LIBRARY
Michigan State
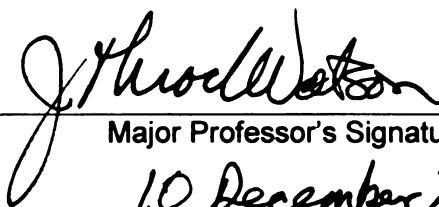University

This is to certify that the
dissertation entitled

APPLICATIONS AND COMPUTATIONAL
CONSIDERATIONS OF THE CYANYLATION-BASED
DISULFIDE MAPPING METHODOLOGY FOR CYSTINYL
PROTEINS

presented by

Wei Wu

has been accepted towards fulfillment
of the requirements for the

_____Ph.D._____ degree in _____Chemistry_____

_____
Major Professor's Signature

_____10 December 2003_____

Date

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

APPLICATIONS AND COMPUTATIONAL CONSIDERATIONS OF THE CYANYLATION-BASED DISULFIDE MAPPING METHODOLOGY FOR CYSTINYL PROTEINS

By

Wei Wu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

2003

ABSTRACT

APPLICATIONS AND COMPUTATIONAL CONSIDERATIONS OF THE
CYANYLATION-BASED DISULFIDE MAPPING METHODOLOGY FOR
CYSTINYL PROTEINS

By

Wei Wu

Disulfide bond conformation is very important to maintain the stability and

function of a cystinyl protein. Cyanylation (CN)-based disulfide mapping methodology is

very robust and versatile to characterize disulfide structures of cystinyl proteins. The

acidic condition for partial reduction and cyanylation efficiently thwarts the thio-disulfide

exchange reaction that generates disulfide artifacts. Cleavage of the amide bond at the N-

terminal side of a cysteine makes it possible to analyze disulfide structures with adjacent

cysteines.

Two large proteins (pea chloroplast fructose bisphosphatase, FBPase, 39 kDa and

cobalamin-independent methionine synthase, MetE from *E. coli*, 84 kDa) were subjected

to the CN-based disulfide mapping. For FBPase, two oxidized structures were identified

by mass spectrometry. For MetE, the mapping experiment showed that the reduced form

of the enzyme contains seven free cysteines, and that the oxidized form of the enzyme

has a glutathione adduct at Cys645.

A quantification method using homologous alkylating reagents was proposed and

evaluated. The problem of over-alkylation was minimized by adjusting the reagent-to-

peptide ratio and reaction time. Concentration ratio of cysteine-containing peptides

between 1:5 and 5:1 can be reliably represented by the peak intensity ratios of two differently alkylated peptides.

A computational analysis was performed to evaluate the CN-based disulfide mapping methodology with generic cystinyl proteins. Signature sets, minimal subsets of a fragment set to uniquely identify a disulfide structure, were generated from fragment sets and differential sets. The conciseness and multiplicity of signature sets ensure the efficiency of the method even when the majority of the expected CN-induced fragments were not detected. Analysis of the fragment sets showed that partial reduction should be directed to maximize the yield of both singly and doubly reduced isoforms. Experimental data from mapping experiments of ribonuclease A (RNase A), human epidermal growth factor (hEGF) and its two disulfide isomers were used to validate the concept of 'signature sets'.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

xi

CHAPTER 1

INTRODUCTION OF DISULFIDE MASS MAPPING

1. Biological Importance of Disulfide Bonds

Disulfide bonds stabilize the 3-dimensional conformation of proteins, thus helping to maintain their biological function. Each one of the disulfide bonds may contribute as much as 5~6 kcal/mol to the stability of a folded protein at optimal temperature (1,2). In protein engineering, efforts have been made to introduce one or several novel disulfide bonds to make the target protein more stable (3-5). Some successfully increase the stability of the target protein after inserting novel, non-native disulfide bonds, while others have the reverse effect. Nonetheless, the characterization of disulfide bond structures, in both the native and the mutant, is a very important part of these studies.

Disulfide bonds are implicated in 'mad cow disease', a transmissible spongiform encephalopathetic disease (6). The disease, unlike other diseases that are induced by bacterial or viral infections, is triggered by a conformational change of a key protein (the prion protein, PrP) from its normal cellular isoform (PrP^C) to an abnormal scrappie isoform (PrP^{Sc}). Recently, there is evidence showing that intermolecular disulfide bond formation may play a role in this conformational change. Disulfide formation in many other proteins is found as a key response to oxidative stress (7-9). In these proteins, the formation and dissociation of disulfide bonds is a 'redox switch' that regulates the protein activity.

In eukaryotic organisms, the secretory and membrane proteins, often objects of protein folding studies, are synthesized in the rough endoplasmic reticulum (ER),

followed by translocation into the cisternal space of the ER where the signal peptide is cleaved and other post-translationally modified, including disulfide bond formation and glycosylation. These assembled proteins are then translocated to the Golgi apparatus for further modification, and then either translocated to the cell surface or packaged into secretory vesicles for secretion (10).

It is a common practice in protein engineering to insert the DNA that encodes therapeutically important proteins into the prokaryotic host's DNA and use the host's transcription and translation machineries to express the proteins (known as the recombinant technology or 'cloning'). Bacteria, such as. *E. coli*, are usually used as the host expression system due to their fast production rate and convenient handling. Due to the lack of cytoplasmic organelles in prokaryotic organisms that facilitate protein folding, the desired over-expression of proteins often leads to the formation of insoluble aggregates, often referred to as inclusion bodies, in the cytoplasmic or periplasmic spaces. To solubilize these inclusion bodies and to restore the biological function of the protein, a refolding process is required. The inclusion body is usually first dissolved in buffer with a high concentration of the denaturant. Then, a step-wise dialysis lowers the concentrations of denaturant and eventually solubilizes the protein under physiological pH and ionic strength. A critical oxidative refolding is followed to establish the disulfide structure by exposing the protein to a redox buffer, typically containing oxidized glutathione and reduced glutathione with a concentration ratio between (1:3 to 1:1), the ratio present in the eukaryotic endoplasmic reticulum to hold ER slightly more oxidizing than the rest of the cytoplasm (11). A slightly basic pH (around 8) in the redox buffer promotes the formation of thiolate anion and assists the thio-disulfide exchange reaction.

Addition of molecular chaperones, e.g., heat-shock proteins and protein disulfide isomerase (PDI) may assist the refolding process (12). Isomers with non-native disulfide structures are often present along with the protein with the native disulfide structure after the folding process. It is important to separate them and determine the disulfide structures of the folding products so that only the one with the correct disulfide structure is subjected to further purification and therapeutic purposes.

2. Disulfide Mapping Approaches

2.1. Enzymatic Digestion-based Disulfide Mass Mapping

Mass Spectrometry (MS)-based disulfide mapping is based on controlled degradation of the target cystinyl protein, followed by the characterization of the degradation products by mass spectrometry. Disulfide connectivity, retained throughout the proteolysis, is explored by analyzing the digestion products, particularly those containing disulfide bonds. Chemical degradations, such as cyanogen bromide digest, specific to a methionine residue, are often employed. Because the sequence of the protein is known, and because the cleavage sites of the controlled degradations are usually specific, the masses of the fragments, including those containing disulfide bonds, can be calculated prior to the mapping experiment.

A typical example of this enzymatically based mass mapping can be found in the recent publication by Overgaard *et al.* (13). They studied the disulfide structure of a protein complex of pregnancy-associated plasma protein-A and the proform of eosinophil major basic protein (a 2:2 complex with a total M.W. of 500 kDa). A combination of

3

cyanogen bromide digestion and trypsin digestion was utilized to fragment the protein complex. Due to the unusually large size of this protein complex, the peptide digest mixture is separated with a 2-dimensional liquid chromatography system (cation exchange + reverse phase) due to its complexity. In Figure 1.1, fraction P31 off the 2nd dimensional reverse phase high performance liquid chromatography (RP-HPLC) was identified as a disulfide-containing peptide by matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS). By identifying the disulfide-containing peptides in the remaining HPLC fractions, the disulfide structure of the protein complex was resolved.
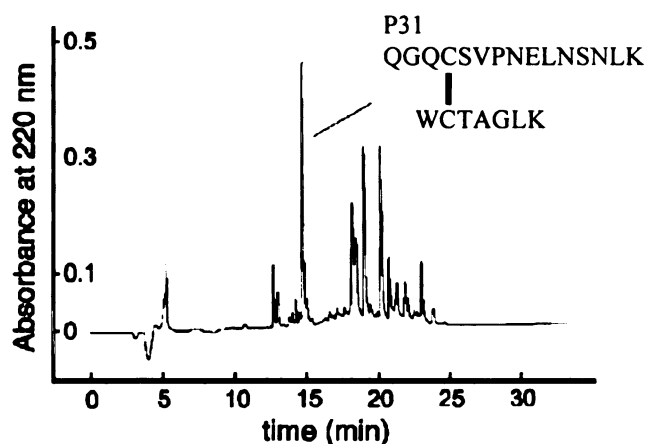
Figure 1.1. Chromatogram of 2nd dimension RP-HPLC run of one of the fractions collected from the 1st dimension cation exchange separation of the digest of the protein complex of pregnancy-associated plasma protein-A and the proform of eosinophil major basic protein. Fraction P31 contains a pair of peptides linked through a disulfide bond.

This protease-based disulfide mapping method has several potential drawbacks. First, the disulfide/cysteine containing peptides are only a very small fraction of the entire protein digest, due to the fact that cysteine has a natural occurrence of 2% among all the

4

20 natural amino acids. Separating and identifying these small fractions of peptides of interest poses a great challenge to analysts. Secondly, most of the specific proteases, such as trypsin, Lys-C, Glu-C, etc., function at a slightly alkaline pH, under which the thiolate anion (RS⁻), a strong nucleophile, is promoted from deprotonation of a sulfhydryl group and may attack a neighboring native disulfide bond to form an artificial disulfide isomer. Control of pH and time of digestion is often critical to retain the disulfide structure of the protein in the fragments. Lastly, in order to get a complete disulfide map, this method requires a sufficient number of cleavage sites distributed between cysteines. However, as two cysteines become closer to each other in the amino acid sequence, the chance of finding an available enzymatic cleavage site reduces accordingly. In the case of adjacent cysteines, cleavage between them by a protease becomes impossible, as no known proteases cleave at a cysteine residue. Protease-based disulfide mass mapping methods generate inconclusive disulfide structures in such cases.

2.2. Partial Reduction/Alkylation-based Methods

This method was proposed by Gray to study disulfide structure of small, highly bridged proteins (14). Those proteins are often resistant to proteolytic digestion due to their highly knotted structure. The method involves step-wise partial reduction, with different alkylating reagents capturing the reduced cysteine pairs at each step, and then using a protein sequence analyzer to identify the differently alkylated cysteines. The method works best for proteins that expose different disulfide bonds sequentially for reduction and alkylation under a successively more severe reducing environment. However, a protein's behavior under reducing conditions is less controlled by the

concentration of the reducing agent than by the protein's intrinsic property, i.e., its sequence. Additional degradation and identification steps of the differentially alkylated protein are needed if the protein size exceeds the limit of a protein sequencer. Variation of this method, in which mass spectrometry (LC/MS/MS) was used to identify the alkylation groups, was developed by Yen *et al.* (15).

2.3. Cyanylation (CN)-induced Cleavage-based Disulfide Mass Mapping

The method is based on selective cyanylation of the sulfhydryl group of a cysteine residue, and subsequent nucleophilic attack by NH_3 to promote cyanylation (CN)-induced cleavage of the peptide backbone. This specific chemical cleavage at the N-terminal side of a cysteine residue involves two steps, as illustrated in Scheme 1.1. In the first step, the hypothetical peptide (a 20mer with a cysteine at residue 10) is cyanylated with the cyanylating reagent, cyanodiaminopyridinium (CDAP) tetrafluoroborate. In the second step, the peptide bond between residue 9 and cyanylated Cys10 (in the middle of the scheme) breaks upon alkaline incubation with ammonia, resulting in a truncated peptide (1-9), and an iminothiazolidine (itz)-terminated peptide (itz10-20).

1

10 — SH

Cyanylation
—————→
CDAP,
pH=3

20

1

10 — SCN

CN-induced
Cleavage
—————→
NH₃

20

1

9

itz10

20
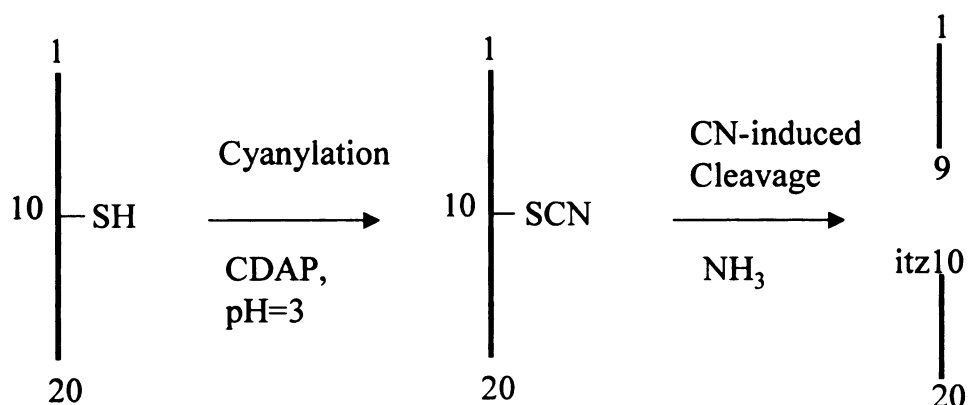
Scheme 1.1. Cyanylation-induced cleavage of a hypothetical cysteine-containing peptide generates two cleavage fragments.

Because cystines (disulfide bonds) are nonreactive with the cyanylating reagent CDAP, a multi-cystinyl protein must be partially reduced (to generate specific pairs of free sulfhydryl groups) using tris-carboxyethylphosphine (TCEP), as illustrated in Scheme 1.2 for a hypothetical 2-disulfide protein. In this example, the four constituent cysteines are located at residues 10, 20, 30 and 40. The four cysteines form two disulfide bonds (Cys10 with Cys30, and Cys20 with Cys40). When subjected to reduction, the disulfide bond between Cys20 and Cys40 may be reduced to form a 'singly reduced isoform', as illustrated on the right side of the equation in Scheme 1.2. The other singly reduced isoform, in which the disulfide between Cys10 and Cys30 is reduced, may be formed at the same time, as well as the doubly/totally reduced isoform, in which both disulfide bonds are reduced (structures not shown). The partially reduced isoforms are immediately exposed to the cyanylyating reagent, CDAP, to covalently modify all the free sulfhydryl groups for the subsequent cleavage reaction. The low pH during both partial reduction and cyanylation suppresses the thiol-disulfide exchange reaction that could lead to disulfide artifacts. Reduction of one disulfide bond generates two nascent

sulfhydryl groups (SH) and causes a +2-Da shift; cyanylation of the cysteine pair results in a +50-Da mass shift ([SH → SCN] ×2, [26-1]×2=50 Da). Thus, the net mass shift in going from an oxidized sulfur (as in a disulfide bond) is 26 Da (25+1) per sulfur involved in the chemical modification. The degree of partial reduction can be determined with a mass measurement of the partially reduced and cyanylated isoform, e.g., a doubly reduced/cyanylated isoform would have a +104-Da mass shift from that of the fully oxidized protein.
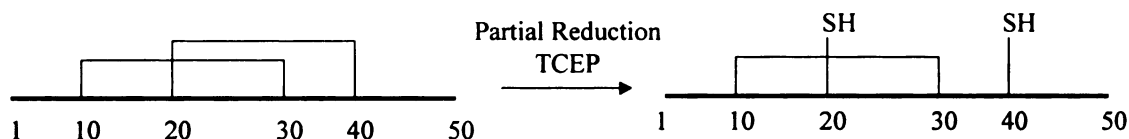
Scheme 1.2. Generation of one of the two possible singly reduced isoforms of a hypothetical 2-disulfide protein.

Once cyanylated and subjected to nucleophilic attack in 1M NH₄OH, the polypeptide backbone of the singly reduced and cyanylated isoform of the hypothetical protein shown in Scheme 1.3 is cleaved at the N-terminal side of the two cyanylated cysteine residues. However, as shown in the middle of Scheme 1.3, a residual disulfide bond, in this example, holds two of the three cleavage products together. Thus, the cleavage reaction mixture represented by the middle of Scheme 1.3 is treated with an excess of reducing reagent (TCEP) to reduce the residual disulfide bond, thereby ensuring that all CN-induced cleavage products, 1-19, itz20-39, and itz40-50, are free as shown at the end of Scheme 1.3. With the mass mapping of these three fragments and the knowledge that they are from a singly reduced isoform, we can deduce that Cys20 and Cys40 are involved in a disulfide bond in the original protein.

8

Similarly, when subjected to cyanylation, cleavage and total reduction, the other possible partially reduced isoform (not illustrated), in which the disulfide bond between Cys10 and Cys30 is reduced, generates three fragments, 1-9, itz10-29 and itz30-50. By detecting them through mass mapping, the other disulfide bond is confirmed; of course, in such a case involving two disulfide bonds in a protein with only four cysteines, a determination of the connectivity of cysteines in one disulfide bond allows the connectivity of the other two cysteines to be established by default.
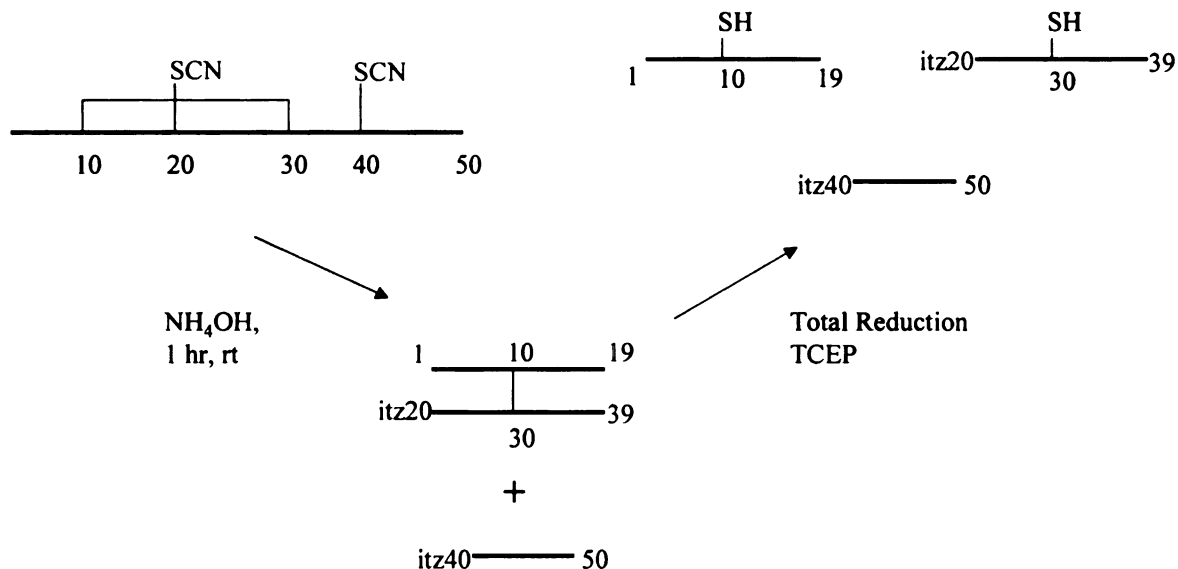
Scheme 1.3. Cleavage of a singly reduced and cyanylated isoform, after total reduction (to break the residual disulfide bond), generates three fragments.

In summary, the original strategy for CN-based mass mapping(16) involves three key steps: partial reduction/cyanylation, HPLC separation of the partially reduced isoforms/mass measurement to determine the reduction status, and alkaline cleavage/total reduction and mass spectral analysis. The overall disulfide structure is established by determining the individual disulfide bonds with the detection of sets of three fragments

from the cleavage reaction of singly-reduced isoforms. This method has been used successfully to solve the disulfide structures of ribonuclease A (13 kDa) and sillucin (3 kDa with three adjacent cysteines) (16,17).

3. Computer-assisted Analysis of Disulfide Bonds

3.1. Homology-based Approaches

A study of the cystinyl proteins in the protein databank has shown that the symmetric and reducible (disulfide structure has at least two non-overlapping domains) disulfide structures occur at a much higher rate than just by chance (18). Attempts have been made to predict the cysteine status (free or disulfide-bonded) based on the amino acid sequence based on the assumption that the local secondary structure affects the propensity of a cysteine to be free or oxidized. Sequence alignment using 7-17 residues around the cysteine residues to all the sequences in the cysteine-containing protein database suggests predictors of that cysteine being free or being oxidized. After tuning the parameters with a set of protein sequences with known structures, the prediction efficiency was reported to be as high as 81% (19). However, this program only predicts whether a cysteine is free or connected, not how it is connected when it is in the form of a cystine.

Another homology-based algorithm using integer linear programming (ILP) predicts both the beta sheet and the disulfide bridges (connectivity) successfully with model proteins such as bovine pancreatic trypsin inhibitor (BPTI) (20). However, the chance of success of these homology-based methods highly depends on the presence of a

highly homologous protein, the structure of which has already been solved either by chemical means or by crystallography data, in the existing protein database.

3.2. Mass Spectrometry (MS)-based Data Analysis

Computer-assisted data analysis has been utilized to analyze the CN-based disulfide mapping experimental data. We observed that the two terminal cysteines that define a particular cleavage fragment could not have been involved in a disulfide bond with the internal cysteines in that fragment. The m/z value of the ion representing the CN-induced cleavage fragment containing internal free cysteines is referred to as a 'negative signature mass (NSM)' (21). A computer algorithm, negative signature mass algorithm (NSMA), was developed to determine protein disulfide structures based on 'ruling out' certain disulfide linkages and reconstructing the disulfide structure with the surviving disulfide linkages. NSMA has been used successfully to solve the disulfide structures of a four-disulfide protein, ribonuclease A, and a six-disulfide protein, transforming growth factor β type II receptor extracellular domain.

4. Summary

In chapter 2 of the dissertation, results from the cyanylation-based disulfide mass mapping experiments of reduced and oxidized forms of two large proteins (chloroplast fructose bisphosphatase, FBPase, 39 kDa; cobalamin-independent methionine synthase MetE from *E. coli,* 85 kDa) are presented and discussed. These applications represented the first attempts to use this method to analyze disulfide structures of proteins over 15 kDa. Both enzymes have a dynamic disulfide structure: the reduced form and the

oxidized form. Change in disulfide structure, as a response to the changes (light in FBPase and redox potential in MetE) in the environment, directly regulates the activity of the enzymes.

Chapter 3 explores a novel mass spectrometry-based quantification method of cysteinyl peptides. The method involves alkylation of a cysteine-containing peptide at two different states with two homologous reagents. The relative quantity ratio of the same cysteine-containing peptide at two different states is represented by area ratio of the mass spectral peaks corresponding to the two differently alkylated peptides. The over-alkylation problem was observed, and minimized by adjusting the reaction conditions after a survey of the kinetics of the two alkylation reactions. This quantification method could be used to quantify the two isomeric structures identified in the oxidized forms of FBPase.

Chapter 4 presents the theoretical aspects of the cyanylation(CN)-based mass mapping method. The concept of 'signature sets' is proposed, followed by a careful study of the fragment set associated with a disulfide structure. The conciseness of the signature sets proves the robustness of this methodology: The disulfide structure of a fully-oxidized cystinyl protein can be determined by the CN-based mapping approach with the detection of a very small fraction of all the possible cleavage fragments. The 'signature sets' concept was validated with the data from the disulfide mapping experiments of human epidermal growth factor (hEGF) and its two disulfide iomers (generated from oxidative refolding) and ribonucleaes A (RNase A).

Reference:

12

1.   Betz, S. F. (1993) *Protein Sci* **2**, 1551-1558

2.   Darby, N., and Creighton, T. E. (1997) *Mol Biotechnol* **7**, 57-77

3.   Futami, J., Tada, H., Seno, M., Ishikami, S., and Yamada, H. (2000) *Journal of Biochemistry* **128**, 245-250

4.   Johnson, C. M., Oliveberg, M., Clarke, J., and Fersht, A. R. (1997) *J Mol Biol* **268**, 198-208

5.   Hinck, A. P., Truckses, D. M., and Markley, J. L. (1996) *Biochemistry* **35**, 10328-10338

6.   Welker, E., Wedemeyer, W. J., and Scheraga, H. A. (2001) *Proc Natl Acad Sci U S A* **98**, 4334-4336

7.   Linke, K., and Jakob, U. (2003) *Antioxidants & Redox Signaling* **5**, 425-434

8.   Eaton, P., Jones, M. E., McGregor, E., Dunn, M. J., Leeds, N., Byers, H. L., Leung, K. Y., Ward, M. A., Pratt, J. R., and Shattock, M. J. (2003) *J Am Soc Nephrol* **14**, S290-S296

9.   Harding, H. P., Zhang, Y. H., Zeng, H. Q., Novoa, I., Lu, P. D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., Stojdl, D. F., Bell, J. C., Hettmann, T., Leiden, J. M., and Ron, D. (2003) *Mol Cell* **11**, 619-633

10. Ruddon, R. W., Sherman, S. A., and Bedows, E. (1996) *Protein Sci* **5**, 1443-1452

11. Gilbert, H. F. (1997) *J Biol Chem* **272**, 29399-29402

12. Thomas, J. G., Ayling, A., and Baneyx, F. (1997) *Appl Biochem Biotechnol* **66**, 197-238

13. Overgaard, M. T., Sorensen, E. S., Stachowiak, D., Boldt, H. B., Kristensen, L., Sottrup-Jensen, L., and Oxvig, C. (2003) *J Biol Chem* **278**, 2106-2117

14. Gray, W. R. (1993) *Protein Sci* **2**, 1732-1748

15. Yen, T. Y., Yan, H., and Macher, B. A. (2002) *J Mass Spectrom* **37**, 15-30

16. Wu, J., and Watson, J. T. (1997) *Protein Sci* **6**, 391-398

17. Qi, J. F., Wu, J., Somkuti, G. A., and Watson, J. T. (2001) *Biochemistry* **40**, 4531-4538

18. Benham, C. J., and Jafri, M. S. (1993) *Protein Sci* **2**, 41-54

19. Fariselli, P., Riccobelli, P., and Casadio, R. (1999) *Proteins-Structure Function and Genetics* **36**, 340-346

20. Klepeis, J. L., and Floudas, C. A. (2003) *Journal of Computational Chemistry* **24**, 191-208

21. Qi, J. F., Wu, W., Borges, C. R., Hang, D. H., Rupp, M., Torng, E., and Watson, J. T. (2003) *Journal of the American Society for Mass Spectrometry* **14**, 1032-1038

CHAPTER 2

CYANYLATION-BASED DISULFIDE MAPPING OF TWO LARGE PROTEINS: CHLOROPLAST FRUCTOSE BISPHOSPHATASE AND COBALAMINE-INDEPENDENT METHIONINE SYNTHASE

1. Chloroplast Fructose Bisphosphatase (FBPase)

1.1. Introduction

The chloroplast fructose bisphosphatase (Enzyme Classification, E.C. 3.1.3.11) participates in the Calvin cycle (photosynthetic CO_2 fixation). The enzyme is reductively activated in the light *in vivo* (1). Oxidative inactivation of fructose bisphosphatase, glyceraldehyde-3-P dehydrogenase, sedoheptulose bisphosphatase, and phosphoribulokinase prevents the futile operation of the Calvin cycle in the dark; light activation reverses the inactivation and allows photosynthetic CO_2 fixation to proceed (2-4). In chloroplast fructose bisphosphatases, there is an insertion that contains redox-sensitive cysteines (Figure 2. 1) in the shape of a loop that protrudes from the surface of the reduced form of the enzyme. Although site-directed mutagenesis experiments indicate that each of the three Cys residues in the loop can form disulfide bonds with the other Cys residues in the loop (5-7), there are only two, Cys153 and Cys173, close enough to be joined by a disulfide bond in the crystal structures of the oxidized pea chloroplast enzyme (8). We reasoned that mass mapping should allow us to determine whether there are additional disulfide bonds in the oxidized form of the enzyme. The results indicate that there is an alternate disulfide bond between Cys173 and Cys178 in some of the molecules of the pea chloroplast enzyme, as isolated from pea leaves, and

15

between the corresponding cysteine residues in the recombinant spinach enzyme, as isolated.

Figure 2. 1. Structure of spinach chloroplast fructose bisphosphatase in protein data bank (PDB) file 1spi (9). Ring D is highlighted. The α-carbons of the Cys residues in the insertion into the chloroplast enzyme are represented as spheres. Clockwise from left to right they are Cys179, Cys174 and Cys155. (For more detail see Figure 2. 6) In the crystal structure of the pea chloroplast enzyme in PDB file 1d9q (8) Cys153 and Cys173 (correspond to 155 and 174 in PDB file 1spi) are involved in a disulfide bond. The figure was constructed using MOLSCRIPT (10).

1.2. Materials and methods

Plant material

Pea (*Pisum sativum* L. var Little Marvel) plants were grown from seed in the University of Illinois at Chicago greenhouse as described previously (11). Seeds were purchased from Old's Seed Company, Madison WI.

Enzyme isolation

All steps in the purification of the pea leaf enzyme were performed at or below pH 6.5, in order to avoid thiol-disulfide exchange and scrambling of disulfide bonds. The enzyme is stabilized by phosphate, hence potassium phosphate was used as the buffer throughout. All operations were carried out on ice or at 0 to 4°C. Shoots (500 to 700 g) from 12 to 14-day pea plants were harvested, washed in deionized water, and frozen overnight at -30°C. The following morning the frozen tissue was blended in twice its weight of 50 mM, pH 5.5 potassium phosphate and 0.4-times its weight of polyvinylpolypyrollidone, filtered through a wire-mesh strainer, 4 layers of cheesecloth and one layer of lutrasil (Lutravil Co., Durham NC) and centrifuged (13,200g, 20 min). The green supernatant solution was slowly made 40% saturated with respect to ammonium sulfate by addition of 0.231 g pulverized $(NH_4)_2SO_4$ per mL, stirred for 25 min and centrifuged (13,200g, 30 min). The straw colored supernatant solution was made 60% saturated with respect to ammonium sulfate (by addition of 0.125 g

17

(NH$_4$)$_2$SO$_4$ per mL), stirred for 15 min and centrifuged (5,860g, 20 min). The pelleted protein was suspended in 50 mM, pH 6.5 potassium phosphate (the phosphate buffer) and dialyzed overnight against 1 L of the buffer with one change of buffer, and frozen. Freezing did not appear to affect the activity of the enzyme at this point. When the isolation was continued, the retentate in the dialysis bag was thawed, centrifuged (27,000g, 15 min), and the supernatant solution was applied to a 2.54 x 9 cm column of Fast-flo DEAE-cellulose in the phosphate buffer. The column was washed with 500 mL of the buffer and then with 5 L of the buffer containing 0.05 M KCl. The enzyme was eluted with a linear 0.05 to 0.4 M KCl gradient (200 mLs + 200 mLs) in the phosphate buffer. An equal volume of 1 M pH 6.5 potassium phosphate buffer that was 40% saturated with respect to ammonium sulfate was added to the combined active fractions and the solution was applied to a 1 mL phenyl Sepharose CL-4B column in 0.5 M potassium phosphate (pH 6.5) that was 20% saturated with respect to ammonium sulfate. The column was washed with 50 mLs of that buffer and then the enzyme was eluted in 0.5 M pH 6.5 potassium phosphate. The active fractions were concentrated on Millipore Centriprep and Centricon 30 filters and applied to a 1 x 22 cm Superose 12 10/20 column in 50 mM pH 6.5 potassium phosphate buffer. Active fractions were collected and concentrated on a Centricon 30 filter. Purity was assessed by SDS gel electrophoresis.

This method was not particularly efficient in terms of enzyme recovery. Reproducibility was a problem. The yield was never more than 8% of the activity in the dialyzed ammonium sulfate fraction that was applied to the DEAE cellulose column. There was considerable loss of activity at the phenyl-Sepharose step. The highest specific activity obtained was about 150 units per one mg of protein. Many preparations

18

were not homogeneous and had to be discarded. This purification procedure did allow purification of the enzyme at a pH unfavorable for thiol-disulfide exchange.

Prior to cyanylation, any extraneous proteins present in the enzyme preparation were separated from the fructose bisphosphatase by reverse phase HPLC on a Waters system (Waters 2695 module with a Waters 2487 dual wavelength detector). 100 μg of protein in 25mM pH 6.5 potassium phosphate buffer was injected onto a Vydac C18 (catalog number 218TP54, 4.6 mm i.d., 250 mm length) column. The gradient was 2% to 90% solvent B in 50 min, with solvent A being 0.1% TFA (trifluoroacetic acid) in water and solvent B being 0.1% TFA, 90% acetonitrile in water. The flow rate was 1mL/min. The detector was set at 214 nm. Fractions were collected manually. The masses of the fractions were determined by MALDI (matrix-assisted laser desorption ionization) mass spectrometry. Fractions that contained proteins with masses corresponding to the mass of fructose bisphosphatase were combined and taken to dryness by lyophilization.

The recombinant spinach fructose bisphosphatase protein was kindly provided by Renate Scheibe, Universität Osnabrück. It was produced and purified as described by Reichert *et al.* (5). All operations were performed at pH values of 6.5 or lower. The protein appeared to be homogeneous on SDS gel electrophoresis.

Activity assay, protein estimation

Activity of the pea leaf enzyme was routinely assayed at pH 8.8. At this pH, changes in the redox status of the protein do not affect activity. Cuvettes contained 100 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM fructose-1,6-bisphosphate, 1 mM NADP, and 0.5 units

glucose-6-phosphate dehydrogenase and 1.5 units glucose-6-phosphate isomerase per mL. Activity was followed by change in absorbance at 340 nm on a Varian Cary 210 spectrophotometer at room temperature (23°C). Protein was assayed according to Bradford (12).

Cyanylation (CN) and CN-induced cleavage

For cyanylation of the pea chloroplast fructose bisphosphatase, 30 µL of 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate solution (150 mM CDAP, 8 M urea, 100 mM sodium citrate at pH 3) was added to the combined lyophilized fructose bisphosphatase fractions. After 20 min at room temperature, the cyanylated protein was separated from the cyanylating reagents by reverse phase HPLC, as described above. The volume of the protein-containing fraction from the HPLC column was reduced to 10 µl under vacuum and 10 µL of 6 M guanidine hydrochloride was added. Cleavage was accomplished by adding 20 µL of 1 M NH_4OH and incubating the mixture at room temperature for an hour. Excess ammonia was removed under vacuum. After addition of 40 µL of 30% (v/v) acetonitrile/0.1% TFA to the dried cleavage products, the samples were analyzed by MALDI-TOF-MS.

The recombinant spinach enzyme (insoluble in water) was washed with 500 µL distilled water to remove the residual sodium acetate from the lyophilization step, and was collected by centrifugation (14,000g, 8 min). 50 µL of 50 mM CDAP, 6M guanidine-HCl, 100 mM sodium citrate buffer (pH 3) was added to the pellet to dissolve and cyanylate the protein. After 15 min, the reaction was terminated by separation of the

cyanylated protein from the reaction mixture on a Bio-Rad 6 spin column that had been

pre-equilibrated with 8 M urea. For the alkaline cleavage reaction, 50 μL of 4 M NH₄OH

was added to the cyanylated protein (50 μL, in 8 M urea) to make the final NH₄OH

concentration 2 M. After incubation for 1 hour at room temperature, the cleavage

reaction was terminated by removing the ammonia by vacuum aspiration. The cleavage

products were dissolved in 40 μL of 0.1% aqueous TFA.

Mass spectrometry

MALDI mass spectra were obtained on a Voyager DE-STR mass spectrometer (Applied

Biosystems, Foster City, CA) equipped with a 337-nm nitrogen laser. The accelerating

voltage in the ion source was set at 25 kV. A solution of saturated α-cyano-4-

hydroxycinnamic acid (Aldrich Chemical Co.) in 1:2 of acetonitrile/water was diluted

with isopropanol (1:4 v/v). 30μL of the diluted solution was applied to the flat MALDI

probe using Cadene's protocol (13) to form an 'ultra-thin layer'. The matrix solution

consisted of a saturated solution of α-cyano-4-hydroxycinnamic acid (Aldrich Chemical

Co.) in 1:1 of acetonitrile/water. 1 μL of the cleaved peptide solution was diluted with 9

μl of matrix solution and 0.5 μL of the peptide/matrix mixture was deposited onto the

probe. Co-crystallization was usually observed in 1 min. Residual liquid above the co-

crystallized peptide/matrix solid was removed by vacuum aspiration and 10 μL of water

was deposited on the crystals; after 15 seconds, the water was removed by vacuum

aspiration. This washing process was repeated two more times in order to remove the

majority of the salt contaminant, the guanidine hydrochloride (pea enzyme) or urea

(spinach enzyme), which, at a high concentration, suppresses the protein signals in MALDI. External calibration was used with both the intact enzymes and their CN-induced cleavage products.

For the ESI-LC/MS experiment with the recombinant spinach protein, the cleavage peptide mixture was mixed 1:1 with 10% acetic acid (v:v). 6 μL of sample was loaded onto a peptide capillary trap (Michrom 004/25108/32) through the auto sampler of the CapLC system (Waters Co., Milford, MA). An auxiliary pump was used to wash the trap with 1% formic acid in water at a flow rate of 20 μL/min for 5 min. At the end of the first 5–min wash, an automated valve on the CapLC system switched allowing the gradient mixture of solvent A and B (Solvent A: 1% formic acid in water; Solvent B: 1% formic acid in acetonitrile) to flow through the trap and onto the nano-spray column (PicoFrit column from NewObjective, PFC7515-AQ-5, 15-μm tip i.d., 75-μm column i.d., 5 cm of C18 packing before the frit at the tip). The gradient started at 5% of B at 5 min and went to 70% in 85 min, then it ramped up to 90% in 5 min and remained at 90% for 20 min. The flow rate from the PicoFrit column tip was 0.1 μL/min. The flow from the pump was 9 μL/min, allowing the gradient to be ramped up or down smoothly. A T-splitter between the CapLC pumps for solvent A/B mixture and the capillary trap interfaced the two flow rates. The flow from the PicoFrit column tip was directly sprayed into a Finnigan LCQ-Deca mass spectrometer (Thermo Finnigan, San Jose, CA). The source voltage was set at 2.7 kV and the heated capillary temperature was set at 200°C. The quadruple ion trap mass spectrometer scanned from m/z 200 to m/z 2000 at a rate of 1.5 sec/scan. No data-dependent scans were used. The tune method was created by infusion of 1 pmol/μL of angiotensin peptide at the same flow rate (0.1 μL/min).

Modeling

In order to evaluate the suitability of sites identified for the formation of disulfide bonds, tertiary structure diagrams were made and displayed with Accelrys (Biosym Technologies, San Diego, CA) on a Silicon Graphics work station. Subsequent energy minimizations using the Discover Molecular Dynamics Package (Biosym Technologies) were performed as previously described (14). Briefly, minimization was initiated with all heavy atoms subject to a tethering force of 2000 kcal/Å^2. The tethering force was gradually reduced during steepest descents minimization, which was followed by subsequent rounds of complex gradient minimization. The distance-dependent dielectric constant directly proportional to atom separation was used throughout.

1.3. Results and Discussion

Mass Mapping

The final reverse phase HPLC step in the purification of the pea chloroplast enzyme yielded a broad peak at the retention time corresponding to fructose bisphosphatase protein. When this component was analyzed by MALDI-TOF-MS, five peaks were detected: mass spectral (MS) peaks at *m/z* 39059.0, 19529.8, 12999.8, 9735.4 and 7791.9, corresponding to the intact enzyme carrying 1 to 5 protons. After

23

cyanylation and HPLC separation, the major protein peak consisted of cyanylated

fructose bisphosphatase, as determined by MALDI-MS analysis.

Peak	Identity	Observed mass (Da)	Calculated mass (Da)	Deviation
1	doubly protonated itz-306-357	2889.33	2891.265	-0.067%
2	itz-153-189 with a disulfide bond between 173-178	3953.98	3955.30	-0.033%
3	itz-49--91	4657.16	4658.19	-0.022%
4	2--48	5183.22	5182.98	0.005%
5	itz-306-357	5781.51	5781.53	0.000%
6	itz-92-152	6426.83	6426.90	-0.001%
7	itz-92-177 with a disulfide bond between 153-173	9164.86	9162.85	0.022%
8	itz-92-189 with a disulfide bond and a β-elimination among 153,173,178	10303.85	10304.20	-0.003%
9	itz-306-357 dimer	11558.61	11562.02	-0.029%
10	itz-190-305	13290.20	13289.27	0.007%

Table 2. 1. Assignment of peaks in the MALDI mass spectrum (Figure 2. 2) after CN-induced cleavage of the pea chloroplast fructose bisphosphatase.
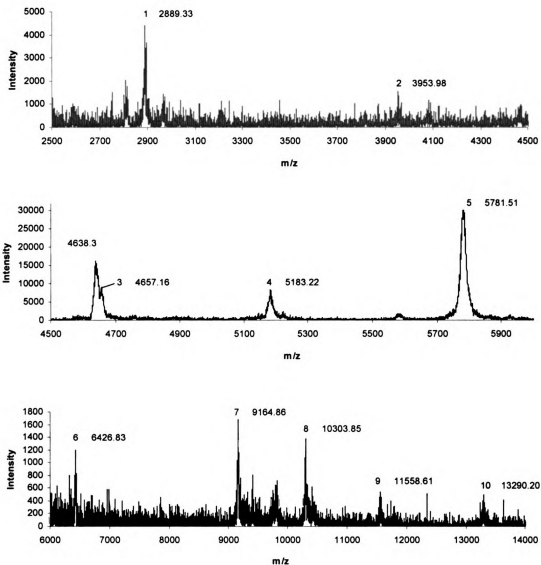
Figure 2. 2. MALDI mass spectrum of CN-induced cleavage products of pea chloroplast fructose bisphosphatase, divided into three traces. Peaks were automatically marked at the centroid by the Data Explorer software. The mass spectrum is divided into three mass regions because the peaks in mass region two (middle panel) are significantly more intense than those in regions one and three. The experimentally determined *m/z* values together with calculated masses for expected protonated CN-induced cleavage fragments appear in Table 2. 1.

Structure A

```
        SCN  SCN  SCN           SCN  SCN     Cleavage, NH₃        1——48        itz49——91
         |    |    |             |    |
—49— 92 —153—│73 —│78 —190 —306 —              ─────────▶  itz92 —152  itz190——305  itz306——357
              ·····

                                                                     itz153——1│3——│78—— 189  *
                                                                           ·····
                                                                     (signature fragment)
```

Structure B

```
        SCN  SCN             SCN  SCN  SCN     Cleavage, NH₃        1——48        itz49——91
         |    |               |    |    |
—49— 92 —153—173 —178 —190 —306 —              ─────────▶  itz178——189  itz190——305  itz306——357
            └──────┘

                                                                     itz92 —153——173—— 177  *
                                                                           └──────┘
                                                                     (signature fragment)
```
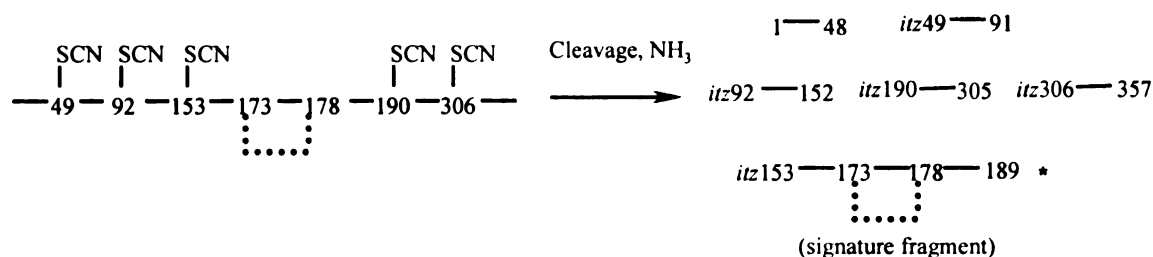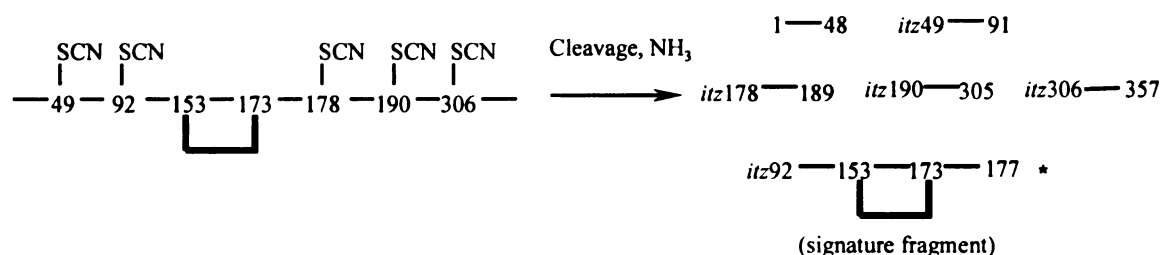
Scheme 2. 1. The oxidized form of the native pea chloroplast fructose bisphosphatase is subjected to cyanylation and alkaline cleavage. All the free cysteines at the starting point are cyanylated and specifically cleaved, resulting in six fragments. In the upper panel, the observation of the signature fragment, marked with an asterisk, supports the existence of Structure A, in which there is a disulfide bond between Cys173 and Cys178. Likewise, the observation of the fragment itz-92-177 (in the lower scheme for Structure B) indicates a disulfide bond between Cys153 and Cys173.

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 (15). If the sequence of the protein is known, mass mapping analysis (comparing the experimentally-determined masses from mass spectrometry against the calculated masses of the expected cyanylation-induced cleavage fragments) will allow identification of the cysteine status. As shown in Figure 2. 2 and Table 2. 1, mass mapping analysis of the CN-induced cleavage products of cyanylated pea chloroplast fructose bisphosphatase

indicated the presence of CN-induced fragments consisting of residues 2-48, itz-49-91, itz-92-152, itz-92-177, itz-92-189, itz-153-189, itz-190-305, itz-306-357, itz-306-357 (doubly charged), and an itz-306-357 dimer (where itz = iminothialzolidine-carboxyl blocked amino terminus). CN-induced cleavage occurred at six cyanylated cysteine residues as illustrated in Scheme 2. 1. Cleavage did not occur at Cys173. One possibility could be that in the original enzyme as isolated, Cys173 is disulfide bonded and is never free, or free Cys173 might be present in the original enzyme, yet its CN-induced cleavage products itz-153-172 and itz-173-177 were not detected, possibly due to the signal suppression from other peptides in MALDI, or from interference from matrix ion background. In itz-92-177, Cys153 must be involved in a disulfide bond with Cys173, because these are the only two otherwise unmodified in this fragment, and they must be oxidized to match the calculated mass value of the fragment. In itz-153-189, Cys173 must be involved in a disulfide bond with Cys178, as both of these residues are oxidized to fit the mass calculation (and neither is available for cyanylation). Itz-92-189 represents a peptide containing a disulfide bond between two Cys residues and a third Cys residue that has undergone β-elimination. Details about β-elimination and other side reactions have been described previously (16). These data are consistent with Cys173 forming two alternate disulfide bonds in this enzyme: Cys153-Cys173 and Cys173-Cys178. Fragments that would have indicated a Cys153-Cys178 disulfide bond, namely itz-92-172 and itz-173-189 linked by the disulfide bond (m/z 10381.1) and its prompt fragmentation(17) products, itz-92-172 (m/z 8636.1) and itz-173-189 (m/z 1748.0), were not observed.

Peak 4 of the MALDI spectrum of the pea chloroplast fructose bisphosphatase (Table 2. 1) corresponds to a CN-induced cleavage product consisting of residues 2 through 48. The transit peptide must then be cleaved after, not before, the methionine shown as residue 1 in PDB file 1dcu (8) when the enzyme is transported into the chloroplast. That amino acid sequence is based on DNA sequencing. This is the first identification of the transit peptide cleavage site and the N-terminal amino acid for pea chloroplast fructose bisphosphatase. To avoid confusion, the numbering from the PDB files has been retained for the crystal structure of the pea enzyme.

Peak	Identity	Observed mass (Da)	Calculated mass (Da)	Deviation
1	doubly-protonated itz-51-93	2285.35	2286.57	-0.053%
2	doubly-protonated itz-307-358	2904.13	2904.30	-0.006%
3[a]	Itz-155-190 with 174 and 179 oxidized	3917.06	3918.21	-0.029%
4	itz-51-93	4571.21	4572.13	-0.020%
5	itz-307-358	5807.80	5807.60	0.003%
6[a]	itz-94-178 with 155 and 174 oxidized	9097.61	9097.76	-0.002%

Table 2. 2. Assignment of peaks in the MALDI mass spectrum after CN-induced cleavage of the recombinant spinach chloroplast fructose bisphosphatase

[a]Peak 6 indicates a disulfide bond between Cys155 and Cys174, consistent with the crystallography data. The alternative disulfide bond between Cys174 and Cys179 is indicated by peak 3.
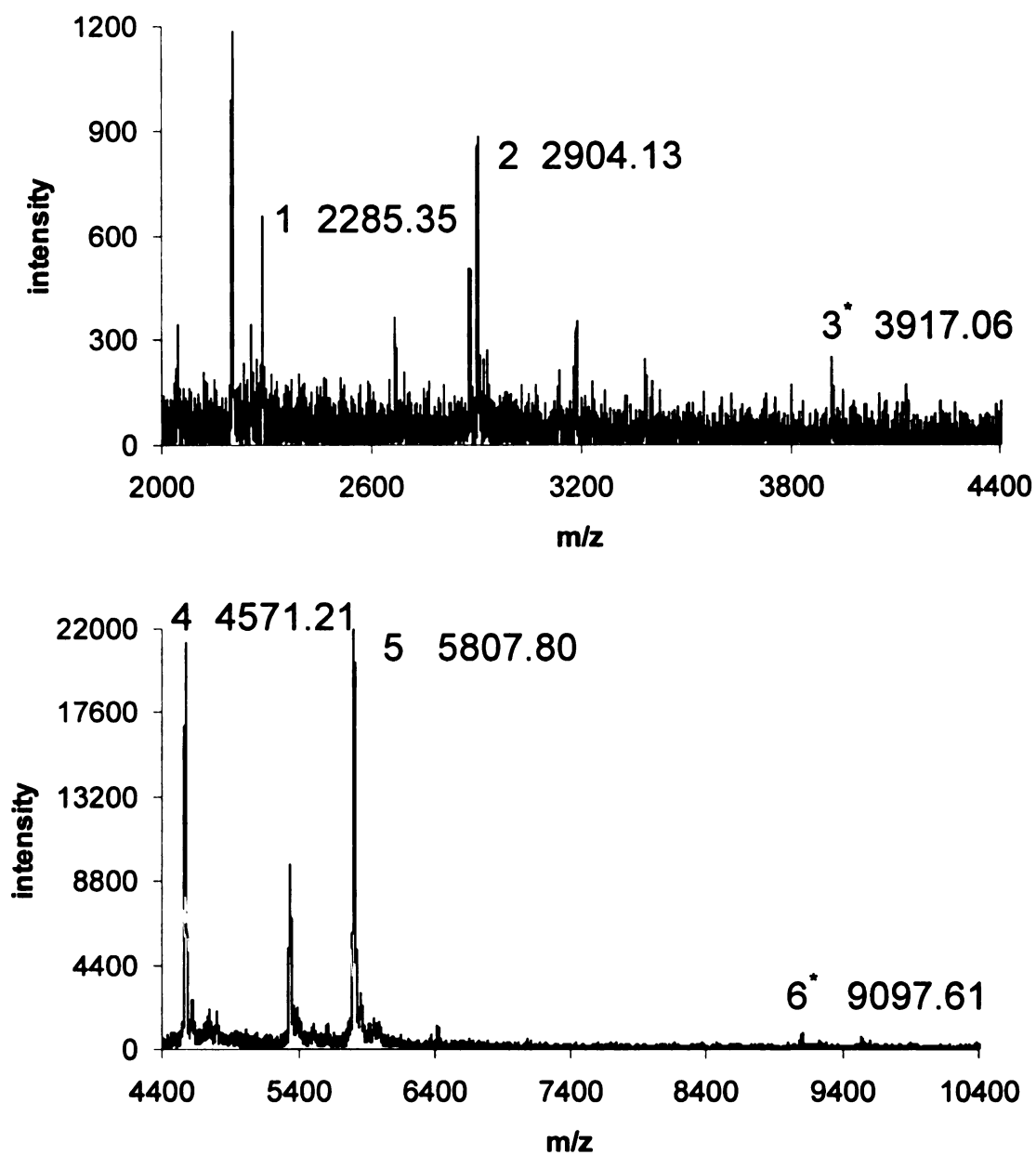
Figure 2. 3. MALDI mass spectrum of CN-induced cleavage products of recombinant spinach fructose bisphosphatase. The mass spectrum is divided into two mass ranges due to the intensity differences. The assignment of the labeled peaks is shown in Table 2. 2.

29

The signal to background ratio for the itz-153-189 peak (peak 2 in Table 2. 1) indicating the alternative disulfide bond between Cys173 and Cys178 in the MALDI spectrum for the pea leaf fructose bisphosphatase cleavage products was relatively low. This could be the result of suppression of the ionization of that particular CN-cleavage fragment by the other peptides during analysis by MALDI. Another possibility is that that fragment was less abundant in the mixture. To seek more convincing evidence, a similar mapping experiment was performed on recombinant spinach chloroplast fructose bisphosphatase, a homologue of the pea fructose bisphosphatase, using ESI-LC/MS (as described below) as well as MALDI-MS. In ESI-LC/MS, the cleavage fragments were separated on a C18 column prior to MS analysis, thereby reducing the chance of signal suppression between the individual components. ESI also has a better detection limit (tens of femtomoles) than MALDI. The ion trap mass spectrometer serves as the detector for the LC. It scans over the m/z range (m/z 200-2000) at a rate of ~0.7 scans/sec. Because the column was relatively short (5 cm), a long elution program (150 min) with a shallow gradient was used to achieve a better separation. A total of 5983 scans was collected during the chromatographic separation.

The MALDI spectrum of the CN-induced cleavage products of the recombinant spinach chloroplast fructose bisphosphatase was shown in Figure 2. 3. The peak assignments are detailed in Table 2. 2. The results are similar to those obtained from analysis of the pea leaf enzyme. Among the CN-induced cleavage fragments are protonated itz-155-190 (analogous to itz-153-189 described above for the pea leaf fructose bisphosphatase cleavage products), with a disulfide bond between Cys174-179

and a calculated mass of 3918.2 Da, and protonated itz-94-178, with a disulfide bond between Cys155-174 and a calculated mass of 9097.76 Da.

The ESI-LC/MS results are shown in Figure 2. 4. The upper frame is the mass chromatogram plotting the relative intensity of the base peak in the entire m/z range (m/z 200-2000) over the time of spectrum acquisition. The middle frame and the lower frame in Figure 2. 4 are mass chromatograms plotting the total ion current in a specified m/z range against the time of spectrum acquisition. The middle frame plots m/z 1305.8-1306.4 that corresponds to the triply charged itz-155-190 fragment with a disulfide bond between Cys174 and 179. The bottom frame plots m/z 1958.6-1959.1 that corresponds to the doubly charged itz-155-190 fragment. The fact that the two chromatographic peaks share the same retention time (51.3 min) suggests that the two peaks represent the same peptide with different degrees of protonation. Figure 2. 5 shows the mass spectrum of the CN-induced cleavage fragments eluted from the column in the chromatographic window indicated in Figure 2. 4. The major peaks correspond to the double and triply charged itz-155-190 fragment. These results indicate that there is a second disulfide bond in the recombinant spinach fructose bisphosphatase, and that bond joins Cys174 and Cys179, consistent with the results obtained from the disulfide mapping of the native pea chloroplast enzyme. Clearly, two alternate disulfide bonds are present in the oxidized form of the spinach enzyme, as well as in the pea chloroplast enzyme, as isolated. As with the spinach enzyme, fragments indicative of a Cys155-Cys179 disulfide bond, namely itz-94-173 and itz-174-190 linked by the disulfide bond (m/z 10344.1), and the prompt fragmentation products itz-94-173 (m/z 8585.1) and itz-174-190 (m/z 1762.0), were not found.
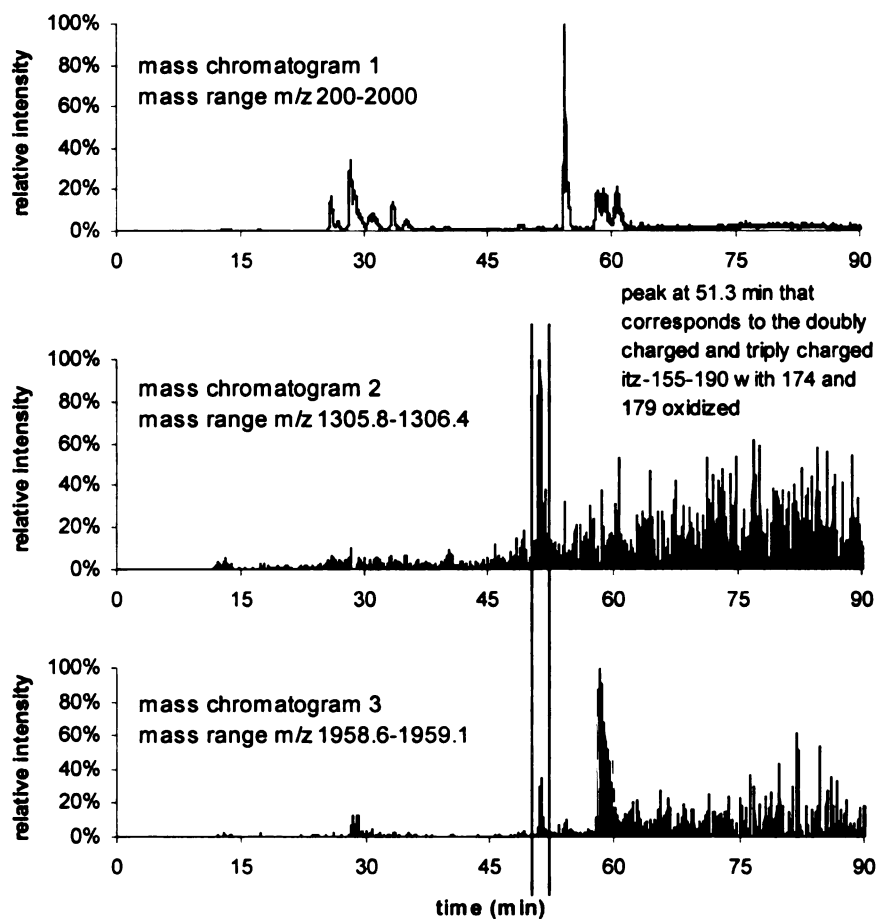
Figure 2. 4. Mass chromatograms corresponding to the analysis of the CN-induced cleavage products of the recombinant spinach fructose bisphosphatase by LC-ESI-MS. The top panel is a chromatogram reconstructed from the entire mass range (*m/z* 200-2000). The second panel is a mass chromatogram of ion current at *m/z* 1306 corresponding to the triply charged itz-155-190 fragment containing a disulfide bond between Cys174 and 179. The third panel is a mass chromatogram at *m/z* 1959 corresponding to the doubly protonated CN-induced cleavage fragment itz-155-190.
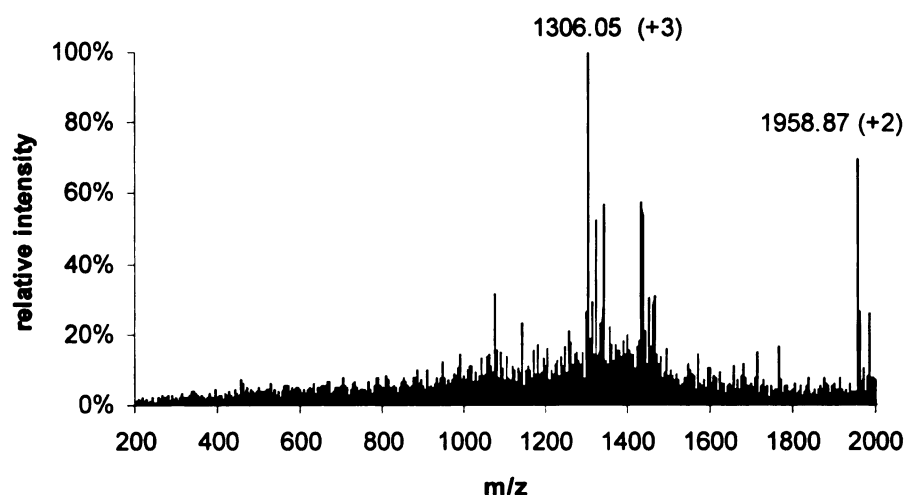
Figure 2. 5. An averaged spectrum (of 40 spectra) acquired during the analysis of the CN-induced cleavage products of the recombinant spinach fructose bisphosphatase by LC-ESI-MS. The two most intense peaks in the mass spectrum correspond to the doubly and triply charged itz-155-190 fragment, respectively, with a disulfide bond between Cys174 and Cys179.

Modeling

We used the coordinates of ring D of the spinach enzyme in PDB file 1spi (9) to estimate the energy values of the reduced enzyme and of the oxidized forms of the enzyme with the two alternate disulfide bonds, and with the 155-179 disulfide bond that was not detected in the wild type pea and spinach enzymes (Table 2. 2). All of these disulfide bonds are energetically possible according to these calculations, but the Cys155-Cys179 disulfide bond is energetically less favorable (by ~270 kcal/mol, enthalpy), and, therefore, it would not be predicted to be present in the predominant oxidized species. The positions of the three Cys residues in the modeled structures are shown in Figure 2. 6. In the crystal structure of the oxidized pea enzyme, the loop has contracted into a short

helix. In ring A of that structure, the Cα separation between Cys173 and Cys178 is 15.7

Å and the Sγ separation is 8.58 Å.

Form of Enzyme	Enthalpy			
	Final	Bond	Non-Bond	Coulombic
	(kcal/mol)			
Fully Reduced	2107.0	1004.3	632.2	-1767.7
155-174 cystine	1443.3	980.4	923.2	-2549.3
174-179 cystine	1439.7	979.3	889.7	-2550.2
155-179 cystine	1713.8	991.7	922.3	-2387.5

Table 2. 3. Summary of energy values of the oxidized Cys155-174, Cys174-179 and Cys155-179 forms of the spinach chloroplast fructose bisphosphatase, and of the reduced form, following energy minimization[a]

[a]The minimization protocol is described earlier in 'methods' section. Each series represents a total of 9000 to 12000 steps.
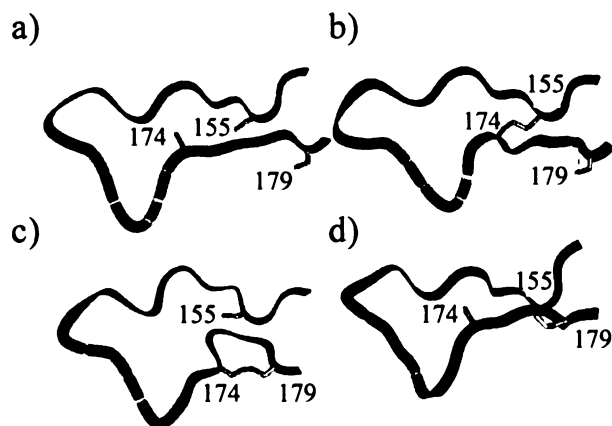
Figure 2. 6. Modeled structures of the loop containing the redox-sensitive cysteine residues in the reduced form of spinach chloroplast fructose bisphosphatase, in the two oxidized forms observed in these experiments ('b' and 'c'), and in the oxidized form containing the 155-179 disulfide bond ('d') that apparently can occur when Cys174 is replaced by mutation. In the oxidized form of the pea leaf enzyme in PDB file 1d9q (8), the loop has contracted into a short helix.

In these cyanylation-based disulfide mass mapping experiments, the protein is exposed to the reagent, 1-cyano-4-dimethylamino-pyridinium (CDAP), which cyanylates all the free cysteine residues in a protein, yet remains non-reactive to the disulfide bonds (cystines). The modified protein, under alkaline conditions, is cleaved specifically at the N-terminal side of the cyanylated cysteine residues, resulting in an iminothialzolidine-carboxyl blocked N-terminus and an amide C-terminus. The acidic condition of the cyanylation reaction effectively thwarts the disulfide scrambling reactions. Blocking of the free cysteines with a cyano group further prevents disulfide scrambling at the alkaline cleavage step. Scheme 2. 1 provides a structural correlation between the sites of the cysteines involved in the putative dual disulfide bonds in pea fructose bisphosphatase and the observed CN-induced cleavage fragments; in this case, deductive reasoning indicates that the native structure of pea fructose bisphosphatase contains a disulfide bond between

35

Cys153 and Cys173. Consideration of the fact that no free cysteines exist after cyanylation, coupled with the capability to differentiate between the mass of a cystine and that of two cysteines, infers the existence of a disulfide bond between the two internal cysteines in the CN-induced cleavage products marked with an asterisk. During the CN-based disulfide mass mapping of pea leaf fructose bisphosphatase, a peak corresponding to itz-92-177 with a disulfide bond between Cys153 and Cys173 was detected in the MALDI mass spectrum of the CN-induced cleavage reaction mixture. In addition, another fragment, itz-153-189 with a disulfide bond between Cys173 and Cys178, was observed. Based on this information, the dual disulfide structure in Scheme 2. 2 is proposed.

The results of analyzing the recombinant spinach enzyme by the CN-based disulfide mapping methodology indicated that this protein has a disulfide bond structure analogous to that of the pea enzyme. The observation of two CN-cleavage fragments, namely itz-94-178, with a disulfide bond between Cys155 and Cys174, and itz-155-190, with a disulfide bond between Cys174 and Cys179, provided direct evidence for the proposed disulfide structure of the recombinant spinach protein.

The Cys153-Cys173 disulfide bond is obvious in the protein crystals used for the determination of the structure of the oxidized enzyme (PDB files 1d9q and 1dcu (8)). The second disulfide bond, between Cys173 and Cys178, is not evident in those crystals. In ring D of the structure of the reduced form of the spinach enzyme (PDB file 1spi (9)) these residues are well positioned to form a disulfide bond. This disulfide bond was also suggested by Balmer *et al.* (7) on the basis of site-directed mutagenesis experiments. We do not find evidence for a third disulfide bond, joining Cys153-Cys178, suggested by Jacquot *et al.* (6), Balmer *et al.* (7), and Reichert *et al.* (5) on the basis of their mutagenesis experiments, nor do we find evidence for the disulfide bond in the interior of the protein between Cys49 and Cys190 suggested by Li *et al.* (18).

Notably, replacement of Cys153 with a serine results in the formation of an enzyme that is permanently active (6,7). This suggests that if the Cys173-Cys178 disulfide bond does form, it does not affect the activity of the enzyme. Balmer *et al.* (7) suggest that the Cys173-Cys178 disulfide bond is non-physiological. It formed slowly in the recombinant enzyme in which Cys153 was replaced by a serine. In the C153S mutant, this disulfide bond can only be formed by oxidation, but in the wildtype enzyme, with three Cys residues in the 170s loop, this bond could be formed by thiol-disulfide exchange.

SH SH   SH SH
| | | |
——49 —— 92 ——153 ——173 ——178 —— 190 ——306 ——
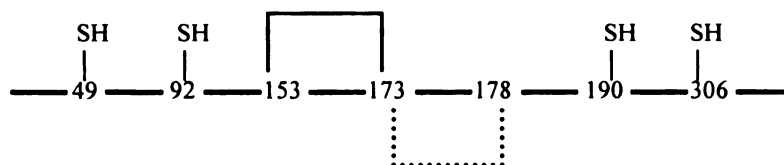
Scheme 2. 2. The proposed disulfide structure of the oxidized form of pea chloroplast fructose bisphosphatase. The protein is a mixture of two species, each with a single disulfide bond. In one species, as illustrated by the solid-line bracket, Cys153 and Cys173 form a disulfide bond. In the second species, as illustrated by the broken-line bracket, Cys173 and Cys178 form a disulfide bond.

It is reasonable to assume that the inactivating Cys153-173 disulfide bond is formed initially and that the enzyme containing this disulfide bond is the predominant species in the chloroplast in the dark. The 170s loop forms a very negative shield that will protect the Cys153-Cys173 disulfide bond from external attack by thiolates and decrease the possibility of the formation of mixed disulfides. Cys178 is buried on the other side of the Cys153-Cys173 disulfide bond. When dissociated to a thiolate, Cys178 should be able to attack the Cys153-Cys173 disulfide bond. The pH in the stroma rises about 1 unit when chloroplasts are illuminated. As the pH rises, the Cys178 sulfhydryl should begin to dissociate, thereby gaining the ability to attack any disulfide bonds in the vicinity. A pH-dependent shift from one disulfide form to the other could account for the difference in reductive activation seen when the enzyme is assayed in the presence of high levels of substrate and Mg^{++} at pH 7.8 (full activation) and at pH 8.8 (only about 20% activation) (19).

We cannot exclude the possibility that the Cys173-178 disulfide bond was formed during enzyme isolation, either by direct oxidation or by thiol-disulfide exchange. It is reasonable to assume that if the Cys173-178 disulfide bond is present in some of the

enzyme molecules in those preparations, that it will also be present in some of the enzyme molecules in the chloroplast in the green leaf after several hours of darkness.

There is no evidence for the presence of the Cys173-Cys178 disulfide bond in the crystal structure of the recombinant pea chloroplast enzyme in Protein Data Bank files 1d9q and 1dcu. The fully reduced form is absent. There are no disulfide bonds apparent in the structure of the reduced form of the spinach enzyme in Protein Data Bank file 1spi, and there is no indication in the description of the purification of the spinach enzyme (9) that it was protected from oxidation. Apparently, fructose bisphosphatase crystallizes out as the reduced form or as the oxidized form, and, in the case of the oxidized enzyme, the two isoforms do not co-crystallize. Neither crystallography nor mass spectrometry provide quantitative data, and therefore the relative amounts of the two disulfide-containing isomers *in vivo*, and in the enzyme as isolated, remain to be determined.

These experiments further illustrate the utility of cyanylation-induced cleavage/mass mapping in determining the location of protein disulfide bonds. Clearly, where alternate disulfide bonds are possible, a crystal structure showing only one of the possible disulfide bonds is not definitive.

2. Cobalamin-independent methionine synthase (MetE)

2.1. Introduction

Cobalamin-independent methionine synthase (MetE, E.C. 2.1.1.14) catalyzes the transfer of a methyl group from 5-methyltetrahydropteroylpolyglutamate (CH_3-$H_4PteGlu_n$, $n \geq 2$) to the thiol group of L-homocysteine (Hcy) to form tetrahydropteroylpolyglutamate ($H_4PteGlu_n$, $n \geq 2$) and L-methionine (1) in the terminal

step of methionine biosynthesis in *Escherichia coli*. The 753-residue enzyme contains seven cysteines, located at residues 323, 353, 516, 560, 643, 645 and 726. It has been established that the MetE enzyme contains one equiv of zinc, which is essential for the enzymatic activity, and that the homocysteine thiol group directly ligates to the active site zinc(20,21). MetE from *E. coli* has high sequence homology to methionine synthases from plants, fungi, and oether eubacteria. Sequence alignment shows that His641, Cys643, and Cys726 are the only conserved amino acid residues throughout the alignment. Binding assays and activity assays of the wild type and the four mutants (Cys726Ser, Cys643Ser, His641Gln, and His641Asn) suggest that these three residues bind with zinc in the native enzyme. It is hypothesized that the zinc in MetE also ligates with a molecule of water, which is replaced by the sulfur in homocysteine during the methyl transfer reaction (22).

$$CH_3 - H_4PteGlu_n +_L\_homocysteine(RSH) \rightarrow H_4PteGlu_n +_L\_methionine(RSCH_3) \quad (1)$$
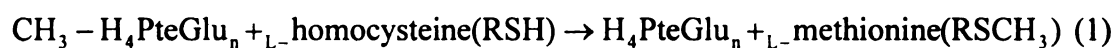
The activity of many proteins is regulated by the redox state in the environment. Zinc finger transcription factors such as members of the Sp1 family have been demonstrated to be redox-regulated both *in vitro* and *in vivo*(23). The zinc finger of replication protein A as well as the zinc ring finger protein SAG have been shown to be highly sensitive toward oxidation and reduction processes (24,25). Metallothionein, an important cystosolic zinc storage protein, has been shown to utilize disulfide bond formation to release its stored zinc upon oxidative stress (26,27). This allows the efficient transfer of zinc from the high-affinity zinc center of metallothionein to proteins with much lower zinc binding affinity. Reversible disulfide bond formation is a very

convenient way to translate environmental changes into changes in protein activity. Here, we investigate the disulfide structure of oxidized MetE using mass spectrometry and correlate the result to the observation of zinc loss during oxidation.

2.2. Materials and Methods

Enzyme isolation

The recombinant MetE was overexpressed by growing the appropriate *E. coli* strains aerobically in Luria-Bertani medium supplemented with 100 µg/mL ampicillin, 0.5 mM zinc sulfate, and 0.4 mM isopropyl β-D-thiogalactopyranoside (IPTG) as previously described (20). Most strains were grown at 37 °C to late log-phase or early stationary phase (A_{420} about 8.0). MetE was then purified to homogeneity by a single step DEAE-Sepharose ion-exchange chromatography using a linear gradient of potassium phosphate buffer (180-500 mM at pH 7.2 containing 500 µM DTT) as previously described (20). Protein purity was analyzed by electrophoresis in 12% polyacrylamide gels in the presence of sodium dodecyl sulfate (SDS-PAGE) and visualized by Coomassie blue staining. Purified proteins were concentrated to about 100 mg/mL and stored at -80 C.

Oxidation and Protein Estimate

Oxidized glutathione (GSSG) was added to the reduced MetE to make a final solution containing 70 µM of MetE, 10 mM of GSSG and 20 mM of tris-buffer (pH 7.2). The Protein/GSSG mixture was incubated at 37 C for 90 min. The GSSG was removed at the end of the incubation by running the mixture through a Bio-Rad 6 spin column that

41

had been equilibrated with 20mM tris buffer. Protein concentration was determined by the absorption coefficients at 280 nm (1.62 mg^{-1} cm^2).

Cyanylation (CN) and CN-induced cleavage and HPLC

For cyanylation of the reduced and oxidized MetE, cyanylation reagent, 20 μL of 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate solution (100 mM CDAP, 8 M urea, 100 mM sodium citrate at pH 3), was added to 50 μL of reduced or oxidized enzyme (70 μM, with 20 mM of tris-buffer). After 20 min at room temperature, the reaction was terminated by separation of the cyanylated protein from the reaction mixture on a Bio-Rad 6 spin column that had been pre-equilibrated with 8 M urea. For the alkaline cleavage reaction, 70 μL of 2 M NH$_4$OH was added to the cyanylated protein (70 μL, in 8 M urea) to make the final NH$_4$OH concentration 1 M. After 1 hour at room temperature, the cleavage reaction was terminated by removing the ammonia by vacuum aspiration. The cleavage mixture was dissolved in 0.1% TFA solution for separation. A Michrom BioResources Ultrafast Microprotein Analyzer (Auburn, CA) (microbore HPLC system) was used to separate the cleavage mixture. A gradient of 15% to 75% B (solvent A: 0.1% trifluoroacetic acid, 95% water, 5% acetonitrile; solvent B: 0.1% trifluoroacetic acid, 5% water and 95% acetonitrile) over 50 min was ramped at a flow rate of 50 μL/min through a Michrom microbore column (150-mm length, 1-mm i.d., packed with 8-u, 4000A polymeric stationary phase, part number 901/00711/00). UV detector was set at 214 nM. The manually collected fractions were lyophilized to dryness. Aqueous solvent (5 μL of 0.1% trifluoroacetic acid) was added to the dried sample for further MALDI analysis. For the fractions identified as with glutathionated peptides, a

42

total reduction was also performed by adding 2 μL of 100 mM TCEP to 1 μL of the reconstituted HPLC fraction.

Mass spectrometry

MALDI mass spectra were obtained on a Voyager DE-STR mass spectrometer (Applied Biosystems, Foster City, CA) equipped with a 337-nm nitrogen laser. The accelerating voltage in the ion source was set at 25 kV. A solution of saturated α-cyano-4-hydroxycinnamic acid (Aldrich Chemical Co.) in 1:2 of acetonitrile/water was diluted with isopropanol (1:4 v/v). The diluted solution (30μL) was applied to the flat MALDI probe using Cadene's protocol (13) to form an 'ultra-thin layer'. The matrix solution consisted of a saturated solution of α-cyano-4-hydroxycinnamic acid (Aldrich Chemical Co.) in 1:1 of acetonitrile/water. The reconstituted cleavage fragment solution (1 μL) from HPLC was mixed with 1 μL of matrix solution, and 0.5 μL of the peptide/matrix mixture was deposited onto the probe. Co-crystallization was usually observed after 1 min. External calibration was used during analysis of the intact enzymes and their CN-induced cleavage products by MALDI-MS.

A nanospray-ESI-MS infusion assay of the intact oxidized and reduced MetE was developed using a Micromass Q-TOF II (Micromass, Wythenshawe, UK.) mass spectrometer equipped with an orthogonal electrospray source (Z-spray). The intact protein solution (2.5 pmol/μL MetE in 5% formic acid, 50% methanol) was directly infused into the mass spectrometer at a flow rate of 0.1μL/min by a syringe pump through an empty nano-spray column (PicoFrit column from NewObjective, PF360-75-15-CE-5, 15-μm tip, 75-μm column i.d.). The source voltage was set at 2.0 kV and source

temperature was set at 160°C. The linear quadruple (Q1) was set to pass ions from m/z 500 to 2000 through into the pusher region of the time-of-flight (TOF) mass analyzer. MassLynx software was used to control the settings and view the spectrum. The final spectrum was an average of over 40 scans for both reduced and oxidized MetE.

2.3. Results and Discussion

Mass Measurement of the Intact Protein

We used ESI infusion to measure the intact enzyme, in both the reduced form and the oxidized form. Multiply protonated enzyme with number of charges up to 120 was observed. An envelop of peaks, corresponds to the same recombinant enzyme associated with different numbers of protons. The number of charges on the peaks in the series, from left to right, corresponds to a series of continuously decreasing integers. The m/z values of the two neighboring peaks can be determined from the following equations:

$$m/z_1 = \frac{M+n}{n} \quad (1)$$

$$m/z_2 = \frac{M+n+1}{n+1} \quad (2)$$

The two neighboring peaks are designated as m/z_1 and m/z_2 (m/z_1 being on the right) in the spectrum. M is the molecular weight of the protein and n is the number of charges on the peak on the right (m/z_1). The mass of the proton is 1 Da for illustrative purposes; in calculations in Table 2. 4, a mass of 1.0079 Da was used for hydrogen. The number of charges and the molecular weight can be solved from the previous two equations:

$$n = \frac{m/z_2 - 1}{m/z_1 - m/z_2} \quad (3)$$

$$M = n \times (m/z_1 - 1) \quad (4)$$

The number of charges (**n**) calculated from (3) is rounded to the closest integer, with a few exceptions when the knowledge of the charge series being a continuous integers forces **n** to the second closest integer. The molecular weight is calculated from equation (4) using the integer numbers of **n**. The expected mass of the reduced form of the enzyme is 84542.4 Da, while the observed mass for reduced MetE is 84548.5 Da, corresponding to residue 2 through residue 753 in the protein sequence from protein database (SwissProt, P25665). The transit pro-enzyme must be cleaved after, not before, the methionine in *E.coli* before the enzyme is expressed into the cytoplasm. To avoid confusion, we have retained the numbering consistent with that in the protein databank. From ESI-MS result, we also noticed that the mass difference between the reduced MetE and oxidized MetE (309.07 Da) corresponds well with the mass shift caused by a single glutathionation event (305.32 Da = reduced glutathione GSH, MW: 307.32 Da – 2H from disulfide bond formation).

S-Glutathionation is a post-translational modification event when a cysteine residue is covalently linked to a reduced glutathione molecule to form a mixed disulfide bond, usually as a cellular response to oxidative stress(28-30) . The enzymatic reduction of the mixed disulfide bond between GSH and protein (reverse reaction of glutathionation) is mediated by thioltransferases. The reversible modification of cysteines with GSH disulfide linkages is believed to serve as a protection mechanism under oxidative stress, i.e., it masks critical protein sulfhydryl groups from more damaging, irreversible oxidation.
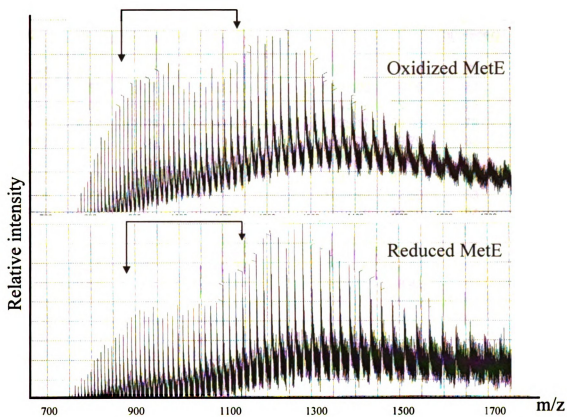
Figure 2. 7. ESI-MS spectra of the oxidized and reduced MetE. The clusters of peaks marked between two arrows are used to calculate the molecular weight (result shown in Table 2. 4).

ESI of reduced MetE				ESI of oxidized MetE			
m/z	n	n(integer)	M (Da)	m/z	n	n(integer)	M (Da)
1113.44	76.52	76	84544.84	1117.53	76.16	76	84855.68
1099.09	76.60	77	84552.32	1103.06	77.05	77	84858.01
1084.94	78.06	78	84546.70	1088.94	78.01	78	84858.70
1071.23	79.17	79	84547.55	1075.17	78.98	79	84858.81
1057.88	79.99	80	84549.77	1061.74	79.79	80	84858.57
1044.83	81.13	81	84549.59	1048.61	81.23	81	84855.77
1032.12	81.75	82	84551.19	1035.87	81.99	82	84858.69
1019.66	82.77	83	84548.12	1023.40	82.80	83	84858.54
1007.50	84.02	84	84545.34	1011.20	84.00	84	84856.14
995.66	85.01	85	84545.60	999.32	85.08	85	84856.19
984.10	86.15	86	84545.66	987.72	85.92	86	84857.07
972.82	87.05	87	84547.39	976.37	86.91	87	84856.15
961.78	88.61	88	84547.94	965.27	88.46	88	84855.15
951.06	88.21	89	84554.46	954.49	88.69	89	84860.08
940.41	90.08	90	84546.01	943.86	90.13	90	84856.78
930.09	91.58	91	84546.84	933.52	91.20	91	84858.15
920.06	91.78	92	84552.70	923.40	92.07	92	84860.17
910.15	92.80	93	84550.49	913.49	92.38	93	84860.84
900.46	93.98	94	84548.59	903.72	94.28	94	84854.75
890.99	95.09	95	84548.39	894.24	95.02	95	84857.43
881.73	95.87	96	84549.23	884.94	96.05	96	84857.58
872.64				875.83			
	average		84548.51		average		84857.58
	standard deviation		2.62		standard deviation		1.70

Table 2. 4. Molecular weight calculation based on the 22 peaks marked between the two arrows in the spectra of the oxidized MetE and the reduced MetE. Average mass and its standard deviation were calculated from the 21 independent calculations from two peaks that correspond to the enzyme with n protons and n+1 protons. (n from 76 to 95)

CN-based Disulfide Mass Mapping

The mixture of cleavage fragments was separated with microbore HPLC. In analysis of $MetE_{red}$, 16 fractions were collected, concentrated and subjected to MALDI-MS analysis. Eight fragments are expected as CN-induced cleavages occur at the seven cyanylated cysteines along the polypeptide chain (shown in Scheme 2. 3). Seven expected fragments were detected by mass spectrometry. The one that was not observed has only two amino acid residues. The ionization of that peptide was very likely being suppressed by the matrix ions, or a peak for a matrix ion interfered with observation of a peak for the cleavage fragment. However, information about the status of the two terminal cysteines that define this fragment, Cys643 and Cys645, can be readily retrieved from the observation of two other cleavage fragments, namely, itz-560-642 and itz-645-726. Observation of itz-560-642 indicates that Cyst560 and Cys643 were cyanylated and, thus, had been free in the original molecule of the reduced enzyme. Similarly, observation of itz-645-725 indicates Cys645 and Cys726 are free in the reduced MetE. The deduction of the cysteine status in the reduced MetE by CN-induced mass mapping is illustrated in Scheme 2. 3. We conclude from the mapping experiment that in the reduced MetE, all seven cysteines are free (not engaged in any kind of disulfide linkages).

Reduced MetE

2 — 322   itz323 — 352   itz353 — 515

itz516 — 559   itz560 — 642

SCN SCN SCN SCN SCN SCN SCN   Cleavage, NH₃
 | | | | | | |
— 323 — 353 —516— 560 — 643 —645—726 —   ⟶   itz643 — 644   itz645 — 725   itz726 — 753

Not Detected

SCN
|
itz645 —726—753

Incomplete Cleavage at Cys726

Scheme 2. 3. In the disulfide mass mapping of reduced MetE, seven out of eight expected fragments were detected. Detection of a CN-induced cleavage fragment, itz-645-753 with incomplete cleavage helps to determine the glutathionation sites between Cys645 and Cys726 in the oxidized form (see Scheme 2. 4).

In the analysis of oxidized MetE, sixteen chromatographic fractions where collected and analyzed by MALDI. Two key fragments: itz-643-753 with one glutathionation and one incomplete cleavage (observed mass: 12855.95, expected mass: 12855.48), and itz-643-753 with one glutathionation and one beta elimination (a side reaction during alkaline treatment in which HSCN is eliminated from the cyanylated cysteine to form a dehydroalanine) (observed mass: 12798.08, expected mass: 12796.48), were identified from mass mapping of the HPLC fraction marked with an asterisk in the lower panel in Figure 2. 8. Only a portion of each mass spectrum (m/z 11800-13800) is shown in Figure 2. 9 because a peak at m/z 9402.00 corresponding to the co-eluted itz-560-642 is about ten times more intense than the key peaks described above.

Beta-elimination reaction is a side reaction that competes with the cleavage reaction during alkaline incubation of the cyanylated peptide. The yield of beta-elimination varies depending on the primary sequence of the cyanylated peptide. Incomplete cleavage, a seemingly sequence-dependent event, was observed and reported

in the analysis of a small, highly knitted protein before. These side reactions, though causing the yield of the desired cleavage products to decrease, do not pose a great threat to the overall mapping strategy, because products from these two reactions have a different mass from the properly-cleaved fragments, and, thus, are discernable by mass spectrometry. In the oxidized MetE disulfide mass mapping, detection of these incomplete-cleavage products essentially provided useful information for disulfide structure elucidation.
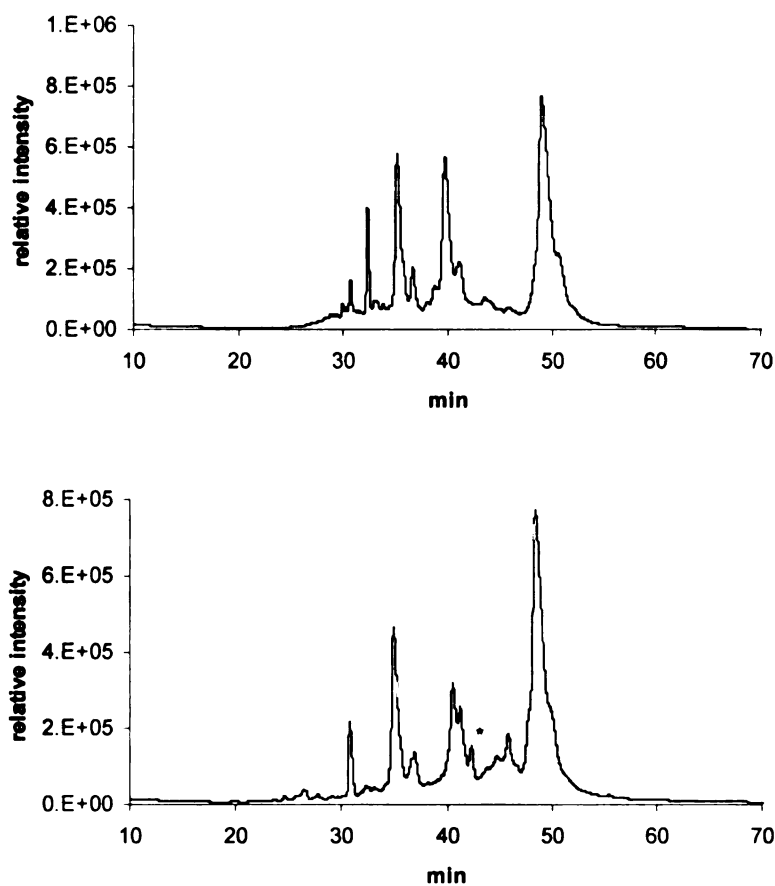
Figure 2. 8. Microbore HPLC separation of CN-induced cleavage fragments from cyanylated reduced (upper panel) and oxidized (lower panel) MetE. The fraction marked with an asterisk was identified as containing two variants of a glutathionated peptide and was subjected to total reduction and further mass spectral analysis.

In addition to direct mass identification, a complete reduction experiment was performed on the same HPLC fraction to validate the glutathionation, as shown in Scheme 2. 4. The loss of a glutathione molecule upon reduction of the mixed disulfide bond would cause a mass shift of –305 Da for each of the two variants of the itz-643-753 fragment described in the previous paragraph. In the analysis after reduction, as expected, the incomplete cleavage itz-643-753 fragment showed –303.27-Da mass shift, while the beta-elimination itz-643-753 fragment showed –305.90-Da mass shift.

Identity	Expected mass (Da)	Observed mass (Da)	Relative Error
2-322	36264.00	36243.91	-0.055%
itz-323-352	3430.93	3429.19	-0.051%
itz-353-515	18411.56	18411.70	0.001%
itz-516-559	4730.66	4729.95	-0.015%
itz-560-642	9400.59	9399.80	-0.008%
itz-643-644	310.34	n.d.	n.d.
itz-645-725	9207.30	9204.06	-0.035%
itz-726-753	3095.57	3095.96	0.013%
itz-645-753 with incomplete cleavage at 726	12283.85	12283.86	0.000%

Table 2. 5. Peak assignment of the fractions collected from HPLC separation of the reduced MetE protein.
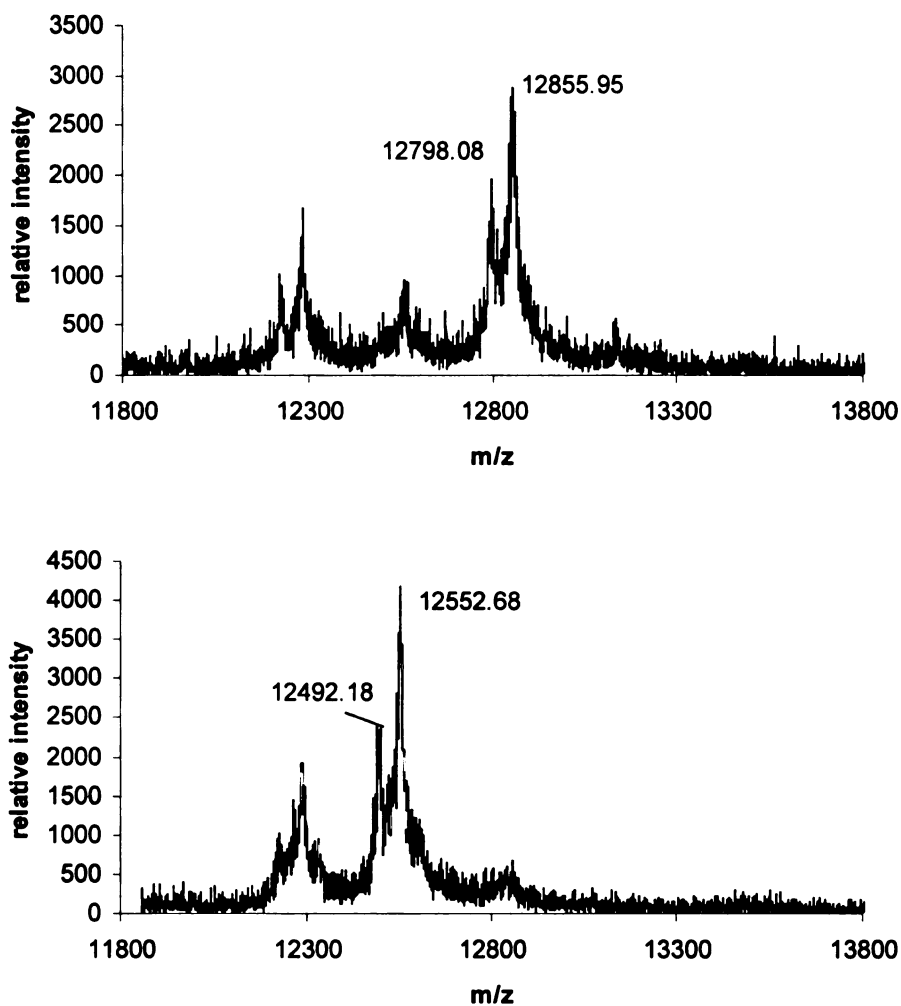
Figure 2. 9. Direct MALDI-MS analysis of one of the chromatographic fractions (marked with * in the lower panel of Figure 2. 8) consisting of mixture of CN-induced cleavage fragments from the CN-induced cleavage of oxidized MetE (upper panel). MALDI analysis of the same fraction after treatment with an excess of TCEP (lower panel).

Oxidized MetE

```
                                                              SSG    SCN
                                                               |      |
                                                        itz643—645——726 ——753

 SCN  SCN  SCN  SCN  SCN  SSG  SCN   Cleavage, NH₃
  |    |    |    |    |    |    |                              SSG   Beta Elimination
—323 —353 —516—560—643 —645—726—       ————————▶              |      |
                                                        itz643—645——726 ——753

                                                                |
                                                                |  Total Reduction with
                                                                |  TCEP
                                                                ▼

                                                               SH     SCN
                                                               |      |
                                                        itz643——645——726 ——753


                                                               SH    Beta Elimination
                                                               |      |
                                                        itz643——645——726 ——753
```
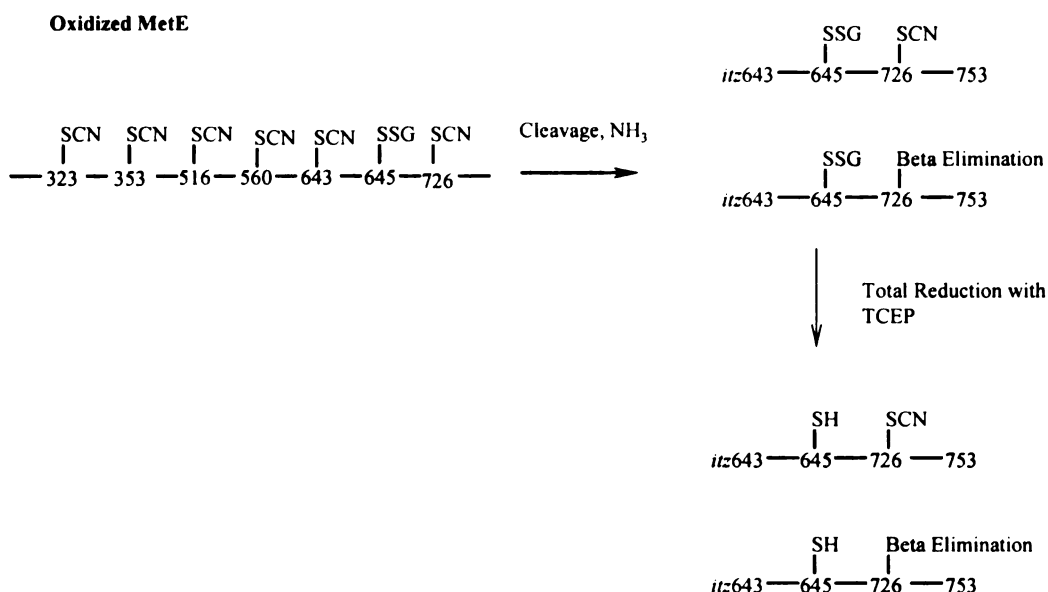
Scheme 2. 4. Detection of two key variants of a CN-induced cleavage fragment itz-643-753, limits the glutathionation site to Cys645 and Cys726. Total reduction experiment confirmed the glutathionation. The disulfide structure of the oxidized MetE was proposed considering the observation from the controlled experiment of the reduced MetE, in which incomplete cleavage at Cys726 was observed (Scheme 2. 3).

From the CN-induced cleavage mapping of the oxidized MetE, candidacy for the glutathionation site is reduced from seven cysteines to the two internal cysteines (Cys645 and Cys725) in fragment itz-643-753. In the mapping experiment on the reduced form of the enzyme, we observed fragment itz-645-753 with an incomplete cleavage at Cys726, but had not observed evidence for fragment itz-643-726 with an incomplete cleavage at Cys645, suggesting that Cys726 is more likely to suffer incomplete cleavage under alkaline condition than Cys645. This controlled experiment strongly suggests that fragment itz-645-753 (before complete reduction) has the glutathionation at Cys645 and the incomplete cleavage at Cys726. However, we do not have direct, conclusive experimental evidence to support this.

Mass spectral analysis of the intact oxidized MetE and reduced MetE suggests that there is one glutathionation event on one of the seven cysteine residues in the oxidized form of the protein. The CN-induced mass mapping experiment limits the single glutathione site to Cys645 or Cys726, but most likely Cys645. There is no experimental evidence to support the formation of a disulfide bond between Cys643 and Cys726, the two zinc binding sites, under oxidative stress. Zinc loss and corresponding loss of protein activity during oxidation may be caused by the glutathionation on Cys645 that disrupts the zinc binding pocket in its vicinity.

3. Conclusion

CN-induced disulfide mass mapping of both reduced MetE and oxidized MetE showed that CN-induced disulfide mass mapping methodology is not only applicable to small, tightly-folded proteins, but also to proteins as big as 84.5 kDa. Use of a high concentration of denaturant is critical to solubilize large, and usually more hydrophobic, proteins. Size exclusion-based spin column proves to be a fast and efficient way to separate the cyanylated proteins (MetE and FBPase) from the cyanylating reagent.

In the analysis of the intact proteins, ESI-MS showed great mass accuracy and could provide useful information as regards to the number of post-translational events (glutathionation) over the entire protein sequence. However, such analysis can not pinpoint the locus of such event. Controlled degradation of the protein is required to map the post-translational events. CN-induced disulfide mass mapping proved to be valuable in analysis of cysteine-related modifications, such as disulfide bond formation and glutathionation, in that all the fragments from specific cleavages at cysteines contain

information regarding to the cysteine status. Fragments from beta elimination and incomplete cleavage reactions can be used in some cases to elucidate disulfide structures.

Reference:

1. Chueca, A., Sahrawy, M., Pagano, E. A., and Gorge, J. L. (2002) *Photosynth Res* **74**, 235-249

2. Kobayashi, D., Tamoi, M., Iwaki, T., Shigeoka, S., and Wadano, A. (2003) *Plant Cell Physiol* **44**, 269-276

3. Anderson, L. E., Huppe, H. C., Li, A. D., and Stevens, F. J. (1996) *Plant J* **10**, 553-560

4. Qi, J. F., Isupov, M. N., Littlechild, J. A., and Anderson, L. E. (2001) *J Biol Chem* **276**, 35247-35252

5. Reichert, A., Dennes, A., Vetter, S., and Scheibe, R. (2003) *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1645**, 212-217

6. Jacquot, J. P., LopezJaramillo, J., MiginiacMaslow, M., Lemaire, S., Cherfils, J., Chueca, A., and LopezGorge, J. (1997) *FEBS Lett* **401**, 143-147

7. Balmer, Y., Stritt-Etter, A. L., Hirasawa, M., Jacquot, J. P., Keryer, E., Knaff, D. B., and Schurmann, P. (2001) *Biochemistry* **40**, 15444-15450

8. Chiadmi, M., Navaza, A., Miginiac-Maslow, M., Jacquot, J. P., and Cherfils, J. (1999) *Embo Journal* **18**, 6809-6815

9. Villeret, V., Huang, S. H., Zhang, Y. P., Xue, Y. F., and Lipscomb, W. N. (1995) *Biochemistry* **34**, 4299-4306

10. Kraulis, P. J. (1991) *Journal of Applied Crystallography* **24**, 946-950

11. Anderson, L. E., Goldhabergordon, I. M., Li, D., Tang, X. Y., Xiang, M. H., and Prakash, N. (1995) *Planta* **196**, 245-255

12. Bradford, M. M. (1976) *Anal Biochem* **72**, 248-254

13. Cadene, M., and Chait, B. T. (2000) *Anal Chem* **72**, 5655-5658

14. Stevens, F. J., Li, A. D., Lateef, S. S., and Anderson, L. E. (1997) *Photosynth Res* **54**, 185-197

15. Wu, J., and Watson, J. T. (1998) *Anal Biochem* **258**, 268-276

16. Qi, J. F., Wu, J., Somkuti, G. A., and Watson, J. T. (2001) *Biochemistry* **40**, 4531-4538

17. Patterson, S. D., and Katta, V. (1994) *Anal Chem* **66**, 3727-3732

18. Li, D., Stevens, F. J., Schiffer, M., and Anderson, L. E. (1994) *Biophys J* **67**, 29-35

19. Ashton, A. R. (1998) *Arch Biochem Biophys* **357**, 207-224

20. Gonzalez, J. C., Peariso, K., PennerHahn, J. E., and Matthews, R. G. (1996) *Biochemistry* **35**, 12228-12234

21. Peariso, K., Goulding, C. W., Huang, S., Matthews, R. G., and Penner-Hahn, J. E. (1998) *J Am Chem Soc* **120**, 8410-8416

22. Zhou, Z. H. S., Peariso, K., Penner-Hahn, J. E., and Matthews, R. G. (1999) *Biochemistry* **38**, 15915-15926

23. Wu, X. S., Bishopric, N. H., Discher, D. J., Murphy, B. J., and Webster, K. A. (1996) *Mol Cell Biol* **16**, 1035-1046

24. Park, J. S., Wang, M., Park, S. J., and Lee, S. H. (1999) *J Biol Chem* **274**, 29075-29080

25. Swaroop, M., Bian, J. H., Aviram, M., Duan, H. J., Bisgaier, C. L., Loo, J. A., and Sun, Y. (1999) *Free Radical Biology and Medicine* **27**, 193-202

26. Maret, W., and Vallee, B. L. (1998) *Proc Natl Acad Sci U S A* **95**, 3478-3482

27. Jakob, U., Eser, M., and Bardwell, J. C. A. (2000) *J Biol Chem* **275**, 38302-38310

28. Klatt, P., and Lamas, S. (2000) *Eur J Biochem* **267**, 4928-4944

29. Thomas, J. A., Chai, Y. C., and Jung, C. H. (1994) in *Oxygen Radicals in Biological Systems, Pt C* Vol. 233, pp. 385-395

30. Thomas, J. A., Poland, B., and Honzatko, R. (1995) *Arch Biochem Biophys* **319**, 1-9

CHAPTER 3

QUANTIFICATION OF CYSTEINE-CONTAINING PEPTIDES WITH

HOMOLOGOUS ALKYLATING REAGENTS

1. Introduction

Sensitive methods to quantify gene expression at the mRNA level have been developed and widely used to identify clusters of genes for which the transcription levels are linked at a specific state (1,2). However, these methods do not reflect directly the expression level of proteins, the ultimate biological effector molecules. Quantitative proteome analysis, especially one that monitors changes in protein expression upon perturbation, may provide insight into the mechanisms of disease and may lead to the discovery of novel therapeutically important proteins that new drugs can target.

Conventional protein quantification methodology is based on 2-D polyacrylamide gel electrophoresis (PAGE), followed by gel excision, in-gel digestion and LC/MS/MS to identify the proteins in a particular gel spot (3-5). The PAGE method is problematic when two proteins cannot be resolved fully from the labor intensive, hard-to-automate 2-D gel separation; furthermore, this method is biased toward highly abundant proteins, but against membrane proteins, very large or very small proteins, very acidic or very basic proteins, and most importantly, proteins with low expression levels, such as transcription factors and protein kinases.

New quantitative methods based on protein modification with amino acid-specific reagents, proteolytic digestion and mass spectrometry have been developed to quantify complex protein mixtures (6,7). Isotope-coded affinity tags (ICAT) reagents were

developed by Gygi *et al.* for such purposes (7). The structure of a representative ICAT reagent is shown in Figure 3. 1. Labeled and unlabeled forms of the reagent can be used to quantitatively compare the same proteins in two different samples. The reagent contains three moieties: a thiol-specific reactive group attaches the reagent to proteins via alkylation, a linker group that can be labeled with eight deuterons to distinguish the reagent pair, and a biotin group to enrich cysteine-containing peptides by affinity capture.

Heavy Reagent: d8-ICAT (X=deuterium)
Light Reagent:d0-ICAT(X=hydrogen)

Biotin   Linker (heavy or light)   Thiol-specific
                                                                      Reactive Group

Figure 3. 1. Structures of ICAT reagent pair used for protein quantification.

The ICAT quantification method includes three steps: (1) Use of the light (or heavy) reagent to label an aliquot of totally reduced protein mixture at state one; use of the other reagent to label an equal aliquot of totally reduced protein mixture at state two. (2) Mix the two aliquots and subject them to proteolytic digestion. (3) Use affinity chromatography based on the biotin-avidin interaction to capture the peptides that are modified with ICAT reagents and use mass spectrometry to analyze the captured cysteine-containing peptides. The protein concentration ratio at the two states is reflected by the relative intensity ratio of the mass spectral peaks corresponding to the differently labeled (d_0 vs. d_8) peptides of the same protein.

58

A similar approach, named MCAT (mass-coded abundance tag) (6), uses a lysine-specific reagent to modify the amino group of the lysine residues in the protein of the first state. In this approach, the proteins in the second state are not modified. Because modification at lysine resides will disable tryptic cleavage at such residues, the proteins of the two different states, unlike in the ICAT approach, are digested separately, prior to being modified. After the modification, the peptides are mixed for MS analysis. The two differently-modified peptides used for quantification have a mass difference of 42 Da, corresponding to the mass shift caused by the modification group.

In both ICAT and MCAT approaches, quantification of particular proteins is accomplished through quantification their differently modified proteolytic peptides. In this chapter, we developed a method that uses two brominated, structurally homologous reagents to quantify cysteinyl peptides. The two reagents used are 2-bromoacetamide (2-BA) and 2-bromopropionamide (2-BP), with their structures shown in Figure 3. 2. A peptide modified by 2-BA has a 57-Da mass difference from the original peptide with a free sulfhydryl group; A peptide modified by 2-BP has a 71-Da mass difference to the original peptide. The net difference of 14 Da between the two modified products arises from the overall elemental composition difference of a CH_2 between the two reagents.

R$_1$—CH—R$_2$
|
CH$_2$
|
SH
+

Cysteinyl Peptide

Br—CH$_2$—C(=O)—NH$_2$

2-BA

$\longrightarrow$

R$_1$—CH—R$_2$
|
CH$_2$
|
S—CH$_2$—C(=O)—NH$_2$

+57 Da

Br—CH(—CH$_3$)—C(=O)—NH$_2$

2-BP

$\longrightarrow$

R$_1$—CH—R$_2$
|
CH$_2$
|
S—CH(—CH$_3$)—C(=O)—NH$_2$

+71 Da
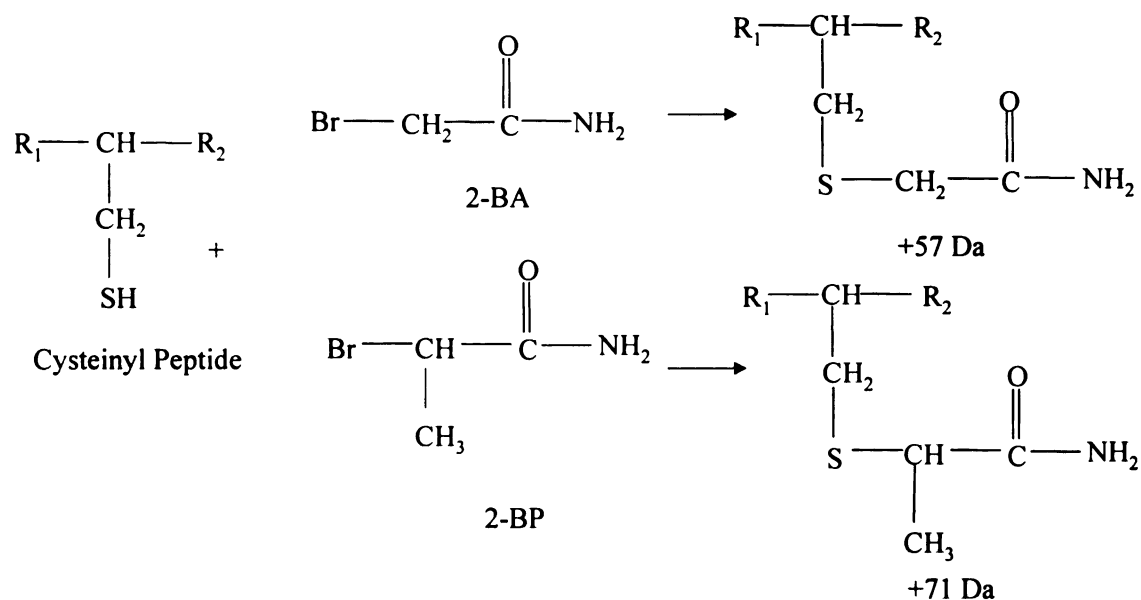
Figure 3. 2. The two reagents 2-BA and 2-BP modify the cysteine residues in a peptide/protein. The reagents have a CH$_2$ as the net difference in their composition, as reflected by a mass difference of 14 Da in their differentially modified products.

To quantify a cysteinyl peptide in two states, a different brominated homologous reagent is used to label the peptide in each state, as illustrated in Scheme 3. 1. 2-BA causes a +57 Da mass shift to the original peptide; 2-BP causes a mass shift of +71Da. After the modification, the two products are mixed together for MALDI-MS analysis. The relative concentration ratio of the two peptides is represented by the intensity ratio of mass spectral peaks representing the two products.

This analytical method, sharing the same working principle as those for ICAT and MCAT, provides a simplified system to explore the potential problems of such quantification processes and to evaluate the various assumptions under which ICAT and MCAT work. This method, successful in quantification at the cysteinyl peptide level, could be utilized to quantify protein complexes by adding a biotin-like affinity tag into the structure of the reagents. A modified method is proposed that could be used to

quantify proteins with different disulfide structures, such as the two oxidized FBPase isomers discussed in chapter two.
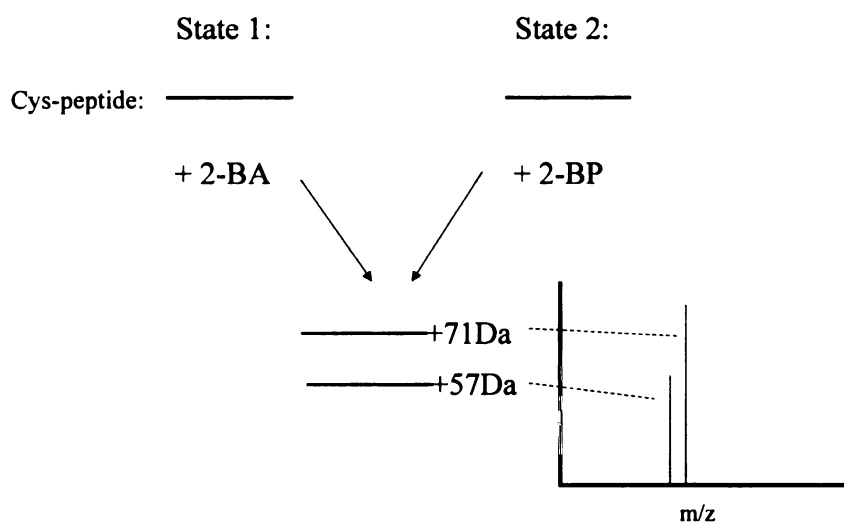
Scheme 3. 1. A cysteine-containing peptide from a protein at two different states is modified with 2-BA and 2-BP, respectively. The two modified peptides are then combined for analysis by MS. The relative intensity ratio of the two mass spectral peaks that correspond to the two products represents the quantity ratio of the peptide/protein in two different states.

2. Materials and Methods

Peptide Modification

Peptides (American Peptides Co., Sunnyvale, CA) were dissolved in aqueous buffer (100 mM tris, pH8.5, 10 mM dithiothreitol). 2-BA (100 mM dissolved in methanol) was added to the peptide solution with a molar ratio of 300: 1. 2-BP (700 mM dissolved in methanol) was added to the peptide solution with a molar ratio of 2000:1.

Alkylation with 2-BA is allowed to proceed at room temperature for 5 min, while 2-BP is allowed to proceed at room temperature for 30 min. The reaction was terminated with the addition of a quencher solution (saturated α-cyano-4-hydroxycinnamic acid in 50% aqueous acetonitrile, 1% acetic acid) to the reaction mixture (1: 1 v/v) after the designated period of time for reaction at room temperature. The pH of the final solution is ~3.5, which efficiently prevents the sulfhydryl groups from deprotonation. MS analysis of the peptide mixture showed that no additional alkylation occurs under such pH condition.

Mass Spectrometry

The MALDI target was coated with a thin layer of α-cyano-4-hydroxycinnamic acid matrix, as described by Cadene *et al.* (8). 0.5 µL of peptide/quencher mixture was directly deposited onto the target. The co-crystallization formed almost instantaneously. The residue liquid evaporated at ambient conditions in about 10 minutes. The target was inserted into a MALDI mass spectrometer (MALDI DE-STR, Applied Biosystems, Foster City, CA) with a reflector. All MS spectra were acquired in reflectron mode and averaged on 100 laser shots (9). Data Explorer software (Applied Biosystems, Foster City, CA) was used to view and analyze data. Information such as peak height and peak area was directly derived from the software.

Alkylation Reaction Yield Estimation

The yield of the desired singly modified peptide (Ys) can be represented by the peak height (monoisotopic) of the unmodified peptide (Hu), the peak height

(monoisotopic) of the singly-modified peptide (Hs) and sum of the peak height (monoisotopic) of the multiply modified peptide (Hm) in the following equation:

$$Ys = \frac{Hs}{Hu + Hs + Hm} \times 100\%$$

Similarly, the yield of the multiply modified peptide (Ym) can be represented by this equation:

$$Ym = \frac{Hm}{Hu + Hs + Hm} \times 100\%$$

3. Results and Discussion

3.1. Optimization of the Alkylation Reaction Conditions

The desired reaction should be specific to cysteine and it should go to completion. Other favorable characters involve a mild pH condition and relatively short incubation time as to minimize peptide decomposition. The reactivity of brominated alkylating reagents, such as 2-BP and 2-BA, is much lower than that of their iodinated counterparts due to the fact that bromide (Br⁻) is a poorer leaving group than iodide (I⁻). Brominated alkylating reagents were reported to work at a higher reagent-to-thiol ratio (400), a longer incubation time (2 hours), and a higher pH (9.0) (10,11). We evaluated the effect of pH on the reaction yield. Within three differently buffered conditions (pH 8.5, 9, and 9.5) studied, we found no significant difference in their effects on the yield. This result can be explained as the thiols are sufficiently deprotonated under those conditions. In later

experiments, pH 8.5 was used. Methanol was used as a solvent to dissolve reagents 2-BP and 2-BA, rather than DMF(*N,N*-dimethylformamide) as reported by Gardner *et al.* (10) for two reasons: First, 2-BA and 2-BP have a good solubility in methanol. Second and more importantly, methanol is more volatile than DMF. The low volatility of DMF was found to interfere with the co-crystallization process between matrices and peptides, resulting in very poor mass spectrometric signals for the peptides. Methanol does not seem to affect the co-crystallization or interfere with the ionization during MALDI, and, thus, the reaction mixture can be directly analyzed by MS without further purification.

Alkylation with a 2000-molar excess of 2-BA and 2-BP was used to modify the peptide HCKFWW (mono-isotopic MH$^+$: 906.4 Da) separately, under above-mentioned conditions for 2 hours. Then, the reaction mixtures were combined and directly analyzed by MALDI-TOF-MS. The 2-BP modification proceeded as expected: the desired mono-alkylation (with +71-Da mass shift) was the predominant product, as shown in the spectrum in the top panel of Figure 3. 3. However, for 2-BA modification under the same condidtion, products with one to five alkylation events (with +57×n-Da mass shift, n: 1~5) were observed. Besides the cysteine residue, the side chains of histidine and tryptophan, as well as the N-terminal amino group, were modified by 2-BA at the same time.

Over-alkylation with iodoacetamide was reported to occur more preferentially at N-terminal NH$_2$ of the peptide and at the side chain of the histidine residue than at the side chain of lysine at relatively low pH (pH 7-7.5), sometimes, even than alkylation at the thiol of cysteine. At pH 7-8.5, alkylation at thiol occurs much faster, yet competition

from lysine side chain increases at the same time (12). Over-alkylation with iodoacetamide at methionine residues was also reported by Lapko *et al.* (13).

Experiments with peptide HCKFWW showed that over-alkylation with 2-BA can occur on tryptophan residues as well under certain conditions. It is reasonable to believe that iodoacetamide, a stronger alkylating reagent than 2-BA, could modify the side chain of tryptophan under similar conditions.

Stoichiometric modification of cysteine residues is very important for both the ICAT method and the brominated homologous reagent method discussed in this chapter, because only peaks representing cysteine-modification products are used to represent the peptide concentration. If over-alkylation happened, we would have underestimated the quantity of the peptide modified by 2-BA by using the intensity ratio of the mono-alkylated peaks.

Alkylation of Peptide HCKFWW (MW:905.4)
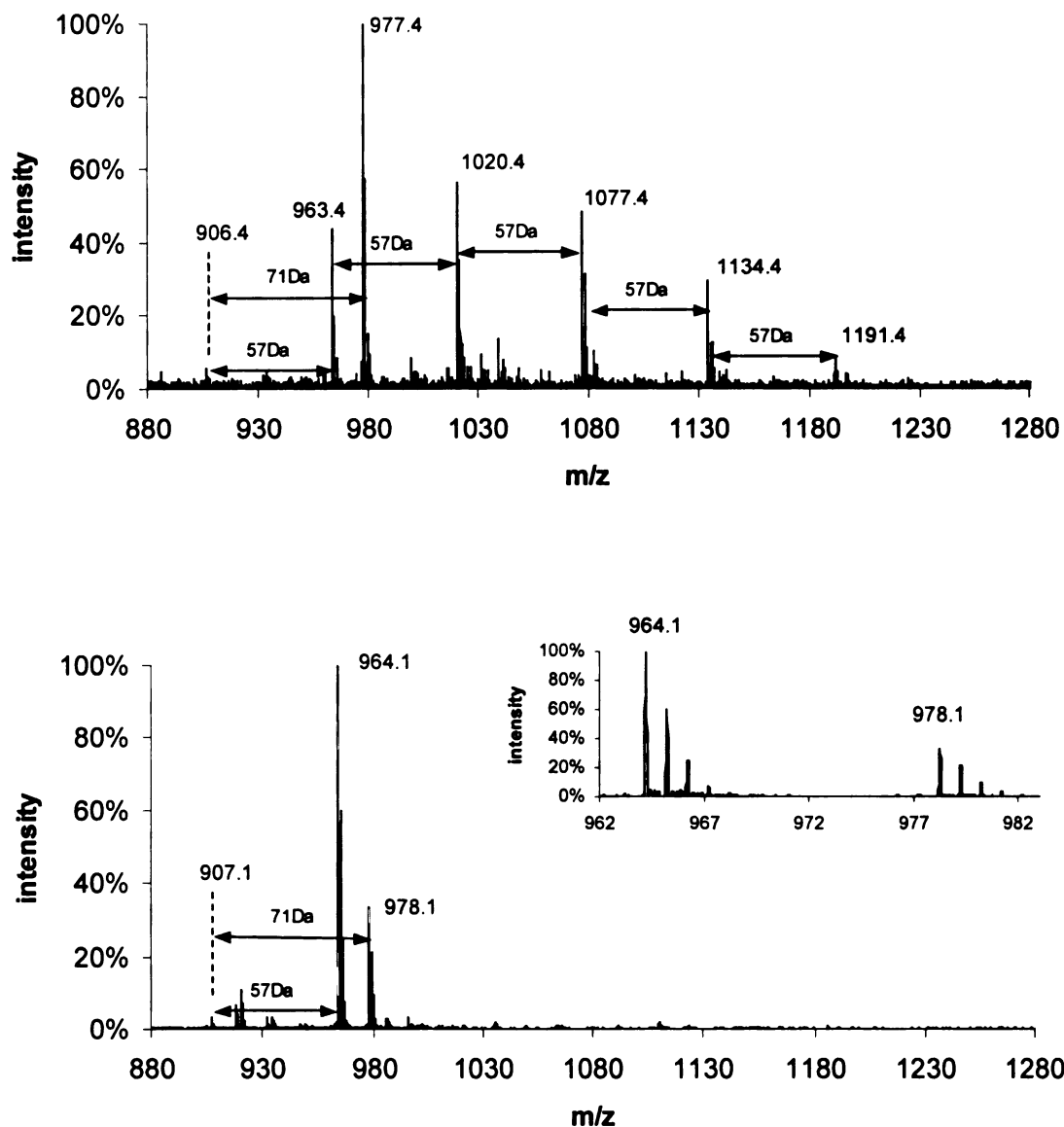
Figure 3. 3. MALDI-TOF mass spectra of the 2-BA/2-BP modification reaction (peptide HCKFWW) mixture. Before optimization (upper panel), up to five modifications by 2-BA were observed for the peptide. There was no obvious over-alkylation from the 2-BP modification under that condition. After optimization (lower panel), the desired products were the predominant species in the spectrum. No over-alkylation was observed. The insert is a 'zoom-in' of the m/z range around the peaks for the desired products. [13]C isotopic peaks were well resolved.

A time-course study was performed for both 2-BA and 2-BP modification of the same peptide HCKFWW to explore the optimal condition for alkylation so that only cysteine residues in a peptide are stoichiometrically modified. Both alkylation reactions proceeded with a reagent-to-peptide ratio of 2000. Aliquots of the reaction mixture were removed and quenched by mixing with acidic matrix solutions at different time intervals. Analysis of the reaction mixture by MALDI-MS was used to assess the reaction yield at a particular time point with the assumption that the intact peptide and its differently-alkylated forms have the same ionization efficiency. Under this condition, 2-BA modification and 2-BP modification took a vastly different pace, as shown in Figure 3. 4. For 2-BP, the desired mono-alkylation reaction proceeded slowly to near-completion after 30 minutes of incubation. A noticeable level of under-alkylation was perceptible within 5 min, yet no significant over-alkylation was observed even after an hour of incubation. For 2-BA modification of HCKFWW, the desired mono-alkylation was achieved almost instantaneously, however, the over-alkylation began to be more significant after 10 min.

Time Course of 2-BP Modification of HCKFWW

Time Course of 2-BA Modification of HCKFWW
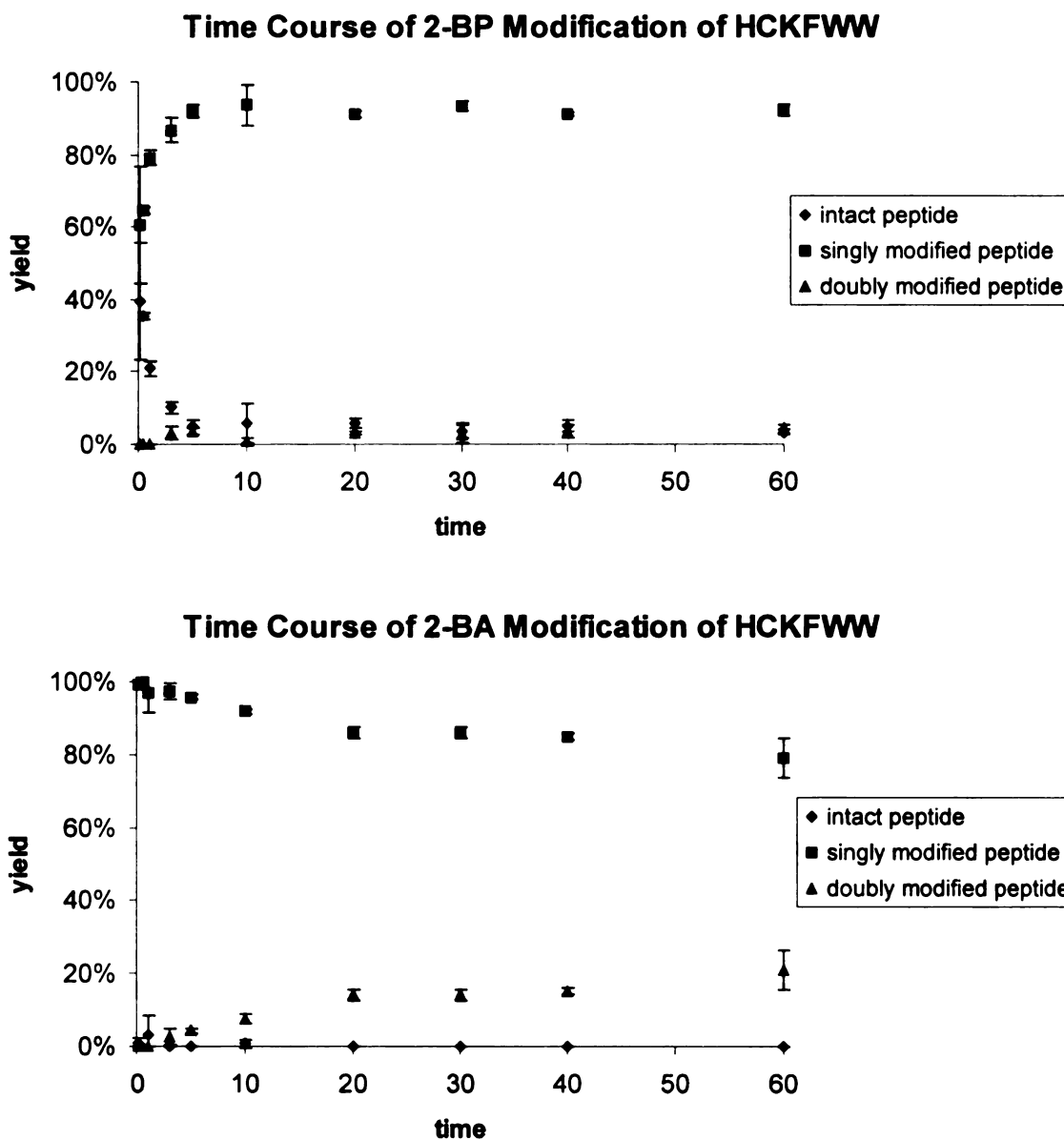
Figure 3. 4. Time course of the two modification reactions of peptide HCKFWW. For 2-BP modification (upper panel), the desired singly modified peptide yield increases relatively more slowly. After 30 minutes, the yield is over 95%. No over-alkylation was observed. For 2-BA modification, the desired singly modified product yield reached 95% within 5 minutes. The over-alkylation becomes significant after 10 minutes.

From the time course analysis, we recognized that different reaction conditions were needed for 2-BA modification and 2-BP modification to achieve stoichiometric conversion to mono-alkylated products. After more experiments, we determined that for 2-BA modification, a 300-fold molar excess of reagent and a 5-minute of incubation time is appropriate to achieve over 95% conversion to the desired product; for 2-BP modification, a 2000-fold molar excess of reagent and a 30-minutes incubation time is sufficient to achieve >95% yield. This modified condition was used to analyze the same peptide HCKFWW at two states with a concentration ratio of 5:2. We used 2-BA to modify the first state and 2-BP to modify the second state. Then, we combined the peptides in the two states for analysis by MALDI-TOF-MS. In the spectrum shown in the lower panel of Figure 3. 3, the problem of over-alkylation for 2-BA was minimized, and the peak intensity ratio closely represented the peptide concentration ratio.

The overall mass difference of a 'CH$_2$' in the two modified peptide products is quite small, compared with the average size of a tryptic peptide (600-1500 Da). However, the net structural difference of a CH$_2$ has a relatively big effect on the reactivity difference between a 2-carbon reagent (2-BA) and a 3-carbon reagent (2-BP). The reactivity difference, more apparent in structural homologous reagents, also exists between isotope-labeled isomeric reagents, but at a much lower level, such as for the ICAT reagent pair. For example, the differently modified peptides by ICAT reagents showed noticeable difference in retention times during HPLC analysis, suggesting the

two ICAT reagents have different hydrophobicity. Besides reactivity difference, another important factor that affects the stoichiometric modification of a specific amino acid residue is the reagent-to-substrate ratio used during modification reactions. A low ratio may cause under-modification, while a high ratio may cause over-modification. As concentrations of peptides/proteins are unknown in these applications, it is impossible to maintain the same reagent-to-substrate ratio for both states (unless the peptide/protein level at two different states is similar). Perturbation (disease)-caused changes in protein expression level may accompany other changes, such as in pH, redox potential, etc. Care should be taken during modification of proteins/peptides in the perturbed state and in the normal state so that these changes would not affect the stoichiometry of the reaction.

Results of the studies described above show that 2-BA is much more reactive than 2-BP, yet the desired result of complete mono-alkylation for both reactions can be achieved by adjusting the reagent-to-thiol ratio as well as incubation time in each case.

3.2. Quantification Analysis of Cysteine-Containing Peptides

Two peptides, PHCKRM (expected MH^+: m/z 770.4 Da) and DRVYIHPCHLL-YYS (expected MH^+: m/z 1777.9 Da) were combined at a 1: 3 molar ratio. An aliquot of the peptide mixture was modified with 2-BA, and an equal aliquot was modified with 2-BP, with aforementioned protocols. The two differently modified aliquots were combined and analyzed by MALDI-MS. For peptide PHCKRM, the peak intensity ratio of the two differently-modified products (2-BA modified product with a calculated monoisotopic mass for MH^+ = 827.37 Da and 2-BP modified product with a calculated monoisotopic

mass for MH$^+$ = m/z 841.37 Da) was 1:1, as expected (shown in Figure 3. 5). MS peaks

representing two modification products of peptide DRVYIHPCHLLYYS (2-BA modified

product with a calculated monoisotopic mass for MH$^+$ = m/z 1834.9 Da and 2-BP

modified product with a calculated monoisotopic mass for MH$^+$ = m/z 1848.9 Da) had an

expected intensity ratio of 1:1 as well (shown in Figure 3. 5). Though the intensities of

the two peaks representing two peptides differ by a factor of 10, the intensity ratio (1:1)

between the differently-alkylated products correctly represented the concentration ratio in

the two aliquots (1:1) in both peptides. Note that the two peaks near m/z 1900 in the mass

spectrum of DRVYIHPCHLLYYS alkylation products (marked with an asterisk in Figure

3. 5) do not represent over-alkylation products. These two peaks have a mass shift of

+119-Da and +133-Da mass shift from that of the protonated intact peptide, respectively.

As the intensity ratio of these two peaks is 1:1 as well, these two peaks are most likely

the 2-BA and 2-BP alkylation products of a modified peptide that has a +62-Da mass

shift from the intact peptide DRVYIHPCHLLYYS. The unknown +62-Da mass shift

may arise from single or multiple modification events, possibly due to prolonged

exposure in air at room temperature.

71
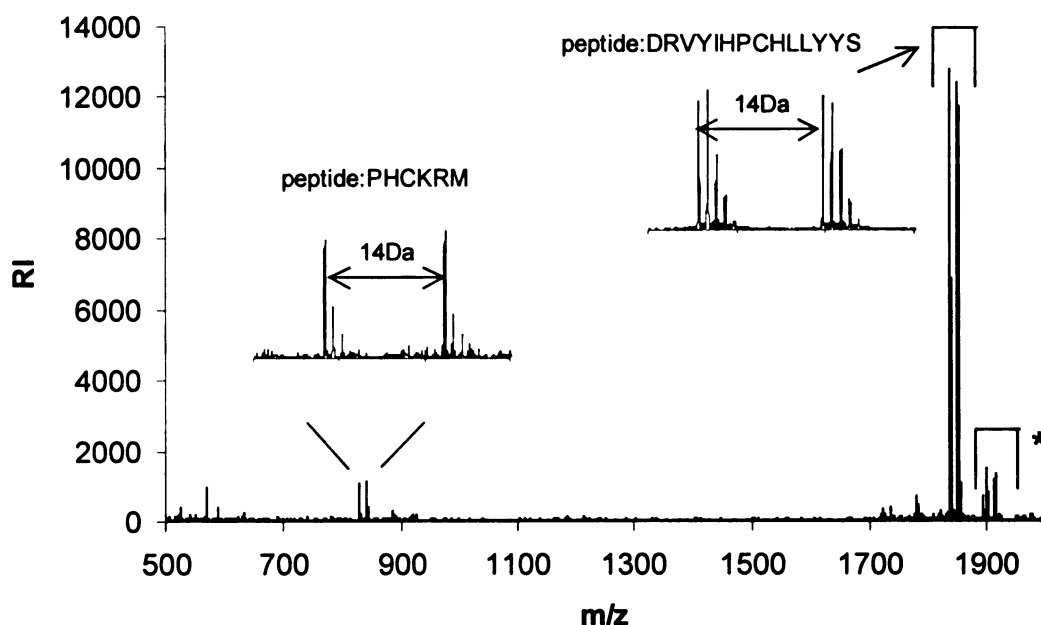
Figure 3. 5. Mass spectrum of 2-BA and 2-BP modified peptides (PHCKRM and DRVYIHPCHLLYYS) at two different states (concentration ratio 1:1).

Peptide CQDSETRTFY (calculated monoisotopic mass of MH^+ = 1249.5 Da) was subjected to modification with 2-BA and 2-BP with the above-mentioned protocol to assess the dynamic range and standard deviation of this method. After quenching of the reaction, the two modified products were mixed to create nine different concentration ratios at 10:1, 5:1, 5:2, 5:3, 1:1, 3:5, 2:5, 1:5 and 1:10. The nine mixtures were analyzed by MALDI-TOF-MS. In Figure 3. 6, the peak intensity ratios corresponded closely to the peptide concentration ratios in each of the standard mixtures of preformed alkylated peptides.
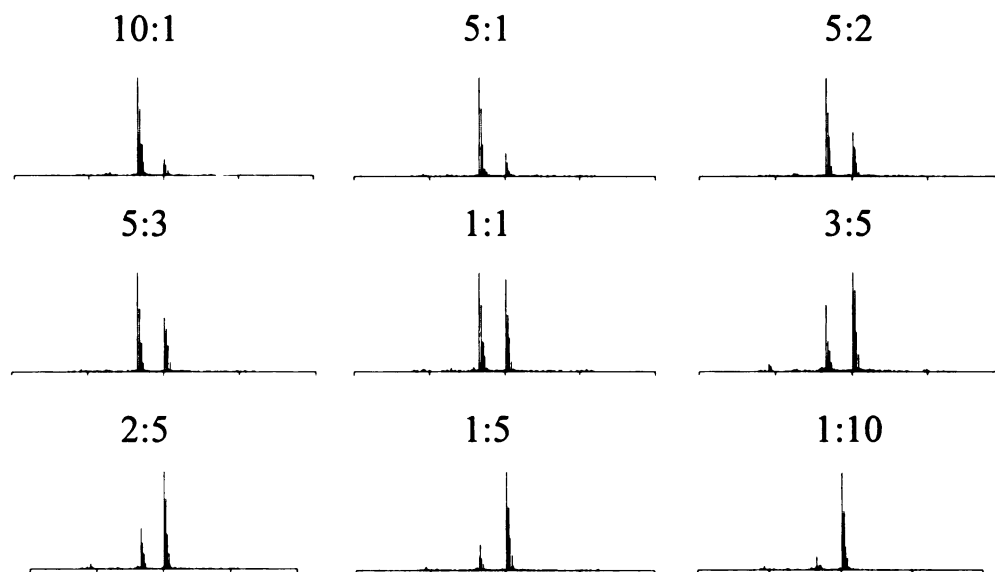
Figure 3. 6. MALDI mass spectra from quantification analysis of a series of standard solutions containing various ratios (10:1, 5:1, 5:2, 5:3, 1:1, 3:5, 2:5, 1:5 and 1:10) of preformed alkylation products (2-BA and 2-BP) of peptide CQDSETRTFY. Each spectrum was averaged over 100 laser shots. The mono-isotopic peaks for the two modified products have masses of 1306.4 Da (2-BA) and 1320.4 Da (2-BP).

In the 'thin layer' MALDI sample preparation method, the peptide and matrix co-crystallize almost instantaneously (<10 seconds) after being spotted on to the probe. As the co-crystallization is homogeneous, laser shots at any position within the sample well generate a good and reproducible signal. For the experiments described here, we randomly chose 5 spots to analyze. The spectra acquired at different spots within one sample well have different base peak intensity and signal to noise ratio for the base peak. There was a 5-fold difference between the highest value to the lowest value for both peak intensity and signal to noise ratio in different spectra acquired across the sample well. However, we noticed that the deviation of the individual peak intensity ratio of the differentially modified peptides from the mean does not correlate to either peak intensity or signal-to-noise ratio in that spectrum. With repetitive measurements, a standard

deviation can be measured (results in shown in Table 3. 1). The intensity ratio of the two differently-modified products is plotted against the concentration ratio of peptides in the standard mixtures, as shown in Figure 3. 7. A standard curve was generated from linear regression. We only graphed the dynamic range from 1:5 to 5:1 in Figure 3. 7, as 1:10 and 10:1 points have relatively big deviations and standard deviations.

concentration ratio (BA/BP)	intensity ratio experimental (BA/BP)	standard deviation (experimental)	relative standard deviation	deviation
0.10	0.14	0.016	11.87%	38.41%
0.20	0.23	0.015	6.57%	17.46%
0.40	0.43	0.036	8.50%	6.85%
0.60	0.62	0.029	4.69%	3.33%
1.00	1.07	0.080	7.50%	6.65%
1.67	1.76	0.032	1.82%	5.43%
2.50	2.53	0.251	9.91%	1.30%
5.00	4.86	0.413	8.49%	-2.79%
10.00	8.12	0.584	7.20%	-18.84%

Table 3. 1. Precision and accuracy of the assay of peptide CQDSETRTFY, over a dynamic range between 1:10 and 10:1. (5 measurements)
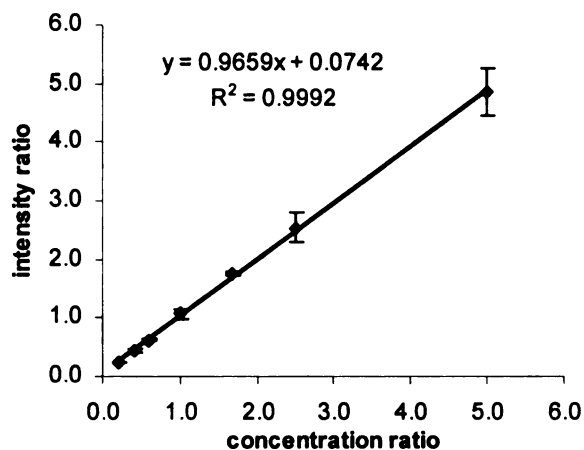
Figure 3. 7. Standard curve of quantification of peptide CQDSETRTFY, over a dynamic range between 1:5 and 5:1 (5 measurements).

The same analysis was performed on two other peptides: (HCKFWW and DRVYIHPCHLLYYS). Two standard curves plotting the peak intensity ratio vs. the concentration ratio are shown in Figure 3. 8. All three curves showed good linearity, revealing that peak intensity ratio from analysis by MALDI-MS can be used to represent the concentration ratio of the modified peptides in standard solutions. However, as two reactions may have different yields for particular peptides, the concentration ratio of the modified peptides does not necessarily reflect the concentration ratio of the intact peptides at two states accurately. The degree of deviation of the slope of the standard curve from one reflects the degree of yield difference from the two modification reactions.
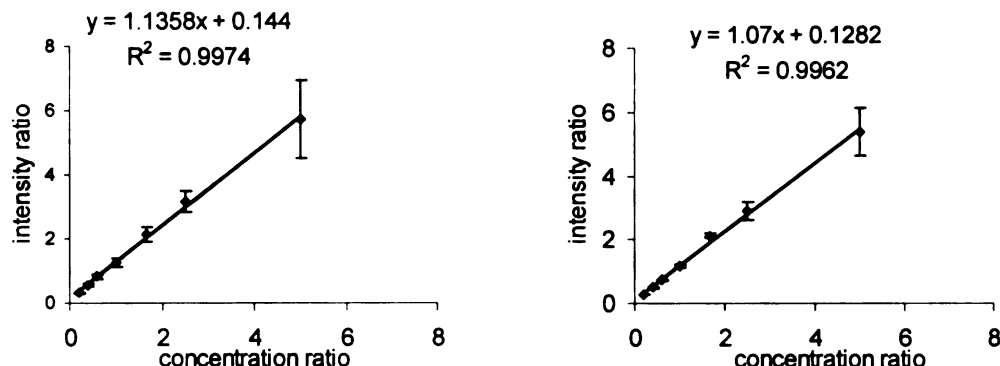
Figure 3. 8. Standard curve for quantification of peptide HCKFWW and peptide DRVYIHPCHLLYYS, over a dynamic range between 1:5 and 5:1. The ratio is the intensity of the peak for the 2-BA modified product over that for the 2-BP modified product.

3.3. Estimation of the Minimal Error and Standard Deviation

We used the ^{13}C isotope peak intensity distribution for a 2-BA modified peptide (DRVYIHPCHLLYYS) to estimate the minimal error and standard deviation of this quantification method. In this analysis, we focus on the ratio of peak height for the monoisotopic form of the peptide and that of the same peptide with n ^{13}C carbons (n = 1~4). The concentration (abundance) ratio of the mono-isotopic species and those with n ^{13}C carbons should be a constant, and can be estimated from the natural abundance of ^{13}C and the number carbons in the peptide. Because there is no modification reaction involved in this estimation and because ionization efficiency difference is minimal between these species, the error and standard deviation from this measurement should

represent a best case that we can accomplish when performing homologous alkylation-based quantification.

Three spectra from the quantification analysis, with known ratio of 5:3, 5:2 and 5:1 (ratio of 2-BA product to 2-BP product) of peptides at two different states, were shown in Figure 3. 9 to demonstrate the deviation of the expected isotopic ratio and the measured isotopic ratios. We shall focus on the isotopic peak intensity ratio among peaks representing a peptide, but not focus on the intensity ratio between the two peaks that represent the two differently-modified peptides. All spectra were acquired by averaging the accumulated ion current from 100 laser shots. The insert spectrum is a computer-generated spectrum with the isotopic distribution calculated from the elemental composition ($C_{85}H_{122}N_{22}O_{22}S$) of the 2-BA modified peptide DRVYIHPCHLLYYS. The isotopic distribution of 2-BP modified peptide is not shown here. However, it should be very similar to that of the 2-BA modified peptide, as the contribution from a 'CH$_2$' to the isotopic distribution of all other atoms is minimal. In Figure 3. 9, the isotopic cluster (a) has a similar distribution to the one in the insert spectrum. Three peak clusters labeled (b) vary from the expected distribution in that the intensity of the monoisotopic peak exceeds that of the peptide with one ^{13}C. Peak cluster (c) shows a big deviation for the 2-^{13}C peptide.

Accuracy and precision was assessed by comparing the experimental isotopic distribution of the 2-BA modified peptide from ten measurements with the calculated isotopic distribution (results shown in Table 3. 2). The concentration ratio of mono-isotopic and multiple-^{13}C should be a constant across the spot on the MALDI plate. The error and standard deviation are large because the relatively small number of ions

produced in each laser shot lead to poor ion-counting statistics and the isotopic ratio in this small population may not represent the bulk isotopic ratio very well. Whatever the reason might be, this error and standard deviation from isotopic ratio analysis represents a best case that can be achieved for homologous quantification analysis.

	expected ratio	experimental ratio	standard deviation	relative standard deviation	deviation
C1/C0	1.067	0.982	0.061	6.25%	-8.02%
C2/C0	0.653	0.608	0.035	5.75%	-6.86%
C3/C0	0.291	0.275	0.025	9.18%	-5.68%
C4/C0	0.104	0.118	0.010	8.21%	13.50%

Table 3. 2. Precision and accuracy of the experimental ratio of the intensity of the monoisotopic peak vs. that of the peaks with 1~4 (^{13}C) atoms in the 2-BA modified peptide. C0 represents the monoisotopic peak while C1~4 represents the peptide with 1~4 ^{13}C atoms. The expected ratio was calculated from the elemental composition of the product. (10 measurements)
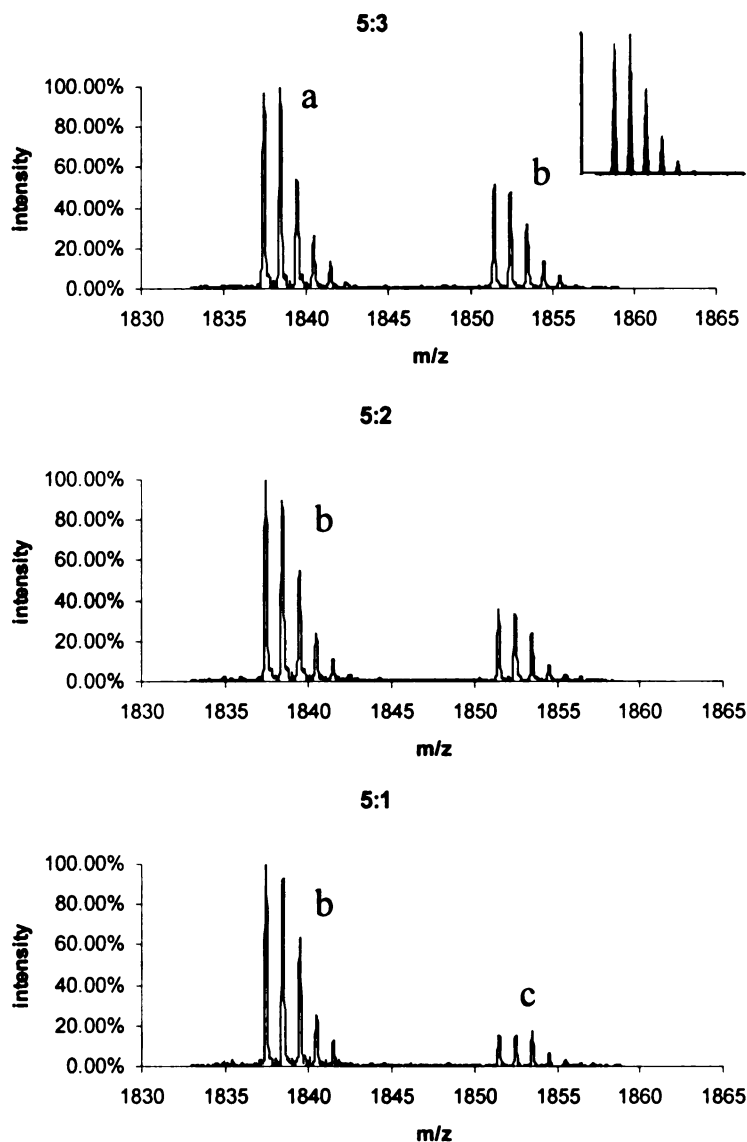
Figure 3. 9. Isotopic distribution variance of the product peaks after 2-BA and 2-BP modification of peptide DRVYIHPCHLLYYS in three experiments (5:3, 5:2 and 5:1). The insert is a computer-generated distribution representing the isotopic variants for the elemental composition of the 2-BA modified product (from http://prospector.ucsf.edu). Peak (a) shows very similar distribution pattern. The three isotopic clusters labeled with (b) show the intensity of the monoisotopic peak higher than that of the mono-^{13}C peak. The isotopic cluster (c) shows a bigger deviation from the expected distribution, especially for the bis-^{13}C peak. All spectra were the average of the accumulation of 100 shots.

3.4. Using Homologous Reagent to Quantify Disulfide Isomers

Homologous alkylating reagent pairs may be used to quantify simple mixtures of disulfide isomers. The analytical strategy is illustrated in Scheme 3. 2, using an example of the two disulfide isomers identified from CN-induced mass mapping of oxidized FBPase. Structure **1** has a sole disulfide bond between Cys155 and Cys174; structure **2** has a sole disulfide bond between Cys174 and Cys179. Assume the concentration ratio of structure **1** and structure **2** in oxidized FBPase is 2:1. On the left, the oxidized enzyme is directly modified with tagA (2-BA), as Cys155 is only free and available for alkylation in structure **2**, one of the three molecules are labeled with tagA. On the right, the same aliquot is totally reduced to generate a second state. All cysteines are free and available for alkylation after total reduction. During modification with tagB (2-BP), Cys155 in all three molecules is labeled. Then, the two aliquots are combined and proteolytically digested. Analysis of the digestion mixture would focus on the differentially-modified peptide containing Cys155. In this example, the tagA-modified peptide and tagB-modified peptides would have an intensity ratio of 1:3. In general, if the intensity ratio of the tagA-modified Cys155 peptide and tagB-modified Cys155 is X, the concentration ratio of structure 1 and structure 2 can be expressed as (1-X)/X.

Scheme 3. 2. Analytical strategy to quantify relative concentration ratio of two disulfide isomers in oxidized FBPase using homologous alkylating regent pair.

4. Conclusion

For the relative quantification of cysteine-containing peptides at two different states to be successful, the modification reaction needs to proceed to completion (14). Cysteine-specific alkylating reagents, such as 2-BP, 2-BA, iodoacetamide and ICAT, may incompletely modify cysteine residues as well as overly modify other non-cysteine residues, such as the N-terminal amino group, the amino group of a lysine residue, or the side chains of histidine, methionine and tryptophan. After adjusting reaction conditions, such as reaction time and reagent-to-substrate ratio, stoichiometric alkylation at cysteines is achieved for the 2-BA and 2-BP reagents.

The lack of knowledge and control over the reagent-to-substrate (peptide or protein) ratio, which may lead to over-modification or under-modification, is recognized as a major threat to the accuracy and the dynamic range of this homologous alkylation-based quantification methodology, as well as to other similar quantification technologies such as ICAT-based methods and MCAT-based methods.

The quantification at the peptide level with the homologous reagent pair (2-BA and 2-BP) was successful in that the peptide concentration ratio at two states can be represented by the peak intensity ratio of the two differently alkylated peptides. An assessment of the minimal error and standard deviation was made by studying the isotopic distribution of a peptide. An analytical strategy based on this homologous alkylating reagent pair method is proposed to analyze the relative concentration ratio of disulfide isomers.

References:

1.   Roth, F. P., Hughes, J. D., Estep, P. W., and Church, G. M. (1998) *Nat Biotechnol* **16**, 939-945

2.   DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) *Science* **278**, 680-686

3.   Link, A. J., Hays, L. G., Carmack, E. B., and Yates, J. R. (1997) *Electrophoresis* **18**, 1314-1334

4.   Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futcher, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., MesquitaFuentes, R., and Payne, W. E. (1997) *Electrophoresis* **18**, 1347-1360

5.   Boucherie, H., Sagliocco, F., Joubert, R., Maillet, I., Labarre, J., and Perrot, M. (1996) *Electrophoresis* **17**, 1683-1699

6.   Cagney, G., and Emili, A. (2002) *Nat Biotechnol* **20**, 163-170

7.   Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) *Nat Biotechnol* **17**, 994-999

8.   Cadene, M., and Chait, B. T. (2000) *Anal Chem* **72**, 5655-5658

9.   Gobom, J., and Nordhoff, E. (2002) *Mass Spectrometry and Hyphenated Techniques in Neuropeptide Research*, 415-429

10. Gardner, J. A., and Matthews, K. S. (1987) *Anal Biochem* **167**, 140-144

11. Uchida, K., Tanaka, K., and Hiratsuka, T. (1972) *Biochim Biophys Acta* **256**, 132-141

12. Boja, E. S., and Fales, H. M. (2001) *Anal Chem* **73**, 3576-3582

13. Lapko, V. N., Smith, D. L., and Smith, J. B. (2000) *J Mass Spectrom* **35**, 572-575

14. Smolka, M. B., Zhou, H., Purkayastha, S., and Aebersold, R. (2001) *Anal Biochem* **297**, 25-31

CHAPTER 4

'SIGNATURE SETS (S_i)', A MINIMAL POSITIVE SIGNATURE FEATURE FOR IDENTIFYING PROTEIN DISULFIDE STRUCTURES WITH CYANYLATION (CN)-BASED MASS MAPPING METHODS

1. Introduction

The conventional methodology for characterizing disulfide structures relies on a combination of Edman sequencing, proteolytic digestion, and mass mapping (1,2). The drawbacks of these conventional methods include the inability to analyze proteins with adjacent cysteines and a risk of forming artifacts due to disulfide scrambling during the alkaline proteolysis. The cyanylation (CN)-based mass mapping methodology avoids the problem of disulfide scrambling by covalently modifying the free sulfhydryls at low pH (3); furthermore, this methodology is amenable to the analytical challenge presented by proteins containing closely spaced cysteines as demonstrated in the analysis of cystinyl protein folding intermediates (4), and of a tightly knotted protein containing three adjacent cysteines among those forming four disulfide bonds (5).

One potential problem of using the CN-based method is that some CN-induced cleavage fragments may be undetected, partly because cleavage yield varies among different local sequences (6), and partly because matrix ions may interfere with detection of a peptide ion during the MALDI process. We introduce the `signature set' (S_i) concept to alleviate this concern. Specifically, we show that the correct disulfide structure can be derived as long as a small, critical fraction of CN-induced cleavage fragments is

obtained. Furthermore, we analyze the structure of signature sets in order to derive the optimal conditions for producing them. For example, we observe the surprising fact that signature sets typically do contain fragments from doubly reduced isoforms, so it is advisable to adjust conditions so that both singly and doubly reduced isoforms will be generated.

2. Development of the Concept of Signature Sets

2.1. Disulfide Structure and Fragment Set

For a cystinyl protein with $2n$ cysteines, when all the cysteines are involved in disulfide bonds, there are $(2n-1)*(2n-3)*...*3*1$ possible, isomeric disulfide structures(7). The number of CN-induced cleavage fragments has an upper bound of $C_{2n+2}^2 = (n+1)*(2n+1)$, a combination of picking two points (two ends of a cleavage fragment) out of $2n+2$ (number of cysteines + 2 termini of the protein) possible points. Note that the number of disulfide structures in Table 4. 1 increases at a faster-than-exponential rate in the number of disulfide bonds ($n!$ grows faster than a^n.), while the number of cleavage fragments increases in a polynomial fashion in the number of disulfide bonds.

number of disulfide bonds	1	2	3	4	5	6	7	8	9	10
number of disulfide structures	1	3	15	105	945	10,395	135,135	2,027,025	34,459,425	654,729,075
number of possible CN-induced cleavage fragments	6	15	28	45	66	91	120	153	190	231

Table 4. 1. Number of disulfide structures and number of CN-induced cleavage fragments as a function of number of disulfide bonds.

In Scheme 4. 1, 'N' represents the N-terminus of the protein, 'C' represents the C-terminus, and '1' represents a token cysteine residue. This new symbolism is proposed because it provides an efficient shorthand code to represent various CN-induced cleavage fragments for the computational considerations described in this chapter. Upon CN-induced cleavage, the structure represented at the left in Scheme 4. 1 produces two fragments: [N,1] and [1,C]: [N,1] represents the N-terminal fragment, which consists of all residues from the N-terminus of the protein up to, but not including the first (in this case, the only) cysteine; [1,C] represents the second fragment consisting of the modified cysteine, in the form of an iminothiazolidine derivative, and all remaining residues through the C-terminus of the protein.

N

1 —SH

Cyanylation

CDAP,
pH=3

N

1 —SCN

CN-induced
Cleavage

NH₃

N

[N,1]

1

1

[1,C]

C
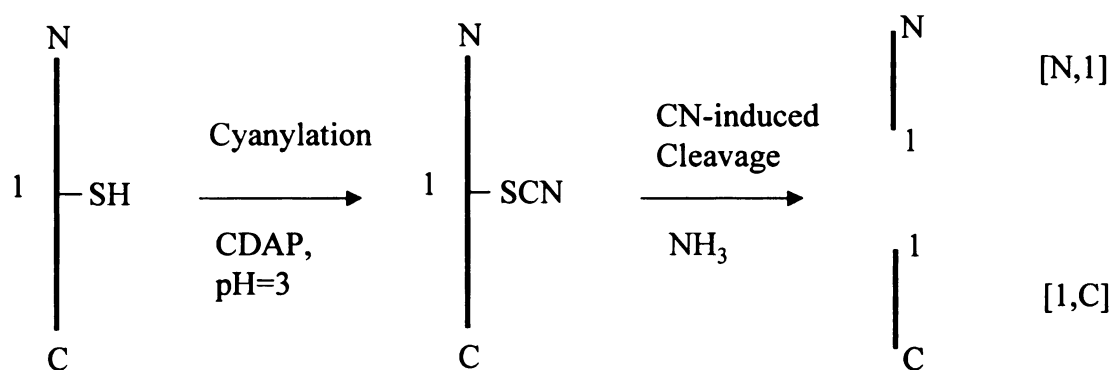
Scheme 4. 1. Symbolism for cleavage products resulting from cleavage of a cyanylated cysteine-containing peptide.

Whereas the symbolism introduced in Scheme 4. 1 is attractive from the standpoint of computational coding, it does not readily represent the well-documented chemistry involved in the cyanylation-based mass mapping methodology (16). For example, the numeral '1' appears twice in Scheme 4. 1 to code two different fragments. When a numeral (corresponding to the relative position of a given cysteine from the N-terminus) is the first entry in a set of two coordinates that code a given fragment, such as '1' in the case of [1,C], that numeral represents the modified cysteine residue that defines the N-terminus of that CN-induced cleavage fragment. When a numeral is the second defining coordinate for a given fragment, such as '1' in the case of [N,1], that numeral refers to the amino acid residue on the N-terminal side of the designated cysteine (#1 in this case); thus, the residue preceding the designated cysteine becomes the C-terminus of the given CN-induced cleavage fragment. When applied to a protein sequence, this generic representation can be decoded to the amino acid residue numbers (see hEGF example in this chapter). In general, when X and Y are two cysteines in the protein sequence (X being the one that is closer to the N-terminus of the protein), fragment [X, Y] represents the amino acid sequence of the CN-induced cleavage fragment starting with

the derivatized CysX residue and extending to the residue on the N-terminal side of CysY (but not including CysY). This generic representation of CN-induced fragments is used throughout this chapter.

There are three possible disulfide structures for a generic 2-disulfide bond protein, all shown in Scheme 4. 2. Disulfide structure A (with two disulfide bonds: Cys1-Cys3 and Cys2-Cys4), upon partial reduction and cyanylation, generates three isoforms: two singly reduced isoforms and one doubly/totally reduced isoform. One of the two singly reduced isoforms in which the disulfide bond between Cys2 and Cys4 is reduced and cyanylated, upon CN-induced cleavage, will generate three fragments: [N,2], [2,4], [4,C]; the other possible singly reduced isoform, in which the disulfide bond between Cys1 and Cys3 is reduced and cyanylated, will generate three other fragments: [N,1], [1,3] and [3,C]. Similarly, the CN-induced cleavage of the doubly/totally reduced isoform generates five fragments: [N,1], [1,2], [2,3], [3,4] and [4,C].

A fragment set associated with a disulfide structure is defined as a set containing CN-induced cleavage fragments, with replicated fragments only appearing once, from all possible partially-/totally reduced isoforms.. The fragment set associated with disulfide structure A in Scheme 4. 2 contains 9 fragments: [N,1], [1,2], [2,3], [3,4], [4,C], [N,2], [2,4], [1,3] and [3,C]. Note that fragment [4,C] may be derived from a singly reduced isoform, as well as from a doubly reduced isoform. Such redundant fragments are only reported once in a fragment set. Mass spectrometry, a tool to determine the identity of a fragment by means of measuring its mass, is incapable of differentiating the pathways from which a fragment is generated. The fragment sets for structures associated with

disulfide structure B and disulfide structure C can be generated in a similar fashion and are listed in Scheme 4. 2.

A.

Fragment Set:
[N,1] [1,2] [2,3] [3,4] [4,C] [N,2] [2,4] [1,3] [3,C]

N   1   2   3   4   C

B.

Fragment Set:
[N,1] [1,2] [2,3] [3,4] [4,C] [N,3] [2,C]

N   1   2   3   4   C

C.

Fragment Set:
[N,1] [1,2] [2,3] [3,4] [4,C] [1,4] [N,2] [3,C]

N   1   2   3   4   C
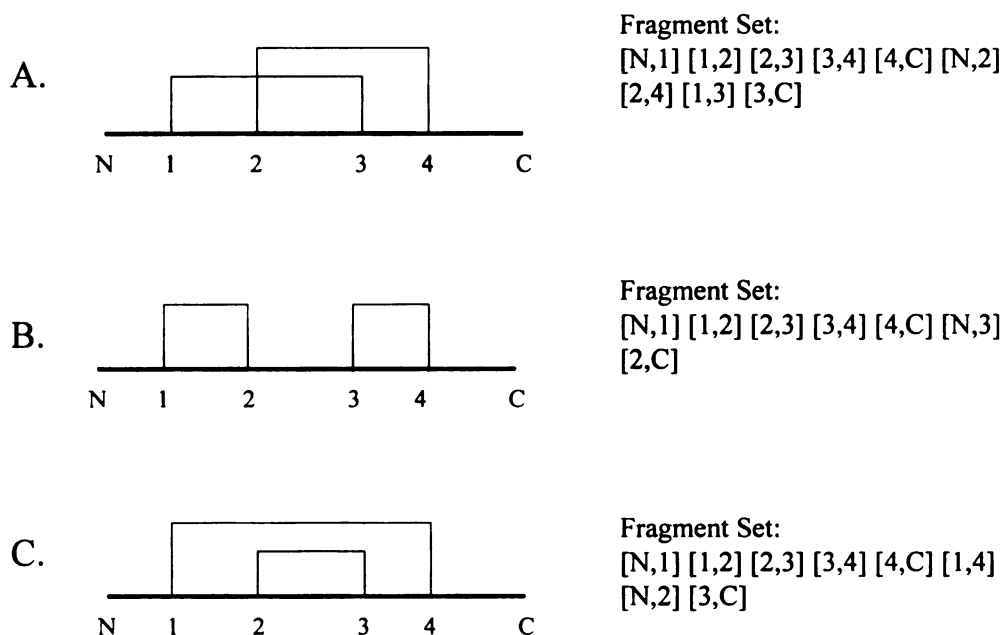
Scheme 4. 2. Three possible disulfide structures for a two-disulfide protein and the corresponding fragment sets.

2.2. Fragments from CN-induced Cleavage of Singly and Doubly Reduced Isoforms Represent the Entire Fragment Set

During partial reduction of a protein containing more than four disulfide bonds, triply reduced isoforms may be generated. As reduction and cyanylation of one disulfide bond generates two cleavage sites along the poly-peptide backbone, a triply reduced isoform has 6 cleavage sites and may generate 7 CN-induced cleavage fragments

.

N———1

4 ———5

SCN SCN SCN SCN SCN SCN   CN-induced   1 ———2

Cleavage   5 ———6

————●——●——●——●——●——●——→ 2 ———3

N 1 2 3 4 5 6 C   6 ———C

3 ———4
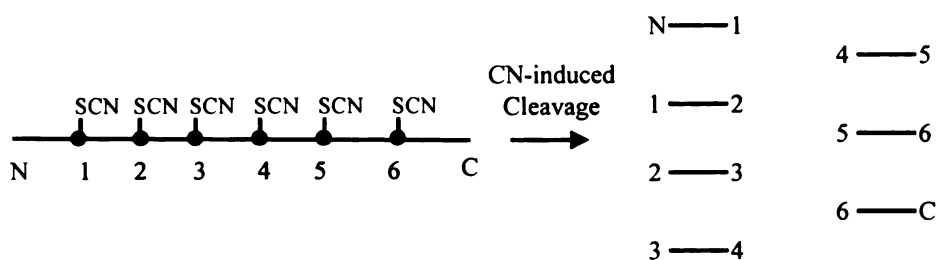
Scheme 4. 3. CN-induced cleavage fragments from a triply reduced/cyanylated isoform of a multi-cystinyl protein ($n \geq 3$). All the seven fragments can be generated from the CN-induced cleavage of a singly or a doubly reduced isoform from the same protein.

The structure on the left in Scheme 4. 3 represents a triply reduced and cyanylated isoform of a multi-cystinyl protein ($n \geq 4$). Note that the residual n-3 disulfide bonds that are not reduced during partial reduction are not shown in this scheme. The six dots along the polypeptide backbone represent the six cysteines that are reduced and cyanylated; these six cysteines had been involved in three disulfide bonds in the original structure. Upon cleavage and total reduction, this isoform generates 7 fragments: [N,1], [1,2], [2,3], [3,4], [4,5], [5,6] and [6,C]. These seven fragments can all be generated from the CN-induced cleavage of a singly or a doubly reduced and isoform from the same structure: Fragment [N,1] can be generated from the cleavage reaction of a reduced isoform in which the disulfide bond involving Cys1 is reduced. By the same reasoning, fragment [6,C] can also be generated from another singly reduced isoform in which the disulfide bond involving Cys6 is reduced. Fragment [1,2] can be generated from either a singly reduced isoform or a doubly reduced isoform depending on how Cys1 and Cys2 are connected: if Cys1 and Cys2 are connected to each other to form a disulfide bond, fragment [1,2] can be generated from a singly reduced isoform when disulfide bond between Cys1 and Cys2 is reduced; if Cys1 and Cys2 are connected

to two of the rest of the four cysteines (Cys3~Cys6) to form two disulfide bonds, fragment [1,2] can be generated from a doubly reduced isoform when those two disulfide bonds are reduced. By the same reasoning, fragments [2,3], [3,4], [4,5], [5,6] can all be generated from the CN-induced cleavage reactions of either a singly or a doubly reduced isoform. Thus, we proved that all fragments from the cleavage of a triply reduced isoform can all be generated from singly/doubly reduced isoforms. The conclusion is applicable to more highly-reduced isoforms by the same reasoning. This theoretical insight provides a guidance for the partial reduction experimental procedure: it is not advisable to overly reduce the cystinyl protein during partial reduction to generate mostly triply or more highly reduced isoforms, as the CN-induced cleavage fragments from these isoforms would provide no more information regarding to the disulfide structure than those from singly/doubly reduced isoforms.

2.3. Differential Sets ($D_{(i, j)}$) for a Disulfide Structure (F_i)

In a comparison between any two fragment sets (F_i and F_j), there are some fragments that are common to both structures. Importantly, there are fragments that are unique to only one structure. We define a subset of the original fragment set (F_i) that contains all the fragments that are present in F_i but absent in F_j, the differential set $D_{(i, j)}$. Similarly, another differential set $D_{(j, i)}$ contains all the fragments that are present in fragment set F_j but absent in fragment set F_i. In a particular pair-wise comparison between two fragment sets (F_i and F_j)generated from two isomeric 2-disulfide proteins, as illustrated in Scheme 4. 4, $D_{(i, j)}$ contains four distinguishing fragments, while $D_{(j, i)}$ contains two fragments.

A.

N   1   2   3   4   C

Fragment Set i:
[N,1] [1,2] [2,3] [3,4] [4,C] [N,2] [2,4]
[1,3] [3,C]

B.

N   1   2   3   4   C

Fragment Set j:
[N,1] [1,2] [2,3] [3,4] [4,C] [N,3] [2,C]

[N,2]   [N,1] [1,2]
[2,4]   [2,3] [3,4]   [N,3]
[1,3]   [4,C]   [2,C]
[3,C]

Differential Set D$_{(i,j)}$ Differential Set D$_{(j,i)}$

Common Fragments
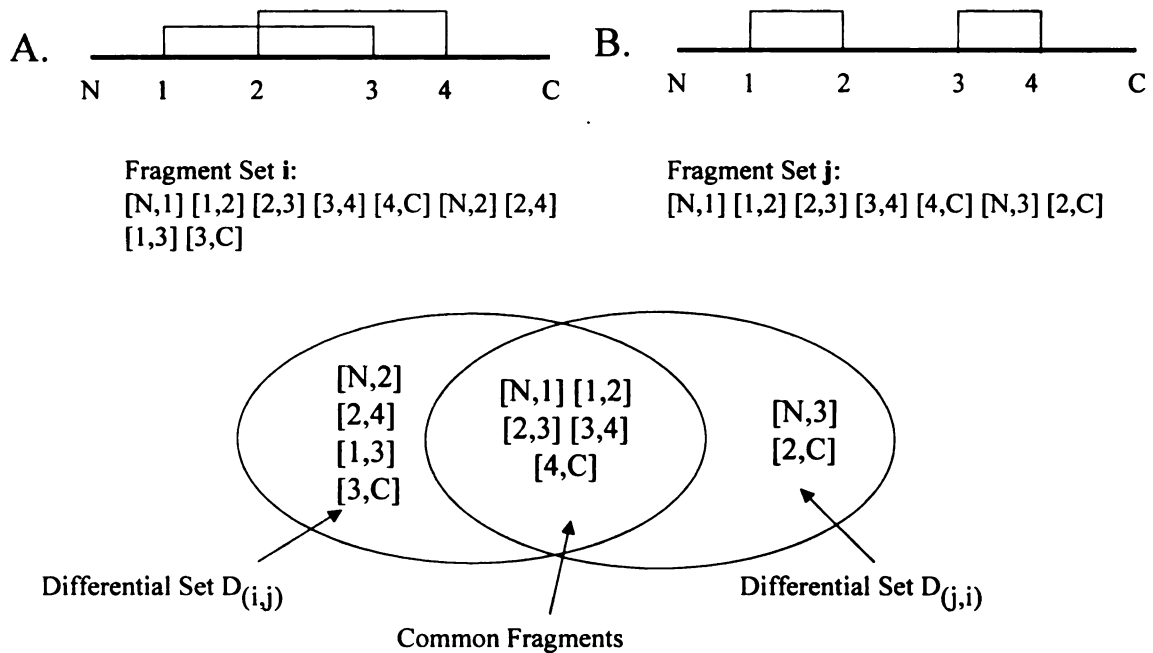
Scheme 4. 4. Two differential sets (**D$_{(i,j)}$** and **D$_{(j,i)}$**) generated in the comparison between the fragment sets of two isomeric disulfide structures for a 2-disulfide protein.

Disulfide
Structure i

d$_2$

d$_1$ d$_f$

N   X   Y   C

Disulfide
Structure j

d$_2$

d$_1$

region1   region2

N   X   Y   C
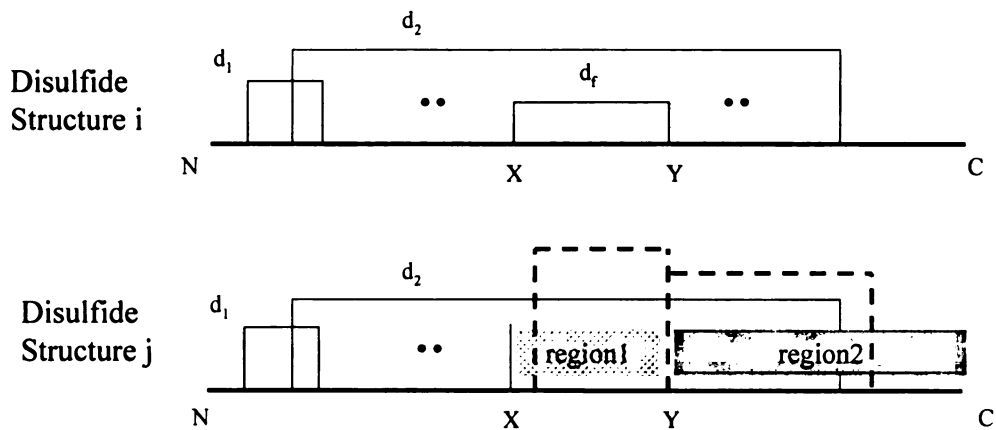
Scheme 4. 5. 'd$_f$' in disulfide structure i represents the first disulfide bond from the N-terminus that connects differently between the two distinct disulfide structures.
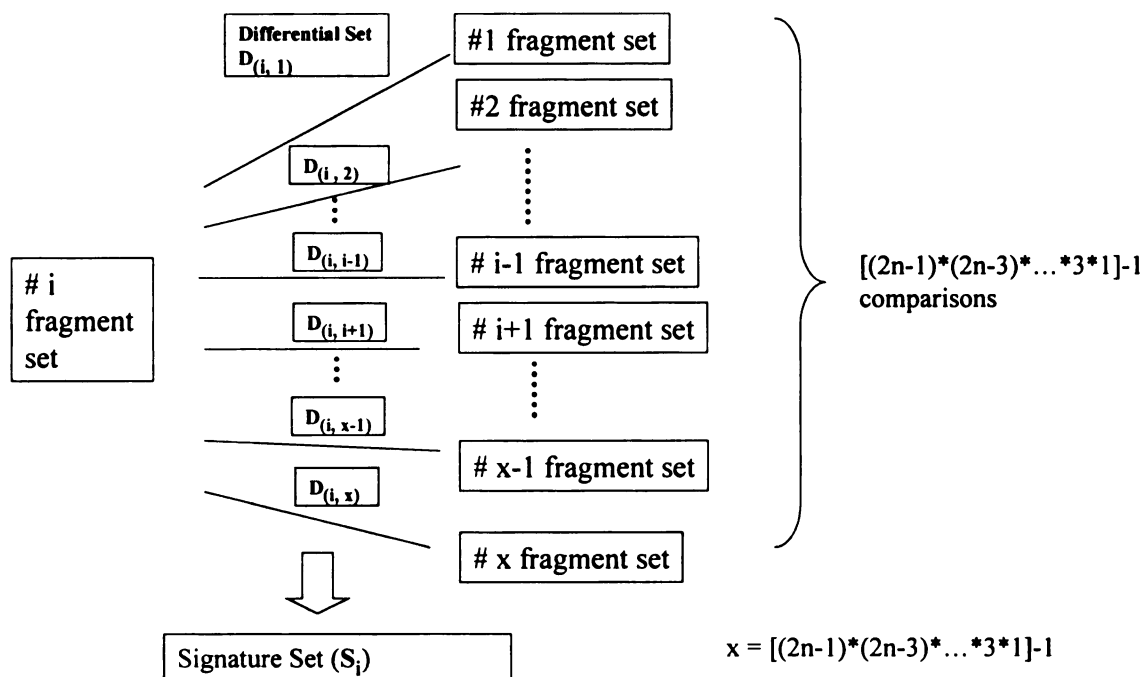
We observed that regardless of how many disulfide bonds are there in a cystinyl protein, the differential set **D$_{(i, j)}$** in any pair-wise comparison is always populated. Two

92

isomeric, fully disulfide bonded structures, structure **i** and structure **j** as shown in Scheme 4. 5, have to have at least two different disulfide bonds. In disulfide structure **i**, $\mathbf{d_f}$ represents the first disulfide bond numbered from the N-terminus that is absent in disulfide structure **j**. **X** and **Y** are the two cysteine residues that join to form $\mathbf{d_f}$. All the cysteines to the N-terminal side of **X** have the same connectivity in both structure **i** and structure **j**. Because none of the aforementioned cysteines is connected to either **X** or **Y** in structure **i**, neither are those cysteines in disulfide structure **j** connected to either **X** or **Y**. In disulfide structure **j**, either **X** or **Y** forms a disulfide bond with another cysteine in the region either between **X** and **Y** (region 1) or to the C-terminal side of **Y** (region 2 in Scheme 4. 5). Consider the cysteine that binds with **Y** to make disulfide structure **j**; if it is a cysteine in region 1, fragment [X, Y], a cleavage product of a singly reduced isoform of disulfide structure **i**, becomes a fragment in the differential set $\mathbf{D_{(i, j)}}$. Disulfide structure **j**, upon partial reduction, cyanylation, cleavage and total reduction, can never produce fragment [X, Y]; any fragment terminated with **Y** for disulfide structure **j** will always be shorter than fragment [X, Y]. The other possibility is for **Y** in structure **j** to bind with a cysteine that is to the C-terminal side of **Y** (region 2). In that case, fragment [Y, C] from structure **i** will be a member of in the differential set $\mathbf{D_{(i, j)}}$ in a pair-wise comparison with disulfide structure **j**: for disulfide structure **j** to generate a CN-induced cleavage fragment terminated with **Y**, the fragment will always be shorter than [Y, C]. In either case, the differential set $\mathbf{D_{(i, j)}}$ contains at least one fragment. For a protein with **n** disulfide bonds, there are [(2n-1)*(2n-3)*...*3*1] different structures when all cysteines are involved in disulfide bonds. For a particular disulfide structure **i**, the differential set $\mathbf{D_{(i, j)}}$ (**j** =1 ~ [(2n-1)*(2n-3)*...*3*1], **j** ≠ **i**) is always populated.

2.4. Generation of Signature Sets for a Particular Disulfide Structure

The detection of any one of the fragments in a differential set is sufficient to distinguish between two structures. However, to pinpoint one disulfide structure from among all theoretical possibilities, a signature set (S_i) can be constructed from comparisons of the fragment set F_i of one disulfide structure with the fragment sets characterizing all the other theoretically possible disulfide structures.

Scheme 4. 6 shows how the signature sets for a particular disulfide structure (i) are generated. For a particular fragment set (i), there are $[(2n-1)*(2n-3)*...*3*1]-1$ differential sets, which are all subsets of fragment set (F_i), generated in the comparisons between F_i and all the other possible fragment sets from possible disulfide structures. We can take one fragment from each differential set to form a new set. The new set, after duplicated are removed, is defined as a signature set (S_i). There are $[(2n-1)*(2n-3)*...*3*1]-1$ fragments in this set, with many replicates. It is noticed that this signature set (S_i), is also a subset of fragment set (F_i), in fact, it is a minimal subset of F_i that leads to an unequivocal identification of disulfide structure i. As there are often multiple fragments in a differential set $(D_{(i, j)})$, there are usually multiple signature sets (S_i) associated with one disulfide structure.

Differential Set $D_{(i, 1)}$

#1 fragment set

#2 fragment set

$D_{(i, 2)}$

$D_{(i, i-1)}$ # i-1 fragment set

i fragment set

$D_{(i, i+1)}$ # i+1 fragment set

$D_{(i, x-1)}$

$D_{(i, x)}$ # x-1 fragment set

x fragment set

$[(2n-1)*(2n-3)*...*3*1]-1$ comparisons

Signature Set (S_i)

$x = [(2n-1)*(2n-3)*...*3*1]-1$

Scheme 4. 6. A signature set (S_i) is formed by combining one fragment from each of the $[(2n-1)*(2n-3)*...*3*1]-1$ differential sets.

3. Evaluation of the Signature Sets Concept with Disulfide Mass Mapping Data

3.1. Output of Signature Sets for Generic Proteins with n Disulfide Bonds

Computer programs are written to generate disulfide structures, fragment sets, differential sets and signature sets, sequentially, for generic cystinyl proteins with n disulfide bonds. An example of the output for all isomeric disulfide structures is shown in shorthand symbols in Table 4. 2 for a generic 3-disulfide protein. The 15 possible structures for a three-disulfide protein are indexed from 0 to 14; likewise, for a four-disulfide protein, the 105 possible disulfide structures are indexed from 0 to 104. Because a cystinyl protein containing three disulfide bonds has only 15 isomeric structures, they are listed in Table 4. 2 for illustrative purposes. Structure 0 in Table 4. 2 has three disulfide bonds: (1, 2) (3, 4) (5, 6). Note that the symbol (1, 2) is different from [1, 2]: the

95

former symbol refers to a disulfide bond formed by linking Cys1 and Cys2; the latter

symbol refers to a CN-induced cleavage fragment consisting of residues from Cys1 up to

and including the residue that is on the N-terminal side of Cys2, as defined earlier.

Structure Index Number	Disulfide Structure		
0	(1, 2)	(3, 4)	(5, 6)
1	(1, 2)	(3, 5)	(4, 6)
2	(1, 2)	(3, 6)	(4, 5)
3	(1, 3)	(2, 4)	(5, 6)
4	(1, 3)	(2, 5)	(4, 6)
5	(1, 3)	(2, 6)	(4, 5)
6	(1, 4)	(2, 3)	(5, 6)
7	(1, 4)	(2, 5)	(3, 6)
8	(1, 4)	(2, 6)	(3, 5)
9	(1, 5)	(2, 3)	(4, 6)
10	(1, 5)	(2, 4)	(3, 6)
11	(1, 5)	(2, 6)	(3, 4)
12	(1, 6)	(2, 3)	(4, 5)
13	(1, 6)	(2, 4)	(3, 5)
14	(1, 6)	(2, 5)	(3, 4)

Table 4. 2. Indexing of all 15 possible isomeric disulfide structures for a 3-disulfide protein containing 6 cysteines.

A second program is responsible to generate fragment set, differential sets and

signature sets for each disulfide structure. For a particular disulfide structure, the program

composes all possible singly reduced and doubly reduced isoforms. For each isoform, the

program performs the cleavage and total reduction operation *in silico* to generate three

CN-induced cleavage fragments from a singly reduced isoform and five fragments from a

doubly reduced isoform; it then arranges these fragments into a fragment set. Before

including a CN-induced cleavage fragment in a fragment set, the program first

determines whether the fragment has already appeared in the fragment set. Thus, all fragment sets are 'non-redundant'; i.e., they contain no replicates. From previous discussion, we showed that triply or more highly reduced isoforms do not contribute new information to a fragment set; thus, they are not considered by the program. A differential set ($D_{(i,j)}$) is generated by comparing fragment set i and fragment set j and contains all the fragments that are present in fragment set i but not in fragment set j. Signature set (S_i) is then generated by taking one fragment from each of the D_{ij} (j from 1 to $[(2n-1)*(2n-3)*...*3*1]$ and $j \neq i$) and combine them together. Repetitive fragments in S_i are only recorded once. As there can be multiple fragments in D_{ij}, there are multiple signature sets (S_i) for each structure (i) and each one of these signature sets (S_i) is equally sufficient to differentiate disulfide structure (i) from all the other possible disulfide structures. The output of all theoretically possible disulfide structures for a generic 3- and 4-disulfide protein (from program 1) and the signature sets associated with these disulfide structures (from program 2) can be obtained from following URL:

http://www.bch.msu.edu/facilities/massspec/disulfide/index.html

## 3.2. Human Epidermal Growth Factor (hEGF) and its Disulfide Isomers

Scheme 4. 7. Primary structure and native disulfide linkages of human epidermal growth factor (hEGF)

97

We re-examined data obtained from analysis of three isomeric forms of a cystinyl protein containing three disulfide bonds using our partial reduction, CN-induced cleavage, and mass mapping methodology. The proteins are human epidermal growth factor (hEGF, structure shown in Scheme 4. 7) and two of its non-native three-disulfide isomers, as produced during an oxidative refolding experiment with reduced hEGF and separated by HPLC as illustrated in Figure 4. 1 (4). In the following, we coordinate the computed signature sets (S_i) with experimentally detected fragments to efficiently identify the disulfide structures of the three isomeric species. In the 53-residue sequence of hEGF, the six cysteines reside at positions 6, 14, 20, 31, 33 and 42. These cysteines in the computer-generated disulfide structures, fragment sets, differential sets and signature sets are enumerated '1-6', according to their relative positions in the sequence. The computer-shorthand representation of the 15 possible disulfide structures are listed and indexed in Table 4. 2. The computer-shorthand representation of the CN-induced cleavage fragments in Table 4. 3 needs to be carefully transformed into a chemical representation for agreement with the mass spectral results of analysis(4). For example, [N,3] corresponds to CN-induced cleavage at the N-terminal side of the 3rd cysteine (residue 20 in hEGF), and thus, corresponds to a peptide containing all residues from the N-terminus of hEGF up to residue 19. Fragment [2,5] corresponds to CN-induced cleavages at the N-terminal side of Cys2 (residue 14) and of Cys5 (residue 33), and thus, consists of residues from the iminothiazolidine-blocked residue 14 up to residue 32 in hEGF.

Cleavage fragment, numbered by Cys residue numbers in hEGF	Cleavage fragments, numbered by relative Cys positions	Observed masses (Da)	Expected masses (Da)	Fragments in Signature Sets
1-32	[N,5]	3692.9	3694.1	X
itz-6-19	[1,3]	1540.1	1541.7	X
itz-14-30	[2,4]	1969.2	1970.3	X
itz-20-53	[3,C]	4218.3	4218.0	X
itz-31-53	[4,C]	2917.1	2916.4	X
1-5	[N,1]	682.7	n.d.	
1-13	[N,2]	1555.6	1554.6	
itz-33-41	[5,6]	1020.1	1020.2	
itz-42-53	[6,C]	1721.0	1722.0	
itz-6-13	[1,2]	915.9	915.9	
itz-14-19	[2,3]	667.7	n.d.	
itz-20-30	[3,4]	1345.6	1345.0	
itz-20-32	[3,5]	1562.8	n.d.	
itz-31-32	[4,5]	260.2	n.d.	

Table 4. 3. Expected fragments from CN-induced mass mapping of native hEGF and those detected by MALDI-MS. The first five fragments are constituents of four signature sets (see Table 4. 4), while the other nine theoretically possible fragments do not contribute in distinguishing the disulfide structure of the native protein from other theoretically possible disulfide structures. (n.d. =not detected)
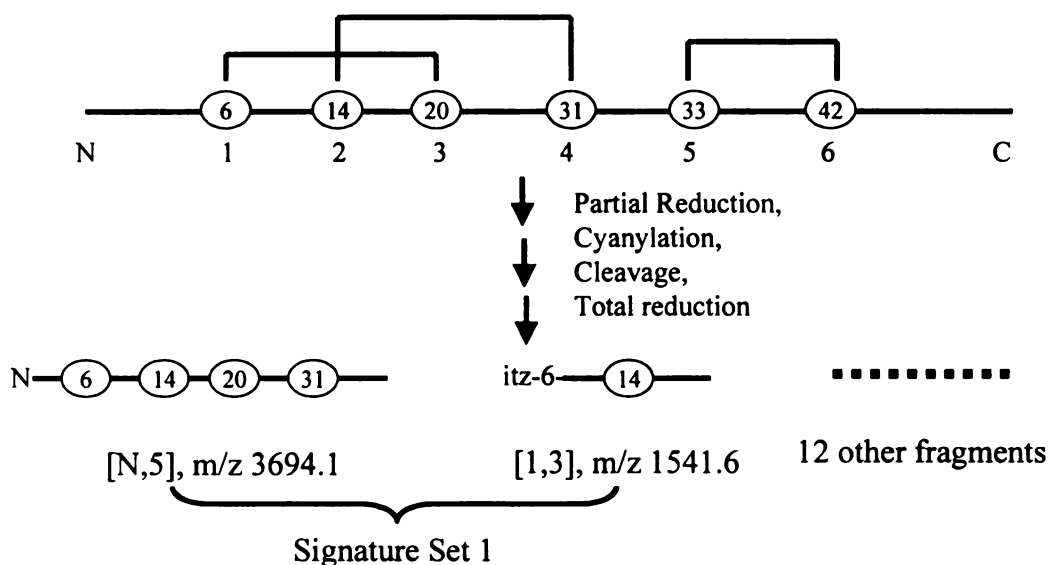
Scheme 4. 8. The two fragments, [N,5] and [1,3], in the first signature set for native hEGF are generated in the CN-induced mass mapping experiment, along with 12 other theoretically possible fragments.

Native hEGF has a disulfide structure identical to structure 3 in Table 4. 2. From computational analysis, there are four signature sets for structure 3, as listed in Table 4. 4, two consisting of two fragments, and the other two consisting of three fragments. Analysis of the experimental data show that peaks representing fragments in all four sets were detected in the MS mapping. The generation of the two fragments in the first signature set from CN-induced mass mapping experiment as well as the shorthand representation of these two fragments are shown in Scheme 4. 8. In the mass spectral analysis of the mixture of CN-induced cleavage fragments, a peak at m/z 3692.9 was observed and assigned to fragment [N,5], consisting residues from the N-terminus of hEGF to the alanine at residue 31 (calculated mass = 3694.1 Da); another peak at m/z 1540.1 was observed and correspondingly mapped to fragment [1,3], consisting of residues from the iminothiazolidine derivative of Cys6 to Val19 (calculated mass =

1541.6 Da). Detection of these two fragments satisfies one of the signature sets, and, thus, constitutes identification of hEGF as disulfide structure 3. Other peaks in the mass spectrum also represent all the fragments in three other signature sets, thereby corroborating the identification of hEGF as structure 3. Whereas, it is possible for some other disulfide structures to generate one of the two fragments listed in the first two signature sets in Table 4. 4, it is impossible for any one of the other 14 possible disulfide structure to generate both fragments simultaneously. From observation of the mass spectra, ten of the fourteen expected fragments in the fragment set for structure 3 were identified by mass mapping, as shown in Table 4. 3. Three of the four undetected fragments have a very low calculated mass. Possible reason is that their mass spectral peaks may have been obscured by those for matrix ions during analysis by MALDI. Ionization of the fourth fragment [3, 5], a larger cleavage fragment from a doubly reduced isoform, might have been suppressed by other peptide components, causing the fragment to be undetected. However, failure to detect these four peaks does not affect the outcome of the disulfide structure determination. Signature sets analysis shows the intrinsic robustness of the cyanylation mapping approach: even when a majority of the cleavage fragments are not detected, one can still unequivocally determine the disulfide structure.

Similarly, when analyzing the experimental data on folding intermediate IIIB, we detected all members of three of the five signature sets, each containing only two of the twelve possible fragments (in Table 4. 4). In the analysis of the data on the folding intermediate IIIA, we detected both fragments for the only signature set.
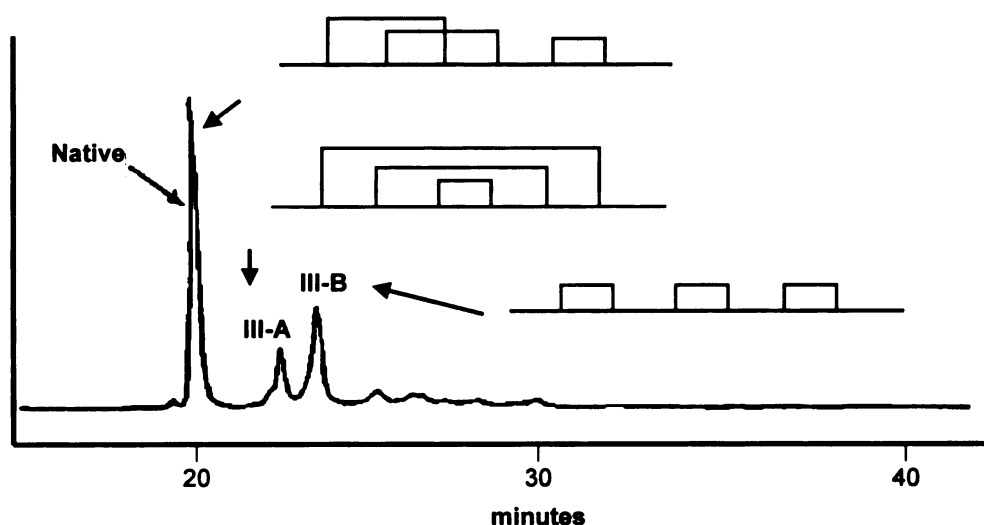
Figure 4. 1. HPLC chromatogram showing three fully oxidized isomers of hEGF after 48 hours of oxidative refolding. 'Native' represents the native hEGF, while 'III-A' and 'III-B' represent the misfolded versions of the protein.

Signature sets for structure 3 (native)	Detected	Signature sets for structure 0 (III-B)	Detected	Signature sets for structure 14 (III-A)	Detected
[N,5] [1,3]	X	[N,3] [N,5]	X	[1,6] [2,5]	X
[N,5] [2,4]	X	[N,5] [2,5]			
[1,3] [3,C] [4,C]	X	[N,5] [2,C]	X		
[2,4] [3,C] [4,C]	X	[2,5] [2,C]			
		[2,C] [4,C]	X		

Table 4. 4. Signature sets for the three isomeric proteins resulting from the refolding of hEGF. The sets marked with 'X' had all constituent fragments represented by peaks in previously recorded mass spectra during disulfide mass mapping.

3.3. Further Evaluation of the Signature Sets with Ribonuclease A (RNase A) Mapping Data

RNase A is consisted of 124 amino acid residues, 8 of which are cysteines that link to one another to form of 4 disulfide bonds (disulfide structure shown in Scheme 4. 9). After partial reduction, cyanylation, cleavage and total reduction, 27 possible CN-induced cleavage fragments can be generated *in silico* to form its theoretical fragment set. The differential sets, generated by comparing this fragment set with the fragment sets of the other 104 theoretically possible disulfide structures, contain from 2 to 15 fragments (results not shown). With the aforementioned algorithm, the computer outputs 25 signature sets of native RNase A in Table 4. 5 (shown in Table 4. 5). Detection of the combinations of fragments in any one of the 25 signature sets is sufficient to distinguish the native disulfide structure of RNase A from the other 104 possible isomeric disulfide structures. Mass spectra obtained during previous mass mapping of RNase A showed three signature sets (marked with 'X' in Table 4. 5) were detected.
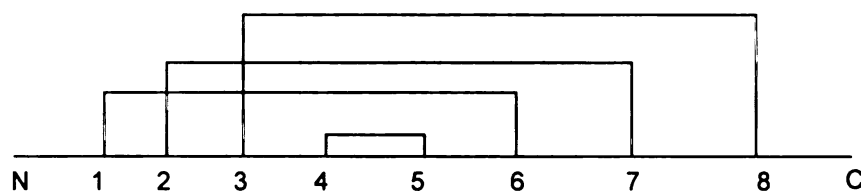
Scheme 4. 9. Representation of the native disulfide structure of RNase A: Eight cysteines at residues 26, 40, 58, 65, 72, 84, 95 and 110 are numerated '1-8' according to their relative position to the N-terminus of the protein.

	Signature Sets for RNase A	Identified in Experiments
1	[1,6] [2,6] [3,8]	X
2	[1,6] [2,7] [3,6]	
3	[1,6] [2,7] [3,7]	
4	[1,6] [2,7] [3,8]	X
5	[1,6] [3,7] [3,8]	
6	[2,6] [2,7] [3,8]	X
7	[2,7] [3,6] [3,8]	
8	[0,3] [2,6] [2,7] [3,6]	
9	[0,3] [2,6] [2,7] [3,7]	
10	[0,3] [2,7] [3,6] [3,7]	
11	[1,4] [2,6] [2,7] [3,6]	
12	[1,4] [2,6] [2,7] [3,7]	
13	[1,4] [2,7] [3,6] [3,7]	
14	[1,6] [2,6] [2,7] [3,6]	
15	[1,6] [2,6] [2,7] [3,7]	
16	[1,6] [2,7] [3,6] [3,7]	
17	[2,6] [2,7] [3,6] [3,8]	
18	[2,6] [2,7] [3,6] [5,8]	
19	[2,6] [2,7] [3,6] [6,9]	
20	[2,6] [2,7] [3,7] [3,8]	
21	[2,6] [2,7] [3,7] [5,8]	
22	[2,6] [2,7] [3,7] [6,9]	
23	[2,7] [3,6] [3,7] [3,8]	
24	[2,7] [3,6] [3,7] [5,8]	
25	[2,7] [3,6] [3,7] [6,9]	

Table 4. 5. Signature sets for RNase A; compositions for fragments are expressed in computer-shorthand according to relative position of Cys residues as explained in text.

3.4. Characteristics of Signature Sets

An analysis of the signature sets from all 105 disulfide structures for a cystinyl protein with eight cysteines shows that most of the CN-induced cleavage fragments in a set are those that contain some internal cysteines. For example, most fragments in a signature set contain at least two free internal cysteines. Another observation is that most of the fragments are internal fragments, i.e., they do not contain either the N-terminus or the C-terminus of the protein. These two observations are in agreement with the 'rule-out' concept of the 'negative signature mass algorithm'(8). Larger fragments with more internal free cysteines rule out more invalid linkages; fragments not terminated by the N- or C-terminus of the protein are capable of ruling out invalid linkages from both ends of the cleavage fragment, as opposed to the N-or C-terminal fragment, which can only rule out invalid linkages from the opposing end of the original molecule. Both of these observations are manifested in the signature sets for native RNase A, as shown in Table 4. 5.

Another observation of the signature sets is that most often, a signature set consists of a combination of fragments from singly and doubly reduced isoforms of the protein. For example, out of the 25 signature sets for RNase A, only one set consists of fragments derived solely from a singly reduced isoform; in all of the other 24 sets, the fragments are derived from both singly and doubly reduced isoforms. This observation suggests that during the partial reduction, it behooves the analyst to adjust conditions so that both singly and doubly reduced isoforms are generated.

4. Frequency Analysis of the Fragments from CN-induced Cleavage of a Four-Disulfide Protein

A particular CN-induced cleavage fragment is characterized by its two 'ends', either a modified (iminothiazolidine derivative) cysteine residue (or a residue to the N-terminal side of a cysteine) or the N- or C- terminus of the protein. For a protein with four disulfide bonds, a total of 10 points (8 cysteines plus the N- and C- terminus) can serve as these 'ends' for a CN-induced cleavage fragment. The number of all the theoretically possible CN-induced cleavage fragments for such a protein is 45, derived from a combination of selecting two points ('ends') out of the 10 available points. For a randomly selected disulfide structure, we examined all four possible singly reduced isoforms and the CN-induced cleavage fragments that could be derived from them; cumulatively, 12 of the 45 possible fragments were derived from all the singly reduced isoforms. We also considered all six possible doubly reduced isoforms and their CN-cleavage products; cumulatively, 30 fragments were derived from the doubly reduced isoforms of the particular disulfide structure. Then, we projected this list of 42 fragments, some of them replicates, onto the list of the 45 theoretically possible fragments and recorded the number of occurrences for each fragment. In analogous fashion, we can assess the occurrences of each of the 45 theoretically possible fragments in cleavage fragment sets for all the other 104 theoretically possible disulfide structures. Totally, $105*42 = 4410$ occurrences are distributed among the 45 fragments.

If a fragment has a high number of occurrences, many different disulfide structures are capable of generating that particular fragment, i.e., such a fragment would not be especially useful in distinguishing between disulfide structures. Fragments with a low occurrence rate, on the other hand, has more distinguishing power and are more

likely to validate a given disulfide structure. The distribution of occurrences among the 45 theoretically possible fragments is shown in Table 4. 6. Fragments [N,8], [1,C] and [N,C] never appear once; this is an expected result because such fragments derive from either no or only one cleavage site, a condition that is incompatible with our partial reduction methodology in which a minimum of two cysteines (cleavage sites) is produced.

From Table 4. 6, we note that 'longer' fragments with more internal free cysteines have a lower number of occurrences, and thus, are more informative in distinguishing between disulfide structures. This observation is graphically illustrated in Figure 4. 2. The fragments with fewer internal free cysteines occur more frequently than ones with more free cysteines, a phenomenon more obvious for the fragments derived from doubly reduced isoforms than from singly reduced isoforms.

Another observation of the results of frequency analysis shown in Table 4. 6 is that a given internal fragment has a lower number occurrences than a terminal fragment with the same number of free cysteines. For example, consider terminal fragments [N,4] and [5,C]; each has three internal free cysteines and each has occurred 96 times. On the other hand, fragments [1,5], [2,6], [3,7] and [4,8], each with three internal free cysteines, have occurred 69 times, and are thus are more informative in distinguishing between disulfide structures. These observations agree with those made from the signature sets analysis.

Fragment	Number of Occurrences	Fragment	Number of Occurrences	Fragment	Number of Occurrences
[N,1]	420	[1,8]	15	[4,5]	150
[N,2]	270	[1,C]	0	[4,6]	105
[N,3]	165	[2,3]	150	[4,7]	69
[N,4]	96	[2,4]	105	[4,8]	42
[N,5]	54	[2,5]	69	[4,C]	54
[N,6]	30	[2,6]	42	[5,6]	150
[N,7]	15	[2,7]	24	[5,7]	105
[N,8]	0	[2,8]	15	[5,8]	69
[N,C]	0	[2,C]	15	[5,C]	96
[1,2]	150	[3,4]	150	[6,7]	150
[1,3]	105	[3,5]	105	[6,8]	105
[1,4]	69	[3,6]	69	[6,C]	165
[1,5]	42	[3,7]	42	[7,8]	150
[1,6]	24	[3,8]	24	[7,C]	270
[1,7]	15	[3,C]	30	[8,C]	420

Table 4. 6. Number of occurrences of fragments from CN-induced cleavage of singly and doubly reduced isoforms of all 105 theoretically possible disulfide structures for a protein containing four disulfide bonds.
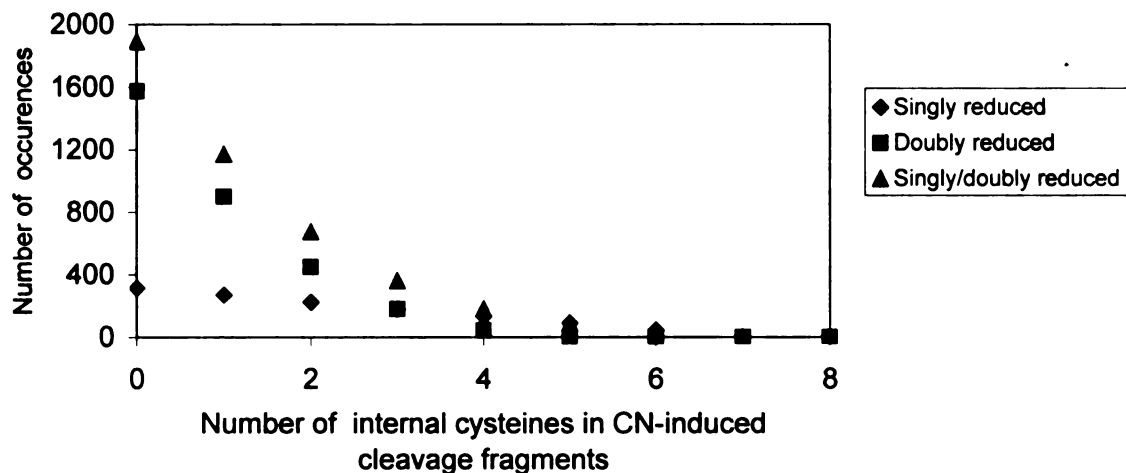
Figure 4. 2. The number of occurrences for a CN-induced cleavage fragment as a function of its constituent free cysteines for a four-disulfide protein.

5. Homogeneity in Disulfide Structures of the Protein under Study

Similar analysis of the fragment sets generated from combinations of two different disulfide structures of a protein with 3 disulfide bonds showed that the distinguishing feature between the sets is lost in some cases. From 15 possible disulfide structures for a protein with three disulfide bonds, 105 different heterogeneous combinations of a pair of disulfide structures can be derived (C_{15}^2=105). We generate the fragment sets for all the 105 structure pairs. Like a fragment set for an individual disulfide structure, the fragment set for a disulfide structure pair does not contain duplicates. Some fragment sets associated with the structure pairs, in at least one pair-wise comparison with other fragment sets, do not have any fragments in a differential set.

In Scheme 4. 10, we generated the fragment sets from two pairs: pair A consisting of disulfide structures (1, 2) (3, 4) (5, 6) and (1, 4) (2, 3) (5, 6); Pair B consisting of

disulfide structures (1, 2) (3, 4) (5, 6) and (1, 5) (2, 3) (4, 6). The fragment set of pair A consists of 16 members: [0,1], [1,2], [2,7], [0,3], [3,4], [4,7], [0,5], [5,6], [6,7], [2,3], [2,5], [4,5], [1,4], [0,2], [3,7] and [3,5]; the fragment set of pair B consists of 20 members: [0,1], [1,2], [2,7], [0,3], [3,4], [4,7], [0,5], [5,6], [6,7], [2,3], [2,5], [4,5], [1,5], [5,7], [0,2], [3,7], [0,4], [4,6], [3,5] and [1,4]. We noticed that fragment set A is a subset of fragment set B. In the comparison between these two sets, set A does not have any fragments in its differential set ($D_{(A, B)}$), while pair B has four fragments: [1,5], [5,7], [0,4] and [4,6] in its differential set ($D_{(A, B)}$). For Pair A, because of the lack of at least one CN-induced cleavage in this differential set, the signature set can not be constructed. Pair B, however, has a populated differential set in every one of the comparisons with all the other 104 sets. Signature sets for structure pair B can be constructed.
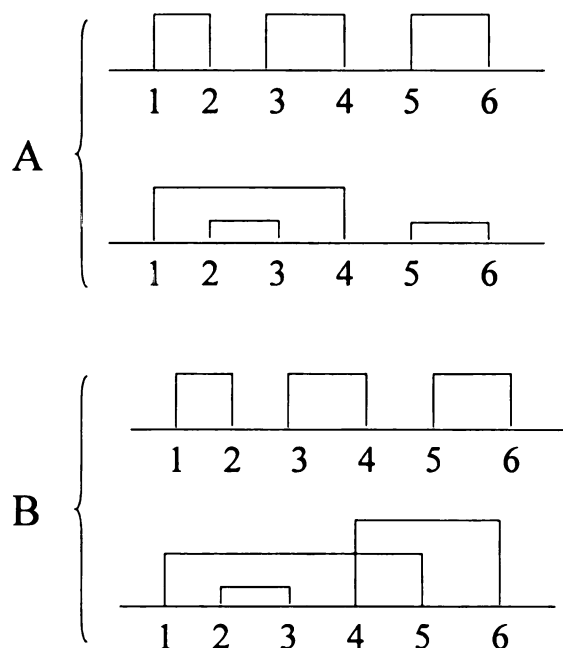
Scheme 4. 10. Pair A does not have any fragments in its differential set when its fragment set is compared with that of Pair B. Pair B has at least one fragment in its differential set when compared with the fragment sets from 104 other possible pairs.

The major limitation of the CN-induced disulfide mass mapping methodology is the requirement for starting with one single, pure disulfide structure, free from other molecules with isomeric disulfide structures. Of course, this is a requirement for all methodologies used to solve disulfide structures because some conflicting evidence (observed or not) will always be produced by isomeric disulfide structures using any distinguishing technique. Two disulfide structures in certain combinations for a generic 3-disulfide protein, such as the two forming pair B in Scheme 4. 10, can be identified simultaneously by CN-based mapping method without any separation of the two isomers.

6. Conclusions

Cyanylation-induced cleavage of the polypeptide backbone of cystinyl proteins generates sequence-specific fragments that can be used to mass-map the linkages of specific cysteines involved in a particular disulfide structure. A proof is given herein for the existence of at least one unique CN-induced cleavage fragment in $D_{(i,j)}$, in a pair-wise comparison between two fragment sets. Further, we have shown that certain combinations of CN-induced fragments form a series of signature sets (S_i), that uniquely characterize a given disulfide structure. By inspection, we noted that signature sets consist of fragments that contain internal cysteines within their sequences. We also note that triply and more highly reduced isoforms of the cystinyl protein do not contribute unique fragments to a fragment set; thus, it behooves the analyst to seek conditions that promote formation of the singly and doubly reduced isoforms of the protein. Selected generic signature sets for a three-disulfide cystinyl protein were transposed to the sequence of hEGF to illustrate the capacity to distinguish the native protein and two mis-

folded isomers based on combinations of as few as three CN-induced cleavage fragments as detected by mass spectrometry. The concept of signature sets provides the basis for designing analytical methodology to unambiguously identify specific disulfide structures for isolated cystinyl proteins.

Reference:

1. Li, J. H., Yen, T. Y., Allende, M. L., Joshi, R. K., Cai, J., Pierce, W. M., Jaskiewicz, E., Darling, D. S., Macher, B. A., and Young, W. W. (2000) *J Biol Chem* **275**, 41476-41486

2. Zhou, Z., and Smith, D. L. (1990) *J Protein Chem* **9**, 523-532

3. Wu, J., and Watson, J. T. (1997) *Protein Sci* **6**, 391-398

4. Wu, J., Yang, Y., and Watson, J. T. (1998) *Protein Sci* **7**, 1017-1028

5. Qi, J., Wu, J., Somkuti, G. A., and Watson, J. T. (2001) *Biochemistry* **40**, 4531-4538

6. Wu, J., and Watson, J. T. (1998) *Anal Biochem* **258**, 268-276

7. Benham, C. J., and Jafri, M. S. (1993) *Protein Sci* **2**, 41-54

8. Qi, J., Hang, D., Rupp, M., Torng, E., Borges, C. R., Wu, W., and Watson, J. T. (2003) *Journal of the American Society for Mass Spectrometry* **14**, 1032-1038

APPENDIX

DISULFIDE INDEX, FRAGMENT SETS, SIGNATURE SETS FOR A GENERIC

3-DISULFIDE BOND PROTEIN

1. Disulfide Structure Index

Disulfide Structure Number	Linkages of the 6 Cysteines, Numbered by their Relative Position within Sequence		
0	(1, 2)	(3, 4)	(5, 6)
1	(1, 2)	(3, 5)	(4, 6)
2	(1, 2)	(3, 6)	(4, 5)
3	(1, 3)	(2, 4)	(5, 6)
4	(1, 3)	(2, 5)	(4, 6)
5	(1, 3)	(2, 6)	(4, 5)
6	(1, 4)	(2, 3)	(5, 6)
7	(1, 4)	(2, 5)	(3, 6)
8	(1, 4)	(2, 6)	(3, 5)
9	(1, 5)	(2, 3)	(4, 6)
10	(1, 5)	(2, 4)	(3, 6)
11	(1, 5)	(2, 6)	(3, 4)
12	(1, 6)	(2, 3)	(4, 5)
13	(1, 6)	(2, 4)	(3, 5)
14	(1, 6)	(2, 5)	(3, 4)

2. Fragment Sets

Fragment Set for Disulfide Structure 0:

[0,1] [1,2] [2,7] [0,3] [3,4] [4,7] [0,5] [5,6] [6,7] [2,3] [2,5] [4,5]

Fragment Set for Disulfide Structure 1:

[0,1] [1,2] [2,7] [0,3] [3,5] [5,7] [0,4] [4,6] [6,7] [2,3] [2,4] [3,4] [4,5] [5,6]

Fragment Set for Disulfide Structure 2:

[0,1] [1,2] [2,7] [0,3] [3,6] [6,7] [0,4] [4,5] [5,7] [2,3] [2,4] [3,4] [5,6]

Fragment Set for Disulfide Structure 3:

[0,1] [1,3] [3,7] [0,2] [2,4] [4,7] [0,5] [5,6] [6,7] [1,2] [2,3] [3,4] [3,5] [4,5]

Fragment Set for Disulfide Structure 4:

[0,1] [1,3] [3,7] [0,2] [2,5] [5,7] [0,4] [4,6] [6,7] [1,2] [2,3] [3,5] [3,4] [2,4] [4,5] [5,6]

Fragment Set for Disulfide Structure 5:

[0,1] [1,3] [3,7] [0,2] [2,6] [6,7] [0,4] [4,5] [5,7] [1,2] [2,3] [3,6] [3,4] [2,4] [5,6]

Fragment Set for Disulfide Structure 6:

[0,1] [1,4] [4,7] [0,2] [2,3] [3,7] [0,5] [5,6] [6,7] [1,2] [3,4] [4,5] [3,5]

Fragment Set for Disulfide Structure 7:

[0,1] [1,4] [4,7] [0,2] [2,5] [5,7] [0,3] [3,6] [6,7] [1,2] [2,4] [4,5] [1,3] [3,4] [4,6] [2,3] [3,5] [5,6]

Fragment Set for Disulfide Structure 8:

[0,1] [1,4] [4,7] [0,2] [2,6] [6,7] [0,3] [3,5] [5,7] [1,2] [2,4] [4,6] [1,3] [3,4] [4,5] [2,3] [5,6]

Fragment Set for Disulfide Structure 9:

[0,1] [1,5] [5,7] [0,2] [2,3] [3,7] [0,4] [4,6] [6,7] [1,2] [3,5] [1,4] [4,5] [5,6] [3,4]

Fragment Set for Disulfide Structure 10:

[0,1] [1,5] [5,7] [0,2] [2,4] [4,7] [0,3] [3,6] [6,7] [1,2] [4,5] [1,3] [3,5] [5,6] [2,3] [3,4] [4,6]

Fragment Set for Disulfide Structure 11:

[0,1] [1,5] [5,7] [0,2] [2,6] [6,7] [0,3] [3,4] [4,7] [1,2] [2,5] [5,6] [1,3] [4,5] [2,3] [4,6]

Fragment Set for Disulfide Structure 12:

[0,1] [1,6] [6,7] [0,2] [2,3] [3,7] [0,4] [4,5] [5,7] [1,2] [3,6] [1,4] [5,6] [3,4]

Fragment Set for Disulfide Structure 13:

[0,1] [1,6] [6,7] [0,2] [2,4] [4,7] [0,3] [3,5] [5,7] [1,2] [4,6] [1,3] [5,6] [2,3] [3,4] [4,5]

Fragment Set for Disulfide Structure 14:

[0,1] [1,6] [6,7] [0,2] [2,5] [5,7] [0,3] [3,4] [4,7] [1,2] [5,6] [1,3] [4,6] [2,3] [4,5]

3. Signature Sets

Signature Sets for Structure 0:

[N,3] [N,5]

[N,5] [2,5]

[N,5] [2,C]

[2,5] [2,C]

[2,C] [4,C]

Signature Sets for Structure 1:

[2,C] [3,5]

[2,C] [4,6]

[N,3] [N,4] [3,5]

[N,3] [N,4] [4,6]

[N,3] [2,C] [3,5]

[N,3] [2,C] [4,6]

Signature Sets for Structure 2:

[2,C] [3,6]

[N,3] [N,4] [3,6]

[N,3] [2,C] [3,6]

Signature Sets for Structure 3:

[N,5] [1,3]

[N,5] [2,4]

[N,5] [1,3] [3,C]

[N,5] [2,4] [3,C]

[1,3] [3,C] [4,C]

[2,4] [3,C] [4,C]

Signature Sets for Structure 4:

[N,4] [2,5]

[2,5] [3,C]

[N,4] [1,3] [2,5]

[N,4] [1,3] [3,5]

[N,4] [1,3] [4,6]

[N,4] [2,4] [2,5]

[1,3] [2,5] [3,C]

[1,3] [3,C] [4,6]

[2,4] [2,5] [3,C]

[2,4] [3,C] [4,6]

[N,2] [N,4] [2,4] [3,5]

[N,2] [N,4] [2,4] [4,6]

[N,4] [1,3] [2,4] [3,5]

[N,4] [1,3] [2,4] [4,6]

[N,4] [1,3] [3,5] [3,C]

[N,4] [2,4] [2,5] [3,5]

[N,4] [2,4] [2,5] [4,6]

[N,4] [2,4] [3,5] [3,C]

[N,4] [2,4] [3,C] [4,6]

[1,3] [2,5] [3,5] [3,C]

[1,3] [3,5] [3,C] [4,6]

[1,3] [3,5] [3,C] [5,C]

[2,4] [2,5] [3,5] [3,C]

[2,4] [3,5] [3,C] [4,6]

[2,4] [3,5] [3,C] [5,C]

Signature Sets for Structure 5:

[N,4] [2,6]

[2,6] [3,6]

[2,6] [3,C]

[N,4] [1,3] [3,6]

[1,3] [2,6] [3,6]

[1,3] [3,6] [3,C]

[2,4] [2,6] [3,6]

[2,4] [3,6] [3,C]

[N,2] [N,4] [2,4] [3,6]

[N,4] [1,3] [2,4] [3,6]

[N,4] [2,4] [2,6] [3,6]

[N,4] [2,4] [3,6] [3,C]

Signature Sets for Structure 6:

[N,5] [1,4]

[N,5] [1,4] [3,C]

[1,4] [3,C] [4,C]

Signature Sets for Structure 7:

[1,4] [2,5]

[2,5] [3,6]

[N,3] [1,4] [3,6]

[N,3] [2,4] [2,5]

[N,3] [2,5] [3,5]

[1,3] [1,4] [3,6]

[1,4] [2,4] [2,5]

[1,4] [2,4] [3,6]

[1,4] [2,5] [3,5]

[1,4] [2,5] [3,6]

[1,4] [3,5] [3,6]

[1,4] [3,6] [4,6]

[1,4] [3,6] [4,C]

[2,4] [2,5] [3,6]

[2,4] [2,5] [4,C]

[2,5] [3,5] [3,6]

[2,5] [3,5] [4,C]

Signature Sets for Structure 8:

[1,4] [2,6]

[2,6] [3,5]

[N,3] [2,4] [2,6]

[1,4] [2,4] [2,6]

[2,4] [2,6] [3,5]

[2,4] [2,6] [4,6]

[2,4] [2,6] [4,C]

Signature Sets for Structure 9:

[N,4] [1,5]

[1,4] [1,5]

[1,5] [3,C]

[N,4] [1,4] [3,5]

[N,4] [1,4] [4,6]

[1,4] [1,5] [3,C]

[1,4] [3,C] [4,6]

[N,4] [1,4] [3,5] [3,C]

[1,4] [1,5] [3,5] [3,C]

[1,4] [3,5] [3,C] [4,6]

[1,4] [3,5] [3,C] [5,C]

Signature Sets for Structure 10:

[1,5] [2,4]

[1,5] [3,6]

[N,3] [1,5] [3,5]

[1,3] [1,5] [3,5]

[1,5] [2,4] [3,5]

[1,5] [3,5] [3,6]

[1,5] [3,5] [4,C]

Signature Sets for Structure 11:

[1,5] [2,5]

[1,5] [2,6]

[2,5] [2,6]

Signature Sets for Structure 12:

[N,4] [1,6]

[1,4] [1,6]

[1,6] [3,6]

[1,6] [3,C]

[N,4] [1,4] [3,6]

[1,4] [1,6] [3,6]

[1,4] [3,6] [3,C]

Signature Sets for Structure 13:

[1,6] [2,4]

[1,6] [3,5]

Signature Sets for Structure 14:

[1,6] [2,5]