

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**FURTHER DEVELOPMENT, APPLICATIONS, AND DATA PROCESSING
ALGORITHMS FOR DISULFIDE MASS MAPPING OF PROTEINS BASED ON
PARTIAL REDUCTION, CYANYLATION, AND CN-INDUCED CLEAVAGE**

By

Jianfeng Qi

A DISSERTATION

Submitted to

Michigan State University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

2001

UMI Number: 3036736

UMI[®]

UMI Microform 3036736

Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

FURTHER DEVELOPMENT, APPLICATIONS, AND DATA PROCESSING ALGORITHMS FOR DISULFIDE MASS MAPPING OF PROTEINS BASED ON PARTIAL REDUCTION, CYANYLATION, AND CN-INDUCED CLEAVAGE

By

Jianfeng Qi

The location of the free cysteines in a protein can be determined by cyanylation/CN-induced cleavage chemistry with subsequent analysis by mass spectrometry. The free cysteines can be cyanylated and the protein backbone can be cleaved in aqueous ammonia at the N-terminal side of the cyanylated cysteines. The mass spectral data of the resulting mixture of cleavage fragments can be mass mapped to the sequence of the protein by comparing the experimental mass of the cleavage fragments with expected mass. The location of the free cysteines can then be deduced.

Partial reduction of a multi-disulfide protein with tris-carboxyethyl-phosphine (TCEP) hydrochloride under kinetically limiting conditions can produce a mixture of singly reduced isoforms of the protein. The location of the two nascent free cysteines in a singly reduced isoform, resulting from reduction of a disulfide bond, can be determined by mass mapping the cleavage fragments; thus, the connectivity of the disulfide bonds can be deduced. Various aspects of this methodology were further investigated. Utilization of doubly reduced isoforms offers flexibility in the chemistry of partial reduction and provides complementary data to the previous partial reduction protocol, which utilizes only singly reduced isoforms. An additional analysis step by mass

spectrometry after the CN-induced cleavage reaction of partially reduced/cyanylated protein isoforms, but before complete reduction of the residual disulfide bonds, may provide important diagnostic information. When CN-induced cleavage reactions are conducted in aqueous ammonia, the resulting CN-induced fragments may have an amide or a free acid C-terminus; thus, the mass of each form must be computed for use in data interpretation. In cases of isomass CN-induced fragments, sequence analysis of such fragments by low-energy collision induced dissociation (CID) -tandem mass spectrometry (MS/MS) can be used to distinguish the isomeric species.

The cysteine-rich and apparently highly knotted anti-microbial peptide sillucin, which is resistant to the proteolytic mapping approach to disulfide mapping, is amenable to partial reduction/cyanylation/CN-induced cleavage mass mapping methodology. One of four possible singly reduced isoforms and one of six possible doubly reduced isoforms of sillucin generated sufficient information from mass mapping of the cyanylation/CN-induced cleavage products to assign the disulfide connectivity (Cys2-Cys7, Cys12-Cys24, Cys13-Cys30, and Cys14-Cys21) and rule out the other 104 isomeric possibilities.

The cysteine status of the chloroplast glyceraldehyde-3-P dehydrogenase subunits was determined by cyanylation, CN-induced cleavage, and mass mapping.

Algorithms for data processing for the disulfide mass mapping based on cyanylation and CN-induced cleavages were proposed. The concept of using negative signature masses (masses that can be used to reject certain disulfide linkages if present) facilitates the disulfide assignment process and, in some cases, allows identification of the disulfide structure even when an incomplete data set is obtained.

To my wife, Tong Pei, for her love, support, and patience that makes my life enjoyable

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my mentor, Dr. J. Throck Watson, for his guidance, encouragement, and support throughout my graduate study at Michigan State University. I would also like to thank Dr. John Allison for his continual encouragement, support and serving as a second reader of my dissertation; Dr. Eric Torng and Mr. Dehua Hang for their guidance and advice on development of the computer algorithms; Dr. Marcos Dantus for serving on my committee.

I especially thank my wife, Tong Pei, for her love, support, and patience that make my life enjoyable. I am also very grateful to my mom, parents-in-law, brother, and brother-in-law, for their caring, encouragement, and understanding during my graduate study at Michigan State University.

To all Watson group members, Mr. Wei Wu, Mr. Yingda Xu, Dr. Ying Yang, Ms. Susan Li, and Mr. Jose-Luis Gallegos-Perez, for your friendship and collaboration.

TABLE OF CONTENTS

LIST OF TABLES.....	xi
LIST OF FIGURES.....	xiii
ABBREVIATIONS.....	xxii
CHAPTER 1	
INTRODUCTION: DISULFIDE MAPPING OF PROTEINS BY MASS SPECTROMETRY	1
I. Introduction	1
II. Disulfide Mass Mapping Based on Proteolysis.....	4
1. Overview of the Method	4
2. Constraints/Caveats.....	7
3. An Experimental Protocol.....	8
4. A Representative Application.....	9
5. Advanced Topics: Using MALDI-PSD or ESI-MS/MS to Obtain Additional Information on Disulfide-Linked Peptides.....	18
6. Conclusions.....	22
III. Determination of Disulfide Bonds by Partial Reduction, Chemical Tagging, Edman Sequencing Analysis and Mass Spectrometry	23
1. Overview of the Method.....	23
2. Constraints/Caveats.....	25
3. An Experimental Protocol.....	26
4. A Representative Application.....	27
5. Conclusions.....	29

IV. Disulfide Mass Mapping Based on Partial Reduction, Cyanylation, and CN-induced Cleavage.....	29
1. Overview of Method.....	29
2. Constraints, Caveats, and Other Aspects.....	33
3. Experimental Protocols.....	35
A. A Protocol for Determining the Location of Free Cysteines by Alkylation/Cyanylation /CN-Induced Cleavage.....	35
B. A Protocol for Determining the Location of Free Cysteines by Cyanylation/CN-Induced Cleavage.....	35
C. A Protocol for Determining the Location of Disulfide Bonds by Partial Reduction/ Cyanylation/CN-Induced Cleavage.....	36
4. A Representative Application: Disulfide Mass Mapping Using Singly Reduced/Cyanylated Isoforms	37
5. Advanced topic: Combining Partial reduction/Cyanylation/CN-induced Cleavage with Proteolytic Mass Mapping.....	41
6. Conclusions.....	44
V. References.....	45
CHAPTER 2	
DEVELOPMENT OF MORE FLEXIBLE PROTOCOLS FOR THE PARTIAL REDUCTION AND SUBSEQUENT CYANYLATION/CN-INDUCED CLEAVAGE/MASS MAPPING METHODOLOGY.....	50
I. Introduction	50
II. Materials and Methods	53
III. Expand Partial Reduction Strategy to Include Doubly Reduced Protein Isoforms	57
IV. Addition of an Analysis Step by Mass Spectrometry after CN-induced Cleavage, but Before Complete Reduction of the Residual Disulfide Bonds	66
V. Further Investigation of the Status of C-terminus (Amide or Free Acid) of CN-induced Cleavage Fragments.....	71

VI. Sequence Analysis of Iminothiazolidine-Blocked Peptides by CID-MS/MS	78
VII. Conclusions.....	81
VIII. References	82
 CHAPTER 3	
DETERMINATION OF THE DISULFIDE STRUCTURE OF SILLUCIN, A HIGHLY KNOTTED, CYSTEINE-RICH PEPTIDE, BY CYANYLATION/CN-INDUCED CLEAVAGE MASS MAPPING	84
I. Introduction.....	84
II. Materials and Methods.....	90
III. Results.....	93
1. Accurate Mass Measurement of Intact Sillucin.....	93
2. Partially Reduced/Cyanylated Species of Sillucin.....	94
3. A Single Reduced/Cyanylated Species.....	96
4. A Doubly Reduced/Cyanylated Species	106
IV. Discussions	111
1. Partial Reduction	111
2. Incomplete Cyanylation, Incomplete CN-induced cleavage, and Disulfide Scrambling.....	113
V. Conclusion	122
VI. References.....	122
 CHAPTER 4	
DETERMINATION OF THE CYSTEINE STATUS OF THE CHLOROPLAST GLYCERALDEHYDE-3-P DEHYDROGENASE SUBUNITS	125
I. Introduction.....	125
II. Materials and Methods.....	131
III. Results.....	134

1. Mass Mapping Based on Cyanylation and CN-induced Cleavage	134
2. Subunit Composition.....	149
IV. Discussions.....	149
V. References.....	151
 CHAPTER 5	
COMPUTER ALGORITHMS FOR DISULFIDE MASS MAPPING OF PROTEINES BASED ON PARTIAL REDUCTION AND CYANYLATION/ CN-INDUCED CLEAVAGE.....	154
I. Introduction.....	154
II. A Naive Algorithm.....	156
III. A Basic Algorithm.....	158
1. Consideration of Singly Reduced Isoforms.....	162
2. Consideration of Doubly Reduced Isoforms.....	167
3. Consideration of Triply Reduced Isoforms.....	170
IV. Negative Signature Masses and an Improved Algorithm.....	171
1. Negative Signature Masses.....	171
2. An Algorithm Based on Negative Signature Masses.....	173
V. Dealing with Imperfect Data.....	180
VI. Implementation and Testing of the Algorithms.....	181
1. Implementation of the Basic Algorithm	181
A. C++ Program Files	181
B. Preliminary Testing using Data from a Model Protein: ribonuclease-A	182
C. The Content of Input and Output Text Files.....	183

2. Implementation of the Algorithm Based on Negative Signature Masses	189
A. C++ Program Files	189
B. Preliminary Testing using Data from a Model Protein: ribonuclease-A	189
C. The Content of Input and Output Text Files	190
VII. Conclusions	197
VIII. References.....	198

LIST OF TABLES

Table 1.1. Calculated m/z values of protonated CN-induced cleavage fragments from all possible singly reduced/cyanylated isoforms of an unknown protein having the same sequence as RnaseA. The boldface entries correspond to native RnaseA, one of the 105 possible isomeric arrangements of the four cystines in a protein with this sequence.	39
Table 1.2. m/z values of CN-cleavage products of two singly reduced/cyanylated isoforms GM2AP(110-142) after reduction of the residual disulfide bond (adopted from reference 17).	44
Table 2.1. m/z values for possible fragments resulting from the cleavage reaction of one of the doubly reduced/cyanylated isoform of ribonuclease-A represented by HPLC peak B in Figure 2.3.	64
Table 2.2. Some itz-fragments studied by ESI-CID-MS/MS on an ion trap.	79
Table 3.1. Calculated m/z values for MH^+ (based on average masses of amino acid residues) of possible fragments resulting from CN-induced cleavage of singly reduced/cyanylated isoforms of all disulfide structural isomers of sillucin.	97
Table 3.2. Calculated m/z values (monoisotopic) for b and y ions of CN-induced fragments itz-13-29-NH ₂ and itz-14-30-NH ₂ from sillucin.	104
Table 3.3. Calculated m/z values (monoisotopic) for b and y ions of CN-induced fragments itz-13-29-NH ₂ and itz-14-30-OH from sillucin.	105
Table 3.4. Identities of the MALDI-MS peaks in Figure 3.5.	115
Table 3.5. Identities of the MALDI-MS peaks in Figure 3.9.	116
Table 3.6. Identities of the MALDI-MS peaks in Figure 3.11.	117
Table 4.1. The experimental and calculated m/z values for the protonated CN-induced cleavage fragments of the cyanylated GAPD subunit A.	139
Table 4.2. The experimental and calculated m/z values for the protonated CN-induced cleavage fragments of the cyanylated GAPD subunit B-oxidized	141

Table 4.3. The experimental and calculated m/z values for the protonated CN-induced cleavage fragments of the cyanylated GAPD subunit B-reduced.	143
Table 4.4. The experimental and calculated m/z values for the fragments from β -elimination and incomplete CN-induced cleavage of subunit A of GAPD.	146
Table 4.5. The experimental and calculated m/z values for the fragments from β -elimination and incomplete CN-induced cleavage of subunit B of GAPD.	147
Table 5.1. Some of the same CN-induced fragments from two singly reduced isoforms of two hypothetical protein isomers with different disulfide structures. The hypothetical protein has 90 amino acid residues, among which 8 cysteines located at positions 10, 20, 30, 40, 50, 60, 70, and 80.	159
Table 5.2. The number of possible disulfide bonds and disulfide structures for a protein having n disulfide bonds.	160
Table 5.3. The CN-induced fragments from all the possible singly reduced/cyanylated isoforms of ribonuclease-A.	164
Table 5.4. Filtering the List B for ribonuclease-A using three negative masses	176
Table 5.5. The shortened disulfide bond List C for ribonuclease-A after filtering using negative masses.	177
Table 5.6. Forming the disulfide structure list D from the shortened disulfide list C (Table 5.5) for ribonuclease-A.	179

LIST OF FIGURES

- Figure 1.1. Different congeners of a hypothetical protein with 4 cysteines: 1 completely reduced form, 6 one-disulfide forms, and 3 two-disulfide forms. The number of congeners increases geometrically with the number of cysteines.2
- Figure 1.2. Conceptual representation of mass changes observed upon reduction of peptides containing intermolecular vs. intramolecular disulfide linkages. “ m/z ” designates mass to charge ratio.3
- Figure 1.3. Conceptual illustration of the different array of fragments that can be used to distinguish the three possible isomeric arrangements of two disulfide bonds in a hypothetical protein containing four cysteines by the general proteolytic approach for disulfide mass mapping. Note that fragments C, C', and C'' have the same mass in each case because they contain no cysteine, whereas the mass of the fragments containing cysteine vary substantially for the different isomeric disulfide arrangements.6
- Figure 1.4. Overall sequence of tACE showing tryptic cleavage sites at K491 and K511; these define the origin of tryptic fragment Tyr492-Lys511 at the lower left that shifts from 2145.4 Da to 2453.7 Da upon alkylation of the free sulfhydryl on Cys496. The mass of the protonated alkylated-peptide (2453.7 Da) is very close to the observed peak at m/z 2453.4 in Figure 1.6.10
- Figure 1.5. HPLC chromatograms of the endoproteinase Lys-C digestion mixture of AEDANS-labeled tACE as recorded by two different detectors. The upper panel was monitored at 340 nm for the fluorescence of the AEDANS fluorophore to selectively detect the peptide with the AEDANS-modified cysteine; the lower panel was monitored at 214 nm to detect all the digestion peptides (adopted from reference 39).12
- Figure 1.6. The MALDI mass spectrum of the HPLC fraction represented by the single peak in the top panel of Figure 1.5. The peak at m/z at 2453.4 represents the protonated proteolytic fragment containing the modified cysteine as illustrated in Figure 1.4 and described in the text. The other two peaks represent other proteolytic peptides that coelute with the alkylated one (adopted from 39).13

Figure 1.7. Conceptual representations of disulfide arrangement among the six cysteines involved in the three cysteine-containing fragments from CNBr digestion of tACE. Category (i) shows the simplest arrangement, which derives from three isolated disulfides in tACE; experimental results indicate that this category represents the disulfide arrangement in native tACE. Categories (ii) and (iii) consist of six and eight isomeric arrangements, respectively, some of which involve overlapping disulfide bonds (indicated, in part for two possibilities, by dashed lines).	15
Figure 1.8. MALDI mass spectrum of the putative peptide of a CNBr fragment (see Figure 1.9) of the tACE protein. Peaks at m/z 856.6 and m/z 4388.0 correspond to the individual peptides connected by the disulfide as cleaved during the MALDI process (adopted from reference 39).	16
Figure 1.9. Conceptual illustration of digestion of tACE by CNBr and trypsin to produce a cysteine-containing fragment, the analysis of which by MALDI-MS identifies the disulfide Cys538-Cys550.	17
Figure 1.10. MALDI-PSD-MS mass spectrum of a disulfide-linked proteolytic fragment from dual digestion of the TNFbp protein with trypsin and thermolysin. The numbers by the structure give the mass of the indicated fragment ions; the peaks marked by an asterisk at m/z 825.6, 940, and 1027 confirm the disulfide linkage pattern shown (adopted from reference 28).	19
Figure 1.11. ESI-CID-MS/MS mass spectrum of a triply charged ion $[M+3H^+]^{3+}$ at m/z 1406.1 corresponding to a disulfide-linked tryptic fragment (in which a disulfide bond links two peptides, S130-K156 and Y171-K181) of bovine beta-1,4-galactosyltransferase (adopted from reference 34).	21
Figure 1.12. A diagrammatic presentation of sequencing results for the partial reduction/alkylated isoforms of AVR9 represented by peaks in the HPLC chromatogram. Peak 2 represents a singly reduced isoform from partial reduction and indicates that AVR9 contains a disulfide bond between Cys2-Cys16. Peak 3 represents a doubly reduced isoform containing a residual disulfide bond between Cys6-Cys19 before complete reduction; peak 4 represents a doubly reduced isoform containing a residual disulfide Cys12-Cys26 before complete reduction. The peaks 5 and 6 represent the completely reduced/NEM-modified species (as determined by mass spectrometry); both species have the same mass and apparently result from epimeric forms of the prochiral alkylating reagent (adopted from reference 44).	28

Figure 1.13. (A) Cyanylation of sulfhydryl group by CDAP and (B) peptide bond cleavage by ammonia. (itz = iminothiazolidine-carboxyl; β -elimination = loss of HSCN from a cyanylated cysteine residue.) (adopted from reference 47)	30
Figure 1.14. Chemical overview of the partial reduction/cyanylation/CN-induced cleavage methodology using singly reduced isoforms of a hypothetical 2-disulfide protein (adopted from reference 16).	32
Figure 1.15. HPLC separation of denatured ribonuclease A and its partially reduced/cyanylated isoforms. The separation was done by reversed-phase HPLC on a Vydac C18 column (5 μ m, 4.6mm $\times$ 250mm) at a flow rate of 1.5 ml/min with a linear gradient 20-40% B in 90 minutes, where A = 0.1% TFA in water and B = 0.1% TFA in 90% CH ₃ CN. The peak marked by IP is the intact protein (ribonuclease A). Peaks 1-4 represent singly-reduced/cyanylated ribonuclease-A isoforms, as determined by MALDI-TOF MS (adopted from reference 16).	38
Figure 1.16. The MALDI mass spectra of four peptide mixtures resulting from the CN-induced cleavage of the singly reduced/cyanylated ribonuclease A isoforms. Spectra (a)- (d) correspond to the HPLC peaks 1-4 in Figure 1., respectively. The symbols # and * represent the doubly charged fragments and protonated β -elimination products, respectively. The double mark #* in (c) shows a doubly charged β -elimination product (adopted from reference 16).	40
Figure 1.17. C18-RP-HPLC separation of partially reduced and cyanylated isoforms of GM2AP(110-142) (adopted from reference 17).	43
Figure 2.1. Overview of chemical reactions involved in the disulfide mass mapping of two doubly reduced/cyanylated isoforms (B1) and (C1) of a hypothetical protein (A). (C2): the cleaved fragments from (C1) before complete reduction; (B3) and (C3): the corresponding cleaved fragments from (B1) and (C1) after complete reduction.	58
Figure 2.2. Disulfide mass mapping using a doubly reduced (B1) and a singly reduced/cyanylated (D1) isoform of a hypothetical protein (A) containing four disulfide bonds. (B3) and (D3): the corresponding cleaved fragments from (B1) and (D1) after complete reduction.	59

Figure 2.3. HPLC chromatogram of partially reduced/cyanylated isoforms of ribonuclease-A: IP represents the intact ribonuclease-A; Peaks 1-4 represent the singly reduced/cyanylated isoforms (as also shown in Figure 1.15 in Chapter 1); Peaks A-F represent the doubly reduced/cyanylated isoforms; Peaks g-j represent the triply reduced/cyanylated isoforms, as determined by MALDI-MS analysis.	61
Figure 2.4. The MALDI mass spectrum of a peptide mixture resulting from the cleavage of a doubly reduced/cyanylated ribonuclease-A isoform corresponding to HPLC peak B in Figure 2.3.	63
Figure 2.5. The MALDI mass spectrum of the peptide mixture resulting from CN-induced cleavage (but before complete reduction) of the doubly reduced/cyanylated ribonuclease-A isoform represented by HPLC peak B in Figure 2.3. Itz-40-64 represents the itz-blocked fragment of ribonuclease-A from residue 40 to 64. Details about prompt fragmentation (or in-source fragmentation) during analysis by MALDI-MS can be found in references 4-8.	68
Figure 2.6. Eliminating a possible disulfide structure (B) by mass mapping analysis of the CN-cleavage products before the complete reduction step. Disulfide structure (A) shows the correct disulfide linkages in the hypothetical protein; structures (B) and (C) are two of 15 possible structures of arranging six cysteines as three disulfide bonds in such a protein.	69
Figure 2.7. The MALDI mass spectrum of the peptide mixture resulting from CN-induced cleavage (but before complete reduction) of the doubly reduced/cyanylated ribonuclease-A isoform represented by HPLC peak A in Figure 2.3. Itz-26-64 represents the itz-blocked fragment of ribonuclease-A from residue 26 to 64.	70
Figure 2.8. A proposed mechanism (9) involving nucleophilic attack by OH ⁻ for the CN-induced cleavage reaction at a cyanylated cysteine residue.	72
Figure 2.9. Production of CN-induced cleavage fragments with an amide C-terminus via a mechanism involving nucleophilic attack by NH ₃ at a cyanylated cysteine residue.	73
Figure 2.10. Isotope peak intensity distribution in expanded ESI mass spectrum of itz-13-29-red-NH ₂ (itz-CCGDSGYWCRQCGIKYT-NH ₂) produced from sillucin. “ M ” represents the molecule and “ MH²⁺ ” represents the doubly protonated molecule.	75

Figure 2.11. Isotope peak intensity distribution in expanded MALDI mass spectrum of itz-12- β @13-23-ox-OH (itz-C(β @C)CGDSGYWCRQ-OH) produced from sillucin.	76
Figure 2.12. Isotope peak intensity distributions for two forms of the CN-induced fragment, 1-12-red (ACLPNSCVSKGC; 1-12-red-NH ₂ and 1-12-red-OH), produced from sillucin, were detected during ESI infusion on an ion trap. “0.3×” in the middle panel indicates that the peak intensity is 30% of the intensity for the corresponding peaks in the bottom panel, respectively.	77
Figure 2.13. ESI-CID-MS/MS of itz-CEGNPYVPVHFDASV (residue 110-124 of ribonuclease-A), a doubly charged ion (m/z 830.4), derived from cyanylation/CN-induced cleavage of a singly reduced ribonuclease-A isoform, obtained on an ion trap MS. The nomenclature of Biemann (13) for b and y ions is used throughout, although for convenience, lower case letters are also used to indicate the peptide bond that is cleaved to produced key fragments (b ₇ and y ₉) in a manner analogous to that introduced by Roepstorff <i>et al.</i> (14) using capital letters.	80
Figure 3.1. Overview of chemical reactions involved in the disulfide mass mapping of a singly reduced/cyanylated species of sillucin. Recognition of the cleaved fragments by mass spectrometry allows the linkage of Cys13-Cys30 to be deduced. itz represents the 2-iminothiazolidine-4-carboxyl blocked N-terminus resulting from CN-induced cleavage by aqueous ammonia on the N-terminal side of the cyanylated cysteines. * The structure in the parenthesis designates a β -elimination product, which results in this case from loss of HSCN at cyanylated Cys 13 (β -elimination) and CN-induced cleavage at cyanylated Cys30.	88
Figure 3.2. Overview of chemical reactions involved in the disulfide mass mapping of a doubly reduced/cyanylated isoform (C1) of sillucin (A) as well as a singly reduced/cyanylated isoform (B1). Recognition of the cleaved fragments (C2) from C1 before complete reduction allows the connectivity of Cys14-Cys21 to be deduced. Recognition of the cleaved fragments (B3) from B2 (shown in Figure 3.1) after complete reduction combined with the determination of Cys14-Cys21 as described above allows the connectivity of Cys12-Cys24 to be deduced; Z could be 2, 7, or 30.	89
Figure 3.3. The MALDI of spectrum of intact sillucin.	94

Figure 3.4. HPLC chromatogram of denatured sillucin (IP) and some of its partially reduced/cyanylated species. Separation was carried out on a Vydac C18 column at a flow rate of 1.0 mL/min with a linear gradient of 15% to 40% B in 50 minutes, where A = 0.1% TFA in water and B = 90% (v/v) acetonitrile/0.1% TFA. Peaks IP, 1, and 2 represent the intact peptide, a singly reduced/cyanylated species, and a doubly reduced/cyanylated species, respectively; peak 3 represents the completely reduced/cyanylated species, as determined from analysis by MALDI-MS.	95
Figure 3.5. The MALDI-MS spectrum of the peptide mixture resulting from CN-induced cleavage and subsequent complete reduction of the singly reduced/cyanylated sillucin species represented by HPLC peak 1 in Figure 3.4. "IS" designates an internal standard.	98
Figure 3.6. Assignment of the disulfide Cys13-Cys30 in sillucin.	99
Figure 3.7. Detection of CN-induced fragment 1-12-OH (two forms: 1-12-ox-OH and 1-12-red-OH) by MALDI-MS after it was separated from itz-12-29-red-NH2 and 1-β@13-29-red-NH2.	102
Figure 3.8. ESI CID-MS/MS spectrum of CN-induced fragment itz-13-29-NH2 from sillucin. The precursor is a doubly protonated molecule; $[MH_2-H_2O]^{2+}$ designates a dehydrated and doubly protonated fragment of the precursor; "B ₁₃ " indicates the peptide bond that is cleaved in the peptide backbone between residue G and I (see definition of the nomenclature in reference 21) to form an amino terminal ion represented by "b ₁₃ " (singly charged) and b ₁₃ ²⁺ (doubly charged); the definition of a, b, and y ions can be found in reference 22.	103
Figure 3.9. The MALDI-MS spectrum of the peptide mixture resulting from CN-induced cleavage (before complete reduction) of the doubly reduced/cyanylated sillucin species represented by HPLC peak 2 in Figure 3.4.	107
Figure 3.10. Assignment of disulfide bonds Cys14-Cys21, Cys12-Cys24, and Cys2-Cys7.	108
Figure 3.11. The MALDI-MS spectrum of the peptide mixture resulting from CN-induced cleavage and subsequent complete reduction of the doubly reduced/cyanylated sillucin species represented by HPLC peak 2 in Figure 3.4.	110

Figure 3.12. Formation of itz-13-SCN@24-29-ox-NH ₂ and itz-13-SCN@24-29-red-NH ₂ because of no CN-induced cleavage at SCN@24.	118
Figure 3.13. Formation of itz-13-30-red-OH due to incomplete cyanylation at Cys 30; Cys30SH is used to designate the free cysteine residue.	120
Figure 3.14. Didactic illustration of possible scrambling in a singly reduced isoform, [sillucin(13SH-30SH)], illustrated for the case in which the disulfide Cys13-Cys30 has been reduced, and a trace amount has dissociated to form the thiolate ion, [sillucin(13S ⁻ -30SH)], which scrambles to [sillucin(12S ⁻ -30SH)] before cyanylation is accomplished.	121
Figure 4.1. Sequence alignment of the pea chloroplast glyceraldehyde-3-P dehydrogenase A (gi ["gi" is the National Center for Biotechnology Information protein identification number] 66025) and B (gi120663) subunits (pea a and pea b), the <i>Sulfolobus solfataricus</i> glyceraldehyde-3-P dehydrogenase (Protein Data Bank (PDB) file 1b7g), and the <i>Bacillus stearothermophilus</i> glyceraldehyde-3-P dehydrogenase (PDB file 1gd1). Identical residues in the B subunit (pea b) and in one or both of the sequences from the Protein Data Bank are bolded. Residues that are superimposable in the two crystal structures are underlined. Cysteines in pea a and pea b are numbered, above the residue for pea a, below for pea b. Dashes indicate gaps introduced to optimize the alignment.	126
Figure 4.2. Conventional methodology for determining the status of the cysteines in a hypothetical protein of 70 amino acid residues. The straight line designates the protein backbone; "SH" designates a free cysteine; "S—S" designates a disulfide bonds; "SR" designates an alkylated cysteine residue; and "E" together with an arrow indicates where the enzyme cleaves.	128
Figure 4.3. Cyanylation and CN-induced cleavage process to identify free cysteines in a hypothetical protein of 70 amino acid residues. The straight line designates the protein backbone; "SH" designates a free cysteine; "S—S" designates a disulfide bond; "SCN" designates a cyanylated cysteine residue; and "itz" designates an iminothiazolidine-carboxyl blocked amino terminus.	130
Figure 4.4. MALDI mass spectrum of the chloroplast glyceraldehyde-3-P dehydrogenase. 1µl of a solution of 9 pmol enzyme subunits in 12 mM pH 7-ammonium acetate buffer was used.	135

Figure 4.5. HPLC separation of the cyanylated A and B subunits of chloroplast glyceraldehyde-3-P dehydrogenase (GAPD). Cyanylated proteins (approximately 3.8 nmol total of the A and B subunits) were separated by reverse phase HPLC on a Vydac C18 column. UV detection was done at 215 nm. For details see 'Methods'. Peaks 1 and 2 represent an 8.3k-Da protein; peaks 4 and 5 represent the cyanylated B and A subunits of GAPD, respectively, as determined by MALDI-TOF-MS. Nothing was detected from HPLC fraction 3 by MALDI-TOF-MS. The arrows indicate the starting and stopping points for collecting fractions for peaks 4 and 5, respectively.136

Figure 4.6. The MALDI spectrum of the CN-induced cleavage fragments of the cyanylated A subunit of chloroplast glyceraldehydes-3-P dehydrogenase represented by HPLC peak 5 in Figure 4.5. Portions 1 and 2 are the low-mass and high-mass portions, respectively. The mark "B" in parenthesis indicates that the peak represents a CN-induced cleavage fragment of the cyanylated B subunit because HPLC peaks 4 and 5 (Figure 4.5) were not completely resolved. The mark "++" designates a peak representing a doubly protonated (also doubly charged) protein fragment. All other peaks without such marks are singly protonated protein fragments. The mark "?" indicates that the source of the peak is not identified.138

Figure 4.7. The MALDI spectrum of the CN-induced cleavage fragments of the cyanylated B subunit of chloroplast glyceraldehyde-3-P dehydrogenase represented by HPLC peak 4 in Figure 4.5. Portions 1 and 2 are the low-mass and high-mass portions, respectively. The mark "A" in parenthesis indicates that the peak represents a CN-induced cleavage fragment of the cyanylated A subunit because HPLC peaks 4 and 5 (Figure 4.5) were not completely resolved. The mark "++" designates a peak representing a doubly protonated (also doubly charged) protein fragment. All other peaks without such marks represent singly protonated protein fragments. The mark "?" indicates that the source of the peak is not identified. The peak at m/z 6962.1 represents both a singly protonated CN-induced cleavage fragment (itz-291-353, marked as "+, B_r"; calculated m/z for MH^+ : 6960.8) from the cyanylated subunit B-reduced and a doubly protonated CN-induced cleavage/ β -elimination fragment (itz-159- β @280-290) [Table 4.2] formed by β -elimination at cyanylated Cys280 instead of cleavage (marked as "++ , B"; calculated m/z for MH_2^{++} : 6962.6).140

Figure 4.8. The CN-induced cleavage fragments from the oxidized and reduced forms of the cyanylated subunit B of GAPD. The straight line designates the protein backbone; "S—S" designates a disulfide bond; "SCN" designates a cyanylated cysteine residue; and "itz" designates iminothiazolidine-carboxyl blocked amino terminus.144

Figure 4.9. The formation of some fragments resulted from β -elimination, incomplete CN-induced cleavage, and incomplete cyanylation of subunit A of GAPD. The straight line designates the protein backbone; “SCN” designates a cyanylated cysteine residue; “SH” designates a free cysteine residue; “ β ” designates β -elimination (loss of HSCN from a cyanylated cysteine residue); and “itz” designates iminothiazolidine-carboxyl blocked amino terminus.	148
Figure 5.1. Information flow for a naïve algorithm for disulfide mass mapping based on partial reduction, cyanylation and CN-induced cleavage. The sequence of the protein to be analyzed and the locations of free cysteines are used as input 1; MS spectra of the CN-induced fragments and m/z measurement accuracy are used as input 2. The output is a ranked disulfide structure list, in which candidate disulfide structures are ranked according to the number of CN-induced fragments matched (Experimental m/z values and calculated m/z values are compared.).	157
Figure 5.2. A basic algorithm for disulfide mass mapping based on partial reduction, cyanylation and CN-induced cleavage. The sequence of the protein to be analyzed and the locations of free cysteines are used as input 1; MS spectra of the CN-induced fragments and m/z measurement accuracy are used as input 2. The output is a ranked partially reduced isoform list, in which candidate partially reduced isoforms are ranked according to matching percentage = (the number of CN-induced fragments matched/the number of such fragments expected for a partially reduced isoform).	161
Figure 5.3. Using a negative signature mass (m/z 6546.4) from ribonuclease-A to reject the existence of 8 disulfide bonds.	172
Figure 5.4. An algorithm based on negative signature masses.	174

ABBREVIATIONS

2-D NMR	two-dimensional nuclear magnetic resonance
4-VP	4-vinyl pyridine
AEDANS	5-([2-(acetyl)amino]ethylamino) naphthalene-1-sulfonic acid group
Ala	alanine residue
Arg	arginine residue
Asn	asparagine residue
Asp	aspartic acid residue
β	β -elimination of HSCN from cyanylated cysteine residue
C	cysteine residue
CDAP	1-cyano-4-dimethylamino-pyridinium
CID	collision induced dissociation
CNBr	cyanogen bromide
Cys	cysteine residue
Cys i - Cys j	the disulfide bond between Cys i and Cys j
D	aspartic residue
Da	dalton
DEAE	diethylaminoethyl
E	glutamic acid residue
ESI	electrospray
F	phenylalanine residue
FAB	fast atom bombardment

FTICR	fourier transform ion cyclotron resonance
G	glycine residue
GAPD	glyceraldehydes-3-P dehydrogenase
gi	National Center for Biotechnology Information protein identification number
Glu	glutamic acid residue
Gln	glutamine residue
GM2AP	GM2-activator protein
H	histidine residue
His	histidine residue
HPLC	high performance liquid chromatography
I	isoleucine residue
IAEDANS	5-([2-(iodoacetyl)amino]ethylamino) naphthalene-1-sulfonic acid
IGFs	insulin-like growth factors
Ile	isoleucine residue
IS	internal standard
itz	iminothiazolidine-4carboxyl terminus, resulting from the CN-induced cleavage by aqueous ammonia on the N-terminal side of the cyanylated cysteine residues
k	kilo
K	lysine residue
L	leucine residue
LC	liquid chromatography
Leu	leucine residue

LR ³ IGF-1	long insulin-like growth factor-1
Lys	lysine residue
M	molecule or mass of a molecule
M	methionine residue
MALDI	matrix assisted laser desorption/ionization
Me	methyl
Met	methionine residue
MH ⁺	protonated molecule
MS	mass spectrometry, mass spectral or mass spectrometer
MH ₂ ²⁺	doubly protonated molecule
MS/MS	mass spectrometry/mass spectrometry or tandem mass spectrometry
<i>m/z</i>	mass to charge ratio
N	asparagine residue
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NEM	N-ethylmaleimide
NH ₂	designates an amide C-terminus
OH	designates a free acid C-terminus
ox	designates that there are residual disulfide bonds within a CN-induced cleavage fragment
P	proline residue
P	protease
PDB	Protein Data Bank

PEG	polyethylene glycol
Phe	phenylalanine residue
Pro	proline residue
PSD	post source decay
PTH	phenylthiohydantoin
Q	glutamine residue
R	arginine residue
R	resolution
red	indicates that the residual disulfide bonds, which survived the partial reduction, have been reduced during the complete reduction step; thus, all the cysteine residues within the fragment are free cysteines
RnaseA	ribonuclease A
rp	reversed phase
S	serine residue
Ser	serine residue
T	threonine residue
tACE	testis angiotensin-converting enzyme
TCEP	tris(2-carboxyethyl)phosphine
Thr	threonine residue
TFA	trifluoroacetic acid
TNFbp	tumor necrosis factor binding protein
TOF	time of flight
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
Trp	tryptophan residue

Tyr	tyrosine residue
UV	ultra violet
V	valine residue
Val	valine residue
W	tryptophan residue
Y	tyrosine residue

CHAPTER 1

INTRODUCTION:

DISULFIDE MAPPING OF PROTEINS BY MASS SPECTROMETRY

I. Introduction

The molecular weight and amino acid sequence are available at an early stage in characterizing a protein, while details of posttranslational modification usually require more time and effort to elucidate. Determination of the cysteine status in a protein is an important part of its structural characterization. Concerning cysteine status, there are usually two simple questions: Are there any disulfide bonds? If there are, what is the connectivity of the cysteines in each disulfide bond? A sense of the complexity of possible disulfide bonding structures that must be considered for a given number of cysteines can be gleaned from Figure 1.1.

The question of whether there are disulfide bonds is relatively easy to answer. The difference in mass between a protein containing two **free cysteines** ("Free cysteines" are cysteines that have free sulfhydryls and are not involved in disulfide bond formation.) and that containing two **disulfide cysteines** ("Disulfide cysteines" are cysteines that are involved in disulfide bonds.) forming a disulfide bond is two Daltons. However, confidence in the experimental mass measurement will depend on the molecular weight of the protein as well as the resolution and mass accuracy of the mass spectrometer. For example, for a protein in excess of 10 kDa and a mass spectrometer with a resolving power of 10,000 and a mass accuracy no better than 1 in 10^4 , the analyst may wish to chemically modify the protein in a way that will exaggerate the mass difference between

the oxidized and reduced state. Chemical reactions with sulfhydryls ranging from alkylation with iodoacetamide to 4-hydroxymercuri-benzoic acid (1) would shift the mass difference between a cystine (disulfide) and two free or modified cysteines from 2 Da to 116.1 Da up to 643.4 Da, respectively.

Possible Redox States in A Protein Containing Four Cysteines

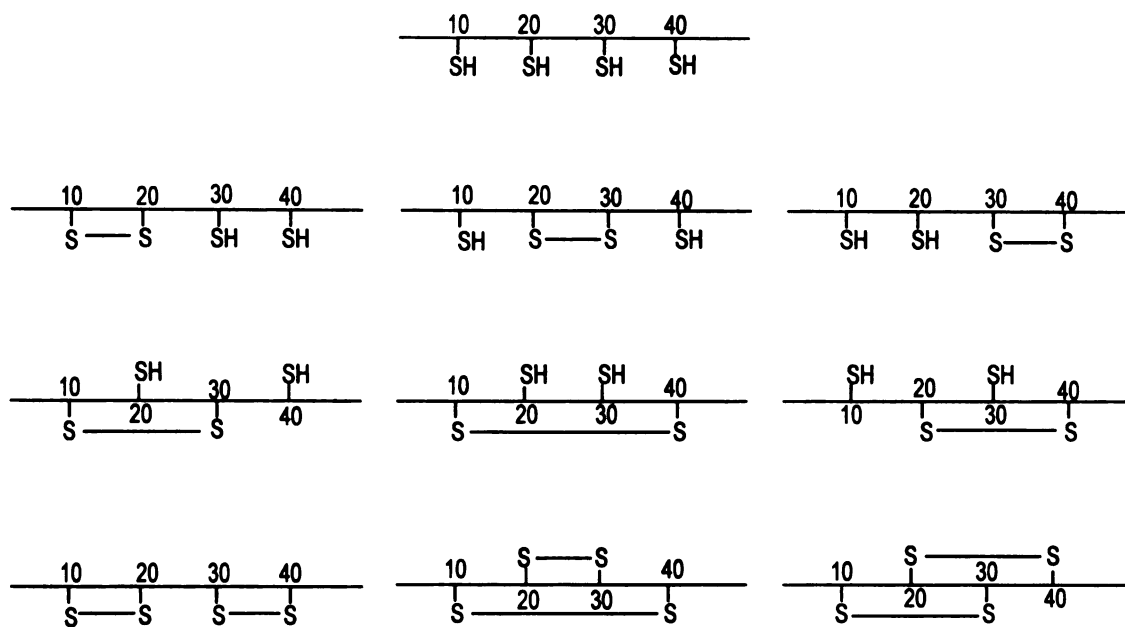


Figure 1.1. Different congeners of a hypothetical protein with 4 cysteines: 1 completely reduced form, 6 one-disulfide forms, and 3 two-disulfide forms. The number of congeners increases geometrically with the number of cysteines. The straight line designates the protein backbone; “S—S” designates a disulfide bond; “SH” designates a free cysteine.

Disulfide bonds come in two forms, intramolecular and intermolecular disulfide bonds. Recognition of intermolecular disulfide bonding can be accomplished by mass analysis of the protein before and after treatment with an excess of a reducing agent. An increase in the mass of the protein by only a few daltons (as each cystine acquires two

hydrogens upon reduction) will indicate intramolecular disulfides exclusively. On the other hand, a dramatic shift (by hundreds to thousands of daltons) to lower mass for the reduced species from the original protein indicates that at least one intermolecular disulfide bond is present (see Figure 1.2, with further explanation in the next section).

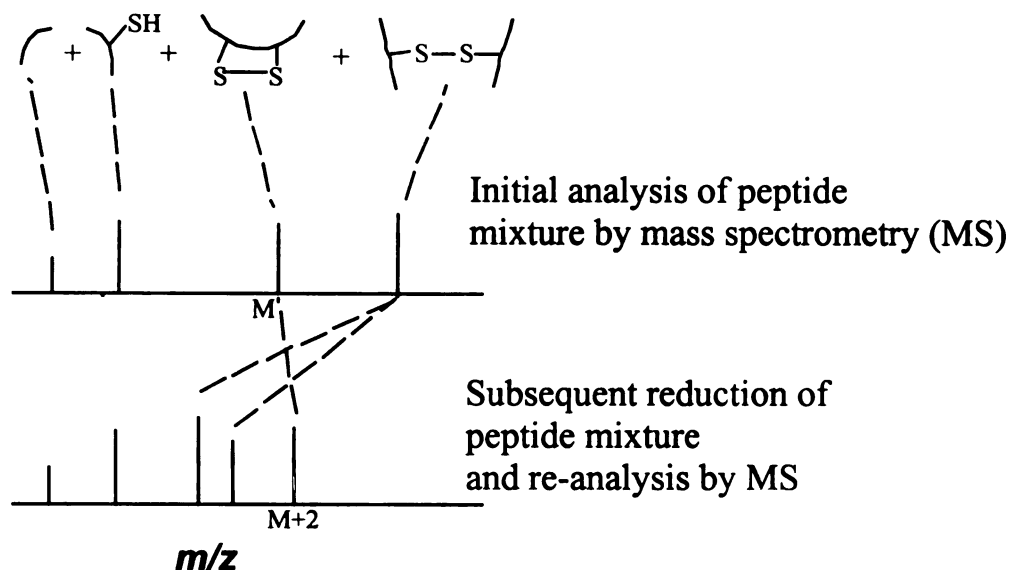


Figure 1.2. Conceptual representation of mass changes observed upon reduction of peptides containing intermolecular vs. intramolecular disulfide bonds. “ m/z ” designates mass to charge ratio.

Here, we will examine different analytical strategies and describe some representative experimental protocols for determining the connectivity of various disulfide bonds. The emphasis in this chapter will be on controlled modification and degradation of the original protein with intervening analyses by mass spectrometry to map the degradation products according to mass shift from that of the primary structure. Three major strategies for controlled degradation of the protein will be highlighted: mass mapping based on proteolysis (2-6), partial reduction coupled with chemical tagging and mass mapping/Edman Sequencing (6-14), and partial reduction coupled with CN-induced

cleavage involving cyanylation followed by mass mapping (15-20). All mapping strategies are based on a knowledge of the amino acid sequence and cleavage of the protein backbone at specific sites; the involvement of particular disulfide cysteines in a given disulfide bond is deduced from mass analysis of the cleaved fragments.

II. Disulfide Mass Mapping Based on Proteolysis

1. Overview of the Method

Given the sequence of amino acid residues in a protein and an enzyme of known specificity, one can anticipate the fragments (and thus, compute their masses) into which the protein would be cleaved upon proteolytic digestion. The analytical process of correlating experimentally detected fragments with those predicted based on a knowledge of the sequence is called '**mapping**'. Experimentally, one subjects the protein to proteolytic hydrolysis, analyzes the digestion mixture, and compares the experimental results with the expected results. The larger the proportion (at least 25% coverage) of the expected fragments detected, the greater the confidence in correct identification of the protein. In early applications of this mapping strategy, the components of the digestion mixture were separated analytically by HPLC with identification by Edman sequencing, but more recently, the digestion mixture has been analyzed directly by mass spectrometry (21), in which case the methodology is called '**mass mapping**'.

Ideally, the protease (P) will cleave the protein at least once between the cysteines as indicated by 'P' in the analytical scheme involving a hypothetical protein, as illustrated Figure 1.3. Proteolytic cleavage gives a different array of fragments depending on the

connectivity of the disulfides. A tedious aspect of proteolytic disulfide mapping is the isolation and identification of the cysteinyl and cystinyl peptide fragments from a usually large group of non-sulfur containing peptides. A cysteinyl fragment usually contains a free cysteine residue; whereas a cystinyl fragment usually contains an intra- or inter-disulfide bond (Figure 1.2). Cysteinyl peptides that are 'missing' in the results of analysis of the digestion mixture (compared to the list of anticipated proteolytic peptides) indicate which cysteines are involved in the disulfides (cystines). Of prime importance is the isolation and identification of the cystinyl proteolytic peptides, as these indicate which two cysteines are involved in a particular disulfide bond. The cystinyl peptides will be represented by unexpected HPLC peaks in a chromatogram of the digestion mixture of the protein or by mass spectral peaks for fragments heavier than anticipated from consideration of individual proteolytic fragments of the completely reduced protein. Recognition of cysteinyl peptides can be facilitated by the clever stable-isotope labeling strategies recently reported by Chen *et al.* (22) and Adamczyk (23).

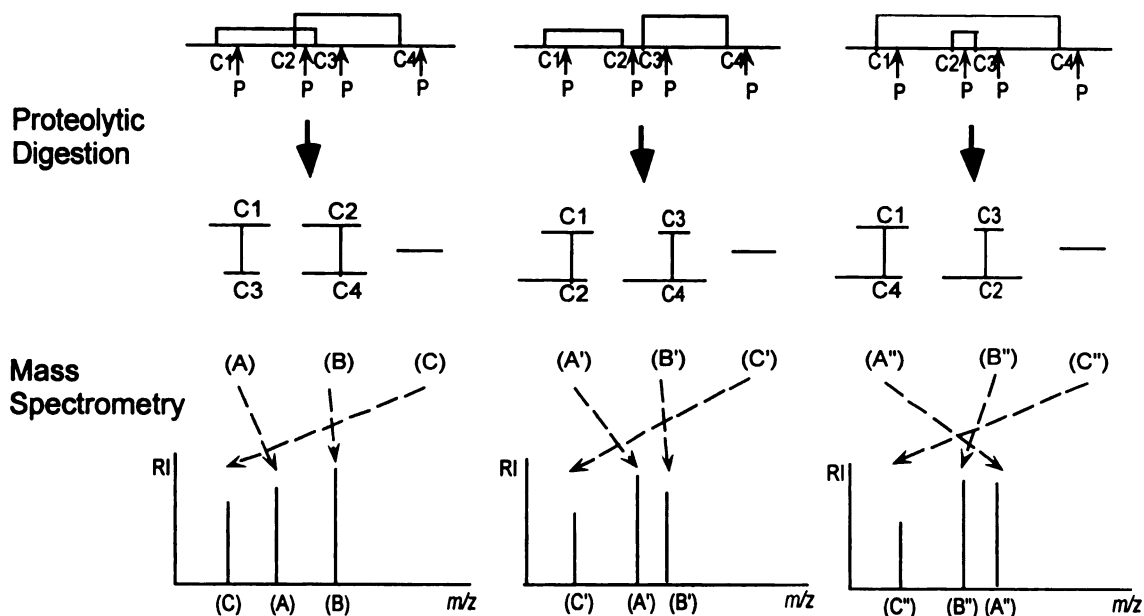


Figure 1.3. Conceptual illustration of the different array of fragments that can be used to distinguish the three possible isomeric arrangements of two disulfide bonds in a hypothetical protein containing four disulfide cysteines by the classic proteolytic approach for disulfide mass mapping. Note that fragments C, C', and C'' have the same mass in each case because they contain no cysteine, whereas the mass of the fragments containing cysteines vary substantially for the different isomeric disulfide arrangements.

To the extent possible, it is prudent to perform confirmatory experiments on the primary mapping results. For example, in Figure 1.2, even though the peak of the highest m/z value in the first mass spectrum is consistent with the mass of the disulfide-linked peptides, subjecting the putative structure to chemical reduction and detecting peaks for the individual peptides, as illustrated in the bottom panel of Figure 1.2, substantially increases confidence in the final results. Similarly, chemical reduction of the structure containing an intramolecular disulfide in the top panel of Figure 1.2 should shift the mass of the putative species by 2 Da, as illustrated in the lower panel. Alternatively, reduction of disulfides during analysis by matrix assisted laser desorption/ionization (MALDI) mass spectrometry (MS) (also called 'prompt' fragmentation; references 24-28) and sequencing analysis by MALDI-post source decay (PSD)-MS (reference 17, 27-30) or

electrospray (ESI) tandem mass spectrometry (MS/MS) (references 31-35) can provide confirmation for preliminary results based solely on mass; examples of these procedures are described below in section 4: “Representative Application” and section 5: “Advanced Topic”.

2. Constraints/Caveats

It is desirable, if not a requirement, that at least one proteolytic site be available between each of the cysteines in the amino acid sequence of the protein. The lack of any such cleavage site could complicate the analysis because it will not produce separate proteolytic peptides for each of the cysteines (or cystines). A single proteolytic peptide containing two or more cysteines (or cystines) will complicate the analysis and/or lead to ambiguous results. The closer the cysteines are to one another in the sequence, the less likely it will be to find a useful proteolytic site between them. In the extreme case of adjacent cysteines, the proteolytic approach is unsatisfactory, if not useless (36).

A frequent concern in using the proteolytic approach to disulfide mapping is the problem of disulfide scrambling or exchange. Most proteolytic digestions are conducted in slightly alkaline conditions (e.g., pH 8), in which a significant fraction of any free sulfhydryl group will be deprotonated to the thiolate form. An existing disulfide is vulnerable to attack by a thiolate group that is in close proximity, resulting in formation of a different disulfide bond; such ‘scrambling’ or interchange of native and non-native disulfides can give rise to artifactual results. Frequently, the free sulfhydryls on free cysteines are alkylated prior to proteolytic digestion to avoid scrambling; however, some alkylation reactions themselves are conducted at high pH, e.g., the use of iodoacetamide

(37). During modification of the free cysteines, it was found that the alkylating reagent with the lower optimal working pH (such as maleimide derivatives) more effectively suppresses disulfide scrambling; the more commonly used iodoacetyl derivative alkylating reagent, which works at more alkaline conditions, is more liable to generate scrambled artifacts. It was also found that addition of the alkylating reagent prior to the denaturant helps prevent scrambling to some extent (34). Disulfide scrambling also can occur during proteolytic digestion. For the same reason, using sample-processing proteases having a low working pH may be better to minimize the scrambling. However, proteases that function well at low pH such as pepsin and thermolysin, usually cleave the substrate with less specificity, which may bring other problems to mass mapping. For those more specific proteases, like trypsin and Glu-C endoproteinase having an alkaline working pH, use of an immobilized protease cartridge for digestion can effectively reduce disulfide interchange because the higher enzyme-to-substrate ratio reduces the digestion time from 2-18 hours to 5 minutes (38).

3. An Experimental Protocol

(1) Dissolve the protein in a suitable buffer, add alkylating reagent followed by denaturant.

(2) Adjust the pH to digestion conditions and incubate the mixture in a thermal bath.

If limited digestion is found, the following approaches might be helpful: a)

Increase the protease concentration. b) Choose a different protease or use a

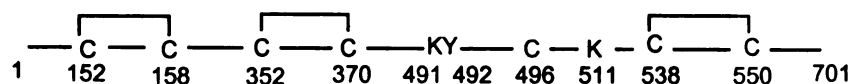
combination of two or more proteases. c) Pre-treat the protein with cyanogen bromide, if methionine residues are present in the protein.

- (3) Use HPLC to separate the digestion mixture. A comparative HPLC run after treating the digestion mixture with an excess of reducing agent is helpful in recognizing the disulfide-linked peptides.
- (4) Use MS (MALDI-MS or ESI-MS) to analyze the HPLC fractions; compare the experimental data with the mass values calculated for expected digestion fractions of the protein. Absent fragments plus unexpected fragments (based on consideration of the fully reduced protein) provide information relating to the connectivity of disulfide cysteines involved in disulfide bonds.
- (5) Reduce the disulfide-linked peptides and re-analyze them by mass spectrometry or sequencing analysis by MALDI-PSD or ESI-collision induced dissociation (CID)-MS/MS. These may help confirm the identification solely based on mass mapping.

4. A Representative Application (39)

Testis angiotensin-converting enzyme (tACE) is a protein with 701 amino acid residues including 7 cysteines, which were asserted by the authors to form three disulfide bonds with one free cysteine as illustrated in the linear structure in Figure 1.4.

Putative disulfide structure of tACE:



Experimentally observed tryptic fragment :
492-511-- incorporating derivatized cys 496



Figure 1.4. Overall sequence of tACE showing tryptic cleavage sites at K491 and K511; these define the origin of the tryptic fragment Tyr492-Lys511 at the lower left that shifts from 2145.4 Da to 2453.7 Da upon alkylation of the free sulfhydryl on Cys496. The mass of the protonated alkylated-peptide (2453.7 Da) is very close to the observed peak at m/z 2453.4 in Figure 1.6.

In general, a protein consisting of an odd number of cysteines will have at least one free cysteine (maybe more) in its native form; a protein consisting of an even number of cysteines might have no free cysteine. For a protein with one free cysteine and three disulfide bonds, one must consider 105 different possible **disulfide structures**. (A disulfide structure is defined as a particular arrangement of all the disulfide bonds in a protein. The formula for calculating the number of possible isomers for this case is as follows:

$$\alpha(M, n) = \frac{M!}{2^n n! (M - 2n)!}$$

where $\alpha(M, n)$ represents the number of isomeric disulfide structures; n is the number of disulfide bonds among M cysteines, and $M \geq 2n$; see reference 40 for details.) All the 105

possible disulfide structures need to be considered when computing the expected masses for all possible proteolytic peptide fragments. Given the sequence of a protein and the specificity of a protease, calculation of the masses of the expected proteolytic fragments can be automated; several websites (<http://prosepector.ucsf.edu>, <http://prowl.rockefeller.edu>, etc.) provide this service.

In an effort to simplify the assessment of cysteine status in a protein, the locus of any free cysteines is usually determined first. In the case of tACE, the protein was digested by endoproteinase Lys-C after the free cysteine was modified by a fluorescent alkylating reagent, 5-([2-(iodoacetyl)amino]ethylamino) naphthalene-1-sulfonic acid (IAEDANS) to facilitate detection of the alkylated cysteine-containing peptide. During analysis of the alkylated digestion mixture by HPLC, because one of the detectors was adjusted to 340 nm to measure the fluorescence of the 5-([2-(acetyl)amino]ethylamino) naphthalene-1-sulfonic acid (AEDANS) group, only a single peak was observed in the chromatogram at the top of Figure 1.5. The fraction corresponding to the single fluorescence peak was collected and analyzed by MALDI-MS. The corresponding mass spectrum (Figure 1.6) of the detected fraction (in the top panel of Figure 1.5) shows three peaks indicating that the HPLC fraction actually consists of at least three coeluting components. Two of the mass spectral peaks are consistent with the calculated masses of two of the expected proteolytic peptides (Gln308-Lys317 and Val119-Lys137). The peak at m/z 2453.4 represents the protonated peptide consisting of residues from Tyr492 to Lys511, which after alkylation of Cys496 with the fluorophore (AEDANS) as illustrated in Figure 1.4, shift to a mass of 2453.7 Da for (calculated mass of the protonated species = 2453.7 Da). Sequencing analysis of this peptide by MALDI-PSD-

MS also confirmed the above identification (data not shown). These data provide convincing evidence that Cys496 in the native tACE is a free cysteine.

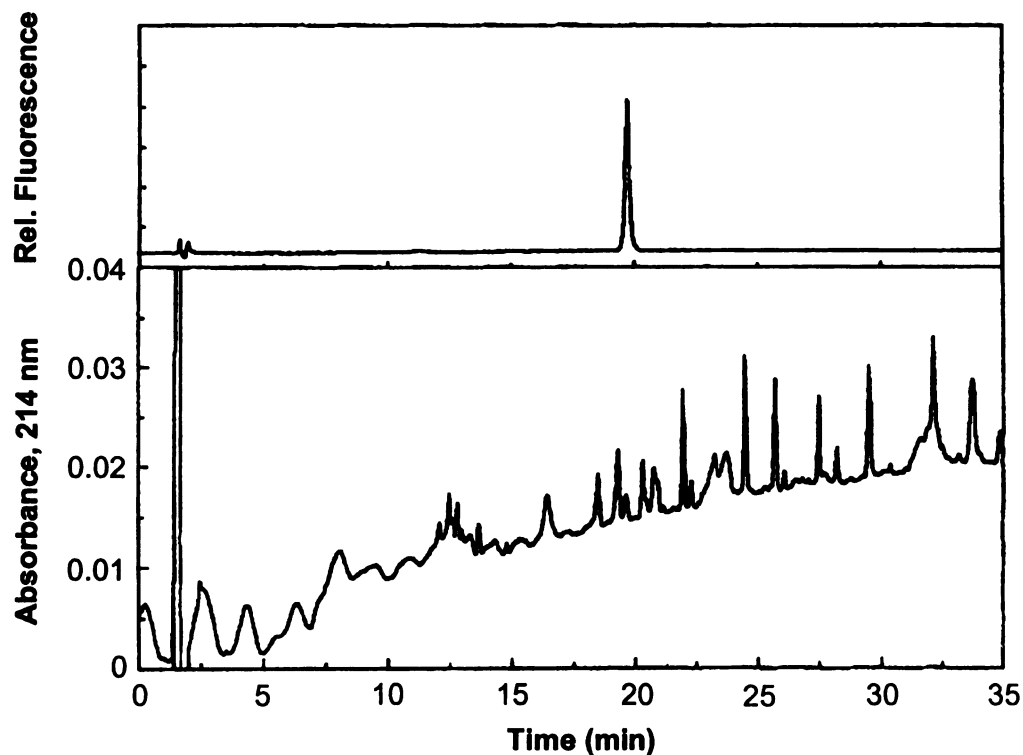


Figure 1.5. HPLC chromatograms of the endoproteinase Lys-C digestion mixture of AEDANS-labeled tACE as recorded by two different detectors. The upper panel was monitored at 340 nm for the fluorescence of the AEDANS fluorophore to selectively detect the peptide with the AEDANS-modified cysteine; the lower panel was monitored at 214 nm to detect all the digestion peptides (adopted from reference 39).

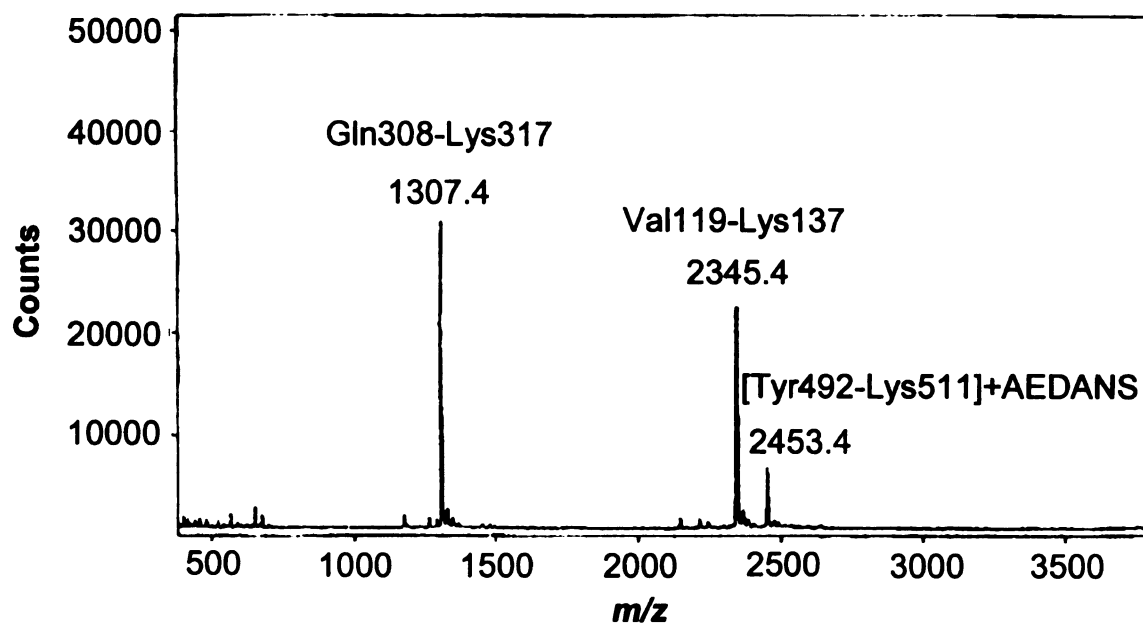
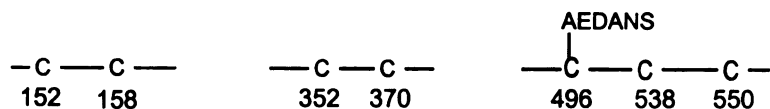


Figure 1.6. The MALDI mass spectrum of the HPLC fraction represented by the single peak in the top panel of Figure 1.5. The peak at m/z 2453.4 represents the protonated proteolytic fragment containing the modified cysteine as illustrated in Figure 1.4 and described in the text. The other two peaks represent other proteolytic peptides that coelute with the alkylated one (adopted from reference 39).

Having identified the locus of the free cysteine in tACE, attention can now be focused on determining the connectivity of the other six disulfide cysteines to form three cystines (recall that there are 15 isomeric possibilities; see also in Figure 1.7). It is desirable to degrade the protein by cleaving the peptide backbone at least once between every two disulfide cysteine residues. In this case, alkylated tACE was digested with cyanogen bromide and, subsequently, further digested with endoproteinase Lys-C. Cyanogen bromide (CNBr) is a chemical reagent that cuts the peptide backbone specifically at the C-terminal side of methionine; because of its small size, CNBr can readily penetrate to interior cleavage sites that sometimes are less accessible to proteases of much larger size. There are 19 methionine residues in tACE; a perfect CNBr digestion with no miscleavages will generate 20 fragments if no disulfide bonds link them together.

As shown at the top of Figure 1.7, three of the CNBr fragments contain cysteines; two of these fragments contain two cysteines each and one contains three cysteines, including Cys496, the alkylated (AEDANS) cysteine. The disulfide cysteines in those three fragments can form disulfide bonds in a variety of ways as shown by three categories in the lower portion of Figure 1.7. In category (i), each of these of the three pairs of disulfide cysteines forms an intra-molecular disulfide bond. In category (ii), two pairs of the disulfide cysteines form two inter-molecular disulfides; there are six possibilities. In category (iii), all three CNBr fragments are linked by three inter-molecular disulfide bonds; there are eight possible isomeric arrangements. Fortunately for the authors, the MALDI-MS mass spectral data showed peaks corresponding to the three separate intra-molecular disulfide peptides represented in category (i), indicating that the three intra-molecular disulfide bonds in tACE are isolated or do not overlap one another.

Three cysteine-containing fragments from CNBr digestion
of tACE without regard for cysteine status:



Possible disulfide structures
(without consideration of cys 496 as it's alkylated):

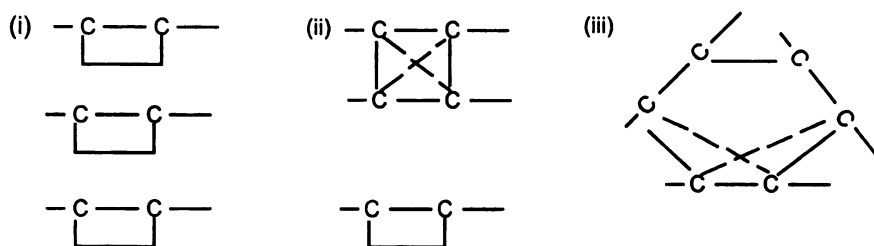


Figure 1.7. Conceptual representations of the disulfide arrangement among the six cysteines in the three cysteine-containing fragments from CNBr digestion of tACE. Category (i) shows the simplest arrangement, which derives from three isolated disulfides in tACE; experimental results indicate that this category represents the disulfide arrangement in native tACE. Categories (ii) and (iii) consist of six and eight isomeric arrangements, respectively, some of which involve overlapping disulfide bonds (indicated, in part for two possibilities, by dashed lines).

Mass spectral data (not shown) from analysis of the reaction mixture from CNBr digestion of AEDANS-alkylated tACE include peaks that correspond to the calculated masses of the three cysteine-containing fragments in Figure 1.7. While these data strongly suggest three isolated disulfides between the pairs of cysteines within the three fragments at the top of Figure 1.7., it is prudent to seek confirmatory evidence. The following text describes an additional experiment to confirm the putative connectivity in the CNBr digestion fragment Ala451-Met566, which contains the alkylated cysteine, as illustrated in Figure 1.9. Subsequent digestion of this fragment (the 2nd structure in Figure 1.9) by trypsin generates a fragment (the 3rd structure in Figure 1.9) consisting of two peptides linked by a disulfide (Cys538-Cys550) and having a calculated mass (for

singly protonated species MH^+) of 5243.1 Da. Analysis of an HPLC fraction (not shown) of the trypsin digest of the putative CNBr fragment (the 2nd structure in Figure 1.9) by MALDI-MS produced the mass spectrum in Figure 1.8. The peak at m/z 5242.3 in Figure 1.8 corresponds well to the calculated mass of the protonated fragment (5243.1 Da). Furthermore, prompt fragmentation (also called in-source fragmentation, see references 24-28) of the disulfide bond in this fragment (bottom of Figure 1.9) during MALDI-MS analysis provided evidence (peaks at m/z 856.6 and m/z 4388.0) for the individual peptides. Overall, these results provide convincing evidence for the structure of tACE at the top of Figure 1.9.

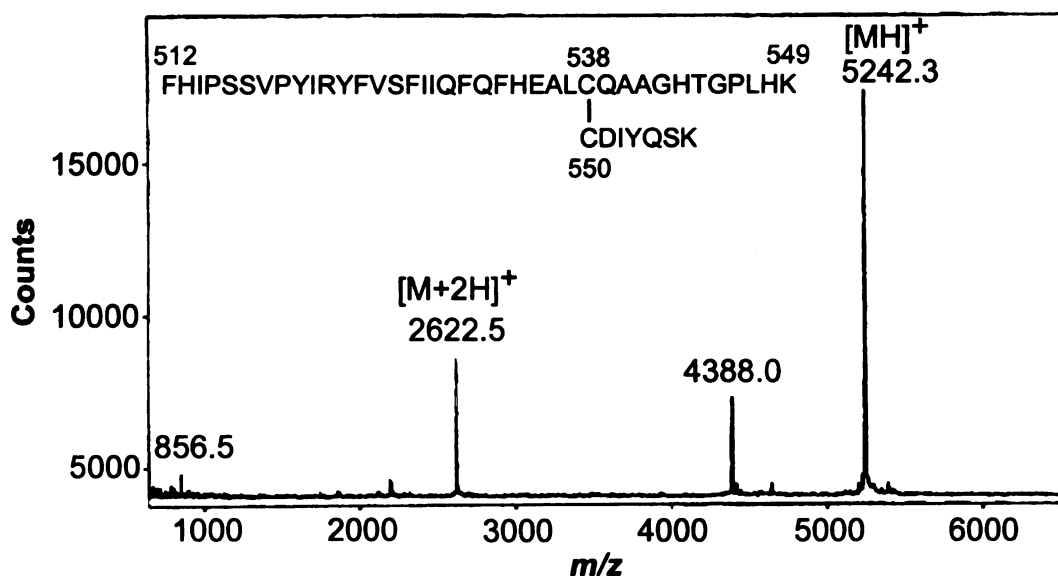


Figure 1.8. The MALDI mass spectrum of the putative peptide of a CNBr fragment (see Figure 1.9) of the tACE protein. Peaks at m/z 856.6 and m/z 4388.0 correspond to the individual peptides connected by the disulfide that was cleaved during the MALDI process (adopted from reference 39).

Putative Disulfide Structure of the Alkylated tACE:

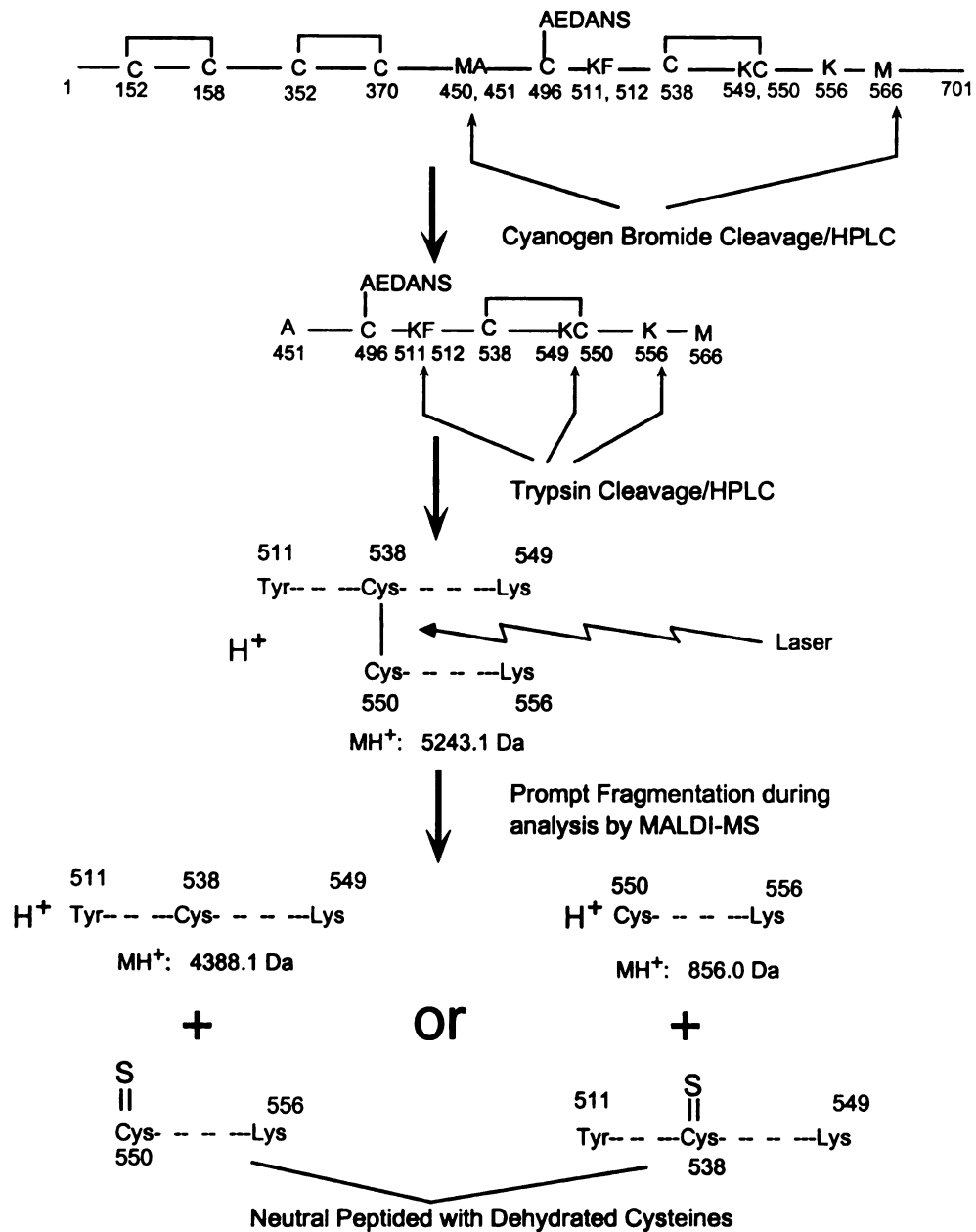


Figure 1.9. Conceptual illustration of the digestion of tACE by CNBr and trypsin to produce two peptides linked by a disulfide bond, the analysis of which by MALDI-MS identifies the disulfide Cys538-Cys550.

5. Advanced Topics: Using MALDI-PSD-MS and ESI-CID-MS/MS to Obtain

Additional Information on Disulfide-Linked Peptides

Fragmentation of the protonated-disulfide-linked peptides by metastable decay during analysis by MALDI-PSD-MS (17, 27-30) or by collision-induced dissociation (CID) during ESI-MS/MS (31-35) sometimes provides sequence information on disulfide-linked proteolytic peptides. This second dimension of mass spectrometric information can be used for confirmation of the identity of some disulfide-linked peptides.

The first example here involves the analysis by MALDI-PSD-MS of a digestion product consisting of three peptides linked by two disulfide bonds as shown by the structure insert in the mass spectrum in Figure 1.10. The proteolytic fragment of interest here derives from the thermolysin digestion of a tryptic fragment of the recombinant human tumor necrosis factor binding protein (TNFbp in reference 28), a cysteine-rich protein with 13 disulfides among 26 cysteines across 162 residues. In this special case, there are two possible disulfide structures; the alternative to the structure shown in Figure 1.10 would involve switching positions of linking the tetrapeptide and the pentapeptide to the central peptide. During analysis of the disulfide-linked peptides by MALDI-PSD-MS, it is likely that some of the disulfide bonds as well as some of the peptide bonds will fragment. While the peaks at m/z 242.9 and m/z 1350 in the mass spectrum (Figure 1.10) do not distinguish the two isomeric disulfide structures under consideration, the peaks at m/z 825.6, 940, and 1027 uniquely correspond to the structure shown in Figure 1.10.

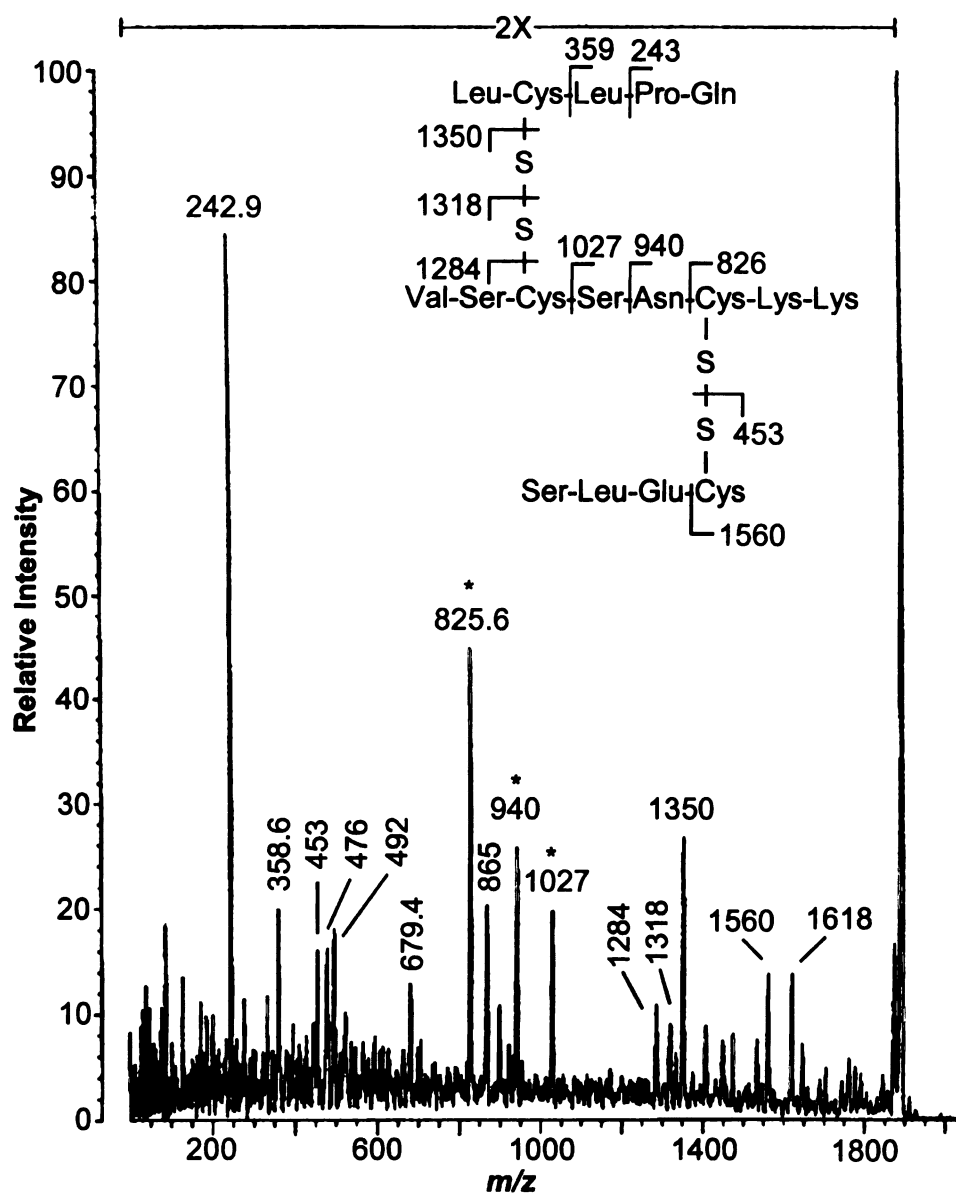


Figure 1.10. The MALDI-PSD-MS mass spectrum of a disulfide-linked proteolytic fragment from dual digestion of the TNFbp protein with trypsin and thermolysin. The numbers by the structure give the mass of the indicated fragment ions; the peaks marked by an asterisk at m/z 825.6, 940, and 1027 confirm the disulfide structure shown (adopted from reference 28).

Cysteines occur in a protein less frequently on average than those residues serving as proteolytic cleavage sites. If cysteines are somewhat evenly distributed within a protein (the above example is an exception to this assertion, as the cysteine content of TNFbp protein is abnormally high), there is often more than one cleavage site between the cysteines. As a result, a less complicated situation involving a disulfide-linked proteolytic fragment consisting of only two peptides linked by one disulfide is usually observed. An example of such a structure is inserted into the ESI-CID-MS/MS mass spectrum (Figure 1.11) of a disulfide-linked tryptic fragment of beta-1,4-galactosyltransferase (34). In describing their mass spectral data, the authors used a hybridization of the Roepstorff *et al.* (41) and Biemann (42) nomenclature to designate ions from the two peptides linked by a disulfide bond. The peaks labeled with an upper case **Y** denote the y-ions from the fragmentation of the lower peptide; peaks with the lower case **y** and **b** denote the y- and b-series of ions from fragmentation of the upper peptide. $(Y_{11}b_{12})^{2+}$ denotes a doubly charged **b** ion of the upper peptide with 12 residues from the N-terminus that was linked by a disulfide bond to the **Y**₁₁ ions from the lower peptide (with 11 residues). Most of the fragmentation occurred on the C-terminal side of the cysteine (residue 5) in the upper peptide, namely, the **y**_n series and **Y**₁₁**b**_n series of ions. Only the ions $[(Y_{11}b_n)^{2+}$ and $(Y_9b_{13})^{2+}$ ions] containing the intermolecular disulfide provide information on the disulfide connectivity. The mass of the lower peptide can be deduced by subtracting the mass of the **b**_n series of ions (to the right of Cys5) of the upper peptide from the mass of the **b**_n**Y**₁₁ ions as observed in the experiment. Ions such as **Y**₉**b**₁₃ are not observed often, possibly because they require fragmentation of two peptide bonds.

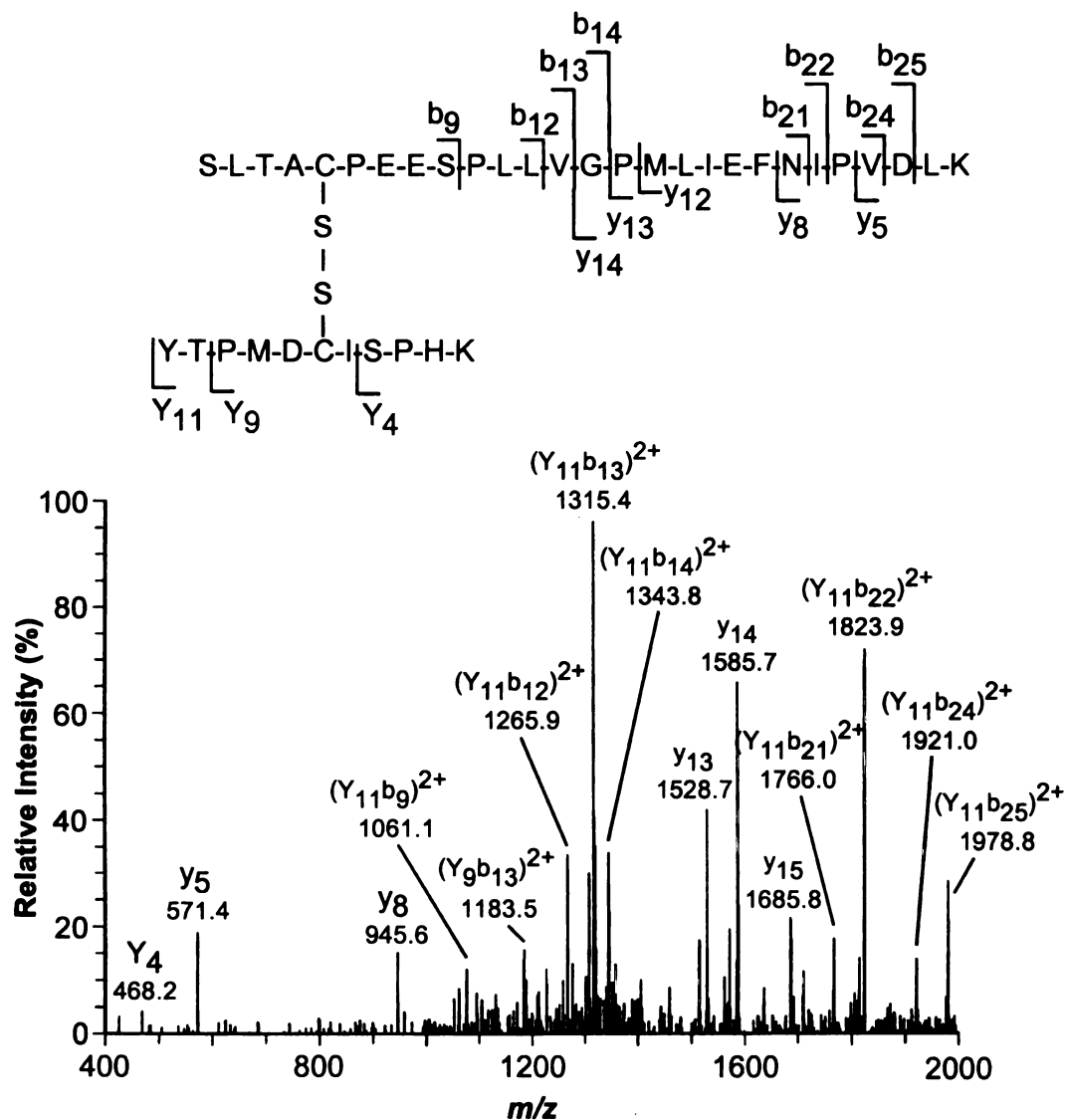


Figure 1.11. The ESI-CID-MS/MS mass spectrum of a triply charged ion $[M+3H]^3+$ at m/z 1406.1 corresponding to a disulfide-linked tryptic fragment (in which a disulfide bond links two peptides, S130-K156 and Y171-K181) of bovine beta-1,4-galactosyltransferase (adopted from reference 34).

It is important to note the following observations about sequencing analysis of disulfide-linked peptides by MALDI-PSD-MS and ESI-CID-MS/MS. First, although several successful examples by both techniques had been shown (17, 27-35), both

techniques failed to provide diagnostic information in other cases (28). In addition, more diverse disulfide-linked peptides need to be studied by both techniques to optimize the experimental parameters and determine the limitations of the techniques. For example, while intermolecular disulfide bonds in the disulfide-linked peptides are usually fragmented during analysis by MALDI-PSD-MS, fragmentation of disulfide bonds does not always occur during analysis by ESI-CID-MS/MS of disulfide-linked peptides. No indication of the fragmentation of the intermolecular disulfide bond, either symmetrical or non-symmetrical was found in the example described above and several other examples (34, 35); whereas fragmentation of such disulfide bonds in other disulfide-linked peptides during ESI-CID-MS/MS was observed (27-32, 43). It seems that what controls the fragmentation at various sites of the disulfide-linked peptides would be the key to solve this problem.

6. Conclusions

The proteolytic approach to disulfide mapping has been used for many years, first with only chromatographic and sequencing analyses, but these have been combined with analyses by mass spectrometry in recent applications. The proteolytic approach can be time- and labor-intensive primarily because of the effort in recognizing and isolating the cysteinyl and cystinyl peptides from the often-large number of other proteolytic peptides generated upon digestion of a large protein.

Disulfide mapping of proteins containing adjacent cysteines is impractical, if not impossible, by the proteolytic approach. Proteins containing nearby cysteines in their

amino acid sequence may also present intractable problems for the proteolytic approach because of the lack of a trustworthy proteolytic site between the cysteine residues.

Concerns about disulfide scrambling are always an issue in the proteolytic approach to disulfide mapping because of the susceptibility of an existing cystine to attack by a thiolate ion produced from significant dissociation of any cysteine under the slightly alkaline conditions required for the enzymatic digestion. Alkylation of free cysteines is effective in some case in suppressing artifact formation due to disulfide scrambling; thiolate-disulfide exchange was observed during some of the thiolate alkylations under alkaline pH (9).

III. Determination of Disulfide Bonds by Partial Reduction, Chemical Tagging, Edman Sequencing Analysis and Mass Spectrometry

1. Overview of the Method

Free cysteines in the native protein are modified with a particular alkylating reagent, e.g., with iodoacetamide, to prevent disulfide scrambling. As the multiple disulfide bonds are progressively reduced by partial reduction with tris-(2-carboxyethyl)-phosphine (TCEP) hydrochloride, the nascent sulfhydryls are modified with different alkylating reagents to chemically tag the products at different stages of reduction. The modified cysteines produce distinguishing responses during separation of the phenylthiohydantoin (PTH) derivatives by HPLC following Edman degradation. Thus, this analytical strategy permits recognition of free cysteines in the original protein as well as the locus of disulfide cysteines that were coupled into specific disulfide bonds.

Partial reduction of the protein is accomplished with an excess amount of reducing agent, but under kinetically limiting conditions (6, 7). In this way, products isolated from the reaction mixture at an early stage will consist principally of one or more of several different singly reduced isoforms of the protein (e.g., a protein containing four disulfides can produce as many as four singly reduced isoforms). At a later stage during the dynamic process of partial reduction, doubly reduced isoforms (e.g., a protein containing four disulfides can produce as many as six doubly reduced isoforms) as well as triply reduced isoforms will appear in the reaction mixture.

This partial reduction/chemical modification is especially advantageous in the analysis of small cysteine-rich proteins, which do not respond well to proteolytic digestion due to inaccessibility of the protease to sites that are in the closely packed core region. For proteins consisting of fewer than 50 residues, Edman degradation performed on the protein after the cysteines have been modified by different alkylating reagents provides information on the cysteine status (free cysteines or disulfide cysteines involved in disulfide bonds) (reference 44). Mass spectrometry is helpful in verifying that proper alkylation has been performed after appropriate stages of partial reduction of the disulfide bonds in the protein (9, 12-14). For proteins consisting of more than 50 residues, partially reduced/alkylated species of the protein are usually cleaved into smaller sized fragments by proteolyses to allow different “tags” on various alkylated cysteines residues to be identified by Edman sequencing and mass spectrometry (12-14).

2. Constraints/Caveats

Because different proteins have different disulfide structures, the reduction yields vary accordingly. Some proteins can be converted into some of each of the possible singly reduced isoforms because all the disulfides are accessible to the reducing agent. For other proteins, however, only certain disulfides are readily accessible to the reducing agent. In such a case, a stepwise reduction/alkylation approach may be useful in which the available singly reduced isoforms of the protein are alkylated before the reducing conditions are strengthened to force some of the protein to the doubly reduced stage (8). Different alkylating reagents can be used after different stages of reduction, thereby chemically 'tagging' specific cysteines as a function of time during the dynamic reduction process (8, 44). The temporal relationship of certain cysteines as manifested via chemical 'tagging' can provide a basis for deducing the connectivity of cysteines involved in a particular disulfide bond.

Many widely used alkylating reagents work under slightly alkaline conditions that bring about the possibility of disulfide scrambling. Usually, a reactive reagent like iodoacetamide is preferred over a less reactive reagent such as iodoacetic acid (7). Nonetheless, scrambling is sometimes observed if the kinetics of the disulfide exchange (scrambling) reaction is faster than that of the alkylation reaction. In those cases, an alkylating reagent, such as N-ethylmaleimide (NEM), that can be used under acid conditions is preferred (44). Other complications may arise during the use of alkylating reagents, such as the formation of isomeric products that give multiple peaks during HPLC separation as has been reported for N-ethylmaleimide, a problem related to the prochiral nature of the reagent (44).

Sometimes there is a problem in separating the partial reduction products by HPLC, due to their being structurally similar. While coelution of structurally similar partially reduced species can complicate data interpretation, the involvement of mass spectrometry can provide useful information regarding the disulfide status (9, 12-14).

3. An Experimental Protocol

- (1) Dissolve the protein in the appropriate buffer to alkylate any free cysteines.
- (2) Use the reducing agent (TCEP) to partially reduce the protein. The amount of reagent and the incubation time as well as the temperature should be controlled to optimize the yield of the partially reduced products. Alkylate the products with a alkylating reagent.
- (3) Subject the mixture to HPLC separation. If a poor yield is observed; repeat step (2) with stronger reducing conditions.
- (4) Use a different alkylating reagent at each progressive stage of reduction/alkylation.
- (5) To each HPLC fraction, add an excess of reducing agent to totally reduce the residual disulfide bonds. Alkylate the nascent free cysteines (from complete reduction of the residual disulfides) with a unique alkylating reagent.
- (6) If the protein is small enough, use Edman degradation to sequence the whole partially reduced/alkylated protein species. If the protein is large (>50 residues), a combination of proteolytic mapping and Edman degradation is preferred. Because there is no disulfide left at this stage, there is little resistance to the proteolytic digestion.

4. A Representative Application (44)

The AVR9 elicitor protein from *Cladosporium fulvum* has three disulfide bonds across the 28-residue structure (44). The AVR9 protein was partially reduced with TCEP and the nascent free cysteines were modified with N-ethylmaleimide (NEM) under acidic conditions. The reaction mixture was then separated by HPLC; major fractions were collected and subjected to complete reduction (of residual disulfides)/alkylation with 4-vinyl pyridine (4-VP). This second alkylating reagent (4-VP) tagged the cysteines in disulfides that survived the partial reduction procedure, while the first reagent (NEM) tagged the cysteines from disulfides that had been reduced during partial reduction. In this way, the nascent cysteines in the singly or doubly reduced species (tagged as NEM) could be distinguished from those cysteines (eventually tagged as 4-VP) remaining as disulfides after the initial step involving partial reduction.

The doubly modified protein was then subjected to Edman sequencing directly due to its small size. The retention times for PTH-NEM-modified cysteine and PTH-4-VP-modified cysteine in HPLC are different from one another as well as different from that for PTH-unmodified cysteine; thus, the sequencing results of the doubly modified protein provide a temporal record of its exposure to reduction/alkylation. The chemical tagging of various cysteines in the sequence allows the connectivity in disulfide bonds to be deduced because certain pairs of cysteines became available for simultaneous alkylation. For example, during the second cycle of Edman sequencing of the partially reduced/alkylated species of AVR9 represented by HPLC peak 2 in Figure 1.12, a peak having the retention time of PTH-NEM-cysteine was observed. As Edman degradation progressed to cycle 6, a 4-VP-cysteine was observed, and so forth. As Edman

degradation progressed to cycle 6, a 4-VP-cysteine was observed, and so forth. Thus, the diagram of sequencing results for the species represented by the HPLC peak 2 (in Figure 1.12) indicate that it is a singly reduced species, from which it can be deduced that AVR9 contains a disulfide bond between Cys2 and Cys16.

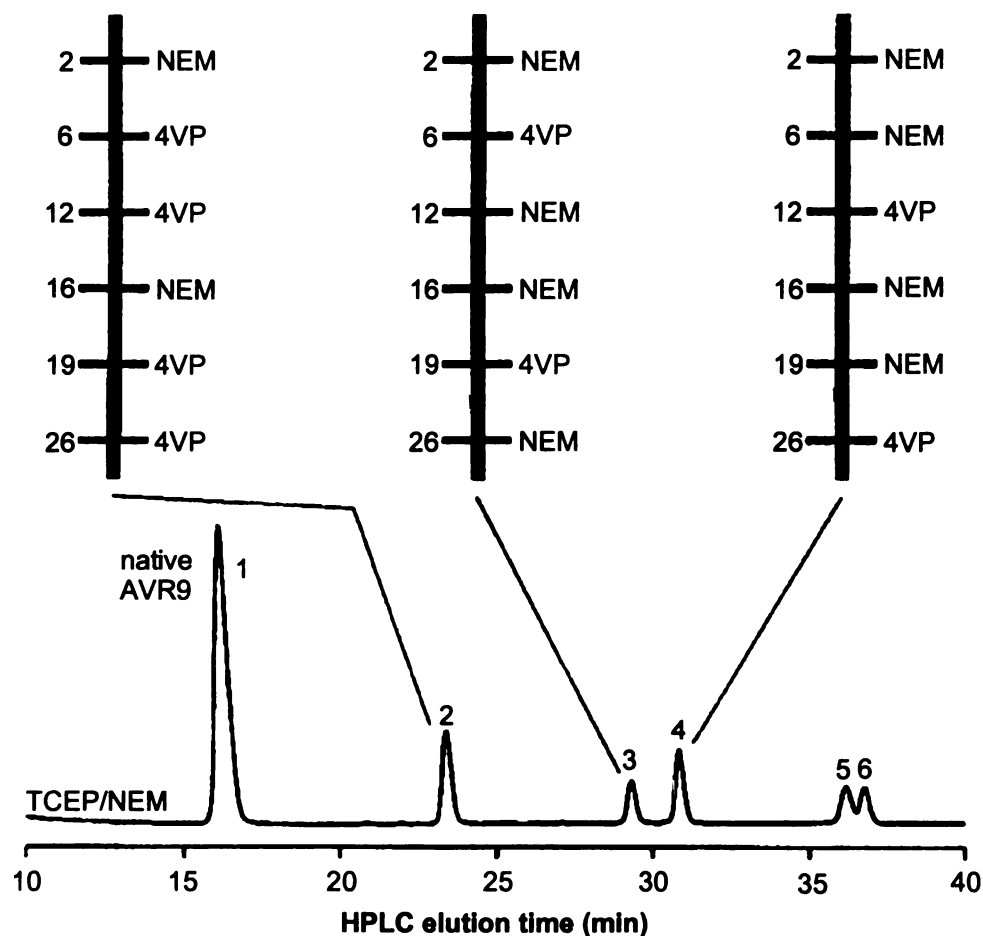


Figure 1.12. A diagrammatic presentation of sequencing results for the partial reduction/alkylated species of AVR9 represented by the peaks in the HPLC chromatogram. Peak 2 represents a singly reduced species from partial reduction and indicates that AVR9 contains a disulfide bond between Cys2-Cys16. Peak 3 represents a doubly reduced isoform containing a residual disulfide bond between Cys6-Cys19 before complete reduction; peak 4 represents another doubly reduced isoform containing a residual disulfide Cys12-Cys26 before complete reduction. Peaks 5 and 6 represent the completely reduced/NEM-modified species (as determined by mass spectrometry); both species have the same mass and apparently result from epimeric forms of the prochiral alkylating reagent (adapted from reference 44).

From the sequencing results for the species represented by the HPLC peaks 3 and 4 (in Figure 1.12), it can be deduced that these are two doubly reduced isoforms. Thus, each of the three fractions represents a different disulfide, thereby identifying the three disulfides as Cys2-Cys16, Cys6-Cys19, and Cys12-Cys26.

5. Conclusions

The partial reduction/alkylation approach is suitable for disulfide assignment in small cysteine-rich proteins. For larger proteins, this approach will require more steps than either the proteolytic approach or the partial reduction/cyanylation approach that will be discussed in the next section.

IV. Disulfide Mass Mapping Based on Partial Reduction, Cyanylation and CN-induced Cleavage

1. Overview of the Method

This novel approach to disulfide mapping is based on chemical cleavage of the peptide backbone on the N-terminal side of cyanylated cysteines (Figure 1.13). The first step in the process is cyanylation of free cysteines by a suitable reagent, e.g., 1-cyano-4-dimethylamino-pyridium (CDAP) tetrafluoroborate (45), as shown in Figure 1.13. The cyanylated cysteine is susceptible to nucleophilic attack, e.g., by ammonia, as shown in the second chemical reaction illustrated in Figure 1.13, producing a truncated peptide and an iminothiazolidine (itz)-carboxyl-blocked peptide (46).

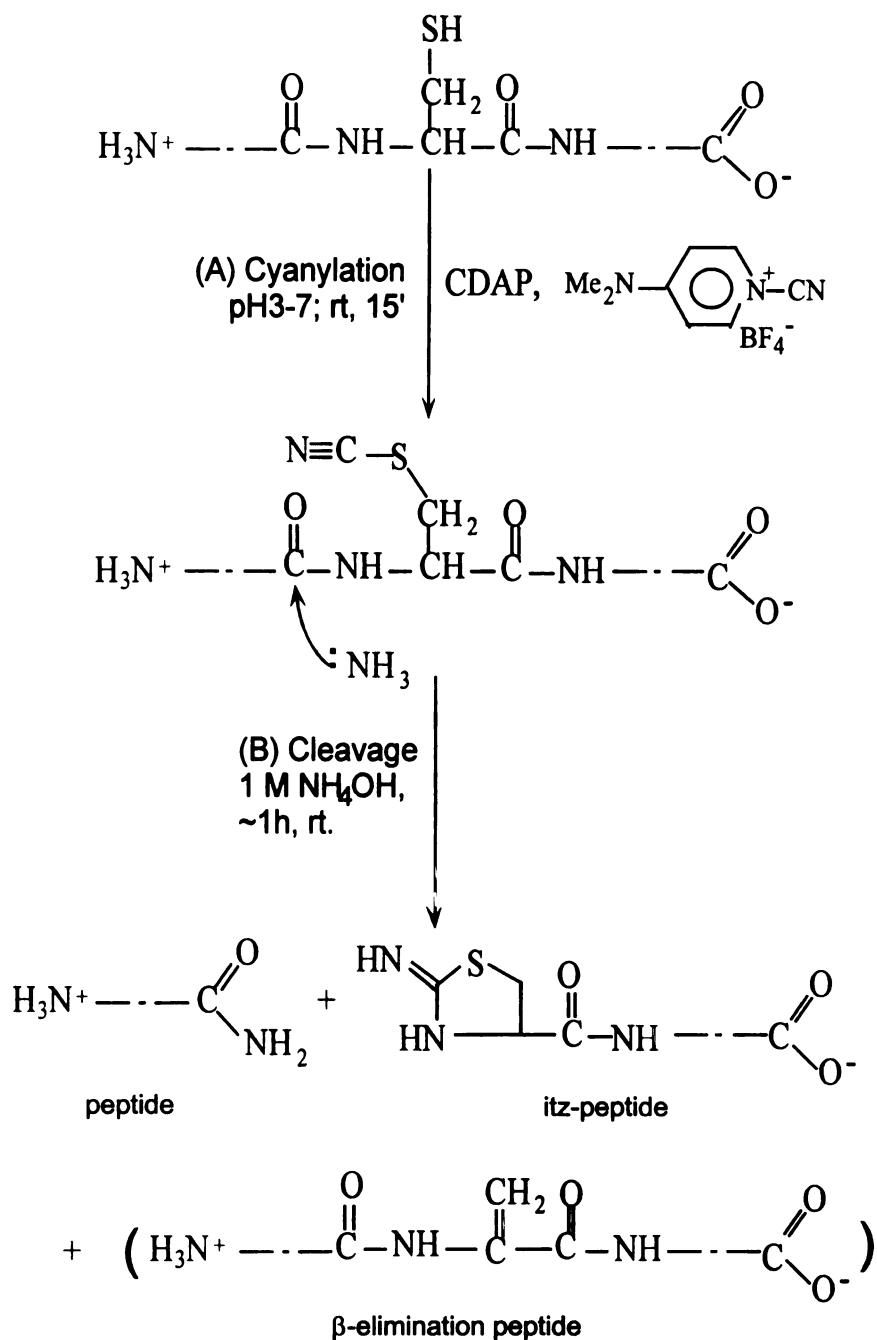


Figure 1.13. (A) Cyanylation of sulfhydryl group by CDAP and (B) peptide bond cleavage by ammonia. (itz = iminothiazolidine-carboxyl; β -elimination = loss of HSCN from a cyanylated cysteine residue.) (adopted from reference 47)

The cyanylation reaction is specific for cysteine; it does not modify methionine, nor does it react with hydroxylated, nor acidic nor basic side chains under analytically useful conditions (47). Furthermore, the cyanylating reagent will not react with a cystine (disulfide bond)! How, then, can the cyanylation reaction be used in disulfide mapping? It is first necessary to reduce a particular disulfide bond, then cyanylate the nascent free cysteines to promote cleavage at the sites previously tied up as a cystine. (Free cysteines in the native protein should be alkylated as described earlier in the previous section, or by use of the CDAP methodology prior to partially reducing the cystines as described below.) The analytical challenge comes with proteins containing more than one cystine, in which case the process of partial reduction (with TCEP) is used to generate a singly reduced isoform corresponding to each cystine (16).

Partial reduction (6, 7, and 16) is achieved under kinetically limiting reduction conditions using a phosphine-derived reducing reagent (e.g., tris-carboxyethyl-phosphine, TCEP). The concept of partial reduction is illustrated in Figure 1.14 with a hypothetical protein containing two cystines. The cyanylation/CN-induced cleavage chemistry is applied to each singly reduced isoform. Note that late in Figure 1.14, two or more of the expected three CN-induced cleavage products of the cyanylated protein might still be connected by residual disulfide bonds; complete reduction at this stage produces three individual fragments. Also note at the bottom of Figure 1.14 that each cystine is ideally represented by a unique set of fragments, the composition of which can be anticipated because the sequence of the original protein is known. For some disulfide proteins, doubly reduced isoforms, in addition to the singly reduced isoforms, may be utilized to determine their disulfide structure (reference 48, Chapter 2, and Chapter 3). Guidelines

for optimizing the partial reduction of disulfide proteins was discussed in references 6, 7, 16 and 48.

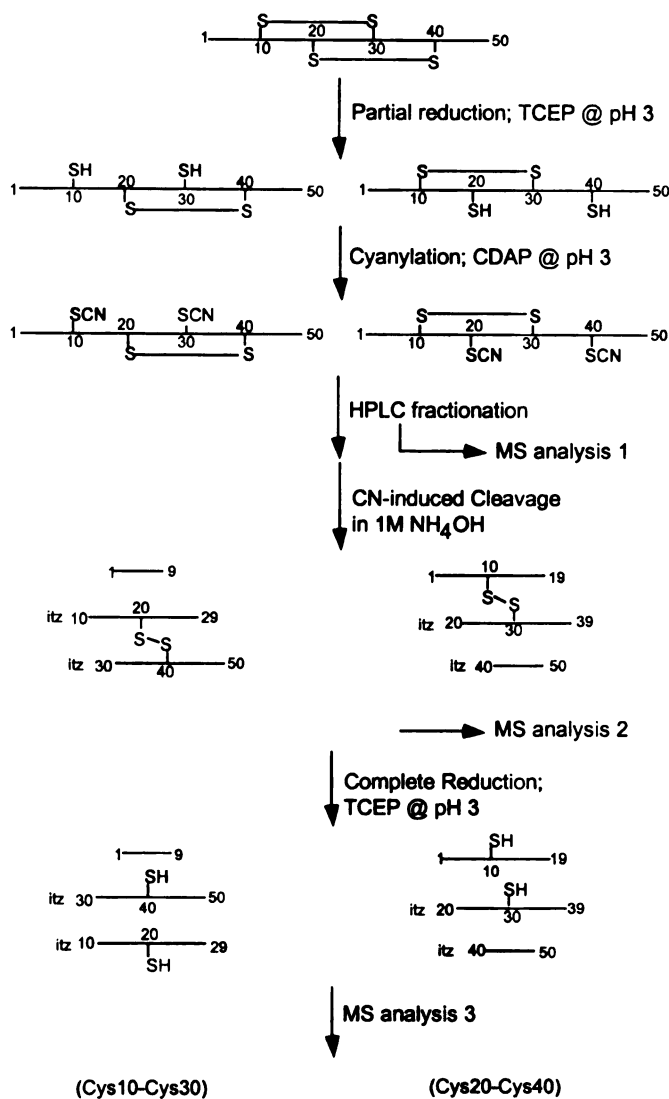


Figure 1.14. Chemical overview of the partial reduction/cyanylation/CN-induced cleavage methodology using singly reduced isoforms of a hypothetical 2-disulfide protein. “SCN” designates a cyanylated cysteine residue. (adopted from reference 16).

The individual partially reduced species of the protein can be isolated (either before, but preferably after cyanylation) usually by reversed phase (rp)-HPLC. In principle, such an isolation of the partially reduced species is not necessary, but doing so can simplify data interpretation (Chapter 5). Furthermore, such a separation leads to less complex mixtures of cleavage products, thereby minimizing the possibility of signal suppression for some components during analysis by MALDI or ESI mass spectrometry.

2. Constraints, Caveats, and Other Aspects

In addition to, and in competition with, the desired cleavage reaction illustrated at the bottom of Figure 1.13, the cyanylated cysteine can undergo β -elimination expelling HSCN producing a dehydroalanine peptide as illustrated. Observing the β -elimination product provides a ‘parity check’ on mass spectral signals by representing the sum of all residues in the expected cleavage fragments for a given cyanylated cysteine, a useful accounting feature when two or more cyanylated cysteines are present. The extent of β -elimination ranges from a minor to a major process depending mainly on the type of amino acid residue on the N-terminal side of the cyanylated cysteine (47); when proline is on the N-terminal side of a cyanylated cysteine, β -elimination is the dominant process.

Other side reactions, such as incomplete cyanylation and incomplete CN-induced cleavage, could occur and may complicate data processing (48). How to deal with such side reactions is discussed in reference 48 and Chapter 3.

In the case where there is an odd number of cysteines or where all the cysteines are not involved in cystines, the positions of the free cysteines in the original protein can be determined by immediate cyanylation and mass analysis of CN-induced cleavage

products after complete reduction of the residual disulfide bonds, which may still link the cleavage products together (49). If only two disulfide cysteines are involved as a cystine, the connectivity of the involved disulfide cysteines can be deduced by default. If four or more disulfide cysteines are involved as cystines, the partial reduction scheme can be invoked as described above. Alternatively, the initially free cysteines can be marked by alkylation as described earlier (37) (a process that renders them unavailable to the subsequent cyanylation reaction).

The strategy for determining the connectivity of disulfide cysteines involved in an intermolecular cystine is the same as that pursued in analysis of intramolecular cystines. That is, given the sequence of the polypeptide chains, one would use initial cyanylation or alkylation (if free cysteines are present), followed by the use of complete or partial reduction (depending on the number of cystines) and cyanylation/CN-induced cleavage/mass mapping.

The classic problem of adjacent cysteines is amenable to the cyanylation/cleavage approach (18), thereby providing a viable opportunity to solve disulfide structures that have hitherto been refractory to the proteolytic approach.

3. Experimental Protocols

A. A Protocol for Determining the Location of Free Cysteines by Alkylation/Cyanylation/CN-Induced Cleavage

- (1) Dissolve the protein in an appropriate buffer and mix it with alkylating reagent to alkylate (37) any free cysteines. Remove the reagents by using rp-Zip tips or rp-HPLC.
- (2) Completely reduce the disulfide bonds in the alkylated protein with TCEP (49).
- (3) Cyanylate the nascent free cysteines in the alkylated protein with CDAP (47).
Because CDAP reacts with TCEP, an excess amount of CDAP is needed.
- (4) Purify the product by rp-HPLC; collect and dry the HPLC fractions under vacuum.
- (5) Reconstitute a solution of the product in 1 M aqueous ammonia to cleave the peptide bond on the N-terminal side of the cyanylated cysteines (47).
- (6) Analyze the mixture by MALDI-MS or liquid chromatography (LC)-MS using ESI.

B. A Protocol for Determining the Location of Free Cysteines by Cyanylation/CN-Induced Cleavage

- (1) Dissolve the protein in an appropriate buffer and mix it with CDAP to cyanylate (47) any free cysteines.
- (2) Purify the cyanylated protein by rp-HPLC; collect and dry the HPLC fractions under vacuum.

- (3) Reconstitute a solution of the cyanylated protein in 1 M aqueous ammonia to cleave the peptide bond on the N-terminal side of the cyanylated cysteines (47).
- (4) Completely reduce any residual disulfide bonds in the cleaved protein with TCEP under strong reducing conditions (49).
- (5) Analyze the mixture by MALDI-MS or LC-MS using ESI.

C. A Protocol for Determining the Location of Disulfide Bonds by Partial

Reduction/ Cyanylation/CN-Induced Cleavage

- (1) Dissolve the protein in an appropriate buffer and mix it with an appropriate reagent to alkylate (37) or cyanylate (47) any free cysteines. Remove the reagents using rp-Zip-tips or rp-HPLC.
- (2) Use the reducing agent (TCEP) to partially reduce the alkylated or cyanylated protein. The amount of TCEP and the incubation time as well as the temperature should be adjusted to optimize the yield of the partially reduced products (6, 7, 16 and 48).
- (3) Cyanylate the products with CDAP (47). Because CDAP reacts with TCEP, an excess of CDAP is needed.
- (4) Separate the mixture by rp-HPLC. If a poor yield is observed, repeat step (2) with stronger reducing conditions. Collect and dry the HPLC fractions.
- (5) Reconstitute a solution of the cyanylated protein species in 1 M aqueous ammonia to cleave the peptide bond on the N-terminal side of the cyanylated cysteines (47).
- (6) Completely reduce any residual disulfide bonds in the cleaved protein mixtures with TCEP under strong reducing conditions (16, 49).

(7) Analyze the mixture by MALDI-MS or LC-MS using ESI.

4. A Representative Application: Disulfide Mass Mapping Using Singly

Reduced/Cyanylated Isoforms (16)

Bovine ribonuclease A (molecular mass for $MH^+ = 13,683$ Da) (50) contains 124 amino acids including eight disulfide cysteines that are linked by four disulfide bonds: Cys26-Cys84, Cys40-Cys95, Cys58-Cys110, and Cys65-Cys72.

When ribonuclease A (RnaseA) was subjected to the partial reduction/CN-cleavage/mass mapping methodology under conditions in reference 16, essentially equal amounts of the four possible singly reduced/cyanylated isoforms were produced. Figure 1.15 is the chromatogram of the residual intact RnaseA and its singly reduced/cyanylated isoforms, which resulted from partial reduction and cyanylation reactions of RnaseA. Only one equivalent of TCEP relative to the total cystine content in RnaseA was used during partial reduction for 15 minutes at room temperature; analysis by MALDI-MS showed that each of the four products produced (represented by HPLC peaks 1-4 in Figure 1.15) had a molecular weight 52 Da higher than the intact RnaseA (2 Da for reducing a cystine plus 25 Da net shift of each of the two nascent free cysteines upon cyanylation; see Figure 1.13.) as expected for singly reduced/cyanylated isoforms of RnaseA.

Table 1.1 lists the calculated m/z values for possible protonated fragments (MH^+) due to CN-induced cleavage of the peptide chains at different sites depending on which disulfide bond was reduced and cyanylated in forming the singly reduced/cyanylated isoform of ribonuclease-A.

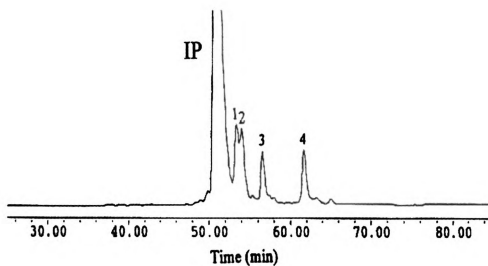


Figure 1.15. The HPLC chromatogram of the denatured ribonuclease-A and its singly reduced/cyanylated isoforms. The separation was done by reversed-phase HPLC on a Vydac C18 column (5 μ m, 4.6mm $\times$ 250mm) at a flow rate of 1.5 ml/min with a linear gradient 20-40% B in 90 minutes, where A = 0.1% TFA in water and B = 0.1% TFA in 90% CH₃CN. The peak marked by IP is the intact protein (ribonuclease-A). Peaks 1-4 represent singly reduced/cyanylated ribonuclease-A isoforms, as determined by MALDI-TOF-MS (adopted from reference 16).

Table 1.1. Calculated m/z values of protonated CN-induced cleavage fragments from all possible singly reduced/cyanylated isoforms of an unknown protein having the same sequence as RnaseA. The boldface entries correspond to native RnaseA, one of the 105 possible isomeric arrangements of the four cystines in a protein with this sequence.

No.	Cys i – Cys j	MH ⁺ [1-(i-1)]	MH ⁺ [itz-i-(j-1)]	MH ⁺ [itz-j-124]
1	Cys26-Cys40	2704.87	1750.05	9322.44
2	Cys26-Cys58	2704.87	3689.27	7383.23
3	Cys26-Cys65	2704.87	4420.1	6652.39
4	Cys26-Cys72	2704.87	5165.91	5906.58
5	Cys26-Cys84	2704.87	6546.42	4526.08
6	Cys26-Cys95	2704.87	7769.74	3302.75
7	Cys26-Cys110	2704.87	9412.68	1659.81
8	Cys40-Cys58	4411.89	1982.25	7383.23
9	Cys40-Cys65	4411.89	2713.09	6652.39
10	Cys40-Cys72	4411.89	3458.89	5906.58
11	Cys40-Cys84	4411.89	4839.4	4526.08
12	Cys40-Cys95	4411.89	6062.73	3302.75
13	Cys40-Cys110	4411.89	7705.66	1659.81
14	Cys58-Cys65	6351.11	773.87	6652.39
15	Cys58-Cys72	6351.11	1519.68	5906.58
16	Cys58-Cys84	6351.11	2900.18	4526.08
17	Cys58-Cys95	6351.11	4123.51	3302.75
18	Cys58-Cys110	6351.11	5766.45	1659.81
19	Cys65-Cys72	7081.94	788.84	5906.58
20	Cys65-Cys84	7081.94	2169.35	4526.08
21	Cys65-Cys95	7081.94	3392.67	3302.75
22	Cys65-Cys110	7081.94	5035.61	1659.81
23	Cys72-Cys84	7828.75	1423.54	4526.08
24	Cys72-Cys95	7828.75	2646.87	3302.75
25	Cys72-Cys110	7828.75	4289.8	1659.81
26	Cys84-Cys95	9208.26	1266.36	3302.75
27	Cys84-Cys110	9208.26	2909.3	1659.81
28	Cys95-Cys110	10431.58	1685.97	1659.81

* “Cys i – Cys j” designate the disulfide bond that was reduced to form the corresponding singly reduced/cyanylated isoform of ribonuclease-A.

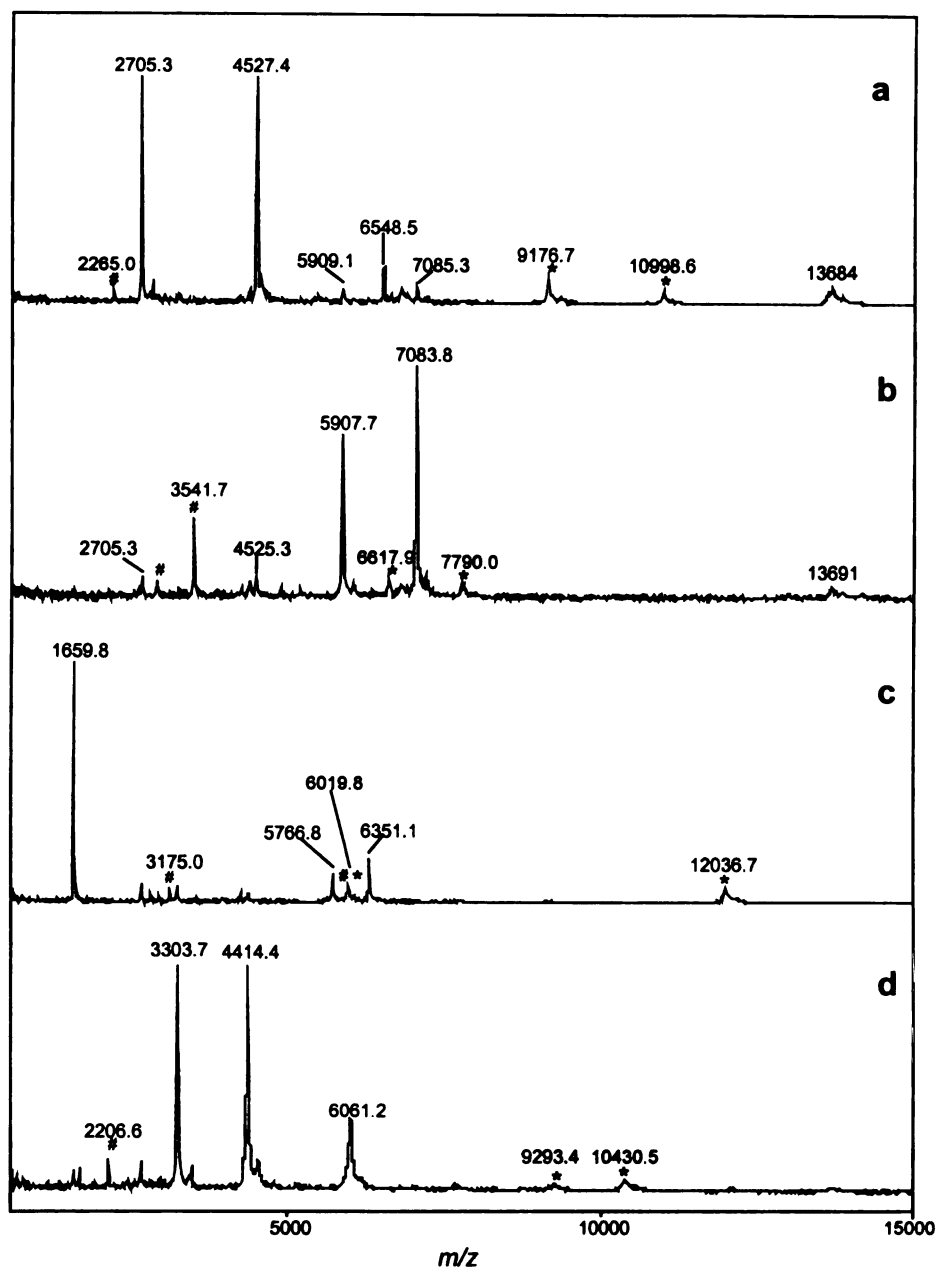


Figure 1.16. The MALDI mass spectra of four peptide mixtures resulting from the CN-induced cleavage of the singly reduced/cyanylated ribonuclease-A isoforms. Spectra **a-d** correspond to the HPLC peaks 1-4 in Figure 1.15, respectively. The symbols # and * represent the doubly charged fragments and protonated β -elimination products, respectively. The double mark #* in spectrum **c** shows a doubly charged β -elimination product (adopted from reference 16).

Figure 1.16a-d are four MALDI-MS spectra of peptide mixtures resulting from CN-induced cleavage of singly reduced/cyanylated isoforms of RnaseA corresponding to HPLC peaks 1-4, respectively, in Figure 1.15. By comparing the experimentally obtained m/z values for the protonated peptide fragments (MH^+) with the calculated values in Table 1.1, assignments of cysteine connectivity can be made. For example, the mass spectrum in Figure 1.16a (three peaks at m/z 2705.3, 6548.5, and 4527.4) agrees well with entry No. 5 in Table 1.1, which corresponds to fragments 1-25, itz-26-83, and itz-84-124, respectively (calculated m/z : 2704.87, 6546.42, and 4526.08) with a relative mass deviation of less than 0.05%. From these data, one can deduce that peptide chain cleavages occur at Cys26 and Cys84. Additionally, the MALDI-MS peak at m/z 9176.7 corresponds to an overlapped peptide, 1- β @26-83 (calculated m/z for MH^+ = 9174.26), resulting from peptide chain cleavage at Cys84, but with β -elimination at Cys26. Likewise, the MALDI peak at m/z 10998.6 is another overlapped peptide, itz-26- β @84-124 (calculated m/z for MH^+ = 10990.45), with cleavage at Cys26, but β -elimination at Cys84. Overall, a disulfide bond between Cys26-Cys84 can be unambiguously deduced.

Using a similar strategy, three other disulfide bond linkages, Cys 65-Cys72, Cys58-Cys110, and Cys40-Cys95, also can be recognized from Figure 1.16b, Figure 1.16c and Figure 1.16d, respectively.

5. Advanced Topic: Combining Partial reduction/Cyanylation/CN-induced Cleavage with Proteolytic Mass Mapping

Proteins that have closely-spaced cysteine residues, are often resistant to proteolysis; even prolonged incubation may produce only big fragments having two or

more intra-disulfide bonds (17) or consisting of three or more smaller digested peptides linked by more than two inter-disulfide bonds (27). We can call these big fragment **proteolytic fragment assemblies**. Analyzing these proteolytic fragment assemblies directly by mass spectrometry may not provide suitable information for deducing the disulfide linkages. In some cases, partial reduction of these large proteolytic fragment assemblies may be used to open one of the disulfide bonds; subsequent cyanylation and CN-induced cleavage of the partially reduced proteolytic fragment assemblies followed by mass mapping could provide useful information about the connectivity of the constituent disulfide bonds.

In one example, a fragment of GM2-activator protein (GM2AP, a protein of 162 amino acids, 8 of them are disulfide cysteines in the form of 4 disulfide bonds), 110-H FC(112)DVLDMLIP TGEPC(125)PEPLR TYGLPC(136)HC(138)PF KE-142, was obtained by *S. aureus* V8 digestion (17). Since there are no free cysteines in the native GM2AP, the presence of this proteolytic fragment indicates that Cys112, Cys125, Cys136, and Cys138 form two disulfide bonds. This proteolytic fragment, GM2AP(110-142), was subjected to partial reduction by TCEP and cyanylation by CDAP. Two singly reduced/cyanylated isoforms, as determined by MALDI-TOF-MS, were isolated (Figure 1.17). The results of analysis by MALDI-TOF-MS also showed that cyanylation was incomplete, because species with two nascent free sulfhydryls cyanylated and with only one cyanylated were both detected (data not shown). Therefore, one would expect singly cleaved as well as doubly cleaved products, and also products of the side-reaction (β -elimination) after CN-induced cleavage in 1 M NH_4OH and complete reduction of the residual disulfide bond. For example, when Cys125-Cys136 was reduced to produce one

isoform, only Cys125 was cyanylated (Cys136 was not cyanylated because of incomplete cyanylation), a singly cleaved product, itz-125-142 (cleaved at cyanylated Cys125, but not at Cys136 because it was not cyanylated), was produced. This was confirmed in the analysis by MALDI-TOF-MS (Table 1.2). From the data in Table 1.2, one can deduce that the two disulfide bonds in GM2AP(110-142) are Cys125-Cys136 and Cys112-Cys138.

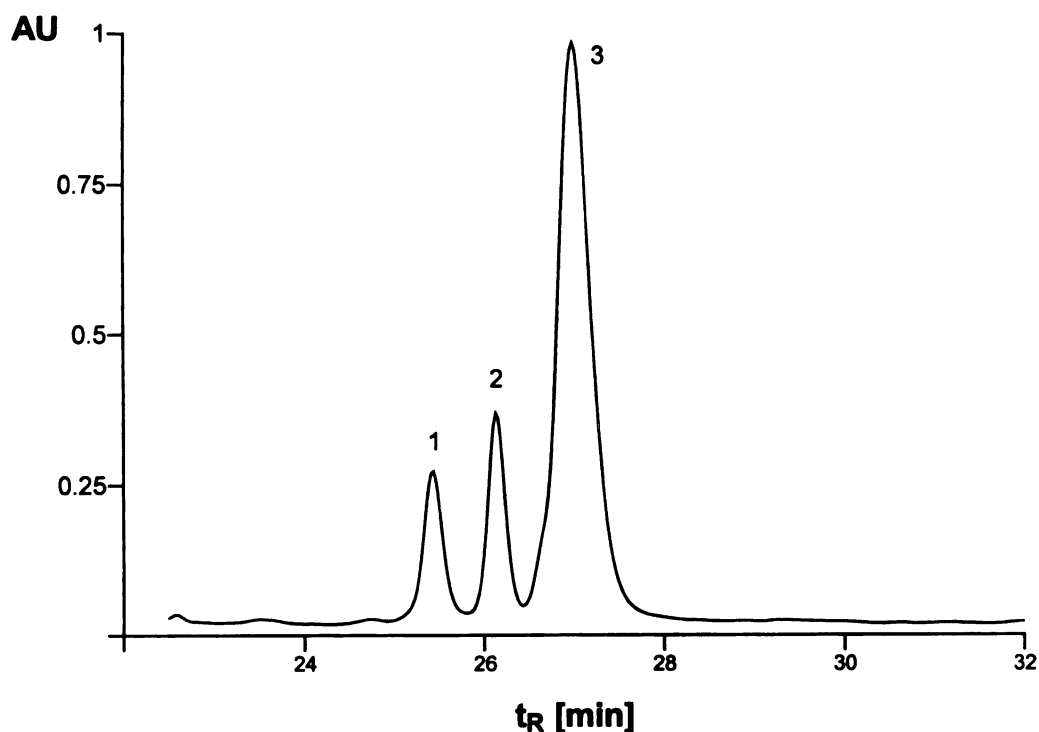


Figure 1.17. C18-RP-HPLC separation of singly reduced and cyanylated isoforms of GM2AP(110-142) (adopted from reference 17).

Table 1.2. m/z values of CN-cleavage products of two singly reduced/cyanylated isoforms GM2AP(110-142) after reduction of the residual disulfide bond (adopted from reference 17).

Reduced/cyanylated Cys125-Cys136(isoform 1)			Reduced/cyanylated Cys112-Cys138(isoform 2)		
CN-cleavage products	Calculated m/z	Experimental ^a m/z	CN-cleavage products	Calculated m/z	Experimental ^a m/z
110-124	1686.79	1686.5	110-111	303.15	-- b
Itz-125-136	1270.63	1269.6	Itz-112-137	2894.33	2893.40
Itz-136-142	888.35	-- b	Itz-138-142	648.28	-- b
110-135 ^c	2913.40	2913.23	110-137 ^c	3153.47	3152.18
Itz-125-142 ^c	2114.96	2114.77	Itz-112-142 ^c	3498.60	3498.04
110- β @125-135	2879.41	2878.30	110- β @112-142	3119.48	3118.48
Itz-125- β @136-142	2080.97	2080.10	Itz-112- β @138-142	3464.62	3464.06

a. Average mass deviation < 500 ppm.

b. Not detected possibly due to signal suppression by matrix molecules used in MALDI.

c. Singly cleaved products from singly reduced but incompletely cyanylated species.

6. Conclusions

The partial reduction/CN-induced cleavage/mass mapping method is simple, specific, and applicable to proteins with adjacent cysteines. With appropriate planning, the cyanylation approach to disulfide mapping can be applied to highly knotted compact systems that are usually resistant to proteolytic attack. Although there has been a preliminary report on the determination of disulfide connectivity from analysis of an intact protein by mass spectrometry (51), it was not elaborated here because the new

technologies of electron capture dissociation and very high resolution ($R = 10^6$) fourier transform ion cyclotron resonance (FTICR)-MS are either in their infancy or not readily available in most laboratories.

V. References

1. Zaluzec, E. J., Gage, D. A. & Watson, J. T. Quantitative assessment of cysteine and cystine in peptides and proteins following organomercurial derivatization and analysis by matrix-assisted laser desorption ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **5**, 359-366 (1994).
2. Schrohenloher, R. & Bennett, J. C. Disulfide Bonds. in *Practical Protein Chemistry -- A Handbook* (ed. Darbre, A.) Vol. 149-163 (John Wiley & Sons, 1986).
3. Smith, D. L. & Sun, Y. Detection and Location of Disulfide Bonds in Proteins by Mass Spectrometry. in *Mass Spectrometry of Peptides* (ed. Desiderio, D. M.) Vol. 275-287 (CRC Press, 1991).
4. Aitken, A. Analysis of cysteine residues and disulfide bonds. in *Methods Mol. Biol. (Totowa, N. J.)* (ed. Walker, J. M.) Vol. **32**, 351-360 (Humana Press, 1994).
5. Sun, Y., Bauer, M. D., Keough, T. W. & Lacey, M. P. Disulfide bond location in proteins. in *Methods Mol. Biol. (Totowa, N. J.)* (ed. Chapman, J. R.) Vol. **61**, 185-210 (Humana Press, 1996).
6. Gray, W. R. Disulfide bonds between cysteine residues. in *Protein Struct. (2nd Ed.)* (ed. Creighton, T. E.) Vol. 165-186 (IRL Press at Oxford University, New York, 1997).
7. Gray, W. R. Disulfide structures of highly bridged peptides: a new strategy for analysis. *Protein Sci.* **2**, 1732-1748. (1993).
8. Gray, W. R. Echistatin disulfide bridges: selective reduction and linkage assignment. *Protein Sci.* **2**, 1749-1755. (1993).
9. Heck, S. D., Kelbaugh, P. R., Kelly, M. E., Thadeio, P. F., Saccomano, N. A., Stroh, J. G. & Volkmann, R. A. Disulfide Bond Assignment of omega-Agatoxins IVB and IVC: Discovery of a D-Serine Residue in omega-Agatoxin IVB. *J. Am. Chem. Soc.* **116**, 10426-10436 (1994).

10. Li, F. & Liang, S. P. Assignment of the three disulfide bonds of Selenocosmia huwena lectin-I from the venom of spider Selenocosmia huwena. *Peptides* **20**, 1027-1034 (1999).
11. Daquinag, A. C., Sato, T., Koda, H., Takao, T., Fukuda, M., Shimonishi, Y. & Tsukamoto, T. A Novel Endogenous Inhibitor of Phenol oxidase from Musca domestica Has a Cystine Motif Commonly Found in Snail and Spider Toxins. *Biochemistry* **38**, 2179-2188 (1999).
12. Bures, E. J., Hui, J. O., Young, Y., Chow, D. T., Katta, V., Rohde, M. F., Zeni, L., Rosenfeld, R. D., Stark, K. L. & Haniu, M. Determination of Disulfide Structure in Agouti-Related Protein (AGRP) by Stepwise Reduction and Alkylation. *Biochemistry* **37**, 12172-12177 (1998).
13. Leal, W. S., Nikonova, L. & Peng, G. Disulfide structure of the pheromone binding protein from the silkworm moth, Bombyx mori. *FEBS Lett.* **464**, 85-90 (1999).
14. Young, Y., Zeni, L., Rosenfeld, R. D., Stark, K. L., Rohde, M. F. & Haniu, M. Disulfide assignment of the C-terminal cysteine knot of agouti-related protein (AGRP) by direct sequencing analysis. *J. Pept. Res.* **54**, 514-521 (1999).
15. Yamashita, H., Nakatsuka, T. & Hirose, M. Structural and functional characteristics of partially disulfide-reduced intermediates of ovotransferrin N lobe. Cystine localization by indirect end-labeling approach and implications for the reduction pathway. *J. Biol. Chem.* **270**, 29806-29812 (1995).
16. Wu, J. & Watson, J. T. A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci.* **6**, 391-398 (1997).
17. Schutte, C. G., Lemm, T., Glombitza, G. J. & Sandhoff, K. Complete localization of disulfide bonds in GM2 activator protein. *Protein Sci.* **7**, 1039-1045 (1998).
18. Yang, Y., Wu, J. & Watson, J. T. Disulfide mass mapping in proteins containing adjacent cysteines is possible with cyanylation/cleavage methodology. *J. Am. Chem. Soc.* **120**, 5834-5835 (1998).
19. Yang, Y., Wu, J. & Watson, J. T. Probing the folding pathways of long R3 insulin-like growth factor-I (LR3IGF-I) and IGF-I via capture and identification of disulfide intermediates by cyanylation methodology and mass spectrometry. *J. Biol. Chem.* **274**, 37598-37604 (1999).
20. Wu, J., Yang, Y. & Watson, J. T. Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci.* **7**, 1017-1028 (1998).

21. Billeci, T. M. & Stults, J. T. Tryptic mapping of recombinant proteins by matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Chem.* **65**, 1709-1716. (1993).
22. Chen, X., Chen, Y. H. & Anderson, V. E. Protein cross-links: universal isolation and characterization by isotopic derivatization and electrospray ionization mass spectrometry. *Anal. Chem.* **273**, 192-203. (1999).
23. Adamczyk, M., Gebler, J. C. & Wu, J. A simple method to identify cysteine residues by isotopic labeling and ion trap mass spectrometry. *Rapid Commun. Mass Spectrom.* **13**, 1813-1817 (1999).
24. Zhou, J., Ens, W., Poppe-Schriemer, N., Standing, K. G. & Westmore, J. B. Cleavage of interchain disulfide bonds following matrix-assisted laser desorption. *Int. J. Mass Spectrom. Ion Processes* **126**, 115-122 (1993).
25. Patterson, S. D. & Katta, V. Prompt fragmentation of disulfide-linked peptides during matrix-assisted laser desorption ionization mass spectrometry. *Anal. Chem.* **66**, 3727-3732 (1994).
26. Crimmins, D. L., Saylor, M., Rush, J. & Thoma, R. S. Facile, in situ matrix-assisted laser desorption ionization-mass spectrometry analysis and assignment of disulfide pairings in heteropeptide molecules. *Anal. Biochem.* **226**, 355-361 (1995).
27. Jones, M. D., Hunt, J., Liu, J. L., Patterson, S. D., Kohno, T. & Lu, H. S. Determination of Tumor Necrosis Factor Binding Protein Disulfide Structure: Deviation of the Fourth Domain Structure from the TNFR/NGFR Family Cysteine-Rich Region Signature. *Biochemistry* **36**, 14914-14923 (1997).
28. Jones, M. D., Patterson, S. D. & Lu, H. S. Determination of Disulfide Bonds in Highly Bridged Disulfide-Linked Peptides by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry with Postsource Decay. *Anal. Chem.* **70**, 136-143 (1998).
29. Gorman, J. J., Ferguson, B. L., Speelman, D. & Mills, J. Determination of the disulfide bond arrangement of human respiratory syncytial virus attachment (G) protein by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Protein Sci.* **6**, 1308-1315 (1997).
30. Merewether, L. A., Le, J., Jones, M. D., Lee, R., Shimamoto, G. & Lu, H. S. Development of Disulfide Peptide Mapping and Determination of Disulfide Structure of Recombinant Human Osteoprotegerin Chimera Produced in *Escherichia coli*. *Arch. Biochem. Biophys.* **375**, 101-110 (2000).

31. Bean, M. F. & Carr, S. A. Characterization of disulfide bond position in proteins and sequence analysis of cystine-bridged peptides by tandem mass spectrometry. *Anal. Biochem.* **201**, 216-226 (1992).
32. Bauer, M., Sun, Y., Degenhardt, C. & Kozikowski, B. Assignment of all four disulfide bridges in echistatin. *J. Protein Chem.* **12**, 759-764 (1993).
33. Badock, V., Raida, M., Adermann, K., Forssmann, W.-G. & Schrader, M. Distinction between the three disulfide isomers of guanylin 99-115 by low-energy collision-induced dissociation. *Rapid Commun. Mass Spectrom.* **12**, 1952-1956 (1998).
34. Yen, T. Y., Joshi, R. K., Yan, H., Seto, N. O., Palcic, M. M. & Macher, B. A. Characterization of cysteine residues and disulfide bonds in proteins by liquid chromatography/electrospray ionization tandem mass spectrometry. *J. Mass Spectrom.* **35**, 990-1002. (2000).
35. Holmes, E. H., Yen, T.-Y., Thomas, S., Joshi, R., Nguyen, A., Long, T., Gallet, F., Maftah, A., Julien, R. & Macher, B. A. in *J. Biol. Chem.* 24237-24245 (2000).
36. Smith, D. L. & Zhou, Z. R. Strategies for locating disulfide bonds in proteins. in *Methods in Enzymology* (ed. McCloskey, J. A.) Vol. **193**, 374-389. (Academic Press, New York, 1990).
37. Sechi, S. & Chait, B. T. Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification. *Anal. Chem.* **70**, 5150-5158. (1998).
38. Hara, S., Katta, V. & Lu, H. S. Peptide map procedure using immobilized protease cartridges in tandem for disulfide linkage identification of neu differentiation factor epidermal growth factor domain. *J. Chromatogr. A* **867**, 151-160. (2000).
39. Sturrock, E. D., Yu, X. C., Wu, Z., Biemann, K. & Riordan, J. F. Assignment of free and disulfide-bonded cysteine residues in testis angiotensin-converting enzyme: functional implications. *Biochemistry* **35**, 9560-9566. (1996).
40. Benham, C. J. & Jafri, M. S. Disulfide bonding patterns and protein topologies. *Protein Sci.* **2**, 41-54 (1993).
41. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601 (1984).
42. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* **16**, 99-111. (1988).

43. John, H. & Forssmann, W. G. Determination of the disulfide bond pattern of the endogenous and recombinant angiogenesis inhibitor endostatin by mass spectrometry. *Rapid Commun. Mass Spectrom.* **15**, 1222-1228 (2001).
44. van den Hooven, H. W., van den Burg, H. A., Vossen, P., Boeren, S., de Wit, P. & Vervoort, J. Disulfide bond structure of the AVR9 elicitor of the fungal tomato pathogen *Cladosporium fulvum*: Evidence for a cystine knot. *Biochemistry* **40**, 3458-3466 (2001).
45. Wakselman, M. & Guibe-Jampel, E. 1-Cyano-4-dimethylamino-pyridinium salts: New water-soluble reagents for the cyanylation of protein sulphydryl groups. *J. C. S. Chem Commun.*, 21-22 (1976).
46. Jacobson, G. R., Schaffer, M. H., Stark, G. R. & Vanaman, T. C. Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues. *J. Biol. Chem.* **248**, 6583-6591. (1973).
47. Wu, J. & Watson, J. T. Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping. *Anal. Biochem.* **258**, 268-276 (1998).
48. Qi, J. F., Wu, J., Somkuti, G. A. & Watson, J. T. Determination of the disulfide structure of sillicin, a highly knotted, cysteine-rich peptide, by cyanylation/cleavage mass mapping. *Biochemistry* **40**, 4531-4538 (2001).
49. Wu, J., Gage, D. A. & Watson, J. T. A strategy to locate cysteine residues in proteins by specific chemical cleavage followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* **235**, 161-174 (1996).
50. Smyth, D. G., Stein, W. H. & Moore, S. The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations. *J. Biol. Chem.* **238**, 227-234. (1963).
51. Zubarev, R. A., Kruger, N. A., Fridriksson, E. K., Lewis, M. A., Horn, D. M., Carpenter, B. K. & McLafferty, F. W. Electron Capture Dissociation of Gaseous Multiply-Charged Proteins Is Favored at Disulfide Bonds and Other Sites of High Hydrogen Atom Affinity. *J. Am. Chem. Soc.* **121**, 2857-2862 (1999).

CHAPTER 2

DEVELOPMENT OF MORE FLEXIBLE PROTOCOLS FOR THE PARTIAL REDUCTION AND SUBSEQUENT CYANYLATION/CN-INDUCED CLEAVAGE/MASS MAPPING METHODOLOGY

I. Introduction

As described in Chapter 1, identification of a particular disulfide bond can be achieved by detecting an anticipated array of fragments that derive from the CN-induced cleavage of its corresponding singly reduced/cyanylated isoform and complete reduction of the residual disulfide bonds (Figure 1.14 in Chapter 1). In principle, partial reduction of a protein with n disulfide bonds can produce as many as n singly reduced isoforms. If the determination of cysteine status is based solely on singly reduced isoforms, at least $n-1$ of them must be isolated and analyzed to deduce the linkage of $n-1$ disulfide bonds; the remaining disulfide can be identified by default. This strategy has been successfully applied to several proteins, such as ribonuclease-A (1) and long insulin-like growth factor-1 (LR³IGF-1) (2). However, certain proteins, especially those with a cysteine-rich knotted arrangement, are resistant to reduction, making it difficult, if not impossible, to prepare even as many as $n-1$ of the n possible (predicted) singly reduced isoforms (3). In these situations, doubly reduced isoforms of the original protein (as well as the available singly reduced isoforms) may be used to solve the disulfide structure of the protein. In addition to n singly reduced isoforms, $n*(n-1)/2$ doubly reduced isoforms are possible for a protein containing n disulfide bonds. The possibility of using both doubly reduced and singly reduced isoforms, provides flexibility in applying the chemistry of partial

reduction, and increases the possibility of successful identification of the linkages of all the disulfide bonds in the protein.

In the previous proposed disulfide mass mapping methodology based on partial reduction and cyanylation/CN-induced cleavage (1), the mixture of CN-induced fragments were analyzed by mass spectrometry only after the complete reduction of the residual disulfide bonds that are either within the fragments or linking some of the fragments together (Figure 1.14 in Chapter 1). If a CN-induced fragment contains one or more intra-molecular disulfide bonds, it is defined as an **oxidized CN-induced fragment**; if two or more CN-induced fragments are linked by one or more inter-molecular disulfide bonds, such an assembly is defined as an **oxidized CN-induced fragment assembly**. If there are no disulfide bonds in a CN-induced fragment (all the cysteine residues within the fragment are free cysteines), the CN-induced fragment is defined as a **reduced CN-induced fragment**. An additional analysis step of the oxidized CN-induced fragments or oxidized CN-induced fragment assemblies by mass spectrometry prior to the complete reduction step may provide extra information about the linkages of the residual disulfide bonds within various oxidized CN-induced fragments or CN-induced fragment assemblies.

Preliminary data (not shown) indicated that CN-induced cleavage on the N-terminal side of the cyanylated cysteine residues might produce CN-induced fragments with either an amide or a free acid C-terminus. The mass of a particular CN-induced fragment having an amide C-terminus is one dalton lower than the one with a free acid C-terminus. Therefore, the status of the terminus (amide or free acid) for CN-induced fragments need to be further investigated.

Iminothiazolidine (itz)-blocked peptides constitute the majority of the final products from the cyanylation/CN-induced cleavage reactions. In the disulfide mass mapping methodology based on cyanylation and CN-induced cleavage, the connectivity of cysteines in the disulfide bonds is deduced from the mass of cleavage products (itz-peptides) expected for a particular partially reduced species (a singly, doubly, or triply reduced species) of the native protein. The identification of the CN-induced fragments (mostly itz-peptides) is usually based on the matching of an experimental mass spectral peak and the corresponding calculated values. In most cases, this is successful because each of the possible itz-peptides has a unique mass. However, we may run into cases of isomass, where two CN-induced fragments have the same or very close masses, but they have different sequences. In those cases, we need to obtain additional evidence to assign the correct fragment to the experimental mass spectral peak. Collision induced dissociation (CID) mass spectrometry/mass spectrometry (MS/MS) of itz-fragments was investigated to determine if it can provide some evidence to distinguish a particular itz-fragment from its isomass isomers.

All of the aspects introduced above are described in detail in the following sections.

II. Materials and Methods

Chemicals

Bovine pancreatic ribonuclease-A type III-A was purchased from Sigma. Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride was purchased from Pierce Chemical Co. (Rockford, IL). Guanidine hydrochloride is a product of Boehringer-Mannheim Biochemicals (Indianapolis, IN). 1-Cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate was purchased from Sigma. Water, acetonitrile, and trifluoroacetic acid (TFA) were of HPLC grade. TCEP and CDAP solutions were freshly prepared in 0.1 M sodium citrate buffer (pH 3.0) prior to use. Because of the instability of CDAP in aqueous solution, solid CDAP was dissolved in pH 3 sodium citrate buffer just before use.

Optimize Partial Reduction to Produce Doubly Reduced/Cyanylated Ribonuclease-A Isoforms

Approximately one equivalent of TCEP for the total disulfide bond content in the proteins was used for the conditions optimized for producing singly reduced/cyanylated protein isoforms (see reference 1 for details). However, a 10-fold equivalent excess of TCEP over the disulfide content in ribonuclease-A was used to increase the yield of doubly reduced/cyanylated isoforms. A solution containing 10 nmol ribonuclease-A was prepared in 10 μ L of 0.1 M pH 3.0 sodium citrate buffer containing 6 M guanidine hydrochloride. Partial reduction of ribonuclease-A was carried out by adding 400 nmol

of TCEP (4 μ L of 0.1 M TCEP solution), followed by incubation at room temperature for 15 minutes.

To the TCEP reaction mixture containing partially reduced ribonuclease-A species were added 2000 nmol of CDAP (20 μ L of 0.1 M CDAP solution). Cyanylation of the nascent free cysteines was accomplished during incubation at room temperature for 15 minutes.

HPLC Separation of Partially Reduced and Cyanylated Protein Species

Mixtures of partially-reduced and cyanylated species were separated by reversed phase HPLC using a linear gradient elution of 20%-40% B in 90 minutes (B is 90% (v/v) acetonitrile/0.1% TFA; A is 0.1% TFA in water.) with Waters model 6000 pumps. UV detection was done at 215 nm. A Vydac C18 (#218TP54, 5- μ m particle size, 300-Å pores in a 4.6 $\times$ 250 mm) column was used. The predominant HPLC fractions were collected manually and the masses of the collected species in each fraction (approximately 0.8~1.0 mL) were determined by MALDI-MS.

CN-induced Cleavage of Partially Reduced and Cyanylated Protein Species

A solution of a dried HPLC fraction was reconstituted in 2 μ L of 6 M guanidine hydrochloride in 1 M aqueous ammonia to dissolve the partially reduced and cyanylated ribonuclease-A species; to this solution was added another 8 μ L of 1 M aqueous ammonia. CN-induced cleavage of the peptide bond on the N-terminal side of cyanylated cysteine residues was accomplished during reaction at room temperature for one hour. Excess ammonia was removed in a vacuum system. Truncated peptides, still linked in

some cases by residual disulfide bonds depending on the extent of partial reduction, were dissolved in 10 μ L of water; 1 μ L of this solution was further diluted with 9 μ L of 50% (v/v) acetonitrile/0.1% TFA, and analyzed by MALDI-MS. The other 9 μ L of the solution were dried in a vacuum system for future use.

Complete Reduction of the Residual Disulfide Bonds

Truncated peptides, from the CN-induced cleavage reaction, some still linked by residual disulfide bonds, were completely reduced by reaction with 200 nmol of TCEP (2 μ L of 0.1 M TCEP solution) at 37 °C for 30 minutes. Samples were then diluted with 90 μ L of 50% (v/v) aqueous acetonitrile/0.1% TFA solution prior to analysis by MALDI-MS.

Production of CN-Induced Fragments from Sillucin

The details of producing the CN-induced fragments (from sillucin) that are used in section V can be found in the “Materials and Methods” section of Chapter 3.

MALDI-MS

MALDI mass spectra were obtained on a Voyager Elite or Voyager DE-STR time-of-flight (TOF) mass spectrometer (Perkin Elmer Biosystems Inc., Framingham, MA) equipped with a 337-nm nitrogen laser. The accelerating voltage in the ion source was set at 25 kV. Except for accurate mass measurement of intact sillucin that was done in positive reflectron mode, all other data were acquired in the positive linear mode of operation. Protein standards of bovine bradykinin (monoisotopic mass of the protonated

molecule (MH^+): 1060.569 Da; average mass of MH^+ : 1061.217 Da), bovine pancreatic insulin chain B oxidized (monoisotopic mass of MH^+ : 3494.651 Da; average mass of MH^+ : 3496.903 Da), bovine insulin (monoisotopic mass of MH^+ : 5730.609 Da; average mass of MH^+ : 5734.6 Da) and Equine Apomyoglobin (average mass of MH^+ : 16952.6 Da) obtained from Sigma Chemical Co. (St. Louis, MO). All experiments were performed using α -cyano-4-hydroxycinnamic acid (Aldrich Chemical co., Milwaukee, WI) as the matrix. Saturated matrix solutions were prepared in a 50% (V/V) aqueous solution of acetonitrile/0.1% TFA. Just prior to analysis, 0.5 μ L of matrix solution was applied to the stainless steel sample plate and allowed to air-dry; then 1 μ L of sample solution was added to the top of it, and finally another 1 μ L of matrix solution was added to the same spot. The mixture was allowed to air-dry before being introduced into the mass spectrometer.

CID-MS/MS of itz-peptides

The itz-peptides resulting from the CN-induced cleavage reaction were separated from salts and reagents by HPLC, and then analyzed by infusion on a Finnigan LCQ-Deca mass spectrometer. The solution of the itz-peptides usually contains 20-40% acetonitrile and 0.5% acetic acid. CID was performed at 60% relative collision energy.

III. Expand Partial Reduction Strategy to Include Doubly Reduced Protein

Isoforms

As discussed in the “Introduction” section, doubly reduced/cyanylated protein isoforms are need to be explored if not enough number of singly reduced/cyanylated protein isoforms are obtained. Mass mapping a doubly reduced/cyanylated isoform of a protein followed by CN-induced cleavage and complete reduction can determine which four cysteine residues form two disulfide bonds (Figure 2.1). This information can be used along with other information from the mapping experiment of another doubly or a singly reduced/cyanylated isoform (Figure 2.1 and Figure 2.2).

In order to identify which four cysteine residues form two disulfide bonds (that had been reduced and cyanylated) in a particular doubly reduced/cyanylated species, the mass spectral data should be compared against all sets of calculated masses of the CN-induced fragments expected from all possible doubly reduced/cyanylated isoforms after the chemical treatment represented in Figure 2.1. The strategies for processing the data from the CN-induced fragments of a doubly reduced/cyanylated isoform is described in Chapter 5.

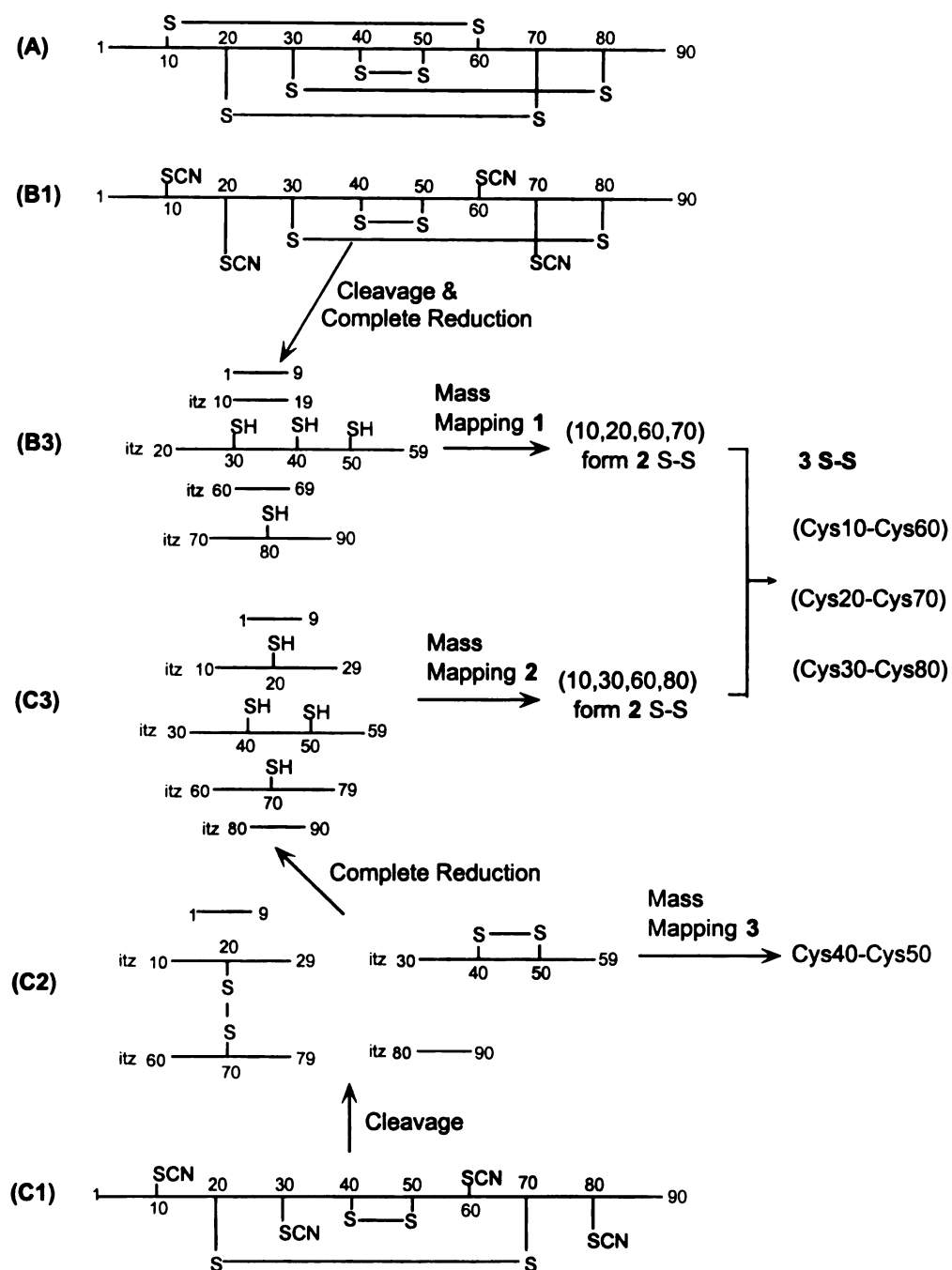


Figure 2.1. Overview of chemical reactions involved in the disulfide mass mapping of two doubly reduced/cyanylated isoforms (B1) and (C1) of a hypothetical protein (A). (C2): the cleaved fragments from (C1) before complete reduction; (B3) and (C3): the corresponding cleaved fragments from (B1) and (C1) after complete reduction.

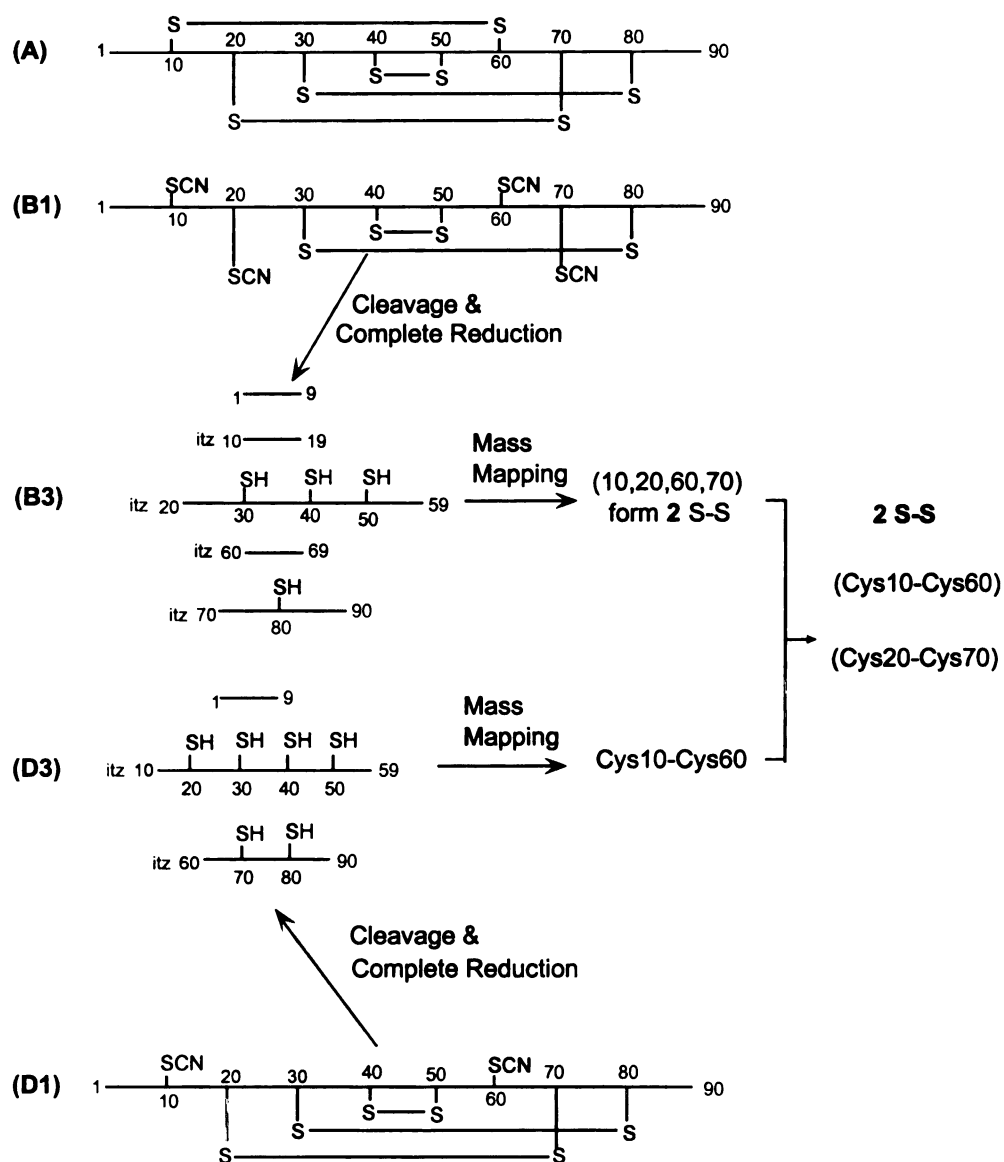


Figure 2.2. Disulfide mass mapping using a doubly reduced (**B1**) and a singly reduced/cyanylated (**D1**) species of a hypothetical protein (**A**) containing four disulfide bonds. (**B3**) and (**D3**): the corresponding cleaved fragments from (**B1**) and (**D1**) after complete reduction.

In experiments using ribonuclease-A as a model compound, we have found that the partial reduction conditions can be adjusted principally via changing the concentration of the TCEP, the reducing agent, to optimize the yield of doubly reduced isoforms. The expected number of doubly reduced isoforms of ribonuclease-A, which has four disulfide bonds, is six. (The formula for computing the number of doubly reduced isoforms for a protein containing n disulfide bonds is $n(n-1)/2$.) Figure 2.3 is the HPLC chromatogram of singly, doubly, and triply reduced/cyanylated isoforms of ribonuclease-A. It is instructive to compare Figure 2.3 with Figure 1.15 in Chapter 1, which is the HPLC chromatogram of the partially reduced mixture resulting from treatment of ribonuclease-A under conditions optimized for production of singly reduced/cyanylated isoforms. Under conditions optimized for producing singly reduced/cyanylated isoforms, four singly reduced/cyanylated isoforms were produced for ribonuclease-A (Figure 1.15 in Chapter 1); whereas under conditions optimized for producing doubly reduced/cyanylated isoforms, six doubly reduced/cyanylated isoforms were produced in addition to the four singly reduced/cyanylated isoforms.

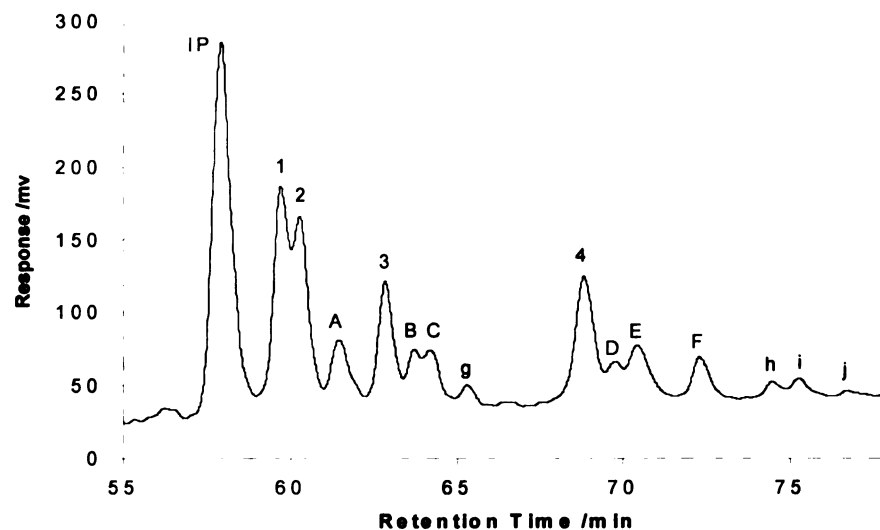


Figure 2.3. The HPLC chromatogram of partially reduced/cyanylated species of ribonuclease-A: **IP** represents the intact ribonuclease-A; Peaks **1-4** represent the singly reduced/cyanylated isoforms (as also shown in Figure 1.15 in Chapter 1); Peaks **A-F** represent the doubly reduced/cyanylated isoforms; Peaks **g-j** represent the triply reduced/cyanylated isoforms, as determined by MALDI-MS.

Figure 2.4 is the MALDI mass spectrum of the cleavage products of one of the partially reduced/cyanylated species represented by HPLC peak **B** in Figure 2.3. From Table 2.1, we can see that four MALDI peaks at m/z 4412.2, 2712.3, 2646.7 and 3302.7 are due to fragments 1-39, itz-40-64, itz-72-94, and itz-95-124, respectively. From these data, one can deduce that CN-induced cleavages occurred at cyanylated cysteines, Cys40, Cys65, Cys72 and Cys95; and thus, these four cysteine residues form **two** disulfide bonds in ribonuclease-A. In addition, the MALDI peak at m/z 7050.5 corresponds to the fragment, 1-(β @40)-64, resulting from cleavage at cyanylated Cys65, but β -elimination at cyanylated Cys40. Likewise, the MALDI peak at m/z 5874.8 is another such fragment, itz-72-(β @95)-124, resulting from cleavage at cyanylated Cys72, but β -elimination at cyanylated Cys95. These two fragments provide additional corroborative information that the four cysteines, Cys40, Cys65, Cys72 and Cys95, form two disulfide bonds in ribonuclease-A.

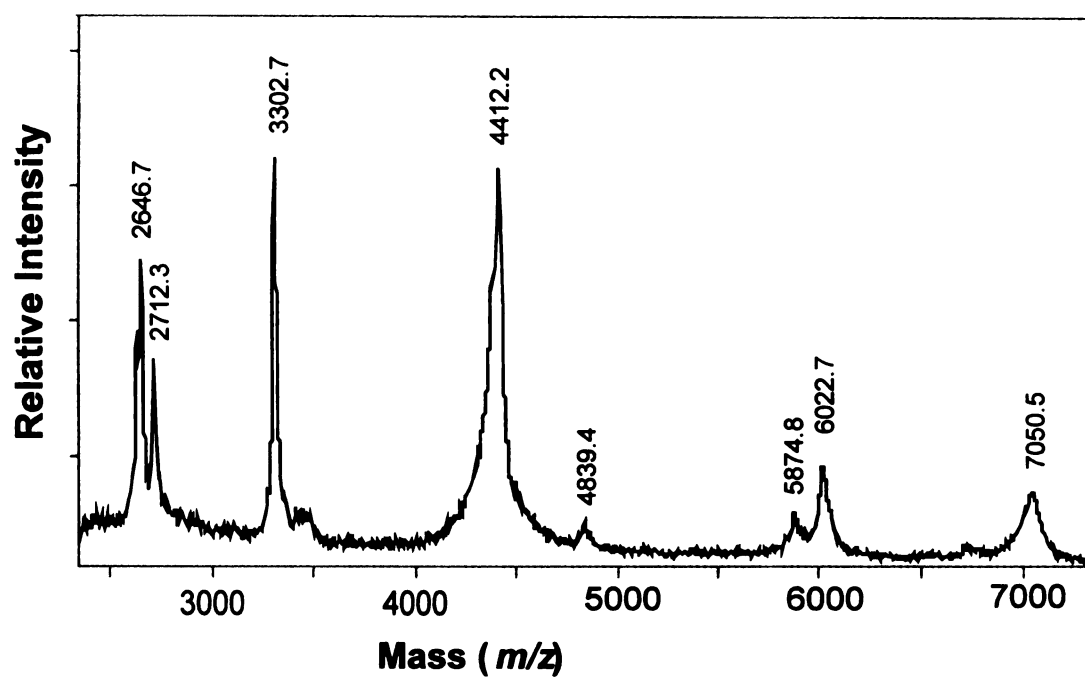


Figure 2.4. The MALDI mass spectrum of a peptide mixture resulting from the cleavage of the doubly reduced/cyanylated ribonuclease-A isoform corresponding to HPLC peak B in Figure 2.3.

Table 2.1. *m/z* values for possible fragments resulting from the cleavage reaction of one of the doubly reduced/cyanylated isoform of ribonuclease-A represented by HPLC peak B in Figure 2.3.

Fragment	Experimental <i>m/z</i>	Calculated MH ⁺ /Da	Cyanylated Cysteines in the doubly reduced/cyanylated isoform
itz-72-94	2646.7	2646.9	40, 65, 72, 95
itz-40-64	2712.3	2713.1	40, 65, 72, 95
itz-95-124	3302.7	3302.7	40, 65, 72, 95
1-39	4412.2	4411.9	40, 65, 72, 95
itz-40-83	4839.4	4839.4	26, 40, 84, 95
itz-72-(β@84)-124	5874.8	5872.6	40, 65, 72, 95
itz-40-(β@84)-94	6022.7	6028.7	26, 40, 84, 95
1-(β@40)-64	7050.5	7047.9	40, 65, 72, 95

Using the same method to analyze the MALDI spectra (data not shown) of the CN-induced cleavage products from another doubly reduced/cyanylated isoform represented by HPLC peak C in Figure 2.3, one can deduce that Cys26, Cys40, Cys84 and Cys95 form **two** disulfide bonds in ribonuclease-A. With the above information about HPLC peak B and C (Figure 2.3), one can deduce that three disulfide bonds, Cys26-Cys84, Cys40-Cys95 and Cys65-Cys72 exist in a ribonuclease-A molecule.

The HPLC peaks B and C in Figure 2.3 were not resolved completely. Careful manual collection of HPLC fractions from one component still contained a small amount of the other component. This is reflected in the MALDI mass spectrum (Figure 2.4) of the CN-induced cleavage products of HPLC fraction B; two MALDI peaks m/z 4839.4 and 6022.7 are due to CN-induced fragments itz-40-83 and itz-40-(β @84)-94, CN-induced cleavage products from another doubly reduced/cyanylated isoform represented by HPLC peak C in Figure 2.3. The four cyanylated cysteines (from reduction of two disulfide bonds and the subsequent cyanylation) in this isoform are listed in bold font in Table 2.1.

Using the strategy described above, we have successfully correlated the disulfide bond connectivity in ribonuclease-A with the masses of the CN-induced fragments from cyanylation and cleavage of the doubly reduced isoforms (peak A in Figure 2.3 represents the isoform resulting from reduction of the disulfide bonds, Cys26-Cys84 and Cys65-Cys72; peak B represents the isoform resulting from reduction of the disulfide bonds, Cys40-Cys95 and Cys65-Cys72; peak C represents the isoform resulting from reduction of the disulfide bonds, Cys26-Cys84 and Cys40-Cys95; peak D represents the disulfide bonds from reduction between Cys26-Cys84 and Cys58-Cys110; peak E represents that

from reduction of Cys58-Cys110 and Cys65-Cys72; peak F represents the isoform resulting from reduction of Cys40-Cys95 and Cys58-Cys110) by analysis of the CN-induced cleavage products by MALDI-MS. Whereas, MALDI data corresponding to the materials represented by peaks g-j indicated (by a mass shift of 156 Da from the molecular weight of ribonuclease-A) that these are triply reduced/cyanylated isoforms (of which, there are predicted mathematically to be four) of ribonuclease-A.

IV. Addition of an Analysis Step by Mass Spectrometry after CN-induced Cleavage, but Before Complete Reduction of the Residual Disulfide Bonds

As described in the “Introduction” section, an additional analysis step of the oxidized CN-induced fragments or assemblies by mass spectrometry prior to the complete reduction step may provide extra information about the linkages of the residual disulfide bonds within various oxidized CN-induced cleavage fragments or assemblies. For example, as shown in Figure 2.1, mass mapping analysis 3 is accomplished before complete reduction of the residual disulfide bonds within each fragment of the CN-induced cleavage reaction mixture, and thus the detected presence of CN-induced cleavage fragment itz-30-59-ox (“ox” designates that there are residual disulfide bonds within a CN-induced cleavage fragment.) indicates the presence of the disulfide Cys40-Cys50. This observation demonstrated that mass mapping analysis done **after CN-induced cleavage (but before complete reduction)** can provide additional useful connectivity information about the residual disulfide bonds (those not reduced during the partial reduction step). Another important observation is that the disulfide information obtained here is not available through mass mapping analysis after the complete

reduction step. This means that this additional mass mapping step may help to solve the disulfide structure in some situations, where fewer than enough partially reduced isoforms are produced. Utilization of this additional analysis by mass spectrometry before the complete reduction step is described in the determination of the disulfide structure of sillucin (Chapter 3) and in the determination of the cysteine status of the glyceraldehyde-3-P dehydrogenase subunit-A (Chapter 4).

If there is only one inter-disulfide bond in an oxidized CN-induced fragment assembly (defined in the “Introduction” section), detection of such an oxidized CN-induced fragment assembly will also allow us to identify the linkage of the inter-disulfide bond. Such an example is shown in Figure 2.5. The MALDI-MS peak at m/z 6012.2 corresponds to an oxidized CN-induced fragment assembly, itz40-64-(58S-S110)-itz-95-124 (“58S-S110” designates the disulfide bond between Cys 58 and Cys 110; calculated mass for MH^+ of the assembly is 6013.8 Da). This provided the evidence for disulfide bond Cys58-Cys110. In addition, prompt fragmentation (references 4-8) of the CN-induced fragment assembly into two CN-induced fragments, itz-40-64 and itz-95-124 during analysis by MALDI, further confirmed the existence of the disulfide bond Cys58-Cys110.

In other cases, mass mapping analysis done after CN-induced cleavage (but before complete reduction) may not provide information by which to identify a disulfide bond. However, such results can help eliminate some possible disulfide structures (Figure 2.6). One example of such an application is shown in Figure 2.7.

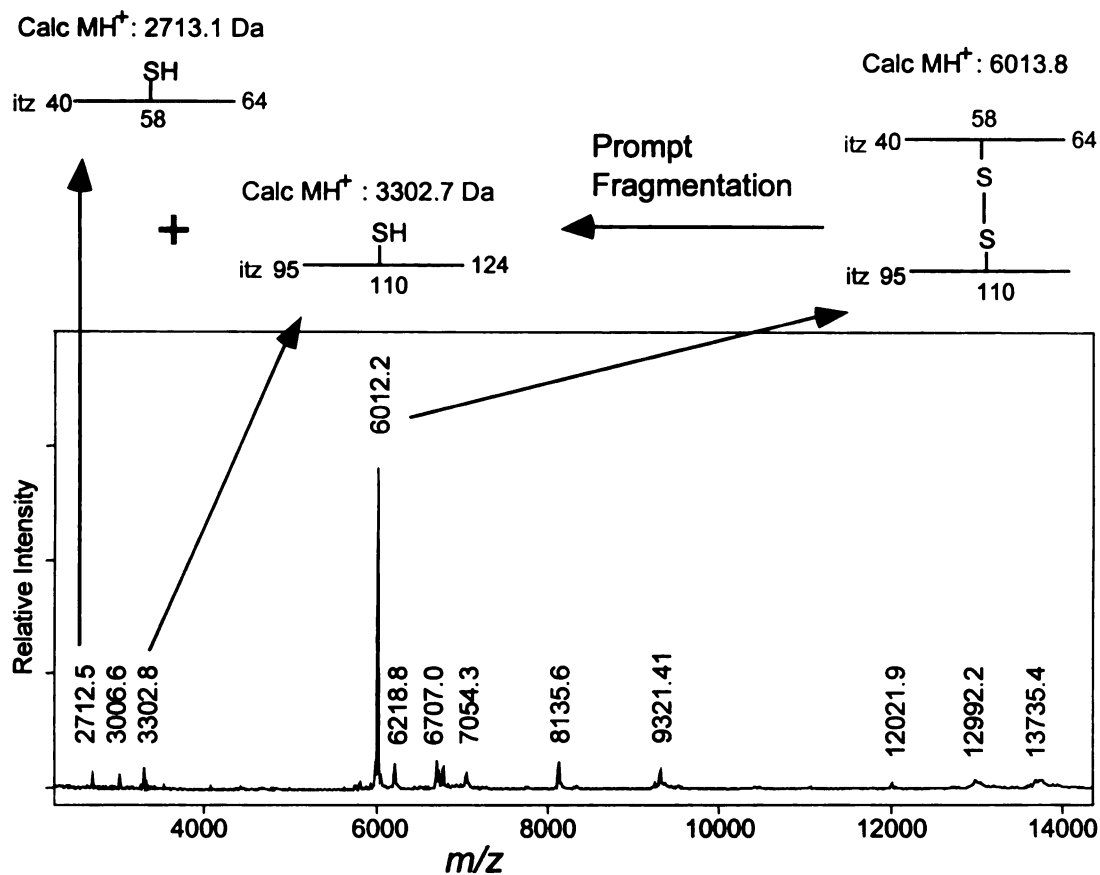


Figure 2.5. The MALDI mass spectrum of the peptide mixture resulting from CN-induced cleavage (but **before complete reduction**) of the doubly reduced/cyanylated ribonuclease-A isoform represented by HPLC peak B in Figure 2.3. Itz-40-64 represents the itz-blocked fragment of ribonuclease-A from residue 40 to 64. Details about prompt fragmentation (or in-source fragmentation) during analysis by MALDI-MS can be found in references 4-8.

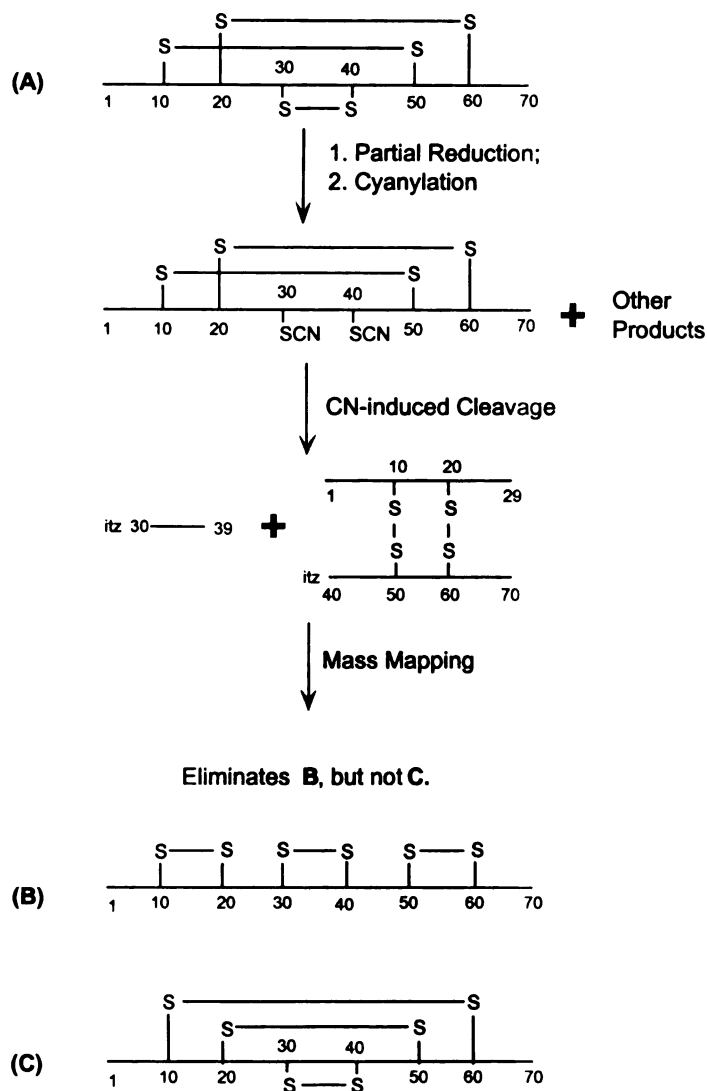


Figure 2.6. Eliminating a possible disulfide structure (**B**) by mass mapping analysis of the CN-cleavage products **before** the complete reduction step. Disulfide structure (**A**) shows the correct disulfide linkages in the hypothetical protein; structures (**B**) and (**C**) are two of 15 possible structures of arranging six cysteines as three disulfide bonds in such a protein.

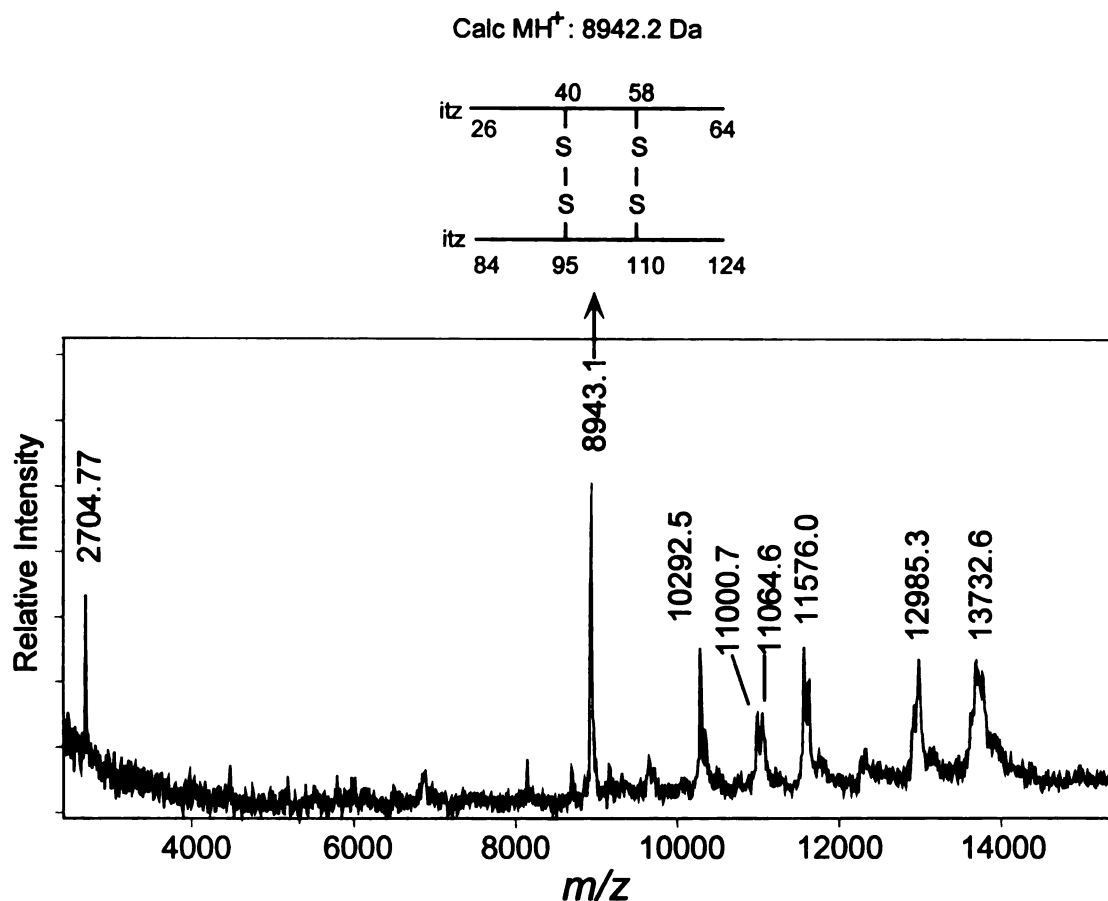


Figure 2.7. The MALDI mass spectrum of the peptide mixture resulting from CN-induced cleavage (but **before complete reduction**) of the doubly reduced/cyanylated ribonuclease-A isoform represented by HPLC peak A in Figure 2.3. Itz-26-64 represents the itz-blocked fragment of ribonuclease-A from residue 26 to 64.

V. Further Investigation of the Status of the C-terminus (Amide or Free Acid) of CN-Induced Cleavage Fragments

Jacobson *et al.* (9) proposed a mechanism for the CN-induced cleavage reaction that occurs on the N-terminal side of the cyanylated cysteine residue under alkaline conditions to form an amino-N-terminal peptide (fragment) to the left of the cleavage site and a 2-iminothiazolidine (itz) -4-carboxyl blocked peptide (fragment) to the right of the cleavage site (Figure 2.8). As shown in Figure 2.8, when the CN-induced cleavage reaction occurs in a buffer that provides OH^- to carry out a nucleophilic attack toward the carbonyl carbon, the resulting fragments, including the amino-N-terminal peptide and the itz-peptide, have a free acid C-terminus.

The CN-induced cleavage reaction was later optimized by Nakagawa *et al.* (10) and Wu *et al.* (11) by using a stronger nucleophile, NH_3 , to replace OH^- . The CN-induced fragments produced using this modified protocol were assumed to have an amide C-terminus (10, 11). As shown in Figure 2.9, nucleophilic attack by NH_3 to the carbonyl carbon would produce a fragment with an amide C-terminus. However, both OH^- ($\sim 0.001 \text{ M}$) and NH_3 ($\sim 1 \text{ M}$) exist in 1 M aqueous ammonia solution, which was used in the modified protocol. Although NH_3 is far more concentrated than OH^- , CN-induced fragments with both a free acid and an amide C-terminus were detected as discussed below. Therefore, CN-induced fragments with either an amide or a free acid terminus must be considered in computing the calculated mass of these fragments.

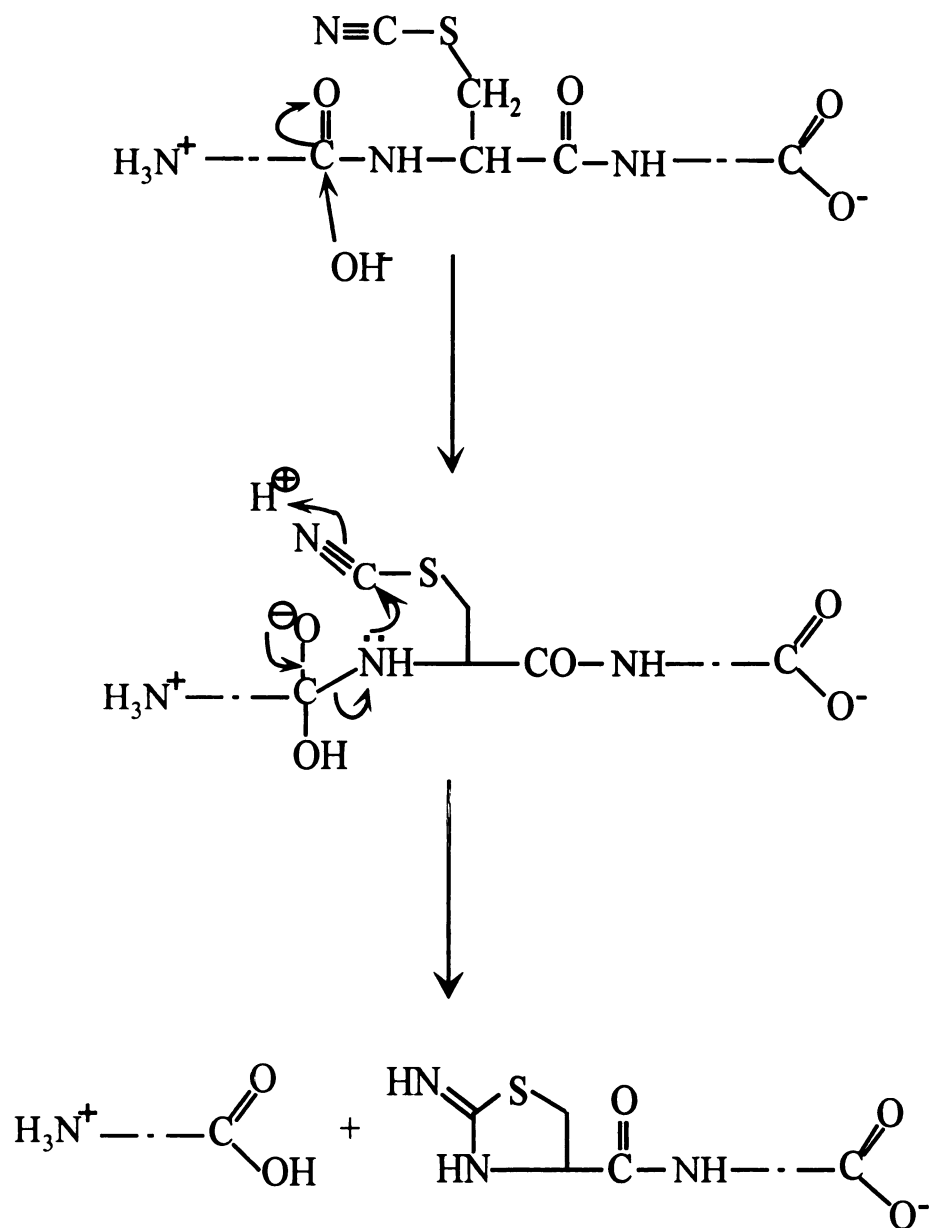


Figure 2.8. A proposed mechanism (9) involving nucleophilic attack by OH^- for the CN-induced cleavage reaction at a cyanylated cysteine residue.

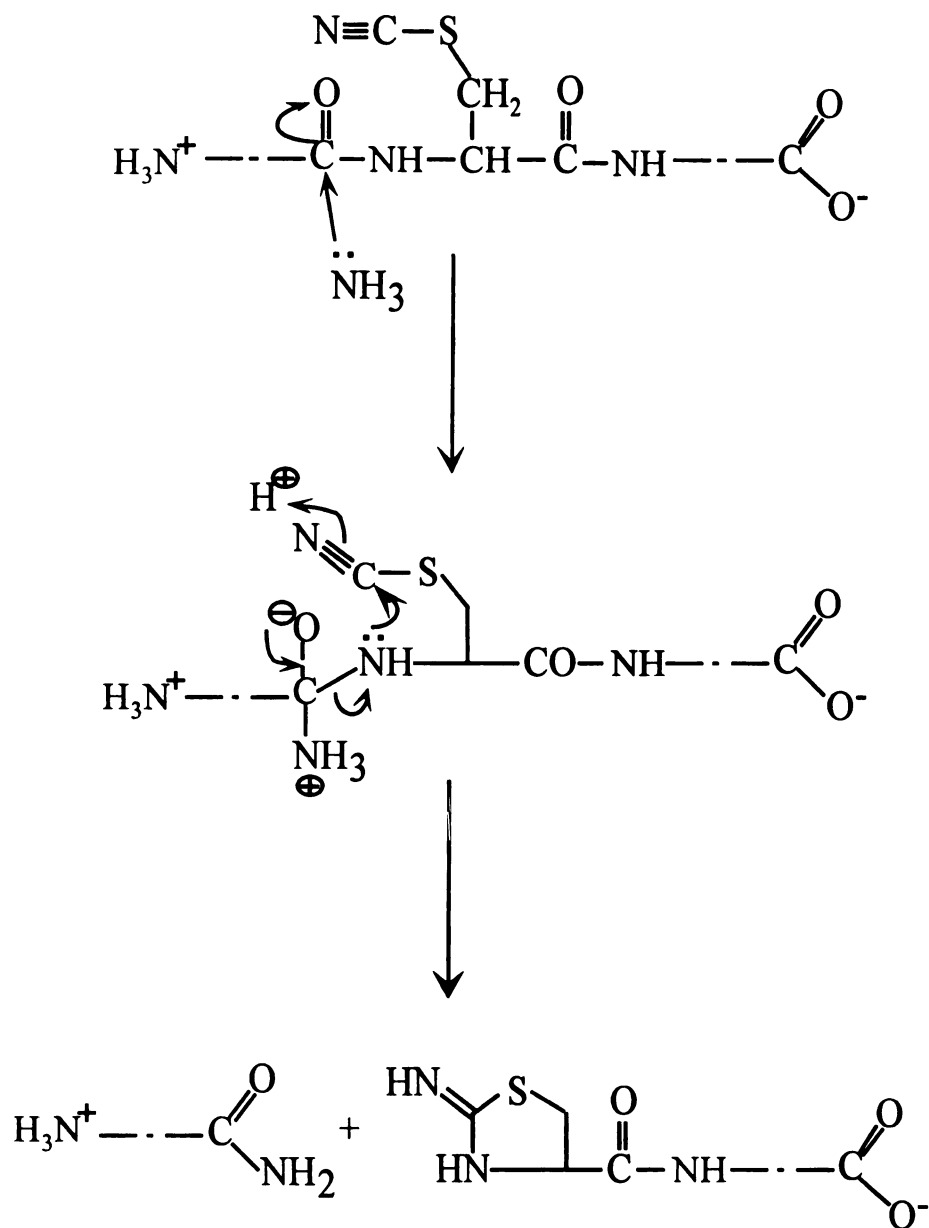
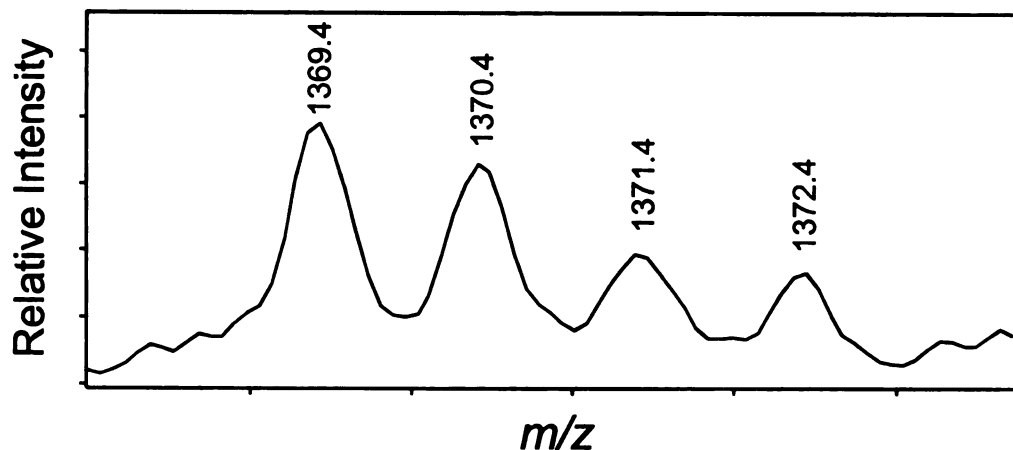


Figure 2.9. Production of CN-induced cleavage fragments with an amide C-terminus via a mechanism involving nucleophilic attack by NH_3 at a cyanylated cysteine residue.

If a CN-induced cleavage fragment is small (usually less than 4,000 Da for an ion trap mass spectrometer (MS) with an electrospray (ESI) ion source or 6,000 Da for a matrix assisted laser desorption/ionization (MALDI)-time of flight (TOF)-MS, its isotopic peaks can be resolved to allow us to determine whether there is an amide C-terminus, a free acid C-terminus or even a mixture of the two. For example, the CN-induced cleavage fragment, itz-13-29-red (itz-blocked fragment from residue 13 to 29: itz-CCGDSGYWCRQCIGIKYT; “red” refers to the species in which the remaining disulfide bond have been reduced during the complete reduction step; thus, all the cysteine residues within the species are free cysteines), produced from sillucin (a peptide having four disulfide bonds), has an amide C-terminus because its ESI spectrum correlates well with its calculated isotopic distribution (Figure 2.10). Another CN-induced cleavage fragment from sillucin, itz-12-β@13-23-ox [itz-C(β@C)CGDSGYWC-RQ; there is a disulfide bond between Cys 14 and Cys 21 within the fragment, as described in Chapter 3] has a free acid C-terminus as indicated in Figure 2.11. In contrast, two forms of another CN-induced cleavage fragment, 1-12-red (ACLPNSCVSKGC), resulted from CN-induced cleavage at cyanylated cysteine 13 in sillucin, one with an amide C-terminus and the other one with a free acid C-terminus were detected. As shown in Figure 2.12, the isotopic peak series beginning at m/z 1180.4 corresponds to the isotopic distribution of fragment 1-12-red-NH₂ (calculated monoisotopic mass for MH⁺ = 1180.5 Da; “NH₂” designates an amide C-terminus); the isotopic peak series beginning at m/z 1181.3 corresponds to the isotopic distribution of fragment 1-12-red-OH (calculated monoisotopic mass for MH⁺ = 1181.5 Da; “OH”

designates a free acid C-terminus); some peaks in the amide series are superimposed on the corresponding peaks for the free acid series.

A) Experimental



B) Calculated

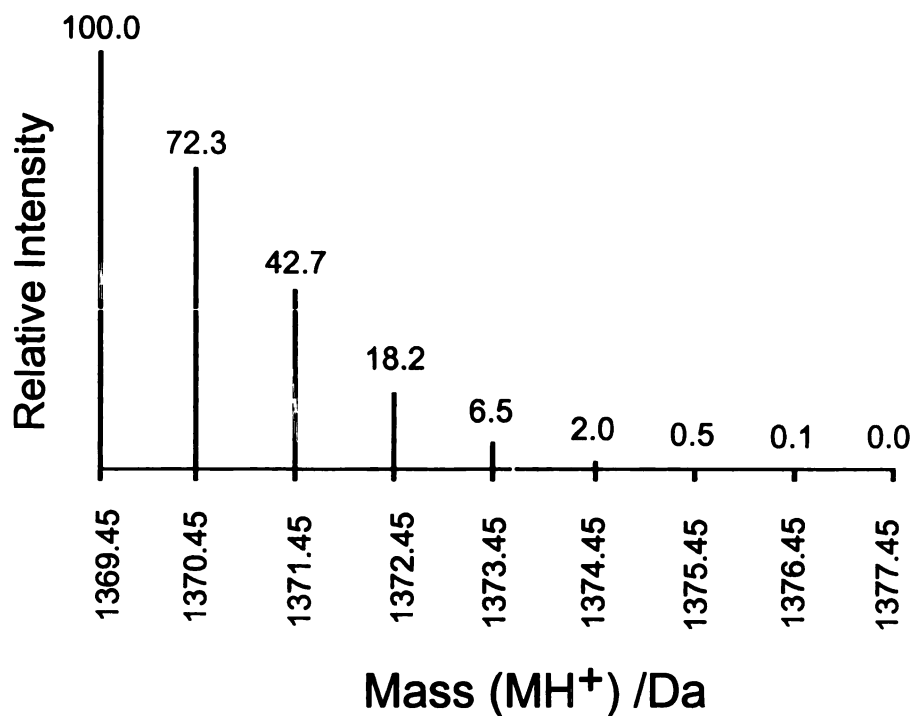
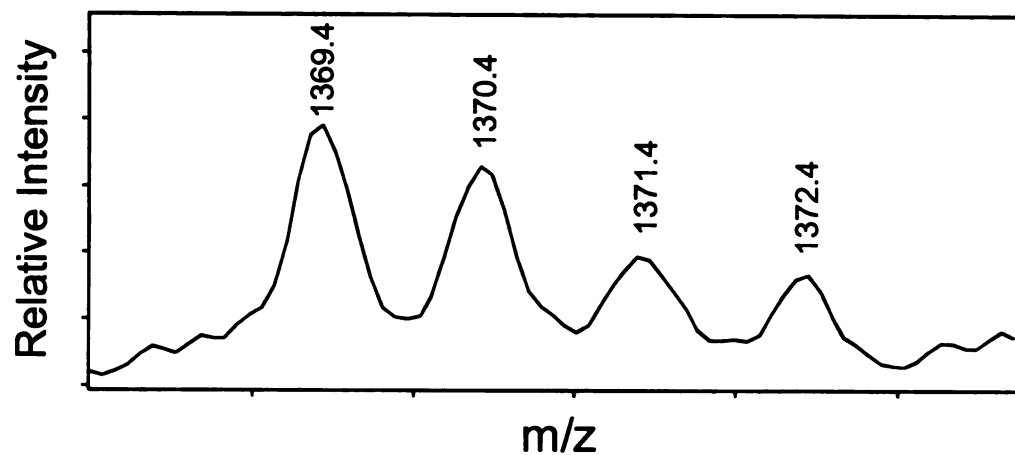


Figure 2.10. Isotope peak intensity distribution in expanded ESI mass spectrum of itz-13-29-red-NH₂ (itz-CCGDSGYWCRQCIGIKYT-NH₂) produced from sillucin. “M” represents the molecule and “MH₂²⁺” represents the doubly protonated molecule.

A) Experimental



B) Calculated

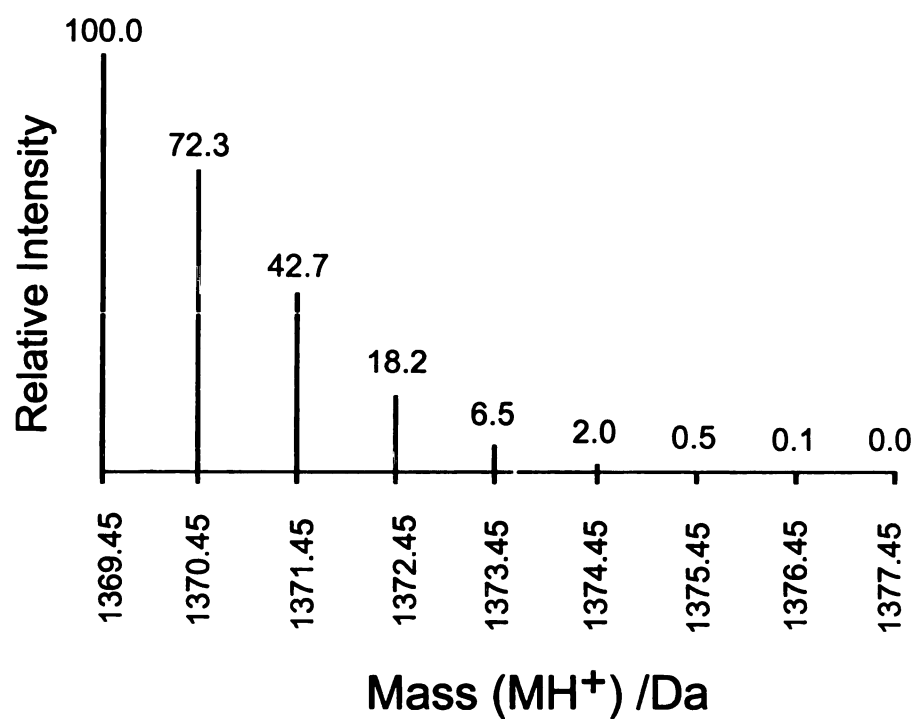
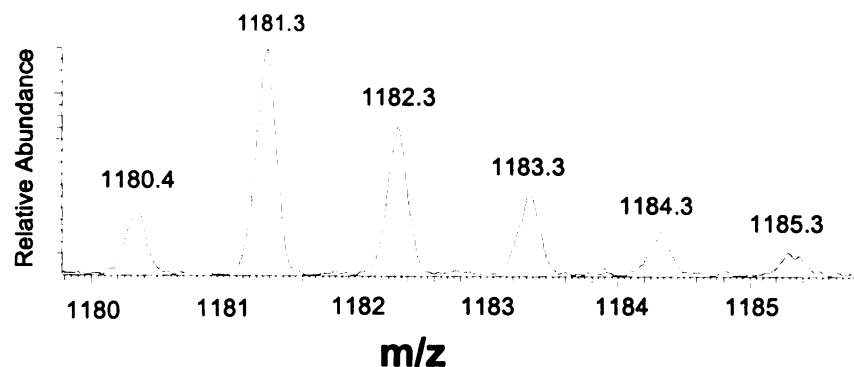


Figure 2.11. Isotope peak intensity distribution in expanded MALDI mass spectrum of itz-12- β @13-23-ox-OH (itz-C(β @C)CGDSGYWCRQ-OH) produced from sillucin.

A) Experimental



B) Calculated

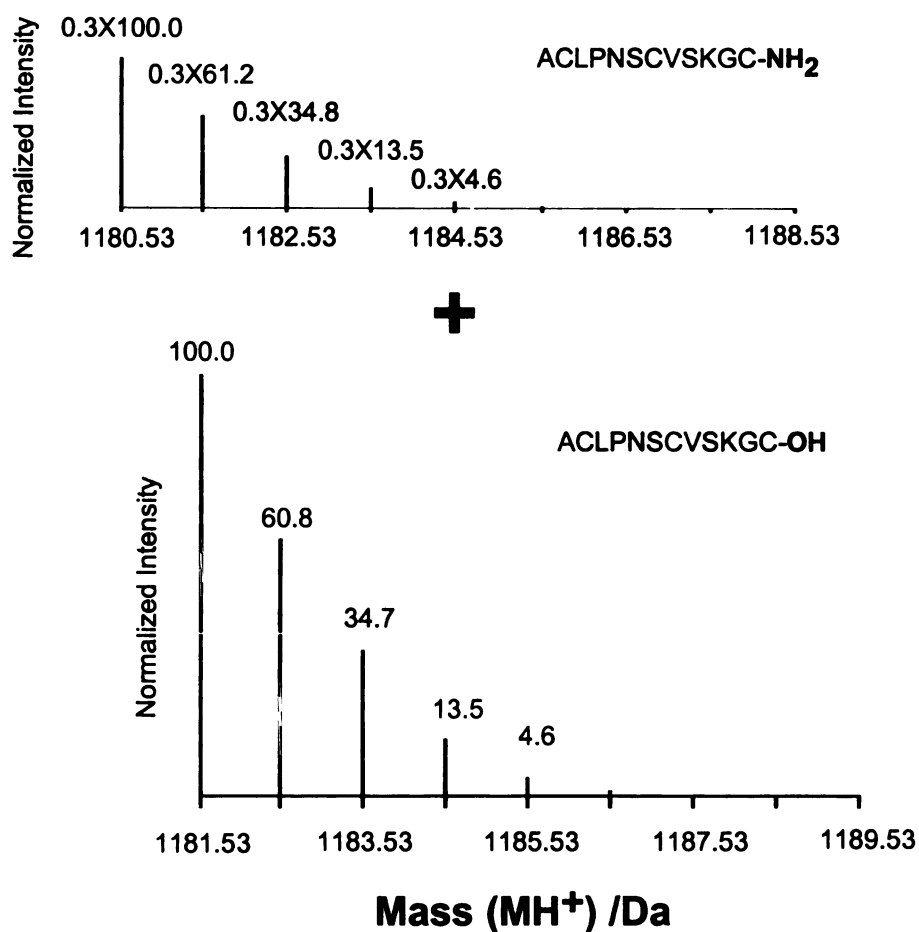


Figure 2.12. Isotope peak intensity distributions for two forms of the CN-induced fragment, 1-12-red (ACLPNSCVSKGC; 1-12-red-NH₂ and 1-12-red-OH), produced from sillucin, were detected during ESI infusion on an ion trap. “0.3×” in the middle panel indicates that the peak intensity is 30% of the intensity for the corresponding peaks in the bottom panel, respectively.

VI. Sequence Analysis of Iminothiazolidine-Blocked Peptides by CID-MS/MS

Papayannopoulos and Biemann (12) showed that high-energy collision induced dissociation (CID) mass spectrometry/mass spectrometry (MS/MS) on a four-sector mass spectrometer could be used to deduce the sequence of some itz-peptides. In our experiments, low-energy CID-MS/MS data from an ion trap mass spectrometer (Finnigan LCQ) were used to distinguish isomass itz-peptides that may arise in mass computations of itz-peptide fragments for the many possible disulfide structural isomers that must be considered.

We have studied the itz-peptides that are listed in Table 2.2 by CID-MS/MS on an ion trap mass spectrometer. The MS/MS data for these itz-peptides are good enough to allow a partial sequence to be identified, and a partial sequence is typically enough for us to distinguish one CN-induced fragment from its isomass isomer. A representative CID spectrum is shown in Figure 2.13.

In determining the disulfide structure of sillucin, a cysteine-rich peptide, a CN-induced fragment, itz-13-29-red-NH₂, was distinguished from its isomass isomer, itz-14-30-NH₂ (see Figure 3.8 in Chapter 3 for details).

Table 2.2. Some itz-fragments studied by ESI-CID-MS/MS on an ion trap.

itz-Fragment	Calculated Average Mass (MH ⁺) /Da
itz-CKRM	561.7
itz-CFGGR	563.6
itz-CFEKGMNYTVR	1372.6
itz-CVEWLRRYLEN	1505.7
itz-CEGNPYVPHFDASV	1659.8
itz-CAYKTTQANKHIIVACEGNPYVPHFDASV	3302.7

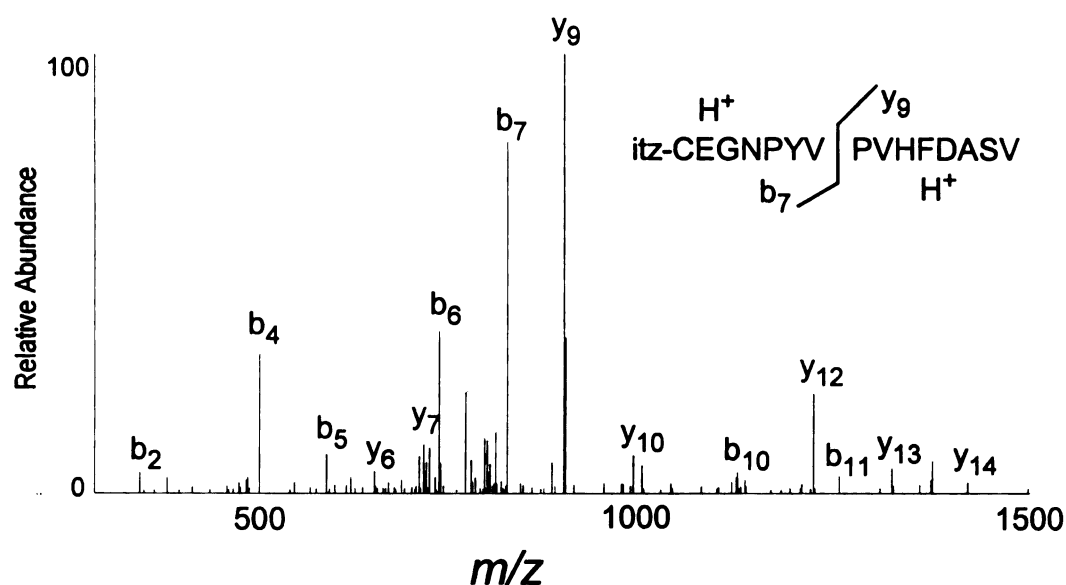


Figure 2.13. The ESI-CID-MS/MS of itz-CEGNPYVPVHFDASV (residue 110-124 of ribonuclease-A), a doubly charged ion (m/z 830.4), derived from cyanylation/CN-induced cleavage of a singly reduced ribonuclease-A isoform, obtained on an ion trap MS. The nomenclature of Biemann (13) for **b** and **y** ions is used throughout, although for convenience, lower case letters are also used to indicate the peptide bond that is cleaved to produced key fragments (**b₇** and **y₉**) in a manner analogous to that introduced by Roepstorff *et al.* (14) using capital letters.

VII. Conclusions

Various aspects of the disulfide mass mapping methodology based partial reduction, cyanylation, and CN-induced cleavage are described. Including disulfide mass mapping of doubly reduced isoforms offers flexibility in the chemistry of partial reduction and provides complementary data to those obtained from the previous partial reduction protocol, which utilizes only singly reduced isoforms. An additional analysis step by mass spectrometry after the CN-induced cleavage reaction of partially reduced/cyanylated species, but before complete reduction of the residual disulfide bonds, may provide diagnostic information that cannot be obtained after all the residual disulfide bonds have been reduced. When CN-induced cleavage reactions are conducted in aqueous ammonia, the resulting CN-induced fragments may have an amide or a free acid C-terminus. Both cases of an amide and a free acid C-terminus must be considered in computing the masses for expected CN-induced cleavage fragments from different disulfide structural isomers of an analyte. In cases of isomass iminothiazolidine-blocked (itz) fragments, sequence analysis of such fragments by low-energy CID-MS/MS can be used to distinguish one CN-induced fragment from its isomass isomers.

VIII. References

1. Wu, J. & Watson, J. T. A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci.* **6**, 391-398 (1997).
2. Yang, Y., Wu, J. & Watson, J. T. Disulfide Mass Mapping in Proteins Containing Adjacent Cysteines Is Possible with Cyanylation/Cleavage Methodology. *J. Am. Chem. Soc.* **120**, 5834-5835 (1998).
3. Qi, J. & Watson, J. T. Determination of the Disulfide Structure of Sillucin, a Highly Knotted, Cysteine-Rich Peptide, by Cyanylation/Cleavage Mass Mapping. *Biochemistry* **40**, 4531-4538 (2001).
4. Zhou, J., Ens, W., Poppe-Schriemer, N., Standing, K. G. & Westmore, J. B. Cleavage of interchain disulfide bonds following matrix-assisted laser desorption. *Int. J. Mass Spectrom. Ion Processes* **126**, 115-122 (1993).
5. Patterson, S. D. & Katta, V. Prompt fragmentation of disulfide-linked peptides during matrix-assisted laser desorption ionization mass spectrometry. *Anal. Chem.* **66**, 3727-3732. (1994).
6. Crimmins, D. L., Saylor, M., Rush, J. & Thoma, R. S. Facile, in situ matrix-assisted laser desorption ionization-mass spectrometry analysis and assignment of disulfide pairings in heteropeptide molecules. *Anal. Biochem.* **226**, 355-361. (1995).
7. Jones, M. D., Hunt, J., Liu, J. L., Patterson, S. D., Kohno, T. & Lu, H. S. Determination of Tumor Necrosis Factor Binding Protein Disulfide Structure: Deviation of the Fourth Domain Structure from the TNFR/NGFR Family Cysteine-Rich Region Signature. *Biochemistry* **36**, 14914-14923 (1997).
8. Jones, M. D., Patterson, S. D. & Lu, H. S. Determination of Disulfide Bonds in Highly Bridged Disulfide-Linked Peptides by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry with Postsource Decay. *Anal. Chem.* **70**, 136-143 (1998).
9. Jacobson, G. R., Schaffer, M. H., Stark, G. R. & Vanaman, T. C. Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues. *J. Biol. Chem.* **248**, 6583-6591. (1973).
10. Nakagawa, S., Tamakashi, Y., Hamana, T., Kawase, M., Taketomi, S., Ishibashi, Y., Nishimura, O. & Fukuda, T. Chemical cleavage of recombinant fusion proteins to yield peptide amides. *J. Am. Chem. Soc.* **116**, 5513-5514 (1994).

11. Wu, J. & Watson, J. T. Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping. *Anal. Biochem.* **258**, 268-276 (1998).
12. Papayannopoulos, I. A. & Biemann, K. Amino acid sequence of a protease inhibitor isolated from *Sarcophaga bullata* determined by mass spectrometry. *Protein Sci.* **1**, 278-288 (1992).
13. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* **16**, 99-111. (1988).
14. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass. Spectrom.* **11**, 601 (1984).

CHAPTER 3

DETERMINATION OF THE DISULFIDE STRUCTURE OF SILLUCIN, A HIGHLY KNOTTED, CYSTEINE-RICH PEPTIDE, BY CYANYLATION/CN- INDUCED CLEAVAGE MASS MAPPING

I. Introduction

The presence of antimicrobial peptides in growth media of *Rhizomucor pusillus* (formerly *Mucor pusillus*) and *Rhizomucor miehei* (formerly *Mucor miehei*) was first observed over thirty years ago (1). Purification and compositional analysis of bioactive peptides isolated from different strains established that *Rhizomucor pusillus* synthesized a 30-residue peptide designated as sillucin, whereas *Rhizomucor miehei* strains produced the antimicrobial peptides designated as mieheins that ranged in size from 56 to 74 amino acids (2). The range of antimicrobial activity of these cationic peptides was found to be limited to Gram+ microorganisms.

A common feature of the antimicrobial peptides from *Rhizomucor species* is the high cysteine content varying from 8 residues in sillucin, and up to 16 residues in mieheins (2). Negative analytical tests for reactive –SH groups in sillucin and mieheins indicated that most likely, all cysteine residues in these peptides were involved in disulfide bonds, resulting in a highly complex tertiary structure. Although the primary structure of sillucin was established (3) as ACLPNSCVSK GCCCGBSGYW CRQCGIKYTC (B indicates either N or D; in our work B is determined to be D by accurate mass measurement, as described later.), the specific disulfide bond pattern between its cysteine residues has remained unknown.

Antimicrobial peptides from eukaryotes have become subjects of intensive research since they may serve as potential sources of new antibiotics for therapeutic applications (4, 5). Because minor variations in peptide structure can influence antimicrobial activity, precise knowledge of the location of disulfide bonds is a prerequisite to future structure/activity studies of these bioactive peptides.

Conventional methodology for determining disulfide linkages involves enzymatically or chemically cleaving the peptide backbone between the cysteine residues and identifying the peptide fragments that contain disulfide bonds or are linked by them. Analysis of proteins containing adjacent cysteines is very difficult, if not possible, by conventional methodology as there is no adequate method to cleave between the adjacent cysteine residues (6, 7). In special cases such as the one described by Poerio *et al.* (8), proteolytic digests can be analyzed by a combination of Edman degradation and fast atom bombardment mass spectrometry (FAB-MS) to determine the linkage of two disulfide bonds involving adjacent cysteines. However, success with this approach is only possible when the residual string of residues on the N-terminal side of the pair of adjacent cysteines in one peptide of the three-peptide proteolytic fragment (which is linked by two disulfide bonds) is shorter than the residual string of residues on the N-terminal side of the cysteine residue in each of the other two peptides (8, 9). A similar fortunate situation occurred during the analysis of insulin (9) and huwentoxin-I (10), where 17 and 6 Edman degradation steps, respectively, on the intact molecule were required to provide evidence for the disulfide bond pattern involving two adjacent cysteines.

In another case, fast atom bombardment mass spectrometry/mass spectrometry (FAB-MS/MS) was used to confirm the Cys6-Cys48/Cys47-Cys52 prediction in insulin-

like growth factors (IGFs) (11). Recently, matrix assisted laser desorption/ionization (MALDI) -post source decay (PSD) was found to successfully fragment the peptide bonds between two adjacent cysteines in a three-peptide proteolytic fragment linked by two disulfide bonds; however, the generality of this procedure remains to be tested (12).

In some cases, if enough sample and good crystals can be obtained, high resolution X-ray crystallography might be used to determine the disulfide structure involving adjacent cysteines (13). In similar situations, two-dimensional nuclear magnetic resonance (2D-NMR) could be applied to this analytical problem, if sufficient sample is available (14). In general, efficient chemical methods are still needed for quality control of recombinant disulfide proteins.

Partial reduction by the water-soluble tris(2-carboxyethyl)phosphine (TCEP) hydrochloride to generate an array of partially reduced protein species containing disulfides and thiols has been widely used to develop new strategies for assigning disulfide bonds (15-18). Such strategies have been especially useful in solving the disulfide structure of peptides and proteins with a cysteine-knotted core or those with adjacent cysteines, which are very difficult to analyze with traditional strategies based on enzymatic digestions. For most of these strategies, the nascent free cysteine residues produced during partial reduction were alkylated and detected by Edman sequencing, as the basis for deducing the disulfide linkages. However, as alkylation is achieved under basic conditions, artifacts due to disulfide scrambling could be a problem. Another shortcoming of these strategies is that only small peptides can be considered because the resulting alkylated peptides finally must be analyzed by Edman sequencing.

A novel analytical strategy developed by Wu and Watson (19) uses partial reduction by TCEP of a multi-disulfide protein to produce singly reduced isoforms, which are then cyanylated by 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate, chemically cleaved by aqueous ammonia on the N-terminal side of the cyanylated cysteine residues, and mass-mapped to the sequence for assignment of disulfide bond linkages in proteins. The feasibility of analyzing disulfide structures involving adjacent cysteines by the cyanylation/CN-induced cleavage procedure was proven by Yang *et al.* (20).

Sillucin is resistant to various proteases, in part because its **three** adjacent cysteines and its tightly folded structure, and attempts to determine its disulfide structure using conventional methods based on enzymatic digestion have not been successful. However, by mass mapping the CN-induced cleavage products from both singly reduced/cyanylated and doubly reduced/cyanylated species, we have been able to deduce the disulfide structure of sillucin as Cys2-Cys7, Cys12-Cys24, Cys13-Cys30, and Cys14-Cys21. The chemical processes and array of CN-induced cleavage products used in deducing the disulfide connectivity in sillucin are illustrated in Figure 3.1 and Figure 3.2.

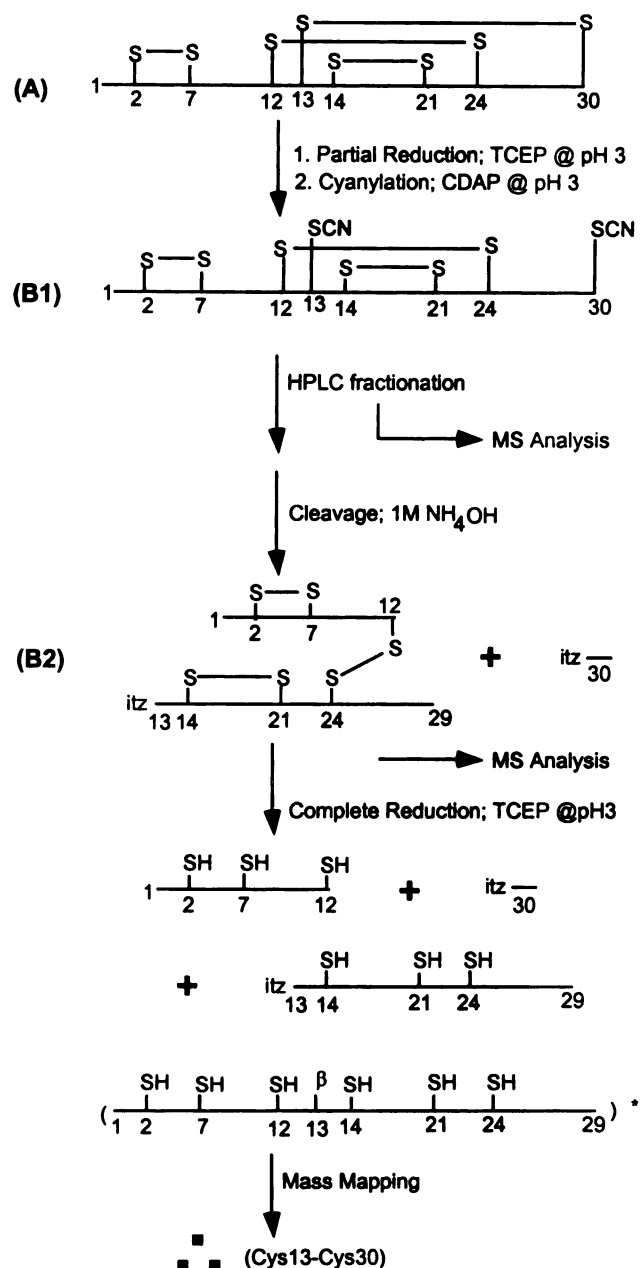


Figure 3.1. Overview of chemical reactions involved in the disulfide mass mapping of a singly reduced/cyanylated species of sillucin. Recognition of the cleaved fragments by mass spectrometry allows the disulfide bond Cys13-Cys30 to be deduced. * The structure in the parenthesis represents a β -elimination product, which results in this case from loss of HSCN at cyanylated Cys 13 (β -elimination) and CN-induced cleavage at cyanylated Cys30.

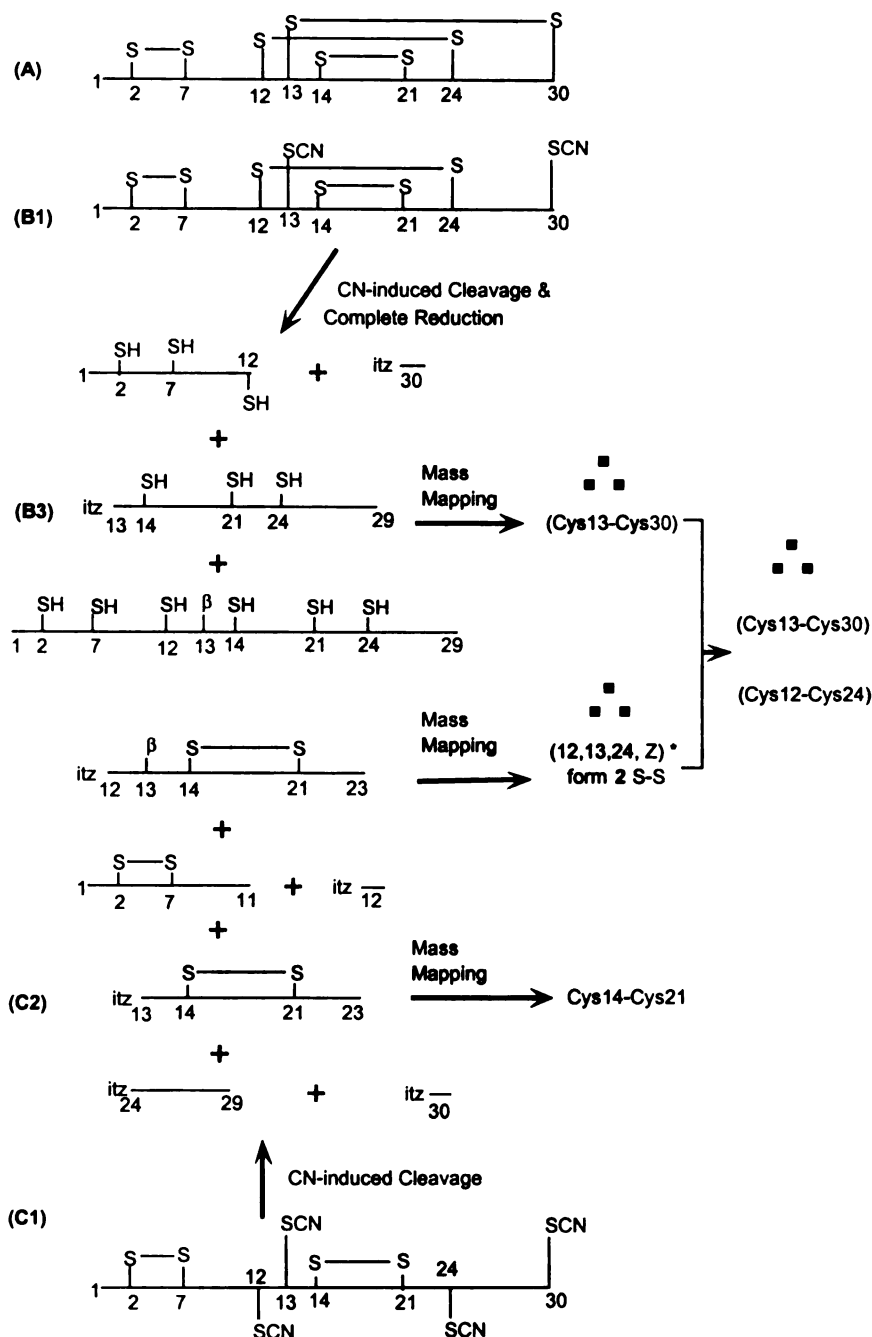


Figure 3.2. Overview of chemical reactions involved in the disulfide mass mapping of a doubly reduced/cyanylated species (C1) of sillicin (A) as well as a singly reduced/cyanylated species (B1). Recognition of the cleaved fragments (C2) from (C1) before complete reduction allows the connectivity of Cys14-Cys21 to be deduced. Recognition of the cleaved fragments (B3) from (B2) (shown in Figure 3.1) after complete reduction combined with the determination of Cys14-Cys21 as described above allows the connectivity of Cys12-Cys24 to be deduced; Z could be 2, 7, or 30.

II. Materials and Methods

Chemicals

Sillucin was isolated and purified according to methods previously described (1, 2). Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride was purchased from Pierce Chemical Co. (Rockford, IL). Guanidine hydrochloride is a product of Boehringer-Mannheim Biochemicals (Indianapolis, IN). 1-Cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate was purchased from Sigma. Water, acetonitrile, and trifluoroacetic acid (TFA) were of HPLC grade. TCEP and CDAP solutions were freshly prepared in 0.1 M citrate buffer (pH 3.0) prior to use. Because of the instability of CDAP in aqueous solution, solid CDAP was dissolved in pH 3 buffer just before use.

Partial Reduction and Cyanylation

A solution containing 20nmol sillucin was prepared in 10 μ L of 0.1 M pH 3.0 citrate buffer containing 6 M guanidine hydrochloride. Partial reduction of sillucin was carried out by adding 1600 nmol of TCEP (16 μ L of 0.1 M TCEP solution), followed by incubation at 50 °C for 10 minutes. Other trials of partial reduction under different conditions of stoichiometry, temperature, and reaction time were conducted similarly to optimize production of singly or doubly reduced species; such a search for reasonably optimum reaction conditions is typical because of the diverse range of stability of disulfide bonds from protein to protein.

To the TCEP reaction mixture containing partially reduced sillucin species were added 4000 nmol of CDAP (20 μ L of 0.2 M CDAP solution; more than a two-fold

equivalent to TCEP because residual TCEP reacts with CDAP.). Cyanylation of the nascent sulfhydryl groups was accomplished during incubation at room temperature for 15 minutes.

HPLC Separation of Partially Reduced and Cyanylated Sillucin Species

Mixtures of partially-reduced and cyanylated species were separated by reversed phase HPLC using a linear gradient elution of 15%-40% B in 50 minutes (B is 90% (v/v) acetonitrile/0.1% TFA; A is 0.1% TFA in water.) with Waters model 6000 pumps controlled by a PC. UV detection was at 215 nm. A Vydac C18 (#218TP54, 5- μ m particle size, 300-Å pores in a 4.6 $\times$ 250 mm) column was used. The predominant HPLC fractions were collected manually and the masses of the collected species in each fraction (approximately 0.8~1.0 mL) were determined by MALDI-MS.

CN-induced Cleavage of Partially Reduced and Cyanylated Sillucin Species

A solution of a dried HPLC fraction was reconstituted in 2 μ L of 6 M guanidine hydrochloride in 1 M aqueous ammonia to dissolve the partially-reduced and cyanylated sillucin species; to this solution was added another 8 μ L of 1 M aqueous ammonia. CN-induced cleavage of the peptide chain on the N-terminal side of cyanylated cysteine residues was accomplished during reaction at room temperature for one hour. Excess ammonia was removed in a vacuum system. Truncated peptides, still linked in some cases by remaining disulfide bonds depending on the extent of partial reduction, were dissolved in 10 μ L of water; 1 μ L of this solution was further diluted with 9 μ L of 50%

(v/v) acetonitrile/0.1% TFA, and analyzed by MALDI-MS. The other 9 μ L of the solution was dried in a vacuum system for future use.

Complete Reduction of the Residual Disulfide Bonds

Truncated peptides, from the CN-induced cleavage reaction, some still linked by residual disulfide bonds, were completely reduced by reaction with 200 nmol of TCEP (2 μ L of 0.1 M TCEP solution) at 37 °C for 30 minutes. Samples were then diluted with 90 μ L of 50% (v/v) aqueous acetonitrile/0.1% TFA solution prior to analysis by MALDI-MS.

MALDI-MS

MALDI mass spectra were obtained on a Voyager DE-STR time-of-flight (TOF) mass spectrometer (Perkin Elmer Biosystems Inc., Framingham, MA) equipped with a 337-nm nitrogen laser. The accelerating voltage in the ion source was set at 25 kV. Except for accurate mass measurement of intact sillucin that was done in positive reflectron mode, all other data were acquired in the positive linear mode of operation. Time-to-mass conversion was achieved by internal calibration using standards of bovine bradykinin (monoisotopic mass of the protonated molecule (MH^+): 1060.569 Da; average mass of MH^+ : 1061.217 Da) and bovine pancreatic insulin chain B oxidized (monoisotopic mass of MH^+ : 3494.651 Da; average mass of MH^+ : 3496.903 Da) obtained from Sigma Chemical Co. (St. Louis, MO). All experiments were performed using α -cyano-4-hydroxycinnamic acid (Aldrich Chemical co., Milwaukee, WI) as the matrix. Saturated matrix solutions were prepared in a 50% (V/V) aqueous solution of

acetonitrile/0.1% TFA. Just prior to analysis, 0.5 μL of matrix solution was applied to the stainless steel sample plate and allowed to air-dry; then 1 μL of sample solution was added to the top of it, and finally another 1 μL of matrix solution was added to the same spot. The mixture was allowed to air-dry before being introduced into the mass spectrometer.

III. Results

1. Accurate Mass Measurement of Intact Sillucin

The sequence of sillucin, determined previously by Bradley and Somkuti (3) is ACLPNSCVSK GCCCGBSGYW CRQCGIKYTC (B stands for N or D). As mass measurement error in the range of 2,000 to 4,000 Da using internal calibration in the reflectron mode on a MALDI-TOF mass spectrometer (PE Voyager DE-STR) is less than 0.03% (data not shown), the uncertainty in the previously determined sequence of sillucin can be resolved by accurate mass measurement. The monoisotopic mass of protonated sillucin (MH^+) is determined as 3200.23 ± 0.04 Da (3200.23 is mean of seven measurements and 0.04 is the confidence value at the 95% level). Figure 3.3 shows one of the seven measurements by MALDI-MS. The data indicate that sillucin has aspartic acid (D) in position 16 (with overall calculated monoisotopic mass of 3200.211 Da for MH^+), instead of asparagine (N) (with overall calculated monoisotopic mass of 3199.227 Da for MH^+). These experimental data also indicate that sillucin has four disulfide bonds because the observed mass of 3200.23 ± 0.04 Da is in agreement with the calculated mass of 3200.211 Da for MH^+ , which is calculated assuming that all cysteines are involved in

disulfide bonds. In addition, the data showed that sillucin has a free acid C-terminus because the experimental data agree with all the corresponding calculated masses with the assumption of a free acid C-terminus in sillucin.

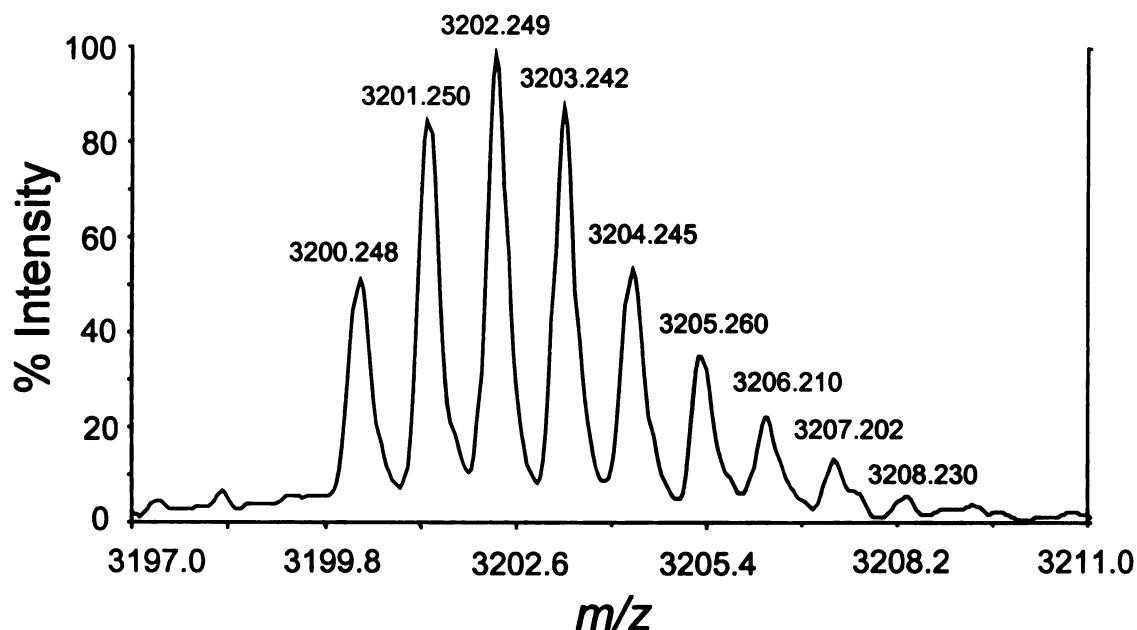


Figure 3.3. The MALDI mass spectrum of the intact sillucin.

2. Partially Reduced/Cyanylated Species of Sillucin

Four major components were separated by HPLC after partial reduction and cyanylation of sillucin (Figure 3.4). Analyses of aliquots of designated fractions by MALDI-MS indicated (data not shown) that the HPLC fraction IP represents the intact peptide (The observed MALDI peak at a mass-to-charge ratio (m/z) 3202.8 represents MH^+ , whose calculated average mass is 3202.8 Da.). HPLC fractions 1, 2, and 3 represent species that are 52, 104, and 208 Da heavier (MALDI peaks at m/z 3254.9, 3306.6, and 3410.8, respectively), respectively, than the intact peptide. Because two CN

groups are added to a singly reduced species upon cyanylation, HPLC fraction 1 corresponds to one of the four possible singly reduced/cyanylated isoforms of sillucin. Similarly, HPLC fraction 2 corresponds to one of the six possible doubly reduced/cyanylated isoforms, and HPLC fraction 3 represents the completely reduced/cyanylated sillucin.

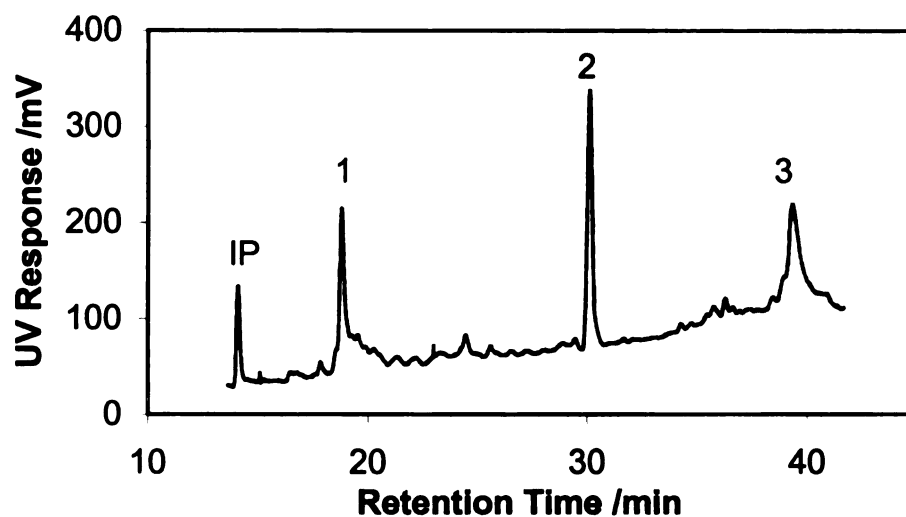


Figure 3.4. The HPLC chromatogram of denatured sillucin (IP) and some of its partially reduced/cyanylated species. Separation was carried out on a Vydac C18 column at a flow rate of 1.0 mL/min with a linear gradient of 15% to 40% B in 50 minutes, where A = 0.1% TFA in water and B = 90% (v/v) acetonitrile/0.1% TFA. Peaks IP, 1, and 2 represent the intact peptide, a singly reduced/cyanylated species, and a doubly reduced/cyanylated species, respectively; peak 3 represents the completely reduced/cyanylated species, as determined from analysis by MALDI-MS.

3. A Single Reduced/Cyanylated Species

Because there are eight cysteine residues in sillucin, 105 disulfide structure isomers are possible. Although there are four possible singly reduced isoforms for each of the 105 disulfide structure isomers, the total number of possible singly reduced isoforms is only 28 because some isomers have the same disulfide bonds as the others. The calculated m/z values for MH^+ (based on average masses of amino acid residues) of each of the possible CN-induced cleavage fragments resulting from all 28 possible singly reduced/cyanylated isoforms is listed in Table 3.1.

The MALDI-MS spectrum in Figure 3.5 was obtained upon analysis of the mixture of products resulting from **CN-induced cleavage** and following **complete reduction** of the singly reduced/cyanylated species (HPLC fraction 1 in Figure 3.4); it shows two major peaks at m/z 1966.8 (monoisotopic mass) and 3072.6 (average mass). As the mass measurement error in the range of 1,000 to 3,500 Da using internal calibration in the linear mode on a MALDI-TOF mass spectrometer (PE Voyager DE-STR) is less than 0.03% (data not shown), a certain MALDI peak can be assigned to the CN-induced cleavage fragment, which has a calculated mass that falls into the range of $\text{mass} \pm 0.03\% * \text{mass}$. According to calculated values in Table 3.1, the peak at m/z 1966.8 corresponds to itz-13-29-red-NH₂ (calculated monoisotopic mass of $MH^+ = 1966.8$ Da). If the C-terminus of the intact sillucin was an amide, the peak at m/z 1966.8 could have also been assigned to the fragment itz-14-30-NH₂ (calculated monoisotopic mass for $MH^+ = 1966.8$ Da).

Table 3.1. Calculated m/z values for MH^+ (based on average masses of amino acid residues) of possible fragments resulting from CN-induced cleavage of singly reduced/cyanylated isoforms of all disulfide structural isomers of sillucin.

C_x-C_y ^a	1-(x-1)-red-NH ₂ ^b	itz-x-(y-1)-red-NH ₂	itz-y-30-red-OH	1-(β@x)-(y-1)-red-NH ₂ ^c	itz-x-(β@y)-30-red-OH
C2-C7	89.1	557.6	2650.1	569.6	3130.7
C7-C12	603.7	517.6	2175.5	1044.3	2616.1
C2-C12	89.1	1032.2	2175.5	1044.3	3130.7
C12-C13	1078.3	146.2	2072.4	1147.4	2141.5
C7-C13	603.7	620.7	2072.4	1147.4	2616.1
C2-C13	89.1	1135.3	2072.4	1147.3	3130.7
C13-C14	1181.4	146.2	<u>1969.2</u> ^d <u>(1967.8)</u>	1250.6	2038.4
C12-C14	1078.3	249.3	<u>1969.2</u>	1250.6	2141.5
C7-C14	603.7	723.9	<u>1969.2</u>	1250.6	2616.1
C2-C14	89.1	1238.5	<u>1969.2</u>	1250.6	3130.7
C14-C21	1284.6	811.8	1200.4	2019.4	1935.2
C13-C21	1181.4	915.0	1200.4	2019.4	2038.4
C12-C21	1078.3	1018.1	1200.4	2019.4	2141.5
C7-C21	603.7	1492.7	1200.4	2019.4	2616.1
C2-C21	89.1	2007.3	1200.4	2019.4	3130.7
C21-C24	2053.4	430.5	813.0	2406.8	1166.4
C14-C24	1284.6	1199.3	813.0	2406.8	1935.2
C13-C24	1181.4	1302.4	813.0	2406.8	2038.4
C12-C24	1078.3	1405.6	813.0	2406.8	2141.5
C7-C24	603.7	1880.1	813.0	2406.8	2616.1
C2-C24	89.1	2394.7	813.0	2406.8	3130.7
C24-C30	2440.8	708.8	147.2	<u>3072.5</u>	779.0
C21-C30	2053.4	1096.3	147.2	<u>3072.5</u>	1166.4
C14-C30	1284.6	1865.1	147.2	<u>3072.5</u>	1935.2
C13-C30	1181.4 (1180.5)	<u>1968.2</u> <u>(1966.8)</u>	147.2	<u>3072.5</u>	2038.4
C12-C30	1078.3	2071.4	147.2	<u>3072.5</u>	2141.5
C7-C30	603.7	2545.9	147.2	<u>3072.5</u>	2616.1
C2-C30	89.1	3060.5	147.2	<u>3072.5</u>	3130.7

^a C_x-C_y designates the disulfide bond that is reduced in a given singly reduced/cyanylated isoform. ^b 1-(x-1)-red-NH₂, itz-x-(y-1)-red-NH₂, and itz-y-30-red-OH designate the CN-induced cleavage fragments. ^c 1-(β@x)-(y-1)-red-NH₂ and itz-(β@y)-30-red-OH designate the β-elimination products. ^d Bold and underlined m/z values indicate matches (or close values) with experimental data. The corresponding monoisotopic masses for average masses 1181.4 Da, 1968.2 Da, and 1969.2 Da are 1180.5, 1966.8 Da and 1967.8 Da, respectively, as shown in the parenthesis.

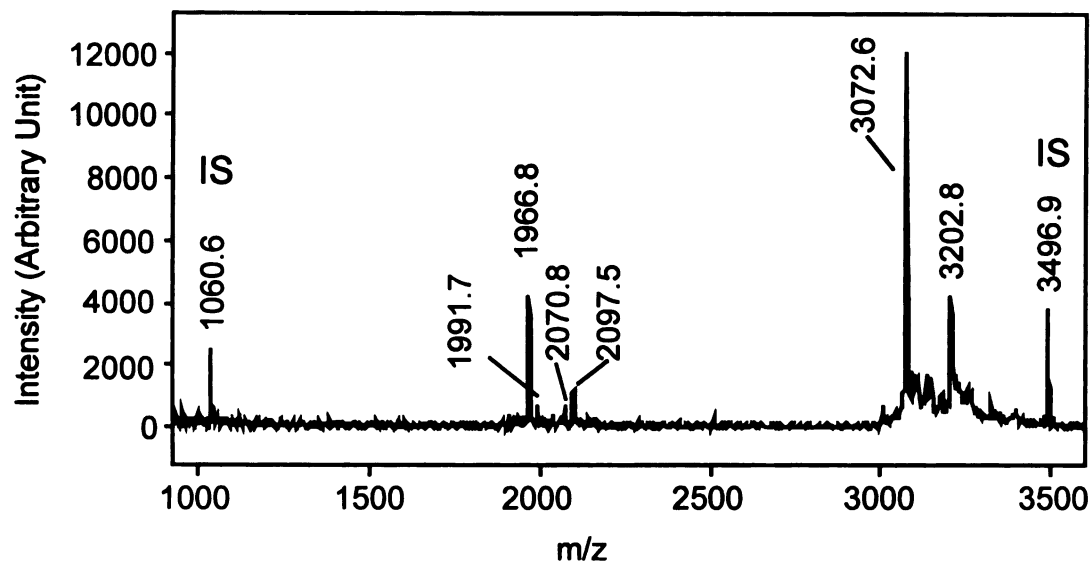


Figure 3.5. The MALDI mass spectrum of the peptide mixture resulting from CN-induced cleavage and subsequent complete reduction of the singly reduced/cyanylated sillucin species represented by HPLC peak 1 in Figure 3.4. "IS" designates an internal standard.

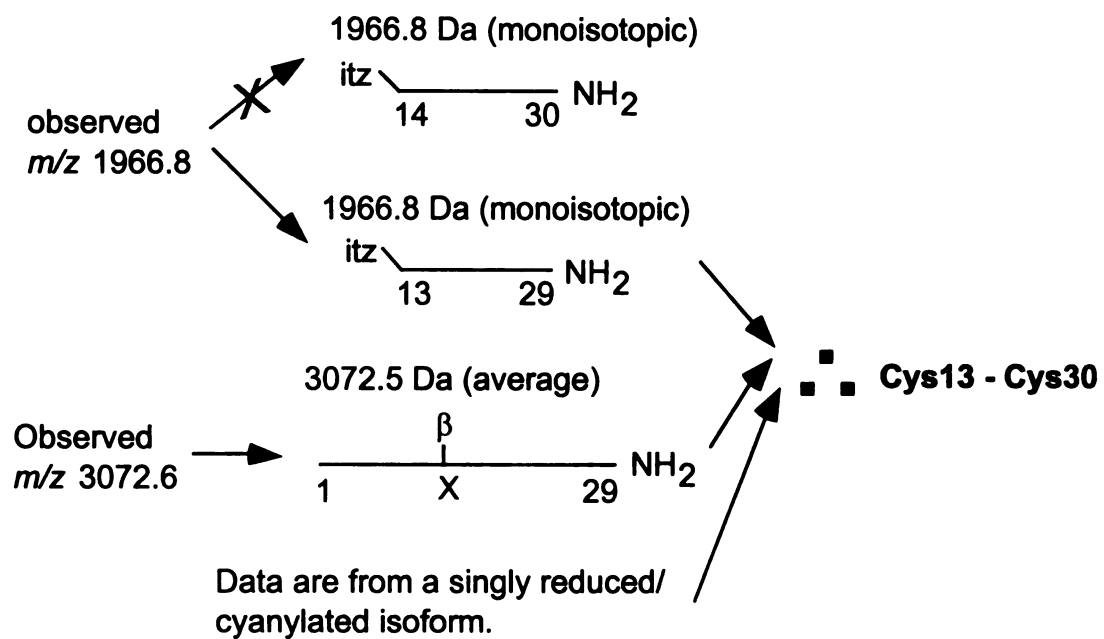


Figure 3.6. Assignment of the disulfide Cys13-Cys30 in sillucin.

However, the accurate mass measurement of intact sillicin (Section A) showed its C-terminus is a free acid, instead of an amide. Therefore, the MALDI peak at m/z 1966.8 should be assigned itz-13-29-red-NH₂. The peak at m/z 3072.6 can be attributed to 1-(β@X)-29-red-NH₂ (calculated average mass of MH⁺ = 3072.5 Da; β@X refers to cyanylation, but subsequent β-elimination (**no CN-induced cleavage**) at cyanylated Cys X, X can be 2, 7, 12, 13, 14, 21, or 24).

Which of the 28 candidate singly reduced isoforms in Table 3.1 corresponds to HPLC peak 1 in Figure 3.4? According to Table 3.1, only the singly reduced/cyanylated isoform represented by Cys13 - Cys30 has a match at MALDI peaks m/z 1966.8 as well as at m/z 3072.6. This coincidence also indicates that the MALDI peak at m/z 3072.6 corresponds to the fragment 1-13@X-29-red-NH₂. The basis for assignment of the disulfide bond, Cys 13-Cys 30, is summarized in Figure 3.6.

Two expected fragments, 1-12-red-NH₂ and itz-30-red-OH (Table 3.1), were not detected. Fragment itz-30-red-OH (calculated average mass of MH⁺ = 147.2 Da) falls into the matrix-peak region and cannot be detected. Another expected fragment, 1-12-red, (With a free acid C-terminus, the calculated average mass of MH⁺ is 1181.4 Da and the calculated monoisotopic mass of MH⁺ is 1180.5 Da.) was not detected because of signal suppression during the initial analysis by MALDI. It was detected in MALDI after it was separated from other species in the mixture by HPLC (Figure 3.7). As shown in Figure 3.7, the MALDI peak at m/z 1181.5 corresponds to 1-12-red-OH; the MALDI peak at m/z 1179.5 represents 1-12-ox-OH, which resulted from oxidization of two of the three free cysteines, Cys 2, Cys 7, and Cys 12, into a disulfide bond. The oxidation occurred because HPLC separated the reducing agent, TCEP, from the fragment 1-12-

red-OH. However, slightly different from the result by offline HPLC-MALDI-MS, another two forms of the CN-induced fragment 1-12-red, 1-12-red-NH₂ and 1-12-red-OH, were detected when the whole mixture including two other fragments, itz-13-29-red-NH₂ and 1-β@13-29-red-NH₂, was subjected to analysis by infusion on an ESI-ion-trap-MS (Figure 2.12 in Chapter 2). The sources of these two forms of the fragment 1-12-red were discussed in detail in chapter 2. Fragment 1-12-red-NH₂ was not detected in offline-HPLC-MALDI-MS because the presence of 1-12-ox-OH, resulted from oxidation, masked the presence of 1-12-NH₂ (Figure 3.7). To the contrary, fragment 1-12-red-NH₂ was detected by ESI infusion (Figure 2.12 in Chapter 2) because the reducing reagent, TCEP, was still present in the mixture, preventing 1-12-ox-OH from forming by oxidation of 1-12-red-OH.

The experimental peak at m/z 1966.8 (monoisotopic mass) in Figure 3.5 is the same as the monoisotopic mass (1966.8 Da for MH⁺) of itz-14-30-NH₂ (if the C-terminus of the intact sillucin is an amide); it is close to the calculated monoisotopic mass (1967.8 Da for MH⁺) of itz-14-30-OH (if the C-terminus of the intact sillucin is OH). As shown in Figure 3.8, Table 3.2, and Table 3.3, electrospray (ESI) tandem mass spectrometry (MS/MS) data by an ion trap further showed that the MALDI peak at m/z 1966.8 should be assigned to itz-13-29-NH₂ rather to itz-14-30-NH₂ or to itz-14-30-OH. Almost a complete series of **b** ions from itz-13-29-NH₂, rather than those from itz-14-30-NH₂ or itz-14-30-OH, were detected (Figure 3.8).

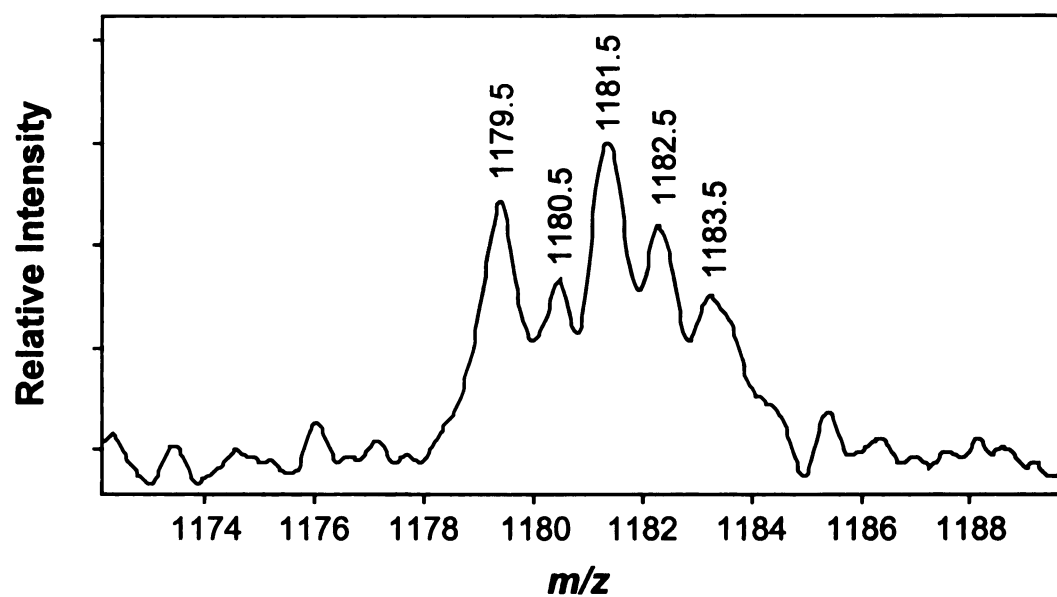
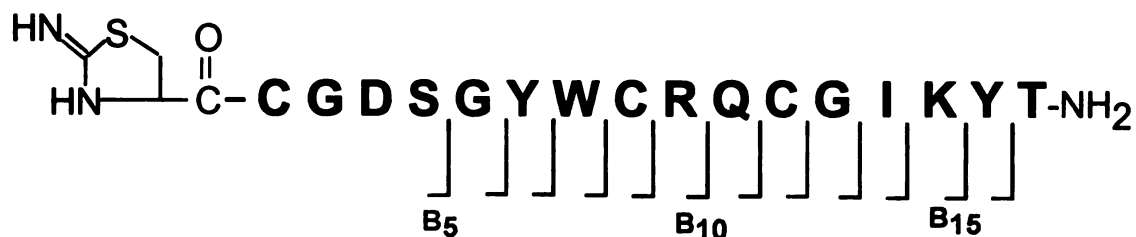
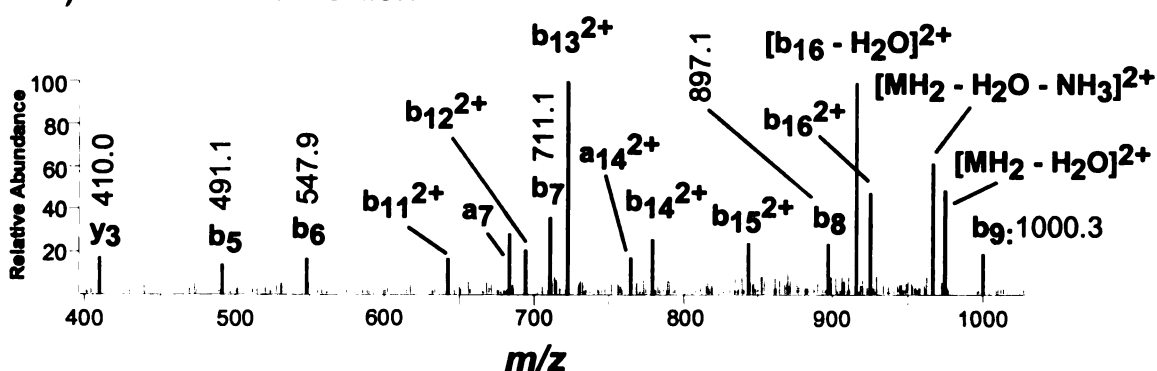


Figure 3.7. Detection of the CN-induced fragment 1-12-OH (two forms: 1-12-ox-OH and 1-12-red-OH) by MALDI-MS after it was separated from itz-12-29-red-NH₂ and 1-β@13-29-red-NH₂.

itz-13-29-NH₂ from Sillucin



a) Lower-Mass Portion



b) Higher-Mass Portion

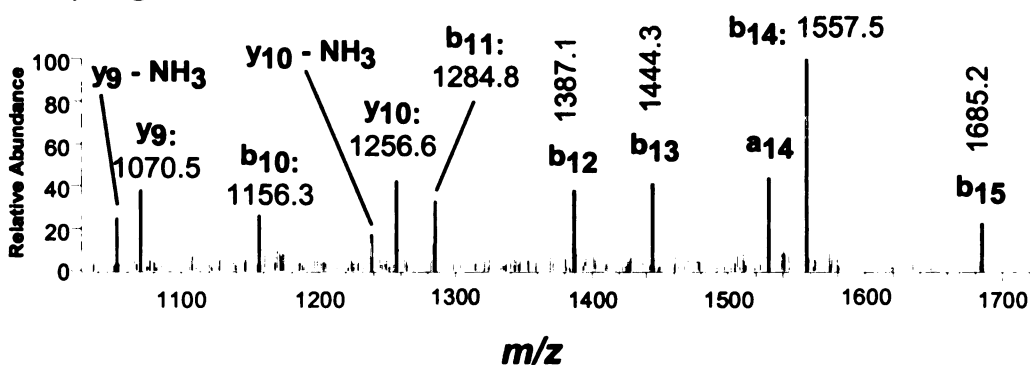


Figure 3.8. The ESI-CID-MS/MS spectrum of CN-induced fragment itz-13-29-NH₂ from sillucin. The precursor is a doubly protonated molecule; [MH₂-H₂O]²⁺ designates a dehydrated and doubly protonated fragment of the precursor; “B₁₃” indicates the peptide bond that is cleaved in the peptide backbone between residue G and I (see definition of the nomenclature in reference 21) to form an amino terminal ion represented by “b₁₃” (singly charged) and b₁₃²⁺ (doubly charged); the definition of a, b, and y ions can be found in reference 22.

Table 3.2. Calculated m/z values (monoisotopic) for **b** and **y** ions of CN-induced fragments itz-13-29-NH₂ and itz-14-30-NH₂ from sillucin.

b_i	itz-13-29-NH₂	itz-14-30-NH₂	y_i	itz-13-29-NH₂	itz-14-30-NH₂
b₁	129.0	129.0	y₁	119.1	121.0
b₂	232.0	186.0	y₂	282.2	222.1
b₃	289.0	301.1	y₃	410.2	385.2
b₄	404.1	388.1	y₄	523.3	513.3
b₅	491.1	445.1	y₅	580.4	626.3
b₆	548.1	608.2	y₆	683.4	683.4
b₇	711.2	794.3	y₇	811.4	786.4
b₈	897.3	897.3	y₈	967.5	914.4
b₉	1000.3	1053.4	y₉	1070.5	1070.5
b₁₀	1156.4	1181.4	y₁₀	1256.0	1173.5
b₁₁	1284.4	1284.4	y₁₁	1419.7	1359.6
b₁₂	1387.4	1341.5	y₁₂	1476.7	1522.7
b₁₃	1444.5	1454.6	y₁₃	1563.7	1579.7
b₁₄	1557.6	1582.6	y₁₄	1678.8	1666.7
b₁₅	1685.6	1745.7	y₁₅	1735.8	1781.8
b₁₆	1848.7	1846.8	y₁₆	1838.8	1838.8

Table 3.3. Calculated m/z values (monoisotopic) for **b** and **y** ions of CN-induced fragments itz-13-29-NH₂ and itz-14-30-OH from sillucin.

b_i	itz-13-29-NH₂	itz-14-30-OH	y_i	itz-13-29-NH₂	itz-14-30-OH
b₁	129.0	129.0	y₁	119.1	122.0
b₂	232.0	186.0	y₂	282.2	223.1
b₃	289.0	301.1	y₃	410.2	386.1
b₄	404.1	388.1	y₄	523.3	514.2
b₅	491.1	445.1	y₅	580.4	627.3
b₆	548.1	608.2	y₆	683.4	684.3
b₇	711.2	794.3	y₇	811.4	787.4
b₈	897.3	897.3	y₈	967.5	915.4
b₉	1000.3	1053.4	y₉	1070.5	1071.5
b₁₀	1156.4	1181.4	y₁₀	1256.0	1174.5
b₁₁	1284.4	1284.4	y₁₁	1419.7	1360.6
b₁₂	1387.4	1341.5	y₁₂	1476.7	1523.7
b₁₃	1444.5	1454.6	y₁₃	1563.7	1580.7
b₁₄	1557.6	1582.6	y₁₄	1678.8	1667.7
b₁₅	1685.6	1745.7	y₁₅	1735.8	1782.7
b₁₆	1848.7	1846.8	y₁₆	1838.8	1839.8

4. A Doubly Reduced/Cyanylated Species

The MALDI-MS spectrum in Figure 3.9 corresponds to the fragments resulting from **CN-induced cleavage** (before complete reduction) of the doubly reduced/cyanylated species (HPLC fraction 2 in Figure 3.4). The MALDI peak series beginning at m/z 1299.4 and at m/z 1300.4 can be attributed to itz-13-23-ox-NH₂ (calculated monoisotopic mass of MH⁺ = 1299.4 Da) and itz-13-23-ox-OH. As this analysis by MALDI-MS was done before **complete reduction** of the sample, the presence of itz-13-23-ox-NH₂ and itz-13-23-ox-OH implies the existence of a disulfide bond, Cys14-Cys21 (see summary in Figure 3.10). With the disulfide bond Cys14-Cys21 identified, the MALDI peak series at m/z 1369.4 can be inferred as representing itz-12-(β@13)-23-ox-OH (calculated monoisotopic mass of MH⁺ = 1369.5 Da); thus, Cys 12, Cys 13, Cys 24, and Cys Z form two disulfide bonds because two disulfide bonds had to be reduced to form the doubly reduced/cyanylated species (HPLC fraction 2 in Figure 3.4). Because the disulfide bond Cys13-Cys30 was identified in the last section, Cys Z must be Cys 30. Thus, another disulfide bond, Cys12-Cys24, can be deduced (Figure 3.10). In addition, MALDI peak at m/z 1331.4 can be attributed to 1-SCN@12-SCN@13-ox-NH₂ (calculated monoisotopic mass of MH⁺ = 1331.5); and thus, Cys2-Cys7 can be deduced because 1-SCN@12-SCN@13-ox-NH₂ contains a disulfide bond (see summary in Figure 3.10).

Analysis by MALDI-MS (data not shown) shows that denatured intact sillucin does not react with CDAP, the cyanylation reagent. This implies that there are no free cysteines in sillucin and that all eight cysteine residues form four disulfide bonds. This observation is in agreement with the accurate mass measurement described above and

literature results (3). With the three disulfide bonds, Cys12-Cys24, Cys13-Cys30, and Cys14-Cys21, having been deduced as described above, the fourth disulfide bond, Cys2-Cys7, can be deduced by default. This deduction is consistent with the presence of a MALDI-MS peak at m/z 1331.4 (Figure 3.9), representing 1-SCN@12-SCN@13-ox-NH₂. These assignments are summarized illustratively in Figure 3.10.

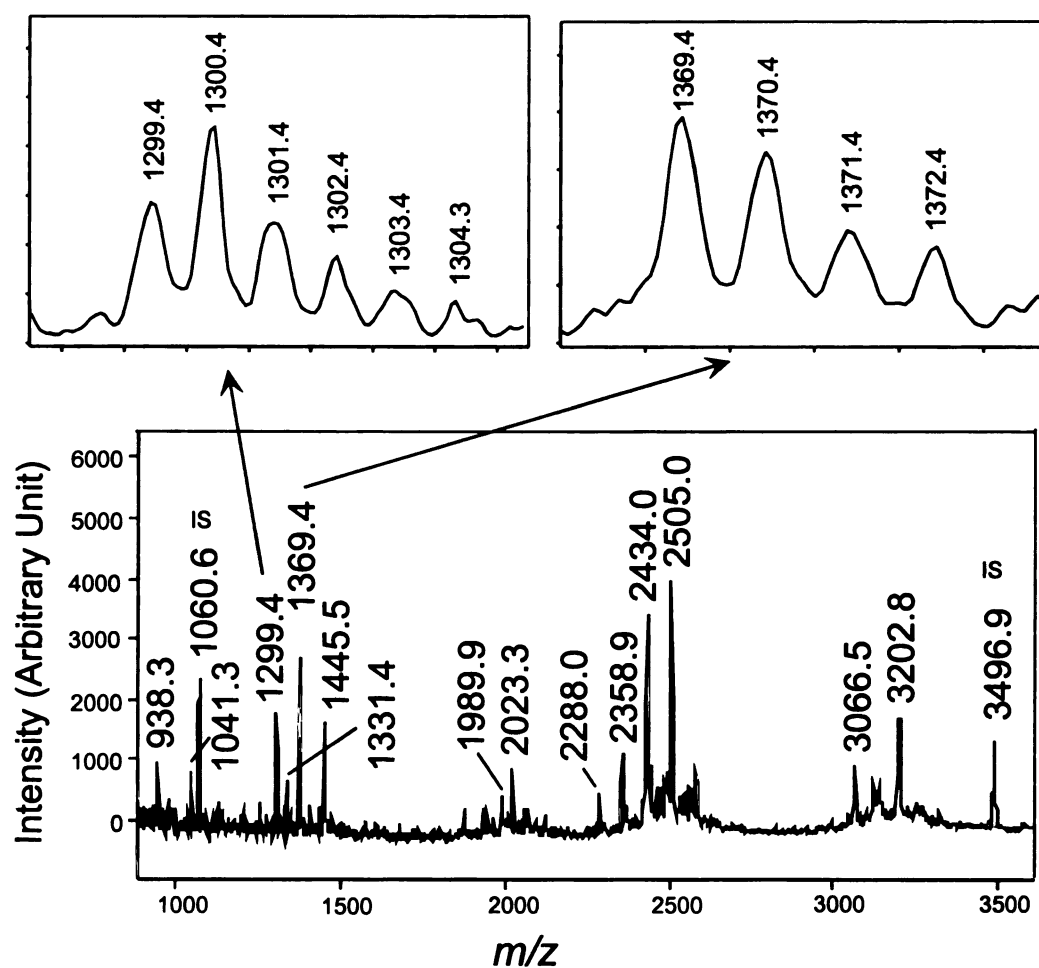


Figure 3.9. The MALDI mass spectrum of the peptide mixture resulting from CN-induced cleavage (before complete reduction) of the doubly reduced/cyanylated sillicin species represented by HPLC peak 2 in Figure 3.4.

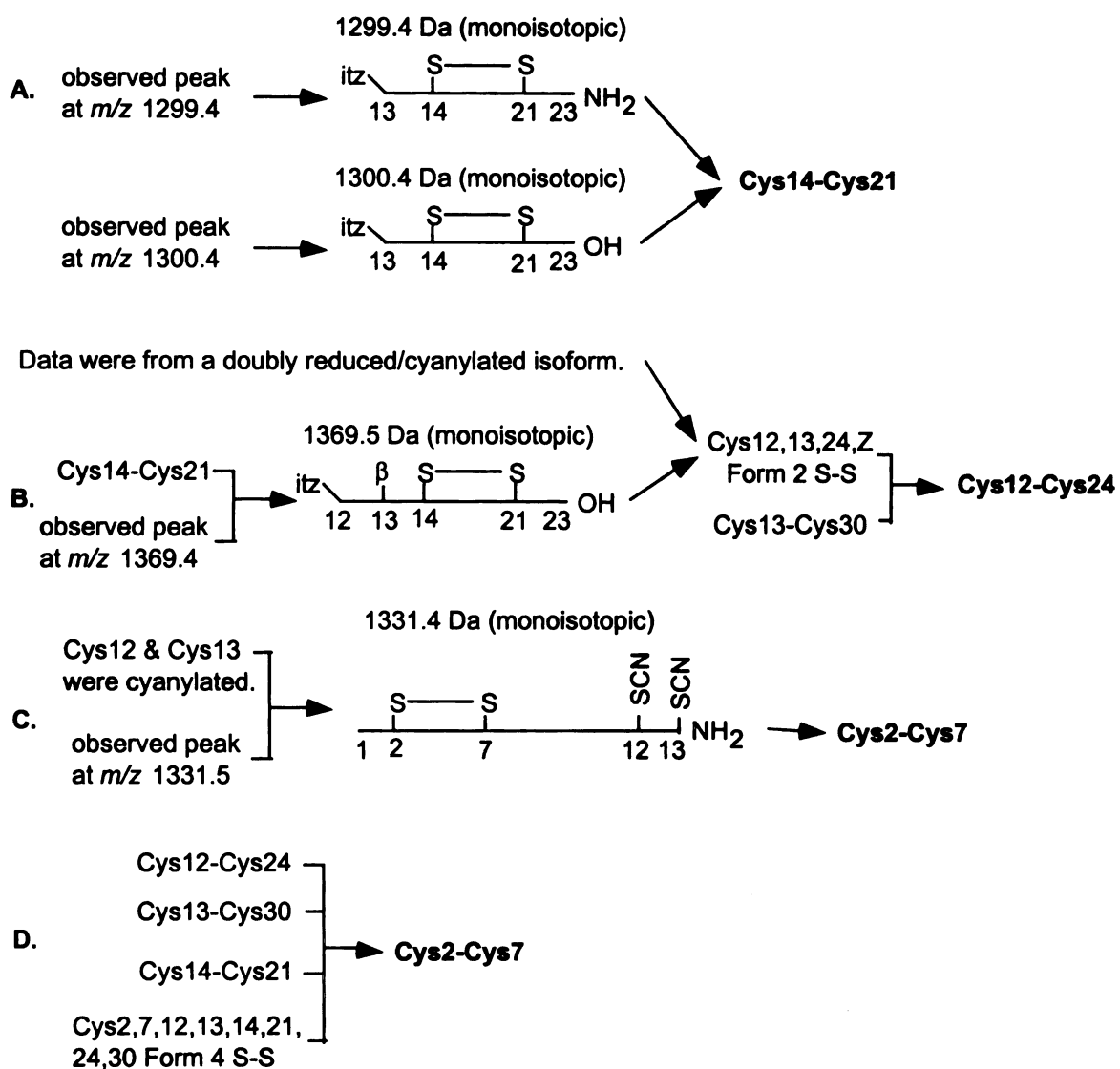


Figure 3.10. Assignment of the disulfide bonds Cys14-Cys21, Cys12-Cys24, and Cys2-Cys7.

The MALDI-MS spectrum in Figure 3.11 was obtained from analysis of the products from **CN-induced cleavage and subsequent complete reductions** of the doubly reduced/cyanylated species (HPLC fraction 2 in Figure 3.4). The MALDI peak series at m/z 1301.5 (Figure 3.11) is 2 mass units higher than the corresponding peak at m/z 1299.4 (Figure 3.9), which represents the oxidized form itz-13-23-ox-NH₂ (calculated mono isotopic mass of MH⁺ = 1299.4 Da); thus, the peak at m/z 1301.5 can be attributed to itz-13-23-red-NH₂ (calculated monoisotopic mass of MH⁺ = 1301.5 Da), in which the disulfide bond Cys14-Cys21 had been reduced during the complete reduction step. This mass spectrometric evidence of a shift of 2 Daltons upon reduction of itz-13-23-ox-NH₂ to itz-13-23-red-OH confirms the existence of the disulfide bond Cys14-Cys21. Similarly, the MALDI peak series at m/z 1371.5 can be assigned as representing itz-12-β@13-23-red-OH (calculated monoisotopic mass of MH⁺ = 1371.5 Da), which results from reduction of the disulfide bond Cys14-Cys21 in itz-12-β@13-23-ox-OH (observed at m/z 1369.4; calculated monoisotopic mass of MH⁺ = 1369.5 Da).

In summary, the linkage of four disulfide bonds in sillucin is identified as Cys2-Cys7, Cys12-Cys24, Cys13-Cys30, and Cys14-Cys21, by partial reduction, cyanylation, and CN-induced cleavage with or without complete reduction, followed by mass mapping analyses.

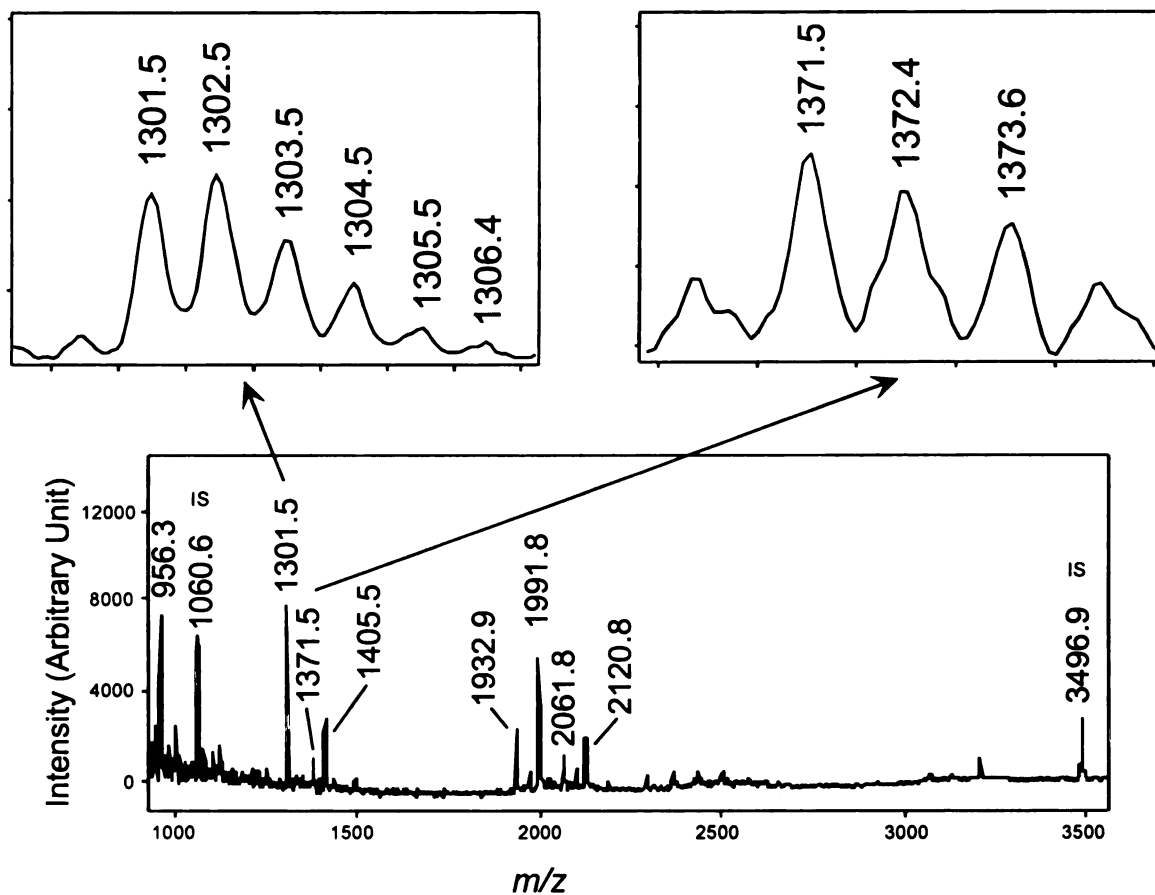


Figure 3.11. The MALDI mass spectrum of the peptide mixture resulting from CN-induced cleavage and subsequent complete reduction of the doubly reduced/cyanylated sillicin species represented by HPLC peak 2 in Figure 3.4.

IV. Discussions

1. Partial Reduction

Obtaining several partially reduced species of a multi-disulfide peptide is essential for successful determination of the disulfide linkages. All the disulfide linkage information is lost by generating the completely reduced peptide species. The rationale for optimizing partial reduction conditions is to generate a reasonable yield of partially reduced peptides with at least one disulfide bond opened, while at least one remains closed (23). For cysteine-rich, knotted peptides, it is best to use a large excess of reducing reagent (TCEP) and control the extent of reduction by varying the time and temperature (23). Partial reduction can be accomplished in a relatively short time (in the range of 2-15 minutes) and stopped by removing the TCEP by HPLC when the reaction mixture is injected while still under kinetic control. As a result, a fair amount of the starting material will be left untouched, while a useful amount of partially reduced peptide species is obtained. The general guidelines for optimization of partial reduction have been established by Gray (23) and Wu (19); usually, only a few experiments are needed (23).

Because eight of the thirty amino acid residues in sillucin are cysteine, including three adjacent cysteines, the peptide appears to be in a knotted arrangement, which offers great resistance to chemical reduction. For this reason, a large excess (10-fold equivalent excess over total disulfide content) was used in the initial trial for optimizing the partial reduction reaction. For the same reason, the temperature of the partial reduction reaction was raised from room temperature to 40°C, and eventually to 50°C. A good yield (25%+) of doubly reduced species of sillucin was produced during the third attempt to optimize the partial reduction reaction by increasing the temperature to 50°C while using a 20-fold

equivalent excess over total disulfide content for 10 min; the relative distribution of products in this reaction mixture is represented by the HPLC chromatogram in Table 3.4. The reaction mixture from the first two optimization trials gave similar chromatograms (data not shown), except with considerably smaller peaks for the singly- and doubly-reduced species.

Information leading to the identification of one disulfide bond (Cys13-Cys30) can be gleaned from mass mapping analysis of the reaction mixture following CN-induced cleavage of a singly reduced species of sillucin after complete reduction of residual disulfide bonds; the key reactions and data are shown in Figure 3.1. Ideally, one would hope to isolate a singly reduced isoform of the intact peptide that corresponds to each of the disulfide bonds; this was the case for ribonuclease A (consisting of eight cysteines in the form of four disulfides), where nearly equal amounts of four singly reduced isoforms were available for mass mapping (19). However, in the case of sillucin, its knotted conformation apparently prevents equal access of the chemical reducing agent to each of the disulfide bonds, and thus only one of the four possible singly reduced isoforms can be readily generated.

In situations where an insufficient number of singly reduced isoforms of the analyte is available, attempts to produce and isolate doubly reduced isoforms of the analyte may solve the problem. Such is the case with sillucin, where as indicated in the HPLC chromatogram in Figure 3.4, a substantial amount of one of the six possible doubly reduced isoforms is available for analysis. As summarized in Figure 3.2, mass mapping analysis of the reaction mixture resulting from CN-induced cleavage before and after subsequent complete reduction to reduce the residual two disulfide bonds allows the

connectivity of three more disulfide bonds (Cys2-Cys7, Cys14-Cys21 and Cys12-Cys24) to be deduced.

2. Incomplete Cyanylation, Incomplete CN-induced cleavage, and Disulfide

Scrambling

Although the efficiency of the cyanylation reaction is usually 95%+ (24), there are various imperfections associated with the cyanylation and CN-induced cleavage reactions, especially when there are three adjacent cysteines to offer extreme steric hindrance for the cyanylation and cleavage reagents to access the middle cysteine residue. In Figure 3.1, a product from a competing side reaction, β -elimination of HSCN from a cyanylated cysteine residue, was shown; depending on the sequence, variable amounts of this β -elimination product are generated in competition with the desired CN-induced cleavage on the N-terminal side of cyanylated cysteine residues. Other side reactions include incomplete cyanylation and incomplete CN-induced cleavage at cysteine residues. Figure 3.12 and Figure 3.13 show some examples of incomplete cyanylation and incomplete CN-induced cleavage reactions. These side reactions, including β -elimination, incomplete cyanylation, and incomplete CN-induced cleavage, can occur in combination. Furthermore, these side reactions contribute additional peaks to the MALDI spectra of the CN-induced cleavage reaction mixture. Sometimes these peaks are minor, indicating little side reaction; in other cases, these peaks dominate as a result of super sensitivity of some of these minor products in the MALDI process or high yields for the side reactions.

Another complication is that the partially reduced/cyanylated species of the peptide are sometimes similar, and carry-over or co-elution of such species with others is possible during HPLC separation. This aspect can also contribute to additional MALDI peaks that are not related to the CN-induced cleavage products of a specific partially reduced/cyanylated species.

Most of the MALDI peaks in Figure 3.5, Figure 3.9, and Figure 3.11, except those described for determination of the disulfide structure of sillucin ("Results" section), can be attributed to the side reactions discussed above. The identities of those peaks are listed in Table 3.4, Table 3.5, and Table 3.6. Because these "extra" peaks can be accounted for, they do not preclude correct data interpretation, but simply complicate it. Some representative MALDI peaks related to side reactions are described below.

Some of the MALDI peaks in Figure 3.5, other than those identified in the "Results" section, can be interpreted as follows. The peak at m/z 1991.7 corresponds to itz-13-SCN@24-29-red-NH₂ (calculated monoisotopic mass of MH⁺ = 1991.8 Da; SCN@24 is used to designate a cyanylated cysteine residue at position 24). Itz-13-SCN@24-29-red could result from incomplete CN-induced cleavage (CN-induced cleavages occur at Cys13SCN, and Cys30SCN, but **NOT** at Cys24SCN) and complete reduction of minor "carry-over" of the doubly reduced/cyanylated species, [sillucin(13SCN-30SCN, 12SCN-24SCN)], in which disulfide bonds Cys13-cys30 and Cys12-cys24 have been reduced and cyanylated (Another partially reduced/cyanylated species is designated similarly in the following discussions.) as represented by HPLC peak 2 in Figure 3.4. Formation of itz-13-SCN@24-29-red-NH₂ is illustrated in Figure 3.12.

Table 3.4. Identities of the MALDI-MS peaks in Figure 3.5.

Experimental <i>m/z</i>	Calculated <i>m/z</i>	CN-induced Fragment	S-S Reduced during Partial Reduction	Side Reactions involved
MALDI Peaks Identified in the “Results” Section				
1966.8 (M) *	1966.8 (M)	itz-13-29-red-NH ₂	13-30	N/A
3072.6 (A)	3072.5 (A)	1-β@13-29-red-NH ₂	13-30	N/A
3202.8 (A)	3202.8 (A)	Intact Sillucin	N/A	N/A
MALDI Peaks Identified in the “Discussion” Section				
1991.7 (M)	1991.8 (M)	itz-13-SCN@24-29-red- NH ₂	13-30, 12-24	No cleavage at SCN@24
2070.8 (M)	2070.8 (M)	itz-13-30-red-OH	13-30	No cyanylation at SH@30
2097.5 (A)	2097.4 (A)	itz-13-SCN@30-red-OH	13-30	No cleavage at SCN@30

* “M” designates monoisotopic mass; “A” designates average mass.

Table 3.5. Identities of the MALDI-MS peaks in Figure 3.9.

Experimental <i>m/z</i>	Calculated <i>m/z</i>	CN-induced Fragment	S-S Reduced during Partial Reduction	Side Reactions involved
MALDI Peaks Identified in the “Results” Section				
1299.4 (M)	1299.4 (M)	itz-13-23-ox-NH ₂ (contains Cys14-Cys21)	12-24, 13-30	N/A
1300.4 (M)	1300.4 (A)	itz-13-23-ox-OH (contains Cys14-Cys21)	12-24, 13-30	N/A
1369.4 (M)	1369.5 (M)	itz-12-β@13-23-ox-OH (contains Cys14-Cys21)	12-24, 13-30	β-elimination at SCN@13
MALDI Peaks Identified in the “Discussion” Section				
1331.4 (M)	1331.5 (M)	1-SCN@12-SCN@13- ox-NH ₂ (contains Cys2 -Cys7)	12-24, 13-30	No cleavage at SCN@12 and SCN@13
1989.9 (M)	1989.8 (M)	itz-13-SCN@24-29-ox- NH ₂ (contains Cys14-Cys21)	12-24, 13-30	No cleavage at SCN@24
3066.6 (A)	3066.5 (A)	1-β@13-29-ox-NH ₂	13-30	β-elimination at SCN@13
3202.8 (A)	3202.8 (A)	Intact sillucin	N/A	N/A

Table 3.6. Identities of the MALDI-MS peaks in Figure 3.11.

Experimental <i>m/z</i>	Calculated <i>m/z</i>	CN-induced Fragment	S-S Reduced during Partial Reduction	Side Reactions involved
MALDI Peaks Identified in the “Results” Section				
1301.5 (M) *	1301.5 (M)	itz-13-23-red-NH ₂	12-24, 13-30	N/A
1302.5 (M)	1302.4 (M)	itz-13-23-red-OH	12-24, 13-30	N/A
1371.5 (M)	1371.5 (M)	itz-12-β@13-23-red-OH	12-24, 13-30	β-elimination at SCN@13
MALDI Peaks Identified in the “Discussion” Section				
1405.5 (M)	1405.5 (M)	itz-12-23-red-OH	12-24, 13-30	No cyanylation at SH@13
1932.9 (M)	1932.8 (M)	itz-13-β@24-29-red- NH ₂	12-24, 13-30	β-elimination at SCN@24
1991.8 (M)	1991.8 (M)	itz-13-SCN@24-29-red- NH ₂	12-24, 13-30	No cleavage at SCN@24
2061.8 (M)	2061.8 (M)	itz-13-β@24-SCN@30- red-OH	12-24, 13-30	β-elimination at SCN@24, no cleavage at SCN@30
		itz-13-SCN@24-β@30- red-OH	12-24, 13-30	β-elimination at SCN@30, no cleavage at SCN@24
2120.8 (M)	2120.8 (M)	itz-13-SCN@24- SCN@30-red-OH	12-24, 13-30	no cleavage at SCN@24 and SCN@30

* “M” designates monoisotopic mass; “A” designates average mass.

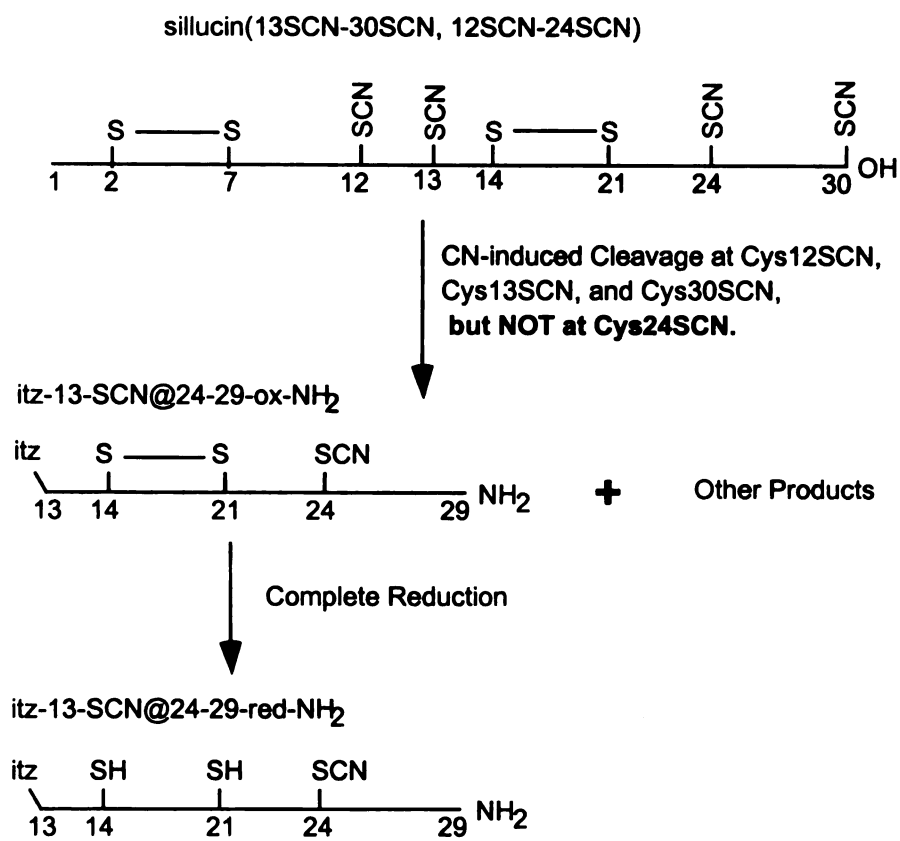


Figure 3.12. Formation of itz-13-SCN@24-29-ox-NH₂ and itz-13-SCN@24-29-red-NH₂ because of no CN-induced cleavage at SCN@24.

The peak at m/z 2070.8 could be attributed to itz-13-30-red-OH or itz-12-29-red-OH (calculated monoisotopic masses of MH^+ ions for both fragments are 2070.8 Da). Itz-13-30-red-OH could result from CN-induced cleavage and complete reduction of the singly reduced/incompletely cyanylated species, **[Sillucin(13SCN-30SH)]** ("SH" designates a free sulfhydryl at Cys 30), which is a minor component in the collected HPLC fraction 1 in Figure 3.4. **[Sillucin(13SCN-30SH)]** (calculated average mass of $MH^+ = 3229.8$ Da; 27 Da heavier than that for intact sillucin.) is represented by a minor MALDI peak at m/z 3229.7 (data not shown). Figure 3.13 illustrates the formation of itz-13-30-red-OH.

Itz-12-29-red-OH can result from CN-induced cleavage and complete reduction of scrambled/cyanylated species, **[Sillucin(12SCN-30SCN)]**, which results from scrambling between Cys 12 and Cys 13 (Figure 3.14).

Further experiments showed that the MALDI peak at m/z 2070.8 disappeared after additional HPLC separation removed the minor component **[Sillucin(13SCN-30SH)]** from HPLC fraction 1 in Figure 3.4; the purified species **[Sillucin(13SCN-30SCN)]** was cleaved and the remaining disulfide bonds were completely reduced (data not shown). This infers that no scrambling occurred; the peak at m/z 2070.8 represents itz-13-30-red-OH, which is a CN-induced cleavage product of the singly reduced/incompletely cyanylated species, **[Sillucin(13SCN-SH)]**. Disulfide scrambling does not occur despite the existence of three adjacent cysteines, probably due to the fact that partial reduction and cyanylation are carried out at pH 3, and, thus, the concentration of thiolate anion is minimal.

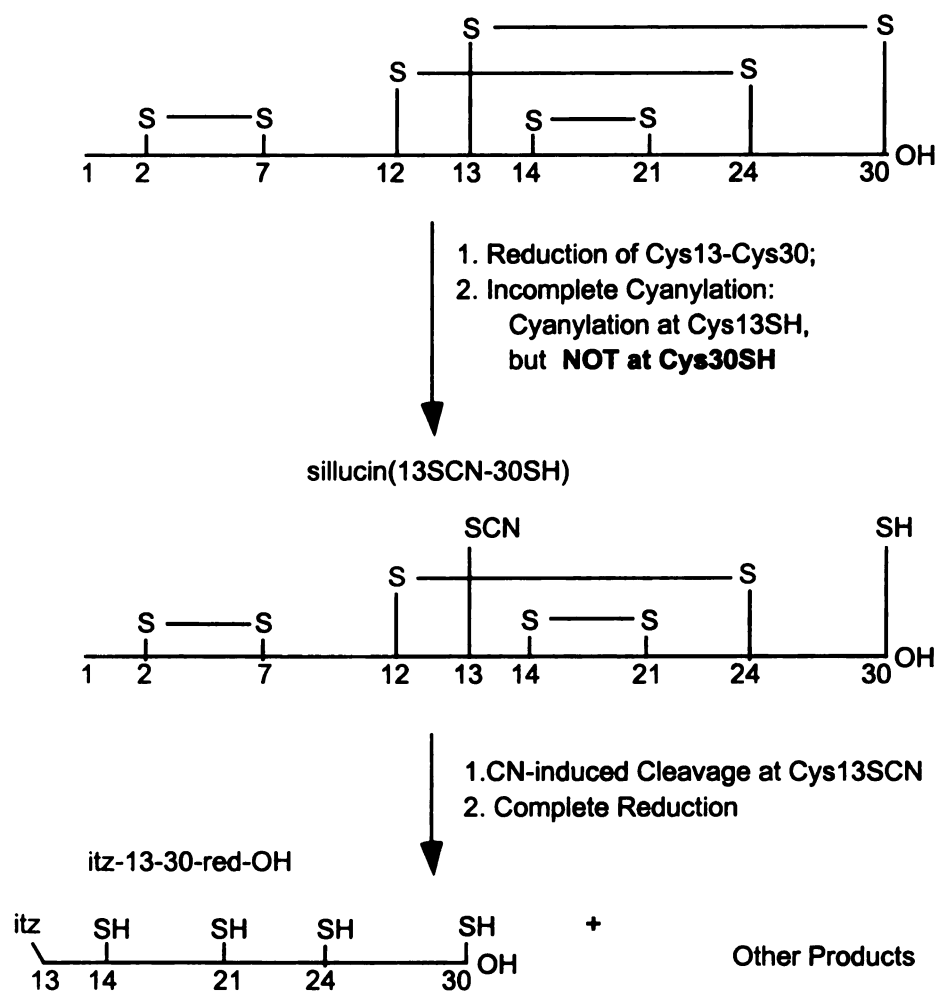


Figure 3.13. Formation of itz-13-30-red-OH due to incomplete cyanylation at Cys 30; Cys30SH is used to represent the free cysteine residue.

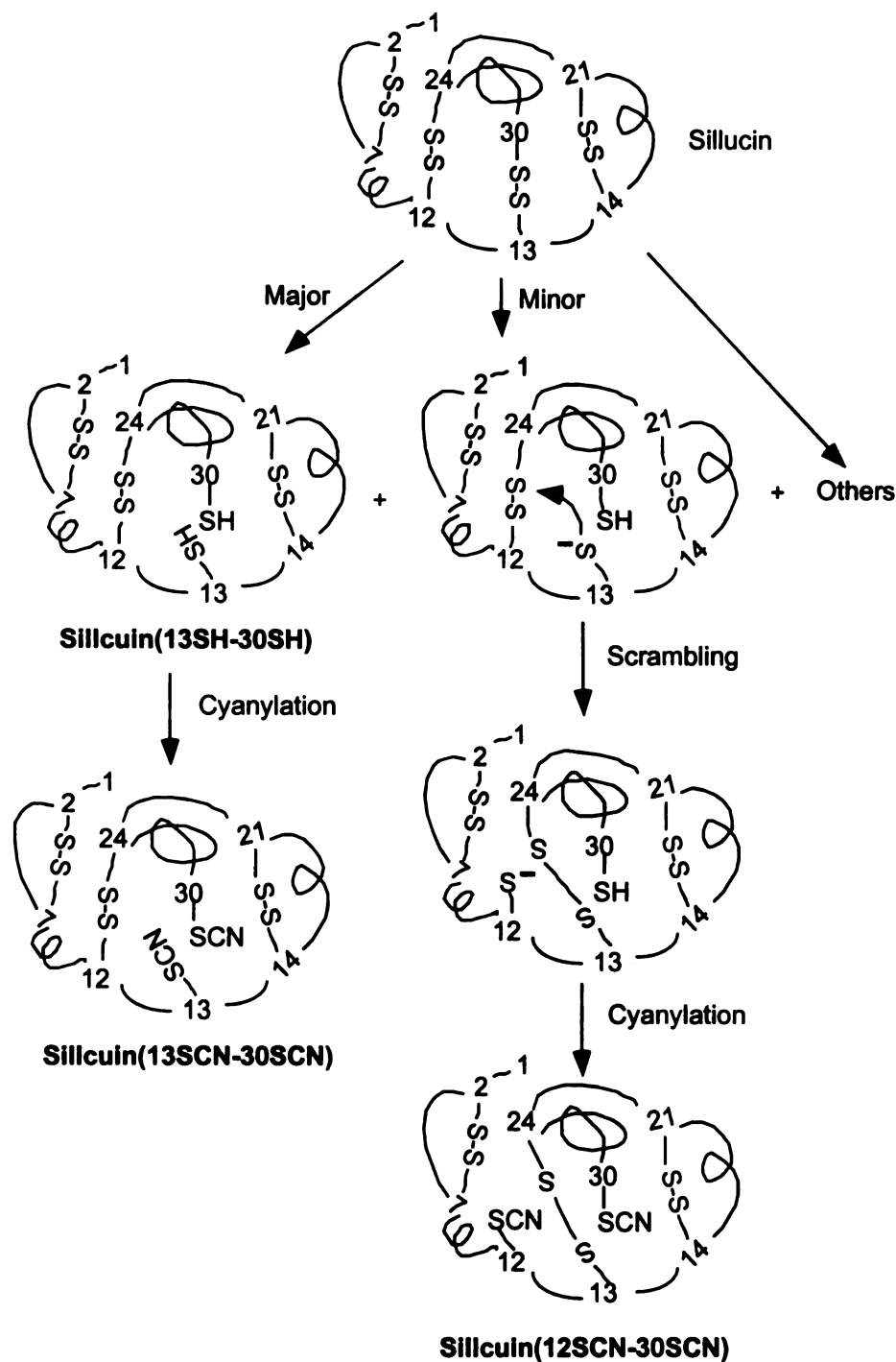


Figure 3.14. Didactic illustration of possible scrambling in a singly reduced species, [sillucin(13SH-30SH)], illustrated for the case in which the disulfide Cys13-Cys30 has been reduced, and a trace amount has dissociated to form the thiolate ion, [sillucin(13S⁻-30SH)], which scrambles to [sillucin(12S⁻-30SH)] before cyanylation is accomplished.

V. Conclusions

In summary, the cysteine-rich and apparently highly knotted anti-microbial peptide sillucin, which is resistant to the proteolytic mapping approach to disulfide mapping, is amenable to partial reduction/cyanilation/CN-induced cleavage mass mapping methodology. One of four possible singly reduced isoforms and one of six possible doubly reduced isoforms of sillucin generated sufficient information from mass mapping of the cyanilation/CN-induced cleavage products to assign the disulfide connectivity (Cys2-Cys7, Cys12-Cys24, Cys13-Cys30, and Cys14-Cys21) and rule out the other 104 isomeric possibilities.

VI. References

1. Somkuti, G. A. & Walter, M. W. Antimicrobial polypeptide synthesized by *Mucor pusillus* NRRL 2543. *Proc. Soc. Exp. Biol. Med.* **133**, 780-785. (1970).
2. Somkuti, G. A. & Greenberg, R. Antimicrobial peptides of thermophilic *Mucor* fungi. *Dev. Ind. Microbiol.* **20**, 661-672 (1979).
3. Bradley, W. A. & Somkuti, G. A. The primary structure of sillucin and antimicrobial peptide from *Mucor pusillus*. *FEBS Lett.* **97**, 81-83. (1979).
4. Hancock, R. E. W. & Lehrer, R. Cationic peptides: a new source of antibiotics. *Trends Biotechnol.* **16**, 82-88 (1998).
5. Ganz, T. & Lehrer, R. I. Antibiotic peptides from higher eukaryotes: biology and applications. *Mol. Med. Today* **5**, 292-297 (1999).
6. Smith, D. L. & Zhou, Z. R. Strategies for locating disulfide bonds in proteins. in *Methods in Enzymology* (ed. McCloskey, J. A.) Vol. **193**, 374-389. (Academic Press, San Diego, CA, 1990).
7. Maeda, K., Wakabayashi, S. & Matsubara, H. Disulfide bridges in an alpha-amylase inhibitor from wheat kernel. *J. Biochem.* **94**, 865-870. (1983).
8. Poerio, E., Caporale, C., Carrano, L., Pucci, P. & Buonocore, V. Assignment of the five disulfide bridges in an alpha-amylase inhibitor from wheat kernel by fast-

- atom-bombardment mass spectrometry and Edman degradation. *Eur. J. Biochem.* **199**, 595-600. (1991).
9. Morris, H. R. & Pucci, P. A New Method For Rapid Assignment of S-S Bridges in Proteins. *Biochem. Biophys. Res. Commun.* **126**, 1122-1128 (1985).
 10. Zhang, D. & Liang, S. Assignment of the three disulfide bridges of huwentoxin-I, a neurotoxin from the spider selenocosmia huwena. *J. Protein Chem.* **12**, 735-740. (1993).
 11. Raschdorf, F., Dahinden, R., Maerki, W., Richter, W. J. & Merryweather, J. P. Location of disulfide bonds in human insulin-like growth factors (IGFs) synthesized by recombinant DNA technology. *Biomed. Environ. Mass Spectrom.* **16**, 3-8 (1988).
 12. Jones, M. D., Patterson, S. D. & Lu, H. S. Determination of disulfide bonds in highly bridged disulfide-linked peptides by matrix-assisted laser desorption/ionization mass spectrometry with postsource decay. *Anal. Chem.* **70**, 136-143 (1998).
 13. Oda, Y., Matsunaga, T., Fukuyama, K., Miyazaki, T. & Morimoto, T. Tertiary and quaternary structures of 0.19 alpha-amylase inhibitor from wheat kernel determined by X-ray analysis at 2.06 angstrom resolution. *Biochemistry* **36**, 13503-13511 (1997).
 14. Hrabal, R., Chen, Z. G., James, S., Bennett, H. P. J. & Ni, F. The hairpin stack fold, a novel protein architecture for a new family of protein growth factors. *Nat. Struct. Biol.* **3**, 747-752 (1996).
 15. Gray, W. R. Disulfide structures of highly bridged peptides: a new strategy for analysis. *Protein Sci.* **2**, 1732-1748. (1993).
 16. Gray, W. R. Echistatin disulfide bridges: selective reduction and linkage assignment. *Protein Sci.* **2**, 1749-1755. (1993).
 17. Yamashita, H., Nakatsuka, T. & Hirose, M. Structural and Functional-Characteristics of Partially Disulfide-Reduced Intermediates Ovotransferrin N-Lobe - Cystine Localization by Indirect End-Labeling Approach and Implications for the Reduction Pathway. *J. Biol. Chem.* **270**, 29806-29812 (1995).
 18. Li, F. & Liang, S. P. Assignment of the three disulfide bonds of Selenocosmia huwena lectin-I from the venom of spider Selenocosmia huwena. *Peptides* **20**, 1027-1034 (1999).
 19. Wu, J. & Watson, J. T. A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci.* **6**, 391-398 (1997).

20. Yang, Y., Wu, J. & Watson, J. T. Disulfide mass mapping in proteins containing adjacent cysteines is possible with cyanylation/cleavage methodology. *J. Am. Chem. Soc.* **120**, 5834-5835 (1998).
21. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass. Spectrom.* **11**, 601 (1984).
22. Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* **16**, 99-111. (1988).
23. Gray, W. R. Disulfide bonds between cysteine residues. in *Protein Structure: A Practical Approach* (ed. Creighton, T. E.), 164-186 (IRL Press at Oxford University, New York, NY, 1997).
24. Wu, J. & Watson, J. T. Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping. *Anal. Biochem.* **258**, 268-276 (1998).

CHAPTER 4
DETERMINATION OF THE CYSTEINE STATUS OF THE CHLOROPLAST
GLYCERALDEHYDE-3-P DEHYDROGENASE SUBUNITS

I. Introduction

The NADP (nicotinamide adenine dinucleotide phosphate)-linked chloroplast glyceraldehyde-3-P dehydrogenases (GAPD) (EC 1.2.1.13) of flowering plants consist of two subunits. The A subunit is similar in length and in sequence to the eucaryotic and eubacterial NAD (nicotinamide adenine dinucleotide) -linked enzyme (EC 1.2.1.12) protomers. The B subunit has a 27- to 29-residue C-terminal extension that contains two invariant cysteines (Figure 4.1). B subunits have only been found in angiosperms. In species ranging from cyanobacteria to angiosperms, NADP-linked glyceraldehyde-3-P dehydrogenases are light activated (1, 2). Activation involves reduction of an inhibiting disulfide bond or bonds. While modeling indicates that two Cys residues in the A subunit are almost certainly responsible for the redox-sensitivity of the enzyme in green algae and non-flowering plants (3), the presence of the conserved Cys pair in the extended angiosperm B subunit suggests that it, too, might be involved in redox-regulation. The purpose of the present experiments was to determine the status of the cysteines in the A and B subunits of pea chloroplast glyceraldehyde-3-P dehydrogenase and which disulfide bond(s) is responsible for the redox-sensitivity of the enzyme.

```

                20
                |
lgd1  AVKVGINGFGRIGRNVFRAAL--KNPDIEVVAVNDLTDANTLAHLLKYDSVHGRLDAEVSVNGNN-
pea a  KQLKVAINGFGRIGRNFLRCWHGRKDSPLDVIAINDTGGVKQASHLLKYDSTLGI FDADV KPVGTDG
pea b  KLKVAINGFGRIGRNFLRCWHGRKDSPLEVIVVNDSGGVKNASHLLKYDSMLGTFKAEVKILNNET
lb7g  MVNVAVNGYGTIGKRVADAII--KQPDMKLVGVAKTSP NYEAFIAHRRG-----
                |
                19

lgd1  LVVNGKEIIVKAERDPENLAWGEI-----GVDIVVEST-GRFTKREDAAKHLEAGAKKVI
pea a  ISVDGKVIKVVSDRNPANLPWKEL-----GIDLVIEGT-GVFVDREGAGRHITAGAKKVL
pea b  ITVDGKPIKVVSSRDPLKLPWAEL-----GIDIVIEGT-GVFVDGPGAGKHIQAGAKKVI
lb7g  -----IRIYVPQQSIKKFEEESGIPVAGTVEDLIKTSDIVVDTT PNGVGAQY-KPIYLQL-QRNAI

                154 158
                |   |
lgd1  ISAPAK--NEDI-TIVMGVNQDKYDPKAHHVISNASCTTNCLAPFAKVLHEQFGIVRGMMTTVHSYT
pea a  ITAPGK--G-DIPTYVVGVNADAYTH-ADDIISNASCTTNCLAPFVKVLDQKFGIIKGTMTTTHSYT
pea b  ITAPAK--GADIPTYVIGVNEQDYGHEVADIISNASCTTNCLAPFAKVLDEEFGIVRGMMTTVHSYT
lb7g  FQGGEKAEVADI-SFSALCNYNEALG--KKYIRVVSCNTTALLRTICTVNKVSKEKVRATIVRRAA

                155 159
                |   |
lgd1  NDQRILDLPHKDLRRAAAAESIIPTT----TGAAKAVALVLPELKGKLNGAMRVPTPNVSVVDLV
pea a  GDQRLLDASHRDLRRAAAAALNIVPTS---TGAAKAVALVLPTLKGKLNGIALRVPTPNVSVVDLV
pea b  GDQRLLDASHRDLRRAAAAALNIVPTS---TGAAKAVSLVLPQLKGKLNGIALRVPTPNVSVVDLV
lb7g  DQKEVKK-----GPINSLVPD PATVPSHHAKDVNS VIRNL-DIATMAVIAPTTLMHMHFIN

                278                289
                |                |
lgd1  AELEKEV-TVEEVNAALKAAEGELKGILAYSEE---PLVSRDYN-----GSTVSSTIDAL
pea a  VQVSKKTF-AEEVNEAFRESAAKELTGILSVCDE---PLVSVDFR-----CTDVSSTVDSS
pea b  VNVAKKGISAEDVNAAFRKAAEGPLKGILDVCDV---PLVSVDFR-----CSDVSTTIDSS
lb7g  ITLKDKV-EKKDILSVLEN-----TPRIVLISS KYDAEATAELVEVARDLKRDRNDIPE-VMIFSD

                280                291
                |                |
lgd1  STMVIDGKMVKVVSWYDNETGYSHRVVDLAAYIASKGL
pea a  LTMVMGDDLVKVIAWYDNEWGYSQRVVDLADIVANNWK
pea b  LTMVMGDDMVKVVAWYDNEWGYSQRVVDLAHLVANKWPGTPKVSGDPLEDFCETNPADEECKVYE
lb7g  SIYVK-DDEVMLMYAVHQESIVVPENIDAIRASMK LMS-----AEDSMRITNESLGILKGYLI

                354                363

```

Figure 4.1. The Sequence alignment of the pea chloroplast glyceraldehyde-3-P dehydrogenase A (gi ["gi" is the National Center for Biotechnology Information protein identification number] 66025) and B (gi120663) subunits (pea a and pea b), the *Sulfolobus solfataricus* glyceraldehyde-3-P dehydrogenase (Protein Data Bank (PDB) file 1b7g), and the *Bacillus stearothermophilus* glyceraldehyde-3-P dehydrogenase (PDB file 1gd1). Identical residues in the B subunit (pea b) and in one or both of the sequences from the Protein Data Bank are bolded. Residues that are superimposable in the two crystal structures are underlined. Cysteines in pea a and pea b are numbered, above the residue for pea a, below for pea b. Dashes indicate gaps introduced to optimize the alignment.

Determining the status of cysteines in proteins includes locating free cysteines that have sulfhydryl groups and disulfide cysteines that are involved in forming disulfide bonds. The conventional methodology usually involve several steps: 1) Modify free sulfhydryl groups in free cysteine residues, usually by an irreversible reaction, such as alkylation; 2) Cleave the protein backbone by enzyme(s) or chemical reagent(s); 3) Separate and identify the cysteine-containing protein fragments by Edman sequencing or mass spectrometry; 4) From the protein fragments that contain modified cysteines, deduce the location of free cysteines; and from the protein fragments that contain intra-molecular disulfide bonds or are connected by inter-molecular disulfide bonds, deduce the disulfide structure (Figure 4.2).

While this methodology was widely used in the past, there are several limitations associated with it. First, since most alkylation's reactions have to be done under alkaline conditions ($\text{pH} > 7$), disulfide/thiol exchange (Figure 3.15 in Chapter 3 and reference 4) might occur and lead to artifactual results. In addition, deducing the location of a certain free cysteine requires the successful identification of the protein fragment containing the modified cysteine. Thus, nearly all of such protein fragments need to be recovered. This may not be always possible due to chemical suppression (Cysteine-containing pieces are typically a very small percentage of a large number of cleaved protein fragments.) and signal suppression during analysis such as mass spectrometry.

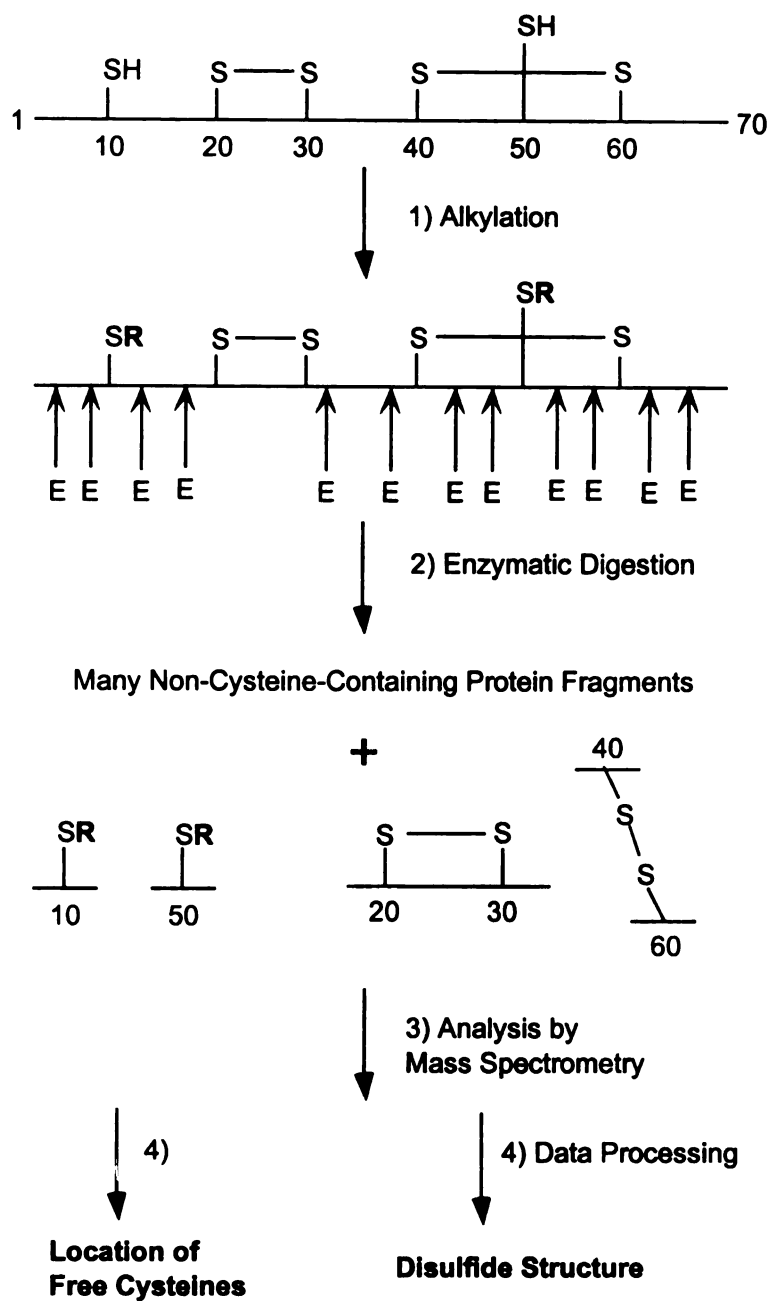


Figure 4.2. The conventional methodology for determining the status of the cysteines in a hypothetical protein of 70 amino acid residues. “SR” designates an alkylated cysteine residue; and “E” together with an arrow indicates where the enzyme cleaves.

An alternative mass mapping methodology based on cyanylation and the following CN-induced cleavage was introduced to determine the location of the free cysteines in proteins (reference 5 and Figure 4.3). Since cyanylation by reagent 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoro-borate can be carried out under acidic conditions (pH 3), disulfide/thiol exchange (reference 4, Figure 3.15) is minimized. Disulfide/thiol exchange usually occurs only after the protein is denatured and free cysteines move more “freely” towards disulfide bonds. Therefore, denaturing the protein with a large excess of the cyanylation reagent CDAP may further minimize disulfide/thiol exchange.

As shown in Figure 2.9 (in chapter 2 and reference 6), cyanylation of the sulfhydryl group of a single free cysteine residue in a peptide followed by CN-induced cleavage in aqueous ammonia will produce two fragments (6-8). Recovery and identification of only one of the two fragments will be sufficient to determine the location of the free cysteine. Furthermore, cyanylation and the following CN-induced cleavage lead to a much simpler mixture compared to that resulting from enzymatic digestion. If there are n free cysteines in the peptide, only $n+1$ CN-induced fragments will be produced by the cyanylation and CN-induced cleavage process. Therefore, the mass mapping methodology based cyanylation and CN-induced cleavage may be less vulnerable to chemical and signal suppressions.

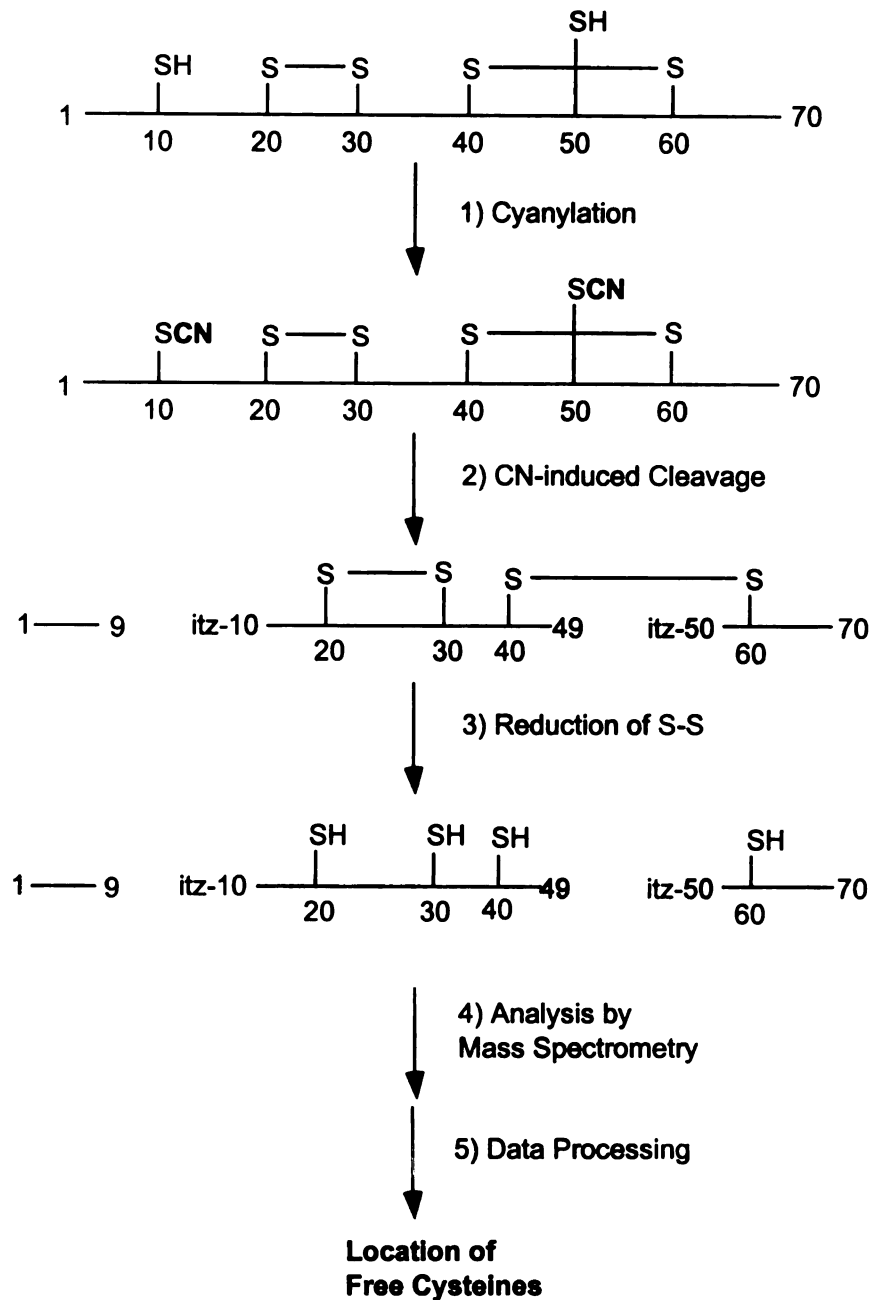


Figure 4.3. Cyanylation and CN-induced cleavage process to identify free cysteines in a hypothetical protein of 70 amino acid residues.

II. Materials and Methods

Enzyme Isolation

The wild type NADP-linked glyceraldehyde-3-P dehydrogenase used in the mass mapping experiments was purified from pea plants by a modification of the method of Anderson et al. (9). Prior to the acetone fractionation step, the suspended MgCl_2 -Polyethylene Glycol (PEG) fraction was passed through a Diethylaminoethyl (DEAE) Sepharose column (10 mL Sigma Fast Flow DEAE) in 10 mM KHPO_4 , pH 7.8 buffer. The glyceraldehyde-3-P dehydrogenase in this fraction was not retained on the column. It emerged in a milky fraction. After acetone fractionation, the cloudy 50% acetone supernatant was applied to a second 10-mL DEAE column in 10 mM Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), pH 7.8, the column was washed overnight with the Tris-HCl buffer and eluted with a linear phosphate gradient (200 mL 10-mM K_2HPO_4 , pH 7.8 and 200 mL 0.4-M K_2HPO_4 , pH 7.0). In some experiments, including those represented by Figure 4.4 and Figure 4.5, 10 mM mercaptoethanol was included in the buffers. In other experiments, including the experiments represented by Figure 4.6 and Figure 4.7, mercaptoethanol was omitted. Essentially identical results were obtained with or without inclusion of mercaptoethanol in the buffers during isolation. Chromatography on a Phenyl Sepharose column (10) was added as a final step in the purification of the enzyme used in the mass mapping experiments reported here. This isolation procedure resulted in the co-purification of the glyceraldehyde-3-P dehydrogenase and an 8.3-kDa protein, which is apparently the small chloroplast protein designated CP12 (11, 12).

Cyanylation, HPLC separation and CN-induced cleavage

To 150 μ g of enzyme in 50 μ L of 50-mM pH 7.0 ammonium acetate buffer (approximately 3.8 nmol of A and B subunits in total) was added 2,400 nmol of 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate (in 24 μ L of 0.1-M pH 3.0 sodium citrate buffer, 12.0 M guanidine hydrochloride; this represents approximately a 120-fold equivalent excess of CDAP to the free sulfhydryl content). Cyanylation of the free sulfhydryl groups was accomplished during incubation at room temperature for 30 minutes.

The mixture of cyanylated A and B subunits of GAPD was separated by reversed phase high performance liquid chromatography (HPLC) at a flow rate of 1.0 mL/min using gradient elution (10% B @0-5 min $\rightarrow$ 40% B @20 min $\rightarrow$ 80% B @60 min) (solvent B was 90% (v/v) acetonitrile/0.1% trifluoroacetic acid (TFA); solvent A was 0.1% TFA in water) with Waters model 6000 pumps. UV detection was done at 215 nm. A Vydac C18 (catalog number 218TP54, 5- μ m particle size, 300-Å pores, 4.6 $\times$ 250 mm) column was used. Fractions (approximately 0.8~1.0 mL) were collected manually. The masses of the collected species in each fraction were determined by matrix-assisted laser desorption/ionization (MALDI) time of flight (TOF) mass spectrometry (MS).

After reduction of the volume (to less than 5 μ L) under partial vacuum, the HPLC fractions containing the cyanylated A and B subunits of GAPD were reconstituted in 20 μ L aliquots of 1.0 M aqueous ammonia. CN-induced cleavage of the peptide chain on the N-terminal side of the cyanylated cysteine residues (8, 6) was accomplished during reaction at room temperature for one hour. Excess ammonia was removed under

vacuum. After addition of 40 μ L of 30% (v/v) acetonitrile/0.1% TFA to the dried cleavage products, the samples were analyzed by MALDI-TOF-MS.

MALDI-TOF-MS

MALDI mass spectra were obtained on a Voyager DE-STR or DE-Elite mass spectrometer (Perkin-Elmer Biosystems Inc., Framingham, MA) equipped with a 337-nm nitrogen laser. The accelerating voltage in the ion source was set at 25 kV. Mass spectra were acquired in the positive linear mode of operation. For analysis of the intact enzyme subunits, the mass-to-charge ratio (m/z) scale was calibrated externally using bovine ribonuclease A (average mass of singly protonated molecule (MH^+): 13683.2 Da) and bovine serum albumin (average mass of MH^+ : 66431 Da) obtained from Sigma Chemical Co. (St. Louis, MO). To analyze mixtures of CN-induced cleavage fragments, bovine ribonuclease A was added to the sample and the m/z scale was internally calibrated on the singly- and doubly- protonated ribonuclease A peaks (m/z values are 13683.2 and 6842.1, respectively). The m/z values of the two most intense peaks were then determined. Because mixing ribonuclease A with the sample caused signal suppression, several less sensitive components in the sample were not detected when the sample was mixed with ribonuclease A. These less sensitive components were detectable if the sample alone was analyzed by MALDI-MS. In order to determine the accurate masses of those less sensitive components, the mass spectrum of the sample alone was acquired and calibrated with the accurate m/z values of the two most intense peaks that were determined as described above.

Analysis of the intact and cyanylated enzyme subunits was performed using sinapinic acid (Aldrich Chemical Co., Milwaukee, WI); all other analyses were performed using α -cyano-4-hydroxycinnamic acid (Aldrich Chemical Co.) as the matrix. Saturated matrix solutions were prepared in a 50% (V/V) aqueous solution of acetonitrile/0.2% TFA. Just prior to analysis, 2 μ L of matrix solution and 2 μ L sample solution were mixed in a vial; then 2 μ L of the mixed solution were applied to a stainless steel sample plate and allowed to air-dry before being introduced into the mass spectrometer.

III. Results

1. Mass Mapping Based on Cyanylation and CN-induced Cleavage

When the enzyme purified as described in 'Materials and Methods' was subjected to MALDI-TOF-MS, three species were detected: mass spectral (MS) peaks at m/z 36218.9, 18101.8, and 12125.2 corresponding to the singly, doubly, and triply protonated A subunit of GAPD (calculated mass: 36185.4), and MS peaks at m/z 39341.3, 19691.8 and 13122.5 corresponding to the singly, doubly, and triply protonated B subunit of GAPD (calculated mass: 39304.1 Da) [Figure 4.4]. Proteins with molecular weights of 8.0 kDa, 8.3 kDa, and 16.5 kDa were also detected. After cyanylation and HPLC separation on a C18 column, four proteins were found, two corresponding to cyanylated products of an 8.3-kDa protein (HPLC peak 1 and 2), one corresponding to the cyanylated B subunit of GAPD (HPLC peak 4) and one corresponding to the cyanylated A subunit (HPLC peak 5) [Figure 4.5] as determined by MALDI-MS analysis. The ratio of the area under the HPLC peaks corresponding to the A and B subunits was 1:1.7.

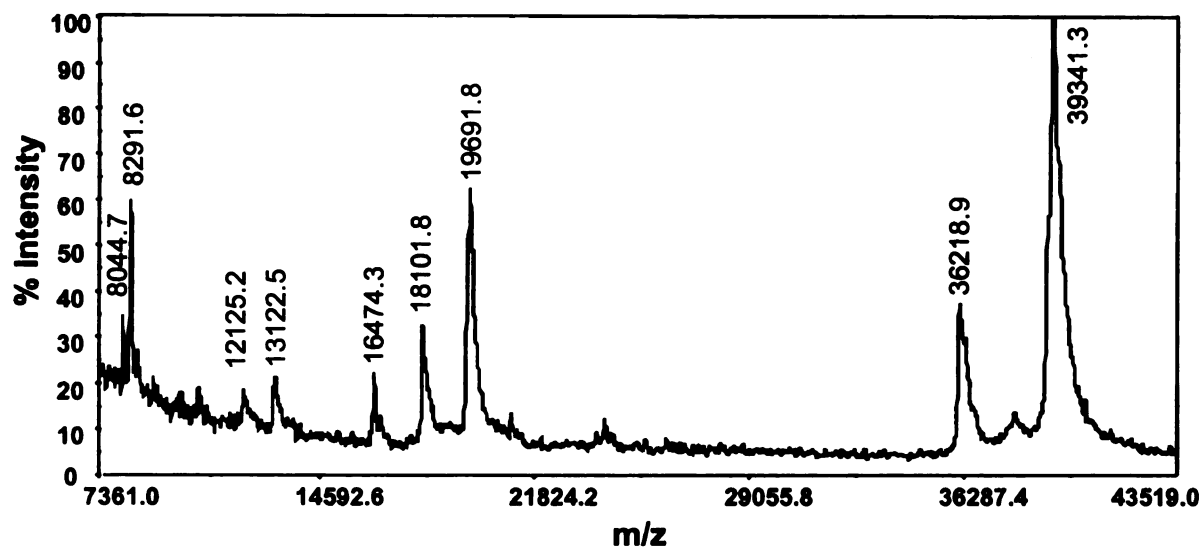


Figure 4.4. The MALDI mass spectrum of the chloroplast glyceraldehyde-3-P dehydrogenase. 1 μ L of a solution of 9 pmol enzyme subunits in 12 mM pH 7-ammonium acetate buffer was used.

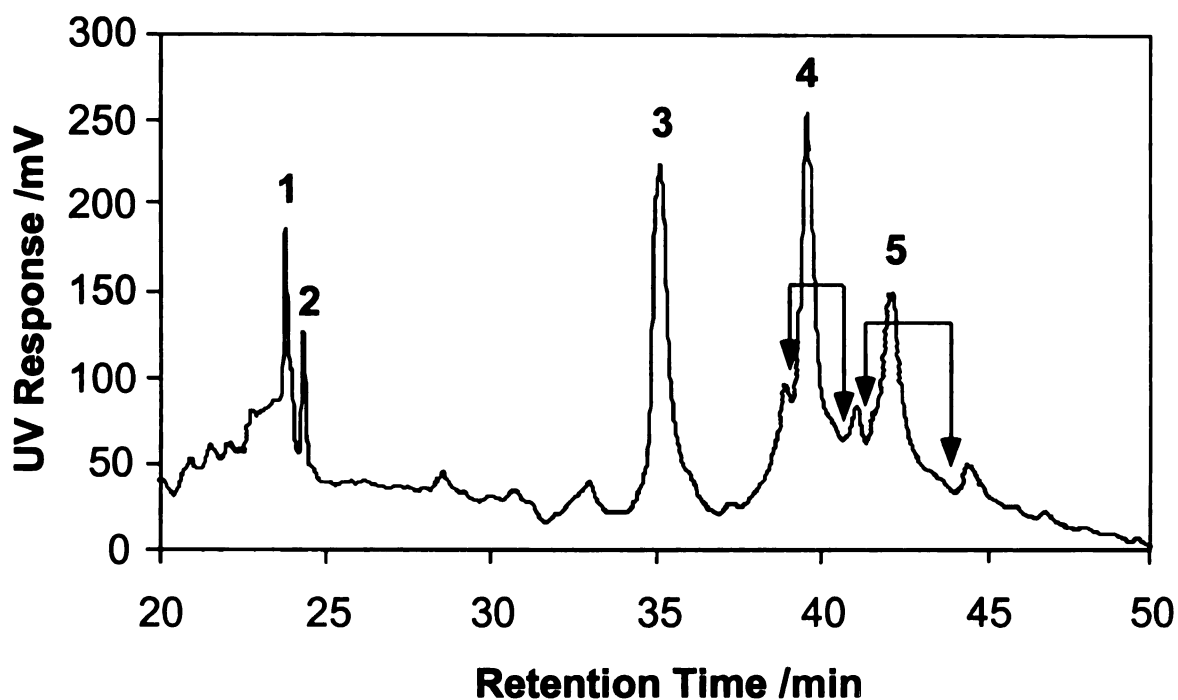


Figure 4.5. The HPLC chromatogram of the cyanylated A and B subunits of chloroplast glyceraldehyde-3-P dehydrogenase (GAPD). The cyanylated proteins (approximately 3.8 nmol total of the A and B subunits) were separated by reverse phase HPLC on a Vydac C18 column. UV detection was done at 215 nm. For details see 'Methods'. Peaks 1 and 2 represent an 8.3k-Da protein; peaks 4 and 5 represent the cyanylated B and A subunits of GAPD, respectively, as determined by MALDI-TOF-MS. Nothing was detected from the fraction (represented by peak 3) by MALDI-TOF-MS. The arrows indicate the starting and stopping points for collecting fractions represented by peaks 4 and 5, respectively.

As shown in Figure 4.6 and Table 4.1, mass-mapping analysis of the cleavage products of the cyanylated A subunit of GAPD indicated the presence of CN-induced fragments 1-19, itz-20-153, itz-158-277, itz-278-288, and itz-289-337 (where itz = iminothiazolidine-carboxyl blocked amino terminus). According to the CN-induced cleavage reaction shown in Figure 2.9 in Chapter 2, 1-19 resulted from CN-induced cleavage on the N-terminal side of cyanylated Cys 20; and thus, in the original A subunit of GAPD, Cys 20 was a free cysteine that was not involved in a disulfide bond. Similarly, Cys 154, Cys 158, Cys 278, and Cys 289, were identified as free cysteines in the original A subunit. In summary, all five cysteines in A subunit are free cysteines.

Mass mapping of the CN-induced cleavage fragments of the cyanylated B subunit indicated that there were two different subunit B species in the original sample, namely the oxidized form and the reduced form. The presence of 1-18, itz-19-154, itz-159-279, itz-280-290, and itz-291-367 (Figure 4.7 and Table 4.2) in the cleavage reaction mixture indicates the presence of the oxidized form of subunit B, which has a disulfide bond between Cys 354 and Cys 363 and free sulfhydryls at five other cysteines residues. The presence of itz-291-367 provides evidence for the Cys 354 - Cys 363 disulfide bond; otherwise, the fragment itz-291-367 would not exist. Cyanylation and CN-induced cleavage would have occurred at Cys 354 and Cys 363 had the cysteines been free. The mass of the singly-protonated fragment, itz-291-367, resulting from CN-induced cleavage at cyanylated Cys 291, was determined to be 8570.5 ± 0.3 (8570.5 is the average of 13 measurements using calibration by internal standards; 0.3 is the confidence interval at the level of 95%). Further, the data confirm that itz-291-367 (see Figure 4.8 for the structure of itz-291-367) contains a disulfide bond, namely Cys 354 - Cys 363 (with overall

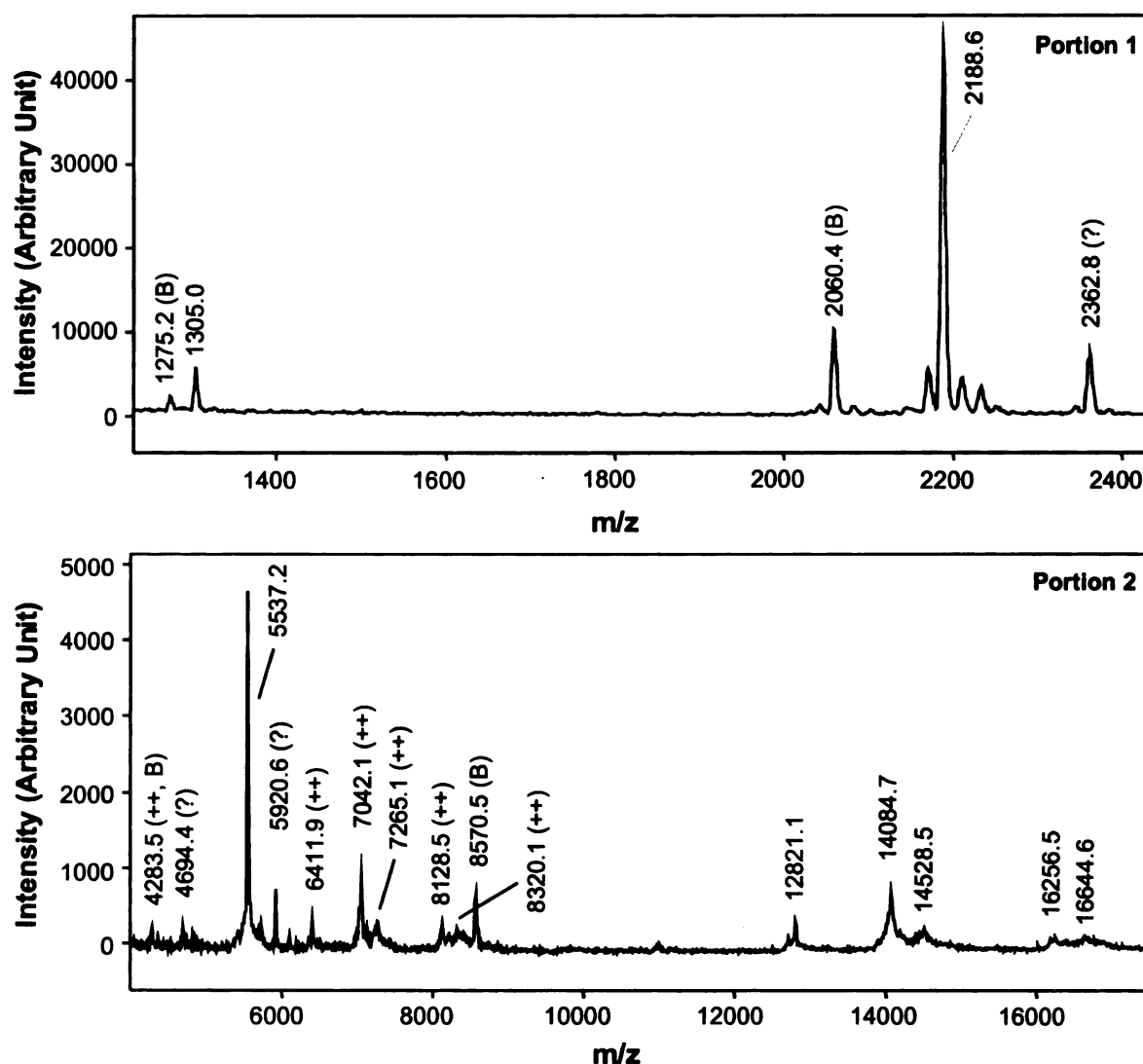


Figure 4.6. The MALDI mass spectrum of the CN-induced cleavage fragments of the cyanylated A subunit of chloroplast glyceraldehydes-3-P dehydrogenase represented by HPLC peak 5 in Figure 4.5. Portions 1 and 2 are the low-mass and high-mass portions, respectively. The mark "B" in parenthesis indicates that the peak represents a CN-induced cleavage fragment of the cyanylated B subunit because HPLC peaks 4 and 5 (Figure 4.5) were not completely resolved. The mark "++" designates a peak representing a doubly protonated (also doubly charged) protein fragment. All other peaks without such marks are singly protonated protein fragments. The mark "?" indicates that the source of the peak is not identified.

Table 4.1. The experimental and calculated m/z values for the protonated CN-induced cleavage fragments of the cyanylated GAPD subunit A.

CN-induced Fragment	Calculated m/z	Experimental m/z
1-19	2188.6 (+1) ^a	2188.6
itz-20-153	14087.9 (+1)	14084.7
	7044.5 (+2)	7042.1
itz-154-157	463.5 (+1)	Not detected ^b
itz-158-277	12824.0 (+1)	12821.1
	6412.5 (+2)	6411.9
itz-278-288	1305.5 (+1)	1305.0
itz-289-337	5537.2 (+1)	5537.2

^a '+1' designates singly-protonated (also singly-charged) species and '+2' designates doubly protonated (also doubly charged) species.

^b Due to signal suppression by matrix molecules in MALDI.

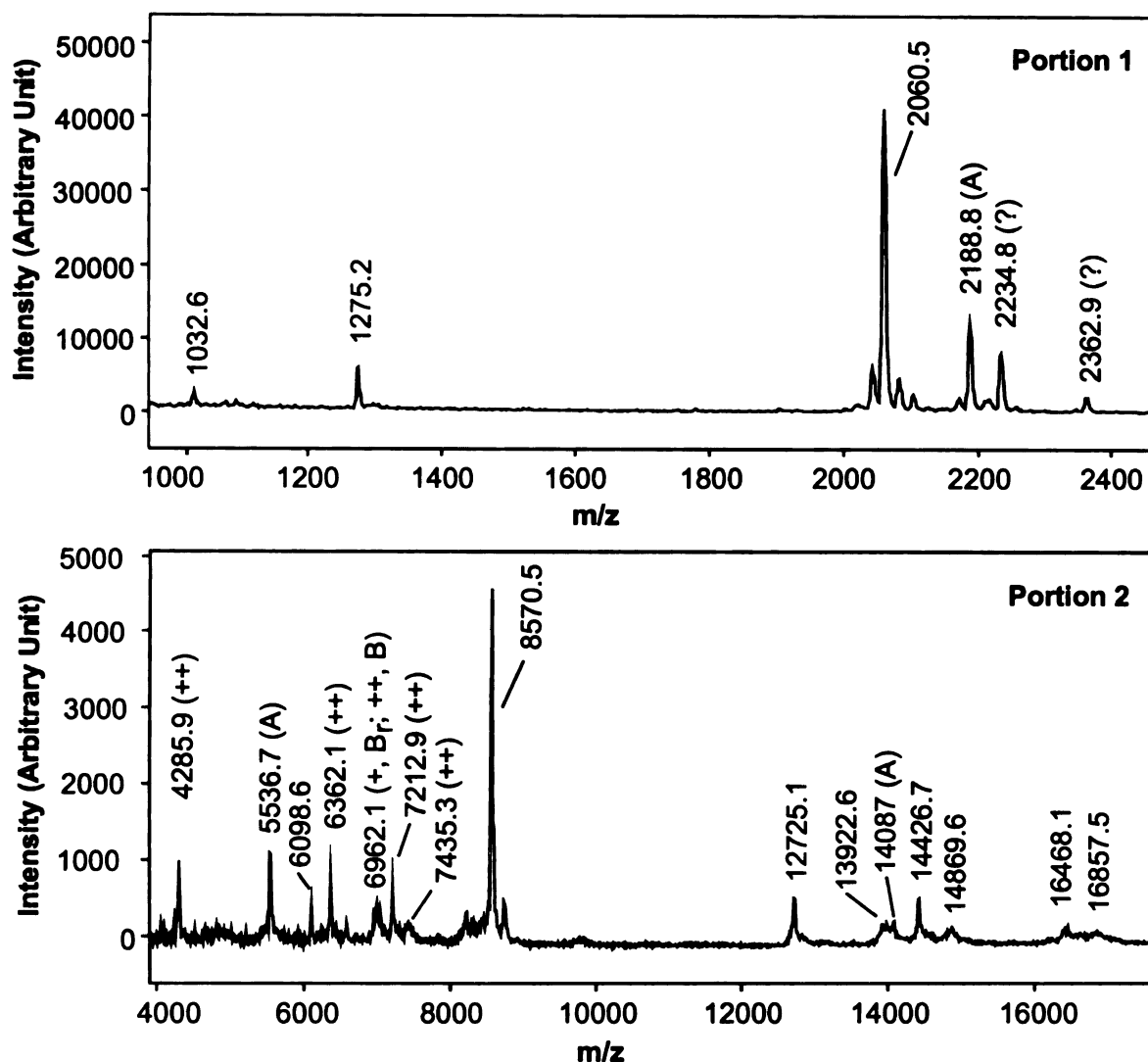


Figure 4.7. The MALDI mass spectrum of the CN-induced cleavage fragments of the cyanylated B subunit of chloroplast glyceraldehyde-3-P dehydrogenase represented by HPLC peak 4 in Figure 4.5. Portions 1 and 2 are the low-mass and high-mass portions, respectively. The mark “A” in parenthesis indicates that the peak represents a CN-induced cleavage fragment of the cyanylated A subunit because HPLC peaks 4 and 5 (Figure 4.5) were not completely resolved. The mark “++” designates a peak representing a doubly protonated (also doubly charged) protein fragment. All other peaks without such marks represent singly protonated protein fragments. The mark “?” indicates that the source of the peak is not identified. The peak at m/z 6962.1 represents both a singly protonated CN-induced cleavage fragment (itz-291-353, marked as “+, B_r”; calculated m/z for MH^+ : 6960.8) from the cyanylated subunit B-reduced and a doubly protonated CN-induced cleavage/ β -elimination fragment (itz-159- β @280-290) [Table 4.2] formed by β -elimination at the cyanylated Cys280 instead of cleavage (marked as “++, B”; calculated m/z for MH_2^{++} : 6962.6).

Table 4.2. The experimental and calculated m/z values for the protonated CN-induced cleavage fragments of the cyanylated GAPD subunit B-oxidized

CN-induced Fragment	Calculated m/z	Experimental m/z
1-18	2060.5 (+1) ^a	2060.5
itz-19-154	14426.5 (+1)	14426.7
	7213.8 (+2)	7212.9
itz-155-158	463.5 (+1)	Not detected ^b
itz-159-279	12726.8 (+1)	12725.1
	6363.9 (+2)	6362.1
itz-280-290	1275.5 (+1)	1275.2
itz-291-367 ^c	8570.5 (+1)	8570.5
	4285.3 (+2)	4285.9

^a '+1' designates singly-protonated (also singly-charged) species and '+2' designates doubly-protonated (also doubly-charged) species.

^b Due to signal suppression by matrix molecules in MALDI.

^c There is a disulfide bond (Cys 354 - Cys363) in itz-291-367.

calculated mass of 8570.5) rather than two free sulfhydryls, at Cys 354 and Cys 363 (with overall calculated mass of 8572.5). Detection of 1-18, itz-19-154, itz-159-279, itz-280-290, itz-291-353, and itz-354-362 (Figure 4.7 and Table 4.3) indicates the presence of the reduced form of subunit B, which has free sulfhydryls at all seven cysteine residues. The CN-induced cleavage reactions of the oxidized and reduced forms of subunit B are illustrated in Figure 4.8.

Because GAPD was simultaneously denatured in 4 M guanidine hydrochloride and cyanylated at acidic pH (pH 3), the possibility of scrambling between thiolate ions (from ionization of the free cysteine sulfhydryl groups) and a disulfide bond was minimized. The experiments reported here demonstrate that it is feasible to determine the cysteine status of various forms (such as partially oxidized and reduced forms of GAPD subunit B) of a protein using methodology based on cyanylation, CN-induced cleavage, and mass mapping, even when these different forms cannot be separated by HPLC.

As is apparent from examining Figure 4.6 and Table 4.1, the expected cleavage fragment itz-154-157 from the cyanylated A subunit was not detected by MALDI-MS because of suppression by matrix molecules used in MALDI. However, determination of the cysteine status was still possible because there were two CN-induced cleavage fragments related to cleavage at one cyanylated cysteine residue: one to the N-terminal side and the other to the C-terminal side of the cleavage site. Detecting only one of the two CN-induced cleavage fragments will allow for the identification of the cyanylated cysteine residue. This methodology, then, does not require complete detection of all the CN-induced cleavage fragments. Similar cases of signal suppression are shown in Figure 4.7, Table 4.2 and Table 4.3.

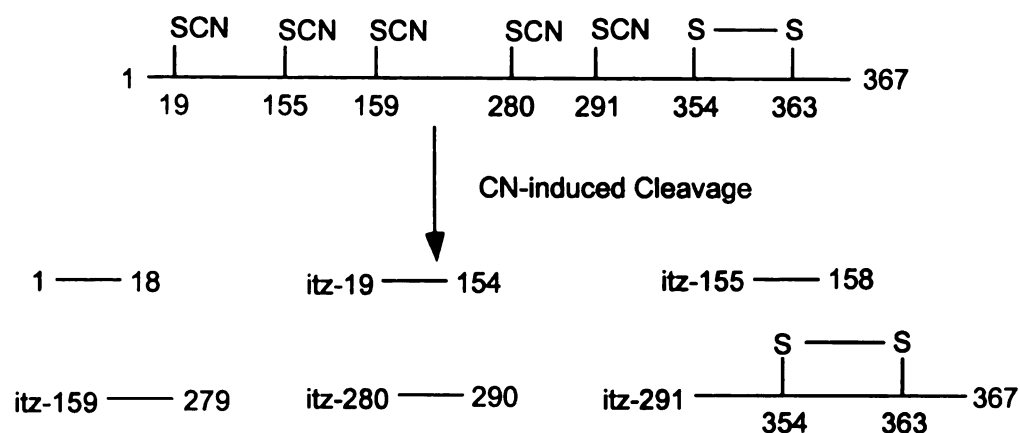
Table 4.3. The experimental and calculated m/z values for the protonated CN-induced cleavage fragments of the cyanylated GAPD subunit B-reduced.

CN-induced Fragment	Calculated m/z	Experimental m/z
1-18	2060.5 (+1) ^a	2060.5
itz-19-154	14426.5 (+1)	14426.7
	7213.7 (+2)	7212.9
itz-155-158	463.5 (+1)	Not detected ^b
itz-159-279	12726.8 (+1)	12725.1
	6363.9 (+2)	6362.1
itz-280-290	1275.4 (+1)	1275.2
itz-291-353	6960.8 (+1)	6962.1
itz-354-362	1033.0 (+1)	1032.6
itz-363-367	666.8 (+1)	Not detected ^b

^a '+1' designates singly-protonated (also singly charged) species and '+2' designates doubly-protonated (also doubly charged) species.

^b Due to signal suppression by matrix molecules in MALDI.

A) The Cyanylated GAPD-Subunit-B-Oxidized-Form



B) The Cyanylated GAPD-Subunit-B-Reduced-Form

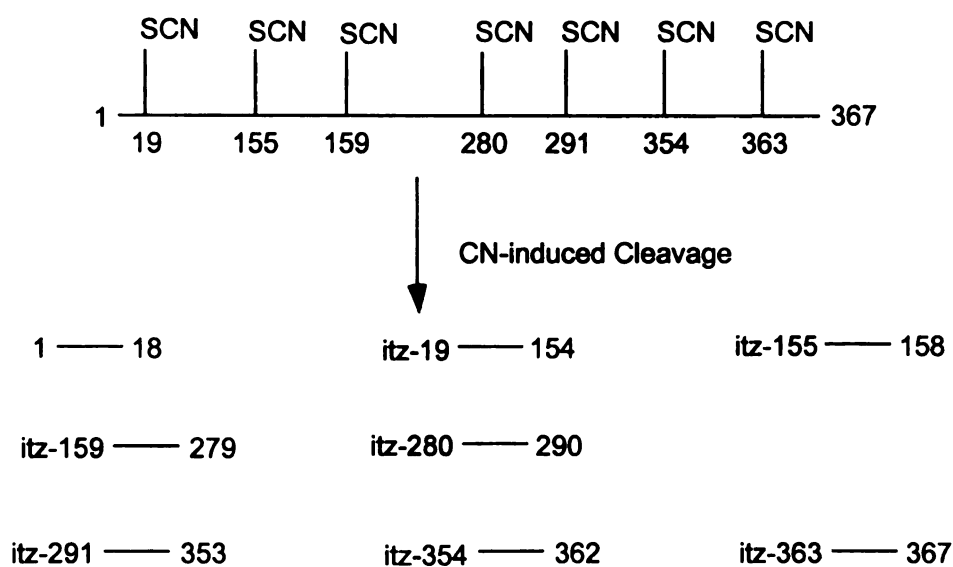


Figure 4.8. The CN-induced cleavage fragments from the oxidized and reduced forms of the cyanylated subunit B of GAPD.

In Figure 4.5, HPLC peaks 4 and 5 (corresponding to the cyanylated B subunit and A subunit, respectively) were not completely resolved. This is reflected in the MALDI spectra [Figure 4.6 and Figure 4.7] of the CN-induced cleavage products of these two fractions, each of which contains minor peaks representing fragments corresponding to the CN-induced cleavage products of the other fraction. Since fractions that are not completely separated do not interfere with the identification of the cysteine status, care was not taken to avoid collecting some of the fraction represented by HPLC peak 5 into the fraction represented by HPLC peak 4 and some of the fraction represented by HPLC 4 into the fraction represented by HPLC peak 5 (Figure 4.5).

Most of the other MALDI peaks in Figure 4.6 and Figure 4.7 that were identified, but not described above, are from side reactions such as β -elimination (Figure 1.13 in Chapter 1), incomplete CN-induced cleavage (cleavage does not occur at all cyanylated cysteines), and incomplete cyanylation (not all free cysteines are totally cyanylated). The identities of those peaks are listed in Table 4.4 and Table 4.5. Formation of some representative fragments in Table 4.4 is described below. Itz-154-SCN@158-SH@278-288 (Table 4.4) resulted from CN-induced cleavage at 154 and 289 but no cleavage at 158 of an incompletely cyanylated species of subunit A of GAPD (Figure 4.9). 1-SCN@20- β @154-157 (Table 4.4) resulted from CN-induced cleavage at 158, no cleavage at 20, and β -elimination at 154 (Figure 4.9). As we can see from Table 4.4 and Figure 4.9, identification of these fragments does not interfere with the determination of the cysteine status of GAPD subunits A and B because the data are consistent with the data that we have discussed early in this section. Details about these side reactions can be found elsewhere (chapter 3 and reference 13).

Table 4.4. The experimental and calculated m/z values for the fragments from β -elimination and incomplete CN-induced cleavage of subunit A of GAPD.

CN-induced Fragment	Calculated m/z	Experimental m/z
itz-154-SCN@158-SH@278-288 ^a	14529.8 (+1)	14528.5
(or itz-154-SH@158-SCN@278-288)	7265.4 (+2)	7265.1
1-SCN@20-153	16257.5 (+1)	16256.5
	8129.3 (+2)	8128.5
1-SCN@20- β @154-157 ^b	16643.0 (+1)	16644.6
(or 1- β @20-SCN@154-157)	8322.0 (+2)	8320.1

^a itz-154-SCN@158-SH@278-288 designates the protein fragment from amino acid residue 154 to 288, in which cysteine 158 is cyanylated (SCN) and cysteine 278 is free (SH).

^b β @154 designates that β -elimination (loss of HSCN) happened on the cyanylated cysteine 154.

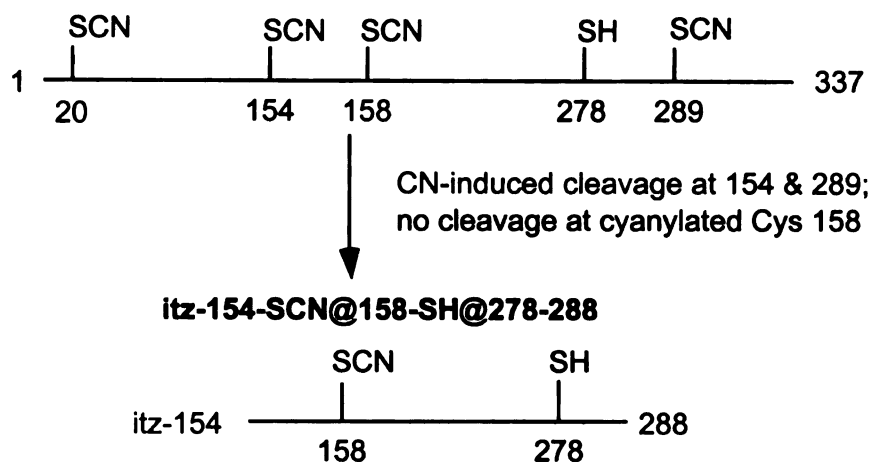
Table 4.5. The experimental and calculated m/z values for the fragments from β -elimination and incomplete CN-induced cleavage of subunit B of GAPD.

CN-induced Fragment	Calculated m/z	Experimental m/z
itz-159- β @280-290 ^a	13924.2 (+1)	13922.6
	6962.6 (+2)	6962.1
itz-19-SCN@155-158 ^b	14871.0 (+1)	14869.6
	7436.0 (+2)	7435.3
1-SCN@19-154	16468.0 (+1)	16468.1
1-SCN@19- β @155-158 (or 1- β @19-SCN@155-158)	16853.4 (+1)	16857.5

^a itz-159- β @280-290 designates the protein fragment from iminothiazolidine-blocked-residue 159 to residue 290, in which β -elimination (loss of HSCN) occurred on the cyanylated cysteine 280.

^b SCN@155 designates that cysteine 155 is cyanylated.

**A) An Incompletely Cyanylated GAPD-Subunit-A
(Cys 278 Was Not Cyanylated)**



B) The Cyanylated GAPD-Subunit-A

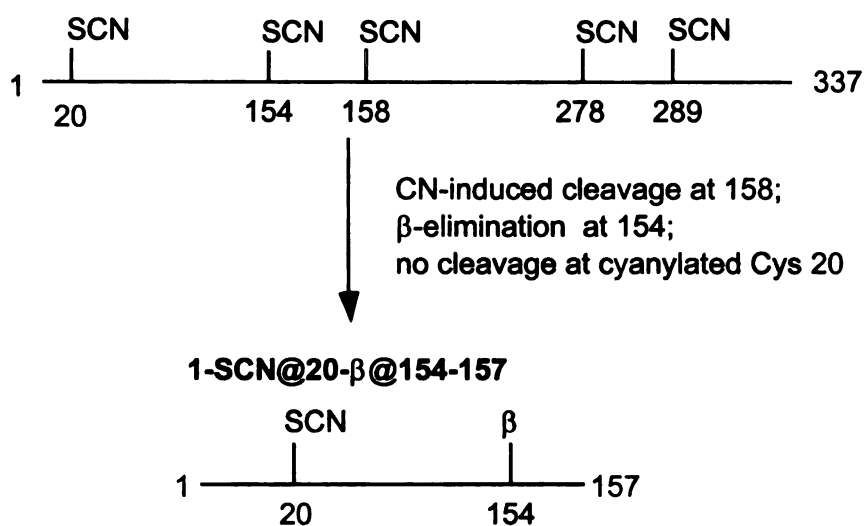


Figure 4.9. Formation of some fragments resulted from β -elimination, incomplete CN-induced cleavage, and incomplete cyanylation of subunit A of GAPD.

2. Subunit Composition

It has been suggested that the chloroplast enzyme is present *in vivo* as a mixture of the A₄ tetramer, the A₂B₂ tetramer and higher oligomeric forms of the A₂B₂ tetramer. The recombinant B₄ tetramer forms in the absence of the A subunit in *Escherichia coli* (10, 14). Our results indicate a 1.7 to 1 ratio of the B to the A subunit in the pea leaf enzyme preparation used in these experiments. There is no indication in our experiments of the presence of degradation products of the B subunit or that the shorter subunit corresponds to a mixture of A subunit and degraded B subunits. We suggest that the wild-type tetramer represents at least one additional species containing at least three B subunits. More likely, it represents the five possible combinations of A and B subunits in a tetramer (B₄, A₁B₃, A₂B₂, A₃B, A₄).

IV. Discussions

The mass mapping analysis based on cyanylation and CN-induced cleavage clearly indicates that the regulatory disulfide bond in the light-activated pea chloroplast glyceraldehyde-3-P dehydrogenase links the two Cys residues in the C-terminal extension of subunit B. Recently, Hutchison et al. (15) determined the redox-potential for the inactivating disulfide bond in the tomato leaf chloroplast enzyme. That value then represents the redox-potential of the C-terminal disulfide bond. The recombinant pea B₄ glyceraldehyde-3-P dehydrogenase expressed in *Escherichia coli* is redox-sensitive (10). Consistent with the present results, when the C-terminal extension is removed genetically, the resultant recombinant mutant enzyme is significantly less redox-sensitive (10, 14). The regulatory disulfide bonds in the redox-regulated chloroplast malate

dehydrogenase (16, 17) and in the chloroplast fructose biphosphatase (18) are also located in extensions or insertions. Modeling suggests that the two Cys residues in the C-terminal extension of the chloroplast glyceraldehyde-3-P dehydrogenase are located in different domains (19). If there is inter-domain movement during catalysis (20) in the NADP-linked chloroplast enzyme, crosslinking the domains would be expected to be inhibitory. In contrast, the disulfide bonds in malate dehydrogenase and fructose biphosphatase are intra-domain disulfide bonds. A crystal structure will be required to confirm the prediction that the C-terminal disulfide bond forms an inter-domain crosslink in the chloroplast glyceraldehyde-3-P dehydrogenase (19).

There is no experimental evidence here for a second disulfide bond in the angiosperm chloroplast enzyme, although modeling (19, 21) suggests a disulfide bond between Cys20 and Cys289 in the A subunit and another one between the corresponding Cys residues in the B subunit. Similarly, in both malate dehydrogenase (21, 22) and fructose biphosphatase (23), at least one additional disulfide bond not seen in the available crystal structures seems possible. Homology modeling, based on the coordinates of the NAD-linked *Bacillus stearothermophilus* enzyme in PDB file 1gd1 (24), implicates the Cys residues corresponding to Cys20 and Cys289 in the dark inactivation of the NADP-linked *Chlamydomonas reinhardtii* enzyme. The algal enzyme has no C-terminal extension, and, as modeled, no other Cys residues positioned to form disulfides (reference 3; PDB files 1nlg and 1nlh). As lower plant and algal chloroplast glyceraldehyde-3-P dehydrogenases are not known to contain B subunits, it seems likely that this disulfide bond can form and does affect activity in many species. Presumably, when the enzyme acquired a more effective disulfide bond in the C-terminal extension,

formation of the obsolete Cys20-Cys289 disulfide was eliminated. This could have been accomplished by a change in structure and/or charge distribution. Interestingly, the *Sulfolobus* enzyme also has an inter-domain disulfide bond (25). This non-regulatory disulfide bond is thought to stabilize the enzyme in the high-temperature environment where the organism exists in nature.

Consistent with the mutational experiments of Baalman *et al.* (14) and of Li and Anderson (10), these results indicate that the Cys residues in the extension are responsible for the redox-sensitivity of the higher plant chloroplast glyceraldehyde-3-P dehydrogenases that contain B subunits.

V. References

1. Anderson, L. E. Light Dark Modulation of Enzyme-Activity in Plants. in *Advances in Botanical Research Incorporating Advances in Plant Pathology* (ed. Callow, J. A.) Vol. 12, 1-46 (Academic Press, New York, NY, 1986).
2. Buchanan, B. B. Role of Light in the Regulation of Chloroplast Enzymes. *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 31, 341-374 (1980).
3. Li, A. D., Stevens, F. J., Huppe, H. C., Kersanach, R. & Anderson, L. E. Chlamydomonas reinhardtii NADP-linked glyceraldehyde-3- phosphate dehydrogenase contains the cysteine residues identified as potentially domain-locking in the higher plant enzyme and is light activated. *Photosynth. Res.* 51, 167-177 (1997).
4. Smith, D. L. & Zhou, Z. Strategies for locating disulfide bonds in proteins. in *Methods Enzymol.* (ed. McCloskey, J. A.) Vol. 193, 374-389 (1990).
5. Wu, J., Gage, D. A. & Watson, J. T. A strategy to locate cysteine residues in proteins by specific chemical cleavage followed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal. Biochem.* 235, 161-174. (1996).

6. Wu, J. & Watson, J. T. Optimization of the cleavage reaction for cyanylated cysteinyl proteins for efficient and simplified mass mapping. *Anal. Biochem.* **258**, 268-276 (1998).
7. Jacobson, G. R., Schaffer, M. H., Stark, G. R. & Vanaman, T. C. Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues. *J. Biol. Chem.* **248**, 6583-6591. (1973).
8. Wakselman, M. & Guibe-Jampel, E. 1-Cyano-4-dimethylamino-pyridinium salts: New water-soluble reagents for the cyanylation of protein sulphydryl groups. *J. Chem. Soc., Chem. Commun.*, 21-22 (1976).
9. Anderson, L. E., Goldhabergordon, I. M., Li, D., Tang, X. Y., Xiang, M. H. & Prakash, N. Enzyme-Enzyme Interaction in the Chloroplast - Glyceraldehyde- 3-Phosphate Dehydrogenase, Triose Phosphate Isomerase and Aldolase. *Planta* **196**, 245-255 (1995).
10. Li, A. D. & Anderson, L. E. Expression and characterization of pea chloroplastic glyceraldehyde-3-phosphate dehydrogenase composed of only the B-subunit. *Plant Physiol.* **115**, 1201-1209 (1997).
11. Wedel, N., Soll, J. & Paap, B. K. CP12 provides a new mode of light regulation of Calvin cycle activity in higher plants. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10479-10484 (1997).
12. Wedel, N. & Soll, J. Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/glyceraldehyde-3-phosphate dehydrogenase complex dissociation. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9699-9704 (1998).
13. Qi, J. & Watson, J. T. Determination of the Disulfide Structure of Sillucin, a Highly Knotted, Cysteine-Rich Peptide, by Cyanylation/Cleavage Mass Mapping. *Biochemistry* **40**, 4531-4538 (2001).
14. Baalman, E., Scheibe, R., Cerff, R. & Martin, W. Functional studies of chloroplast glyceraldehyde-3-phosphate dehydrogenase subunits A and B expressed in *Escherichia coli*: formation of highly active A4 and B4 homotetramers and evidence that aggregation of the B4 complex is mediated by the B subunit carboxy terminus. *Plant Mol. Biol.* **32**, 505-513. (1996).
15. Hutchison, R. S., Groom, Q. & Ort, D. R. Differential effects of chilling-induced photooxidation on the redox regulation of photosynthetic enzymes. *Biochemistry* **39**, 6679-6688 (2000).

16. Carr, P. D., Verger, D., Ashton, A. R. & Ollis, D. L. Chloroplast NADP-malate dehydrogenase: structural basis of light-dependent regulation of activity by thiol oxidation and reduction. *Struct. Fold. Des.* **7**, 461-475 (1999).
17. Johansson, K., Ramaswamy, S., Saarinen, M., Lemaire-Chamley, M., Issakidis-Bourguet, E., Miginiac-Maslow, M. & Eklund, H. Structural basis for light activation of a chloroplast enzyme: The structure of sorghum NADP-malate dehydrogenase in its oxidized form. *Biochemistry* **38**, 4319-4326 (1999).
18. Chiadmi, M., Navaza, A., Miginiac-Maslow, M., Jacquot, J. P. & Cherfils, J. Redox signaling in the chloroplast: structure of oxidized pea fructose-1,6-bisphosphate phosphatase. *Embo J.* **18**, 6809-6815 (1999).
19. Qi, J., Isupov, M. N., Littlechild, J. A. & Anderson, L. E. Chloroplast Glyceraldehyde-3-P Dehydrogenase Contains a Single Disulfide Bond Located in the C-terminal Extension to the B Subunit. *J. Biol. Chem.* **in press** (2001).
20. Skarzynski, T. & Wonacott, A. J. Coenzyme-induced conformational changes in glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus*. *J. Mol. Biol.* **203**, 1097-1118. (1988).
21. Li, D., Stevens, F. J., Schiffer, M. & Anderson, L. E. Mechanism of Light-Modulation - Identification of Potential Redox-Sensitive Cysteines Distal to Catalytic Site in Light- Activated Chloroplast Enzymes. *Biophys. J.* **67**, 29-35 (1994).
22. Muslin, E. H. & Anderson, L. E. Identification of one pair of domain-locking, redox-sensitive cysteine residues in maize chloroplast malate dehydrogenase. in *Photosynthesis, Mechanisms and Effects* (ed. Garab, G.) Vol. **5**, 3583-3586 (Kluwer, Boston, MA, 1999).
23. Jacquot, J. P., Lopez-Jaramillo, J., Miginiac-Maslow, M., Lemaire, S., Cherfils, J., Chueca, A. & Lopez-Gorge, J. Cysteine-153 is required for redox regulation of pea chloroplast fructose-1,6-bisphosphatase. *FEBS Lett.* **401**, 143-147 (1997).
24. Skarzynski, T., Moody, P. C. E. & Wonacott, A. J. Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* at 1.8 Å resolution. *J. Mol. Biol.* **193**, 171-187 (1987).
25. Isupov, M. N., Fleming, T. M., Dalby, A. R., Crowhurst, G. S., Bourne, P. C. & Littlechild, J. A. Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* **291**, 651-660 (1999).

CHAPTER 5

COMPUTER ALGORITHMS FOR DISULFIDE MASS MAPPING OF PROTEINES BASED ON PARTIAL REDUCTION AND CYANYLATION/CN- INDUCED CLEAVAGE

I. Introduction

The number of possible disulfide structural isomers of a given protein containing several cysteines increases rapidly with the number of cysteines. For example, in the case of ribonuclease-A, which has eight **disulfide cysteines** ("Disulfide cysteines" are cysteines that are involved in disulfide bonds.), 105 isomers are possible. For a protein having ten disulfide cysteines, 945 isomeric forms are possible. In general, the total number of all possible disulfide structures (A **disulfide structure** is defined as a particular arrangement of all the disulfide bonds in a protein.) is $(2n-1)*(2n-3)...3*1$ for a protein that has $2n$ disulfide cysteine residues that are involved in forming n disulfide bonds. Clearly, in proteins containing multiple disulfide cysteines, computational assistance is essential to calculate the mass of CN-induced cleavage fragments from all possible disulfide structural isomers (of a given sequence) that have been processed by partial reduction, cyanylation and CN-induced cleavage.

In principle, for a given cyanylated protein having two free sulfhydryls on two **free cysteines** ("Free cysteines" are cysteines that have free sulfhydryls and are not involved in disulfide bond formation; reducing a disulfide bond in a protein will produced two nascent free cysteines.), after cyanylation and CN-induced cleavage, only

three cleavage products are expected for the protein because two cyanylated cysteines provide two cleavage sites. In practice, analysis of a protein of unknown disulfide structure (but known sequence) requires the computation of the mass of each of all possible CN-induced fragments of each of all possible disulfide structural isomers of the analyte. For example, if the disulfide bonding in ribonuclease-A (containing 8 disulfide cysteines residues) were unknown, the analyst would need to compute the masses of the expected three cleavage fragments for each of all four possible cyanylated singly reduced isoforms of each of the 105 possible disulfide structural isomers. Mass spectrometric analysis of a cyanylated and cleaved singly reduced isoform of the protein generates three mass spectral peaks that must be compared with the array of computed masses corresponding to the three fragments expected from each of the four possible singly reduced isoforms of each of the 105 different disulfide structural isomers of the unknown protein. Since there are four possible singly reduced isoforms of each of the 105 different isomers obtained in the process of partial reduction, $105 \times 4 \times 3 = 1260$ computations must be made. A match between the three experimentally observed mass spectral peaks with the mass values of three computed fragments would constitute identification of one of the connectivities involved in one of the four disulfide bonds in one of the 105 possible isomeric forms involving four disulfide bonds. Clearly, the development of a suitable algorithm to compute these masses, to coordinate a given mass set with a given connectivity for a given isomer, and to match the experimental mass set with a calculated one to determine the disulfide connectivity is essential to the success of applying the partial reduction/cyanylation/CN-induced cleavage/mass mapping methodology to proteins containing several cysteines.

II. A Naive Algorithm

In the computational analysis phase, the computer program must determine the connectivity of the disulfide bonds (disulfide structure) given the mass spectral data, the location of the original free cysteines, and the sequence of the protein. A naive algorithmic approach to solving this problem is the following. First, the program will compute all possible sets of masses of cleaved fragments; that is, for each disulfide structure in the **complete disulfide structure list A** (such a list can be computed given the sequence of the protein to be analyzed), the program computes the entire set of masses of all the CN-induced fragments that such a protein would be cleaved into using our partial reduction, cyanylation, and CN-induced cleavage methodology. These are then compared against the experimental mass spectral data using a **disulfide structure search module**. Given the accuracy of the mass spectrometric measurement, we should be able to match each experimentally measured mass spectral peak with its corresponding calculated value. The correct disulfide structure is then established by determining which set of computed fragments matches the actual data. For the correct disulfide structure, there exists a perfect matching between its set of calculated values of m/z and the list of experimentally obtained mass spectral peaks (represented by m/z values); that is, each calculated m/z value will match exactly one experimentally obtained peak (unless the m/z values for two fragments differ by less than the mass spectrometric error), and each element in both the correct set of calculated values and the list of experimental peaks will be matched. For other disulfide structures, there will not exist a perfect matching between the set of calculated values and the list of experimentally obtained peaks. Thus,

the correct disulfide structure can be easily identified because it is the one with the highest number of CN-induced fragments matched in the **ranked disulfide structure list**.

The information flow for such an algorithm is illustrated in Figure 5.1.

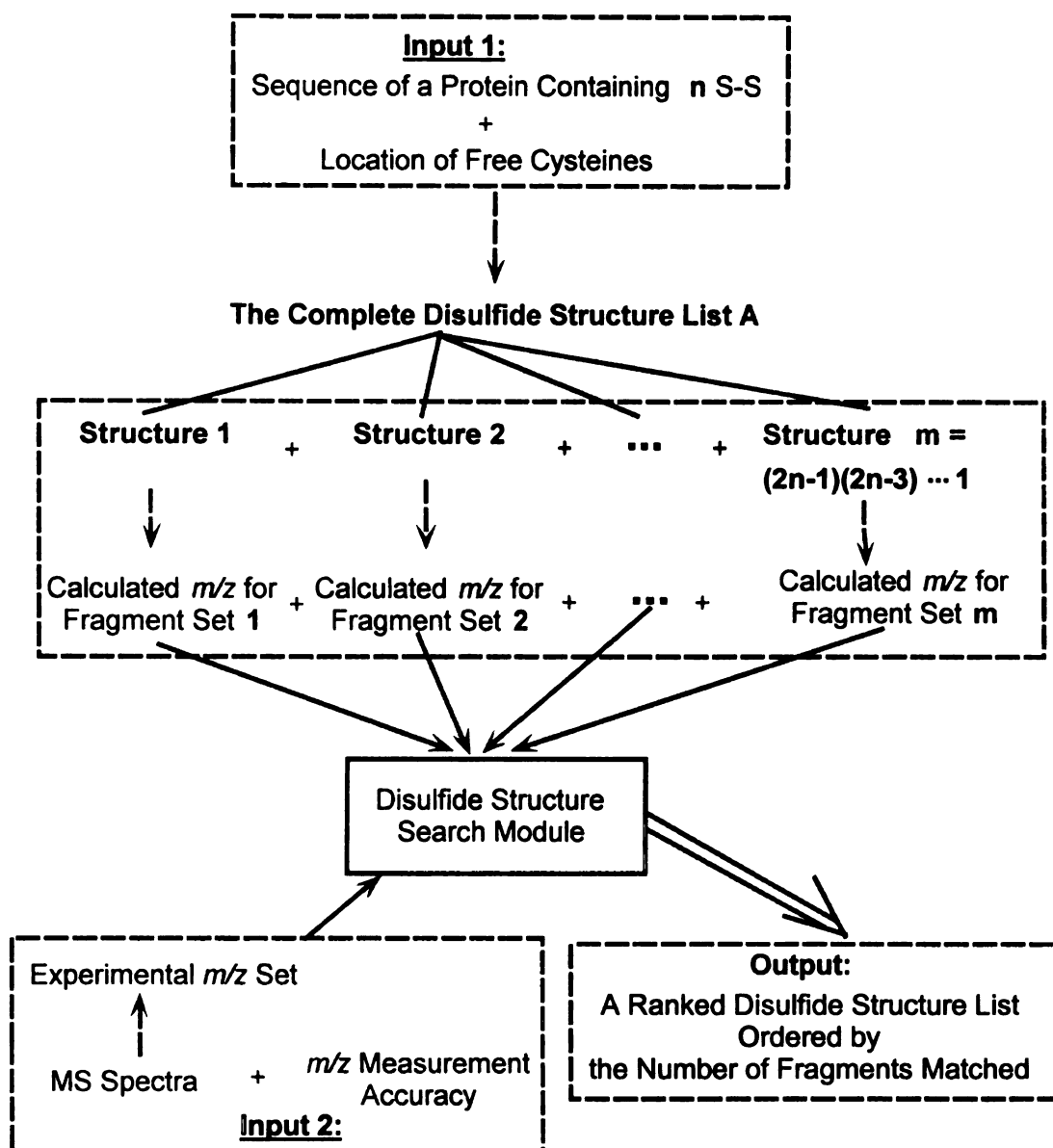


Figure 5.1. Information flow for a naïve algorithm for disulfide mass mapping based on partial reduction, cyanylation and CN-induced cleavage. The sequence of the protein to be analyzed and the locations of free cysteines are used as input 1; MS spectra of the CN-induced fragments and m/z measurement accuracy are used as input 2. The output is a ranked disulfide structure list, in which candidate disulfide structures are ranked according to the number of CN-induced fragments matched (Experimental m/z values and calculated m/z values are compared.).

III. A Basic Algorithm

As discussed in the first section of this chapter, the disulfide structure **list A** is extremely long as the number of candidate structures, $((2n-1)*(2n-3)*...*3*1)$, is exponential in the number of disulfide cysteines ($2n$) in the original protein containing n disulfide bonds. Thus, the matching step in the naïve algorithm will take considerable time as the experimental data are compared against each structure in **list A**. Furthermore, this is not a preprocessing step as this matching phase cannot occur until the experimental data have been gathered. In addition, many repeated calculations and comparisons take place in this algorithm because some of all the CN-induced fragments from the chemical processing (partial reduction, cyanylation, and CN-induced cleavage) of protein isomers of different disulfide structures will be the same. An example of this is shown in Table 5.1. Therefore, it is desirable to develop alternative matching strategies that do not require us to compare the experimental data against the entire **list A**.

One observation shows that while the number of possible disulfide structures is large, the number of possible disulfide linkages is much smaller (Table 5.2). Thus, a better strategy is to deal with disulfide linkages instead of disulfide structures.

Many times, partially reduced/cyanylated species of the original protein are isolated by reversed phase HPLC (1, 2). Considering this, we can develop a basic algorithm (Figure 5.2) to analyze the mass spectral data of the CN-induced fragments from each isolated species to find out which disulfide bond(s) in the original protein is reduced to form the species. Then, we can combine all the individual results to obtain the correct disulfide structure. Next, the algorithms for singly reduced, doubly reduced, and triply reduced isoforms are discussed in detail.

Table 5.1. Some of the same CN-induced fragments from two singly reduced isoforms of two hypothetical protein isomers with different disulfide structures. The hypothetical protein has 90 amino acid residues, among which 8 disulfide cysteines are located at positions 10, 20, 30, 40, 50, 60, 70, and 80.

Disulfide Structure	(Cys10-Cys20, Cys30-Cys40, Cys50-Cys60, Cys70-Cys80)	(Cys10-Cys20, Cys30-Cys40, Cys50-Cys70, Cys60-Cys80)
S-S Opened in Singly	Cys10-Cys20	Cys10-Cys20
Reduced Isoform 1		
CN-induced Fragments from	1-9,	1-9,
Singly Reduced Isoform 1	itz-10-19, itz-20-90	itz-10-19, itz-20-90
S-S Opened in Singly	Cys30-Cys40	Cys30-Cys40
Reduced Isoform 2		
CN-induced Fragments from	1-29,	1-29,
Singly Reduced Isoform 2	itz-30-39, itz-40-90	itz-30-39, itz-40-90

Table 5.2. The number of possible disulfide linkages and disulfide structures for a protein having **n** disulfide bonds.

Number of Disulfide Bonds in the Protein	Number of Possible Disulfide Linkages	Number of Possible Disulfide Structures
2	6	3
3	15	15
4	28	105
5	45	945
6	66	10,395
7	91	135,135
8	120	2,027,025
...	...	...
n	$n*(2n-1)$	$(2n-1)*(2n-3)...3*1$

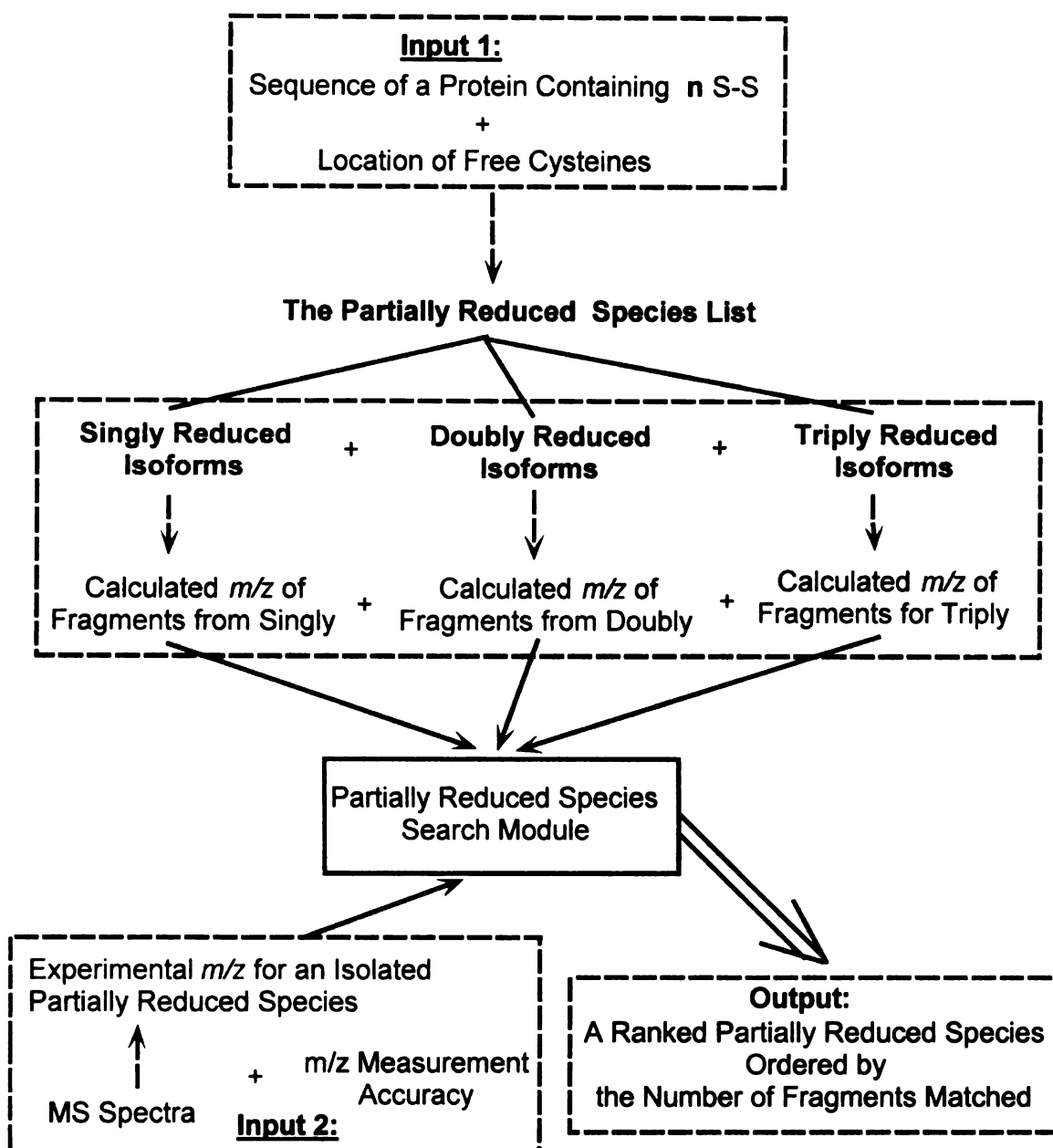


Figure 5.2. A basic algorithm for disulfide mass mapping based on partial reduction, cyanylation and CN-induced cleavage. The sequence of the protein to be analyzed and the locations of free cysteines are used as input 1; MS spectra of the CN-induced fragments and m/z measurement accuracy are used as input 2. The output is a ranked partially reduced species list, in which candidate partially reduced species are ranked according to matching percentage = (the number of CN-induced fragments matched/the number of such fragments expected for a partially reduced species).

1. Consideration of Singly Reduced Isoforms

A given singly reduced/cyanylated isoform of the protein followed by CN-induced cleavage and complete reduction of the residual disulfide bonds can provide information leading to identification of one particular disulfide bond, which was reduced during partial reduction. In order to identify this disulfide bond, the mass spectral data of the CN-induced cleavage fragments should be compared against the calculated masses of the CN-induced fragments expected from each of the possible singly reduced isoforms, which has been cyanylated, then subjected to CN-induced cleavage followed by complete reduction of the residual disulfide bonds. Although there are $[(2n-1)*(2n-3)*\dots*3*1]$ possible disulfide structures for a protein containing n disulfide bonds, only $n(2n-1)$ singly reduced/cyanylated isoforms need be considered because of the degeneracy associated with completely reducing *the $n-1$ residual disulfide bonds* in these isoforms during chemical processing. For example, there are only 28 $(=4*(2*4-1))$ possible singly reduced/cyanylated isoforms (Table 5.3) for a protein having four disulfide bonds, such as bovine ribonuclease-A, in spite of the fact that there are 105 $(=(2*4-1)*(2*4-3)*\dots*3*1)$ possible disulfide structures for the protein.

All the singly reduced isoforms can be computed as follows. A hypothetical protein is used to illustrate the computation process. The protein has 100 amino acid residues, among which $2n$ disulfide cysteines (forming n disulfide bonds) are at positions C1, C2, ..., C(2n) ["C" represents a cysteine residue.].

Step 1. Compute all singly reduced isoforms in which the disulfide bond reduced is C1-C(j) ($j = 2, 3, \dots, \text{or } 2n$). There are $2n-1$ such singly reduced isoforms.

Step 2. Compute all singly reduced isoforms in which the disulfide bond reduced is C2-C(j) ($j = 3, 4, \dots, \text{or } 2n$). There are $2n-2$ such singly reduced isoforms.

...

Step i. Compute all singly reduced isoforms in which the disulfide bond reduced is C(i)-C(j) ($j = i+1, i+2, \dots, \text{or } 2n$). There are $2n-i$ such singly reduced isoforms.

...

Step $2n-1$. Compute the singly reduced isoform in which the disulfide bond reduced is C($2n-1$)-C($2n$). There is only one such singly reduced isoform.

Counting from Step 1 to Step $2n-1$, there are a total of $2n*(2n-1)/(2*1)$ singly reduced isoforms.

Once all the singly reduced isoforms are computed, all the CN-induced cleavage fragments from such singly reduced/cyanylated isoforms can be easily computed. For example, when the disulfide bond C(i)-C(j) is reduced to form a singly reduced/cyanylated isoform, its corresponding CN-induced fragments are 1-[C(i)-1], itz-C(i)-[C(j)-1], itz-C(j)-100 ["1-[C(i)-1]" designates the fragment from residue 1 to residue C(i)-1.].

As an example, all sets of CN-induced fragments from 28 ($=4*(2*4-1)$) possible singly reduced/cyanylated isoforms of ribonuclease-A (consisting of 124 amino acid residues, 8 disulfide cysteines at positions 26, 40, 58, 65, 72, 84, 95, and 110 forming 4 disulfide bonds) are represented in Table 5.3.

Table 5.3. The CN-induced fragments from all the possible singly reduced/cyanylated isoforms of ribonuclease-A.

S-S No.	S-S reduced * Cys i – Cys j	Fragment 1 1-(i-1)	Fragment 2 itz-i-(j-1)	Fragment 3 itz-j-124
1	Cys26–Cys40	1-25	itz-26-39	itz-40-124
2	Cys26–Cys58	1-25	itz-26-57	itz-58-124
3	Cys26–Cys65	1-25	itz-26-64	itz-65-124
4	Cys26–Cys72	1-25	itz-26-71	itz-72-124
5	Cys26–Cys84	1-25	itz-26-83	itz-84-124
6	Cys26–Cys95	1-25	itz-26-94	itz-95-124
7	Cys26–Cys110	1-25	itz-26-109	itz-110-124
8	Cys40–Cys58	1-39	itz-40-57	itz-58-124
9	Cys40–Cys65	1-39	itz-40-64	itz-65-124
10	Cys40–Cys72	1-39	itz-40-71	itz-72-124
11	Cys40–Cys84	1-39	itz-40-83	itz-84-124
12	Cys40–Cys95	1-39	itz-40-94	itz-95-124
13	Cys40–Cys110	1-39	itz-40-109	itz-110-124
14	Cys58–Cys65	1-57	itz-58-64	itz-65-124
15	Cys58–Cys72	1-57	itz-58-71	itz-72-124
16	Cys58–Cys84	1-57	itz-58-83	itz-84-124
17	Cys58–Cys95	1-57	itz-58-94	itz-95-124
18	Cys58–Cys110	1-57	itz-58-109	itz-110-124
19	Cys65–Cys72	1-64	itz-65-71	itz-72-124
20	Cys65–Cys84	1-64	itz-65-83	itz-84-124
21	Cys65–Cys95	1-64	itz-65-94	itz-95-124
22	Cys65–Cys110	1-64	itz-65-109	itz-110-124
23	Cys72–Cys84	1-71	itz-72-83	itz-84-124
24	Cys72–Cys95	1-71	itz-72-94	itz-95-124
25	Cys72–Cys110	1-71	itz-72-109	itz-109-124
26	Cys84–Cys95	1-83	itz-84-94	itz-95-124
27	Cys84–Cys110	1-83	itz-84-109	itz-95-124
28	Cys95–Cys110	1-94	itz-95-109	itz-110-124

* A singly reduced/cyanylated isoform is designated by a numeric representation of the two disulfide cystines residues constituting the disulfide bond that was reduced.

Once the masses (m/z values) of the CN-induced fragments from all singly reduced/cyanylated isoforms are calculated, mass spectral data from analysis of the CN-induced fragments from singly reduced/cyanylated isoforms can be compared against the calculated values to identify the disulfide bonds in the original protein. For example, the MALDI-MS spectrum of the CN-induced fragments from a singly reduced/cyanylated isoform of ribonuclease-A shows major peaks at m/z 2705.3, 4527.4, and 6548.5 (1). These three peaks match with the calculated m/z values of singly protonated CN-induced fragments, 1-25 (m/z 2704.9), itz-84-124 (m/z 4526.1), and itz-26-83 (m/z 6546.4), respectively, within the mass measurement error of 0.1%. According to Table 5.3, the disulfide bond Cys26-Cys84 was reduced to produce the singly reduced/cyanylated isoform of ribonuclease-A.

Note that while seven different singly reduced/cyanylated isoforms (No. 1 to No. 7 in Table 5.3) would produce the CN-induced fragment 1-25 and five singly reduced/cyanylated isoforms (No. 5, 11, 16, 20, and 23 in Table 5.3) would produce the CN-induced fragment itz-26-83, only one (No. 5 in Table 5.3) of these singly reduced isoforms would generate 1-25, itz-26-83 and itz-84-124, a pattern that matches the experimental data. What if some of the expected CN-induced fragments were not detected because of chemical instability of the fragments or signal suppressions in mass spectrometry? Will we be able to identify the disulfide bond that was reduced to produce the singly reduced isoform? The answer is 'yes' in some cases, 'no' in other cases. We introduce a term "**matching percentage**" to determine when a singly reduced isoform can be identified with the detection of an incomplete set of CN-induced cleavage fragments. Matching percentage (MP) is defined as the number of detected CN-induced

cleavage fragments divided by the total number of the expected such fragments for a certain partially reduced/cyanylated species of the original protein. When a complete set of CN-induced fragments was detected, MP is 100%; otherwise, MP is less than 100%. A candidate isoform with a MP of 0 is no longer considered. The candidate isoform with the highest MP is the correct one for the experimental data.

Case 1. 1-25 and itz-26-83 were detected. According to Table 5.3, the matching percentages for all the singly reduced isoforms are 2/3 for No. 5; 1/3 for No. 1, 2, 3, 4, 6, and 7; and 0/3 for all other singly reduced isoforms. In this case, MP for No. 5 is the highest. The correct isoform, No. 5, is identified.

Case 2. Itz-26-83 was detected. According to Table 5.3, the matching percentages for all the singly reduced isoforms are 1/3 for No. 5; and 0/3 for all other singly reduced isoforms. In this case, MP for No. 5 is the highest. The correct isoform, No. 5, is identified.

Case 3. 1-25 was detected. According to Table 5.3, the matching percentages for all the singly reduced isoforms are 1/3 for No. 1, 2, 3, 4, 5, 6, and 7; and 0/3 for all other singly reduced isoforms. In this case, the correct isoform, No. 5, is not identified.

If there are some free cysteines in the protein, they need to be pre-modified (alkylated or cyanylated) and their positions need to be determined before the protein is subjected to the partial reduction/cyanylation/CN-induced cleavage/complete reduction protocol. The data processing strategies can be easily adjusted for the calculated masses of the species affected by the pre-modifications.

2. Consideration of Doubly Reduced Isoforms

As described in chapter 2, the possibility of using doubly reduced isoforms, in addition to singly reduced isoforms, provides flexibility in applying the chemistry of partial reduction, and increases the possibility of successful identification of the linkage of all the disulfide bonds in the protein. Mass mapping a doubly reduced/cyanylated isoform of the protein followed by CN-induced cleavage and complete reduction can determine which four disulfide cysteine residues form two disulfide bonds (Figure 2.1 in Chapter 2). This information can be used along with other information from the mapping experiment of another doubly or a singly reduced/cyanylated isoform (Figure 2.1 and Figure 2.2 in Chapter 2).

In order to identify which four disulfide cysteine residues form two disulfide bonds (that have been reduced and cyanylated) in a particular doubly reduced/cyanylated isoform, the mass spectral data should be compared against all sets of calculated masses of the CN-induced fragments expected from all possible doubly reduced/cyanylated isoforms. For a protein consisting of n disulfide bonds, there are $[(2n-1)*(2n-3)*...*3*1]$ possible disulfide structures and $n*(n-1)/2$ possible doubly reduced/cyanylated isoforms for each structure, but only $2n*(2n-1)*(2n-2)*(2n-3)/24$ doubly reduced/cyanylated isoforms need be considered because of the degeneracy associated with completely reducing the $n-2$ residual disulfide bonds in these species during chemical processing. For example, only 70 $(=(2*4)*(2*4-1)*(2*4-2)*(2*4-3)/24)$ possible doubly reduced/cyanylated isoforms for a protein having four disulfide bonds, such as bovine ribonuclease-A, even though there are 105 $(=(2*4-1)*(2*4-3)*...*3*1)$ possible disulfide structures for the protein.

All sets of calculated masses of the CN-induced fragments from all possible doubly reduced/cyanylated isoforms can be computed and tabulated in an analogous manner to those from all possible singly reduced/cyanylated isoforms, as listed in Table 5.3. We use a hypothetical protein to illustrate the computation process. The protein has 100 amino acid residues, among which $2n$ disulfide cysteines (forming n disulfide bonds) are at positions $C1, C2, \dots$, and $C(2n)$.

Step 1. Compute all doubly reduced isoforms in which the four nascent free cysteines (from reduction of two disulfide bonds) are $C1, C(j), C(k)$, and $C(m)$ ($1 < j, k, m < 2n$). We can choose $C(j)$ from $C2, C3, \dots, C(2n)$ first; then choose $C(k)$ and $C(m)$ from $C(j+1), C(j+2), \dots$, and $C(2n)$ for each $C(j)$ using the strategy described for computing singly reduced isoforms in the last section. There are $(2n-1)*(2n-2)*(2n-3)/(3*2*1)$ such doubly reduced isoforms.

Step 2. Compute all doubly reduced isoforms in which the four nascent free cysteines are $C2, C(j), C(k)$, and $C(m)$ ($2 < j, k, m < 2n$). We can choose $C(j)$ from $C3, C4, \dots, C(2n)$ first; then choose $C(k)$ and $C(m)$ from $C(j+1), C(j+2), \dots$, and $C(2n)$ for each $C(j)$ using the strategy described for computing singly reduced isoforms in the last section. There are $(2n-2)*(2n-3)*(2n-4)/(3*2*1)$ such doubly reduced isoforms.

...

Step i. Compute all doubly reduced isoforms in which the four nascent free cysteines are $C(i), C(j), C(k)$, and $C(m)$ ($i < j, k, m < 2n$). We can choose $C(j)$ from $C(i), C(i+1), \dots, C(2n)$ first; then choose $C(k)$ and $C(m)$ from $C(j+1), C(j+2), \dots$, and $C(2n)$

for each C(j) using the strategy described for computing singly reduced isoforms in the last section. There are $(2n-i)*(2n-i-1)*(2n-i-2)/(3*2*1)$ such singly reduced isoforms.

...

Step 2n-3. Compute the doubly reduced isoform in which the four cysteines that are involved in two disulfide bonds are C(2n-3), C(2n-2), C(2n-1), and C(2n). There is only one such doubly reduced isoform.

Counting from Step 1 to Step 2n-3, there are a total of $2n*(2n-1)*(2n-2)*(2n-4)/4*3*2*1$ doubly reduced isoforms.

Once all the doubly reduced isoforms are computed, all the CN-induced cleavage fragments from such doubly reduced/cyanylated isoforms can be easily computed. For example, when a doubly reduced/cyanylated isoform contains four cyanylated cysteines C(i), C(j), C(k), and C(m), its corresponding CN-induced fragments are 1-[C(i)-1], itz-C(i)-[C(j)-1], itz-C(j)-[C(k)-1], itz-C(k)-[C(m)-1], itz-C(m)-100.

Once the masses (m/z values) of the CN-induced fragments from all doubly reduced/cyanylated isoforms are calculated, mass spectral data from analysis of the CN-induced fragments from doubly reduced/cyanylated isoforms can be compared against the calculated values to identify the four cysteines that are involved in two of the disulfide bonds in the original protein. We can use the concept of matching percentage in the same way as we did for the singly reduced isoforms. The candidate doubly reduced isoform with the highest matching percentage is the correct one for the experimental data.

3. Consideration of Triply Reduced Isoforms

A very similar strategy can be used to compute all the triply reduced isoforms and to search the correct candidate isoform for the experimental data. Therefore, its details are not described here.

The algorithms discussed in the last three sections assume that we know how many disulfide bonds were reduced and cyanylated in the partially reduced/cyanylated species. Namely, we know whether the species is a singly reduced, doubly reduced or triply reduced isoforms. However, we may not know this in some cases. We can still use the algorithms discussed above, but we need to compare the experimental data against the calculated values for singly reduced, doubly reduced, and triply reduced isoforms. A candidate singly reduced isoform list ranked by descending matching percentage, a candidate doubly reduced isoform list ranked by descending matching percentage, and a candidate triply reduced isoform list ranked by descending matching percentage will be produced. The top candidate in each group (singly reduced, doubly reduced, or triply reduced) needs to be considered by the analyst to determine the disulfide structure of the original protein.

Only up to triply reduced isoforms are considered because normally only singly, doubly, and triply reduced isoforms are used in the disulfide mass mapping strategy.

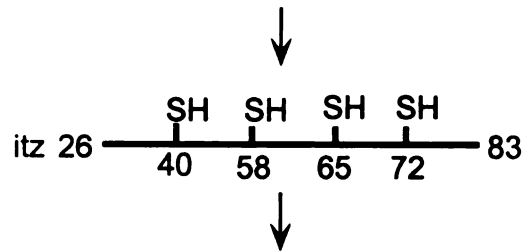
IV. Negative Signature Mass and an Improved Algorithm

1. Negative Signature Mass

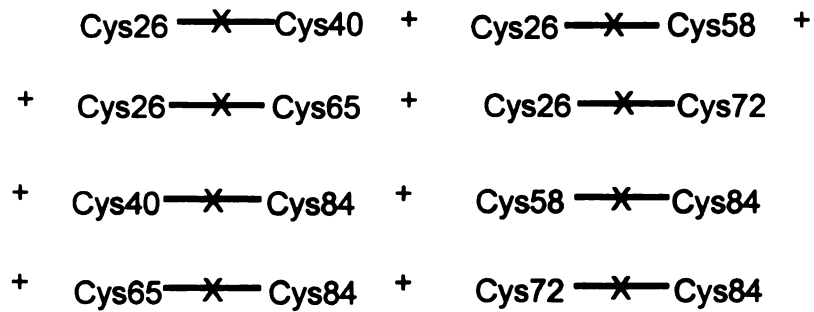
As discussed above, it is desirable to deal with disulfide linkages rather than disulfide structures. Partially reduced/cyanylated species of the original protein need to be separated to allow application of the basic algorithm described in the last section. An alternative algorithm based on negative signature mass (defined later) that does not require this condition is described as follows.

The presence of a **negative signature mass** in the experimental data insures that a candidate disulfide bond cannot be present in the correct disulfide structure. For example, a MALDI-MS peak at m/z 6546.4 indicates the presence of the fragment itz-26-83 from ribonuclease-A (contains 8 disulfide cysteine residues at positions 26, 40, 58, 65, 72, 84, 95, and 110); and thus, this fragment shows that cysteine 26 cannot be connected to cysteine 40 for example; that is, Cys 26 and Cys 40 cannot form a disulfide bond (linkage) in the protein. Let us see why this is true. If the disulfide bond Cys26-Cys40 had been in the original protein, it would have been reduced; and thus, itz-26-39 would have been produced by the following cyanylation and CN-induced cleavage reactions (Figure 5.3). Similarly, the presence of itz-26-83 allows us to reject the existence of seven other disulfide linkages: Cys26-Cys58, Cys26-Cys65, Cys26-Cys72, Cys40-Cys84, Cys58-Cys84, Cys65-Cys84, and Cys72-Cys84 (Figure 5.3).

Detection of m/z 6546.4



The following disulfide linkages could not have been present in the original protein:



For example, if

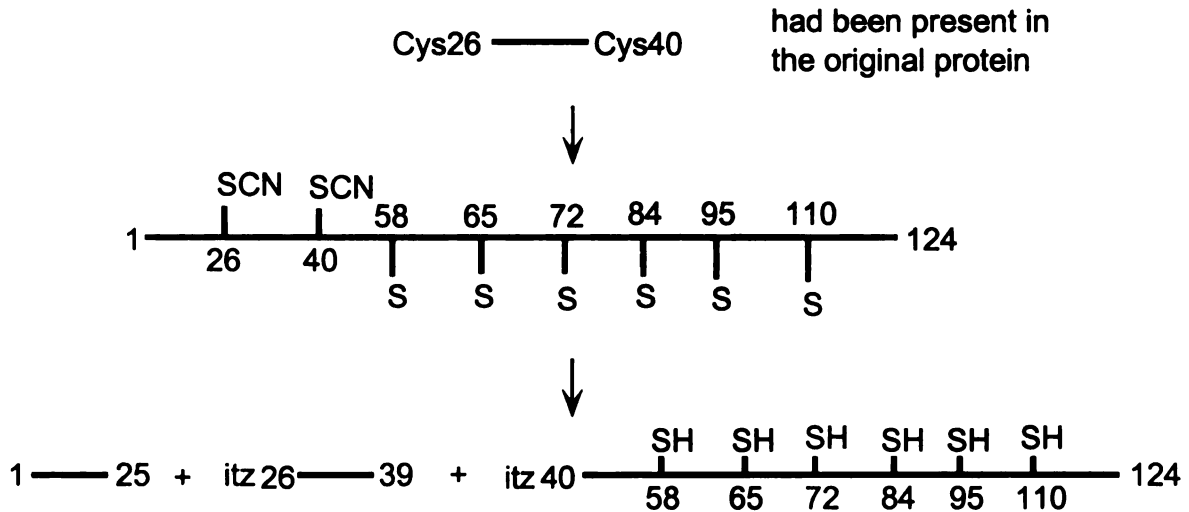


Figure 5.3. Using a negative signature mass (m/z 6546.4) from ribonuclease-A to reject the existence of 8 disulfide linkages.

2. An Algorithm Based on Negative Signature Masses

We use the list of negative signature masses as follows. Rather than comparing the experimental data against **list A**, we scan the experimental data looking for the presence of the negative signature masses. Then we can use the presence of some of the negative signature masses to reject certain disulfide linkages in **list B** to obtain a shortened disulfide linkages **list C**. A shortened disulfide structure **list D** can then be produced from list C. If enough negative signature masses are found in the experimental data, the list D will only contain the correct disulfide structure. Such an improved algorithm can be illustrated by Figure 5.4.

All the negative signature masses (fragments) can be computed as the following. We use a hypothetical protein to illustrate this computation process. The protein has 100 amino acid residues, among which $2n$ disulfide cysteines (forming n disulfide bonds) are at position $C1, C2, \dots, C(2n)$.

Step 1: Compute negative signature fragments beginning from position 1. They are $1-(C2-1), 1-(C3-1), \dots, 1-(C(2n)-1)$ ["1-(C2-1)" designates the CN-induced fragments from residue 1 to residue $C2-1$]. There are $2n-1$ such fragments.

Step 2: Compute negative signature fragments beginning from position $C1$. They are $itz-C1-(C3-1), itz-C1-(C4-1), \dots, itz-C1-(C(2n)-1), itz-C1-100$ ["itz-C1-(C3-1)" designates the CN-induced fragments from residue $C1$ to residue $C3-1$]. There are $2n-1$ such fragments.

...

Step $2n$: Compute negative signature fragments beginning from position $C(n-1)$. There is only one such fragment, $itz-C(2n-1)-100$.

Counting from Step 1 to Step n , there are a total of $(2n-1)(n+1)$ such fragments.

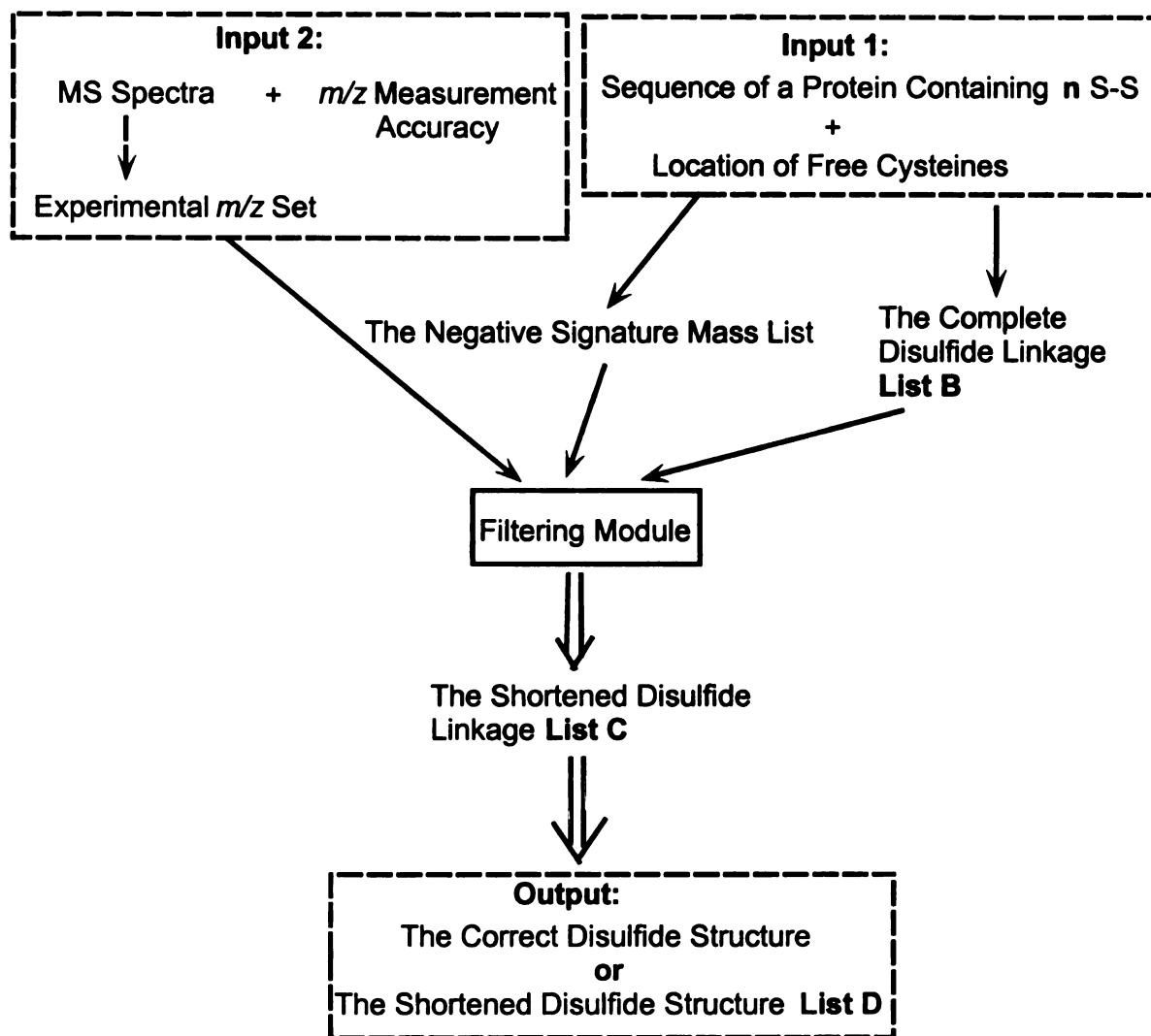


Figure 5.4. An algorithm based on negative signature masses.

Once negative signature fragments are computed, their mass (negative signature masses) can be easily calculated, as the sequence of the original protein is known.

Next, we are going to use ribonuclease-A as an example to see how this algorithm can be applied. Suppose the chemical process using the partial reduction/cyanylation/CN-induced cleavage/complete reduction protocol for ribonuclease-A yields only three MALDI-MS peaks, m/z 6548.5, 5766.8, and 6061.2, which correspond to fragments itz-26-83, itz-58-109, and itz-40-94 with calculated m/z value 6546.4, 5766.4, and 6062.7, respectively. Actually, instead of three MALDI-MS peaks, the experimental data consist of eleven identified peaks from four singly reduced/cyanylated isoforms (1). Only a subset of the experimental data is selected in order to illustrate the effectiveness of the improved algorithm. As ribonuclease-A has 8 disulfide cysteine residues, which are involved in 4 disulfide bonds, there are 28 possible disulfide linkages in its **list B**. The three experimental peaks can be used as negative signature masses to shorten list B into **list C**. Please see Table 5.4 for details.

After the filtering process using the negative masses, there are only seven possible disulfide linkages left in **list C** (Table 5.5).

Table 5.4. Filtering the List B for ribonuclease-A using three negative masses

S-S No.	S-S in the List B	Identified Negative Masses		
		6548.5 (itz-26-83)	5766.8 (itz-58-109)	6061.2 (itz-40-94)
1	Cys26–Cys40	×		
2	Cys26–Cys58	×		
3	Cys26–Cys65	×		
4	Cys26–Cys72	×		
5	Cys26–Cys84			
6	Cys26–Cys95			
7	Cys26–Cys110			
8	Cys40–Cys58			×
9	Cys40–Cys65			×
10	Cys40–Cys72			×
11	Cys40–Cys84	×		×
12	Cys40–Cys95			
13	Cys40–Cys110			
14	Cys58–Cys65		×	
15	Cys58–Cys72		×	
16	Cys58–Cys84	×	×	
17	Cys58–Cys95		×	×
18	Cys58–Cys110			
19	Cys65–Cys72			
20	Cys65–Cys84	×		
21	Cys65–Cys95			×
22	Cys65–Cys110		×	
23	Cys72–Cys84	×		
24	Cys72–Cys95			×
25	Cys72–Cys110		×	
26	Cys84–Cys95			×
27	Cys84–Cys110		×	
28	Cys95–Cys110		×	

× means that a certain disulfide linkage cannot exist in the analyte because of the presence of a particular negative mass .

Table 5.5. The shortened disulfide linkage List C for ribonuclease-A after filtering using three negative masses.

Disulfide Linkage Group i	1st Cys in the S-S	S-S with the Same First Disulfide Cys		
		S-S No. i.1	S-S No. i.2	S-S No. i.3
1	Cys58	Cys58-Cys110		
2	Cys65	Cys65-Cys72		
3	Cys40	Cys40-Cys95	Cys40-Cys110	
4	Cys26	Cys26-Cys84	Cys26-Cys95	Cys26-Cys110

The shortened disulfide structure **list D** can be computed as follows.

Step 1. The disulfide linkages in list C are classified and sorted as in Table 5.5. The disulfide linkages are first divided into groups according to the first disulfide cysteine in each disulfide linkage (bond) (The first disulfide cysteine in a disulfide linkage is the one with the lower numeric value for the position of the residue in the protein sequence.). The groups are then sorted according to the number of disulfide linkages in each group in an ascending order.

Step 2. Compute disulfide structure list D from the disulfide linkages in list C. In the case of ribonuclease-A, a disulfide structure consists of four disulfide bonds (linkages). Thus, four disulfide linkages need to be chosen from list C to form a disulfide structure. Note that we can only select one disulfide linkage from each of the four groups for forming a disulfide structure. As shown in Table 5.6, the positions of disulfide cysteines in the original protein are filled with numbers according to the constituent disulfide linkages. For example, if the first disulfide linkage filled for a disulfide structure is Cys58-Cys110, position 58 and 110 will be both filled with number 1. For one

disulfide structure, if all positions of disulfide cysteines are filled and each of them is only filled once, then that structure is a valid structure. Otherwise, if any position is filled twice with two different numbers, the corresponding structure is not valid. The process of forming disulfide structures continues until all possible combinations are enumerated. Such a typical process is illustrated in Table 5.6.

As shown in Table 5.6, there is only one disulfide structure, (Cys58-Cys110, Cys65-Cys72, Cys40-Cys95, Cys26-Cys84), in the **list D** because only one disulfide structure can be formed by the allowable disulfide linkages in list C (Table 5.6).

One of the advantages of this algorithm is that the number of candidate disulfide linkages is quadratic in the number of disulfide cysteines in the original protein, so the matching phase should be more efficient than that of the naive algorithm where the number of disulfide structures is exponential in the number of disulfide cysteines. The other advantage of this algorithm is that, in some cases, even if only an incomplete data set is obtained, the correct disulfide structure can be deduced. However, there is a condition for using this algorithm. That is, cyanylation needs to be complete for the nascent free cysteines after partial reduction of the original protein. Otherwise, incorrect deduction may result from an incorrectly identified negative signature mass. Let us examine what would happen if the condition of complete cyanylation cannot be met for the previous example of using itz-26-83 as a negative signature fragment for ribonuclease-A. Without the condition of complete cyanylation, any of the eight previously rejected disulfide linkages may be wrongly ruled out. For example, Cys26-Cys40 could be valid because itz-26-83 could result from the CN-induced cleavage of a doubly reduced and incompletely cyanylated isoform of ribonuclease-A, in which two

disulfide bonds Cys26-Cys40 and Cys84-Cys95 were reduced, but only Cys 26, Cys84, and Cys95 were cyanylated.

Table 5.6. Forming the disulfide structure list D from the shortened disulfide list C (Table 5.5) for ribonuclease-A.

S-S No.	Disulfide Cysteine Positions							
	C26	C40	C58	C65	C72	C84	C95	C110
1.1: Cys58-Cys110			1					1
2.1: Cys65-Cys72				2	2			
3.1: Cys40-Cys95		3					3	
4.1: Cys26-Cys84	4					4		
	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1.1: Cys58-Cys110			1					1
2.1: Cys65-Cys72				2	2			
3.1: Cys40-Cys95		3					3	
4.2: Cys26-Cys95	4						4	
							No	
1.1: Cys58-Cys110			1					1
2.1: Cys65-Cys72				2	2			
3.1: Cys40-Cys95		3					3	
4.3: Cys26-Cys110	4							4
								No
1.1: Cys58-Cys110			1					1
2.1: Cys65-Cys72				2	2			
3.2: Cys40-Cys110		3						3
								No

V. Dealing with Imperfect Data

Assuming we have perfect experimental data, there will be a perfect matching between the list of experimental peaks for a particular partially reduced/cyanylated species and the list of calculated masses of fragments for that species. However, we cannot assume that the list of experimental peaks will be perfect. It may include extra mass spectral peaks and/or some expected mass spectral peaks are missing. This leads to a situation where we will not get a perfect matching between the experimental peaks and the correct set of the calculated masses.

One possible way to deal with the imperfect data is to rank candidate partially reduced/cyanylated species by their matching percentage using the basic algorithm (Section III). An incomplete experimental data set could be obtained due to signal suppression and interference from matrix peaks (1, 2). As demonstrated in section III, in some cases, ranking the candidate partially reduced species may allow us to identify the correct species and thus, the correct disulfide bond(s).

Utilizing negative signature masses (Section IV) is another way to deal with incomplete experimental data sets. The algorithm based on negative signature masses does not even require partially reduced/cyanylated species of the original protein to be separated prior to CN-induced cleavage and complete reduction of the residual disulfide bonds.

Other sources of imperfect data are products of side reactions, such as incomplete cyanylation, β -elimination, and incomplete CN-induced cleavage (2). CN-induced fragments due to β -elimination and incomplete CN-induced cleavage can normally be identified by mass mapping; and sometimes provide complementary information for

determining disulfide connectivities (1, 2). Therefore, these side reactions need to be considered in computing the expected values. On the other hand, incomplete cyanylation may prevent us from using negative signature masses (Section IV). If incomplete cyanylation happened, the algorithm based on negative masses may reject all candidate disulfide structures of the original protein; the results from the basic algorithm based on matching percentages may contradict each other for identified partially reduced/cyanylated species. These can then be used as an indication that incomplete cyanylation occurred. In such cases, additional chemical or mass spectrometric analyses may be needed to determine where incomplete cyanylation occurred if they cannot be deduced using the described algorithms.

VI. Implementation and Testing of the Algorithms

The algorithms are coded in C++. The program code and its test results are listed as follows.

1. Implementation of the Basic Algorithm

A. C++ Program Files:

SSbyCyan1.cpp, the source file for the main program's source code

InputOutput.cpp, the source file for the Input and Output library functions to input the protein sequence, output results, etc.

InputOutput.h, the header file for the Input and Output library functions

DisulfideProtein_extern.h, the header file for declaration of all the extern variables

All these C++ files are available electronically.

B. Preliminary Testing Using Data from a Model Protein: Ribonuclease-A

Input from Keyboard:

Standard Error Limit: 2

Text Files for Input:

ExpMass1.txt, the text file that stores the experimental m/z values of the CN-induced cleavage fragments (MH^+) from a partially reduced/cyanylated species of ribonuclease-A

ProteinSequence.txt, the text file that stores the name of the protein, its sequence, etc.

Text Files for Output:

Result1.txt, the text file that stores the results: ranked lists of candidate singly, doubly, and triply reduced species of ribonuclease-A

C. The Content of Input and Output Text Files

The Content of ExpMass1.txt:

2705.3 4527.4 6548.5

The Content of ProteinSequence.txt:

A template for input file protein sequence, by Jianfeng (Frank) Qi, 6/27/2001.

NOTE: ALL ENTREES BEGIN WITH "*" AND TERMINATE WITH "#".

1) Name of the Protein

*Ribonuclease A, Bovine#

2) Sequence of the Protein (Amino Acids, Capital Single Letter Representation
except lower case "c" for free cysteines and upper case "C" for disulfide
cysteines that are involved in disulfide bonds)

NOTE: Space and return can be used within the sequence.

*KETAAAKFER QHMDSSSTSAA SSSNYCNQMM KSRNLTKDRC KPVNTFVHES
LADVQAVCSQ KNVACKNGQT NCYQSYSTMS ITDCRETGSS KYPNCAYKTT
QANKHIIVAC EGNPYVPVHF DASV#

3) User-defined Amino Acids

(Format: "AA_user monoisotopic_residue_mass average_residue_mass)

NOTE: Up to Num_AA_User (6 in this case) user-defined amino acids

(B, J, O, U, X, and Z) can be used.)

*B 0 0

J 0 0

O 0 0

U 0 0

X 0 0

Z 0 0#

4) Status of Cyanylation of Free Cysteines (1 (yes) or 0 (no), default is 1)

*1#

5) Status of Non-cyanylation Modification of Free Cysteines (1 (yes) or 0 (no),
default is 0)

*0#

6) Added Mass to Free Cysteines by the Non-Cyanylation Modification
(Format: "monoisotopic_mass average_mass", default is "0 0")

*0 0#

The Content of Result1.txt:

The Protein Name is: Ribonuclease A, Bovine. 124 AAs.

The AA sequence is:

KETAAAKFERQHMDSS TSAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLAD
VQAVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHII
VACEGNPYVPVHFDASV.

It has 4 Disulfides.

It has 8 CysS.

The positions of them are:

26 40 58 65 72 84 95 110

The experimental data are:

2705.3 4527.4 6548.5

The standard error limit for searching is: 2 Da

NOTE1: MatchedMassTypes

type 0: monoisotopic with -OH as C-terminus,

type 1: monoisotopic with -NH2 as C-terminus,

type 2: average with -OH as C-terminus,

type 3: average with -NH2 as C-terminus

Search Result from Singly Reduced Isoforms:

The following is the name of each of the items in each of the data entries and their appearing order in the results.

**NO: Cysi-Cysj has (N) matches: Mass1 FragID; MatchedMassTypes ... Mass_N
FragID; MatchedMassTypes**

1: 26 - 84 has 3 matches: 2705.3 1 to 25 ; 0 1 2 3 6548.5 26 to 83 ; 2 3 4527.4
84 to 124; 2 3
2: 26 - 40 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3
3: 26 - 58 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3
4: 26 - 65 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3
5: 26 - 72 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3
6: 26 - 95 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3
7: 26 - 110 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3
8: 40 - 84 has 1 matches: 4527.4 84 to 124; 2 3
9: 58 - 84 has 1 matches: 4527.4 84 to 124; 2 3
10: 65 - 84 has 1 matches: 4527.4 84 to 124; 2 3

Search Result from Doubly Reduced Isoforms:

1: 26 - 40, 58 - 84 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 4527.4 84 to 124; 2 3
2: 26 - 40, 65 - 84 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 4527.4 84 to 124; 2 3
3: 26 - 40, 72 - 84 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 4527.4 84 to 124; 2 3

4: 26 - 58, 65 - 84 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 4527.4 84 to 124; 2 3

5: 26 - 58, 72 - 84 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 4527.4 84 to 124; 2 3

6: 26 - 65, 72 - 84 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 4527.4 84 to 124; 2 3

7: 26 - 84, 95 - 110 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 6548.5 26 to 83 ; 2 3

8: 26 - 40, 58 - 65 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

9: 26 - 40, 58 - 72 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

10: 26 - 40, 58 - 95 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

Search Result from Triply Reduced Isoforms:

1: 26 - 40 , 58 - 65 , 72 - 84 has 2 matches: 2705.3 1 to 25 ; 0 1 2 3 4527.4
84 to 124; 2 3

2: 26 - 40 , 58 - 65 , 72 - 95 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

3: 26 - 40 , 58 - 65 , 72 - 110 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

4: 26 - 40 , 58 - 65 , 84 - 95 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

5: 26 - 40 , 58 - 65 , 84 - 110 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

6: 26 - 40 , 58 - 65 , 95 - 110 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

7: 26 - 40 , 58 - 72 , 84 - 95 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

8: 26 - 40 , 58 - 72 , 84 - 110 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

9: 26 - 40 , 58 - 72 , 95 - 110 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

10: 26 - 40 , 58 - 84 , 95 - 110 has 1 matches: 2705.3 1 to 25 ; 0 1 2 3

2. Implementation of the Algorithm Based on Negative Signature Masses

A. C++ Program Files:

SSbyCyan2.cpp, the source file for the main program's source code

InputOutput.cpp, the source file for the Input and Output library functions to input the protein sequence, output results, etc.

InputOutput.h, the header file for the Input and Output library functions

Signature.cpp, the source file for the signature library functions that deals with computations related to negative signature mass

Signature.h, the header file for the signature library functions that deals with computations related to negative signature mass

All these C++ files are available electronically.

B. Preliminary Testing using Data from a Model Protein: ribonuclease-A

Text Files for Input:

ExpMass2.txt, the text file that stores the experimental m/z values of the CN-induced cleavage fragments (MH^+) from one or more partially reduced/cyanylated species of ribonuclease-A

ProteinSequence.txt, the text file that stores the name of the protein, its sequence, etc.

Text Files for Output:

Result2.txt, the text file that stores the results: found negative signature masses, a list of bad disulfide bonds that were rejected by the found negative signature masses, a list of good disulfide bonds that are still possible, the possible good disulfide structures, etc.

C. The Content of Input and Output Text Files

The Content of ExpMass2.txt:

A template for input file ExpMass2.txt for the algorithm based on negative signature masses, by Jianfeng (Frank) Qi, 9/8/2001.

NOTE: THE MASS DATA BEGIN WITH "*".

-----	-----
Exp. Mass	Mass Deviation
(MH+)	(/Da)
-----	-----
6548.5	2
5766.8	2
6061.2	2

The Content of ProteinSequence.txt:

A template for input file protein sequence, by Jianfeng (Frank) Qi, 6/27/2001.

NOTE: ALL ENTREES BEGIN WITH "*" AND TERMINATE WITH "#".

1) Name of the Protein

*Ribonuclease A, Bovine#

2) Sequence of the Protein (Amino Acids, Capital Single Letter Representation
except lower case "c" for free cysteines and upper case "C" for disulfide
cysteines that are involved in disulfide bonds)

NOTE: Space and return can be used within the sequence.

*KETAAAKFER QHMDSSTSAA SSSNYCNQMM KSRNLTKDRC KPVNTFVHES
LADVQAVCSQ KNVACKNGQT NCYQSYSTMS ITDCRETGSS KYPNCAYKTT
QANKHIIVAC EGNPYVPVHF DASV#

3) User-defined Amino Acids

(Format: "AA_user monoisotopic_residue_mass average_residue_mass)

NOTE: Up to Num_AA_User (6 in this case) user-defined amino acids

(B, J, O, U, X, and Z) can be used.)

*B 0 0

J 0 0

O 0 0

U 0 0

X 0 0

Z 0 0#

4) Status of Cyanylation of Free Cysteines (1 (yes) or 0 (no), default is 1)

*1#

5) Status of Non-cyanylation Modification of Free Cysteines (1 (yes) or 0 (no),
default is 0)

*0#

6) Added Mass to Free Cysteines by the Non-Cyanylation Modification
(Format: "monoisotopic_mass average_mass", default is "0 0")

*0 0#

The Content of Result2.txt:

1) Name of the Protein

Ribonuclease A, Bovine

2) Sequence of the Protein

KETAAAKFERQHMDSSSTAASSSNYCNQMMKSRNLTKDRCKPVNTFVHESLAD
VQAVCSQKNVACKNGQTNCYQSYSTMSITDCRETGSSKYPNCAYKTTQANKHII
VACEGNPYVPVHFDASV

3) The Positions of Disulfide Cysteines in the Protein: Ribonuclease A, Bovine

26 40 58 65 72 84 95 110

4) A list of bad disulfide linkages that can be rejected by each of the matched
negative signature masses for the protein: Ribonuclease A, Bovine

NOTE1: MatchedMassTypes

type 1: monoisotopic with -OH as C-terminus,

type 2: monoisotopic with -NH₂ as C-terminus,

type 3: average with -OH as C-terminus,

type 4: average with -NH2 as C-terminus

NOTE2: FragID: (LeftAA, RightAA, Frag_Type)

Frag_Type = 0: N-terminal fragment, residue 1 is not Cys, 1-(Ci-1)

Frag_Type = 1: itz-fragment between two cysteines, itz-Ci-(Cj-1)

Frag_Type = 2: C-terminal fragment, residue Num_AA is not Cys, itz-Cj-

Num_AA

Frag_Type = 3: N-terminal fragment, residue 1 is CysSH, 1-(Ci-1)

Frag_Type = 4: N-terminal fragment, residue 1 is CysSCN, 1-(Ci-1)

Frag_Type = 5: C-terminal fragment, residue Num_AA is CysH, itz-Cj-Num_AA

The following is the name of each of the items in each of the data entries and their appearing order in the results.

ExpMass

MatchedCalMass; MatchedMassType; MatchedSigFragID (LeftAA, RightAA, Frag_Type);

BadSS_linkages (1st_CysS, 2nd_CysS)

6548.5

6547.47; 3; 26 83 1

26 40 26 58 26 65 26 72 40 84 58 84 65 84 72 84

5766.8

5767.49; 3;

5766.49; 4; 58 109 1

58 65 58 72 58 84 58 95 65 110 72 110 84 110 95 110

6061.2

6059.72; 1;

6062.77; 4; 40 94 1

40 58 40 65 40 72 40 84 58 95 65 95 72 95 84 95

5) The good disulfide linkage list for the protein: Ribonuclease A, Bovine

26 84

26 95

26 110

40 95

40 110

58 110

65 72

6) The good disulfide structures for the protein: Ribonuclease A, Bovine

Cys26-Cys84 Cys40-Cys95 Cys58-Cys110 Cys65-Cys72

VII. Conclusions

Computer-aided data processing for disulfide mass mapping based on partial reduction and cyanylation/CN-induced cleavage is possible. The concept of using negative signature masses facilitates the disulfide assignment process and, in some cases, allows identification of the disulfide structure even when an imperfect data set, resulting from side reactions and/or signal suppression, is obtained.

VIII. References

1. Wu, J. & Watson, J. T. A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci.* **6**, 391-398 (1997).
2. Qi, J. & Watson, J. T. Determination of the Disulfide Structure of Sillucin, a Highly Knotted, Cysteine-Rich Peptide, by Cyanylation/Cleavage Mass Mapping. *Biochemistry* **40**, 4531-4538 (2001).

MICHIGAN STATE LIBRARIES



3 1293 02305 5142