Improving Our Charged World: Development of Peptide Charge State Enhancement Methodologies

Marshall Paul Chaffee
Fullerton, California

B.S., Chemistry, University of Redlands, 2010
B.S., Biology, University of Redlands, 2010

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Master of Science

Department of Chemistry

University of Virginia
May, 2013

Donald F Hunt

B. Jill Venton

Preface

Acknowledgements

By writing this thesis, I am charting a radically different course than the vast majority of students who have come through our lab over the past several decades. My decision to get to this point was not an easy one, and I am very grateful for Dr. Donald Hunt's willingness to offer me the flexibility in choosing the course that is right for me. I learned many valuable skills during my time in Dr. Hunt's lab, and I am sure these will serve me well as I move to the next adventure in life. Dr. Jeffery Shabanowitz also played a large part in building this skill set I would not have been able to acquire elsewhere. I am especially grateful to the fun loving group of people whom I had the pleasure of sharing my lab space with, and whom put up with my constant vigilance as lab safety officer (which I am sure was annoying at times).

I would like to specifically thank Dr. Michelle English, and Dr. Weihan Wang for their guidance and thought-provoking discussions while I was carrying out the method development work for my thesis. The vast majority of experiments I tried in this arena ended with little success, and it was always very helpful to have a second opinion when trying to determine what went wrong and figure out the next experiment. I would also like to thank Dr. Joseph Strukl for his mentorship and willingness to provide assistance, even when it was outside his area of expertise. I learned something from each and every person in the lab (Dr. Jenn Abelin, Lissa Anderson, Dr. Dina Bai, Dr. Aaron Bailey, Dr. Jessica Chapman, Dr. Jenny Cottine, Andrew Dawdy, Dr. Sushmit Maitra, Stacy Malaker, Amanda Wriston, Lichao Zhang, Dr. Linda Zhang), and for this I am very

As is the case for most grad students, we all find things to do outside of the lab to help clear our minds from long days and late nights in the lab. To this end, I have had the pleasure of being a member of the Charlottesville-Ablemarle Rescue Squad. The camaraderie and friendship I found with my fellow SDC crewmembers was second to none, and really helped provide a reprieve from the day-to-day routine of being a grad student. My experiences with the rescue squad added a unique flavor to my experience here, and I will take this with me wherever I go.

I also wish to thank my family for their love and encouragement during my time in grad school. Without their support, this thesis would not have been possible. I know it has been difficult being on opposite coasts, and I really appreciate all the trips you all made to Charlottesville during my time here.

Caitlin, your companionship means the world to me. I appreciate you putting up with me these past couple months as I put this thesis together!

Table of Contents

Preface ..... **ii**
Acknowledgements ii
Table of Contents iv
List of Tables and Figures vi
Abbreviations viii
Abstract xii

Chapter 1 – Introduction to the Dissertation and Research Objectives **1**
1.1 Introduction 1
1.2 Why Study Proteins? 1
1.3 Key Methodologies and Workflow in MS-Based Proteomics 4
1.4 Peptide Sequence Determination Using Mass Spectrometry 6
1.5 Developments in Peptide Charge State Enhancement 13
1.6 Aims of Research 16
1.7 References 16

Chapter 2 – Method Development **19**
2.1 Background **19**
2.2 Materials & Reagents **24**
2.2.1 Reagents 24
2.2.2 Column Making Materials 26
2.2.3 Laboratory Equipment 27
2.3 Methods **28**
2.3.1 Peptide Derivitization Attempt with Benzimidazole 28
2.3.2 Acetone Wash Procedure for Benzimidazole Reaction 28
2.3.3 Attempt to Adapt Peptide Derivatization Method by Xu et al[1] 28
2.3.4 Modified EDC/HOAt Peptide Derivatization Method 29
2.3.5 Tryptic Digest of Bovine Serum Albumin (BSA) 30
2.3.6 Assembly of HPLC Columns 30
2.3.7 Loading and Rinsing of Samples 31
2.3.8 Cleanup of Digested Protein Samples 32
2.3.9 Charge Enhancement Chemistry Reaction with BSA Digest Peptides 32
2.3.10 Analysis of Samples by Mass Spectrometry 33
2.4 Results and Discussion **33**
2.4.1 Benzimidazole Reaction Cleanup with Acetone Wash 33
2.4.2 Attempt to Adapt EDC/HOAt-Reaction as Described by Xu et al[1] 35
2.4.3 Examination of reaction pH 38
2.4.4 Adaption EDC/HOAt-Reaction as Described by Xu et al[1] Using Microelectrode 39
2.4.5 Investigation of pH and Its Effect on Reaction Yield 40
2.4.6 Examination of Data from Optimized Yield Conditions 42
2.4.7 Study of Reaction Solvent Conditions and Necessity of HOAt 45
2.4.8 Study of Reaction Yield Over Time 48
2.4.9 Trial of Optimized EDC/HOAt Reaction on a Whole Protein Digest 50
2.4.10 Evaluation of Methyl Benzimidazole as a Peptide Derivatization Reagent 55
2.4.11 Comparison of Derivatized Peptide Retention Time 57
2.4.12 Evaluation of Methyl Benzimidazole as a Charge Enhancement Reagent 61

2.4.13 Further Examination of Incubation Time on Reaction Outcome..... 63
2.5 Conclusion.....65
2.6 References 67

List of Tables and Figures

Chapter 1

Figure 1.1. ETD spectra of vasoactive intestinal peptide 1-12 (vaso). (8)

Figure 1.2. Proposed mechanism for the loss of phosphate during CAD-based peptide fragmentation. (9)

Figure 1.3. Proposed ETD mechanism on a multiply protonated peptide, following the capture of an electron. (10)

Figure 1.4. Diagram showing bond cleavage sites along a peptide backbone. (12)

Scheme 1. Derivatization of a free carboxylic acid using carbodiimide cross-coupling chemistry. (14)

Figure 1.5. Extracted ion chromatogram comparing retention time and charge state distributions of fully histamine-derivatized and underivatized vaso (HSDAVFTDNYTR). (15)

Chapter 2

Figure 2.1. Comparison of three amines used for charge enhancement chemistry. (19)

Figure 2.2. Resulting chromatographic data from the analysis of a 1 pmol aliquot from a benzimidazole-derivatization reaction of vaso (HSDAVFTDNYTR). (21)

Scheme 2.1. Derivatization of a carboxylic acid using EDC/HOAt-mediated chemistry. (23)

Table 2.1. Comparison of final reagent concentrations used in cross-coupling reactions for two different protocols (Xu *et al* and Hunt Lab). (23)

Figure 2.3. Resulting chromatographic data from the analysis of a 1 pmol aliquot from a combined benzimidazole derivatization/acetone wash procedure of vaso. (34)

Figure 2.4. Labeled base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from an EDC/HOAt/PP derivatization reaction at pH 7.5 (measured using test strips). (36)

Table 2.2. Comparison of pH measurements of test strips vs. a microelectrode. (38)

Figure 2.5. Resulting base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from an EDC/HOAt/PP derivatization reaction at pH 7.72 (measured using microelectrode). (39)

Figure 2.6. Graph illustrating the effect of pH on the EDC/HOAt cross-coupling reaction of YGGFL with histamine. (41)

Figure 2.7. Labeled base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from a two hour EDC/HOAt/Histamine derivatization reaction at pH 5.53 (measured using microelectrode). (43)

Table 2.3. Summary of observed reaction products. (43)

Figure 2.8. Comparison of cross-coupling reaction yield over several different conditions. (47)

Figure 2.9. Examination of reaction yield of three different peptides over time. (49)

Figure 2.10. Labeled base peak chromatogram from the analysis of 500 fmols of BSA digested with trypsin and derivatized with histamine at pH 5.5 for 40 minutes. (52)

Figure 2.11. Sequence of BSA showing regions covered using only histamine-derivatized peptides from tryptic digest. (53)

Table 2.4. Digest derivatization yield estimate. (54)

Figure 2.13. Labeled base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from a two hour EDC/HOAt/methyl benzimidazole derivatization reaction at pH 5.38 (measured using microelectrode). (56)

Table 2.5. Summary of observed reaction products. (53)

Figure 2.14. Analysis of effects on peptide retention time upon derivatization with histamine, benzimidazole, and methyl benzimidazole. (59)

Figure 2.15. ETD spectra of the +3 ion DFNKFHpTFPQTAIGV (underivatized). (62)

Figure 2.16. ETD spectra of the +4 ion DFNKFHpTFPQTAIGV (derivatized with methyl benzimidazole). The chart at the top of the figure lists all possible +1 c and z fragment ions. (63)

Table 2.6. Summary of reaction products from 30 second and 5 minute methyl benzimidazole/EDC/HOAt reactions. (64)

Abbreviations

ºC	Degrees Celcius
·	Radical Species
Å	Angstroms
A	Solvent A, 0.1 M acetic acid in water
ACN	Acetonitrile
Ala, A	Alanine
Angio	Angiotensin I, human
Ambic	Ammonium bicarbonate (buffer)
Arg, R	Arginine
Asn, N	Asparagine
B	Solvent B, 0.1 M acetic acid in water
BSA	Bovine serum albumin
c	Centi (1×10^{-2})
C-term	Carboxyl terminus of a peptide
C18 ODS	Octadecyl silane
CAD	Collision-activated dissociation
Cys, C	Cysteine
D	Aspartic acid
Da	Dalton

DMF	N,N-Dimethylformamide
DTT	Dithiothreitol
E	Glutamic acid
EDC	1-ethyl-3-(3-dimiethylaminopropyl) carbodiimide
ESI	Electrospray ionization
ETD	Electron-transfer dissociation
F	Phenylalanine
f	Fempto ($1x10^{-15}$)
FETF	Front-end electron transfer dissociation
FT-ICR	Fourier transform ion cyclotron resonance
FT-MS	Fourier transform mass spectrometry
g	Gram
Gln	Glutamine
Glu	Glutamic acid
h, hr	Hour
HCl	Hydrochloric acid
His, H	Histidine
HOAc	Acetic Acid
HOAt	1-Hydroxy-7-azabenzotriazole
HPLC	High-performance liquid chromatography
i.d	Inner diameter
IAM	Iodoacetamide

Ile, I	Isoleucine
K	Lysine
Kasil	Potassium silicate solution
L	Liter
LC	Liquid chromatography
Leu, L	Leucine
Lys	Lysine
µ	Micro ($1x10^{-6}$)
m	Mili ($1x10^{-3}$)
M	molar
M + nH	Molecular ion with n charge(s)
MeOH	Methanol
Met, M	Methionine
min	Minute
mol	mole (6.022×10^{23})
MS	Mass spectrometry
MS/MS	Mass spectrometry
MS1	Full mass spectrum
MS2	Tandem mass spectrum
MW	Molecular weight
n	Nano ($1x10^{-9}$)
N-term	Amimo terminus of a peptide

NaOH	Sodium hydroxide
o.d	Outer diamater
p	Pico ($1x10^{-12}$)
PP	1-(2-pyrimidyl) piperazine
Phe	Phenylalanine
Pro, P	Proline
psi	Pounds per square inch
pS, pT, pY	Phosphorylated residue (serine, threonine, and tyrosine, respectively)
PTM	Post-translational modification
RF	Radio frequency
s	Seconds
Ser, S	Serine
Thr, T	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val, V	Valine
Vaso	Vasoactive intestinal peptide fragment 1-12, human
W	Tryptophan
Y	Tyrosine

Abstract

In mass spectrometry (MS), electron transfer dissociation (ETD) is often employed to generate fragments used for determining a peptide's sequence. ETD is a preferred fragmentation technique because labile post-translational modifications (PTMs) such as phosphorylation are left intact, which allows for straightforward localization. However, the ability to successfully utilize ETD is highly dependent on a peptide's charge state, which excludes some peptides of interest. To help get around this problem, our lab has been able to develop a carbodiimide-based derivatization strategy to covalently link histamine molecules to free carboxylic acid residues found on peptides. Once attached, the imidazole ring on histamine provides an additional site for positive charge to be retained. One drawback to using histamine is that upon derivatization peptides become substantially more hydrophilic, which can lead to problems retaining peptides on the reverse-phase chromatography columns typically utilized in conjunction with MS-based detection systems.

The work presented for this defense is an attempt to resolve the problem of increased hydrophilicity. To accomplish this task we have identified several molecules with similar functionality to histamine, but which also contain hydrophobic functionalities. Unfortunately, these chemicals contain undesirable contaminants, and in order to use them, we had to develop a new derivatization strategy. We ultimately developed a cross-coupling strategy, which made use of 1-Hydroxy-7-azabenzotriazole (HOAt) and N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC) to covalently attach 2-(7-Methyl-1H-benzimidazol-2-yl) ethanamine (Methyl

Benzimidazole) free carboxylic acid residues. Once attached, we were able to demonstrate that methyl benzimidazole increases peptide charge while also increasing hydrophobicity. We believe this charge enhancement reagent shows great promise for improving ETD-based sequencing of peptides containing PTMs of interest, which have previously been difficult to localize.

Chapter 1 – Introduction to the Dissertation and Research Objectives

1.1 Introduction

In addition to presenting the basis for peptide charge enhancement method development research, this introduction also serves to give the reader a general overview of mass spectrometry and its applications for studying proteins. This background will help to provide relevant context, and point out current limitations of mass spectrometry in determining protein sequence information. These limitations highlight the need for new methods to be developed, and thus serve as the primary basis for the research being presented. This work builds upon previous developments in our lab, and attempts to address technical hurdles that pose challenges in applying these methodologies to studying complex protein samples. Towards the end of the dissertation, a practical example will be presented which demonstrates that methodologies presented here can be used to study a complex sample.

1.2 Why Study Proteins?

In a cell, proteins are the biological effector molecules responsible for carrying out many functions necessary for life such as metabolism, transport, genetic regulation, and structural support. The instructions for producing these proteins are coded into an organism's DNA in the form of genes. In each human cell it is estimated there are 20,000-25,000 genes, which implies a great deal of complexity in cellular protein

populations.[2] Identifying the role of each protein, understanding how proteins interact and associate with one another, and determining which proteins are present in a given cell type are on-going studies in modern biology. The rewards of answering these fundamental questions can have many implications from spurring the development of new biotechnologies to developing novel strategies for disease treatment.

The area of biology working to answer some of these questions can be grouped into a field now known as proteomics. All cellular proteins comprise what is known as the proteome, which is defined as "the protein complement of a given cell at a specified time, including the set of all protein isoforms and protein modifications.[1]" this general definition implies an even greater amount of complexity than was originally implied by the number of genes within an organism or cell because each protein encoded by a gene can exist in many different forms. From the perspective of an analytical chemist attempting to separate and sequence individual proteins, this is an immense challenge.

If we assume that each gene produces one protein, a crude cellular protein extract will contain a heterogeneous mixture of tens of thousands of different proteins. However, as pointed out by the definition of a proteome, this is only the first level of sample complexity. At any given time during a cell's life a subset of proteins may exist in a covalently modified form that results from a process known as post-translational modification (PTM), or a truncated form may be present as a result of alternative splicing.[3] To add a further level of complexity, a cell's development stage has a large impact on the protein populations within it. For example, a developing embryonic stem cell will have a significantly different protein makeup compared to a mature hepatocyte.

It is also important to consider if a particular cell of interest happens to be under environmental stress, or in a disease state. This list is by no means comprehensive, but it does bring to light many of the possible factors that need to be considered when designing a proteomics experiment and interpreting the results.

Among the PTMs that can exist on a given protein, a type known as phosphorylation is particularly interesting because of its widespread presence though out the proteome. Of the 20 amino acid building blocks for proteins, this particular modification is known to occur on serine (S), threonine (T), and tyrosine residues (Y). It has been known for some time that phosphorylation plays a key role in regulating cell signaling processes.[4] In fact, it is estimated that 30% of proteins within a cell can be phosphorylated at any given time.[5] Of recent interest to many researchers is the means by which these phosphorylation-based signaling processes can transform a cell from a healthy to a diseased state. Cancer is one example where cell signaling has gone awry, and there is a mounting body of evidence that phosphorylation (or lack thereof) plays a key role in the disease process.[6] Given this finding, the ability to detect and localize a phosphorylation event on a protein in addition to determining its entire sequence would be highly desirable.

With all of the different factors involved, studying a proteome presents an immense challenge to the modern protein scientist. Fortunately, there have been key technological developments over the past several decades that have greatly aided proteomic studies. Initially, electrophoretic separations and associated blotting/staining techniques allowed scientists to separate and visualize complex protein mixtures, and

Edman sequencing has allowed for sequence determination of abundant proteins.[7,8] More recently, mass spectrometry (MS) and associated liquid chromatography (LC) separation technologies have allowed scientists to study ever more complex protein systems. The high dynamic range and sensitivity offered by MS, coupled with the resolving power of LC now allows scientists separate complex protein mixtures and determine the sequence of these proteins (including those with PTMs) all in one experiment.[9,10] The next sections will discuss in more detail some of these key methodologies and MS instrumentation used to study proteins.

1.3 Key Methodologies and Workflow in MS-Based Proteomics

To determine the sequence of an unknown protein, researchers often use a technique known as "bottom up" proteomics. This strategy employs enzymes to digest a whole protein into smaller peptide pieces. This mixture of peptides can then be separated using LC techniques, and subsequently introduced into a mass spectrometer for analysis as each peptide elutes from the LC column. By determining the sequence of each peptide using MS data, and repeating the digest experiment with different enzymes, overlapping peptide sequences from each digest can be used to deduce the sequence of an unknown protein.

One of the key breakthroughs that have enabled MS analysis of peptides is the development of an ionization technique known as electrospray ionization (ESI). Developed in 1985 by Fenn & coworkers, this technique facilitates ionization and desolvation of liquid-phase biomolecules (such as proteins and peptides), and has

successfully been shown to ionize molecules in excess of 100 kDa.[11,12] This very high molecular weight threshold easily accommodates peptides generated from a digest. Prior to development of this technology, chemical ionization (CI) and electron impact (EI) were the primary techniques used to generate ions for MS analysis, which are not suitable for ionizing peptides resulting from a digest. The large excess of energy imparted to molecules undergoing these ionization processes causes premature dissociation before a mass measurement can be taken. Furthermore, neither ionization technique is readily compatible with ionization of liquid-phase analytes (the preferred environment for peptides). Given ESI's very generous molecular weight range and ease of compatibility with liquid-phase chromatography systems, complex peptide mixtures from an enzymatic digest can be readily analyzed.

A typical analysis workflow starts with reverse phase high performance liquid chromatography (RP-HPLC) separation of digested peptides, which separates molecules based on hydrophobicity. As peptides elute in an acidic solvent, basic sites such as lysine (K), arginine (R), or the N-terminus become ionized. Immediately following elution from a chromatography column, a desolvation process begins after peptides pass through an ESI nozzle under atmospheric pressure. The voltage applied to the ESI nozzle results in the formation of charged solvent droplets (containing the peptides), which are directed towards an inlet to the mass spectrometer. The desolvation process is continued as droplets enter a heated capillary drawn under vacuum. Eventually, Coulombic repulsion in a rapidly evaporating droplet causes fission to occur, and results in the complete

removal of all solvent molecules from a peptide.[13] At this point, peptides now exist in the mass spectrometer as gas-phase ions.

These peptide ions are then introduced into a mass analyzer for an initial mass to charge (*m/z*) measurement to be taken. However, this initial measurement is not sufficient for the determination of peptide sequence information. In order to obtain this information, additional energy can be imparted to the now gas-phase peptides to induce a predictable set of fragmentation events within the peptide. The means by which these bond cleavages can be induced and interpreted are discussed in the next section.

1.4 Peptide Sequence Determination Using Mass Spectrometry

Once completely desolvated and in the gas phase, ions are focused and directed into a linear ion trap (which can store ions as well as function as a mass analyzer) for brief storage where they undergo cooling with helium gas. Next, these ions are ejected from the ion trap and moved into an Orbitrap mass analyzer where a high resolution *m/z* of each intact peptide ion is measured through the image current produced from the orbiting ions (MS1). Using a data dependent acquisition method, the most abundant ions are then isolated from the concurrently eluting sample and fragmentated back in the ion trap to produce an MS2 spectrum. With properly optimized instrument parameters, this process of obtaining an MS1, followed by an MS2, is easily compatible with a chromatographic time scale.

The information from the MS2 contains a series of ions resulting from predictable fragmentation of many copies of the same original peptide. In most cases, these

fragments result from cleavage of individual amino acid subunits. Since the mass of all 20 amino acids are known and the mass of the intact peptide is known, the difference between overlapping fragment ions can be used to determine a sequence of the peptide. From data obtained in the MS1 scan, an investigator is given sufficient information to determine the monoisotopic mass of an intact peptide ion. Based on the previously determined amino acid sequence, a predicted monoisotopic mass can be calculated and compared to the value obtained in the MS1 scan. If these values match, this provides a very high measure of confidence that a sequence has been determined correctly.

An example of the process for determining a peptide sequence is shown in **Figure 1.1**. In modern MS-based peptide sequencing, there are two primary methods by which these fragments can be generated for peptide sequencing: Collision-Activated Dissociation (CAD) and Electron-Transfer Dissociation (ETD).

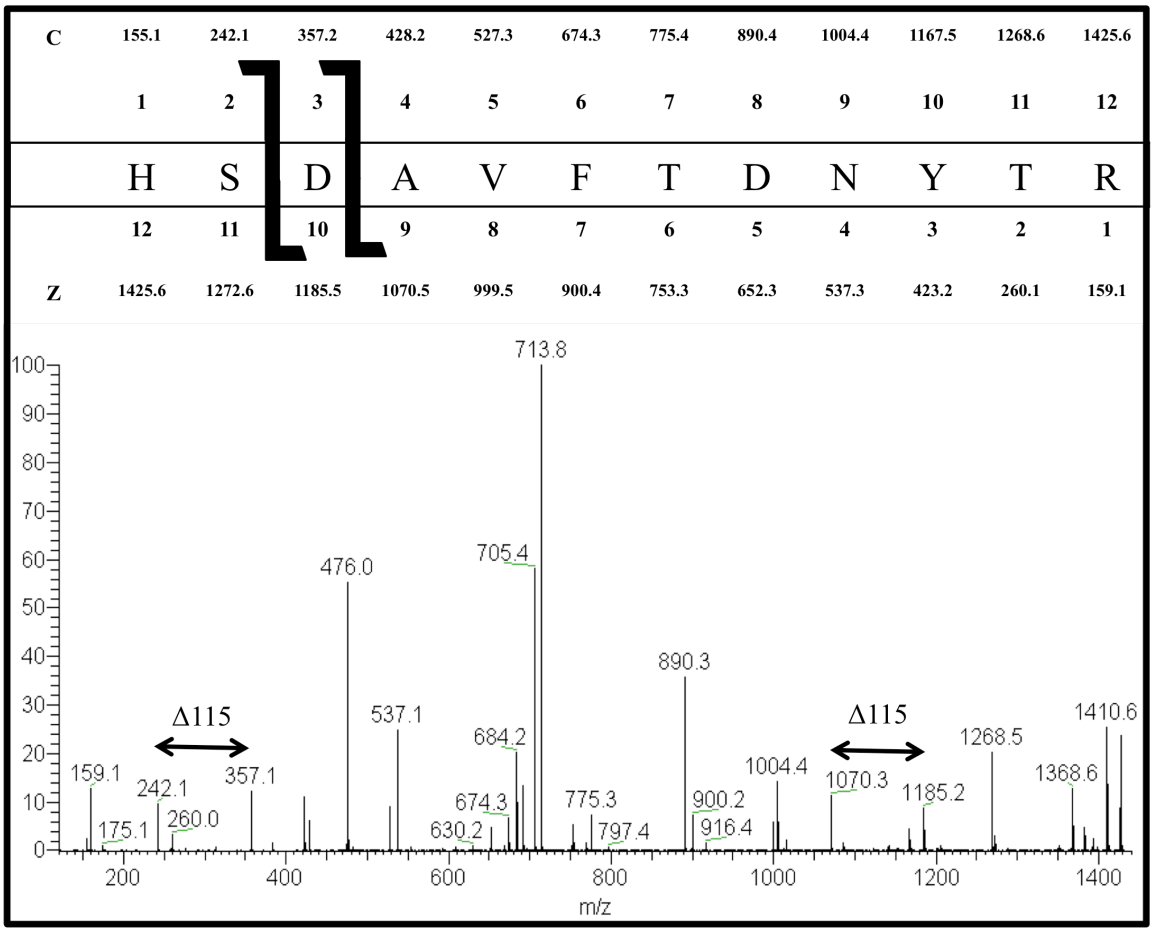


Figure 1.1. ETD spectra of vasoactive intestinal peptide 1-12 (vaso). This fragmentation spectra is an example of how a series of overlapping fragments can be used to determine a peptide sequence. In this example, fragments C_2 and C_3 have a mass difference of 115 which corresponds to the mass of aspartic acid (D). Aspartic acid is further verified by the Z-ions 9&10, which also have a mass difference of 115.

To fragment a peptide via CAD, all ions except the peptide of interest are ejected from the ion trap. Next, the peptide is subjected to a large number of low-energy collisions with helium gas in the ion trap. The resulting vibrational excitation causes random fragmentation of amide bonds along each peptide's backbone.[14] This produces a series of complementary ions referred to as type b (containing the N-terminus) and type y (containing the C-terminus). Once the fragmentation interval is complete, an *m*/*z* is obtained for all remaining ions in the trap. The resulting spectrum can then used to determine a sequence for the peptide as exemplified by **Figure 1.1**. However in addition

to bond cleavage along the peptide backbone, cleavage of some PTMs such as phosphorylation on serine/threonine (if present) also occurs readily.[15] A possible mechanism for this process is shown in **Figure 1.2**.

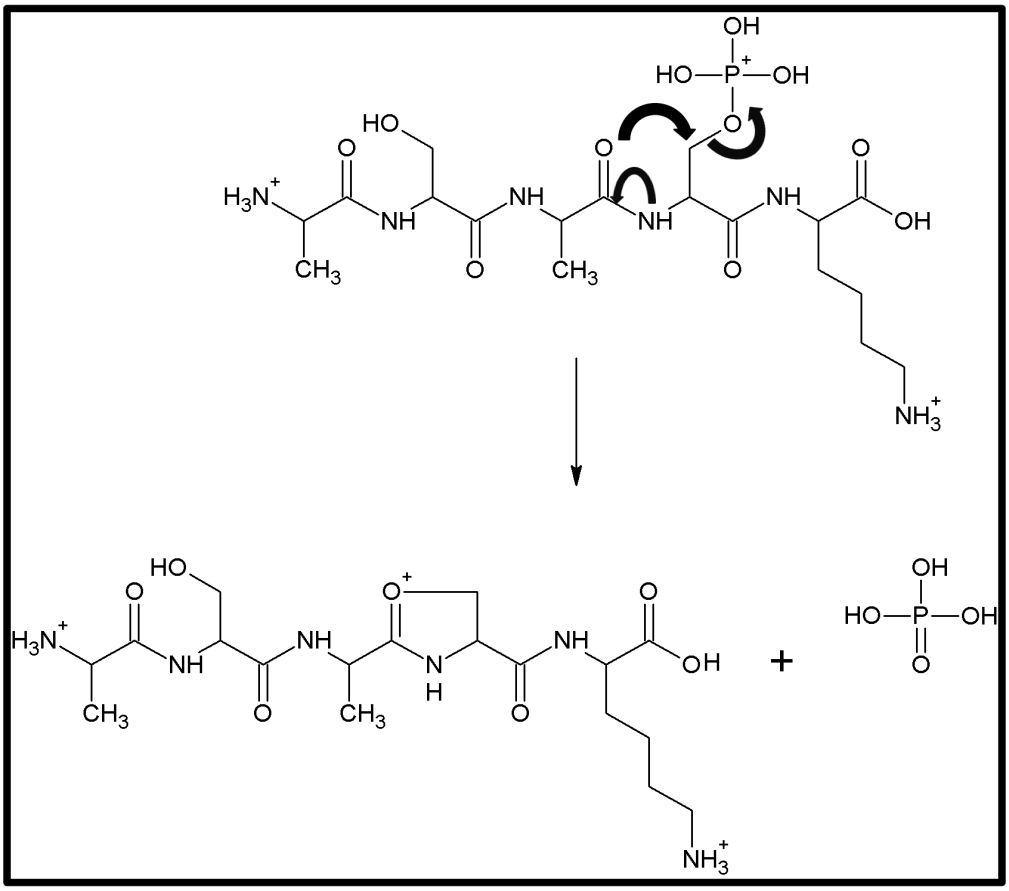


Figure 1.2. Possible mechanism for the loss of phosphate during CAD-based peptide fragmentation.[16]

If the peptide has multiple phosphorylation sites, the resulting fragmentation spectra is likely to be dominated by the loss of phosphate from the intact peptide ion, and contain few informative fragment ion. Additionally, CAD struggles to yield informative sequence information on highly charged peptides (>+3 charge state). Non-uniform distribution of basic sites on a peptide often causes fragmentation to occur in close proximity to these areas likely to contain charge, which seldom produce enough

informative fragments to determine a sequence. In this case of a highly charged peptide, ETD can be used as an alternative fragmentation option.

ETD is also performed in the ion trap by ejecting all ions except the ion of interest identified in the MS1. Next, the precursor ion is mixed with a radical, anionic form of a polyaromatic hydrocarbon (such as azulene). A radical electron is then transferred from the anion to the carbonyl group on the peptide backbone.[16] Once accepted by a peptide, the radical electron ultimately causes cleavage of the C-N_{α} bond on the peptide via a free-radical mechanism (**see Figure 1.3**).[16,17]

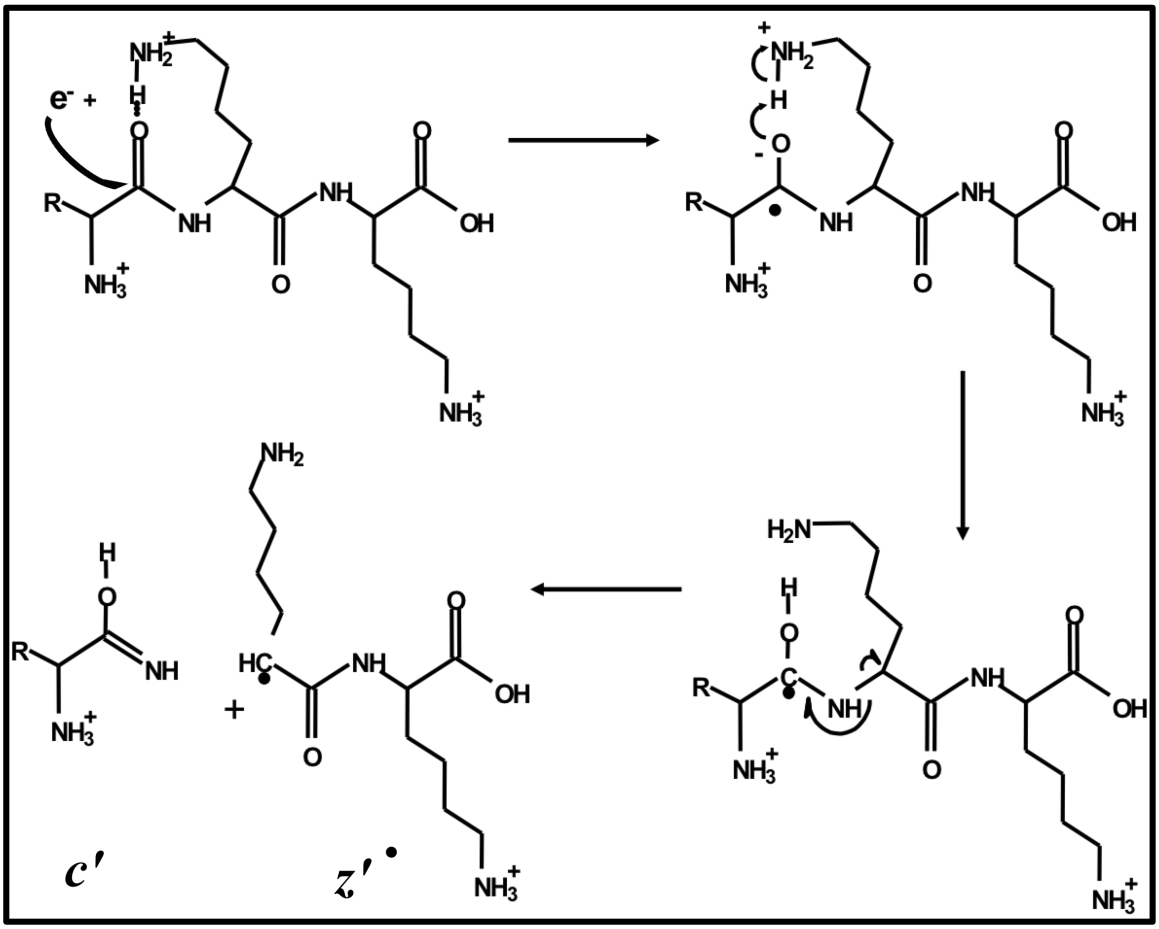


Figure 1.3. Proposed ETD mechanism on a multiply protonated peptide, following the capture of an electron. This process results in the formation of a c-ion (contains N-terminal fragment) and a z-ion (contains C-terminal fragment).[17]

As this process occurs along the peptide backbone, a complementary series of c (containing the N-terminus) and z ions (containing the C-terminus) are produced. Once a fragmentation cycle is complete, an *m*/*z* spectra is obtained for all remaining ions. By comparing the known mass of each amino acid to the mass of each fragment ion in the spectra, overlapping sequence information can be obtained. This will allow for a peptide sequence to be determined in much the same way as would be done for a CAD spectrum. However, because fragmentation is not based on collisional activation, labile post-translational modifications (PTMs) such as phosphorylation are left intact by the free-radical cleavage process.[15] Unfortunately, as shown by the fragmentation mechanism in **Figure 1.3**, one charge is consumed each time a bond cleavage occurs. In order to detect the resulting fragment ions, the peptide must have at least one other positively charged group present on each resulting fragment. For this reason, ETD is best carried out on peptides with two or more positive charges. Otherwise, half of the peptide fragments resulting from ETD are unlikely to be detected by the mass spectrometer, and sequence coverage of the peptide will be limited. A summary of cleavages resulting from CAD and ETD is shown in **Figure 1.4**.

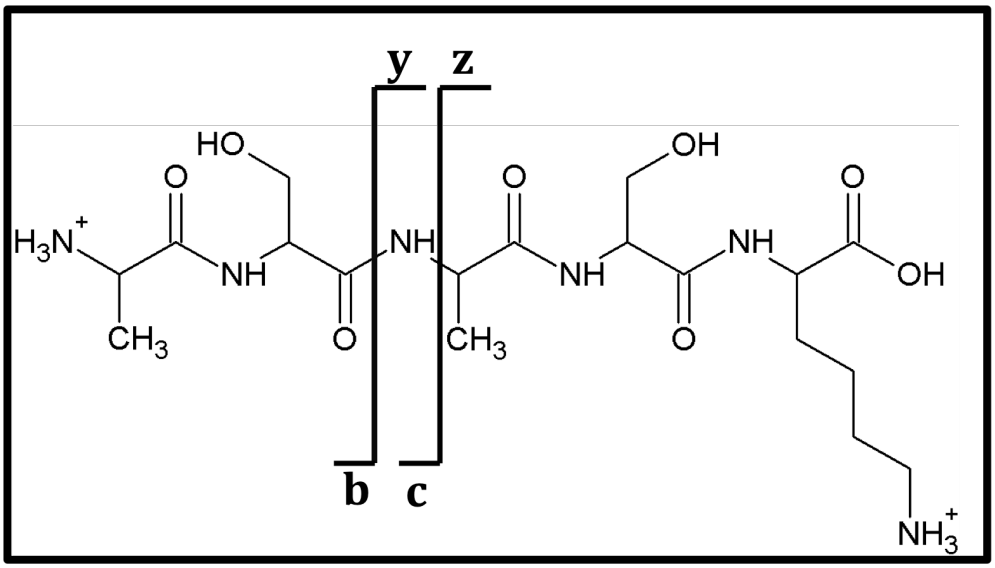


Figure 1.4. Diagram showing bond cleavage sites along a peptide backbone. 'b' and 'y' ions result from CAD-based fragmentation. 'c' and 'z' ions result from ETD-based fragmentation.

A principle difference between these two techniques is the fact that CAD performs well on peptides having a low charge state while ETD performs best on peptides with higher charge states. However, if the goal of an experiment is to identify and localize labile PTMs such as a phosphorylation, ETD is the preferred fragmentation method provided sufficient charge is present on the peptides of interest. If sufficient charge is not available on the phosphorylated peptide(s) of interest, researchers are often left in a quandary. CAD-based fragmentation may yield some useful information, but it will most likely not be able to provide enough fragments for researchers to confidently identify sequence and localize phosphorylation sites. In this situation, researchers would like to have a means of imparting additional charge onto charge-deficient peptides. The next section will focus on progress made in enhancing peptide charge state.

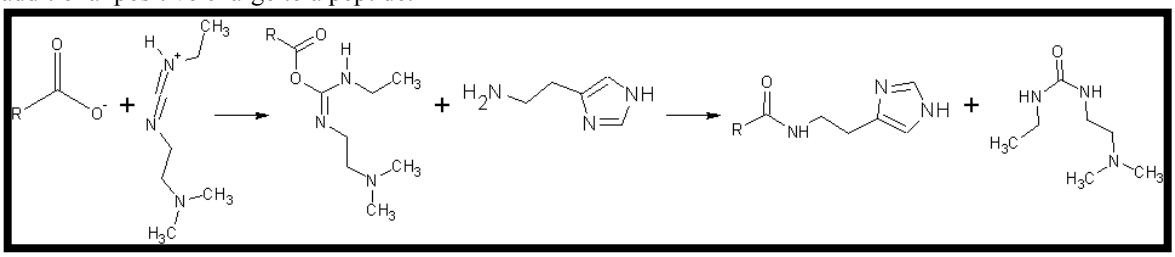
1.5 Developments in Peptide Charge State Enhancement

Over the past several years, several developments have been made to help impart additional positive charge to peptides. One strategy is the modification of RP-HPLC solvents by addition of m-nitrobenzyl alcohol, which facilitates increased protonation as peptides move through a chromatography column.[18] While effective at enhancing peptide charge, this additive alters peptide separations by decreasing peptide retention time and increasing peak width.[18] In analysis of complex samples, neither of these effects is desirable. Another strategy is to covalently modify peptides themselves by incorporating small molecules capable of retaining additional charge. Peptide and protein chemists have long used free sulfhydryl groups on cysteine residues as a site for modification, and recent work has shown that incorporation of charged prosthetic groups to these residues can be effective in enhancing ETD fragmentation.[19] However such methods do not have widespread utility because cysteine residues only account for 2% of the amino acids in a natural protein; most peptides resulting from a digest are thus unlikely to contain a cysteine residue. Given this low abundance of sulfhydryl groups, free carboxylic residues present a much more attractive target for modification. Every peptide will have at least one free carboxylic residue on the C-terminus, and the combined natural abundance of aspartic acid (D) and glutamic acid (E), 10%, means there is likely to be at least two modifiable sites on an average ten-amino acid digest fragment. Many groups have made use of these residues as a site for adding additional charge on peptides and improving ETD fragmentation.[20–23] Adding charge to the C-terminus is of particular value because

every z-ion resulting from ETD fragmentation will have at least one charge, which greatly increases the chance for each of these ions to be detected.

Additional charge can be incorporated by covalently attaching a charged molecule such as a quaternary amine (fixed charge), or a secondary amine (adds charge upon protonation). These reactions are typically carried out using carbodiimide cross-coupling chemistry (**see Scheme 1**). This dissertation will focus on the use of secondary amines groups to enhance peptide charge. We believe this functional group facilitates better ETD fragmentation than a fixed charge because the accepted proton on these groups is free to move between basic sites on a peptide and actively contribute to the fragmentation process.

Scheme 1. Derivatization of a free carboxylic acid using carbodiimide cross-coupling chemistry. This diagram shows a carboxylic acid group first reacting with a protonated carbodiimide (EDC) to form a o-acyl urea intermediate. This intermediate then reacts with a primary amime (histamine is shown as an example) to form an amide linkage at the former carboxylic acid site and an iso-urea derivative. The secondary amine-containing imidazole ring, now attached, can accept an additional proton and contribute additional positive charge to a peptide.



Previous work in the Hunt lab has determined that histamine can be covalently attached to carboxylic acid residues on peptides to increase charge state and improve ETD fragmentation.[23] As shown in **Figure 1.5**, the predominant charge state of vasoactive intestinal peptide 1-12 (vaso) increases from +3 to a +5 when all carboxylic acid groups are derivatized with histamine.

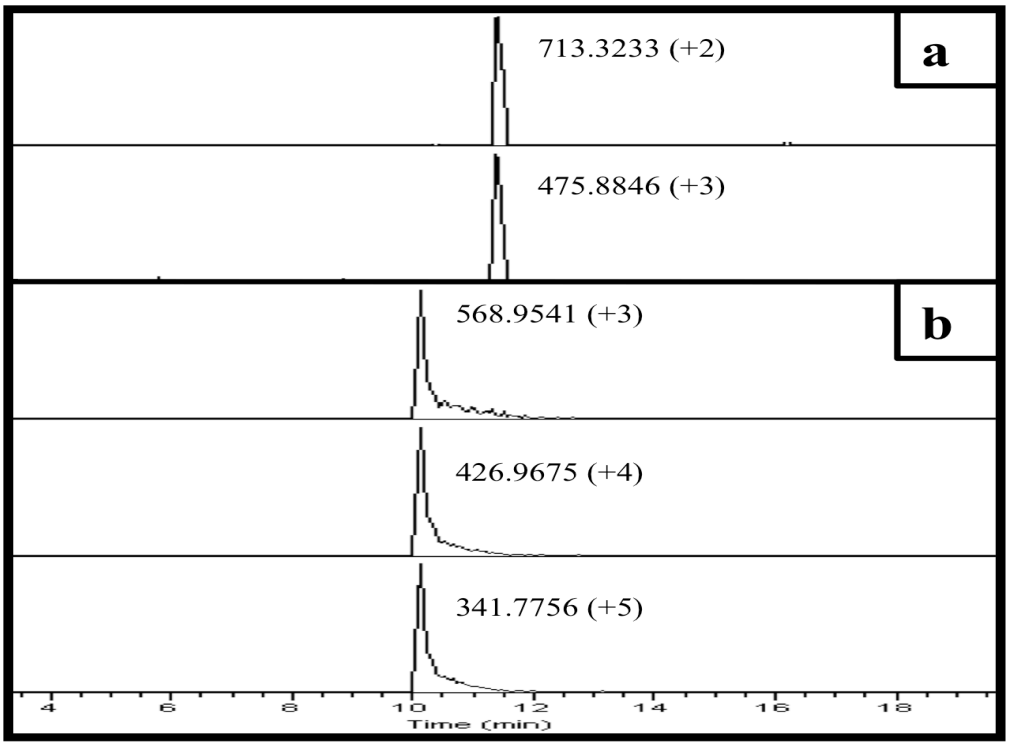


Figure 1.5. Extracted ion chromatogram comparing retention time and charge state distributions of fully histamine-derivatized and underivatized vaso (HSDAVFTDNYTR). Fully derivatized means all carboxylic acid side chains (c-terminus included) are derivatized with histamine. The number listed next to each peak is the monoisotopic m/z of the ion making up each corresponding peak, and the overall charge is listed in parentheses. a: Charge state distribution of underivatized vaso. b: Charge state distribution of fully histamine-derivatized vaso. Gradient: 0-100% B in 17 min., hold at 100% B for 2min, drop back to 0% B in 3min.

This increase in charge state arises from protonation of the imidazole ring systems on attached histamine molecules. It is also worth noting that upon derivatization, the peptide becomes substantially more hydrophilic compared to the native form. Aside from causing an undesirable shift in peptide retention time, the methodology developed by the Hunt lab has been very useful allowing for the sequencing of peptides and localization of PTMs using ETD that were previously not amenable to this fragmentation technique.

1.6 Aims of Research

The goal of this project is largely to improve and refine currently used charge enhancement methods in our lab. Specifically, we would like to develop a derivatization method that increases peptide charge state while leaving peptide hydrophobicity unchanged or slightly increased. Past attempts to accomplish this task have identified molecules similar to histamine, which also contain hydrophobic functionalities to help counteract hydrophilic qualities afforded by a protonated amine. Unfortunately, past use of these more hydrophobic amines has resulted in an extremely poor quality chromatogram where non-peptidic contaminants are found to completely overshadow most peptide species. A key goal in developing a new derivatization method is to eliminate or reduce the contaminant problem either by developing a post-reaction cleanup method, or by devising a synthesis strategy that makes use of minimal amine reagent from the start. The work presented here will primarily focus on the latter in hopes of allowing ETD to be used in the sequencing of post-translationally modified peptides with low or uneven charge distributions.

1.7 References

1. De Hoog, C. L. & Mann, M. Proteomics. *Annual Review of Genomics and Human Genetics* **5,** 267–293 (2004).
2. Consortium, I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
3. Walsh, C. T., Garneau-Tsodikova, S. & Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angewandte Chemie International Edition* **44,** 7342–7372 (2005).
4. Hubbard, M. J. & Cohen, P. On target with a new mechanism for the regulation of protein phosphorylation. *Trends in Biochemical Sciences* **18,** 172–177 (1993).

5. Google Scholar Linked Page. at <http://www.aseanbiotechnology.info/Abstract/21009904.pdf>
6. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).
7. Graves, P. R. & Haystead, T. A. J. Molecular biologist's guide to proteomics. *Microbiol. Mol. Biol. Rev.* **66,** 39–63; table of contents (2002).
8. James, P. Protein identification in the post-genome era: the rapid rise of proteomics. *Q. Rev. Biophys.* **30,** 279–331 (1997).
9. Hunt, D. F., Zhu, N.-Z., Shabanowitz, J. & Cooks, R. G. Oligopeptide sequence analysis by collision-activated dissociation of multiply charged ions. *Rapid Communications in Mass Spectrometry* **3,** 122–124 (1989).
10. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).
11. Whitehouse, C. M., Dreyer, R. N., Yamashita, M. & Fenn, J. B. Electrospray interface for liquid chromatographs and mass spectrometers. *Anal. Chem.* **57,** 675–679 (1985).
12. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **246,** 64–71 (1989).
13. Kebarle, P. A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *Journal of Mass Spectrometry* **35,** 804–817 (2000).
14. Hoffmann, E. de & Stroobant, V. *Mass spectrometry: principles and applications*. (John Wiley & Sons, 2007).
15. Mikesh, L. M. *et al.* The utility of ETD mass spectrometry in proteomic analysis. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* **1764,** 1811–1822 (2006).
16. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 9528 –9533 (2004).
17. Chen, X. & Tureček, F. The Arginine Anomaly: Arginine Radicals Are Poor Hydrogen Atom Donors in Electron Transfer Induced Dissociations. *J. Am. Chem. Soc.* **128,** 12520–12530 (2006).
18. Kjeldsen, F., Giessing, A. M. B., Ingrell, C. R. & Jensen, O. N. Peptide Sequencing and Characterization of Post-Translational Modifications by Enhanced Ion-Charging and Liquid Chromatography Electron-Transfer Dissociation Tandem Mass Spectrometry. *Analytical Chemistry* **79,** 9243–9252 (2007).
19. Vasicek, L. & Brodbelt, J. S. Enhanced Electron Transfer Dissociation through Fixed Charge Derivatization of Cysteines. *Analytical Chemistry* **81,** 7876–7884 (2009).
20. Zhang, L., Xu, Y., Lu, H. & Yang, P. Carboxy group derivatization for enhanced electron‐transfer dissociation mass spectrometric analysis of phosphopeptides. *PROTEOMICS* **9,** 4093–4097 (2009).
21. Xu, Y., Zhang, L., Lu, H. & Yang, P. Mass Spectrometry Analysis of Phosphopeptides after Peptide Carboxy Group Derivatization. *Anal. Chem.* **80,** 8324–8328 (2008).

22. Ko, B. J. & Brodbelt, J. S. Enhanced Electron Transfer Dissociation of Peptides Modified at C-terminus with Fixed Charges. *J. Am. Soc. Mass Spectrom.* **23,** 1991–2000 (2012).
23. English, A. Development and Application of Methodologies to Facilitate Peptide Sequencing with Mass Spectrometry. (2011).

Chapter 2 – Method Development

2.1 Background

As mentioned in the previous chapter, efforts to increase peptide charge state while increasing hydrophobicity have proved challenging. Two promising amines with more hydrophobic character than histamine are 2-(2-Aminoethyl) benzimidazole (benzimidazole) and 2-(7-Methyl-1H-benzimidazole-2-yl) ethanamine (methyl benzimidazole). Chemical structures of these three amines are shown in **figure 2.1**.

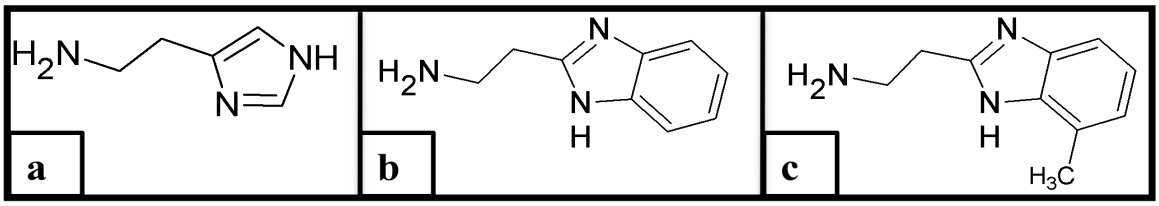


Figure 2.1. Comparison of three amines used for charge enhancement chemistry. a: chemical structure of histamine. b: chemical structure of benzimidazole. c: chemical structure of methyl benzimidazole.

Unlike histamine, both benzimidazole and methyl benzimidazole contain a fused aromatic ring, which affords significant hydrophobic character. In the case of methyl benzimidazole, the methyl group imparts additional hydrophobic functionality to the aromatic ring system. For the purposes of our method development process, methyl benzimidazole did not become known/available to us until the later stages of our study. Our original efforts were directed towards development of benzimidazole as a charge enhancement reagent.

Initial experiments utilizing our established derivatization protocol (see methods 2.3.1) demonstrated that successful derivatization with benzimidazole results in an increased peptide charge state (**see figure 2.2a**). However as shown by the chromatogram

in **figure 2.2b**, significant amounts of contaminating species were found despite on-column washing of the derivatized peptide prior to analysis. These contaminating species were present in such high amounts that underivatized peptide standards added prior to analysis were not detected; the only visible peptide in the base peak was the derivatized form of vaso. It was thus impossible to determine how benzimidazole alters peptide retention time. Furthermore as a practical matter, a reaction that leaves behind such an excessive amount of contaminants would not be amenable to analyzing a complex sample such as a whole protein digest. We know from experience that when histamine is used in an identical reaction, very few contaminant peaks result and the predominant species in the base peak are derivatized peptides. The origin of these contaminants must therefore be from the amine reagent itself.

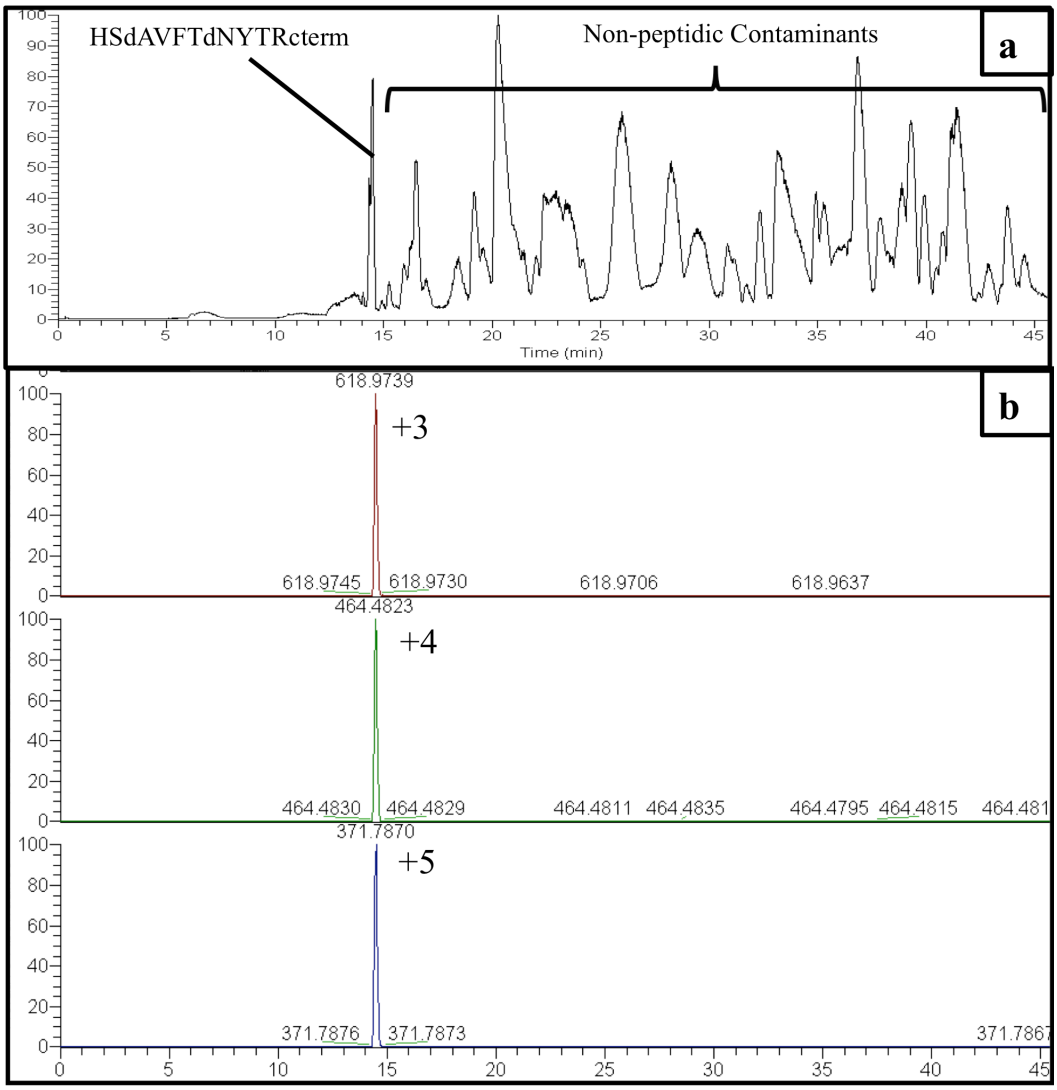


Figure 2.2. Resulting chromatographic data from the analysis of a 1 pmol aliquot from a benzimidazole-derivatization reaction of vaso (HSDAVFTDNYTR). Lower case letters and the suffix "cterm" indicate benzimidazole has been attached to these positions on the peptide. a: Base peak chromatogram showing all detected species from the reaction. Aside from the labeled peak, all other peaks are non-peptidic contaminants. b: Selected ion chromatogram showing an increased charge state distribution of derivatized HSDAVFTDNYTR. For reference, the native form has a charge distribution of (+1 to +3). Gradient: 0-100% B in 17 min., hold at 100% B for 2min, drop back to 0% B in 3min. Standards added: 100 fmol (each) of vaso (underivatized) and angio.

Given this reality, we needed to develop a new derivatization methodology in order to evaluate benzimidazole and other hydrophobic reagents. There are two possible routes to accomplish this goal. Option one: we could keep the same synthesis strategy and devise a means of removing contaminants from the reaction once completed. Option

two: we could employ an alternative synthesis strategy that requires significantly lower amounts of amine from the start.

The first option we tested is an adaptation of a commonly used biochemical purification technique known as acetone precipitation. This technique relies on the fact that proteins/peptides are virtually insoluble in acetone, while many organic molecules readily dissolve in this solvent. When a mixture consists of peptides and organic contaminants, the two can be separated by vigorously washing with acetone followed by high-speed centrifugation. Removal of the supernatant will leave behind a pellet containing a higher proportion of peptides compared to the original mixture. As this process is repeated many times, peptides will gradually become more pure. We attempted to employ this method in hopes of being able to remove contaminants resulting from a benzimidazole derivatization reaction.

The second option we tested is an attempt to reproduce a peptide derivatization procedure developed by Xu *et al*, which utilizes 1-Hydroxy-7-azabenzotriazole (HOAt) to promote high yields (see **scheme 2.1**).[1] As shown in **scheme 2.1**, HOAt is believed to help facilitate the reaction by helping to bring an amine nucleophile in close proximity to the activated carboxylic through a hydrogen bonding interaction.[2] An attractive quality of this protocol is the fact that investigators claim high yields were obtained (~90% range) in a couple seconds, while using roughly two orders of magnitude less amine reagent (see **table 2.1**), compared to the protocol developed in our lab. Given the problems faced when contaminants are present in hydrophobic amines such as benzimidazole, being able to accomplish the same reaction with more than a hundred times less amine is highly

desirable. In theory, such a reduction would leave peptide signals above any reaction contaminants during HPLC/MS analysis of the reaction, and allow complex samples to be derivatized with more hydrophobic reagents.

Scheme 2.1. Derivatization of a carboxylic acid using EDC/HOAt-mediated chemistry. a: Initially activated with EDC (shown in **scheme 1.1**), a carboxylic acid residue is made to react with HOAt and form an HOAt intermediate. b: The HOAt intermediate acts as a hydrogen bond donor to attract an amine nucleophile.[2] Once interfaced with the pyridine ring system, the amine can proceed to attack the carbonyl carbon and form an amide bond and displace the HOAt moiety.

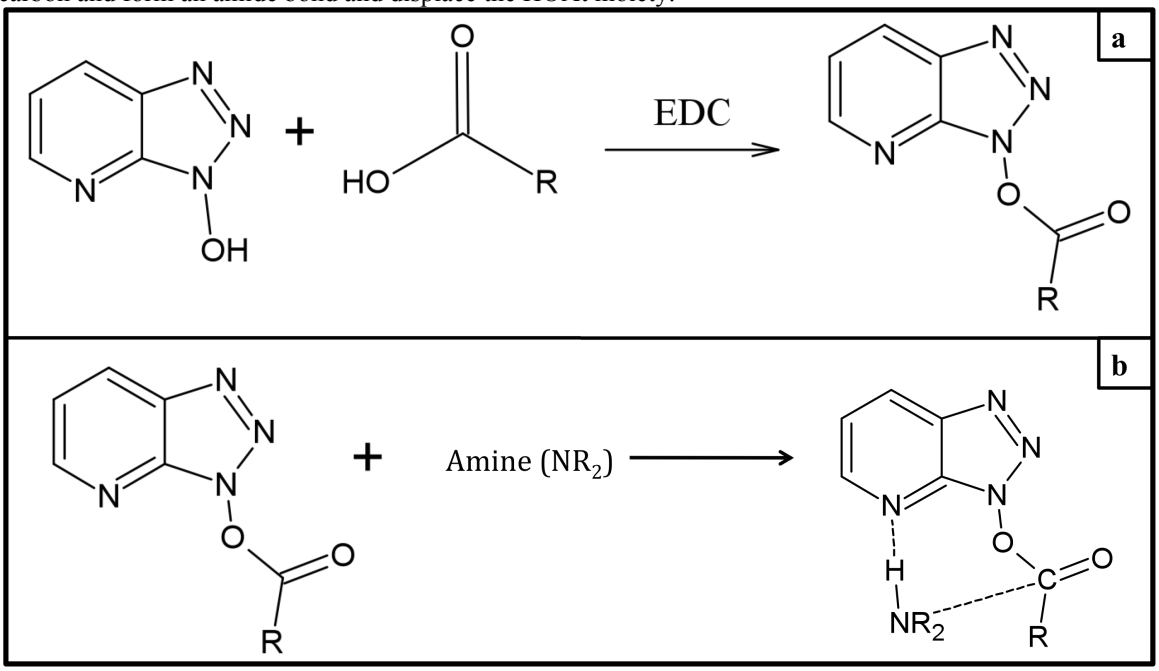


Table 2.1. Comparison of final reagent concentrations used in cross-coupling reactions for two different protocols (Xu *et al* and Hunt Lab). EDC = N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide. HOAt = 1-Hydroxy-7-azabenzotriazole.

Reagent	Xu *et al*[1] (mM)	Hunt Lab[3] (mM)
Amine	3	800
EDC	1	20
HOAt	1	n/a

In their paper, Xu *et al* were not clear about the total amount of peptides or carboxylic acid residues present in their derivatization reactions. When setting up this reaction for ourselves, we decided to carry out the reaction on a mixture of five

picomoles each of five different peptides (HSDAVFTDNYTR, YGGFL, DRVpYIHPF, IKNLQpSLDPSH, DFNKFHpTFPQTAIGV). Taking into account the number of modifiable carboxylic acid residues and the final volume of the reaction, the final concentration of carboxylic acids in the reaction is roughly 0.1 μM. Unless otherwise specified, all trials of this reaction will be carried out on this mixture. In comparison to the reagent concentrations listed in **table 2.1**, all reagents are present in roughly thousand-fold excess compared to the amount of modifiable carboxylic acid residues. From a stoichiometric point of view, this makes the peptides themselves to be the limiting reagent.

The work presented here is an attempt to test and evaluate these two options, and determine if either method is able to help produce a relatively "clean," derivatized, and more hydrophobic sample with enhanced charge. Once we evaluated our options and optimized a methodology, we demonstrated the its utility for improving ETD fragmentation on a complex sample.

2.2 Materials & Reagents

2.2.1 Reagents

Advanced ChemTech, Loiusville, KY

1-Hydroxy-7-azabenzotriazole (HOAt)

Honeywell Inc, Muskegon, MI

Acetonitrile, LC-MS grade (ACN)

Promega, Madison, WI

Endoproteinase trypsin, sequencing grade

Protea, Frederick, MD

Phosphopeptide Standard I

Contains peptides: DRVpYIHPF, IKNLQpSLDPSH, DFNKFHpTFPQTAIGV

Sigma-Aldrich, St. Louis, MO

1-(2-pyrimidyl) piperazine (PP)

1-phenylpiperazine hydrochloride (Phenylpiperazine)

2-(2-Aminoethyl) benzimidazole dihydrochloride (Benzimidazole)

2-(7-Methyl-1H-benzimidazol-2-yl) ethanamine dihydrochloride (Methyl Benzimidazole)

Acetic Acid, glacial, 99.99% (HOAc)

Albumin from bovine serum, lyophilized powder, ≥96% (BSA)

Aminopropyl imidazole

Ammonium bicarbonate (Ambic)

Angiotensin I human acetate hydrate, ≥90% purity (Angio)

Azulene

Dithiothreitol (DTT)

Formamide

Histamine dihydrochloride

Iodoacetamide (IAM)

Leucine Enkphalin trifloroacetate hydrate (YGGFL)

N,N-Dimethylformamide (DMF)

N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC)

Pyridine

Sodium Hydroxide (NaOH)

N-Hydroxysulfosuccinimide sodium salt (Sulfo-NHS)

Vasoactive intestinal peptide fragment 1-12, human ≥97% (Vaso)

Thermo Fisher Scientific, Waltham, MA

Acetone, anhydrous

Hydrochloric acid (HCl)

Methanol, anhydrous (MeOH)

Water, LC-MS grade

2.2.2 Column Making Materials

New Objective Inc., Woburn, MA

PicoClear™ Union for 360 μm OD fused silica (PCU-360)

PolyMicro Technologies, Inc., Phoenix, AZ

Polyimide coated fused silica fused capillary

360 μm O.D. x 50 μm I.D.

360 μm O.D. x 75 μm I.D.

360 μm O.D. x 150 μm I.D.

PQ Corporation, Valley Forge, PA

Kasil®- Potassium silicate solution

Lithisil® 829- Lithium silicate solution

Sutter Instrument Co., Novato, CA

P-2000 Laser puller with fused silica adapter

YMC Corp., Morris Plains, NJ (available through Waters Corp.)

ODS-AQ, C18, 5 μm spherical silica particles, 120 Å pore size

ODS-AQ, C18, 15 μm spherical silica particles, 120 Å pore size

ODS-AQ, C18, 5-20 μm spherical silica particles, 120 Å pore size

2.2.3 Laboratory Equipment

Agilent Technologies, Palo Alto, CA

1100 Agilent HPLC

Branson, Danbury, CT

Branson 1200 Ultrasonic Bath

EMD Milipore, Darmstadt, Germany

colorpHast® pH Test Strips (non-bleeding)

Eppendorf Corp., Hamburg, Germany

Microcentrifuge 5415R

Labconco, Kansas City, MO

CentriVap Vacuum Concentratior System

Microelectrodes Inc., Bedford, NH

MI-410 Micro-Combination pH Probe

Thermo Fisher Scientific, Waltham, MA

LTQ FT mass spectrometer (custom modified with front-end ETD)

LTQ-Qrbitrap mass spectrometer (custom modified with front-end ETD)

2.3 Methods

2.3.1 Peptide Derivitization Attempt with Benzimidazole

A 10 pmol aliquot of a test peptide was taken to dryness in a vacuum centrifuge. The following solutions were then added in the order listed. 20 μL 1 M benzimidazole (in 1 M pyridine buffer pH=5.5), 5 μL 0.1 M EDC (in 1 M pyridine buffer pH=5.5). The mixture was placed in a sonnicator bath for 10 minutes, and allowed to incubate at room temperature for 1.5 hours before being taken to dryness in a vacuum centrifuge. This protocol was adapted from Dr. Michelle English's dissertation.[3]

2.3.2 Acetone Wash Procedure for Benzimidazole Reaction

Once the previous reaction (section 2.3.1) was taken to dryness, 1 mL of ice-cold acetone was added and mixed vigorously to resuspend the pellet. The reaction tube was then placed in a 4°C tabletop centrifuge and spun at maximum speed (~20,000 x g) for 15 minutes. 90% of the supernatant was then discarded, and the process was repeated a total of six times. After the final wash, the remaining pellet was taken to dryness in a vacuum centrifuge.

2.3.3 Attempt to Adapt Peptide Derivatization Method by Xu et al[1]

To a 50 μL solution of water (containing 5 pmol of each of the following peptides: HSDAVFTDNYTR, YGGFL, DRVpYIHPF, IKNLQpSLDPSH, DFNKFHpTFPQTAIGV), the following solutions were added in the order listed. 6 μL

0.5% 1-(2-pyrimidyl) piperazine (in DMF), 4 μL 2 mg/mL EDC (in DMF), 3 μL 2 mg/mL HOAt (in DMF), ~3 μL 0.1% TFA (in H_2O). The solution was vortexed for several seconds, and immediately pulsed on a centrifuge to collect any droplets on the side of the reaction vessel (an Eppendorf tube). Depending on the experiment, the reaction was either immediately placed in the vacuum centrifuge or allowed to incubate for a defined amount of time. A duplicate reaction was also created to allow for measurement of pH using the test strips described in materials section. If necessary, the amount of acid was adjusted to achieve a pH between 7.5 and 7.8. Depending in the experiment, pH was measured using either colorpHast® test strips of the appropriate range, or a microelectrode.

2.3.4 Modified EDC/HOAt Peptide Derivatization Method

To a 45-55 μL solution of water (containing 5 pmol of each of the following peptides: HSDAVFTDNYTR, YGGFL, DRVpYIHPF, IKNLQpSLDPSH, DFNKFHpTFPQTAIGV), the following solutions were added in the order listed. 6 μL 3.5×10^{-2} M amine (depending on the experiment), 6 μL DMF, 4 μL 2 mg/mL EDC (in DMF), 3 μL 2 mg/mL HOAt (in DMF), 2-6 μL 0.125 N NaOH or HCl (depending on desired final pH). In all cases, the final volume of the reaction was 66 μL. The solution was vortexed for several seconds, and immediately pulsed on a centrifuge to collect any droplets on the side of the reaction vessel (an Eppendorf tube). The reaction pH was then measured with microelectrode. The reaction was then allowed to incubate for a set amount of time, and placed in a vacuum centrifuge. In some cases, 10 μL of a 5% TFA

solution was added after incubation to immediately quench the reaction. Note: it was often necessary to setup several "test" reactions to determine the precise amount of HCl or NaOH that was required to obtain the desired pH.

2.3.5 Tryptic Digest of Bovine Serum Albumin (BSA)

An aliquot of BSA was dissolved in 100 mM Ammonium bicarbonate (pH = 8), and reduced for one hour at room temperature by adding a 20-fold molar excess (compared to the total moles of cysteine present) of dithiothreitol (DTT). Following the reduction, cysteine residues were acetylated for one hour in darkness at room temperature by adding a 3-fold molar excess (compared to the total moles of DTT added previously) of iodoacetamide. The resulting protein was then mixed with trypsin at an enzyme to protein ratio (by mass) of 1:20. The digestion reaction was incubated at 37°C for 7 hours, followed by termination with glacial acetic acid. Once terminated, the digestion reaction was taken to dryness on a vacuum centrifuge.

2.3.6 Assembly of HPLC Columns

Precolumns were constructed out of 360 μm O.D. x 75 μm I.D. fused silica. A kasil® frit was created in one end of the capillary by quickly (<2 seconds) dipping it into a 3:1 kasil®/formamide solution. The capillary was then placed in a 70°C oven and allowed to incubate overnight. The following day, excess kasil® contained in the capillary was trimmed to leave behind a ~1 mm frit. The precolumn was then packed with 7 cm of 15 μm C18 resin using a pressure bomb at 50 psi, and equilibrated to

relevant HPLC solvents. Sample cleanup columns were prepared similarly, except a larger capillary (150 μm I.D.) was utilized in conjunction with 5-20 μm C18 resin (packed to 6 cm).

Analytical columns were constructed out of 360 μm O.D. x 50 μm I.D. fused silica. A lithisil® frit was constructed by burning a 4 cm window at one end of the capillary and dipping the burned side it into a premixed fritting solution (50 μL tetramethylammonium silicate, 100 μL Lithisil, and 10 μL formamide). The capillary was removed from the solution once the liquid was within ~0.5 cm of the top burn mark. A 375°F soldering tip was placed in the middle of the window until a ~1 mm frit was created. Unpolymerized fritting solution was washed away using a 0.1% acetic acid solution loaded into a pressure bomb at 1000 psi. The frit was then briefly (<2 seconds) placed over a butane flame to ensure complete polymerization of the frit. The analytical column was then packed with 10 cm of 5 μm C18 resin using a pressure bomb at 500 psi, and equilibrated to relevant HPLC solvents. An emitter tip was then constructed after the column frit.[4]

2.3.7 Loading and Rinsing of Samples

All peptide samples were reconstituted in 0.1% acetic acid solution and loaded onto a precolumn using a pressure bomb. The precolumn was then attached to an HPLC pump and washed with 100% solvent A (0.1 M acetic acid) for 20 minutes at 20 bar. The precolumn was then connected to an analytical column using a PicoClear™ union. Relevant standards were then loaded to the combined column using a pressure bomb.

2.3.8 Cleanup of Digested Protein Samples

1 pmol of digested protein was loaded onto a cleanup column and washed with 100% solvent A (0.1 M acetic acid) for 20 minutes at 20 bar on an HPLC. A gradient was then started in which solvent B (70% acetonitrile/30% 0.1 M acetic acid) was ramped from 0% to 80% in 40 minutes with a hold for 30 minutes. The flow-through for this entire 70 minute period was collected and taken to dryness in a vacuum centrifuge.

2.3.9 Charge Enhancement Chemistry Reaction with BSA Digest Peptides

To 45 μL water (containing 500 fmol from cleaned up BSA digest, and 200 fmols of each of the following peptides: HSDAVFTDNYTR, YGGFL, DRVpYIHPF, IKNLQpSLDPSH, DFNKFHpTFPQTAIGV), the following reagents were added. 6 μL 3.5×10^{-2} M histamine, 6 μL DMF, 4 μL EDC (2 mg/mL in DMF), 3 μL HOAt (2 mg/mL in DMF), 2 μL 0.125 M NaOH. The mixture was vortexed for several seconds and pulsed on a centrifuge. The reaction pH was then measured with a microelectrode to confirm the pH was close to 5.5 ±0.1. After 40 minutes of incubation at room temperature, the reaction was taken to dryness in a vacuum centrifuge. Note: we estimated there were approximately 90 pmols of modifiable carboxylic acids present in this reaction.

2.3.10 Analysis of Samples by Mass Spectrometry

Peptides were analyzed by nanoflow-HPLC/microelectrospray ionization coupled directly to a Thermo Orbitrap or FT-ICR mass spectrometer with a home built frontend ETD (FETD) source.[5] Briefly, a pre-column holding bound sample was connected via a PicoClear™ union to a reversed-phase C18 HPLC analytical column rinsed with 0.1 M acetic acid and gradient eluted through a laser-pulled electrospray tip directly into the mass spectrometer with an Agilent 1100 series binary LC pump at a flow rate of ~60 nL/min. The elution gradients utilized solvent A: 0.1M acetic acid in H2O and solvent B: 70% acetonitrile and 0.1 M acetic acid in H2O. The ETD reagent was azulene and the ion-ion reaction times were 30 or 50 ms depending on the reagent used. The instrument method used for analysis of standard peptides was data dependant fragmentation of the 5 most abundant peaks.

2.4 Results and Discussion

2.4.1 Benzimidazole Reaction Cleanup with Acetone Wash

A benzimidazole cross-coupling reaction was setup as described in *methods 2.3.1* with vaso as the peptide substrate. After drying down the reaction, the resulting mixture was washed with cold acetone for six wash cycles (*methods 2.3.2*), and taken to dryness following the last cycle. An aliquot from this combined process was then taken for analysis. Chromatographic data from this experiment is shown in **Figure 2.3**.

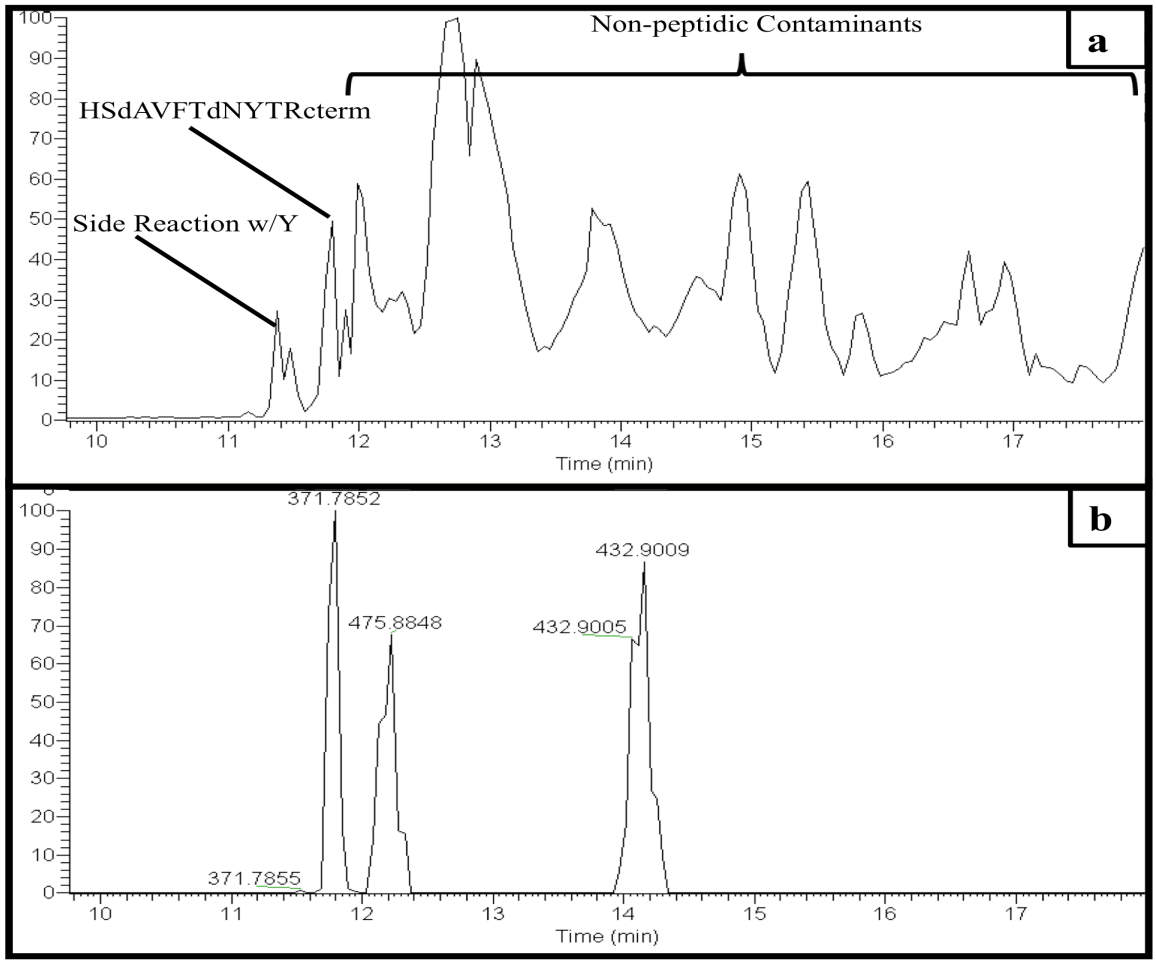


Figure 2.3. Resulting chromatographic data from the analysis of a 1 pmol aliquot from a combined benzimidazole derivatization/acetone wash procedure of vaso. a: Base peak chromatogram showing all major species present from after washing the reaction products. Aside from the labeled peaks, all other peaks in the chromatogram are non-peptidic contaminants. In one peak, a side reaction was observed whereby an EDC molecule reacted with a tyrosine (Y) residue on an otherwise fully derivatized peptide. Lower case D and the suffix "cterm" indicate benzimidazole has been attached to these positions on the peptide. b: Selected ion chromatogram showing predominant monoisotopic m/z of three peptides. 371: fully derivatized vaso (+5 charge). 475: underivatized vaso. 432: underivatized angio. Note: the peak heights were individually scaled in (b) to allow for retention time comparison on the same axis. The height of each peak is not reflective of actual abundance. Gradient: 0-100% B in 17 min., hold at 100% B for 2min, drop back to 0% B in 3min. Standards added: 100 fmol (each) of vaso (underivatized) and angio.

As shown in **figure 2.3(a)**, the resulting chromatogram still contains numerous non-peptidic impurities. However as shown in **figure 2.3(b)**, both added standards were found after manual searching and are shown in the selected ion chromatogram along with the fully derivatized form of vaso. By comparing the fully derivatized form of vaso to the

unreacted peptide, there is a slight decrease in retention time by approximately 0.4 minutes. This hydrophilic shift is somewhat surprising given the presence of a fused aromatic ring system in each derivatized site on the peptide. However, this shift is considerably less dramatic compared to the same peptide derivatized with histamine. Despite the improvement of being able to detect standards during reaction analysis, the overwhelming majority of the chromatogram is filled with numerous contaminant peaks, and renders the method unsuitable for analysis of complex samples.

While this hydrophilic shift is less pronounced than would be upon derivatization with histamine, benzimidazole does not meet out criteria of causing an increase on hydrophobicity. For this reason, it was not utilized for future study as a charge enhancement reagent. However the massive amounts of non-peptidic contaminants resulting from the reaction highlighted the need to purse an alternative derivatization strategy before other hydrophobic amines could be tested.

2.4.2 Attempt to Adapt EDC/HOAt-Reaction as Described by Xu et al[1]

A cross-coupling reaction was setup using 1-(2-pyrimidyl) piperazine (PP) as the charge enhancement reagent (see *methods 2.2.3*) with equimolar amounts of five test peptides (HSDAVFTDNYTR, YGGFL, DRVpYIHPF, IKNLQpSLDPSH, DFNKFHpTFPQTAIGV). For reference, a structure of PP is shown in the inset of **Figure 2.4**. After vortexing the mixture for several seconds, the mixture was immediately placed in a vacuum centrifuge and taken to dryness. The resulting chromatogram is shown in **Figure 2.4**.

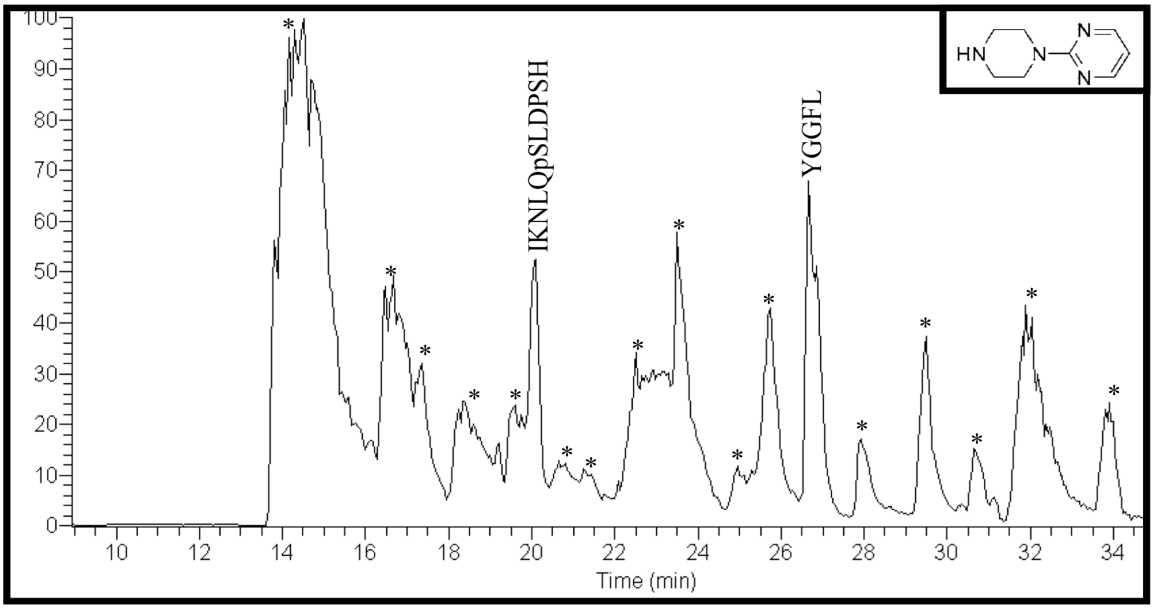


Figure 2.4. Labeled base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from an EDC/HOAt/PP derivatization reaction at pH 7.5 (measured using test strips). The inset shows a chemical structure of PP, which was used as the charge-enhancing reagent. The asterisk denotes the predominant species under a particular peak is a non-peptidic contaminant. Gradient: 0-100% B in 17 min., hold at 100% B for 2min, drop back to 0% B in 3min. Standard added: 250 fmol of underivatized and angio.

After manual analysis of these data, the major peptidic species found in the chromatogram were underivatized starting material. The mass area of each of these peptides was at expected levels, which suggests very few peptide carboxylic acid derivatization reactions took place. Trace amounts of PP-derivatization were detected on the YGGFL peptide, but it represented less than 1% of the total mass area for that peptide. As shown by the chromatogram, there is also a significant amount of non-peptidic material in the sample. The almost non-existent yield, and high degree of contaminants are in direct contradiction to what Xu *et al*[1] claimed could be accomplished using this methodology. However the fact that reagents were present in a thousand-fold molar excess suggests there was a problem with the reaction conditions.

To ensure incubation time was not the problem, the reaction was setup again and allowed to incubate at room temperature for two hours. Even though Xu *et al* claimed a high yield was obtained after only a few seconds, it was not specified how long it took their reaction to dry down in a vacuum centrifuge. Since nothing was added to terminate the reaction, it is perfectly conceivable that the bulk of their reaction took place during the dry down process. By letting our reaction incubate to a longer period of time, we believed there was a chance of obtaining a higher yield. Unfortunately such was not the case. After letting the reaction stand for a longer period of time, a nearly identical result was obtained to that shown in **Figure 2.4**.

The reaction was also attempted using histamine instead of PP. Histamine had recently been used by members of our lab for other purposes, and was known to be a viable reagent. In addition, a melting point analysis of EDC and HOAt was also conducted to confirm each was still viable (both melting points matched expected values). However even with a different amine and reagents confirmed to be viable, no evidence was found to suggest a reaction with peptides had occurred.

We speculated there might be a problem with the reaction pH that prevented a reaction from starting. The total volume of this reaction was approximately 66 microliters (μL), and the pH was measured using a strip designed to measure solutions measured in the milliliter range. In order to fully wet all the tiles on the strip, a duplicate reaction tube had to be prepared for the sole purpose of measuring pH. Even with this duplicate, the amount of liquid is barely enough to confidently measure pH. Furthermore, the reaction mixture contains roughly 20% DMF by volume. It is not known whether or not the pH

strips available to us are able to give accurate pH measurements with this amount of organic solvent. In their paper, Xu *et al* specified a pH range of 7.5 and 7.8, but did not specify how this range was measured.[1] Given the precarious nature of how pH was measured for this test and the fact that no pH instrumentation or equipment is specified in our reference paper, it is highly conceivable the reaction failed to work as expected due to unfavorable pH conditions.

2.4.3 Examination of reaction pH

In order to get a more definitive measure of the reaction pH, we obtained a microelectrode that is capable of measuring pH in the small volumes present in this reaction. A simple experiment was conducted in which the reaction presented in the previous sections (*results section 2.4.2*) was prepared again and measured using the electrode. The result from this experiment is shown in **Table 2.2**.

Table 2.2. Comparison of pH measurements of test strips vs. a microelectrode. The solution was prepared as described in methods 2.2.3.

Measurement Device	pH
Electrode	8.27
pH Strip	7.5

Based on the data presented in **Table 2.2**, there is a clear disparity between the value given by a pH strip compared to a calibrated pH meter. In their paper, Xu *et al*[1] specify a pH range between 7.5 and 7.8 is critical for the reaction to proceed. Based on data from the electrode, this requirement was not being satisfied. The apparent disparity in pH readings could explain why derivatization was not observed in the preceding results.

2.4.4 Adaption EDC/HOAt-Reaction as Described by Xu et al[1] Using Microelectrode

Based on the data from the previous section, it is clear our initial attempts to reproduce this reaction were not achieving the specified pH range. In order to lower the reaction pH, the volume of added TFA solution was increased while maintaining the same total volume of solution as indicated in *methods 2.3.3*. After empirically determining the precise amount of TFA solution to obtain a pH between 7.5 and 7.8, the reaction was allowed to incubate for 40 minutes at room temperature. The resulting chromatogram from an aliquot of this reaction is shown in **Figure 2.5**.

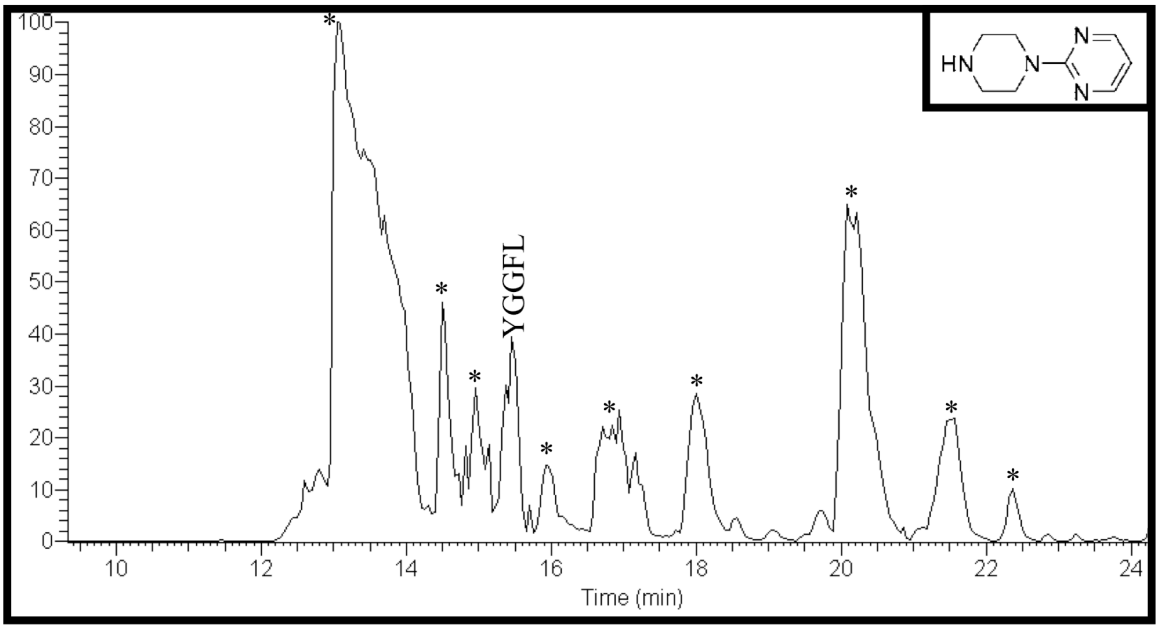


Figure 2.5. Resulting base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from an EDC/HOAt/PP derivatization reaction at pH 7.72 (measured using microelectrode). The inset shows a chemical structure of PP, which was used as the charge-enhancing reagent. In this analysis, only one peptide was visible in the base peak (see label). The asterisk denotes the predominant species under a particular peak is a non-peptidic contaminant. Gradient: 0-100% B in 17 min., hold at 100% B for 2min, drop back to 0% B in 3min. Standard added: 250 fmol of underivatized and angio.

As shown in **Figure 2.5**, the predominant species in the chromatogram are non-peptidic contaminants. Only one peptide was found in the base peak. By searching the data manually, the other four reaction peptides and the standard were found at expected levels. Only trace amounts of PP-derivatized peptides were found. Aside from the difference in pH, this is virtually the same result obtained when the reaction was first attempted. Even though pH was accurately measured, and set to the specified range of 7.5-7.8, the peptides still emerged from the reaction as though nothing had happened.

To again test whether or not the amine has an effect on reaction outcome, the reaction was repeated in the prescribed 7.5-7.8 pH range using histamine as the charge enhancement reagent. Aside from a cleaner chromatogram, we still only found trace amounts of derivatized peptides (< 5% yield). The overwhelming majority of the peptides present were still unreacted.

These repeated failures to reproduce published methodology brought into question whether or not the prescribed pH range is in fact valid. In our own experience, cross-coupling reactions are typically performed in the acidic range. It is thus conceivable that Xu *et al* used test strips to measure pH, and incorrectly determined the pH of their solution as was previously demonstrated in **Table 2.2**. As a result, their published method may not have incorrectly reported the reaction pH.

2.4.5 Investigation of pH and Its Effect on Reaction Yield

We set out to determine if a reasonable yield could be obtained using EDC/HOAt chemistry by optimizing the reaction pH. We choose a range of 3.7 – 7.6, and studied the

effect of reaction yield across this range. The reactions were setup as described in *methods 2.3.4*. Of the five peptides present in the reaction, YGGFL was chosen to monitor reaction yield. Unlike the other peptides, YGGFL contains only one modifiable carboxylic residue and does not undergo side-reactions or incomplete derivatization, as do some of the other peptides. Also, to minimize contamination/degradation of chromatography columns, histamine was used as the charge enhancement reagent instead of PP. The results from this experiment are summarized in **Figure 2.6**.

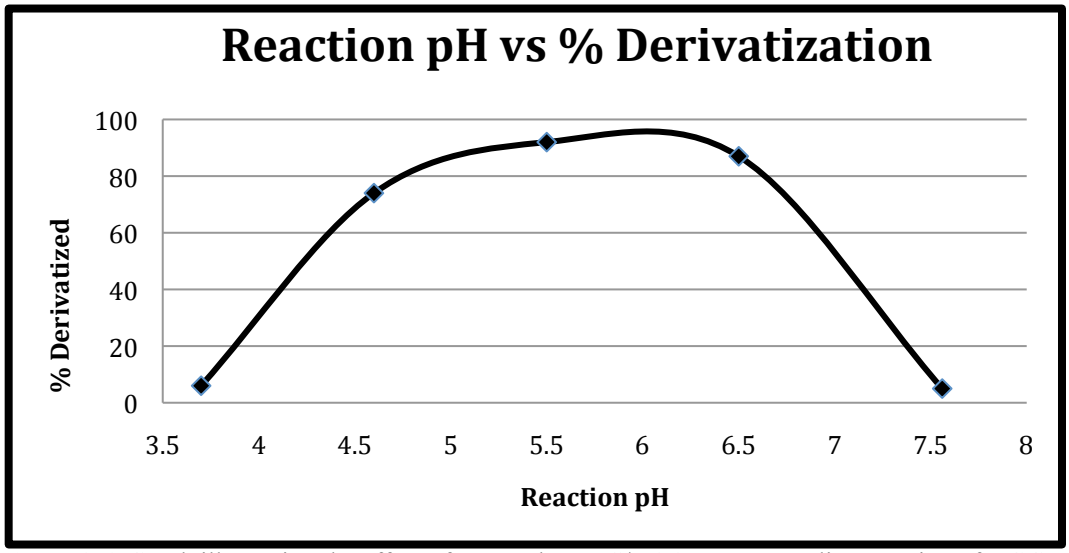


Figure 2.6. Graph illustrating the effect of pH on the EDC/HOAt cross-coupling reaction of YGGFL with histamine. All reactions were allowed to react for two hours at room temperature before being taken to dryness and analyzed. The derivatized form of YGGFL exists when histamine is attached to the c-terminus of the peptide. Thus, the percentage of fully derivatized YGGFL is calculated by comparing the mass areas of the derivatized form to the unreacted starting material.

The data in **Figure 2.6** clearly show pH has a dramatic influence on the outcome of this reaction. These data show the optimal pH for this reaction is roughly between 5.5 and 6, which is dramatically different from the 7.5-7.8 range specified by Xu *et al*.[1] According to our data, the cross-coupling reaction hardly proceeds at all under basic conditions. It is clear that mild acidic conditions are required for this reaction to produce

a meaningful yield, which is in agreement with previous work on carbodiimide-based derivatization reactions in our lab.[3] Under basic conditions, it is likely the carbodiimide nitrogen fails to become protonated in sufficient concentration to activate carboxylic acid carbons for nucleophilic attack.

2.4.6 Examination of Data from Optimized Yield Conditions

In order to get an accurate perspective of how this optimized reaction could be applied to more complex samples, data from the pH 5.5 reaction were analyzed further to determine the incidence of side reactions/incomplete derivatization reactions. The resulting chromatogram is shown in **Figure 2.7**, and detailed information regarding reaction yields for each peptide is shown in **Table 2.3**.

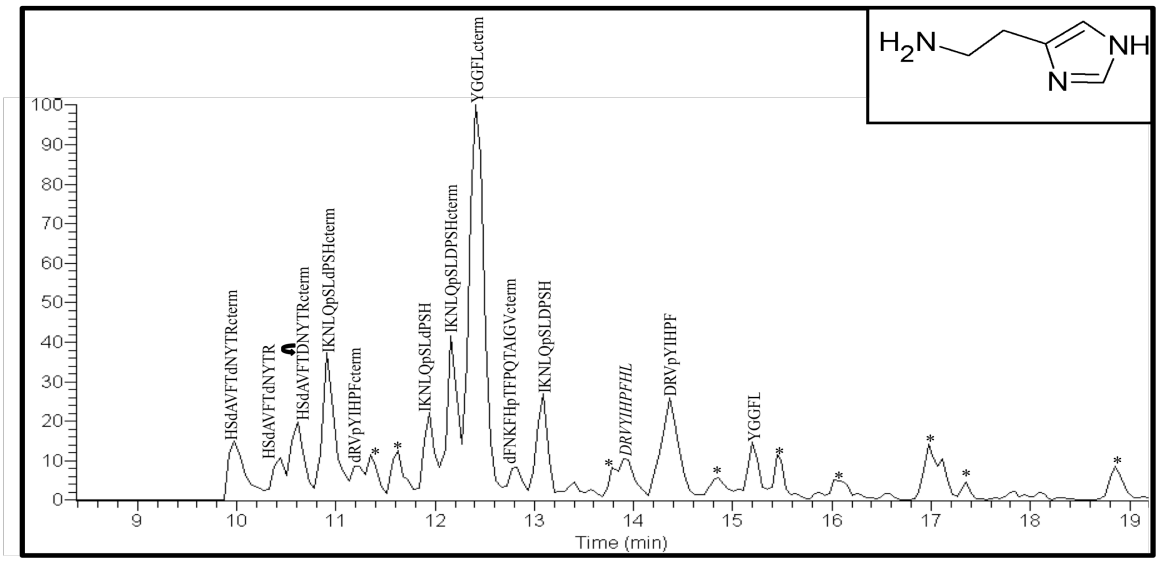


Figure 2.7. Labeled base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from a two hour EDC/HOAt/Histamine derivatization reaction at pH 5.53 (measured using microelectrode). The inset shows a chemical structure of histamine, which was used as the charge-enhancing reagent. Lower case D or the suffix "cterm" indicate a histamine molecule has been attached to that site on the peptide. The asterisk denotes the predominant species under a particular peak is a non-peptidic contaminant. The circular arrow above aspartic acid (D) residues indicates an intramolecular cyclization reaction took place. Gradient: 0-100% B in 17 min., hold at 100% B for 2min, drop back to 0% B in 3min. Standard added: 250 fmol of underivatized and angio (labeled in italic).

Table 2.3. Summary of observed reaction products. The listed percentages were calculated based on the mass area of all observed forms for the given peptide. "C-term only" refers to a form of the peptide in which only the c-terminal COOH group was derivatized without any side reactions. "Incomplete: no D cyclizations" refers to a summation of all partially derivatized forms of the peptide in which no aspartic acid cyclizations occurred. "Incomplete: with D cyclizations" refers to a summation of all partially derivatized forms of the peptide in which one or more aspartic acid cyclizations took place (which prevents that residue from reacting further). "Completely unreacted" refers to the relative amount of the peptide that did not undergo derivatization. "Fully derivatized" refers to the relative amount of peptide in which all COOH groups underwent derivatization. ND: None detected. N/A: Not applicable.

Peptide	c-term only	Incomplete: no D cyclization	Incomplete: with D cyclization	Completely Unreacted	Fully Derivatized
HSDAVFTDNYTR (vaso)	ND	25%	35%	6%	34%
YGGFL	-	N/A	N/A	8%	92%
DRVpYIHPF	36%	10%	10%	17%	28%
IKNLQpSLDPSH	21%	11%	7%	20%	41%
DFNKFHpTFPQTAIGV	35%	4%	3%	5%	53%

Compared to previous chromatograms shown in this section, data presented in **Figure 2.7** is a dramatic improvement. Instead of non-peptidic contaminants dominating the base peak, reaction peptides are now the dominant species. Upon closer inspection, most of these peptides are the expected products from the cross-coupling reaction. This is a clear demonstration the optimized methodology can be used to help sequence peptides in a complex sample (such as a whole protein digest). However the detailed reaction analysis presented in **Table 2.3** highlights several caveats to utilizing this derivatization strategy.

Chief among these caveats is the fact that peptides containing more than one carboxylic acid group are not likely to react in quantitative yield. Indeed, the percentage of a peptide found in it's fully derivatized form falls to the 50% level or less when one or more carboxylic acid groups are present. This leaves behind partially derivatized forms and peptides that have undergone internal cyclization reactions, both of which increase the complexity of the original sample. Of all the potential carboxylic acid groups to derivatize, c-terminal residues confer the greatest benefit for the purpose of peptide sequencing via ETD (ensures both N and C–terminal fragments will be charged). Unfortunately based on these data, a disparity exists between peptides tested in this reaction. Unlike the other four peptides, the c-terminus of vaso does not appear to react as readily as the two other aspartic acid residues on the peptide.

The yields obtained from this optimized reaction do not even come close to the 90% yields reported by Xu *et al*.[1] It is not immediately clear why this disparity exists. All of the reagents are in substantial excess, compared to the number of modifiable

carboxylic acid residues, and the reaction was allowed to incubate for two hours instead of only a couple seconds. Both of these factors should be promoting high yields, but this is unfortunately not the case.

If these observations hold true with a complex sample, conducting a *de novo* sequencing experiment will prove challenging (though not impossible). However these data do highlight potential uses for diagnostic purposes. In cases where the sequence of a protein is known, this methodology can be used to derivatize digest fragments thought to contain PTMs of interest. Because the amino acid sequences of these fragments are known, it is relatively straightforward to predict the *m/z* ratios of derivatized forms. A directed search can then be performed on the resulting LC/MS data to find an MS2 containing sequence information about a region of interest.

2.4.7 Study of Reaction Solvent Conditions and Necessity of HOAt

With an optimized reaction pH, we next moved to determine if comparable yields could be obtained using different solvent systems. It has been a longstanding goal of this laboratory to develop on-column compatible charge enhancement chemistry whereby a sample could be loaded onto a column, derivatized, washed and subsequently eluted into a mass spectrometer for analysis. As it stands, the methodology presented in this dissertation makes use of high amounts of organic solvent (~20% DMF) in the reaction. Such a high amount of solvent is not likely to allow for peptide retention on a reverse-phase chromatography column while undergoing derivatization. In this study, we tested three different solvent conditions: completely aqueous, 20% methanol (MeOH), and 20%

acetonitrile (ACN). As with the previous optimization study, the peptide YGGFL was used to monitor the effectiveness of derivatization under each solvent condition. In carrying out this study, all reactions were prepared as described in *methods 2.3.4* (pH= 5.5 ±0.1), except DMF was substituted for an equal volume of the particular solvent under study.

When preparing reagent solutions for this study, it was noticed that HOAt was completely insoluble in water, and only sparingly soluble in ACN. As a result, it was first necessary to study the effect of HOAt on reaction outcome and determine if its presence is actually necessary for high yields to be obtained. To this end, a reaction was prepared according to *methods 2.3.4*, except pure DMF was added in place of the 2 mg/mL solution of HOAt. The results from this comparative analysis are shown in **Figure 2.8**

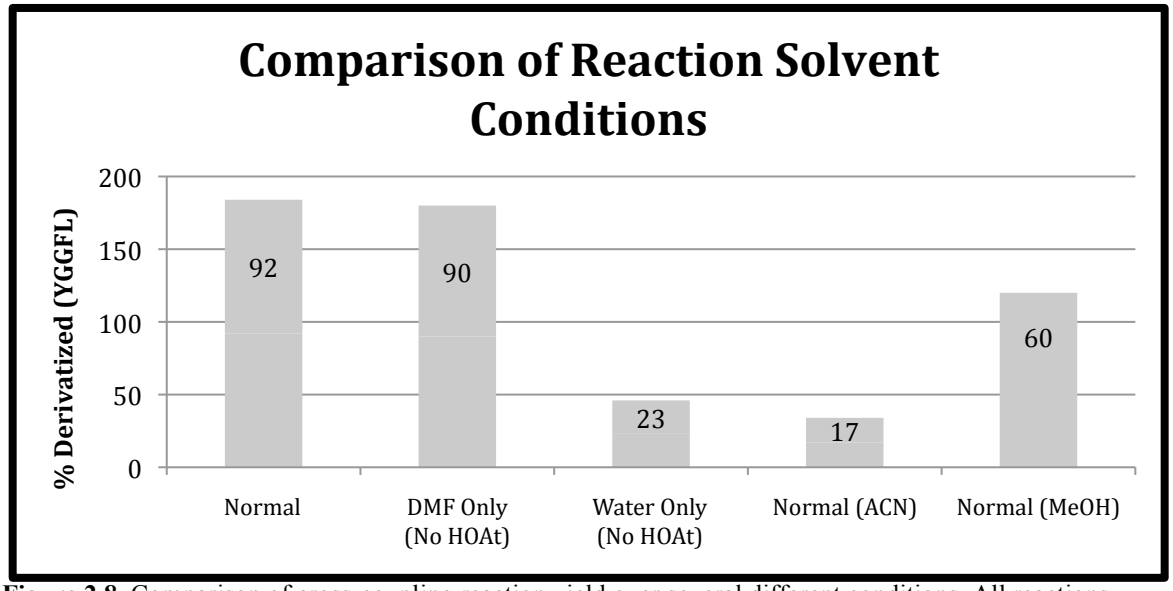


Figure 2.8. Comparison of cross-coupling reaction yield over several different conditions. All reactions were setup at pH 5.5 ±0.1 and allowed to react for two hours. Histamine was used as the charge-enhancement reagent. Normal: Prepared as described by *methods 2.3.4*. DMF Only (No HOAt): Same as normal, except HOAt was not added to reaction. Water Only (No HOAt): Reaction was prepared with all aqueous solutions, no HOAt was added (due to insolubility). Normal (ACN): Prepared same as normal, except DMF was replaced with ACN (HOAt was not fully soluble in ACN solution). Normal (MeOH): Prepared same as normal, except DMF was replaced with MeOH.

Based on the data presented in **Figure 2.8**, it appears the presence of HOAt may not be required for high yields to be obtained. When HOAt was removed from the reaction completely, the yield appears to be comparable to the "normal" conditions when it is present in the reaction. These data show solvent has more of an influence on reaction outcome than the presence of HOAt alone. Removing DMF completely and replacing it with water or ACN produced a dismal yield (23% and 17%, respectively). Under MeOH conditions more than half of YGGFL was derivatized (60%), but the yield was not nearly as high as could be obtained when DMF was present.

The difference between the normal reaction conditions, and conditions where HOAt is removed may or may not be statistically significant. However, when these data

are considered as a whole it appears the presence of DMF has the largest impact on reaction outcome. HOAt may help contribute to a high yield, but it is likely to be a minor rather than a major contributor. The synthetic chemistry procedure[6] from which the Xu *et al*[1] protocol was derived makes use of comparatively more concentrated reagents to carry out cross-coupling reactions. It is possible the relatively dilute conditions presented in this dissertation negate any benefit afforded by HOAt. However, the presence of HOAt does not seem to cause any undesirable effects on chromatography during a reaction analysis. Until more studies are done to demonstrate otherwise, we did not see any reason to remove it from the reaction.

Unfortunately, the finding that organic solvent is needed for this reaction will make it difficult to attempt on-column chemistry. Initial experiments found that a 20% solution of DMF caused peptides to elute from a C18 reverse-phase chromatography column. Given this reality it appears the methodology presented here is not going to be compatible with on-column derivatization as long as reverse-phase columns are employed.

2.4.8 Study of Reaction Yield Over Time

As part of our optimization process, we were curious to know how long the EDC/HOAt reaction needed to incubate at room temperature before maximum yield was achieved. In the initial attempts to reproduce the methods from Xu *et al*, two hours of incubation time was set in a somewhat arbitrary manor. We were interested to know if a two-hour incubation time is actually necessary, or if the reaction can be taken to dryness

sooner. To answer this question, several duplicate reactions were setup as described in *methods 2.3.4*, set to pH 5.5 ±0.1 with histamine as the charge enhancement reagent, and allowed to react for a set amount of time. Results from this reaction are summarized in **Figure 2.9**.

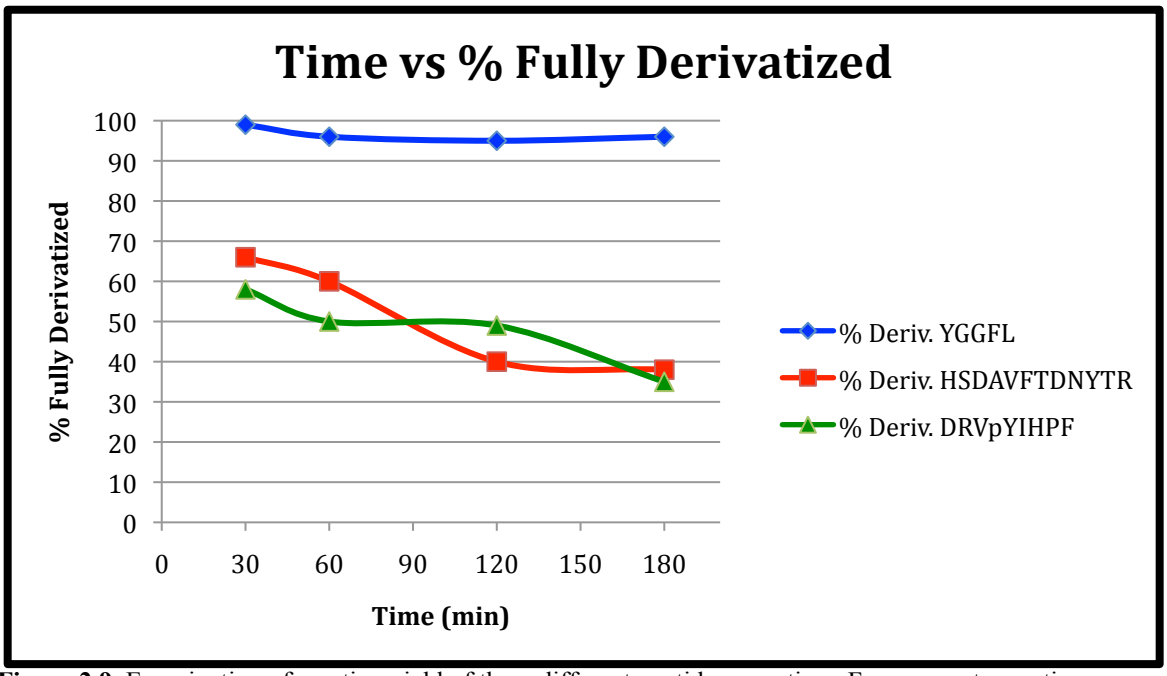


Figure 2.9. Examination of reaction yield of three different peptides over time. Four separate reactions were setup at pH 5.5 with histamine as the charge enhancement reagent and allowed to incubate at room temperature for 30, 60, 120, and 180 minute intervals before being taken to dryness. The amount fully derivatized is calculated as a percentage of the total mass area of all detected forms of the peptide.

The data presented in **Figure 2.9** demonstrate that maximal reaction yield is achieved after only 30 minutes. Letting the reaction stand for two hours does not seem to have an appreciable effect on reaction yield. A study on the hydrolysis of EDC found that after 60 minutes in aqueous solution at pH 5.8, half of the original EDC is converted into a non-reactive urea derivative.[7] At pH 4.5, total conversion into the urea species occurs after only 15 minutes.[7] As soon as an EDC molecule is hydrolyzed into this urea derivative, it is no longer able to participate in the cross-coupling reaction. Given the

findings from the hydrolysis study, it is not surprising to see that peptide derivatization remains relatively unchanged after 30 minutes of incubation. After this time most of the EDC reagent has undergone hydrolysis, which prevents any further peptide reactions from taking place.

In **Figure 2.9** there appears to be a slight downward trend over time for the peptides HSDAVFTDNYTR and DRVpYIHPF. Because each time point was not conducted in triplicate, it is not possible to determine if this trend is significant. Given the solid nature of an amide bond and relatively mild acidic conditions present in the reaction hydrolysis of the attached histamine molecule is not likely. During analysis of the 180 minute time point, many non-peptidic species were found in the resulting chromatogram which could have suppressed ionization of some peptides and resulted in low area counts of HSDAVFTDNYTR and DRVpYIHPF. The observation that percent conversion of YGGFL stayed relatively constant over time suggests ion suppression is the likely scenario for the downward trend observed in the other two peptides.

After considering these findings, we realized there was no added benefit to letting the reaction incubate on the bench top for two hours. 30 minutes of incubation appears to be sufficient to achieve an optimal yield.

2.4.9 Trial of Optimized EDC/HOAt Reaction on a Whole Protein Digest

With an optimized protocol for peptide derivatization, we decided to scale up the experiment and attempt a derivatization reaction on a whole protein digest. To this end, a tryptic digest of bovine serum albumin (BSA) was chosen as a test protein. BSA has a

known sequence, which enables us to fully evaluate the efficiency of the derivatization reaction. Based on our estimates, the final concentration of free carboxylic acids (including added standards) in this reaction was approximately 1.5 μM (see *methods 2.3.9*), which still leaves all reagents in a thousand-fold excess.

Initial attempts of this reaction were carried out on a BSA digest taken to dryness without any prior cleanup. Unfortunately, were unable to find any evidence of derivatization taking place. We speculated the reaction failed due to the presence of excess Dithiothreitol (DTT) from the sample preparation process (see *methods 2.3.5*). The thiol group on DTT is a very strong nucleophile and can react preferentially with EDC to prevent any cross-coupling reactions from taking place. By cleaning up the BSA digest prior to setting up a reaction (see *methods 2.3.8*), we were able to detect successful peptide derivatization on BSA digest peptides. The resulting chromatographic data from this experiment is shown in **Figure 2.10**.

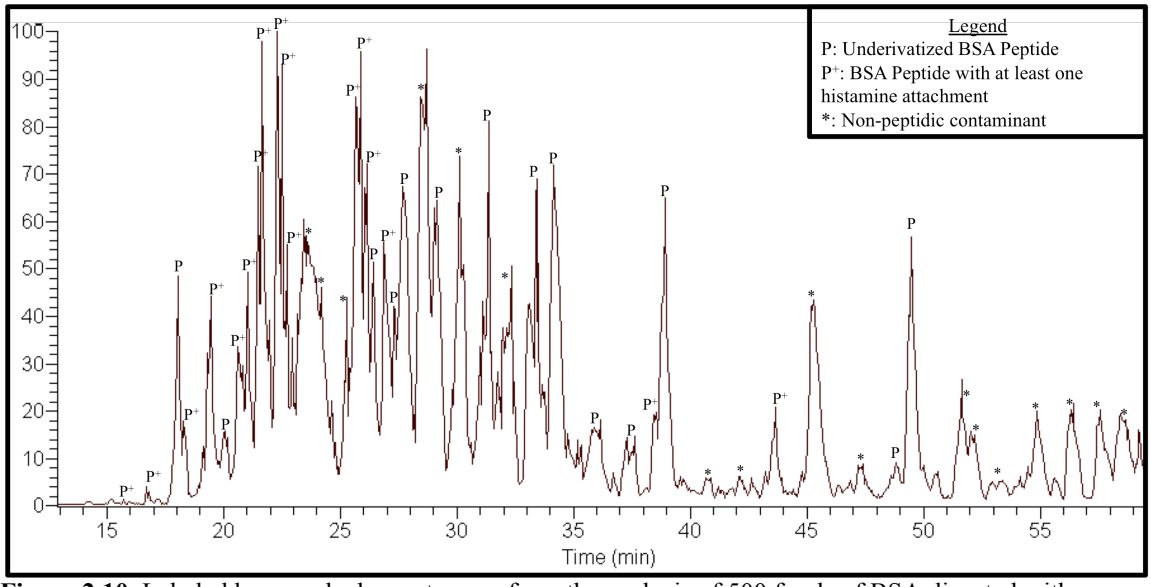


Figure 2.10. Labeled base peak chromatogram from the analysis of 500 fmols of BSA digested with trypsin and derivatized with histamine at pH 5.5 for 40 minutes. Peptides visible in the base peak are labeled according to the legend (see inset). Gradient: 0-60% B in 60 minutes. Standard added prior to analysis: 250 fmol of underivatized and angio (not visible in base peak).

The data presented in **Figure 2.10** clearly show a derivatization reaction occurred on tryptic BSA peptides. Unfortunately, the standard peptides added to the reaction were not detected in our analysis. Their absence suggests the precolumn may have been overloaded, and some peptides were lost during the wash prior to analysis. Thorough a combination of computer-assisted database searching, manual searching for expected peptides, and *de novo* sequencing a list of BSA peptides containing at least one histamine molecule was compiled. This information was then used to determine how much of the BSA protein as a whole could be covered using only these derivatized peptides. Results from this analysis are shown in **Figure 2.11**.

MKWVTFISLLLLFSSAYSRGVFRRDTHKSEIAHR*FKDLGEEHFK*GLVLIAFSQYLQQCPFDEHVK

*LVNELTEFAK*TCVADESHAGCEK*SLHTLFGDELCK*VASLRETYGDMADCCEKQEPERNECFLSHK

DDSPDLPKLKPDPNTLCDEFKADEKKFWGK*YLYEIARRHPYFYAPELLYYANK*YNGVFQECCQAED

KGACLLPKIETMREKVLASSARQRLRCASIQKFGER*ALKAWSVAR*LSQKFPK*AEFVEVTK*LVTDLT

KVHKECCHGDLLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVEKDAIPENLPP

LTADFAEDKDVCKNYQEAK*DAFLGSFLYEYSRRHPEYAVSVLLR*LAKEYEATLEECCAKDDPHACY

STVFDKLK*HLVDEPQNLIK*QNCDQFEK*LGEYGFQNALIVR*YTR*KVPQVSTPTLVEVSR*SLGKVGTRC

CTKPESER*MPCTEDYLSLILNRLCVLHEKTPVSEK*VTKCCTESLVNRRPCFSALTPDETYVPKAFDE

KLFTFHADICTLPDTEKQIK*KQTALVELLK*HKPKATEEQLK*TVMENFVAFVDK*CCAADDKEACFAV

EGPK*LVVSTQTALA* (607)

Figure 2.11. Sequence of BSA showing regions covered using only histamine-derivatized peptides from tryptic digest. Italic/underlined: Regions of BSA sequence covered using only histamine-derivatized fragments.

Based on the sequence coverage presented in **Figure 2.11**, 32% of the entire BSA sequence can be accounted for using peptides derivatized with histamine. Further data mining would most likely reveal that 32% is a conservative estimate. Given the number of peptides generated from a tryptic digest, there is great potential for many combinations of intramolecular cyclizations and incomplete derivatization to occur. Unless the sequence of a region of interest is known before an experiment, it is very difficult to manually sequence and identify these fragments. Also, the fact that reaction standards were lost prior to analysis suggests some histamine-derivatized BSA peptides may have been lost as well. A repeat of this experiment using a column with a higher capacity would likely provide improved sequence coverage.

Failure to detect the added reaction standards makes it impossible to precisely determine the overall yield of the reaction. However it was possible to obtain a rough

estimate using BSA peptides that were detected in the reaction chromatogram. This estimate is presented in **Table 2.4**.

Table 2.4. Digest derivatization yield estimate. The listed percentages were calculated based on the mass area of all observed forms for the given peptide. "C-term only" refers to a form of the peptide in which only the c-terminal COOH group was derivatized without any side reactions. "Incomplete: no side reaction" refers to a summation of all partially derivatized forms of the peptide in which no side reactions occurred. "Incomplete/side-reactions" refers to a summation of all partially derivatized forms of the peptide in which at least one side reaction took place (D cyclization or acetylation). "Completely unreacted" refers to the relative amount of the peptide that did not undergo derivatization. "Fully derivatized" refers to the relative amount of peptide in which all COOH groups underwent derivatization. ND: None detected. N/A: Not applicable.

Peptide	c-term only	Incomplete: no side reaction	Incomplete/side-reactions	Completely Unreacted	Fully Derivatized
DAFLGSFLYEYSR	20%	37%	ND	34%	8%
YLYEIAR	36%	36%	ND	27%	0.4%
ALKAWSVAR	N/A	N/A	6%	92%	2%

Based on these three peptides, the yield obtained from the digest reaction is considerably lower than was obtained from the mix of five peptides used to develop our derivatization methodology. This may suggest the concentration of carboxylic acid present in the reaction can impact reaction outcome despite an overwhelming excess of reagents. This BSA derivatization reaction represents more than a ten-fold increase in the concentration of carboxylic acid compared to previous reaction experiments. However the fact that standards were not recovered, makes it difficult to definitively determine why low yields were obtained. It is also possible that fully derivatized peptides were simply not retained as well as the unreacted forms, which would give the appearance of a low yield during analysis. More tests with BSA are needed to determine which is scenario is responsible for the low observed yield.

2.4.10 Evaluation of Methyl Benzimidazole as a Peptide Derivatization Reagent

At this point in our method development process a new potential charge enhancement reagent, methyl benzimidazole, became available through the Sigma-Aldrich catalog. A structure of this molecule is shown in **Figure 2.1(c)**, along with histamine and benzimidazole. Comparison of the three molecules in **Figure 2.1** demonstrates a relative increase in hydrophobicity starting at histamine and moving towards methyl benzimidazole. From previous experiments shown in this dissertation, we know that peptides derivatized with histamine and benzimidazole become more hydrophilic than their respective underivatized forms. However the extent to which this shift occurs is far less pronounced when benzimidazole is used. We hypothesized the extra methyl group present on methyl benzimidazole would be enough to reverse this shift and cause peptides to become more hydrophobic than their underivatized forms. In order to test this hypothesis, we first needed to determine if methyl benzimidazole could be used to derivatize peptides in high enough yield to make this comparison.

This reaction was carried out as described in *methods 2.3.4* with a final pH of 5.4 and an incubation time of 120 minutes. The results from this reaction are presented in **Figure 2.13** and **Table 2.5**.

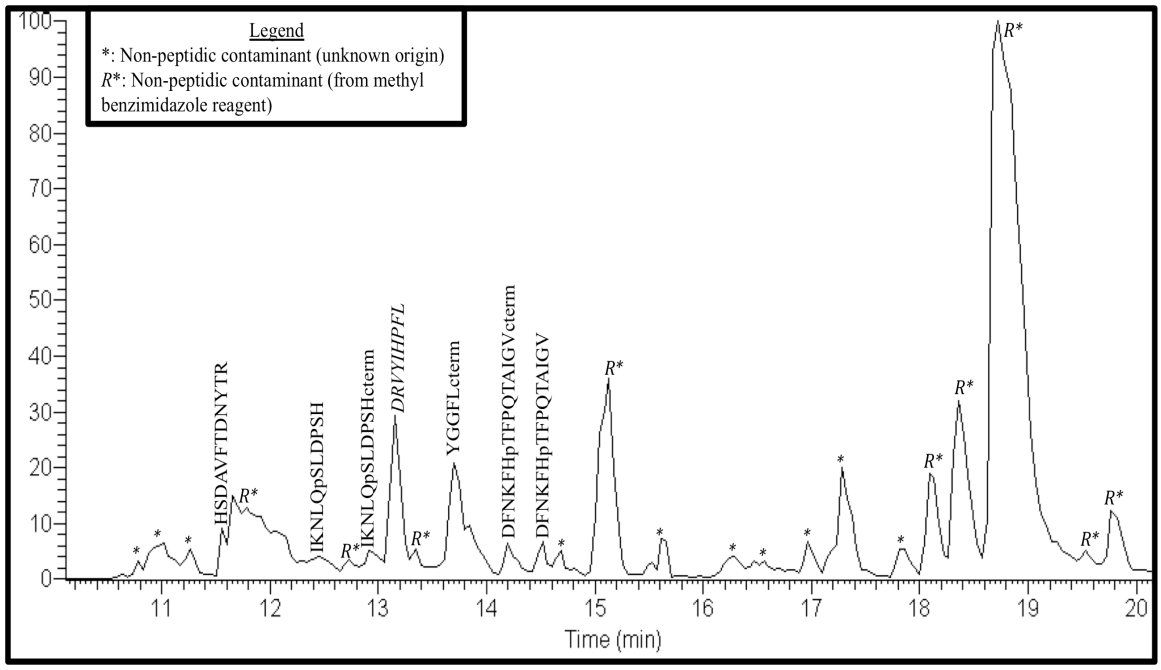


Figure 2.13. Labeled base peak chromatogram from the analysis of a 1 pmol (of each peptide) aliquot from a two hour EDC/HOAt/methyl benzimidazole derivatization reaction at pH 5.38 (measured using microelectrode). Lower case D or the suffix "cterm" indicate a methyl benzimidazole molecule has been attached to that site on the peptide. A separate reagent blank experiment identified contaminants unique to the methyl benzimidazole reagent. Gradient: 0-100% B in 17 min., hold at 100% B for 2min, drop back to 0% B in 3min. Standard added: 250 fmol of underivatized and angio (labeled in italic).

Table 2.5. Summary of observed reaction products. The listed percentages were calculated based on the mass area of all observed forms for the given peptide. "C-term only" refers to a form of the peptide in which only the c-terminal COOH group was derivatized without any side reactions. "Incomplete: no D cyclizations" refers to a summation of all partially derivatized forms of the peptide in which no aspartic acid cyclizations occurred. "Incomplete: with D cyclizations" refers to a summation of all partially derivatized forms of the peptide in which one of more aspartic acid cyclizations took place (which prevents that residue from reacting further). "Completely unreacted" refers to the relative amount of the peptide that did not undergo derivatization. "Fully derivatized" refers to the relative amount of peptide in which all COOH groups underwent derivatization. ND: None detected. N/A: Not applicable. Asterisk denotes a peptide product, which was not confirmed by MS2.

Peptide	c-term only	Incomplete: no D cyclization	Incomplete: with D cyclization	Completely Unreacted	Fully Derivatized
HSDAVFTDNYTR (vaso)	15%*	24%	22%	15%	24%
YGGFL	-	N/A	N/A	7%	93%
DRVpYIHPF	56%	ND	ND	32%	12%
IKNLQpSLDPSH	14%	15%	ND	58%	13%
DFNKFHpTFPQTAIGV	57%	ND	ND	20%	23%

The data presented in **Figure 2.13** show peptides can be successfully derivatized with methyl benzimidazole. However many non-peptidic contaminants remain, despite an

on-column wash prior to analysis. By conducting a separate "reagent blank" experiment, many of these peaks were found to originate from the methyl benzimidazole reagent itself. This particular reagent is new to the Sigma-Aldrich catalog, and there is no available data to determine reagent purity. Several of these contaminants elute in at similar time to the peptides used for this study, which makes design of a removal strategy challenging. Re-crystallization and/or further purification of the methyl benzimidazole reagent may be needed before this reaction is tested on a more complex sample.

In terms of overall yield, methyl benzimidazole does not appear to fully derivatize peptides as effectively as histamine. However the yield is high enough to setup a comparative study of a given reagent's effect on peptide retention time (see results section 2.4.12). The amount of YGGFL found in the fully derivatized form is comparable to what was obtained using histamine. There also appears to be a slight preference for derivatizing C-terminal residues on DRVpYIHPF and DFNKFHpTFPQTAIGV, compared to what was obtained using histamine. Given the comparatively bulky nature of methyl benzimidazole compared to histamine, the less hindered C-terminal residue may be a preferred (though not exclusive) site of reactivity.

2.4.11 Comparison of Derivatized Peptide Retention Time

With sufficient amounts of peptides fully derivatized with methyl benzimidazole, we setup a comparison of three different derivatization agents. Peptides derivatized with histamine, benzimidazole, and methyl benzimidazole were all loaded onto the same column in order to determine the relative effect each reagent has on retention time

compared to the native form of each peptide. Selected ion chromatograms from this study are shown in **Figure 2.14**.

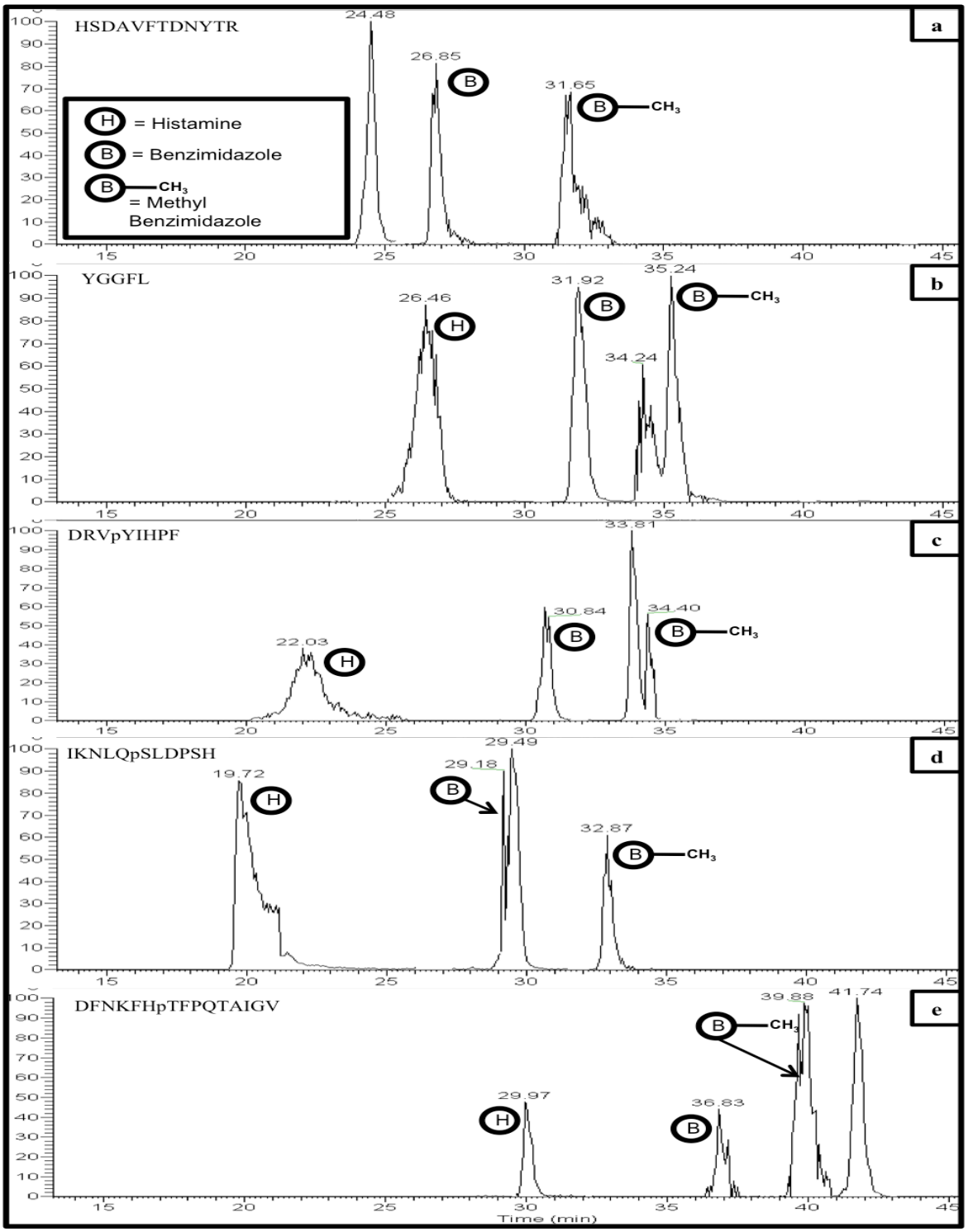


Figure 2.14. Analysis of effects on peptide retention time upon full derivatization with histamine, benzimidazole, and methyl benzimidazole. **a-e**: selected ion chromatograms for the five test peptides used in our derivatization reactions. The box in (**a**) correlates each symbol with its representative charge enhancement reagent. Each labeled peak represents the fully derivatized from of the listed peptide in each chromatogram. The unlabeled peak in each chromatogram is the underivatized form of the listed peptide, which serves as the basis for comparison of each derivatization reagent. Note: the histamine-derivatized form of HSDAVFTDNYTR was not retained on the column during analysis of this mixture of reaction products. Gradient: 0-60% B in 60 minutes.

The data presented in **Figure 2.14** shown an overall trend among the three reagents tested in this comparison. Histamine-derivatized peptides elute first, followed by those derivatized with benzimidazole and methyl benzimidazole, respectively. In four out of the five peptides tested, derivatization with benzimidazole and histamine produces a more hydrophilic peptide (to a greater extent with histamine). Interestingly, HSDAVFTDNYTR (vaso) became more hydrophobic when derivatized with benzimidazole (**Figure 2.14a**). This result is in direct contradiction to the data presented in **Figure 2.3**, where benzimidazole-derivatized vaso was found to be more hydrophilic than its underivatized form. It is not immediately clear as to why this trend is only observed with vaso, and not with any of the other peptides. A minor shift in packing material may be the culprit behind this unexpected change in retention time.

When comparing peptides derivatized with methyl benzimidazole to the underivatized form, there is an overall increase in retention time in all but one of the peptides. This confirms our hypothesis that adding an extra methyl group to the fused aromatic ring system is enough cause an increase in peptide hydrophobicity, following derivatization.

The fact that all peptides derivatized with benzimidazole and methyl benzimidazole were retained on the HPLC column demonstrates that a fused aromatic ring system has a stronger binding affinity to reverse-phase packing material compared to an imidazole ring alone. This provides a possible explanation as to why the sequence of coverage of BSA was only 32% in **Figure 2.11**. Many of the histamine-derivatized BSA peptides may have been lost during the wash prior to analysis. If comparable yields can

be obtained with methyl benzimidazole, it is likely a higher proportion of BSA could be covered using only derivatized peptides.

2.4.12 Evaluation of Methyl Benzimidazole as a Charge Enhancement Reagent

Given the promising trend of increasing peptide retention time upon derivatization with methyl benzimidazole, we next moved to evaluate its potential as a charge enhancement reagent for improving sequence coverage using ETD. To this end, the peptide DFNKFHpTFPQTAIGV was used as a basis for comparison. As shown in **Figure 2.15**, fragmentation of this peptide in its underivatized form using ETD results in mostly N-terminal fragment ions (c-type ions). If an attempt was made to *de novo* sequence this or a similar peptide, the lack of overlapping sequence coverage would not provide enough informative fragments to confidently identify the peptide.

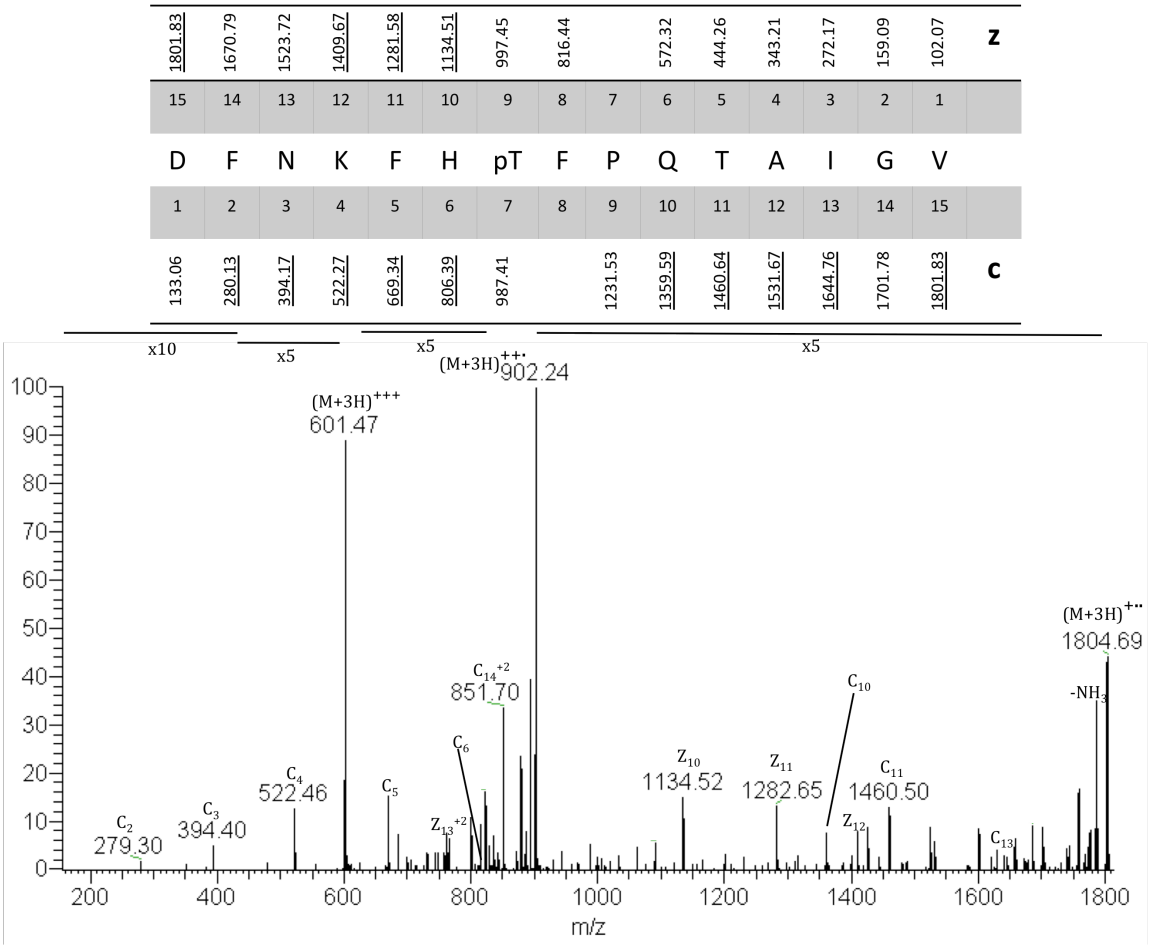


Figure 2.15. ETD spectra of the +3 ion DFNKFHpTFPQTAIGV (underivatized). The chart at the top of the figure lists all possible +1 c and z fragment ions. Underlined numbers indicate those particular fragment ions were found in the ETD spectra. ETD reagent: azulene, 40 ms ion-ion reaction time.

With the underivatized form of the peptide (**Figure 2.15**) as a basis for comparison, ETD fragmentation was carried out on the methyl benzimidazole-derivatized form. A sample spectra from this experiment is shown in **Figure 2.16**. Compared to the underivatized form of the peptide, ETD fragmentation is significantly improved. The additional charge on the c-terminus allows for nearly complete coverage using z-type ions, which produces overlapping sequence coverage. This result clearly demonstrates potential for methyl benzimidazole as a charge-enhancement reagent. On a novel peptide,

this improvement in sequence coverage would be of great help to confidently *de novo* sequence a peptide.

2116.03	1827.89	1680.82	1566.77	1438.68	1291.61	1154.55	973.54		729.42	601.36	500.31	429.27	316.19	259.17	**z**
15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
d	F	N	K	F	H	pT	F	P	Q	T	A	I	G	V	cterm
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
290.16	437.23	551.27	679.37	826.44	963.49	1144.51		1388.63	1516.69	1617.74	1688.77	1801.86	1858.88	2116.03	**c**

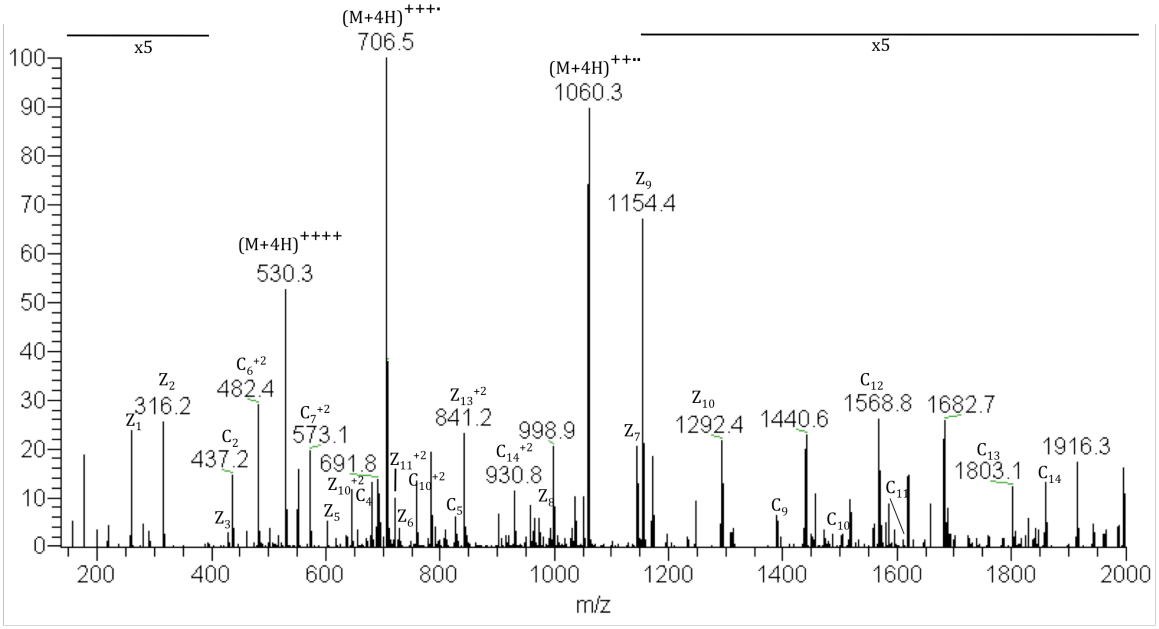


Figure 2.16. ETD spectra of the +4 ion DFNKFHpTFPQTAIGV (derivatized with methyl benzimidazole). The chart at the top of the figure lists all possible +1 c and z fragment ions. Underlined numbers indicate those particular fragment ions were found in the ETD spectra. ETD reagent: azulene, 40 ms ion-ion reaction time. Note: the underivatized form of the peptide exists in a +2/+3 charge state. Upon derivatization with methyl benzimidazole, this charge state distribution is increased to +3/+4.

2.4.13 Further Examination of Incubation Time on Reaction Outcome

Based on reaction yield data from the methyl benzimidazole/EDC/HOAt reaction shown in **Table 2.5**, we were curious to know if incubating the reaction for a short period of time could afford some degree of selectivity for derivatizing only the c-terminus of

peptides. In previous trials of EDC cross-coupling reactions, termination was accomplished by placing samples in a vacuum centrifuge. This method does not immediately terminate the reactions because it usually requires at least 30 minutes to take a sample to dryness. In order to terminate the reaction immediately following incubation, we dramatically lowered the pH by adding trifluoroacetic acid to the sample as described in *methods* 2.3.4. Using this method, we investigated the reaction yield after 30 seconds (pH = 5.53) and five minutes (pH=5.51) of incubation. Results from these experiments are shown in **Table 2.6(a)** and **Table 2.6(b)**, respectively.

Table 2.6. Summary of reaction products from 30 second and 5 minute methyl benzimidazole/EDC/HOAt reactions. The listed percentages were calculated based on the mass area of all observed forms for the given peptide. "C-term only" refers to a form of the peptide in which only the c-terminal COOH group was derivatized without any side reactions. "Incomplete: no D cyclizations" refers to a summation of all partially derivatized forms of the peptide in which no aspartic acid cyclizations occurred. "Incomplete: with D cyclizations" refers to a summation of all partially derivatized forms of the peptide in which one of more aspartic acid cyclizations took place (which prevents that residue from reacting further). "Completely unreacted" refers to the relative amount of the peptide that did not undergo derivatization. "Fully derivatized" refers to the relative amount of peptide in which all COOH groups underwent derivatization. ND: None detected. N/A: Not applicable.

a: Summary of observed reaction products from a 30 second reaction.

Peptide	c-term only	Incomplete: no D cyclization	Incomplete: with D cyclization	Completely Unreacted	Fully Derivatized
HSDAVFTDNYTR (vaso)	1%	ND	6%	93%	ND
YGGFL	-	N/A	N/A	99%	1%
DRVpYIHPF	1%	3%	ND	96%	ND
IKNLQpSLDPSH	1%	1%	2%	95%	ND
DFNKFHpTFPQTAIGV	ND	ND	ND	100%	ND

b: Summary of observed reaction products from a 5 minute reaction.

Peptide	c-term only	Incomplete: no D cyclization	Incomplete: with D cyclization	Completely Unreacted	Fully Derivatized
HSDAVFTDNYTR (vaso)	1%	ND	4%	95%	ND
YGGFL	-	N/A	N/A	99%	1%
DRVpYIHPF	2%	ND	ND	98%	ND
IKNLQpSLDPSH	0.5%	0.5%	1%	98%	ND
DFNKFHpTFPQTAIGV	ND	ND	3%	97%	ND

Based on the data presented in **Table 2.6**, the selectivity we were hoping to observe for reacting with the c-terminus was not found to be a dominant product. Instead, the result was a practically non-existent yield after letting the reaction incubate for five minutes or less. This is in stark contrast to the data presented in **Table 2.5**, where the reaction was allowed to proceed for approximately two hours and produced exceeding higher yields. Taken together these data suggest the methyl benzimidazole/EDC/HOAt reaction takes between five minutes and two hours to produce a meaningful yield. More trials are necessary to determine when optimal yields are achieved in this interval.

2.5 Conclusion

The original goals in this thesis were to develop a "clean" peptide derivatization methodology that promotes enhanced ETD fragmentation, and increases overall peptide hydrophobicity. The latter of these goals has proved to be the most challenging due to contaminants present in some of the charge enhancement reagents. We initially attempted to cleanup a benzimidazole derivatization reaction via an acetone precipitation, which helped remove some of the contaminating species. From this, we were able to learn that peptides derivatized with benzimidazole are still more hydrophilic compared to their native forms. However this shift was far less pronounced compared to what has been observed with histamine. The results from this reaction also highlighted the need to pursue an alternative derivatization method, given the relatively high amounts of non-peptidic material still remaining in the sample. Thus, the acetone precipitation strategy

was abandoned in favor of a previously reported alternative synthesis method, which utilized minute quantities of amine reagent and reported high yields.

After many attempts to re-create and optimize reaction conditions, we were able to develop an adaptable derivatization method that can be used to study several charge enhancement reagents. We started with a broken derivatization method presented by Xu *et al*,[1] and carried out systematic tests to determine the correct conditions for modifying peptide carboxylic acid residues using EDC/HOAt cross-coupling chemistry. To this end, we presented data that clearly demonstrates an acidic pH and solvent composition are of critical importance when using EDC chemistry to modify carboxylic residues. We also demonstrated the means by which this pH is measured (ie: test strips vs electrode) can be equally important. Unfortunately, the yields obtained from our optimized reaction did not reach the 90% level for all peptides as had been advertised. However, we were able to demonstrate our optimized reaction conditions could be used to study a whole protein digest. In addition, we were able to use our methodology to conduct a comparative analysis of several different charge enhancement reagents and ultimately find one that meets our desired criteria of affording increased hydrophobicity and enhancing peptide charge.

This thesis has demonstrated that peptides can be derivatized with methyl benzimidazole. Upon derivatization we were able to demonstrate improvement in ETD fragmentation for a charge-deficient peptide, combined with an overall increase in hydrophobicity. The protocol we developed here uses substantially lower concentrations of amine reagent, compared to the method our lab has utilized in the past. Although the

resulting chromatograms from methyl benzimidazole reactions are not completely free of contaminants, the levels are low enough that peptides can still be detected. This should allow for our methods to be utilized for diagnostic purposes.

2.6 References

1. Xu, Y., Zhang, L., Lu, H. & Yang, P. Mass Spectrometry Analysis of Phosphopeptides after Peptide Carboxy Group Derivatization. *Anal. Chem.* **80,** 8324–8328 (2008).
2. Carpino, L. A. 1-Hydroxy-7-azabenzotriazole. An efficient peptide coupling additive. *J. Am. Chem. Soc.* **115,** 4397–4398 (1993).
3. English, A. Development and Application of Methodologies to Facilitate Peptide Sequencing with Mass Spectrometry. (2011).
4. Martin, S. E., Shabanowitz, J., Hunt, D. F. & Marto, J. A. Subfemtomole MS and MS/MS Peptide Sequence Analysis Using Nano-HPLC Micro-ESI Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **72,** 4266–4274 (2000).
5. Earley, L. Implementation of a Glow Discharge Reagent Ion Source for the Introduction of ETD Reagent Anions into a Mass Spectrometer. (2010).
6. Ho, G.-J., Emerson, K. M., Mathre, D. J., Shuman, R. F. & Grabowski, E. J. J. Carbodiimide-Mediated Amide Formation in a Two-Phase System. A High-Yield and Low-Racemization Procedure for Peptide Synthesis. *J. Org. Chem.* **60,** 3569–3570 (1995).
7. Paula Lei, Q. *et al.* Kinetic studies on the rate of hydrolysis of N-ethyl-N'-(dimethylaminopropyl)carbodiimide in aqueous solutions using mass spectrometry and capillary electrophoresis. *Analytical Biochemistry* **310,** 122–124 (2002).