Protein Sequencing Using a Combination of Mass Spectrometric Methods.

by

Vladimir V. Papov, Jr.

B.A., Connecticut College

(1987)

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy at the

Massachusetts Institute of Technology

February 1996

© 1996 Massachusetts Institute of Technology All rights reserved

Signature of Author ..................................................

Department of Chemistry, 23 January 1996

Certified by ..............................................................

Klaus Biemann
Thesis Supervisor

Accepted by ..............................................................

Dietmar Seyferth
Chairman, Departmental Committee on Graduate Students

MAR 04 1996 Science
This doctoral thesis has been examined by a Committee of the Department of Chemistry as follows:

Professor Alexander M. Klibanov ......................................................

Chairman

Professor Klaus Biemann .....................................................

Thesis Supervisor

Professor Peter T. Lansbury, Jr. ..............................................
Protein Sequencing Using a Combination of Mass Spectrometric Methods.

by

Vladimir V. Papov, Jr.

Submitted to the Department of Chemistry on January 23, 1996 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Chemistry

ABSTRACT

The complete primary structure of a thioltransferase isolated from human erythrocytes was determined using a combination of matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) and fast atom bombardment (FAB) high performance tandem MS. The amino acid sequence of this human glutaredoxin (hErGrx) was assembled from tryptic and chymotryptic peptides whose sequences were deduced from collision-induced-dissociation (CID) mass spectra.

An adhesive protein, mefp-3, isolated from the foot of the marine mussel *M. edulis*, presented unique challenges to the determination of the amino acid sequence due to its unusual composition. The complete amino acid sequence of mefp-3 was finally determined by the application of mass spectrometry to this problem. Specifically, mefp-3 peptides were digested with carboxypeptidase P and analyzed by MALDI-TOF-MS to obtain the C-terminal sequences. The combined approach of Edman degradation and MALDI-TOF-MS serves as an example of the complementarity of mass spectrometry with more traditional sequencing methods. In addition, mefp-3 contains a basic amino acid which was determined to be 4-hydroxyarginine by FAB tandem mass spectrometry.

*E. coli* inosine monophosphate dehydrogenase (IMPDH) inactivated with 5-ethynyl-1-β-D-ribofuranosylimidazole-4-carboxamide monophosphate (EICARMP) forms a covalent adduct with Cys-305 as determined by MALDI-TOF-MS. Sequence corrections, identified and confirmed by MALDI-TOF-MS, showed that the native protein consists of 488 amino acids, has 5 Cys, and that Arg is at position 206.

A method for specific, partial cleavage of proteins at locations occupied by Cys was developed. Reaction with cyanide ion and 4-vinylpyridine generates overlapping sequences from proteins containing disulfide bonds. The addition of 4-vinylpyridine was shown to drive the reaction by trapping the free sulfhydryl groups formed by cyanide attack of the disulfide bond.

Thesis Supervisor: Klaus Biemann
Title: Professor of Chemistry
To Carol
and my parents
# Table of Contents

<table>
<thead>
<tr>
<th>Section I.</th>
<th>Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A Brief History of Mass Spectrometry of Peptides and Proteins</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Fast Atom Bombardment Mass Spectrometry</td>
<td>20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section II.</th>
<th>Amino Acid Sequencing by Mass Spectrometry</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introduction</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>FAB Tandem Mass Spectrometry</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Reflectron and Post-Source-Decay Mass Spectrometry</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>MALDI-TOF-MS and &quot;Ladder&quot; Sequencing</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Summary of Applications Discussed in Sections III - VI</td>
<td>52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section III.</th>
<th>Determination by Mass Spectrometry of the Primary Structure of the Glutaredoxin Isolated from Human Erythrocytes</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introduction</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Experimental Procedures</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Results and Discussion</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Section IV.</th>
<th>Determination of the Primary Structure of a Mussel Adhesive Protein from <em>Mytilus edulis</em> and Identification of a Novel Amino Acid</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introduction</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Experimental Procedures</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Structure Determination of a Modified Arginine</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>Amino Acid Sequence of Mefp-3F</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>Conclusion</td>
<td>105</td>
</tr>
</tbody>
</table>

| Section V. | | Page |
|------------| | |
Section V.  The Determination of the Active Site of Inosine Monophosphate Dehydrogenase of *Escherichia coli* and Correction of the Amino Acid Sequence of this Enzyme ........................................ 129

V.1 Introduction ................................................. 129
V.2 Experimental Procedures ..................................... 131
V.3 Identification of the Active Site Modification ................... 135
V.4 Correction of the Amino Acid Sequence .......................... 138
V.5 Conclusion ................................................... 149

Section VI.  Generation of Peptides with Overlapping Sequences using Cleavage at Cysteine with Cyanide Ion in Combination with 4-Vinylpyridine .... 151

VI.1 Introduction .................................................. 151
VI.2 Specific Cleavage of Proteins by S-Cyanylation ...................... 152
VI.3 A New Strategy for the Optimization of the Cyanide Cleavage Reaction to Generate Overlapping Peptide Fragments .................. 158
VI.4 Experimental Procedures ...................................... 161
VI.5 Results and Discussion ....................................... 163
VI.6 Conclusion ................................................... 195

Section VII.  References ............................................. 200

Section VIII. Appendices ............................................ 218

Section IX.  Biographical note, publications and abstracts ..................... 222

Section X.  Acknowledgements ....................................... 225
Table of Figures

Section I.

Figure I.1  Schematic of a FAB ion source illustrating the FAB ionization process ................................................. 21
Figure I.2  Structures of some common MALDI matrix compounds used for the analysis of proteins and peptides ....................... 27
Figure I.3  Schematic of the VT 2000 MALDI-TOF mass spectrometer ............ 31

Section II.

Figure II.1  Schematic diagram of the JEOL HX110/HX110 high performance tandem mass spectrometer ...................... 38
Figure II.2  Principle of protein sequencing by tandem mass spectrometry ........ 39
Figure II.3  Schematic diagram of the Voyager Elite reflectron mass spectrometer ................................................... 45
Figure II.4  Principle of MALDI-PSD mass spectrum acquisition ................... 48
Figure II.5  MALDI-PSD mass spectrum of human angiotensin I (M+H)^+ = m/z 1296.7 ..................................................... 49
Figure II.6  MALDI-TOF mass spectrum of human renin substrate after 15 min with CpY ............................................. 55

Section III.

Figure III.1  MALDI-TOF mass spectra of hErGrx ........................................ 68
Figure III.2  FAB CID mass spectra of hErGrx peptides (M+H)^+ = m/z 1037.3 and m/z 1356.7 .................................. 75
Figure III.3  FAB CID mass spectra of hErGrx peptides (M+H)^+ = m/z 1127.8, m/z 1164.6, and m/z 1395.9 .......................... 76
| Figure III.4 | FAB CID mass spectra of hErGrx peptides (M+H)^+ = m/z 629.4 and m/z 672.5 | 77 |
| Figure III.5 | Amino acid sequence of hErGrx determined by tandem mass spectrometry | 78 |
| Figure III.6 | Comparison of pig liver, calf thymus, rabbit bone marrow, and human erythrocyte glutaredoxin amino acid sequences showing their alignment and homology | 79 |
| Figure III.7 | FAB CID mass spectrum of the N-terminal tryptic peptide from hErGrx (M+H)^+ = m/z 1085.7 | 81 |
| Figure III.8 | FAB CID mass spectrum of the His containing peptide from hErGrx (M+H)^+ = m/z 856.3 | 83 |
| Figure III.9 | FAB CID mass spectrum of the hErGrx peptide (M+H)^+ = m/z 1559.0 | 84 |
| Figure III.10 | MALDI-TOF mass spectra of S-ethylpyridylated hErGrx proteolytic digests | 89 |
| Figure III.11 | FAB CID mass spectrum of tryptic peptides (M+H)^+ = m/z 1001.2 and m/z 1422.3 | 92 |

Section IV.

| Figure IV.1 | RP-HPLC of dialysis precipitated mefp-3 | 100 |
| Figure IV.2 | RP-HPLC chromatograms of mefp-3F digested with Endo-Lys-C for 4 hrs | 101 |
| Figure IV.3 | FAB-CID mass spectra of Arg and ^15^N-Arg | 106 |
| Figure IV.4 | Structures of the fragment ions from the FAB-CID mass spectra of Arg and ^15^N-Arg | 107 |
| Figure IV.5 | FAB-CID mass spectra of Arg-OH from mefp-3 and authentic 4-hydroxyarginine | 108 |
| Figure IV.6 | Fragment ions of Arg and mefp-3 derived Arg-OH from which the structure of the latter was elucidated | 109 |
Figure IV.7  MALDI-TOF mass spectrum of the major component of mefp-3F  

Figure IV.8  MALDI-TOF mass spectrum of an aliquot from an Endo-Lys-C 
   digest of mefp-3F  

Figure IV.9  FAB-CID mass spectra of (M+H)$^+$ = m/z 384.1 and the dipeptide 
   YW  

Figure IV.10  FAB-CID mass spectrum of peptide GWNNGWNR from mefp-3F  

Figure IV.11  MALDI-TOF mass spectra of mefp-3F peptide (M+H)$^+$ = m/z 1606.7 
   incubated with AmM at various timepoints  

Figure IV.12  MALDI-TOF mass spectra of mefp-3F peptide (M+H)$^+$ = m/z 1606.7 
   incubated with CpP at various timepoints  

Figure IV.13  MALDI-TOF mass spectra of mefp-3F peptide (M+H)$^+$ = m/z 4182.9  
   incubated with CpP at various timepoints  

Figure IV.14  MALDI-TOF mass spectra of mefp-3F peptide (M+H)$^+$ = m/z 4182.9  
   digested with CpP after longer incubation times  

Figure IV.15  MALDI-TOF mass spectrum of an Endo-Lys-C peptide (HPLC 
   fraction 3) after 30 min  

Section V. 

Figure V.1  Possible mechanisms for the IMPDH catalyzed conversion of 
   IMP to XMP  

Figure V.2  MALDI-TOF mass spectra of E. coli IMPDH  

Figure V.3  MALDI-TOF mass spectra of Endo-Lys-C digests of 
   S-ethylpyridylated E. coli IMPDH  

Figure V.4  MALDI-TOF mass spectrum of a tryptic digest of S- 
   ethylpyridylated E. coli EICARMP inactivated IMPDH  

Figure V.5  MALDI-PSD mass spectrum of tryptic peptide (M+H)$^+$ = 
   m/z 1509.0  

9
Figure V.6  MALDI-PSD mass spectrum of Endo-Lys-C peptide
(M+H)$^+$ = m/z 731.4 ............................................... 147

Figure V.7  MALDI-PSD mass spectrum of Endo-Lys-C peptide
(M+H)$^+$ = m/z 1121.9 ........................................ 148

Section VI.

Figure VI.1  The reaction of cyanide with a protein containing a disulfide bond .... 154
Figure VI.2  Non-native disulfide bond formation due to the reversibility of
S-cyanylation with cyanide ........................................ 156
Figure VI.3  New strategy for optimization of the S-cyanylation step with cyanide
by concomitant formation of S-ethylpyridylated (EP) peptides ........... 160
Figure VI.4  MALDI-TOF mass spectra of reaction mixtures containing aprotinin
after two hours. A. NaCN only. B. NaCN and 4-vinylpyridine ........... 165
Figure VI.5  Primary structure of aprotinin showing disulfide bond structure ...... 166
Figure VI.6  MALDI-TOF mass spectrum of aprotinin digested with NaCN and
4-vinylpyridine after three hours .................................. 167
Figure VI.7  Graphical representation of peptides resulting from digestion
of aprotinin with NaCN and 4-vinylpyridine
showing overlapping sequences ........................................ 168
Figure VI.8  The primary structure of \textit{E. coli} thioredoxin indicating disulfide
bond location and peptides expected from a digest with cyanide
and 4-vinylpyridine ....................................................... 175
Figure VI.9  MALDI-TOF mass spectra of reaction mixtures containing \textit{E. coli}
thioredoxin and NaCN after one hour. Ten-fold (A) and 100-fold (B)
molar excesses of 4-vinylpyridine, respectively .......................... 176
Figure VI.10 MALDI-TOF mass spectra of ribonuclease A digested with NaCN
and 4-vinylpyridine after two hours. A. pH 6.5. B. pH 8.5 ............... 180
Figure VI.11 Graphical representation of peptides resulting from digestion of ribonuclease A with NaCN (pH 8.5) and 4-vinylpyridine showing overlapping sequences ........................................ 181

Figure VI.12 Graphical representation of peptides resulting from digestion of lysozyme with NaCN (pH 8.5) and 4-vinylpyridine showing overlapping sequences ........................................ 187

Figure VI.13 Graphical representation of peptides resulting from digestion of trypsinogen with NaCN (pH 8.5) and 4-vinylpyridine showing overlapping sequences ........................................ 189

Figure VI.14 Urotensin II digested with NaCN and 4-vinylpyridine after two hours. A. MALDI-TOF mass spectrum. B. HPLC chromatogram ........ 191

Figure VI.15 PSD mass spectra of peptides derived from urotensin II digestion with NaCN and 4-vinylpyridine. A. (M+H)$^+$ = m/z 1267.5. B. (M+H)$^-$ = m/z 1249.5 ........................................... 193

Figure VI.16 PSD mass spectrum of a peptide containing an N-terminal iminothiazolidine group, (M+H)$^+$ = m/z 1078.5 ...................... 196
Table of Tables

Section III.

Table III.1 FABMS data for peptides resulting from the tryptic digestion (and further digestion with Endo-Asp-N) of S-carbamidomethylated hErGrx ........................................ 72
Table III.2 FABMS data for peptides resulting from the chymotryptic digestion of S-ethylpyridylated hErGrx ........................................ 73
Table III.3 FABMS data for peptides resulting from the Endo-Lys-C digestion of S-ethylpyridylated hErGrx ........................................ 74
Table III.4 MALDI-TOF mass spectral data from Endo-Lys-C digestion of S-ethylpyridylated hErGrx from Figure III.10 .......................... 90
Table III.5 MALDI-TOF mass spectral data from Endo-Asp-N digestion of S-ethylpyridylated hErGrx from Figure II.10 .......................... 91

Section IV.

Table IV.1 Peptides from AmM digestion of Endo-Lys-C peptide (M+H)^+ = m/z 1606.7 .............................................. 119
Table IV.2 Peptides from CpP digestion of Endo-Lys-C peptide (M+H)^+ = m/z 1606.7 .............................................. 119
Table IV.3 Peptides from CpP digestion of Endo-Lys-C peptide (M+H)^+ = m/z 4182.9 ............................................ 123
Table IV.4 Peptides from further chymotryptic digestion (Figure IV.15) of the Endo-Lys-C peptide in HPLC Fraction 3 ........................ 125
Table IV.5 Peptides derived from Endo-Lys-C digestion of mefp-3F ............ 126
Section V.

Table V.1 Peptides resulting from Endo-Lys-C digestion of *E. coli* native and EICARMP inactivated IMPDH ........................................ 140

Table V.2 Peptides resulting from tryptic digestion of EICARMP inactivated *E. coli* IMPDH .................................................. 143

Section VI.

Table VI.1 Peptides resulting from the digestion of aprotinin with NaCN and 4-vinylpyridine .......................................................... 169

Table VI.2 Peptides resulting from the digestion of ribonuclease A with NaCN and 4-vinylpyridine at pH 8.5 for 2 hours ................................ 182

Table VI.3 Peptides resulting from the digestion of lysozyme with NaCN and 4-vinylpyridine at pH 8.5 for 2 hours ............................... 186

Table VI.4 Peptides resulting from the digestion of trypsinogen with NaCN and 4-vinylpyridine at pH 8.5 for 2 hours ............................... 188

Table VI.5 HPLC fractionated peptides resulting from digestion of urotensin II with NaCN and 4-vinylpyridine after two hours .............. 192
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Acronym</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACN</td>
<td></td>
<td>Acetonitrile</td>
</tr>
<tr>
<td>AmM</td>
<td></td>
<td>Aminopeptidase M</td>
</tr>
<tr>
<td>Arg-OH</td>
<td></td>
<td>4-hydroxyarginine</td>
</tr>
<tr>
<td>CAD</td>
<td></td>
<td>Collisionally activated dissociation (decomposition)</td>
</tr>
<tr>
<td>CID</td>
<td></td>
<td>Collisionally induced dissociation (decomposition)</td>
</tr>
<tr>
<td>CpX</td>
<td></td>
<td>Carboxypeptidase X; where X is A, B, Y, or P</td>
</tr>
<tr>
<td>DE</td>
<td></td>
<td>Delayed extraction</td>
</tr>
<tr>
<td>Dopa</td>
<td></td>
<td>3,4-dihydroxyphenylalanine</td>
</tr>
<tr>
<td>DTT</td>
<td></td>
<td>1,4-Dithiothreitol</td>
</tr>
<tr>
<td>DTE</td>
<td></td>
<td>1,4-Dithioerythritol</td>
</tr>
<tr>
<td>EDTA</td>
<td></td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EICAR</td>
<td></td>
<td>5-Ethynyl-1-β-D-ribofuranosylimidazole-4-carboxamide</td>
</tr>
<tr>
<td>EICARMP</td>
<td></td>
<td>EICAR-5'-monophosphate</td>
</tr>
<tr>
<td>EIMS</td>
<td></td>
<td>Electron ionization mass spectrometry</td>
</tr>
<tr>
<td>ESI</td>
<td></td>
<td>Electrospray ionization</td>
</tr>
<tr>
<td>FAB</td>
<td></td>
<td>Fast atom bombardment</td>
</tr>
<tr>
<td>Grx</td>
<td></td>
<td>Glutaredoxin</td>
</tr>
<tr>
<td>GSH</td>
<td></td>
<td>Glutathione - SH</td>
</tr>
<tr>
<td>GTP</td>
<td></td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>hErGrx</td>
<td></td>
<td>Human erythrocyte glutaredoxin</td>
</tr>
<tr>
<td>Hb</td>
<td></td>
<td>Hemoglobin</td>
</tr>
<tr>
<td>HEDS</td>
<td></td>
<td>Hydroxyethyldisulfide</td>
</tr>
<tr>
<td>HFBA</td>
<td></td>
<td>Heptafluorobutyric acid</td>
</tr>
<tr>
<td>HPLC</td>
<td></td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>IMP</td>
<td></td>
<td>Inosine 5'-monophosphate</td>
</tr>
<tr>
<td>IMPDH</td>
<td></td>
<td>Inosine 5'-monophosphate dehydrogenase</td>
</tr>
</tbody>
</table>
FABMS  Fast atom bombardment mass spectrometry
LSIMS  Liquid secondary ionization mass spectrometry
(M+H)+  Protonated molecular ion
MALDI  Matrix-assisted laser desorption ionization
MCP    Microchannel plate
Mefp   Mytilis edulis foot protein
MeOH   Methanol
MS     Mass spectrometry
MS-1/MS-2  First and second mass spectrometer in a tandem mass spectrometer
MS/MS  Tandem mass spectrometry
m/z    Mass to charge ratio (z is assumed to be +1 unless otherwise stated)
NTCB   2-Nitro-5-thiocyanobenzoic acid
PIC    Phenylisocyanate
PITC   Phenylisothiocyanate
PSD    Post-source-decay
PTH    Phenylthiohydantoin
S/N    Signal to noise ratio
RP-HPLC Reversed-phase-HPLC
TFA    Trifluoroacetic acid
TFEITC Trifluoroethylisothiocyanate
TFETH  Trifluoroethylhydantoin
TOF    Time-of-Flight
Tris-HCl Tris (hydroxymethyl) aminomethane-HCl
Trx    Thioredoxin
XMP    Xanthosine 5'-monophosphate
Section I

Introduction

I.1 A Brief History of Mass Spectrometry of Peptides and Proteins

Ionization of a given analyte, whether it be a small organic molecule or a large protein such as serum albumin, is a prerequisite to mass spectrometric analysis. This basic property of mass spectrometry, independent of the type of instrument, confined early protein sequencing efforts to the analysis of small peptides. Electron ionization mass spectrometry (EIMS), probably the most common method of ionization in the earlier days of mass spectrometry, worked extremely well for the analysis of small non-polar and generally more volatile organic compounds. On the other hand, EIMS of even the smallest peptides was a challenge because of the polar (and hence, non-volatile) nature of the peptide structure. One of the first successful strategies for enhancing peptide ionization involved increasing peptide volatility through chemical derivatization. Another innovation was the use of gas chromatography to separate the derivatized peptides and deliver them into a mass spectrometer to determine the mass to charge ratio (m/z) of the molecular and fragment ions. The amino acid sequence of these small peptides could then be deduced by analysis of the fragments generated.

The discovery and development of new ionization methods played a crucial role in
the maturation of mass spectrometry as a technique for the analysis of proteins and peptides. With EI, a so-called "hard" ionization method, high energy deposition allowed the detection of only the most stable molecular ions. New desorption ionization techniques such as $^{252}$Cf plasma desorption and field desorption, enjoyed early success due to the "softer" ionization processes that produced predominantly protonated molecules, $(M+H)^+$, which are very stable even when the molecules are large. In addition, the derivatization necessary to increase peptide volatility for EIMS was no longer required. While field desorption is rarely employed today, $^{252}$Cf plasma desorption instruments still enjoy routine use in many biochemistry laboratories to this day largely due to the fact that they were the first instruments capable of the generation and detection of very high mass molecular ions.

In 1981, a new ion source and ionization method were described by Barber et al. which allowed for the first time the facile ionization of very polar, thermally labile, and nonvolatile molecules without any prior derivatization. The procedure involved dissolution of the analyte in a liquid matrix that was then bombarded with energetic (fast) atoms giving rise to protonated molecular ions with little or no fragmentation. The discovery of fast atom bombardment (FAB) ionization ushered in a new era in the mass spectrometry

---

*The terms liquid secondary ionization mass spectrometry (LSIMS) and FAB mass spectrometry (FABMS) are both found in the literature and may cause some confusion for the uninitiated reader. The latter term simply refers to the ionization achieved by bombardment with neutral atoms while ions (usually metal cations) are used in LSIMS. The fast atoms or ions are accelerated to achieve a translational energy on the order of eV to several keV. Whether analyte ionization is achieved by neutral atoms or ions, the mechanism of ionization is almost certainly the same. Furthermore, the mass spectra arising from use of either method are essentially*
spectrometry of peptides and proteins (as well as other biomaterials and involatile molecules). Large peptides and even small proteins such as insulin were now capable of being ionized without derivatization and with little or no fragmentation. FAB ionization also allowed the development of protein sequencing methodology through the use of collisionally activated dissociation (discussed further in Section 11.2) using different types of mass spectrometers. The success of FAB as an ionization technique is demonstrated by its widespread implementation since the early 1980s and the resultant unprecedented growth in the volume of biological research using mass spectrometry for peptides and proteins that still continues to this day.

Although FAB is still widely used, the more recent development of two additional "soft" ionization techniques has revolutionized the field of biological mass spectrometry even further. Electrospray as an ionization method for mass spectrometry was first described in 1984 by Fenn and Yamashita. No doubt the most interesting characteristic of electrospray ionization (ESI) is its ability to generate numerous multiply charged ions. Due to the complexities involved with interfacing a liquid spray to a high vacuum system, the first mass spectrometers to be used with ESI were the simpler quadrupole instruments that have only a limited mass range. Fortuitously, with ESI the ion signal due to large species (such as proteins) is divided over multiple \((M+nH)^n\) ions, most of which are identical. It is thus the practice of this laboratory to refer to both techniques simply as FAB for simplicity and in deference to the original discoverers.
observed at m/z values under 2000 (which is within the mass range of quadrupole mass spectrometers). ESI has also been successfully coupled with high performance magnetic sector instruments to increase the resolution and mass range.\textsuperscript{15-18} However, the design of ion sources to interface ESI with the high vacuum systems and high accelerating voltages such as those found in magnetic sector instruments is quite a challenge and consequently, still in the process of refinement. This is demonstrated by the fact that most of the work with ESI and magnetic sector instruments to date has dealt with fundamental studies using known compounds.

The remaining "soft" ionization method, matrix-assisted-laser-desorption-ionization (MALDI), was somewhat serendipitously discovered by Karas and Hillenkamp in 1987.\textsuperscript{19} While experimenting with simple laser desorption, these workers made the crucial observation that very large molecules could be ionized without fragmentation using ultraviolet (UV) laser desorption by introducing a small UV absorbing molecule as a matrix. MALDI differs from ESI in that the MALDI experiment is generally performed with a crystalline sample preparation (rather than a liquid spray at atmospheric pressure) which facilitates operation at high vacuum. Furthermore, the best results are obtained with pulsed lasers, the nature of which is particularly amenable to the use of time-of-flight mass spectrometers which (in principle) have an unlimited mass range. MALDI is also capable of generating multiply charged ions but not to the extent that ESI does. Finally and most importantly, ESI and MALDI have in common an increase of several orders of magnitude
for both detectable mass range and sensitivity as compared to FAB ionization. This large increase in sensitivity and mass range has rapidly placed MALDI and ESI methods in the forefront of mass spectrometric research with proteins and peptides.

I.2 Fast Atom Bombardment Mass Spectrometry

Ionization of peptides or proteins by FAB involves dissolution of the analyte in a non-volatile liquid matrix such as glycerol and bombarding the liquid with a beam of neutral or charged molecules (see footnote on page 17).\textsuperscript{7,20,21} Figure I.1 is a schematic drawing illustrating the FAB ionization process using a beam of Cs\textsuperscript{+} ions (which have been determined to increase the sensitivity as compared to bombardment with neutral atoms such as those from gaseous Xe).\textsuperscript{22,23} As shown in Figure I.1, both positively and negatively charged analyte and matrix ions as well as uncharged neutral molecules are desorbed from the liquid as a result of the bombardment. The precise mechanism by which an analyte is ionized by FAB is not thoroughly understood; the liquid matrix apparently absorbs most of the energy imparted by the fast atoms and only a portion of the introduced energy is absorbed by the analyte molecules. The result is the production of intact molecular ions
Figure 1.1. Schematic of a FAB ion source illustrating the FAB ionization process. Ion slits and focussing lenses are not shown for clarity and the drawing is not to scale. The fast atom (Cs') ion beam is represented by the dark circles. The desorbed secondary ion beam consists of positive, negative, and matrix ions denoted by "+", "-", and "M", respectively. Neutrals are also produced and are indicated by "N".
(with the addition or removal of a proton)\(^a\) with almost no fragmentation. In positive ion mode FAB, only the positively charged analyte (and matrix) ions are accelerated out of the ion source and into the mass analyzer. Although matrix ions are detected along with the peptide ions, they rarely interfere with the peptide analysis since the signals due to the matrix ions are strongest only at the lower \(m/z\) values.

The effect of the liquid matrix has been investigated by a number of laboratories.\(^{24-27}\)

For peptides, glycerol with the addition of an acid such as acetic acid or trifluoroacetic acid (TFA) is usually the first choice of matrix because of its high viscosity and low volatility. Thioglycerol alone or mixed with glycerol also tends to work slightly better with increasing size and hydrophobic character of the peptide. Although seldom used in this laboratory, a eutectic mixture of dithiothreitol and dithioerythritol 5:1 (a.k.a. "magic bullet") has also been shown to increase the ion yield of larger peptides.\(^{28}\) The drawback to using the thiol matrices is that they are more volatile and thus evaporate from the ion source more readily which reduces the time available for acquiring a mass spectrum. This can be particularly problematic with tandem mass spectrometry where the signal to noise (S/N) is routinely enhanced through the summation of several individual scans. The instrument used for FABMS as well as protein sequencing using FAB and tandem mass spectrometry is

\(^a\)Both positive and negative FABMS, [with detection of (M+H)\(^+\) and (M-H)\(^-\) ions, respectively], have been shown to successfully generate peptide molecular ions. However, positive ion FAB is more sensitive for peptide ions (probably because peptides contain many sites that are easily protonated) by at least an order of magnitude and consequently, the overwhelming majority of work in this area has been done using positive ion mode FAB.
discussed further in Section II.2.

The best FABMS results are obtained with purified peptides since contaminants, particularly inorganic salt cations, can severely reduce the ion yield. Metal cations, for example sodium and potassium, also decrease ion yield in part by distributing the total ion current among several ions, e.g., \((M+H)^+\), \((M+Na)^+\), and \((M+K)^+\), rather than just the protonated species. For this reason, peptides destined for FABMS analysis are almost always purified with reversed phase high performance liquid chromatography (RP-HPLC) to remove the bulk of the contaminants. HPLC is also useful for the partial fractionation of peptide mixtures, (e.g., a protein digest), especially for the separation of hydrophilic from hydrophobic peptides. Since peptides containing hydrophobic groups tend to stay closer to the surface of the liquid matrix, they are more easily ionized by the FAB ion beam. In some cases, this may result in the suppression of the ion signal of more hydrophilic peptides. A partial fractionation of a protein digest followed by FABMS analysis of the resultant fractions alleviates this problem.

1.3 Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry

The use of laser desorption in mass spectrometry dates back to the early 1960s and is very similar conceptually to FAB ionization (Figure I.1) except that the analyte is
bombarded directly (no matrix was used initially) with photons instead of fast atoms from an ion gun.\textsuperscript{30-32} Similar to FAB ionization, an advantage of laser desorption lies in its ability to ionize non-volatile, polar compounds.\textsuperscript{33,34} However, the upper limit to the size of biopolymer molecules that can be ionized directly as intact molecular ions is only about 1 kDa. Analyte ionization with laser desorption occurs by the energy transfer from the laser photons to the analyte via resonant absorption of the analyte molecule at the laser light wavelength. However, in the case of larger molecules, for example, proteins, the increase in degrees of freedom (which provides additional pathways for energy dissipation) requires higher laser power to achieve enough resonant absorption for ionization. Unfortunately, the higher laser power also causes photodissociation and sometimes thermal decomposition of the analyte. A major breakthrough occurred when Hillenkamp and Karas et al. from the Münster group decided to mix the analyte with a large amount of a small organic molecule ("matrix") resulting in the technique known today as matrix-assisted-laser-desorption ionization (MALDI).\textsuperscript{19} The incorporation of a matrix with the analyte was found to permit laser desorption ionization of very large (> 100 kDa) protonated molecular ions with little or no fragmentation.\textsuperscript{35} The first successful MALDI mass spectrum of a protein was acquired using nicotinic acid as the matrix compound, which was selected on the basis of its high molar absorptivity at the wavelength (266 nm) of the frequency-quadrupled Nd:YAG laser.\textsuperscript{36}

To minimize thermal decomposition, the energy transfer from laser to analyte must
be executed as rapidly as possible. For this reason, pulsed lasers with short pulse widths of 10 nsec or less are utilized for MALDI. The pulsed nature of the lasers makes them ideally suited for use with time-of-flight (TOF) mass spectrometers\textsuperscript{37} which have a very high mass range.\textsuperscript{a} The optimal matrix to analyte molar ratio has been empirically determined for a broad range of sample types and is normally somewhere between 100:1 and 50,000:1\textsuperscript{35} (the observation of this laboratory is that the optimum ratio for the analysis of proteins and peptides is closer to the latter values). It has been proposed that the presence of the MALDI matrix serves at least three functions.\textsuperscript{39,40} The first is to absorb the energy of the laser photons through resonant absorption. Secondly, the large molar excess of the matrix evidently isolates the analyte molecules and reduces the strong intermolecular forces between analyte molecules thereby allowing ionization of individual molecules and preventing excessive formation of large cluster ions. Finally, the matrix effects ionization of the analyte molecule through partial transfer of the energy gained by resonant absorption without causing fragmention or decomposition of the analyte molecular ion. The ionization of the analyte is believed to proceed through proton transfer from the matrix.

The structures of the MALDI matrices used in the experimental work described in the following sections are shown in Figure I.2. These are the most commonly used

\textsuperscript{a}In theory, the mass range of a time-of-flight mass spectrometer is limited only by the laser pulse frequency. Thus, mass range increases with lower frequency. In practice, this is not the only limitation. For example, high mass ions sometimes do not have sufficient kinetic energy to be efficiently detected especially when lower accelerating voltages are employed.\textsuperscript{38} This particular problem can be alleviated through the use of post-acceleration at the detector.
matrices, especially for protein and peptide analysis. All successful matrices have in common a relatively low mass (M, < 250 u) and the ability to resonantly absorb energy at the wavelength of the laser light. However, not all compounds with these characteristics are necessarily good matrices. Analyte incorporation into the crystalline matrix has been found to be essential for successful ionization of the analyte.\textsuperscript{39,41} The first step in MALDI sample preparation is to mix solutions containing analyte and matrix and then apply a small drop (0.5 - 1 μL) of the resultant solution to the MALDI target plate (usually polished stainless steel). As the liquid dries, (prior to insertion into the mass spectrometer), a circular crystalline film forms. Matrices such as 2,5-dihydroxybenzoic acid (DHB) and 3-hydroxypicolinic acid (HPA) successfully co-crystallize with the analyte to form long needle-like crystals about 100-150 μm in length.\textsuperscript{41,42} Although both DHB and HPA both provide satisfactory sample preparations, the spots are heterogenous due to the crystal size and the tendency of the matrix/analyte crystals to form closer to the edges of the sample preparation (i.e., near the rim of the sample spot). This surface heterogeneity can sometimes make it more difficult to find so-called "hot spots" (positions on the sample preparation surface from which the sample ions are easily generated). On the other hand, sinapinic acid (SA)\textsuperscript{43} and α-cyano-4-hydroxycinnamic acid (ACCA)\textsuperscript{44} co-crystallize with the analyte to generate much smaller crystals (about 5-30 μm) and thus provide a much more homogeneous surface.\textsuperscript{45} Furthermore, the analyte concentration gradient from center to edge found with DHB and HPA is not observed when SA and ACCA are used as the matrix compounds. The more uniform surface allows the acquisition of good quality mass
Figure I.2. Structures of some common MALDI matrix compounds used for the analysis of proteins and peptides.
spectra from virtually every position on the sample spot. In addition, more homogeneous sample preparations also tend to increase the sensitivity thereby requiring less laser power for analyte ionization which enhances resolution (vide infra). The exploration of different methods for preparing more homogeneous and thinner films of matrix and analyte mixtures in an effort to improve sensitivity, resolution, mass accuracy, and reproducibility is currently a very active area of MALDI research.46-49

Although a number of compounds have been evaluated as matrices for MALDI,50 in practice, only a handful of these are routinely used for the analysis of proteins and peptides. Different matrix properties such as molar absorptivity, size and type of crystal formation, solubility, etc., provide for unique matrix behaviors. The results with mixtures such as peptide digests, can be quite different when using different matrices.51 In general, it has been observed that smaller peptides are more easily ionized using either ACCA or DHB matrices. On the other hand, larger peptides and proteins that are poorly ionized with ACCA or DHB can often be ionized very easily simply by switching to the SA matrix. In the experimental work related in the following sections, MALDI-TOF-MS analysis of proteins and peptides was performed almost exclusively with either the SA or the ACCA matrix because of the resulting more homogeneous sample surface. Although these matrices usually provide the best results, occasionally it is still necessary to use another matrix, (e.g., one of the other three shown in Figure I.2) to obtain the optimal result. Another potential problem is the tendency of certain matrices to form adduct ions that can
broaden peaks at high mass where the protonated molecular ion is not resolved from the adducted species.\textsuperscript{43,52} Although the use of SA generally does produce a small adduct ion at \( m/z = (M+H+206)^+ \) [where M is the molecular weight of the analyte], ACCA has the advantage of not generating any analyte adduct ions. Some matrices are very sensitive to the analyte concentration while others may tolerate a much wider range. For example, while HABA is known to work optimally at analyte concentrations of 1 pmol/\( \mu \)L or less, other matrices tend to be less sensitive and can still work quite well at analyte concentrations above 10 pmol/\( \mu \)L. Thus, selection of the MALDI matrix is largely dependent on the type of application and necessitates a familiarity with a variety of different types (and behaviors) of matrix compounds. Finally, it is highly unlikely that a "super" matrix will ever be discovered that can provide a superior result for every analyte type.

The energy deposition as a result of the photon bombardment of the MALDI matrix leads to heating and expansion of the crystal surface generating a plume of particles in the ion source containing both ionized and neutral matrix and analyte molecules.\textsuperscript{53} A jet expansion model has been proposed to explain this ejection of matrix and analyte molecules.\textsuperscript{54,55} Higher laser irradiance increases the size and density of the plume, leading to greater energy dispersion and also raises the potential for collisional activation of the analyte.\textsuperscript{54,56} Therefore, maximal resolution and minimal fragmentation is obtained by maintaining the laser beam at the minimum irradiance required for detection of the analyte ion.
A schematic diagram of the first MALDI-TOF instrument acquired by this laboratory (in 1990) is shown in Figure I.3. This instrument, the VT 2000 (Vestec Corp., Houston, TX), was the first commercially available instrument designed exclusively for MALDI-TOF-MS. Initial experiments with this instrument were carried out using the nicotinic acid matrix with the frequency-quadrupled Nd:YAG (which has a 9 nsec pulse width) at 266 nm. However, the introduction of cinnamic acid derivatives by Beavis and Chait\textsuperscript{43} and DHB by Karas and co-workers\textsuperscript{41} were found to work as well or better at the higher wavelength, 355 nm, achievable with the third harmonic of the Nd:YAG laser.\textsuperscript{57} This quickly led to the use of the much simpler and less expensive N\textsubscript{2} laser which emits light of wavelength 337 nm. Furthermore, the pulse width of the N\textsubscript{2} laser is much narrower (3 nsec) and the shot-to-shot reproducibility superior to that of the Nd:YAG laser which rendered the latter laser essentially obsolete for this application. Consequently, all of the MALDI experiments that are described in the following sections were carried out using the N\textsubscript{2} laser.

Referring to Figure I.3, the N\textsubscript{2} laser (generally operated at a repetition rate of 5 Hz for the analysis of peptides and proteins) emits a photon beam that is guided through a UV transparent aperture into the instrument and focused onto the sample target in the ion source. The laser beam is manually attenuated using a variable beam attenuator to achieve the "threshold" value, (i.e., the minimum laser irradiance required to produce a signal due to the analyte) which is facilitated by a real-time display of the ion signal throughout the entire
Figure I.3. Schematic of the VT 2000 MALDI-TOF mass spectrometer.
mass range on the oscilloscope. On the VT 2000, the laser trigger for the N₂ laser is the same one originally used for the Pockels cell for the Q-switch on the Nd:YAG laser. The trigger signal is simultaneously sent to the laser and to the fast transient digitizing oscilloscope (LeCroy 9450A) to begin acquisition of data. The ion source of the TOF mass spectrometer consists of the sample target and a dual set of acceleration plates to enhance field homogeneity. The ions are accelerated into the field free flight tube (2 m in length) and allowed to drift toward the detector. The VT 2000 uses high acceleration voltages (typically 30 or 35 kV) to obtain good resolution since the initial energy distribution of the ions in the ion source (a principal source of peak broadening) as a percentage of the kinetic energy decreases with increasing acceleration voltage. The detection of the accelerated ions is achieved with a hybrid detector consisting of a single microchannel plate (MCP) followed by a 20-stage secondary electron multiplier with focusing mesh Cu/Be dynodes. A potential of -3.4 kV is applied to the first (MCP) detector plate which results in a gain of approximately 10⁸. This amplified ion signal is then directed to the digitizing oscilloscope. The analog signal is digitized at 200 MHz (5 nsec time slices) using an 80 MHz bandwidth filter. Typically 30 shots (at 5 Hz) are signal averaged by the digitizer for each acquired mass spectrum. The digitized information is then downloaded to a PC and the mass spectra are further processed by software developed in-house on a MicroVAX system.

---

*aSince most proteins and peptides are easily protonated, the best sensitivity is usually achieved with positive ion acceleration.
Since the accelerated ions all have the same initial kinetic energy, different masses will have different velocities and the mass of a detected ion can thus be related to time with calibration standards using the following equation:

\[
\frac{m}{z}^{1/2} = At + B
\]

(where \(m/z\) is the mass to charge ratio, \(t\) is the time, and \(A\) and \(B\) are calibration constants that are dependent on the acceleration voltage and ion drift length). 59

The original VT 2000 instrument came equipped with a set of pulsed deflection plates between the ion source and the detector which were installed for deflection of lower mass (matrix) ions to prevent saturation of the detector. The deflection plates were later replaced with a system utilizing a guide wire which was installed in the center of the flight tube. The guide wire voltage is synchronized with the trigger signal to allow deflection of the matrix ions. The guide wire is at +100 V potential at the start of acquisition which deflects the majority of the low mass ions away from the detector. Shortly after the known arrival time of the matrix ions at the detector, the guide wire potential is switched to -100 V where it remains until the next trigger signal is received. Use of the guide wire for deflection of matrix ions was found to greatly increase the sensitivity of this instrument for detection of higher mass ions. For the analysis of smaller molecules such as peptides, (where saturation of the detector is not usually problematic), the voltage on the guide wire
was usually kept constant at ~100 V throughout the mass range.

The VT 2000 is also capable of generating IR MALDI spectra using the Er-YAG laser ($\lambda = 2.94 \mu m$). The use of IR lasers with MALDI was first explored due to the potential of greatly expanding the range of available matrix compounds.\textsuperscript{60,61} In contrast to UV-MALDI where ionization occurs via electronic transitions, IR-MALDI ionization is due to vibrational transitions. Some of these transitions, e.g., the O-H and N-H stretching modes can be quite energetic at $\sim 3 \mu m$ wavelength. Liquids such as water and glycerol are thus potential matrices using IR lasers. The drawback to IR lasers is that they tend to be difficult to use and usually have very long pulse widths (the Er-YAG laser pulse width is 120 nsec). The primary application of IR lasers to date has been in the MALDI analysis of blotted proteins on membranes such as polyvinylidene difluoride (PVDF).\textsuperscript{62,63}
II.1 Introduction

The most common strategy for sequencing proteins involves specific cleavage of the unknown protein with an endoproteinase followed by HPLC fractionation of the digest in order to individually separate the peptides for sequencing by the Edman degradation. Alignment of the peptides in the proper order necessitates a second digest with an endoproteinase of a different specificity. The second digest generates peptides which, after purification and sequencing, provide overlapping sequence information which then allows the complete primary structure of the protein to be deduced. When sequencing proteins by mass spectrometry, this general strategy remains the same except that the sequencing is performed using a different principle of analysis. A major advantage of mass spectrometry is the capability of sequencing individual components of a mixture without the prior separation into single peptides required for the Edman method. Nevertheless, protein sequencing by mass spectrometry is most properly viewed not as a replacement of traditional methods such as Edman sequencing, but rather as a powerful complementary technique (especially in those cases where the traditional methods have failed).

For a successful digest, the protein to be sequenced is denatured and any disulfide
bond linkages are reduced and the resulting sulfhydryl groups S-alkylated. Typically only a handful of endoproteinases such as Endo-Lys-C, Endo-Asp-N, and trypsin are used because of their high degree of specificity. In addition to the enzymatic methods, there are also a number of chemical cleavage reagents that have also been employed for the purpose of generating peptides for sequence analysis. A variety of enzymes as well as chemical reagents for protein cleavage are commonly available and their use has been described in the literature. In addition, a method for the specific chemical cleavage of proteins at cysteine with concomitant generation of peptides with overlapping sequences is described in Section VI.

Section II.2 FAB Tandem Mass Spectrometry

A schematic diagram of the JEOL HX110/HX110 high performance tandem mass spectrometer used for the FAB tandem mass spectrometry (MS/MS) experiments is shown in Figure II.1. Both HX110 mass spectrometers are identical magnetic sector instruments of EB configuration that can be operated together or independently. The first sector (E) is an energy filter that utilizes a radial electric field to reduce the energy dispersion of the ions generated in the ion source. The energy focused ions leaving the electric sector are then focused according to their momentum by the magnetic (B) sector. Thus, much higher resolution is achieved than is possible with a simpler one sector magnetic instrument due to
the double focusing by both energy and momentum. For protein sequencing, the first step is to determine the molecular weight of peptides generated by the digests. After partial HPLC fractionation, the peptides are ionized by FAB and the \( m/z \) values for the protonated peptides \([(M+H)^+\)] are determined using the first mass spectrometer (MS-1). For this type of experiment the electric field of the first E sector is held constant while the magnetic field of the B sector is scanned in order to successively focus the ions on Detector 1 according to their \( m/z \) values. At the 10 kV maximum acceleration voltage of this instrument, ions up to \( m/z \) 14,500 can be detected. Although the width of the main slit of MS-1 can be narrowed (at the expense of sensitivity), to achieve very high mass resolution the much lower resolution of 1:3000 is sufficient for the separation of an isotopic multiplet and allows selection of the \(^{12}\)C-only species for the majority of the peptides.\(^a\) Peptide (M+H)\(^+\) ions generated by FAB ionization are generally very stable although fragmentation for sequence information can be increased somewhat by introducing higher peptide concentrations into the mass spectrometer. However, tandem mass spectrometry (MS/MS) offers a much better solution.\(^{68,69}\) One of the major advantages of MS/MS is that it precludes the need for careful fractionation by HPLC because MS-1 is now used as a mass filter (illustrated in Figure II.2) to introduce specific ions into the second mass spectrometer (MS-2). Furthermore, the high

\(^a\)Operation of the magnetic sector instrument under these conditions allows the isotopic multiplets of peptides with \( m/z \) of at least 3000 to be easily resolved. Thus, the italicized \( m/z \) values in the following text and tables denote monoisotopic mass, i.e., the sum of \(^{12}\)C, \(^1\)H, \(^{14}\)N, \(^{16}\)O, etc., isotopes only. This distinguishes these measurements from polyisotopic \( m/z \) measurements (for example, the \( m/z \) values determined by linear, continuous [non-DE] MALDI-TOF-MS) indicated in normal type.
Figure II.1. Schematic diagram of the JEOL HX110/HX110 high performance tandem mass spectrometer.
Figure II.2. Principle of protein sequencing by tandem mass spectrometry. Peptides obtained from proteolytic digests are ionized to $P_n$ and mass analyzed by MS-1 (first mass spectrum). In the tandem experiment, the $^{12}$C-only components of each peptide are individually selected by MS-1 and dissociated (via CID) in the collision cell. The resulting fragments ($F_n$) are mass analyzed by MS-2 (second mass spectrum).
resolution (the mass resolution of MS-1 is doubled when B is kept constant) of MS-1 allows the selection of only the $^{12}$C component of the peptide ion without sacrificing any sensitivity. For the MS/MS experiment, the ion source of MS-2 is replaced with a collision cell that is filled with an inert gas with a high ionization potential (such as He) and usually floated to some potential above ground. The $^{12}$C-only molecular ions from MS-1 with several keV translational energy encounter the He atoms in the collision cell and become collisionally activated. The molecular ions fragment via collisionally induced dissociation (CID) and the resultant ions are detected by MS-2. Since the velocities of the precursor (molecular) ions and fragment ions are the same, the separation can be achieved by simply scanning the electric sector of MS-2. A well resolved fragment ion spectrum is obtained by the simultaneous scanning of MS-2 in such a manner that $E_2$ and $B_2$ always have the proper values for optimal transmission of fragment ions (i.e., a linked scan).

The type of fragment ions (the structures and nomenclature appear in Appendices I - III) produced by CID of peptide molecular ions is now well established. The majority of the fragments arise from cleavage of the peptide backbone. Those ions containing the amino terminus are designated $a_n$, $b_n$, and $c_n$, while $x_n$, $y_n$, and $z_n$ are those ions that contain the C-terminus. In addition, some of these ions (also resulting from cleavage of the peptide backbone) contain neither the N-terminus nor the C-terminus in which case they are termed

---

*The subscript "n" refers to the position of the amino acid in the peptide from the N-terminus and C-terminus for N-terminal and C-terminal ion types, respectively.
internal fragment ions. Due to the high acceleration voltage capability of the JEOL HX110/HX110 instrument, the peptide fragmentation is usually performed using high energy CID (so called because the collisions occur at kilovolt energies). The principal advantage of using high energy CID (as opposed to low energy CID\(^7\)) for peptide sequencing is that the higher energy of the collisionally activated species results in additional fragmentation of the amino acid side chain; these ion types are termed \(d_n\) for the N-terminal containing and \(v_n\) and \(w_n\) for the C-terminal containing ions, respectively. For example, the fragments produced due to the different side chain structures of the isomeric Leu and Ile now allows these two amino acids to be distinguished.\(^7\)\(^4\)\(^7\)\(^5\) Since the side chain cleavage occurs between the \(\beta-\gamma\) carbons, residues with \(\beta\)-branched side chains can generate more than one product. For example, the two possible \(d_n\) ions from the side chain cleavage of Ile (at position \(n\)) are shown below:

\[
\begin{align*}
\text{H}^+ & \quad \text{CH}_3 \\
R_{n-1} & \quad \text{CH}_2 \\
H-(\text{NH}-\text{CH}-\text{CO})_{n-1}-\text{NH}-\text{CH} & \quad \text{CH} \\
\end{align*}
\]  
\[a_n + 1\]

\[
\begin{align*}
\text{H}^+ & \quad \text{CH}_3 \\
R_{n-1} & \quad \text{CH}_2 \text{CH}_2 \text{CH}_3 \\
H-(\text{NH}-\text{CH}-\text{CO})_{n-1}-\text{NH}-\text{CH} & \quad \text{CH} \\
\end{align*}
\]  
\[a_n + 1\]

\[
\begin{align*}
\text{H}^+ & \quad \text{CH}_3 \\
R_{n-1} & \quad \text{CH} \\
H-(\text{NH}-\text{CH}-\text{CO})_{n-1}-\text{NH}-\text{CH} & \quad \text{CH} \\
\end{align*}
\]  
\[d_{na}\]

\[
\begin{align*}
\text{H}^+ & \quad \text{CH}_3 \\
R_{n-1} & \quad \text{CH}_2 \text{CH}_2 \text{CH}_3 \\
H-(\text{NH}-\text{CH}-\text{CO})_{n-1}-\text{NH}-\text{CH} & \quad \text{CH} \\
\end{align*}
\]  
\[d_{nb}\]
By convention, the fragment of lower m/z is designated as \( d_{na} \) and the higher one \( d_{nb} \). The C-terminal \( w_n \) ions are designated similarly. Finally, the presence of these additional ions in high energy CID mass spectra provides supplementary information and thereby increases the confidence of the sequence assignment.

Both low and high energy CID mass spectra contain immonium ions (Appendix I) which serve to confirm the presence (or absence) of individual amino acids. With high energy CID, the additional fragmentation of the side chains leads to high mass ions (Appendix IV) whose mass differences from the precursor ion are specific for individual amino acids. These ions are labeled as "-X" in the high energy CID mass spectra in the following sections and indicate loss of the side chain of amino acid X.

For any given peptide, not all of the possible fragment ions types are generated. However, this is not a problem and in fact aids in the interpretation of the spectrum since the fragmentation mechanisms are known to arise mainly from charge-remote processes. Thus, the presence and location of basic amino acids, which are easily protonated and thus positively charged, strongly affects the type of fragmentation that is observed. For example, N-terminal basic sites lead to the formation of the N-terminal ion types while the C-terminal ion types occur when the basic site is localized on the C-terminal portion of the fragment. Fragmentation can also be influenced by other (non-basic) amino acids. For
example, the side chain specific formation of "d" ions, (an N-terminal ion type with the charge therefore localized on the N-terminal portion of the fragment), tends to be most prominent when the side chain of the amino acid is not aromatic and contains at least two carbons. Similarly, proline has been observed to form abundant $y_n$ and $y_{n-2}$ ions and residues preceding threonine tend to form more abundant $c_n$ ions. The understanding of the observed fragmentation in relation to a specific peptide sequence is continually being refined. For example, the two species $a_{n-46}$ and $z_{n-46}$ have recently been discovered to occur (due to loss of -CH$_3S^+$) when Met is at position n - 1 relative to the N-terminus and C-terminus, respectively.

II.3 Reflectron and Post-Source-Decay Mass Spectrometry

The existence of an initial kinetic energy distribution for the ions produced in the MALDI ion source thereby contributing to peak broadening and poorer resolution has already been mentioned in Section I.3. Although increasing the acceleration voltage reduces this effect, the resultant higher electric fields in the ion source can also lead to more prompt fragmentation (i.e., in the ion source) of the molecular ions. A much better alternative is to construct a secondary acceleration field after the drift region to arrest the ions and re-accelerate (reflect) them back toward the ion source. This reflection corrects for the time dependent dispersion caused by the initial kinetic energy distribution. Figure II.3
is a schematic diagram of the Voyager Elite MALDI-TOF reflectron mass spectrometer (PerSeptive Biosystems, Framingham, MA). This instrument can be operated both in the linear and reflector mode and thus has two (circular) MCP detectors. The ions generated in the ion source reach both detectors by passing through an orifice in the center of the reflector detector. In the reflector mode, the ions are prevented from reaching the linear detector by the electrostatic mirror (reflector) which reverses the peak broadening due to the initial kinetic energy distribution. The result is an improvement in resolution by at least an order of magnitude which allows isotopic species to be easily resolved for peptides up to about 3 kDa. A drawback of using the instrument in the reflectron mode is the decrease in sensitivity in comparison to an identical experiment performed in linear mode. This is due to the fact that both the intact molecular ions as well as any resultant fragment ions contribute to the ion signal in the linear mode since both species have essentially the same flight time. Conversely, only intact molecular ions retain sufficient kinetic energy to be well focused onto the reflector detector in reflector mode. Fragment (metastable) ion species that formerly contributed to the signal in the linear mode no longer have the same flight time as the molecular ion and are not well resolved. This feature of reflectron mass spectra is exploited in post-source-decay (PSD) mass spectrometry (vide infra).

The decrease in the signal to noise ratio (S/N) observed in reflector mass spectra has

---

*a* The dissociation of molecular ions in the drift tube region is another source of peak broadening and contributes to the lower resolution in linear mass spectra.
Figure II.3. Schematic diagram of the Voyager Elite reflectron mass spectrometer.
led to a reexamination of the factors leading to lower resolution in linear mode and some solutions. MALDI ions have been successfully cooled by storage in an ion trap prior to extraction into a TOF mass spectrometer. More recently, several groups have explored the application of a pulsed extraction field in the MALDI ion source in an effort to reduce the initial kinetic energy distribution. Assuming that production of gas-phase ions occurs instantaneously in MALDI and that the primary source of poor resolution is due to energy dispersion in the ion source, the initial kinetic energy distribution of the ions can be decreased by applying the extraction field after ionization has occurred (delayed extraction). The Voyager instrument has recently been modified to allow the acquisition of mass spectra by delayed extraction. The exact time delay is empirically determined and is normally between 50-500 nsec. This modification now allows routine MALDI analysis of a peptide such as angiotensin I (M, 1295.7) with mass resolution of at least 1:3000 in the linear mode. Furthermore, sensitivity remains excellent and peptide isotope peaks can easily be resolved using delayed extraction in the linear mode.

In early 1991, Spengler et al. identified fragment ions in reflector mass spectra of peptides and proteins due to metastable decay of molecular ions and suggested that these fragment ions could provide structural information. Further work by the same laboratory demonstrated that it was indeed possible to obtain sequence information from peptides by PSD using a reflectron mass spectrometer. The Voyager data acquisition software is equipped with an option for instrument control in the PSD mode. The concept behind PSD
mass spectrum acquisition is illustrated in Figure II.4. In this hypothetical case, a precursor peptide ion having (M+H)$^+$ of $m/z$ 1000 undergoes dissociation to form two fragment ions, A$^+$ and B$^+$:

\[
\begin{array}{c}
\text{B}^+ \quad \text{----------} \quad (\text{M+H})^+ \quad \text{----------} \quad \text{A}^+ \\
\text{m/z 300} \quad \text{m/z 700}
\end{array}
\]

All three ions have different kinetic energies and will therefore penetrate the electrostatic mirror to different depths. At a mirror ratio \((V_{\text{ref}} / V_{\text{acc}})\) of 1.00, maximal penetration (and therefore maximal resolution in time) is achieved only for the precursor ion. Although both A$^+$ and B$^+$ are detected at this mirror ratio, the reflector penetration of these ions is very shallow and results in broad or unfocused peaks. Focusing of the fragment ions is accomplished by reducing the reflector voltage to increase the flight time (and hence the penetration depth) of a fragment ion such that it follows the trajectory of the precursor ion (i.e., it has the same flight time and penetration depth) at mirror ratio 1.00. By stepwise reduction of the reflector voltage, metastable fragment ions of differing energy can successively be brought into focus. Thus, in this example, fragments A$^+$ [m/z 700] and B$^+$ [m/z 300] are optimally focused when the PSD mirror ratios are adjusted to 0.70 and 0.30, respectively. The individual PSD mass spectra are concatenated into a composite mass spectrum using the same software. The \textit{m/z} scale of the PSD spectrum of an unknown is calibrated using the fragment ions in the PSD mass spectra of authentic peptides of known
Figure II.4. Principle of MALDI-PSD mass spectrum acquisition. The shape of the reflected ion trajectories is conical. The circular reflector detector (cutaway view shown) has an opening in the center to allow ions to pass through into the electrostatic reflector region. Only molecular ions, (M+H)^+, are correctly focused (solid lines) when the mirror ratio equals 1.00. Fragment ions A' and B' are brought into focus (solid lines) by successive lowering of the mirror ratio. In this example the m/z of (M+H)^+, A^+, and B^+ is 1000, 700, and 300, respectively.
Figure II.5. MALDI-PSD mass spectrum of human angiotensin I (M+H)$^+$ = m/z 1296.7. The peaks in the low mass range labeled with capital letters corresponding to one letter amino acid abbreviations are the immonium ions. Internal ion fragments are indicated by the sequence. Refer to Appendix 1 for an explanation of the nomenclature.
fragmentation according to the following equation:

\[
m/z_{\text{fragment}} = m/z_{\text{precursor}} \left( \frac{t_{\text{fragment}} / t_{\text{precursor}} - C_1}{(1-C_1)} \right) \left( (MR - C_3) (1 - MR)^{C_4} \right)
\]

(where MR is the mirror ratio, \(t\) is the flight time, and \(C_1, C_3,\) and \(C_4\) are the calibration constants). The calibration constants are calculated with the aid of another computer program. An example of a calibrated composite PSD mass spectrum is shown in Figure II.5. The molecular ion \((M+H)^+ = m/z\) 1296.7 is due to human angiotensin I which has the amino acid sequence DRVYIHPFHL. MALDI-PSD mass spectra are very similar to the CID mass spectra acquired using tandem mass spectrometers (Section II.2) and a comparison has already been reported.\(^8\) The nomenclature in Figure II.5 is the same used for identification of fragments derived from CID mass spectra.\(^7\) The structures of these fragment ions are shown in Appendix I. MALDI-PSD mass spectra generally contain \(a_n, b_n,\) and \(y_n\) type as well as immonium and internal fragment ions. Some additional ions (Appendix II) encountered in MALDI-PSD mass spectra include the \(a_n - 17, b_n - 17, b_n + H_2O,\) and \([\text{internal fragment} - 28]\) ions.

The PSD spectra are similar in many respects to the MS/MS spectra obtained from CID experiments.\(^8\) Furthermore, the use of MALDI-PSD for protein sequencing is very attractive since very little sample is required (the complete mass spectrum in Figure II.5 was acquired from only one pmol of the peptide) and mixtures such as the protein digests
do not present a problem due to the introduction of timed deflection gates that allow individual selection of molecular ions as long as they differ in mass by at least ± 0.2 %. The Voyager instrument is equipped with a timed ion selector (Figure 11.3) that consists of a pair of deflection plates near the entrance of the reflector region. The timed ion selector deflects all but the desired molecular ions-- since this deflection occurs prior to ion reflection, some of these ions are also metastable fragments-- allowing MALDI-PSD to be performed even with mixtures (such as peptide digests). Although it is possible in principle to determine the complete sequence of an unknown peptide by analysis of the fragments that result from PSD, the strength of this technique (at the present time) lies in providing confirmatory rather than complete sequence information. While the molecular ions of some peptides fragment easily producing many ions, others exhibit fewer cleavages and some molecular ions appear to not fragment at all. Due to its infancy as a method for protein sequencing, PSD is still in the process of refinement.

Ion detection on the Voyager instrument is accomplished by dual microchannel plates. Similar to the VT 2000, the Voyager also has a guide wire to enhance sensitivity. In addition, saturation of the detectors by matrix ions is prevented by a low mass gate which consists simply of a time delay before application of the detector voltage. On the Voyager instrument, the ion source acceleration voltage is generally between 20 and 25 kV. The analog detector signal is digitized with a 300 MHZ Tektronics scope in the same manner as with the VT 2000 instrument except that data processing is performed using the PerSeptive
Biosystems GRAMS software. Between 64 and 256 scans are summed for each mass spectrum.

II.4 MALDI-TOF-MS and "Ladder" Sequencing

Conventional SDS-PAGE for obtaining protein molecular weight information is a crude technique compared to MALDI-TOF-MS and is sometimes an unreliable indicator of $M_r$. A principle advantage of MALDI-TOF-MS is its ability to provide accurate molecular weight information of very high molecular weight species (such as proteins) with very low sample consumption (typically 1 pmol or less). In the course of protein structure determination, the extremely accurate $M_r$ information (ascertained by measurement of the $m/z$ value for the protonated molecular ion, $[(M+H)^+]$) provided by MALDI-TOF-MS is very useful by confirming or ruling out certain sequences. Protein sequencing by MALDI-TOF-MS is similar to the other methods (e.g., Edman degradation, FAB MS/MS) in that cleavage of the protein to generate peptides for individual sequence analysis is still required. Unlike FAB, the relative insensitivity of MALDI to buffers and salts allows aliquots from proteolytic digest solutions to be analyzed directly without any prior purification. Thus, MALDI-TOF-MS provides a method for rapid and accurate molecular weight determination of peptides derived from protein digests. As with the protein $M_r$ determination, MALDI analysis of several digests using enzymes of different specificity
can either confirm or rule out specific sequences.

Following protein cleavage, the resultant peptides can be sequenced by MALDI-TOF-MS in several ways. Although it has been known for some time that molecular ions generated by MALDI are prone to fragmentation while in the field free drift regions of a TOF mass spectrometer, it is only very recently\textsuperscript{84,85} that experiments have been designed and carried out expressly to exploit this aspect of MALDI-TOF-MS. This type of MALDI reflectron experiment, (PSD), has already been discussed (Section II.3). Another promising development is the application of CID to MALDI-TOF-MS of peptides and it has been demonstrated that some of the resultant fragments are the d\textsubscript{n} and w\textsubscript{n} types which occur only in high energy CID experiments.\textsuperscript{88} These types of fragments facilitate interpretation of the mass spectrum by providing additional information and also allow the isomeric Ile and Leu residues to be distinguished.\textsuperscript{74}

MALDI-PSD on most reflectron instruments\textsuperscript{*} requires the acquisition of individual reflectron spectra obtained using different mirror ratio voltages which are then concatenated into a single composite mass spectrum. On the other hand, MALDI-TOF experiments performed in the linear mode are much faster, simpler and provide increased sensitivity (for the molecular ions) since the molecular and resultant fragment ions are detected

\textsuperscript{*} Acquisition of MALDI-PSD mass spectra of peptide standards has been demonstrated on a TOF instrument utilizing a coaxial curved field reflectron which obviates the need to scan the reflectron or step the voltage.\textsuperscript{89}
simultaneously. For these reasons, it would be desirable to have a method for protein sequencing using a simpler linear TOF instrument. Exopeptidase digestion, (carboxypeptidase Y^{90,91}), has been used to obtain the C-terminal sequence of peptides by analysis of the resulting systematically degraded peptides by FABMS.^{92} More recently, carboxypeptidases Y and B (CpY and CpB^{93}, respectively) were used together with linear MALDI-TOF-MS to obtain C-terminal sequence information of a peptide from synthetic human parathyroid hormone.^{94} The much lower sample consumption of MALDI-TOF-MS compared to FABMS makes the former the method of choice for these types of experiments. For example, the MALDI-TOF mass spectrum of 1 pmol of human renin substrate, a tetradecapeptide with sequence DRVYIHPHLLVYS, is shown in Figure II.6 after incubation with CpY for 15 min. The six peaks below the molecular ion of renin substrate, [(M+H)^+ = m/z 1760.0] correspond to losses in mass consistent with the sequential removal of S, Y, V, L, L, and H, respectively. The mass accuracy of MALDI-TOF-MS, (± 0.01% with an internal standard^{59}), is sufficient to unambiguously distinguish residues whose masses differ by 1 Da (e.g., Asp and Asn; see Appendix IV). An additional advantage of using carboxypeptidases is that the resulting C-terminal sequence information is complementary to the N-terminal sequence data provided by traditional Edman degradation.

A number of exopeptidases have been isolated and characterized but only a few of these are commercially available.^{95,96} Among the carboxypeptidases, CpA^{93}, CpB, CpY and
Figure II.6 MALDI-TOF mass spectrum of human renin substrate after 15 min with CpY. Annotations correspond to residues cleaved off the C-terminus of the peptide.
CpP are easily obtainable, however, only CpY and CpP are non-specific in that they are capable of catalyzing the hydrolysis of all of the commonly known amino acids. With CpY, optimal peptide cleavage occurs near neutral pH while CpP digests are generally carried out at lower pH values (pH 4-6). Both CpY and CpP release hydrophobic amino acids rapidly and charged residues somewhat less rapidly while Gly, especially in the penultimate position, slows the release of the terminal amino acid. Although Pro is cleaved by both CpP and CpY, the reaction rate is very slow. It is the experience of this author that although CpY and CpP behave similarly, CpP exhibits higher enzymatic activity. In addition, peptides can also be cleaved sequentially from the N-terminus using aminopeptidases. Aminopeptidase M (AmM) isolated from hog kidney is commercially available and like CpY and CpP is also non-specific. The complete sequence of a peptide can thus be determined by separate carboxypeptidase and aminopeptidase experiments using MALDI-TOF-MS to analyze digest aliquots removed at timed intervals.

In the ideal case, exopeptidases sequentially cleave each amino acid and the resultant degraded peptides are detected by MALDI-TOF-MS which then allows deduction of the sequence. Unfortunately, different amino acids are released at different rates which can lead to gaps in the sequence. In addition, the reaction often appears to stop at certain residues (especially Pro) without yielding any further information. The hydrolysis rate of the terminal amino acid is also affected by neighboring residues and especially by the penultimate one. Although chemical reagents for the C-terminal sequential degradation of
peptides exist, (e.g., thiocyanate ion to form C-terminal thiohydantoins\textsuperscript{100-102}), these chemical degradation methods suffer from their own unique problems.

Undoubtedly the most successful sequential cleavage method is the one used for Edman degradation sequencing where phenylisothiocyanate (PITC) reacts with the peptide N-terminus to form (after addition of TFA) the phenylthiohydantoin (PTH) derivative. A "ladder" sequencing technique\textsuperscript{103}, based on the Edman degradation reaction, has been developed in conjunction with MALDI-TOF-MS analysis using PITC with a small amount (5 \%) of phenylisocyanate (PIC). The latter functions as a terminating agent by generating peptides having phenylcarbamyl (PC) N-termini (these are stable in TFA and do not lead to cleavage of the next peptide bond). Each Edman cycle is performed in the same reaction vessel, the contents of which, are then analyzed by MALDI-TOF-MS. Sequence interpretation is straightforward since the resulting "ladder" mass spectrum (in this case for six cycles) is conceptually similar to the one shown in Figure II.6. A drawback of this method is that with each successive cycle, PIC consumes a portion of the starting peptide which ultimately limits the number of cycles. This problem has been overcome by a technique (also based on the Edman degradation reaction) developed recently by Pappin et al. utilizing trifluoroethylisothiocyanate (TFETC) rather than PITC to form the thiohydantoin derivative (TFETH).\textsuperscript{104} The use of volatile reagents eliminates the need for HPLC since the reagents (TFETH and TFA [or HFBA]) can be removed by evaporation. Instead of the terminating reagent, a fresh aliquot of starting peptide is introduced at the
beginning of each Edman cycle. Like the PITC/PIC reaction, all of the cycles are
performed in the same reaction vessel. After completion of the final cycle, the "ladder"
mass spectrum of the vessel contents is generated by MALDI-TOF-MS.

The modified Edman degradation methods described above are too tedious to
perform manually and are well suited to automation. On the other hand, exopeptidase
digestion of peptides coupled with MALDI-TOF-MS can, in principle, be carried out very
rapidly and with much less effort. As already noted, the major disadvantage to
exopeptidase digestion is due to the variability of amino acid cleavage rates. This problem
has been addressed quite recently by attachment of quasi-terminating groups to the C-
terminus of peptide standards in the course of carboxypeptidase digestion.\textsuperscript{105} By
introducing high concentrations of lysinamide to CpY digests, the former functions as a
nucleophilic reagent that competes with H\textsubscript{2}O in the hydrolysis reaction. Since the
hydrolysis rate of the amino acid amide is very slow, a "ladder" MALDI mass spectrum can
be generated that contains a set of ions due to sequentially degraded, lysinamide terminated
peptides.
II.5 Summary of Applications Discussed in Sections III - VI.

This section has provided an overview of the different methods used for the mass spectrometric analysis of proteins and peptides presented in the following sections. The complete primary structure of a protein (glutaredoxin isolated from human erythrocytes) solely determined using FAB ionization and high performance tandem mass spectrometry is described in Section III. FAB tandem mass spectrometry was also used to deduce the structure of a novel amino acid (Section IV) isolated from a mussel protein. The amino acid sequence of the latter, isolated from *M. edulis*, was finally determined (after incomplete results by Edman degradation) using MALDI-TOF-MS in combination with exopeptidase digestion (Section IV). The active site of the *E. coli* inosine monophosphate dehydrogenase enzyme has been determined and the amino acid sequence of the protein was corrected using MALDI-TOF-MS (Section V). Finally, the reaction of cyanide with disulfide bond containing proteins was optimized using MALDI-TOF-MS analysis of the reaction mixtures (Section VI).
Section III

**Determination by Mass Spectrometry of the Primary Structure of the Glutaredoxin Isolated from Human Erythrocytes**

III.1 Introduction

Glutaredoxins (Grx), also known as thioltransferases,\textsuperscript{106} are essential for the glutathione (GSH) dependent synthesis of deoxyribonucleotides.\textsuperscript{107,108} Ribonucleoside diphosphates are known to be reduced to deoxyribonucleoside diphosphates via an electron transport chain with NAPDH, FAD, thioredoxin (Trx) and Trx reductase.\textsuperscript{109} However, an alternate electron transport system involving Grx and Grx reductase was discovered following the observation that mutant \textit{E. coli} not expressing Trx was still capable of producing deoxyribonucleotides.\textsuperscript{110-112} The Grx and Trx electron transport systems are very similar although the former system differs from the latter in having glutathione as a prosthetic group in Grx reductase. Glutaredoxins are distinguished from thioredoxins in their ability to be reduced by GSH. Although not homologous, the similar function of Trxs and Grxs has been supported by X-ray data which show these two proteins to have similar conformations.\textsuperscript{108} Thioredoxins and glutaredoxins are also thought to play a role in cellular sulfhydryl homeostasis.\textsuperscript{106,108,113} A Grx has been isolated from human red blood cells or erythrocytes (hErGrx) that has been shown to play a role in the maintenance of reduced
sulfhydryl groups in these cells.\textsuperscript{114}

Circulating red blood cells respond in a number of ways to oxidative stresses. For example, the conformation of hemoglobin (Hb) can change and affect O$_2$ and heme binding due to the presence of a reactive sulfhydryl group on each $\beta$-subunit of the protein.\textsuperscript{115,116} Similarly, energy production in erythrocytes is affected by inactivation of phosphofructokinase due to oxidation of the sulfhydryl groups.\textsuperscript{117} The hErGrx protein was identified as a thioltransferase due to its ability to catalyze the GSH-dependent reduction of Hb-S-S-glutathione mixed disulfides and reanimate oxidized phosphofructokinase \textit{in vitro}.\textsuperscript{114,118} In addition, the same protein was found to catalyze the regeneration of red blood cell membrane sulfhydryl groups modified by oxidation.\textsuperscript{119} In this section, the application of FAB in conjunction with high performance tandem mass spectrometry (as well as MALDI-TOF-MS) for the complete primary structure determination of this thioltransferase is discussed.

All mammalian Grxs of known primary structure\textsuperscript{120-122} exhibit very close homology and possess an acetylated amino terminus. Because of this close homology, the primary structure of hErGrx is of interest in evolutionary studies. The hErGrx also contains a blocked amino terminus\textsuperscript{114} and is thus an excellent candidate for sequencing by mass spectrometry since, unlike amino acid sequencing by Edman degradation, blocked amino termini do not present any special problems in mass spectrometric protein sequencing.\textsuperscript{123}
The identity of this thioltransferase as a human Grx was confirmed by the high degree of sequence homology to the mammalian Grxs, however, the human homologue is distinguished by unique amino acid differences.

III.2 Experimental Procedures

For both the FABMS and MALDI-TOF-MS experiments, the isolated hErGrx was first S-alkylated. S-ethylpyridylated hErGrx was prepared by dissolving approximately 60 µg (5 nmol) of lyophilized protein in denaturing buffer (8 M urea, 100 mM Tris-HCl, pH 8.4) for 30 min at 37 °C. Reduction and S-ethylpyridylation were performed concurrently by simultaneous addition of 10 µL of each reagent solution (1% (v/v) solutions in n-propanol of triethylphosphine and 4-vinylpyridine, respectively) followed by purging with argon and incubation of the reaction mixture at 37 °C for 2.5 h. S-carbamidomethylated glutaredoxin was prepared by dissolving about 240 µg (20 nmol) of the protein in a denaturing buffer (6 M guanidine-HCl, 100 mM Tris-HCl, 1 mM EDTA, pH 8.4). After 1 h at 37 °C, 17 µL of a 0.04 µg/µL aqueous solution of DTT were added, the solution purged with argon, and incubated at 37 °C for an additional hour. Thirteen microliters of a 0.17 µg/µL aqueous solution of iodoacetamide were then added, the solution again purged with argon, and incubated for 1 h at 37 °C. The resulting S-alkylated protein was purified using reversed phase HPLC (RP-HPLC).
Peptides for FABMS analysis were prepared by proteolytic digestion of the S-alkylated Grx. S-carbamidomethylated Grx was digested with trypsin (100:1 substrate:enzyme) in digestion buffer of 2 M urea, 100 mM Tris-HCl (pH 8.4) for 12 h at 37 °C. The digest was partially fractionated by RP-HPLC. Two of the S-carbamidomethylated fractions (containing larger peptides (M+H)^+ = m/z 2404 and 3291, respectively) were further digested with Endo-Asp-N in 100 mM ammonium acetate (pH 7.8) for 6 h followed by HPLC fractionation. S-ethylpyridylated Grx (from 5 nmol of protein) was digested with Endo-Lys-C (200:1 substrate:enzyme) in digestion buffer (2 M urea, 100 mM Tris-HCl, pH 9.0) at 25 °C for 22 h. S-ethylpyridylated Grx prepared from 10 nmol of protein was digested with trypsin using the same conditions as described for the S-carbamidomethylated analogue. Finally, S-ethylpyridylated Grx (from about 20 nmol of protein) was digested with α-chymotrypsin using the same conditions as for trypsin. In all cases, partial fractionation was accomplished by RP-HPLC.

Peptides (i.e., protein digests) for MALDI-TOF-MS were prepared from 1 nmol or less of the S-alkylated Grx. The S-ethylpyridylated Grx was digested with Endo-Glu-C (100:1 substrate:enzyme) in 2 M urea, 100 mM ammonium acetate (pH 4.0) for 12 h at 37 °C. However, Endo-Glu-C was observed to be less specific than the Endo-Lys-C, trypsic, and Endo-Asp-N digests. The S-ethylpyridylated Grx was also digested with Endo-Asp-N in 2 M urea, 100 mM ammonium acetate (pH 7.8) for 12 h at 37 °C. Aliquots from the reaction solutions were diluted with MALDI matrix solution and analyzed (no HPLC).
fractionation) directly by MALDI-TOF-MS. The Endo-Lys-C digest described in the previous paragraph was analyzed similarly.

The molecular weight of the proteolytic peptides was determined by MS-1 of the tandem high resolution mass spectrometer (JEOL HX110/HX110). Single scans were acquired at a scan speed of 2.2 min to scan the range from \( m/z \) 100 to 6000. The main (source) slit of MS-1 was adjusted to provide approximately 1:3000 resolution. At this resolution, the isotope multiplets resolved at least up to \( m/z \) 3000. Thus, the \( m/z \) values correspond to the monoisotopic values. The ions were accelerated using 10 kV accelerating voltage and the cesium gun operated at 20-25 kV. For tandem mass spectrometry, MS-1 was used to mass select individual peptides as their \( ^{12}\text{C}\)-only protonated molecules, \((\text{M+H})^+\). Fragmentation of the molecular ions occurred via CID with He in the field-free region between MS-1 and MS-2. The resulting fragments were mass analyzed by MS-2 using linked scans (0.8 min from \( m/z \) 50 to 2000). The collision cell was floated at 3 kV above ground. The resolution of both instruments was set to 1:1000 (i.e., the static resolution of MS-1 was 1:2000 and dynamic resolution of MS-2 was 1:1000). Since all CID fragments are derived from the \( ^{12}\text{C}\)-only precursor ion they also are monoisotopic. Both MS-1 and CID profile scans (using 300 Hz and 100 Hz filtering, respectively) were collected using the JEOL DA5000 data system. A more detailed description of FAB ionization and the JEOL tandem mass spectrometer can be found in Sections I.2 and II.2, respectively.
Matrix solutions of either SA or ACCA were prepared in mixtures of ACN and H₂O. The ACCA matrix solution was approximately 10 g/L in 50 % ACN while SA was prepared as a saturated solution in 30 % ACN. Sample preparation for MALDI-TOF-MS involved dilution of the intact Grx or Grx digest with one of the matrix solutions to give a final concentration between 1-10 pmol/µL. About 1 µL of the analyte/matrix solution was applied to the probe target and allowed to evaporate. MALDI-TOF mass spectra were acquired using the modified Vestec VT 2000 instrument (Section 1.3). A Laser Science N₂ laser (LSI, Inc. Cambridge, MA) was operated at a repetition rate of 5 Hz. Laser irradiation of the sample produced protonated (mostly singly and doubly charged) ions which were analyzed by the TOF mass spectrometer using 30 kV accelerating voltage. The isotope multiplets were not resolved (the resolution is generally between 1:100 and 1:300) and the resulting m/z values are thus derived from the polyisotopic envelope at full-width of the peak at half-height.

RP-HPLC was carried out using Waters model 510 HPLC pumps, a 680 gradient controller, a Rheodyne 7125 injector, and a Waters model 490 UV multiwavelength detector monitoring at 214 nm. For fractionation of digests, a Vydac C₄ column (25 cm x 4.6 mm I.D.) was used at a flow rate of 0.5 mL/min with a linear gradient of 100 % H₂O/TFA (1:0.0005) to 40 % H₂O/TFA (1:0.0005) and 60 % ACN/TFA (1:0.00035) over 60 min. For purification after reduction and alkylation of Grx, a short Vydac C₄ guard column (2 cm x 4.6 mm I.D.) was used at a flow rate of 2 mL/min initially at 100 %
H₂O/TFA (1:0.0005) changing directly to 80 % ACN/TFA (1:0.00035) after the solvent front had passed the detector.

In order to resolve the ambiguity between Gln or Lys in the final position of the C-terminal peptide, the digest fraction containing this peptide was dried and redissolved in 40 μL glacial acetic acid and 10 μL acetic anhydride and kept at 37 °C for 1 h. The solvent was then removed and the peptide analyzed by FABMS.

All water used in these experiments was filtered and deionized using a Milli-Q™ system. ACN and solvents other than water were chromatographic grade and were obtained from either Fisher or Baker. A Fisher Accumet model 15 pH meter and standard Ag/AgCl reference combination glass electrode were used to measure the pH of mL quantities, e.g., buffer solutions. Otherwise "colorpHast" (EM Science) indicator strips were used to check pH. Urea, guanidine-HCl, glycerol, and thioglycerol were purchased from Fluka, Trizma base and EDTA from Sigma, enzymes from Boehringer Mannheim, a solution of 100 mM ammonium acetate (pH 7.8) from Pierce, acetic anhydride from Mallinckrodt, glacial acetic acid and concentrated HCl from Fisher, triethylphosphine, 4-vinylpyridine, iodoacetamide, and all MALDI matrix compounds from Aldrich. The solvent from collected HPLC fractions was removed using a Savant speed-vac concentrator.
III.3 Results and Discussion

The primary structure of hErGrx was determined solely by tandem mass spectrometry following a strategy used previously in this laboratory to determine the amino acid sequence of rabbit bone marrow Grx.\textsuperscript{122} In addition, this work with hErGrx also provided an opportunity to demonstrate the utility of MALDI-TOF-MS in protein sequencing. The Grx protein, (11.3 kDa by SDS-PAGE), was isolated and purified from human erythrocytes at Case Western Reserve University.\textsuperscript{114} A more accurate molecular weight determination of the native protein was performed using MALDI-TOF-MS which typically requires only about 1 pmol of material.\textsuperscript{a} Sinapinic acid (SA) was chosen as the matrix for these MALDI-TOF experiments because it provided the best signal intensity and resolution for the Grx protein at the 337 nm wavelength generated by the N\textsubscript{2} laser. The resultant protonated molecular ions, (M+H)\textsuperscript{+} and (M+nH)\textsuperscript{+n}, were mass analyzed by the TOF mass spectrometer. The MALDI-TOF mass spectrum of the isolated hErGrx (shown in Figure III.1A) exhibited singly and doubly charged protonated ions, [(M+H)\textsuperscript{+} and (M+2H)\textsuperscript{2+}, respectively], of the protein molecule. The structure of these peaks (in contrast to the sharper and better resolved peaks of the horse heart myoglobin used as an internal standard), implied that the material was not homogeneous and that the major component

\textsuperscript{a}While FABMS can also be used to determine the molecular weight of proteins, (for example, the M\textsubscript{r} of a similar protein, Trx (M, 11,750) isolated from Chromatium vinosum was determined by FABMS\textsuperscript{125}), it is a difficult experiment to perform and sample consumption (typically several nmol) is very high due to the decrease in sensitivity with increasing mass.
Figure III.1. MALDI-TOF mass spectra of hErGrx. A. Material isolated and purified from RBC. B. Reduced with triethylphosphine. C. Reduced and S-ethylpyridylated. The ions labeled "Std (M+H)" and "Std (M+2H)" at m/z 16,952.5 and m/z 8476.8, respectively, are due to the internal standard (horse heart myoglobin). The matrix was sinapinic acid and the low intensity peaks adjacent to the major peaks are adducts of the matrix.
had a $M_r$ of 11,841. Reduction of Grx with either triethylphosphine or DTT resulted in a decrease in the mass of the protein to 11,688 Da. The mass spectrum (shown in Figure III.1B) exhibits a somewhat broadened peak because the reduced glutaredoxin ionizes poorly (i.e., requires higher laser power). On the other hand, the mass spectrum of the S-ethylpyridylated Grx (shown in Figure III.1C) produces a single sharp peak for the $(M+H)^+$ ion at $m/z$ 12,215. The amino acid analysis performed prior to the primary structure determination predicted that the hErGrx protein, (consistent with the close homology of the other mammalian Grx proteins to each other$^{120-122}$), contained only four cysteines.$^{114}$ However, based on the $M_r$ of the reduced Grx (11,688), the increase in mass for the S-ethylpyridylated protein (525 Da) corresponded to the value calculated for alkylation by five ethylpyridyl groups ($5 \times 105 = 525$). Thus, the hErGrx contains five cysteine groups. This conclusion was confirmed by reducing and alkylating hErGrx (with DTT and iodoacetamide, respectively) to give the S-carbamidomethylated product which generated a $(M+H)^+$ ion at $m/z$ 11,974 by MALDI-TOF-MS. Once again, the increase in mass (285 Da) of the S-carbamidomethylated compared to the reduced protein agreed very well with the calculated value for alkylation by five carbamidomethyl groups (285 Da).

The heterogeneity of the isolated Grx (Figure III.1A) is attributed to the modification of some of the cysteines in the isolation and purification process. The fact that the reduced Grx ionized poorly in contrast to the S-alkylated and native protein provides additional support for this hypothesis. The reduction in mass by 153 Da corresponds to the
conversion of two -S-S-CH₂-CH₂-OH groups to -SH (calculated difference: 2 x 76 = 152 Da). However, the shape of the peak in Figure III.1A does not exclude the possibility that one, two, and three cysteines have reacted in a 1:2:1 ratio rather than two completely modified cysteines. The scheme shown below provides a plausible explanation for these observations. Treatment of the reduced Grx with hydroxyethyldisulfide (HEDS), routinely performed at the last step of purification, could give the distribution of protein forms shown.

The S-carbamidomethylated Grx was digested with trypsin and the digest partially fractionated by RP-HPLC. The molecular weight of the peptides in each HPLC fraction was determined by FABMS and the amino acid sequence by tandem mass spectrometry.
However, three of the tryptic peptides \((M+H)^+ = m/z\) 2309, \(m/z\) 2854, and \(m/z\) 3292 were too large to be unambiguously sequenced and were further digested with Endo-Asp-N. The sequence of the resulting peptides permitted (together with overlapping sequences derived separately from chymotryptic peptides) the assembly of the sequences of these three tryptic peptides. Overlapping sequence information was furnished by digestion of the S-ethylpyridylated Grx with \(\alpha\)-chymotrypsin. In a separate experiment, larger peptides were produced by digesting the S-ethylpyridylated protein with Endo-Lys-C. The latter digest provided confirming information for the alignment deduced from the tryptic and chymotryptic peptides. As with the tryptic digest, the chymotryptic and Endo-Lys-C digests were partially fractionated and the molecular weights of the peptides were determined by FABMS. The amino acid sequences (except for the larger Endo-Lys-C peptides) of the peptides were also deduced from the CID mass spectra. The FABMS data resulting from the tryptic, chymotryptic, and Endo-Lys-C digests of hErGrx is shown in Tables III.1, III.2, and III.3, respectively. Figures III.2, III.3, and III.4 provide examples of CID mass spectra of Grx peptides and the interpretation of these spectra (as shown by the annotations) leading to the sequences displayed at the top of each spectrum. The complete amino acid sequence of hErGrx derived from the peptides generated by the various enzymatic digests of the protein is shown in Figure III.5.

A comparison of pig liver\textsuperscript{121}, calf thymus\textsuperscript{120}, rabbit bone marrow\textsuperscript{122}, and the human Grx amino acid sequences is shown in Figure III.6. Like the other mammalian Grxs,
Table III.1. FABMS data from the tryptic digest (and further digestion with Endo-Asp-N) of S-carbamidomethylated hErGrx.

<table>
<thead>
<tr>
<th>HPLC Frac.</th>
<th>Tryptic Peptide(^a) (M+H)(^+)</th>
<th>Sequence (from CID)</th>
<th>Tryptic/Asp-N(^a) (M+H)(^+)</th>
<th>Sequence (from CID)</th>
<th>Position(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>472.1</td>
<td>TVPR</td>
<td></td>
<td></td>
<td>68-71</td>
</tr>
<tr>
<td>1</td>
<td>542.1</td>
<td>IQPGK(^c)</td>
<td></td>
<td></td>
<td>9-13</td>
</tr>
<tr>
<td>6</td>
<td>563.4</td>
<td>VFIGK</td>
<td></td>
<td></td>
<td>72-76</td>
</tr>
<tr>
<td>7</td>
<td>629.4</td>
<td>QIGALQ</td>
<td></td>
<td></td>
<td>100-105</td>
</tr>
<tr>
<td>10</td>
<td>870.4</td>
<td>LKQIGALQ(^d)</td>
<td></td>
<td></td>
<td>98-105</td>
</tr>
<tr>
<td>10</td>
<td>1037.3</td>
<td>Ac-AQEFVNCK</td>
<td></td>
<td></td>
<td>1-8</td>
</tr>
<tr>
<td>13</td>
<td>1638.7</td>
<td>VVVFIKPTCPYCR</td>
<td></td>
<td></td>
<td>14-26</td>
</tr>
<tr>
<td>14</td>
<td>1395.9</td>
<td>RAQEILSQLPIK</td>
<td></td>
<td></td>
<td>27-38</td>
</tr>
<tr>
<td>15</td>
<td>1239.7</td>
<td>AQEILSQLPIK</td>
<td></td>
<td></td>
<td>28-38</td>
</tr>
<tr>
<td>15</td>
<td>1322.6</td>
<td>VVVFIKPTCPY(^e)</td>
<td></td>
<td></td>
<td>14-24</td>
</tr>
<tr>
<td>16</td>
<td>2309</td>
<td></td>
<td>768.3</td>
<td>DCIGGCS</td>
<td>77-83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1559.0</td>
<td>DLVSLQSGELLTR</td>
<td>84-97</td>
</tr>
<tr>
<td>17</td>
<td>2854</td>
<td></td>
<td>563.2</td>
<td>VFIGK</td>
<td>72-76</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>768.3</td>
<td>DCIGGCS</td>
<td>77-83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1558.8</td>
<td>DLVSLQSGELLTR</td>
<td>84-97</td>
</tr>
<tr>
<td>22</td>
<td>3293</td>
<td></td>
<td>520.3</td>
<td>DITAT</td>
<td>46-50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>805.5</td>
<td>QGLLEFV</td>
<td>39-45</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>856.3</td>
<td>DHTNEIQ</td>
<td>51-57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1164.7</td>
<td>DYLQQLTGAR</td>
<td>58-67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1356.7</td>
<td>DITATNHTNEIQ</td>
<td>46-57</td>
</tr>
</tbody>
</table>

\(^a\)Monoisotopic (M+H)\(^+\) = m/z (in italics) determined by FABMS.

\(^b\)According to the final sequence determined in this work.

\(^c\)Ile determined from the CID mass spectrum of chymotryptic peptide (M+H)\(^+\) = m/z 1535.9.

\(^d\)The N-terminal Leu was determined from the CID mass spectrum of the chymotryptic peptide (M+H)\(^+\) = m/z 1127.8.

\(^e\)The Lys preceding Pro was identified from the CID spectrum of an Endo-Lys-C peptide having this Lys at its C-terminus.
Table III.2. FABMS data for peptides resulting from the chymotryptic digestion of S-ethylpyridylated hErGrx.

<table>
<thead>
<tr>
<th>HPLC Fraction</th>
<th>Chymotryptic Peptide (M+H)⁺</th>
<th>Sequence Derived by CID</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>757.6</td>
<td>KQIGALQ</td>
<td>99-105</td>
</tr>
<tr>
<td>7</td>
<td>501.2</td>
<td>LQQL⁺</td>
<td>60-63</td>
</tr>
<tr>
<td>7</td>
<td>536.2</td>
<td>Ac-AQEF</td>
<td>1-4</td>
</tr>
<tr>
<td>7</td>
<td>573.4</td>
<td>AQEIL</td>
<td>28-32</td>
</tr>
<tr>
<td>7</td>
<td>629.5</td>
<td>QIGALQ</td>
<td>100-105</td>
</tr>
<tr>
<td>7</td>
<td>729.3</td>
<td>RAQEIL</td>
<td>27-32</td>
</tr>
<tr>
<td>7</td>
<td>885.7</td>
<td>RRAQEIL</td>
<td>26-32</td>
</tr>
<tr>
<td>7</td>
<td>926.5</td>
<td>IKPTCPY⁺</td>
<td>18-24</td>
</tr>
<tr>
<td>7</td>
<td>1093.8</td>
<td>CRRAQEIL</td>
<td>25-32</td>
</tr>
<tr>
<td>8</td>
<td>870.6</td>
<td>LKQIGALQ⁺</td>
<td>98-105</td>
</tr>
<tr>
<td>8</td>
<td>960.5</td>
<td>VSLQQSGEL⁺</td>
<td>86-94</td>
</tr>
<tr>
<td>8</td>
<td>1127.8</td>
<td>TRLKQIGALQ</td>
<td>96-105</td>
</tr>
<tr>
<td>9</td>
<td>718.5</td>
<td>TVPRVF</td>
<td>68-73</td>
</tr>
<tr>
<td>9</td>
<td>1103.8</td>
<td>TGARTVPRVF</td>
<td>64-73</td>
</tr>
<tr>
<td>10</td>
<td>1174.6</td>
<td>VSLQQSGELL⁺</td>
<td>86-96</td>
</tr>
<tr>
<td>11</td>
<td>1073.5</td>
<td>VSLQQSGELL⁺</td>
<td>86-95</td>
</tr>
<tr>
<td>11</td>
<td>1096.8</td>
<td>SQLPIKQGILL</td>
<td>33-42</td>
</tr>
<tr>
<td>11</td>
<td>1353.9</td>
<td>VNCKIQPGKVVVF</td>
<td>5-17</td>
</tr>
<tr>
<td>14</td>
<td>1372.7</td>
<td>SQLPIKQGLLEF</td>
<td>33-44</td>
</tr>
</tbody>
</table>

aMonoisotopic (M+H)⁺ = m/z (in italics) determined by FABMS.
bGln and Lys were differentiated based on this specificity of Endo-Lys-C (see Table III.3).
cAccording to the final sequence determined in this work.
dThe N-terminal Leu was determined from the CID spectrum of tryptic/Asp-N peptide (M+H)⁺ = m/z 1164.7.
eIle was determined from the CID spectrum of tryptic peptide (M+H)⁺ = m/z 1638.7.
fThe N-terminal Leu was determined from the CID spectrum of chymotryptic peptide (M+H)⁺ = 1127.8.
gThe Leu at position 88 was differentiated from Ile from the CID mass spectrum of
tryptic/Endo-Asp-N peptide (M+H)^+ = m/z 1559.0.

Table III.3. FABMS data for peptides resulting from the Endo-Lys-C digestion of S-ethylpyridylated hErGrx.

<table>
<thead>
<tr>
<th>HPLC Frac.</th>
<th>Endo-Lys-C Peptide^a (M + H)^+</th>
<th>Sequence Derived by CID</th>
<th>Position^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>542.2</td>
<td>IQPGKe</td>
<td>9-13</td>
</tr>
<tr>
<td>3</td>
<td>629.3</td>
<td>QIGALQd</td>
<td>100-105</td>
</tr>
<tr>
<td>4</td>
<td>1085.6</td>
<td>Ac-AQEFVNCK</td>
<td>1-8</td>
</tr>
<tr>
<td>5</td>
<td>704.5</td>
<td>VVVFIK</td>
<td>14-19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HPLC Frac.</th>
<th>Endo-Lys-C Peptide^a (M + H)^+</th>
<th>Assigned Sequence^e</th>
<th>Calc'd^a (M+H)^+</th>
<th>Position^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>2426.2</td>
<td>PTCPYCRRAQEILSQLPIK</td>
<td>2426.3</td>
<td>20-38</td>
</tr>
<tr>
<td>8</td>
<td>2645.0</td>
<td>DCIGGCSDLVSLQGSELLTRLK</td>
<td>2645.3</td>
<td>77-99</td>
</tr>
<tr>
<td>8</td>
<td>3113.3</td>
<td>VVVFIKPTCPYCRRAQEILSQLPIK</td>
<td>3111.7</td>
<td>14-38</td>
</tr>
<tr>
<td>11</td>
<td>4289</td>
<td>QGLLEFVDITAT(D/N)HTNEIQDYLQQLTGARTVPRVFIGKf</td>
<td>4288.8g</td>
<td>39-76</td>
</tr>
</tbody>
</table>

^a Monoisotopic (M+H)^+ = m/z values are indicated in italics.
^b According to the final sequence determined in this work.
^c Ile was determined from the CID spectrum of chymotryptic peptide (M+H)^+ = m/z 1535.9.
^d The C-terminal Gln was determined by N-acetylation and determination of the mass difference by FABMS (see text).
^e The peptide sequences were assigned on the basis of molecular weight, the specificity of the Endo-Lys-C protease, and CID derived data from the tryptic and chymotryptic digests of hErGrx (Tables III.1 and III.2).
^f The signal due to the (M+H)^+ ion of this peptide was not sufficiently strong to permit resolving the 1 Da difference between Asp and Asn at position 51.
^g The calculated (polyisotopic) value is based on Asn at position 51.
Figure III.2. FAB CID mass spectra of hErGrx peptides (M+H)^+ = m/z 1037.3 and m/z 1356.7. Ac-AQEFVNCK and DITATNHTNEIQ are tryptic and tryptic/Endo-Asp-N peptides, respectively from S-carbamidomethylated hErGrx.
Figure III.3. FAB CID mass spectra of hErGrx peptides (M+H)$^+$ = m/z 1127.8, m/z 1164.6, and m/z 1395.9. These spectra are examples of Ile/Leu differentiation due to the formation of the side chain specific cleavage ions, (d$_n$ and w$_n$ in high energy CID).
Figure III.4. FAB CID mass spectra of hErGrx peptides (M+H)$^+ = m/z$ 629.4 and $m/z$ 672.5. The deduced peptide sequence is the same for both peptides (QIGALQ). While the $m/z$ values of the C-terminal ions (e.g., $y_n$) are identical, the masses of the N-terminal ions in the lower mass spectrum are 43 Da higher. The abundant $b_1$ ion at $m/z$ 172.3 (absent in the [top] mass spectrum of the unadducted peptide) confirms that this peptide is the N-terminally carbamylated peptide.
Figure III.5. Amino acid sequence of hErGrx determined by tandem mass spectrometry. a. Endo-Lys-C peptides. b. tryptic peptides. c. chymotryptic peptides. d. peptides resulting from digestion of tryptic HPLC fractions with Endo-Asp-N. e. An additional peptide from tryptic/Endo-Asp-N digestion having Asn instead of Asp as indicated by the asterisk. The (M+H)^+ value (obtained by FABMS and thus monoisotopic) is displayed above the line segments representing the peptides. Heavy underlining denotes sequence data obtained from CID mass spectra. Light underlining represents peptides for which the molecular weight was determined.
Figure III.6. Comparison of pig liver, calf thymus, rabbit bone marrow, and human erythrocyte glutaredoxins amino acid sequences showing their alignment and homology. Identification of Asn as well as Asp at position 51 in the human sequence is indicated by the asterisk. Identical residues are framed.
hErGrx also contains an acetylated Ala at the N-terminus which was determined from the CID mass spectrum of the N-terminal tryptic peptide \([(\text{M+H})^+ = \text{m/z} \ 1085.7]\) of S-ethylpyridylated hErGrx (Figure III.7). The presence of the \(b_1\) ion at \( \text{m/z} \ 114.0 \) indicated that acetylated Ala (rather than un-acetylated Leu or Ile which have the same mass) is at the N-terminus since \(b_1\) ions in CID mass spectra of peptides with a free (basic) N-terminal amino acid are generally not observed.\(^{122}\) In addition, the absence of Leu or Ile in the peptide sequence is further confirmed by the lack of a peak at \( \text{m/z} \ 86 \) which corresponds to the immonium ion for these isomeric amino acids. The CID mass spectrum of the tryptic peptide from S-carbamidomethylated Grx (top of Figure III.2) also contains a \(b_1\) ion at \( \text{m/z} \ 114.0 \) and served to confirm the deduced sequence.

Despite the similarities of hErGrx to the other mammalian Grx proteins, there are a number of differences. Of particular interest is the discovery of a fifth cysteine near the N-terminus of hErGrx. The N-terminal tryptic peptide from the S-ethylpyridylated Grx (Figure III.7) also contains an abundant ion at \( \text{m/z} \ 106.0 \) corresponding to the ethylpyridyl group and a high mass ion at \( \text{m/z} \ 979.2 \) which arises from the loss of the ethylpyridyl moiety from the \( (\text{M+H})^+ \) precursor ion. Tryptic digestion of the S-carbamidomethylated Grx gave an analogous \( (\text{M+H})^+ \) ion at \( \text{m/z} \ 1037.3 \) and the subsequent CID mass spectrum (top of Figure III.2) of this peptide provided additional confirmatory evidence for the

\(^{*}\text{The m/z values in italic type denote monoisotopic mass as discussed previously in this section.}\)
Figure III.7. FAB CID mass spectrum of the N-terminal tryptic peptide from hErGrx (M+H)$^+$ = m/z 1085.7.
position of this fifth cysteine.

The hErGrx contains a His whose position in the sequence is not homologous with the His-89 in the Grx proteins from pig liver and calf thymus. The CID mass spectrum of the tryptic/Endo-Asp-N peptide (M+H)$^+$ = \text{m/z} 856.3 (shown in Figure III.8) exhibits an abundant ion at \text{m/z} 110.3, the immonium ion of His. The a$_7$ and b$_7$ ions, (\text{m/z} 225.4 and \text{m/z} 253.4, respectively), correspond to a peptide with Asp and His at the amino terminus. The Asp was identified as the N-terminus of this peptide (finally assigned to position 51) on the basis of the specificity of Endo-Asp-N and the presence of the y$_6$ and z$_6$ $^+$ ions, (\text{m/z} 741.4 and \text{m/z} 725.4, respectively). Thus, His is the second residue in this peptide and was assigned to position 52 in the final amino acid sequence. A second peptide generated from the tryptic/Endo-Asp-N digest, (M+H)$^+$ = \text{m/z} 1356.7 (bottom spectrum of Figure III.2), provided confirmatory evidence for the position of His in hErGrx. In addition, the presence of a strong signal at \text{m/z} 770.4 for the c$_7$ ion provided further confirmation for Thr after His since c$_n$ ions are generally not very abundant in CID mass spectra except at positions preceding Thr.$^{76}$ This peptide (together with the peptide of (M+H)$^+$ = \text{m/z} 856.3) revealed an apparent microheterogeneity in the protein since Asn instead of Asp was identified at position 51 for the peptide in Figure III.2. Further confirmation of this finding came from the fact that the measured \text{m/z} of this peptide was 0.9 Da less than that expected from a peptide with Asp at position 51 that resisted cleavage. (The residue masses of Asp and Asn are 115.0 and 114.0, respectively). The assignment of Asn rather than Asp (except
Figure III.8. FAB CID mass spectrum of the His containing peptide (M+H)+ = m/z 856.3.
Figure III.9. FAB CID mass spectrum of the tryptic peptide $(M+H)^+ = m/z 1559.0$. 
at the N-terminus) in this peptide is consistent with the fact that Endo-Asp-N cleaves at the 
peptide bond N-terminal to Asp but not Asn. Furthermore, the results of isoelectric 
focusing of hErGrx in agarose indicated the presence of two pl forms whose observed pH 
values corresponded to those predicted with either Asp or Asn at position 51. However, 
subsequent cloning and sequencing of the cDNA of hErGrx identified Asn as the 
amino acid at position 51. Thus, a plausible explanation is that the Asp-51 is due to 
deamidation of the Asn presumably occurring during the isolation and purification of the 
protein. On the other hand, this was surprising in light of the fact that none of the other 
eleven Asn and Gln residues (including a near neighbor at Asn-54) were similarly 
affected.

The hErGrx protein is unique among the other Grx proteins in the absence of Met at 
position 88 which is in agreement with the amino acid analysis and lack of CNBr reactivity 
which indicated the absence of this amino acid. The CID mass spectrum of the 
tryptic/Endo-Asp-N peptide (M+H)\(^+\) = m/z 1559.0 (shown in Figure III.9) has a nearly 
complete \(w_n\) ion series. The \(b_2\) and abundant \(w_{10}\) ions (m/z 528.6 and m/z 1085.7, 
respectively) together identify Leu as the amino acid at position 88. (The side chain 
specific cleavage resulting in the formation of the \(w_{10}\) ion also serves to differentiate Leu 
from Ile). The identity of all twenty-one Ile and Leu residues (which are isomeric) could be 
determined based on the \(d_n\) or \(w_n\) ions which involve cleavage of the side chains at the \(\beta,\gamma\)-
carbon bond. Further examples of CID mass spectra where Ile and Leu were
differentiated on the basis of $d_n$ or $w_n$ ions are shown in Figure III.3. The top and bottom mass spectra also demonstrate how the presence of a basic amino acid (Arg) near the N-terminus directs the fragmentation to produce mainly N-terminal ions of which many are $d_n$ ions. Conversely, the middle spectrum has Arg at the C-terminus and consequently, the C-terminal $w_n$ and $v_n$ ions predominate.

The isobaric Lys and Gln were differentiated on the basis of the specificity of Endo-Lys-C. The only exception was the C-terminal Gln which was identified by N-acetylation of the C-terminal tryptic peptide $(M+H)^+ = m/z \ 629.4$. The $m/z$ of this peptide after N-acetylation shifted to $671.4$ (i.e., monoacetylation) due to the reaction of the free N-terminal amino group, rather than to $m/z \ 713.4$ (i.e., diacetylation), the value expected if a Lys was part of this molecule. The proteolytic digests were typically carried out over longer time periods (12 - 24 h) to insure complete cleavage and therefore proper identification of internal Lys and Gln residues. A side effect of the longer incubation times was the formation of a number of N-terminally carbamylated peptides due to the presence of the denaturant urea. Fortunately, these peptides are easily distinguished on the basis of their mass (the carbamylated peptides are 43 Da higher in mass) and from their CID mass spectra. An example of a tryptic peptide and its corresponding N-terminally carbamylated peptide is shown in Figure III.4. The $m/z$ values for the C-terminal ions are identical in both mass spectra while the N-terminal ions are 43 Da higher for the adducted peptide. Furthermore, the carbamylated peptide generates an abundant $b_1$ ion (absent in the spectrum
of the unadducted peptide) at m/z 172.3 which, as already noted above, is indicative of N-terminal modification. Thus, the unintended consequence of the presence of these carbamylated peptides was to provide additional supporting evidence for the assigned sequences.

The top CID mass spectrum in Figure III.4 of the uncarbamylated peptide (QIGALQ) is unique in containing at least four N-terminal fragment ions of the type a$_n$-17 and b$_n$-17. The source of these ions is the particularly facile deamidation of the N-terminal Gln to form pyroglutamic acid$^{129}$ as shown below:

\[
\begin{align*}
\text{H}_2\text{N}-\text{C} & \text{CH}_2 \\
\text{CH}_2 \text{O} & \\
\text{H-N-CH-C---} & \rightarrow \\
\text{H} & \text{O} \\
\end{align*}
\]

The experimentally determined M$_r$ (11,688) of the reduced hErGrx by MALDI-TOF-MS was in close agreement with the calculated M$_r$ (11686.6) based on the mass spectrometrically derived sequence (for Asn at position 51). Similarly, the M$_r$ (also determined by MALDI-TOF-MS) of the S-carbamidomethylated (11,973) and of the S-
ethylpyridylated (12,214) proteins also agrees well with the calculated values of 11,971.9 and 12,212.3, respectively. These results establish the mass of the human Grx protein and clarify the ambiguity that existed previously because of the earlier estimation of 6500 Da for the molecular weight of human placental thioltransferase. \(^{130}\) MALDI-TOF-MS was also used to confirm the hErGrx sequence by the analysis of unfractionated proteolytic digests. For example, the MALDI-TOF mass spectra of Endo-Lys-C and Endo-Asp-N digests of S-ethylpyridylated hErGrx are shown in Figure III.10. These digests generate larger peptides which are more efficiently ionized by MALDI. The data from Figure III.10 for the Endo-Lys-C and Endo-Asp-N digests is compiled in Tables III.4 and III.5. The sequences listed were not determined by these experiments but are shown for the convenience of the reader. Although some of the smaller peptides were not detected by the MALDI experiments, taken together, the information from the two digests provided confirmation for the entire hErGrx sequence.

In the course of sequencing hErGrx, a few peptides were observed whose masses did not fit the final sequence. These peptide ions, (which tended to be weaker signals), could have resulted from non-specific cleavage of hErGrx. Two of the tryptic peptides happened to ionize well enough to allow acquisition of good quality CID mass spectra (shown in Figure III.11). Clearly, the sequences of these peptides is not homologous with any of the Grx sequences shown in Figure III.6. The masses and sequences of these "unidentified" peptides were searched using a protein database library. \(^{131}\) The results of
Figure III.10. MALDI-TOF mass spectra of S-ethylpyridylated hErGrx proteolytic digests. A. Endo-Lys-C digest using SA as the matrix. B. Endo-Asp-N using ACCA. In the upper mass spectrum, the doubly charged ions, [(M+2H)⁺], due to peptides 77-99, 14-38, and 39-76 are indicated by "++". The peak annotated with an asterisk in the lower mass spectrum is due to a contaminant from porcine insulin (indicated by "Ins") which was used as an internal standard.
Table III.4. MALDI-TOF mass spectral data from Endo-Lys-C digestion of S-ethylpyridylated hErGrx from Figure III.10.*

<table>
<thead>
<tr>
<th>Position</th>
<th>Calc'd (M+H)$^+$</th>
<th>Obs'd (M+H)$^+$</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-8</td>
<td>1086.26</td>
<td>1086.26*</td>
<td>( ) Ac-AQEFVNCK</td>
</tr>
<tr>
<td>9-13</td>
<td>542.66</td>
<td>-</td>
<td>(K) IQPGK (V)</td>
</tr>
<tr>
<td>14-19</td>
<td>704.93</td>
<td>-</td>
<td>(K) VVVFIK (P)</td>
</tr>
<tr>
<td>14-38</td>
<td>3113.84</td>
<td>3113.8</td>
<td>(K) VVVFIKPTCPYCRRAQEILSQLPIK (Q)</td>
</tr>
<tr>
<td>20-38</td>
<td>2427.94</td>
<td>2427.6</td>
<td>(K) PTCYCRRAQEILSQLPIK (Q)</td>
</tr>
<tr>
<td>39-76</td>
<td>4288.85</td>
<td>4288.85*</td>
<td>(K) QGLLEFVDIT...QLTGARTVPRVFIGK (D)</td>
</tr>
<tr>
<td>77-99</td>
<td>2647.09</td>
<td>2646.5</td>
<td>(K) DCIGGCSDLVSLQQSGELLTRLK (Q)</td>
</tr>
<tr>
<td>100-105</td>
<td>629.73</td>
<td>-</td>
<td>(K) QIGALQ ( )</td>
</tr>
</tbody>
</table>

*Shown are the expected Endo-Lys-C cleavage products and the peptides detected. The $m/z$ values marked with an asterisk were used as internal standards. The smaller peptides (ionized by FABMS) are more difficult to ionize by MALDI-TOF-MS and were not detected.
Table III.5. MALDI-TOF mass spectral data from Endo-Asp-N digestion of S-ethylpyridylated hErGrx from Figure III.10.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Position</th>
<th>Calc'd (M+H)$^+$</th>
<th>Obs'd (M+H)$^+$</th>
<th>Amino Acid Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-45</td>
<td>5491.65</td>
<td>5491.7</td>
<td>( ) Ac-AQEFVNCKIQPG...QGLLEFV (D)</td>
</tr>
<tr>
<td>46-57</td>
<td>1357.42</td>
<td>-</td>
<td>(V) DITATNHTNEIQ (D)</td>
</tr>
<tr>
<td>58-76</td>
<td>2163.53</td>
<td>2164.6</td>
<td>(Q) DYLQQLTGARTVPRVFIGK (D)</td>
</tr>
<tr>
<td>77-83</td>
<td>865.02</td>
<td>865.02*</td>
<td>(K) DCIGGCS (D)</td>
</tr>
<tr>
<td>84-105</td>
<td>2411.81</td>
<td>2412.1</td>
<td>(S) DLVSLQQSGELLTRLKQIGALQ ( )</td>
</tr>
<tr>
<td>93-105</td>
<td>1483.80</td>
<td>1484.2</td>
<td>(G) ELLTRLKQIGALQ ( )</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Shown are the expected Endo-Asp-N cleavage products and the peptides detected. Although the cleavage rate is generally much faster at the N-terminus of Asp, one N-terminal cleavage at Glu (93-105) was observed. Porcine insulin, [(M+H)$^+$ = \textit{m/z} 5778.60], and peptide 77-83 (denoted by the asterisk) were used to internally calibrate the spectrum. The presence of Asp rather than Asn at position 51 would explain the absence of peptide 46-57.
Figure III.11. FAB CID mass spectra of tryptic peptides $(M+H)^+ = m/z$ 1001.2 and 1422.3 which were found to be due to contamination of the hERGrx preparation by other proteins or fragments thereof (see text).
this search identified the sequence, GIPTLLLFK, [(M+H)\(^+\) = m/z 1167.6] as a tryptic peptide from *E. coli* thioredoxin while three other peptides (M+H)\(^+\) = m/z 765.5, m/z 1422.3 (VHVGDEDFVHLR as deduced by CID-MS) and m/z 2457.4 (whose ion signal was too weak to allow unambiguous sequence determination) corresponded to tryptic peptides from human cystatin-B. Therefore, the preparation used for the mass spectrometric sequence determination of hErGrx contained fragments from at least two other proteins, (human cystatin-B and *E. coli* thioredoxin\(^4\)), although the concentration of these contaminants was probably quite low in comparison to hErGrx. In retrospect, it is not at all surprising that only the *E. coli* peptide GIPTLLLFK was found, because it is easily ionized since it consists primarily of hydrophobic amino acids (and has Lys at the C-terminus) and therefore would be expected to ionize well by FAB even at lower levels.\(^{29}\) The peptide VHVGDEDFVHLR produced a strong signal for the same reasons. The CID mass spectra in Figure III.11 also serve as additional examples of ion types known to arise from specific amino acids or sequences. Thus, in the CID spectrum of peptide GIPTLLLFK, the \(y_{-2}\) ion provides further evidence for Pro at that position and the relatively higher abundance of the \(c_3\) ion confirms that the Pro precedes Thr. This peptide is also somewhat unusual in generating abundant internal fragment ions indicated by the "PT" and "IPT/PTL" annotations. In addition, both of these internal fragment ions also undergo loss of H\(_2\)O (18 Da) presumably from the Thr side chain. The lower CID mass spectrum consists

\(^4\)These particular proteins were used as molecular weight markers for SDS-PAGE of hErGrx in the same laboratory where the isolation of hErGrx was performed.
predominantly of C-terminal ions due to the basic Arg at the C-terminus. The abundant ion at \( m/z \ 110.2 \) in the immonium ion region of the spectrum indicates that His is in the sequence. It is interesting to note that in both of the His containing mass spectra (Figures III.8 and III.11 [bottom]), His is in the second position and that a secondary ion, \(^{75}\) \( ^{d_{2D}} \) and \( ^{d_{2V}} \), respectively, is formed. A secondary \( ^{d_{nX}} \) ion results from the fragmentation of the \( n - 1 \) side chain (where \( X \) stands for the single letter code for the \( n - 1 \) residue).

III.4 Conclusion

The complete amino acid sequence of a human thioltransferase (Grx) isolated from erythrocytes was determined solely by high performance tandem mass spectrometry. In addition, molecular weight determination of the native and S-alkylated hErGrx was accomplished by MALDI-TOF-MS which established that hErGrx has five instead of four cysteines. The MALDI molecular weight determination of hErGrx (M, 11,688) also served to clarify the ambiguity caused by the previous estimate (M, 6500 Da) for the molecular weight of a human thioltransferase.\(^{130}\) The final primary structure of hErGrx was assembled from tryptic and chymotryptic peptides whose sequences were deduced by CID mass spectrometry. Finally, unfractionated protein digests of hErGrx were also analyzed by MALDI-TOF-MS. The resultant mass spectra provided additional confirmatory evidence for the assigned primary structure.
The hErGrx protein was found to be quite homologous to the other mammalian Grx proteins of known structure and also contains an acetylated N-terminus and lacks Trp. Like the pig liver and calf thymus Grx proteins hErGrx contains a His residue, however, its location is at position 52 rather than 89. In contrast to the other Grx proteins, hErGrx does not contain any Met. Of particular interest is the discovery of a fifth cysteine near the N-terminus of hErGrx which could be due to a silent mutation or a functionally significant change. It is noteworthy that the human thioredoxin (Trx) also has an "extra" fifth Cys residue compared to other mammalian Trx proteins. It has been hypothesized that this Cys could serve to anchor the hErGrx to the red blood cell membrane while retaining catalytic activity at its active cite (Cys-22/Cys-25).
Section IV

Determination of the Primary Structure of a Mussel Adhesive Protein

from *Mytilus edulis* and Identification of a Novel Amino Acid

IV.1 Introduction

The marine mussel, *Mytilus edulis*, is particularly adept at attaching its foot to underwater surfaces through the byssus, a thread-like structure ending in adhesive pads or plaques.\(^1\)\(^{33}\) The process of this surface bonding is of considerable technological and scientific interest since it is quite strong, rugged, and persists in spite of prolonged immersion in seawater. Furthermore, the mussel foot proteins involved in this adhesion process are naturally occurring and presumably environmentally benign. Mature byssal adhesive plaques exhibit extensive protein cross-linking which renders isolation of individual protein components from these plaques difficult or impossible. However, the cross-linking of freshly secreted plaque precursor proteins can be prevented by lowering the temperature (cold shock)\(^1\)\(^{34}\) which then allows their isolation. At least four distinct families of polymorphic proteins named mefp (for *Mytilus edulis* foot protein) have previously been isolated in this manner from the foot and byssal adhesive plaques of *Mytilus edulis*.\(^1\)\(^{34}\)

The first mefp protein to be characterized (mefp-1)\(^1\)\(^{35}\) was found to have a mass of
ca. 120 kDa, contain 3,4-dihydroxyphenylalanine (or Dopa), and consist of tandemly repeated decapeptides each containing at least one residue of Lys, Dopa, as well as 4-hydroxyproline, and 3,4-dihydroxyproline.\textsuperscript{136\textendash}138 The second foot protein (mefp-2), another Dopa containing constituent of the mussel foot plaques of mass ca. 46 kDa, has also been previously characterized.\textsuperscript{139} The remaining two proteins, mefp-3 and mefp-4 are also major components of the cold-shocked plaques, however, until recently there was no detailed information concerning these proteins.

The mefp-3 protein is synthesized and stockpiled in the mussel foot of \textit{M. edulis} and then specifically deposited into the adhesive plaques of the byssus.\textsuperscript{140} Like the other plaque precursor proteins, mefp-3 also has a high Dopa content but is distinguished from the others by high levels of a new post-translational modification of arginine. The structure of this modified arginine was determined to be 4-hydroxyarginine by FAB tandem mass spectrometry (Section IV.2). Although the conversion of Tyr to Dopa in mefp-3 was essentially complete, the modification of arginine to the hydroxylated analogue varies between 40 and 80\%. The initial primary structure determination of mefp-3 was carried out using Edman degradation sequencing, however, complete sequence analysis of the protein was hampered by the coelution of related peptides (since every position containing Arg also contained Arg-OH) and the presence of Dopa. Consequently, only the N-terminal portion of these mefp-3 derived peptides were clearly defined by the Edman method. Another factor contributing to the failure of initial sequencing attempts was the notorious
difficulty in sample handling presented by the unique structural characteristics of these proteins (a high content of Arg-OH and Dopa with the latter known to form reactive quinones\textsuperscript{[41]}). The complete primary structure of mefp-3 was finally determined by the application of exopeptidase digestion in combination with MALDI-TOF-MS. The sequence determination of the mefp-3 protein described in this section serves to illustrate the utility of mass spectrometry in providing solutions to problems that are difficult or impossible to solve using traditional methods alone.

IV.2 Experimental Procedures

The mefp-3 protein was isolated from the feet of fresh exsanguinated mussels by Waite et al. at the University of Delaware (the procedure is described elsewhere\textsuperscript{[40]})\textsuperscript{40}. Like the other mefp proteins, mefp-3 represents a family of adhesive proteins with similar structure and is thus a heterogeneous mixture. Isolation of the different polymorphs (at least nine variants were observed) of the crude mefp-3 was accomplished at the University of Delaware using C\textsubscript{8} RP-HPLC and the resultant peaks analyzed by acid-urea-polyacrylamide gel electrophoresis (Figure IV.1). Fractions 37-39 appeared to be the most electrophoretically homogeneous and were initially pooled for sequence analysis. The mefp-3 variant in these three fractions (37-39) was designated mefp-3F and the sequence determination of mefp-3 described in this section is based on this variant.
The mefp-3 protein is extremely prone to oxidation and very reactive in large part due to the presence of Dopa. Oxidation of mefp-3 was minimized by performing the protein digests in a solution containing ascorbic acid as an anti-oxidant. For each digest, a fresh buffer solution was prepared by adding crystalline ascorbic acid to a 1 M solution of Tris base with constant stirring until the pH, (measured using a Fisher Accumet Model 15 pH meter), was 9.0. Prior to digestion, the mefp-3 protein was denatured by dissolution in 8 M urea for 30 min at 37 °C. Endo-Lys-C digestion of mefp-3F (1:100 enzyme:substrate by weight) was carried out in 2 M urea, 250 mM Tris-Ascorbate at pH 9.0 under Ar for 4 h at 25 °C. The progress of the digest was monitored by analyzing aliquots from the Endo-Lys-C digests by MALDI-TOF-MS.

Peptides from the Endo-Lys-C digest were fractionated by RP-HPLC (Figure IV.2) for the exopeptidase digests. Surprisingly, the chromatography of the mefp-3 peptides using longer (25 cm x 4.6 mm I.D.) C₄ and C₁₈ RP-HPLC columns yielded unsatisfactory results. The best separation of the peptides was achieved using a shorter, 5 cm x 4.6 mm I.D. Vydac C₄ column, with gradient elution of 100 % A to 25 % B in 30 min (where A and B were the HPLC mobile phases, 0.005% aqueous TFA and 0.0035% TFA/ACN, respectively) on a HP-1050 chromatograph using a Waters 490 multiwavelength UV detector with the detector output captured with GRAMS/386 chromatography software. Approximately 500 pmol of the mefp-3F protein digested with Endo-Lys-C was injected onto the HPLC chromatograph after four hours (Figure IV.2). The peptides in the collected
Figure IV.1. RP-HPLC of dialysis precipitated mefp-3. The inset is the acid-urea polyacrylamide gel of protein aliquots taken from fractions 30-45 under the elution profile. The electrophoretic variants are denoted A to J. The most electrophoretically homogeneous variant, mefp-3F (fractions 37-39), was selected for sequence analysis. The work represented in this figure was performed by Waite et al. at the University of Delaware.
Figure IV.2. RP-HPLC chromatograms, (monitored at 214 and 280 nm absorbance), of approximately 500 pmol mepf-3F digested with Endo-Lys-C for 4 hrs. Three HPLC fractions (indicated in the 214 nm trace) were collected.
fractions were lyophilized using a centrifugal evaporator and later redissolved in distilled water for analysis. Further chymotryptic digestion of the largest Endo-Lys-C peptide (approximately 1:100 enzyme:substrate by weight) was performed in 1.5 M urea, 50 mM Tris-ascorbate buffer (pH 8.5) at 25 °C for 30 min (and then analyzed by MALDI-TOF-MS).

MALDI-TOF-MS analysis of the mefp-3F protein or peptides derived thereof were dissolved in ACCA matrix solution to give a final concentration between 1-10 pmol/μL. About 1 μL of this solution was applied to the target plate and allowed to evaporate. The MALDI-TOF-MS experiments were carried out using the Vestec VT-2000 mass spectrometer (Section I.3) using the N₂ laser (337 nm) at a repetition rate of 5 Hz. The resulting protonated molecular ions were accelerated to 30 or 35 keV kinetic energy.

For exopeptidase digestion, the mefp-3F protein or peptides generated from enzymatic digestion of mefp-3F were dissolved in either 100 mM ammonium acetate (pH 4.0) or 100 mM ammonium citrate (pH 5.5) and digested with CpP. Alternatively, AmM in 50 mM sodium phosphate (pH 7.0) buffer was used to obtain N-terminal sequence information. The reactions with CpP were performed at room temperature and those with AmM at 37 °C. The enzyme to substrate ratio was between 1:10 and 1:100 by weight, respectively. Aliquots from the reaction solutions were taken at timed intervals and dissolved in ACCA matrix solution before being analyzed by MALDI-TOF-MS.
FAB-CID mass spectra of Arg, Arg uniformly labelled with $^{15}$N ($^{15}$N-Arg for short), and Arg-$^{14}$N-Arg-OH were measured to determine the structure of the CID fragment ions from Arg. N-acetylation of these compounds was accomplished by dissolving several $\mu$g in 80 $\mu$L glacial acetic acid and 20 $\mu$L acetic anhydride at 37 °C for 1 hr. Butyl esters of Arg, $^{15}$N-Arg, and Arg-$^{14}$N-Arg-OH were prepared separately by incubation in 200 $\mu$L of 3 N HCl/n-butanol at 90 °C for 30 min. The reagents were removed using a Savant speed-vac concentrator prior to mass spectrometric analysis. Arg, the modified and unmodified Arg standards, the unknown Arg-OH from mefp-3, (and later, the authentic 4-hydroxyarginine) were dissolved in distilled H$_2$O and ionized by FAB using glycerol as the matrix in the JEOL high performance tandem mass spectrometer (Section II.2). Similarly, the peptide in HPLC fraction 1 (Figure IV.2) was analyzed by FABMS. Both two sector and four sector mass spectra were acquired using the JEOL Complement data acquisition software. The operating conditions were identical to those used for FABMS analysis of the Grx peptides (Sections III.2).

The sources for many of the reagents used for the work in this section have been described previously in Section III.2. In addition, ascorbic acid, citric acid, and sodium phosphate were obtained from Mallinckrodt, 100 mM ammonium acetate (pH 4.0) buffer and CpY from Pierce, Arg and the dipeptide YW from Sigma, 3 N HCl/n-butanol reagent from Supelco, and $^{15}$N-Arg from Schwartz Chemical Co. Both the endopeptidases (Endo-Lys-C and $\alpha$-chymotrypsin) and exopeptidases (AmM and CpP) used for structure
determination of mefp-3F were purchased from Boehringer Mannheim. Finally, the Arg-
N\textsuperscript{G}-OH and authentic 4-hydroxyarginine were kindly supplied by P.L. Feldman (Glaxo
Research Institute, Research Triangle Park, NC) and E.A. Bell (King's College, London),
respectively.

IV.3 Structure Determination of a Modified Arginine

An unidentified basic amino acid from mefp-3 hydrosylates was isolated using ion
exchange chromatography. This amino acid gave an intensely scarlet color after exposure
to Sakaguchi spray which indicated the presence of a guanidino group.\textsuperscript{140} The amino acid
was ionized by FAB and produced an abundant (M+H)\textsuperscript{+} ion of \textit{m/z} 191.2 indicating a
molecular weight of 190.2. This is 16 Da higher than that of Arg and, together with the
Sakaguchi reaction, suggested that the unknown amino acid was a hydroxyarginine. In
order to elucidate the structure of the Arg-OH, Arg and \textsuperscript{15}N-Arg were first analyzed by
FAB-CID-MS (mass spectra shown in Figure IV.3). Comparison of these two CID spectra
allowed deduction of the structures of the fragments (Figure IV.4) due to Arg since each
nitrogen atom increased the \textit{m/z} value of the corresponding \textsuperscript{15}N-Arg fragment by 1 Da. The
\textit{m/z} values in Figure IV.4 are due to the Arg fragments while those in parentheses are from
the \textsuperscript{15}N-Arg derivative. In addition, a CID mass spectrum of N\textsuperscript{G}-hydroxyarginine was also
obtained. The structure assignments in Figure IV.4 were confirmed by CID mass spectral
analysis of N-acetylated and butyl esters of Arg, \(^{15}\text{N}-\text{Arg}\) and \(\text{Arg}-\text{N}^\text{G}-\text{OH}\). The CID mass spectrum of the unknown amino acid from mefp-3 is shown in Figure IV.5A and its similarity to the Arg and \(^{15}\text{N}-\text{Arg}\) CID mass spectra in Figure IV.3 provided further confirmation for its identity as a hydroxylated Arg. Two abundant ions \((m/z\ 70\ \text{and}\ 87)\) in the CID mass spectrum of Arg are not present to any appreciable extent in the CID mass spectrum of Arg-OH whereas the latter spectrum exhibits two new peaks at \(m/z\ 86\ \text{and}\ 103\) which correspond to a mass shift of 16 Da, respectively. On the other hand, both spectra contain an ion at \(m/z\ 73\). From the structures of these ions (Figure IV.6) one can deduce that the unmodified Arg must be 4-hydroxyarginine. This assignment was later confirmed by obtaining a CID mass spectrum of authentic 4-hydroxyarginine (Figure IV.5B) which is identical to the CID mass spectrum of the Arg-OH from mefp-3 (Figure IV.5A).

IV.4 Amino Acid Sequence of Mefp-3F

The mefp-3 protein isolated from the mussel foot of \(M.\ edulis\) was fractionated by \(C_8\ \text{RP-HPLC}\) and the fractions analyzed by acid-urea-polyacrylamide gel electrophoresis (Figure IV.1) by Waite et al. at the University of Delaware.\(^{140}\) The mefp-3 protein family consists of at least nine variants of which the sixth, mefp-3F (fractions 37-39), appeared to be the most electrophoretically homogeneous. The mefp-3 primary structure reported in this section is based on this mefp-3F variant (for sequence analysis fractions 37-39 were
Figure IV.3. FAB-CID mass spectra of (A) Arg and (B) $^{15}$N-Arg. The $m/z$ values of the (M+H)$^+$ ions are 175.1 and 179.1, respectively.
<table>
<thead>
<tr>
<th>Structure</th>
<th>m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{NH}_4$</td>
<td>18 (19)</td>
</tr>
<tr>
<td>$\text{H}_2\text{N}=\text{CH}$</td>
<td>28 (29)</td>
</tr>
<tr>
<td>$\text{H}_2\text{N}=\text{CH}_2$</td>
<td>30 (31)</td>
</tr>
<tr>
<td>$\text{H}_3\text{N}--\text{C}=\text{N}$</td>
<td>43 (45)</td>
</tr>
<tr>
<td>$\text{H}_2\text{N}=\text{CH}=\text{NH}$</td>
<td>44 (46)</td>
</tr>
<tr>
<td>$\text{H}_2\text{N}=\text{CH}$</td>
<td>56 (57)</td>
</tr>
<tr>
<td>$\text{H}_2\text{N}=\text{CH}_2$</td>
<td>70 (71)</td>
</tr>
<tr>
<td>$\text{H}_2\text{N}=\text{CH}$</td>
<td>73 (76)</td>
</tr>
<tr>
<td>$\text{H}_2\text{N}=\text{CH}_2$</td>
<td>87 (90)</td>
</tr>
<tr>
<td>$\text{H}_3\text{N}=\text{CH}_2$</td>
<td>100 (103)</td>
</tr>
</tbody>
</table>

Figure IV.4. Structures of the fragment ions from the FAB-CID mass spectra of Arg and $^{15}$N-Arg. The m/z values in parenthesis are of the fragments from $^{15}$N-Arg. All m/z values are monoisotopic and are from the mass spectra in Figure IV.3.
Figure IV.5. FAB-CID mass spectra of (A) Arg-OH from mefp-3F and (B) authentic 4-hydroxyarginine.
Figure IV.6. Fragment ions of Arg and mcfp-3 derived Arg-OH from which the structure of the latter was elucidated.
The initial molecular weight determination of mefp-3F (11 kDa) was performed by Waite et al. using discontinuous SDS-PAGE. Prior to the application of mass spectrometry to this problem, the electrophoretically determined $M_r$ of mefp-3F caused considerable confusion in the attempt to reconcile the data obtained by Edman sequencing. The MALDI-TOF mass spectrum of the major component of the intact mefp-3F is shown in Figure IV.7. Bovine ubiquitin, $(M+H)^+ = m/z$ 8565.9 and $(M+2H)^{++} = m/z$ 4283.4 for the singly and doubly charged protonated molecular ions, respectively, was used as an internal standard. The singly charged region of mefp-3F is expanded in the inset and shows that mefp-3F produces a cluster of peaks that differ from the most highly hydroxylated form, $(M+H)^+ = m/z$ 6135.2, of the protein in multiples of -16 mass units. The mefp-3F protein and peptides derived thereof are unique in forming these ion clusters due to the presence of both Arg and Arg-OH.

The unusual amino acid composition of mefp-3 complicated conventional approaches to sequencing this protein. For example, the incomplete post-translational hydroxylation of mefp-3 coupled with numerous short sporadic sequence repeats introduced uncertainties into the Edman degradation derived sequence data. In addition, endoproteinase digestion of mefp-3 was complicated by the tendency of this protein to oxidize, polymerize, and degrade while in solution, especially at alkaline pH. Tryptic
Figure IV.7. MALDI-TOF mass spectrum of the major component of mefp-3F. The peaks labeled "STD (M+H)" and "STD (M+2H)" at m/z 8568.9 and m/z 4283.4 are due to the singly and doubly charged ions, respectively, of the internal calibrant bovine ubiquitin. The inset is an expansion of the (M+H) region of mefp-3F which reveals the different hydroxylation states of the protein (the individual peaks in the cluster are 16 mass units apart).
digestion in particular resulted in an inordinate amount of peptides making analysis very
difficult. On the other hand, Endo-Lys-C retained its specificity producing a very simple
digest containing apparently only two peptides, based on the MALDI-TOF mass spectrum
of this digest (Figure IV.8). It contains two major ion clusters with \((M+H)^+ = m/z\ 1606.1\)
and \(m/z\ 4182.9\) for the most highly hydroxylated component in each cluster, respectively.
The digest was then fractionated by RP-HPLC to separate the peptides. Fractions 2 and 3
(Figure IV.2) corresponded to these two peptides, respectively. Although no peptide ion
signal was detected from MALDI analysis of HPLC fraction 1, FABMS analysis of the
same fraction produced an \((M+H)^+\) ion of \(m/z\ 384.1\) (which is 16 Da higher than the
\((M+H)^+\) ion of the dipeptide YW). This provided an explanation for the absence of a third
peptide in the MALDI mass spectrum of the digest (Figure IV.8) since small peptides are
not efficiently ionized by MALDI and the low mass region is often obscured by peaks due
to the matrix. The FAB-CID mass spectra of the \((M+H)^+ = m/z\ 384.1\) ion and the dipeptide
YW (Figure IV.9) are clearly very similar. The ions annotated with "Y" and "W" in Figure
IV.9B are fragment ions of Tyr and Trp, respectively. In addition, the two C-terminal ions
\((z_i\text{ and } y_i)\) indicated in this mass spectrum establish that W is at the C-terminus (and that Y
is at the N-terminus). While the peaks labeled with "W" have the same \(m/z\) values in both
spectra, the peaks labeled with "Y" are all 16 mass units higher in Figure IV.9A. Thus, the
\((M+H)^+\) ion of \(m/z\ 384.1\) is due to the dipeptide YW where the Y indicates Dopa.
(Similarly, a bold "R" was later used to identify Arg-OH found in mefp-3).
Figure IV.8. MALDI-TOF mass spectrum of an aliquot from an Endo-Lys-C digest of mefp-3F. The inset is an expansion of the (M+H)^+ region of the larger peptide showing peaks that are separated in mass by 16 mass units according to different hydroxylation states.
Figure IV.9. FAB-CID mass spectra of (A) (M+H)^+ = m/z 384.1 and (B) the dipeptide YW. The fragment ions annotated with "W" and "Y" are fragment ions of Trp and Tyr, respectively. The "Y" ions (in bold) in the upper spectrum are due to Dopa. Ions marked "G" are from the glycerol matrix.
The high content of Arg, Arg-OH, and Dopa in mefp-3 contributed to the hydrophilic nature of this protein and rendered FABMS analysis difficult. Only one peptide (isolated from a tryptic digest by Waite et al.) was sufficiently hydrophobic to be ionized by FAB. The sequence of this peptide, GWNNGWNR, was deduced from the CID mass spectrum (shown in Figure IV.10). Fortuitously, the mefp-3 protein and peptides ionized well by MALDI-TOF-MS without requiring any chemical modifications.

The C-terminal sequences of the two larger Endo-Lys-C peptides were unambiguously determined by digesting each peptide with CpP and directly analyzing aliquots at different time points by MALDI-TOF-MS. In addition, the peptide in HPLC fraction 2 was digested with AmM which together with the CpP data, allowed its complete sequence to be determined. The MALDI-TOF mass spectra of the AmM digest of this peptide at various timepoints are shown in Figure IV.11 while those of the CpP digest are shown in Figure IV.12. The partial peptide sequence determined from the resulting data (Tables IV.1 and IV.2) was GWNNGWN-GRRGK. The missing residue was determined to be R from the difference between the molecular weight of the above sequence and the measured (M+H)⁺ of the undigested peptide. Thus, the full sequence of this peptide must be GWNNGWNRGRRGK. Further support for this conclusion was provided by the sequence of the tryptic peptide deduced by FAB-CID (Figure IV.10) and discussed previously.
Figure IV.10. FAB-CID mass spectrum of peptide GWNNNGWR from mepf-3F.
Figure IV.11. MALDI-TOF mass spectra of mefp-3F peptide (M+H)^+ = m/z 1606.7 incubated with AmM at various timepoints. The data is summarized in Table IV.1.
Figure IV.12. MALDI-TOF mass spectra of mefp-3F peptide (M+H)^+ = m/z 1606.7 incubated with CpP at various timepoints. The data is summarized in Table IV.2.
Table IV.1. Peptides from AmM digestion of Endo-Lys-C peptide (M+H)^+ = m/z 1606.7.

<table>
<thead>
<tr>
<th>Peptide (M+H)^+ Observed^a</th>
<th>Mass Difference Observed^b</th>
<th>Mass Difference Calculated</th>
<th>Partial Sequence^c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1606.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1549.7</td>
<td>57.0</td>
<td>57.05</td>
<td>G...</td>
</tr>
<tr>
<td>1363.3</td>
<td>186.4</td>
<td>186.21</td>
<td>GW...</td>
</tr>
<tr>
<td>1249.1</td>
<td>114.2</td>
<td>114.10</td>
<td>GWN...</td>
</tr>
<tr>
<td>1135.0</td>
<td>114.1</td>
<td>114.10</td>
<td>GWNN...</td>
</tr>
<tr>
<td>1077.8</td>
<td>57.2</td>
<td>57.05</td>
<td>GWNNG...</td>
</tr>
<tr>
<td>891.7</td>
<td>186.1</td>
<td>186.21</td>
<td>GWNNGW...</td>
</tr>
<tr>
<td>777.7</td>
<td>114.0</td>
<td>114.10</td>
<td>GWNNGWN...</td>
</tr>
</tbody>
</table>

Table IV.2. Peptides from CpP digestion of Endo-Lys-C peptide (M+H)^+ = m/z 1606.7.

<table>
<thead>
<tr>
<th>Peptide (M+H)^+ Observed^a</th>
<th>Mass Difference Observed^b</th>
<th>Mass Difference Calculated</th>
<th>Partial Sequence^c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1606.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1478.6</td>
<td>128.1</td>
<td>128.17</td>
<td>...K</td>
</tr>
<tr>
<td>1421.4</td>
<td>57.2</td>
<td>57.05</td>
<td>...GK</td>
</tr>
<tr>
<td>1249.3</td>
<td>172.1</td>
<td>172.18</td>
<td>...RGK</td>
</tr>
<tr>
<td>1077.1</td>
<td>172.2</td>
<td>172.18</td>
<td>...RRGK</td>
</tr>
<tr>
<td>1019.9</td>
<td>57.2</td>
<td>57.05</td>
<td>...GRRGK</td>
</tr>
</tbody>
</table>

^aThe (M+H)^+ = m/z 1606.7 peptide and the ACCA matrix dimer [(M+H)^+ = m/z 379.1] were used as internal calibrants.

^bMass difference from the peptide in the row above.

^cResidues in this row have been cleaved. The bold Arg (R) represents Arg-OH.
The remaining ambiguity, the C-terminal sequence of the peptide in HPLC fraction 3, was resolved by CpP digestion with the resultant MALDI-TOF mass spectra (at different time points) shown in Figures IV.13 and IV.14. The data from this exopeptidase digest (Table IV.3) provided the following C-terminal sequence for this peptide:

...GNYNRYNYRGGYK. The same peptide was also digested with chymotrypsin and analyzed by MALDI-TOF-MS (Figure IV.15). The resulting data (Table IV.4) confirmed the peptide sequence assigned by the combination of Edman degradation and MALDI-TOF-MS. Two peptides from this chymotryptic digest, (M+H)$^+$ = m/z 1786.1 and m/z 1486.1, were isolated by RP-HPLC and each separately digested with AmM and CpP. The results of these four exopeptidase experiments provided further confirmation of the deduced sequence. Most significant was the AmM digestion of the (M+H)$^+$ = m/z 1786.1 peptide which resulted in the consecutive loss of A, D, Y, and Y, but did not proceed further due to the GP residues which follow, and the CpP digestion of the peptide (M+H)$^+$ = m/z 1486.8 indicating the consecutive loss of K, Y, and G from the C-terminus.

The data from analysis of the Endo-Lys-C peptides is summarized in Table IV.5. The calculated (M+H)$^+$ values expected from an Endo-Lys-C digest of mefp-3F agree well with the actual measured values. Final sequence alignment of the Endo-Lys-C peptides was accomplished by MALDI-TOF-MS by CpP treatment of intact mefp-3F protein which established the C-terminus as ...RRGKYW. In addition, it was discovered serendipitously that addition of guanidine hydrochloride reduced the activity of Endo-Lys-C sufficiently to
Figure IV.13. MALDI-TOF mass spectra of mepf-3F peptide (M+H)$^+$ = m/z 4182.9 incubated with CpP at various timepoints. This figure is continued on the next page (Figure IV.14).
Figure IV.14. Continuation of the previous figure. MALDI-TOF mass spectra of mep3F peptide (M+H)^+ = m/z 4182.9 with CpP after longer incubation times. The data is summarized in Table IV.3.
Table IV.3. Peptides from CpP digestion of Endo-Lys-C peptide (M+H)$^+$ = m/z 4182.9.

<table>
<thead>
<tr>
<th>Peptide (M+H)$^+$ Observed$^a$</th>
<th>Mass Difference Observed$^b$</th>
<th>Mass Difference Calculated</th>
<th>Partial Sequence$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4183.25*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4055.1</td>
<td>128.2</td>
<td>128.17</td>
<td>...K</td>
</tr>
<tr>
<td>3875.89*</td>
<td>179.2</td>
<td>179.18</td>
<td>...YK</td>
</tr>
<tr>
<td>3761.7</td>
<td>114.2</td>
<td>114.10</td>
<td>...GGYK</td>
</tr>
<tr>
<td>3410.4</td>
<td>351.3</td>
<td>351.36</td>
<td>...RYGGYK</td>
</tr>
<tr>
<td>3238.2</td>
<td>172.2</td>
<td>172.18</td>
<td>...RRYGGYK</td>
</tr>
<tr>
<td>3180.8</td>
<td>57.4</td>
<td>57.05</td>
<td>...GRRYGGYK</td>
</tr>
<tr>
<td>3002.0</td>
<td>178.8</td>
<td>179.18</td>
<td>...YGRRYGGYK</td>
</tr>
<tr>
<td>2829.4</td>
<td>172.6</td>
<td>172.18</td>
<td>...RYGRRYGGYK</td>
</tr>
<tr>
<td>2715.72*</td>
<td>113.7</td>
<td>114.10</td>
<td>...NRGGGGYK</td>
</tr>
<tr>
<td>2536.3</td>
<td>179.4</td>
<td>179.18</td>
<td>...YNRGRRYGGYK</td>
</tr>
<tr>
<td>2364.2</td>
<td>172.1</td>
<td>172.18</td>
<td>...RYNRYGRRYGGYK</td>
</tr>
<tr>
<td>2250.1</td>
<td>114.1</td>
<td>114.10</td>
<td>...NRYNRYGRRYGGYK</td>
</tr>
<tr>
<td>2071.2</td>
<td>178.9</td>
<td>179.18</td>
<td>...YNRYNRYGRRYGGYK</td>
</tr>
<tr>
<td>1957.2</td>
<td>114.0</td>
<td>114.10</td>
<td>...NYNRYNRYGRRYGGYK</td>
</tr>
<tr>
<td>1900.2</td>
<td>57.0</td>
<td>57.05</td>
<td>...GNYNRYNRYGRRYGGYK</td>
</tr>
<tr>
<td>1434.3</td>
<td>-</td>
<td>1434.46$^d$</td>
<td>ADYYGPNYPGR$^d$</td>
</tr>
</tbody>
</table>

$^a$The peptides marked with an asterisk were used as internal calibrants in some of the spectra.

$^b$Mass difference from the peptide in the row above.

$^c$Residues in this column (except for the last row) have been cleaved. The bold Arg (R) and Tyr (Y) represent Arg-OH and Dopa, respectively.

$^d$N-terminal peptide sequence determined by Edman degradation.
Figure IV.15. MALDI-TOF mass spectrum of further chymotryptic digestion of an Endo-Lys-C peptide (HPLC Fraction 3) after 30 min. Peptides marked with an asterisk were used as internal standards. The data is summarized in Table IV.4.
Table IV.4. Peptides from further chymotryptic digestion (Figure IV.15) of the Endo-Lys-C peptide of HPLC fraction 3 (Figure IV.2).

<table>
<thead>
<tr>
<th>(M+H)⁺ Obsvdᵃ</th>
<th>(M+H)⁺ Calcd</th>
<th>Sequenceᵇ</th>
<th>Positionᶜ</th>
</tr>
</thead>
<tbody>
<tr>
<td>949.9</td>
<td>949.96</td>
<td>(Y) NRY NRY (G)</td>
<td>20-25</td>
</tr>
<tr>
<td>1021.08*</td>
<td>-</td>
<td>(Y) GRRY GGYK ( )</td>
<td>26-33</td>
</tr>
<tr>
<td>1486.8</td>
<td>1486.54</td>
<td>(Y) NRY GRRY GGYK ( )ᵈ</td>
<td>23-33</td>
</tr>
<tr>
<td>1786.1</td>
<td>1785.82</td>
<td>() ADY YGPN YGPP RRY (G)ᵉ</td>
<td>1-14</td>
</tr>
<tr>
<td>1952.7</td>
<td>1952.01</td>
<td>(Y) NRY NRY GRRY GGYK ( )</td>
<td>20-33</td>
</tr>
<tr>
<td>2250.7</td>
<td>2250.26</td>
<td>() ADY YGPN YGPP RR YGG GNY (N)</td>
<td>1-19</td>
</tr>
<tr>
<td>2416.4</td>
<td>2416.45</td>
<td>(Y) GGG N YR NRY GRR Y GGYK ( )</td>
<td>15-33</td>
</tr>
<tr>
<td>2716.1</td>
<td>2715.73</td>
<td>() ADY YGPN YGPP RR YGG GNY NRY (N)</td>
<td>1-22</td>
</tr>
<tr>
<td>3181.2</td>
<td>3181.19</td>
<td>() ADY YGPN YGPP RR YGG GNY NRY NRY (G)</td>
<td>1-25</td>
</tr>
<tr>
<td>4183.25*</td>
<td>-</td>
<td>() ADY YGPN YGPP RR YGG GNY - NRY NRY GRR Y GGYK ( )</td>
<td>1-33</td>
</tr>
</tbody>
</table>

ᵃPeptides marked with asterisks in the first column were used as internal standards.
ᵇThe bold Arg (R) and Tyr (Y) represent Arg-OH and Dopa, respectively. The sequences shown are those corresponding to the measured molecular weights and the final sequence of mefp-3F.
ᶜPosition in the final sequence.
ᵈResults of CpP digestion underlined.
ᵉResults of AmM digestion underlined.
Table IV.5. Peptides derived from Endo-Lys-C digestion of mefp-3F. The Y and R residues in bold represent Dopa and 4-hydroxyarginine, respectively. The observed (M+H)+ values are for the most highly hydroxylated component.

<table>
<thead>
<tr>
<th>RP-HPLC Fraction</th>
<th>Position</th>
<th>Edman Sequence</th>
<th>Final Sequence</th>
<th>(M+H)+ Calc.</th>
<th>(M+H)+ Obsvd.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47-48</td>
<td><strong>YW</strong></td>
<td><strong>YW</strong></td>
<td>384.16</td>
<td>384.1</td>
</tr>
<tr>
<td>2</td>
<td>34-46</td>
<td>GWNNGWNR[R$<em>2$G$</em>{23}$]K</td>
<td>GWNNGWNRGRRGK</td>
<td>1606.70</td>
<td>1606.7</td>
</tr>
<tr>
<td>3</td>
<td>1-33</td>
<td>ADYYGPNYGPPRRYGGG-NYNR[Y$<em>4$R$</em>{23}$NG$_{43}$]K</td>
<td>ADYYGPNYGPPRRYGGG-NYNRNYNRYGRRYGGYK</td>
<td>4183.25</td>
<td>4182.9</td>
</tr>
</tbody>
</table>

*From Figure IV.2.

*Position in the final sequence.

*Sequence from Edman degradation performed by Waite et al at the University of Delaware. Amino acids in brackets represent residues determined by amino acid analysis.

*Deduced sequence after combined sequencing effort by Edman degradation and mass spectrometry.

*The monoisotopic (M+H)+ value (in italics) was determined by FABMS. The remaining (M+H)+ values are polyisotopic and were determined by MALDI-TOF-MS using an internal standard.

126
avoid cleavage at Lys-46 resulting in a peptide of \((M+H)^+ = m/z\) 1972.3 (calculated \(m/z\) 1972.09) using bovine insulin as an internal calibrant. Thus, the mass spectrometric data together with the data from Edman degradation give the following sequence for mefp-3F:

\[
\text{ADYYGPNYGPPRRYGGNYNRYNGRRYYGGYKGWNNGWNRGRGRKYW}
\]

The experimentally determined \((M+H)^+\) value for the most highly oxidized form of mefp-3F, \(m/z\) 6135.2, agrees well with the \((M+H)^+\) value 6136.31 calculated for this sequence.

IV.5 Conclusion

An adhesive protein, mefp-3, isolated from the foot of the marine mussel *M. edulis*, presented unique challenges to the determination of the amino acid sequence due to its unusual composition. Mefp-3 represents a family of similar polymorphic proteins that vary in the extent of post-translational modification and are composed almost exclusively of Dopa, Gly, Asn, Trp, Arg, and a previously unknown basic amino acid. (The latter was determined to be 4-hydroxyarginine by FAB tandem mass spectrometry). Although at least nine variants of mefp-3 were discovered, only one variant (mefp-3F) appeared electrophoretically homogeneous. The primary structure determination of mefp-3 discussed in this section is based on this variant.
In contrast to the essentially complete conversion of Tyr to Dopa, the post-translational modification of Arg was incomplete resulting in multiple hydroxylation states. This complicated the initial attempts to sequence mefp-3 using conventional Edman degradation. The complete sequence of the variant mefp-3F was finally determined by using a combined approach of Edman degradation sequencing and mass spectrometry with the C-terminal sequences of Endo-Lys-C peptides ultimately determined by MALDI-TOF-MS analysis of CpP digests. This work serves as an additional example of the complementarity of mass spectrometry with more traditional sequencing methods.
V.1 Introduction

The oxidation of inosine 5'-monophosphate (IMP) to xanthosine 5'-monophosphate (XMP) is catalyzed by the enzyme inosine-5'-monophosphate dehydrogenase (IMPDH) and is the rate limiting step in the synthesis of guanine nucleotides. XMP is first converted to guanosine 5'-monophosphate (GMP) by GMP synthetase and ultimately yields the guanosine 5'-triphosphate (GTP) necessary for nucleic acid synthesis. The reaction involving IMPDH is of particular interest due to the discovery that IMPDH is highly expressed in tumor cells and in rapidly growing tissues. For example, northern blot analysis of steady-state mRNAs indicated that IMPDH was up to 5-fold more active transcriptionally in leukemic cells than in normal cells. Inhibitors of IMPDH have been shown to reduce intracellular GTP levels and to suppress biological processes that depend on cellular proliferation. In addition, IMPDH inhibitors have been found to have anticancer, antiviral, and immunosuppressive activity. Early studies with the broad-spectrum antiviral drug Virazole (I), [a synthetic nucleotide having the structure 1-β-D-ribofuranosyl-1,2,4-triazole-3-carboxamide-- also known as Ribavirin], suggested that its
antiviral activity may be due to the inhibition of GMP synthesis at the step involving conversion of IMP to XMP.\textsuperscript{151} More recently, 5-ethynyl-1-\textbeta-D-ribofuranosylimidazole-4-carboxamide [EICAR (II)]\textsuperscript{152,153}, an analogue of Virazole, was shown to be an extremely potent antiviral agent.\textsuperscript{154} Adenosine kinase converts EICAR to the active 5'-monophosphate metabolite EICARMP\textsuperscript{155} (III) which has been shown to inhibit IMPDH in murine leukemia L1210 and human lymphocyte CEM cells.\textsuperscript{156} The \textit{in vitro} experiments with EICARMP and IMPDH isolated from \textit{E. coli} were carried out to directly assess the role of IMPDH as a target enzyme for EICARMP.\textsuperscript{157}

\begin{center}
\begin{tabular}{ccc}
\includegraphics[width=0.3\textwidth]{Virazole.png} & \includegraphics[width=0.3\textwidth]{EICAR.png} & \includegraphics[width=0.3\textwidth]{EICARMP.png} \\
Virazole (I) & EICAR (II) & EICARMP (III)
\end{tabular}
\end{center}

At least two mechanisms (Figure V.1) have been proposed for the IMPDH mediated conversion of IMP to XMP.\textsuperscript{145,158} The first involves IMPDH as a catalytic base to add hydroxide (by deprotonation of water) directly at position 2 (C2) of IMP. Hydride transfer to NAD then yields XMP. In the alternative mechanism, attack at C2 occurs due to base activation of a nucleophilic residue (Nuc) from IMPDH. The resultant intermediate
contains an oxyanion that is stabilized by an acidic residue of the enzyme. Collapse of the oxyanion occurs with concomitant reduction of NAD$^+$ to NADH. Finally, a hydrolysis step regenerates the nucleophilic residue and yields XMP. The *E. coli* IMPDH was observed to be irreversibly inactivated by EICARMP\textsuperscript{157} which suggested that EICARMP forms a covalent adduct with IMPDH at the active site. In this section, this hypothesis was investigated by using MALDI-TOF-MS to analyze both the native and the EICARMP inactivated *E. coli* IMPDH enzymes. In addition, in the course of these experiments, the resulting mass spectral data indicated that the original IMPDH primary structure predicted from the DNA sequence\textsuperscript{159} was in error. Corrections to the IMPDH amino acid sequence were ultimately confirmed using MALDI-PSD mass spectrometry.

V.2 Experimental Procedures

The *E. coli* IMPDH enzyme was expressed, isolated and incubated with the EICARMP reagent by Wang and Hedstrom at Brandeis University (the procedure is described elsewhere).\textsuperscript{157} Approximately 30 pmol of aqueous IMPDH was mixed with an equal volume of denaturing buffer (8 M guanidine hydrochloride, 0.025 M Tris-HCl pH 8.3, 0.002 M EDTA). Enough DTT (14 μg/μL) was added to provide a 2000-fold molar excess compared to the enzyme and the solution incubated at 37 °C for 3 h. The S-alkylation reagent, 4-vinylpyridine, was then added (5-fold molar excess over DTT) and the
Figure V.1. Possible mechanisms for the IMPDH catalyzed conversion of IMP to XMP.
reaction mixture kept at 37 °C overnight. Desalting was accomplished by injecting the reaction mixture onto a 2 cm Vydac C4 reversed phase guard column and washing with 0.05% aqueous TFA. The protein was eluted from the column with 0.035% TFA/ACN. The fraction containing the eluted protein was dried using a centrifugal evaporator.

About 15 µg of S-ethylpyridylated IMPDH (both native and EICARMP inactivated) were dissolved in 2 M urea and 50 mM Tris-HCl (pH 9.0) for Endo-Lys-C digestion. About 0.2 µg of Endo-Lys-C was added and the digest solution kept at room temperature for 3 h. Similarly, the S-ethylpyridylated EICARMP inactivated IMPDH was digested with trypsin in 2 M urea and 100 mM Tris-HCl (pH 8.5) for 12 h. The tryptic digests were partially fractionated using a 25 cm x 4.6 mm I.D. Vydac C4 RP-HPLC column. The peptides were eluted using a 99 % A to 60 % B gradient in 60 min (where A and B are the 0.05 % TFA/H2O and 0.035 % TFA/ACN mobile phases, respectively) using the HP-1050 chromatograph. The peptides were collected in fractions and dried using a centrifugal evaporator.

A saturated solution of SA matrix in 30 % ACN was used for MALDI-TOF-MS analysis of the intact IMPDH protein. The enzyme was dissolved in this matrix solution to give a final protein concentration between 1-10 pmol/µL. For the digests and fractionated peptides, ACCA prepared in 50 % ACN was used as the matrix. About 1 µL of the analyte/matrix solution was applied to the target plate and allowed to evaporate. The
Voyager Elite (Section II.3) MALDI-TOF-MS mass spectrometer was used to obtain linear, reflectron, and PSD mass spectra. The sample spots were irradiated using the N₂ laser at a repetition rate of about 4 Hz. The resulting ions were accelerated to 30 kV kinetic energy (20 kV for reflectron and PSD mass spectra). To increase mass accuracy, BSA was added as an internal standard to the IMPDH/matrix solution. At least six separate mass spectra (sum of 128 laser shots) were obtained for each experiment with the undigested enzyme to ensure good measurement precision (≤ 0.1 %). The reported m/z values for the IMPDH (M+H)⁺ ions in this section are the averaged results.

Aliquots from the endoproteinase digests were taken directly and diluted with the ACCA matrix solution. The initial calibration was performed using bovine insulin as an external standard. At least 64 (up to a maximum of 256) laser shots were summed for each mass spectrum. PSD mass spectra of Endo-Lys-C peptides were obtained from the unfractionated digest using the timed ion selector (Section II.3) to deflect all but the desired molecular and fragment ions. However, the relatively lower ion current due to the tryptic peptide containing Cys-305 (the residue modified by EICARMP) required HPLC fractionation prior to the PSD analysis. A more detailed description of MALDI-PSD can be found in Section II.3. The final PSD spectra were generated using the Perceptive Biosystems GRAMS computer program to concatenate at least twelve separate reflector mass spectra each acquired at different mirror ratios in order to optimally focus fragment ions with differing kinetic energies.
The sources for the enzymes and reagents used in this work have already been described in Sections III.2 and IV.2.

V.3 Identification of the Active Site Modification

The IMPDH enzyme is known to be inactivated by sulfhydryl reagents such as iodoacetamide and methanethiosulphonate, however, the presence of IMP protects the enzyme from inactivation. This suggested that the active site near C2 of IMP (Figure V.1) may contain a cysteine residue. The observed irreversible inactivation of IMPDH with the EICRMP reagent may thus be due to a covalent modification of the E. coli protein. The hypothetical reaction between IMPDH and one molecule of EICRMP is shown below:
Covalent modification of IMPDH by addition of EICARMP (M, 347.2) would result in a shift to higher mass for the protein by 347.2 Da. The MALDI-TOF mass spectra of the native and EICARMP inactivated *E. coli* IMPDH are shown in Figures V.2A and V.2B, respectively. The difference (306 Da) between the measured values for the native and EICARMP inactivated enzyme, (*m/z* 52,016 and *m/z* 52,322, respectively), corresponds to the modification of IMPDH with one EICARMP molecule (within MALDI-TOF-MS experimental error of ± 0.1%). Both the native and EICARMP inhibited IMPDH were reduced, S-ethylpyridylated and digested with Endo-Lys-C in separate experiments. The MALDI-TOF mass spectra of these two digests are shown in Figure V.3. The measured and calculated masses as well as the sequences of the IMPDH peptides resulting from these two digests are listed in Table V.1. Although the mass spectra of the digests are similar, they are not identical. The major qualitative difference is the shift in mass by the ion of *m/z* 5227.0 in Figure V.3A to *m/z* 5470.3 in Figure V.3B. The former corresponds to peptide 297-349 (Table V.1) in which the Cys-305 is S-ethylpyridylated. On the other hand, the peptide ion at *m/z* 5470.3 in Figure V.3B is consistent only with the addition of one molecule of EICARMP (M, 347.2) instead of vinylpyridine (M, 105.1). Consequently, the EICARMP modification must be at position 305, the only cysteine in this peptide. The EICARMP modified peptide, (M+H)⁺ = *m/z* 5470 peptide was too large for efficient fragmentation by MALDI-PSD. Therefore, a shorter peptide was generated by tryptic digestion of the S-ethylpyridylated EICARMP inactivated protein. The peptides resulting from this digest (MALDI-TOF mass spectrum shown in Figure V.4) provided additional
Figure V.2. MALDI-TOF mass spectra containing singly and doubly charged ions of *E. coli* IMPDH. A. Native. B. EICARMP inactivated. The peaks labeled "Std (M+H)⁺" at *m/z* 66,431.0 and "Std (M+2H)⁺⁺" at *m/z* 33,216.0 are due to the singly and doubly charged ions, respectively, of the internal calibrant bovine serum albumin (BSA).
confirmation for the EICARMP modification site and IMPDH primary structure (Section V.4). The data from the tryptic digest is summarized in Table V.2. The S-ethylpyridylated tryptic peptide 297-308, VGIGPGSICTTR [calculated m/z 1266.51], was not observed. However, the ion at m/z 1509.0 corresponds to this peptide with an EICARMP modified cysteine [calculated m/z 1508.57] rather than S-ethylpyridylated. The MALDI-PSD mass spectrum of this tryptic peptide (Figure V.5) contains three abundant fragment ions at m/z 1427.8, m/z 1295.9, and m/z 1126.9. These ions correspond to the loss of phosphate, ribose-5'-phosphate, and S-EICARMP, respectively, indicating that EICARMP is indeed a component of this peptide. The latter fragment ion provided further evidence for cysteine as the modified residue. Although the facile cleavage of the phosphate and ribose groups prevented the acquisition of a PSD spectrum with complete sequence information, the lower mass region provided enough confirmatory data for confident sequence assignment of this peptide and of the site of EICARMP adduction.

V.4 Correction of the Amino Acid Sequence

Based on the amino acid sequence deduced from the DNA sequence, E. coli IMPDH contains six cysteines and has a calculated M, of 54,575.7. However, MALDI-TOF-MS analysis of the native E. coli protein produced an (M+H)+ ion at m/z 52,016 (Figure V.2A) rather than the expected value (54,576.7). Furthermore, the (M+H)+ ion of
Figure V.3. MALDI-TOF mass spectra of Endo-Lys-C digests of S-ethylpyridylated *E. coli* IMPDH. A. Native. B. EICARMP inactivated. The peak labels represent the positions of the peptides in the corrected amino acid sequence. The peptides containing Cys-305 are denoted by an asterisk.
Table V.1. Peptides resulting from Endo-Lys-C digestion of *E. coli* native and EICARMP inactivated IMPDH. (Footnotes are on the following page).

<table>
<thead>
<tr>
<th>Position</th>
<th>Native(^a) (M+H)(^+)</th>
<th>Modified(^a) (M+H)(^+)</th>
<th>Calc. (M+H)(^+)</th>
<th>Amino acid sequence(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-6</td>
<td>731.4</td>
<td>-</td>
<td>731.98</td>
<td>( ) MLRIAK (E)</td>
</tr>
<tr>
<td>7-36</td>
<td>3213.0</td>
<td>3213.0</td>
<td>3211.64</td>
<td>(K) EALTFDDVLVPAHSTVL-PNTADLSTQTLTK (T)</td>
</tr>
<tr>
<td>37-72</td>
<td>3825.0</td>
<td>3824.8</td>
<td>3824.55</td>
<td>(K) TIRLNIPMLSAAMDVTE-ARLAIALAQEGGIHG(N)</td>
</tr>
<tr>
<td>73-87</td>
<td>1846.12(^*)</td>
<td>1846.12(^*)</td>
<td>1846.12</td>
<td>(K) NMSIERQAAEEVRRVK (K)</td>
</tr>
<tr>
<td>88-110</td>
<td>2537.3</td>
<td>2537.0</td>
<td>2536.90</td>
<td>(K) KHESGVVTDPQTVLPTTT-LREVK (E)</td>
</tr>
<tr>
<td>89-110</td>
<td>2408.5</td>
<td>-</td>
<td>2408.72</td>
<td>(K) KHESGVVTDPQTVLPTTT-LREVK (E)</td>
</tr>
<tr>
<td>111-156</td>
<td>5157.7</td>
<td>5157.3</td>
<td>5156.83</td>
<td>(K) ELTERNGFAGYPVVTENELVGITGRDVRVTDLNQ-PVSVYMTPK (E)</td>
</tr>
<tr>
<td>157-174</td>
<td>2055.4</td>
<td>2055.5</td>
<td>2055.39</td>
<td>(K) ERLTVTGREGAEV-LAK (M)</td>
</tr>
<tr>
<td>183-199</td>
<td>1900.9</td>
<td>1901.9</td>
<td>1901.28</td>
<td>(K) ALVVDDEFHLMITVK (D)</td>
</tr>
<tr>
<td>204-212</td>
<td>1121.9</td>
<td>1122.3</td>
<td>1122.34</td>
<td>(K) AERKNACK (D)</td>
</tr>
<tr>
<td>213-267</td>
<td>5708.33(^*)</td>
<td>5708.33(^*)</td>
<td>5708.33</td>
<td>(K) YPDLOQIGGNVATAAGAR-ALAEAGCSAVK (V)</td>
</tr>
<tr>
<td>268-296</td>
<td>2894.7</td>
<td>2894.3</td>
<td>2894.3</td>
<td>(K) VGIGPGSICCTTRIVTGVG-PQITAVADAVEALEGTGIPV-ADGDIRFGDIK (A)</td>
</tr>
<tr>
<td>297-349</td>
<td>5227.0</td>
<td>5470.3(^c)</td>
<td>5227.05</td>
<td>(K) AIAAGASAVMVGSMLAG-TEESPGEIYQGRSYK (S)</td>
</tr>
<tr>
<td>350-383</td>
<td>3446.6</td>
<td>3446.5</td>
<td>3445.90</td>
<td>(K) SYRGMGLGAMS (G)</td>
</tr>
<tr>
<td>384-396</td>
<td>1345.4</td>
<td>1345.1</td>
<td>1345.59</td>
<td>(K) GSSDRYFQSDNAADK (L)</td>
</tr>
<tr>
<td>397-411</td>
<td>1661.6</td>
<td>-</td>
<td>1661.68</td>
<td>(K) LVPEGIERVAYK (G)</td>
</tr>
<tr>
<td>412-424</td>
<td>1431.5</td>
<td>1431.4</td>
<td>1431.68</td>
<td>(K) EIHHQOMGGLRSCMGLT-GCGTIDELRTK (A)</td>
</tr>
<tr>
<td>429-456</td>
<td>3260.3</td>
<td>3259.9</td>
<td>3259.87</td>
<td>(K) AEVRISGAGIQESHVHD-VITIK (E)</td>
</tr>
<tr>
<td>457-479</td>
<td>2496.1</td>
<td>2496.0</td>
<td>2495.80</td>
<td>(K) ESPNYRLGS ( )</td>
</tr>
<tr>
<td>480-488</td>
<td>1022.8</td>
<td>-</td>
<td>1023.09</td>
<td></td>
</tr>
</tbody>
</table>
Footnotes for Table V.1.

a(M+H)+ ions marked with asterisks were used as internal standards for mass calibration.

bAmino acids in parenthesis refer to those preceding and following the Endo-Lys-C cleavage sites.

cCalculated 5469.14 for the addition of EICARMP (M, 347.22) to Cys-305 instead of 4-vinylpyridine (M, 105.14).
Figure V.4. MALDI-TOF mass spectrum of a tryptic digest of S-ethylpyridylated *E. coli* EICARMP inactivated IMPDH. The peak labeled with the asterisk in bold contains Cys-305. The annotations correspond to the first column in Table V.2.
Table V.2. Peptides resulting from tryptic digestion of EICARMP inactivated *E. coli* IMPDH. (Footnotes are on the following page).

<table>
<thead>
<tr>
<th>Pep&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Position</th>
<th>Obs.&lt;sup&gt;b&lt;/sup&gt; (M+H)&lt;sup&gt;+&lt;/sup&gt;</th>
<th>Calc. (M+H)&lt;sup&gt;+&lt;/sup&gt;</th>
<th>Amino acid sequence&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7-36</td>
<td>3212.4</td>
<td>3211.64</td>
<td>(K) EALTFDDVLLVPAPAHLTVLPTNT-ADLSTQQLTK (T)</td>
</tr>
<tr>
<td>2</td>
<td>40-56</td>
<td>1834.4</td>
<td>1834.17</td>
<td>(R) LNIPMLSAAMDVTVEAR (L)</td>
</tr>
<tr>
<td>3</td>
<td>57-72</td>
<td>1639.1</td>
<td>1638.95</td>
<td>(R) LAIAALAEQEGGIGFIHK (N)</td>
</tr>
<tr>
<td>4</td>
<td>79-85</td>
<td>887.9</td>
<td>887.97</td>
<td>(R) QAEVRR (V)</td>
</tr>
<tr>
<td>5</td>
<td>88-107</td>
<td>2180.8</td>
<td>2180.47</td>
<td>(K) KHESGVTDPQTVLPTTTLR (E)</td>
</tr>
<tr>
<td>6</td>
<td>89-107</td>
<td>2052.7</td>
<td>2052.29</td>
<td>(K) HESGVTDPQTVLPTTTLR (E)</td>
</tr>
<tr>
<td>7</td>
<td>111-137</td>
<td>2965.7</td>
<td>2965.29</td>
<td>(K) ELTERNGFAGYPVVTEENELVGIITGR (D)</td>
</tr>
<tr>
<td>8</td>
<td>116-137</td>
<td>2337.4</td>
<td>2336.61</td>
<td>(R) NGFAGYPVVTEENELVGIITGR (D)</td>
</tr>
<tr>
<td>9</td>
<td>141-156</td>
<td>1840.7</td>
<td>1840.15</td>
<td>(R) FVTDLNQPVSVMTPK (E)</td>
</tr>
<tr>
<td>10</td>
<td>183-199</td>
<td>1901.5</td>
<td>1901.28</td>
<td>(K) ALVVDDHEHLIMTIVK (D)</td>
</tr>
<tr>
<td>11</td>
<td>207-217&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1351.6</td>
<td>1351.53</td>
<td>(R) KPNACKDEQGR (L)</td>
</tr>
<tr>
<td>12</td>
<td>220-260</td>
<td>3998.40&lt;sup&gt;*&lt;/sup&gt;</td>
<td>3998.40</td>
<td>(R) VGAAVAGAGAGNEERVDALVA-AGVDVLLIDSHHGSEGVLQR (I)</td>
</tr>
<tr>
<td>13</td>
<td>234-260</td>
<td>2759.3</td>
<td>2759.09</td>
<td>(R) VDALVAAGVDVLLIDSHHGSEGVLQR (I)</td>
</tr>
<tr>
<td>14</td>
<td>266-285</td>
<td>1987.6</td>
<td>1987.27</td>
<td>(R) AKYPDLQIIGGNVATAAGAR (A)</td>
</tr>
<tr>
<td>15</td>
<td>286-296</td>
<td>1125.33&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1125.33</td>
<td>(R) ALAEAGCSAVK (V)</td>
</tr>
<tr>
<td>16</td>
<td>297-308</td>
<td>1509.0</td>
<td>1508.57&lt;sup&gt;e&lt;/sup&gt;</td>
<td>(K) VGIGPGSICTTR (I)</td>
</tr>
<tr>
<td>17</td>
<td>309-342</td>
<td>3261.1</td>
<td>3260.76</td>
<td>(R) IVTGVGVPQITAVADADEVAGLEGTGIP-VIADGGR (F)</td>
</tr>
<tr>
<td>18</td>
<td>350-380</td>
<td>3067.5</td>
<td>3067.47</td>
<td>(K) AIAAGASAVMGSMLAG-TEESPEIETYQGR (S)</td>
</tr>
<tr>
<td>19</td>
<td>397-420</td>
<td>2613.0</td>
<td>2612.77</td>
<td>(K) GSSDRYFQSDNAADKLVPPEGIEGR (V)</td>
</tr>
<tr>
<td>20</td>
<td>402-420</td>
<td>2110.7</td>
<td>2110.29</td>
<td>(R) YFQSDNAADKLVPPEGIEGR (V)</td>
</tr>
<tr>
<td>21</td>
<td>427-439</td>
<td>1523.9</td>
<td>1523.84</td>
<td>(R) LKEIIHQQMGGLR (S)</td>
</tr>
<tr>
<td>22</td>
<td>429-439</td>
<td>1282.6</td>
<td>1282.51</td>
<td>(K) EIIHQQMGGLR (S)</td>
</tr>
<tr>
<td>23</td>
<td>440-454</td>
<td>1767.2</td>
<td>1767.10</td>
<td>(R) SCMGLTGCCTIDELR (T)</td>
</tr>
<tr>
<td>24</td>
<td>455-461</td>
<td>851.0</td>
<td>850.99</td>
<td>(R) TKAEFVR (I)</td>
</tr>
<tr>
<td>25</td>
<td>462-479</td>
<td>1893.5</td>
<td>1893.11</td>
<td>(R) ISGAGIQESHVHDVTITK (E)</td>
</tr>
</tbody>
</table>
Footnotes for Table V.2

\( ^a \) Peptide annotations as in Figure V.4.

\( ^b \) (M+H)\(^+\) ions marked with asterisks were used as internal standards for mass calibration.

\( ^c \) Amino acids in parenthesis refer to those preceding and following the tryptic cleavage sites.

\( ^d \) Tryptic peptide (11) providing supporting evidence for Arg at position 206 rather than Ala (see Section V.4).

\( ^e \) The calculated mass for peptide 16 is for the addition of EICARMP (M, 347.22) at Cys-305 instead of 4-vinylpyridine (M, 105.14).
Figure V.5. MALDI-PSD mass spectrum of tryptic peptide (M+H)+ = m/z 1509.0. (The monoisotopic (M+H)+ value was m/z 1507.8). The annotations "-5'-P" and "-Rib-5'-P" refer to loss of the phosphate and ribose-5'-phosphate, respectively. The Cys labeled with an asterisk in the internal fragment ions indicates loss of the Rib-5'-P.
m/z 52,241 due to reduced and S-ethylpyridylated E. coli IMPDH corresponded to the addition of vinylpyridine to five rather than six cysteine residues. These discrepancies were resolved by incubating the S-ethylpyridylated protein with Endo-Lys-C and analyzing the resulting digest using MALDI-TOF-MS (Figure V.3A). The N-terminal peptide MQSVTLCIMPRQYLLTTLVEILPMLRIAK (calculated (M+H)$^+$ = m/z 3482.41) expected from the DNA sequence was not observed in this mass spectrum. Instead, an ion at m/z 731.4, whose mass did not correspond to any predicted Lys-C peptides was in agreement with the calculated m/z value (731.98) for the protonated molecular ion of peptide MLRIAK. The MALDI-PSD mass spectrum of the (M+H)$^+$ = m/z 731.4 ion is shown in Figure V.6 and confirms that amino acid sequence. Edman degradation sequencing (5 cycles) of intact IMPDH provided supporting evidence that the N-terminal sequence of native E. coli IMPDH begins with MLRIA$^{157}$ and thus, protein synthesis must begin at the third methionine of the original DNA-derived sequence.$^{159}$ Finally, the measured m/z values due to the (M+H)$^+$ ions of native E. coli IMPDH (52,016) and the reduced and S-ethylpyridylated protein (52,541) are consistent with the calculated values 52,023.6 and 52,549.3, respectively. Consequently, the E. coli IMPDH gene product is only 488 amino acids long and contains five rather than six cysteines.

The MALDI mass spectra of the Endo-Lys-C digests in Figure V.3 both contained an initially unidentified ion at m/z 1121.9 and m/z 1122.3 (Figures V.3A and V.3B, respectively). The MALDI-PSD mass spectrum (Figure V.7) of the (M+H)$^+$ ion of m/z
Figure V.6. MALDI-PSD mass spectrum of Endo-Lys-C peptide (M+H)$^+$ = m/z 731.4 from S-ethylpyridylated *E. coli* IMPDH. (The monoisotopic (M+H)$^+$ value was m/z 731.2).
Figure V.7. MALDI-PSD mass spectrum of Endo-Lys-C peptide (M+H)^+ = m/z 1121.9 from S-ethylpyridylated *E. coli* IMPDH. (The monoisotopic (M+H)^+ value was m/z 1121.6). The annotations "VP" and "-EP" refer to vinylpyridine and loss of ethylpyridine, respectively.
1121.9 allowed the complete sequence (AERKPNACK) of this peptide to be deduced. The calculated m/z for the protonated (M+H)^+ ion of AERKPNACK (1122.3) agrees well with the measured values (1121.9 and 1122.3) and corresponds to peptide 204-212 with Arg instead of Ala at position 206. Furthermore, the tryptic peptide 207-217 (Table V.2) also provided further evidence for Arg rather than Ala at this position. The mass spectrometric data thus confirmed another error in the published DNA sequence that was discovered upon re-sequencing the *E. coli* IMPDH gene.157

V.5 Conclusion

The *E. coli* IMPDH was shown to be irreversibly inactivated by EICARMP.157 MALDI-TOF-MS analysis of the native and inactivated enzymes confirmed that EICARMP forms a covalent adduct with IMPDH. Endoproteinase digestion of the proteins in combination with MALDI-TOF-MS analysis allowed identification of the site of modification. Furthermore, the mass spectrometric results indicated that the deduced IMPDH primary structure from the DNA sequence159 was in error. The sequence corrections, which were identified and confirmed by MALDI-PSD, showed that the native *E. coli* protein consists of 488 amino acids, has five cysteines, and the amino acid in position 206 is Arg, not Ala.
The site of covalent modification of *E. coli* IMPDH with EICARMP is at Cys-305 and strongly suggests that this residue is near C2 of IMP in the IMP/IMPDH complex. This result is consistent with earlier work with human type II IMPDH where 6-chloropurine riboside 5'-monophosphate (6-Cl-IMP—another IMP derivative that inhibits IMPDH) was shown to form a covalent adduct with Cys-331. In *E. coli* IMPDH, Cys-305 (305 in *E. coli* is homologous to 331 in the human enzyme) apparently acts as a catalytic nucleophile in the IMPDH mediated conversion of IMP to XMP. However, other roles for Cys-305 cannot be eliminated.

Finally, this work with the *E. coli* IMPDH enzyme demonstrates the applicability of MALDI-TOF-MS and especially MALDI-PSD to important biological research. It is particularly noteworthy that all of the mass spectrometric work was performed with less than 500 pmol of total material and of this amount, only a small fraction (a few pmol) was actually used in the course of MALDI analysis. In conclusion, the experimental work related in this section provides yet another example of the complementarity of mass spectrometry with traditional biochemical methods for protein structure analysis.
VI.1 Introduction

Chemical reagents for protein cleavage are infrequently used in large part due to the exquisite specificity of enzymes. Strategies involving most chemical cleavage reagents can be regarded as "brute force" methods that, in addition to the desired cleavages, often generate unwanted side products which can further complicate the sequence determination of an unknown protein. A notable exception is cyanogen bromide which specifically cleaves at the C-terminus of methionine and is often used in situations where larger peptides are desired. Commonly used proteases such as trypsin, chymotrypsin, and endoproteinase-Lys-C cleave at sites that are abundant in most proteins resulting in numerous smaller peptides. This is advantageous for Edman degradation and FAB mass spectrometry methodologies where peptide sequencing becomes increasingly more difficult with larger peptides. While smaller peptides are generally more difficult to ionize by MALDI the opposite is almost always true for larger peptides. Specific cleavage of proteins at cysteine in most cases will result in a smaller number of larger peptides since cysteine is one of the less abundant amino acids in proteins. Consequently, MALDI-TOF mass spectrometry is especially well suited to the analysis of the larger peptide products that
would result from a cleavage reaction of this specificity.

It has been known for some time now that disulfide bond containing proteins can be cleaved with cyanide ion at cysteine residues. Yet, unlike cyanogen bromide, cyanide ion is seldom used as a cleavage agent in modern laboratories interested in protein characterization. Protein biochemists generally rely on only a handful of sequencing "tools" such as trypsin or cyanogen bromide to cut proteins into smaller, more manageable peptides. For this reason, additional methods for the specific cleavage of proteins are highly desirable since they would provide more flexibility especially when dealing with difficult sequencing problems. The potential of cyanide to specifically generate larger peptides with overlapping sequences was a compelling reason to examine this reaction more closely. For a quick and accurate measure of the molecular weights of the products, MALDI-TOF mass spectrometry turned out to be an ideal method.

VI.2 Specific Cleavage of Proteins by S-Cyanylation

Almost exactly a century ago, Lang reported that mammals react to cyanide poisoning by excreting thiocyanate. Decades later, Wood and Cooley confirmed this result by noting that rats fed methionine-\textsuperscript{35}S prior to injection with cyanide produced radiolabeled thiocyanate in their urine. Interestingly, when they repeated the same
experiment using cystine-$^{35}$S instead of methionine-$^{35}$S, the major radioactive metabolite upon isolation was determined to be 2-iminothiazolidine-4-carboxylic acid (II) which is formed by cyclization of β-thiocyanoalanine (I) as shown below.\textsuperscript{166}

\[
\begin{align*}
\text{(I)} & \quad \xrightarrow{\text{N--C}} \quad \text{(II)} \\
\text{HN} & \quad \text{CH--C--OH} & \quad \text{HN} & \quad \text{S--C--O} & \quad \text{CH} \\
\text{S} & \quad \text{CH--C--OH} & \quad \text{HN} & \quad \text{S--C--O} & \quad \text{CH}
\end{align*}
\]

These findings suggested the existence of an alternative (non-enzymatic) pathway for cyanide detoxification involving endogenous cystines or protein disulfide bonds and prompted a number of investigators to explore the reaction of cyanide with a variety of different disulfide bond containing peptides and proteins \textit{in vitro}.\textsuperscript{167-172} The collective conclusion was that cyanide ion can be utilized as a specific chemical reagent for cleaving proteins at the N-terminus of cysteines participating in disulfide bonds.\textsuperscript{173}

The first step of the reaction (shown in Figure VI.1) involves nucleophilic attack by cyanide ion to open the disulfide bond resulting in the formation of thiocyanoalanine and cysteine. In the second step, thiocyanoalanine cyclizes at alkaline pH to form an iminothiazolidine ring with concomitant peptide backbone bond cleavage. In the ideal case, each cysteine in any given disulfide has an equal probability of getting cyanylated in the first step. Since cyanide does not react with free thiol groups the resulting peptides should
Figure VI.1. The reaction of cyanide with a protein containing a disulfide bond.
contain overlapping sequence information as shown by the products of step 2 in Figure VI.1. Unfortunately, preservation of the overlapping sequence regions is hampered by the tendency of free sulfhydryl groups to reoxidize to form new disulfide bonds under basic conditions. Furthermore, a major problem lies in the reversibility of the S-cyanylation step (requiring large excesses of cyanide ion) since free sulfhydryls can react with thiocyanoalanine to release cyanide ion and form a new disulfide bond. The mechanism is illustrated in Figure VI.2 and provides an explanation for the reversible nature of the first step leading to the scrambling of disulfide bonds and eventually to the formation of peptides lacking any overlap. For example, when ribonuclease, which contains four disulfide bonds, was incubated with cyanide, nine peptides were detected (all iminothiazolidine blocked except for the N-terminal peptide) covering the entire sequence yet none of them contained overlapping regions.

Cyclization of the β-thiocyanoalanine and consequent cleavage of the peptide backbone is induced by performing the reaction at alkaline pH. However, prolonged incubation of the protein in alkaline solution can produce dehydroalanine via β-elimination of the thiocyanoalanine releasing thiocyanate and preventing cyclization and hence, backbone cleavage. In addition, other possible side products may complicate the analysis of the reaction mixture but these seem not to have been investigated in any detail.

Electrophilic reagents, such as 2-nitro-5-thiocyanobenoic acid (NTCB) which is a
Figure VI.2. Non-native disulfide bond formation due to the reversibility of S-cyanylation with cyanide.
source of CN⁺, were developed to overcome some of these problems and were found to be generally more reliable. The reaction of a peptide sulfhydryl group with NTCB to form cyanylated cysteine is shown below:

\[
\text{Peptide + NTCB} \rightarrow \text{Cyanylated Peptide + NTCB}
\]

The drawback to the use of electrophilic cyanide reagents is that three separate reaction conditions are required for the disulfide reduction, sulfhydryl cyanylation, and backbone cleavage steps in contrast to the single step nucleophilic cyanide method. Although the S-cyanylated product predominates, NTCB has been observed to form a mixed disulfide (with mercaptanitrobenzoate) preventing S-cyanylation and peptide bond cleavage. More importantly, it is no longer possible to generate overlapping peptides since electrophilic reagents do not react with disulfide bonds. In practice, neither method has gained widespread acceptance because the iminothiazolidine peptides produced are N-terminally blocked and thus not amenable to Edman degradation sequencing although the N-terminal iminothiazolidine can be reduced to alanine using a nickel catalyst. It has also been suggested that a pyrrolidone carboxylyl peptidase may provide a future method for the enzymatic removal of these N-terminal blocking groups.
blocked peptides, regardless of type, are easily sequenced using mass spectrometry.\textsuperscript{123,180} For example, the primary sequence of the protease inhibitor from \textit{Sarcophaga bullata} was determined directly from the collision induced dissociation (CID) mass spectra of the shorter, NTCB derived, N-terminally blocked iminothiazolidine peptides.\textsuperscript{181} The nucleophilic cyanide ion method can in principle produce larger N-terminal iminothiazolidine peptides containing internal cysteines. These larger peptides (containing overlapping sequence regions) would be more easily ionized by MALDI than by FAB.

VI.3 A New Strategy for the Optimization of the Cyanide Cleavage Reaction to Generate Overlapping Peptide Fragments

The simplicity of the nucleophilic cyanide reaction and its potential to generate larger peptide fragments with overlapping sequence information prompted the reexamination of this reaction using MALDI-TOF-MS-- a method particularly well suited to the analysis of larger peptides and proteins. As mentioned in the previous section, the main problem with using cyanide as a protein cleavage reagent is apparently due to the reversibility of the first step of the reaction (Step 1 in Figure VI.1) and consequently, the need for a large molar excess of cyanide in the reaction mixture. Based on this hypothesis, a new strategy was developed in an attempt to shift the equilibrium of the S-cyanylation reaction toward products. Thus, free sulfhydryl groups formed by cyanide disulfide bond

158
scission were specifically targeted to prevent them from reoxidizing to form new disulfide bonds. This was accomplished by the addition of an S-alkylation reagent to the reaction mixture in order to "trap" the free sulfhydryls as they formed. Figure VI.3 illustrates this concept with 4-vinylpyridine as the S-alkylation agent. Reaction of 4-vinylpyridine with a free sulfhydryl group results in the formation of the S-ethylpyridylated derivative which is denoted in Figure VI.3 with the symbol "EP". 4-Vinylpyridine was chosen as the preferred S-alkylation reagent due to its unique specificity toward free sulfhydryl groups and rapid reaction at alkaline pH.$^{182,183}$

Formation of a stable covalent cysteine adduct should in principle drive the equilibrium of the first step of the cyanide reaction to the product side. Therefore, the amount of cyanide required should decrease and the overall reaction rate should increase. The generated peptides would contain stable internal S-ethylpyridylated cysteines and thus preserve overlapping sequence information. Again, MALDI-TOF-MS should provide the ideal analysis method for these larger peptides. Finally, the effect of the addition of the S-alkylation agent on the reaction as a whole, including the monitoring of dehydroalanine production and any other side product formation, (should it occur), could all be simultaneously accomplished by MALDI-TOF-MS.
Figure VI.3. New strategy for optimization of the S-cyanylation step with cyanide by concomitant formation of S-ethylpyridylated (EP) peptides.
VI.4 Experimental Procedures

Solutions containing cyanide ion (typically 0.1 M) were prepared separately by dissolving solid NaCN or KCN in deionized water. The pH of the cyanide solution was carefully lowered by dropwise addition of 5-10% acetic acid using constant stirring. Care was exercised to not allow the pH of the cyanide solution to fall below pH 7 at any time during the addition of the acetic acid to prevent formation and escape of hydrogen cyanide at lower pH levels. Following pH adjustment, enough deionized water was added to achieve the desired final concentration. The total volume of the cyanide solution was intentionally rather large (usually 50 mL) to facilitate pH measurement.

Optimization of the cyanide cleavage reaction was performed empirically by selectively varying individual variables in the reaction conditions until the best possible result (as determined by MALDI-TOF-MS) was achieved. In general, the protein to be cleaved was dissolved in aqueous solution containing a denaturant such as guanidine hydrochloride and a buffer. An aliquot of the cyanide ion solution was then added with nearly simultaneous addition of an S-alkylation agent (4-vinylpyridine was added as a 2% solution in n-propanol). The majority of the experiments were performed using 2 M guanidine hydrochloride, a \(10^2\)-fold molar excess of 4-vinylpyridine, a \(10^3\)-fold molar excess of cyanide ion, and 100 mM Tris-HCl buffer (pH 8.5). The reaction mixtures (total volumes ranged from 50 to 200 \(\mu\)L) were normally incubated in a heating block at 60°C for
2 hours. The reaction was terminated by lowering the pH of the reaction mixture by adding glacial acetic acid. In some cases, a portion of the reaction solution was injected onto a reversed phase HPLC column. The chromatography methodology (RP-HPLC) has been already described in detail in Section III.2.

Aliquots from the reaction solutions were analyzed directly by MALDI-TOF-MS without any prior purification. The MALDI matrix solution was usually 10 g/L ACCA prepared in 50/50 (v/v) acetonitrile/water. A saturated solution of SA in 30/70 (v/v) acetonitrile/water was also used for some of the experiments. The reaction mixtures, which contained denaturants and buffers, were diluted with larger amounts of matrix solution to give a final concentration (for the protein and peptides) between 1-10 pmol/μL. Although MALDI is generally insensitive to contaminants such as buffers and denaturants, it is easier to achieve higher resolution and hence, mass accuracy by minimizing the final concentration of these contaminants. Consequently, the initial protein concentration of the reaction mixtures was usually about $1 \times 10^{-4}$ M which then allowed the dilution of the aliquots (with matrix solution) by at least one to two orders of magnitude.

MALDI-TOF-MS experiments were performed using two different instruments. The modified Vestec VT 2000 LD-TOF, (described previously in Section I.3), was used to acquire linear mode mass spectra with 30 kV accelerating voltage. Thirty laser shots were averaged for each scan with the resolution generally between 1:100 and 1:300. The
PerSeptive BioSystems Voyager Elite mass spectrometer was used to acquire linear, reflector, and post-source-decay (PSD) mode mass spectra. A more detailed description of this instrument can be found in Section II.3. Most of the mass spectra generated by the Voyager Elite were acquired in the linear mode with delayed extraction (DE) and 20 kV accelerating voltage to take advantage of the enhanced resolution (between 1:2000 and 1:4000). At least 64 (up to a maximum of 256) laser shots were summed for each mass spectrum. PSD spectra were acquired in the continuous (non-DE) mode using 25 kV acceleration. Complete PSD spectra were generated as described in Section III.3.

The sources for many of the reagents used for this work have already been described in Sections III.2 and IV.2. In addition, NaN, KCl, and guanidine thiocyanate were purchased from Fluka, an 8 M solution of guanidine hydrochloride from Pierce, all proteins and peptides (except E. coli thioredoxin which was obtained from Calbiochem), buffers from Sigma, and S-alkylation reagents including iodoacetate and N-ethylmaleimide from Aldrich.

VI.5 Results and Discussion

Preliminary experiments were performed with disulfide bond containing peptides and proteins using varying concentrations of reactants, denaturants, and different
temperatures in an attempt to optimize the reaction conditions. The MALDI-TOF mass spectra of the reaction of the protein aprotinin with cyanide after 2 h at 60 °C is shown in Figures VI.4A and VI.4B. Reaction conditions were identical for both (2 M guanidine-HCl, 100 mM Tris-HCl, pH 8.5, and a 1000-fold molar excess of NaCN) except that a 100-fold molar excess of 4-vinylpyridine was also added to the reaction mixture in Figure VI.4B. Clearly, the addition of 4-vinylpyridine was successful in shifting the equilibrium of the reaction to the right since Figure VI.4A contains many products while essentially no products were detected for the reaction mixture containing only cyanide ion (Figure VI.4B).

After three hours, nearly all of the aprotinin had been digested using the NaCN and 4-vinylpyridine combination. The primary structure of native aprotinin, which contains three disulfide bonds,184 is shown in Figure VI.5. The MALDI-TOF mass spectrum in Figure VI.6 is of the same reaction mixture as in Figure VI.4B after three hours and displays only the relevant peptide region. Figure VI.7 is a graphical representation of the peptides from this digest that shows the overlapping sequences. The numbered peptides in the mass spectrum (Figure VI.6), their position and sequence in the aprotinin amino acid sequence, and the calculated and observed m/z values are listed in Table VI.1. The peptide labels on the left of Figure VI.7 correspond to the peptide annotations in the mass spectrum (Figure VI.6) and to the entries in the first column of Table VI.1.

The MALDI-TOF mass spectrum in Figure VI.6 contains several interesting
Figure VI.4. MALDI-TOF mass spectra of reaction mixtures containing aprotinin after two hours.
A. NaCN only. B. NaCN and 4-vinylpyridine.
Figure VI.5. Primary structure of aprotinin showing disulfide bond structure.
Figure VI.6. MALDI-TOF mass spectrum of aprotinin digested with NaCN and 4-vinylpyridine after three hours. (The numbered peptide labels correspond to the annotations in Figure VI.7 and Table VI.1. Mass differences, denoted by the arrows, and the origin of the peaks labeled with "G" or asterisks are explained in the text).
Figure VI.7. Graphical representation of peptides resulting from digestion of aprotinin with NaCN and 4-vinylpyridine showing overlapping sequences. (The numbers on the left correspond to the peptide annotations in Figure VI.6 and Table VI.1. The position of specific cysteines are noted at the top [the numbers 1 and 58 refer to the aprotinin N-terminus and C-terminus, respectively]. The large dots on the peptide N-termini represent iminothiazolidine rings and "S-EP" denotes an S-ethylpyridylated cysteine at that position.)
Table VI.1. Peptides resulting from the digestion of aprotinin with NaCN and 4-vinylpyridine.

<table>
<thead>
<tr>
<th>Pep.*</th>
<th>Pos.</th>
<th>Calc. m/z&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Obs. m/z&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Sequence&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-13</td>
<td>1597.8*</td>
<td>1597.9</td>
<td>( ) RPDFCLEPPYTGP (C)</td>
</tr>
<tr>
<td>2</td>
<td>1-29</td>
<td>3554.2</td>
<td>3554</td>
<td>( ) RPDFCLEPPYTGCARKIIRYFYNAKAGL (C)</td>
</tr>
<tr>
<td>3</td>
<td>1-37</td>
<td>4515.4</td>
<td>4514.7</td>
<td>( ) RPDFCLEPPYTGCARKIIRYFYNAKAGLCTFVYGG (C)</td>
</tr>
<tr>
<td>4</td>
<td>5-29</td>
<td>2958.5</td>
<td>2958.7</td>
<td>(F) CLEPPYTGCARKIIRYFYNAKAGL (C)</td>
</tr>
<tr>
<td>5</td>
<td>5-37</td>
<td>3919.6*</td>
<td>3919.6**</td>
<td>(F) CLEPPYTGCARKIIRYFYNAKAGLCQTFVYGG (C)</td>
</tr>
<tr>
<td>6</td>
<td>14-29</td>
<td>1895.3</td>
<td>1895.3</td>
<td>(P) CARKIIRYFYNAKAGL (C)</td>
</tr>
<tr>
<td>7</td>
<td>14-37</td>
<td>2856.4</td>
<td>2856.6</td>
<td>(P) CARKIIRYFYNAKAGLCQTFVYGG (C)</td>
</tr>
<tr>
<td>8</td>
<td>30-50</td>
<td>2507.8</td>
<td>2508.2</td>
<td>(L) CQTFVYGGCRAKRNFKS-AED (C)</td>
</tr>
<tr>
<td>9</td>
<td>30-58</td>
<td>3516.1*</td>
<td>3516.3</td>
<td>(L) CQTFVYGGCRAKRNFKS-AEDCMRTCGGA ()</td>
</tr>
<tr>
<td>10</td>
<td>38-50</td>
<td>1546.7</td>
<td>1547</td>
<td>(G) CRAKRNFSAED (C)</td>
</tr>
<tr>
<td>11</td>
<td>38-54</td>
<td>2143.5</td>
<td>2144.1</td>
<td>(G) CRAKRNFSAEDCMRT (C)</td>
</tr>
<tr>
<td>12</td>
<td>38-58</td>
<td>2555.0*</td>
<td>2555.3</td>
<td>(G) CRAKRNFSAEDCMRTCGGA ()</td>
</tr>
<tr>
<td>13</td>
<td>51-58</td>
<td>929.1*</td>
<td>929.1**</td>
<td>(D) CMRTCGGA ()</td>
</tr>
</tbody>
</table>

<sup>a</sup>Peptide annotations from Figures VI.6 and VI.7.
<sup>b</sup>The position of the peptide in the amino acid sequence of aprotinin.
<sup>c</sup>The calculated masses for the dehydration product, (M+H)⁺ -18, except for those peptides marked with an asterisk where the calculation is for the (M+H)⁺ ion.
<sup>d</sup>Peptides marked with a double asterisk were used as internal calibrants.
<sup>e</sup>N-terminal cysteines are cyclized iminothiazolidine groups. Internal cysteines (in bold) are S-ethylpyridylated. Amino acids in parentheses before and after the peptide sequence correspond to the preceding and following residues, respectively.
features. First of all, ions at calculated m/z values for (M+H)^+ ions for most of the peptides were either very weak or absent completely. Instead, the dehydration product, m/z = (M+H)^+ - 18 was observed. Peptide #11 was an exception with significant intensities for both the (M+H)^+ and (M+H)^+ - 18 ions as indicated in Figure VI.6. In addition, although the observed m/z for peptide #1 was predominantly due to the (M+H)^+ ion, the (M+H)^+ - 18 was also detected for this peptide. On the other hand, three of the peptides (#9, #12, and #13) were detected solely as (M+H)^+ ions. A quick glance at Figure VI.7 shows that peptides #9, #12, and #13 are all C-terminal peptides, i.e., they contain the C-terminus of aprotinin. Therefore, it would appear from this experiment that any peptide that is present wholly or in part as the dehydration product cannot contain the C-terminus of the original protein.

Only two (#6 and #10) of the thirteen peptides detected were the result of complete cleavage and therefore did not contain internal cysteines. Although it is possible that the remaining five possible complete cleavage products may have been too small in size to be efficiently ionized by MALDI, peptides 5-13 and 30-37 (complete cleavages) should have been large enough (calculated m/z 1002.1 and 900.0, respectively) to be easily detected by MALDI-TOF mass spectrometry. Therefore, approximately 85% of the peptides generated from digesting aprotinin with cyanide and 4-vinylpyridine contained overlapping sequences.
Peptides containing at least two internal S-ethylpyridylated cysteines were always accompanied by a dehydroalanine analogue formed by elimination of the 4-(thioethyl)-pyridine:

This was not problematic for two reasons. First of all, the dehydroalanine containing peptide was never the major product as compared to the S-ethylpyridylated peptide and secondly, dehydroalanine products (although arising from cyanylated cysteines rather than S-ethylpyridylated ones) can easily be identified in the mass spectrum as (M+H)^+ - 139 as shown above. The dehydroalanine containing peptides, [(M+H)^+ - 139], in the mass spectrum (Figure VI.6) are 139 Da lower in mass than the corresponding S-ethylpyridylated peptide and are indicated by the arrows. It should be emphasized that these arrows represent chemical cleavage products that are 139 Da lighter than the corresponding expected peptides and therefore are not fragment ions. Apart from dehydration, dehydroalanine formation was the major side product of the reaction with five of the thirteen peptides exhibiting this derivative. Two of these five peptides also formed
lanthionine analogues but are again minor products in comparison with the desired S-ethylpyridylated derivative. The lanthionines are characterized by a mass difference of -244 Da from the fully S-ethylpyridylated derivative product as indicated in Figure VI.6. Lanthionines form via the attack of a free sulphydryl group on thiocyanoalanine with release of thiocyanate as shown below:\textsuperscript{168,173}

\[ \text{Lanthionine} \]

\[ \text{S-Ethylpyridylated peptide (M+H)}^+ -244 \]

The (M+H)$^+$ - 244 arrows indicating lanthionine containing peptides are similar to the notation used for the (M+H)$^+$ - 139 peaks discussed earlier in that they represent a chemical cleavage product rather than fragment ions.

The high specificity of the cyanide/4-vinylpyridine cleavage reaction is
demonstrated by the fact that only two peptides (marked with asterisks in Figure VI.6) are due to non-specific cleavage. The first, at $m/z$ 1443.8 corresponds to the peptide sequence 1-11 and the second ($m/z$ 2201.7) to peptide sequence 37-54. The peak at $m/z$ 1639.9 (labelled G) turned out to be the N-terminal peptide #1 (1-13) guanidylated by reaction with guanidinium chloride. This side reaction occurs predominantly with unblocked, free N-terminal peptides, and leads to the following product:

\[
\begin{align*}
\text{H}_2\text{N} & \text{C} - \text{NH} - \text{CH} - \text{C} & \ldots & \text{NH} - \text{CH} - \text{C} - \text{OH} \\
| & \text{R}_1 & | & \text{R}_2 \\
\text{Guanidyl moiety} & & \text{Peptide}
\end{align*}
\]

*E. coli* thioredoxin was one of the initial larger test proteins selected due to its commercial availability and more importantly, because it contained only one disulfide bond. Despite the unique reactivity of the thioredoxin cysteines (*thioredoxin is a potent protein disulfide reductase*\(^{185,186}\)), the protein resisted cyanlation under non-denaturing conditions even after incubation with cyanide and 4-vinylpyridine for 24 hours at 60 °C. This lack of reactivity in the absence of a denaturant was a consistent result for all of the proteins tested. Urea, although an excellent denaturant, was found to carbamylate primary amino groups very easily and was therefore not used. Heat, non-aqueous solvents, and guanidine thiocyanate were also evaluated as potential denaturation agents, however,
guanidine hydrochloride (2 M) proved to be the best overall reagent for this purpose. Although some guanidylation products were observed, these occurred primarily with N-terminal peptides of proteins and could be readily identified by the presence of the (M+H)+ + 42 ion as already discussed above. An added benefit of using a denaturant such as guanidine hydrochloride was its ability to solubilize proteins that ordinarily have poor solubility in alkaline solution.

The primary sequence of *E. coli* thioredoxin and the overlapping peptides expected from a cyanide/4-vinylpyridine digest of this protein are illustrated in Figure VI.8. As in Figure VI.7, the dots symbolize N-terminal iminothiazolidine groups and "S-EP" denotes an S-ethylpyridylated cysteine. Figures VI.9A and VI.9B are MALDI-TOF mass spectra of *E. coli* thioredoxin reaction mixtures after 1 hour at 60 °C containing a 10-fold and 100-fold molar excess of 4-vinylpyridine, respectively. Both solutions are otherwise identical, each containing 2 M guanidine hydrochloride and a 1000-fold molar excess of cyanide ion at pH 8.5. Figure VI.9A is dominated by the singly and doubly charged peaks, (M+H)+ at m/z 11,674 and (M+2H)++ at m/z 5838, respectively, due to the native, unreacted *E. coli* thioredoxin. In addition, the most prominent peaks are at m/z 3538 and 7925 which correspond to the (M+H)+ ions of peptides 1-31 and 35-108, respectively. However, at the larger excess of 4-vinylpyridine, the reaction is more complete (Figure VI.9B). The ions at m/z 3539 and 7925 are now abundant and there are additional peaks at m/z 3900 and 8288 which correspond to the (M+H)+ ions of the overlapping peptides 1-34 and 32-108 as
Figure VI.8. The primary structure of *E. coli* thioredoxin indicating disulfide bond location and peptides expected from a digest with cyanide and 4-vinylpyridine. (The dots on the N-termini indicate cyclized iminothiazolidine groups and "S-EP" denotes S-ethylpyridylated cysteine).
Figure VI.9. MALDI-TOF mass spectra of reaction mixtures containing *E. coli* thioredoxin and NaCN after one hour. Ten-fold (A) and 100-fold (B) molar excess of 4-vinylpyridine, respectively. (See text for annotations and reaction conditions).
predicted in Figure VI.8. An artifact of this reaction with thioredoxin is the unspecific cleavage of the protein between residues 33 and 34 giving rise to peptides marked with the asterisks in Figure VI.9B at m/z 3426 and 8102 ((M+H)^+ ions for peptides 1-33 and 34-108, respectively). The unexpected formation of these peptides probably is due to the unusually close proximity of the two cysteines (positions 32 and 35) in the primary structure and the unique characteristics of that disulfide bond.\textsuperscript{186} The extent of this cleavage of a specific bond not involving cysteine has not been found in similar experiments with other proteins and peptides. On the other hand, in analogy with the aprotinin experiment, neither of the C-terminal peptides generated from thioredoxin exhibited an (M+H)^- - 18 ion. Conversely, such dehydration products were detected for both the 1-31 and 1-34 peptides. No dehydroalanine or lanthionine containing peptides were observed with thioredoxin. Finally, repeating the same experiment with more 4-vinylpyridine (10\textsuperscript{3}-fold molar excess) did not increase the rate of reaction and served only to complicate the mass spectrum by introducing additional nonspecific ethylpyridylation adducts.

Catsimpoulas and Wood, who were among the first to investigate protein cleavage with cyanide proposed that formation of iminothiazolidine increases with proton concentration.\textsuperscript{168} On the basis of their experiments with glutathione, they suggested that the cyanylation reaction be performed with cyanide in the pH range 6-8 to increase the rate of S-cyanylation and to retard the formation of β-elimination products. The conclusion that lower pH favors S-cyanylation is surprising since one would expect \textit{a priori} that the
reaction rate should be slower because hydrogen cyanide is a poorer nucleophile than cyanide ion. Catsimpoulas and Wood also hypothesized that thiocyanoalanine forms a stable cyclic iminothiazolidine intermediate in this pH range without concomitant peptide bond cleavage. Jacobson et al., however, were unable to detect an internal (i.e., uncleaved) cyclic iminothiazolidine intermediate after S-cyanylation\textsuperscript{176}. Their conclusion was that iminothiazolidine formation and peptide bond cleavage is a concerted process that is catalyzed by hydroxide ion as shown below:

\[
\begin{align*}
\text{O} & \quad \text{H} \\
\text{N} & \quad \text{C} \\
\text{S} & \quad \text{O} \\
\text{O} & \quad \text{C} \quad \text{N} \\
\text{H} & \quad \text{N} \\
\end{align*}
\]

Both \textit{E. coli} thioredoxin and ribonuclease A were reacted with cyanide and 4-vinylpyridine at different pH values (6.5, 7.5, 8.5, and 9.5) to investigate the effect of pH on the generation of overlapping peptides. For thioredoxin, all of the pH values resulted in the production of the expected peptides (Figure VI.8), however, the reaction rate was much slower at the lower pH values (6.5 and 7.5) with the majority of the thioredoxin still
uncleaved after long incubation times at elevated temperatures. The effect of pH was more noticeable with bovine ribonuclease A, a protein containing four disulfide bonds. Figure VI.10 compares the MALDI-TOF mass spectra of ribonuclease A digested with cyanide and 4-vinylpyridine at pH 6.5 (Figure VI.10A) and pH 8.5 (Figure VI.10B) after two hours at 60 °C. In each spectrum, the numbered peaks represent expected cleavage products and correspond to the peptide labels column in Table VI.2. (Table VI.2 is a compilation of the mass spectrometric data from cleavage of ribonuclease A at pH 8.5). The peptides listed in Table VI.2 are also graphically represented in Figure VI.11 to emphasize the overlapping sequences produced. Similar to thioredoxin, the rate of the ribonuclease digest increased with increasing pH. However, with ribonuclease A, eight of the nineteen peptides detected at pH 8.5 were not observed when the identical reaction was performed at pH 6.5.

Although three contaminant or non-specific cleavage peaks (marked by asterisks) in Figure VI.10A are relatively easily ionized at pH 6.5 they are barely detectable at pH 8.5 (Figure VI.10B). This observation can be attributed to the higher product peptide concentration at pH 8.5. Since more of the protein is digested at pH 8.5, the product peptide concentration is also increased. A distinguishing feature of the mass spectrum in Figure VI.10A is the presence of the peaks with a mass difference of -80 mass units as denoted by the arrows. Similar to the ones containing dehydroalanine, these peptides do not arise from the fragmentation of the S-ethylpyridylated peptide but are related by having the same amino acid sequence. At pH 6.5, (M+H)+ - 80 ions are observed that correspond to peptides with an internal S-cyanylated rather than S-ethylpyridylated cysteine. Two of the fully S-
Figure VI.10. MALDI-TOF mass spectra of ribonuclease A digested with NaCN and 4-vinylpyridine after 2 hours. A. pH 6.5. B. pH 8.5. (See text for annotations and reaction conditions). Some of the peaks not labeled in (B) are due to (M+H - H₂O)^+ ions.
Figure VI.11. Graphical representation of peptides resulting from digestion of ribonuclease A with NaCN (pH 8.5) and 4-vinylpyridine showing overlapping sequences. (The numbers on the left correspond to the peptide annotations in Figure VI.10B and Table VI.2. The position of specific cysteines are noted at the top [the numbers 1 and 124 refer to the ribonuclease A N-terminus and C-terminus, respectively]. The large dots on the peptide N-termini represent iminothiazolidine rings and "S-EP" denotes an S-ethylpyridylated cysteine at that position).
Table VI.2. Peptides resulting from the digestion of ribonuclease A with NaCN and 4-vinylpyridine at pH 8.5 for 2 hours. (Table continues on the next page).

<table>
<thead>
<tr>
<th>Pep.</th>
<th>Pos.</th>
<th>Calc. m/z</th>
<th>Obs. m/z</th>
<th>Sequence d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-25</td>
<td>2705.9</td>
<td>2707.1</td>
<td>( ) KETAAKFERQHMDSTSAASSSNY (C)</td>
</tr>
<tr>
<td>2</td>
<td>1-39</td>
<td>4518.1**</td>
<td>4518.1</td>
<td>( ) KETAAKFERQHMDSTSAASSSNYCNQMMKSRNLTKDR (C)</td>
</tr>
<tr>
<td>3</td>
<td>26-39</td>
<td>1751.1</td>
<td>1751.4</td>
<td>(Y) CNQMMKSRNLTKDR (C)</td>
</tr>
<tr>
<td>4</td>
<td>40-57</td>
<td>1983.3</td>
<td>1984.1</td>
<td>(R) CKPVNFTVHESLADVQAV (C)</td>
</tr>
<tr>
<td>5</td>
<td>40-64</td>
<td>2819.3</td>
<td>2820.9</td>
<td>(R) CKPVNFTVHESLADVQAVCSQKNVA (C)</td>
</tr>
<tr>
<td>6</td>
<td>40-71</td>
<td>3670.2</td>
<td>3671.6</td>
<td>(R) CKPVNFTVHESLADVQAVCSQKNVACKNGQTN (C)</td>
</tr>
<tr>
<td>7</td>
<td>58-94</td>
<td>4440.0</td>
<td>4438.7</td>
<td>(V) CSQKNVACKNGQTNCYQSYTMSITDCYQSYTMSITDCRETGSSKYPN (C)</td>
</tr>
<tr>
<td>8</td>
<td>65-83</td>
<td>2275.5</td>
<td>2276.2</td>
<td>(A) CKNGQTNCYQSYTMSITD (C)</td>
</tr>
<tr>
<td>9</td>
<td>65-94</td>
<td>3604.0</td>
<td>3605.6</td>
<td>(A) CKNGQTNCYQSYTMSITDCYQSYTMSITDCRETGSSKYPN (C)</td>
</tr>
<tr>
<td>10</td>
<td>72-83</td>
<td>1424.6</td>
<td>1425.7</td>
<td>(N) CYQSYTMSITD (C)</td>
</tr>
<tr>
<td>11</td>
<td>72-94</td>
<td>2753.0</td>
<td>2754.1</td>
<td>(N) CYQSYTMSITDCRETGSSKYPN (C)</td>
</tr>
<tr>
<td>12</td>
<td>72-109</td>
<td>4501.1</td>
<td>4500.5</td>
<td>(N) CYQSYTMSITDCRETGSSKYPNCAYKTTQANKHIIVA (C)</td>
</tr>
<tr>
<td>13</td>
<td>72-124</td>
<td>6222.1</td>
<td>6217.6</td>
<td>(N) CYQSYTMSITDCRETGSSKYPNCAYKTTQANKHIIVACEGNPSYVPHF-DASV ( )</td>
</tr>
<tr>
<td>14</td>
<td>84-94</td>
<td>1267.4</td>
<td>1267.3</td>
<td>(D) CRETGSSKYPN (C)</td>
</tr>
</tbody>
</table>
Table VI.2. (continued from the previous page). Peptides resulting from the digestion of ribonuclease A with NaCN and 4-vinylpyridine at pH 8.5 for 2 hours.

<table>
<thead>
<tr>
<th>Pep.*</th>
<th>Pos.*</th>
<th>Calc. m/z</th>
<th>Obs. m/z</th>
<th>Sequence*</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>84-109</td>
<td>3015.4</td>
<td>3017.4</td>
<td>(D) CRETGSSKYPNCAYKTTQANKHIIVA (C)</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>84-124</td>
<td>4736.4</td>
<td>4736.1</td>
<td>(D) CRETGSSKYPNCAYKTTQANKHIIVACEGNPYVPVHFDASV ( )</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>95-109</td>
<td>1687.0</td>
<td>1687.2</td>
<td>(N) CAYKTTQANKHIIVA (C)</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>95-124</td>
<td>3407.9</td>
<td>3409.1</td>
<td>(N) CAYKTTQANKHIIVACEGNPYVPVHFDASV ( )</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>110-124</td>
<td>1659.8**</td>
<td>1659.8</td>
<td>(A) CEGNPYVPVHFDASV ( )</td>
<td></td>
</tr>
</tbody>
</table>

*Peptide annotations from Figures VI.10 and VI.11.

*The position of the peptide in the amino acid sequence of ribonuclease A.

*The calculated masses are for the (M+H)+ ion of the peptide. Peptides marked with a double asterisk were used as internal calibrants.

*N-terminal cysteines are cyclized iminothiazolidine groups. Internal cysteines (in bold) are S-ethylpyridylated. Amino acids in parentheses before and after the peptide sequence correspond to the preceding and following amino acids, respectively.
ethylpyridylated peptides, at m/z 4736.4 and m/z 3407.9, were accompanied by corresponding (M+H)^+ - 80 ions as indicated by the arrows in Figure VI.10A.

Again, the arrows represent a cleavage product that is 80 Da lighter than the expected S-ethylpyridylated peptide and are not meant to imply mass spectrometric fragmentation. The β-elimination of thiocyanalanine was successfully prevented at pH 6.5 since no (M+H)^+ - 139 ions indicating dehydroalanine formation were detected. Despite the fact that dehydroalanine formation was suppressed, there is no advantage to performing the reaction at this lower pH since the reaction seems to be quite incomplete. Furthermore, the (M+H)^+ ions due to the fully S-ethylpyridylated peptides were much more abundant in the mass spectrum of the pH 8.5 reaction mixture. There were no (M+H)^+ - 80 ions detected at pH 8.5 and the dehydroalanine products, although present, produced weak signals in comparison to the S-ethylpyridylated peptide ions. The MALDI mass spectrum of the ribonuclease A reaction at pH 9.5 was almost identical to the one at pH 8.5. Consequently,
raising the pH to 9.5 was not particularly helpful since no additional peptides were observed (data not shown). Furthermore, undesirable dehydroalanine formation was enhanced at pH 9.5 and the intensity of signals due to lanthionine containing peptides increased significantly.

In summary, optimal conditions for cleavage and generation of overlapping peptides were achieved by using a 10^3-fold molar excess of cyanide (NaCN), 10^2-fold molar excess of 4-vinylpyridine, guanidine hydrochloride (2 M), 100 mM Tris-HCl at pH 8.5 and 60 °C for 2-4 hours. For all proteins and peptides tested, the pH range between 8-9 was the best compromise in that the reaction was rapid and produced easily identifiable products. Lysozyme (from hen egg white)^189 and trypsinogen (from bovine pancreas)^190 were also cleaved using the cyanide/4-vinylpyridine methodology. The data for the lysozyme cleavage is tabulated in Table VI.3 and graphically represented in Figure VI.12. Similarly, the data due to trypsinogen cleavage is shown in Table VI.4 and Figure VI.13. In both cases, peptides with overlapping sequences were successfully generated as illustrated by Figures VI.12 and VI.13.

As mentioned at the beginning of this section, systematic experiments were also performed with disulfide bond containing peptides including oxytocin, brain natriuretic peptide, somatostatin, melanin concentrating hormone and urotensin II to provide simpler models. Overlapping peptides were successfully generated for each of these peptides using
Table VI.3. Peptides resulting from the digestion of lysozyme with NaCN and 4-vinylpyridine at pH 8.5 for 2 hours.

<table>
<thead>
<tr>
<th>Pept.</th>
<th>Pos.</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-29</td>
<td>( ) KVFGRCELAAMKRHGLDNYRGYSGLGNWV (C)</td>
</tr>
<tr>
<td>2</td>
<td>1-63</td>
<td>( ) KVFGRCELAAMKRHGLDNYRGYSGLGNWVCAAKF-ESNFNTQATNRNTDGSTDYGILQINSRWW (C)</td>
</tr>
<tr>
<td>3</td>
<td>6-29</td>
<td>(R) CELAAAMKRHGLDNYRGYSGLGNWV (C)</td>
</tr>
<tr>
<td>4</td>
<td>6-63</td>
<td>(R) CELAAAMKRHGLDNYRGYSGLGNWVCAAKF-ESNFNTQATNRNTDGSTDYGILQINSRWW (C)</td>
</tr>
<tr>
<td>5</td>
<td>30-63</td>
<td>(R) CAAKFESNFNTQATNRNTDGSTDYGILQINSRWW (C)</td>
</tr>
<tr>
<td>6</td>
<td>64-75</td>
<td>(W) CNDGRTPGSRNL (C)</td>
</tr>
<tr>
<td>7</td>
<td>64-79</td>
<td>(W) CNDGRTPGSRNLCNIP (C)</td>
</tr>
<tr>
<td>8</td>
<td>64-93</td>
<td>(W) CNDGRTPGSRNLCNIPCSALLSSDITASVN (C)</td>
</tr>
<tr>
<td>9</td>
<td>76-93</td>
<td>(L) CNIPCSALLSSDITASVN (C)</td>
</tr>
<tr>
<td>10</td>
<td>76-114</td>
<td>(L) CNIPCSALLSSDITASVNCAKKIVSDGDGMNAWVAWRNR (C)</td>
</tr>
<tr>
<td>11</td>
<td>80-129</td>
<td>(P) CSALLSSDITASVNCAKKIVSDGDGMNAWVAWRNRCKGT-DVQAIRGCRL ()</td>
</tr>
<tr>
<td>12</td>
<td>94-114</td>
<td>(N) CAKKIVSDGDGMNAWVAWRNR (C)</td>
</tr>
<tr>
<td>13</td>
<td>94-126</td>
<td>(N) CAKKIVSDGDGMNAWVAWRNRCKGTDVQAIRGCRL ()</td>
</tr>
<tr>
<td>14</td>
<td>94-129</td>
<td>(N) CAKKIVSDGDGMNAWVAWRNRCKGTDVQAIRGCRL ()</td>
</tr>
<tr>
<td>15</td>
<td>115-126</td>
<td>(R) CKGTDVQAIRG (C)</td>
</tr>
<tr>
<td>16</td>
<td>115-129</td>
<td>(R) CKGTDVQAIRGCRL ()</td>
</tr>
</tbody>
</table>

Peptide annotations from Figure VI.12.

The position of the peptide in the amino acid sequence of lysozyme.

N-terminal cysteines are cyclized iminothiazolidine groups. Internal cysteines (in bold) are S-ethylpyridylated. Amino acids in parentheses before and after the peptide sequence correspond to the preceding and following amino acids, respectively.
Figure VI.12. Graphical representation of peptides resulting from digestion of lysozyme with NaCN (pH 8.5) and 4-vinylpyridine showing overlapping sequences. (The numbers on the left correspond to the peptide annotations in the first column of Table VI.3. The position of specific cysteines are noted at the top [the numbers 1 and 129 refer to the lysozyme N-terminus and C-terminus, respectively]. The large dots on the peptide N-termini represent iminothiazolidine rings and "S-EP" denotes an S-ethylpyridylated cysteine at that position).
Table VI.4. Peptides resulting from the digestion of trypsinogen with NaCN and 4-vinylpyridine at pH 8.5 for 2 hours.

<table>
<thead>
<tr>
<th>Pept.*</th>
<th>Pos.</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-30</td>
<td>( ) VDDDDKIVGGYTCGANTVPYQVSLNSGYHF (C)</td>
</tr>
<tr>
<td>2</td>
<td>1-46</td>
<td>( ) VDDDDKIVGGYTCGAVT...YHFCGGSLINSQWVVSAAH (C)</td>
</tr>
<tr>
<td>3</td>
<td>13-30</td>
<td>(T) CGANTVPYQVSLNSGYHF (C)</td>
</tr>
<tr>
<td>4</td>
<td>13-46</td>
<td>(T) CGANTVPYQVSLNSGYHFCGGSLINSQWVVSAAH (C)</td>
</tr>
<tr>
<td>5</td>
<td>47-114</td>
<td>(H) CYKSGIQVRLGQ...ASLNSRVASISLPTS (C)</td>
</tr>
<tr>
<td>6</td>
<td>47-121</td>
<td>(H) CYKSGIQVRLGQ...ASLNSRVASISLPTSCASAGTQ (C)</td>
</tr>
<tr>
<td>7</td>
<td>115-142</td>
<td>(S) CASAGTQCLISGWNTKSSGTSYPDVLK (C)</td>
</tr>
<tr>
<td>8</td>
<td>115-153</td>
<td>(S) CASAGTQCLISGWNTKSSGTSYPDVLKCLKAPILSNSS (C)</td>
</tr>
<tr>
<td>9</td>
<td>122-142</td>
<td>(Q) CLISGWNTKSSGTSYPDVLK (C)</td>
</tr>
<tr>
<td>10</td>
<td>143-167</td>
<td>(K) CLKAPILSNSSCKSAYPGQITSNMF (C)</td>
</tr>
<tr>
<td>11</td>
<td>143-178</td>
<td>(K) CLKAPILSNSSCKSAYPGQITSNMFQGDSGPPV (C)</td>
</tr>
<tr>
<td>12</td>
<td>154-167</td>
<td>(S) CKSAYPGQITSNMF (C)</td>
</tr>
<tr>
<td>13</td>
<td>154-178</td>
<td>(S) CKSAYPGQITSNMFQGDSGPPV (C)</td>
</tr>
<tr>
<td>14</td>
<td>179-202</td>
<td>(S) CQGDGGPPVVCSEKQGIVSGW (C)</td>
</tr>
<tr>
<td>15</td>
<td>179-229</td>
<td>(S) CQGDGGPPVVCSEKQGIVSGWGCQKNKPGVYTKV (C)</td>
</tr>
<tr>
<td>16</td>
<td>203-229</td>
<td>(G) CAQKNKPGVYTKVCNYVSWIKQTIASN ( )</td>
</tr>
<tr>
<td>17</td>
<td>216-229</td>
<td>(V) CNYVSWIKQTIASN ( )</td>
</tr>
</tbody>
</table>

*aPeptide annotations from Figure VI.13.

bThe position of the peptide in the amino acid sequence of trypsinogen.

cN-terminal cysteines are cyclized iminothiazolidine groups. Internal cysteines (in bold) are S-ethylpyridylated. Amino acids in parentheses before and after the peptide sequence correspond to the preceding and following amino acids, respectively.
Figure VI.13. Graphical representation of peptides resulting from digestion of trypsinogen with NaCN (pH 8.5) and 4-vinylpyridine showing overlapping sequences. (The numbers on the left correspond to the peptide annotations in the first column of Table VI.4. The position of specific cysteines are noted at the top [the numbers 1 and 229 refer to the trypsinogen N-terminus and C-terminus, respectively]. The large dots on the peptide N-termini represent iminothiazolidine rings and "S-EP" denotes an S-ethylpyridylated cysteine at that position).
the reaction conditions outlined in the preceding paragraph. For example, urotensin II has the sequence AGTADCFWKYCV with a disulfide bond between Cys-5 and Cys-11.191 Although purification of the reaction mixture is not necessary for MALDI-TOF-MS (Figure VI.14A), an HPLC fractionation (Figure VI.14B) was performed to obtain an independent quantitative estimate of the products. The peptides in each HPLC fraction were identified by MALDI-TOF-MS and the data is summarized in Table VI.5. Four major HPLC peaks were observed corresponding to fractions 1, 2, 4, and 5. The three fractions containing the peaks with the highest absorbance, fractions 2, 4, and 5, corresponded to the expected cleavage products with overlapping sequences. Unexpectedly, the fraction representing the fourth largest HPLC peak (fraction 1) contained a peptide due to non-specific cleavage. PSD mass spectra of the major products were obtained to confirm the assigned sequences in Table VI.5. The PSD mass spectra of urotensin II peptides $m/z$ 1267 [(M+H)$^+$] and $m/z$ 1249 [(M+H)$^+$ - H$_2$O] are shown in Figures VI.15A and VI.15B, respectively. The nomenclature for the peptide fragmentation observed in PSD spectra has already been discussed in detail in Section II.3. The sequence of the peptide (M+H)$^+$ = $m/z$ 1267, (AGTADCFWKY; Cys is S-ethylpyridylated) was confirmed by the nearly complete N-terminal b ion series in Figure VI.15A. The lower mass region contains expected immonium ions denoted by the capital letter abbreviations for specific amino acids ("VP" denotes the ion due to vinylpyridine). At slightly higher mass range, there are numerous peaks (mostly left unlabeled for clarity) due to internal fragment ions that provide further evidence for the assigned sequence. Finally, loss of ethylpyridine from the precursor ion is
Figure VI.14. Urotensin II digested with NaCN and 4-vinylpyridine after 2 hours. A. MALDI-TOF mass spectrum. B. HPLC chromatogram. (Collected fractions and retention time for native urotensin II are indicated).
Table VI.5. HPLC fractionated peptides resulting from digestion of urotensin II with NaCN and 4-vinylpyridine after 2 hours.\textsuperscript{a}

<table>
<thead>
<tr>
<th>HPLC Fraction</th>
<th>Position</th>
<th>Observed m/z</th>
<th>Calculated m/z</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1-11</td>
<td>1370.6</td>
<td>1370.6</td>
<td>( ) AGTADCFWKYC (V)</td>
</tr>
<tr>
<td>2</td>
<td>1-10</td>
<td>1267.3</td>
<td>1267.5</td>
<td>( ) AGTADCFWKY (C)</td>
</tr>
<tr>
<td>3</td>
<td>6-10</td>
<td>771.7</td>
<td>771.9</td>
<td>(D) CFWKY (C)</td>
</tr>
<tr>
<td>4</td>
<td>1-12</td>
<td>1574.9</td>
<td>1574.9</td>
<td>( ) AGTADCFWKYCV ( )</td>
</tr>
<tr>
<td>5</td>
<td>6-12</td>
<td>1079.2</td>
<td>1079.3</td>
<td>(D) CFWKYCV ( )</td>
</tr>
<tr>
<td>6</td>
<td>1-10</td>
<td>1249.4</td>
<td>1249.4</td>
<td>( ) AGTADCFWKY (C) - H\textsubscript{2}O</td>
</tr>
<tr>
<td>7</td>
<td>1-12</td>
<td>1435.6</td>
<td>1435.7</td>
<td>( ) AGTADCFWKYCV ( )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( ) AGTADCFWKYCV ( )</td>
</tr>
<tr>
<td>8</td>
<td>1-10</td>
<td>1309.3</td>
<td>1309.5</td>
<td>( ) guanidyl.- AGTADCFWKY (C)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The amino acids in parentheses before and after the peptide sequences denote the preceding and following amino acids, respectively. N-terminal cysteines are cyclized iminothiazolidine groups. Underlined cysteines (fraction 7) correspond to the dehydroalanine product while cysteines in bold are S-ethylpyridylated. Fractions 6 and 7 contain the dehydration and guanidylated products, respectively.
Figure VI.15. PSD mass spectra of peptides derived from urotensin II digestion with cyanide and 4-vinylpyridine. A. (M+H)+ = m/z 1267.5. B. (M+H)+ = m/z 1249.5. (Annotations are explained in the text.)
indicated by the "-EP". Most of the fragment ions in the PSD mass spectrum of the dehydration product (Figure VI.15B) are the same as those of the normal peptide [i.e., (M+H)$^+$ = m/z 1267] in Figure VI.15A. However, the mass spectrum of the dehydration product differs notably from that of the normal peptide in the large increase in the number of internal fragment ions throughout the mass spectrum. Since the loss of H$_2$O must occur N-terminal to the peptide backbone cleavage (no dehydration products are observed corresponding to peptides that contain the original C-terminus), a possible mechanism for the formation of the dehydration product (M+H)$^+$ = m/z 1249.5 is shown below:

![Diagram](https://example.com/diagram.png)

The $y_{n-18}$ ions (which are not usually observed in PSD spectra), a and b ions, and the absence of the $y_2$ ion together provide supporting evidence for the structure of the dehydration product shown above.
Finally, a discussion of this reaction is not complete without pointing out that the sequence of shorter peptides containing N-terminal iminothiazolidine groups (such as those derived from disulfide bond containing peptides) can be easily verified by obtaining a PSD mass spectrum of the peptide. For example, the PSD mass spectrum of the urotensin II peptide \((\text{M}+\text{H})^+ = m/z \, 1079.3\) (shown in Figure VI.16) confirms the sequence CFWKYCV where Cys-1 is the N-terminally blocked iminothiazolidine and Cys-6 is S-ethylpyridylated. In contrast to the peptides in Figure VI.15, the presence of the \(b_1\) ion provides further confirmation for the N-terminal iminothiazolidine group since peptides with a free N-terminus generally do not produce \(b_1\) ions.\(^{122}\) Furthermore, the sensitivity of the PSD technique is illustrated by the fact that only approximately one picomole of peptide was used to obtain the complete spectrum.

VI.6 Conclusion

Specific cleavage of proteins at cysteine via S-cyanylation is certainly not a new reaction and has been studied since at least the early 1960's. As already mentioned in Section VI.2, it failed to gain much use principally due to the production of N-terminally blocked peptides that were not directly amenable to sequencing by Edman degradation. A renewed interest in this reaction was sparked by the demonstration that the iminothiazolidine N-terminally blocked peptides could easily be sequenced by mass
Figure VI.16. PSD mass spectrum of a peptide with an N-terminal iminothiazolididine group, $(M+H)^+ = m/z$ 1078.5. (The internal cysteine is S-ethylpyridylated. The peptide was generated from a NaCN/4-vinylpyridine digest of urotensin II. Annotations are explained in the text).
spectrometry. On the other hand, the potential of the nucleophilic cyanide reaction to produce peptides with overlapping sequences and thus identify native disulfide bond linkages was never fully realized. This failure was attributed to the tendency of free sulfhydryl groups to react with original disulfide bonds to form new disulfide bonds. The problem was exacerbated by the similar reaction of free sulfhydryls reacting with cyanylated cysteines to release cyanide with consequent disulfide bond rearrangement. Therefore, the next logical step was to attempt to block these unfavorable reactions.

Sequencing methods generally require at least two separate protein cleavage experiments utilizing enzymes or chemical cleavage agents with different specificities in order to obtain overlapping sequence information which then allows proper alignment of the peptides. The reaction of cyanide with proteins is unique in that it can take place in two different directions. Thus, for each disulfide bond, at least two different peptides can be formed each overlapping with each other. The addition of 4-vinylpyridine as an S-alkylation agent proved to be successful in blocking free sulfhydryl groups and preserving peptides with overlapping sequences. As predicted, addition of the S-alkylation agent served to shift the equilibrium of the reaction toward products reducing the reaction time and lowering the amount of cyanide required. The cyanide/4-vinylpyridine reaction, despite being a chemical cleavage method, rivals some enzymatic methods in its specificity. Although side reactions do occur, the products are easily identifiable and can actually be informative. Dehydroalanine formation by β-elimination of thiocyanate is never
the main product and is identified by its mass deficiency of 139 mass units as compared to
the fully S-ethylpyridylated product. Lanthionine containing product ions arising from the
expulsion of thiocyanate by a sulfide ion (see page 172) without formation of
dehydroalanine are considerably rarer and are observed only with larger peptides containing
multiple internal cysteines. Finally, dehydration products are fortuitously informative since
it has been observed that they only occur when the peptide does not contain the original C-
terminus of the original protein.

Earlier workers often had to rely on indirect methods such as calorimetry,
colorimetry, or amino acid analysis to establish and quantitate the components of the
cyanide/protein reaction mixtures. MALDI-TOF-MS with its tolerance for contaminants
such as buffers and salts and low detection limits (in the low picomole to femtomole range)
is especially well suited to the analysis of unpurified reaction mixtures containing proteins
and larger peptides. Since cysteines are generally not particularly abundant in proteins, the
peptides generated tend to be fewer and larger making cyanide digestion an ideal
application for MALDI-TOF-MS analysis. The low sample requirement due to the
sensitivity of MALDI coupled with the simplicity and rapidity of the cyanide/4-
vinylpyridine experiment makes this an easy experiment to perform without sacrificing
precious material. For example, it can provide a quick method for sequence confirmation of
proteins containing disulfide bonds. With unknowns, the newly emerging technique of
post-source-decay (PSD) is especially promising since sequence information can routinely
be obtained from just a few picomoles of peptide although the use of this method, in this case, is probably limited to the smaller peptides.

In conclusion, since overlapping peptides are generated by this method and MALDI-TOF-MS provides accurate mass measurement of the peptides, it should be possible, in principle, to align the peptides in their proper order (preferably using a computer program) without any prior knowledge of the amino acid sequence.
References


56. Zhou, J., Ens, W., Standing, K., and Verentchikov, A. "Kinetic Energy


66. Allen, G. "Specific Cleavage of Proteins by Chemical Methods" In Sequencing of Proteins and Peptides, Burdon, R.H., van Knippenberg, P.H. Eds., Elsevier,
Amsterdam, (1989), 95-104.


78. Chien, B.M., Michael, S.M., and Lubman, D.M. "Enhancement of Resolution in


100. Stark, G.R. "Sequential Degradation of Peptides from Their Carboxyl Termini with
Ammonium Thiocyanate and Acetic Anhydride" (1968) Biochemistry. 7, 1796-1807.


212


157. Wang, W., Papov, V.V., Minakawa, N., Matsuda, A., Biemann, K., and Hedstrom, L. "Inactivation of Inosine 5'-Monophosphate Dehydrogenase by the Antiviral Agent 5-Ethynyl-1-β-D-Ribofuranosylimidazole-4-Carboxamide 5'-Monophosphate" Biochemistry. [In Press].


Appendices

Appendix I.

Structures and nomenclature for fragment ions in high energy CID and PSD

*The "R" refers to the amino acid side chain. $R_a$ and $R_b$ in the structure of the internal fragment ion are not the N-terminal residue and C-terminal residue side chains, respectively.
Appendix II.

Additional fragment ions formed via high energy CID

\[ \text{H}^+ \quad \text{R} \quad \text{CHR'} \quad \text{H}^+ \quad \text{R} \quad \text{CHR'} \]

\[ \text{H}^- \quad (\text{NH}^- \quad \text{CH}^- \quad \text{CO})_{n-1}^- \quad \text{NH}^- \quad \text{CH} \]

\[ \quad \text{a}_n + 1 \]

\[ \text{H}^- \quad (\text{NH}^- \quad \text{CH}^- \quad \text{CO})_{n-1}^- \quad \text{NH}^- \quad \text{CH} \]

\[ \quad \text{d}_n \]

\[ \text{R} \quad \text{H}^+ \quad \text{CHR'} \quad \text{R}_{n-1} \quad \text{H^-} \quad (\text{NH}^- \quad \text{CH}^- \quad \text{CO})_{n-1}^- \quad \text{OH} \]

\[ \quad \text{z}_n + 1 \]

\[ \text{CHR'} \quad \text{R}_{n-1} \quad \text{H}^+ \quad \text{CH}^- \quad (\text{NH}^- \quad \text{CH}^- \quad \text{CO})_{n-1}^- \quad \text{OH} \]

\[ \quad \text{w}_n \]

\[ \text{H}^+ \quad \text{R}_{n-1} \quad \text{HN} = \text{CH}^- \quad \text{CO}^- \quad (\text{NH}^- \quad \text{CH}^- \quad \text{CO})_{n-1}^- \quad \text{OH} \]

\[ \quad \text{v}_n \]
Appendix III.

Additional fragment ions commonly found in PSD spectra

\[ \begin{align*}
\text{a}_n - 17 & \\
\text{b}_n - 17 & \\
\text{b}_n + \text{H}_2\text{O} & \\
\text{Internal fragment ion} - 28 (-\text{CO}) & \\
\end{align*} \]

*The above ions are more prevalent in PSD than in high energy CID mass spectra. In addition, dehydration occurs with certain residues to form \( a_n - 18 \) and \( b_n - 18 \) ions.*
Appendix IV.

Amino acid codes, nominal mass of residue\(^a\), corresponding immonium ions\(^b\) and side chain (R) loss\(^c\).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>Ala</td>
<td>A</td>
<td>71</td>
<td>44</td>
<td>-15</td>
</tr>
<tr>
<td>Arginine</td>
<td>Arg</td>
<td>R</td>
<td>156</td>
<td>129</td>
<td>-100</td>
</tr>
<tr>
<td>Asparagine</td>
<td>Asn</td>
<td>N</td>
<td>114</td>
<td>87</td>
<td>-58</td>
</tr>
<tr>
<td>Aspartic Acid</td>
<td>Asp</td>
<td>D</td>
<td>115</td>
<td>88</td>
<td>-59</td>
</tr>
<tr>
<td>Cysteine</td>
<td>Cys</td>
<td>C</td>
<td>103</td>
<td>76</td>
<td>-47</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Gln</td>
<td>Q</td>
<td>128</td>
<td>101</td>
<td>-72</td>
</tr>
<tr>
<td>Glutamic Acid</td>
<td>Glu</td>
<td>E</td>
<td>129</td>
<td>102</td>
<td>-73</td>
</tr>
<tr>
<td>Glycine</td>
<td>Gly</td>
<td>G</td>
<td>57</td>
<td>30</td>
<td>-1</td>
</tr>
<tr>
<td>Histidine</td>
<td>His</td>
<td>H</td>
<td>137</td>
<td>110</td>
<td>-81</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Ile</td>
<td>I</td>
<td>113</td>
<td>86</td>
<td>-57</td>
</tr>
<tr>
<td>Leucine</td>
<td>Leu</td>
<td>L</td>
<td>113</td>
<td>86</td>
<td>-57</td>
</tr>
<tr>
<td>Lysine</td>
<td>Lys</td>
<td>K</td>
<td>128</td>
<td>84</td>
<td>-72</td>
</tr>
<tr>
<td>Methionine</td>
<td>Met</td>
<td>M</td>
<td>131</td>
<td>104</td>
<td>-75</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Phe</td>
<td>F</td>
<td>147</td>
<td>102</td>
<td>-91</td>
</tr>
<tr>
<td>Proline</td>
<td>Pro</td>
<td>P</td>
<td>97</td>
<td>70</td>
<td>-41</td>
</tr>
<tr>
<td>Serine</td>
<td>Ser</td>
<td>S</td>
<td>87</td>
<td>60</td>
<td>-31</td>
</tr>
<tr>
<td>Threonine</td>
<td>Thr</td>
<td>T</td>
<td>101</td>
<td>74</td>
<td>-45</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Trp</td>
<td>W</td>
<td>186</td>
<td>204</td>
<td>-130</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Tyr</td>
<td>Y</td>
<td>163</td>
<td>107</td>
<td>-107</td>
</tr>
<tr>
<td>Valine</td>
<td>Val</td>
<td>V</td>
<td>99</td>
<td>72</td>
<td>-43</td>
</tr>
</tbody>
</table>

\(^a\)Residue structure NHCH(R)CO where R is the amino acid side chain.
\(^b\)Mass of H\(_2\)N'\(^+=CHR\) where R is the amino acid side chain.
\(^c\)The loss in mass from the precursor ion due to side chain cleavage between the \(\beta-\gamma\) carbons. (The resultant high mass ions are observed only in high energy CID mass spectra).
Section IX

Biographical note

Vladimir V. Papov, Jr. was born in New York City on January 25, 1965, the only child of Professor Vladimir V. Papov, Sr. and Mrs. Katherine Papov. He matriculated at Connecticut College in September 1982 where he received a B.A. in Chemistry *Cum Laude* in May 1987. In the fall of 1985, he withdrew for one year to work first as an analytical chemist for Pfizer, Inc. (Groton, CT) and then enrolled at the Université de Grenoble, France, the following spring. Upon graduation from Connecticut College, he returned to Pfizer to where he received his first exposure to mass spectrometry under Dr. Justin Stroh. He subsequently enrolled in the Department of Chemistry at the Massachusetts Institute of Technology as a Doctoral Candidate and worked under the supervision of Professor Klaus Biemann. Shortly after entering MIT he was married to Carol Papov (née Simpson), M.D.
Publications


Abstracts


Section X

Acknowledgements

I would like to express my sincere gratitude to Professor Biemann for his support and careful guidance in my research and for the opportunity to work in his laboratory. I particularly appreciate his selflessness and willingness to devote many hours of his time to helpful discussions despite numerous other commitments. I am also grateful for the financial support provided by the National Institutes of Health (grants RR0317 and GM05472 to Klaus Biemann).

I would also like to thank all the members of Professor Biemann's laboratory whose scientific competence and encouragement were most helpful in completing this thesis. I owe a great debt to Dr. Joe Zaia for his help especially at the beginning of my career as a graduate student at MIT and whose experience on HPLC systems, mass spectrometry, and a variety of other matters scientific or otherwise have been invaluable to me. I thank Drs. Ioannis A. Papayannopoulos and James A. Hill for providing advice and practical information particularly on matters concerning the JEOL tandem mass spectrometer. In addition, the willingness and ability of Mr. Heinrich J. Köchling to offer assistance and advice, his wide knowledge, and his friendship are greatly appreciated.

I especially would like to thank Ms. Kathleen F. O'Brien for her assistance on a
variety of administrative matters and for never being too busy to listen. Her kind words, cheerfulness, and enthusiasm will not be forgotten. Thank you also to Dr. Catherine E. Costello for the many helpful discussions concerning some of the thornier problems relating to my research and also for her faith in my abilities. In addition, I would like to acknowledge the help and expertise of Dr. James E. Biller who wrote the data acquisition software for the VT 2000 mass spectrometer and ensured that all of the laboratory computers functioned smoothly.

Finally, I could not have accomplished this work without the support of my parents and my wife Carol.