that seen with *E. Coli.* with larger proteins generally eluting later. The largest protein, BSA (66 kDa), does elute second, indicating that the separation is more than simply sized-based.

The broad peaks seen in the standards separation were of primary interest. It is known that *E. Coli*. lacks the mechanism for glycosylation post-translation modifications.[18] This is important because the number of protein isoforms that exist is reduced, making the separation less complex. Ovalbumin, for example has ~4% of its mass from carbohydrates, leading to multiple isoforms that can complicate the separation. It is also known that commercial standards, such as the ones used here, exhibit a certain degree of heterogeneity that can convolute the separation process. Looking at Figure 2-12 closer reveals that BSA, for instance, appears to have multiple components that are closely eluting. To improve resolution, the gradient was slowed to 0.6%B/min in order to use more of the separation space. The slow gradient did not affect RNaseA, but improved the resolution of the remaining proteins. A zoomed view of that region of the chromatogram is shown in Figure 2-13.

Seven peaks from BSA were easily resolved while Ova eluted with 4 distinct peaks. From this, it appears evident that a significant degree of heterogeneity exists in the sample and that the initially observed peak width may not be a true indication of the separation performance. Since *E. Coli* eliminates much of this heterogeneity, the sample is essentially "cleaner" giving the better performance.

An advantage the MS data collection was that each peak present could be individually integrated and analyzed with the hopes of identifying a molecular weight at each separate

retention times. Unfortunately, the low signal of each peak made this rather difficult. Integrating the whole peak gave enough signal for post-processing, but each individual peak was rather weak making post-processing difficult. An example of this is shown by Figure 2-14a & b. First (A) shows the results of the MaxEnt routine on the first BSA peak, t_r=31.69-31.85 mins. The result is rather noisy, although a major component peak at 65.4 kDa does appear. Analysis of the same spectrum by the Transform routine is shown in B. The results here indicate that the major peak is closer to the actual MW of 66.6 kDa. A similar analysis of each peak from BSA gives similar, inconclusive results. In some cases, the envelope deconvolution from one method indicates a difference from the reported mass of ±3 kDa, but the other indicates a much closer match. While it seems clear that multiple components are present in the BSA sample, it is difficult to identify the exact mass from the MS data. Since this is related to the limited signal, a more detailed analysis may be possible in the future using a more concentrated sample or more sensitive mass spectrometer.

Reduced Protein Separations. It was desirable to perform a preliminary study on the potential improvement on chromatographic performance by the reduction/alkylation of the protein. The presence of cysteine residues and disulfides bonds could complicate the chromatographic separation, and it may be useful to break these interactions to alter the structure of the protein. Ideally, from a chromatographic perspective, each protein in solution would simply be a long chain of amino acids with little secondary or tertiary structure. This would result in more straightforward and predictable separation. A straight forward reduction/alkylation was performed on the four standard proteins, and the separation results are shown in Figure 2-15. It is clear from this separation that reduction of the standard proteins does not drastically improve the separation performance. In fact, RNaseA

appears to have broadened and gotten worse from the reduction/alkylation. A few interesting results are still present. Most noticeably is the shift in retention time for each protein. The shift is summarized in Table 2-2 and is roughly 8 mins for each protein. This seems reasonable for Ova, BSA and RNaseA as each has been reduced, allowing the protein to unfold and have more interactions with the stationary phase. Myo does not contain any Cys residues, however, and is not reduced during the reaction, yet shows a similar shift it t_r. The reason for this is unknown, although it appears that Myo and BSA are still the same temporal distance apart suggesting that protein-protein interactions may be altering the separation.

Using the MS signal, the ΔM for each protein can be calculated and the alkylation reaction verified. Myo, as expected, does not change in MW while the other three proteins show a shift consistent with the addition of the alkylation reagent. Additionally, analysis of the RNaseA peaks, known to have 8 Cys residues, found that the main peak at t_r 18min was consistent with a ΔM for all 8 Cys being reacted, while the peaks at 19 and 21 mins were consistent with only 7 and 6 residues being reacted, respectively. While the protocol indicated that amounts given should be universally in excess for proteins, other protocols have called for correct stoichiometric ratios of the reagents to the cys residues present in each protein. It has also been noted that some proteins are more resistant to the reduction and need a longer reaction time. It is likely these factors are related to the incomplete reduction and alkylation that was observed. This is encouraging since the peak width for fully reacted RNaseA (t_r=18mins) is quite narrow. If the remaining residues had been properly reacted, the chromatography may have improved. Since Ova and BSA have even more Cys residues, a more complete reduction reaction may yet improve the chromatographic behavior.

2.4 CONCLUSIONS

An improved gradient UHPLC system has been developed and successfully implemented for the separation of intact proteins using small packing material. An *E. Coli.* lysate was shown to be a good sample to evaluate, although the wide range in concentration of proteins made detection problematic and reduced the overall coverage. The separation of standard protein samples proved less successful, but there is evidence that the commercially available standards are not of high enough purity for performance characterization purposes. Overall, this work provided initial evidence that high peak capacities were possible using UHPLC and that significant benefits existed when using smaller particles.

2.5 REFERENCES

(1) MacNair, J. E. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 1998.

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

(3) Patel, K. D. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2001.

(4) Monroe, M., University of North Carolina, Chapel Hill, NC, 2002.

(5) Link, J. C. Doctoral Dissertation, Univ. of North Carolina, Chapel Hill, NC, 2004.

(6) Bevan, C. D.; Mutton, I. M. *Journal of Chromatography, A* **1995**, *697*, 541-548.

(7) Dai, J., University of North Carolina, Chapel Hill, NC, 2004.

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

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

(10) Wyndham, K. D.; O'Gara, J. E.; Walter, T. H.; Glose, K. H.; Lawrence, N. L.; Alden, B. A.; Izzo, G. S.; Hudalla, C. J.; Iraneta, P. C. *Analytical Chemistry* **2003**, *75*, 6781-6788.

(11) 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.

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

(13) 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.

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

(15) Ferrige, A. G.; Seddon, M. J.; Skilling, J.; Ordsmith, N. *Rapid Communications in Mass Spectrometry* **1992**, *6*, 765-770.

(16) Ferrige, A. G.; Seddon, M. J.; Jarvis, S. *Rapid Communications in Mass Spectrometry* **1991**, *5*, 374-377.

(17) Aitken, A.; Learmonth, M. In *The Protein Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 1996, pp 339-340.

(18) Brooks, S. A. *Molecular Biotechnology* **2004**, *28*, 241-255.

2.6 TABLES

Table 2-1: Retention time, deconvoluted MW and base peak width (w_b) for gUHPLC separation of Giddings *E. Coli*. sample. Unlabeled t_r indicates multiple proteins eluting in same peak. Duplicate masses highlighted. * denotes peak too weak to deconvolute.

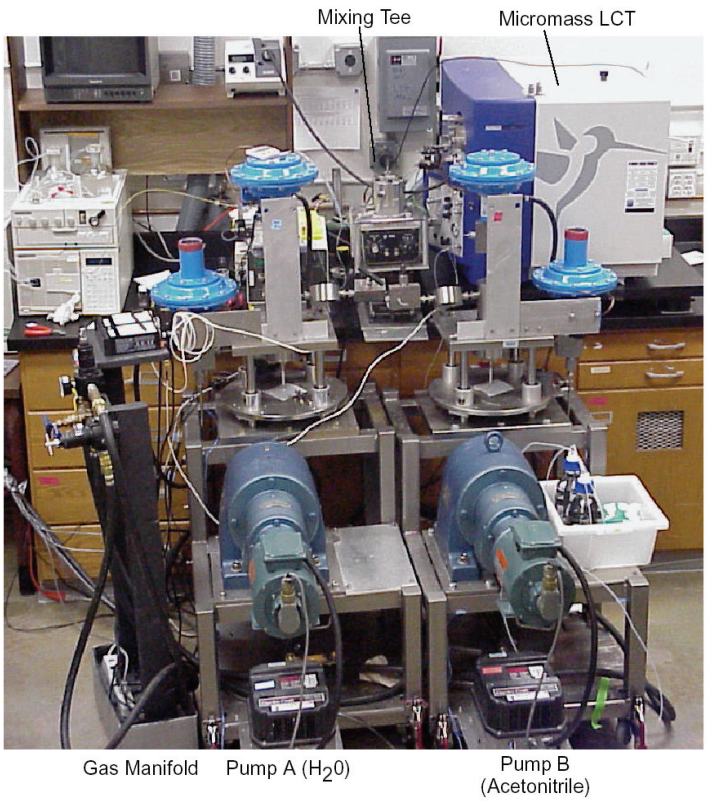
#	t_r (mins)	MW (Da)	W_b (sec)
1	12.6	5749±12	3.6
2	12.9	5091±7	3.6
	13.0	10471±5	3.6
	13.1	6256±2	2.4
3	22.4	7709±3	6
4	26.9	11898±5	8.4
5	29.6	18165±1	21.6
6	34.4	8327±1	21.6
7	34.8	8899±1	9.6
8	37.0	6411±1	12
9	37.7	14697±5	10.8
10	38.0	9574±1	19.2
11	38.5	9573±1	7.2
12	39.4	7274±2	12
13	42.1	9068±2	12
14	42.3	9068±2	9.6
	42.3	10343±1	9.6
15	42.8	9068±2	10.8
	42.8	10343±1	10.8
16	46.1	22260±2	7.2
17	46.9	7335±2	3.6
18	48.4	9745±4	13.2
	48.6	9745±4	7.2
19	48.8	9745±4	18
20	49.2	9745±4	14.4
21	49.7	9192±1	7.2
22	53.3	10107±1	18
	53.3	18080±2	18
23	53.4	11979±4	9.6
	53.4	12655±2	9.6
24	58.2	10653±1	10.8
25	58.9	9536±1	9.6
26	59.7	9387±1	12
27	60.5	14982±1	21.6
	60.5	48775±2	21.6
28	60.7	14982±1	10.8
29	61.8	17517±1	26.4
	61.8	15411±1	26.4
	61.8	9123±2	26.4
30	62.8	15415±9	6
31	63.1	15411±1	21.6
	63.1	15772±1	21.6
32	63.5	16816±1	30
33	65.9	11219±10	15.6
34	66.3	*	19.2
35	66.7	12727±4	24
	66.7	19577±6	24
36	67.0	14870±8	16.8
	67.0	19576±2	16.8
37	67.6	*	24
38	68.6	28484±10	21.6
	68.6	9493±1	21.6
39	71.6	21735±2	25.2
40	72.7	33372±14	31.2
	72.7	43575±25	31.2
41	74.6	13519±6	20.4
42	75.6	40721±18	20.4
43	76.2	*	18
	76.4	*	27.6
44	76.9	*	28.8
45	77.7	8394±2	8.4
	77.7	9277±6	8.4
46	79.0	20718±14	21.6
47	79.5	9227±1	25.2
	79.5	15937±1	25.2
48	80.9	*	26.4
49	81.4	24353±9	20.4
50	81.8	*	28.8
51	82.3	18123±2	26.4
52	84.6	13487±9	66
53	85.2	13482±1	18
	85.2	57431±49	18
54	86.0	*	15.6
55	86.9	*	25.2
56	87.3	17584±3	20.4
57	88.0	18147±16	21.6
58	89.5	30821±8	30
	89.5	35680±32	30
59	91.0	30819±9	27.6
60	91.6	15939±2	9.6
	91.6	18500±4	9.6
61	92.2	*	12
62	93.4	*	36
63	95.7	18507±9	28.8
64	96.9	43245±13	18
65	97.4	*	30
66	97.8	43247±18	33.6
67	99.2	*	19.2
68	100.0	*	18
69	101.8	32344±5	48
70	103.0	46559±54	18
71	103.2	18565±22	21.6
72	103.8	18567±2	26.4
73	104.6	18568±3	30
74	106.0	18567±2	28.8
	106.1	18179±1	14.4
75	107.9	*	54
76	109.6	22777±14	16.8
77	110.4	35104±15	13.2
	110.6	45541±17	12
78	112.9	*	6
79	114.0	*	6
80	115.1	*	4.8
81	117.0	21979±28	4.8
82	119.0	12173±15	4.8
83	121.0	*	2.4
84	122.0	21672±8	3.6
85	123.1	*	3.6
86	124.9	12215±7	8.4
87	125.9	20242±28	3.6
88	128.6	32918±8	10.8

Table 2-2: Summary of retention times for reduced and unreduced protein samples.

Protein	Unreduced t_r (mins)	Reduced t_r (mins)	T_r Shift (mins)
RibA	7.7	18.2	10.5
BSA	32.7	40.4	7.7
Myo	36.7	46.1	9.4
Ova	49.4	58.1	8.7

2.7 FIGURES

A.



B.

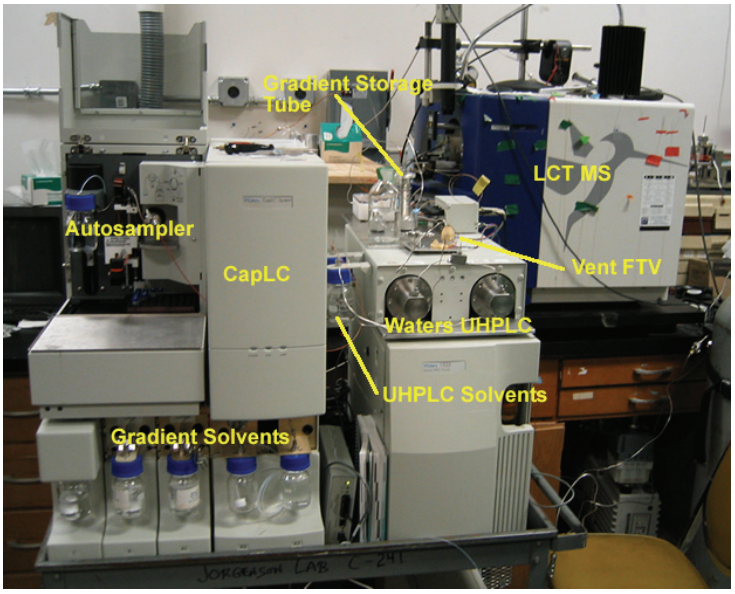


Figure 2-1: Picture of A) Patel's dual-syringe UHPLC[3] and B) Waters preloaded gradient UHPLC. Note LCT in back right for scale.

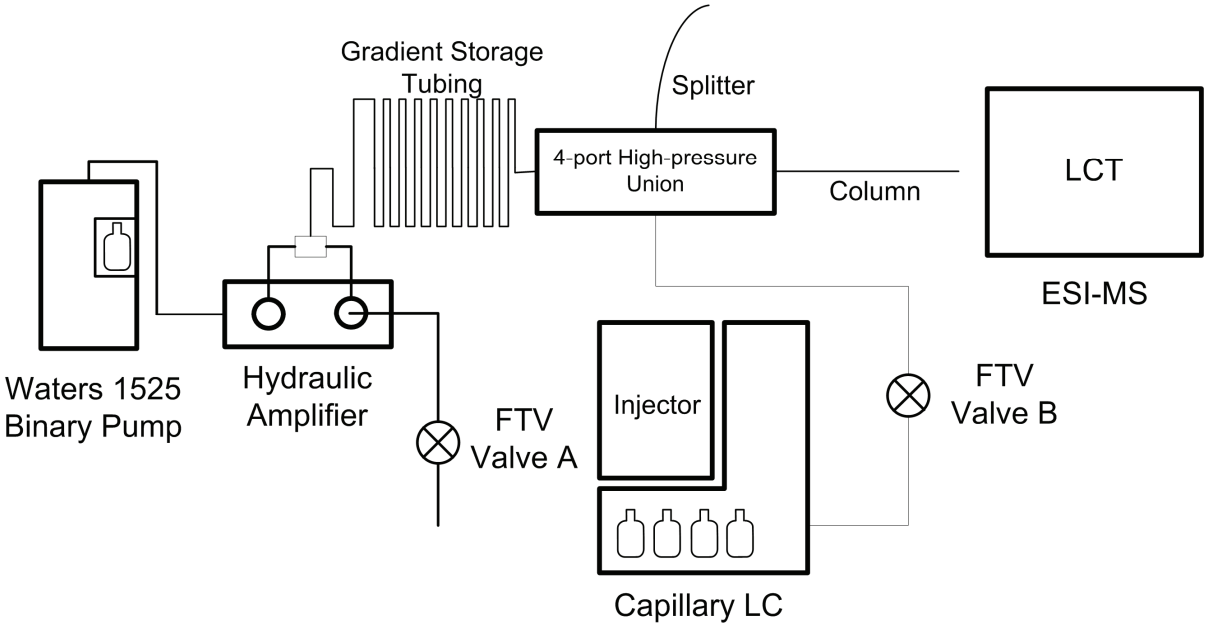


Figure 2-2: Schematic of Waters Preloaded Gradient UHPLC system.

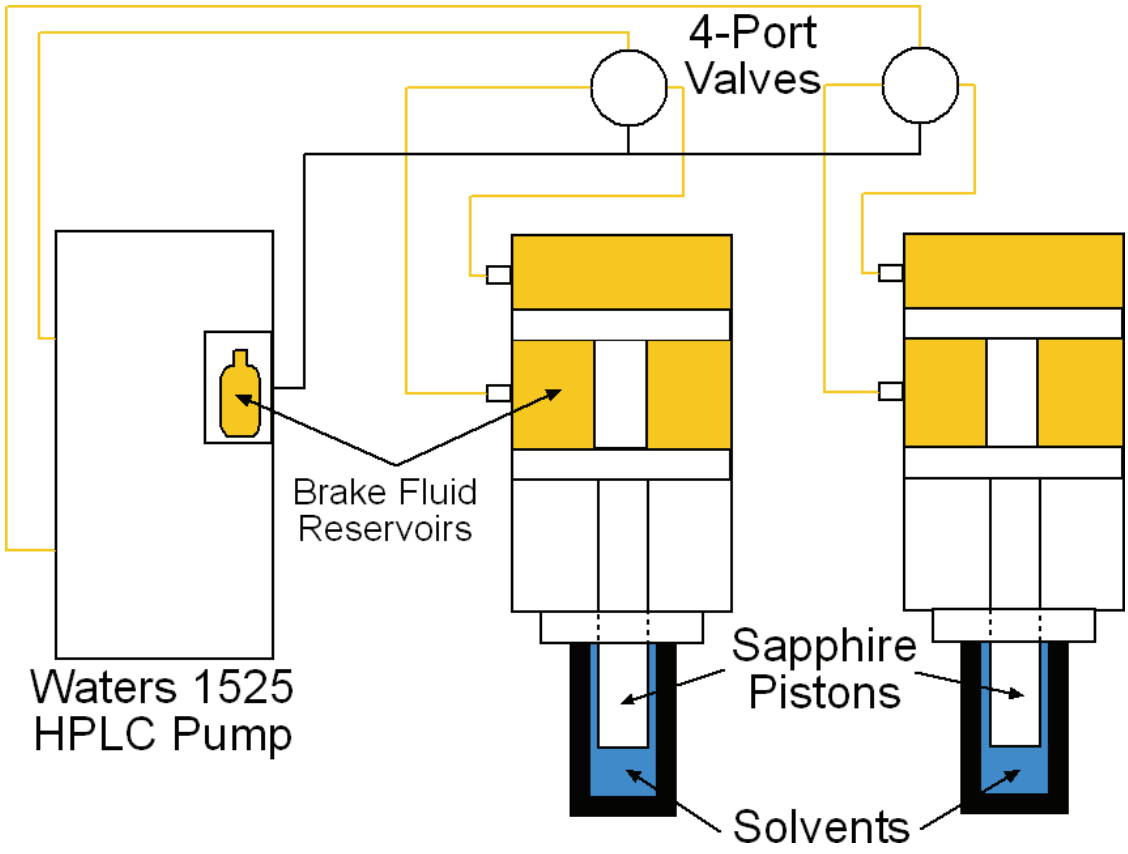


Figure 2-3: Diagram of brake fluid hydraulic amplifier used for generation of ultrahigh pressures in gradient UHPLC system.

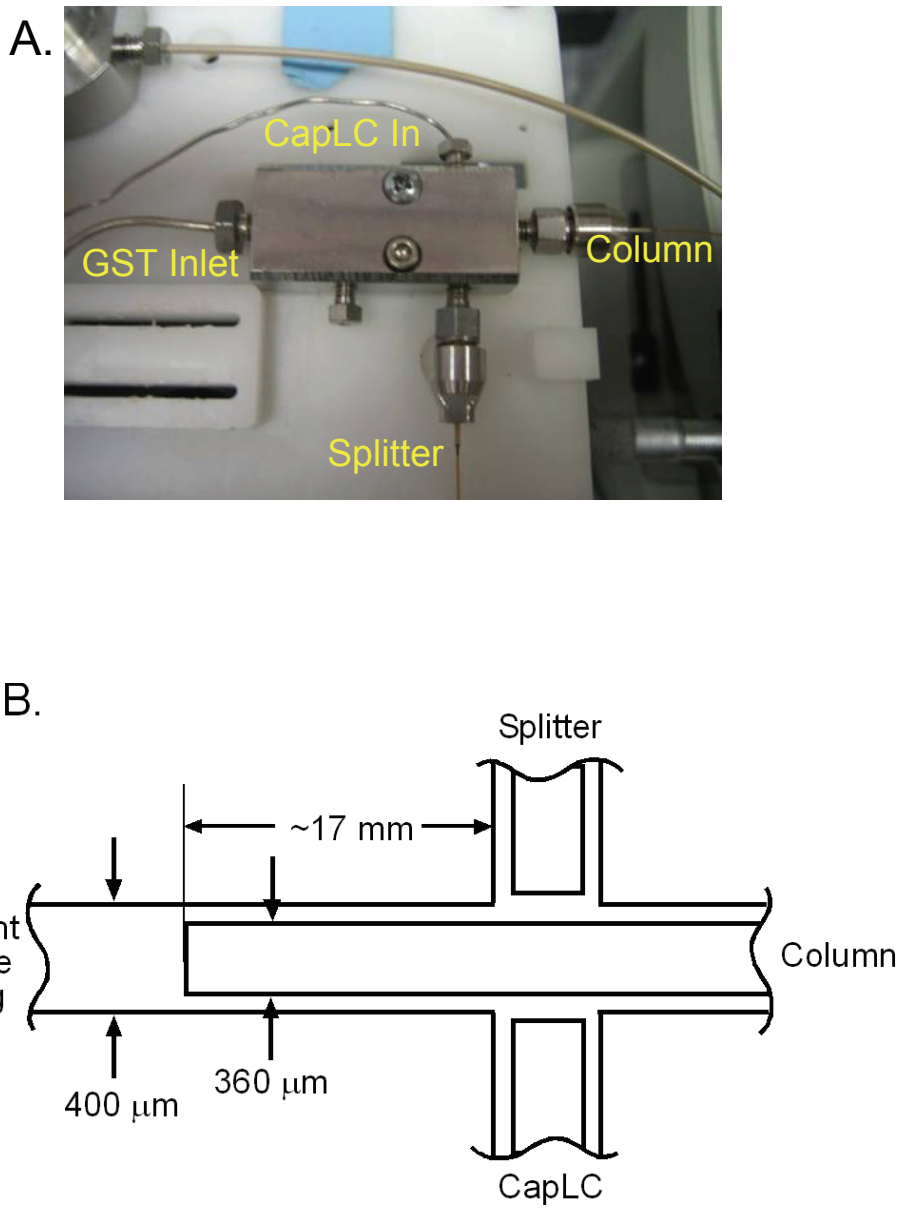


Figure 2-4: A) Picture of 4-port high-pressure union. B) Exploded internal view of high-pressure union as used with UHPLC system. Not to scale.

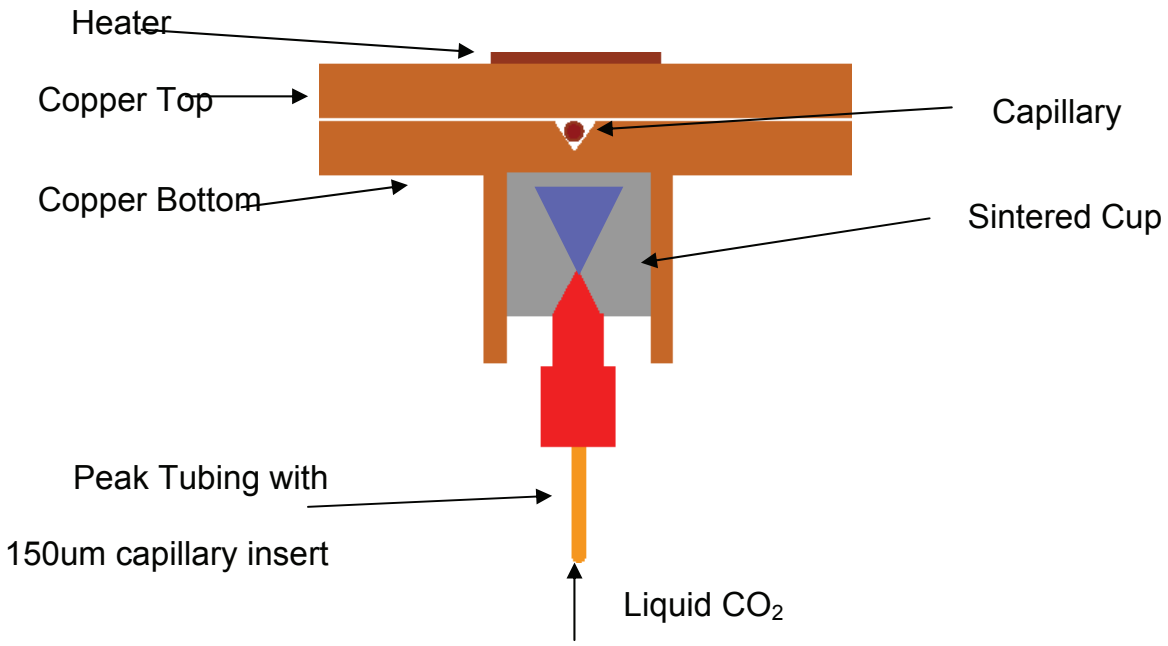


Figure 2-5: Diagram of freeze-thaw valve (FTV) used to control fluid flow in gradient UHPLC system.

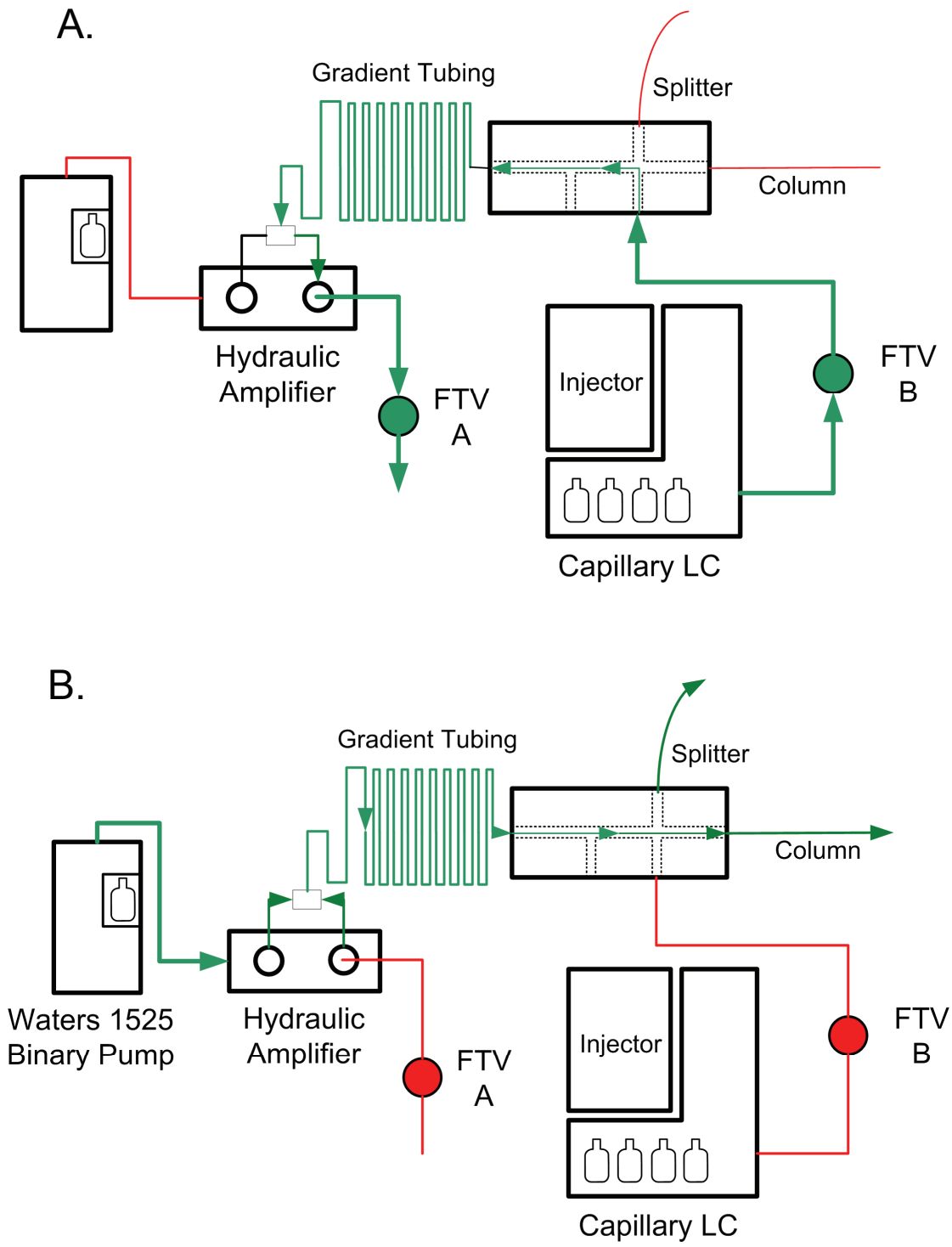


Figure 2-6: Flow diagram of A) loading & B) running Gradient UHPLC system.

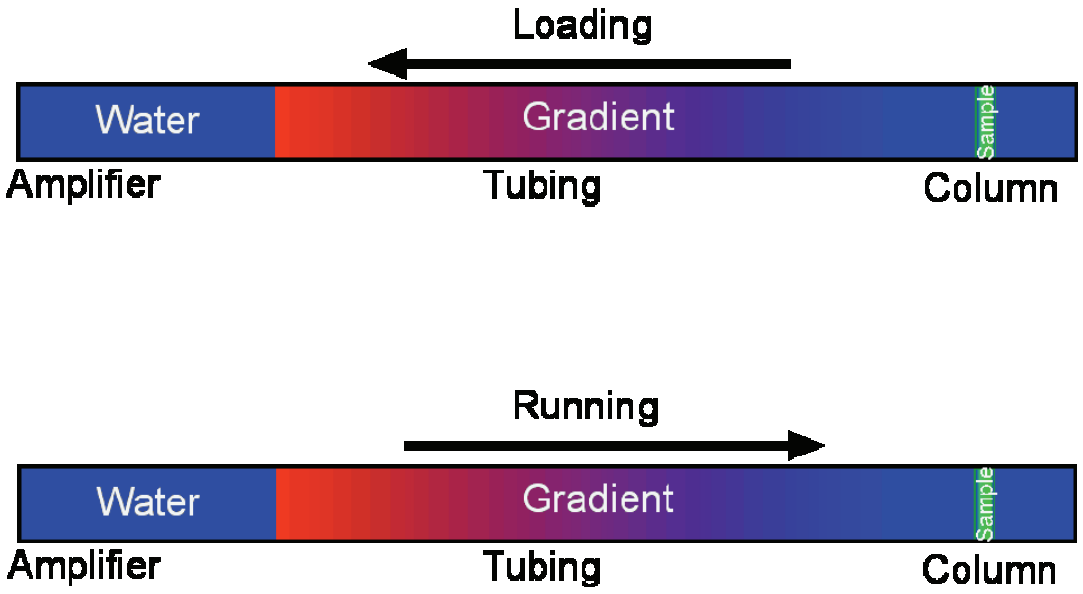


Figure 2-7: Illustration of loading and running the gradient storage tubing (GST) with preloaded gradient UHPLC.

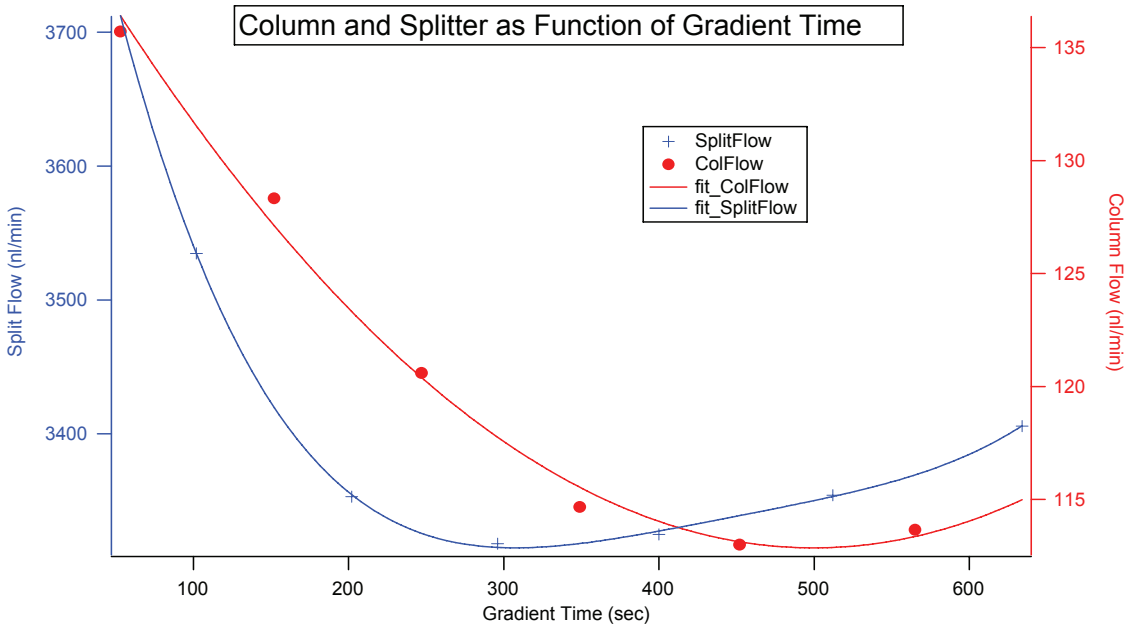
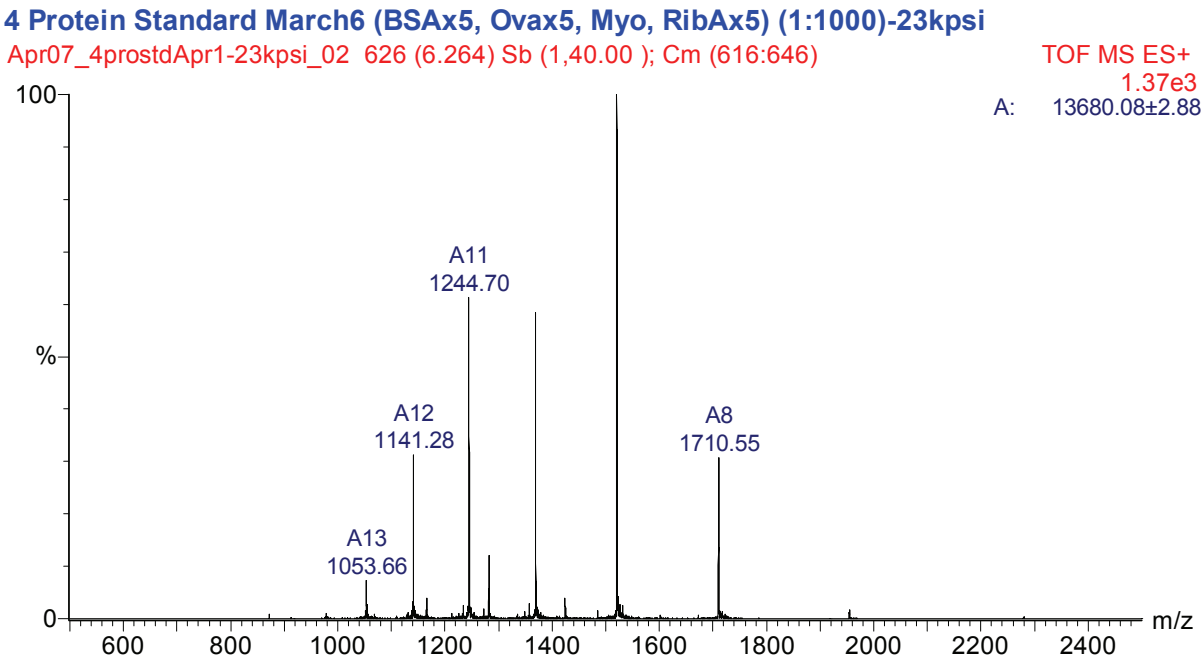


Figure 2-8: Flow differences present in gUHPLC system caused by mismatch of column-splitter deadtimes. Proper splitter would require tens of meters of capillary which is not practical.

A



B

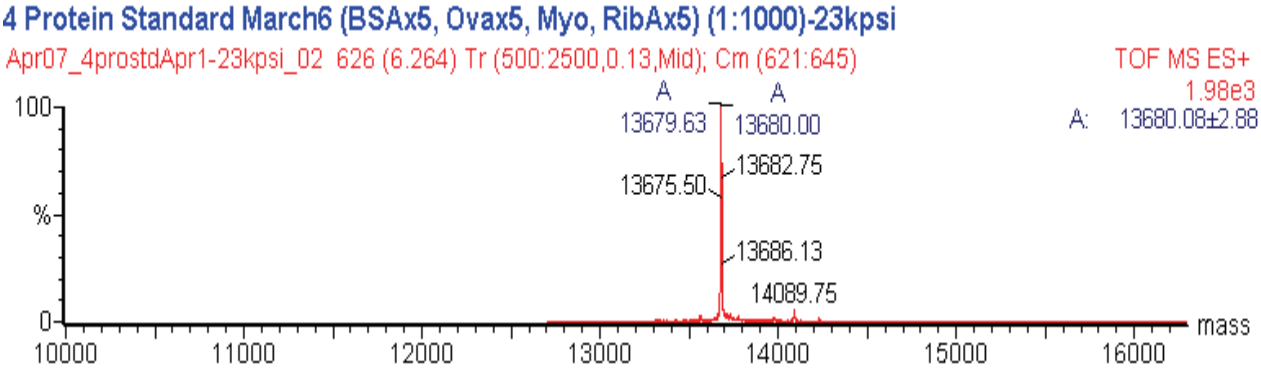


Figure 2-9: Example of protein deconvolution for Ribonuclease A. A) Raw mass envelope from electrospray ionization. B) Deconvoluted data using MassLynx Transform algorithm.

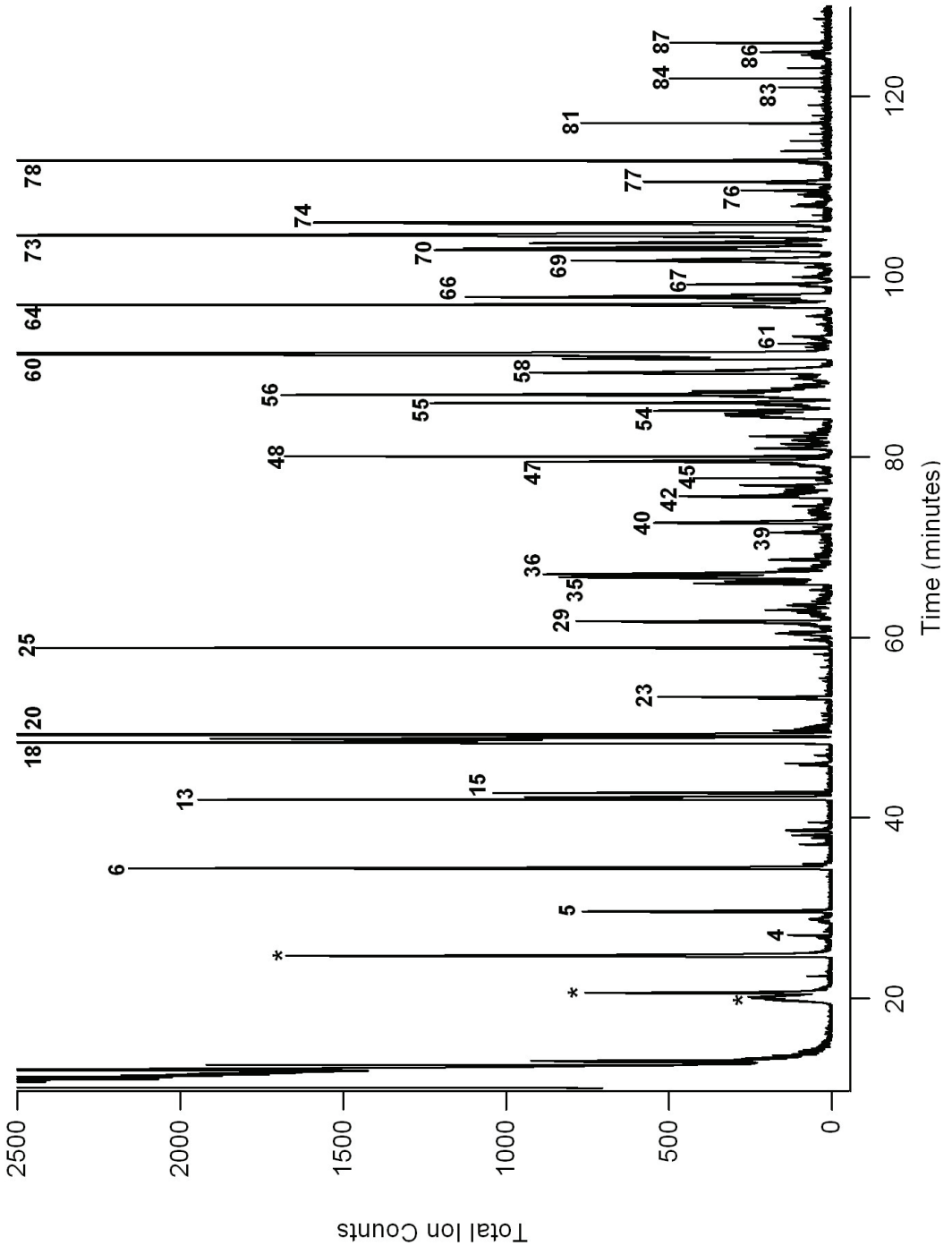


Figure 2-10: Gradient UHPLC separation of Giddings *E. Coli.* sample. Conditions: Gradient: 5%-60% ACN w/ 0.2% formic acid over 140 minutes. Column: L=43 cm x 50 μm i.d, d_p =1.5 μm porous particles.

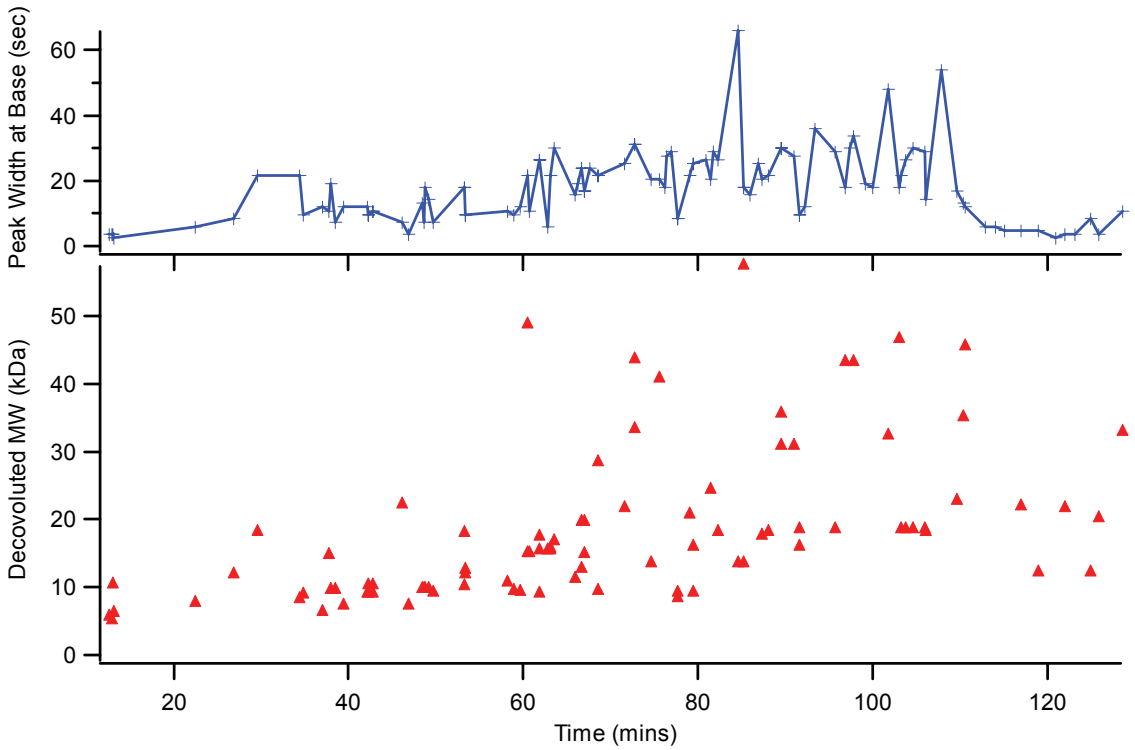


Figure 2-11: Plot of deconvoluted MW and peak width at base for labeled peaks in Figure 2-10.

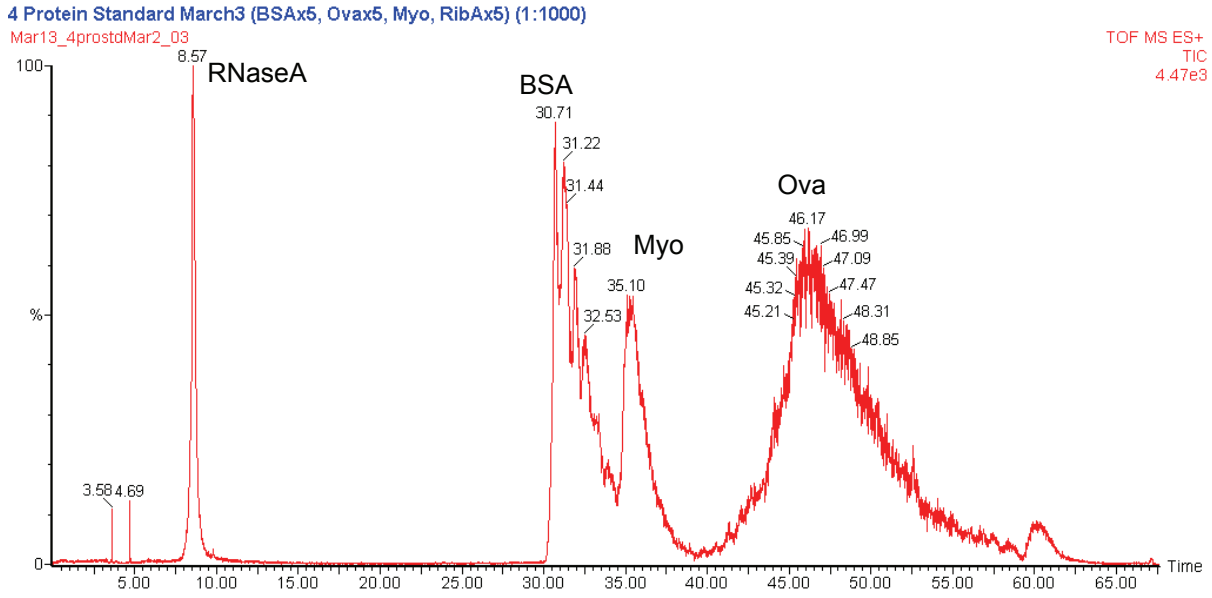


Figure 2-12: Example chromatogram with separation of four standard proteins. Conditions: 1%B to 90%B over 55 mins, 1.6%B per min. Column: L=35 cm, d_p= 1.5 µm Waters BEH

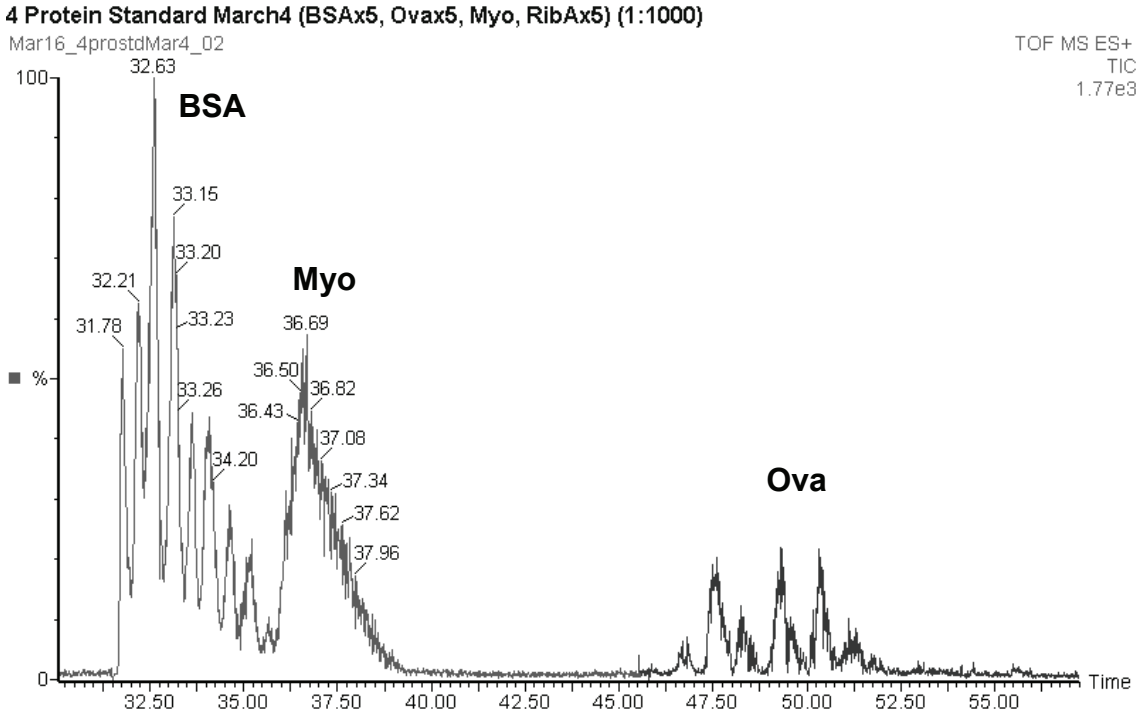


Figure 2-13: Exploded region of slow gradient for four standard proteins. Conditions: 40%B to 90%B over 80 mins, 0.6%B per min. Column: L=35 cm, d_p= 1.5 µm Waters BEH

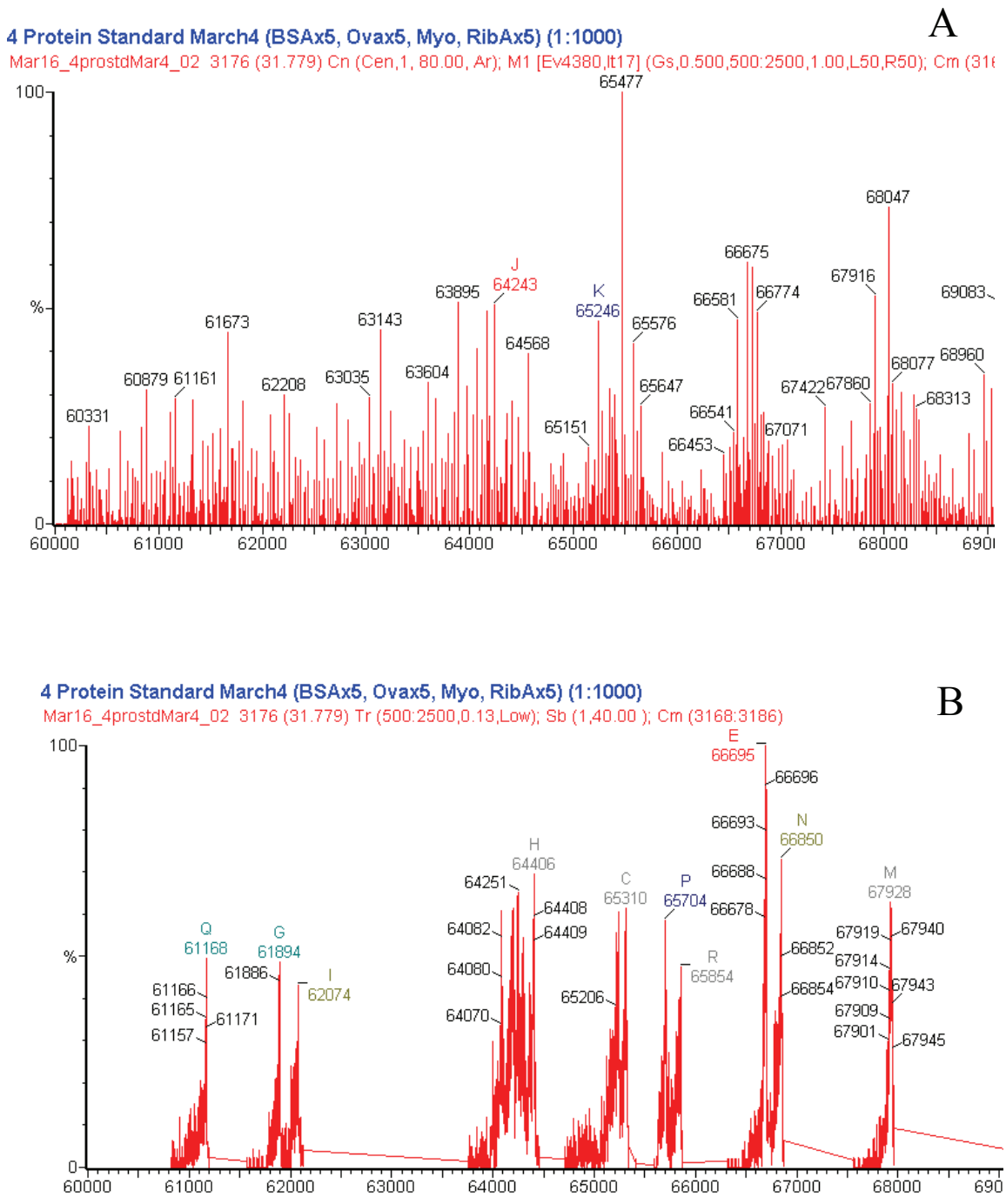


Figure 2-14: Analysis of first BSA peak from slow gradient, 31.69-31.85 mins. A) MaxEnt and B) Transform.

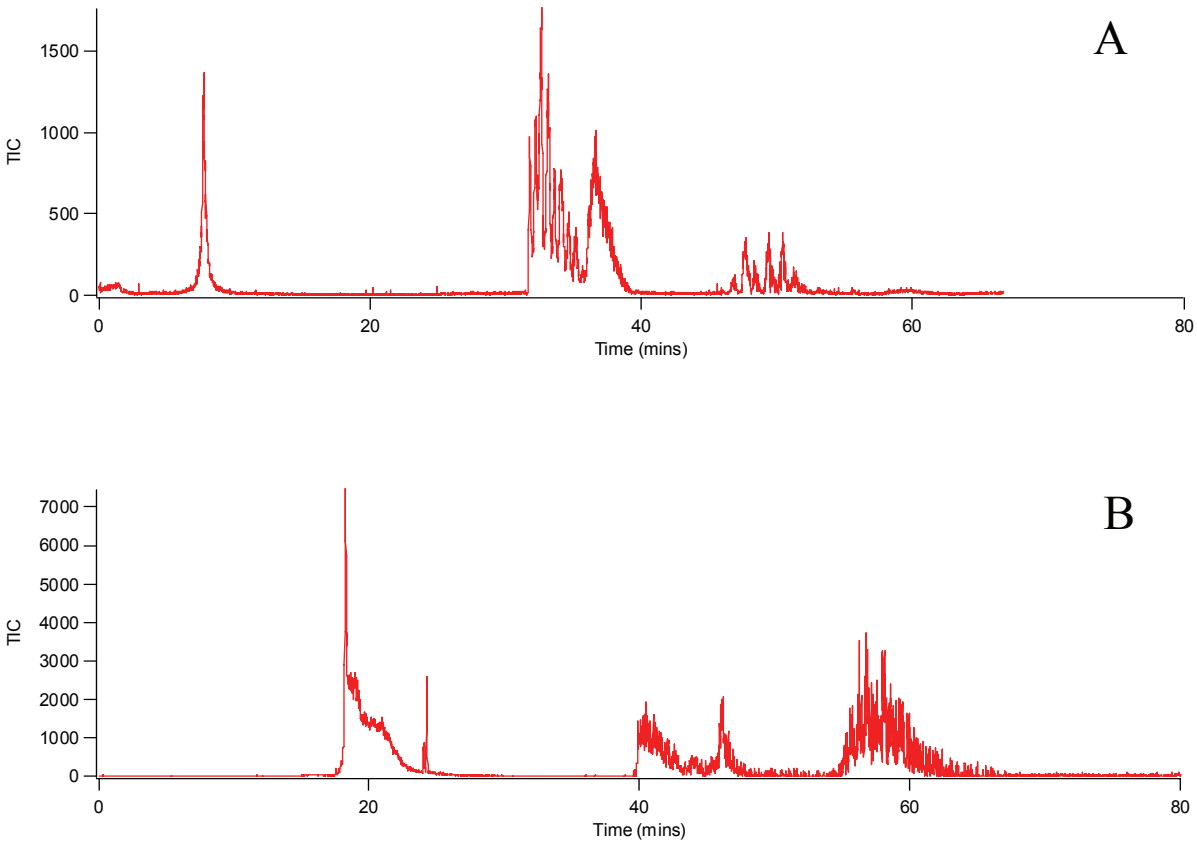


Figure 2-15: Chromatograms showing A) unreduced and B) reduced protein separations. Conditions: A)~3-6 ng Injected, 20-60%, 0.7%/min, 21kpsi; B) ~3-6 ng Injected, 20-80%, , 0.7%/min, 23kpsi

CHAPTER 3
ENHANCED PROTEIN RECOVERY IN RPLC WITH THE USE OF ULTRAHIGH PRESSURES

3.1 INTRODUCTION

In the previous chapter, a new design for a gradient UHPLC system was used to explore separations of intact proteins at ultrahigh pressures. This chapter will focus on the potential side effects that ultrahigh pressures will have on large biological molecules, such as proteins.

3.1.1 Proteins and UHPLC

It has long been known that hydrostatic pressure has various effects on intact proteins in solution[1], however, the exact nature of these effects are still being explored and continue to be an active area of research within protein chemistry[2]. Hawley first introduced the concept of an elliptical pressure-temperature (P-*t*) phase diagram to describe the pressure and temperature conditions for which a protein would favor the native or denatured conformation.[3] While the transition region at 25°C is typically reported in the range of 72 - 100 kpsi, above the usual operational pressures of gradient UHPLC, it is important to note that this earlier work was done under native solvent conditions of the protein and in the absence of a hydrophobic support which would be present in RPLC. Scharnagl recently addressed the role of the cosolvent in protein stability studies and proposed the idea of a P-μ phase diagram, where μ is the chemical potential of the cosolvent in water, for future studies

to explore the effect of the solvent.[2] This effect has yet to be fully investigated, and may prove important to gradient UHPLC, where the solvent is changing throughout the analysis. Finally, Meersman has also recently proposed a secondary region to the P-*t* diagram that addresses the aggregated state of proteins in solution.[4] The pressures proposed for the transition to the monomeric state is significantly lower, around 29 kpsi, than that of denatured state and in the range of the operational pressures for UHPLC.

3.1.2 Scope of Studies

While it is of note that the pressures required for UHPLC are a consequence of using smaller diameter particles, the use of these higher pressures in conjunction with dynamic molecules such as proteins cannot be altogether ignored. Chapter 2 has focused on the use of high pressure to improve the overall separation efficiency through the use of smaller column packing material. This study instead explored the effect operational pressures of UHPLC have on intact proteins. Early results indicated that UHP played a significant role in the carryover of the proteins which led to need for protein recovery data. The effect that UHPLC has on protein carryover along with calibrated protein recovery from the column will be presented.

3.2 GENERAL METHODS

Two separate studies relating to protein carryover and recovery are presented in this chapter. The methods used for each are unique, however, several similarities exist. The overlap between the methods is presented here.

3.2.1 Mobile Phase & Sample Preparation

The mobile phases and protein samples were the same as those used in the previous chapter (Section 2.3.1.2) For review, mobile phase A consisted of 5% HPLC-grade

acetonitrile , 95% water and 0.2% (v/v) formic acid. Mobile phase B was prepared with 90% acetonitrile, 10% water and 0.2% (v/v) formic acid.

The four proteins were Ribonuclease A (RNaseA) Myoglobin (Myo), Ovalbumin (Ova) and Bovine Serum Albumin (BSA). Stock solutions for each protein were prepared separately at a concentration of ~5 mg/ml in mobile phase A and stored at 0°C until use. Samples for protein carryover studies were prepared by combining each of the four proteins into a sample vial and diluting to the target concentration, typically 300 ng/µl- 35 ng/µl, with mobile phase A before analysis. Each sample was prepared fresh daily and stored at 10°C until analysis. Proteins used here were in the unreduced form, since results from Section 2.3.3.2 did not indicate a clear advantage for proteins with reduced disulfides, and showed similar carryover characteristics.

3.2.2 Column Preparation

All columns were prepared by using fused-silica capillaries (Polymicro, Inc., Phoenix, AZ) with a 360 µm outer diameter (o.d.), 50 µm inner diameter (i.d.) and approximately 35 cm in length (specifications given in Table 3-1). Columns were slurry packed with either 1.5-µm particles, identical to those used in Chapter 2, or 5-µm (d_p ($50\%_{vol}$)= 4.3, $90/10_{vol}$=1.45) diameter spherical C18 modified ethyl-bridged hybrid 150 Å porous particles (Waters Corp., Milford, MA) obtained from the same synthesis batch as the 1.5-µm diameter particles. The packing procedure is described in greater detail in Chapter 2 (Section 2.3.1.1), and is used here without changes.

The experiments performed called for a variation in pressure at the head of the column. This was achieved by preparing separate columns with different particle diameters, and running at a constant volumetric flow rate. Alternatively, the flow rate of a given column could be varied to control the backpressure at the column head. The parameters used are

summarized in Table 3-1. As an example, two columns were prepared to study protein carryover at elevated pressure. Both columns were approximately the same length and run at a flow rate of ~130 nl/min. However, by varying the particle diameter from 1.5-µm for the high pressure column to 5-µm for the low pressure column, the backpressure was reduced 10-fold while keeping other experimental conditions constant.

3.2.3 UHPLC & Conventional Pressure Gradient System

The gradient UHPLC system described in detail in Chapter 2 was used for all of the UHPLC ghosting and recovery studies. In order to generate conventional pressures, a few modifications to the overall system were required.

The conventional pressure gradient system utilized the CapLC to generate a real-time gradient in place of a pre-loaded UHPLC gradient. The CapLC was used directly since the check-valves of the UHPLC system were not designed to be operated at pressures below 5 kpsi and were prone to leaks at that flow rate. The modified system layout is shown in Figure 3-1A. The UHPLC side of the pump was removed from the system and replaced with a plug to reduce the volume the CapLC would need to pressurize during the run. FTV valve B was left in-line, but remained open during all runs. An internal view of the column layout for the 4-port union is additionally shown in Figure 3-1B. Of note is that the inlet of the column was moved to a position near the cross. This eliminated the column inlet from occupying the static region of the union which could have lead to poor injections. The CapLC was used to directly perform the injections with the amount being determined from the combination of split ratio, plug volume and concentration.

3.2.4 Gradient Program

The run gradient consisted of a 3 min delay at 1% mobile phase B after the sample had been injected, before the gradient began. A linear gradient from 1%B to 90%B over 55 mins,

1.6%B per min, was used followed by a 5 min hold at 90%B. The reverse, reconditioning gradient of 90%B to 1%B over 5 mins, 18%B per min, was then used before the following run would be injected. The complete gradient program was loaded at the start of the run, maintaining the target pressure of the system over the entire run and reconditioning.

The sample injection was followed immediately by replicate blanks consisting of similar 5 µl loop injections of the starting mobile phase composition for the gradient, 1%B. Blanks would be injected until no protein carryover was detected, and then a final reconditioning blank gradient would be run.

3.2.5 MS Detection

Detection was primarily accomplished via electrospray time-of-flight mass spectrometry (ESI-TOF-MS). The ESI interface consisted of platinum coated 20-µm pulled to 5-µm i.d. fused-silica PicoTips™ (Model #FS360-20-5-CE, New Objective, Woburn, MA) which were butt-connected via a Teflon® sleeve to the outlet of the column. A LCT-TOF (Micromass, Ltd., Milford, MA) was used as the mass spectrometer. The column-tip assembly was positioned manually at 90° to the inlet cone of the MS, and a spray voltage of ~2 kV was typically used. Sample and extraction cone were set to 40 V and 10 V, respectively, to favor the higher masses of intact proteins, but limit fragmentation. Other mass spectral parameters were not modified from their default system values.

3.3 REDUCED PROTEIN CARRYOVER FROM UHPLC

Carryover of proteins is a known problem in RPLC. While not the focus of the *E. Coli.* work in Chapter 2, it was often noticed that protein carryover was minimally present, if at all. This led to the need for a more systematic study to determine if UHP was playing a role in the carryover improvement. By utilizing the gradient UHPLC and CapLC systems with

different diameter packing materials, a wide range of pressures can be generated and used to evaluate the carryover of proteins in RPLC.

3.3.1 Protein Carryover at Elevated Pressure

3.3.1.1 Methods

Column parameters for the study of protein carryover at elevated pressure are given in Table 3-1, experiment A. The variable of interest is the backpressure on the head of the column, which was 23 kpsi and 2.3 kpsi for the ultrahigh and conventional pressure column, respectively. Pressure was controlled by the use of a capillary splitter of the appropriate dimensions. The split ratio was approximately 1:25, leading to 4% of the volumetric flow being directed onto the analytical column.

A 5 µl loop injection from the autosampler was used to inject the sample. The concentration of the four proteins was 80 ng/µl for BSA, RNaseA and Ova, and 8 ng/µl for Myo. Taking into account the split ratio, the amount injected onto the column was 16 ng for BSA, RNaseA and Ova, and 1.6 ng for Myo. Detection was accomplished via ESI-MS, as described, over a m/z range of 500-2500 Da.

3.3.1.2 Results

Separations of intact proteins were performed at both ultrahigh and conventional pressures by preparation of two separate columns with differing diameter particles (d_p), as described in the experimental section. The results of these two columns are shown in Figure 3-2. A column packed with 5-µm particles was run at conventional pressures of 2.3 kpsi and chromatograms are shown in Figure 3-2A. It is clear from the figure that carryover is present for BSA and Ova, while RNaseA and Myo appear to behave slightly better. Examining the mass spectra over the peak elution time, however, reveals that all four proteins exhibit carryover to some degree. RNaseA and Myo, which appear to have been

removed from the column by the end of the second blank in Figure 3-2A, still show a clear mass spectral protein envelope over the elution time. In order to clean this column entirely of protein from carryover of the first run, a total of six blanks were required until a protein mass spectral envelope was no longer detected over the background noise of the mass spectrometer.

A separation of the same four proteins at ultrahigh pressure of 23 kpsi on a column packed with 1.4 μm particles is shown in Figure 3-2B. Both the injection and the first blank are shown. It is evident from the figure that protein carryover is not observed. Examining the mass spectrum over a similar elution window also shows that no protein envelope is detectable, indicating that there is no protein present above the S/N of the mass spectrometer. An additional test was performed to verify that no experimental bias was introduced by the larger dead volume of the high-pressure union and different column configuration of the conventional pressure system. This consisted of performing a conventional pressure injection, then removing the column from the union and flushing it for several minutes. After flushing, the column was put back in the union and a normal blank was performed. If the carryover were an artifact of the extra dead volume that is present, flushing would remove any residual sample and eliminate the carryover. The post-injection flush did not show any difference in the carryover results, further validating that the experimental configuration was valid.

Figure 3-2 shows that while the absolute retention time between conventional and ultrahigh pressure did vary by roughly 10 mins, the relative retention between adjacent peaks was similar. This difference in retention time is caused by the extra dead volume that is associated with using the CapLC in real time. The fact that relative retention time did not

significantly change at ultrahigh pressure is also of note since it indicates that ultrahigh pressures are not prohibitive for the separation of large bimolecular under gradient conditions.

3.3.2 Protein Carryover with Varying Pressures

While Figure 3-2 indicates that protein carryover is affected by the column backpressure, it does not explain to what degree pressure can control this effect. To investigate this potential dependence, a series of runs at reduced pressure increments was performed.

3.3.2.1 Methods

A single column was prepared to explore carryover with decreasing pressure. The volumetric flow rate of the column was used to control the pressure, summarized in Table 3-1, experiment B. A sample and blank injection were first performed at ultrahigh pressure, 23 kpsi. This was followed by two sets of injection/blank at both the very high, 13 kpsi , and conventional, 6 kpsi, pressures. A final conditioning run was completed at the initial conditions of the ultrahigh pressure run. This resulted in two sample injections each at very high and conventional pressure before the column was again exposed to the ultrahigh pressure conditions.

The gradient program was modified slightly from the general program. For the ultrahigh pressure column, a program of 40%B to 90%B over 30 mins, 1.7%B per minute, was used with a similar hold and reconditioning. The very high and conventional pressure columns were run at reduced flow rates, ½ and ¼ the flow rate of the ultrahigh pressure experiment, respectively. To compensate for the lower flow rate and maintain constant elution volumes, the gradient slope was reduced by the same factor as the flow rate, resulting in 60 min and 120 min gradient times.

A second column was prepared to perform work at the maximum pressure of the system. Backpressure was again controlled by the flow rate, however, the main focus was to perform an injection at ultrahigh pressure and then perform a blank at the system limit, shown in Table 3-1, experiment C. An injection and gradient as described above was completed at this pressure. A blank was then performed at 40 kpsi to evaluate the protein carryover at this increased pressure. As before, the gradient slope was modified to compensate for the higher flow rate, resulting in a 17 min gradient time.

A 2 µl plug was loaded onto the gradient storage tubing resulting in injection amounts of 6 ng for BSA, and RNaseA and 0.6 ng for Myo. Ova was not injected for this experiment. Also, as described above, mass spectrometry was used for detection in the same m/z range.

3.3.2.2 Results

Protein Carryover with Decreasing Pressure. In addition to separations at ultrahigh pressure (23 kpsi), flow rate was decreased by a factor of 2 for runs at very high pressures (13 kpsi) and by a factor of 4 for runs at conventional pressures (6 kpsi). While the gradient program was modified, the carryover results at ultrahigh pressure were similar to Figure 3-2B with no indication of carryover detected. Results at 13 and 6 kpsi are shown in Figure 3-3A & B, respectively. Ovalbumin showed some signs of carryover, however, this was not as noticeable as seen in the conventional pressure work (Figure 3-2A). The other three proteins did not show any carryover that was detectable within the S/N of the mass spectrometer.

After the completion of the injections at lower pressures, a blank gradient run was performed at ultrahigh pressure. This blank run is shown in Figure 3-3C. Immediately evident are BSA and Myo. These proteins are the result of a carryover effect from the previous injections at 13 kpsi and 6 kpsi pressures. As noted in the experimental section, the column was not equilibrated at ultrahigh pressure between lower pressure runs. This

indicates that the carryover seen in Figure 3-3C is a build-up of protein on the column from the injections at 13 and 6 kpsi . It also may suggest that the dependence of protein carryover on pressure is binary in nature, where below some threshold pressure, carryover will be significant.

Carryover at 40 kpsi . The results of carryover at decreasing pressure indicate that below a threshold pressure, carryover is not eliminated. The UHPLC system is not limited to 23 kpsi , the pressure used in Figure 3-2A and Figure 3-3C, allowing exploration of somewhat higher pressures. To verify that 23 kpsi was in fact enough pressure to eliminate the carryover effect, an injection was performed at 23 kpsi followed by a blank at 40 kpsi, the limit of the system. Figure 3-4 shows the results from this sequence of injections. It is clear from this figure that no adsorbed protein was eluted by increasing pressure to the system limit. Therefore the threshold pressure for eliminating protein carryover appears to be somewhere between 13 and 23 kpsi for the test samples and column.

3.3.3 Column Performance

The results shown in Figure 3-2 were intended to demonstrate the improvement of ultrahigh pressures on protein recovery, but can also be used to loosely compare the performance between 1.5-µm and 5-µm particles. It is important to note that extensive work was not undertaken to optimize the best possible chromatographic separation as there are numerous parameters that would have been beyond the scope of the study. Additionally, the micro-heterogeneity present in research grade protein samples make an accurate assessment of column performance difficult since the heterogeneity will inevitably lead to broadened peaks. With these factors in mind, it is still of note that the use of smaller particles leads to a slight improvement for BSA in Figure 3-2B. While BSA is still quite broad the protein variants that were seen by use of a slower gradient in Chapter 2 (Section 2.3.3) of BSA are

beginning to be resolved. Overall, though, standard proteins appear to perform poorly with respect to peak width and resolution, even at ultrahigh pressures.

It is also often noted that 300 Å pores are preferable for protein separations over the 150 Å material that was used for this study. Initially, it was only possible to obtain material with a 150 Å pore size for this study, but towards the completion of the work, 300 Å material became available. Preliminary results did not indicate any improvement in column performance but showed similar characteristics to the 150 Å material with regards to reduced protein carryover. Nonetheless, 300 Å pore size material may provide a useful avenue for future work.

3.3.4 Conclusions

It is clear from the data presented in this section that UHPLC reduces protein carryover above a certain threshold pressure. The data suggests this threshold is rather finite, and below it the carryover drastically increases. Within the limits of the UHPLC system, currently 40 kpsi, carryover does not appear present. This does not indicate that recovery is 100% as some amount protein could be irreversibly lost or be below a second threshold region. Therefore, a method for determining the protein recovery is needed.

3.4 CALIBRATED PROTEIN RECOVERY

The work at elevated pressures indicated that above a certain threshold pressure, the protein is affected and carryover is reduced. Since the above work provided only a qualitative view, it does not allow the conclusion that carryover has been totally eliminated. In order to explore this possibility, quantitative protein recovery was investigated. Some of the more common methods (Bradford, Lowry assays) of quantifying intact protein recovery involve staining and spectroscopic detection. These methods share the need for discrete

sample fractions to be collected at significantly higher flow rates to maintain resolution. Radiometric detection methods have also been used,[5, 6] but would be impractical since the initial sample would need to be several orders of magnitude more radioactive to get an accurate response with the nano-volumes being used.

As an alternative, protein recovery can be calculated from the peak area after UV detection, and is described below. This method is advantageous as it requires relatively little adaptation from detection by MS used in the previous section. While the sensitivity is less than with MS, the response is directly related to protein concentration and can be used to calculate recovery. This methodology will be used to determine degree of protein recovery at conventional and ultrahigh pressures.

3.4.1 Methods

For the calibrated recovery work, absorbance detection was used due to the well known linear response related to Beer's law. A Linear UVIS 200 (Linear Technologies, Milpitas, CA) absorbance detector with capillary flow cell was used at 215nm for the detection of proteins. Capillaries with 360-µm o.d. and 200-µm i.d. were used to increase the path length and improve sensitivity. The capillaries were butt-connected to the column in the same manor that the ESI spray tips were attached.

Four standard concentrations of RNaseA, BSA and Ova were separately prepared from dilutions of the stock protein solutions to generate absorbance vs. concentration calibration curves for the detector. Myo samples were not prepared due to the poor resolution between Myo and BSA. Samples were prepared in fresh 2 ml glass vials. The concentration for the calibration curves was in the range of the actual amount of sample injected onto the column after the split ratio. Concentrations used were as follows: 25, 12.5, 6.3 and 3.1 ng/µl. To account for potential changes in molar absorptivity of the protein under different solvent

conditions, stock solutions were diluted into the approximate mobile phase composition at peak elution. The mobile phase at elution was approximately 35%B, 58%B and 72%B for RNaseA, BSA and Ova, respectively.

A generalized calibration scheme is shown in Figure 3-5. Calibration samples were infused into the detection capillary by the use of a Helium pressurized reservoir at 15 psi and a 50 cm x 50-µm fused-silica transfer capillary (Figure 3-5, step 1). A blank consisting of the mobile phase used to dilute the standards was then infused. Three fronts were recorded for each sample and an average of the response was plotted against concentration. A standard linear regression with the x-intercept fixed at 0 ng/µl was then used to calculate the response factor from the slope of the line for the detector (Figure 3-5, step 2).

As in the earlier protein carryover experiment, two separate columns were prepared to produce both ultrahigh and conventional pressures (Table 3-1, experiment D). The split ratio for the columns was equalized as closely as possible to equalize injection amounts, however, there were always small differences resulting from the use of different splitter capillaries. Split ratios for these two columns were 1:28 for the ultrahigh pressure column and 1:25 for the conventional pressure column. This resulted in roughly 15% more being injected for a given conventional pressure run. Since the purpose of this work was to examine absolute recovery, the slight difference in split ratios was compensated for in the final results.

Six standard protein samples in the range of 300 ng/µl – 37 ng/µl were prepared from fresh stock by dilution, producing injection amounts of 12 ng – 1.5ng onto the column. A gradient program of 20%B to 75%B over 40 mins, 1.5%B/min, was used to elute the samples at the desired pressure with absorbance detection at 215 nm. As noted in the general methods, blanks were injected after each sample until no carryover could be detected.

Additionally, sample injections were alternated between low and high concentrations to limit potential sample-to-sample crosstalk. The final order of sample injections used here was 37, 200, 100, 150, 75, and 300 ng/μl. Peak areas were integrated by Igor Pro 4.08 (Wavemetrics, Lake Oswego, OR) using a built-in trapezoidal fit routine. Peak area was converted to units of concentration·time by applying the calibration curve response factor. The peak volume was then calculated from the known flow rate, using the open tube capillary flow measurement method described in section 2.2.2.2, over the elution time and used to calculate the final amount recovered (Figure 3-5, step 3). Finally, recovery curves for each protein were generated by performing a standard linear regression of the amount injected versus the amount recovered (Figure 3-5, step 4).

3.4.2 Calibrated Recovery Results

Known concentrations were infused into the detector flow cell in order to generate detector response factors for each protein. Sample infusion fronts are shown in Figure 3-6. The UV absorbance calibration data is shown in Table 3-2a, which indicates similar response factors for the three proteins. This is not surprising since 215 nm is probing the amide bond region, which will be similar for most proteins. A sample recovery curve for both conventional (2.3 kpsi) and ultrahigh (23 kpsi) pressures of RNaseA is shown in Figure 3-7. The most noticeable difference is the slopes at the two pressures. It is evident that the recovery response at high pressure is roughly 60% greater than at low pressure. Additionally, the x-intercept (Amount Injected) for both columns is roughly 1 ng. This may be indicative of external losses caused by sample preparation, injection and transfer to the column or losses on the column, but is consistent for both columns. Ideally, for complete recovery, the slope would equal 1. The deviation from an x-intercept of 0 caused by losses, however, will artificially inflate the slope of the recovery curve. A summary of the three

proteins and their recovery curves is shown in Table 3-2b. It is again of note that the x-intercept is similar at both the conventional and ultrahigh pressures for all three proteins and that the slope is close to 1 for all three proteins recovered at ultrahigh pressure.

A summary of the three proteins examined in this study is shown in Figure 3-8 in the form of percent recovery. The x-axis is plotted as the Sample Concentration. It would have been preferable to plot amount injected, but due to the slight difference in split ratios it would be misleading to group the plots by injection amount. Instead, each sample that was injected onto the column is grouped together for the low and high pressures and the percent recovery is reported. It is clear from Figure 3-8 that at ultrahigh pressure RNaseA and Ova exhibit high recovery across all concentration ranges, and essentially complete recovery for concentrations above 100 ng/µl. While recovery is slightly reduced at the lower ranges, this can be attributed to small losses that occur during sample preparation and handling, as seen from the recovery curves (Figure 3-7 & Table 3-2b). Recovery at conventional pressure, however, is noticeably worse. Even at the highest concentration, the percent recovery does not exceed 50% for RNaseA and 70% for Ova. The trend is also similar to ultrahigh pressure, with recovery being worse at lower concentrations for reasons discussed above and the percent recovered being almost constant above a concentration of 100 ng/µl.

BSA exhibits worse results and the data indicates that full recovery is not taking place, even at ultrahigh pressures. While it does show an improvement over the conventional pressure recovery, the improvement is not as great as is seen with RNaseA and Ova. Additionally, the trend is not as evident as the data shows that recovery is only consistently higher with concentrations above 150 ng/µl at ultrahigh pressure. The peak shape of BSA was the worst of the four proteins studied which lead to a poor S/N at the lowest

concentrations and ultimately the inability to detect BSA above the baseline for the lowest concentration at conventional pressures. This may also attribute to the poor overall recovery since some of the peak area is inevitably lost in the baseline during peak integration. Additionally, since BSA is larger in MW than both RNaseA and Ova, this could be an indication that perhaps a higher threshold pressure is needed in order to achieve full recovery.

3.5 COMMENTS ON MECHANISM

Protein loss in RPLC has been historically attributed to protein adsorption on the column frit, packing material and column walls.[6-8] While the study cannot distinguish the location of the loss in the column, it is clear that pressure influences protein recovery in RPLC; however, the mechanism for this change is far less obvious. From the data, it is evident that the kinetics of desorption at conventional pressures are significantly slow such that the protein cannot fully desorb during the initial gradient. This agrees with previous studies that have shown protein desorption from both functionalized and non- functionalized silica surfaces to be essentially irreversible within the chromatographic time scale.[6-9] Ultrahigh pressures, however, appear to affect the protein either before or during adsorption to the column and subsequently enhance the kinetics of desorption of the protein from the column, leading to a more complete recovery.

3.5.1 Protein Denaturation & Deaggregation

It is known that increasing hydrostatic pressures destabilize proteins by compressing intra-molecular voids.[10] This compression leads to a decrease in volume ($-\Delta V$) for the protein of a few percent and subsequent unfolding to the denatured state. The pressure at which this transition occurs varies for each protein, but is typically >45 kpsi in water solvent.

Thus it is unlikely that UHPLC fully denatures a protein based on pressure alone. The partially unfolded state does still lead to an enhanced interaction with the hydrophobic stationary phase as the hydrophobic core of the protein is exposed. Guiochon and coworkers demonstrated this by studying insulin variants and showing that the retention factor (*k*) increased approximately 2-fold when pressure was increased to 3.5 kpsi from 0.7 kpsi, a relatively small change.[11, 12] It is therefore possible that partial unfolding could play a role in the improved recovery.

Elevated hydrostatic pressure has also been shown to have a deaggregating effect on proteins in solution.[13-15] Unlike the pressures needed for denaturation, the threshold for protein deaggregation in water is 14.5-30 kpsi and within the operational range of UHPLC. In a study of recombinant human growth hormone (rhGH), it was shown that hydrostatic pressures near 30 kpsi lead to increased protein solubility and allowed for near complete recovery from solution.[13] This provides another potential route for the improved recovery seen by the use of UHPLC. Pressures needed for UHPLC may lead to deaggregation and enhance protein solubility at the head of the column. This in turn results in a more efficient recovery from the column at ultrahigh pressures. Conventional pressures are not high enough to deaggregate the protein, leading to lower solubility and poor recovery.

3.5.2 UHPLC and Proteins

The transition regions for denaturation and deaggregation can be theoretically considered by plotting known values on a Pressure-temperature (P-*t*) protein phase diagram. As discussed above, the transition region is significantly different for denaturation and deaggregation. A theoretical plot based on collected values from the literature is shown in Figure 3-9.[2-4]

A final important point to consider is the pressure drop along the column. Since the pressure inside the column changes as a function of position along the column, pressure has the greatest effect in the injection block and at the head of the column. Earlier studies have indicated that the relaxation time for protein refolding from the denatured state is relatively short, roughly 2 mins.[16, 17] Thus, as the protein elutes through the column and experiences a decreasing pressure, it is possible that it could refold to the native state, which would eliminate any benefits caused by pressure induced denaturation. Reaggregation from the pressure deaggregated state, however, has been shown to have a time constant more than an order of magnitude longer, roughly 40 mins,[18] indicating this process is much slower and essentially permanent from the standpoint of gradient UHPLC. Deaggregation, therefore, may be a more likely explanation since it is less likely to be affected by the pressure drop of the column. Further work is needed to fully understand these two potential mechanisms, and the role pressure plays with protein denaturation in general.

3.6 FUTURE STUDIES

The results presented in this chapter indicate that pressures required for UHPLC improve the carryover and ultimately the recovery of proteins. To more fully understand the reasons for this improvement, a few potential follow up studies are briefly considered.

Since it is know that the relaxation times for renaturation and re-aggregation are significantly different, it may be possible to isolate the various effects. If ultrahigh pressure were applied to a protein before injection, and the protein was then run at conventional pressures, the degree of carryover could indicate whether deaggregation was a dominant factor. The lack of carryover in this case might indicate that because the protein was deaggregated off-column, aggregation is indeed playing a major role in the mechanism. A

similar concept could be explored using a chemical denaturant or deaggregant to the protein before injection onto a conventional RPLC column. This approach also has a practical aspect as not all laboratories are equipped with UHPLC equipment at this time. An off-column approach based on deaggregation could provide a more universal solution.

A further point of interest is the carryover seen by the decreasing pressure studies. This indicated that a threshold pressure was present at which point carryover was essentially eliminated. It would be ideal to control the column head pressure using the UHPLC over a wide range of pressures at small increments to measure the exact threshold pressure. Unfortunately, the only way to vary column head pressure using the current system is by column length or flow rate, neither of which are convenient for collecting measurements at varying pressures. Additionally, the pressure drop that occurs in any HPLC experiment makes it difficult to determine the pressure that is actually causing the final affect. A potential solution could utilize a dual, in-line column setup. An analysis column consisting of larger packing material could be coupled to a restrictor column that would control the overall system pressure. Most of the pressure drop would occur across this second restrictor column, and the pressure in the analysis column would be relatively constant. If detection was performed on-column at the end of the analysis column, pressure would not be changing as rapidly, and the run pressure could be readily varied. More specific information about the mechanism could be obtained since a wide range of pressures would be accessible and easily varied.

3.7 CONCLUSIONS

The experimental work has shown several effects that ultrahigh pressures have on the behavior of model proteins in a chromatographic system. The most significant effect is the

elimination of protein carryover by the pressure required for UHPLC. With conventional pressures, it has been well documented that the carryover effects lead to incomplete recovery and column fouling. Simply by running at significantly increased pressures, the carryover effect is greatly reduced if not eliminated. More importantly, for the model proteins studied, recovery was nearly complete at ultrahigh pressures, but significantly worse at conventional pressures. This indicates that pressures of 23 kpsi are sufficient for this purpose and higher pressures still are not needed to improve the recovery. The mechanism of this improvement is not understood, but is likely related to deaggregation in part.

3.8 REFERENCES

(1) Bridgman, P. W. *Journal of Biological Chemistry* **1914**, *19*, 511-512.

(2) Scharnagl, C.; Reif, M.; Friedrich, J. *Biochimica et Biophysica Acta* **2005**, *1749*, 187-213.

(3) Hawley, S. A. *Biochemistry* **1971**, *10*, 2436-2442.

(4) Meersman, F.; Smeller, L.; Heremans, K. *Helvetica Chimica Acta* **2005**, *88*, 546-556.

(5) Kessler, M. J. *Journal of chromatographic science* **1982**, *20*, 523-527.

(6) Sadek, P. C.; Carr, P. W.; Bowers, L. D.; Haddad, L. C. *Analytical biochemistry* **1986**, *153*, 359-371.

(7) Trumbore, C. N.; Tremblay, R. D.; Penrose, J. T.; Mercer, M.; Kelleher, F. M. *Journal of Chromatography* **1983**, *280*, 43-57.

(8) Sadek, P. C.; Carr, P. W.; Bowers, L. D.; Haddad, L. C. *Analytical Biochemistry* **1985**, *144*, 128-131.

(9) Buijs, J.; Vera, C. C.; Ayala, E.; Steensma, E.; Haakansson, P.; Oscarsson, S. *Analytical Chemistry* **1999**, *71*, 3219-3225.

(10) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 1552-1555.

(11) Szabelski, P.; Cavazzini, A.; Kaczmarski, K.; Liu, X.; Van Horn, J.; Guiochon, G. *Journal of Chromatography, A* **2002**, *950*, 41-53.

(12) Liu, X.; Zhou, D.; Szabelski, P.; Guiochon, G. *Analytical Chemistry* **2003**, *75*, 3999-4009.

(13) St. John, R. J.; Carpenter, J. F.; Balny, C.; Randolph, T. W. *Journal of Biological Chemistry* **2001**, *276*, 46856-46863.

(14) St. John, R. J.; Carpenter, J. F.; Randolph, T. W. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 13029-13033.

(15) Foguel, D.; Silva, J. L. *Biochemistry* **2004**, *43*, 11361-11370.

(16) Vidugiris, G. J. A.; Markley, J. L.; Royer, C. A. *Biochemistry* **1995**, *34*, 4909-4912.

(17) Panick, G.; Herberhold, H.; Sun, Z.; Winter, R. *Spectroscopy* **2003**, *17*, 367-376.

(18) Seefeldt, M. B.; Webb, J. N.; Carpenter, J. F.; Randolph, T. W.; Hesterberg, L. K. *American Biotechnology Laboratory* **2005**, *23*, 18, 20.

3.9 TABLES

Table 3-1: Column parameters for experimental work.

Experiment		Column Type	dp (um)	Length (cm)	Flow Rate (nl/min)	Backpressure (bar)
A:	Carryover at Ultrahigh Pressure	Ultrahigh	1.4	33	132	1580
		Conv.	5	36	129	160
B:	Carryover with Decreasing Pressure	Ultrahigh	1.4	33	140	1580
		Very high	1.4	33	70	900
		High	1.4	33	30	410
C:	Carryover at Maximum Pressure	Ultrahigh	1.4	35	140	1580
		Max	1.4	35	230	2760
D:	Recovery at Ultrahigh Pressure	Ultrahigh	1.4	32.5	141	1580
		Conv.	5	33	140	160

Table 3-2a: Calibrated detector response factors.

Protein	Response Factor (AU/(ng/ul))	R^2
Ribonuclease A	1.22 ± 0.03 x 10-4	0.994
Ovalbumin	1.17 ± 0.08 x 10-4	0.982
BSA	1.1 ± 0.2 x 10-4	0.971

Table 3-2b: Linear regression recovery curve parameters.

	Slope (Recovered/Injected)		**X-intercept** (ng Injected)	
Protein	*Conventional*	*Ultrahigh*	*Conventional*	*Ultrahigh*
Ribonuclease A	0.61	1.07	0.97	1.00
Ovalbumin	0.69	1.18	0.23	0.72
BSA	0.51	0.82	1.80	1.11

3.10 FIGURES

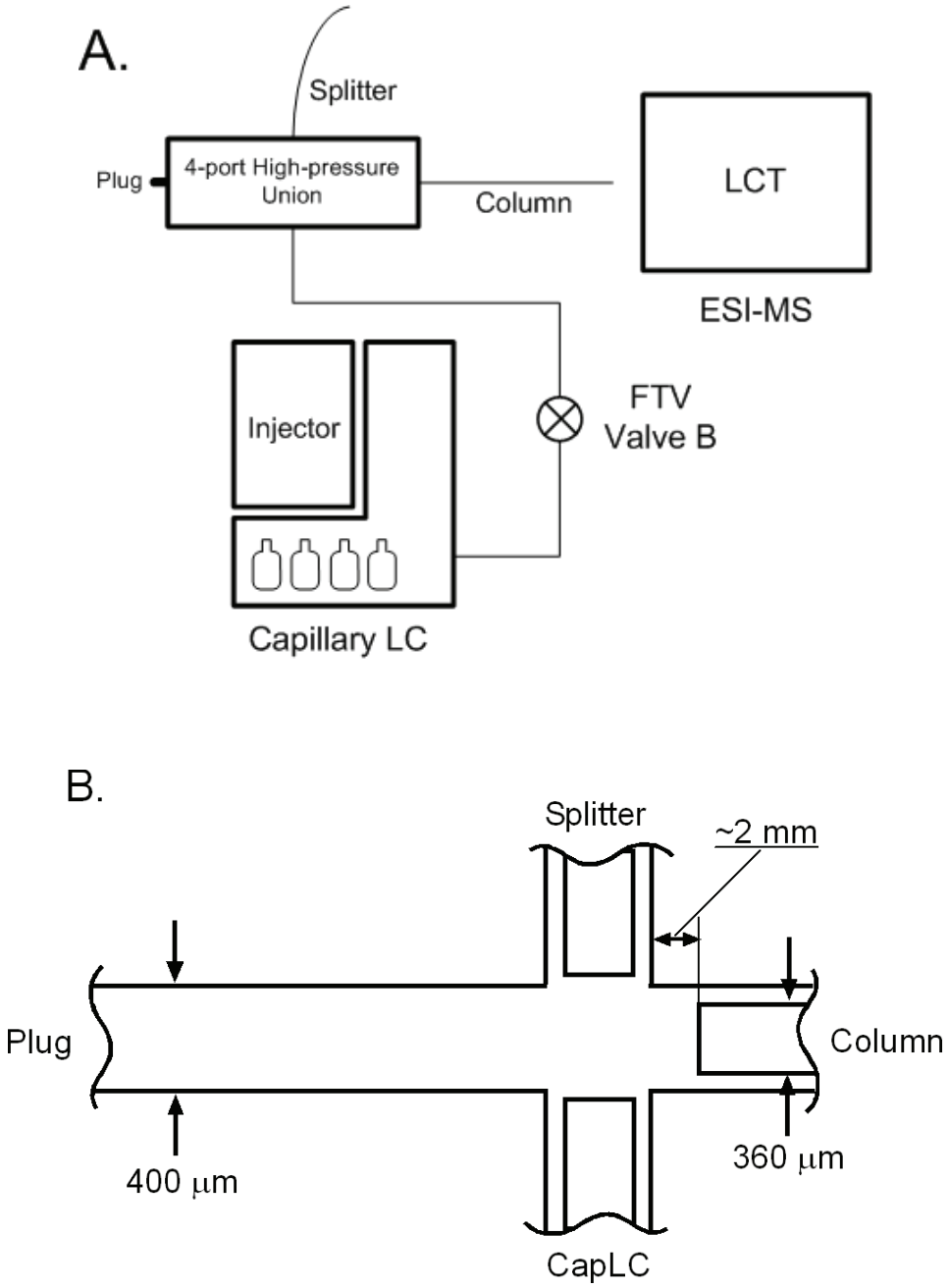


Figure 3-1: A: Schematic of gradient system used for conventional pressure work. Not to scale. B: Internal view of high-pressure union as used with conventional CapLC system. Not to scale.

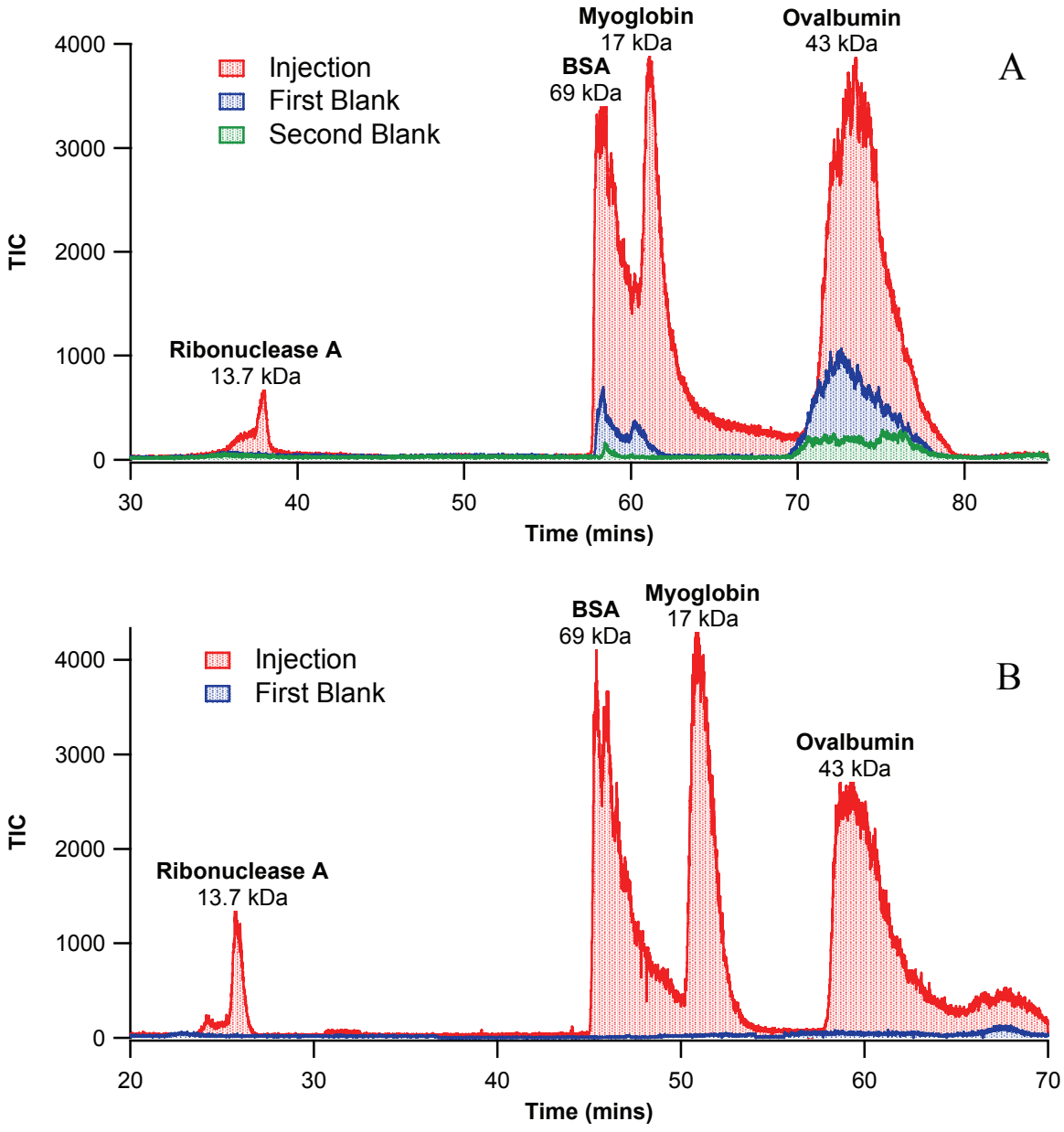


Figure 3-2: Protein carryover at elevated pressure during gradient chromatography at (A) conventional, 2.3 kpsi, and (B) ultrahigh, 23 kpsi, pressure. Proteins: Ribonuclease A, Bovine Serum Albumin, Myoglobin and Ovalbumin.

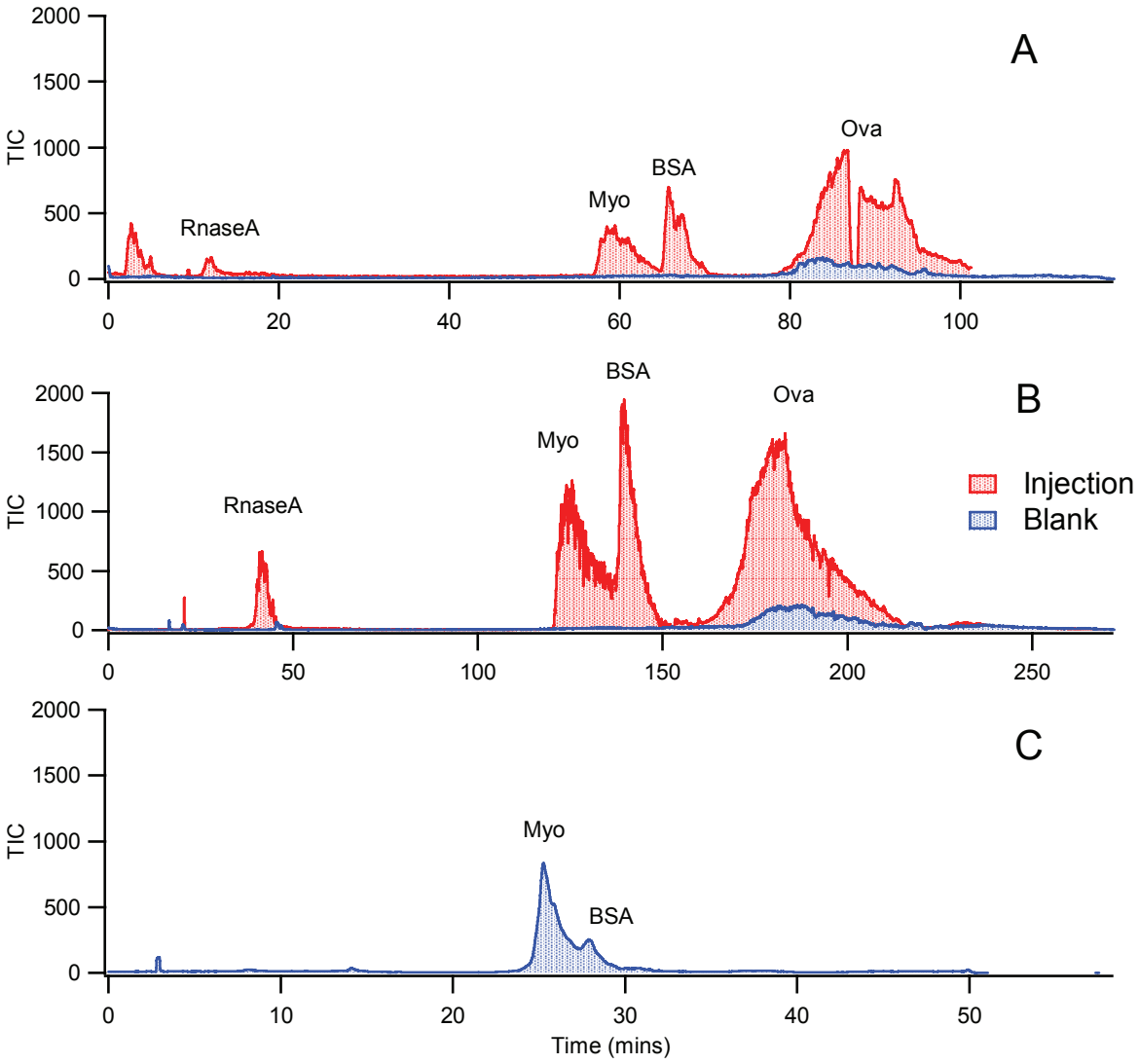


Figure 3-3: Protein carryover with decreasing pressure. A: Very high pressure (13 kpsi), 70 nl/min flowrate. B: High pressure (6 kpsi), 30 nl/min flowrate. C: Post-runs column conditioning at ultrahigh pressure (23 kpsi), no injection. Proteins: Ribonuclease A, Bovine Serum Albumin, Myoglobin and Ovalbumin.

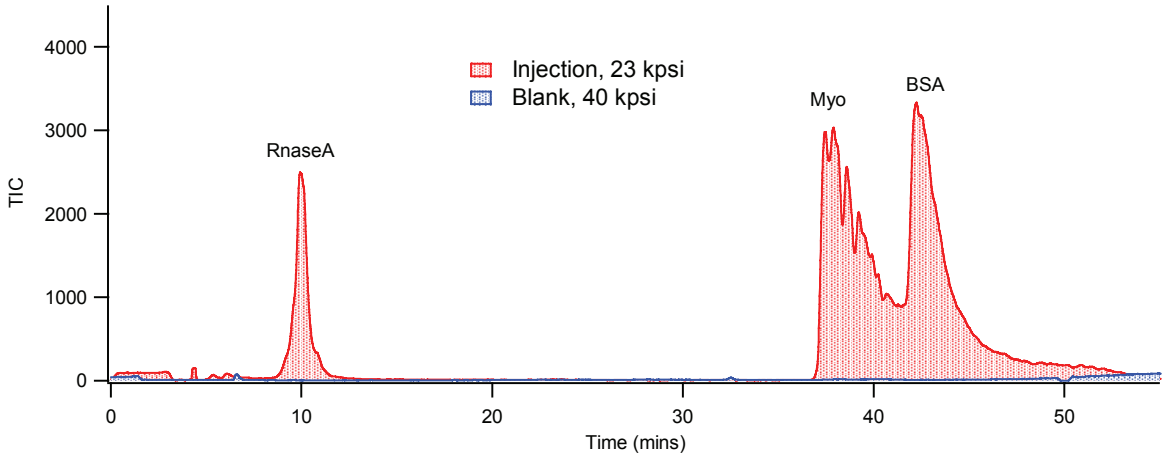


Figure 3-4: Carryover of proteins at ultrahigh pressures with an elevated pressure blank. Injection: Ribonuclease A, Bovine Serum Albumin, Myoglobin at ultrahigh pressure (23 kpsi). Blank: Mobile phase A at 40 kpsi.

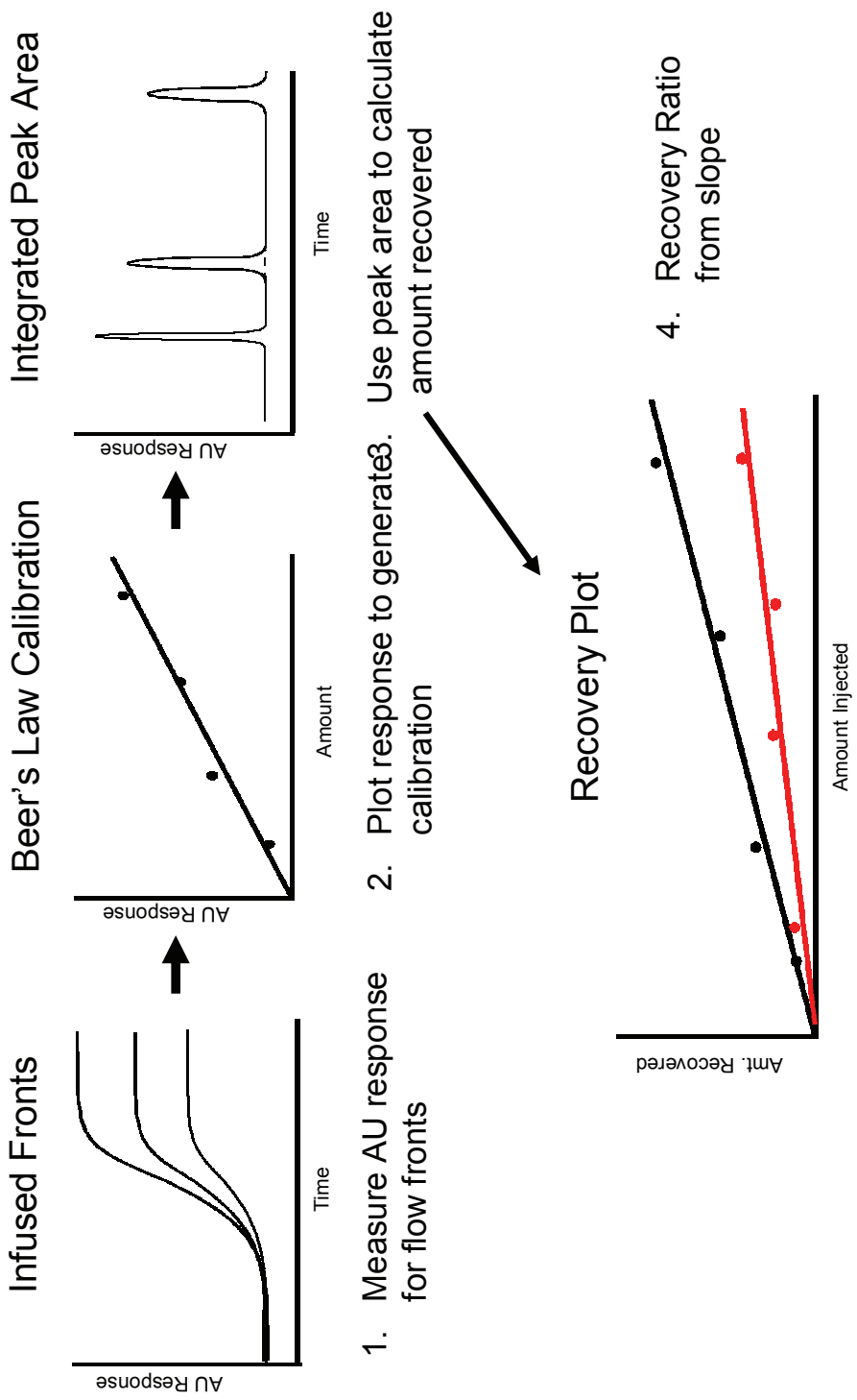


Figure 3-5: Calibration scheme for recovery of proteins at UHPLC by UV-Vis detection.

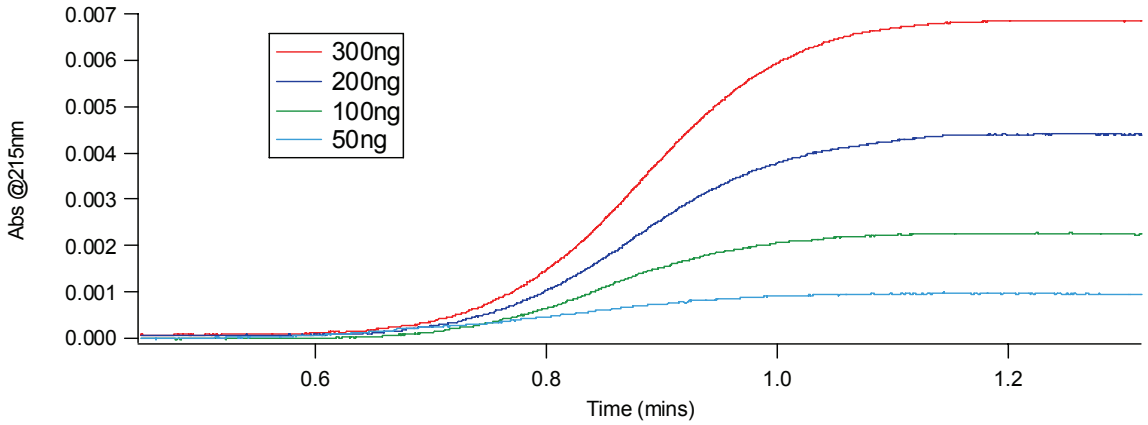


Figure 3-6: Sample of several infused fronts used for calibrating detector.

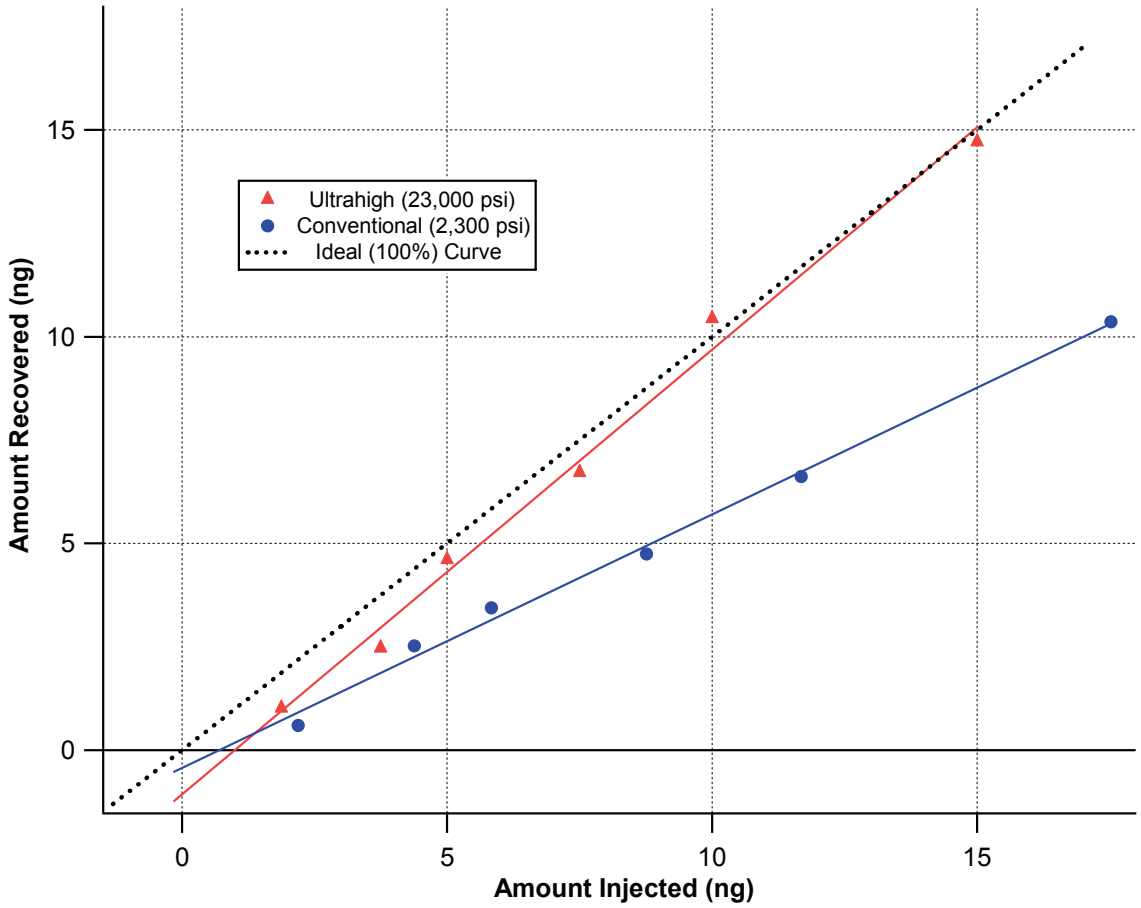


Figure 3-7: Recovery curve for Ribonuclease A at conventional, 150 bar, (●) and ultrahigh, 1580 bar, (▲) pressure. An ideal curve (····) representing 100% recovery is also plotted.

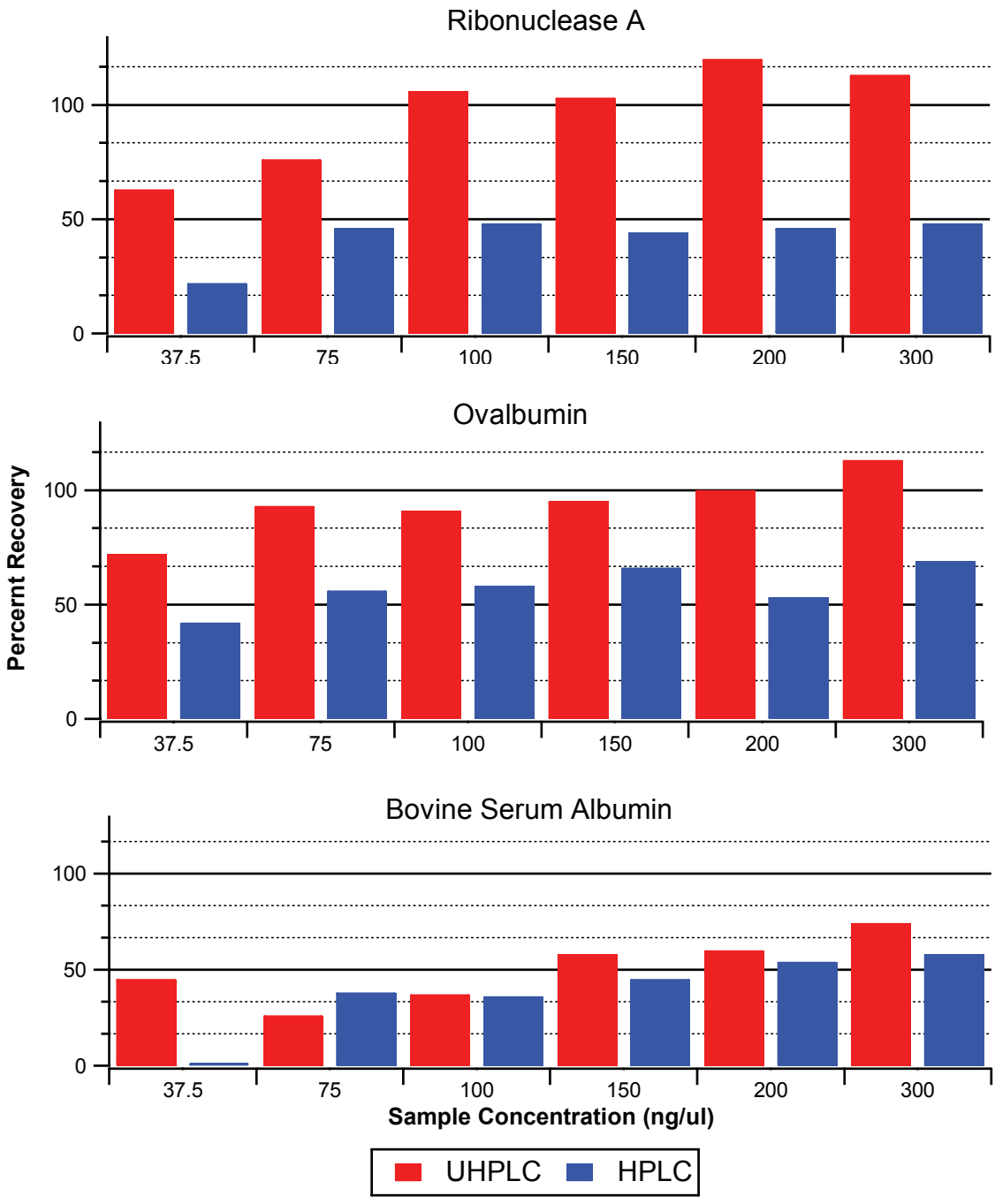


Figure 3-8: Percent recovery for Ribonuclease A, Bovine Serum Albumin, Ovalbumin. The x-axis show the sample concentration before split injection. Split ratio: 1:25 at conventional pressure and 1:28 at ultrahigh pressure. (*) Not detectable above baseline noise.

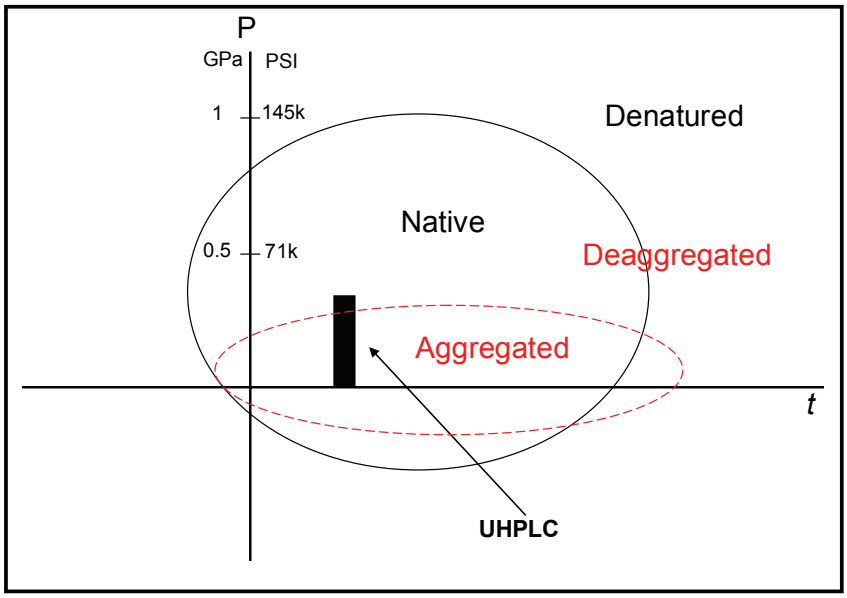


Figure 3-9: Theoretical pressure-temperature (P-*t*) protein phase diagram. Native/Denatured and aggregated/aggregated regions are shown. Transition region values estimated from collected literature values.[2-4]

CHAPTER 4
ULTRAHIGH PRESSURE HYDRODYNAMIC CHROMATOGRAPHY OF BIO- AND SYNTHETIC POLYMERS USING SUB-MICRON NONPOROUS SILICA

4.1 INTRODUCTION

Separation of proteins by HPLC methods is not limited to reverse-phase based approaches. While is it generally accepted that RPLC has the highest resolving power and is best suited for proteomic applications, several other HPLC methods are commonly used either as first dimensions or as a general sample clean-up.[1] While RPLC offers great resolution for proteins, much of this comes from the use of slow gradients, which leads to slow analysis times. Additionally, the high organic mobile phase required for elution makes it difficult to characterize protein behavior in a biological system. The main advantage for considering these alternative methods is faster analysis time and the use of native solvent conditions.

Of these non-RPLC methods, ion exchange and size-exclusion are the most common. Additionally, these two methods are less dependent upon the gradient slope to determine retention time and also allow the use of native solvent systems. Typically, however, these methods show almost a 10-fold decrease in peak capacity. Recalling equation 1.6, $N \propto 1/d_p$, the improvement in efficiency is theoretically independent of the separation mechanism. This indicates that use of smaller particles or longer columns in ion exchange or size-exclusion (SEC) by UHPLC could yield vast improvements in efficiency.

Unfortunately, our lab has not historically had much success trying to adapt these separation methods to UHPLC. McNair briefly explored SEC by UHPLC, but found that the mechanical stability of the particles was not enough to withstand forces associated with flow at pressures above 10kpsi, ultimately limiting its usefulness.[2] Sousa later saw evidence that the kinetics of ion exchange chromatography were too slow for potential gains from UHPLC to be realized.[3] As an alternative to these methods, a lesser know separation technique known as hydrodynamic chromatography (HDC) may be well suited for UHPLC. In HDC, small, mechanically stable particles can be utilized and there are no kinetic limitations since a stationary phase is not required, problems that limited size-exclusion and ion exchange UHPLC. Additionally, the particle dimensions required for HDC of proteins is a good fit for the pressures UHPLC can generate. This chapter focuses of the application of UHPLC to HDC for the separation of proteins and synthetic polymers.

4.1.1 HDC Background

Hydrodynamic chromatography is an unusual separation method, originally developed for separation of colloidal mixtures.[4, 5] This technique utilizes the parabolic flow profile that develops as a result of laminar Poiseuille flow through a narrow channel. A generalized schematic of both the flow profile and the resulting HDC mechanism is shown in Figure 4-1. In HDC, particles with a larger radius are excluded from the low flow velocity regions that are present at the walls of the channel. Smaller particles can sample a larger range of flow velocities, and therefore have a lower average velocity. This results in a SEC-type elution order in which larger particles elute before smaller particles. The slowest elution time for any particle will be the dead time, t_m, of the column since this represents the elution time for an "infinitely" small particle. As a result, HDC separations are relatively fast since samples spend a short time in the column.

Original HDC work was performed in open tube capillaries using the inner diameter as the separation channel. Channels can also be formed in packed beds using the interstitial space between particles as the separation channel. As will be shown, capillaries are not readily available with channel sizes narrow enough for protein separations, and packed beds are required. An advantage of HDC over SEC is that the particles can simply be solid silica spheres and do not require intraparticle pores for separation as SEC does. This leads to a much greater mechanical stability that can withstand UHPLC pressures. Additionally, as smaller particles are used, both the column efficiency from van Deemter theory and the resolving power from HDC theory increase, leading to a potential for extremely fast, high resolution separations.

4.1.2 Theory

Various numerical models have been developed to describe the migration of particles via the HDC mechanism.[6, 7] Since analytes will elute before the column dead time, t_m, the ratio τ defined as:

$$\tau = \frac{t_p}{t_m} \qquad \textbf{(4.1)}$$

where t_p is the retention time of the particle, is used to describe migration behavior. In HDC, $\tau \leq 1$, with a value of 1 indicating that essentially no HDC separation mechanism is occurring. This ratio, τ, can be further defined as a function of the aspect ratio, λ, by:[7]

$$\tau = \frac{1}{1 + 2\lambda - C\lambda^2} \qquad \textbf{(4.2)}$$

where $\lambda = r_i / R_c$, r_i is the radius of the solute in solution, and R_c is the effective radius of the channel. The value C is a constant used to account for secondary effects such as particle rotation, permeability, or deformation, and typically has a value between 1 and 5. A value of

1 is the most idealized case in which only particle exclusion and the flow profile are considered in the mechanism. The typically accepted value for polymers is 2.7, but other values have been reported.[7]

Various values of λ and the corresponding τ calculated from Eq. 4.2 are listed in Table 4-1. Additionally, retention times for a range of column dead times are calculated to show that the separation window for HDC is rather narrow in time. Even for a 10 min column dead time, the peaks will only elute over a 2.5 min window. Thus, complex samples require high column efficiencies.

Theoretical plots of τ(λ) with varying values of the constant *C* are shown in Figure 4-2a. From this plot, it can be seen that for large aspect ratios (λ > 0.4), the HDC mechanism begins to exhibit inversion and analyte overlap will occur. Thus, the generally accepted range for HDC is λ=0.02-0.4 and τ = 0.75-1. Figure 4-2a is more intuitive from a theoretical aspect since the dependent variable τ is plotted along the y-axis. Historically, however, HDC data is plotted in form shown by Figure 4-2b which displays the more common "calibration plot". This is simply Figure 4-2a with the axes swapped, and can be used to calculate the analyte size from a experimentally measured value of τ.

In order to correlate the aspect ratio to particle size, both R_c and r_i must be known. For packed capillaries, the interstitial channels can be treated as an array of capillaries with an effective channel radius defined by[6]:

$$R_c = \frac{d_p}{3}\frac{\varepsilon_i}{(1-\varepsilon_i)} \qquad \textbf{(4.3)}$$

where d_p is the particle diameter and ε_i is the interparticle porosity, usually 0.4. This shows that smaller packing material leads to smaller channels, causing higher aspect ratios, λ. It should be noted that R_c is derived from the hydraulic radius, R_h, and is equal to $2R_h$.

Calculation of the effective radius for polymers has been shown to be a function of the radius of gyration, R_g (μm), previously defined by Venema as[8]:

$$R_{eff} = \frac{\sqrt{\pi}}{2} R_g \qquad \textbf{(4.4)}$$

where $R_g = 1.39 \times 10^{-5} \cdot M_w^{0.588}$ and M_w is the weight of the polymer in g/mol, for polystyrene in THF, as measured by light scattering.

Similarly, a relationship between the Stokes Radius, R_A, of a protein and its M_w can be described by[9]:

$$R_a = \left(3\overline{V}/4\pi N_a\right)^{1/3} \cdot M_w^{1/3} \qquad \textbf{(4.5)}$$

where $\overline{V}$ is the specific volume and N_a is Avagadro's number. This assumes that proteins in solution act as independent solid spheres. Since $\overline{V}$ will vary for proteins, it is more useful to empirically fit Eq. 4.5 to reference values with the general form $aM_w^{1/3}$ in order predict R_a. A fit to reference data, shown in Table 4-2, results in $R_a = 0.884 \cdot M_w^{1/3}$. The Stokes Radius is a hydraulic radius, however, and it is more useful to define it in terms of an effective radius which is used for R_c in Eq. 4.3. Since R_c=2R_h, the Stokes Radius can be written as:

$$R_{a,eff} = 2R_a = 1.78 \cdot M_w^{1/3} \qquad \textbf{(4.6)}$$

when using R_a in relation to R_c for HDC. This function, along with Eq. 4.4 are used in calculation of the aspect ratio, λ, and ultimately τ.

Finally, the resolution R_{ij} between two closely eluting particles, *i* and *j*, can be defined as[10]:

$$R_s = 2(\alpha - 1) \cdot (1 - C\lambda)\lambda\tau \cdot \sqrt{N} \qquad \textbf{(4.7)}$$

where $\alpha=\lambda_i/\lambda_j$ and N was defined in section 1.6 as the number of theoretical plates. This equation also demonstrates an interesting potential advantage of HDC. It shows that the improvement in resolution will depend on both the kinetic term, N, and the retention term, $(1-C\lambda)\lambda\tau$, which is entirely dependent on the particle aspect ratio, λ. Since a decrease in d_p not only increases N, but also increase λ, the improvement in R_s will be greater than just the kinetic term would predict.

If it is assumed that pressure is not a limitation and the H_{min} for a given column can be reached, then the resolution factor (R/R_0) can be plotted to show the theoretical improvement gained from using smaller particles in HDC. This is shown in Figure 4-3 for a 100 and 90 Å particle, which would correspond to a very large protein or medium sized polymer. It can be seen that both N and λ are improving the R_s, and that the contribution from λ is actually greater until the HDC inversion point. Practically, this means that the column dimensions must be chosen based on the analyte sizes, and for a wide range of analytes, smaller packing material may not necessarily yield better results.

4.1.3 Motivations for sub-micron HDC

Hydrodynamic chromatography is a size-based separation method that may prove useful for separation of proteins in native conditions. Due to the complex elution mechanism, however, it is important to first consider the parameters needed for the HDC mechanism. Table 4-2 and Eq. 4.5 show that most proteins below 1,000 kDa have a Stokes Radius between 10 and 100 Å in solution. This relatively small radius indicates the need for small channels in order for the HDC mechanism to be effective. For a hypothetical column packed with 5 μm particles, the resulting channel radius from Eq. 4.3 is ~1000 nm, estimating a $\lambda<0.01$ for even the largest protein. At this aspect ratio, essentially no separation will occur

with even the best performing column. Clearly, smaller particles are required for HDC of proteins. The optimal d_p for protein HDC can be determined by calculating the R_c that gives the biggest range of λ to maximize the use of the separation space. For proteins, as shown by Figure 4-4, this optimum d_p occurs at < 0.5 μm. Calibration plots for proteins can also be generated based on the combination of Eq. 4.2 and 4.5. Plots for R_c of 115 nm, 80 nm and 60 nm are shown in Figure 4-5. These plots additionally point to a need for particles < 0.5 μm in order to obtain high enough value of τ to achieve sufficient separation.

A survey of the literature reveals that there has been only one reported attempt at HDC of proteins to date.[10] This example looked at only the largest proteins, and utilized 2 μm particles, which are simply too big to separate proteins effectively. The researchers instead used the ionic strength of the mobile phase to create a thick double layer and effectively reduce the R_c. While it is more ideal to use a smaller d_p, the back pressures required for particles <2 μm are extreme and would have been problematic for earlier researchers. The UHPLC technology developed by previous work in our lab is well suited for sub-micron sized particles, and thus has a great potential to overcome previous limitations experienced by early HDC research.

Finally, sub-micron HDC may prove useful for separation of polymers in addition to proteins. HDC of polymers using columns packed with 1 μm particles has been reported with over 100,000 plates, however the column length was <15 cm due to pressure restrictions.[8] UHPLC methods offer the ability to use longer columns, or slightly smaller particles for the separation of low MW polymers. Application of these UHPLC methods to sub-micron particles with both proteins and polymers therefore holds great potential.

4.2 ULTRAHIGH PRESSURE HDC

4.2.1 Isocratic UHPLC Instrumentation

The gradient UHPLC system presented in previous chapters is unfortunately not easily adaptable to isocratic UHPLC due to the way the injection is made. Since injections made under isocratic conditions do not refocus the sample band at the head of the column, significant band broadening would occur in the gradient storage tubing and 4-port union. This band broadening would carry over to the actual run, ultimately limiting column efficiency. Methods for isocratic UHPLC, however, have been well characterized and are under continued development in our lab.[2, 11-14] The system, described here in brief, is unchanged except for a few minor details as noted. Pressures for UHPLC are generated by the use of two constant pressure pneumatic amplifier pumps manufactured by Haskel, Inc. (Burbank, CA). These pumps consist of an injection pump, typically operated around 1-3 kpsi, and an elution pump, capable of 75 kpsi, connected in series. A schematic showing general pump layout is shown by Figure 4-6.

A custom injector and ultrahigh pressure fitting for capillary columns was designed and built in-house to make direct injections onto the column in order to limit pre-column band broadening. Injections were performed by filling the injector with a sample and then applying pressure with the injection pump for a few seconds. The pressure is then released and the injector flushed with mobile phase before finally applying the desired run pressure with the elution pump. This method produces a narrow injection plug directly on the head of the column and has a short delay between injection and elution, all important advantages over the gradient UHPLC system.

The biggest modification to the system was in the method of detection. Previous work utilized electrochemical detection; however, polymers and proteins are not electrochemically

active and better suited for detection by UV absorbance. Detection at 215 nm was accomplished with a Linear UVIS 200 (Linear Technologies, Milpitas, CA) absorbance detector with capillary flow cell. Since this work called for such small sized particles, the column was necessarily short (<15 cm) making it difficult to arrange a UV detection cell and the UHPLC injector in close proximity to each other. As a solution, the detector was mounted on a labjack vertical positioner, shown in Figure 4-6. This allowed the column to be mounted in the flow cell and then lowered into the UHPLC injector which was also oriented in the vertical direction. Even with this configuration, the injector, fitting and flow cell still required columns of at least 10 cm. Significant redesign to the injector and UHPLC fittings would be necessary to accommodate shorter columns.

4.2.2 Data Acquisition and Analysis

All data was acquired and saved at 21 Hz using a 0.1 sec time constant on the Linear UV detector. Detector gain was typically 0.001 AUFS, but could be varied depending sample concentration and peak height. Data was in analyzed using custom routines written in Igor Pro 5.1 (Wavemetrics Inc, Lake Oswego, OR). The number of theoretical plates (N), peak width at base (w_b, 4σ) and retention time (t_r) were calculated by the method of iterative statistical moments (ISM).[2, 15] This method requires a flat baseline, so a subtraction was first performed on either a 100-point moving average or a 3-term polynomial fit to the baseline region.

4.2.3 HDC Particles

The separation of proteins and small synthetic polymers requires particles <1 μm in order to produce through channels on the order required for the HDC mechanism to be predominant. Unfortunately, finding particles in this size range with a high degree of monodispersity was not possible. Instead, the hydrolysis of tetraethyl orthosilicate (TEOS), a

method known as the Stöber Process, was used to grow highly monodispersed sub-micron nonporous silica (NPS) particles.[16] This process is optimal for producing NPS particles in the 0.1-0.7 μm size range, an ideal size range for HDC of proteins. All synthesized particles were used without a hydrophobic stationary phase and consisted of only bare silica. Specific conditions for each synthesis batch follow in the appropriate sections, but Table 4-3 summarizes the dimensions used and the intended function. Protein HDC was performed on 0.3 and 0.5 μm particles.

Two different particles were used for HDC of polymers, also summarized in Table 4-3. Using a modified Stöber Process, 0.9 μm particles were synthesized in house. Commercially synthesized and C_{18} bonded 1 μm NPS particles obtained from Eichrom Technologies (formerly Mcira Scientific, Northbrook, IL) were also utilized as a benchmark for performance. These particles have been well characterized by our lab and shown to perform with high column efficiencies.[13, 14]

4.2.4 Column Fabrication

All columns were packed in fused silica capillary tubes obtained from Polymicro Technologies (Phoenix, AZ). A number of different internal diameter (i.d.) dimensions were considered, however, all work presented here utilized 75 μm i.d. x 360 μm o.d. capillaries. Previous work in by our group has shown that 30 μm i.d. was a favorable dimension for 1 μm particles, but also used electrochemical detection.[14] Since polymers and proteins are better suited for detection by UV absorbance, a larger i.d. was desired to increase the optical path length and maximize signal response. Additionally, the minimum i.d. of the column was limited to 75 μm by the diameter of the glass fiber used during frit fabrication. Larger i.d. columns up to 150 μm were briefly explored but found to be too fragile for routine work.

4.2.4.1 In-line Capillary Fritting

Standard fritting procedures previously employed by our lab used a plug of 2.5 μm NPS pushed into a capillary with a ~50 μm gap for electrochemical detection, then arced in place to hold the column bed.[17] This has the advantage of eliminating any post-column broadening that may occur before detection. A similar detection point was desired for use with the UV absorbance detector. Unfortunately, the configuration of the fittings on the capillary flow cell requires a minimum of 3 cm after the detection point to properly hold the column in the cell. This called for a method to integrate the frit into the column, but also leave a > 3 cm gap after the frit to allow for proper installation in the capillary flow cell.

The general fritting procedure is shown in Figure 4-7. A 67 μm o.d. fused silica fiber (Polymicro Technologies, Part # 2001596) was fed into a 75 μm i.d. capillary while viewing under magnification to create a "pusher" capillary. This outer 75 μm sleeve was necessary to prevent the fiber bending or breaking during use. The pusher was used to move a 2.5 μm NPS plug ~5 cm into a 75 μm i.d. capillary column, shown by Figure 4-7a. Several small plugs were needed to form the final frit and the pusher was used to pack these plugs together until the frit could no longer be moved inside the capillary. Once the frit was in a fixed position, the arcer was used to sinter the particles to the capillary walls and form the final frit for packing, shown by Figure 4-7b. A magnified image (4x objective) of a packed column with post-frit detection region is shown in Figure 4-8.

An inherent disadvantage of the arcing process is that the polyimide coating is removed from the arced section of the capillary, which can also be seen in Figure 4-8. This causes the capillary to become brittle and fragile, especially with a larger i.d. Teflon sleeves with a 380 μm i.d were used to protect the window during packing and installing in the UV flow cell.

4.2.4.2 Sub-micron Packing

Methods for packing capillary columns with particles ~1-2 μm have been well established by previous research in our lab.[11, 12, 14, 18] A few modifications were made to the procedure in order to adapt it to sub-micron particle packing and HDC detection. First, significantly higher particle slurry concentrations are needed due to the slower packing rate caused by the use of the sub-micron particles. Since volumetric flow, F, is proportional to d_p^2 (Equation 1.7), a 2-fold decrease in d_p results in a 4-fold decrease in flow, or increase in packing time, at a constant pressure. Higher slurry concentrations are therefore necessary to decrease packing time. Second, the particle synthesis was found to produce large particle agglomerates (shown in Figure 4-9) that were only noticed after column packing. Attempts at particle filtering were not successful because they reduced the slurry concentration too drastically, making column packing too time consuming. The settling method was instead used since it did not adversely affect the final concentration.

All unbonded particle (0.3, 0.5 and 0.9 μm) slurries were prepared in acetone pre-filtered through a 0.2 μm Teflon® syringe filter (Part No. 199-2020, Nalgene, Rochester NY). Slurry concentrations of 30-50 mg/ml, about a 10-fold increase over previous in-house procedures, were found to produce an acceptable packing rate while limiting aggregation. At this concentration, roughly 1 hr of sonication was necessary to fully deaggregate the particles in the slurry. Packing slurries were prepared from 2 x 1.5 ml slurries prepared in 2 ml Eppendorf centrifugation tubes at the target concentration and settled overnight. A volume of 750 μl from each slurry was decanted after settling and recombined to make the final packing slurry. Slurries were not stirred during packing to limit the introduction of larger agglomerates to the column. Bonded 1 μm Micra particles were prepared in 67% hexane/33% acetone at a concentration of 5 mg/ml, according to previous procedures.[13, 14]

Various methods of applying the pressure were explored, however, no noticeable differences in column performance were seen. Therefore, ~3,000 psi was first applied to start the bed packing and ramped to the final pressure of ~25,000 psi over 30 mins. Packing was initially monitored using a 100x oil immersion microscope, and the column then left to pack overnight. After packing was complete, the pressure was allowed to bleed off over a few hours via a high-pressure release valve while the bed was monitored with the microscope for any expansion. Final column lengths were ~15 cm for protein HDC particle columns, and >30 cm for polymer columns particles. Previous methods have then called for pre-pressurizing the column and making a temporary frit by electric heating to prevent the bed from expanding. This was attempted, but ultimately abandoned, because the heating tended to produce large gaps instead of a stable frit. Additionally, <50 μm of bed expansion was measured while watching the column with the microscope during the post-packing pressure bleed. This was deemed acceptable and the column was instead cut to the desired length, typically removing the last 3-5 cm packed, and allowed to dry overnight before fabricating inlet frits.

Protein HDC columns were inlet fritted by arcing a plug of 2.5 μm NPS into a ~50 μm gap at the head of the column. The gap was created by applying a low voltage arc for < 1sec, which tended blow particles out of the inlet. Attempts at directly fritting the packing particles were generally unsuccessful without the additional use of the 2.5 μm particles. Protein HDC columns were first flushed with ten dead volumes to DI water before flushing the column to the desired buffer mobile phase and allowing it to equilibrate overnight.

Polymer HDC columns were vented in a similar manor and then immediately flushed with THF mobile phase before fritting. The 0.9 μm particles were large enough to directly frit

using the arcing process and a 2.5 μm frit plug was not used. After fritting, columns were again flushed with THF before overnight equilibration.

4.3 SUB-MICRON HDC OF PROTEINS

4.3.1 Protein Samples and Mobile Phase Preparation

All proteins samples were obtained from Sigma (St. Louis, MO) and used without further purification. Since the Stokes radius of the protein can be related to molecular weight by Eq. 4.4, proteins with a wide MW range were needed to explore the HDC separation region. The four standard proteins: Myoglobin (Myo, 17.5 kDa), Ovalbumin (Ova, 43 kDa), bovine serum albumin (BSA, 67 kDa) and Thyroglobulin (Thy, 670 kDa)) are listed in Table 4-4a with the corresponding Stokes radius determined from previous work[9, 19, 20] The proteins span a 40-fold MW range, but only a 4-fold range in Stokes radius. Samples were prepared at a concentration of 5 mg/ml in the mobile phase and injected directly on the column.

Protein HDC was performed using 20 mM Borate Buffer at pH 9.0, prepared from Boric Acid (Fisher Scientific) in DI water and titrated to the desired pH with NaOH. High pH mobile phase was necessary to prevent protein-silica interactions since the protein HDC work uses only bare silica spheres in the packed column (see section 4.3.2). At this pH, most proteins (and all protein samples prepared here) are above the isoelectric point, and therefore negatively charged. Additionally, above pH 2, bare silica has a net negative charge from the unprotected silanol groups. Thus, the high pH limits protein-silica interactions that could alter the HDC mechanism.

4.3.2 0.3 and 0.5 μm Particle Synthesis

The Stöber Process[16] was used to prepare particles in the target range of 0.5 μm by adding 7 mL TEOS (Aldrich) dropwise to 21 mL DI H_2O (18 MΩ NanoPure) and 10 mL NH_4OH

(Fisher Scientific, Assay 29.4% NH_3) in 62 mL absolute EtOH for a total volume of 100 mL. This produced a solution of roughly 0.3 M TEOS, 15 M DI H_2O and 1.6 M NH_3 prepared from NH_4OH, but the final concentration was dependent upon excess water present in the NH_4OH and EtOH. The solution was allowed to react for 30 mins before being concentrated by centrifugation and washed with fresh EtOH. After the initial wash, the solution was concentrated and washed two more times before decanting the EtOH and drying the particles overnight in a vacuum oven at 80°C and a final pressure of 0.09 torr. Typically, this synthesis produced ~2 g of particles. Final particles and size distribution, shown in Figure 4-10, were 0.51 ± 0.03 μm.

Later studies of the Stöber Process generated more precise size distribution profiles with a wide range of starting conditions.[21] This data was used to synthesize particles in the 0.3 μm target range by reacting 1 M NH_3, 15 M H_2O and 0.17 M TEOS in 200 proof EtOH. For starting reagents used in this work, the reaction was 4 ml TEOS added to 7 ml NH_4OH and 28 ml H_2O in 61 ml EtOH. This synthesis produced particles of $d_{p,n}$ 0.33 ± 0.04 μm, shown in Figure 4-11.

4.3.3 Protein HDC Results

Two different particle dimensions, 0.3 and 0.5 μm d_p, were used to explore the potential of HDC as a separation mechanism for proteins in biological solvents. From Table 4-3, these particles produce a R_c of ~80-115nm, which is optimal for a wide range of proteins.

Data was first obtained using proteins <100 kDa (Myo & BSA) on the 0.5 μm HDC column. Unfortunately, no separation from the dead time marker was detected. This was primarily due to problems with column efficiency and low aspect ratios. The $\lambda_{0.5um}$ calculated in Table 4-4a is only 0.04, which would result in ~10 sec of separation from the dead time marker during a 2 min run (see Table 4-1). A poorly performing column could

easily mask any separation that would occur. Instead, separation of a very large protein (Thy, 660 kDa) was analyzed. The $\lambda_{0.5um}$ for Thy (Table 4-4a) is ~0.15, which corresponds to a much greater τ and therefore shorter retention time. The result of a separation of Thy from the dead time marker is shown in Figure 4-12a and separate injections of the protein and marker are overlayed in Figure 4-12b. While the column performance is rather poor, the HDC mechanism can clearly be seen.

In order to resolve proteins < 100 kDa, significantly greater column performance would be required, or smaller particles would be needed to decrease τ in order to use more of the separation space. Using 0.3 μm particles, HDC separations of three different proteins are shown in Figure 4-13. It can be seen that both Myo (A) and Ova (B) can be separated and resolved from the ascorbic acid dead time marker. Thy (C), however, appears to actually perform worse on the smaller particles, perhaps indicating that R_c is too small for such large proteins. From these results, it is evident that 0.3 um particles are much more effective at separating proteins < 100 kDa, as theory would predict, since the R_c has been reduced to 80nm.

Using Eq. 4.2 and 4.5, calibration plots for proteins, shown in Figure 4-5, can be generated and used to estimate the MW of the protein from its τ in HDC. Results of τ, the calibrated MW and literature values are summarized in Table 4-5a. The calibrated MW matches the literature values quite closely, indicating that the proteins are performing fairly ideally in HDC. Additionally, the rotational constant C from Eq. 4.2, the main variable in estimating τ, appears to be close to that of polymers. As noted, it can typically range from 1-5, and is dependent on the analyte of interest. Since proteins are more compact in solution than polymers tend to be, and the proteins evaluated in this work are globular in nature, it

seems likely that the behavior will mimic that of polymers. Also of note is the performance of Thy in Figure 4-13c. It is believed that this is evidence of the onset of HDC inversion for the column. The measured τ, 0.9, has no physical meaning, and the presence of the apparent broad peak after the marker may indicate column overloading or clogging from the channel size being too small. It is thus important to choose particle size based on the desired proteins to be separated.

While the HDC mechanism appeared to be performing as predicted by previous developed theory, the fundamental column performance was quite poor. The van Deemter curves for ascorbic acid, summarized in Table 4-6, were generated for both particle sizes. Several different columns were packed and characterized, but none were measured to have H_{min} < 11 μm (N≈10,000), a 30-fold difference from the theoretical performance.

Columns also tended to have a significant decrease in performance over time. This was typically traced to gaps (~500 μm - 1mm) in the bed that would form within the first cm of the column head. Gaps were also noticed in the region of the column normally obscured by the UHPLC fitting (~ 2-4 cm from inlet) once the column was removed from the fitting, however no gaps were found past ~ 4cm from the column inlet. This could be indicative of bed collapse caused by the UHP, however, very little bed movement (<50 μm) was typically seen while monitoring packing and pressure bleed. While these packing pressures were limited to 30 kpsi, the lack of significant detectable movement makes such large gaps unlikely. It is possible that the silica itself was becoming unstable from the use of high pH mobile phases, but gaps were sometimes noticed after <8 hours of use. Silica degradation over this relatively short time would not account entirely for the number or size of gaps found, indicating other factors are present. Attempts were made to keep the bed solvated by

not drying the column before fritting since it was noticed that small cracks (10 μm) formed readily in the dried column, however, fritting proved to be extremely tedious and no improvement in results were noticed.

Early attempts were also made at packing columns in water (instead of acetone) resulted in significant bed shrinkage as the column dried. A 150 μm i.d. capillary packed with 0.3 μm particles, end-on view shown in Figure 4-14, forms a ~7 μm gap between bed and wall after being completely dried. Since ~500 particles would fit across the diameter of 150 μm capillary, the particle-particle repulsion in water may be significant enough to cause the bed to be artificially swollen during packing. Similar effects were not seen with acetone packed columns, but it is important note that since the columns are ultimately used in aqueous mobile phase, undesired effects may result and reduce column efficiency.

With such poor performance, the potential column resolution (R_s, Eq. 4.7) is severally limited. An inherent disadvantage to HDC is that it requires high resolution and efficiency since the separation space is so narrow. Although HDC of proteins proved to be possible, packing of sub-micron particles required for HDC appears to be ineffective and needs more extensive studies.

4.3.4 Protein HDC Preliminary Conclusions

The goal of this experimental work was to evaluate the applicability of the mechanism of HDC for protein separations. It has been shown that the HDC mechanism does indeed apply to proteins when the correct particle dimensions are chosen. Practically, for proteins < 100 kDa, particles < 0.3 μm are required. If column efficiency is extremely high, slightly larger particles may prove effective, but the separation window is nonetheless limited to around $0.9t_m$ making complex sample analysis difficult. Additionally, it has been shown that basic HDC theory previously developed on larger model analytes, such as polymers, can be

applied to proteins with little modification. More extensive work is required to appropriately define the *C* value for proteins, but the tightly packed structure of a protein appears to make it behave fairly similar to a synthetic polymer. Until presumed difficulties with the packing of sub-micron particles can be further explored, HDC of proteins will remain interesting from a theoretical standpoint, but will not be practical with respect to proteomic applications. Also, the required use of particles <0.3 μm for most protein applications will continue to require custom instrumentation for the foreseeable future, further limiting the potential use. Finally, since this work was performed on bare NPS at non-native pH ranges, the use of a hydrophilic coating to prevent protein-silica interaction would be of interest in future work.

4.4 SUB-MICRON HDC OF POLYMERS

Results of sub-micron particle HDC of proteins was encouraging in terms of the HDC mechanism, but extremely poor with regards to column efficiency. In an attempt to gain more information about packing and performance characteristics of sub-micron columns, it is useful to explore larger particles as applied to HDC of polymers. Since, to our knowledge, these columns represent some of the smallest particles ever packed in capillary columns for HPLC, unknown side effects may be present during the packing that reduce column efficiency. Additionally, while proteins are of great interest, they are historically quite difficult to work with in regards to HPLC. Thus, a more predictable and characterized analyte, such as polymers, is useful when evaluating column performance.

4.4.1 Polymer Standards and Mobile Phase

Polymer standards were prepared from a Supelco (Bellefonte, PA) Polystyrene Standard Kit (Cat. No. 4-8938 and 4-8937). Sample weights are listed in Table 4-4b and span the range from 2.5 to 1800 kDa, with a 50-fold change effective radius (R_{eff}). Samples were

dissolved in THF at a concentration of 0.5 mg/ml and allowed to swell overnight before use. Toluene was used as a dead time marker for data shown.

Polymer HDC was completed in unstabilized THF (Fisher Scientific) to eliminate the high UV background present from the stabilization reagent. THF was stored under dry N_2 during the runs and was prepared fresh each day to limit the inherent dangers involved with using unstabilized THF.

4.4.2 0.9 μm Particle Synthesis

The 0.9 μm particles needed for polymer HDC had been previously synthesized in our lab by using a modified two-step Stöber Process. This synthesis has been previously described in detail by Mellors.[22] The synthesis involves first making 0.5 μm seed particles using the standard Stöber Process and then using these seeds to grow larger particles via a second Stöber reaction. The resulting particles used in this work were 0.92 μm ± 0.03 μm ($d_{p,n}$ SEM) and were used without further modification.

4.4.3 Polymer HDC Separations

4.4.3.1 0.9 μm Unbonded NPS Columns

Columns packed with 0.9 μm particles result in a R_c of ~ 200nm. This is optimal for polymers between 10^3-10^6 da, and results for several polystyrenes (PS 2.5, 17.5, 50, 110, 400 and 1800 kDa) separated using 0.9 μm particles is shown in Figure 4-15. The HDC mechanism can again easily be seen as the large polymers elute first, and in a predictable order. Results obtained at low linear velocities (<2 kpsi) had enough resolution to calculate τ for all six polymers analyzed, while faster velocities were only able to resolve the two largest polymers, PS 400 and PS 1800. Best-fits to the data (shown in Figure 4-18a) can be performed using Eq. 4.2 to evaluate how closely the column matches theory. The τ and C

value fits are summarized in Table 4-5b. The C values fall within the theoretical range of 1-5, however both differ from the accepted value of 2.7 for polymers.

Perhaps more interesting is the trend in τ for the largest PS samples, 400 and 1800. Previous work has reported that τ will begin to increase, indicating a decrease in R_{eff}, for very large polymers at high linear velocities due to polymer sheer.[8, 10] This data shows the opposite trend with τ decreasing as linear velocity increases, indicating that the polymer is effectively increasing in radius. The alternative is that PS 400 and PS 1800 are in the HDC inversion region. In this case, a decrease in R_{eff} would have the opposite effect on τ since the polymer would fall on the increasing quadratic side of the curve. Data generated using the 1 μm HDC column (see section 4.4.3.2 that follows) further indicates this is the likely scenario. More data points would be needed to accurately predict the C value and inversion region in order to determine which effect is actually occurring.

While the HDC mechanism appears to be behaving well, it is again clear that column efficiency is poor at all but the slowest velocities. The backpressure requirements for 0.9 μm particles in THF is roughly 20-fold less than the protein HDC columns run in aqueous conditions due to the larger d_p and reduced viscosity, leading to the potential for longer columns. Results reported here are using a 45 cm column, about 3x longer than those used for protein HDC. The resolution has therefore been improved compared to the protein HDC columns, but most of this improvement comes from the ability to use a longer column as H_{min} is essentially no different.

4.4.3.2 1 μm Micra ODSII Bonded NPS Columns

Columns prepared with 1 μm particles should behave fairly similar to the 0.9 μm particles used in the previous section for HDC. The main advantage of these particles is that our lab

has experience packing and successfully using them for UHPLC, eliminating several variables introduced by the previous particles.[13, 14]

The PS sample for this column was similar to that used on the 0.9 μm column, however to limit HDC inversion errors, PS 1800 was replaced with PS 900 and PS 220 was also added for a mixture of seven PS standards. Columns packed with these particles performed noticeably better than previous particles and several results are shown in Figure 4-16. Separations were carried out up to 25kpsi (u=0.25 cm/sec) and all polymers were easily resolved, most with R_s>1 indicating a good separation.

A van Deemter plot for several different polymers is shown in Figure 4-17. Polymers were typically measured to have N>15000, with the smallest polymers having N>30000. The H_{min} ranged from 6-12 μm, which was clearly an improvement over any performance previously measured. Even with these improvements, the H_{min} is still 3-4x higher than theory would predict. Some of this can be attributed to the inherent polydisperstiy of polymer samples, especially for the high-MW PS samples. Venema *et al.* estimated that the contribution to H from a high dispersity, high-MW (>800 kDa) polymer could account for a ~4-fold increase for a column such as the one used here.[8] No information on polydispersity was provided from the manufacturer, so only speculative conclusions can be drawn. Since it appears evident from Figure 4-17 that H is increasing with PS MW, it does seem likely that polydispersity is partially contributing to the decreased performance. The polydispersity present in PS 2.5, however, should be negligible and therefore provide a fairly direct measure of column performance, which still seems to be underperforming when compared to theory.

The trend that is seen for the B-term with the polymers is also of note. Diffusion of PS, estimated in Table 4-4b, in THF is rather slow, and decreases with increasing molecular

weight. As such, the B-term in the van Deemter equation should be greatly reduced since longitudinal diffusion is much slower. The results, however, show that the B-term is greater with larger MW samples. This reversed B-term trend for HDC of polymers has not previously been reported, and may instead be a side-effect of poor column performance as it seems unlikely that HDC could selectively increase the axial diffusion of high MW polymers. It should be finally noted that even though performance is less than desired, the separations show in Figure 4-15 are among the fastest examples of HDC of PS reported to date.

Similar to the analysis completed with 0.9 μm HDC columns, the τ values for each PS can be calculated and used to generate calibration plots. Since the column performance has been improved, τ for all seven PS can be calculated over a wide range of linear velocities. Values for u=0.05-0.25 cm/sec are shown in Table 4-5b, and plotted with corresponding C-fits in Figure 4-18b. When compared to τ for the 0.9 μm column, the values have all increased slightly, which is consistent with the larger R_c that results from 1 μm particles. C-values range from 2.4-4.5, and consistently increase with linear velocity. Further examination of the data shows that the higher linear velocities may be altering the C-value. Figure 4-19 shows the relative increase in τ for several different PS standards as u increases. It is evident that below 400 kda, τ is independent of the velocity. The 900 kDa standard, however, shows an increasing trend that follows u. At the highest velocities, τ has increased over 6%, indicating the PS is effectively getting smaller. This is now consistent with the polymer deformation trend that has previously been reported for large PS with HDC and also indicates that the C-values are being artificially skewed by this increase. A value of 2.5-2.7 is probably more

accurate, which follows theory quite closely. It also further indicates that the reverse trend seen with HDC of PS 1800 on 0.9 μm is likely a result of HDC inversion.

Overall, the 1 μm Micra column performed significantly better and more consistent with previously developed HDC theory. Unfortunately, an explanation as to the improvement is not straightforward. It is likely that the column is simply packed more efficiently, but it is difficult to identify the reason for improvement without further experiments. The C_{18} coating that is present on the 1 μm particles is an obvious modification, but other factors are likely present since the column still does not completely meet performance theory.

4.4.4 Column Evaluations

The drastic improvement in polymer separation efficiency when changing from a 0.9 μm bare NPS to a 1 μm ODS-bonded NPS packing material is difficult to identify. It was therefore useful to perform a fundamental analysis of column performance. Since polymers will artificially increase *H* from their inherent polydispersity, it is more useful to evaluate columns using a small organic molecule as a dead time marker. Toluene was chosen due to the high UV response and was used at concentration of 0.1% (v/v) in THF. Several different 0.9 μm columns were prepared from lengths of 20 to 45 cm. Additionally, one 0.9 μm column was packed in MeOH since the solvent has been successfully used for packing well performing 1 μm HDC columns.[8]

The van Deemter results for 0.9 μm and 1 μm columns are shown in Figure 4-20 and summarized by Table 4-6. It is again immediately evident that the 1 μm bonded-phase column is superior in performance to any of the 0.9 μm bare NPS columns. Typical H_{min} for 0.9 μm particles is ~12 μm while < 6 μm for the 1 μm particles. This is similar to results seen from polymers and also indicates that the polydispersity of PS 2.5, the smallest PS used,

does not contribute significantly to H. Of greater interest is the high C-term for 0.9 μm columns and the measured u_{opt} for both columns.

For the 0.9 μm columns, the observed C-term is not only significantly higher than predicted for such particles, but appears to be increasing with u, further decreasing column performance. This helps explain the poor resolution that is seen at higher pressures with the 0.9 μm columns, but does not have an obvious cause other than a further indication of poor packing structure. The 1 μm column behaves closer to what theory would predict, but more work is needed to evaluate the difference in C-terms between the two columns.

The measured value of u_{opt} is also of note. The diffusion coefficient (D_m) for toluene in THF can be estimated from its viscosity and known coefficients for similar molecules[23] to be ~2.5×10^{-5} cm^2/sec., a fairly high value for UHPLC. With this value of D_m, the u_{opt} is predicted to be ~0.6 cm/sec for 1 μm particles and 0.8 cm/sec for 0.9 μm. Additionally, the B-term (also plotted from theory in Figure 4-20) should affect column efficiency at low linear velocities due to the high longitudinal diffusion. The u_{opt} has been measured at ~0.1 cm/sec for the 0.9 μm particles, a 6-fold decrease from the predicted value, and the region where B-term effects are dominant. The artificially high C-terms for 0.9 μm columns could skew the u_{opt} to lower values, but the 1 μm column shows a similar decrease in u_{opt} when compared to theory. Interestingly, this has been previously reported in earlier work where the u_{opt} was measured to be at an even lower value than ones measured here.[8] This indicates that the columns used here are partially consistent with HDC results, even if not following van Demeter theory.

Overall the column evaluation further indicates that 0.9 μm particles do not pack or perform well with current methods. The 1 μm particles provide a benchmark for the system

and indicate that small inconsistencies in packing parameters could lead to large decreases in performance. Finally, the 1 μm particles still seem to be underperforming in terms of van Deemter theory and more work is needed to explore both 0.9 μm and 1 μm particle HDC using UHP.

4.4.5 Polymer HDC Conclusions

Polymer HDC was performed using polymer standards in the $10^3 – 10^6$ da range, an appropriate range for ~1 μm packed bed HDC. The HDC mechanism appeared to behave similarly to previously developed theory for polystyrenes. For currently unknown reasons, 1μm NPS bonded with a C_{18} stationary phase performs at a greater than 2-fold improvement over the 0.9 μm bare NPS. These bonded particles allow seven PS standards to be resolved in < 2min using UHP, some of the fastest HDC separations seen to date. Even with this improvement, the 1 μm particles are still underperforming in terms of column efficiency. It is hypothesized that the improvement of the 1 μm particles over the 0.9 μm comes from the ability to pack them more efficiently, but the exact cause of this improvement in unknown. Further studies in packing particles of this dimension will be required.

4.5 PERFORMANCE CONSIDERATIONS

Due to the poor theoretical performance of all columns explored, it is important to consider a number of factors that could contribute to band broadening and artificially increase *H*. The most obvious are pre- and post-column effects. This injector has been well characterized and pre-column effects have not been shown to be significant.[2, 13] The detection method used here is different than previous UHPLC methods and should be considered. Additionally, factors including packing efficiency and solvent compression

could adversely affect performance. Three factors of band broadening: post-column effects, packing efficiency and solvent compression, will therefore be briefly considered.

4.5.1 Post-Column effects

Standard UHPLC detection methods utilizing electrochemical detection do not significantly contribute to *H*, however, the UV method used here could lead to artificially high *H* and must be considered.

While the in-line, post-column frit is advantageous for maximizing sensitivity, it will lead to extra-column broadening. First, the detector itself has a finite volume that will add to the overall peak variance. Detector contribution to *H* can be calculated from the temporal variance contribution of a given detector by[24]:

$$H_{\text{det}} = \frac{\sigma^2_{t,\text{det}} u^2}{L} = \frac{\frac{1}{12}\left(\frac{V_{eff}}{F}\right)^2 u^2}{L} \qquad \textbf{(4.8)}$$

where V_{eff} is the volume of the detector, F is volumetric flow rate through the detector and L is the column length, including detection region. The Linear UV detector used has a 100 μm slit for detection, and when used with a 75 μm o.t. capillary, the detector volume (V_{eff}) is ~440 pl. Accounting for the decrease in linear velocity from the packed to o.t. region, the contribution to *H* for this detector is ~0.02 μm, clearly an insignificant amount that can be ignored for practical purposes.

An additional cause of broadening is that of the open tube region itself that immediately follows the frit. The *H* from the flow cell can be calculated from the Golay equation as:

$$H_{cell} = \frac{d_c^2 L_{\text{det}} u_{\text{det}}}{96 D_m L} \qquad \textbf{(4.9)}$$

where d_c is the column diameter, D_m is the diffusion coefficient of the analytes in the mobile phase, u_{det} is liner velocity in the detection region and L_{det} is the length between the column

and the detector. While the volumetric flow, F, remains constant throughout the run, the linear velocity will decrease by ~60%, equivalent to $1-\varepsilon_i$, which will reduce H. Additionally, this distance was typically < 1mm, to minimize this contribution. For toluene, H_{cell} was < 0.2 μm at all but the highest flow rates, which is less than 3% of the measured H for the column. This value is low enough to be considered insignificant at all but the highest flow rates. It may partially explain the discrepancy in measured vs. theoretical u_{opt} since the contribution will become even greater as u_{opt} is approached, but it clearly does not account for most of the performance decrease that was observed.

The contribution from H_{cell} for polymers and proteins compared to toluene, however, is theoretically an order of magnitude higher due to the low D_m of these large molecules. Values of D_m for the polymers analyzed are calculated from previous studies and shown in Table 4-4b.[25] There is roughly a 30-fold decrease in D_m from PS 2.5 to PS 900, which would account for a similar increase of H_{min}. Typical H_{cell} for PS 2.5 were <0.5 μm, while H_{cell} for PS 900 was ~5 μm for the highest flow rates, indicating that the contribution is significant at these high velocities and MW. This gives an additional explanation as to why the higher MW PS samples consistently gave worse results, as seen in Figure 4-17.

Finally, the response of toluene is high enough to detect on column, which would eliminate all of the post-column effects discussed above. Measurements of toluene through the packed-bed differed by < 1 % from post-frit measurements, indicating that the post-column detection method is performing well and not contributing greatly to H for toluene. PS does not have a high enough absorbance in the concentrations used, and was only detected by the post-column method.

Overall, post-column effects add an insignificant amount to H for toluene, indicating that the difference from expected values is caused by other factors. Additionally, the low D_m for PS may account for some of the increase in H and gives additional evidence as to why smaller standards perform more ideally, but does completely account for the discrepancy.

4.5.2 Column Packing Effects

The parameters used to pack a column can have a drastic effect on the column performance. Both the column packing density and aspect ratio have been shown to be important in overall column performance. A brief discussion of these factors is useful to evaluate possible column packing effects that could degrade performance.

While no direct measurements of packing density were performed, it is possible to estimate the relative packing density between columns used with empirical data. The column flow resistance factor, Φ, calculated by[26]:

$$\Phi = \frac{\Delta P d_p^2}{uL\eta\varepsilon_i} \qquad \textbf{(4.10)}$$

where η is solvent viscosity, can be used to estimate column packing density. Ideally, ε_i and P would be known in calculating Φ, but these values can only be estimated for this system without additional calibration procedures. Instead, data was collected at the same applied pressure for various columns. It is therefore accurate to compare Φ between various columns used in this work, but more difficult to compare these columns to others in the literature. Values for Φ are shown in Table 4-6. It is immediately evident that the 0.3 μm column has a significantly higher resistance factor than the 0.9 and 1 μm columns. Additionally, the lowest Φ measured was on the best performing column. This may indicate that packing structure is indeed a factor that needs to be optimized for HDC and that small a change in the bed structure can drastically reduce column performance.

A second factor that has been proposed to account for packing effects is that of the column aspect ratio, or the column wall effect. Knox was the first to recognize that the aspect ratio, $\rho = d_{colum}/d_p$, played an important role in the column performance. High values of ρ would degrade performance due to an inhomogeneous structure across the bed, which results from particles packing more densely along the column wall.[27] The use of sub-micron particles will result in extremely high values of ρ which could decrease column performance. For the columns used in this work, values of ρ are shown in Table 4-6. A generally accepted optimum ρ is ~6-8, significantly lower than that of the columns used for sub-micron HDC. While high values of ρ have been shown to decrease performance, Patel found that there was only a 2-fold increase in H over a ρ of 10-150.[14] The high ρ used in this work is a point of interest for future work, but cannot fully account for the drastic difference in performance that is seen between columns.

4.5.3 Solvent Compression Effects

Work in our lab by Jerkovich showed that compression of the mobile phase from application of the UHP in our UHPLC system accounted for as much as a 50% increase in the measured C-term band broadening at pressures up to 90kpsi.[28, 29] This occurs because the analyte initially experiences a surge in linear velocity as the mobile phase compresses, which artificially increases to the overall measured H. The solvent used in that work was 90/10 water/acetone. That solvent is about half as compressible as the THF mobile phase used for polymers, indicating that the run pressures used here could exhibit similar increases.[30] The high compressibility, therefore, is a possible source of the C-term behavior that was seen when using the 0.9 μm particles. It was shown by Jerkovich that the compression effect is greatest in the first 15% of the column. If the column performance were measured for the final 75% of the column, it would be expected to perform better than the column overall

since the broadening from compression would already have occurred. It is possible to create a "virtual" column by taking simultaneous measurements at two points along the column and then calculating the *H* for the virtual column. A similar concept has been used previously by our lab to account for pre- and post- column effects.[13, 14] Calculation of *H* for the virtual column can be described by:

$$H_{virt} = \frac{\left(\sigma_B^2 - \sigma_A^2\right)L}{\left(t_{r,B} - t_{r,A}\right)^2} \qquad \textbf{(4.11)}$$

where σ_A and σ_B are temporal variances for detection points on the column and t_r is the retention time and those points. This method was used to evaluate possible compression effects with THF on the 0.9 μm HDC column.

4.5.3.1 Instrumentation Modifications

To measure potential compressibility effects, a second UV detector was added to the UHPLC setup shown in Figure 4-6. Detector A was placed at 11.8 cm and detector B at 43.3 cm on a 45.5 cm column packed with 0.9 μm bare NPS. The virtual column was from point A to B, and was 31.5 cm in length. The UV response of toluene allows detection to be achieved through the packed bed without the need for a special frit. The column was only modified by arcing a small section to remove the polyimide before placing the column in the UV detector.

4.5.3.2 Results

Data for the bi-point detector setup was obtained at run pressures from 1-30kpsi. Results are shown in Figure 4-21. Several interesting features can be seen from this data. The effects of compression seem to be a major factor in the overall curve. This is evident because position A is performing noticeably worse than position B. The C-term is following a similar increasing trend with pressure to the data obtained in Figure 4-20, and appears to be even

more severe at position A. This increase in C-term is likely due to compression. As the analyte moves from position A to position B, the contributions to *H* from compression are no longer occurring and only contributions from the column in that region are present. Measurements at position B show an overall decrease, but also still contain the initial contributions from the velocity surge. A decrease indicates the column is performing better in this region and the surge artifacts are essentially being averaged out. Using Eq. 4.11 , the performance of the column in between the two detectors can be used to calculate only the contributions to *H* from this region. The overall effect, shown in Figure 4-21, is that the van Deemter curve for the virtual column has decreased when compared to either position. The calculated improvement is better than values obtained for the full 45.5 cm HDC column, but still not as good as the column packed with 1 μm ODS bonded particles.

Effects of compression are more directly evident by examining the linear velocities at the various points along the column. At a constant pressure, linear velocity should be constant through the length of the column. Isobar lines are also plotted on Figure 4-21 for the various data sets. In the absence of mobile phase compression, the isobar lines would be vertical as linear velocity would be constant at all points along the column. Instead, the slope of the isobar lines indicates that the mobile phase is moving faster at position A than position B; additional evidence of compression.

Jerkovich proposed the concept of a static injector in which the rapid pressurization that is present from our UHPLC system could be eliminated. Such a system would be of use here as well, but would require major design changes to the current system. His work also briefly explored improvements by applying a controlled pressure ramp, and found that this was

moderately successful. A similar approach may be necessary when working with highly compressible solvents like THF.

While compression of THF seems a likely cause of high values of *H* for the HDC columns explored by UHPLC, it should be noted that the performance of the 1 μm bonded column was still superior to compression-correct data that was collected. Additionally, the increasing C-term that is seen for the 0.9 μm columns is not evident from the limited 1 μm data collected. This is yet further evidence that other factors are contributing to the poor column performance.

4.5.4 Summary

Several different factors that could contribute to poor column performance were considered. While no one factor was an obvious reason for the decrease in performance, several pointed to potential work for further research. First, the contribution to *H* from the on-column frit becomes more significant with larger proteins and polymers and could account for some of the differences between samples that were measured. An on-column method of detection or more efficient way to minimize distance from column outlet to detector would be useful to eliminate these contributions. Next, it appears that the 1 μm column was packed much less densely than the 0.3 and 0.9 μm columns. How this improves column performance can only be speculated, but nonetheless is a likely cause for the difference in column performance. Additionally, future work to minimize the column aspect ratio by using smaller i.d. capillary columns could improve column performance based on earlier research. Finally, there is strong preliminary evidence that the high compressibility of THF will be problematic for UHP-HDC using the current instrumentation. Future designs would ideally incorporate a static injector in which the pressure would be held constant in order to minimize injection compression artifacts.

4.6 SUB-MICRON HDC CONCLUSIONS

Hydrodynamic chromatography has been performed using a variety of column parameters at pressures up 30 kpsi. This work has shown early evidence that HDC of proteins using NPS <0.5 μm is possible for samples < 100 kDa, a target range of many proteomic applications. It is evident that HDC of proteins is extremely fast with separation occurring in under 4 minutes. Column performance, however, is quite poor, especially considering the small size of packing material utilized and will need to be optimized before the technique can have any useful applications.

HDC of polymers at ultrahigh pressures additionally showed that separations could be carried out faster than previously reported methods. It was found that 1 μm particles bonded with ODS performed significantly better than 0.9 μm bare silica particles. Reasons for this improvement are not immediately identifiable; however, it is likely the packing methodology and structure of the bed are simply more efficient. More work is needed to understand the packing of sub-micron particles as it appears column performance can drastically suffer when using particles in this size range without proper packing conditions.

4.7 REFERENCES

(1) Shen, Y.; Smith, R. D. *Electrophoresis* **2002**, *23*, 3106-3124.\

(2) MacNair, J. E. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 1998.

(3) Sousa, J. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2004.

(4) Small, H. *Journal of Colloid and Interface Science* **1974**, *48*, 147-161.

(5) Small, H.; Langhorst, M. A. *Analytical Chemistry* **1982**, *54*, 892A-894A, 896A-898A.

(6) McHugh, A. J. *Critical Reviews in Analytical Chemistry* **1984**, *15*, 63-117.

(7) Tijssen, R.; Bos, J.; Van Kreveld, M. E. *Analytical Chemistry* **1986**, *58*, 3036-3044.

(8) Venema, E.; Kraak, J. C.; Poppe, H.; Tijssen, R. *Journal of Chromatography* **1996**, *740*, 159-167.

(9) Wasyl, Z.; Luchter, E.; Bielanski, W., Jr. *Biochimica et Biophysica Acta, Protein Structure* **1971**, *243*, 11-18.

(10) Kraak, J. C.; Oostervink, R.; Poppe, H.; Esser, U.; Unger, K. K. *Chromatographia* **1989**, *11/12*, 585-590.

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

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

(13) Patel, K. D. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2001.

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

(15) Hsieh, S.; Jorgenson, J. W. *Analytical Chemistry* **1996**, *68*, 1212-1217.

(16) Stoeber, W.; Fink, A.; Bohn, E. *Journal of Colloid and Interface Science* **1968**, *26*, 62-69.

(17) Hoyt, A. M. J.; C., B. S.; Larmann J. P., J.; Jorgenson, J. W. *Journal of Microcolumn Separations* **1993**, *5*, 325-330.

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

(19) Cabre, F.; Canela, E. I.; Canela, M. A. *Journal of Chromatography* **1989**, *472*, 347-356.

(20) Wong, P.; Barbeau, A.; Roses, A. D. *Analytical Biochemistry* **1985**, *146*, 191-198.

(21) Bogush, G. H.; Tracy, M. A.; Zukoski, C. F. I. *Journal of Non-Crystalline Solids* **1988**, *104*, 95-106.

(22) Mellors, J. S. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2005.

(23) Chang, P.; Wilke, C. R. *Journal of Physical Chemistry* **1955**, *59*, 592-596.

(24) Sternberg, J. C. In *Advances in chromatography*; Giddings, J. C., Kelleher, F. M., Eds.; M. Dekker: New York, 1965, pp 248.

(25) Schimpf *Journal of Polymer Science, Part B* **1989**, *27*, 1317-1332.

(26) Knox, J. H. *Journal of Chromatographic Science* **1980**, *18*, 453-461.

(27) Knox, J. H. *Analytical Chemistry* **1969**, *41*, 1599-1606.

(28) Jerkovich, A. D., University of North Carolina, Chapel Hill, NC, 2003.

(29) Jerkovich, A. D.; Mellors, J. S.; Thompson, J. W.; Jorgenson, J. W. *Analytical Chemistry* **2005**, *77*, 6292-6299.

(30) Abdel-Azim, A. A. A.; Munk, P. *Journal of Physical Chemistry* **1987**, *91*, 3910-3914.

4.8 TABLES

Table 4-1: Retention times for various HDC Parameters (λ,τ) at various column dead times, t_m.

HDC Parameters		Particle t_r for Various Column Dead Times		
Lambda	Tau	t_m= 2 min	5 min	10 min
0.02	0.96	1.93	4.81	9.63
0.05	0.91	1.83	4.57	9.15
0.1	0.85	1.71	4.26	8.53
0.15	0.81	1.61	4.03	8.07
0.2	0.77	1.55	3.87	7.74

Table 4-2: Selected proteins and corresponding Stokes Radius in water. Data used in estimating Eq. 4.5 fit parameters. Values from Wong *et al*.[20] except as noted.

Protein	Molecular Weight (Da)	Stokes Radius, R_a (Å)
ribonuclease A	13,700	16[a]
chymotrypsinogen A	25,000	20[a]
ovalbumin	43,000	30
albumin (BSA)	67,000	35
aldolase	158,000	48
catalase	232,000	52
ferritin	440,000	61
thyroglobulin	660,000	86

[a] from Cabré *et al.*[19]

Table 4-3: Summary of particle dimensions used in polymer and protein HDC.

Application	Synthesis Method	$d_{p,n}^{(SEM)}$ (μm)	R_c (nm)[b]
Protein HDC	0.3 μm One-step Stöber	0.33 ± 0.04	80 ± 11
Protein HDC	0.5 μm One-step Stöber	0.51 ± 0.03	115 ± 7
Polymer HDC	0.9 μm Two-step Stöber[a]	0.92 ± 0.03	200 ± 6
Polymer HDC	1 μm Micra Commercial	1.01 ± 0.06	222

[a] Previous performed by Mellors.[22]
[b] Calculated from Eq. 4.3.

Table 4-4a: Stokes Radius (R_a) and corresponding Aspect Ratio (λ) for proteins and particles used in HDC experiments.

Protein	Sigma Stock Number	Molecular Weight (kDa)	Stokes Radius (Å)	$\lambda_{0.5um}$ *	$\lambda_{0.3um}$ *
Myoglobin	M-0630	17	20.8[a]	0.04	0.05
Ovalbumin	A2512-250	43	27.3[a,c] (30)[b]	0.05	0.07
Albumin (BSA)	A0281	67	35.5[a,b,c]	0.06	0.09
Thyroglobulin	T1001	660	86[a,b]	0.15	0.22

[a] from Cabré *et al.*[19]
[b] from Wong *et al.*[20]
[c] from Wasyl *et al.*[9]
* by convention, $2R_a$ is used when calculating λ from R_c

Table 4-4b: Effective radius of polystyrene samples in THF

Polymer	Molecular Weight (kDa)	Effective Radius, R_{eff} (Å) [a]	$\lambda_{0.9um}$	$D_{m,}$ [b] $cm^2/sec \times 10^6$
PS 2.5	2.5	12.3	0.01	4.43
PS 17.5	17.5	38.5	0.02	1.48
PS 50	50	71.4	0.04	0.80
PS 110	110	113	0.06	0.51
PS 220	220	170	0.09	0.34
PS 400	400	242	0.12	0.24
PS 900	900	391	0.19	0.15
PS 1800	1800	587	0.29	0.10

[a] Calculated for polystyrene in THF from Eq. 4.4.
[b] Calculated for PS in THF from Scimpf, *et al.*[25]

Table 4-5a: Calibration results for protein HDC.

Protein	Molecular Weight (kDa)	$\tau_{0.5um}$	Calibrated MW	$\tau_{0.3um}$	Calibrated MW
Myoglobin	17	*(0.94)*[*]	--	0.92	13
Ovalbumin	43	--	--	0.89	49
Albumin (BSA)	67	*(0.92)*[*]	--	--	--
Thyroglobulin	660	0.85	640	0.92[**]	--

[*] Theoretical; No separation detected.
[**] Possible HDC inversion

Table 4-5b: Calibration data for polymer HDC.

d_p	ΔP (kpsi)	u (cm/sec)	C-fit	τ for PS Standards							
				2.5	17.5	50	110	220	400	900	1800
0.9	1	0.01	3.2	0.990	0.963	0.931	0.896	--	0.836	--	0.764
	2	0.02	2.3	0.990	0.963	0.932	0.895	--	0.824	--	0.716
	6	0.05	**	*	*	*	*	--	0.806	--	0.693
	10	0.08	**	*	*	*	*	--	0.794	--	0.726
1	5	0.05	2.46	0.991	0.970	0.944	0.910	0.868	0.839	0.781	--
	10	0.10	2.53	0.989	0.968	0.945	0.906	0.867	0.833	0.778	--
	13	0.13	2.92	0.989	0.969	0.941	0.907	0.868	0.835	0.785	--
	17	0.17	3.20	0.990	0.967	0.942	0.908	0.869	0.836	0.791	--
	20	0.20	3.44	0.993	0.976	0.947	0.913	0.874	0.842	0.803	--
	22	0.22	3.61	0.994	0.980	0.954	0.916	0.880	0.844	0.812	--
	25	0.25	4.79	0.991	0.980	0.956	0.921	0.875	0.847	0.835	--

[*] Peak resolution too low for accurate τ fit.
[**] Not enough data points for τ fit.

Table 4-6: Summary of column performance parameters for ascorbic acid and toluene, used to characterize protein and polymer HDC columns, respectively.

Column	Packing Material	$H_{min, ther}$ (μm)	H_{min} (μm)	U_{opt} (cm/sec)	Φ **	ρ **
0.5 Protein HDC	0.5 μm NPS	1	30	0.05-0.2*	--	145
0.3 Protein HDC	0.3 μm NPS	0.6	11	0.3	1100	227
0.9 Polymer HDC	0.9 μm NPS	1.8	11	0.05	530	82
1 μm Polymer HDC	1 μm NPS, ODS	2	5.5	0.15	494	75

* No detectable curve minimum or optimal point.

** Φ and ρ are column packing density and aspect ratio, respectively, discussed in Section 4.5.2

4.9 FIGURES

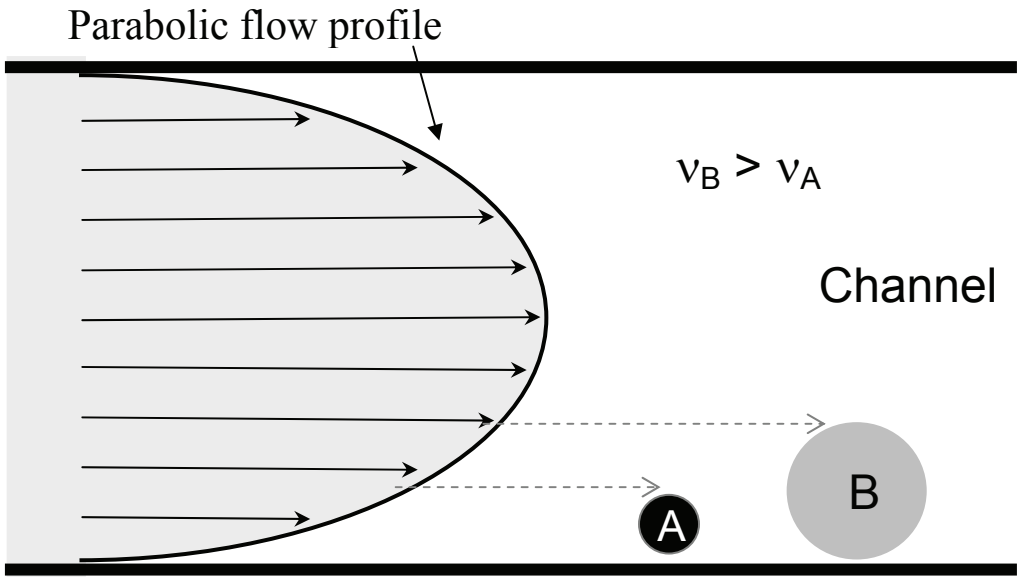


Figure 4-1: Schematic of HDC mechanism. Parabolic flow results from laminar flow through a narrow channel. Larger particles are excluded from channel walls and penetrate deeper in laminar flow, resulting a faster linear velocity and shorter elution time. In general, larger particles elute first.

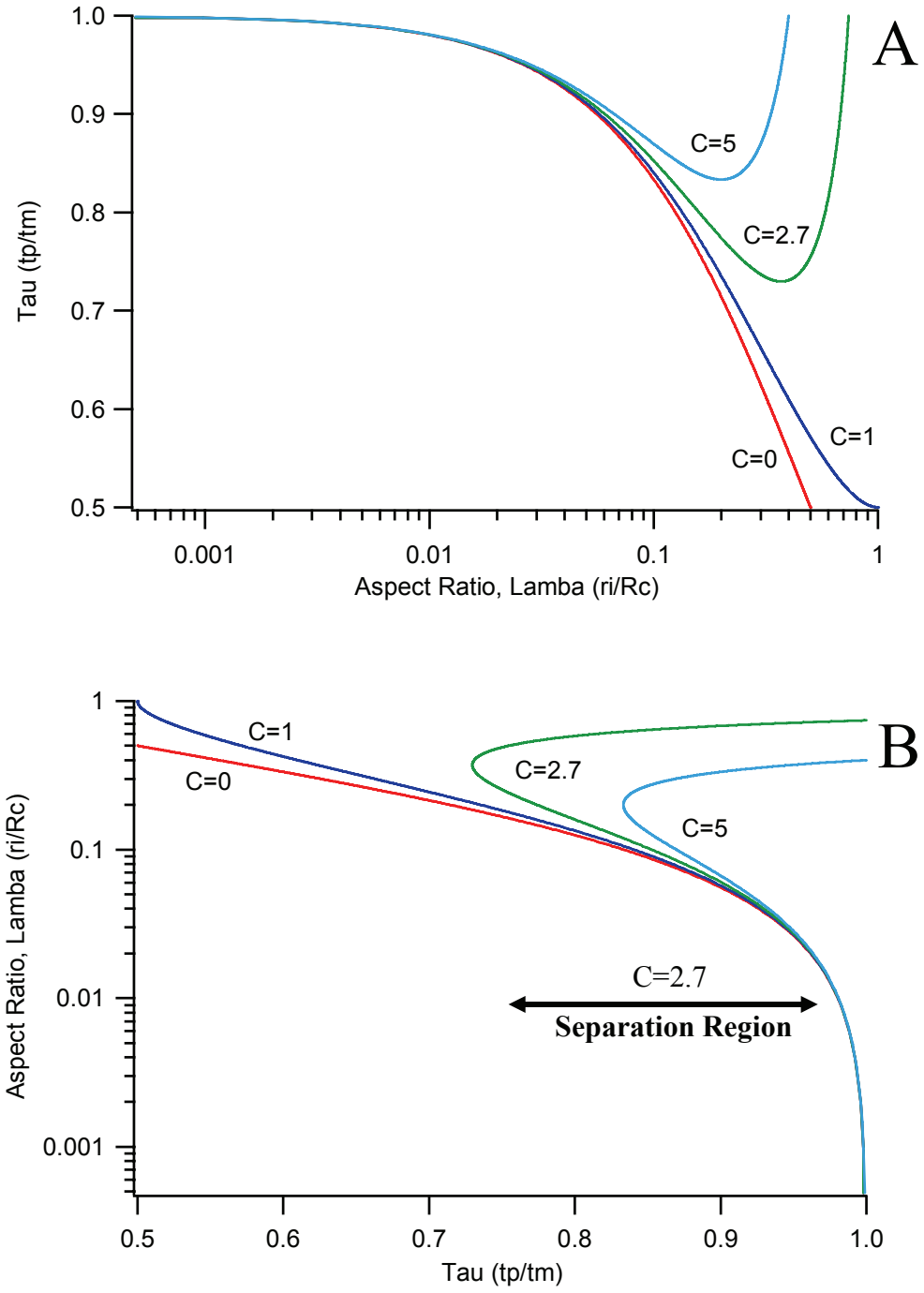


Figure 4-2: Theoretical plots of migration times in HDC. A) Plot of $\tau(\lambda)$. B) Swapped axes "calibration plot", chosen by convention, showing $\lambda(\tau)$. Both demonstrate that, for C=2.7, the useful separation region is τ =1-0.75 and λ= 0.01-0.4

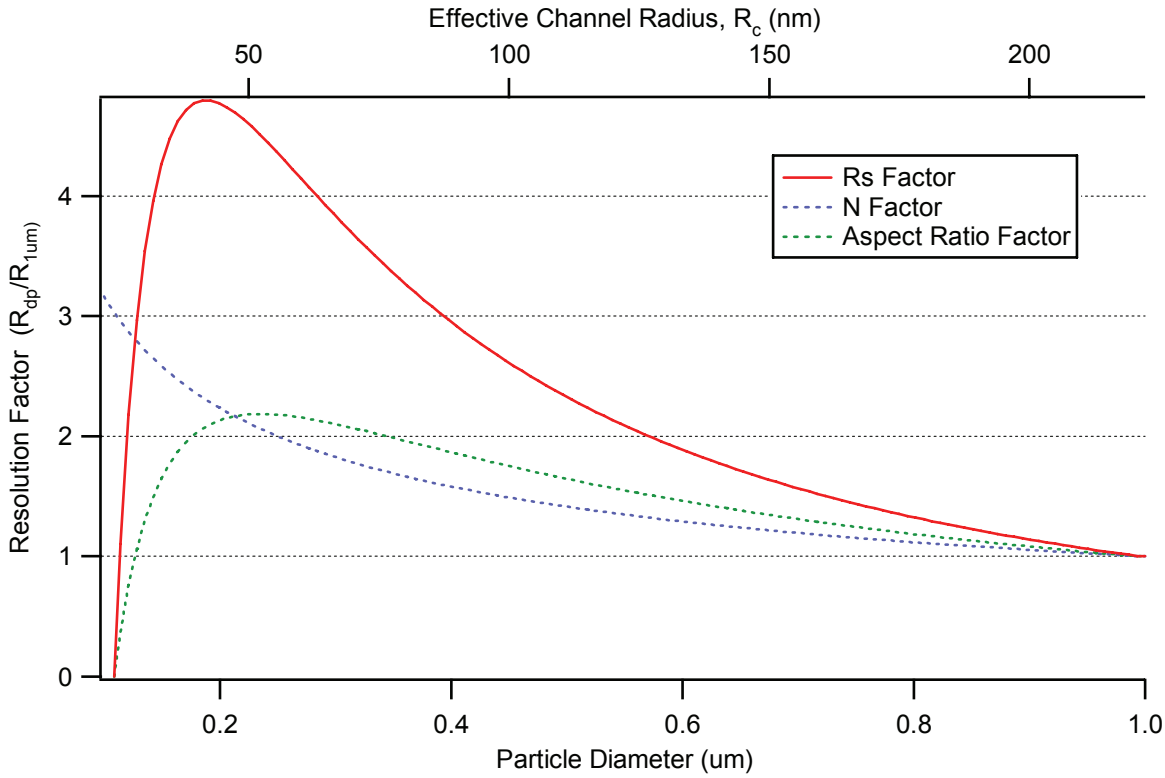


Figure 4-3: Plot of theoretical improvement in resolution (R_s Factor) between 100 Å and 90 Å particle as d_p decreases from 1 µm to 0.1 µm. Contributions of N and λ plotted along with R_s. Assumptions: C=2.7, H = H_{min}.

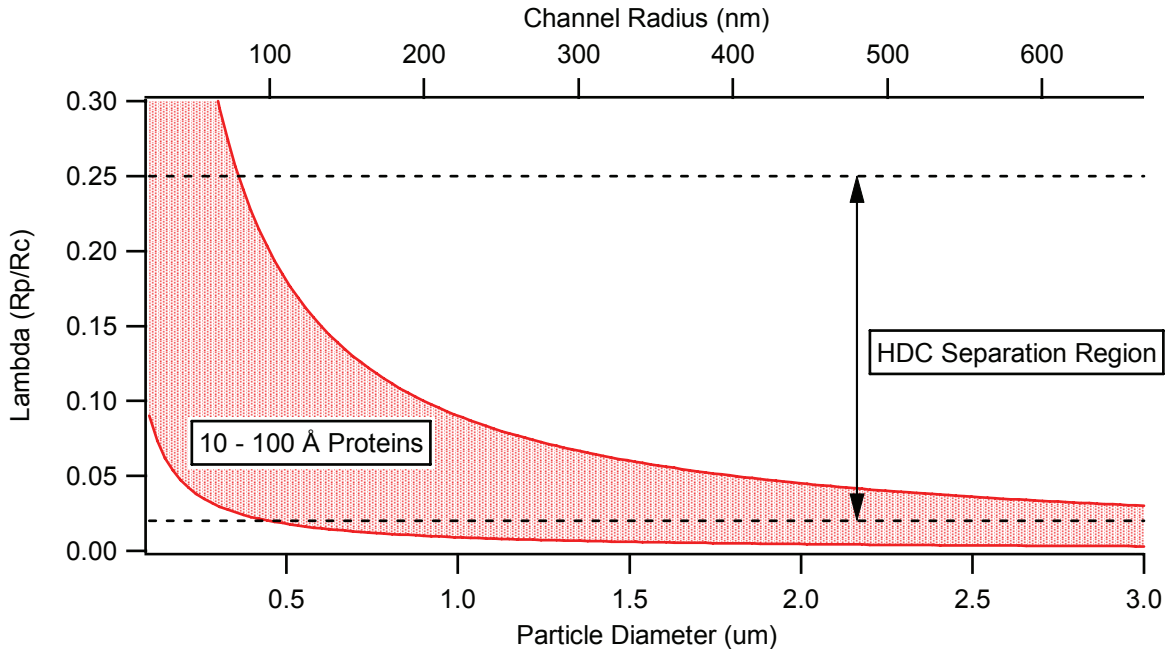


Figure 4-4: Theoretical range of lambda for 10-100 Å proteins in packed column HDC. $\lambda<0.02$ results in essentially no separation and above 0.25 gives HDC inversion. Choosing particle diameter to maximize λ range over the HDC region gives the best separation.

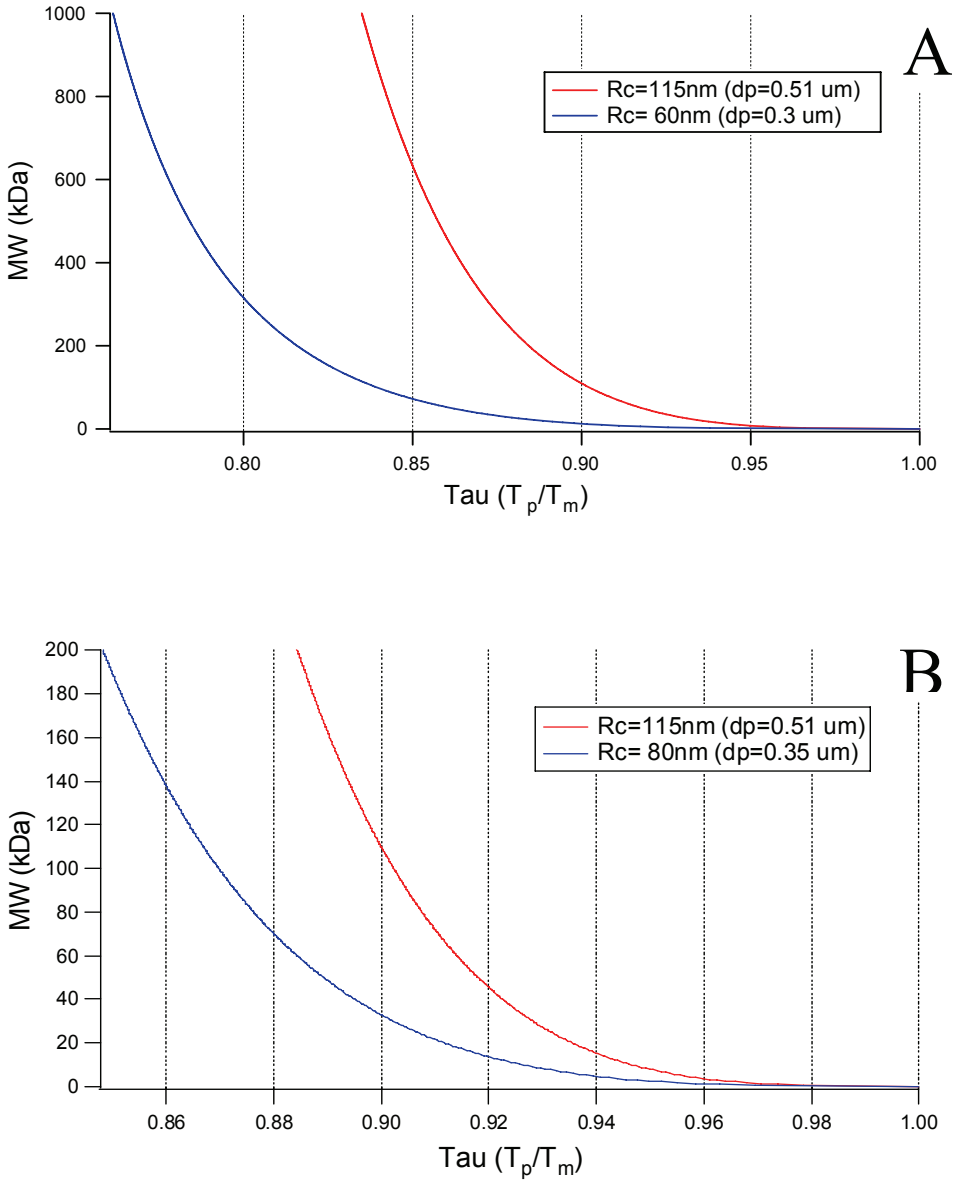


Figure 4-5: Calibration plots for proteins over various MW ranges for 115, 80 and 60 nm channels. Calculated from Eq. 4.2, and 4.5. A) 0-1000 kDa MW range with 115 and 60 nm channels. B) 0-200 kDa range with 115 and 80 nm channels. Plots are only shown before HDC inversion would occur. C= 2.7 in fits.

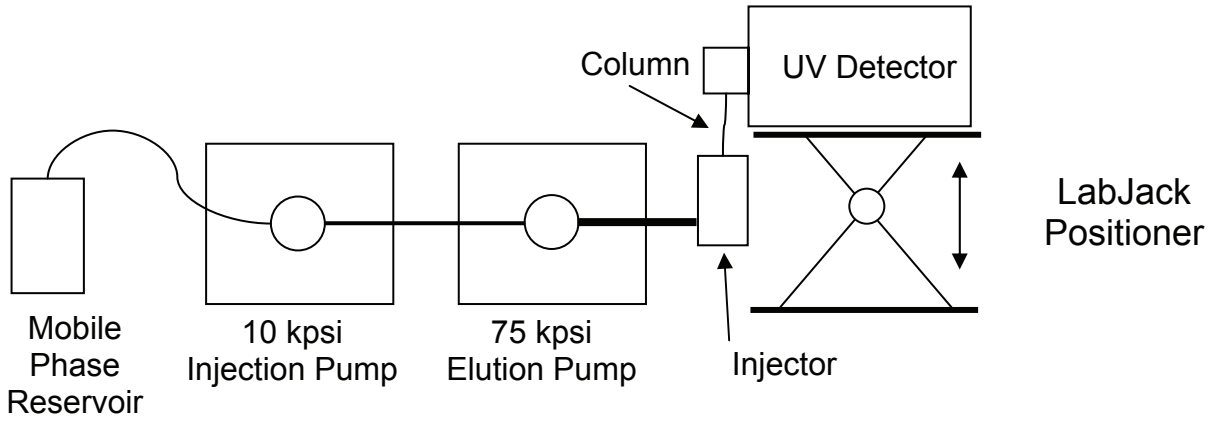


Figure 4-6: Schematic of UHPLC Instrumentation layout for short, on-column UV detection.

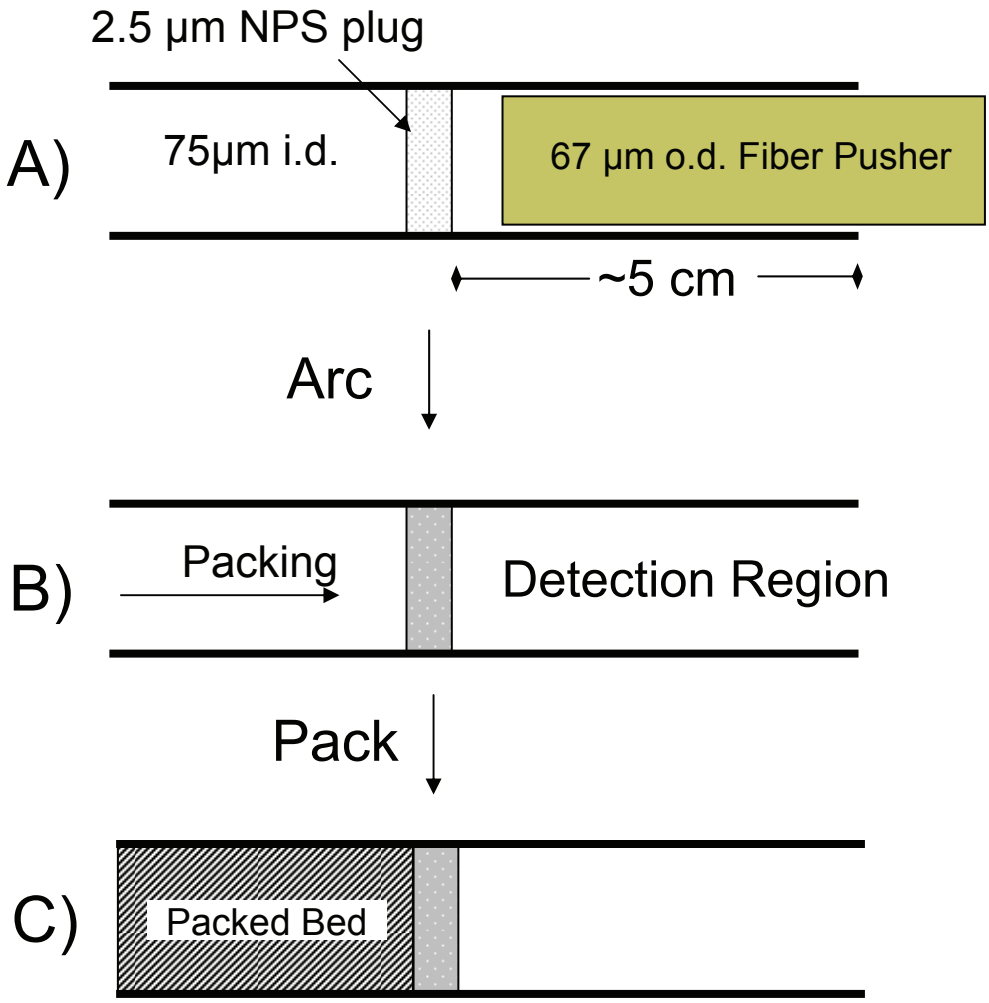


Figure 4-7: In-line capillary fritting procedure diagram.

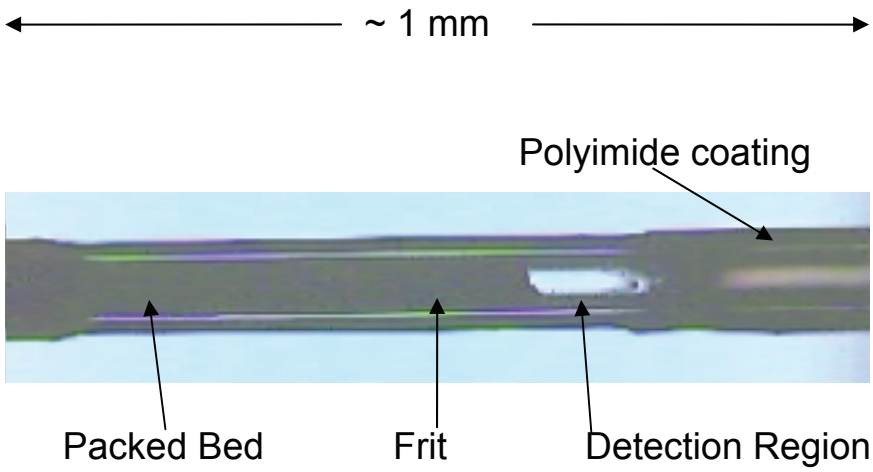


Figure 4-8: View of in-line capillary frit and post-frit detection region. Image taken through a 4x objective microscope lens and shows ~ 1mm of column.

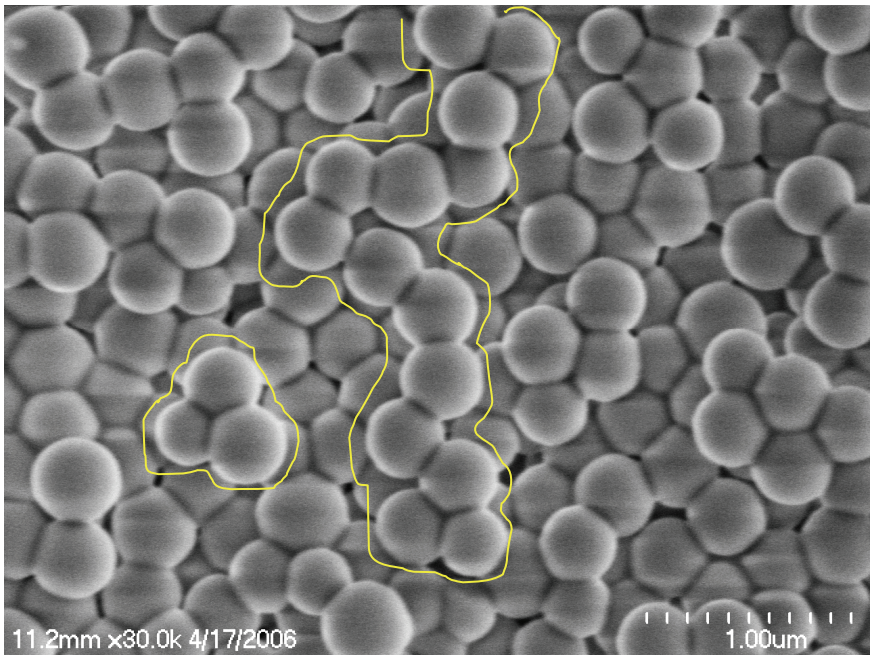
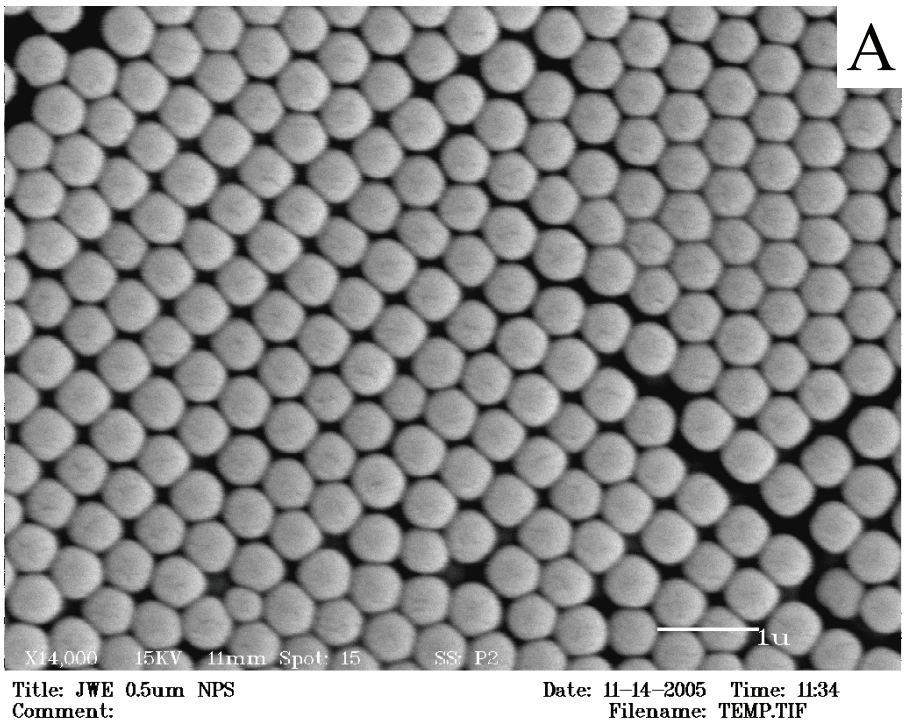


Figure 4-9: SEM of extruded bed showing particle agglomerates formed during 0.5 μm particle synthesis. Agglomerates were only found along the column walls after packing. Two agglomerates are highlighted by (---).



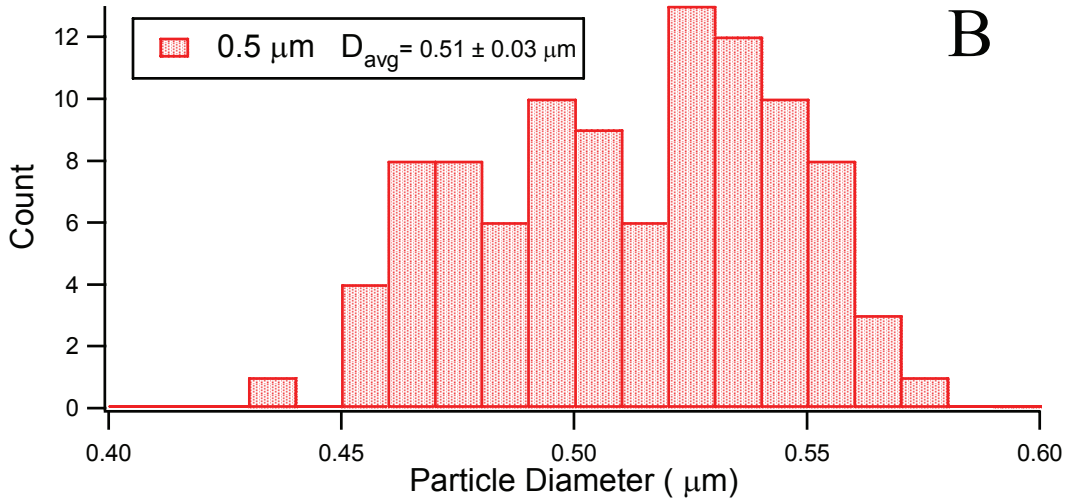
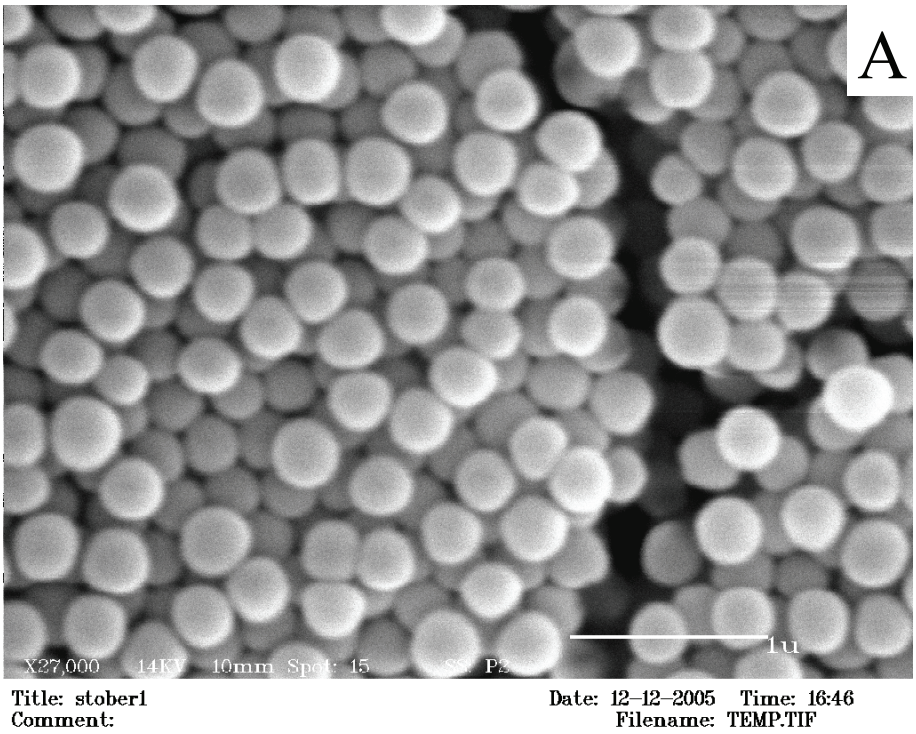


Figure 4-10: A) SEM and B) size distribution of 0.5 µm Stöber synthesis. $d_{p,n}$= 0.51 µm.



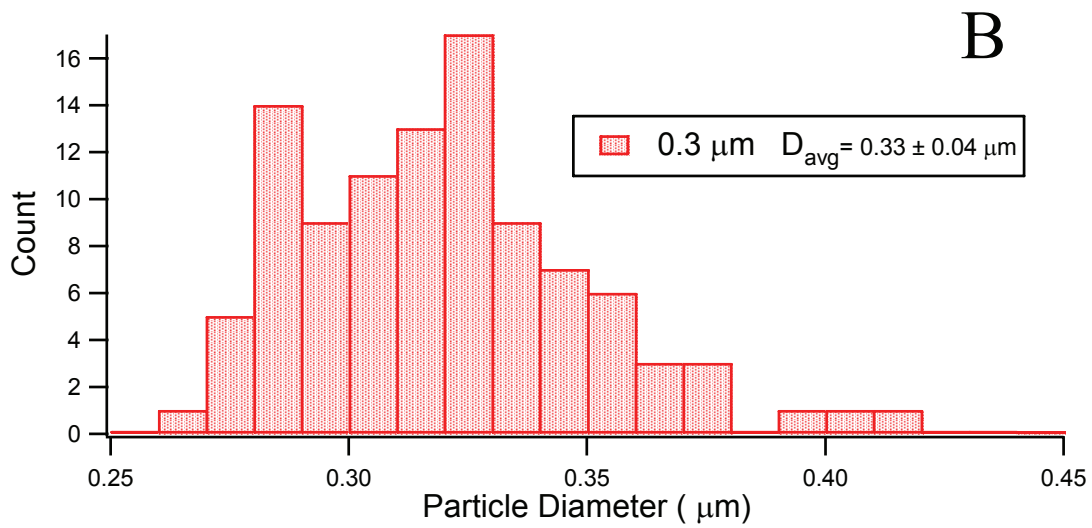


Figure 4-11: A) SEM and B) size distribution of 0.3 μm Stöber synthesis. $d_{p,n}$= 0.33 μm.

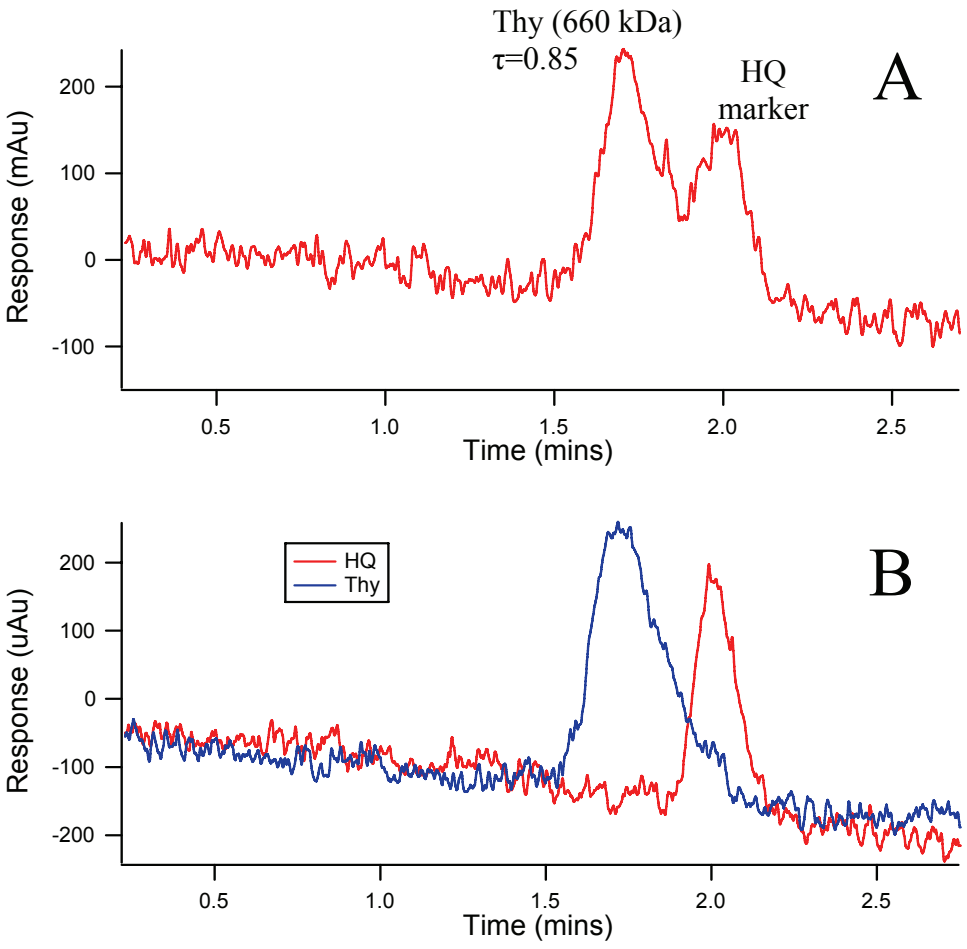


Figure 4-12: 0.5 μm particle hydrodynamic chromatographic separation of Thyroglobulin (660 kDa) from a hydroquinone dead time marker in 20 mM Borate Buffer. A) Mixture of 100 mMHQ + 0.5mg/ml Thy in buffer. B) Separate injections of HQ and Thy. Column parameters: d_p= 0.5 μm, L= 12.5cm, P= 28 kpsi

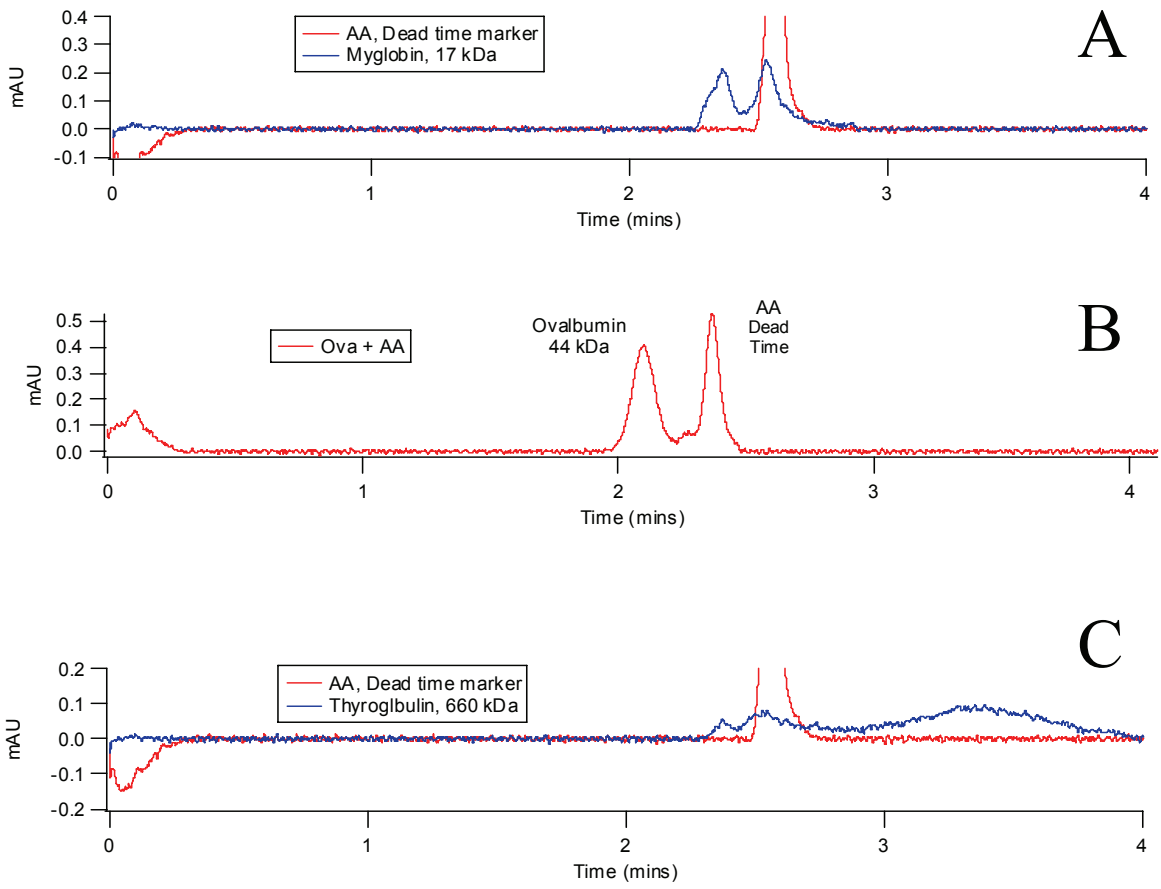
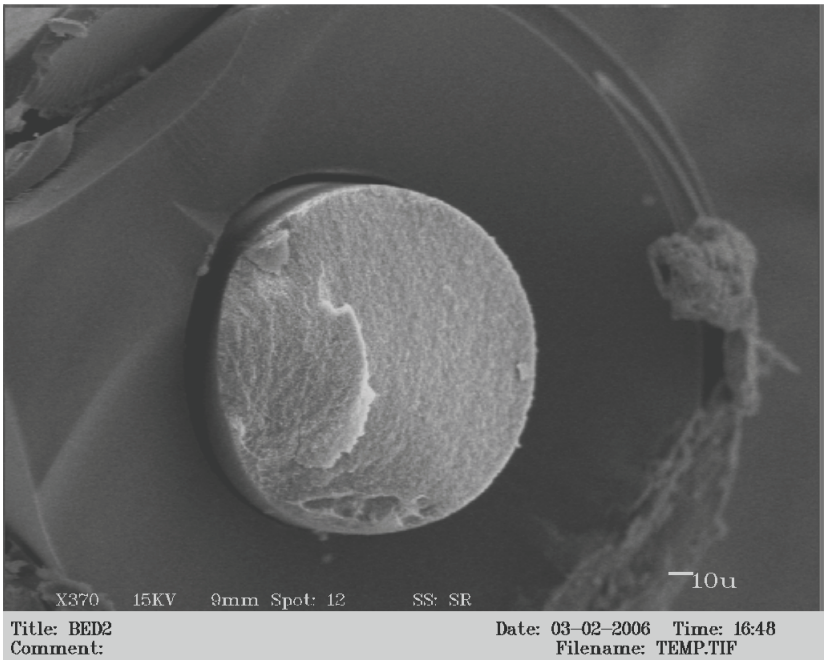


Figure 4-13: Separation of A) Myoglobin, B) Ovalbumin and C) Thryroglobulin using HDC 0.3 µm with 20 mM borate buffer. Column parameters: d_p= 0.3 µm, L= 15 cm, P= ~25 kpsi



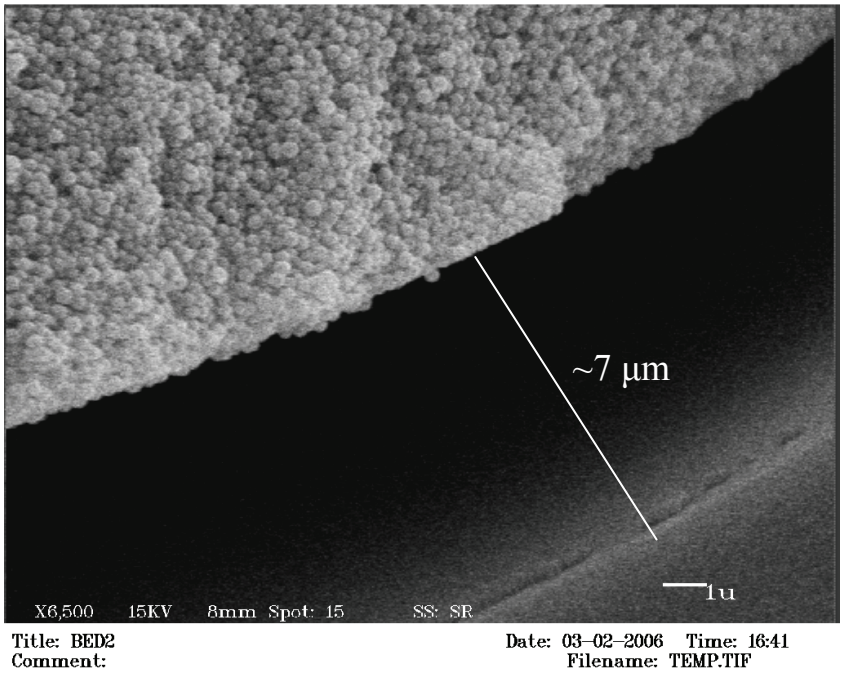


Figure 4-14: End on SEM images of 0.3 µm particles slurry packed with water in 150 µm i.d. capillary. Column dried for 48 hrs before imaging. Apparent bed shrinkage results in ~7 µm gap between bed and wall.

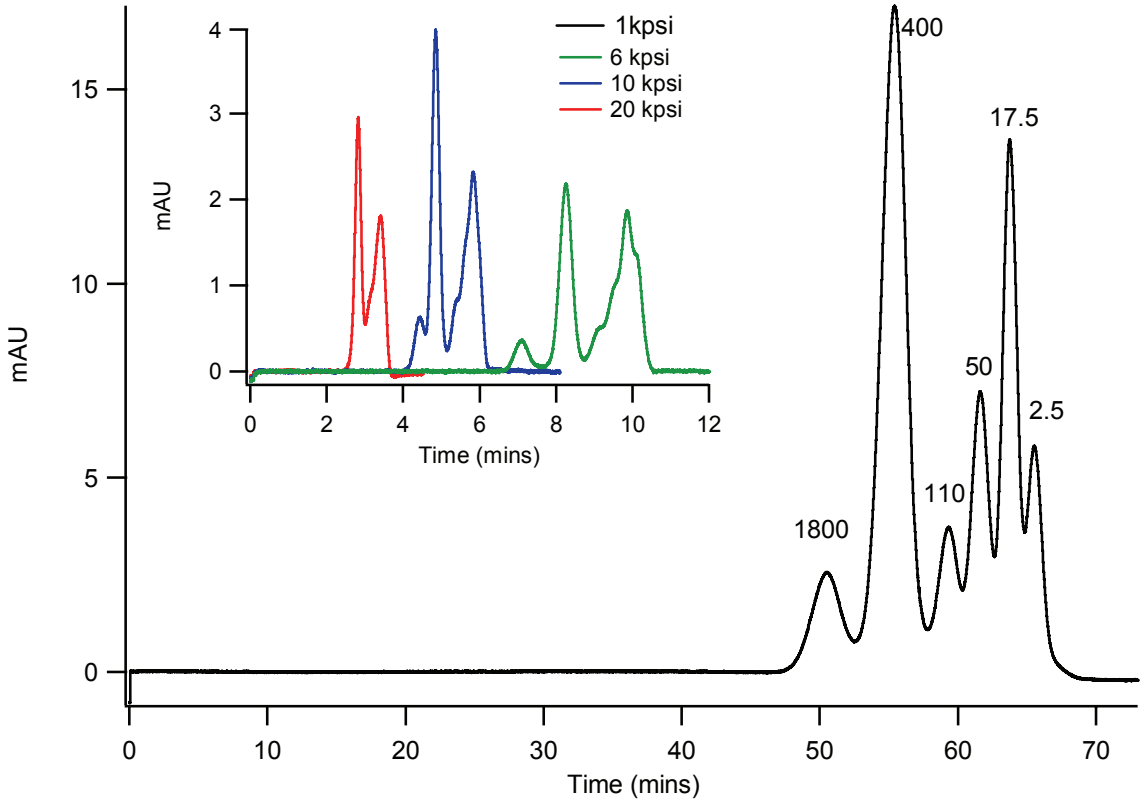


Figure 4-15: Separation of polymers by HDC using 0.9 μm NPS column. Column Parameters: dp= 0.9 μm, L= 45.5 cm, P= 1-20 kpsi

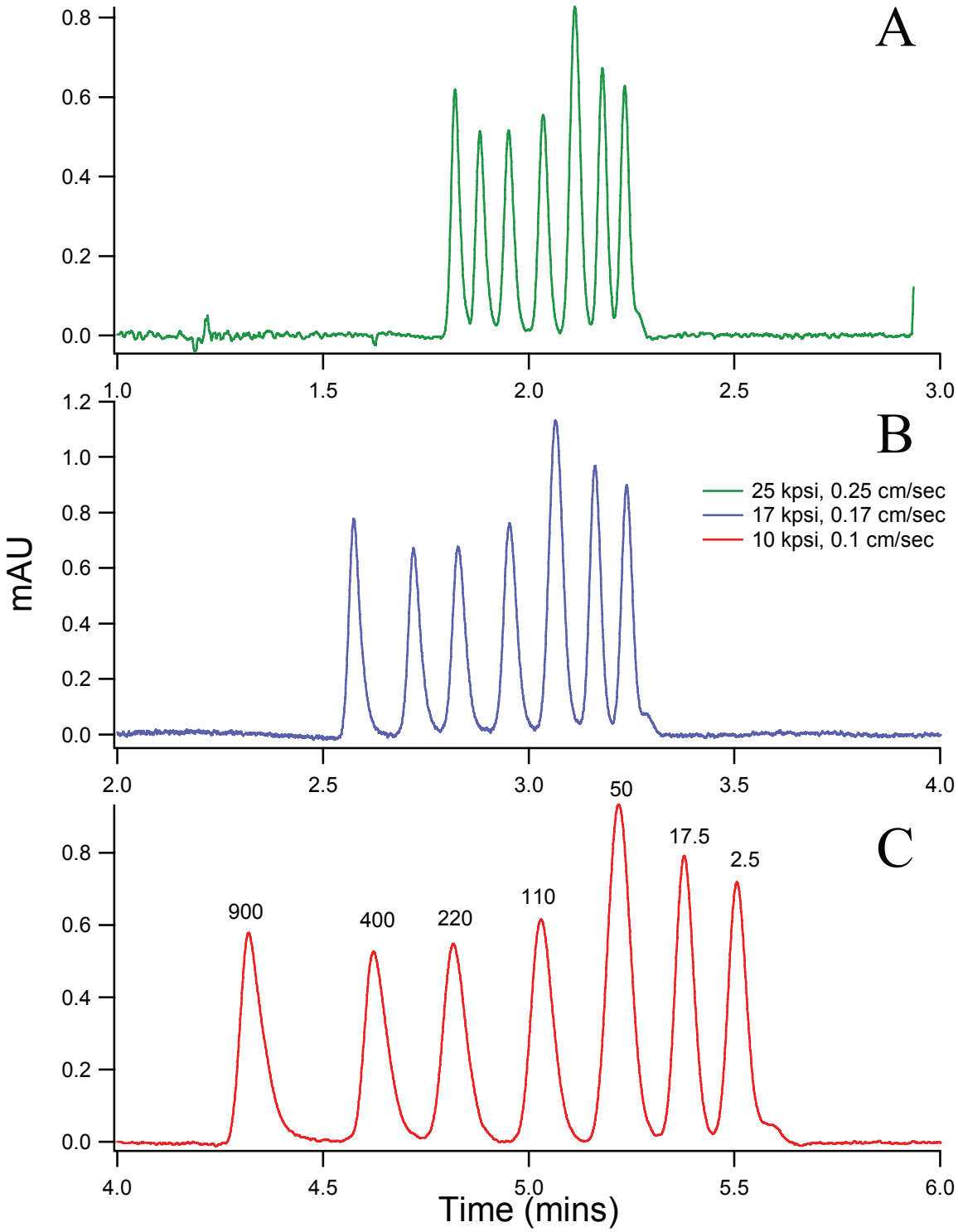


Figure 4-16: HDC separation of polymers using 1µm Micra particles. Chromatograms plotted in 2-min time segments. A) 25 kpsi, 0.25 cm/sec. B) 17 kpsi, 0.17 cm/sec and C) 10 kpsi, 0.1 cm/sec. Column Parameters: dp= 1 µm, ODSII bonded, L= 33.1 cm, P= 10-25 kpsi

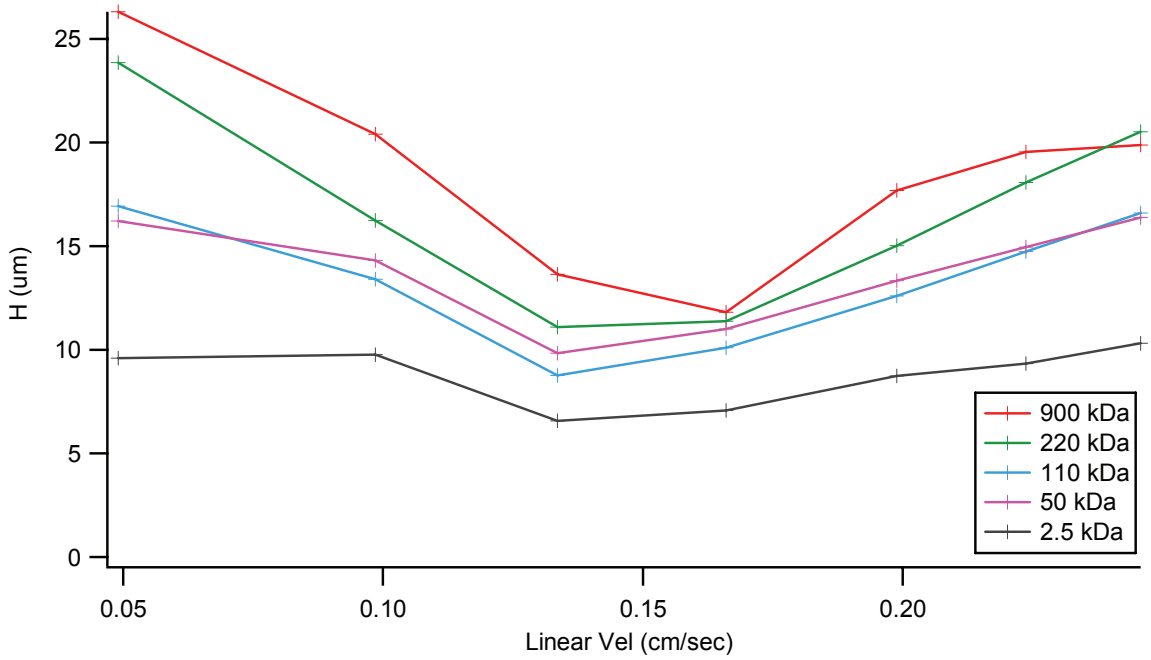


Figure 4-17: Van Deetmer plots for several different polymers separated using 1µm HDC column. Column Parameters: dp= 1 µm, ODSII bonded, L= 33.1 cm, P= 10-25 kpsi

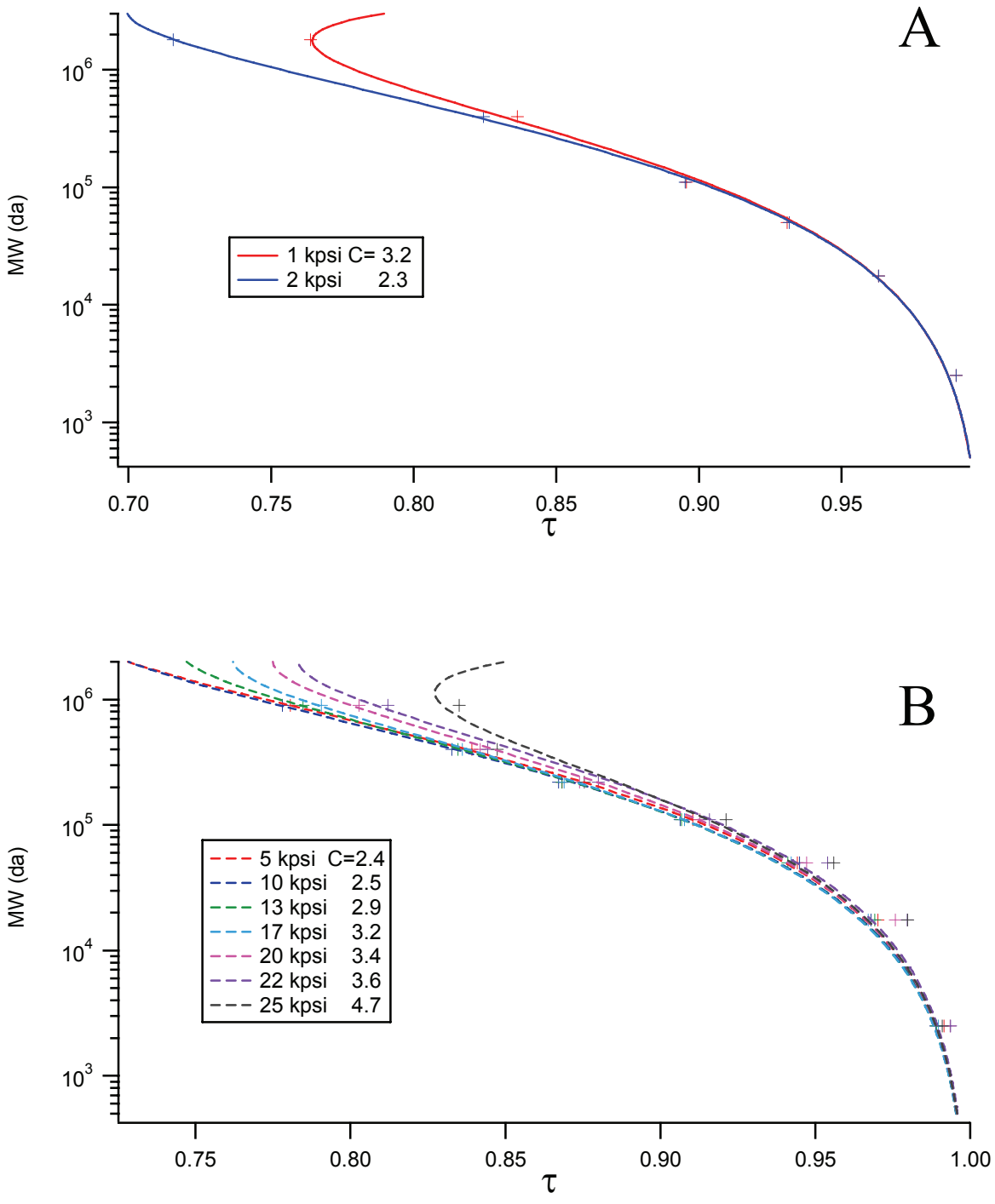


Figure 4-18: Fits of τ for HDC calibration. A) 0.9 µm column. B) 1 µm column.

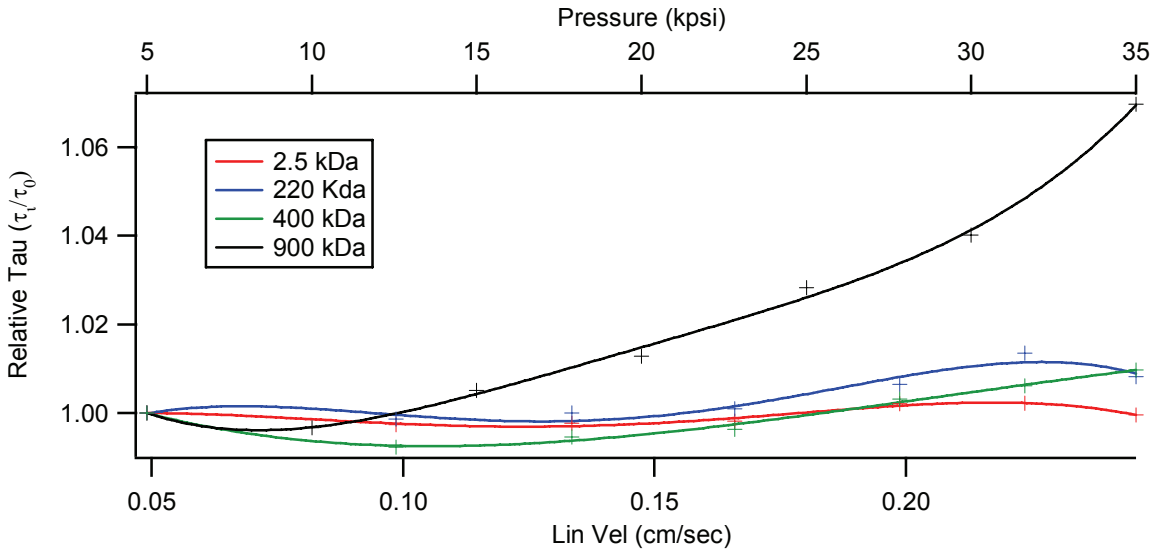


Figure 4-19: Comparison of τ with increasing linear velocity for various MW polymers.

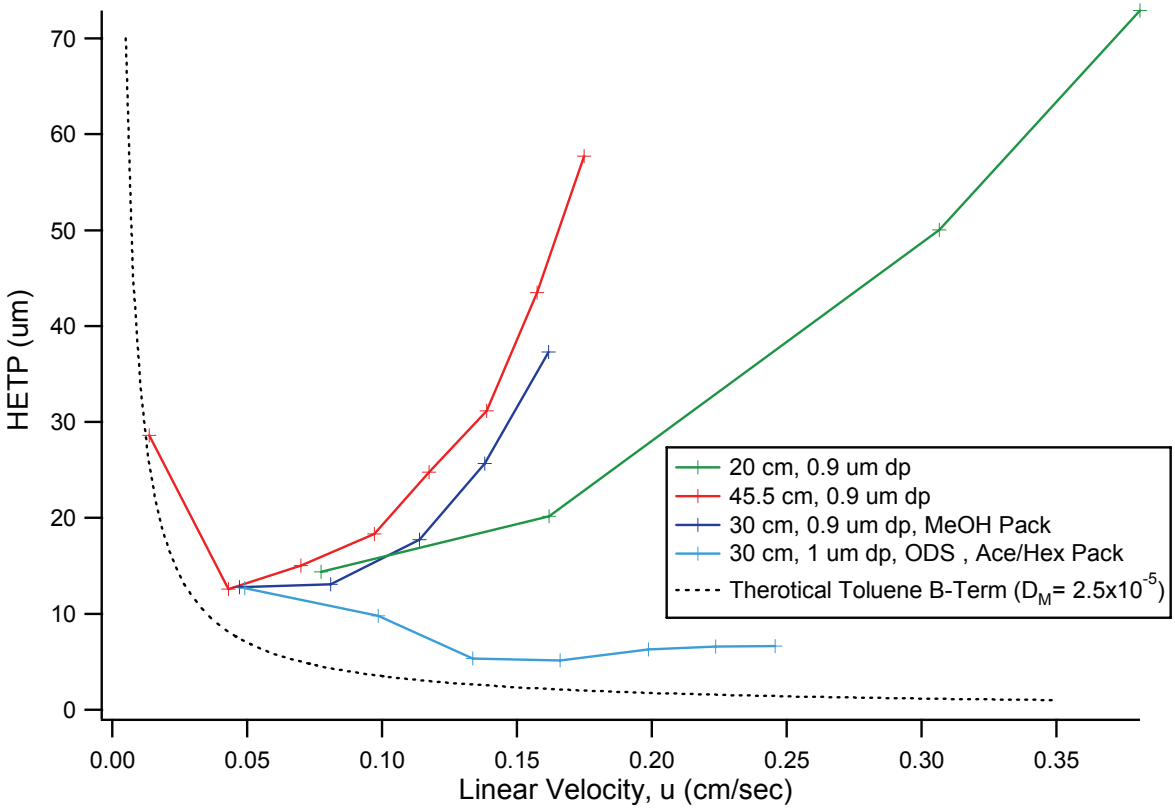


Figure 4-20: van Deemter analysis of 0.9 µm and 1 µm HDC columns at various lengths and packing solvents.

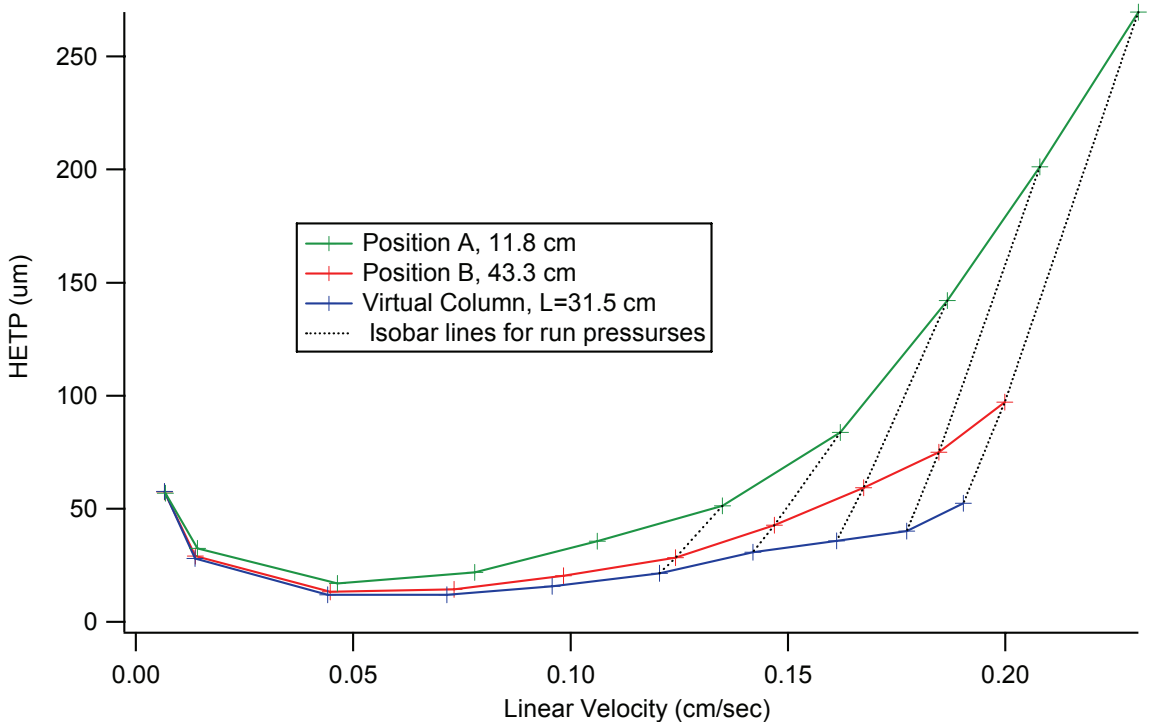


Figure 4-21: van Deemter curves for bi-point on-column detection. Virtual column is H of region from position A to position B. Isobar lines shown for 30, 25, 20 and 15kpsi run pressures. Detection Points: A: 11.8 cm; B: 43.3 cm. Column Parameters: dp= 0.9 μm, L= 45.5 cm, P= 1-30 kpsi

CHAPTER 5
METHODS FOR VISUALIZATION OF THE ELECTROSPRAY PLUME IN THREE-DIMENSIONAL SPACE

5.1 INTRODUCTION

5.1.1 Background

The previous chapters have all addressed improving proteomics from a chromatographic approach using mass spectrometry as a detection method. Any proteomics method based on chromatography, however, will only be as good as the detection method. Chapter 1 discussed that, in addition to problems caused by the number of components in proteomic samples, the dynamic range of the concentrations could be equally problematic. Thus, any chromatographic approach to proteomics will depend on the ability of the MS to detect analytes over a wide concentration range. While MS design is beyond the scope of our research, Thompson identified a potential improvement in ESI-MS sensitivity by improving the ion transmission into the MS.[1] He proposed using an electrostatic system to concentrate the ion beam during the ESI process to increase the number of ions that enter the MS analyzer, thereby increasing sensitivity. In order to develop this focusing system, it was necessary to develop the concept of an ESI Current Density Profiler that could directly probe the ESI plume. The Profiler could then be used to visualize the shape of the ESI plume in order to optimize the electrostatic focuser design. This chapter addresses methods to visualize the ESI Plume in three-dimensional space using the ESI Profiler.

5.1.2 ESI Visualization

Visualization of electrospray plume geometry is commonly performed using microscopy. This approach is limited to optical resolutions, and as the droplets become smaller, the spray plume is no longer visible. Zhou and coworkers have used fluorescence techniques to monitor analyte intensity as a function of location in an ESI plume, and more recently to monitor pH changes that occur in the plume.[2, 3] Although this technique provided the first direct measurement of ion intensities in an ESI plume, the spatial resolution was limited to 1 mm. The ESI Profiler used here monitors the density of ions produced during ESI across a given X-Y plane. By changing the location of the plane relative to the plume, the density can be monitored in three dimensions. When monitoring the local current density with a very small probe electrode, the ESI Profiler can detect current density changes with a resolution of roughly 0.5 mm. To our knowledge, this is the highest resolution measurement of the ESI Plume in real-time. Further, the ability to measure multiple X-Y planes allows for a complete, direct three-dimensional measurement of the ESI Plume.

5.1.3 Goals

The motivation behind this project is to obtain a greater understanding of how the ion density changes inside of an electrospray plume in three-dimensional space. Several methods have been developed and are presented. As expected, each method highlights certain features of the plume while preventing others from being seen. A potential solution is to combine the various techniques into a single view that allows for simultaneous visualization of each approach.

The primary goal of this project was to observe the shape of the ion density in the ESI plume at varying locations relative to the ESI source. This should be done using a single user interface that will allow multiple views of different volumes or simultaneous views of

various two-dimensional planes. The secondary goal of this project will involve characterizing the intensity falloff of the ion density in space.

5.2 ESI PROFILER

5.2.1 General Description

In order to characterize the electrospray ionization (ESI) plume, the concept of an ESI Profiler has been introduced and developed by our labratory.[4] The fabrication & operation of the device has been covered in great detail by Thompson[1], but a general description is useful for understanding the method of data collection.

A top-down view of the profiler is shown in Figure 5-1. The profiler was constructed by attaching a large brass plate to an X-Y motion control system. By applying a ground potential to the plate, and a positive potential to the ESI needle, an electric field gradient could be generated and used to produce ESI. A single electrode was then embedded in the center of the plate and used to measure current density. Using the X-Y positioner, the electrode could then be raster scanned through the ESI plume to monitor the local current density at any given X-, Y-coordinate.

5.2.2 Data Collection

To collect a profile, the ESI needle was first set at the desired distance from the ground plane ("Z"-distance). The brass plate was then translated under the electrospray plume so that the probe electrode traversed a square area in a flat s-fashion. This process was carried out in the following steps: The probe electrode was positioned in the left, lower corner of the plane to be profiled, and then scanned from left to right (+X direction). Next, the probe electrode was moved closer to the center of the ESI plume in a Y-step of previously determined distance. The probe was then scanned back from right to left (-X direction),

followed by another Y-step. This process was repeated until the desired area was covered. Data was collected at 20 Hz only during the ±X-directional scans, and each line scan was saved to an individual file. The typical rate of translation was 0.05 inches (0.13 cm) per second. For normal operation, a survey scan with a Y-step of 1 mm was often used to make sure the profiled area captured the entire electrospray plume. The survey scan typically took less than 2 minutes. Provided the profiled area did not need to be adjusted, this could then be followed with a scan at high resolution, which typically utilized a 0.250 mm Y-step. A 10-mm x 10-mm scan at this resolution could be performed in about 7 minutes. The maximum reproducible resolution in the Y-direction for this instrument is roughly 0.05 mm. For profiles taken at high resolution (Y-step less than 1 mm), only the +X-directional scans were used, in order to eliminate the mechanical actuator hysteresis of about 100 μm.

Three-dimensional current density profiles were collected at 1-mm increments from 2 to 12 mm from the ESI needle, and then stacked in Z-space to create a three-dimensional current density map. An important property of this experiment was that as distance was increased, the voltage applied to the ESI needle was increased to maintain a constant electric field at the ESI tip of roughly 1.1×10^6 V/cm.

5.3 THREE-DIMENSIONAL DATA PROCESSING

5.3.1 Visualization Software

The entire system was written using AVS/Express Visualization Edition 6.2. AVS uses libraries programmed in a proprietary V-code format, although custom developed *C++* modules can be utilized. For this work, all modules used were standard to the AVS package. The modules used and corresponding functions are presented in Appendix C in greater detail. Data import is handled by LabView and AVS. The system was designed on a Dell

Dimensions 2.4 GHz Pentium 4 running Windows XP. All images are rendered by software as a suitable graphics card was not available.

5.3.2 Volume Formation

Data is obtained as single slices and must be combined into the full volume before importing into the visualization methods. This process is shown schematically by Figure 5-2. Each slice is a 2D array in the form:

$$S = Y_j, X_i, i_i \qquad \textbf{(5.1)}$$

where X_i and Y_j represent the X,Y coordinates in Cartesian space and i is the electrospray ion density mapped to current. The volume is formed by stacking the slices to give

$$V = Z_k, Y_j, X_i, i_i \qquad \textbf{(5.2)}$$

where Z_k is the Z-Cartesian coordinate in space. The overall dimensions of the three-dimensional volume are therefore defined by $i, j,$ and k. A LabView application handles the acquisition of the individual slices and combines them into the three-dimensional volume that can be visualized with AVS. This process, summarized by Figure 5-3, is somewhat tedious for the user, but eliminates the need for array manipulation inside AVS/Express.

5.3.3 Data Space, Sampling and Sets

The data is embedded in a 3D volume. It is collected as 2D slices and is then regenerated back into 3D space inside AVS. The data was sampled in a regular grid (X-Y) for each slice. Spacing of the slice grid ranges from 1 mm-0.1 mm over a 10x10 mm area. Z-spacing is 1 mm over 10 mm of travel. This results in a 50x50x50 grid at the highest resolution.

A single point reading is taken at each (X-Y) position in the Slice Plane. This consists of an instantaneous voltage reading from the DAQ card in the computer. Position is read from a motion control board at the same time by a serial port. The signal is first passed through a

current amplifier with a gain of 10^8 V/A and then low pass filtered at 25 HZ to remove high frequency jitter or slew noise.

Data quantization as a result of analog to digital conversion, is present in this data set. The data collection system utilizes a 12-bit A/D converter which results in 4096 (2^12) discrete values that can be represented in the data. Data points were obtained from the entire data sampling gird (discussed above) without any missing values.

The electrospray plume is a dynamic field of ions. During data collection of each plane (on the order of minutes) minor fluctuations in the intensity exist. These fluctuations are random in nature and signal averaging is a potential solution if necessary. It is assumed that the fluctuations are low enough in magnitude to have no effect the visualization as a whole. When regenerating the volume, the field will be considered to be at a steady-state condition.

5.4 VISUALIZATION PROTOTYPING

Three methods were initially prototyped. They consist of an isovolume rendering of the data in 3-space, a direct volume rendering in 3-space and an isoline plot of an orthoslice plane through one of the three Cartesian axis. Sample images and a brief discussion of what the image advantages and disadvantages and why they were ultimately inadequate are presented below. The sample data was collected over a 10x10x10 mm area with an irregular sample grid of 150x20x10 points.

5.4.1 Isovolume Rendering

A sample isovolume rendering of the ESI plume is shown in Figure 5-4. The density of the ions is mapped to color, while the isovolume displays the shape of the plume for values greater than or equal to the isovalue. Densities that are less than the isovalue are mapped to null space and are transparent. A rainbow color map is presented since only the outer shell is

of interest for the image. It is not intended to show the intensity falloff inside the volume, but instead the shape of various ion densities.

Interpolation is accomplished by conversion to a uniform mesh. The isovolume then maps to this mesh and essentially connects the dots to give a single surface. Interpolation settings have a drastic effect on the final image and are therefore cause for concern.

Figure 5-5 shows three renderings from the same viewpoint with different interpolation settings. It is evident that the image is significantly changed, making it difficult to draw conclusions about the fine detail of the plume. A higher resolution data set hopefully will smooth some of these errors, but this has not been examined to date.

The isovolume shows shape in a certain region, as desired by the primary goal but limits the user to that particular view. It is not possible to explore the shape at high and low density levels simultaneously. Information about the density falloff is also non-existent. The color scales, as presented in Figure 5-4, are also not ideal since a scale with better gradation would prevent artifacts from appearing.

5.4.2 Direct Volume Rendering

Rendering of the ESI plume using a direct composite ray tracer volume rendering is shown by Figure 5-6. Data channels are similar to the isovolume as ion density is mapped to a rainbow color map and shape is conveyed by the rendering. The image is further modified to make certain values transparent. This allows the core of the plume to be seen while retaining some information about the shape of the lower density areas. A rainbow color map is used across all values. An ideal map would vary only one color at a time to prevent banding and show a smooth transition between various values.

The primary drawback of the DVR is the rendering time. It does not present any more information then the isovolume about shape, but prevents the user from interacting in real

time. Additionally, the DVR makes it difficult to tell which surface is being viewed at the outer levels due to the transparency. Density falloff is also limited by occlusion and a poor use of the color map.

5.4.3 Isoline Orthoplanes

A final method for visualizing the ESI plume is shown in Figure 5-7. This shows isoline images for two different orthoslice planes. As in the previous methods, density is mapped to color while shape is shown by the isolines. A rainbow color map was used since the data is being presented as intervals. Interpolation is achieved using the scat_to_unif function (see Appendix C), and results in similar hazards as discussed previously. The isoline plots main advantage is an easy ability to see falloff from a qualitative view. Areas with close line spacing indicate a rapid decay of ion current, while larger spacing signifies an almost constant value.

The drawback of the isoline plot is the inability to see the three-dimensional volume. While some information can be obtained, it is limited to a 2D plot and forces one to assume symmetry. Another drawback is limiting the user to a qualitative view of the density area. The ability to probe certain isolines, or plot the quantitative falloff offers additional advantages.

5.5 VISUALIZATION FINAL DESIGNS

The final design for the project consists of three different visualization applications, each aimed at answering a specific question. First, an approach to view the ESI plume in three dimensions while still viewing the interior is presented. Next, an application to generate isosurfaces in both two- and three-space from a user controlled probe location is used to answer the main question regarding ESI plume shape. Finally, a slice and extraction program

is used to extract various lines from the ESI plume and plot intensity versus distance. This helps to answer the secondary question by providing a quantitative representation of different areas in the ESI plume.

5.5.1 Cropped Isovolume View

The isovolume presented in Figure 5-4 (Section 5.4.1) was useful to display the overall shape of the ESI plume in three dimensions. It was limited because there was no way to directly view the interior of the plume and gain simultaneous information about ion current density. While it was possible to select different maximum intensity levels, this still was not very practical.

A modified version of the isovolume rendering is presented in Figure 5-8. This visualization uses a cropping box to remove a portion of the exterior so that the interior of the plume can be viewed. From this view, the relative intensity inside the plume is apparent along with the overall shape of the plume. Since the user can control the location of the cropping box, maximum intensity of the isovolume and orientation of the volume in three-dimensional space, it is more practical to interact with than simply looking at the static isovolume. We have found this view most useful when displaying collected ESI plume data

The ion density is mapped to a white-red-black step scale, with red being the middle of the dataset values (564.2 in Figure 5-10). This datamap was chosen because it provided a smooth transition between values, something that the rainbow map did not offer. Using only three colors (white-red-black) and removing the orange that is present in a hotmetal map provided an advantage when rescaling the datamap to the probe value. This was done by setting the red value to the particular probe value that was chosen. Attempts to rescale the datamap to a hotmetal map were less successful since three ranges need to be controlled.

5.5.2 Multi-Isosurface Viewer

This viewer, shown in Figure 5-9 A & B, is used to answer the main question. It consists of two windows: the 2D probe window (Figure 5-9a) and the 3D volume window (Figure 5-9b). When the user selects various density values in the probe window, the corresponding isosurface is drawn in three-space in the volume window. Additionally, an isoline as the same value is drawn in the probe window in order to provide a point of reference. Consequently, there are three main visualization methods being utilized: Isosurface of the slice plane (probe window), Isosurface of the probe value in three-space (volume window) and the isoline of the probe value on the slice plane (probe window). These three methods are briefly discussed below.

Slice Plane Isosurface. This is a 2D orthoslice plane that is extracted from the full volume. The primary use of the plane is to give the user an easy way to pick values for mapping to three-space. Additionally, this plane is used to give a qualitative view of density falloff. The plane is present in the probe window and the volume window. This allows the user to have a reference to both the probe location and volume location on the plane.

While the prototyping suggested that an isoline plot would serve the goal of density falloff best, the isolines did not give the user a full view of the slice plane. Since the primary goal of this slice plane was to allow the user to select different probe values, the isosurface was more effective and was implemented in place of the isolines.

Finally, grid lines are displayed at 2 mm intervals in white to give an overall reference to the location in the volume. White was chosen so that the lines would blend in with the surface, but still be visible if the user was interested in obtaining position information.

Volume Isosurface. The primary goal of this viewer is to show shape of the ESI plume in three-space, which is also the primary question being asked. This is accomplished with a

volume isosurface rendering of the probe value selected from the probe window. The surface is colored based on the datamap being used in the probe window, and this should match the isosurface orthoplanes at the points where it intersects. The isosurface was chosen over the prototyped volume because it prevented occlusion of the orthoplane from occurring, but still provided the same shape information. Additionally, it was chosen in place of the DVR since there is no rendering time and the isosurface provides similar information about shape. The DVR was still better at showing two different shapes of the plume simultaneously, but the ease and speed of the volume isosurface make it a better choice in the final design. White grid lines are also used to give the user reference points. The main reference is the Z-axis start value, which indicates how far the tip is from the plate when data collection began.

Probe Isoline. A single isoline at the user chosen probe value is drawn on the isosurface in the 2D probe window. This gives the user a point of reference when making further selections and provides a 2D representation of the shape. It is drawn as a solid line, mapped to the color blue. This color was easily seen on top of the color map used for the isosurface. The isoline is also useful to show the relative ion intensities in proximity to the line. A rapid change in the gradient near the line indicates a high rate of intensity falloff, while a relatively constant value indicates little intensity decay. This feature helps to answer the secondary question in conjunction with the isosurface plot.

5.5.3 Profile Extraction Viewer

In order to answer the secondary question, a separate viewer for examining cut lines was developed. While the multi-isosurface viewer can give a qualitative representation of the density falloff, the best method is to plot intensity versus distance. Such a viewer has an additional advantage since it could allow for a more quantitative analysis of the density falloffs through best-fit lines to the data.

As with the multi-isosurface viewer, the profile extraction viewer utilizes two windows and different methods for visualization. The viewer consists of a 2D slice plane view (Figure 5-10a) and 1D intensity versus distance profile view (Figure 5-10b). Using the slice plane modules in AVS, the user selects which line is of interest to view and plots that in the 1D view. The slice plane view utilizes the same features as discussed above and is used as a reference for the user to determine which slice line should be viewed. The profile graph is discussed below.

It should be noted that the main drawback of this viewer is that there is no visual feedback to which line the user has selected on the slice plain. The plains are selected via a slider, but the user does not have any immediate feedback. Instead, the slider is adjusted until an interesting area is obtained in the profile window. It is possible to back-calculate based on the slice number and total number of slices taken, but this does not result in an immediate answer for the user.

Profile Graph. As shown in Figure 5-10, ion intensity versus distance plot is generated from the selected slice plane. This provides a direct answer to the secondary goal of the project by giving the user a more quantitative answer to the question of density decay. This type of plot is ideal for this question, but has two major disadvantages as implemented. First, the "distance" axis is improperly mapped as slice index. This is a result of the way AVS handles the referencing to slice planes. A potential answer would be to generate a more traditional scatter plot, but attempts at such a plot were never successful. Further work with AVS is necessary to map the correct coordinates. However, since the initial interpolation was to a uniform grid, the points are plotted correctly even if the coordinates are not accurate.

The second limitation to this graph is the inability to output the collected data to any kind of ASCII or binary format that would be usable inside a traditional data analysis program. This, unfortunately, is a limitation of AVS that cannot be avoided without a custom export filter. Additionally, AVS does not have any obvious fitting functions that could be used in place of exporting the data. This limits the user to again forming approximations of the intensity decay. The limitations, unfortunately, lie in AVS as an application, making any solution much more difficult.

5.5.4 Implementation Analysis

Overall, the final design serves to answer the two main questions. The primary question, analysis of shape, is well answered and has been attempted in a number of ways. It could, however, be useful to explore various ways to utilize the datamap. Since the data values exist over three orders of magnitude (3-1000), the range the map must display is fairly large, while most values are less than 300. Perhaps finding the average or median of the slice and scaling the data to that value would help to more fully use the datamap. It may also be possible to implement a design with isocontours that would show the shape at several levels simultaneously. This would loose some information regarding the density falloff, as discussed above, but would be the best way to examine several values simultaneously. Alternatively, if further work shows the plume is mostly symmetric, then it may be possible to cut along the center axis and display an isosurface on one side with contours on the other. This would preserve the simultaneous shape information, but also give a better idea of density falloff.

Reimplementation of the secondary question centers on the ability to analyze it. This could involve writing fit functions inside AVS, or working to export the data to a traditional data analysis program. Improvements in the slicing routine are also needed. Most important

is a representation of where the slice in the plane is being taken. Finally, correct coordinates on the X vs. Y plot are necessary to ensure that false conclusions are not reached due to improper plotting.

5.6 SUMMARY

Several methods have been presented that allow visualization of the ESI in three-dimensional space. As presented, these represent some of the first high-resolution three-dimensional images of the ESI plume that also incorporates ion density information into the visualization. While we have found that a two-dimensional approach is most practical for lens design, the three-dimensional views have proven useful from a theoretical aspect. Of interest is the high ion density in close proximity to the ESI needle that rapidly falls off as the distance is increased.

It is finally of note that ion density is not necessarily related to ion transmission or sensitivity in the MS. Thompson found that sensitivity actually suffered when the needle was placed too close to the MS inlet due to insufficient time for desolvation.[1] Thus, while the Profiler and Visualization approach have proven successful to optimize lens designs and ESI plume shape, we have had only moderate success in actually improving the sensitivity of ESI-MS for proteomics.

5.7 REFERENCES

(1) Thompson, J.W. Doctoral Dissertation, University of North Carolina, Chapel Hill, NC, 2006.

(2) Zhou, S.; Edwards, A. G.; Cook, K. D.; Van Berkel, G. J. *Anal. Chem.* **1999**, *71*, 769-776.

(3) Zhou, S.; Prebyl, B. S.; Cook, K. D. *Anal. Chem.* **2002**, *74*, 4885-4888.

(4) Thompson, J. W.; Eschelbach, J. W.; Wilburn, R. T.; Jorgenson, J. W. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 312-323.

5.8 FIGURES

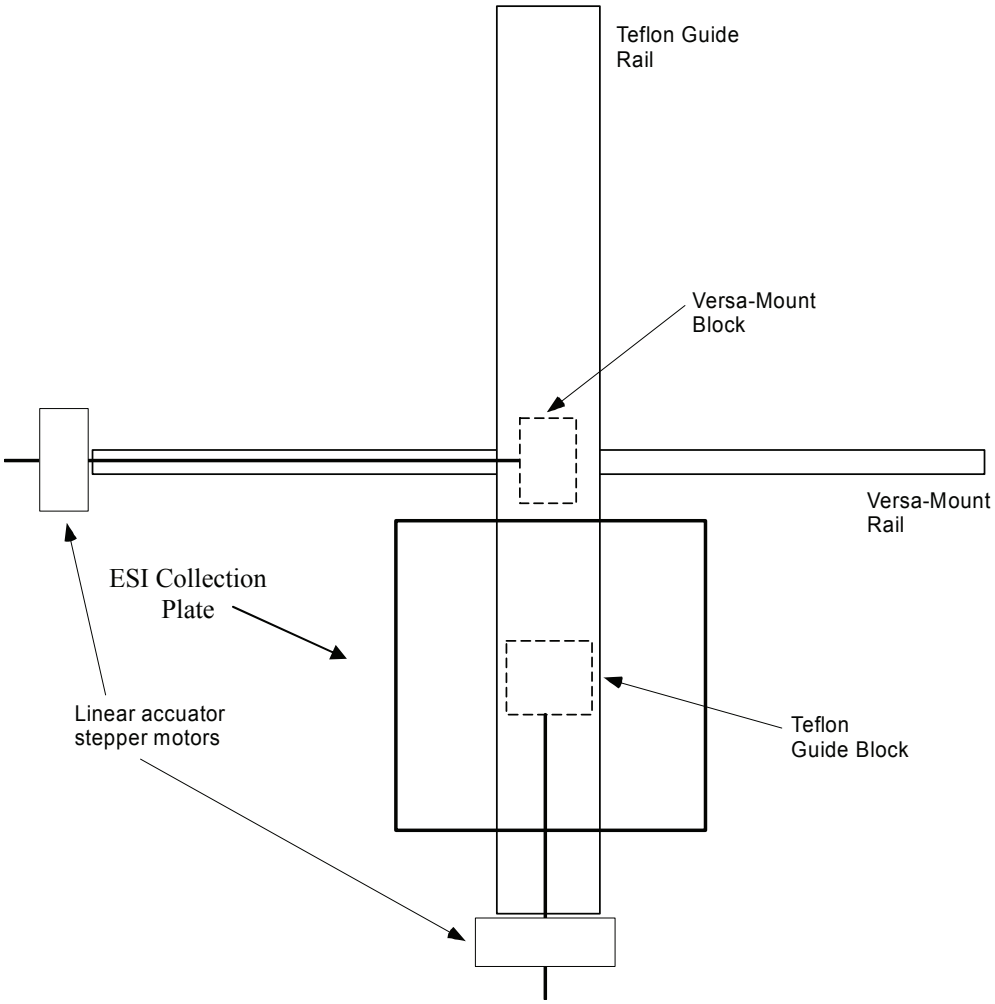


Figure 5-1: Top-down view of ESI Profiler. ESI source (not shown) is orthogonal to brass collection plate on the Z-axis. Sampling electrode (not shown) is embedded in plate.

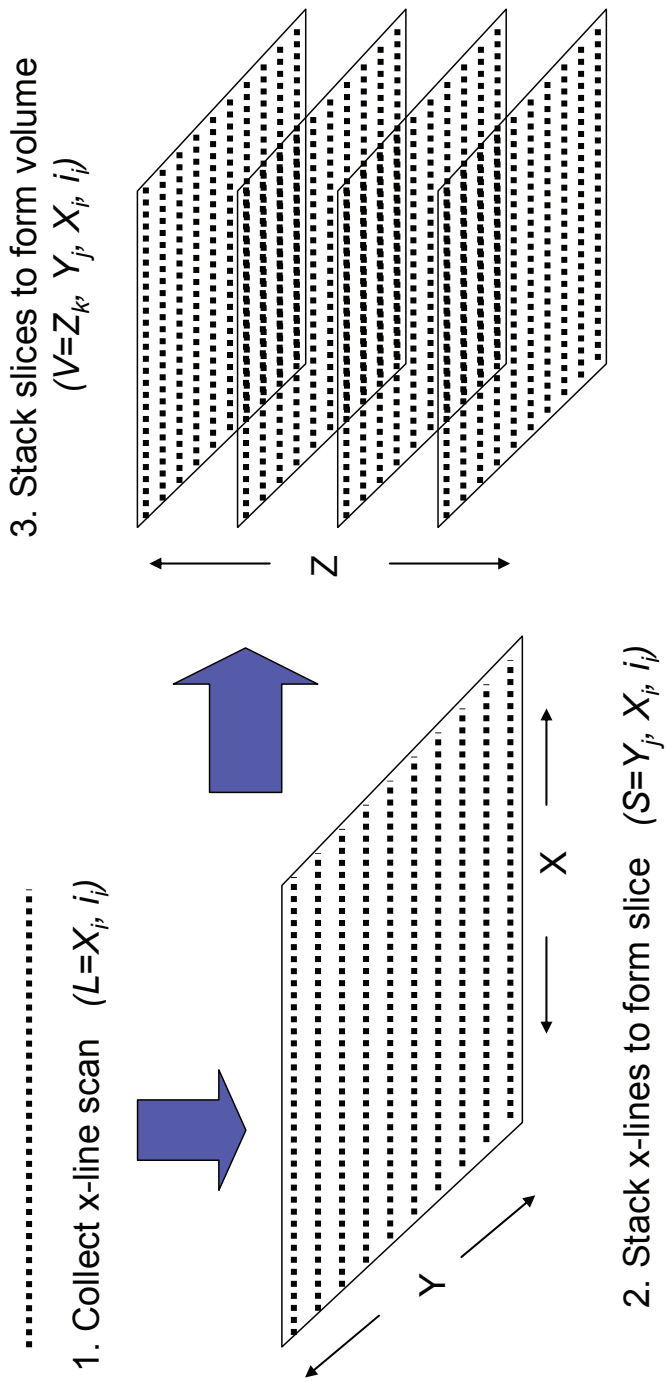


Figure 5-2: Volume formation process. 1) Line scans collected in single X-dimension. 2) Multiple line scans combined across the Y-axis to form Slice Plane. 3) Several Slice Planes combined to form Volume.

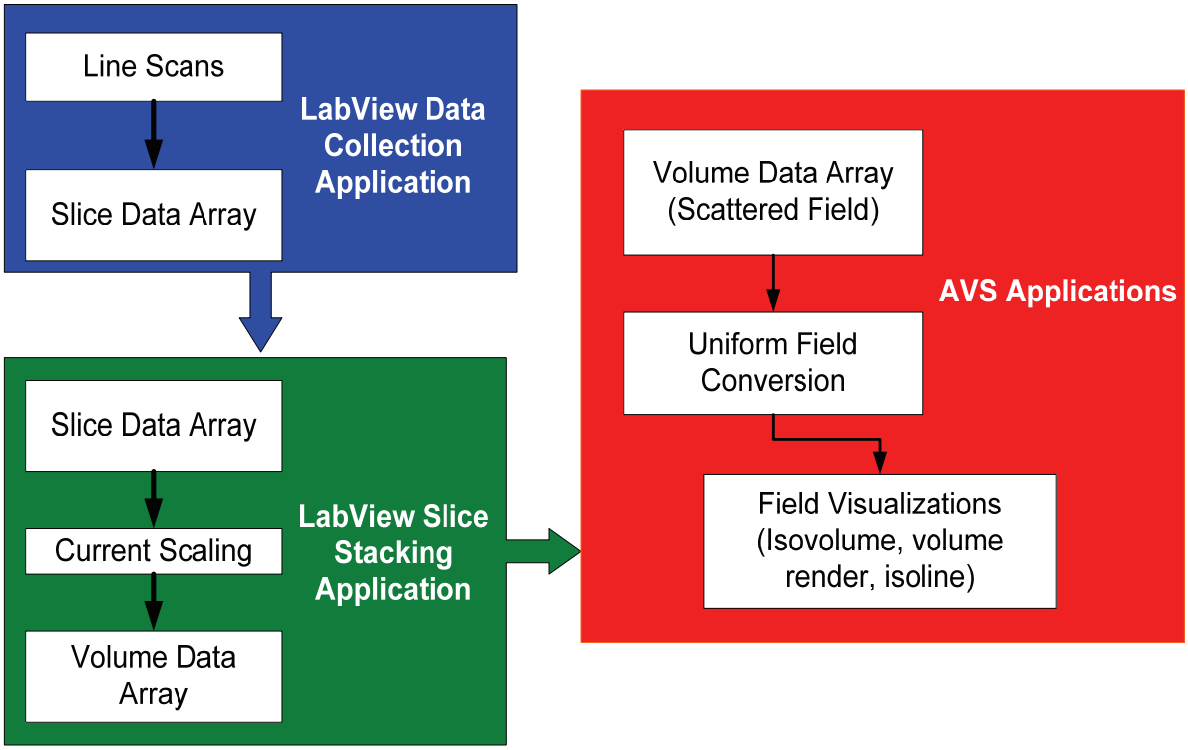


Figure 5-3: Data collection flow chart.

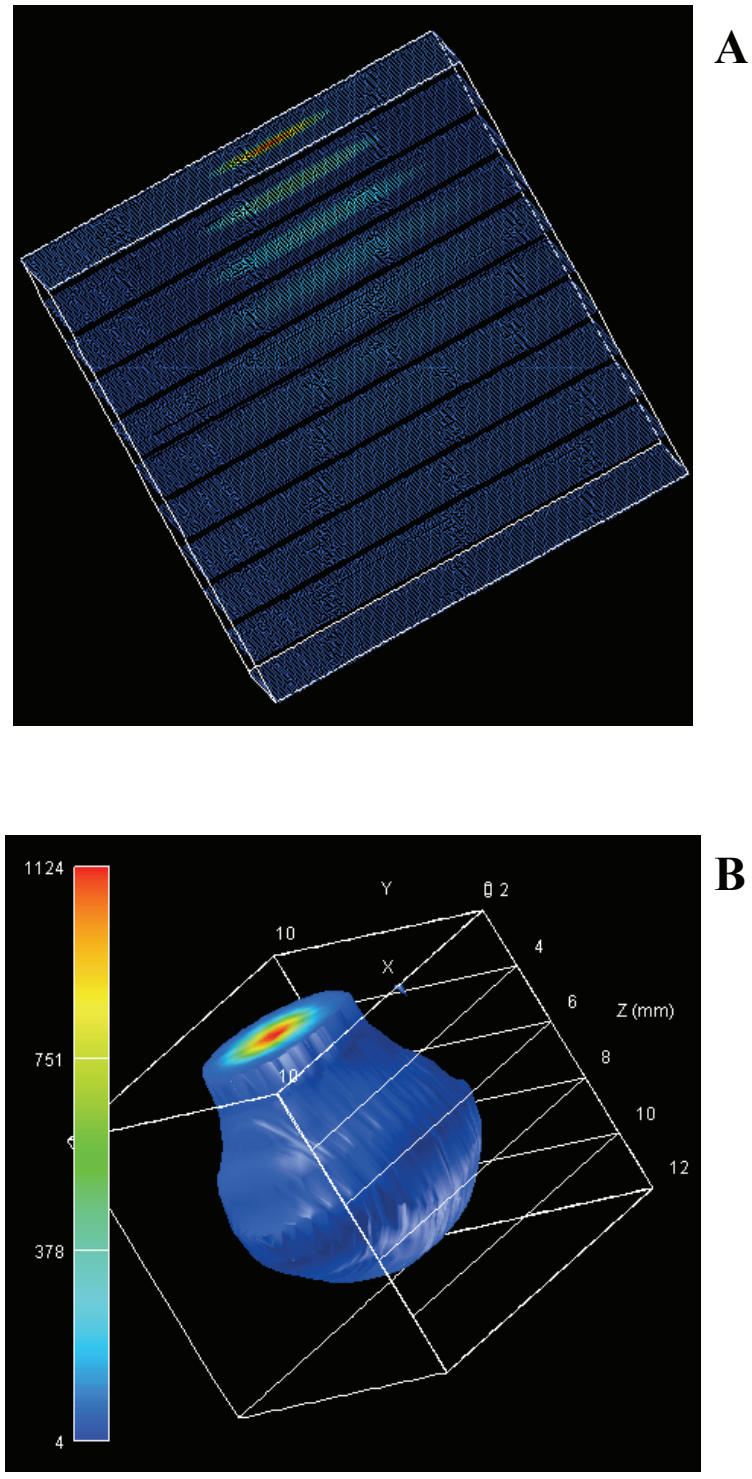
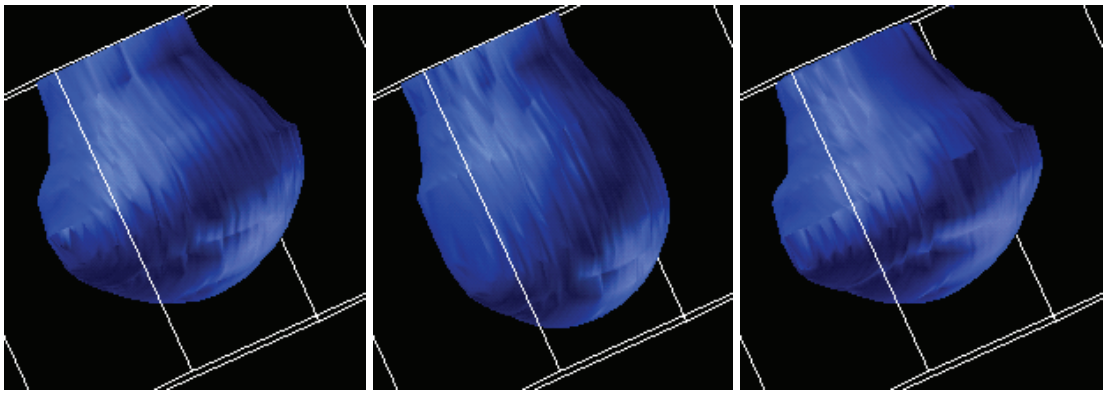


Figure 5-4: Isovolume rendering of ESI plume at iso level of 43. Grid interpolated to 1st order 100x20x20 with range set to 2, 1, and 3 respectively.



A. B. C.

Figure 5-5: Comparison on interpolation settings. A: 1st order 100x20x10; range= 2,1,1. B: range = 2,1,4. C: Same as A, but using 3rd order.

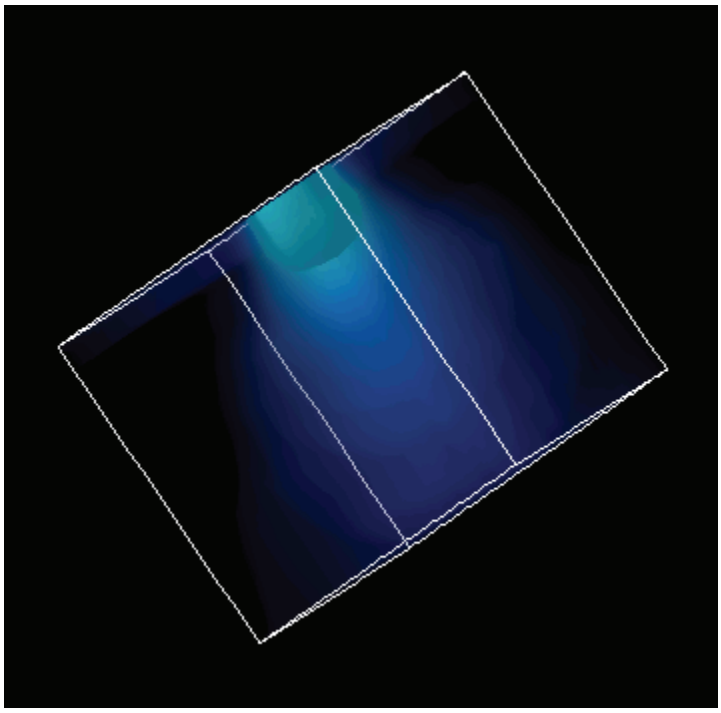


Figure 5-6: Direct volume rendering of ESI plume. The dense, cone-like core region is surrounded by a much larger area of lower density. Direct composite ray tracer rendering with trilinear interpolation. Range control=248, alpha= linear from 0-0.06 over 0-248 range.

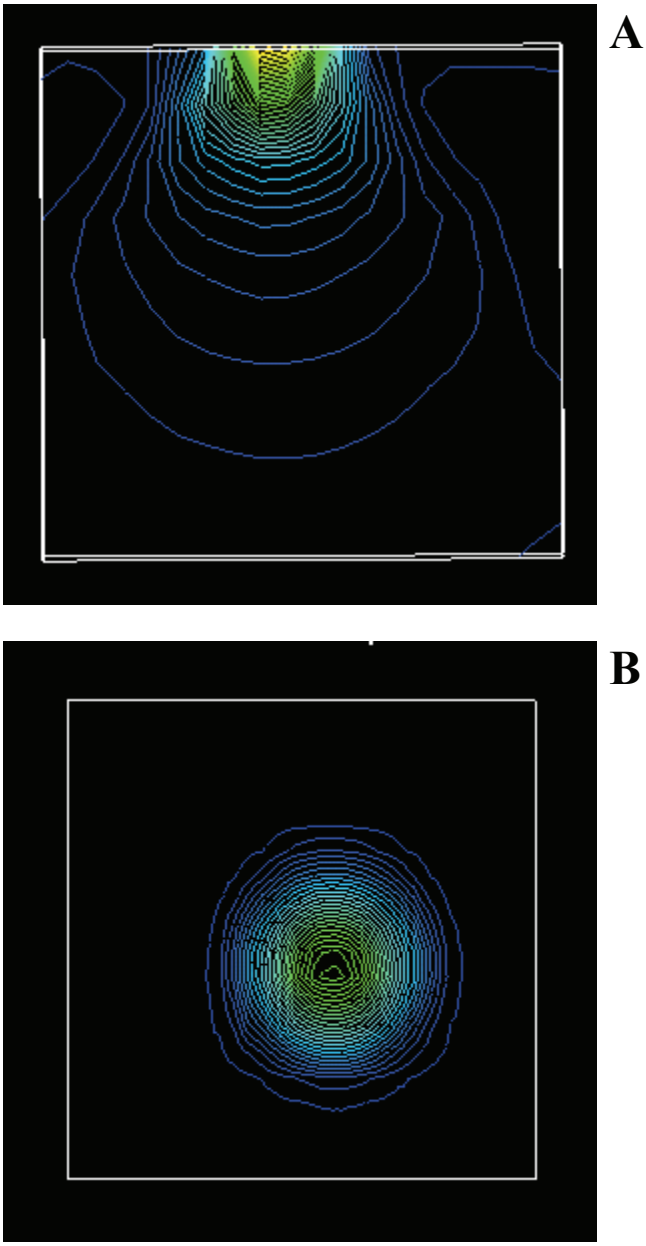


Figure 5-7: Orthoplane isoline views for ESI plume. A) Axis 0 (ZY), slice plane 5. B) Axis 2 (XY), slice plane 3.

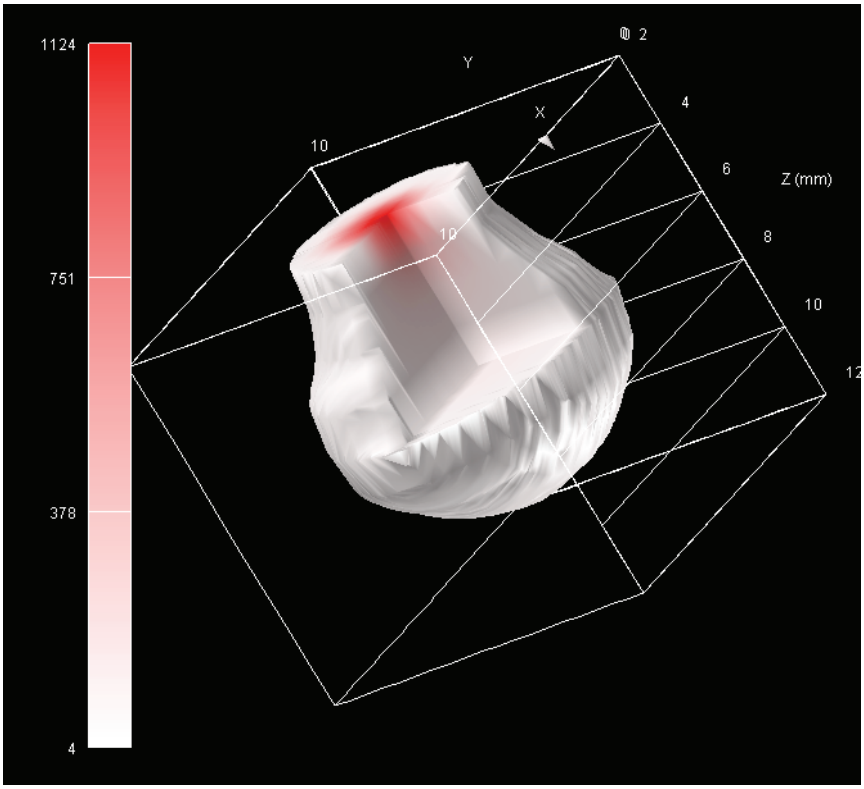


Figure 5-8: Cropped isovolume rendering with center cube crop. Displays overall shape along with interior current density.

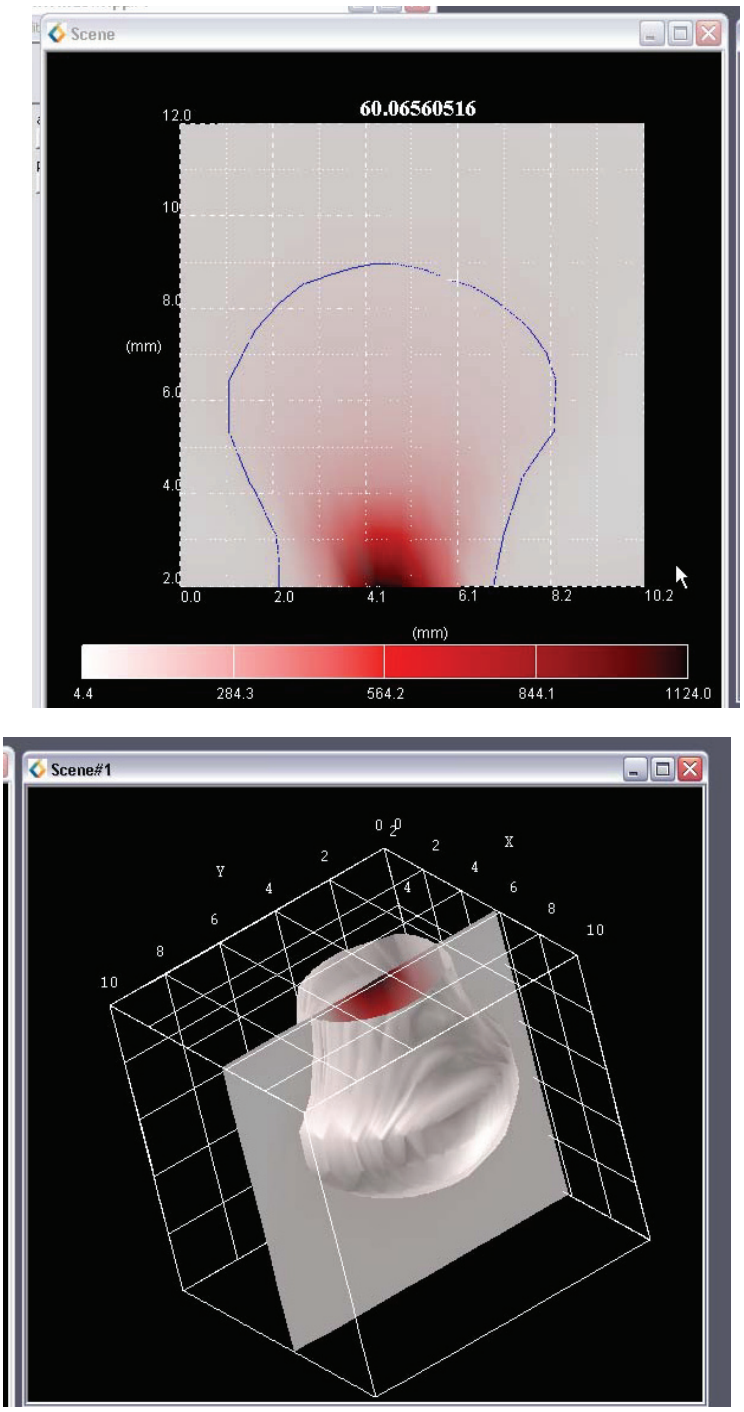


Figure 5-9: Multi-Isosurface Window Views. A) Probe View: Isosurface is user selectable to choose the probe value. Isoline indicates currently selected probe value, as does number on top. B) Volume View: Volume isosurface is displayed at value selected from probe window. Slice plane isosurface present in probe window is also displayed for reference.

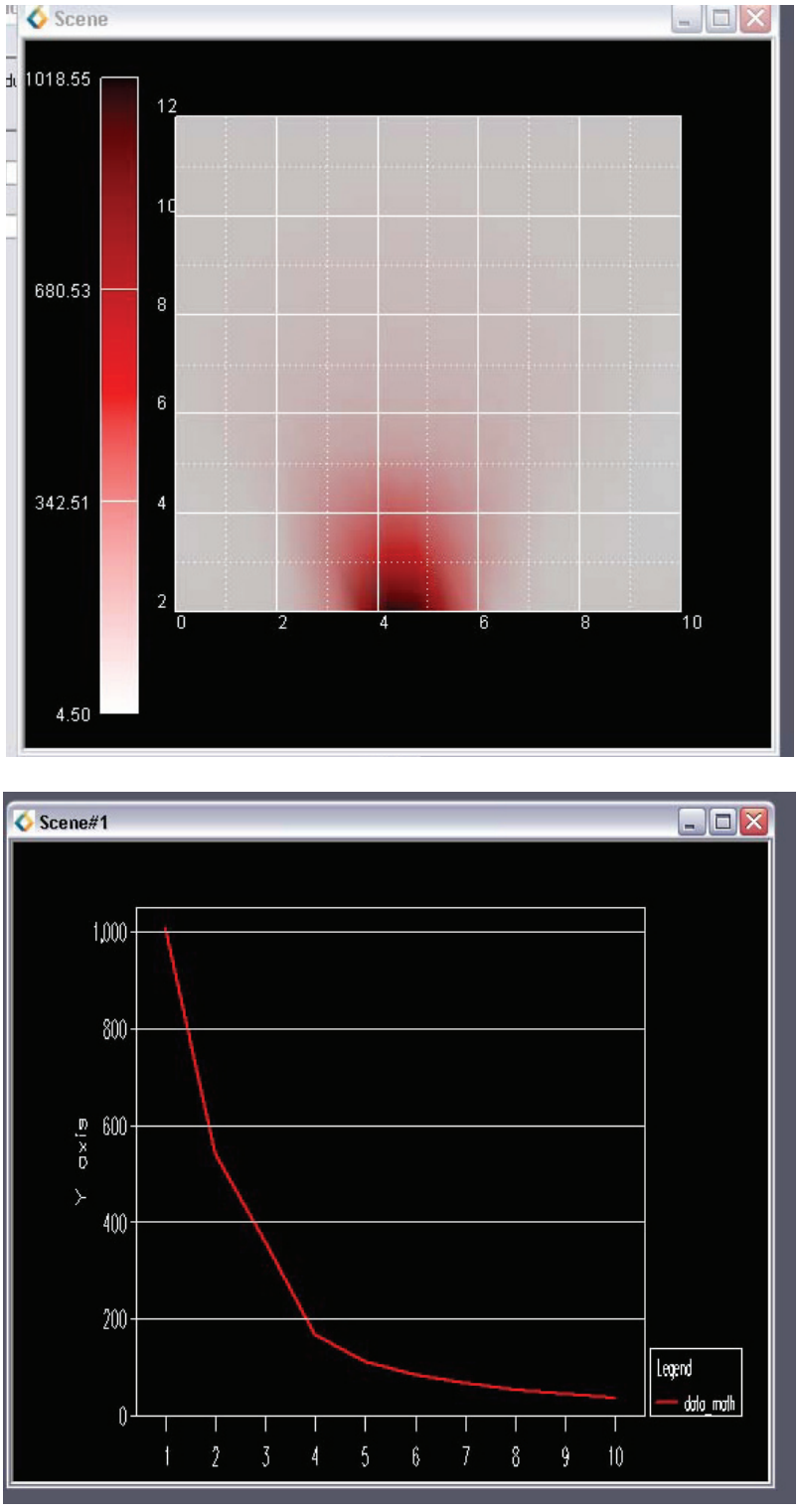


Figure 5-10: A) Probe View: Isosurface is user control the slice ling. B) Profile Extraction Viewer: 1D plot of extracted line.

APPENDIX A
DERIVATION OF OPTIMUM THEORETICAL PLATES

A.1 DERIVATION

To calculate column parameters for a fixed pressure, it is first useful to write Kozeny-Carman (Eq.1.7) in terms of linear velocity, μ:

$$u = \frac{d_p^2 P}{405\eta L} \qquad \textbf{(A.3)}$$

where ε from Eq. 1.7 has been estimated as 0.4, a typical value for packed spheres. Next, the retention time (t_r) for a given analyte can be defined as

$$t_r = \frac{L}{u}(1+k') \qquad \textbf{(A.4)}$$

Where k' is the retention factor. For simplification only the case of column dead time can be considered. In this case, k'=0 and the previous two equations can be combined as:

$$t = \frac{405L^2\eta}{d_p^2 P} \qquad \textbf{(A.5)}$$

From this equation, the dead-time of a column is now a function of both column length and particle diameter. This relationship can be simplified further and written in terms of either L or d_p:

$$d_p = L\left(\frac{405\eta}{tP}\right)^{1/2} \textbf{(a)} \qquad\qquad L = d_p\left(\frac{tP}{405\eta}\right)^{1/2} \textbf{(b)} \qquad \textbf{(A.6)}$$

At fixed viscosity, pressure and dead time, d_p and L are now a direct function of each other. Selecting a d_p dictates the column length, just as a fixed column length dictates the d_p.

With these relationships defined, it is now useful to combine H (Eq. 1.5) and u (Eq. A.3) to write plate height as a function of column parameters:

$$H = \frac{d_p}{2} + \frac{810\eta D_m L}{d_p^2 P} + \frac{d_p^4 P}{2025 D_m \eta L} \qquad \textbf{(A.7)}$$

Further, since the goal is to calculate the optimal parameters to maximize N, the previous Eq. can be combined with N and simplified to:

$$N = \frac{L}{H} = \frac{2{,}000\eta D_m P L^2 d_p^2}{1{,}000\eta D_m P L d_p^3 + (1.6x10^6)\eta^2 D_m^2 L^2 + P d_p^6} \qquad \textbf{(A.8)}$$

Finally, since L was previously defined in Eq. A.6b, it can be substituted into the previous Eq. and simplified to:

$$N_{opt} = \frac{5 D_m P d_p^2 t}{50\eta^{1/2} D_m P^{1/2} d_p^2 t^{1/2} + 4{,}000\eta D_m^2 t + P d_p^4} \qquad \textbf{(A.9)}$$

This now defines N in terms of a variety of physical parameters of the analyte and solvent. Since we have previously decided to operate at certain fixed pressure, and dead times, Eq. A.9 can be more generally written as

$$N(P, d_p) = \frac{\alpha P d_p^2}{\beta P^{1/2} d_p^2 + \chi + P d_p^4} \textbf{ (a)} \qquad N(t, d_p) = \frac{\alpha' t d_p^2}{\beta' t^{1/2} d_p^2 + \chi' t + \rho' d_p^4} \textbf{ (b)} \quad \textbf{(A.10)}$$

where α,β,χ are constants for a fixed dead time and α',β',χ',ρ' are constants for a fixed pressure.

APPENDIX B
GENERATION OF CHROMATOSPECTRA USING SEGMENTED AUTOME DATA

B.1 Automated MaxEnt

B.1.1 Time-Segmented AutoME

The Maximum Entropy algorithm was used in Section 2.3.2.2 for automated deconvolution of acquired ESI-MS data. This was accomplished by manually selecting peaks and summing the MS signal during the peak elution window, and then applying the MaxEnt routine to the summed MS data. Alternatively, the MS software could automatically find and sum chromatographic peaks followed by MaxEnt deconvolution. This approach, known as peak-based MaxEnt, is advantageous because each peak can be identified as a certain protein and the run readily summarized by the components present. A traditional peak-based approach to MaxEnt deconvolution is shown in Figure B-1a. Each highlight peak could be summed and processed by the MaxEnt algorithm for identification. The main drawback, however, is that for complex samples, peak overlap can make the MS signal summation convoluted and difficult to assign to specific proteins. As an alternative to peak-based MaxEnt, the concept of time-segment based MaxEnt has been introduced by Waters Corp. with an automated routine known as AutoME.

The time-segmented approach to MaxEnt deconvolution divides the entire chromatogram into discrete time segments, typically < 0.5 min. Each time segment is then summed and MaxEnt deconvolution performed. A simple illustration is show in Figure B-2b to compare to the traditional peak-based approach. This method is advantages because it allows the data to be processed in uniform steps. The proteins and intensity can then be determined across the entire chromatogram, independent of peak shape and resolution. This approach,

unfortunately, is quite time consuming. A single 60 min run could be divided into 600 discrete segments using 0.1 min segment width. This would compare to 50 peaks that may be present, meaning that the processing time is around 10-fold longer. Additionally, since the segment width is narrower then a given peak width, the summed signal intensity is reduced making MaxEnt deconvolution prone to greater error. Overall, though, the time-segment based approach is still better suited for complex, unknown samples as more data can be extracted at the expense of processing time.

B.1.2 Input Parameters & Post-processing

The AutoME routine consists of a VisualBasic macro that automates the use of MassLynx MS Software for processing of LC-MS data. Each step performed by the routine can be manually recreated inside MassLynx, but would be impractically tedious for a large data set. Several input parameters are required for the AutoME processing. An example of these input parameters is shown in Table B-1. Most important to note is that the segmentation is handled in terms of MS scans, not time. On our mass spectrometers, data is acquired at 2 Hz, meaning that each scan is 0.5 sec of time. Therefore, combining 10 scans gives a segment width of 5 sec, or roughly 0.1 mins. This is a typical segment width for AutoME. Higher widths begin to sacrifice resolution, while lower widths are of limited use due to the decreased signal and increased processing time. The other important parameter to note is the unitless threshold variable. Since MaxEnt requires good S/N from the summed segment, each segment can be thresholded to determine if enough signal is present to perform MaxEnt. This variable determines the minimum required signal for further processing, but is also unitless as it is not correlated to the actual intensity measured. We have found it best to optimize the threshold empirically, minimizing the value to not overly discard segments, but also prevent MaxEnt errors from occurring due to low S/N. The remaining parameters

control how MaxEnt is performed and how the data is saved once processing is complete. Table B- displays typical values for these parameters.

Once AutoME is complete, data is saved in the form of a retention time with all the deconvoluted protein MWs from MaxEnt and their corresponding intensity. This file has a "*.ConCat" extension, and is a simple tab-delimited ASCII text file. This data must be further processed using a custom Igor routine to load and filter. The dialog for loading the ConCat output is shown Figure B-2a. Three variables need to be defined before loading. First, the minimum MaxEnt component intensity to load must be set by the "Intensity Threshold" variable. Since a MaxEnt output typically produces several major components, but also a low-level background noise, the component list would be overly complex if all peaks were loaded. The component list can be simplified further by only loading the *n* most intense peaks using the "Number of Peaks to pick" variable. A value of 0 loads all components above the threshold, while a value of *n* would load the first *n* most intense peaks. Finally, the number of scans per segment, set by the initial AutoME parameter file, is input for proper loading. After loading is complete, three Igor waves are generated that contain retention time, deconvoluted MW and intensity. These waves are then used for further analysis and data plotting.

B.2 Application of AutoME

We have developed two main uses for the Waters AutoME routine. First, the collected MS data can be deconvoluted and re-plotted in the form of a Base Peak Intensity (BPI), a common plot for peptide MS data. Second, two-dimensional LC-MS image plots have been generated. These plots are similar to Figure 2-11, but add a Z-dimension that is mapped to intensity. A brief discussion and example plots follow.

B.2.1 Deconvoluted Protein BPI

Normal MS chromatograms are generally viewed in the form of a Total Ion Count, or TIC. This is a nonselective plot off all the ions that the MS detects at a given elution time. This approach tends to produce a somewhat lower S/N since the baseline intensity is susceptible to the solvent composition and background ions. As a solution, MS data has been plotted in the form of a Base Peak Index, or BPI. In this form only the most intense ion for any given scan is displayed. High signal intensity then indicates that a dominant ion is eluting during the run. This approach is especially useful when analyzing peptides as an increase in the BPI signal almost certainly is the result of an eluting peptide. Unfortunately, the MS envelope produced by ESI of proteins makes it difficult to plot protein data in the form of a BPI since many charge states occur. Using AutoME, it is possible to deconvolute the protein charge states to the original protein MW, so that it can be more conveniently plotted for analysis.

Deconvoluted protein BPI (dBPI) plots are produced by first processing the LC-MS data with a time-segmented AutoME routine. The data can then be processed by the Igor load routine with only the most intense MW for each time-segment being loaded. This is accomplished by setting the "Number of Peaks to Pick" dialog value (Figure B-2b) to "1", meaning that only the most intense peak for each segment will be loaded and subsequently plotted.

A sample dBPI compared to the original TIC is shown in Figure B-3. The figure shows the RP-gUHPLC separation of an *E Coli* lysate fraction collected from an anion exchange analysis. A few advantages can be seen from this plot. First, the background for the dBPI has been reduced. This occurs because the low intensity background ions that increase the TIC have been removed from the dBPI. Next, by plotting only the most intense

deconvoluted species, the peaks can be readily correlated to actual proteins. A MW filter can be applied to the dBPI so that only large deconvoluted ions are displayed. This is more convenient for data analysis since all the peaks are then the result of an eluting protein.

The main disadvantage to analyzing data by this method is the required AutoME post-processing. In order for the dBPI peak area to be related to intensity, MaxEnt must be allowed to process to convergence. Because each subsequent iteration takes progressively longer, the time required to process a single file commonly grows to over 24 hrs for normal 60 min analysis times. This can become tedious for larger, routine data sets and unfortunately is only dependent on the MaxEnt algorithm which is not readily optimized. Nonetheless, the dBPI has still proven useful for analyzing top-down proteomic data.

B.2.2 ChromatoSpectra Plots

The second way in which AutoME has been used for analyzing top-down proteomic data is through the concept of a ChromatoSpectra plot. Bottom-up LC proteomic data has long be viewed via a "peptide fingerprint" map which plots the LC retention time versus the peptide *m/z* intensity. For the most part, a protein will give a unique peptide fingerprint when analyzed via a bottom-up method. Two different proteins can be easily compared from the resulting fingerprint. This approach was extended to top-down data and the idea of a ChromatoSpectra plot developed.

Since AutoME generates a data file with retention time, deconvoluted MW and intensity, this data can be used to make a two-dimensional image plot LC-deconvoluted MS data. A second custom Igor routine, shown in Figure B-2b, is used to redimension and plot the AutoME data. Both the MW range and RT range control the range of the plot, while the bin size controls the resolution of the deconvoluted MW. The bin size can only be as small as

the MaxEnt output resolution (OutputMassIncrement, Table B-1), and is typically much larger for all but the highest resolution plots.

Ideally, the ChromatoSpectra could be plotted with the same bin size as the processed data from AutoME. Unfortunately, this is not practical. For a 5-25 kDa MW range and 10-55 min RT range, a 20,000 x 450 point matrix would be produced at 1 da resolution. This produces an array with 9×10^6 points, which is impractical to work with. Additionally, the resolution is actually too high to "see" the components unless a specific area is focused on. Instead, binning is used to produce a low resolution ChromatoSpectra over a wide MW and RT range or a high resolution plot over a much narrower region. A series of ChromatoSpectra with increasing resolutions are shown in Figure B-4 through Figure B-6. First, a low resolution "survey spectra" is shown with 100 da bin size. A number of deconvoluted protein peaks are evident, and a region of interest is highlighted, from which a second, higher resolution plot can be produced. This plot with 10 da bin size is shown in Figure B-5 and several more deconvoluted proteins are evident. Zooming in on particular regions (callout, Figure B-5) begins to display the parent protein and several examples of the protein adducts. A final plot with 1 da resolution is shown on a narrower Figure B-6. This resolution is now equal to mass increment used for MaxEnt during the AutoME processing. Since the region is sufficiently small, the data does not need to be binned and can be viewed in its raw form. From the figure, a parent protein appears at 11.89 kDa with several adduct masses every ~0.02 kDa. The adduct peaks appear to drift with RT, but this is likely an artifact caused by a poor MaxEnt fit due to the lower signal intensity as the parent protein RT is quite constant. While being the smallest range shown, the plot also contains the most number of data points due to the high resolution. This allows for a lot of information to be

gathered about the particular region of interest, but limits the region to a fairly small area without higher processing power.

B.3 TABLES

Table B-1: Examples AutoME processing parameters.

[Version]					
Version	3.0.9				
[Chromatogram]	**Value**	**Units**	**[SpectralThresholding]**	**Value**	**Units**
Interval	10	scans	Mode	Abs	
UseAll	FALSE		SpecThresh	0	
From	1000	scans	TopX	10	
To	15100	scans			
AdvancedThresholding	TRUE		**[Outputs]**		
Threshold	250	unitless	OutputRaw	FALSE	
			OutputCentroid	TRUE	
[SpeMaxEnt]			ThresholdCentroid	TRUE	
InputRangeLow	Acquisitio		AppendCentroid	TRUE	
InputRangeHigh	-NA-		Bin	TRUE	
OutputMassIncrement	1	da	BinFrom	2500	da
MinLeftIntensityRatio	33	unitless	BinTo	65000	da
MinRightIntensityRatio	33	unitless	BinSize	1	da
OutputMassRanges	3000:60000	da	BinZeroFill	TRUE	
GaussianHalfWidth	0.75	unitless	Deharmonize	TRUE	
SpectrometerHalfWidth	FALSE		DeharmTolerance	5	
MaxIterations	5	unitless	DeharmPresOriginal	FALSE	
			Location	C:\AutoME	
[SpeSubtractPost]			Centerby	HEIGHT	
Dosubtract	FALSE				
MaxPolyOrder					
BelowBackground					

B.4 FIGURES

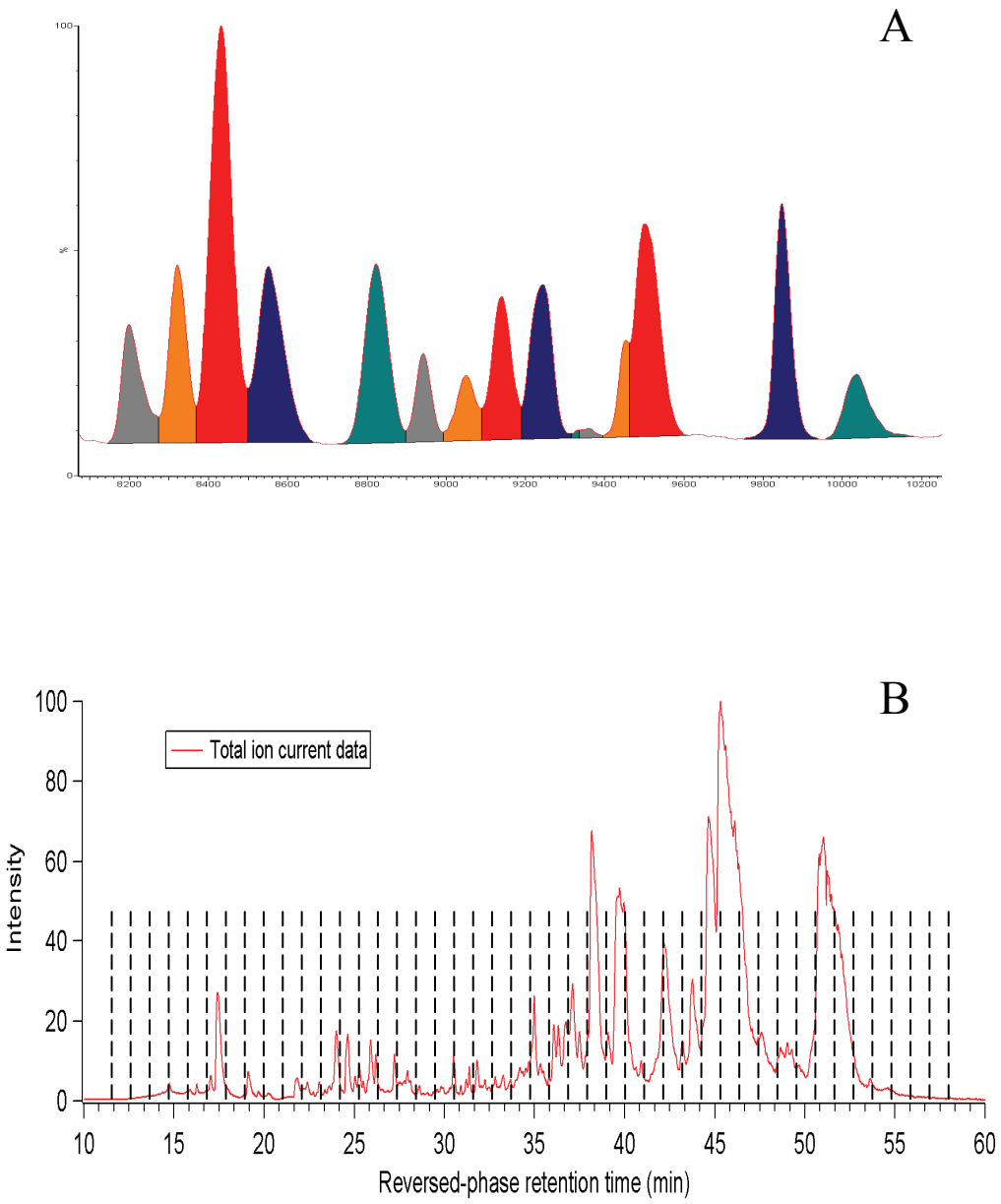


Figure B-1: MaxEnt data processing approaches. A) Illustration of peak based AutoME Approach (Courtesy Waters Corp). B) Time Segment based approach. Segments are enlarged for illustration purposes; typical segment width ~0.1 min

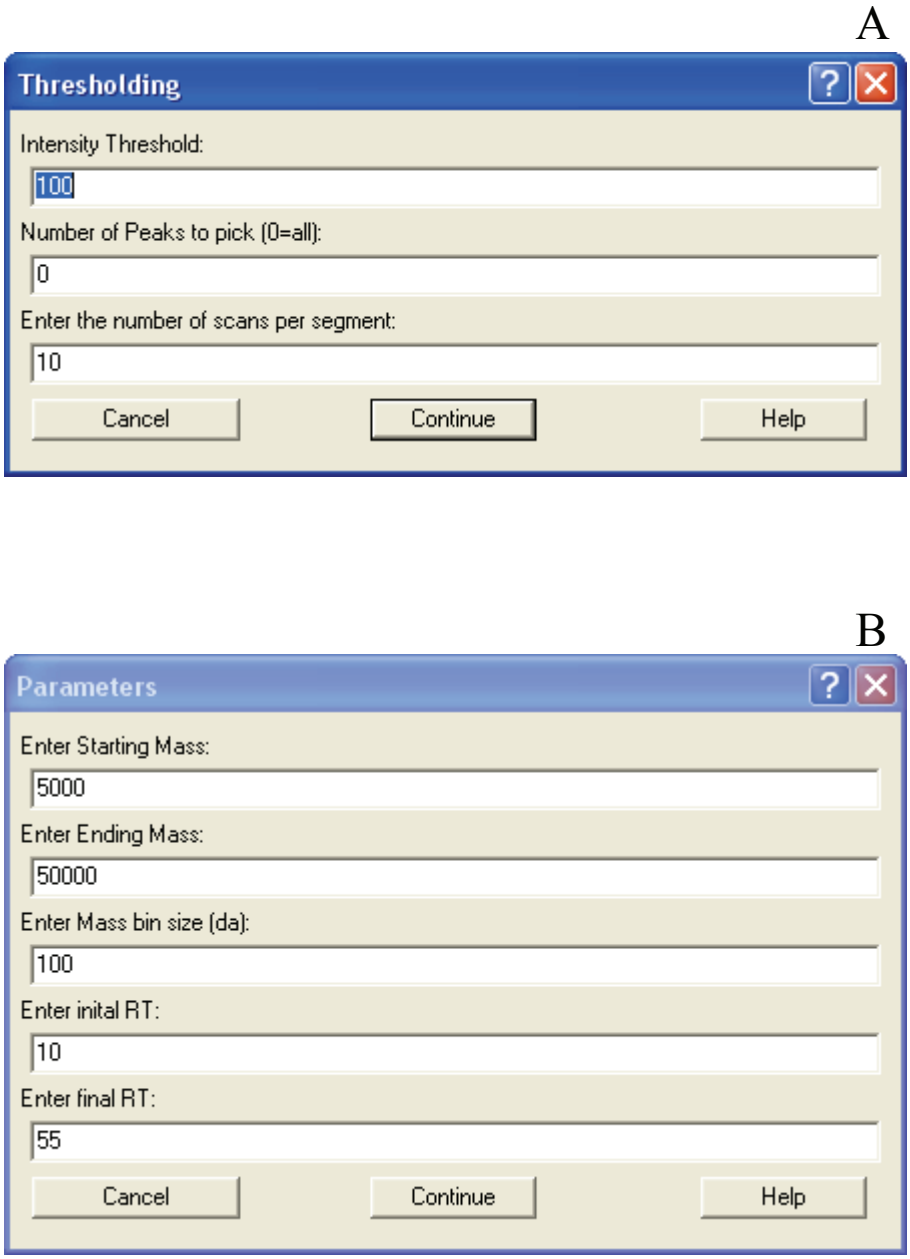


Figure B-2: Igor dialogs for A) Loading and B) Plotting AutoME data.

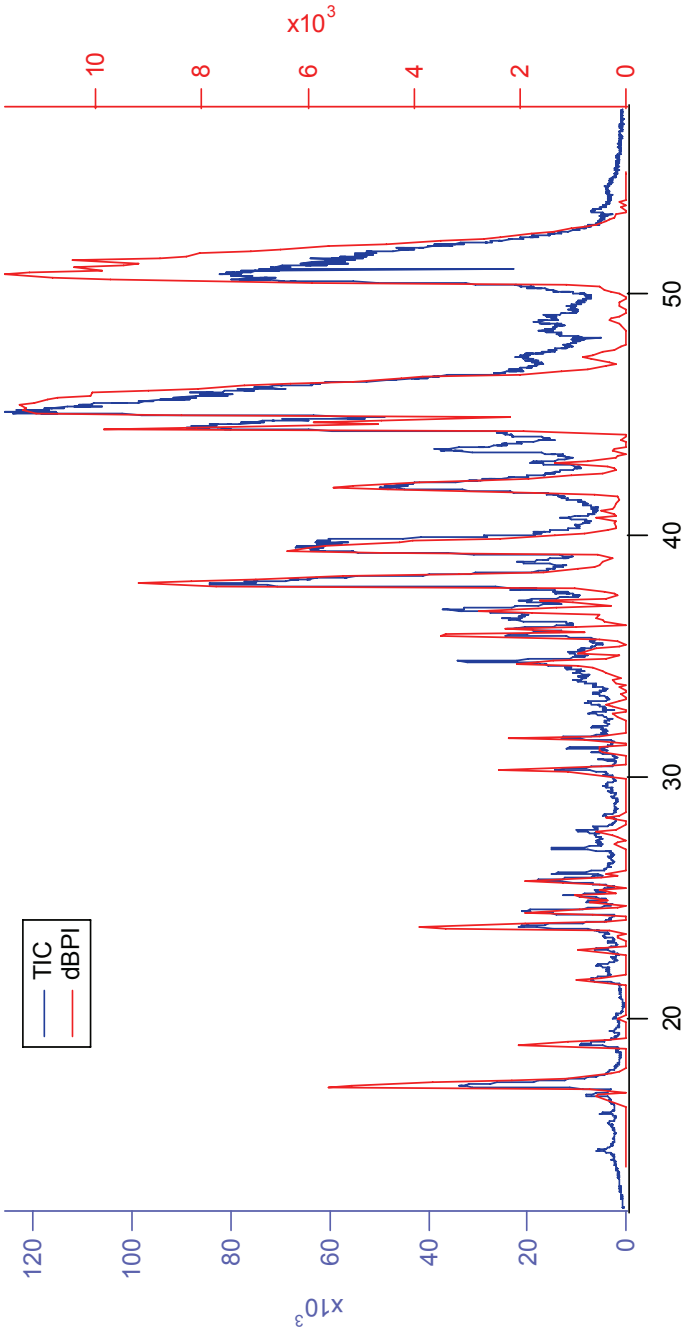


Figure B-3: Overlay of original TIC and processed dBPI. Sample is an *E. Coli* lysate fraction that was first collected from a conventional anion exchange run before analysis by RP-gUHPLC

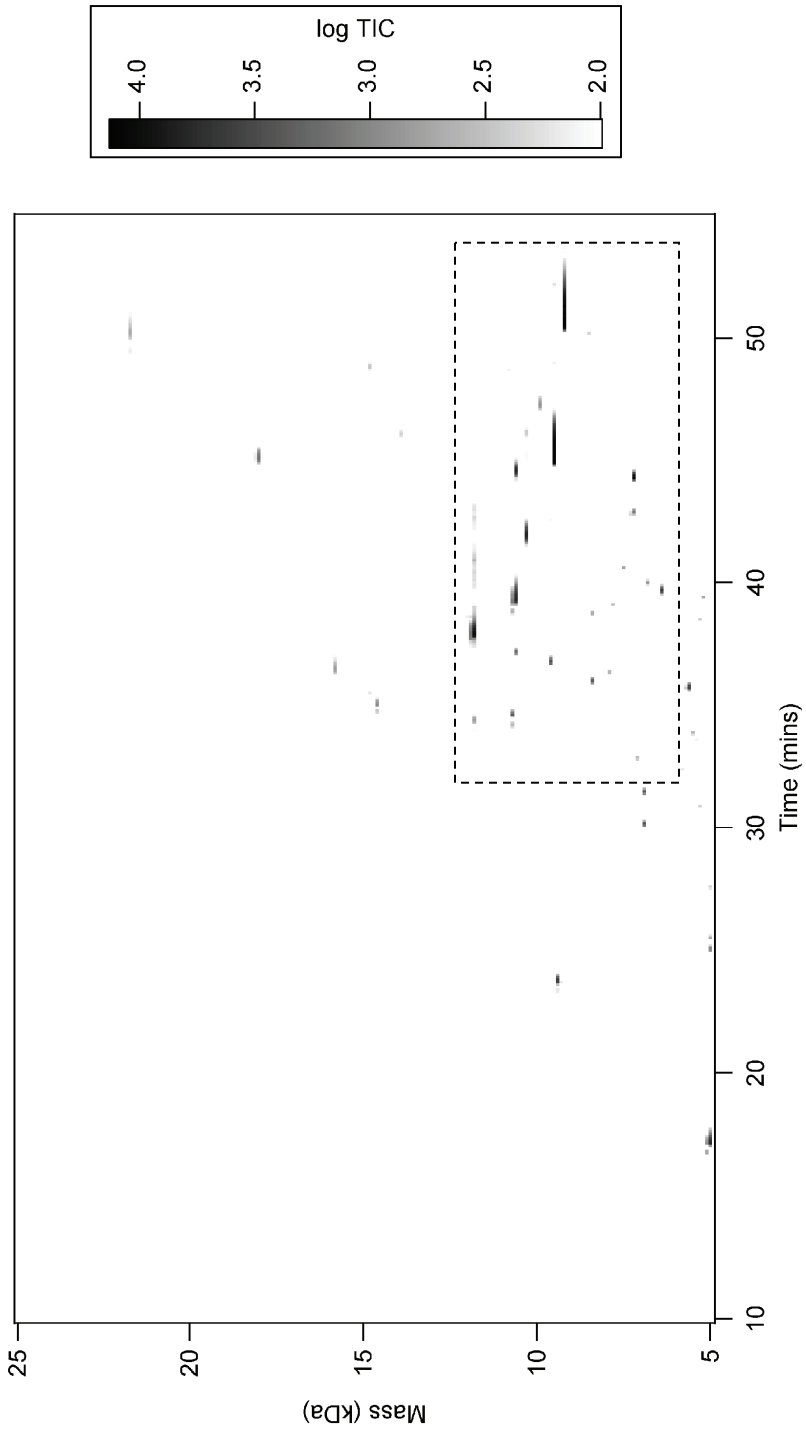


Figure B-4: Low resolution, survey ChromatoSpectra. Highlted area is show in Figure B-5 at higher resolution. MW range: 5-25 kDa, RT range: 10-55 mins, bin size: 100 da. Data points: 90,000

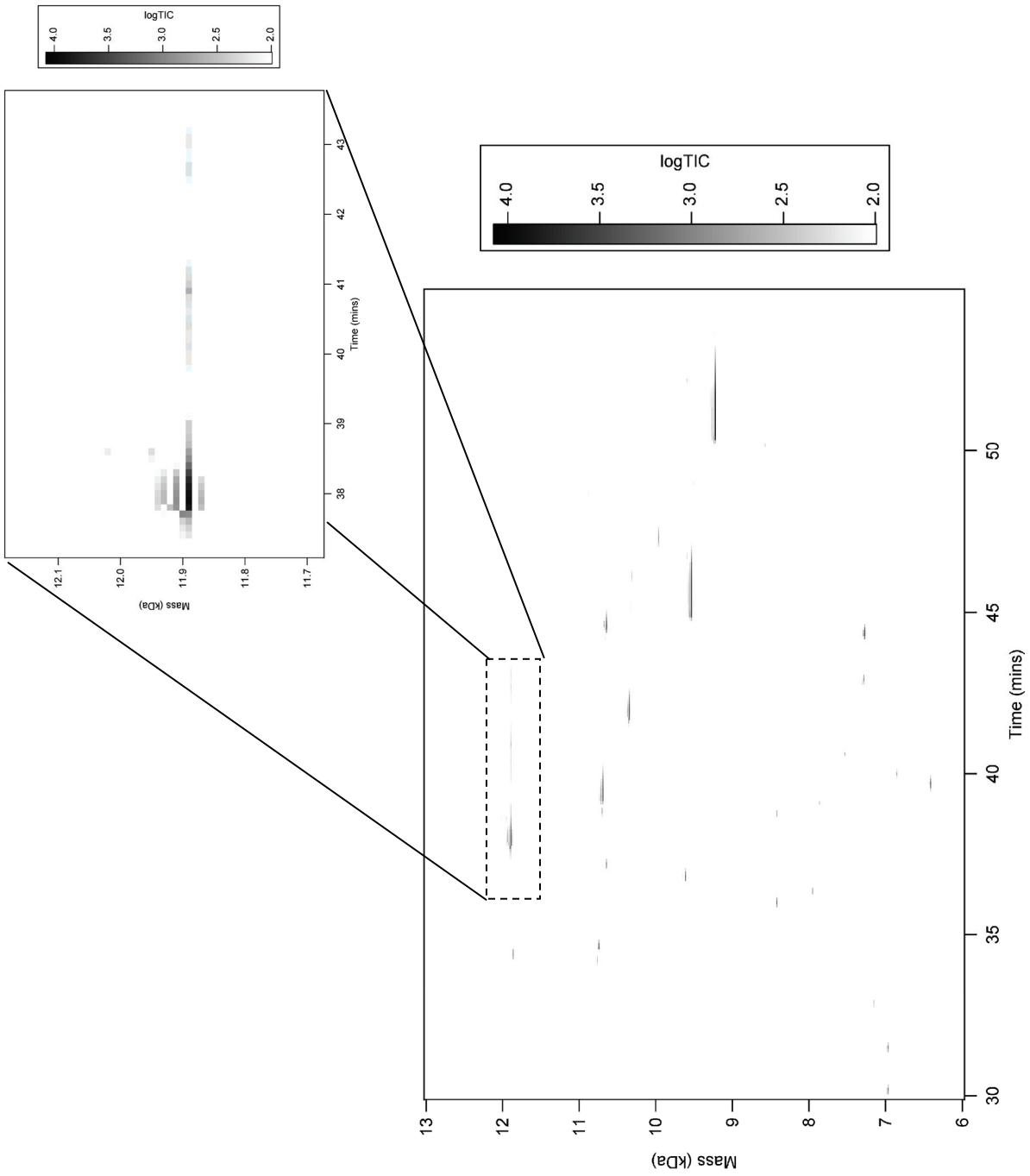


Figure B-5: Medium resolution ChromatoSpectra. Callout is zoomed area at same resolution. Higher resolution shown in Figure B-6. MW range: 6-13 kDa, RT range: 30-55 mins, bin size: 10 da. Data points: 175,000

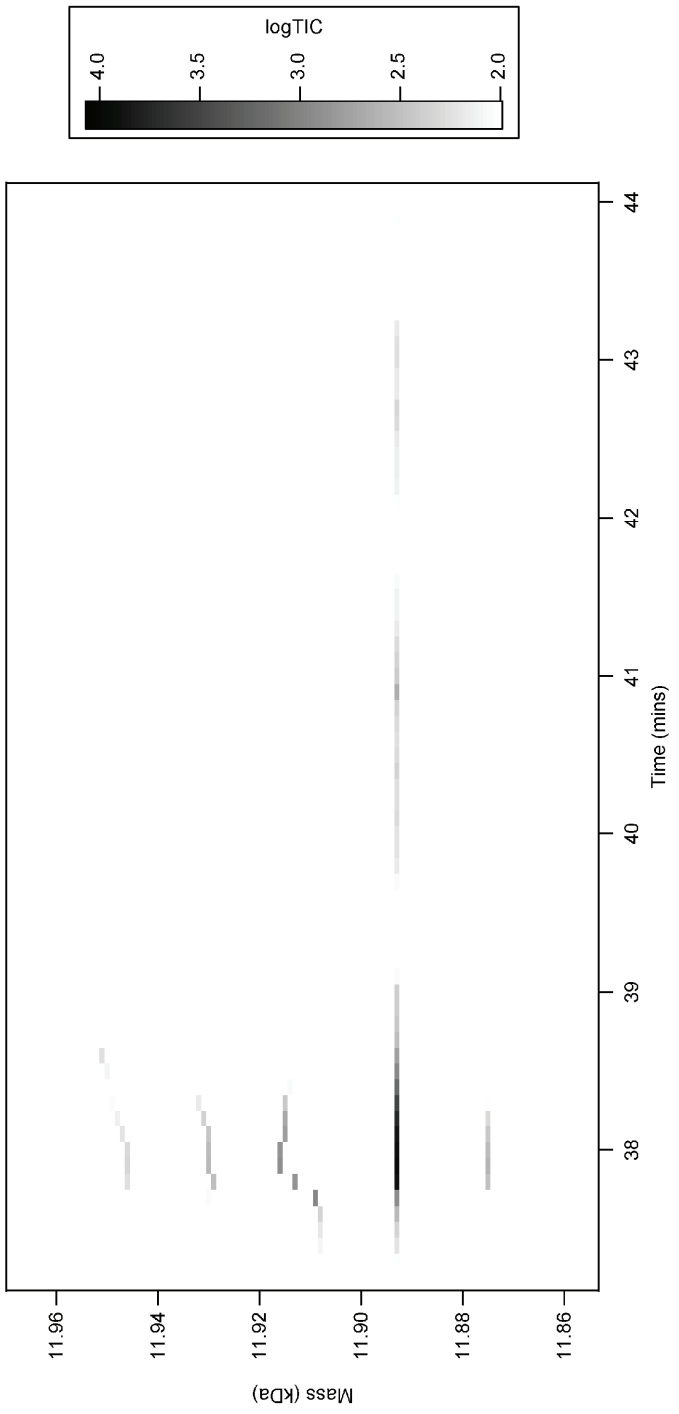


Figure B-6: High resolution ChromatoSpectra of zoomed area in Figure B-5. MW range: 10-13 kDa, RT range: 37-44 mins, bin size: 1 da. Data points: 210,000

APPENDIX C
AVS/EXPRESS PROGRAM DIAGRAMS FOR ESI VOLUME VISUALIZATION

C.1 AVS PROGRAMS

Program diagrams for Cropped Isovolume View (Figure C-1), Multi-Isosurface Viewer (Figure C-2) and Profile Extraction Viewer (Figure C-3) are shown below. Table c-1 summarizes the various AVS modules used to create the visualizations.

Each visualization method shares a common data import procedure. Data is saved by LabView as a tab-delimited text file for import into AVS. All imported data is first converted to an AVS Uniform Mesh before it can be used by the various visualization modules.

C.1.1 Uniform Mesh Creation

Data is obtained as a scattered field when collected as defined by volume V and must be converted into a uniform mesh before mapping in AVS. The array created by V is stored as a 2D text array consisting of i x j x k points that is read into AVS as a table and subsequently converted to a scattered field. A scattered field does not have any regular mesh or grid associated with it and cannot be used for more advanced AVS functions such as volume rendering. The scattered field must be converted into a uniform mesh with a regular spacing. This conversion is accomplished with an AVS mapper module (scat_to_unif) that allows the user to define the grid dimensions and method of interpolation. Although this module was able to successfully convert the field into a uniform mesh, significant artifacts were present

as a result of the conversion due to the interpolation parameters. It was empirically decided that it is best to not try to interpolate more points then were actually collected. This required some additional analysis of the data file, but resulted in the best interpolation possible.

C.1.2 Cropped Isovolume

The cropped isovolume is created by applying a crop box to a standard AVS isovolume. Once the data has been converted to an Uniform Mesh, it can be directly plotted as an isovolume. The crop box is overlayed on the volume and used to remove a given portion.

C.1.3 Isosurface Viewer

The orthoslice function can be applied to create the 2D slice plane view (Figure 5-9a) from a standard AVS Mesh. Using this view in conjunction with the Probe module and user interaction, a single value can be extracted. This value is then passed to the isosurface_trace and isoline_trace functions which handle drawing the surface in volume window and the line in surface window. The xform functions are used through in order to properly synchronize and desynchronize the view as necessary.

C.1.4 Profile Extraction

To create an extracted 1D profile, the AVS Mesh is first converted to a 2D orthoslice as in the previous section. By taking a second orthoslice of this file, a 1D slice results which displays the extracted profile. Simply, the profile is a “slice of a slice”.

C.2 TABLES

Table C-1: AVS modules used and corresponding function.

Module	General Function	Algorithm (if used)
Rd_Txt_Columns	Reads an ASCII file into AVS as a table	None
Table_to_scatter_filed	Converts ASCII table into a AVS field file	None
Scat_to_unif	Interpolates the scattered field to a uniform grid	User controlled order function
Data_math	Used to invert sign of data (negative to positive)	User controlled input function
Bounds	Creates bounding box	None
Orthoslice	Creates ortho-axis slice plane	None
Reset_xform	Reset the AVS field transformation for proper display (desynchronize)	None
Set_xform	Set the AVS field transformation to a reference field (synchronize)	None
Probe	User interactive data probe for selecting values	None
Isosurface_trace	Creates an isosurface at a given probe value	Marching Cubes
Isoline_trace	Creates an isoline trace at a given probe value	Marching Cubes

C.3 FIGURES

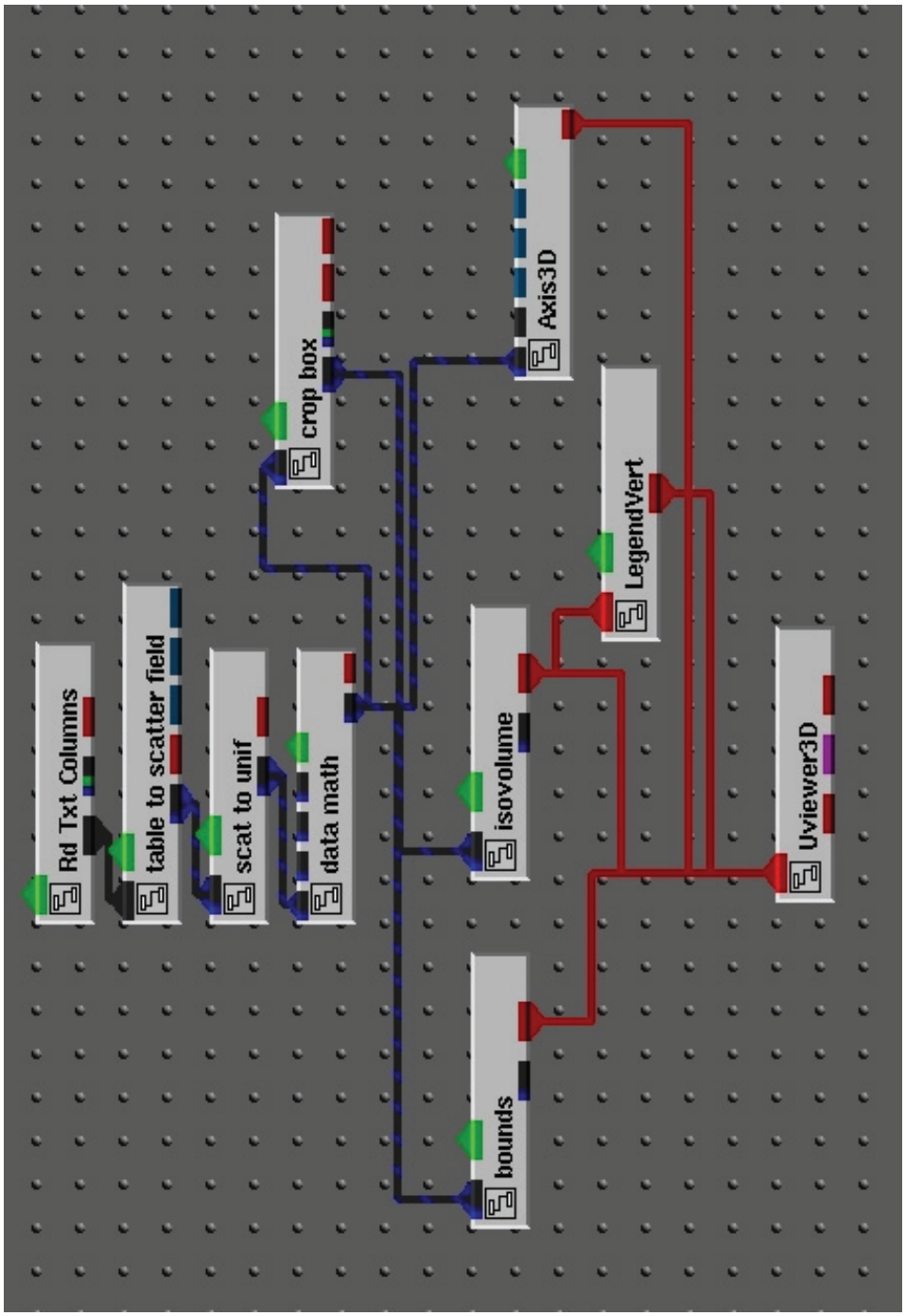


Figure C-1: Crop Isovolume Box (Section 5.5.1)

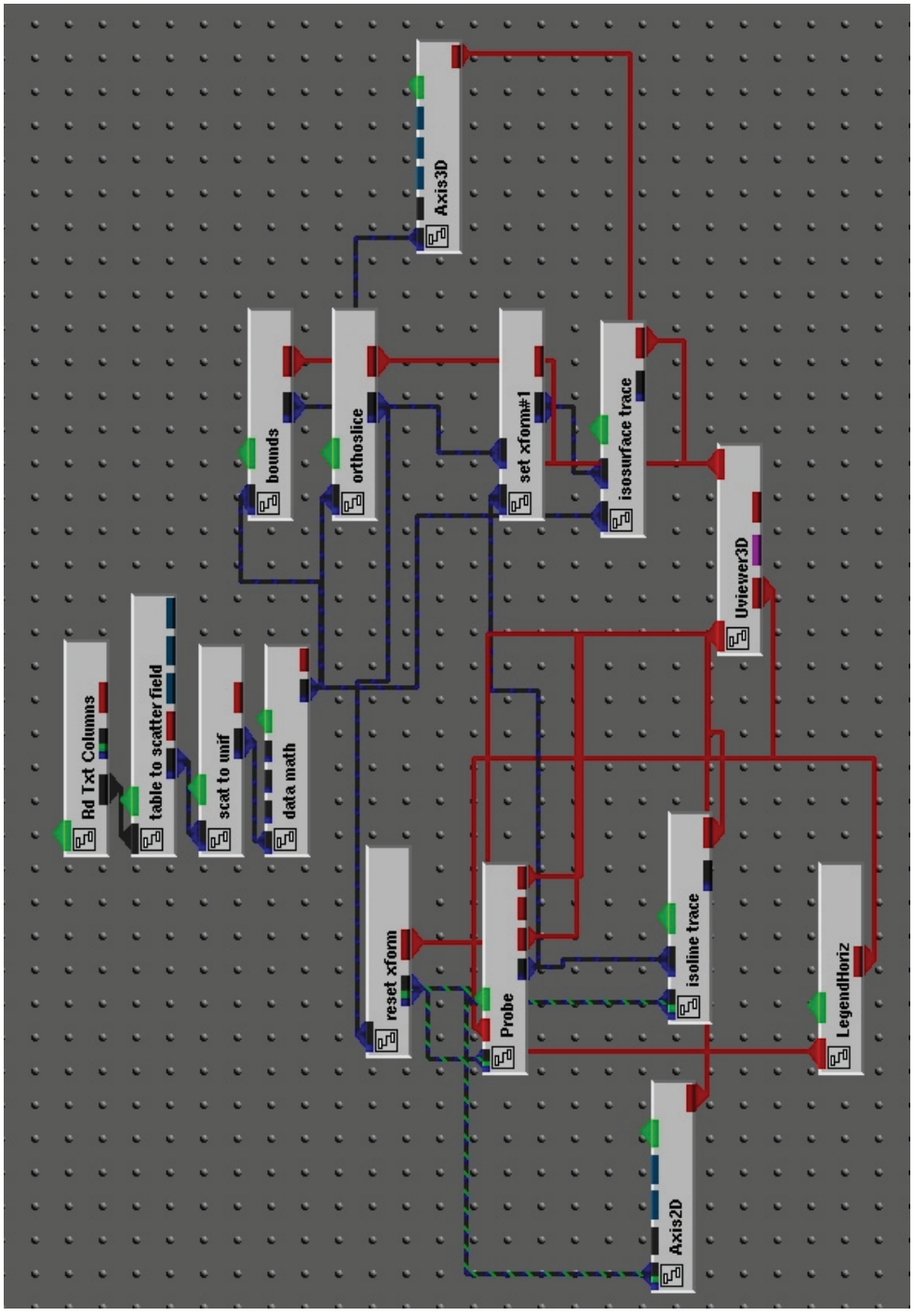


Figure C-2: Plane Viewer (Section 5.5.2)

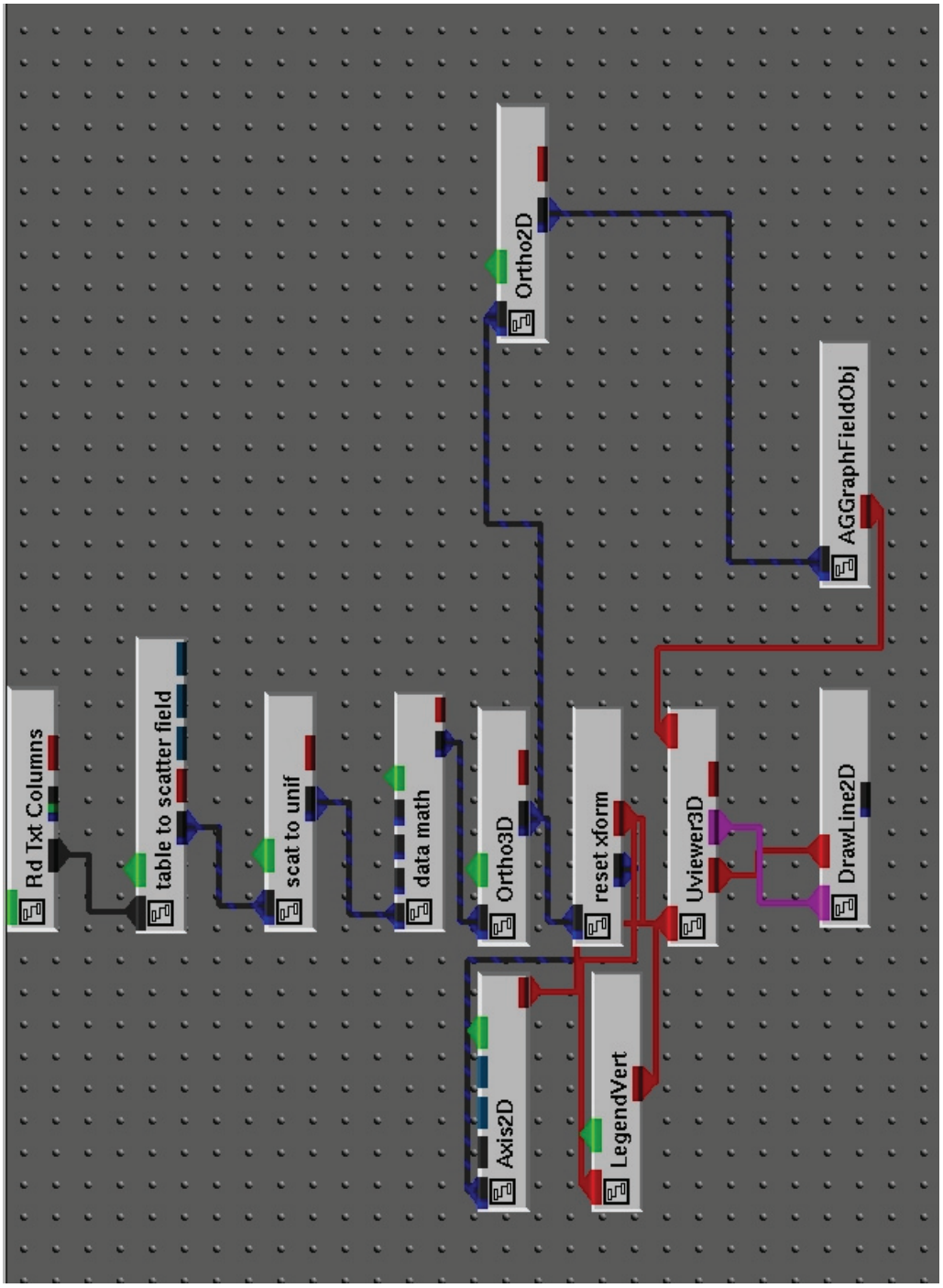


Figure C-3: Profile Extractor (Section 5.5.3)