DEVELOPMENT OF NOVEL MASS SPECTROMETRIC

PROTOCOLS FOR RAPID PROTEIN CHARACTERIZATION

by

DANIEL ALLEN KING

(Under the direction of Dr. Ron Orlando)

ABSTRACT

Mass spectrometry (MS) has found an indispensable roll in the analysis of biological molecules. In the wake of the completion of the human genome, proteomics work in academic, government, and pharmaceutical laboratories has grown dramatically and the need for fast protein analyzing protocols is a high priority. Mass spectrometry is often the final instrument of analysis after a long and arduous series of separation and processing steps. Although mass spectrometry is often a very quick technique, the analysis time is limited by the slow separation, purification, and processing procedures. Presented here are novel mass spectrometric procedures for characterizing proteins that attempt to increase the speed of analysis and provide complementary approaches to standard protein protocols.

New on-probe procedures have been developed that provide an expedient alternative to standard protocols. For the identification of a protein, a proteolytic digestion must be performed. These enzymatic digestions can typically require 10-20 hours under standard conditions. The performance of proteolytic digestions directly on a MALDI probe has proven in the past to be an effective way to shorten the digestion time to less than an hour. Presented here are two digestion procedures that provide some unique advantages over the typical trypsin digest in addition to their increased speed. Pepsin is shown to work in the presence of MALDI matrix and can be used after previous mass spectra have been generated, and acid hydrolysis is shown to generate very valuable sequence ladders.

Initially used to study protein folding, amide exchange-MS is continually being adapted to the study of new types of protein systems. Presented here is the first adaptation of amide exchange-MS to the study of a protein-carbohydrate binding system. Aside from the speed of amide exchange-MS, the many difficulties that are often encountered in x-ray crystallography and nuclear magnetic resonance (NMR) are avoided. Amide exchange-MS is less sensitive to protein size, limited sample quantities, and varying solubility and flexibility properties of protein and carbohydrate components. Amide exchange-MS is shown to be an essential, complementary technique to x-ray crystallography and NMR.

INDEX WORDS: Proteolysis, On-probe, MALDI-MS, Pepsin, Acid hydrolysis, Deuterium, Amide exchange-MS, Protein conformation, Polygalacturonase, Protein-carbohydrate interaction

DEVELOPMENT OF NOVEL MASS SPECTROMETRIC PROTOCOLS FOR RAPID PROTEIN CHARACTERIZATION

by

DANIEL ALLEN KING

B.S., Huntington College, 1998

A Dissertation Submitted to the Graduate faculty of The University of Georgia in Partial Fulfillment of the Requirements for the degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2001

DEVELOPMENT OF NOVEL MASS SPECTROMETRIC

PROTOCOLS FOR RAPID PROTEIN CHARACTERIZATION

by

DANIEL ALLEN KING

Approved:

Major Professor: Ron Orlando

Committee: Jonathan Amster
James Anderson
John Stickney
Rob Woods

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
August 2002

DEDICATION

This work is dedicated to my love, Carie Ann, for her support and encouragement and to my Lord and Savior, Jesus Christ, for the gift of grace and hope. May God be known through his creation and the work of scientists who hope to better understand Him.

ACKNOWLEDGEMENTS

I would like to extend my appreciation to everyone with whom I have had valuable scientific discussions and to those who have provided much needed breaks from all things intellectual. Several individuals have had significant impacts on my research and have been essential to my success. First, I thank Jenn Colangelo and Maria Warren for their time spent teaching and training me when I was a new graduate student. I hope I have repaid that debt fairly through my efforts to train new students in the lab. I would like to thank Carl Bergmann for his daily support and his willingness to take time to discuss my projects. He has always encouraged my free-thinking and inspires me with his passion for learning. I would like to thank Scot Weinberger and Ciphergen Biosystems, Inc. for the opportunity to collaborate and participate in achieving a patent. Lastly, I thank Ron Orlando for giving me such a variety of experiences during my time here. I appreciate his creativity and the freedom for me to try new things even when they fail.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

Mass spectrometry has become an important instrumentation for many protein biochemists. This rapid analysis technique is often used in connection with isolation or purification techniques like gel-electrophoresis and liquid chromatography. These much slower preparatory methods often negate the effective speed of the mass spectrometer in terms of the total experiment time. The works presented here focus on adapting mass spectrometry procedures to circumvent other slower methodologies while retaining the speed and sensitivity of mass spectrometry.

On-Probe Protein Digestions for Modified MALDI Probe Surfaces:

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry has gained wide acceptance as a quick and robust tool for determining the molecular mass of bio-polymers such as proteins, polynucleotides, and carbohydrates. The analyte is mixed with an acidic matrix that absorbs at the wavelength of the laser and is dried on a MALDI target. After the laser is fired at the target the matrix desorbs from the surface and protonates the analyte. The molecular ions now formed can be accelerated by an electric field down a flight path. Because the acceleration voltage, often ~20-30keV, imparts the same kinetic energy to each charged molecule, their flight times are related to their masses according to the equation: $KE = \frac{1}{2} mv^2$. [1]

There is a growing enthusiasm over proteomics, interested in the rapid identification of proteins, and differential protein expression, interested in monitoring the changes in protein expression by a organism or cell that is stressed or sick. Novel protocols that involve the modification of standard MALDI probes are currently being developed to allow for these exciting and difficult fields of study to be brought into the swift and robust realm of MALDI technology.[2-5]

Typical protocols using MALDI-MS require samples to be processed prior to analysis. If the sample is too complex or contaminated with salts or detergents, gel electrophoresis or liquid chromatography (LC) are often employed.[6, 7] Proteins from cell extracts are commonly separated by 2D-gel electrophoresis. The proteins are subsquently extracted and prepared for analysis. If a protein is to be identified, the protein must first be enzymatically digested so that a peptide spectrum can be generated. Trypsin is usually used because of its high specificity, cleaving at the C-terminal side of arginines and lysines if not followed by proline.[6] Another advantage of trypsin for the MALDI user is that lysines and arginines represent approximately 10% of all amino acids, statistically. This yields an average peptide length of about 10 residues, corresponding to a mass/charge (m/z) ratio of between 1,000 and 1,500 Daltons (Da). This is the optimal range for most MALDI instruments since most matrix interference will occur below 500 Da and resolution is poorer as m/z increases. A protein can be identified by comparing the tryptic peptide masses to a theoretical digest of all known proteins from a database, also called database mining. Gel electrophoresis, trypsin digestions, and liquid chromatography can be time consuming and may negate or avoid the advantage in speed of MALDI-MS. Therefore, work has been done to chemically modify MALDI probe surfaces such that these common processing steps can be carried out directly on the probe.

The most elementary of these modifications is in the method called self-assembled monolayer mass spectrometry (SAMS) or solid-phase extraction (SPE)/MALDI. Molecules with thiol end groups will covalently bind to a gold probe via a sulfur-gold linkage. This technology appeared long ago as a method for electrochemists

to modify the surfaces of gold electrodes.[8] One of the initial applications to MALDI-MS was using alkane thiols to create a hydrophobic surface, for which proteins will have a relatively high affinity (Fig. 1.1.a). This attraction of protein to the hydrophobic surface allowed for samples to be rinsed, removing salts and detergents that are commonly found in biological samples.[2] Alkane thiols with ionic head groups such as an alcohol or carboxylic acid were also prepared to create different surface characteristics (Fig. 1.1.b,c). Similar success in immobilizing proteins was found with these ionic surfaces.[3] Spectra of insulin from a saturated sodium acetate solution were generated after washing with water on strong cation-pairing and strong anion-pairing surfaces. Further success was achieved by generating spectra from solutions containing Triton X-100 and 8M urea. Additional tests were able to verify that during washing, sample was not migrating from one spot to another.[3]

One limitation of SAMS is that their binding capacity is limited by the two-dimensional surface area of the target. A further consideration should be made that the laser is often limited to firing at spots in a one-dimensional line across the target. Any sample captured on the surface outside of the laser path will never be analyzed. To increase the binding capacity over that of the monolayers, high molecular weight poly-lysine was attached to the surface. The large number of ionic sites on this polymer surface allowed the capture of analyte in 3-dimensions, leaving 2-demensions worth of sample along the laser path.[4] To accomplish this dithiobis(succinimidyl propionate), DSP, was used as a linker for the primary amine containing poly-lysine to the gold surface of a probe (Fig. 1.2) DSP chemistry first became popular with electrochemists as a tool for modifying gold electrodes.[8] The DSP binds to the gold probe via a sulfur-gold bond similar to the self-

assembled monolayers. The bound DSP, however, has a carbonile that is susceptible to nucleophilic attack causing an addition/elimination reaction. Primary amines will readily attack and bind.

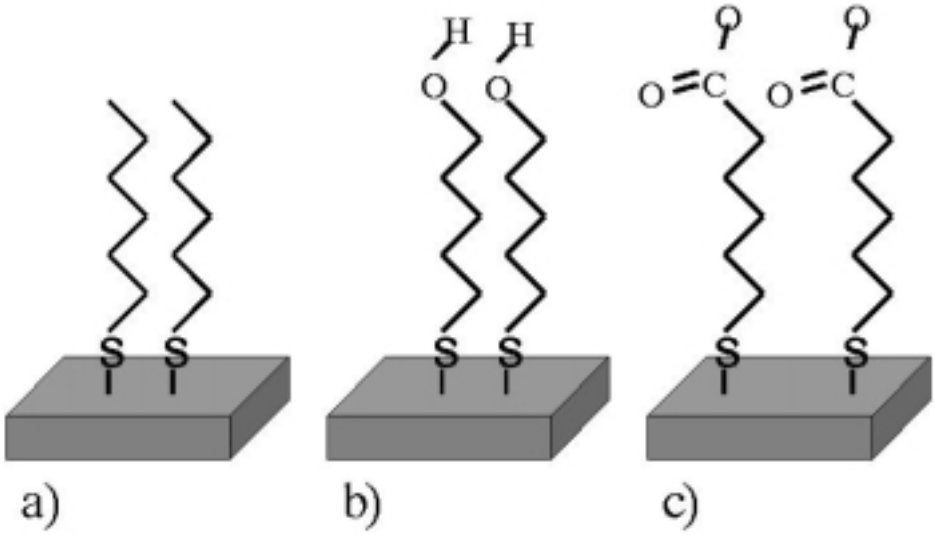

Figure 1.1 Three types of self-assembled monolayers are displayed a) methyl, b) alcohol, and c) carboxylic acid terminated.

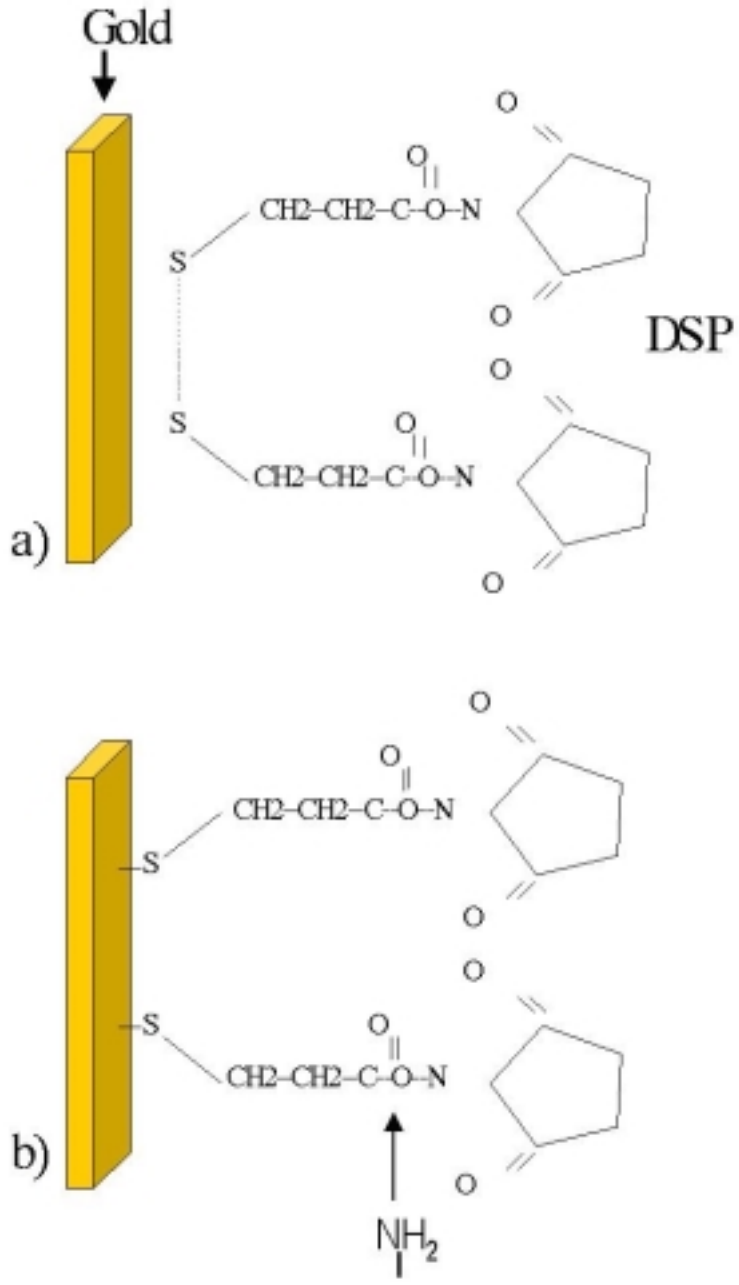

Figure 1.2 The immobilization chemistry for DSP to a gold surface is shown. DSP contains a di-sulfide bond a) that will break and bind to the gold surface b). The carbonyl b) is susceptible to nucleophilic attack by primary amines resulting in an addition-elimination reaction attaching the amine-containing compound to the surface.

Therefore, most primary amine containing molecules, including proteins, may be covalently bound to a gold surface through DSP. The versatility of DSP allows for a variety of complex surfaces to be created on the MALDI probe. In the initial study, the poly-lysine surface was able to increase the binding capacity significantly, more than 100-fold.[4] Additional results indicate the poly-lysine surface is capable of immobilizing >60% of a protein from a highly contaminated solution.[4]

DSP has also been used to immobilize one of an antigen-antibody pair. Primary amines are common in proteins and they will, therefore, bind to a prepared DSP surface. In work done by Brockman, an antibody was bound to a DSP prepared MALDI probe and was capable of capturing its corresponding antigen from a mixture of proteins.[5] This method, termed probe affinity mass spectrometry (PAMS), showed great potential for an antigen immobilized probe to be able to capture its specific antibody out of a mixture of proteins or a complex biological extract. Spectra were generated of a heterogeneous mixture of biotinylated insulin using a conventional MALDI probe and a probe with immobilized anti-biotin.[5] The non-biotinylated species was not retained on the PAMS surface after washing and the tri-biotinylated species is only observed on the PAMS surface.[5] This demonstrates the merger of the high specificity of immobilized affinity chromatography with the high sensitivity of MALDI-MS.

Similarly, biomolecular interaction analysis (BIA) has exploited the immobilization of interactive proteins. Intrinsic Bioprobes, Inc. has developed a chemistry for immobilizing antigens onto BIAchips that are used to capture specific proteins. A high molecular weight dextran is used as a linking agent to achieve increased binding capacity. Using this technology Biacore has developed an instrumentation using BIAchips to capture

protein partners from a flowing system and detect them by surface plasmon resonance. BIA/MS, therefore, required an elution process to release the isolated analyte from the chip for MALDI analysis. Recently, BIAchips have been constructed that can interface into a MALDI instrument so spectra can be taken without eluting.[9-12] Nedelkov, et al,[9] have shown that some BIAchip surface chemistries are more compatible with MALDI analysis than others. For example, streptavidin prepared on a BIAchip will readily dissociate during the MALDI experiment and will be present in the mass spectra. In addition to considering a stable surface for MALDI analysis the geometry of the BIAchips will have to be altered to fit each specific MALDI mass spectrometer.[9]

The specific and non-specific binding of proteins by modified MALDI probes have been helpful in comparing the levels of protein expression between cells under differing conditions.[13-15] Differential protein expression has been a popular and effective method for the discovery of biomarkers. When an organism is stressed, an immune response usually involves a characteristic increase or decrease in the expression of a protein or set of proteins. If a characteristic pattern can be discovered between protein levels in healthy and sick cells, then those key proteins that change significantly can be used as biomarkers for that disease. Therefore, if a protein profile from a patient matches a known characteristic pattern for a disease this technique may serve as a method for early detection of illnesses. This technique may also be used for therapeutics if the identification of key proteins can aid in drug development.[13-15]

2-dimensional gel electrophoresis has been used almost exclusively to perform differential protein experiments. 2-dimensional gels will separate a protein mixture by molecular weight and isoelectric point (pI). Ideally, individual proteins are spread about the

gel to unique locations. For the differential protein profiling experiment, lysates from healthy and sick cells are run on two separate gels and the spots are compared.[16-18] Any spots (proteins) that appear, disappear, or change dramatically in intensity are excised and analyzed by various techniques including MALDI-MS. The gel process is an effective technique but does have some limitations. In addition to the fact that proteins often migrate together, and less abundant proteins are not seen, 2D-gel electrophoresis is time consuming and not conducive to rapid analysis.[13]

If a cell lysate is used directly for MALDI analysis, only a hand full of the most abundant proteins would be seen and the rest would fall into the background. The separation of proteins is necessary for a comprehensive MALDI analysis. Ciphergen Biosystems, Inc. has created an array of surfaces on their probes, called ProteinChips, for this purpose. In the technique called surface-enhanced laser desorption/ionization mass spectrometry (SELDI-MS) each surface is designed to capture a different subset of the total protein extract. Only the portion of proteins that has an affinity for that surface will be preferentially immobilized and detected. By decreasing the population of proteins bound to each chip, the mass spectra are less complex and protein profiles can be generated. Now, through the use of the ProteinChip arrays, protein profiles can be immediately generated from healthy and sick cells, and the spectra can be directly compared.[13-15] In work published by Paweletz, et al, the disease progression of a tumor was detected by monitoring the changes in protein levels.[13] The full scope of the ProteinChip array was demonstrated when comparing the protein profiles of typical strains of mice to the MRL strain that has shown high degrees of tissue regeneration. Proteins that may play key roles in enhancing tissue regeneration were detected.[14]

At this point the protein spectrum can be used as a diagnostic tool, but for drug development, it becomes necessary to know the identity of the key proteins from the differential protein expression experiments. Typically, to identify a protein by mass spectrometry an enzymatic digestion must be performed.[6, 7] After proteolytic digestion, a peptide spectrum is generated, and the protein is identified by protein database mining as described above.

Some digests can be performed directly on MALDI probes, and realize a dramatic decrease in digestion time due to the small sample volume, approximately 1 μL.[19-24] However, some enzymes (including trypsin) have difficulty digesting immobilized proteins. A technique devised by Randy Nelson with Intrinsic Bioprobes, Inc. covalently binds trypsin to a gold probe through DSP. Immobilizing the enzyme prevents the autolysis of trypsin and allows for a higher enzyme-protein ratio for faster digestions.[25] Although the methodology is effective for digesting a protein on-probe, to take advantage of any of the previously discussed techniques for protein isolation, the digestion must take place after the initial capture and rules out the use of an enzyme (trypsin) coated probe. Two on-probe digestion techniques were developed and are described in chapters 2 & 3 demonstrating effective methods for digesting proteins immobilized on any MALDI probe.

Protein Conformational Studies by Hydrogen/Deuterium Exchange MS:

As the known databank of protein sequences continues to grow, there will be an ever-increasing demand for rapid protein conformational studies. The primary structure (amino acid sequence) of any protein is unique, distinguishing it from other proteins much like the letters of the alphabet can be arranged to form unique words. However, it is the logical arrangement of these words into a coherent sentence that gives the letters

meaning or function. Similarly, it is the secondary, tertiary, and quaternary structures of proteins that give them function. After protein creation there are a number of steps required before it is a useful tool for the organism and released to its duty. Proteins are often modified by phosphorylation or glycosylation, and these various forms of post-translational modifications (PTM) have been shown to be crucial for proper activity.[26] After modifications occur, chaperone proteins often aid in folding the protein into its proper functional conformation.[26] Di-sulfide bonds are typically formed to lock the protein in its folded state. Finally, many proteins are only functionally active when associated with other proteins. Hemoglobin, for example, is comprised of four associated sub-units.[26]

In addition to understanding the 3-dimensional conformation of a protein, it is of great importance to learn how proteins interact with other biological molecules. Protein interactions are fundamental in most biological pathways. For example, cell recognition is often a result of an interaction between exterior cell bound proteins and carbohydrates.[27-31] Plant pathogenesis is an extreme example of protein-carbohydrate interaction. In order to gain entry into the plant cell, the pathogen must first breakthrough the cell wall, which consists of various carbohydrates.[30] Understanding the pathogenesis of plants cells has significant agricultural and industrial implications.[27,32-35] Therefore it is necessary to develop quick and effective methodologies for studying protein interactions as well as the 3-dimensional conformations of proteins. X-ray crystallography has achieved great success in determining the conformations of proteins. Reaching ~1.5Å resolution, crystallography has imaged thousands of proteins.[36-38] Solubility, flexibility, and homogeneity within the protein sample are all factors that

affect uniform crystal formation. Also, if the protein is altered in anyway while the crystal is forming, the image may give a misleading description of the protein as it is in solution. To image a protein-protein system, the proteins must co-crystallize, making the task much more difficult. The prospect of imaging a protein-carbohydrate pair is even less favorable. Carbohydrates are much more flexible and often have very different solubilities than proteins. Limited to small oligosaccharides, relatively few protein-carbohydrate complexes have been imaged, emphasizing the need for a complementary technique.[36-38]

Nuclear magnetic resonance (NMR) has also found a significant role in characterizing proteins. When coupled with isotopic labeling, NMR can map the accessibility of amino acids of a protein to the solvent, which is related to the protein's folding or binding. Amide hydrogens along the protein backbone are labile and exchange with protons or deuterons in the solvent. Often for NMR analysis, proteins are exchanged with deuterium and then placed in H_2O while being monitored by H^1-NMR. Blind to deuterium, the instrument detects only hydrogen as it slowly becomes incorporated back into the protein.[39] The rate at which an amino acid incorporates deuterium is related to its position within the protein structure. For example, exterior amino acids will exchange protons with the solvent very quickly, where as interior amino acids can have exchange rates as long as a month or more.[39] Despite the obvious success of NMR, there are limitations. Because of the complex nature of the NMR analysis, studies are typically limited to small proteins or peptides, < 10,000Da, and require a significant amount of time. Analysis typically requires milli-molar concentrations, much more than is available for many biological samples.[38]

Amide hydrogen/deuterium exchange-MS (amide exchange-MS) has been developing over the last several years into a viable complementary technique to crystallography and NMR.[35,40,41] The exchange of amide hydrogen/deuterium as described above for NMR experiments is easily monitored by mass spectrometry, typically electrospray ionization-MS (ESI-MS). ESI is a liquid phase ionization technique often coupled to an LC. The liquid analyte is passed through a narrow metal needle which has a large voltage applied. The most widely excepted mechanism for ionization by ESI is that the droplets take on a charge as they pass through the needle and as the droplets desolvate the charge density increases until a coulombic explosion releases the molecular ions from the droplet (Fig. 1.3).[42] Because the exchange between hydrogen and deuterium affects the mass of the protein, the exchange process can be monitored in real time by ESI-MS.

The first amide exchange-MS experiments involved exchanging a protein with deuterium and monitoring the incorporation over time. Protein folding studies are performed by comparing the protein incorporation in various native, unfolded, or denatured states. Eventually, the effect of parameters like pH and temperature on protein folding could be measured.[35] Protein interaction studies followed subsequently. As ligands or substrates will protect regions of the protein from the solvent, they will cause a noticeable change to the amount of deuterium incorporation. Figure 1.4 illustrates how a bound ligand might block amino acids from exchange with the solvent. Several protein-protein interaction studies have been reported.[38,43-45] Reported here in chapters 4 and 5 is, to our knowledge, the first adaptation of amide exchange-MS to the study of protein-carbohydrate systems. Requiring as little as pico-moles of protein, amide exchange-MS

overcomes many of the limitations encountered with protein-carbohydrate systems by other techniques, and is demonstrated to be a valuable complement to X-ray crystallography and NMR spectroscopy.

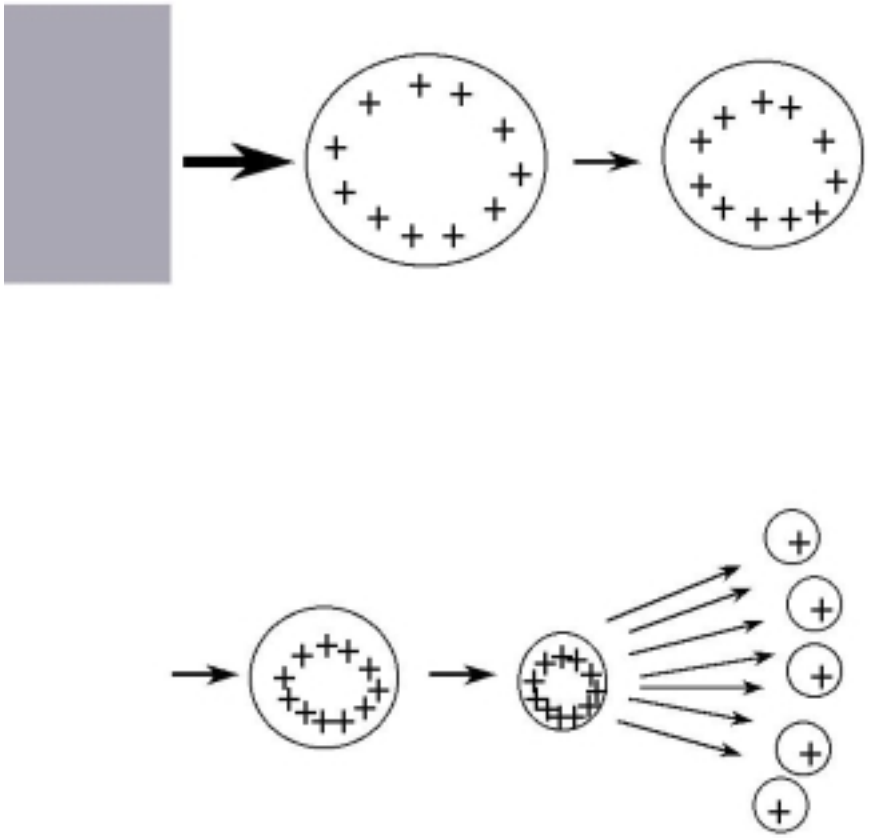

Figure 1.3 Illustration of the electrospray ionization process. As the charged droplets desolvated, the charge density increased until there is a coulombic explosion to release the charged molecules

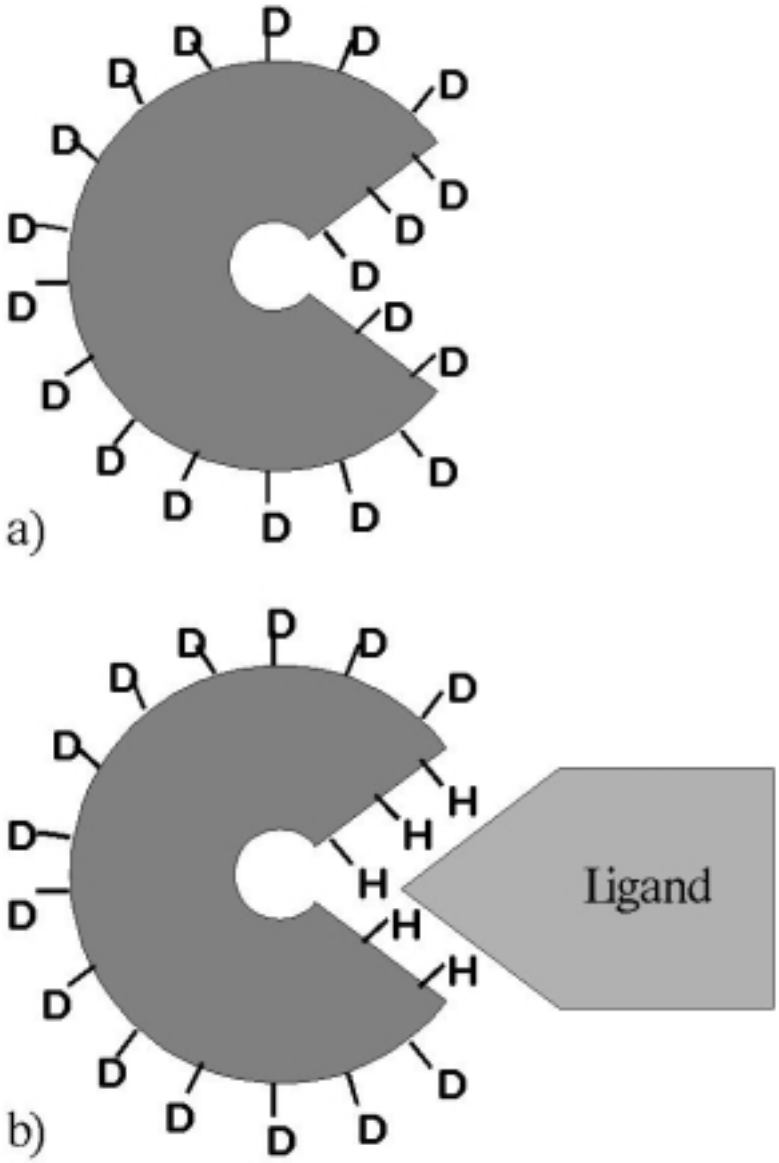

Figure 1.4 Illustration of the effect of a ligand on deuterium incorporation. Deuterium in incorporated a) over the entire exterior of the protein, but with the ligand bound b) deuterium is no longer incorporated into the region of the binding.

References

1. M. Vestal, in *Selected topics in Mass Spectrometry in the Biomolecular Sciences*, R. M. Caprioli, A. Malorni, and G. Sindona (Eds.), Kluwer Academic Publishers, Boston p. 239 (1997).

2. A. H. Brockman, B. S. Dodd, R. Orlando, *Anal. Chem*. **69**, 4716 (1997).

3. M. E. Warren, A. H. Brockman, R. Orlando, *Anal. Chem*. **70**, 3757 (1998).

4. L. Zhang, R. Orlando, *Anal. Chem*. **71**, 4753 (1999).

5. A. Brockman, R. Orlando, *Rapid Commun. Mass Spectrom*. **10**, 1688 (1996).

6. V. Egelhofer and K. Bussow, *Anal. Chem*. **72**, 2741 (2000).

7. M. A. Winkler, J. Uher, *Anal. Chem*. **71**, 3416 (1999).

8. E. Y. Katz, *J. Electroanal. Chem*. **291**, 257 (1990).

9. D. Nedelkov, R. W. Nelson, *J. Mol. Recog*. **13**, 140 (2000).

10. R. W. Nelson, D. Nedelkov, *Anal. Chem*. **72**, 404 (2000).

11. R. W. Nelson, J. R. Krone, *Anal. Chem*. **67**, 1153 (1995).

12. J. R. Krone, R. W. Nelson, D. Dogruel, P. Williams, and R. Granzow, *Anal. Biochem*. **244**, 124 (1997).

13. C. P. Paweletz, J. W. Gillespie, D. K. Ornstein, N. L. Simone, M. R. Brown, K. A. Cole, Q. H. Wang, J. Huang, N. Hu, T. T. Yip, W. E. Rich, E. C. Kohn, W. M. Linehan, T. Weber, P. Taylor, M. R. Emmert-Buck, L. A. Liotta, and E. F. Petricoin, III, *Drug Develop. Res*. **49**, 34 (2000).

14. X. Li, S. Mohan, W. Gu, N. Miyakoshi, and D. J. Bayling, *Biochim. Biophys. Acta*. **1524**, 102 (2000).

15. E. A. Dalmasso, *International Technology Laboratory*. **24**, (2000).

16. M. R. Emmert-Buck, J. W. Gillespie, C. P. Paweletz, D. K. Ornstein, V. Berserk, E. Appella, Q. H. Wang, J. Huang, N. Hu, P. Taylor, and E. F. Petricoin, III, *Mol. Carcinog*. **27**, 158 (2000).

17. R. E. Banks, M. J. Dunn, M. A. Forbes, A. Stanley, D. Pappin, T. Naven, M. Gough, P. Harnden, and P. J. Selby, *Electrophoresis*. **20**, 689 (1999).

18. S. M. Leung, *Ciphergen Biomarker Discovery*, **20**, (1999).

19. Z. Xuegong and C. Borchers, *Biochem*. **39**, 11194 (2000).

20. B. Kuster, and M. Mann, *Anal. Chem*. **71**, 1431 (1999).

21. J. E. Coligan, and D. W. Speicher, in *Current Protocols in Protein Science*. J. E. Coligan, B. M. Dunn, H. L. Ploegh, D. W. Speicher, and P. T. Wingfield (Eds.), John Wiley & Sons, New York (1995).

22. R. D. Holland, and C. R. Duffy, *Anal. Chem*. **71**, 3226 (1999).

23. H. Geyer, S. Schmitt, M. Wuhrer, and R. Geyer, *Anal. Chem*. **71**, 476 (1999).

24. Y. Yang, and R. Orlando, *Anal. Chem*. **68**, 570 (1996).

25. D. Dogruel, P. Williams, and R. W. Nelson, *Anal. Chem*. **67**, 4343 (1995).

26. C. Starr and R. Taggart (Eds.), in *Biology: the Unity and diversity of Life*. Wadsworth Publishing Company, Belmont CA (1998).

27. R. M. Cooper, in *Biochemical Plant Pathology*. J. A. Callow (Ed.), John Wiley & Sons, New York (1995).

28. N. C. Carpita, and D. M. Gibeaut, *Plant J.* **3**, 1 (1993).

29. T. M. Jones, A. J. Anderson, and P. Albersheim, *Physiol. Plant Pathol.* **2**, 153 (1972).

30. M. G. Hahn, P. Bucheli, F. Cervone, S. H. Doares, R. A. O'Neill, A. Darvill, and P. Albersheim, in *Plant-Microbe Interactions. Molecular and Genetic Perspectives*. T. Kosuge and E. W. Nester (Eds.), McGraw Hill Publishing Co., (1989).

31. C. Grassin, and P. Fauquemberque, in *Pectins and Pectinases*. J. Visser and A. G. J. Voragen (Eds.), Elsevier Science B.V., (1996).

32. H. P. Heldt-Hansen, L. V. Kofod, G. Budolfsen, P. M. Nielsen, S. Hüttel, and T. Bladt, in *Pectins and Pectinases*. J. Visser and A. G. J. Voragen (Eds.), Elsevier Science B.V., (1996).

33. C. W. Bergmann, B. Cook, A. G. Darvill, P. Albersheim, D. Bellincampi, and C. Caprari, in *Pectins and Pectinases*. J. Visser and A. G. J. Voragen (Eds.), Elsevier Science B.V., (1995).

34. B. R. Thakur, R. K. Singh, and A. K. Handa, *Crit. Rev. Food Sci. and Nutri*. **37**, 47 (1997).

35. V. Katta, and B. T. Chait, *J. Am. Chem. Soc*. **115**, 6317 (1993).

36. W. I. Weis, K. Drickamer, and W. A. Hendrickson, *Nature*. **360**, (1992).

37. E. A. Merritt, S. Sarfaty, F. van den Akker, C. L'Hoir, J. A. Martial, and W. G. J. Hol, *Protein Sci.* **3**, 166 (1994).

38. H. Ehring, *Anal. Biochem*. **267**, 252 (1999).

39. J. S. Milne, L. Mayne, H. Roder, A. J. Wand, and S. W. Englander, *Protein Sci*. **7**, 739 (1998).

40. Z. Q. Zhang, and D. L. Smith, *Protein Sci*. **2**, 522 (1993).

41. Y. Q. Liu, and D. L. Smith, *J. Am. Soc. Mass Spectom*. **5**, 19 (1994).

42. S. J. Gaskell, M. S. Bolgar, I. Riba, S. G. Summerfield, in *Selected Topics in Mass Spectrometry in the Biomelecular Sciences*, R. M. Caprioli, A. Malorni, and G. Sindona (Eds.), Kluwer Academic Publishers, Boston (1997).

43. J. B. Smith, Y. Q. Liu, and D. L. Smith, *Exp. Eye Res*. **63**, 125 (1996).

44. F. Wang, J. S. Blanchard, and X. J. Tang, *Biochem*. **36**, 3755 (1997).

45. S. Akashi, and K. Takio, *Protein Sci*. **9**, 2497 (2000).

CHAPTER 2

IN MATRIX/ON-PROBE (IMOP) PROTEIN DIGESTION WITH PEPSIN[1]

[1]King, D. and Orlando, R. To be submitted to *Rapid Communications in Mass Spectrometry*.

Abstract

It has long been accepted that after generating a MALDI spectrum, the remaining analyte on the probe is unsalvageable, and therefore wasted. We propose a method for performing an additional enzymatic digestion and generating a subsequent peptide spectrum after an initial protein spectrum has been collected. The method explores the activity of the proteolytic enzyme, pepsin, in the presence of several common matrices. The resulting protocol allows for further analysis to be performed from left over analyte that would typically be disposed of.

Introduction

The standard procedures for mass spectrometric analysis of biomolecules are well established.[1] Typical digestion procedures require pH specific enzyme buffers and hours of incubation. Some form of chromatography is often used to clean up the sample before analysis by mass spectrometry can begin.[1, 2] One method to speed up the analysis is to perform the enzymatic digestions directly on the MALDI target. Digests performed on MALDI probes offer some advantages over the traditional in-solution digests. For example, only small amounts of sample are needed and essentially none is wasted. But, the technique focuses on increasing the speed of the analysis process. The digest can be completed very quickly because of the small sample volume, ~1μl, and when the digest has progressed sufficiently, matrix is added and the probe is analyzed.[3, 4, 5, 6, 7, 8]

Despite the apparent speed that on-probe digests offer, there are drawbacks that may have discouraged its widespread use. One drawback is its incompatibility with common sample purification processes. Most proteolytic enzymes are only active in very specific environments. Therefore, pH buffers are used despite that they have an inherently negative effect on MALDI signal. As a result, some clean-up techniques are employed, typically liquid chromatography, before a mass spectrum is generated. The chemical modification of MALDI probe surfaces has been shown to aid the cleanup of analytes on probes. For example, a poly-lysine surface will bind proteins or peptides through ionic interactions and allow the target to be cleaned by rinsing with water.[9, 10] Nevertheless, it would be advantageous for quicker protein identification and characterization if buffers were not needed, avoiding poorer signal intensities caused by the salt buffers or the need for alternative cleanup processes.

Another significant drawback to on-probe digests is that typically only one digest and one spectrum can be generated. It is thought that once the matrix is added no additional chemistry can be done, particularly because the pH of typical MALDI matrices are too acidic for most proteolytic enzymes. On the other hand, a digest performed in solution allows for aliquots to be removed for enzymatic treatment and many spectra can be generated. Most analyses start with the removal of one aliquot and acquiring a mass spectrum of the original analyte. Subsequent aliquots can be removed for cleaning, analysis by various instruments and techniques, as well as performing further digests if so desired.[1] Multiple digestions are often needed to properly characterize proteins, especially post-translationally modified proteins. If a protein contains a post-translational modification, one or more proteolytic peptides will not match database mass values unless the modification is first removed. For example, when studying glyco-proteins the first step is usually to remove the glycosylation using PNGase-F or Endo-H to generate an accurate protein mass spectrum. The effective application of multiple digestion analysis emphasizes the need for generating multiple spectra.[11]

We propose the effective use of the proteolytic enzyme, pepsin, in the presence of a typical MALDI matrix. MALDI matrices usually contain approximately 0.1% TFA yielding a pH of around 2. Pepsin is optimized at this pH, eliminating the need for additional buffers and subsequent cleanup steps. Pepsin is well known as a very quick and efficient enzyme. The proteins do not typically need to be denatured, and sufficient digests can be completed in only minutes.[12, 13] The drawback with using pepsin is that it is less specific, cleaving at most hydrophobic residues, than the more commonly used proteolytic enzymes like trypsin or chymotrypsin. However, the liability from pepsin's lack in specificity is far outweighed by its robustness.

If pepsin is active in the presence of matrix, it could be added to a target following the acquisition of a protein spectrum. After the digest has proceeded, a peptide spectrum could then be generated. Also, the degree of digestion could be controlled by the lifetime of the solvation droplet, the digest will continue until the target is dried. The digest would likely resume upon the addition of another water droplet to the target. This would yield a simple tool for turning the reaction on and off. The most significant implication is that a pepsin digest could follow any other on-probe procedures including other on-probe digests.

Experimental Section

Methods:

The calibration standards angiotensin 2 and insulin were purchased from Sigma (St. Louis, MO, USA) The proteins cytochrome C and myoglobin, purchased from Sigma, and BSA, purchased from Fluka (Milwaukee, WI, USA), were solvated in deionized water to a concentration of 0.1mg/ml and frozen until used. A purified recombinant protein, *endo*polygalacturonase 1 (EPG-1), overexpressed in *A. niger* was acquired from Dr. Jaap Visser (Wageningen Agricultural University, The Netherlands) and also frozen until its use. The matrices used in these experiments, 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) from Sigma, α-cyano-4-hydroxy-cinnamic acid (α-cyano) and 2,5-dihydroxybenzoic acid (DHB) from Aldrich (Milwaukee, WI, USA), were each prepared as saturated solutions of the matrix in 70% acetonitrile from Fisher Scientific (Fair Lawn, NJ, USA), 30% water and 0.1% trifluoroacetic acid (TFA) from Sigma. Pepsin, purchased from Sigma, was solvated in 10mM HCl to a concentration of 1mg/ml and frozen until used.

On-Target Procedures:

The general procedure by which we performed on-probe pepsin digests began by mixing 0.5μL of matrix and 0.5μL of protein on a MALDI target. After the matrix/protein mixture had dried, a MALDI spectrum of the protein was generated. Then, 0.5μL of pepsin (1μL/μg) was added to the target. Any time a mass spectrum was desired, the target was dried and inserted into the mass spectrometer. The digests were resumed by adding subsequent 0.5μL aliquots of water to the target.

MALDI-TOF Measurements

Some MALDI-MS experiments were performed on a KRATOS Kompact SEQ time-of-flight mass spectrometer. The samples were prepared on a stainless steel probe provided by Kratos. Other experiments were performed on a Hewlett Packard (HP) G2025A time-of-flight mass spectrometer. These samples were prepared on a gold probe provided with the instrument. Both instruments use a nitrogen laser, emitting at 337nm, to desorb/ionize the analyte from the target. The instruments were operated at an accelerating voltage of 28kV and an extractor voltage of 7kV. The mass range was calibrated using an external standard of angiotensin 2, insulin, and cytochrome C.

Results and Discussion

Enzymatic digestions allow for the identification and characterization of biopolymers like proteins, polynucleotides and carbohydrates. When coupled with matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), these digests can provide a rapid method for identifying proteins by searching protein databases with peptide masses.[14, 15] Also, post-translational modifications such as glycosylation and phosphorylation can be located and characterized by monitoring the changes is mass as these modifications are removed by various enzymes.[16, 17] The routine use of proteolytic

enzymes on MALDI targets has shown potential for being a valuable technique in identifying and characterizing proteins. This technique avoids transferring analyte between containers or instruments resulting in less sample loss and less possibility for contamination. Only the small amount of sample placed on the target is needed and series of experiments could be carried out on one MALDI probe (several targets) and stored for later reference. Endeavoring to further the promise of on-probe analysis, we have explored the activity of the proteolytic enzyme, pepsin, in the presence of common matrices.

Initially, to verify the activity of pepsin in the presence of MALDI matrix, 0.5μL of myoglobin and 0.5μL of DHB were added to two HP MALDI targets and dried. Water (0.5μL) was added to the first target, serving as the control (Fig. 2.1.a), while 0.5μL of pepsin was added to the second target (Fig. 2.1.b). The targets were allowed to air dry, taking approximately 2 minutes, and mass spectra were acquired. After this short period of time, it was clear that pepsin was certainly active in the DHB matrix. Under the same mass spectrometry conditions the protein peak fades away and newer peptide peaks appear. A spectrum of the pepsin solution was acquired to verify that the peptides were not from the enzyme.

A time course study was performed monitoring the digestion of cytochrome C over 10 minutes to better understand parameters related to the lifetime of the solvation droplet. Cytochrome C (0.5μL) and α-cyano (0.5μL) were allowed to dry on a Kratos probe. A protein spectrum was generated and then 0.5μL of pepsin was added. After 2 minutes, the target was dry and another spectrum was acquired. At that time, not much digestion was observed. The probe was removed and 0.5μL of water was added. The probe air dried in

about 2 minutes and was then reinserted into the mass spectrometer for analysis. Then, after 4

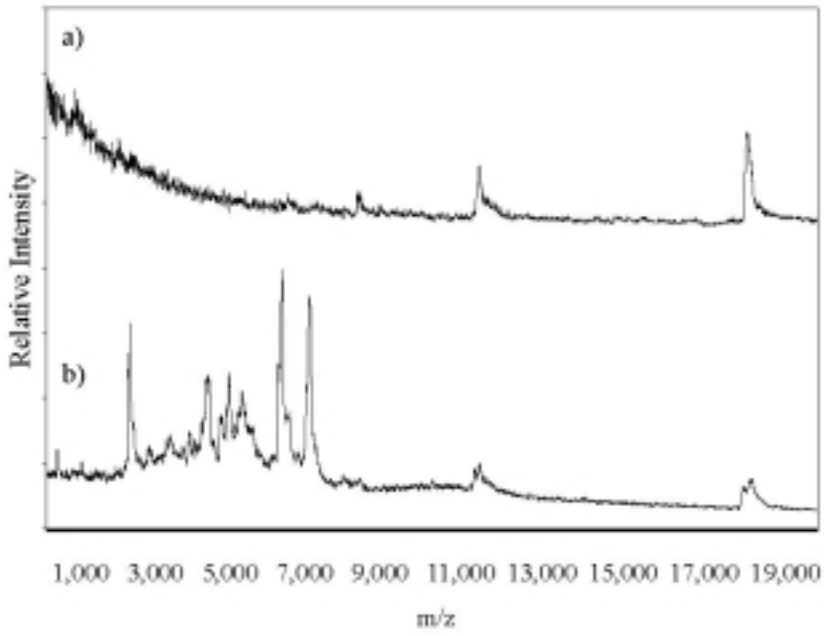

Figure 2.1 2.8 pmoles of myoglobin (a) with DHB and (b) after a 2 minute digestion with pepsin.

minutes of digestion, peptides began to appear. This process was repeated 3 additional times to a final digestion time of 10 minutes. Figure 2.2 shows the shift from protein to high mass peptides to low mass peptides over time. This illustrates how the digest might be controlled by drying or wetting the target. If desired, larger drops of water or a humidity chamber could be used to enable the digestion to progress further with fewer manipulations.

Cytochrome C was digested in α-cyano, DHB, and sinapinic acid, each on different targets of the same Kratos probe, to compare the activity of pepsin in various matrices. Figure 2.3 shows that there is a dramatic difference between matrices. As described earlier the digestions were monitored every 2 minutes. After 2 minutes, before the first addition of water, peptides were only visible in the DHB trial. At 4 minutes, after the first addition of water, peptides began to appear in the α-cyano trial, as well. The digest performed in DHB appears to be quicker than α-cyano, and both are much quicker than sinapinic acid. Therefore, it is concluded that both α-cyano and DHB are fine for this digestion technique, and the selection of matrices should be based on usual considerations like the size of the protein. While α-cyano works well with small proteins, DHB can be used when it is necessary to look at larger proteins like BSA, approximately 67kDa. Using one HP probe and DHB, the digestion of BSA occurred rapidly. Peptides began to appear prior to the addition of any water (Fig. 2.4).

The ability to turn a digest on or off allows for a scientist to tailor the digestion to serve a specific purpose. For example, there are situations when a partial digestion would be preferred over a complete digestion. Peptide masses that match theoretical digest values indicate peptides that do not contain post-translational modifications. In contrast, all post-translationally modified peptides will not match theoretical digest values and are not directly identifiable. Therefore, as peptides are identified possible

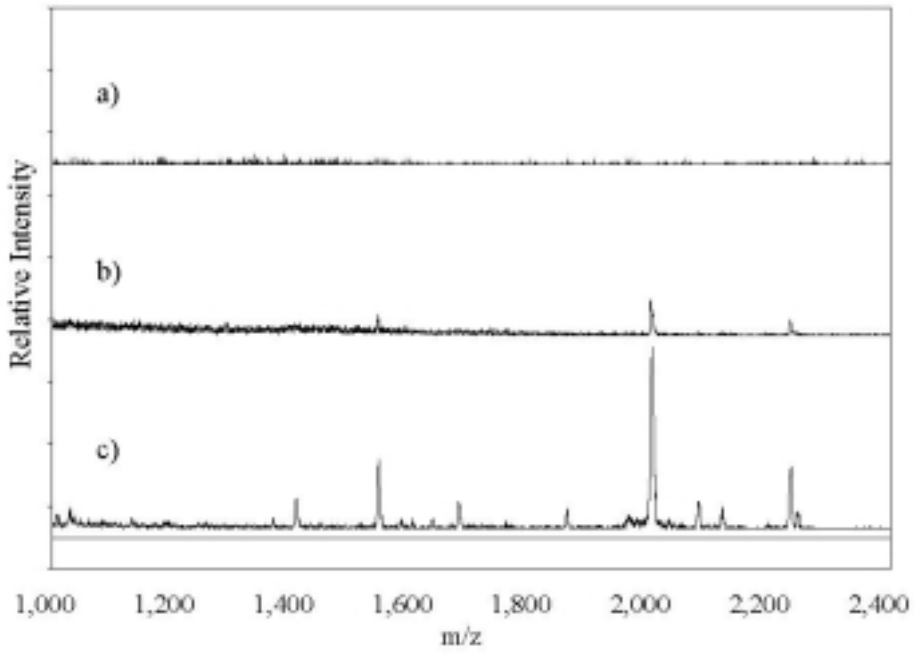

Figure 2.2 4 pmoles of cytochrome C at different digestion times in α-cyano; (a) t = minutes, (b) t = 4 minutes, and (c) t = 8 minutes.

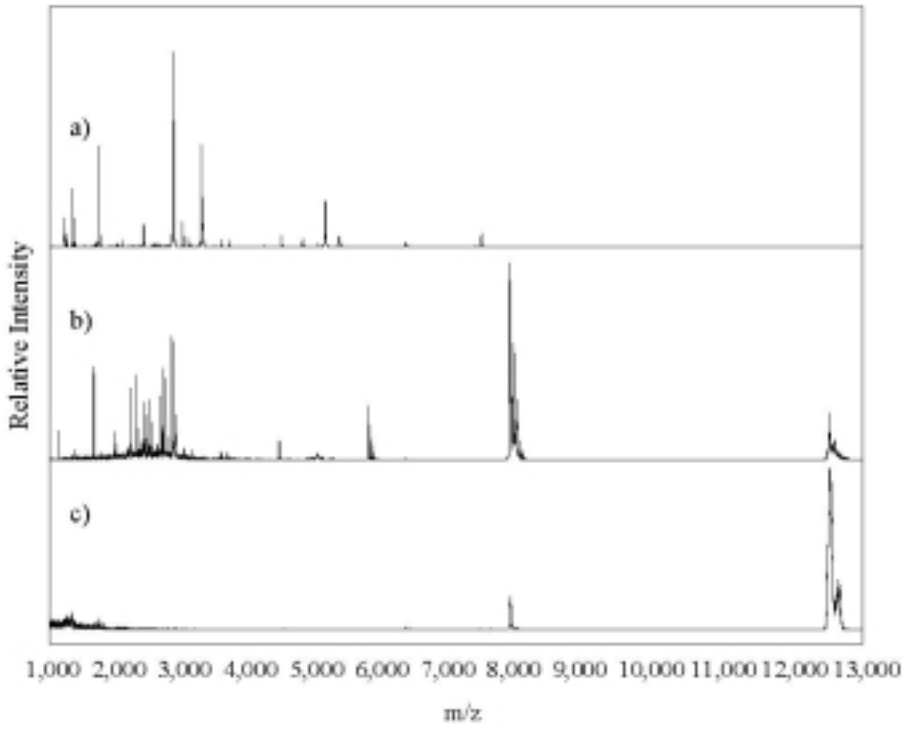

Figure 2.3 4 pmoles of cytochrome C digested by pepsin (a) in DHB for 2 minutes, (b) in α-cyano for 4 minutes, and (c) in sinapinic acid for 4 minutes.

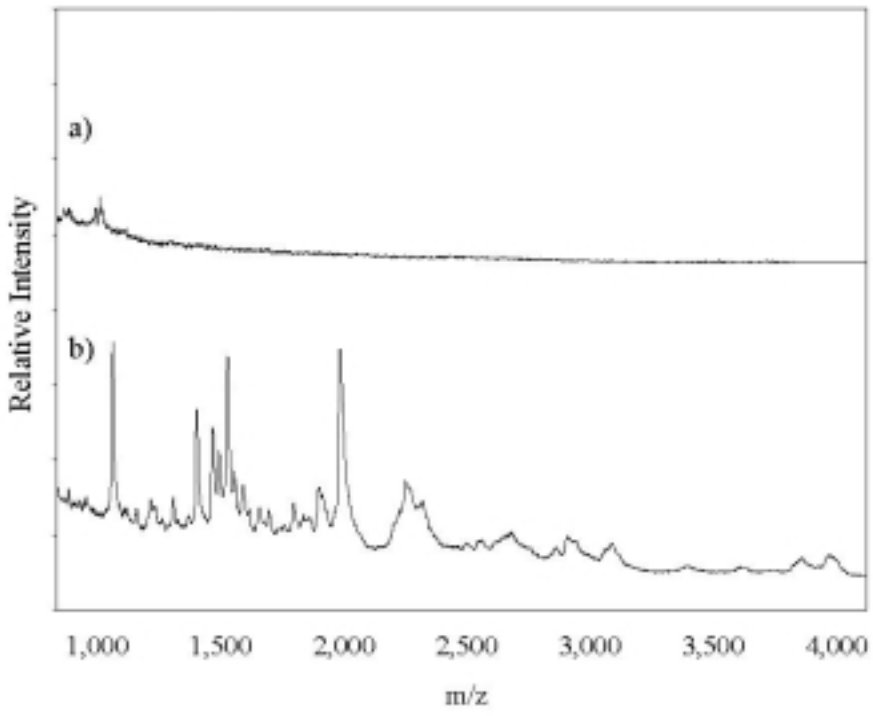

Figure 2.4 750 fmoles of BSA (a) in DHB and (b) after a 2 minute digestion with pepsin.

modification sites are eliminated. In theory, if a modified protein is cleaved in half and one half is identified, all modifications are limited to the second unidentified half of the protein. Under typical digestion conditions, there are too many peptides to search for modification sites in this elimination manner, not to mention that digest spectra will rarely cover 100% of the protein. However, performing a partial digest may limit the number of peptides in such a way that makes locating the modification sites possible. To illustrate and demonstrate this concept, a quick 2 minute digest was performed on EPG-1 on a KRATOS probe using DHB. EPG-1 has two consensus sequences for possible N-linked glycosylation, NXT/S where X is any amino acid but P, at N^{246}VT and N^{298}GS (4). In figure 5, the indicated peaks correspond to theoretical mass values of peptides containing N^{298}. Concluding that N^{298} is therefore not glycosylated, implies that N^{246} must be, which is in agreement with previously published results.[11]

Conclusions

Pepsin is active in the presence of matrix, although the activity varies with the matrix used. This allows for the acquisition of a protein spectrum and then a pepsin digestion spectrum. Once dried, the digestion can be resumed by the addition of water to the target. This procedure can be repeated indefinitely until the digest has progressed as desired. As demonstrated above in locating the EPG-1glycosylation site, there are situations when a partial digestion with larger peptides would be preferred. Another significant implication of pepsin's activity in the presence of matrix is that a pepsin digestion could likely follow any other on-probe procedure, including other on-probe digests or cleanup steps. For example, the possibility of multiple digestions to be performed on the same target allows the removal of post-translational modifications from the protein by an N-glycanase prior to the pepsin digest. Previous work in our lab has

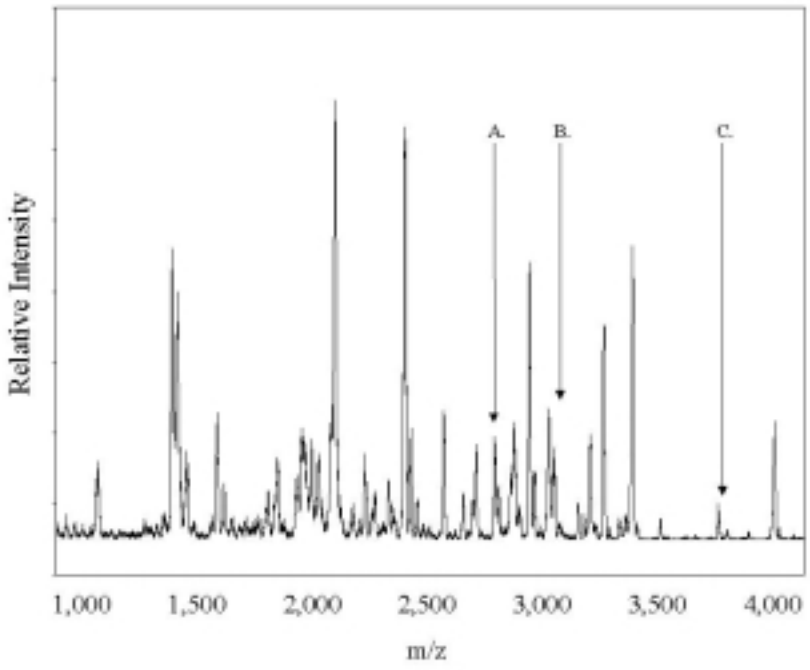

Figure 2.5 1.5 pmoles of EGP-1 digested for 2 minutes in DHB. Peaks with mass/charge values corresponding to peptides that contain N^{298}, indicated with an (*), are as follows; (a) 2515.63 (T^{287}-I^{310}), (b) 2772.86 (S^{284}-I^{310}), and (c) 3553.31 (E^{294}-A^{328}).

demonstrated the wide variety of enzymes that may be used successfully in an on-probe fashion.[4, 11] Samples could also be cleaned or selectively bound using specialty probes like an SPE/MALDI target,[18] analyzed, and then digested with pepsin. Therefore, analyte left on the target after generating a spectrum no longer has to be abandoned. This process is simple and quick, and no cleanup steps are needed. The ability to follow any MALDI analysis with a pepsin digest will provide additional data without any additional material, an ideal situation for the biological chemist for whom analyte is often at a premium.

References

1. J. E. Coligan, and D. W. Speicher, in *Current Protocols in Protein Science*. J. E. Coligan, B. M. Dunn, H. L. Ploegh, D. W. Speicher, and P. T. Wingfield (Eds.), John Wiley & Sons, New York (1995).

2. R. D. Holland, C. R. Duffy, *Anal. Chem*. **71**, 3226 (1999).

3. H. Geyer, S. Schmitt, M. Wuhrer, and R. Geyer, *Anal. Chem*. **71**, 476 (1999).

4. Y.Yang and R. Orlando, *Anal. Chem*. **68**, 570 (1996).

5. B. Kuster, T. J. P. Naven, and D. J. Harvey, *J. Mass Spectrom*. **31**, 1131 (1996).

6. R. W. Nelson, *Mass Spectrom. Rev*. **16**, 353 (1997).

7. R. W. Nelson, D. Dogruel, J. R. Krone, and P. Williams, *Rapid Commun. Mass Spectrom*. **9**, 1380 (1995).

8. D. Dogruel, P. Williams, and R. W. Nelson, *Anal. Chem*. **67**, 4343 (1995).

9. L. Zhang and R. Orlando, *Anal. Chem*. **71**, 4753 (1999).

10. M. E. Warren, A. H. Brockman, and R. Orlando, *Anal. Chem*. **70**, 3757 (1998).

11. J. Colangelo, V. Licon, J. Benen, C. Bergmann, R. Orlando, *Rapid Commun. Mass Spectrom*. **13**, 1448 (1999).

12. G. Marie and L. Serani, *Anal. Chem*. **72**, 5423 (2000).

13. Z. Zhang and D. L. Smith, *Protein Sci*. **2**, 522 1993.

14. V. Egelhofer and K. Bussow, *Anal. Chem*. **72**, 2741 (2000).

15. M. A. Winkler and J. Uher, *Anal. Chem*. **71**, 3416 (1999).

16. Z. Xuegong and C. Borchers, *Biochem*. **39**, 11194 (2000).

17. Bernhard Kuster and Matthias Mann, *Anal. Chem*. **71**, 1431 (1999).

18. M. Merchant and S. R. Weinberger, *Electrophoresis*. **21**, 1164 (2000).

CHAPTER 3

FEASIBILITY STUDY FOR PARTIAL ACID HYDROLYSIS OF PROTEINS PERFORMED ON-PROBE FOR MALDI-MS ANALYSIS

Abstract

Recent successes in digesting proteins by partial vapor phase acid hydrolysis has introduced the possibility of performing such digestions on protein samples that are dried on a MALDI target. The results reported here demonstrate the most valuable MALDI spectra for proteomics that can be generated, a sequence ladder. Some digestions generated very intense peptide signals while others produced sequence ladders that can be used like MS/MS spectra to positively identify proteins. This feasibility study demonstrates the first vapor phase acid hydrolysis process performed on a protein dried on a MALDI target. This investigation has illuminated great inconsistencies within the acid hydrolysis process and has identified several parameters that will need to be addressed to generate desired spectra. Despite the inconsistent results, the value of the data generated will encourage further work to optimize this procedure.

Introduction

Hydrolysis of proteins has been a useful step in the characterization of peptides/proteins for more than half a century.[1] A hydrolysis mechanism with high specificity can be used to locate and identify amino acids within a peptide or protein. One such method, acid hydrolysis, was reported by Partridge and Davis to cleave only aspartic acid residues. Not until the late 1970's did reports begin to surface of partial acid hydrolysis causing multiple backbone cleavages in addition to aspartic acid.[2]

A recent paper by Johan Gobom, et al,[3] described the use of a vapor phase acid hydrolysis protocol to digest proteins for analysis by mass spectrometry. In this work the hydrolysis procedure is characterized by varying many parameters. Specific cleavage at aspartic acid residues was observed as well as at serine, threonine, and glycine residues. In addition to the specific cleavage locations, processive hydrolysis of the protein backbone generated significant N-terminal and C-terminal ladders.[3] Although, specific cleavage of a protein backbone has always been necessary for identification of proteins by MALDI-MS, the peptide ladders generated by partial acid hydrolysis may prove to be the most significant finding. A peptide sequence ladder generates the same information as an MS/MS spectrum. MS/MS is currently relied upon to positively identify proteins by matching a portion of the amino acid sequence with known sequences in a database. Generating amino acid sequence information with the simplicity and sensitivity of MALDI as compared to ESI-MS/MS would be a significant step forward in the area of proteomics. Another article has demonstrated the effectiveness of vapor phase acid hydrolysis to generate sequence ladders so that O-linked glycosylations may be located on a peptide.[4]

The primary cleavage site for partial vapor phase acid hydrolysis is reported to be at aspartic acid residues.[3] The mechanism described in figure 3.1 demonstrates that the ring intermediate can be formed with the carbonyl on either side of the side chain, forming a 5 or 6 member ring intermediate. This variability results in cleavage on either side of the aspartic acid.[5] Mechanisms for the secondary cleavage sites of serine and threonine have previously been described as shown in figure 3.2. The alcohol within the side chain attacks the adjacent carbonyl carbon resulting in the loss of H_2O. Then by the addition of a first H_2O, the protein backbone is cleaved, and by the addition of a second the previously formed ring is cleaved. The net reaction is the addition of one H_2O.[6]

For some time, our lab has been interested in developing on-probe procedures for faster and more efficient MALDI analysis. One area of focus is developing methods for fast digestion of protein samples directly on a MALDI target. Often, by working with small sample volumes, ~1μL, very fast digestions can be achieved on-probe. The following work is our attempt to adapt the partial vapor phase acid hydrolysis procedure for digestion of protein samples dried on a MALDI target. The vapor phase acid hydrolysis results are compared with liquid phase acid hydrolysis performed by adding acid directly to the dried protein spot.

Materials and Methods

Materials:

The proteins lysozyme, bovine serum albumin, and cytochrome C were purchased from Sigma Chemical (St. Louis, Missouri). Pentafluoropropionic acid (PFPA) from Aldrich Chemical (Milwaukee, Wisconsin) and dithiothreotol (DTT) from Sigma were used to perform the acid hydrolysis. The MALDI matrix, dihydroxybenzoic acid (DHB)

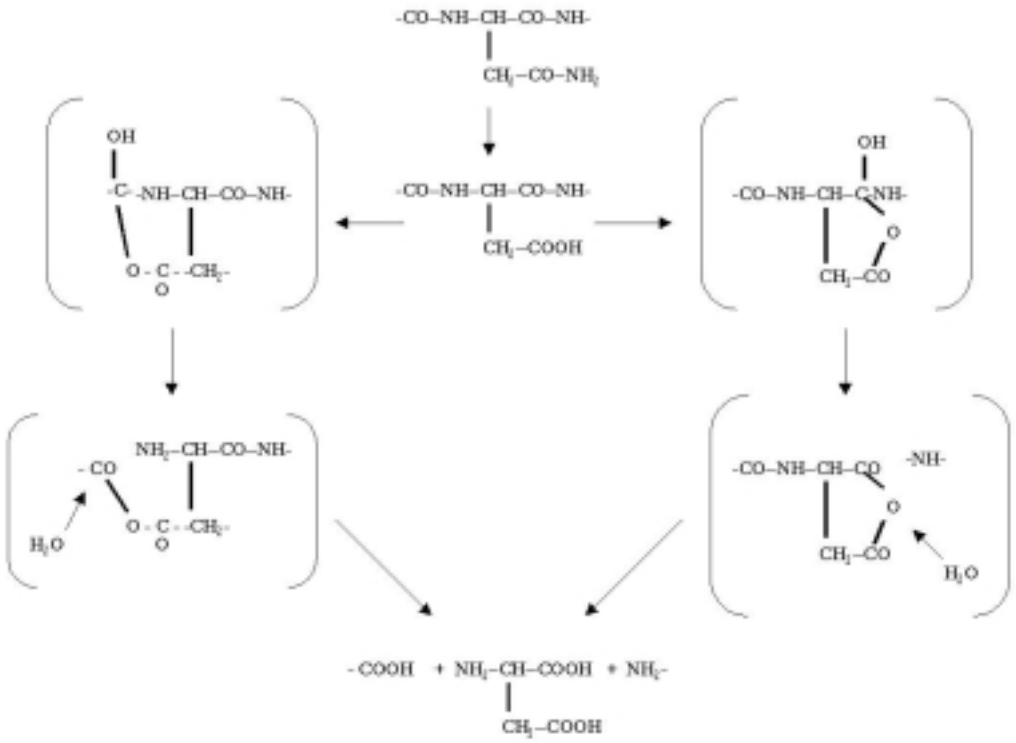

Figure 3.1 Possible mechanisms for N or C-terminal cleavage of aspartic acid.

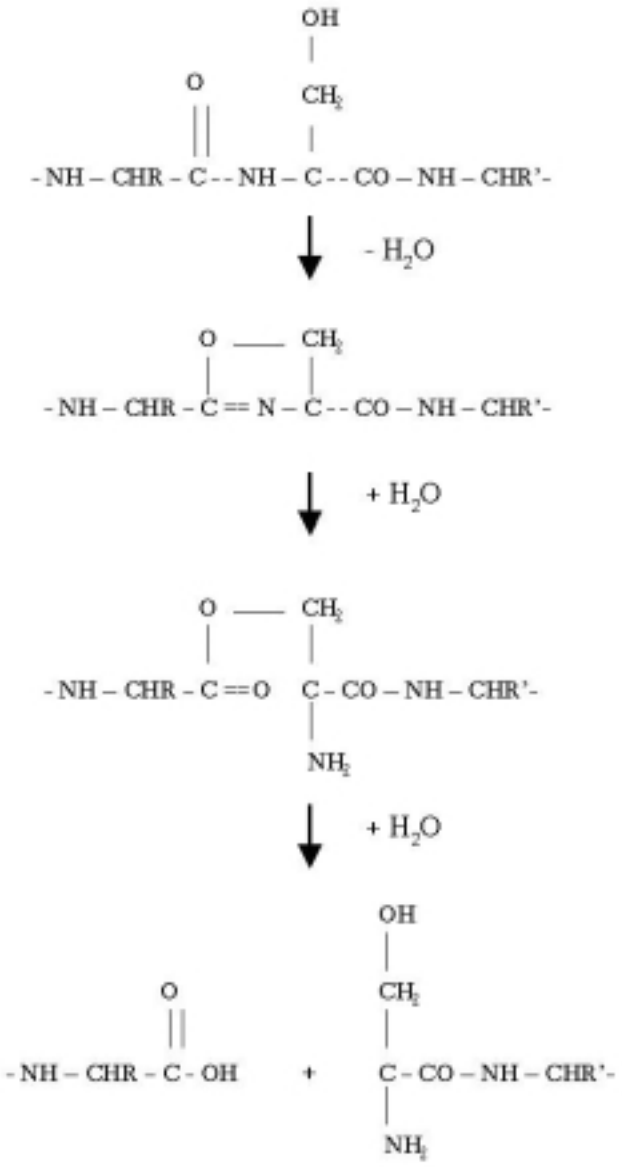

Figure 3.2 Possible mechanism for hydrolysis of serine or threonine.

from Sigma, was prepared by dissolving 1mg of DHB in a 50% acetonitrile (Fisher Scientific, Fair Lawn, New Jersey) solution containing 0.1% trifluoroacetic acid (TFA).

Vapor Phase Acid Hydrolysis:

The initial adaptation of the technique to our on-probe study required the construction of a vacuum chamber that would hold our Kratos MALDI (Manchester, UK) probe. A chamber was constructed out of a test tube fitted with an o-ring sealed cap with a small nozzle to which hoses could be attached (Fig. 3.3). Protein samples are dried on the probe and then placed into the vacuum chamber. 50 μL of 90 % PFPA with 500 μg DTT is pipetted into the bottom of the test tube chamber. Because of the round bottom of the test tube, the probe does not contact the acid solution. The chamber is flushed with He, attached to a vacuum, and placed in a hot water bath at 80°C for 60 minutes. After the incubation period, matrix is added to the targets and analyzed by MALDI-MS.

On-Probe Liquid Phase Acid Hydrolysis:

For on-probe liquid phase acid hydrolysis, a protein sample is dried on a Kratos MALDI target. PFPA is then added to the spot and allowed to digest the sample. The hydrolysis time can be extended by adding aliquots of H_2O to keep the target solvated or by placing the probe into a humidity chamber. Once the target has dried, matrix is added to the target, and the sample can be analyzed by MALDI-MS.

Results and Discussion

Current efforts to employ partial acid hydrolysis to study proteins by mass spectrometry come from modifying previously published work that describes the vapor phase acid hydrolysis of proteins dried in a small vial.[3] The procedure should be capable

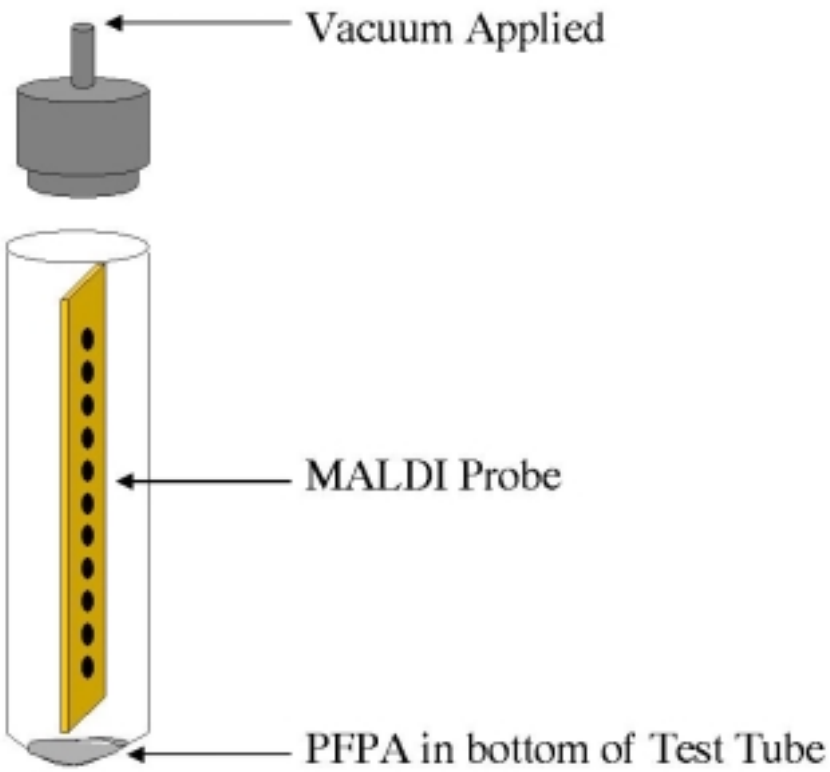

Figure 3.3 The vacuum chamber for the vapor phase acid hydrolysis was constructed from an o-ring sealed test tube. The orifice in the cap allowed for the tube to be evacuated during the hydrolysis process.

of hydrolyzing dried protein samples on any surface, including a MALDI probe. Hydrolysis of proteins on the surface of a MALDI probe could be done in the gas phase or in the liquid phase by the addition of the acid directly to the target. Both approaches were worth investigating, as the gas phase has been shown to work and the liquid phase is likely to be quicker and more robust.

Vapor Phase Acid Hydrolysis

The vapor phase study began by varying the amounts of different proteins to be digested. The results were sporadic and those that were successful yielded weak signals. Because no standard denaturation procedure is being performed on the proteins prior to analysis, each protein may require unique parameters for optimal vapor phase acid hydrolysis. Therefore, varying the temperature and the % of PFPA became necessary for each sample to achieve a reasonable digestion. An example of the variability within the results is shown in figure 3.4. 100 pmol cytochrome C and lysozyme were placed on separate spots on the same probe and dried. The vapor phase acid hydrolysis was carried out at 70°C with a 70% PFPA solution for 30 minutes.

A few very intense peptides were generated in the cytochrome C sample and a very weak sequence ladder was formed in the lysozyme sample. The purpose for the analysis may determine which type of spectrum is more useful. For the purpose of protein identification the sequence ladder will be preferred. Identifying proteins by MALDI-MS in the past has required the generation of a number of specifically cleaved peptides. Trypsin is often preferred because of its high specificity, cleaving the C-terminal side of arginine and lysine residues. The partial acid hydrolysis has been reported and confirmed in our present study to hydrolyze

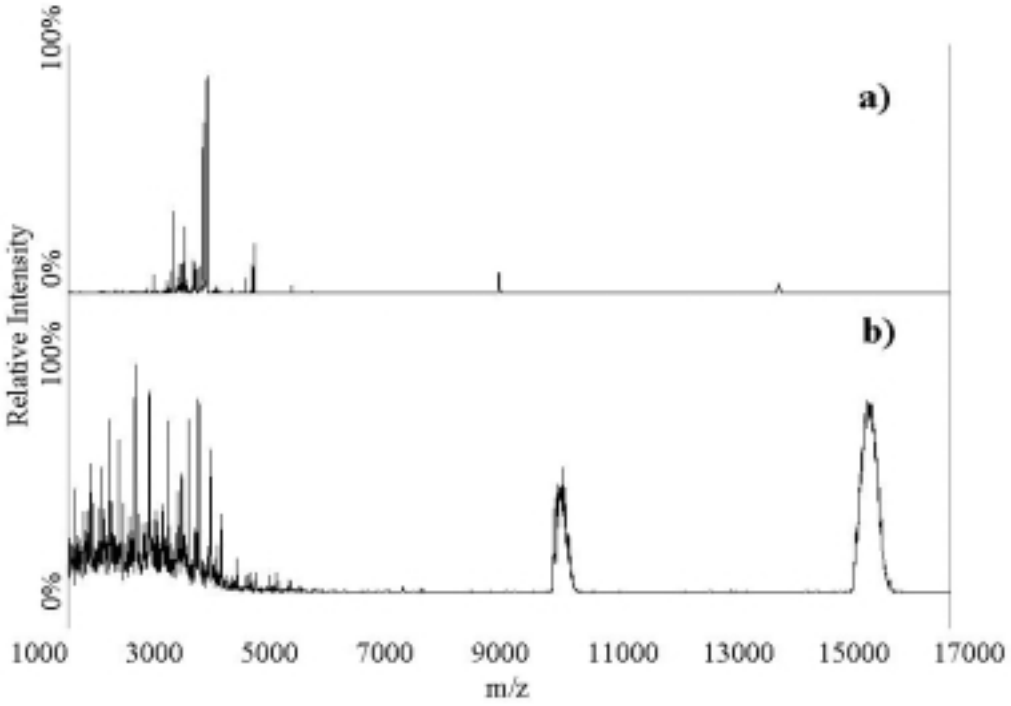

Figure 3.4 Vapor phase acid hydrolysis spectra of a) cytochrome C and b) lysozyme. The hydrolysis is more complete with cytochrome C, but a sequence ladder is generated from lysozyme.

predominantly the C-terminal side of aspartic acids and the N-terminal side of serine and threonine residues. Currently used on-line protein databases like Protein Prospector (http://prospector.ucsf.edu) do not allow the user to define cleavage on both the N- and C-terminal side of residues for the same digestion. As a result most searches take the

compiling of several simulated digestions and are very inconclusive.

A sequence ladder, as generated with lysozyme, indicates a portion of the protein sequence and results in a highly confident identification. Although lysozyme is well known as being one of the more difficult proteins to digest, a sequence ladder 11 residues in length is generated. The spacing between the peaks is shown in figure 3.5 as well as the residues to which the mass differences correspond.

The successful implementation of vapor phase acid hydrolysis to routine protein identification would require significant improvements to the presently reported protocol. Detectable spectra were generated for only 20% of the trials. In addition to the inconsistency of generating results, the quality and degree of digestion seemed to be dependent upon many factors including the protein itself, protein concentration, acid concentration, temperature, and time of the incubation. Also, the vapor phase approach is not capable of hydrolyzing the bulk protein deposit, possibly only small amounts from the surface.

On-Probe Liquid Phase Acid Hydrolysis

The general approach for liquid phase acid hydrolysis begins with drying various amounts of protein on the MALDI probe. Initially, 1mg/ml of BSA and lysozyme were dried on separate targets. Then, 1μL of various concentrations of PFPA, 90%, 60%, and 30%, were deposited directly on the protein spots. 1μL of H_2O was added to each spot after 2 minutes to keep the spots solvated. The targets dried after approximately 4

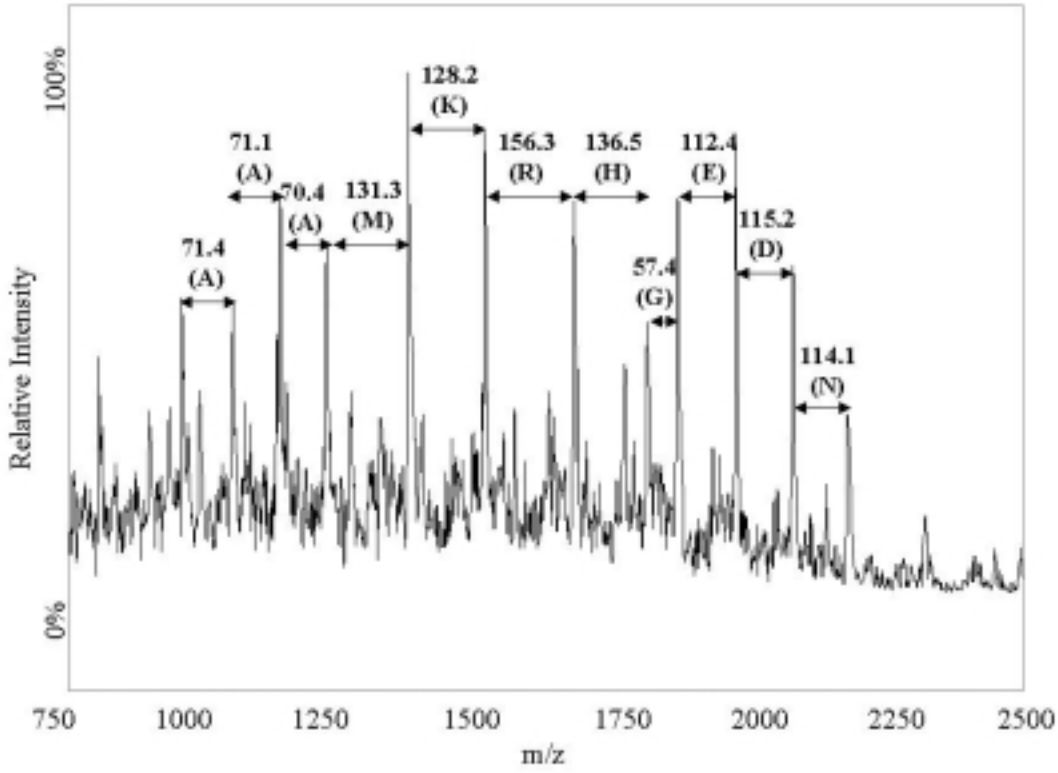

Figure 3.5 A sequence ladder 11 residues in length is generated from lysozyme.

minutes. DHB matrix was added to the targets and the samples were analyzed. The comparison of the BSA digestion in various PFPA percentage solutions are shown in figure 3.6. For 90% and 60% PFPA solution, several intense peptides were formed, while in 30% PFPA a dramatic sequence ladder is generated. Figure 3.7 shows an expanded view of the sequence ladder. The spacing and corresponding amino acids are labeled. For all PFPA concentrations used with lysozyme a few intense peptides were generated (data not shown) but no sequence ladders.

Conclusion

These initial experiments yielded inconsistently successful results. The most complete hydrolysis generating a sequence ladder occurred with 30% PFPA added directly to a dried BSA sample. The ease at which a protein can be identified with such a sequence ladder emphasizes the potential for this technique. The hydrolysis appeared to be limited by the lifetime of the droplet on the chip surface. Also, attempts to hydrolyze protein in the presence of matrix failed. A "perfect" protein-acid ratio or concentration that would optimize the reaction was never determined. At this point the acid hydrolysis of proteins occurs with decreasing efficiency and reproducibility as the concentration of protein is decreased. Because so often no spectra were obtained, we speculate that the high concentration of PFPA may be harmful to the MALDI signal. Thus, future work will explore using lower concentrations of acid and other analogous "MALDI friendly" acids like trifluoroacetic acid. Further development of these experiments and procedures were published[7,8] and patented in collaboration with Ciphergen Biosystems, Inc.

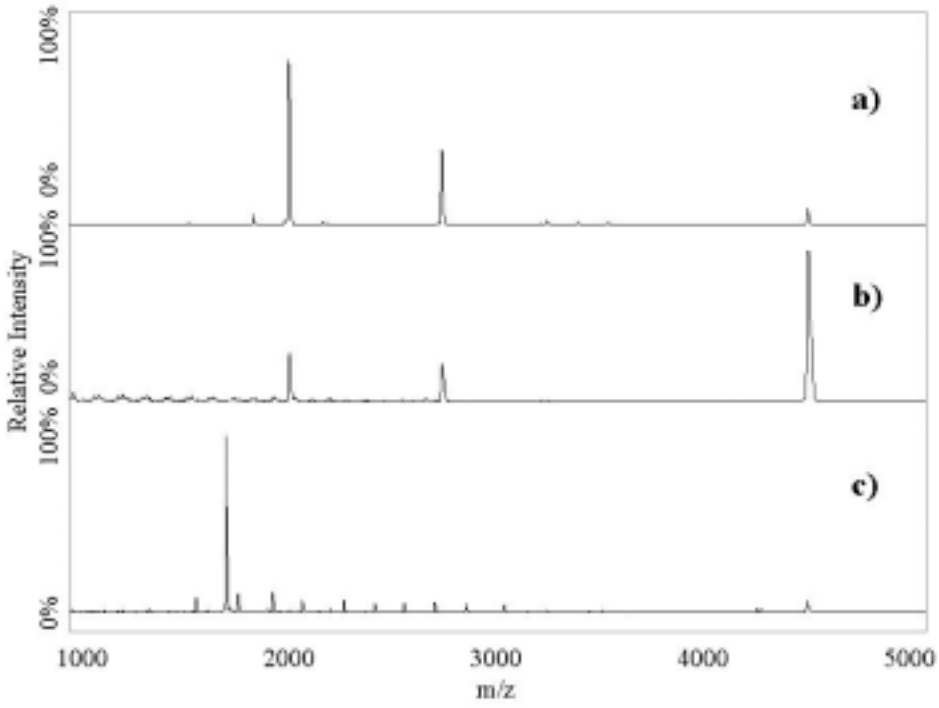

Figure 3.6 On-probe liquid phase acid hydrolysis of BSA with various acid concentrations; a) 90%, b) 60%, and c) 30% PFPA.

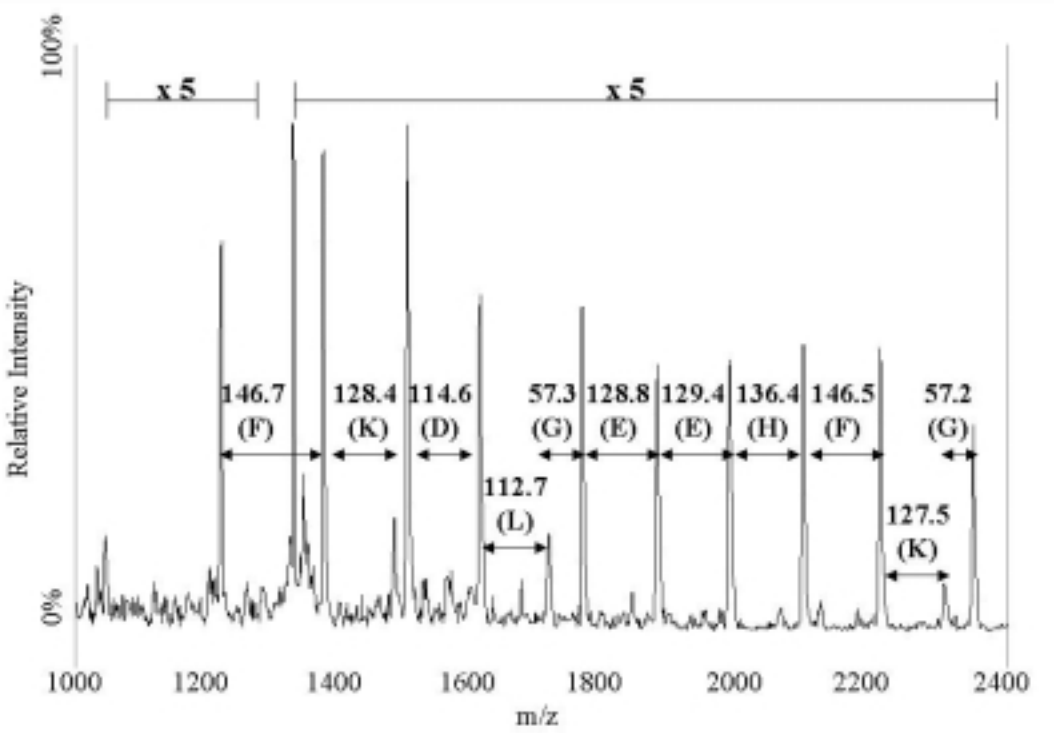

Figure 3.7 On-probe liquid phase acid hydrolysis of BSA with 30% PFPA yielded a very intense sequence ladder.

Acknowledgments

Our appreciation is expressed to Ciphergen Biosystems, Inc. for funding that helped to support this work, and to Scot Weinbeger and Shanua Lin of Ciphergen for there productive collaboration. An additional thank you is extended to Denny Warrenfeltz of the CCRC for his help in assembling the vacuum chamber.

References

1. S. M. Partridge and H. F. Davis, *Nature*. **165**, 62 (1950).

2. W. G. Crewther, A. S. Inglis, and N. M. Mckern, *Biochem. J.* **173**, 365 (1978).

3. J. Gobom, E. Mirgorodskaya, E. Nordhoff, P. Hojrup, and P. Roepstorff, *Anal. Chem.* **71**, 919 (1999).

4. E. Mirgorodskaya, H. Hassan, H. H. Wandall, H. Clausen, and P. Roepstorff, *Anal. Biochem.* **269**, 54 (1999).

5. A. S. Inglis, *Methods Enzymol*. **91**, 330 (1983).

6. R. L. Hill, in *Advances in Protein Chemistry,* C. B. Anfinsen (Ed.), Academic Press, 37 (1963).

7. S. Lin, P. Tornatore, S. R. Weinberger, D. King, and R. Orlando, *Eur. J. Mass Spectrom*. **7**, 131 (2001).

8. S. Lin, P. Tornatore, D. King, R. Orlando, S. R. Weinberger, *Proteomics*. **1**, 1172 (2001).

CHAPTER 4

STUDYING PROTEIN-CARBOHYDRATE INTERACTIONS BY AMIDE HYDROGEN/DEUTERIUM EXCHANGE MASS SPECTROMETRY[1]

[1] King, D., Lumpkin, M., Bergmann, C., and Orlando, R.. To be submitted to *Rapid Communications in Mass Spectrometry*.

Abstract

Protein-carbohydrate interactions play a significant role in biological processes. Presented here is the novel application of amide hydrogen/deuterium exchange-mass spectrometry (amide exchange-MS) to the study of the interaction between a protein and its carbohydrate substrate. The degree of deuterium incorporation into hen egg lysozyme was monitored with and without substrate to verify that a carbohydrate can provide sufficiently stable protection of the amide hydrogens in a protein's backbone from exchange with deuterated solvent. The substrate protected a number of amide hydrogens from exchange, implying that protein-carbohydrate binding systems will be compatible with amide exchange-MS. Endo-polygalacturonase-II (EPG-II) from *Aspergillus niger*, a pectin degrading enzyme, was chosen as the first carbohydrate binding system to be extensively studied using quenched amide exchange-MS. Monitoring the changes in deuterium incorporation of EPG-II in the presence and absence of an oligomer of galacturonic acid implied the location of substrate binding. This study demonstrates the ability of amide exchange-MS to investigate protein-carbohydrate interactions.

Introduction

Protein-carbohydrate interactions play critical roles in cell recognition systems ranging from the fertilization of egg cells to the degradation of plant cell walls.[1-5] While many classes of proteins are well understood and characterized, carbohydrates are much less so and are often difficult to work with. The study of carbohydrates presents unique challenges, and thus the techniques typically employed to study protein-protein binding have not been as readily applied to protein-carbohydrate systems. The principal successes in the characterization of protein-protein interactions have been achieved through X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. X-ray crystallography can image a bound complex only if the ligand and substrate will co-crystallize. In a protein-carbohydrate system, the two are much less likely to crystallize under the same conditions, as carbohydrates are usually more flexible and often have different solubilities than proteins. Although some proteins and small carbohydrate ligands have been co-crystallized, the difficulties encountered with protein-carbohydrate systems spur the efforts to find complimentary techniques.[6,7] NMR has also been used to characterize protein-carbohydrate binding but requires a long experiment time, much more sample than is currently available for many of these systems, is limited to small proteins or peptides, and typically requires isotopic labeling of one or both species.[8] Requiring as little as picomoles of protein, amide hydrogen/deuterium exchange-MS (amide exchange-MS) is not subject to the limits of size and solubility discussed above and may provide a relatively quick alternative for studying protein-carbohydrate interactions.[8]

Amide exchange-MS provides a way to monitor the rates at which amide hydrogens on a protein backbone exchange with hydrogen or deuterium in the solvent. Amide hydrogens are labile and will freely exchange with the protons or deuterons in solution. When a protein is dissolved in D_2O or a mixture of D_2O and H_2O, solvent accessible (exterior) amino acids will incorporate deuterium. Each incorporated deuteron will result in an increase of 1 Dalton to the protein's overall mass, which is easily monitored by modern mass spectrometers. The maximum possible number of deuterons that could be incorporated into the protein is approximately equal to the number of amino acid residues. Using electrospray-ionization mass spectrometry (ESI-MS) the mass of the protein can be monitored during the exchange until the mass stops increasing and becomes relatively constant. The difference between the theoretical maximum incorporation and the experimental result indicates how many amide hydrogens are being protected from exchange with the solvent by the secondary, tertiary, and quaternary structure of the protein.[9-11] In this way, amide exchange-MS has been used for some time to study the three-dimensional conformations of proteins.

In a recent variation of amide exchange-MS, the deuterium exchange is quenched to allow for an enzymatic digestion step.[8,12-17] In addition to determining the total number of amino acids protected or exchanged, digestion of the exchanged protein allows the locations of the incorporated deuterons to be assigned to specific peptides. Quenching is achieved by lowering both the temperature and pH, as amide hydrogen exchange rates decrease with decreasing temperature and with pH to a minimum at pH 2.5.[8] A few experiments have reported the successful use of amide exchange-MS in studying proteins interacting with small ligands.[8,13,14,17] Just as the secondary and tertiary

structure of a protein will protect some backbone amide hydrogens from exchange with solvent, the presence of a bound ligand will also protect exterior amino acids from exchange.

We report here the novel application of amide exchange-MS to study the interaction between protein and carbohydrate systems. Initially, to test the feasibility of studying a protein-carbohydrate system with amide exchange-MS, the technique was used to monitor the deuterium incorporation of hen egg lysozyme in the presence and absence of its carbohydrate substrates, chitobiose and chitotetraose. Subsequently, deuterium incorporation studies of the pectinase endopolygalacturonase-II (EPG-II), from the fungus *Aspergillus niger*, were carried out in the presence and absence of an octamer of galacturonic acid, $(GalA)_8$.

Experimental Methods

Materials:

Lysozyme, pepsin, sucrose, chitobiose, chitotetrose, and D_2O were purchased from Sigma (St Louis, MO). The D201E mutant form of EPG-II was a gift of the laboratory of Jaap Visser of Wageningen Agricultural University, The Netherlands, and was prepared as published.[18,19] The octamer of galacturonic acid, $(GalA)_8$, was a kind gift of Stefan Eberhard of the CCRC. The liquid chromatography (LC) buffers were made with acetic acid from J.T. Baker (Phillipsburgh, NJ) and acetonitrile from Fisher Scientific (Pittsburgh, PA). Buffer A consists of 94% H_2O and 6% acetic acid by volume. Buffer B consists of 77% acetonitrile, 17% H_2O, and 6% acetic acid by volume.

Procedures:

Lysozyme

50.0μg aliquots of lysozyme were placed in 4 vials. Equimolar amounts of chitobiose, chitotetraose, or sucrose (negative control) were added to the first 3 vials, the 4th was left as a control. 500μL of 50% D_2O was added to all 4 vials to generate a final protein concentration of approximately 0.1mg/mL. Immediately after mixing the samples were infused into a Perkin Elmer Sciex API-III biomolecular mass analyzer (Norwalk, CT, USA) at a flow rate of 2μL/min and the molecular mass was monitored for 1hour.

EPG-II

Three trials were performed with the EPG-II sample; EPG-II in H_2O (1), EPG-II in 50%D_2O (2), and EPG-II bound with $(GalA)_8$ in 50%D_2O (3). The mutant EPG-II sample was approximately 1mg/mL in concentration, and 10μL aliquots were placed into three micro-centrifuge tubes. A 100 molar excess of $(GalA)_8$ was added to tube 3 and allowed to incubate at room temperature for 24 hours. Subsequently, 10μL of H_2O was added to tube 1, and 10μL of D_2O to tubes 2 and 3. The samples were allowed to exchange for 24 hours before they were digested and analyzed by LC/MS. At the end of the incubation, the tubes were placed in an ice bath and 10μL of 0.01M HCl was added to quench the exchange. 4μL of 1mg/mL pepsin was then added and the digestion progressed for 6 minutes.

It is important that the deuterium exchange be quenched during both the pepsin digestion and the subsequent LC separation. Therefore, an LC apparatus was constructed in-house such that the majority of its components could be immersed in an ice bath (Fig. 4.1). A styrofoam box was used to house the reverse-phase HPLC column, a 25 mm long

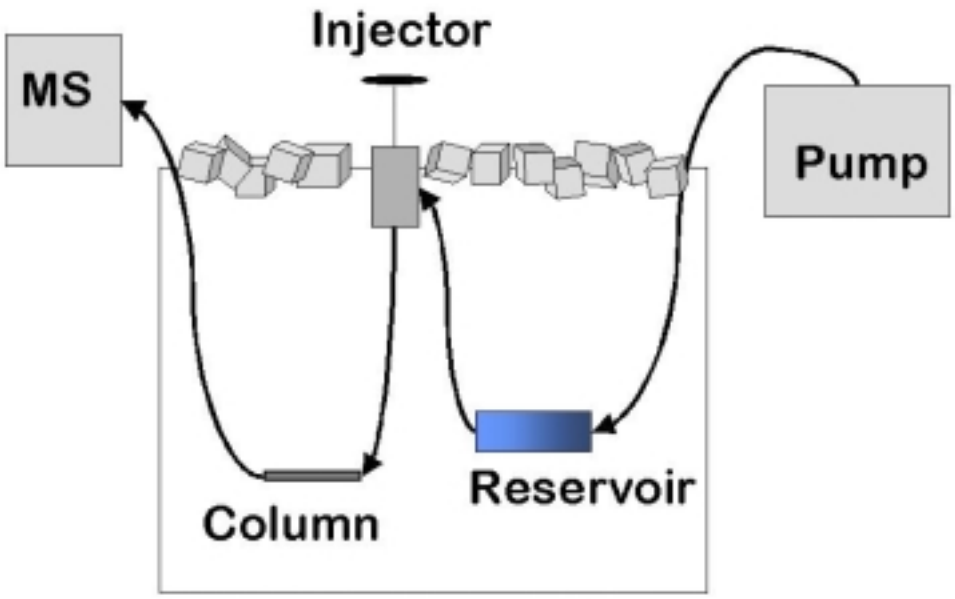

Figure 4.1 Diagram of the ice cooled LC system.

50 micron MAGIC C18 Bullet (Michrom Bioresources, Inc., Auburn, CA), as well as the injector, injection loop, and solvent reservoir. An Applied Biosystems (Foster City, CA, USA) 140B solvent delivery system delivered a gradient of 10% to 80% buffer B over 8 minutes. It was impossible to cool the pumps directly, therefore a solvent reservoir, consisting of a 2mL stainless steel tube (.76mm ID), was inserted in the line prior to the injector. At a flow rate of 200μL/min for these experiments, the 2mL metal reservoir insured cooling of the solvent in the ice bath for 10 minutes before contacting the sample in the injection loop. Metal tubing was used for all plumbing inside the ice bath to encourage heat transfer, and peak tubing was used from the bath to the mass spectrometer for insulation. The peptide mixture was analyzed using a Micromass Q-tof 2 ESI-MS (Manchester, UK). Spectra were obtained for the mass to charge ratio (m/z) of 500 to 2000 at a rate of 1 scan/second in positive ion mode.

To determine the amount of back exchange, the molecular mass of completely deuterated angiotensin-II was monitored with this cooled LC system. After exchanging angiotensin-II with 99.96% deuterium overnight, the fully deuterated sample was analyzed with this LC-MS apparatus. The results demonstrated that minimal back exchange had occurred. The system was later used to identify the exterior amides on cytochrome C, which yielded results in excellent agreement with H^1-NMR data reported for cytochrome C[21] as well as with the exterior amides predicted from the x-ray crystal structure.[20]

Results & Discussion

The study of protein-carbohydrate interactions by NMR and x-ray crystallography has encountered limitations, thus we were eager to adapt amide exchange-MS to study the

interaction between fungal EPGs and plant cell wall polygalacturonic acid. This relatively young technique has found success in monitoring changes in deuterium incorporation as a result of protein folding, unfolding, and ligand binding.[8-17] To our knowledge, amide exchange-MS has not previously been used to characterize a protein-carbohydrate system. In this work we demonstrate the ability of amide exchange-MS to be used as a tool for such studies.

Protection from Exchange by Carbohydrate Substrates:

Several experiments were performed on the well-studied lysozyme-chitin system to test the ability of amide exchange-MS to characterize this type of interaction. In these studies, lysozyme and equimolar amounts of lysozyme with chitobiose, chitotetraose, or sucrose (a negative control) were exchanged with D_2O and monitored for one hour by ESI-MS (Fig. 4.2).

These results clearly show that chitobiose and chitotetraose prevented amide hydrogens on lysozyme from exchange with the solvent. Furthermore, doubling the size of the sugar ligand (biose to tetraose) doubled the number of amide hydrogens protected; approximately 15 were protected by chitobiose and 30 by chitotetraose. These numbers are in excellent agreement with the structure of these complexes determined by X-ray diffraction, which indicate that chitobiose and chitotetraose will protect 16 and 28 amide hydrogens from exchange with the solvent respectively.[22, 23] This data strongly suggests that the carbohydrate substrate selectively protects amide hydrogens at the site of interaction. These experiments provide the first evidence that the interactions between carbohydrates and carbohydrate-binding proteins can be studied by amide exchange-MS.

This system was not used for the quenched exchange experiments as pepsin can not digest lysozyme without reduction/carboxymethylation.

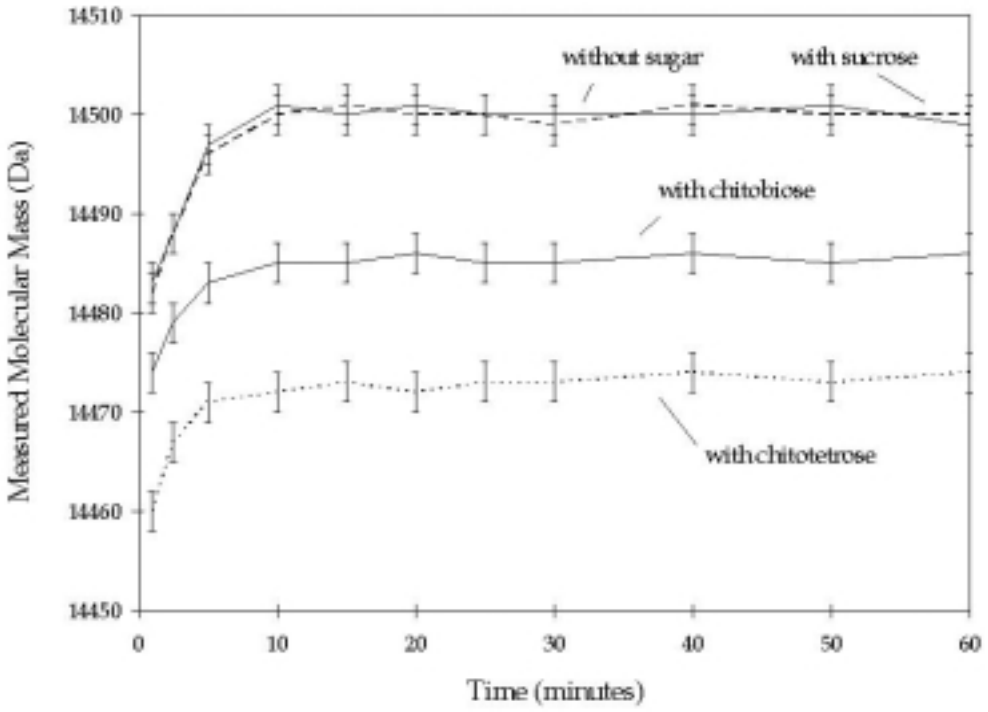

Figure 4.2 Molecular masses of lysozyme, as determined by ESI-MS, for a one-hour period by itself and with chitobiose, chitotetraose, or sucrose (negative control) dissolved in a 1:1 mixture of H_2O to D_2O.

Quenched Exchange for Locating Regions of Protein-Carbohydrate Binding:

Following the successful experiments with lysozyme, we began our study of the EPG-II/$(GalA)_8$ binding interaction, as EPG-II is susceptible to pepsin digestion. Polygalacturonases, like EPG-II, hydrolyze plant cell wall carbohydrates and are released by microbial pathogens as part of the first wave of attack during pathogenesis. To properly adapt the experiment to EPG-II, a few considerations had to be made. First of all, EPG-II is a hydrolytic enzyme with a reasonably high k_{cat} [24] which will not remain bound to the oligomeric substrate for the length of the experiment. Our collaborators have prepared a number of EPG-II mutants, and the D201E mutant was selected for having a similar binding constant to the wildtype enzyme while demonstrating negligible hydrolytic activity (k_{cat} < 0.01% of wildtype).[19] Secondly, the EPG-II/PGA system has a pH optimum between pH 4-5 with activity dropping to essentially zero by pH 6.[24] The sample was therefore exchanged for 24 hours, which is considerably longer than typical pH neutral protocols, because amide hydrogen exchange rates decrease with a decrease in pH (with a minimum at ~ pH 2.5).[8] An oligomeric substrate, $(GalA)_8$, was used instead of the native polymeric substrate because the length of the octamer is similar to that of the EPG-II cleft, eliminating any incidental contact between the extremities of the polymer and the outside of the EPG-II. Finally, while X-ray crystallography and site specific mutagenesis have implicated the EPG-II cleft as the location of carbohydrate binding,[18] it is important to note from the crystal structure that a large percent of the protein (including the entire active site cleft) consists of β-pleated sheets (Fig. 4.3). Amide hydrogens involved in hydrogen bonding within these sheets have greatly reduced rates of exchange

and it is, therefore, unlikely that much deuterium incorporation will occur in these regions.[25]

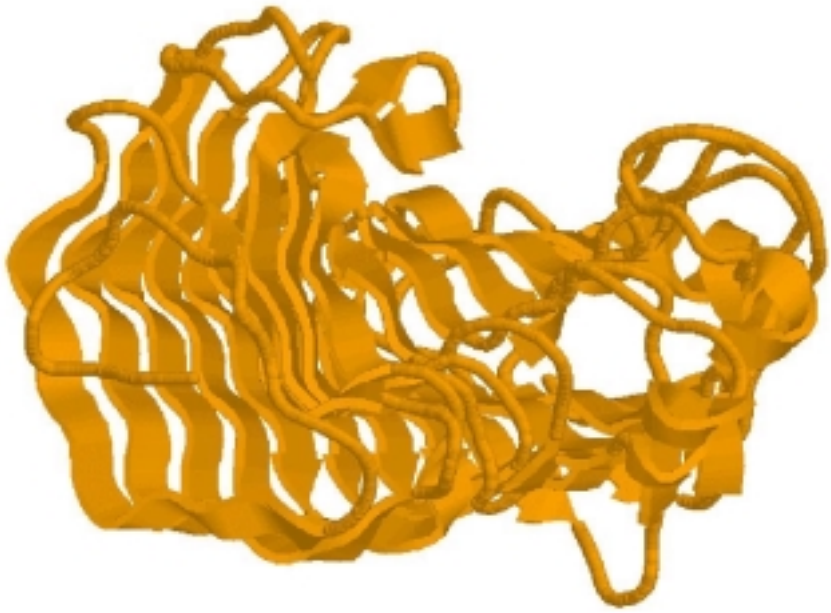

Figure 4.3 A cartoon representation of the X-ray crystal structure of EPG-II from *A. niger*.[18]

The peptide peaks from trial 1(control) were assigned to specific regions of the amino acid sequence by matching the peptide masses to those generated by a theoretical pepsin digest using the on-line program Prospector from the University of California-San Francisco (http://prospector.ucsf.edu). After these peaks were assigned, a comparison of the spectra from trials 1 and 2 indicated that deuterium had been incorporated into some of the peptides (Fig. 4.4).

Traditional methods for quantifying the amount of deuterium incorporation involves finding the centroid of the entire isotope envelope.[8, 16] The shift of the centroid

as a result of deuterium exchange corresponds to the mean deuterium incorporation. As reported previously in the literature, the determination of absolute values of deuterium incorporation are unnecessary for binding studies, as any reduction in the level of deuterium incorporation is sufficient to indicate ligand binding and protection from the solvent.[8] In the present study, the amount of deuterium incorporated into each peptide was determined by multiplying the change in the m/z of the most abundant isotope following deuterium exchange by the charge of the peptide. Therefore, a single integer value, which is an estimate of the most abundant deuterium incorporation, was used to describe the deuterium incorporation into a peptide. Because many of the peptide sequences overlap, it was possible to further deduce the location of incorporated deuterium into smaller sections of the amino acid sequence. This approach is less sensitive than using centroid values because a minimum amount of deuterium (approximately 10% incorporation for an average 10-residue peptide) must be incorporated before the most abundant isotope will shift. However, the sensitivity is sufficient for these studies and therefore desirable because of its simplicity.

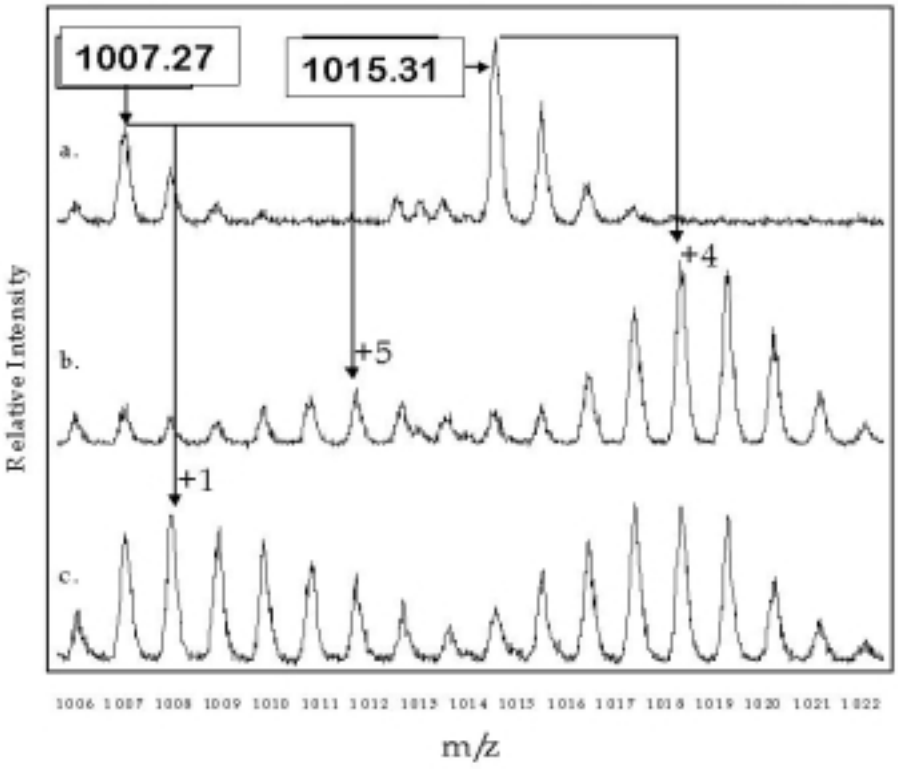

Figure 4.4 Comparison of mass spectra collected from (a) free EPG-II in H_2O, (b) free EPG-II in 50% D_2O, and (c) EPG-II/$(GalA)_8$ complex in 50% D_2O. The arrows indicate shifts in the most abundant isotope peak as a result of deuterium incorporation.

Figure 4.4 shows 2 peptides with m/z ratios of 1007.265 and 1015.307 Da. When the free EPG-II mutant is exchanged with deuterium approximately 5 and 4 deuterons are incorporated into each peptide respectively. However, in the presence of the substrate only one deuteron was incorporated into peptide 1007.265 Da. implying the substrate is contacting this peptide and protecting several residues from the solvent. Despite the presence of β-pleated sheets in the cleft, several peptides were deuterated in the free EPG-II mutant and were subsequently protected when in the presence of substrate (Fig. 4.5).

Interestingly, the distance between the most distant protected residues within the cleft (Tyr_{283}-Gly_{65}) is ~38 Å, similar to the length of the $(GalA)_8$ substrate, estimated using the crystal structures of other oligosaccharides. The octamer of chitin $(GlcNAc)_8$, published by Y. Papanikolau, et al.,[26] and tetramer of hyaluronic acid, $(GlcA\text{-}GlcNAc)_4$, published by J. M. Guss, et al.,[27] are both approximately 36 Å long. A tetramer of galacturonic acid, $(GalA)_4$, was determined to be ~17 Å as published by R. D. Scavetta, et al.[28] These results suggest that the substrate must lie more or less linearly along the entirety of the cleft to provide protection from Tyr_{283}-Gly_{65}. Of the eight amino acids strictly conserved throughout known fungal, bacterial, and plant polygalacturonases, site-directed mutagenesis experiments have identified two amino acids which have no effect on hydrolysis but which may be crucial for substrate binding, Arg_{256} and Lys_{258} (Fig. 5).[18, 19] These crucial residues are located within a region shown to be protected from exchange by the substrate binding. Two β-pleated sheets on either side of these residues were determined to be protected (Fig. 4.5). The agreement between these studies implies that amide exchange-MS may be a valuable complement to mutagenesis research.

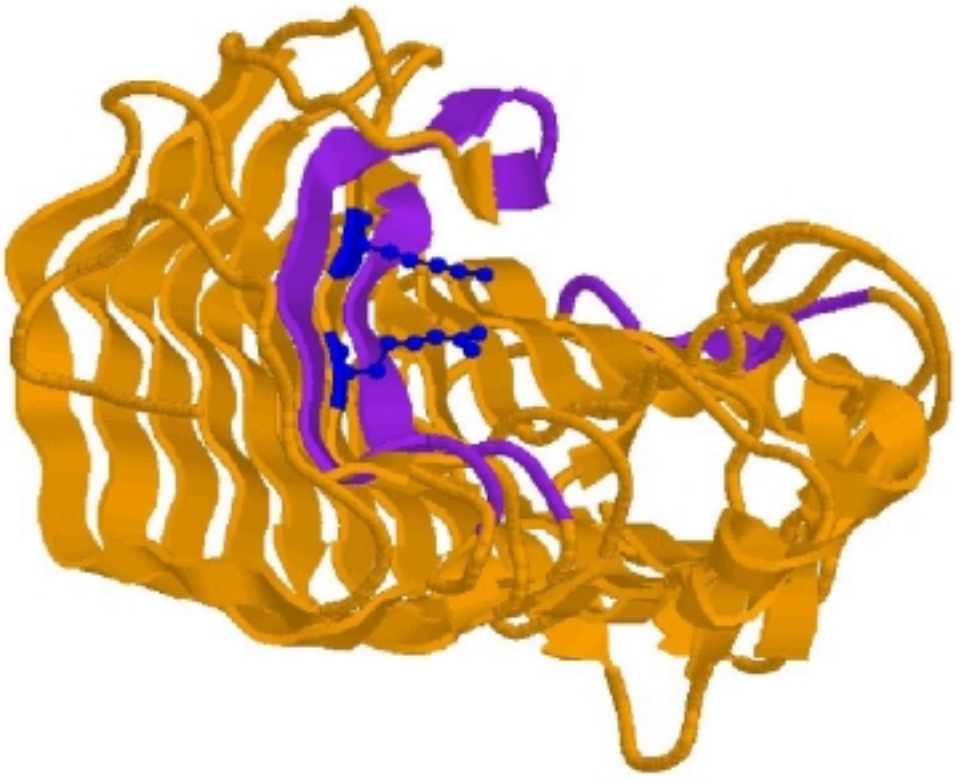

Figure 4.5 A cartoon representation of the X-ray crystal structure of EPG-II from *A. niger*.[18] β-pleated sheets within EPG-II that showed protection from deuterium exchange by the presence of substrate are shown with darker shaded sheets. Arg_{256} and Lys_{258}, shown by site specific mutagenesis to be critical for substrate binding are displayed as ball and stick models.

Conclusion

The results presented here suggest that carbohydrate ligands can effectively protect amide hydrogens from exchange with deuterated solvent despite their natural flexibility. The degree of secondary structure should be considered when choosing amide exchange-MS as an analysis technique. In this study of EPG-II, minimal binding site data was acquired, presumably due to the high degree of β-sheets within the binding cleft. Regions of protection by the substrate were found to be near the residues (Arg_{256} and Lys_{258}) determined to be involved in substrate binding by mutation experiments. The agreement between the amide exchange-MS data and the site specific mutagenesis data demonstrates the potential of amide exchange-MS as a complementary technique to other more common protein characterization methods. Furthermore, amide exchange-MS is shown to be capable of characterizing protein-carbohydrate complexes and will be a valuable tool for future studies.

Acknowledgements

The authors would like to greatfully thank the National Science Foundation (MCB - 0115132), the National Institutes of Health (2 P41 RR053531), the Mizutani Foundation for Glyco Science (17), and the Department of Energy (DE-FG02-93ER20097) and (DE-FG02-96ER20221) for grants that supported this work. We would also like to express our appreciation to the laboratory of Jaap Visser at Wageningen Agricultural University, The Netherlands, for supplying our project with the EPG-II mutant and Stefan Eberhard of the CCRC for the $(GalA)_8$ substrate.

References Cited

1. R. M. Cooper, in *Biochemical Plant Pathology*, J.A. Callow (Ed.), John Wiley and Sons Ltd., 1995.

2. N. C. Carpita, and D. M. Gibeaut, *Plant J.* **3**, 1 (1993).

3. T. M. Jones, A. J. Anderson, and P. Albersheim, *Physiol. Plant Pathol.* **2**, 153 (1972).

4. M. G. Hahn, P. Bucheli, F. Cervone, S. H. Doares, R. A. O'Neill, A. Darvill, and P. Albersheim, in *Plant-Microbe Interactions. Molecular and Genetic Perspectives*, T. Kosuge and E.W. Nester, (Eds.), McGraw Hill Publishing Co., 1989.

5. C. Grassin, and P. Fauquemberque, in *Pectins and Pectinases*, J. Visser and A. G. J. Voragen (Eds.), Elsevier Science B.V., 1996.

6. W. I. Weis, K. Drickamer, and W. A. Hendrickson, *Nature*. **360**, (1992).

7. E. A. Merritt, S. Sarfaty, F. van den Akker, C. L'Hoir, J. A. Martial, and W. G. J. Hol, *Protein Sci.* **3**, 166 (1994).

8. H. Ehring, *Anal. Biochem.* **267**, 252 (1999).

9. V. Katta, and B. T. Chait, *J. Am. Chem. Soc*. **115**, 6317 (1993).

10. Z. Q. Zhang, and D. L. Smith, *Protein Sci*. **2**, 522 (1993).

11. Y. Q. Liu, and Smith, D. L. *J. Am. Soc. Mass Spectom*. **5**, 19 (1994).

12. D. L. Smith, and Z. Q. Zhang, *Mass Spectrom. Rev.* **13**, 411 (1994).

13. J. B. Smith, Y. Q. Liu, and D. L. Smith, *Exp. Eye Res*. **63**, 125 (1996).

14. F. Wang, J. S. Blanchard, and X. J. Tang, *Biochem*. **36**, 3755 (1997).

15. Y. Deng, and D. L. Smith, *Anal. Biochem.* **276**, 150 (1999).

16. L. Wang, L. C. Lane, and D. L. Smith, *Protein Sci*. **10**, 1234 (2001).

17. S. Akashi, and K. Takio, *Protein Sci*. **9**, 2497 (2000).

18. Y. van Santen, J. A. E. Benen, K. H. Schroer, K. H. Kalk, S. Armand, J. Visser, and B. W. Dijkstra, *J. Biol. Chem*. **274**, 30474 (1999).

19. S. Armand, M. J. M. Wagemaker, P. Sanchez-Torres, H. C. M. Kester, Y. van Santen, B. W. Dijkstra, J. Visser, and J. A. E. Benen, *J. Biol. Chem*. **275**, 691 (2000).

20. J. A. Feng, R. C. Johnson, R. E. Dickerson, http://pdb.ccdc.cam.ac.uk/oca-bin/ccpeek?id=1HCR, (1993).

21. J. S. Milne, L. Mayne, H. Roder, A. J. Wand, and S. W. Englander, *Protein Sci*. **7**, 739 (1998).

22. V. B. Vollan, E. Hough, S., Karlsen, http://pdb.ccdc.cam.ac.uk/oca-bin/ccpeek?id=1BB7, (1998).

23. K. Maenaka, M. Matsushima, H. Song, F. Sunada, K. Watanabe, I. Kumagai, http://pdb.ccdc.cam.ac.uk/oca-bin/ccpeek?id=1LZC, (1995).

24. J. A. E. Benen, H. C. M. Kester, and J. Visser, *Eur. J. Biochem*. **259**, 577 (1999).

25. D. L. Smith, Z. Q. Zhang, and Y. Q. Liu, *Pure Appl. Chem*. **66**, 89 (1994).

26. Y. Papanikolau, G. Prag, G. Tavlas, C. E.Vorgias, A. B. Oppenheim, K. Petratos, http://pdb.ccdc.cam.ac.uk/oca-bin/ccpeek?id=1EHN (2000).

27. J. M. Guss, D. W. L. Hukins, P. J. C. Smith, W. T. Winter, S. Arnott, R. Moorhouse, D. A. Rees, http://pdb.ccdc.cam.ac.uk/oca-bin/ccpeek?id=2HYA (1977).

28. R. D. Scavetta, S. R. Herron, A. T. Hotchkiss, N. Kita, N. T. Keen, J. A. Benen, H. C. Kester, J. Visser and F. Jurnak, *Plant Cell* **11**, 1081 (1999).

CHAPTER 5

THE USE OF AMIDE EXCHANGE-MASS SPECTROMETRY TO STUDY CONFORMATIONAL CHANGES WITHIN THE ENDOPOLYGALACTURONASE II / HOMOGALACTURONAN / POLYGALACTURONASE-INHIBITING PROTEIN SYSTEM[1]

[1] King, D., Bergmann, C., Orlando, R., Benen, J.A.E., Kester, H.C.M., and Visser, J. To be submitted to *Biochemistry*.

Abstract

Amide exchange-mass spectrometry (MS) was used to study the enzyme endopolygalacturonase II (EPG-II) from *A. niger* as it binds to an oligosaccharide substrate. A localized decrease in deuterium incorporation in the EPG-II of the EPG-II/oligosaccharide complex relative to the free EPG-II identified the location of substrate contact, which is in agreement with published site specific mutation studies. In addition, when bound with substrate, regions of the EPG-II remote from the substrate binding site became exposed to the solvent, as revealed by an increase in incorporated deuterium, indicating a conformational change in the enzyme. Fluorescence experiments were performed to provide additional evidence for an altered conformation of the EPG-II as a result of substrate binding. This novel application of amide exchange-MS to the study of protein/carbohydrate binding has, for the first time, described in detail the conformational changes associated with EPG-II when binding a substrate. Amide exchange-MS was also used to study the interactions of EPG-II and polygalacturonase inhibitor protein (PGIP). Mass spectral data of the EPG-II/oligosaccharide complex in the presence of *P. vulgaris* PGIP indicates that the inhibitor contacts the EPG-II at a site remote from the substrate binding cleft, and is restricting the conformational changes of the EPG-II. Fluorescence experiments also revealed that upon binding of PGIP, the conformational changes mentioned above for the EPG-II/substrate complex are minimized. These results, together with previously reported data, point to a location on EPG-II for interaction with PGIP as well as a possible mechanism for non-competitive inhibition of EPG-II.

Introduction

A major goal of plant pathogenesis research is the thorough characterization and understanding of the interactions between pathogen derived plant cell wall degrading enzymes and plant cell wall carbohydrate substrates. Plant cell walls are composed of two interacting networks, a cellulose/hemicellulose network and a pectin network.[1] The pectin network includes several related acidic polysaccharides such as rhamnogalacturonan-I (RG-I), rhamnogalacturonan-II (RG-II), and homogalacturonan, also known as polygalacturonic acid (PGA).[2] Pectins form a major component of the primary cell walls of dicotyledons and non-graminaceous monocotyledons and are found in high concentrations in the middle lamella.[3] Among the microbial phytopathogens are viruses, bacteria, and fungi. Fungi often gain entry into the plant via the cell wall, often through the middle lamella.[4] Endopolygalacturonases (EPG) are a major component of the pectin-degrading activity of phytopathogenic fungi and are among the first degradative enzymes to be secreted upon fungal infection.[5,6] EPGs hydrolyze deesterified regions of wall-bound homogalacturonans, solubilizing the RG-I and RG-II, and open up the wall to the action of other exo- and endoglycanases, including cellulases, glucanases, xyloglucanases, and arabinoxylanases.[3,7]

During pathogenesis there is a potential for interaction between EPGs and plant cell wall derived EPG inhibitors known as polygalacturonase inhibiting proteins (PGIPs).[3,8] PGIPs are soluble, leucine rich repeat (LRR) glycoproteins, found in the cell wall.[9] PGIPs form high-affinity complexes with EPGs in a reversible, stoichiometric manner. The rate of hydrolysis of homogalacturonan by an EPG/PGIP complex is, depending on the source of the EPG and PGIP, between one and two orders of magnitude

slower than by the free EPG.[10] The inhibition of EPGs by PGIPs may not only slow down the gross action of EPGs on the solubilization and fragmentation of homogalacturonan, but will also extend the lifetime of biologically active oligogalacturonides (fragments of homogalacturonan) released from the cell wall by the action of EPGs. Since oligogalacturonides have been implicated in the defense response of plants, extending their lifetime likely contributes to a successful defense response.[5,6,10,11]

EPGs from a single strain of fungus may exist in a variety of isoforms, each of which may consist of a series of glycoforms. There is evidence that the heterogeneity among the EPGs allows for variability in their mode of action as well as in their ability to interact with, and be inhibited by, PGIPs.[11,12] In addition, the PGIPs of a single plant species may also be present as a set of isoforms and their associated glycoforms ([13], unpublished data of this lab). This isoform/glycoform variability provides the potential for a wide spectrum of specificity of EPG/PGIP interactions within any plant/pathogen pairing. There are a number of examples in which different EPG/PGIP pairings demonstrate different degrees of inhibition. Additionally, both competitive and non-competitive types of inhibition have been reported. Such data indicate that the location of interaction may differ, and is, at least in some cases, not at the active site.[14] The mode of action of a particular fungal EPG and its inhibition by PGIPs may be critical factors in determining whether the fungus is a viable pathogen.

A proposed model for the structure of a bean PGIP, based on its membership in the plant-specific LRR class of proteins, has been recently published.[15] In addition, site-specific mutation experiments indicate a role for several specific amino acids within the

PGIP during a fungal EPG/PGIP interaction.[16] A recent study identified sites of nine amino acids in PGIPs and nine amino acids in EPGs that are likely candidates for natural mutation, thus altering the specificity of interaction of the two proteins.[17] This study supported the conclusions from the site-specific mutation experiments mentioned above,[16] but also indicated other regions on the PGIP which may be of importance for binding to EPGs.

Crystal structures for a bacterial[18] and a fungal EPG[19] have been determined and serve as a model for EPGs. All EPGs are members of family 28 of the glycosyl hydrolases, as classified by the Henrissat structural classification scheme,[20] thus it is not surprising that the overall structure of the bacterial EPG resembles that of the fungal EPG. In spite of these similarities, PGIPs inhibit fungal EPGs but do not inhibit bacterial EPGs. To date no complete model of the EPG/PGIP complex has been proposed. Due to the difficulty in co-crystallizing two relatively large proteins of approximately 35,000 Da each, both of which show heterogeneity of glycosylation, it is unlikely a crystal structure of the complex will be available in the near future.

The degradation of the plant cell walls plays a significant role in both the ripening and rotting of fruits and vegetables. The importance, both agriculturally and commercially, of the action of EPGs on pectin has led to a large body of research that attempts to fully understand, and thus exploit EPGs.[21-23] Agricultural communities, interested in prolonging the lifetime of crops, are concerned about inhibiting EPG activity.[24] On the other hand, there are many industrial uses for an enzyme capable of degrading plant cell walls. For example, EPGs are used in the clarification of fruit juices, the removal of color from paper, and in detergents to improve the removal of stains.[22,25]

We have begun to study the EPG/homogalacturonan interaction using mass spectrometry coupled with amide hydrogen/deuterium exchange. Amide exchange-MS observes the rate at which amide hydrogens on a protein backbone exchange with hydrogen or deuterium in the solvent. Amide hydrogens are labile and will freely exchange with the protons in solution if they are on the exterior of the protein, accessible to the solvent. When a protein is immersed in D_2O or a mixture of D_2O and H_2O, amide hydrogens will be replaced by deuterons, each resulting in a mass increase of 1 Dalton, a change easily monitored by mass spectrometry.[26-28]

In a recent variation of amide exchange-MS, the protein is enzymatically digested to determine specifically where deuterium is being incorporated.[29-31] Typically, after less than a minute, at neutral pH, all exterior amide hydrogens will be able to exchange with deuterium in the solvent. The protein is then digested, and the peptides are evaluated. The deuterium exchange must therefore be quenched to prevent both new deuterium from being added to interior amino acids exposed as a result of digestion and back-exchange of the deuterium during the LC-MS process. The amide hydrogen/deuterium exchange rate decreases with temperature and pH (minimum at ~2.5). By placing the system in an ice bath and lowering the pH to 2.5, the half-life time of the deuterium on the protein can be extended to 40-50 minutes.[32] Therefore, any changes within the binding of the system due to the quenching conditions will not have a significant effect on the deuterium incorporation levels as the exchange rates are slow during this time. The extended half-life times are typically long enough to allow for enzymatic digestion and LC-MS analysis.[33,34] Pepsin is used for proteolytic digestion because its optimal activity is at low pH. To study protein-substrate binding, a protein is analyzed in the presence and absence

of substrate. The substrate will protect exterior amino acids from deuterium incorporation in the region of its interaction. This approach has been used previously to investigate sites of protein-protein and protein-ligand interactions.[29-31]

In addition to analyzing the enzyme/substrate complex, amide exchange-MS was also used to describe the structure of the *A. niger* EPG/*P. vulgaris* PGIP complex, as well as provide insight into the mechanism of inhibition. Using the methods developed here, other EPG/PGIP complexes with different attributes may be studied in the future as a means toward understanding the variations in EPG/PGIP interactions and the possible role of this protein-protein interaction in pathogenicity.

Materials and Methods

Materials:

Pepsin and D_2O were purchased from Sigma (St Louis, MO). The D201E mutant form of EPG-II was prepared as published (hereafter denoted as mEPG-II).[35] The octamer of galacturonic acid $(GalA)_8$ was a kind gift of Stefan Eberhard at the CCRC. *P. vulgaris* PGIP was prepared as described previously.[36] Recombinant PGIP-II was a kind gift of F. Cervone, University of Rome "La Sapienza". Acetic acid from J.T. Baker (Phillipsburgh, NJ) and acetonitrile from Fisher Scientific (Pittsburgh, PA) were used to prepare the LC buffers. Buffer A consists of 94% H_2O and 6% acetic acid (v:v). Buffer B consists of 77% acetonitrile, 17% H_2O, and 6% acetic acid.

Amide exchange-MS:

A total of four types of experiments were performed: mEPG-II in H_2O (1), mEPG-II in 50% D_2O (2), mEPG-II bound with $(GalA)_8$ in 50% D_2O (3), and mEPG-II bound with PGIP and then incubated with $(GalA)_8$ in 50% D_2O (4). The mutant EPG-II stock

was 1mg/mL, and 10μL aliquots were placed into four micro-centrifuge tubes. An equimolar amount of PGIP was added to tube 4 and 10μL of distilled water was added to tubes 1, 2, and 3 as a blank. All four tubes were incubated at room temperature overnight to ensure mEPG-II /PGIP interaction. Next, a 100 molar excess of $(GalA)_8$ was added to tubes 3 and 4, and all four tubes were incubated at room temperature for 24 hours. Subsequently, 20 μL of H_2O was added to tube 1, and 20 μL D_2O was added to tubes 2, 3, and 4 to reach approximately a 50% D_2O concentration. The samples were left to exchange for a further 24 hours. At the end of the incubation, the exchange was quenched by cooling the sample and lowering the pH. The tubes were placed in an ice bath and 4 μL of 0.1 M HCl was added to each tube to reach a final pH of 2.0. 4 μL of 1 mg/mL pepsin was then added and the digestion progressed for 6 minutes. The solution was then analyzed by HPLC-MS.

To ensure against any back exchange of the deuterium, an ice bath was constructed to house the injection loop and reverse-phase HPLC column (25 mm long, 50 micron MAGIC C18 Bullet, Michrom Bioresources, Inc., Auburn, CA). A Micromass Q-Tof-II (Manchester, UK), an electrospray ionization mass spectrometer, was used to record mass spectra continuously during the LC gradient program, which ran linearly from 20% buffer B to 65% buffer B over 8 minutes. This instrumental configuration was previously determined to yield only minimal backexchange using this experimental protocol.[37]

UV-fluorescence:

Emission spectra were generated for free mEPG-II (1), mEPG-II in the presence of $(GalA)_8$ (2), and mEPG-II /PGIP in the presence of $(GalA)_8$ (3) using a Shimadzu

Spectrofluorophotometer RF-5301 PC. Three 30 nM samples of mEPG-II were prepared. An equimolar amount of PGIP was added to sample 3 and allowed to incubate overnight. A 100 molar excess of $(GalA)_8$ was added to samples 2 and 3. All samples were then left at room temperature for 24 hours. The samples were excited at 290 nm and their maximum emission was recorded between 293-294 nm.

BIAcore:

PGIP was immobilized on a BIAcore 3000 Chip following standard protocols provided from the manufacturer. After immobilization, EPG-II samples were passed across the chip. First, the BIAcore chip was exposed to the wild-type EPG-II, and binding of EPG-II to PGIP was detected by surface plasmon resonance (SPR). Then in a similar mEPG-II was passed across the PGIP immobilized chip and the SPR signals were compared.

Results and Discussion

EPG-II/Homogalacturonan:

The use of pectin degrading enzymes (PDEs) for industrial purposes has heightened the interest in understanding the interactions between the PDEs released by pathogens and plant cell wall carbohydrates. The polygalacturonases from various plant pathogens of both fungal and bacterial origin share regions of significant amino acid and structural homology. The X-ray crystal structures of the EPG from the bacteria *E. carotovora* and of EPG-II from the fungus *A. niger* have been determined.[18,19] The β-barrel structure found within EPG-II is very similar to that found in the crystal structures determined in the past for pectate lyases that, for some time, have served as a model for the pectin degrading enzymes.[6] The prominent cleft along the barrel suggests a location

for substrate binding in the active site. Site-specific mutation experiments have been performed in an attempt to locate the key amino acids of the active site. Aided by homology searches, the mutation experiments identified several amino acids that are important for substrate binding and hydrolysis.[35] In spite of these advances, there are many questions that still remain regarding the mechanism of hydrolysis of homogalacturonan by EPGs. The study of PDE-carbohydrate interactions has proven difficult if not impossible by traditional techniques. Therefore, the adaptation of amide exchange-MS to study the binding of EPGs with homogalacturonan seemed highly appropriate, due to the recent success of amide exchange-MS in studying protein-protein interactions.[30-32,38]

The enzyme EPG-II from the fungus *A. niger* was chosen for the initial experiments because of the availability of an EPG-II inhibitor protein, a mutant form of EPG-II, and a crystal structure for EPG-II. A number of concerns regarding the enzyme's compatibility with amide exchange-MS were encountered and will be discussed here, as many of these concerns will apply to future studies of other PDEs. The EPG-II is a hydrolase, and as a result will not remain bound to the oligomeric substrate for the duration of the experiment. It was therefore necessary for a mutant of EPG-II to be prepared that maintained the binding capability without the hydrolytic activity. A number of mutants had been previously prepared and analyzed for activity and binding, and the D201E mutant was selected for having a K_m approximately that of the wild type while showing negligible hydrolytic activity.[35] Preliminary fluorescence (discussed later) and BIAcore experiments (results not shown) revealed that the D201E mutant has binding properties similar to that of the wild type and further support the appropriateness of this

mutant for the experiments described. A second concern is that the active pH range for EPG-II lies between pH 4-5, far off the optimal pH of 7.0 for amide hydrogen/deuterium exchange experiments. The sample must therefore be allowed to exchange for longer than the few minutes typically required of pH neutral protocols because exchange rates decrease with a decrease in pH, with a minimum exchange rate at ~pH 2.5. Finally, a large percent of the protein, including the entire active site cleft, consists of β-pleated sheets which further decreases the rate of amide exchange for many amino acids.[32,39] Internal hydrogen bonding between pleated sheets makes it unlikely that much deuterium incorporation will occur in this region. However, as will be seen, the large degree of internal hydrogen bonding within pleated sheets proved to be advantageous.

Work has been published describing the activity of EPG-II with its natural substrate, polymeric homogalacturonan.[40] The octamer of homogalacturonan, $(GalA)_8$, was selected as the substrate in this study because modeling indicated it would be approximately the length of the cleft in the EPG-II. A substrate that is too long might provide non-specific protection to amino acids on sites outside the cleft, while a substrate that is too short would not indicate how much of the cleft is occupied by the substrate and may not participate in binding representative of homogalacturonan. The length of $(GalA)_8$ was estimated using the crystal structures of other octamer oligosaccharides. The octamer of chitin $(GlcNAc)_8$, published by Y. Papanikolau, et al.,[41] was approximately 36 Å and a tetramer of galacturonic acid, $(GalA)_4$, published by R. D. Scavetta, et al.,[42] was approximately 17 Å in length.

Three experiments were performed to study EPG-II-homogalacturonan binding: mEPG-II in H_2O (control), mEPG-II in 50% D_2O, and mEPG-II in the presence of

$(GalA)_8$ in 50% D_2O following the procedure presented in detail elsewhere.[37] The deuterium exchange was quenched by dropping the temperature and pH. The mEPG-II samples were digested with pepsin, and the peptides were separated and detected by LC-MS. The peptides from the control trial were identified by matching their masses to those of a computer generated peptic digest of mEPG-II, and only those yielding unambiguous matches were used for analysis. Pepsin was observed to consistently cleave most hydrophobic residues, in agreement with other reports.[30,43] The amount of deuterium incorporation into each peptide was then determined by comparing the spectra of the deuterated trials in the presence and absence of substrate (Fig. 5.1). Figure 5.1(a) shows a singly charged peptide with a mass/charge ratio (*m/z*) of 886. Figure 5.1(b) has the same isotope pattern indicating a non-detectable amount of deuterium was incorporated into this peptide when the free mEPG-II was incubated in D_2O. A shift of the most abundant isotope peak can be used as an estimate of the mean deuterium incorporation for a peptide,[32,43] and figure 5.1(c) shows such a change in the standard isotope pattern. The most abundant isotope increased by 2 indicating that the mean deuterium incorporation was approximately 2 deuterons. After evaluating the amide exchange-MS data, identifying the peptide and estimating the amount of deuterium incorporation into each, a map of deuterium incorporation throughout the protein was constructed. The percent of deuterium incorporation was determined for small segments of mEPG-II. Figure 5.2(a) presents the changes in percent deuterium incorporation in short segments of mEPG-II caused by the presence of substrate. There were three regions of the protein that underwent interesting changes in deuterium incorporation: the binding cleft, an α-helix from Asp_{110} to Trp_{114}, and the β-sheets on the underside of the β-barrel (Fig. 5.3).

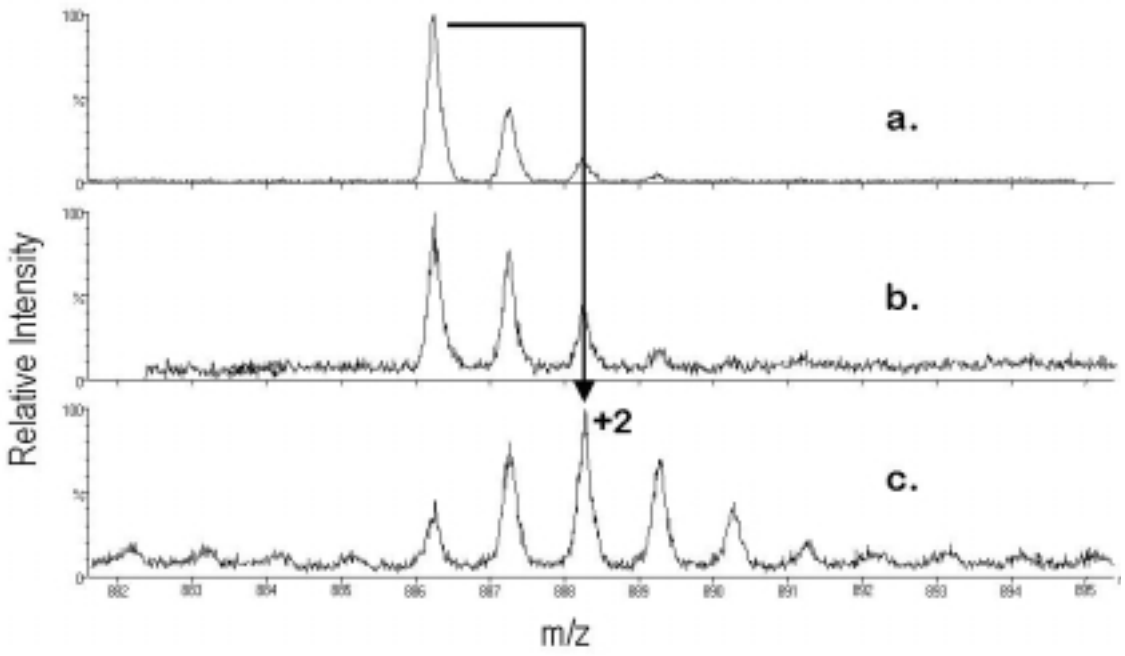

Figure 5.1 Mass spectra of a peptide (residues 131-139) from mEPG-II a.) in H_2O, b.) in D_2O, and c.) in the presence of substrate in D_2O.

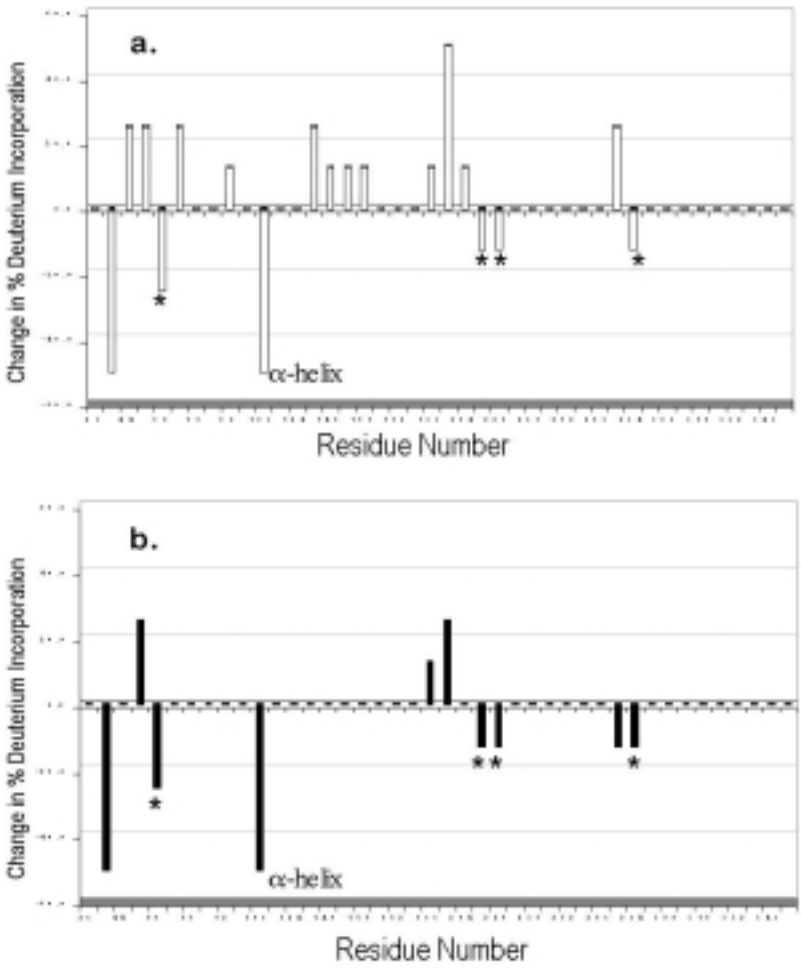

Figure 5.2 In the initial experiment some deuterium was incorporated into the free mEPG-II. From the mass spectral data, the amount of deuterium incorporated into segments of the mEPG-II sequence approximately 8 amino acids long was deduced. The percent incorporation into each segment was calculated by dividing the number of incorporated deuterons by the number of amino acids in the peptide and multiplying by 100. Shown here are the changes in percent deuterium incorporation of mEPG-II a.) caused by the presence of $(GalA)_8$ and b.) caused by the presence of both $(GalA)_8$ and PGIP. * indicate regions in the binding cleft that are protected by the substrate.

Deuterium incorporation into peptides in the cleft area with and without the substrate indicated that the few residues that had been deuterated were protected from exchange by the presence of the oligosaccharide (Fig. 5.3a). It is interesting to note that the distance between the most N- and C-terminal protected residues (Gly_{65} and Tyr_{283}) is similar to the length of the $(GalA)_8$ substrate. Gly_{65} and Tyr_{283} are approximately 38 Å apart in the EPG-II crystal structure suggesting that the substrate lies somewhat linearly along the entire cleft. For the free mEPG-II, deuterium was incorporated somewhat unexpectedly into an α-helix around Asp_{110} (Fig. 5.3a), although it is well documented that like β-pleated sheets, α-helices are slow to incorporate deuterium due to the presence of hydrogen bonding.[44,45] This data implies that in the absence of substrate the α-helix is either very loosely formed or not present at all, in contradiction to the X-ray data. A loose α-helix may become more structured during the crystallization protocol, emphasizing the need for developing techniques complementary to X-ray crystallography. Just as unexpected as the incorporation into this α-helix was the apparent protection of the α-helix by the substrate (Fig. 5.3a). The helix is clearly located on the outside of the β-barrel and should not be directly protected by the substrate. Therefore, it is reasonable to suggest that a conformational change in the mEPG-II upon substrate binding is responsible for the protection. There appear to be two possible causes for the indirect protection by the oligosaccharide; either the α-helix is pulled into the interior of the protein, or the very loose α-helix becomes more structured and the resulting increased level of hydrogen bonding within the helix prevents the exchange.

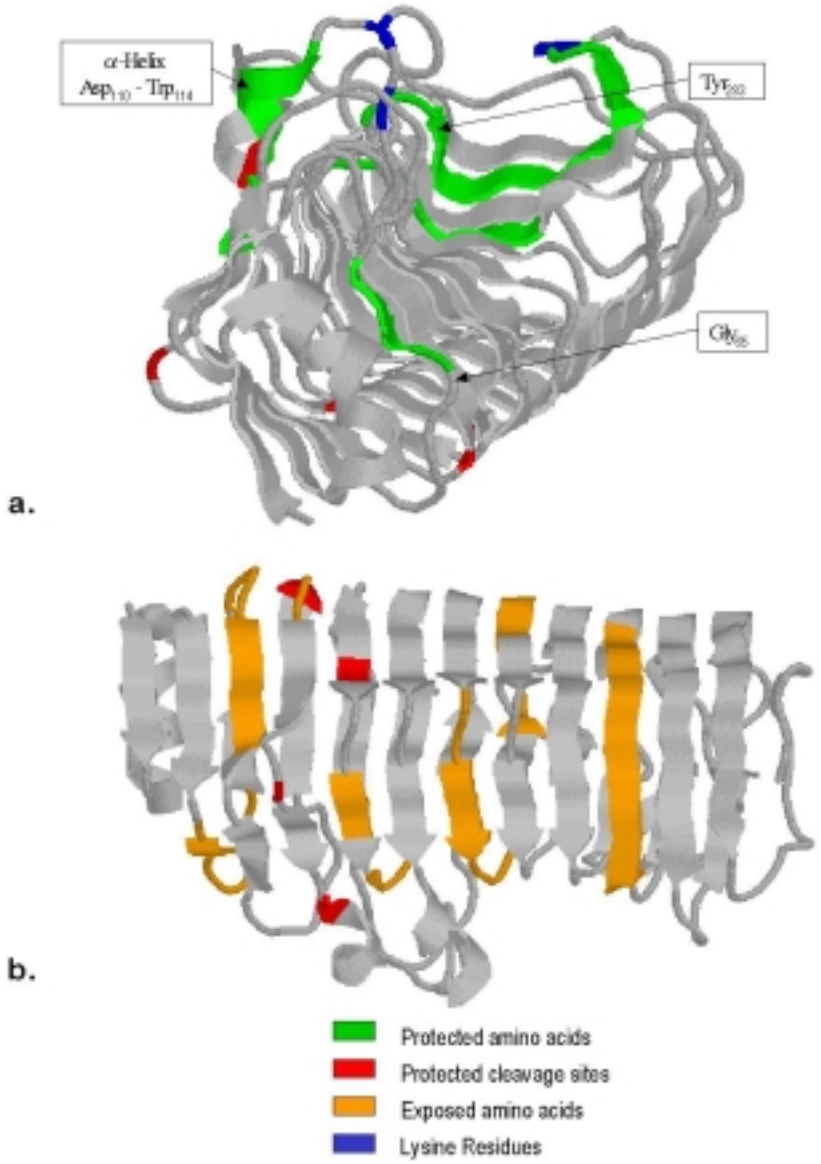

Figure 5.3 The crystal structure of EPG-II from *A. niger* from the (a) top and (b) bottom. Amino acids in the binding cleft that were protected from exchange by substrate binding are shown in green (a). Amino acids of β-pleated sheets that were exposed to exchange due to substrate binding are shown in gold (b). The helix at Asp_{110} is labeled and shown in green (a). The four residues that were protected from pepsin hydrolysis by PGIP binding are shown in red.[19]

Due to the high percentage of hydrogen bonding (β-pleated sheets) within the protein, the majority of peptides identified were from these regions and did not incorporate a detectable level of deuterium. Surprisingly, when the mEPG-II was bound to the substrate, deuterium was incorporated into the β-sheets on the underside of the β-barrel (Fig. 3b) implying a conformational change, most likely the disruption of these sheets. The detection of a local increase in deuterium incorporation as a result of ligand binding far from the region of initial interest (the binding cleft) may point toward a possible mechanism for the activity of EPG-II and also serves to demonstrate the value of the technique.

PDEs often have lysine residues that stabilize the carboxylates on the carbohydrate substrate.[46-50] The α-helix at Asp_{110} to Trp_{114} is directly below a loop that hangs predominantly over the cleft. It should be noted that there is a similar loop on the opposite side of the cleft. There are a total of four lysine residues on the two overhanging loops, and homology mapping has revealed that at least one lysine is conserved within the family of EPGs (Fig. 3a). If these two overhanging loops are required to "lean" into the cleft to allow the lysines to interact with and stabilize the substrate, then much like a spring, the EPG-II may flex the β-sheets on the underside of the barrel to accommodate the necessary movement. Modeling of this system is underway to evaluate this possibility. The flexing may, alternatively, be due to entropic considerations. The complex is at a lower entropy state than the unbound EPG-II if no conformational changes occur, which is unfavorable with respect to Gibb's free energy. This phenomenon has been discussed previously in the literature.[51] NMR relaxation experiments have indicated that while protein flexibility typically decreases with binding

of a ligand, in a few cases it has been noted to increase.[51] Our system seems to fall into this latter class. Major urinary protein-I (MUP-I) from mouse has been shown to increase in flexibility when bound to the pheromone 2-sec-butyl-4,5-dihydrothiazole. The MUP-I, similar to EPG-II, is almost entirely comprised of β-sheets and the increase in flexibility was seen throughout the β-barrel upon binding of the pheromone. It has been suggested that increases in protein backbone flexibility would likely be associated with a corresponding decrease in flexibility of another region.[51] Both of these conditions are observed in the mEPG-II /homogalacturonan complex.

The mEPG-II/homogalacturonan complex was also analyzed by UV-fluorescence. Tryptophan and tyrosine residues can both fluoresce. The fluorescence signal can be quenched by interactions with the solvent, neighboring amino acids, and other prosthetic groups. These various neighbors can allow alternatives to fluorescence as a pathway for relaxation, i.e. internal conversion. Therefore, general changes in protein structure that perturb the microenvironment of fluorescing residues may be detected if a noticeable change in fluorescence intensity is observed.[52,53] While exciting at 290 nm, such that only the tryptophans will fluoresce, an increase of intensity was observed in the presence of the oligosaccharide (Fig. 4). Of the seven tryptophans, five are on the underside far from the binding cleft. The remaining two are on the loop above the D_{110} α-helix. Which tryptophans are most responsible for the change in fluorescence cannot be fully assessed, but the change in fluorescence is consistent with the structural changes defined by the amide exchange data.

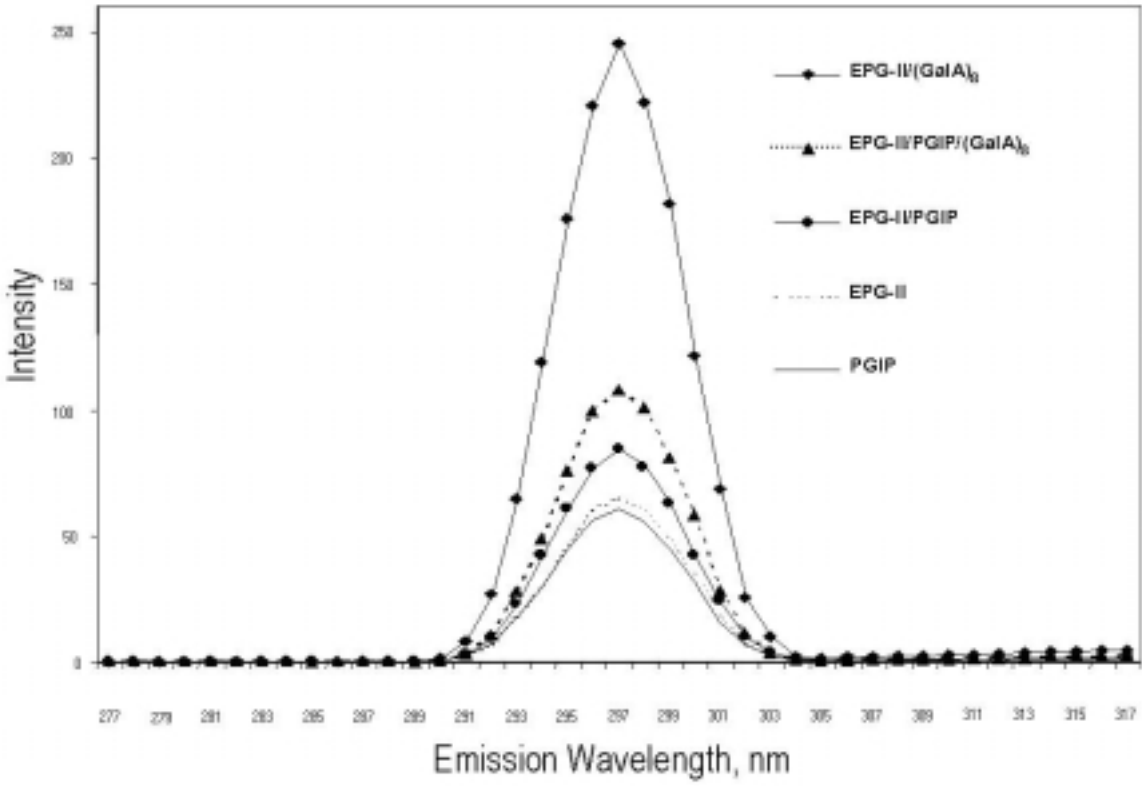

Figure 5.4 Fluorescence spectra of various mEPG-II and PGIP complexes. The emission spectra were excited at 290 nm.

EPG-II/PGIP/Homogalacturonan

The PGIP proved to be highly resistant to proteolysis. The reason for this is unknown, but may be related to the LRR structure of the protein.[15,16] This resistance proved to be extremely advantageous to the experiments described here for two reasons. First, the mass spectra of the mEPG-II peptides were not further complicated by the appearance of PGIP peptides. Secondly, the intact PGIP blocked regions of mEPG-II from pepsin digestion allowing the location on mEPG-II of the mEPG-II /PGIP interaction to be probed by differential peptide mapping. When comparing the mEPG-II peptides, four residues, Glu_{95}, Gly_{104}, Asp_{110}, and Ile_{139} (Fig. 3b), that had consistently been cleavage sites in the mEPG-II and mEPG-II /homogalacturonan samples were now protected from pepsin by the presence of PGIP. Quite conclusively, the four residues lie closely together around the underside of the barrel near the region of the D_{110} α-helix, opposite to the binding site and clearly pointed to the location where the inhibitor was interacting with the mEPG-II. This positioning is perfectly consistent with the other experimental results and with reports of PGIP acting as a non-competitive.[14]

The presence of the PGIP also resulted in a marked change in the pattern of deuterium incorporation in the mEPG-II /substrate complex. As described above, deuterium was not incorporated into the β-sheets in free mEPG-II. In the presence of substrate, the β-sheets are apparently disrupted, allowing deuterium to be incorporated along the underside of the barrel indicated by upward pointing bars in Figure 2(a). When PGIP is added to form the mEPG-II /PGIP/homogalacturonan complex, most of the upward pointing bars are eliminated indicating that the deuterium incorporation into the β-sheets along the underside of the mEPG-II is greatly reduced (Fig. 2b). There are at

least two explanations for this protection from exchange. Because the PGIP is binding within this region, it is certainly directly protecting some of these residues from the solvent. In addition, this interaction may be preventing disruption of β-sheets not directly contacted by the PGIP.

The incorporation around the Asp_{110} α-helix is also significant. In free mEPG-II, deuterium is incorporated into the α-helix implying that the α-helix must be only loosely formed despite the appearance of a tightly wound helix in the crystal stucture. Subsequently, in the presence of substrate no deuterium is incorporated, implying that the α-helix may be forming more tightly during the binding process. In the presence of both substrate and inhibitor, we also find deuterium is not incorporated into the α-helix indicated by a downward bar in Figure 2(b). The inhibitor may be unable to prevent the tightening of the helix in the presence of the substrate, indicating that flexing of the backbone and tightening of the α-helix are not necessarily concerted or even related. Another possibility is that the PGIP is binding near and directly protecting the helix from exchange.

The amino acids in the binding cleft which are protected from exchange by substrate binding remain protected in the presence of the inhibitor providing strong evidence that a stable tertiary system, mEPG-II /PGIP/homogalacturonan, has been formed. For additional verification, fluorescence spectra of mEPG-II /PGIP and mEPG-II /PGIP/homogalacturonan were generated and compared. As shown in Figure 4, the spectra are different, implying that the inhibitor is not simply displacing the substrate. As described above, PGIP has for some hosts been demonstrated to be a non-competitive inhibitor, slowing the hydrolysis process without preventing the binding of substrate.[14] In

contrast to our amide exchange and fluorescence data, a recent article[54] has found evidence for competitive inhibition for the *P. vulgaris* PGIP/*F. moniliforme* EPG pairing. While this makes a general concensus more difficult, it emphasizes the complex nature of these systems and the need for evaluating each enzyme/inhibitor pair independently.

The fluorescence experiment also provided additional evidence for conformational changes within the mEPG-II /PGIP/homogalacturonan system. Figure 4 shows the fluorescence of the free mEPG-II, mEPG-II /homogalacturonan, and mEPG-II /PGIP/homogalacturonan. The fluorescence, as described earlier, increases dramatically with the binding of the substrate. The presence of the PGIP has a remarkable effect on the level of fluorescence, lowering it almost back to the level of free mEPG-II. The fluorescence data implies that the PGIP may be able to prevent at least some of the conformational changes caused by the presence of the substrate. The amide exchange-MS experiments have pointed out two conformational changes that occur within mEPG-II when binding to $(GalA)_8$ that may be important, a disruption of β-sheets in the backbone and the formation of an α-helix. Therefore, an increase in fluorescence intensity corresponds to a change in the structured environment of the tryptophans. The dramatic decrease in fluorescence caused by the presence of the PGIP clearly implicates the involvement of PGIP in changing the tryptophans' environments. The similarity between this fluorescence intensity and that of the free mEPG-II suggests that the inhibitor may be restraining the conformational changes that occurred in the mEPG-II /homogalacturonan system. If the conformational change is required for substrate binding, preventing that change may allow the substrate to diffuse away from the enzyme before hydrolysis of the glycosidic bond can occur, thus effectively inhibiting the activity of the enzyme.

Conclusion

The amide exchange-MS results indicate that the presence of PGIP had a clear effect on the deuterium incorporation as compared to the mEPG-II /homogalacturonan experiments. The most significant change involved the β-sheets on the underside of the barrel structure. In the presence of $(GalA)_8$, these β-sheets incorporated deuterium. β-sheets should not incorporate deuterium because they are involved in hydrogen bonding. So, clearly a conformational change is occuring during substrate binding, specifically the disruption of the β-sheets. In the presence of PGIP, the β-sheets no longer incorporate deuterium, much like the free mEPG-II. The results suggest that PGIP is binding to mEPG-II on the underside of the barrel and below the Asp_{110} α-helix. This location for PGIP interaction is consistent with non-competitive inhibition.

The other region of interest is the α-helix near residue 110. In the free mEPG-II, deuterium was incorporated into a region shown to be an α-helix in the crystal structure. A likely explanation is that the α-helix is not tightly formed in the free system. The deuterium incorporation is not an artifact of the procedure, as in the presence of the substrate, deuterium is no longer incorporated into the helix. Thus, the α-helix may be tightening as part of the function of the EPG-II binding its substrate. The PGIP appears to have no effect on the incorporation of deuterium in the α-helix peptide.

The mechanistic picture that is drawn from these experiments not only demonstrates the value of the amide exchange-MS technique, but the possible complex and specific nature of the EPGs and PGIPs. The apparent flexing of the EPG-II barrel, whether caused by the lysine loops bending into the cleft to interact with the acidic substrate or not, is a pathway that likely decreases the activation energy for

homogalacturonan hydrolysis, and the non-competitive inhibition of this conformational change by PGIP effectively slows the hydrolysis rate by maintaining a higher energy barrier for hydrolysis. Future studies will include other EPG/PGIP pairings. Only by characterizing additional complexes can general conclusions regarding polygalacturonases and polygalacturonase-inhibiting proteins begin to be made.

Acknowledgements

The authors acknowledge and are grateful for financial support of this research provided by the National Institutes of Health (NIH grant number P41RRR05351), the North Atlantic Treaty Organization (NATO grant number CRG973086), the Department of Energy (DOE Center for Plant and Microbial Complex Carbohydrates, DOE grant number DE-FG02-93ER20097, and DOE grant number DE-FG02-96ER20221) and the National Science Foundation (NSF grant MCB-0115132).

References

1. M. A. O'Neill, A. G. Darvill, and P. Albersheim, *Encyclopedia of Life Sciences*, Macmillan Reference Ltd., London, in press (2000).

2. T. M. Jones, A. J. Anderson, and P. Albersheim, *Plant Pathol.* **2**, 153 (1972).

3. A. L. Karr, Jr., and P. Albersheim, *Plant Physiol.* **46**, 69 (1970).

4. D. F. Bateman, and H. G. Basham, *Physiological Plant Pathology*. **4**, 316 (1976).

5. M. G. Hahn, P. Bucheli, F. Cervone, S. H. Doares, R. A. O'Neill, A. Darvill, and P. Albersheim, *Plant-Microbe Interactions. Molecular and Genetic Perspectives*. **3**, 131 (1989).

6. F. Cervone, G. De Lorenzo, G. Salvi, C. Bergmann, M. G. Hahn, Y. Ito, A. Darvill, and P. Albersheim, *Signal Molecules in Plants and Plant-Microbe Interactions* NATO ASI Series, **H36**, 85 (1989).

7. J. K. C. Rose, M. A. O'Neill, P. Albersheim, and A. Darvill, *Oligosaccharides in Chemistry and Biology-A Comprehensive Handbook*, 783 (2000).

8. A. Darvill, C. Bergmann, F. Cervone, G. De Lorenzo, K. Ham, M. D. Spiro, W. S. York, and P. Albersheim, *Bioch. Soc. Symp.* **60**, 89 (1994).

9. F. Cervone, G. De Lorenzo, B. Aracri, D. Bellincampi, C. Caprari, A. Devoto, F. Leckie, B. Mattei, L. Nuss, and G. Salvi, *Biology of Plant-Microbe Interactions*, International Society for Molecular Plant-Microbe Interactions. 93 (1996).

10. B. J. Cook, R. P. Clay, C. W. Bergmann, P. Albersheim, and A. G. Darvill, *Mol. Plant Microbe Interact.* **12**, 703 (1999).

11. F. Cervone, M. G. Hahn, G. De Lorenzo, A. Darvill, and P. Albersheim, *Plant Physiol.* **90**, 542 (1989).

12. A. Di Pietro, and M. I. G. Roncero, *Phytopath*. **86**, 1324 (1996).

13. A. Desiderio, B. Aracri, F. Leckie, B. Mattei, G. Salvi, H. Tigelaar, J. S. C. Van Roekel, D. C. Baulcombe, M. S. Melchers, G. De Lorenzo, and F. Cervone, *Phaseolus vulgaris*, *Mol. Plant-Microbe Interact*. **10**, 852 (1997).

14. H. U. Stotz, J. G. Bishop, C. W. Bergmann, M. Koch, P. Albersheim, A. G. Darvill, and J. M. Labavitch, *Physiol. Mol. Plant Pathol*. **56**, 117 (2000).

15. A. V. Kajava, *J. Mol. Biol*. **277**, 519 (1998).

16. F. Leckie, B. Mattei, C. Capodicasa, A. Hemmings, L. Nuss, B. Aracri, G. De Lorenzo, and F. Cervone, *EMBO J*. **18**, 2352 (1999).

17. H. U. Stotz, J. Bishop, C. W. Bergmann, M. Koch, P. Albersheim, A. G. Darvill, and J. M. Labavitch, *Physiol. Mol. Plant Pathol*. **56**, 117 (1999).

18. R. Pickersgill, D. Smith, K. Worboys, and J. Jenkins, *J. Biol. Chem*. **273**, 24660 (1998).

19. Y. Van Santen, J. A. E. Benen, K. H. Schroer, K. H. Kalk, S. Armand, J. Visser, and B. W. Dijkstra, *J. Biol. Chem*. **274**, 30474 (1999).

20. P. M. Coutinho, and B. Henrissat, *Genetics Biochemistry and Ecology of Cellulose Degradation*. 15 (1999).

21. C. Grassin, and P. Fauquembergue, *Pectins and Pectinases*. 453 (1996).

22. H. P. Heldt-Hansen, L. V. Kofod, G. Budolfsen, P. M. Nielsen, S. Hüttel, and T. Bladt, *Pectins and Pectinases*. 463 (1996).

23. Cooper, R. M. *Biochemical Plant Pathology*. 135 (1995).

24. C. W. Bergmann, B. Cook, A. G. Darvill, P. Albersheim, D. Bellincampi, and C. Caprari, *Pectins and Pectinases*. 275 (1995).

25. B. R. Thakur, R. K. Singh, and A. K. Handa, *Crit. Rev. Food Sci. and Nutri*. **37**, 47 (1997).

26. V. Katta and B. T. Chait, *J. Am. Chem. Soc*. **115**, 6317 (1993).

27. Z. Q. Zhang and D. L. Smith, *Protein Sci*. **2**, 522 (1993).

28. Y. Q. Liu and D. L. Smith, *J. Am. Soc. Mass Spectom*. **5**, 19 (1994).

29. A. Satoko and K. Takio, *Protein Sci*. **9L**, 2497 (2000).

30. A. M. Falick and E. A. Komives, *Proc. Natl. Acad. Sci. USA*. **95**, 14705 (1998).

31. T. B. Farmer and R. M. Caprioli, *J. Mass Spectrom*. **33**, 697 (1998).

32. H. Ehring, *Anal. Biochem*. **267**, 252 (1999).

33. D. L. Smith and Z. Q. Zhang, *Mass Spectrom. Rev*. **13**, 411 (1994).

34. J. B. Smith, Y. Q. Liu, and D. L. Smith, *Exp. Eye Res*. **63**, 125 (1996).

35. S. Armand, M. J. M. Wagemaker, P. Sanchez-Torres, H. C. M. Kester, Y. van Santen, B. W. Dijkstra, J. Visser, and J. A. E. Benen, *J. Biol. Chem*. **275**, 691 (2000).

36. P. Toubart, A. Desiderio, G. Salvi, F. Cervone, L. Daroda, G. De Lorenzo, C. Bergmann, A. G. Darvill, and P. Albersheim, *Plant J*. **2**, 367 (1992).

37. D. King, M. Lumpkin, C. Bergmann, and R. Orlando, *Rapid Commun. Mass Spectrom*. In press (2002).

38. J. G. Mandell, A. Baerga-Ortiz, S. Akashi, K. Takio, and E. A. Komives, *J. Mol. Biol*. **306**, 575 (2001).

39. J. S. Milne, L. Mayne, H. Roder, A. J. Wand, and S. W. Englander, *Protein Sci*. **7**, 739 (1998).

40. J. A. E. Benen, H. C. M. Kester, and J. Visser, *Eur. J. Biochem.* **259**, 577 (1999).

41. Y. Papnikolau, G. Prag, G. Tavlas, C. E. Vorgias, A. B. Oppenheim, and K. Petratos, http://pdb.ccdc.cam.ac.uk/oca-bin/ccpeek?id=1EHN (2000).

42. R. D. Scavetta, S. R. Herron, A. T. Hotchkiss, N. Kita, N. T. Keen, J. A. Benen, H. C. Kester, J. Visser, and F. Jurnak, *Plant Cell* **11**, 1081 (1999).

43. L. Wang, L. C. Lane, and D. L. Smith, *Prot. Sci*. **10**, 1234 (2001).

44. T. M. Raschke and S. Marqusee, *Curr Opin Biotechnol*. **9**, 80 (1998).

45. M. F. Jeng, W. Englander, G. A. Elove, A. J. Wand, and H. Roder, *Biochem*. **29**, 10433 (1990).

46. H. U. Stotz, C. W. Bergmann, A. L. T. Powell, J. J. Contos, P. Albersheim, A. G. Darvill, and J. M. Labavitch, *7th International Symposium on Molecular Plant-Microbe Interactions*, Edinburgh, UK (1994).

47. J. R. Fromm, R. E. Hileman, E. E. Caldwell, J. M. Weilere, and R. J. Linhardt, *Arch. Biochem. Biophys.* **343**, 92 (1997).

48. J. R. Fromm, R. E. Hileman, E. E. O. Caldwell, J. M. Weiler, and R. J. Linhardt, *Arch. Biochem. Biophys*. **323**, 279 (1995).

49. F. A. Quiocho, *Ann. Rev. Biochem*. **55**, 287 (1986).

50. V. Monchois, R. M. Willemot, and P. Monsan, *FEMS Microbiol. Rev*. **23**, 131 (1999).

51. L. Zidek, M. V. Novotny, and M. J. Stone, *Natutre Struct. Biol*. **6**, 12, 1118 (1999).

52. N. M. Nichols and K. S. Matthews, *Biochem. Biophys. Res. Commun*. **288**, 111 (2000).

53. H. H. Willard, L. L. Merritt, Jr., J. A. Dean, and F. A. Settle, *Instrumental Methods of Analysis*, Wadsworth Publishing Company, 197 (1988).

54. L. Federici, C. Caprari, B. Mattei, C. Savino, A. Di Matteo, G. De Lorenzo, F. Cervone, and D. Tsernoglou, *Proc. Natl. Acad. Sci. USA*. **98**, 13425 (2001).

CHAPTER 6

CONCLUSION

The research presented here demonstrates the development of mass spectrometric procedures which increase the speed of protein analysis. Chapters 2 and 3 describe the successful implementation of on-probe protein digestions which are much faster than typical proteolytic digestions. These digestion procedures are also compatible with various MALDI surface chemistries. Chapters 4 and 5 display the use of amide exchange-MS as a complement to NMR and crystallography. Amide exchange-MS is shown to be rapid and robust with respect to carbohydrate samples which are typically difficult for NMR and crystallography to work with.

On-probe protein digestion

Chapter 2:

Pepsin is shown to be a quick and robust enzyme, capable of digesting various proteins that have been previously dried on a MALDI probe surface. Pepsin, being a stomach enzyme, is optimally active in acidic environments, similar to the pH of typical MALDI matrices. The experimental results demonstrate that pepsin is capable of digesting proteins in the presence of MALDI matrices. This property allows for multiple digestions to be performed on-probe. Additionally, pepsin seems to regain its activity after being dried and analyzed simply by the addition of water. Therefore, digestions can be monitored periodically during a digestion.

Chapter 3:

Preliminary experiments show that vapor phase acid hydrolysis and liquid phase acid hydrolysis have the potential for digesting proteins dried on a MALDI probe, yielding either peptide fragments or sequence ladders. Several parameters must be optimized for each protein that is to be analyzed. The percent acid, length of time for the

digestion, and temperature for the digestion can be varied depending upon which type of spectra is desired (peptides or ladders). Initially, no general relationship was discovered between the parameters and the specific proteins analyzed. Further work to lower the sample amount needed is described in Appendix A. This work included the successful use of TFA instead of PFPA because it is much better for MALDI signal.

Amide exchange-MS

Chapter 4:

For the first time amide exchange-MS has been used to study the interaction of a protein and carbohydrate. Initially, the carbohydrate substrates of lysozyme were able to protect a portion of the protein from incorporating deuterium. Additionally, the magnitude of the protection was directly proportional to the size of the substrate. Subsequently, the procedure was successfully applied to a fungal enzyme, EPG-II, and its carbohydrate substrate, PGA. Some protection was observed in the same region as was indicated to be the active-site for hydrolysis by site-directed mutagenesis. The technique is demonstrated to be a viable alternative to NMR or crystallography when working with small sample quantities of carbohydrate substrates.

Chapter 5:

Amide exchange-MS data provided not only binding site information, but the detection of conformational changes that are associated with enzyme-substrate binding within the EPG-II/PGA system. The β-sheets along the underside of the β-barrel within EPG-II appear to be disrupted during carbohydrate binding. Subsequently, this conformation change is reduced in the presence of PGIP suggesting a mechanism for non-competitive inhibition. Additionally, changes in the digestion pattern of EPG-II when in

the presence of PGIP indicate the location for PGIP binding, along the underside of the barrel, further suggesting non-competitive inhibition.