This is to certify that the

dissertation entitled

Capture and Identification of the Intermediates of Refolding and Reductive Unfolding of Proteins by Cyanylation Methodology and Mass Spectrometry

presented by

Ying Yang

has been accepted towards fulfillment
of the requirements for

Doctor of Philosophy degree in Chemistry

J. Throckmorton Watson
Major professor

Date June 16, 1998

CAPTURE AND IDENTIFICATION OF THE INTERMEDIATES OF REFOLDING AND REDUCTIVE UNFOLDING OF PROTEINS BY CYANYLATION METHODOLOGY AND MASS SPECTROMETRY

By

Ying Yang

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Chemistry

1998

ABSTRACT

CAPTURE AND IDENTIFICATION OF THE INTERMEDIATES OF REFOLDING AND REDUCTIVE UNFOLDING OF PROTEINS BY CYANYLATION METHODOLOGY AND MASS SPECTROMETRY

By

Ying Yang

Cyanylation chemistry of cysteine residues combined with mass spectrometry provides promising methodology for providing structural information of cysteine residues in the proteins. The sulfhydryl groups of proteins can be specifically cyanylated by a cyanylation reagent, the N-terminal side of cyanylated cysteine residues can be cleaved in alkaline condition, and the masses of peptide fragments (one N-terminal side peptide and a series of iminothiazolidine-blocked peptides) can be analyzed by mass spectrometry. Matching the masses of the cleaved peptides can be used to recognize the structural information of cysteine residues in the proteins.

Two types of mass spectrometry have been used to analyze proteins. One is matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS); the another is electrospray ionization mass spectrometry (ESI-MS). We found that MALDI-MS showed frequent, and sometimes complete, suppression of a signal from iminothiazolidine-blocked peptides; because such a phenomenon can compromise our analytical strategy, we have investigated the use of ESI for these analyses. The results indicate that the responses of our blocked peptides to ESI and MALDI are frequently complementary.

Conventional methodology for determining the pairing in the disulfide bonding structure of proteins relies on the use of appropriate proteases to cleave the peptide backbone in between the cysteine residues. The possibility of a proteolytic cleavage site becomes less likely when proteins contain closed or adjacent cysteins. We have investigated the applicability of a novel methodology involving partial reduction of disulfide bonds, cyanylation of sulfhydryl groups, chemical cleavage of partially reduced/cyanylated protein, and mass mapping of cleavage products to the structural analysis of proteins containing adjacent cysteines in two model protein, long R^3 insulin-like growth factor-I (LR^3IGF-I) and insulin-like growth factor-I (IGF-I). These proteins contain three disulfide bonds, two of which involve adjacent cysteines. The disulfide structures of the proteins were unambiguously determined by the methodology.

A new methodology based on cyanylation chemistry of cysteine residue and mass spectrometry has been used to capture and identify folding intermediates of long insulin-like growth factor I (LR^3IGF-I) and insulin-like growth factor I (IGF-I). The refolding was quenched at different time points by adjusting the pH to 2.0-3.0 with a 1N HCl solution of 1-cyano-4-dimethylamino-pyridinium (CDAP) which trapped intermediates via cyanylation of free sulfhydryl groups. The disulfide structure of the intermediates was determined by partial reduction/cyanylation/chemical cleavage/mass mapping. The folding intermediates of LR^3IGF-I and IGF-I were successfully trapped and characterized. The time-dependant distribution and disulfide structure of the folding intermediates of LR^3IGF-I and IGF-I were compared based on their structural features.

ACKNOWLEDGMENTS

I wish to thank my mentor, Dr. J. Throck Watson, for his assistance and advice. I especially would like to thank my daughter, J. Jinjing Yang, my husband, Jie Yang, for their support and understanding during this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

I. Introduction

Recent advances in mass spectrometry have created new technological capabilities that are applicable to the study of peptides and proteins involved in biological processes. The discovery and development of ionization techniques for large bio-oligomers at significantly higher ionizing efficiencies permits the detection and measurement of very large intact or fragmented biopolymeric substances (1). These tools are electrospray ionization (ESI) (2) and matrix-assisted laser desorption/ionization (MALDI) (3). They have become among the most powerful methods yet available for the macromolecular characterization of living systems ranging from the measurement of the molecular weights and purity of large polar biopolymeric substances to delineation of the detailed sequence and structure of components of the complex mixtures obtained from their selective enzymatic or chemical degradation. These ionization techniques have been revolutionary in spurring research in protein biochemistry, glycobiology, and biotechnology, and most recently in DNA sequence analysis. Over the next few years these technologies are destined to become as commonplace in the conduct of biomedical research as HPLC did in the early 1980s.

The synthesis of proteins in organisms starts by transcription of the DNA sequences to messenger RNA (mRNA). In eukaryotic organisms, mRNA is processed to remove introns and is then translated on the ribosomal complex to synthesize the protein. To be biologically active, all proteins must adopt specific folded three-dimensional structures

after they depart from the ribosome (4, 5). The genetic information for the protein specifies only the primary structure, the linear sequence of amino acids in the polypeptide backbone. Many purified proteins can spontaneously refold *in vitro* after being completely unfolded, so the three-dimensional structure must be determined by the primary structure (6). How this occurs has come to be known as 'the protein folding problem'. It had been primarily of academic interest, but with the advent of the recombinant DNA revolution and its potential for producing therapeutic proteins, often in an insoluble, unfolded, inactive, and useless form, has made it also of great practical importance. For many years, scientists have endeavored to understand the process by which protein folding occurs, that is, the folding pathway.

In order to understand the folding pathway of a protein, it is necessary to trap and characterize the folding intermediates (7-9). A particularly difficult aspect in the study of the protein folding is the fact that intermediates may be short-lived and therefore hard to isolate and analyze structurally and functionally. Disulfide-containing proteins provide an opportunity to isolate and characterize trapped intermediates by quenching the disulfide pairing chemically during the time course of folding (4,10-12). Furthermore, the study of folding process of disulfide-containing protein is also important because disulfide formation often confounds the high-level expression and renaturation of recombinant proteins (5).

The sufhydryl groups of cysteine residues are among the most reactive side chains in proteins. In 1973, Jacobson *et al.* (13) introduced specific cyanylation of sulfhydryl groups of cysteine residues followed by chemical cleavage at cyanylated cysteine residues. This specific reaction with the combination of MALDI-MS has been recently

used to characterize the cysteine status of proteins (14, 15). The methodology has the advantages of being simple, fast, and sensitive.

In this dissertation, several aspects related to the analysis of cysteine status in a protein, as well as in folding intermediates, based on the combination of cyanylation methodology and mass spectrometry will be discussed. In chapter 2, two ionization technologies, MALDI and ESI, will be compared for their capabilities of analyses of iminothiazolidine-blocked peptides resulting from the chemical cleavage reaction at cyanylated cysteine residues. In chapter 3, assignment of disulfide linkages of proteins containing adjacent cysteines will be discussed. In chapter 4, the trapping method and disulfide structures of intermediates in folding and reductive unfolding processes including the folding pathway of Long R^3 insulin-like growth factor-I will be investigated. The purpose of this chapter is to introduce basic aspects of MALDI and ESI-MS, and describe current methodologies for trapping folding intermediates, locating cysteines and cystines in proteins, as well as recognizing disulfide linkages.

II. Oxidative Folding and Trapping of Intermediates

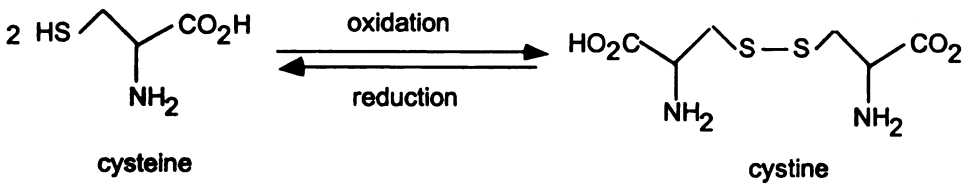
The general rules governing protein folding will be established only by characterizing for many proteins the conformations and energetics of structures at particular points along their folding pathways. Description of the properties of a native state is relatively simple because it is a single, long-lived conformation. Intermediates and unfolded states are much more difficult to characterize because they may be short-lived or exist as a distribution of possible conformers. Two useful methods are currently available to tackle this problem. A newly developed technique of pulsed-label NMR (16-

19) permits trapping and identification of amide groups that are engaged in the structured elements. This technique can, in principle, be applied to all types of proteins. However, intermediates trapped by this method are not amenable to chromatographic purification. Therefore, the intermediates cannot be further characterized structurally and functionally.

Disulfide bonds introduce nonlocal topological restraints in the folding of proteins, and their formation provides valuable signals in tracing the folding pathway of cysteine-containing proteins (20, 21) based on two important phenomena. First, it is known that the formation of disulfides is coordinated with the folding of protein (22). That is, during folding, disulfides reshuffle (break and reform) readily in order to adapt to more favorable conformations. Second, disulfide-folding intermediates, trapped either reversibly or irreversibly, can be separated and isolated by chromatographic systems. This allows examination of the number, diversity, structure, and properties of intermediate species. Data obtained from these analyses are indispensable for the construction of protein folding pathways (23). In this method, the disulfide-bonded intermediates which form during the folding process are chemically trapped in a time course manner. The trapped intermediates are isolated and structurally characterized, the kinetics of the interconversion is determined, and the different components are coordinated with a folding pathway. Therefore, the main concerns in a study of disulfide-containing protein folding are trapping of folding intermediates by quenching the disulfide formation as well as the determination of disulfide structure of intermediates.

A. Disulfide Bonds in Protein Folding

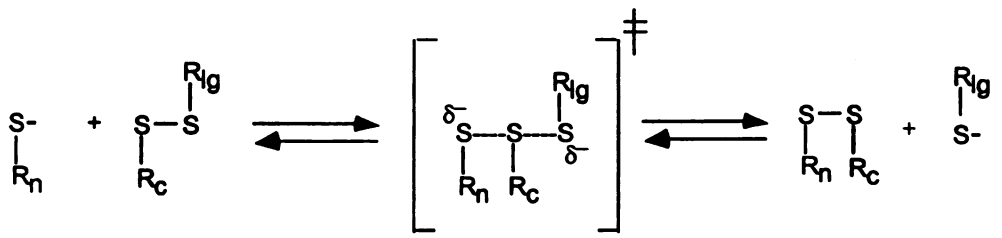
Disulfide bonds are comparatively labile and cleave readily to thiols by reduction, and thiols are readily oxidized to disulfides. Cystine can be readily formed by oxidation of cysteine which is reversibly obtained by reduction (Equation 1.1)



Equation 1.1. Cysteine and cystine

1. Disulfide Formation

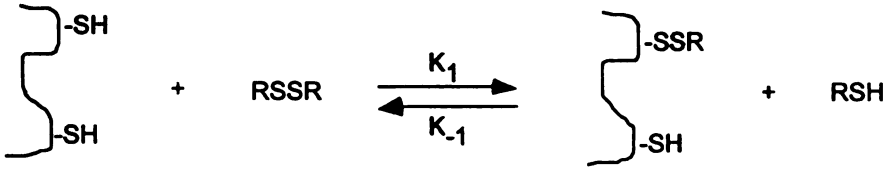
Disulfides do not form spontaneously between thiols, even when in close proximity, unless there is an appropriate oxidant. While molecular oxygen (in the presence of trace amounts of heavy metals) may adequately serve this role for some proteins, the most common source of oxidizing equivalents is a low molecular weight disulfide. Protein thiol oxidation to the disulfide results from a thiol/disulfide exchange process (24) that transfers oxidizing equivalents from the low molecular weight disulfide to the protein (Equation 1.2).



Equation 1.2. The mechanism of thiol/disulfide exchange

Thiol/disulfide exchange occurs via direct attack of a nucleophilic thiolate anion on one of the sulfurs of the disulfide bond (25, 26). The transition state is rather symmetrical with approximately equal bond formation to the nucleophilic sulfur (R_nS^-) and the leaving sulfur ($R_{lg}S^-$) and with a relatively small amount of negative charge on the central sulfur (R_cS) (27). The rate constant for the reaction increases as the basicity of the attacking thiolate (R_nS^-) nucleophile increases (pKa increases) and as the basicity of the leaving thiolate ($R_{lg}S^-$) decreases (pKa decreases). The pKa of the typical cysteine sulfhydryl group is about 8.6; however, this may vary considerably from protein to protein due to the effects of the local environment. At pH of 8.6, the typical intermolecular step of thiol/disulfide exchange between small molecules or unhindered protein thiols and disulfides will occur with a second-order rate constant of about 20 $M^{-1}s^{-1}$ (11). Since thiol/disulfide exchange occurs via the attack of a nucleophilic thiolate anion, the rate will increase with increasing pH until the attacking thiol is predominantly in the thiolate form.

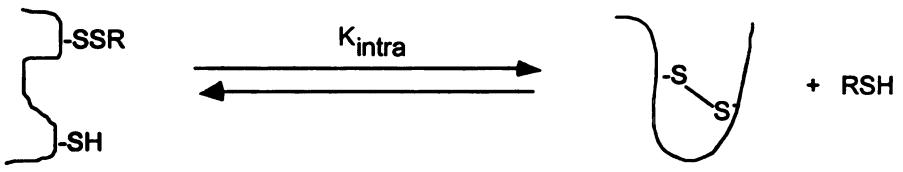
Upon adding a disulfide reagent to a reduced protein, two sequential thiol-disulfide exchange reactions are required to form one protein disulfide bond. The first is the simple chemical reaction between the disulfide reagent and one of the cysteine thiol groups, to generate the mixed disulfide (Equation 1.3).



Equation 1.3. The formation of mixed disulfide bond

The formation reaction of a mixed disulfide is bimolecular in either direction, so the rates should be dependent upon the concentrations of the disulfide and thiol forms of the reagent.

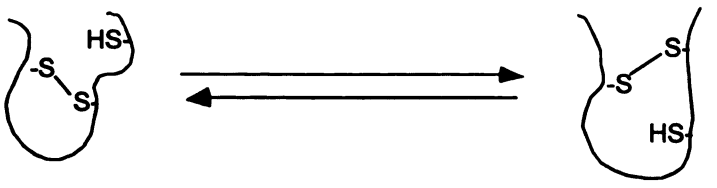
The second step is the one in which a second cysteine thiol group reacts with the mixed disulfide to form the protein disulfide bond (Equation 1.4).



Equation 1.4. The formation of disulfide bond

This step is intramolecular thiol/disulfide exchange in the forward direction, and its rate depends primarily upon the tendencies of the thiols of different cysteine residues to come into proximity of the mixed disulfide, which is primarily controlled by the protein conformation. Therefore, this step provides the most useful information about protein conformation and folding, and K_{intra} can be calculated from the observed rates of disulfide formation and reduction with different reagents (11).

Once a disulfide bond has formed in a protein with more than two cysteine residues, this disulfide bond can undergo intramolecular rearrangements via processes analogous to the intramolecular step in disulfide formation. Again, the rate of a rearrangement reaction depends on the extent to which the conformation of the polypeptide favors an attack of the thiolate on the existing disulfide bond (Equation 1.5).



Equation 1.5. Intramolecular rearrangement of a disulfide bond

The observed rate of disulfide formation depends on several factors, including the degree to which the protein thiols are ionized, the rate of reaction of thiolate with reagent, the stability of the mixed-disulfide intermediate, and the rate of formation of the protein disulfide from the mixed disulfide (28). A redox buffer of a mixture of a low molecular weight disulfide and its corresponding thiol usually is used for the formation of disulfide bonds. During oxidative folding, the disulfide component of the redox buffer serves to provide oxidizing equivalents for protein disulfide formation. The thiol component of the redox buffer serves to reduce native or non-native disulfide bonds that may otherwise lock the protein in misoxidized forms or to catalyze thiol/disulfide rearrangements. Since the reaction of disulfide formation is reversible, the capacity to form a disulfide bond will depend on the equilibrium and the rate constants for the individual steps and the concentration and oxidation potential of the redox buffer (24).

There are two kinds of redox buffer that can be used for disulfide formation, linear disulfide and cyclic disulfide. Two of them are most commonly used, glutathione and dithiothreitol. Glutathione is a tripeptide linear disulfide reagent. It exists in two forms: reduced (GSH) and oxidized (disulfide, GSSG). Glutathione universally occurs in the tissues of animals, plants, and microorganisms at comparatively high concentration (from

0.4 to 12 mM, predominantly in the form GSH) (29). GSH/GSSG system attributes an important role in the formation of disulfide bonds in proteins during their synthesis *in vivo* (30). Therefore, it is a most pertinent (*in vivo*) electron donor and acceptor to be used for the formation of disulfide bonds *in vitro*. The formation of disulfide with GSSG has the advantage of being rapid at alkaline pH, simple, extremely specific, and easy to control (31).

Oxidized dithiothreitol (DTT), a cyclic disulfide, has been employed in studies of the mechanism of protein folding. Disulfide formation using DTT is much more discriminating than with a linear reagent. The mixed disulfide is energetically unfavorable and rapidly dissociates intramolecularly, so it should not accumulate significantly (32). Only very favorable protein disulfides are formed with this reagent, at a rate that should be proportional to the rate of the intramolecular step of the reaction (33).

2. Oxidative-Folding Problem

In the late 1960s and early 1970s, a number of investigators began to determine the mechanisms by which unfolded proteins regain their native conformations *in vitro*. These studies have largely been motivated by the recognition that the number of possible conformations accessible to a polypeptide chain is so great that finding the native conformation by a completely random search would be improbably slow. This suggested that there might be specific pathways of protein folding and that characterizing partially folded intermediates associated with these pathways might provide a key to

understanding the relationship between the amino acid sequence and the three-dimensional structure of a protein.

In 1962, Haber and Anfinsen (34) first used the oxidative folding of ribonuclease and other small, disulfide-containing proteins to show that the primary sequence contained enough information to specify the correct three-dimensional structure of a protein and that the driving force for protein folding was the thermodynamic stability of the native structure. Two cysteines must be within covalent bonding distance to form a disulfide bond; therefore, identifying the intermediates that are actually populated during the folding reaction gives clues about the structural features and interactions that guide the unfolded protein to its native structure (35). At least for the disulfide-containing proteins that spontaneously refold, the overall driving force is the thermodynamic stability of the native structure--native disulfide pairings give the most stable protein (7). The interactions that specify which cysteines will be paired and in what temporal order could be similar to those in the fully folded protein (native interactions), or they may represent interactions that are not found in the native protein (non-native interactions).

Compared to the necessary rate of protein synthesis and secretion *in vivo*, uncatalysed *in vitro* oxidative folding is often a very slow and inefficient process. Consequently, *in vivo*, oxidative folding is catalysed (36, 37) and possibly coupled to protein translation/translocation (38). The rules that govern the oxidative folding of a protein *in vitro* may or may not be the same as those that guide *in vivo* folding.

The folding of even simple disulfide-containing proteins illustrates the potential complexity of oxidative folding. The mechanisms that direct the formation of a three-dimensional native structure from an unfolded protein must now accommodate the

chemistry of disulfide formation. For a protein with six cysteine residues that form three disulfide bonds, there are 15 ways of pairing the cysteines to generate molecules with three disulfide bonds. The complexity rapidly increases when more disulfides are involved.

After more than 30 years since the first study of the oxidative folding of ribonuclease, the knowledge of the mechanism of protein folding is still incomplete. There are several central questions concerning the oxidative folding of proteins.

(a) Do disulfide bond formation determine what structures will form or do other interactions in the protein dictate which disulfide bonds will form?

(b) Are all of the specific interactions that guide the folding process present in the native structure or do non-native interactions play a role?

(c) Does oxidative folding proceed through one pathway or do multiple, alternative pathways exist?

(d) Why do some disulfide-containing proteins fold with high efficiency into their native structures while others do not?

(e) Do catalysts of oxidative folding alter the mechanism of folding or simply increase the rate and extent of native structure formation?

(f) Is the mechanism of oxidative folding *in vivo* the same as that *in vitro*?

3. Folding Pathway of Bovine Pancreatic Trypsin Inhibitor (BPTI)

The most detailed and informative folding pathway elucidated thus far is that of bovine pancreatic trypsin inhibitor (BPTI), a small single-domain 58-residue amino acid protein stabilized by three disulfide bonds. The folding pathway of BPTI (21) is

illustrated in Figure 1.1. The reduced BPTI is a random coil, and it can efficiently refold by sequentially passing through one-disulfide intermediates, two-disulfide intermediates, and finally to the native protein. The folding experiments of BPTI (4, 11, 39, 40) revealed two striking features of the BPTI folding mechanism. First, the process is distinctly non-random: only a small fraction of the possible disulfide-bonded intermediates are detected during folding or unfolding. Second, disulfide formation is not a simple sequential process: the kinetically preferred mechanism includes intramolecular rearrangements involving intermediates with non-native disulfide bonds (i.e, those not present in the native protein). This finding raised the possibility that the amino acid sequence might specify the structures of folding intermediates which have different conformations from that of the final folded protein. This possibility would add new complications to the already formidable problem of predicting the three-dimensional structure of a protein.

B. Trapping of the Folding Intermediates

To identify disulfide-bonded folding intermediates, rapid quenching of the folding reaction is necessary to trap the intermediates in chemically stable forms that can be physically separated and analyzed. The criteria for a good trapping agent include that it block the thiolates of an intermediate rapidly, completely, and without modifying the protein at sites other than thiols. Several methods have been used to trap the intermediates.

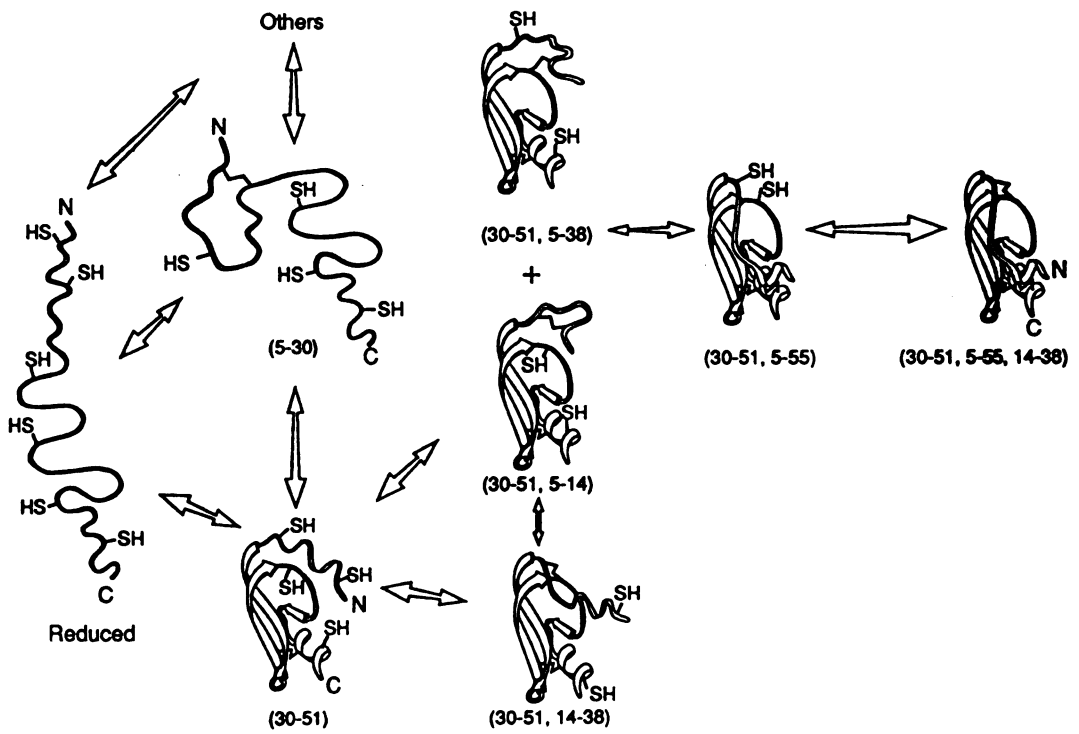
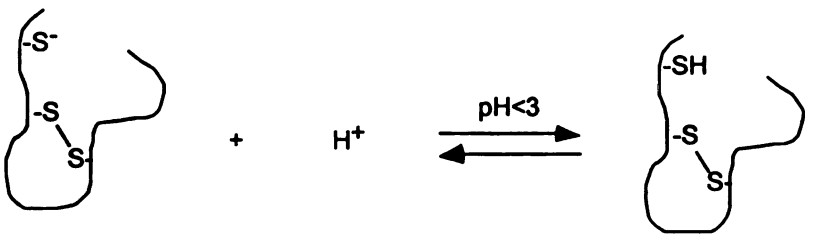


Figure 1.1. Folding pathway of BPTI

1. Acid Trapping

The folding of proteins are conducted under alkaline condition, usually at around pH 8.5. The pKa of the typical cysteine sulfhydryl group is about 8.6. Therefore, the thiolate of the intermediates can be rapidly protonated by adding acid in the folding solution to a pH less than 3 and the further thiol/disulfide exchange can be stopped (Equation 1.6).



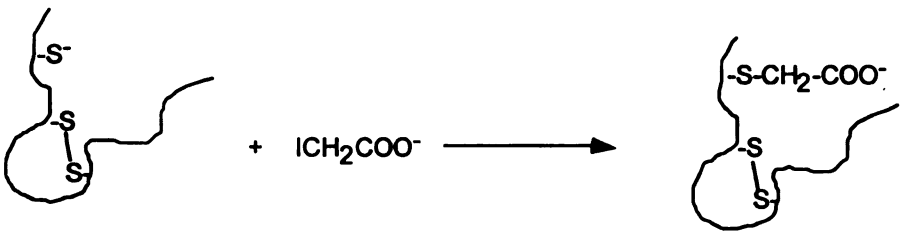
Equation 1.6. Trapping folding intermediates by acidification

Quenching by acidification is rapid and occurs at the diffusion control rate (10^9 $M^{-1}s^{-1}$) (41). Subsequent reversed-phase HPLC at pH 2 is well suited for the separation of acid-quenched intermediates without introducing significant rearrangements during the separation. A practical advantage of acid quenching is its reversibility. As a result, it is possible to purify an acid-quenched intermediate and subsequently to allow further rearrangement or folding to occur so that more detailed information of the folding pathway can be obtained (4, 42, 43). The folding intermediates of several proteins have been trapped by this method (4, 44-47). While trapping by acidification is rapid, it does not completely stop thiol/disulfide exchange; it just slows it down by reducing the concentration of the reactive thiolate anion. Assuming that the pKa of the thiol is 8.5, an intramolecular thiol/disulfide exchange occurring with a half-life of 1ms at pH 8.5 would

occur with a half-life of approximately 1600s at pH 2 (41). Therefore, samples must be analyzed promptly after the acidification and the intermediates trapped by acid must be chemically modified before the disulfide bond structure can be defined. In this respect, it is a cumbersome procedure.

2. Trapping by Alkylation

Folding intermediates can also be trapped by addition of an alkylation reagent, such as iodoacetate, iodoacetomide, and 4-vinyl pyridine (48-51). Iodoacetate is the most commonly used. The thiol groups of the intermediate will be alkylated by the alkylating reagent and further thiol/disulfide exchange will be stopped (Equation 1.7).



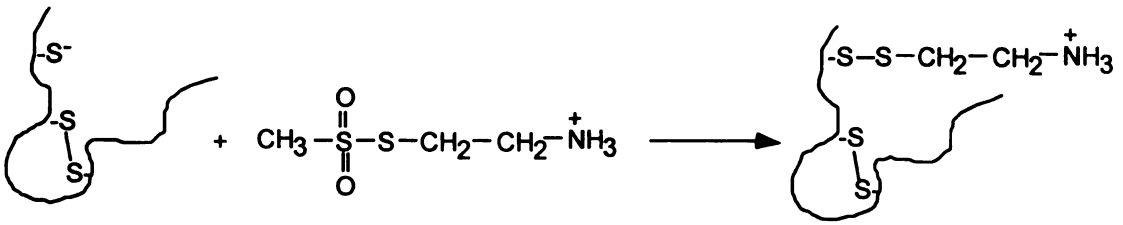
Equation 1.7. Trapping folding intermediates by alkylation

At a concentration of 0.2 M and pH 8.7, iodoacetate reacts with sterically accessible sulfhydryl groups with a half-life of 0.3 s (23). This is much faster than many of the intermolecular thiol/disulfide exchange reactions during oxidative folding; at 1mM GSSG, the half-life for reaction of GSSG with an exposed SH at pH 8.7 should be about 30 s. However, rearrangement of intermediates during trapping with iodoacetate has been observed for both BPTI (4) and ribonuclease (52). Quenching by 0.2 M iodoacetate is comparable to or, in some cases, slower than the rate of some of the intramolecular

rearrangements that interconvert intermediates of the same oxidation state (4, 53). When quenching by alkylation is slow compared to rearrangement, and the alkylating reagent reacts preferentially with one of the intermediates, alkylation will lead to overestimation of the amount of the intermediate that reacts fastest with the alkylation reagent. This could be a more severe problem for partially structured intermediates where steric hindrance would retard the rate of alkylation of some thiols (54). Although a high concentration of iodoacetate can be applied to mininize such a side reaction, modification of other functional groups by a high concentration of iodoacetate may provoke other problems (55).

3. Other Trapping Methods

Acid and alkylation trapping described above were attacked by Rothwarf and Scheraga (56, 57), who insisted that the criteria for a good blocking agent are that it block quickly, completely, and without modifying the protein at sites other than thiols. They proposed a new alkylating reagent, 2-aminoethyl methanethiosulfonate [$(NH_2)C_2H_5SSO_2CH_3$] (AEMTS) (Equation 1.8). AEMTS is a thiosulfonate reagent (58) which reacts with free thiols in the solution rapidly; even thiols that are essentially buried should still be accessible by local or global fluctuations of the protein within the 2 min blocking time (59).

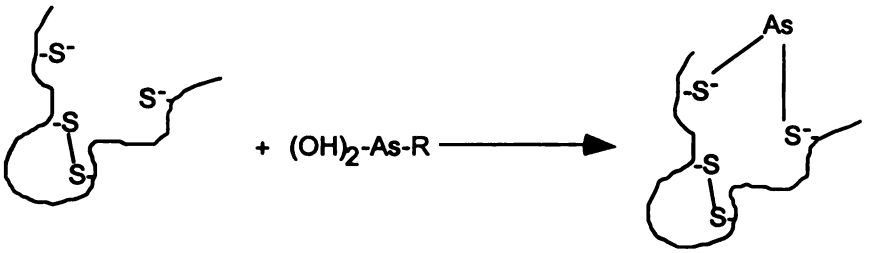


Equation 1.8. Trapping folding intermediates by AEMTS

There are several advantages of trapping by AEMTS (57). It can reversibly block the thiol groups in the intermediates so that it permits the refolding process to be restarted from isolated intermediates, an invaluable tool in the determination of refolding pathways. In addition, blocking with a reagent such as AEMTS provides a basis for separation of intermediates in a predicable manner due to the addition of a positive charge to a thiol group by AEMTS. It is especially important for very complicated refolding processes containing large numbers of chemically distinct species. Recently, Thannhauser *et al.* combined AEMTS trapping and amino acid analysis to quantitatively determine the number of cysteines and cystines in the folding intermediates of hirudin (60). While trapping by AEMTS is rapid and specific, disulfide scrambling is still questionable since the blocking of a thiol group by AEMTS is a thiol/disulfide exchange reaction.

Recently, a new trapping reagent was suggested to study the folding of macrophage colony stimulating factor (M-CSF) (61). The protein-folding intermediates were trapped by selective modification of bis-cysteine sufhydryls with phenylarsonous acid derivatives (PAA) (Equation 1.9). PAA-derivatives form cyclic dithioesters in the presence of two suitably spaced sulfhydryl groups of the intermediates. Although the trapping method

was found to have the advantage of a broad derivatizing pH range, it depends too much on the structural accessibility of bis-cysteine SH groups, therefore, it is limited in practical use.



Equation 1.9. Trapping folding intermediates by PAA

III. Recognizing the Location of Cysteines and Cystines

Practically, all the methods used to recognize the location of cysteine and cystine residues in proteins can be used for their location in the intermediates. Localization of cysteines and cystines in proteins is the oldest and best-studied field in protein chemistry. A large number of reagents and procedures have been developed and used to address this problem and there is no indication that the search for better methods is over yet.

A. Classical Approach

In the classical approach, one derivatizes free sulfhydryl groups with UV absorbing, radioactive or fluorescent groups under the condition to prevent sulfhydryl-disulfide exchange, and then cleaves the protein by chemical or enzymatic method. The resulting peptide fragments are separated by and collected from chromatography. The collected fractions showing specific UV, radioactive or fluorescent detection are recognized,

mapped to sequence by the Edman technique and /or mass spectrometry, and related to specific segments of protein.

A key to the success of the classical approach is to choose an appropriate derivatization reagent for sulfhydryl groups. The reagent should be soluble in the reaction medium and label sulfhydryl groups selectively, rapidly, and irreversibly under mild conditions (ideally under weak acidic conditions to avoid sulfhydryl/disulfide exchange which is minimized at pH 2-6.5). Furthermore, the reagent should possess strong UV or fluorescent absorption which does not overlap with the maximum absorption of proteins, or easily attach a radioactive element to facilitate the recognition of derivatized peptides after HPLC separation. Finally, the derivative of cysteine should be distinguishable from other amino acids by the Edman degradation technique in which the identification of amino acids exclusively relies on the retention time of PTH-derivatives.

Iodoacetate and pyridylethylation are the most extensively used alkylating derivatization methods for sequencing of peptides (62-68). Recently, Lee *et al.* (67) identified the position of the sulfhydryl group of human glucocerebrosidase by modifying the sulfhydryl group with [^{14}C] iodoacetic acid without reduction of disulfide bonds, digesting the protein with both protease-V8 at pH 4.0 followed by α-chymotrypsin at pH 7.5, and finally sequencing the peptide based on detected radioactivity. Krieglstein *et al.* (68) determined the location of sulfhydryl groups and disulfide bridges of botulinum neurotoxin type A. They converted five Cys residues to S-pyridylethyl cysteine residues; after mercaptolysis converted all nine Cys residues to the S-pyridylethylated form. After confirming the predicted number of Cys residues by amino acid analysis, the positions of the five Cys residues carrying sulfhydryl groups and the four involved in disulfide

bridges were determined by comparing the elution patterns in HPLC of the cyanogen bromide mixtures of the exclusively alkylated and the mercaptolyzed-alkylated neurotoxin. The chromatographically isolated components were identified by N-terminal amino acid sequence analysis.

While iodoacetate has the disadvantages of coelution with PTH-Gln and PTH-Glu on HPLC after Edman degradation of modified protein and makes identification problematic, pyridylethylation results the problem of N-terminal alkylation of protein because of the high reactivity of 4-vinylpyridine (69, 70). The other alkylating reagents have been developed for determination of cysteine residues (69-73). The alkylating reagent, 3-Bromopropylamine (72, 73), offers the advantage of easy identification of its derivative of cysteine residue by protein sequencing and amino acid analysis over other alkylating reagents in the modification and subsequent identification of cysteine residues by protein sequencing. Jue & Hale have developed the conditions for concomitant on-sequencer reduction and alkylation of cysteines in proteins with tri-n-butylphosphine and 3-bromopropylamine before sequencing the proteins (71).

While alkylation by iodoacetates continues to be useful, there has been far greater interest in the use of this chemistry as a mechanism for introducing a larger molecule which can serve as structural probe. Examples include the fluorescent reagents, 5-(iodoacetamido)fluorescein (IAF) (74) and 5-[2-((iodoacetyl)amino)-ethyl]naphthalene-1-sulfonic acid (1,5-IAEDANS) (75-77).

Bishop *et al.* has used IAF to selectively label two cysteine residues on the surface of Ca^{+}-ATPase of sacroplasmic reticulum (SR), digest with trypsin, purify the tryptic peptides by size-exclusion and reverse-phase HPLC, and identify the cysteine residues by

amino acid sequencing (78). Recently, a combination of 1,5-IAEDANS labeling of cysteine residues, enzymatic digestion, and MALDI-PSD has been developed to assign the location of free cysteine residue of testis angiotensin-converting enzyme (tACE) (79). The free cysteine residue was located by labeling with 1,5-IAEDANS, isolating the fluorescent peptide from enzymatic digests by HPLC, and analyzing its sequence by MALDI-PSD.

Labeling with IAEDANS or IFA suffers from less selectivity and reactivity toward sulfhydryls, as well as the fact that the reagents themselves are fluorescent before reaction with thiols. A highly reactive and specific reagent for protein thiols, 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD-F) has been developed (80). The selectivity of ABD-F is superior to that of other sulfhydryl-reactive fluorophores. In addition, unlike most other reagents previously utilized, ABD-F is nonfluorescent before reaction with protein thiols. The ABD-Cys adduct absorbs maximally at 385 nm and has a fluorescence emission maximum at 520 nm, allowing simple and sensitive detection in the presence of an excess of unlabeled protein.

ABD-F has been utilized for analysis of cysteine residues in the proteins (81-83). The sulfhydryl and disulfide bond locations in the proteins were established using tributylphosphine (Bu_3P) as the reducing reagent and ABD-F as the sulfhydryl reagent (83). The reaction was first run in the absence of reducing reagent, and all the peaks, both fluorescent and nonfluorescent ones were collected from HPLC of the tryptic peptides. An aliquot of each of the nonfluorescent peaks was then treated simultaneously with Bu_3P and ABD-F since the two reagents do not react with each other. Sequencing the fluorescent peptides established the positions of the cysteine residues and disulfide bonds.

The ability to reduce disulfides and block the resulting thiols simultaneously in a single reaction, the stability of ABD-Cys to acid hydrolysis, and the recovery, albeit in low yield, of PTH-(ABD-)Cys on sequencing, along with the characteristic fluorescence of ABD-thiol derivatives, make it a favorable methodology.

B. Affinity Chromatography

The sulfhydryl group is the only functional group in a protein for which conditions are available for the formation of a stable covalent bond which can be split under mild conditions (84). A thiol-containing proteins can be attached on an affinity column. After the immobolized protein is cleaved chemically or enzymatically on the affinity column, the nonsulfhydryl-containing peptides do not bind to the column and can be washed away. The attached sulfhydryl-containing fragments are then removed from the column and subjected to further structural characterization by Edman degradation and/or mass spectrometry. The affinity interactions that can be used for the sulfhydryl-containing proteins include (85): (i) binding to affinity media that contain heavy metals, (ii) binding to affinity media with reactive disulfides that undergo facile thiol-disulfide exchange, and (iii) binding to affinity media that contain chelated zinc. While these methods were originally used to purify sulfhydryl-containing proteins, this approach was recently used to locate cysteine residues in proteins by mass mapping of bound sulfhydryl-containing peptides (86). This approach is powerful for locating sulfhydryl groups in high-mass proteins because the HPLC separation of the digests is unnecessary. But conditions must be established and optimized under which all the sulfhydryls can be attached to the column.

C. Cleavage at Cysteine Residue

The cleavage of the peptide bond at cysteine residue after cyanylation was first introduced by Catsimpoolas and Wood (87). In this reaction, the conversion of an SH group into an SCN group was achieved in two stages by successive treatment of a protein with Ellman's reagent and then with cyanide. Jacobson and Stark (88, 89) showed that 2-nitro-5-thiocyanobenzoic acid (NTCB) specifically cyanylates cysteine sulfhydrls, which subsequently cleave at N-terminal side of the cyanylated cysteinyl residues under mildly alkaline conditions to form an amino-terminal peptide and a series of 2-iminothiazolidine-4-carboxylyl (itz) peptides. Walkselman and Guibe-Jampel proposed another cyanylating reagent, 1-cyano-4-dimethylamino-pyridinium (90). This reagent has the advantages of water solubility and reactivity at acidic condition.

The cyanylation/chemical cleavage reaction of cysteine residue can be utilized to recognize structural information for the protein. If a protein contains **n** cysteine residues, the cleavage reaction results in the formation of **n+1** peptide fragments, mass alignment of which indicates the number and location of cysteine residues. While SDS-PAGE can be used to mass map the peptide fragments (91-93), it suffers the poor mass accuracy (error > 5%) (89, 94). Daniel *et al.* used electrospray mass spectrometry to determine the masses of the NTCB cleavage products of *E. coli* dihydroorotase to identify the location of the active cysteine residues in the protein (95).

Recently, Wu, *et al.* (96) combined the cyanylation reaction, specific chemical cleavage at cysteine residues, and MALDI-MS to quantitatively determine the number and location of cysteine and cystine residues in the proteins. This method has the advantages of being precise, simple and highly sensitive. However, MALDI-MS has a

signal suppression phenomenon which may result in incomplete assignments; the cyanylation reaction with NTCB is performed at alkaline condition, at which sulfhydryl-disulfide exchange may proceed for some of the proteins.

IV. Assignment of Disulfide Bonds

Disulfide bonds between cysteine residues, to generate cystine, are a common post-translational modification of proteins, principally those that operate in the extracellular milieu. A full description of the covalent structure demands that the connectivity of the bridged cysteines be analyzed. Although the amino acid sequences of proteins can be deduced from the corresponding cDNA sequences, the state and connectivity of the cysteine residues after the translation cannot be predicted accurately. Characterization of disulfide linkages in proteins has become an essential part of the analysis of recombinant DNA products, because molecules having incorrect disulfide linkages may have much lower activity than that of the desired product. Therefore, it is important to verify that recombinant materials have the correct disulfide linkages.

Although disulfides are chemically rather simple, two factors can make the analysis difficult. First, the number of possible isomers grows rapidly as the number of bridges increases—1, 3, 15, 105, 945, etc. Second, base and reducing agents can catalyse exchange among partners, obscuring the original connectivity (97). Native proteins under physiological conditions are deceptively stable, but once they are denatured or nicked by proteases, disulfide exchange occurs when even a trace of thiol is present.

A. Conventional Fragmentation Strategy

The general strategy for locating disulfide bonds in proteins by conventional methods involves several steps:

(1). Digest proteins by chemical reagents and/or enzymes between the half-cystinyl residues avoiding disulfide exchange.

(2). Separate fragments also under conditions that stabilize the disulfide bonds.

(3). Detect disulfide-containing peptides.

(4). Identify subfragments, and hence individual cysteine residues that are connected.

The conventional fragmentation approach has been well established for recognizing disulfide bond structure of proteins (98-100). The characteristic aspects of the strategy will be discussed below.

B. Fragmentation of Proteins

The aim of fragmentation of a protein for disulfide mapping is to produce by enzymatic or chemical digestion a definitive set of peptide fragments only containing one disulfide bond. If a protein is not cleaved between every half-cystinyl residue by an enzyme or a chemical reagent, the peptides obtained by the cleavage must be cleaved again using another enzyme or chemical reagent. It is easy to identify the peptide generated by highly specific cleavages of proteins (101-103), because the N- or C-terminal residues of the peptides are known. The general risk of disulfide exchange occurs during digestion; even though the protein, enzymes, and reagents may be initially free of thiols, the latter can arise by base catalyzed degradation of disulfides. Under any reasonable conditions very little thiol will be produced, but only catalytic amounts are

needed to cause extensive disulfide rearrangement during long incubations. Therefore, the choice of a proper enzyme should satisfy three requirements: the specificity of the enzyme; the capability of cleaving between every half-cystinyl residue; and the conditions to minimize disulfide scrambling.

Cyanogen bromide, trypsin, and *Staphylococcus aureus* V8 protease are available to cleave proteins specifically. Specific chemical cleavage at the C-terminal side of Met by cyanogen bromide (CNBr) has been widely applied for initial cleavage of proteins due to its weak acidic reaction condition and volatile reagent and by-products. Since Met is not a particularly abundant constituent of proteins, cleavage with CNBr usually gives large peptides. Subsequent digestion by other enzyme(s) is required to produce smaller peptides (101). Trypsin has also been used frequently because it cleaves on the C-terminal side of Arg or Lys residues with high specificity. Unfortunately, trypsin has maximum activity at pH 8.3, and is not active in acid. As a result, disulfide scrambling may occur during enzymolysis. Despite this shortcoming, trypsin is still the most widely used enzyme for the digestion of proteins because of its high specificity. Greatly increased use is now being made of the V 8 protease from *S. aureus*. This enzyme specifically cleaves on the C-terminal side of Glu or Asp residues in the pH range of 4.0 to 7.8.

Unfortunately, if proteins cannot be cleaved by specific cleavage reagents, they must be cleaved by non-specific reagents. Pepsin, thermolysin and partial acid hydrolysis are useful for this purpose. The major advantage of using pepsin and partial acid hydrolysis to generate peptides is that disulfide bonds in proteins are fairly stable under mild acidic conditions (104, 105). However, the protein digests using such nonspecific

cleavages are complex mixtures, which make the separation and identification much more difficult.

C. Purification of Cystinyl Peptides

The purification of disulfide-bonded peptides constitutes a special problem. The difficulties in this area relate to the instability of disulfide bonds. Under the condition of mild acidity, i.e, pH 2.0-6.5, the disulfide bond is relatively stable. This requirement restricts the choice of isolation techniques. Reversed-phase HPLC under acidic conditions (a gradient of acetonitrile in 0.1% trifluoroacetic acid, for example), is ideal for this purpose, being both rapid and employing conditions favoring stability of disulfide bonds. However, the purification of digests from high-mass proteins is still a challenging task even though microbore or capillary HPLC columns are use because so many peptide fragments might be produced after hydrolysis.

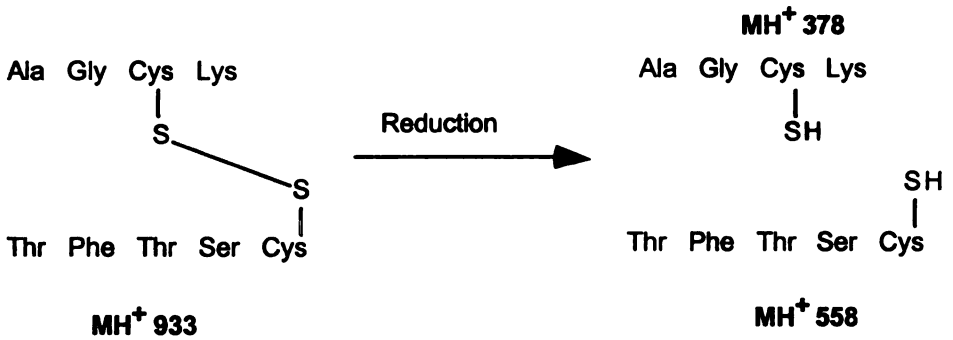
Gel filtration and ion-exchange chromatographic technologies are applicable for purification of digestion mixtures. The techniques are based on the chemical modification of cysteinyl residues to cysteic acids after the reduction of disulfide bonds in cystine peptides. Non-cystine peptides may be separated with great ease from the more highly charged cysteic acid peptides generated from cystine peptides.

D. Identification of Cystine-Containing Peptides

The conventional approach for the identification of cystine-containing peptides after the purification is very tedious and time consuming. The peptides, which may contain

intra- or intermolecular disulfide bonds, are reduced, alkylated, isolated by HPLC again, and identified by their sequence with Edman degradation (106, 107).

The soft ionization methods in mass spectrometry have proven useful for locating disulfide bonds in proteins. One of the methods is by fast atom bombardment (FAB-MS) (108-112). Disulfide-containing peptides suitable for analysis by FAB-MS can be produced by treating proteins with enzyme. An aliquot of the digest can be analyzed by FAB-MS to determine the molecular weights of the peptides which are related to specific segments of the native protein; then following reduction or oxidization of the disulfide bonds with chemical reagents, such as the reducing reagent dithiothreitol or the oxidizing reagent performic acid. The sample is analyzed again by FAB-MS. The disulfide assignment can be completed by comparing the two FAB-MS results. An example is shown in the following scheme:



This method soon became an accepted method and was successfully applied to model proteins such as insulin (109), hen egg-white lysozyme and bovine ribonuclease A (111), and duck egg-white lysozyme (110). However, it needs specific enzymes to cleave the protein to form peptides only containing one disulfide bond, and the fragments should

not be large since FAB-MS is not suitable for high mass analysis, and it cannot be used for close or adjacent cysteine analysis.

The another MS method is tandem mass spectrometry (MS/MS) employing high-energy collision-induced dissociation (CID) or low-energy CID, which produces peptides by enzyme digestion. The peptides containing a disulfide bond are selected as the precursor ion which is sequenced by MS/MS, and after reduction the peptides are sequenced again by MS / MS; the disulfide linkage can be assigned by comparing the two results (113-117).

Electrospray ionization (ESI) (2) and matrix-assisted laser desorption ionization (MALDI) (3) have revolutionized the role of MS for the analysis of proteins and peptides. Peptide-mapping by ESI (118, 119) or MALDI-MS (120, 121) are being used increasingly in the characterization and identification of disulfide bonds. The strategy used here is exactly the same as in FAB-MS. However, they provide much more sensitivity (low picomole to high femtomole). Besides, the easy interface of ESI to HPLC has made the purification of the disulfide-containing peptides an easy task. MALDI-MS provides the capacity of rapidly analyzing small quantities of mixtures of proteins and peptides. It is particularly well suited to the determination of peptide mass mapping without the previous isolation or fractionation.

Recently, MALDI-MS with postsource decay was used to generate fragment ions from peptide fragments containing heteropeptides linked together by two disulfide bonds (122). Postsource decay analysis of these peptide fragments generates a series of singly charged fragment ions that, in addition to the peptide sequence ions, provide useful information for assigning disulfide arrangement in highly bridged disulfide-linked

peptides. The assignment was made possible by fragmentation at peptide bonds between two Cys residues in a peptide that constitutes the highly bridged fragment, while retaining the disulfide linkage to the other peptide. Fragmentation using other types of instruments, such as quadrupole ion-trap mass spectrometry with CID dissociation, usually did not generate such fragment ions. The method was used to facilitate the assignment of disulfide structure in a tumor necrosis factor binding protein, which contains 162 amino acids and 13 disulfide bonds.

The conventional methodology requires the cleavage between half-cystinyl peptides by enzyme(s). If proper enzyme(s) cannot be find for this purpose, partial hydrolysis by acid can be used (105). Partial acid hydrolysis is performed under controlled conditions in which only limited peptide fragments are formed. Partial acid hydrolysis is attractive as disulfide bonds are particularly stable in dilute or weak acids. However, it generates extremely complex digests. Unless one is dealing with a relatively small molecule, the problem is almost intractable without complicated chromatographic separation and mass spectrometry technique.

E. S-S Assignment of Adjacent Cysteines

The major problem of the conventional methodology is that it is unlikely to find a cleavage reagent that can cleave between the two adjacent or close spaced cysteine residues and produce peptide fragments that only contain one disulfide bond. In this case, alternative methodologies should be used.

Different approaches have been reported to solve the problem, which include the combination of organic synthesis of the possible isomers of a disulfide-containing peptide

and high-performance tandem mass spectrometry (123) and multiple steps of Edeman degradation (124-129).

The selective reduction of disulfide bonds can be applied to the problem (130-133). The selective reduction is based on the observation that disulfide bonds in a protein can often be reduced sequentially to thiols. The disulfide bonds of the protein are reduced stepwise by controlling the reduction conditions (temperature, concentration of reducing reagent, time, and denaturing conditions). The thiols so formed are labeled with a reagent, such as alkylating and fluorescent reagents. The protein is then reduced completely, and subsequently cleaved enzymatically. The location of the disulfide bond is determined by isolating and identifying the labeled peptides. The selective reduction only can be applied for the proteins containing disulfide bonds that have different stability.

An alternative method for analyzing highly bridged proteins is based on the partial reduction of disulfide bonds by tris-(2-carboxyethyl)-phosphine (TCEP) at pH 3, purification of the reduced protein by HPLC, alkylation of free thiols, analysis by sequencing, providing explicit assignment of disulfide bonds that had been introduced (134,135). This method has the advantages of being suitable for analyzing adjacent cysteines and tightly clustered cysteines, and preventing the sulfhydryl-disulfide exchange when performing partial reduction. The disadvantage is the limit of the analysis of the alkylated protein by sequencing. It is time consuming and hard to analyze large proteins.

Recently an approach using a combination of the partial reduction by TCEP and cyanylation of free thiols by 1-cyano-4-dimethylamino-pyridinium (CDAP)

tetrafluoroborate at acidic condition was reported (136). Following the cleavage of the peptide bond on the N-terminal side of cyanylated thiols, the mass mapping of peptide fragments was performed by MALDI-MS. This method has the advantages of minimizing disulfide exchange, the possibility to analyze the proteins containing close or adjacent cysteines, and being fast, simple, and sensitive.

V. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS) and Electrospray Mass Spectrometry (ESI-MS)

In the last twenty years, important achievements in organic mass spectrometry have been obtained by the development of a number of new ionization techniques for the analysis of large, polar and thermally labile compounds. The general problem to be solved was how to generate intact gas-phase ions of the intact molecule from compounds that are usually solids and degrade or decompose when being exposed to thermal heating. Progress was achieved with the discovery of the desorption ionization techniques; plasma desorption mass spectrometry (PDMS) (137) and fast atom bombardment mass spectrometry (FAB MS) (138). The introduction in 1988 of two new techniques, matrix-assisted laser desorption/ionization (MALDI) (3) and electrospray ionization (ESI) (139), has extended the applicability of mass spectrometry in the biological science far beyond the scope of PDMS and FAB-MS. Although MALDI and ESI utilized very different ideas and principles, both new techniques have overcome the main difficulties with MS methods, i. e., the desorption and ionization of large and labile biomolecules, and shown considerable promise in characterizing biomolecules by accurate determination of

molecular mass up to a few hundred thousand daltons. In combination with biochemical techniques, numerous applications have been achieved (1).

A. Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS)

The first attempts to make use of laser ionization for the analysis of organic compounds date back to the 1970s. Especially, the desorption of ions containing the intact molecule of complex compounds, like stachyose and digitonin (140), may be regarded as a milestone. However, significant dissociation occurs in larger molecules because of the higher energy needed to desorb the large molecules, and the energy goes directly into the analyte. The major advance in laser desorption was the introduction of matrices (MALDI) by Hillenkamp and Karas in the late 1980s (141). The key to the MALDI process is to embed the analyte macromolecule in a high molar excess of a suitable matrix of molecules having a strong absorption at the laser wavelength. This induces an efficient transfer of the laser-pulse energy to the analyte and results in a soft desorption process, i. e., little or no fragmentation. The matrix also provides photoexcited acid or base sites for ionization of sample molecules during ion/molecule collisions (142).

Although the mechanism of desorption and ion formation is complicated and unresolved, the formation of ions to nearly 10^6 Da and the high sensitivity of MALDI (subfemtomole) has led to the explosive commercialization of instruments in not only mass spectrometry laboratories, but also in biochemical facilities and laboratories where bench scientists are able to use the instrumentation.

1. Matrix and Sample Preparation

The matrix is the key to the successful analysis of high-mass biomolecules. While several hundred of organic compounds have been investigated (143, 144), only a few of these are widely applicable for peptide and protein analysis: α-cyano-4-hydroxycinnamic acid (4-HCCA), 1,5-dihydroxyoxybenzoic acid (DHB), and 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) (145). It was considered important that the laser wavelength match the desorption maximum of the matrix compound. All the matrices above have strong UV absorption in the 320-350 nm range and can be used with a much cheaper nitrogen laser (337 nm) or frequency-doubled Nd-YAG laser (355 nm). Other important matrix characteristics include miscibility with the analyte in the solid phase, solubility in the same solvents required for the dissolution of the analyte, vacuum compatibility (low vapor pressure), a chemical composition that promotes the ionization of matrix substituents that can donate protons to the analyte, nonreactivity with the analyte, and other physical properties such as the low heat of sublimation and a capacity to crystallize readily (146).

A number of different methods have been developed for growing analyte-matrix crystals. The original method described by Karas and Hillenkamp has been named the 'dried-droplet' method (141). It entails drying a droplet of a solution containing the matrix (1-10 mmol) and the protein (1-10 μmol). A simple variant on the "dried droplet" method was proposed by Vorm, *et al.* (147). They first placed a thin layer of the matrix compound onto a metal surface and then put a droplet of the analyte containing solution on top of the layer of matrix compound. As the droplet dries it dissolves some of the matrix which crystallizes onto the substrate matrix layer when the droplet dries

completely. The crystallized matrix layer is very thin, allowing for more reproducible mass measurements and producing ions for < 10 laser shots on a spot.

It is possible to grow large, analyte-matrix crystals under near equilibrium condiitons, rather than a rapidly drying droplet (148). Supersaturated matrix solutions containing analyte will form crystals that can be used directly in an ion source. Supersaturation can be achieved either by heating and cooling or by slow evaporation.

A recently developed technique of producing crystals for use in MALDI ion sources involves growing a thick film of matrix crystal (149). The substrate for the crystals (usually a piece of metal) is first covered with the matrix material, by rapid drying of a solution of the matrix dissolved in an organic solvent. The small matrix crystals that cover the surface are then crushed and smeared over the substrate surface, resulting in a well-adhered layer of stressed, crystalline matrix. A saturated solution containing the matrix and analyte is then placed onto the surface and allowed to dry slightly. The stressed crystalline material on the substrate acts to seed crystal formation at many sites on the surface, resulting in the rapid growth of a rather uniform polycrystalline film of analyte-matrix crystals over the surface. The liquid drop can then be removed by blotting. The resulting film is very strongly adhered to the substrate and it can be washed thoroughly to remove any contaminants that were present in the analyte solution.

The mechanism of "matrix assistance" for the transfer of analyte molecules to the gas phase and their ionization remains incompletely understood; the choice of matrix and method application is still empirical. Sample preparation by different matrices and solvents (150), matrix additives (151), and evaporation rate (147, 152) affects the

resolution and sensitivity of MALDI. Optimal results require parallel analyses under different conditions.

2. Instrumentation

To date, MALDI of large molecules has been obtained mainly in combination with the field-free time-of-flight (TOF) mass analyzer. A schematic drawing of MALDI-TOF instrument is shown in Figure 1.2.

The solid deposit of sample-matrix is irradiated by a pulsed laser, the most common one is the Nd-YAG laser, which is directed and focused by a prism and optical lens. The irradiance is controlled by an attenuator and is increased gradually until the threshold is reached. The interaction of photons with the matrix and analyte molecule results in the desorption and ionization of the co-crystallized sample/matrix from a metal surface. The initial kinetic energy spread of ions generated by MALDI is large, so either a linear TOF with a high accelerating voltage or a reflectron with an ion mirror (or both) is used to improve mass resolution. Experimentally, a static electric field is imposed upon ions generated from the sample by application of a two-stage high voltage (typically ± 25KV) to the sample probe with regard to a closely spaced accelerating electrode. The ions are thus accelerated to the same kinetic energy (assuming the initial kinetic energy of all ions is zero) by the electric field toward a long (1-2 m) field-free time-of-flight (TOF) analyzer.

The time-of-flight mass analyzer is simple, cheap, and well-suited to the pulsed nature of laser desorption in MALDI (153). It has virtually no upper mass range and is therefore compatible with MALDI, which can produce very high m/z ions. Another

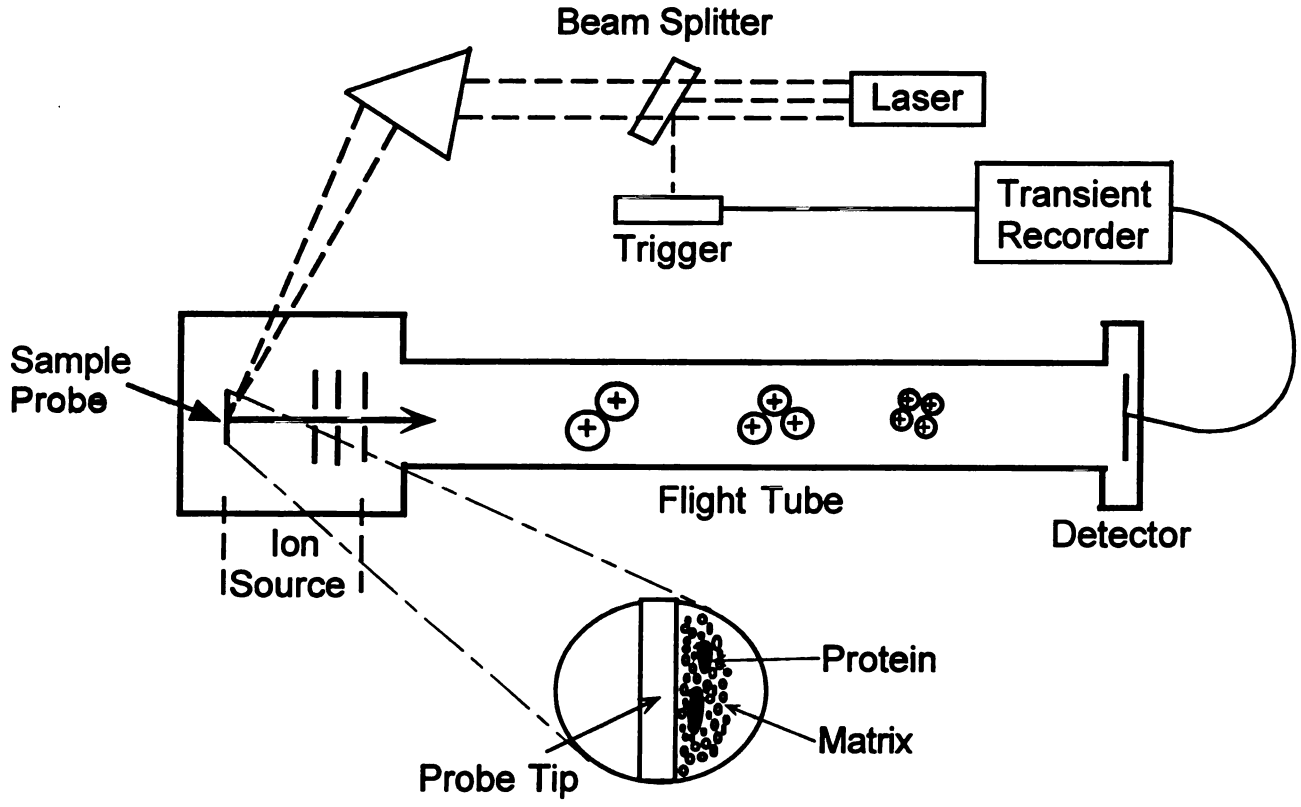


Figure 1.2. MALDI-TOF-MS instrument

advantage of TOF mass analyzer is its capacity to generate the entire spectrum from every single laser shot without losing information, in contrast to a scanning mass analyzer.

The time required for ions to traverse the flight tube, tof, is dependent on their massses and is described by the relationship: $tof=L/v=L(m/2z\ eV)^{1/2}$, where L is the length of flight tube, v is the ion velocity, m is the mass of the ion, and V is the acceleration potential. Thus, low-mass ions have a shorter flight time than heavier ions. They are separated into a series of spatially discrete individual ion packets, each traveling with a velocity characteristic of its m/z ratio. A detector positioned at the end of the field-free flight-tube produces a signal as each ion packet strikes it. A recording of the detector signal as a function of time yields a tof spectrum. The difference between the start time, triggered by the laser pulse and common to all ions, and the arrival time of an individual ion at the detector is proportional to $(m/z)^{1/2}$ and can be used to convert the x-axis of the spectrum (ion arrival time, tof) into a m/z ratio axis (a conventional mass spectrum). The TOF analyzer is highly efficient because all ions of different m/z ratio arising from a single laser shot are measured; they simply arrive at the ion detector at different times. However, the tof recorded for a particular ion reflects many different initial conditions experienced in the ion source (such as the time/location of ion formation and initial kinetic energy distributions). The poor resolving power is reflected by peak broadening which increases with the mass of the ionized macromolecules. This limits the capacity to detect certain protein modifications and protein sequence variations, especially at high mass. In spite of this, it is possible to achieve a mass accuracy of 0.1-0.01% for proteins

with molecular masses between 1 and 40 KDa, and with somewhat poorer accuracy for proteins above 40 KDa (153).

Two techniques were introduced to solve the low-resolution problem of linear TOF mass analyzer, reflectron-TOF (reTOF) (154) and delayed ion extraction (DE) (155). The reflectron-TOF consists of two linear field-free region and an ion mirror, which compensates for different, nonzero initial kinetic energies of ions of the same m/z. Delayed ion extraction allows the ions to disperse in the source region (acceleration region just above the surface) owing to their initial velocity, while the density of neutrals is decreased by pumping them away. The number of ion-molecular collisions before extraction of the ions into the field-free drift region is therefore reduced, which decreases the width of the translational energy distribution. Simultaneously, ions having higher initial kinetic energy move further away from the extraction field in the source region and thus are given less "kick"; those of lower kinetic energy experience a higher extraction field. This approach has proven so successful that limitations in mass resolution are now focused on detection systems (the time resolution).

B. Electrospray Ionization Mass Spectrometry (ESI-MS)

Electrospray ionization-mass spectrometry (ESI-MS) has its origins in research that long preceded the current flurry of activity. The study of the electrospray phenomena extends back perhaps over two and one-half centuries to the work of Bose (156), and certainly to that of Zeleny early in this century (157). The seminal research into the use of electrospray as an ionization method for macromolecules was due to Malcom Dole and co-workers (158, 159), who performed extensive studies into the electrospray process and

defined many of the experimental parameters. Experimental evidence was presented by Dole for ionization of zein (Mr. ~50,000) (160) and lysozyme (161). However, interpretation of their results was problematic because a mass spectrometer was not available, and only ion retardation and ion mobility measurements were obtained.

In 1984, ESI combined with mass spectrometry was first reported, essentially simultaneously, by both Yamashita and Fenn (162) and Aleksandrov *et al.* (163, 164). In 1988, Fenn and co-workers (165) first reported ESI-MS spectra of intact multiply protonated molecules of proteins up to Mr 40,000. Since its introduction in 1988 as a tool for the ionization of large biomolecules, the practice of ESI-MS in medical, biotechnological, and pharmaceutical arenas has grown impressively. One obvious reason for this is the speed with which commercial instrumentation became available, and particularly the ease of retrofiting existing quadrupole mass spectrometers for ESI-MS. As evidenced by the growing number of ESI-MS related publications, routine analysis by ESI-MS is becoming increasely common in basic research, quality control, and bioanalytical laboratories.

1. The Electrospray Process

Electrospray ionization requires continuous flow of solution (generally from a small diameter capillary tube) in the presence of a high electric field. Liquid is expelled from the tube as charged droplets which have a charge-to-volume ratio that can approach the physical maximum allowed for droplet stability, i. e., the Rayleigh limit (166), which by definition is reached when droplet surface charge density exceeds the liquid's surface tension (167). Droplet instability is manifested in a "coulomb explosion" or "droplet

fission" event. It is generally believed solvent is lost through evaporation until Coulomb explosions produce smaller droplets. The small droplets evaporate rapidly and the protonated molecules are released into the gas phase. It is unclear whether ions escape from droplets (i. e., field ionization) or solvent evaporates to leave ions (i. e., droplet evaporation), and the two mechanisms will likely be experimentally undistinguishable. In contrast to all other mass spectrometric ionization methods, the change from the liquid into gas phase is very gentle and probably occurs with at least one solvation shell still surrounding the analyte molecule. Fragmentation during this 'desorption event' has hardly been observed, and it is now thought that even aspects of the tertiary structure of a protein survive into the gas phase, at least under certain circumstances (168).

2. Instrumentation

The generic representation of an ESI source shown in Figure 1.3 incorporates several features that are found on many ESI sources. The electrospray capillary typically consists of a small i.d. (<100 μm) metal or glass capillary biased at ±1-5 KV (positive potential for cations, negative potential for anions) relative to the desolvating capillary. Recent studies suggest that significant improvements in sensitivity can be realized with very small diameter (<20 μm) electrospray capillaries and the use of very low flow rates (169). A solution containing the analyte of interest is forced into the electrospray capillary at a flow rate of 0.01-10 μl/min (depending on analyte concentration and solution conditions) and is nebulized by the resulting electrospray plume. A common solvent for the analysis of peptide cations is 50/50 MeOH/H_2O (or acetonitrile/H_2O) with 1-2% HOAc (or formic acid) and a protein/peptide concentration of 10^{-5} – 10^{-7} M.

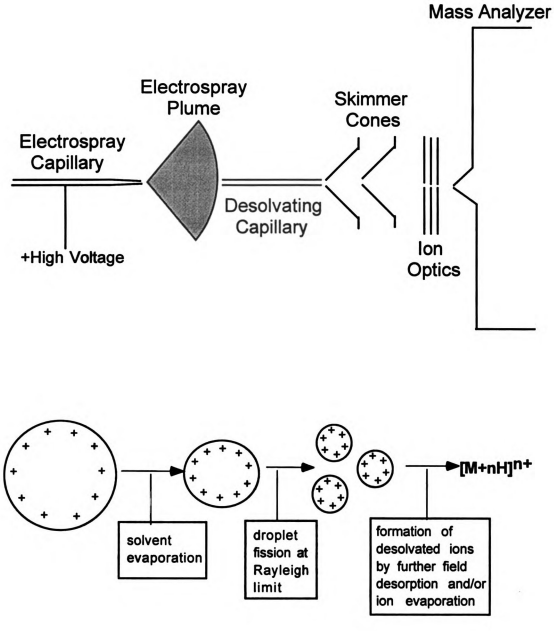


Figure 1.3. ESI-MS instrument and droplet production in ESI interface

Electrospray devices also are used to couple separation techniques, including capillary electrophoresis (170) and high performance liquid chromatography (171) to mass spectrometry. Direct coupling of the separation technique and the ionization technique is accomplished in a number of ways including the addition of a conductive sheath liquid or the application of an electrically conductive coating to the terminus of the separation capillary (171).

It has been recognized that the electrospray ionization process is driven by droplet desolvation and that desolvation at the molecular level is essential to obtain high quality spectra. Thus, a desolvation capillary often is added to expedite desolvation of the highly charged droplets generated in the electrospray plume. This capillary typically has an internal diameter of 400 μm to 1 mm, and a length of 10-25 cm. Glass inlet capillaries are employed widely in conjunction with a countercurrent flow of heated, dry nitrogen to enhance the rate of droplet desolvation and prevent large amounts of material from entering the spectrometer. Both ends of the glass capillary are gold coated and electrically biased to define a finite potential difference between the capillary ends and the neighboring source elements.

Next, a skimmer cone, or more commonly, a series of skimmer cones with a 300μm to 1.5mm diameter aperture is used to transmit to the mass spectrometer as many ions from the sample as possible while decreasing the pressure in each subsequent vacuum stage of the chamber. An electrical bias on the skimmer cone(s) serves to focus the ion beam and defines the average kinetic energy of the ion beam. Ion transmission from the source to the mass analyzer is enhanced greatly by the addition of a few simple ion lenses that guide and focus the ion beam into the mass analyzer.

In conjunction with electrospray ionization a quadrupole mass spectrometer is commonly used as a result of relatively low cost and ease of interfacing with many commercial systems. A quadrupole mass analyzer consists of four equidistant rod electrodes arranged as two pairs to which time-varying ac and dc potentials are applied. Only ions having a mass-to-charge ratio within a very narrow m/z range are transmitted by the quadrupole. Ions having a m/z outside this range move in unstable orbits and collide with the electrodes. Ions striking the detector transfer charge that registers as a current that is proportional to ion abundance and is amplified subsequently. Correlating ion abundance values with the values of applied dc bias and amplitude of ac voltage gives a mass spectrum. The relatively high sensitivity and rapid scan rates obtainable with commercial quadrupole mass spectrometers resulted in their widespread use.

3. Interpretation of ESI-MS Spectra

As a majority of ionization techniques today produce single charged ions (and, thus, mass spectra in which the measured m/z is equivalent to the mass of analyte), the multiple-charge states common to ESI-MS spectra might seem a bit confusing. However, their interpretation often is straightforward once the appropriate procedure is applied.

As first demonstrated by Mann, Meng, and Fenn (172), a simple algorithm can be applied to single component spectra based on the assumption that each peak differs from its neighboring peak by only charge. The value of the charge state of a selected lower m/z peak (z_{low}) is determined according to the equation: $Z_{low} = [(m/z)_{hi}]/\{[(m/z)_{hi}-(m/z)_{low}]/n\}$, where n is the number of peaks between the selected lower m/z peak $[(m/z)_{low}]$ and a selected higher m/z peak $[(m/z)_{hi}]$ being considered. From the known charge state of a

single peak calculated by the equation, the charge states of all the remaining peaks are directly calculated by simple addition or subtraction. From measured m/z and the calculated charge z, the ionic mass, m_i, is obtained easily, $m_i = m/z \times z$. The mass of the corresponding neutral analyte, Mr: $Mr = [(m/z) \times z] \pm nMa$, Ma is the mass of charged adduct. As each charge state yields an independent mass measurement, mass accuracy often is improved considerably by averaging over the entire charge envelope: $Mr = (1/N) \Sigma_i Mr$, where ${}_iMr$ is the calculated molecular weight from a single charge state and N is the number of charge states being averaged.

C. The Comparison of MALDI and ESI

While MALDI and ESI are widely used techniques in analysis of large biomolecules, they have different and complementary natures (Table 1.1), which has made it increasingly important that researchers have access to both types of instrumentation. Independently, ESI and MALDI-MS can help answer many questions; yet together they present a formidable research tool with new levels of sensitivity, accuracy, and mass range. Their utility for mass measurement meets the needs of chemists and biologists alike, facilitating routine characterization in small molecule synthesis, protein synthesis, and compounds obtained directly from biological matrices. The use of ESI and MALDI mass spectrometry extends beyond simple characterization. Noncovalent interactions, protein and peptide sequencing, DNA sequencing, protein folding, *in vitro* drug analysis, and drug discovery are among the areas to which ESI and MALDI-MS have been applied.

Table 1.1. Comparison of ESI-MS and MALDI-MS

	ESI-MS	MALDI-MS
Mass limit, Da (practical)	~200,000 (70,000)	>3,000,000 (150,000)
Advantages	HPLC/MS capable Multiple charging Capable of observing noncovalentcomplexes directly from water Femtomole-to-picomole sensitivity	Tolerant of mM concentrations of salts Highest mass capability Tolerant of mixtures Femtomole sensitivity Being developed as a tool for sequence analysis
Disadvantages	Multiple charging can be confusing with mixtures Typically need < mM salt concentrations for good signals Not tolerant of mixtures	Typically low resolution (<500) linear TOF without DE Not amenable to LC/MS
Suitable compounds	Peptides Proteins Carbohydrates Nucleotides Oligonucleotides Phosphoproteins Small chargeable molecules	Peptides Proteins Glycoproteins Carbohydrates Nucleotides Oligonucleotides Phosphoproteins Small chargeable molecules Heterogeneous samples

VI. References

1. Burlingame, A. L., Boyd, R. K., and Gaskell, S. J., *Anal. Chem.*, 68, 599R (1996).
2. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M., *Science*, 246, 64 (1989).
3. Karas, M., and Hillenkamp, F., *Anal. Chem.*, 60, 2299 (1988).
4. Weissman, J. S., Kim, P. S., *Science*, 253, 1386 (1991).
5. Pain, R. H., *Trends Biochem. Sci.*, 12, 309 (1987).
6. Scheraga, H. A., Konishi, Y., Ooi, T., *Adv. Biophys.*, 18, 21 (1984).
7. Kim, P. S., and Baldwin, R. L., *Annu. Rev. Biochem.*, 59, 631 (1990).
8. Matthews, C. R., *Annu. Rev. Biochem.*, 62, 653 (1993).
9. Ptitsyn, O. B., *Curr. Opin. Struct. Biol.*, 5, 74 (1995).
10. Creighton, T. E., *J. Mol. Biol.*, 87, 563 (1974).
11. Creighton, T. E., Goldenberg, D. P., *J. Mol. Biol.*, 179, 497 (1984).
12. Scheraga, H. A., Konishi, Y., Ooi, T., *Adv. Biophys.*, 18, 21 (1984).
13. Jacobson, G. R., Schaffer, M. H., Stark, G. R., and vanaman, T. C., *J. Biol. Chem.*, 248, 6583 (1973).
14. Wu, J., Gage, D., and Watson, J. T., *Anal. Biochem.*, 235, 161 (1996).
15. Wu, J., and Watson, J. T., *Protein Sci.* 6, 391 (1997).
16. Roder, H., Elove, G., and Englander, S. W., *Nature*, 335, 700 (1988).
17. Balbach, J., Forge, V., Van Nuland, N. A. J., Winder, S. L., Hore, P. J., and Dobson, C. M., *Nat. Struct. Biol.*, 2, 865 (1995).
18. Udgaonkar, J. B., and Baldwin, R. L., *Nature*, 335, 694 (1988).
19. Jennings, P. A., and Wright, P. E., *Science*, 262, 892 (1993).

20. Freedman, R. B., *Trends Biochem. Sci.*, 9, 438 (1984).

21. Creighton, T. E., *Biochem. J.*, 270, 1 (1990).

22. Haber, E., and Anfinsen, C. B., *J. Biol. Chem.*, 237, 1839 (1962).

23. Creighton, T. E., *Methods Enzymol.*, 131, 83 (1986).

24. Gilbert, H. F., *Adv. Enzymol.*, 63, 69 (1990).

25. Houk, J., and Whitesides, G. M., *J. Am. Chem. Soc.*, 109, 6825 (1987).

26. Wilson, J. M., Bayer, R. J., and Hupe, D. J., *J. Am. Chem. Soc.*, 99, 7922 (1977).

27. Szajewski, R. P., and Whitesides, G. M., *J. Am. Chem. Soc.*, 102, 2011 (1980).

28. Goldenberg, D., *Trends Biochem. Sci.*, 17, 992 (1992).

29. Barron, E. S. G., *Adv. Enzymol.*, 11, 201 (1951).

30. Larsson, A., Orrenius, S., Holmgren, A., and Mannervik, B., in *Functions of Glutathione: Biochemical, Physiological, Toxicological, and Clinical Aspects*, Raven, New, York, 1983.

31. Iyer, K. S., and Klee, W. A., *J. Biol. Chem.*, 248, 707 (1973).

32. Creighton, T. E., *Methods in Enzymology*, 107, 305 (1984).

33. Misra, H. P., *J. Biol. Chem.* 249, 2151 (1974).

34. Anfinsen, C. B., Haber, E., Sela, M., and White, F. H., *Proc. Natl. Acad. Sci. USA*, 47, 1309 (1961).

35. Creighton, T. E., *Prog. Biophys. Mol. Biol.*, 33, 231 (1978).

36. Goldberger, R. F., Epstein, C. J., and Anfinsen, C. B., *J. Biol. Chem.*, 238, 628 (1963).

37. Venetianer, P., and Straub, F. B., *Biochim. Biophys. Acta*, 67, 166 (1963).

38. Bergman, L. W., and Kuehl, W. M., *J. Biol. Chem.*, 254, 8869 (1979).

39. Darby, N. J., Norin, P. E., Talbo, G., and Creighton, T. E., *J. Mol. Biol.*, 249, 463 (1995).

40. Weissman, J. S., and Kim, P. S., *Proc. Natl. Acad. Sci. USA*, 89, 9900 (1992).

41. Gilbert, H. F., in *Mechanisms of Protein Folding*, Pain, R. H. (Editor), Oxford Press, Oxford, p112 (1994).

42. Chang, J. Y., *Biochemistry*, 35, 11702 (1996).

43. Chatrenet, B., and Chang, J. Y., *J. Biol. Chem.*, 268, 20988 (1993).

44. Chang, J. Y., *J. Biol. Chem.*, 270, 25661 (1995).

45. Chang, J. Y., Canals, F., Schindler, P., Querol, E., and Aviles, F. X., *J. Biol. Chem.*, 269, 22087 (1994).

46. Hua, Q. X., Narhi, L., Jia, W., Arakawa, T., Rosenfeld, R., Hawkins, N., Miller, J. A., and Weiss, M. A., *J. Mol. Biol.*, 259, 297 (1996).

47. Chatrenet, B., and Chang, J. Y., *J. Biol. Chem.*, 268, 20988 (1993).

48. Wilkins, J. A., Cone, J., Randhawa, Z., Wood, D., Warren, M. K., and Witkowska, H. E., *Protein Science*, 2, 244 (1993).

49. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B., *Biochemistry*, 31, 1749 (1992).

50. Glocker, M. O., Arbogast, B., Milley, R., Cowgill, C., and Deinzer, M. L., *Proc. Natl. Acad. Sci. USA*, 91, 5868 (1994).

51. Creighton, T. E., *Proc. Natl. Acad. Sci., USA*, 85, 5082 (1988).

52. Rothwarf, D. M., Scheraga, H. A., *J. Am. Chem. Soc.*, 113, 6293 (1991).

53. Konishi, Y., Ooi, T., and Scheraga, H. A., *Biochemistry*, 21, 4734 (1992).

54. Goto, Y., and hamaguchi, *J. Mol. Biol.*, 146, 321 (1981).

55. Torella, C., Ruoppolo, M., Marino, G., Pucci, P., *FEBS Letters*, 352, 301 (1994).

56. Rothwarf, D. M., Scheraga, H. A., *J. Am. Chem. Soc.*, 113, 6293 (1991).

57. Rothwarf, D. M., and Scheraga, H. A., *Biochemistry*, 32, 2671 (1993).

58. Bruice, T., and Kenyon, G. L., *J. Protein Chem.*, 1, 47 (1982).

59. Roberts, D. D., Lewis, S. D., Ballou, D. P., Olseon, S. T., and Shafer, J. A., *Biochemistry*, 25, 5595 (1986).

60. Thannhauser, T. W., Sherwood, R. W., and Scheraga, H. A., J. Protein Chem. 17, 37 (1998).

61. Happersberger, P., Cowgill, C., Deinzer, M. L., and Glocker, M. O., *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, Oregon, May, 1996, p1050.

62. Makinen, A. L., and Nowak, T., *J. Biol. Chem.*, 264, 12148 (1989).

63. Wang, Z. X., Preiss, B., and Tsou, C. L., *Biochemistry*, 27, 5095 (1988).

64. Mak, A. S., and Jones, B. L., *Anal. Biochem.*, 84, 432 (1978).

65. Andrews P. C., and Dixon, J. E., *Anal. Biochem.*, 161, 524 (1987).

66. Kruft, V., Kapp U., and Wittmann-Liebold, B., *Anal. Biochem.*, 193, 306 (1991).

67. Lee, Y.,Kinoshita, H., Radke, G., Weilder, S., Barranger, J. A., and Tomich, J. M., J. *Protein Chem.*, 14, 127 (1995).

68. Krieglstein, K. G., DasGupta, B. R., and henschen, A. H., *J. Protein Chem.*, 13, 49 (1994).

69. Jue, R. A., and Hale, J. E., *Anal. Biochem.*, 221, 374 (1994).

70. Krutzsch, H., and Inman, J. K., *Anal. Biochem.*, 209, 109 (1993).

71. Brune, D., *Anal. Biochem.*, 207, 285 (1992).

72. Jue, R. A., and Hale, J. E., *Anal. Biochem.*, 210, 39 (1993).

73. Hale, J. E., Beidler, D. E., and Jue, R. A., *Anal. Biochem.*, 216, 61 (1994).

74. Seifried, S. E., Wang, Y., and Von Hippel, P. H., *J. Biol. Chem.*, 263, 13511 (1988).

75. Pardo, J. P., and Slayman, C. W., *J. Biol. Chem.*, 264, 9373 (1989).

76. Miyanishi, T., and Borejdo, *J., Biochemistry*, 28, 1287 (1989).

77. First, E. A., and Taylor, S. S., *Biochemistry*, 28, 3598 (1989).

78. Bishop, J. E., Squier, T. C., Bigelow, D. J., and Inesi, G., *Biochemistry*, 27, 5233 (1988).

79. Sturrock, E. D., Yu, X. C., Wu, Z., Biemann, K., and Riordan, J. F., *Biochemistry*, 35, 9560 (1996).

80. Toyooka, T., and Imai, K., *Anal. Chem.*, 56, 2461 (1984).

81. Kirley, T., *J. Biol. Chem.*, 264, 7185 (1989).

82. Kirley, T., *Anal. Biochem.*, 180, 231 (1989).

83. Chin, C. C. Q., and Wold F., *Anal. Biochem.*, 214, 128-134 (1993).

84. Jason, J., and Ryden, L., *Protein Purification: Principles, High Resolution methods, and Applications,* VCH Press, (1993), p252.

85. Kalef, E., Walfish, P. G., and Gitler, C., *Anal. Biochem.*, 212, 325 (1993).

86. Zaluzec, E. J., Gage, D. A., and Watson, J. T., in *Proceeding of the 42th ASMS Conference on Mass Spectrometry and Allied Topics,* Chicago, (1994), p20.

87. Catsimpoolas, N., and Wood, J. L., *J. Biol. Chem.*, 241, 1790 (1966).

88. Jacobson, G. R., Schaffer, M, H., Stark, G. R., and Venaman, T. C., *J. Biol. Chem.*, 248, 6583, (1973).

89. Stark, G. R., *Methods in Enzymology*, vol. 47, 129 (1977).

90. Wakselman, M., and Guibe-Jampel, E., *J. C. S. Chem. Comm.*, 21 (1976).

91. Mikami, B., and Morita, Y., *J. Biochem.*, 93, 777 (1983).

92. Nika, H., and Hultin, T., *Eur. J. Biochem.* 142, 521 (1984).

93. Minard, E., Desmadril, M., Ballery, N., Perahia, D., Mouawad, L., Hall, L., and Yon, J. M., *J. Biochem.* 185, 419 (1989).

94. Darbre, A., *Practical protein Chemistry—A Handbook*, John Wiley & Sons, (1985), p149.

95. Daniel, R., Caminade, E., Martel, A., Goffic, F., canosa, D., Carrascal, M., and Abian, J., *J. Biol. Chem.*, 272, 26934 (1997).

96. Wu, J., Gage, D. A., and Watson, J. T., *Anal. Biochem.*, 235, 161 (1996).

97. Ryle, A. P., and Sanger, F., *Biochem. J.*, 60, 535 (1955).

98. Biemann, K., in *Methods in Protein Sequence Analysis*, Plenum Press, New York, (1993), p119.

99. Staswick, P. E., Hermodson, M. A., and Nielsen, N. C., *J. Biol. Chem.*, 259, 13431 (1984).

100. Haeffner-Gormley, L., Parente, L., and Wetlaufer, D. B., *Int. J. Pept. Protein Res.*, 26, 83 (1985).

101. Yazdanparast, R., Andrews, P. C., Smith, D. L., and Dixon, J. E., *J. Biol. Chem.*, 262, 2507 (1987).

102. Sun, Y., Bauer, M. D., Keough, T. W., and Lacey, M. P., *Methods in Molecular Biology, Vol. 61: Protein and peptide Analysis by Mass Spectrometry*, edited by Chapman, J. R., Humana press Inc., Totowa, NJ, (1996).

103. Smith, D. L., and Zhou, Z., *Methods in Ezymol.*, Vol. 193, 374 (1990).

104. Sun, Y., Zhou, Z., and Smith, D. L., in *Techniques in Protein Chemistry*, (hugli, T. E., ed.), Academic, New, York, p176 (1989).

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

106. Wetlaufer, D. B., Parente, L., and Haeffner-Gormley, L., *Int. J. Prpt. Protein Res.*, 26, 83 (1985).

107. Thannhauser, T. W., McWherter, C. A., and Scheraga, H. A., *Anal. Biochem.*, 149, 322 (1985).

108. Morris, H. R. and Pucci, P., *Biochem. Biophys. Res. Commun.*, 126, 1122, 1985.

109. Takao, T., Yoshida, M., Hong, Y., Aimoto, S., and Shimonishi, Y., *Biomed. Mass Spectrom.*, 11, 549, 1984.

110. Yazdanparast, R., Andrews, P. C., Smith, D. L., and Dixon, J. E., *J. Biol. Chem.*, 262, 2507, 1987.

111. Stults, J. T., Bourell, J. H., Canova-Davis, E., Ling, V. T., Laramee, G. R., Winslow, J. W., Griffin, P. R., Rinderknecht, E., and Vandlen, R. L., *Biochem. Environ. Mass spectrom.*, 11, 549-556, 1984.

112. Vaccaro, A. M., Salvioli, R., Barca, A., Tatti, M., and Pucci P., *J. Biol. Chem.*, 270, 9953-9960, 1995.

113. Yoshida, S., Takamatsu, T., Kanda, F., *Rapid commun. Mass Spectrom.* 10, 641-648, 1996.

114. Beam, M. F., and Carr, S. A. *Anal. Biochem.* 201, 216,1992.

115. Sun, Y., and Smith, D. L. *Anal. Biochem.* 172, 130, 1988.

116. Hunt, D.F., Yates, J. R., Shabanowitz, J., Winston, S., and Hauser, C. R., *Proc. Natl. Acad. Sci. USA*, 83,6233, 1986.

117. Bauer, M., Sun, Y., Degenhardt, C., and Kozikoski, B., *J. Protein Chem.* 12, 759,1993.

118.Krishnamurthy, T., and Hauer, C. C., *Biol. Mass spetrom.* 23, 719, 1994.

119. Bayer, M., Sun, Y., Degenhardt, C., and Kozikowsiki, B., *J. Protein Chem.* 12, 759, 1993.

120. Stewart, A. E., Raffioni, S., and Chandhery, T., *Protein Science*, 1, 777, 1992.

121. Patterson, S., and Katta, V., *Anal. Chem.* 66, 3727, 1994.

122. Jones, M. D., Patterson, S. D., and Lu, H. S., *Anal. Chem.* 70, 136 (1998).

123.Stults, J. T., Bourell, J. H., Canova-Davis, E., Ling, V. T., Laramee, G. R., Winslow, J. W., Griffin, P. R., Rinderknecht, E., and Vandlen, R. L., *Biomed. Environ. Mass Spectrom.*, 19, 655 (1990).

124. Fox, J. W., Elzinga, M., and Tu, A. T., *Biochemistry*, 18, 678 (1979).

125. Serensen, H. H., Thomsen, J., and Bayne, S., *Biomed. Environ. Mass Spectrom.* 19, 713 (1990).

126. Nokihara, K., Morita, N., Yamaguchi, M., and Watanabe, M., *Anal. Lett.*, 25, 513 (1992).

127. Zhang, D., and Liang, S., *J. Protein Chem.*, 12, 735 (1993).

128. Krishnamurthy, T., Hauer, C. R., Prabhakaran, M., Freedy, J. G., and Hayashi, K., *Biol. Mass Spectrom.*, 23, 719 (1994).

129. Tang, Y., and Selsted, M. E., *J. Biol. Chem.*, 268, 6649 (1993).

130.Gray, W. R., *Protein Science*, 2, 1749 (1993).

131. Haniu, M., Hsieh, P., Rohde, M. F., and Kenney, W. C., *J. Biochem. Biophys.*, 310, 433 (1994).

132. Yamashita, H., Nakatsuka, T., and Hirose, M., *J. Biol. Chem.*, 270, 29806 (1995).

133. Fairlie, W. D., Stanton, P. G., and Hearn, M. T. W., *Biochem. J.*, 314, 449 (1996).

134. Gray, W. R., Luque, F. A., Galyean, R., Atherton, E., Sheppard, R. C., Stone, B. L., Reyes, A., Alford, J., Mcintosh, M., Olivera, B. M., Cruz, L. J., and rivier, *J., Biochemistry*, 23, 2796 (1984).

135. Gray, W. R., *Protein Sci.*, 2, 1732 (1993).

136. Wu, J., and Watson, J. T., *Protein Science*, 6, 391 (1997).

137. Thorgersen, D. F., Skowronski, R. P., and Macfarlane, R. D., *Biochem. Biophys. Res. Commun.*, 60, 616 (1974).

138. Barber, M., Bordoli, R. S., Sedgwick, R. D., and Tyler, A. N., *Chem. Commun.*, 325 (1981).

139. Meng, C. K., Mann, M., and Fenn, J. B., *Proc. 36th ASMS Conference on Mass Spectrometry and Allied Topics*, San Francisco, CA, (1988), p771.

140. Posthumus, M. A., Kistemaker, P. G., Meuzelaar, H. L. C., deBrauw, M. C., *Anal. Chem.*, 50, 985 (1978).

141. Karas, M., Bachmann, D. B., Hillenkamp, F., *Int. J. Mass Spectrom. Ion. Proc.*, 78, 53 (1987).

142. Hillenkamp, F., Karas, M., Beavis, R. C., Chait, B. T., *Anal. Chem.*, 63, 1193A (1991).

143. Ehring, H., Karas, M., and Hillenkamp, F., *Org. Mass Spectrom.*, 27, 472 (1992).

144. Fitzgerald, M. C., Parr, G. R., and Smith, L. M., *Anal. Chem.*, 65, 3204 (1993).

145. Beavis, R. C., and Chait, B. T., *Rapid Commun. Mass Spectrome.*, 3, 233 (1989).

146. Juhasz, P., Costello, C. E., and Biemann, K., *J. Am. Soc. Mass Spectrom.*, 4, 399 (1993).

147. Vorm, O., Roepstorff, P., and Mann, M., *Anal. Chem.*, 66, 3281 (1994).

148. Xiang, F., and Beavis, R. C., *Org. Mass Spectrom.*, 28, 1424 (1993).

149. Xiang, F., and Beavis, R. C., *Rapid Commun. Mass Spectrom.*, 8, 199 (1994).

150. Cohen, S. L., and Chait, B. T., *Anal. Chem.*, 68, 31 (1996).

151. Preston, L. M., Murray, K. K., and Russsell, D. H., *Biol. Mass Spectrom.*, 22, 544(1993).

152.Vorm, O., and mann, M., *J. Am. Soc. Mass Spectrom.*, 5, 955 (1994).

153. Beavis, R. C., and Chait, B. T., *Proc. Natl. Acad. Sci. USA*, 87, 6873 (1990).

154. Mamyrin, B. A., Karateev, V. I., Shmikk, D. V., Zagulin, V. A., *Sov. Phys. JETP*, 37, 45 (1973).

155. Anderegg, R. J., Wagner, D. S., Stevenson, C. J., Borchardt, R. T., *J. Am. Soc. Mass Spectrom.*, 5, 425 (1994).

156. Bose, G. M., *Recherches sur le cause et sur la veritable theorie de l'electricite*, Wittenberg, 1745.

157. Zeleny, J., *J. Phys. Rev.*, 10, 1 (1917).

158. Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., Alice, M. B., *J. Chem. Phys.* 49, 2240 (1968).

159. Mack, L. L., Kralik, P., Rheude, A., Dole, M., *J. Chem. Phys.*, 52, 4977 (1970).

160. Clegg, G. A., Dole, M., *Biopolymers*, 10, 821 (1971).

161. Gieniec, J., Mack, L. L., Nakama, K., Gupta, C., Kumar, V., Dole, M., *Biomed. Mass Spectrom.*, 11, 259 (1984).

162. Yamashita, M., Fenn, J. B., *J. Phys. Chem.*, 88, 4451 (1984).

163. Aleksandrov, M. L., Gall, L. N., Krasnov, V. N., Nikolaev, V. I., Pavlenko, V. A., Shkurov, V. A., *Dokl., Akad. Nauk. SSSR*, 277, 379 (1984).

164. Aleksandrov, M. L., Gall, L. N., Krasnov, V. N., Nikolaev, V. I., Pavlenko, V. A., Shkurov, V. A., Baram, G. I., Grancher, M. A., Knorre, V. D., Kusner, Y. S., *Bioorg. Kim.*, 10, 710 (1984).

165. Meng, C. K., Mann, M., Fenn, J. B. Z., *Phys. D.*, 10, 361 (1988).

166. Gomez, A., Tang, K., *Phys. Fluids*, 6, 404 (1994).

167. Rayleigh, J. W., *Philos. Mag.*, 14, 182 (1882).

168. Siuzdak, G., *Proc. Natl. Acad. Sci. USA*, 91, 11290 (1994).

169. Gale, D. C., Smith, R. D., *Rapid Comm. Mass Spectrom.*, 7, 1017 (1993).

170. Smith, R. D., Wahl, J. H., Goodlett, D. R., Hofstadler, S. A., *Anal. Chem.*, 65, A574 (1993).

171. Griffin, P. R., Coffman, J. A., Hood, L. E., Yates, J. R., *Int. J. Mass Spectrom. Ion. Proc.*, 111, 131 (1991).

172. Mann, M., Meng, C. K., Fenn, J. B., *Anal. Chem.*, 61, 1702 (1989).

CHAPTER 2

A COMPARISON OF MALDI AND ESI FOR ANALYSIS OF IMINOTHIAZOLIDINE-BLOCKED PEPTIDES IN DISULFIDE MAPPING

I. Introduction

Cleavage at cysteine residues of peptide chains under alkaline conditions after cyanolysis of disulfide bonds was first observed by Catsimpoolas and Wood (1). However, because of the side reactions, such as the reversibility of the cyanolysis and the elimination of thiocyanylate, the yields were low and the cleavage reaction was not acceptable to protein chemists. Jacobson *et al.* (2) showed that 2-nitro-5-thiocyanobenzoic acid (NTCB) specifically cyanylates cysteine thiols under alkaline conditions. Subsequent cleavage occurs on the N-terminal side of the cyanylated cysteinyl residue under mildly alkaline conditions to form an amino-terminal peptide and a series of 2-iminothioazolidine (itz)-4-carboxyl peptides. Although the original paper claimed that the cleavage reaction can come to completion with only minor side reactions for most of the peptides and proteins tested, Digani and Patchornik (3) found β-elimination, occurring also under alkaline conditions, competes with the cleavage as an adverse reaction. Experimentally, a large excess of NTCB and a low total concentration of thiol groups must be applied to avoid the side reaction of displacement of CN- from S-cyanocysteine residues by the unreacted thiol groups (3).

Wu & Watson (4) conducted a systematic study to elucidate the effects of peptide structure and reaction conditions on the kinetics of the cleavage reactions of cyanylated

peptides and proteins and the yields of cleavage products. The optimal results were obtained in 1M ammonium hydroxide solution in which cleavage is complete within an hour at ambient temperature (4). Ammonia is a stronger small nucleophile but has smaller proton affinity than hydroxyl ion so that nucleophilic attack of ammonia on the carbonyl carbon facilitates the cleavage process giving an α-amidated N-terminal peptide but low capacity for catalyzing the β-elimination reaction. Additionally, the excess volatile ammonia can easily be removed from the mixture after cleavage, an advantage for subsequent analysis of the products by MALDI or ESI-MS (4).

Other reagents for cyanylation of SH groups have been suggested (5-8). Among them 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate (8) has the advantage of the minimization of thiol/disulfide exchange because of the acidic condition of cyanylation reaction of sulfhydryl groups.

Cyanylation and cleavage reactions of the cysteine residue have the capacity be used for its structural analysis. If a protein contains **n** cysteine residues, the cleavage reaction will result in the formation of **n+1** peptide fragments; mass analysis of the fragments indicates the number and location of cysteine residues. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) has been used to mass map the cleavage products (9-28). However, peptide assignments using this approach are often complicated by its poor mass accuracy (error>5%). Papayannopoulos and Biemann (29) used CID tandem mass spectrometry to sequence the NTCB cleavage reaction products of a protease inhibitor isolated from *Sarcophaga bullata*. Their work demonstrated that mass spectrometry could be used to sequence peptides from the NTCB cleavage reaction in spite of the blocked N-terminus. However, tandem MS is practically limited to low-

mass peptides produced from cysteine-rich proteins, because of the effective mass limit of CID (<3000 Da).

Recently, Wu *et al.* (30) developed a methodology to recognize the number and location of both cystines and free sulfhydryls in peptides and proteins using the specific chemical cleavage reaction described above, followed by mass mapping of the resulting peptides by MALDI-MS. This approach provides the advantages of fast analysis, easy operation, high mass accuracy, and high sensitivity.

Both electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry allow for the analysis of large biomolecules through "mild" desorption and ionization methods (31, 32), each having unique capabilities, as well as some fundamental similarities. They are truly complementary in nature, almost a mirror image being projected between the disadvantages of one technique and the advantages of the other (33). During our original work, analyses of resulting mixtures of iminothiazolidine-blocked peptides from the cleavage of cyanylated protein by MALDI showed frequent, and sometimes complete, suppression of a signal for some fragments. Since such a phenomenon can compromise the structural analysis of cysteine residues by the methodology based on the cyanaylation chemistry, an alternative mass analysis technique should be used. ESI-MS will be a reasonable choice because of its capability of analysis of large molecules and its complementary nature for MALDI.

In this chapter, we investigated the use of ESI for the analysis of iminothiazolidine-blocked peptides when MALDI had the signal suppression. We used off-line HPLC to separate the mixture of iminothiazolidine-blocked peptides and then analyzed the HPLC fractions by ESI-MS. Two proteins, β-lactoglobulin A and ovalbumin, were used as

model proteins to investigate the complementary nature of MALDI and ESI in separate experiments. The results indicate that the responses of blocked peptides to ESI and MALDI are frequently complementary.

II. Analytical Methodology

A. Cyanylation and Chemical Cleavage Reaction

Cyanylation of cysteine residue employs a one-step selective chemical reaction between free sulfhydryl and NTCB (3, 30, 34) or CDAP (8). After selectively cyanylating the sulfhydryls, the cysteinyl peptide bonds can be cleaved under alkaline condition to form an amino-terminal peptide and a series of iminothiazolidine-blocked peptides. The mechanism of cyanylation and cleavage reactions is illustrated in Figure 2.1.

The side reaction is that the cyanylated cysteine undergoes a base-catalyzed β-elimination to form thiocyanate and dehydroalanine. This side reaction results in lower yield of cleavage.

The cyanylation of sulfhydryl groups was traditionally accomplished by 2-nitro-5-thiocyanobenzoic acid (NTCB) under mildly alkaline conditions (pH 8-10). The reagent was believed to be specific to sulfhydryl groups, although side reactions, such as the formation of mixed disulfide bonds between the NTCB and protein SH groups, were also reported (3, 35, 36). Another reagent, 1-cyano-4-dimethylamino-pyridinium (CDAP) tetrafluoroborate, was proposed for the cyanylation of SH groups under slightly acidic conditions (pH 3-7) (5). CDAP is advantageous over NTCB for specific cyanylation of protein SH groups in the presence of disulfide bonds, because the acidic condition can effectively minimize sulfhydryl/disulfide exchange. Furthermore, the cyanylated proteins

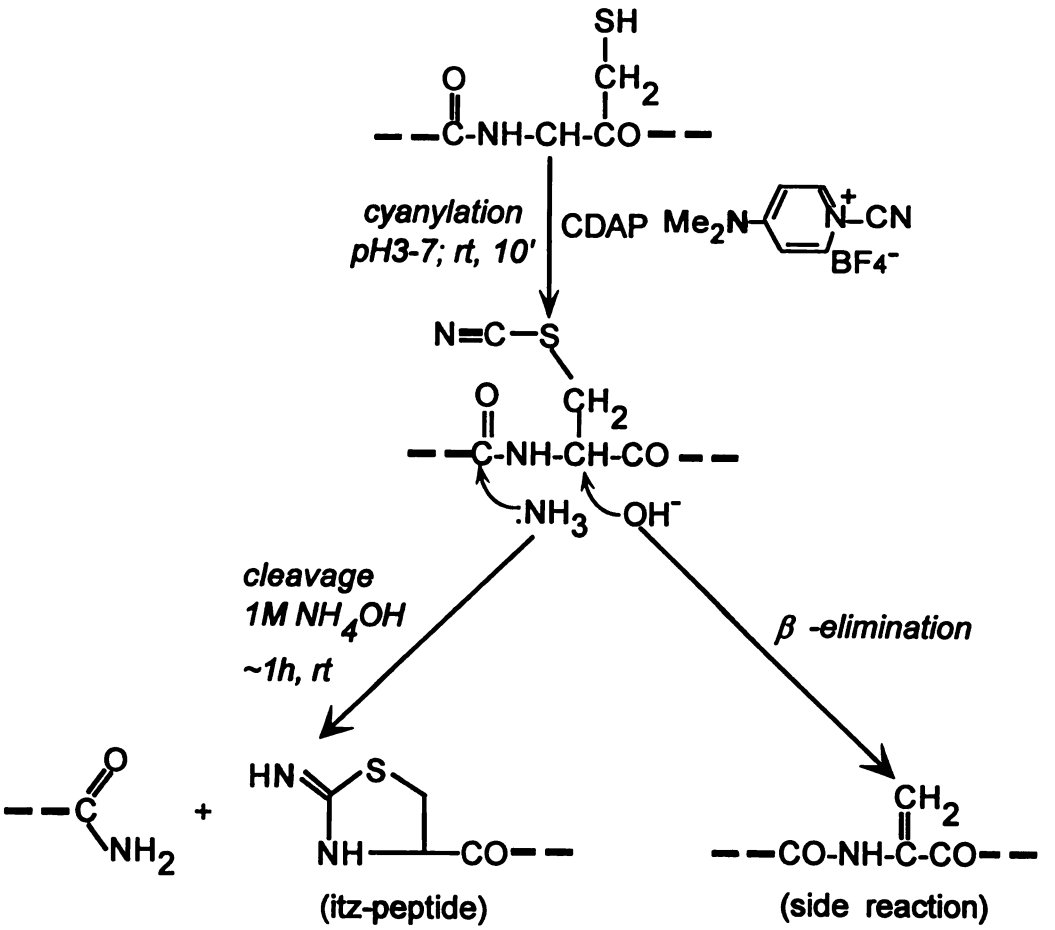


Figure 2.1. Mechanisms of cyanylation and cleavage reactions

or peptides are stable under acidic conditions. While complete cyanylation by CDAP can be carried out using three to five-fold molar excess of the reagent over free SH groups under very mild conditions at room temperature (5, 37, 38), the optimal condition should be 10 to 20-fold molar excess of CDAP (over sulfhydyl groups) in pH 3.0 buffer for 15 min at ambient temperature (4). A large excess of CDAP (~50 fold over SH group) and excessive incubation time (>2 hours) could result in the formation of an unidentified side reaction product (4). Therefore, like cyanylation by NTCB, the cyanylation by CDAP needs to be performed under controlled conditions to minimize the side reaction.

The another characteristic of the reaction is that the resulting peptide fragments may vary greatly in size because of the widely variable frequency of occurrence of cysteine in proteins. Large or small sized fragments may be obtained. Both MALDI and ESI-MS can be used to measure the masses of fragments. While MALDI-MS usually has the suppression phenomenon at low mass range, ESI-MS can be used to determine the mass. Therefore, we can expect that the combination of MALDI-MS and ESI-MS can be promising for cysteine analysis.

B. Localization of Cysteines and Cystines by ESI-MS

The number and location of cysteines and cystines in a protein can be determined by a specific chemical cleavage process, which involves cyanylation of free sulfhydryls by CDAP at pH 3 or NTCB at pH 8.7, then cleavage in 1M ammonia. Two steps can be used to locate cysteines and cystines (30), one is called NTCB or CDAP/TCEP (cyanylation/reduction) procedure which involves first cyanylating the free sulfhydryls in non-reducing proteins, then cleaving the cyanylated proteins following reduction of the remaining disulfide bonds. This step can recognize the location of free cysteines. Another

step is called TCEP/NTCB or CDAP (reduction/cyanylation) procedure, which first reduces all the disulfides by tris-2-carboxyethyl)phosphine (TCEP), then cyanylates the reduced protein, and then cleaves the peptide bond on the N-terminal side of cyanylated cysteine. This step can locate total cysteines. The location of cystines can be deduced by comparing the results of the two steps (scheme 2.1).

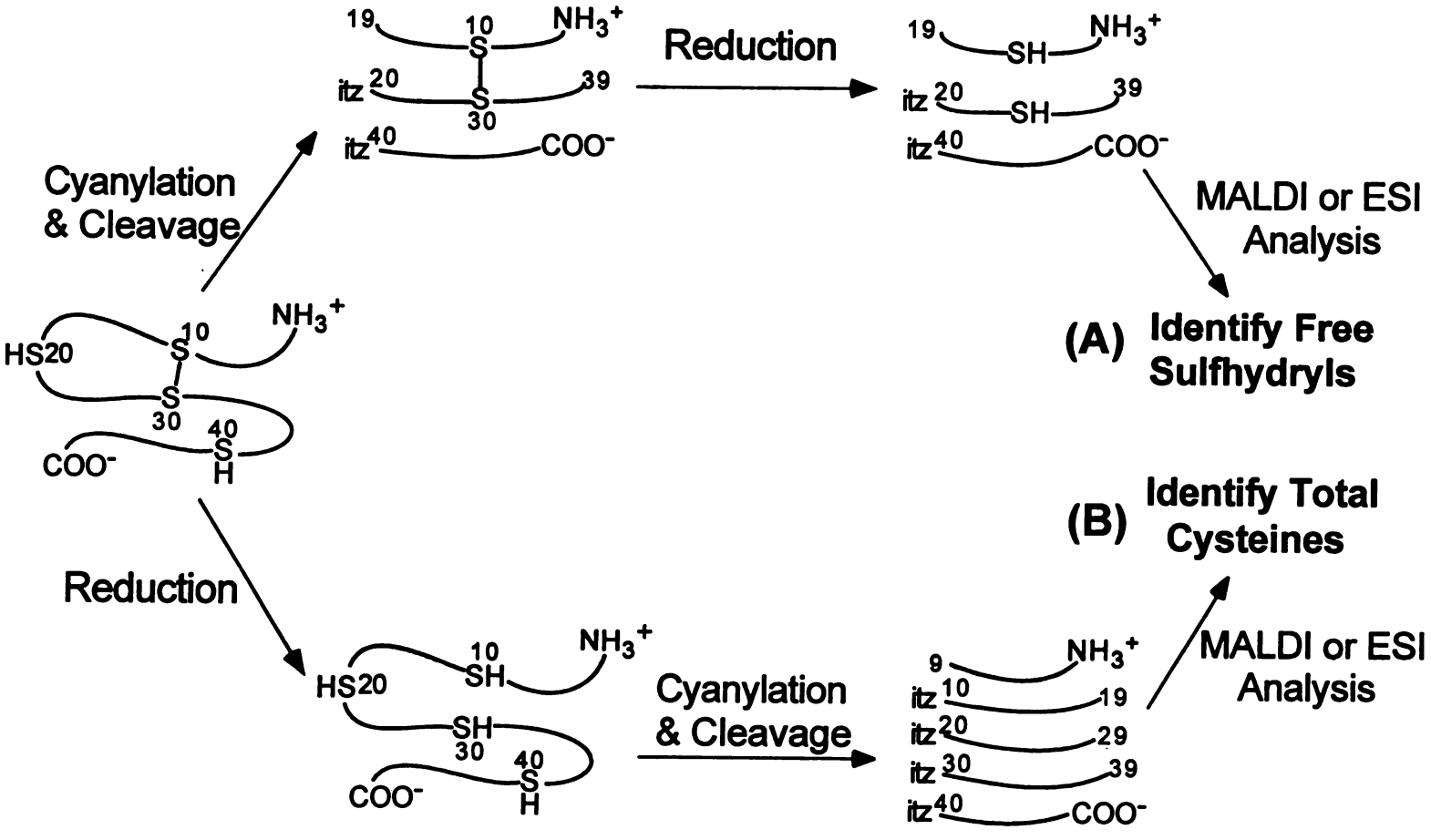
III. Experimental Section

ESI-MS

ESI mass spectra were obtained on a Fisons VG Platform ESI mass spectrometer. This instrument is equipped with a single quadrupole mass analyzer. The flow rate was set at 10μl/min for infusion injection mode, and the solvent is 50:50 acetonitrile/H_2O containing 1 % formic acid. The capillary voltage is set at 3.25KV, cone voltage at 37V, and source temperature at 100^0C.

MALDI-TOF MS

MALDI mass spectra were obtained on a Voyager Elite time-of-flight (TOF) mass spectrometer (PerSeptive Biosystems Inc., Framingham, MA) equipped with delayed extraction and a model VSL-337ND nitrogen laser (Laser Science, Newton, MA). The accelerating voltage in the ion source was set to 20 KV. Grid voltage and guide wire voltages were 93.6% and 0.2% of the accelerating voltage, respectively. Data were acquired in the positive linear DE mode of operation. Time-to-mass conversion was achieved by external and/or internal calibration using standards of bradykinin (m/z



Scheme 2.1. Recognizing the location of cysteines and cystines

1061.2), bovine pancreatic insulin (m/z 5734.5), and horse skeletal myoglobin (m/z 16952) obtained from Sigma Chemical Co. (St. Louis, MO). All experiments were performed using α-cyano-4-hydroxycinnamic acid (Aldrich Chemical Co., Milwaukee, WI) as the matrix. Two types of saturated matrix solutions were prepared. Type A contains 70% (v/v) solution of acetonitrile/aqueous 1% TFA. Type B contains 50% (v/v) solution of acetonitrile/aqueous 1% TFA. The protein or peptide samples were applied with a sandwich mode to a stainless-steel sample plate: the first layer was 0.7 μl matrix solution A, second layer was 0.6 μl B, the third layer was 1 μl sample, and the last layer was 0.5 μl matrix solution A again. The sample sandwich was allowed to air dry before being introduced into the mass spectrometer.

HPLC

HPLC is equipped with Waters model 6000 pumps controlled by a PC computer. UV detection was at 215nm. The column was a Vydac C18 (#218TP54, 10μm particle size, 300- pore, 4.6×250mm).

Chemicals

Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Pierce, Rockford, IL. Guanidine hydrochloride was a product of Boehringer-Mannheim Biochemicals (Indianapolis, IN). Proteins, 2-nitro-5-thiocyanobenzoic acid (NTCB), 1-cyano-4-dimethylamino-pyridinium tetrafluoroborate (CDAP), were purchased from Sigma and used without further purification. Acetonitrile and TFA were HPLC grade. The TCEP solution in 0.1 M citrate buffer at pH 3 or in 0.1 M tris-HCl buffer at pH 8.2

was prepared as 0.1 M stock solution and stored under N_2 at -20^0C for weeks with little deterioration. The 0.01 M solution of NTCB was prepared in 0.1 M tris-HCL buffer and the pH of the solution was adjust to pH 8.7. The 0.1 M CDAP solution in pH 3.0, 0.1 M citrate buffer was freshly prepared prior to use.

Analysis of Free Sulfhydryl Groups in Proteins (CDAP or NTCB/TCEP Procedure)

About 10nmol protein under study is dissolved in 10μl 0.1N tris-HCl buffer (pH 8.0) for NTCB reaction or in 0.1N citrate buffer (pH 3.0) for CDAP reaction, containing 6N guanidine-HCl as denaturing reagent. To the solution, add 5-10 fold molar excess of NTCB or 10-20 fold CDAP (over the free sulfhydryls). Allow the reaction to proceed at 37^0C in a water bath for 30min. Adjust pH to 12 by 1N NH3, let it stay in room temperature for 1 hour. Add 5-10 fold molar excess TCEP (over the disulfide bond). The mixture is put in a water bath at 37^0C for 30min to reduce the disulfide bond. The masses of the cleavage fragments were analyzed by MALDI-MS. Comparatively, the mixture of the cleavage product was separated by reversed-phase HPLC, the peaks were collected from HPLC and analyzed by ESI-MS.

Confirmation of Total Cysteines (TCEP/CDAP or NTCB Procedure)

The protein solution is allowed to react at pH 8.7 with 5-10 fold NTCB (over free sulfhydryl groups) or at pH 3.0 with 10-20 fold CDAP (over free sulfhydrl groups) at 37^0C for 30min after it is reduced with 5-10 fold TCEP (over disulfide bonds) at 37^0C for 30min, and then adjust pH to 12 by 1N NH_3, following reaction at room temperature for 1 hours, analyzed by MALDI-MS or off-line HPLC/ESI-MS.

IV. Results and Discussion

A. Recognizing the Location of Cysteine and Cystines of β-Lactoglobulin A

β-lactoglobulin A has 162 amino acids and molecular weight 18368 Da. It contains two disulfide bonds and one free cysteine at Cys 121 (Figure 2.2). The expected masses of cleavage peptide fragments after two procedures are shown in Table 2.1.

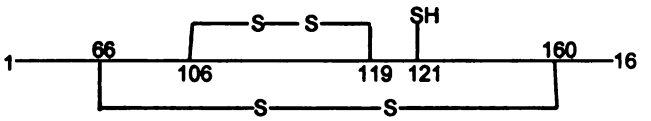


Figure 2.2. Structure of β-Lactoglobulin A

The expected peaks for the fragments 1-65(m/z 7247.4), itz106-118(m/z 1461.6), itz119-120(m/z 274.3), and itz160-162(m/z 396.5) for the procedure reduction/cyanylation (TCEP/NTCB) are undetected in the MALDI spectrum (Figure 2.3 A) which clearly indicates the signal suppression in MALDI-MS. The abbreviation "itz" represents "iminothiazolidine" derivative at the specified residue. Only two peaks of itz66-105 and itz121-159 were identified by MALDI-MS. Because these two peptide fragments indicate the cyanylation and cleavage only at Cys 66 and Cys 121, the numbers and location of other cysteine residues couldn't be recognized. Therefore, MALDI-MS gave the incomplete mass mapping of cysteine residues of β-lactoglobulin A.

Table 2.1. Expected and Observed Masses of Cleavage Product of β-Lactoglobulin A

Reduction/Cyanylation

Fragments	Expected Mass (Da)	ESI(Da)	MALDI(Da)
1-65	7247.4	7243.4	no
itz66-105	4651.6	4651.1	4653.5
itz106-118	1461.6	no	no
itz119-120	274.3	no	no
itz121-159	4551.2	4550.4	4548.1
itz160-162	396.5	397.0	no
*1-105	11822.0	11823.0	no
itz*121-162	4870.7	4870.7	no

Cyanylation/Reduction

Fragments	Expected Mass(Da)	ESI(Da)	MALDI(Da)
1-120	13506.7	no	no
itz121-162	4905.7	4904.7	4901.1

* β-elimination
itz = iminothiazolidine derivative.

Instead of the expected two peaks, one peak at m/z 4901.1 (1-120) was identified by MALDI-MS for the cleavage product resulting from the procedure cyanylation/reduction (NTCB/CDAP) (Figure2.3B).

Based on the above results, we used off-line HPLC separation/ESI-MS to analyze the cleavage products. The HPLC chromatogram of the mixture of cleavage products resulting from procedure TCEP/CDAP of β-lactoglobulin A is shown in Figure 2.4. The ESI spectra corresponding to the HPLC peaks in Figure 2.4 are illustrated in Figure 2.5. ESI-MS identified peptide fragments of 1-65, itz66-105, itz121-159, and itz160-162, indicating that the location of cysteine residues should be at Cys 66, Cys 106, Cys 121, and Cys 160. The location of one cysteine residue at Cys 119 cannot be recognized because of two unidentified fragment peaks of itz106-118 and itz119-120, which is probably due to the difficulty of cleaving the peptide bond at the N-terminal side of cysteine residue at Cys119 because of the closeness of two cysteines at Cys 119 and Cys 121.

The HPLC chromatogram and corresponding ESI spectra of the cleavage product resulting from procedure CDAP/TCEP of β-lactoglobulin A is shown in Figure 2.6 and Figure 2.7, respectively. One fragment of itz121-162 was identified, indicating that the cleavage occurred at Cys 121, thus Cys 121 involves in the pairing of the disulfide bond. However, the cysteine residue paired with Cys 121 cannot be located because of the unidentified fragment of 1-120.

Comparing the results between MALDI and ESI of β-lactoglobulin A, we can conclude:

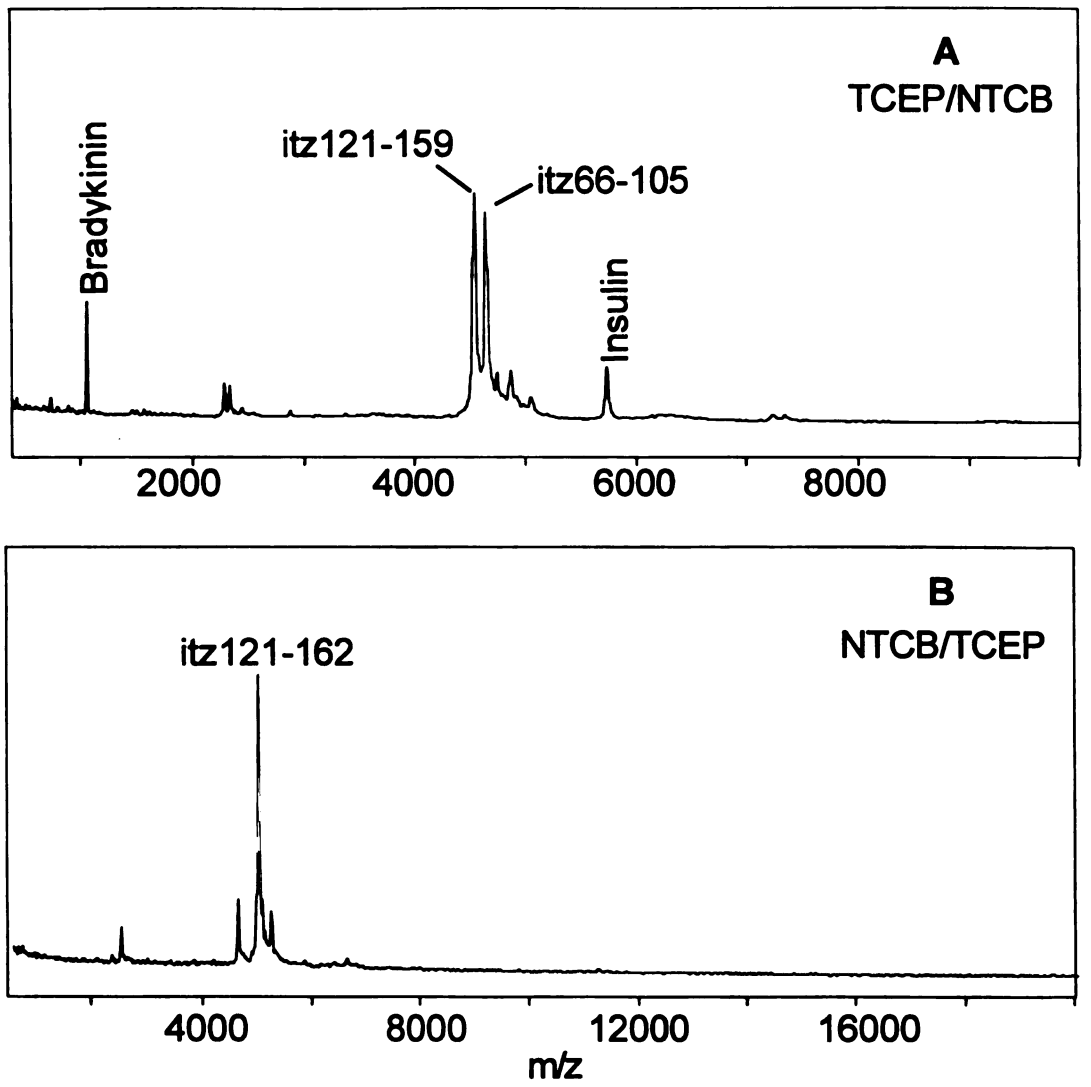


Figure 2.3. MALDI spectra of cleavage products of β-lactoglobulin A following reduction/cyanylation/cleavage (panel A) or cyanylation/cleavage/reduction (panel B).

itz = iminothiazolidine derivative.

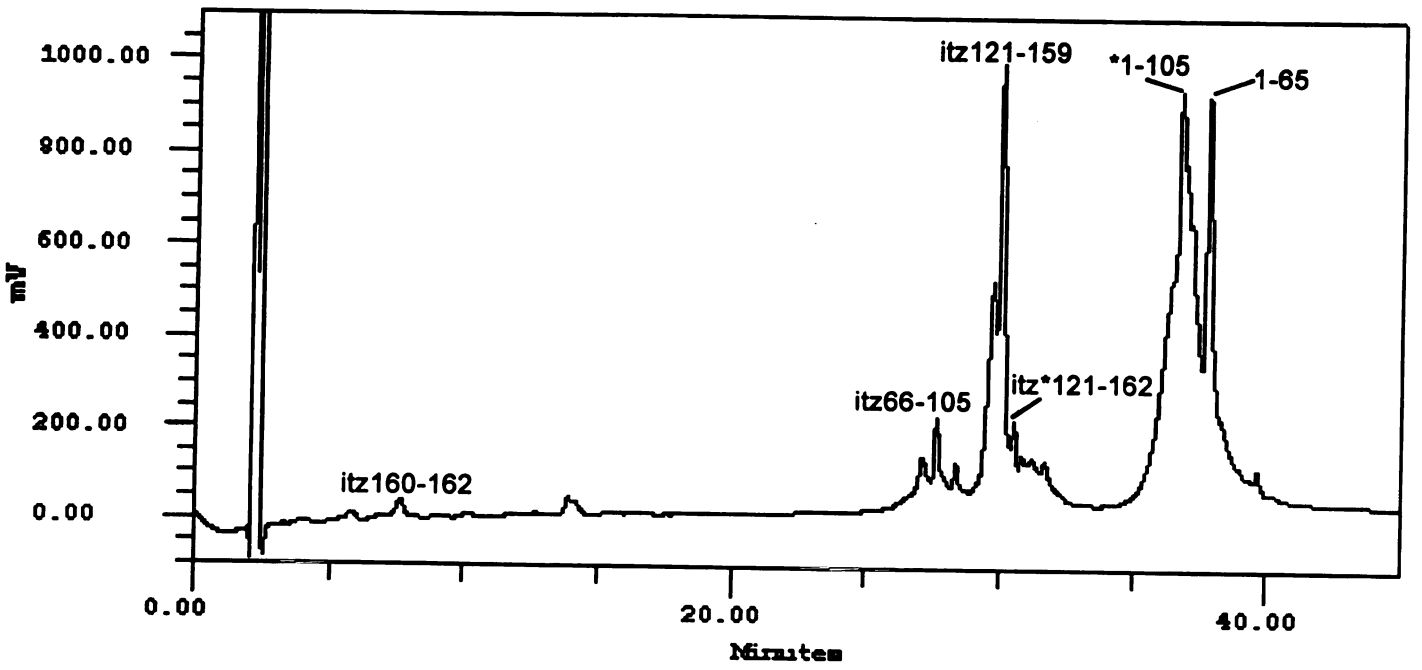


Figure 2.4. HPLC chromatogram of cleavage products following procedure TCEP/CDAP.

Solven A: 0.1% TFA;

Solvent B: 90% acetonitrile containing 0.1% TFA.

Gradient is 10%-70% B linear in 50min.

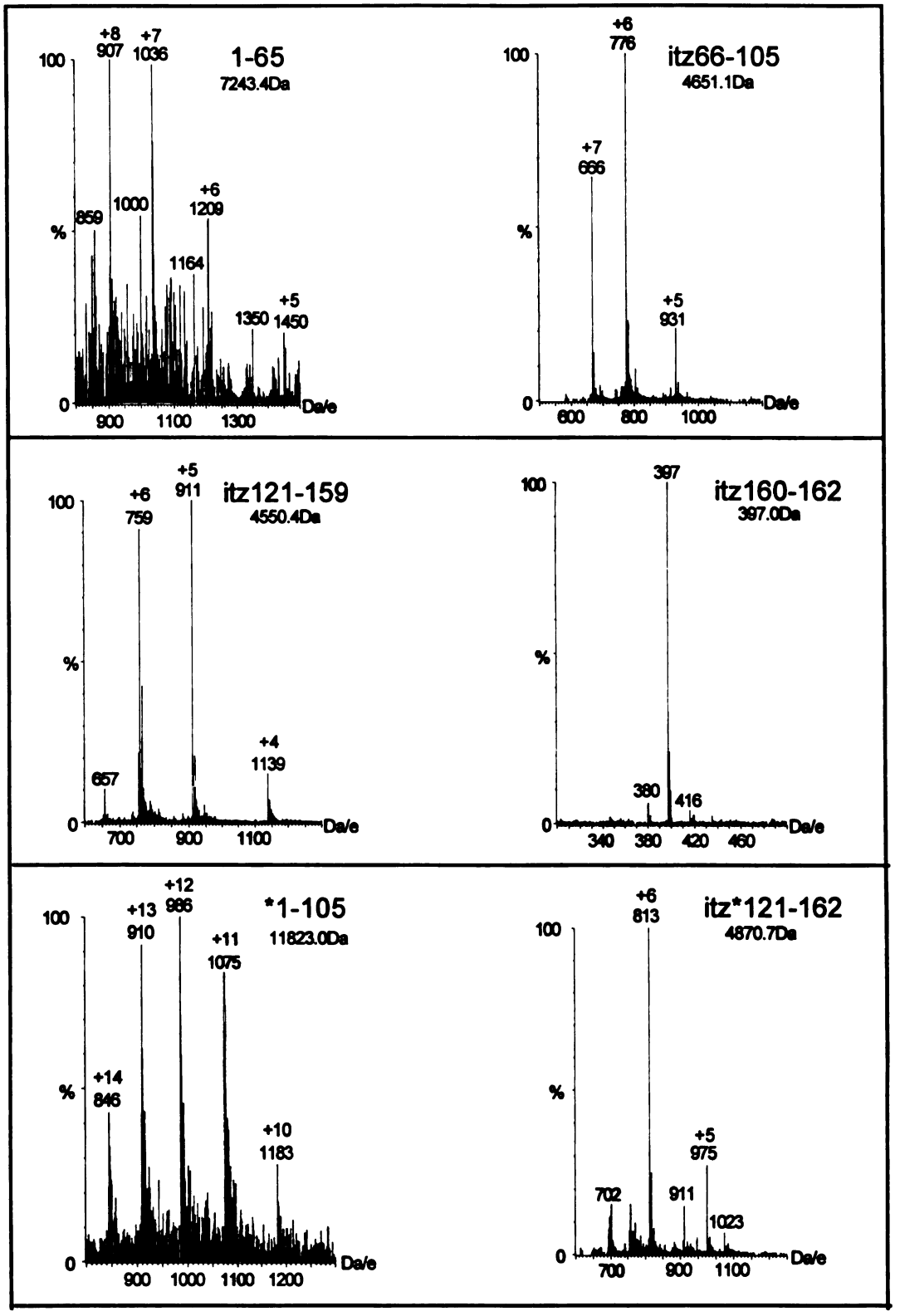


Figure 2.5. ESI spectra of cleavage products from procedure TCEP/CDAP of β-lactoglobulin A

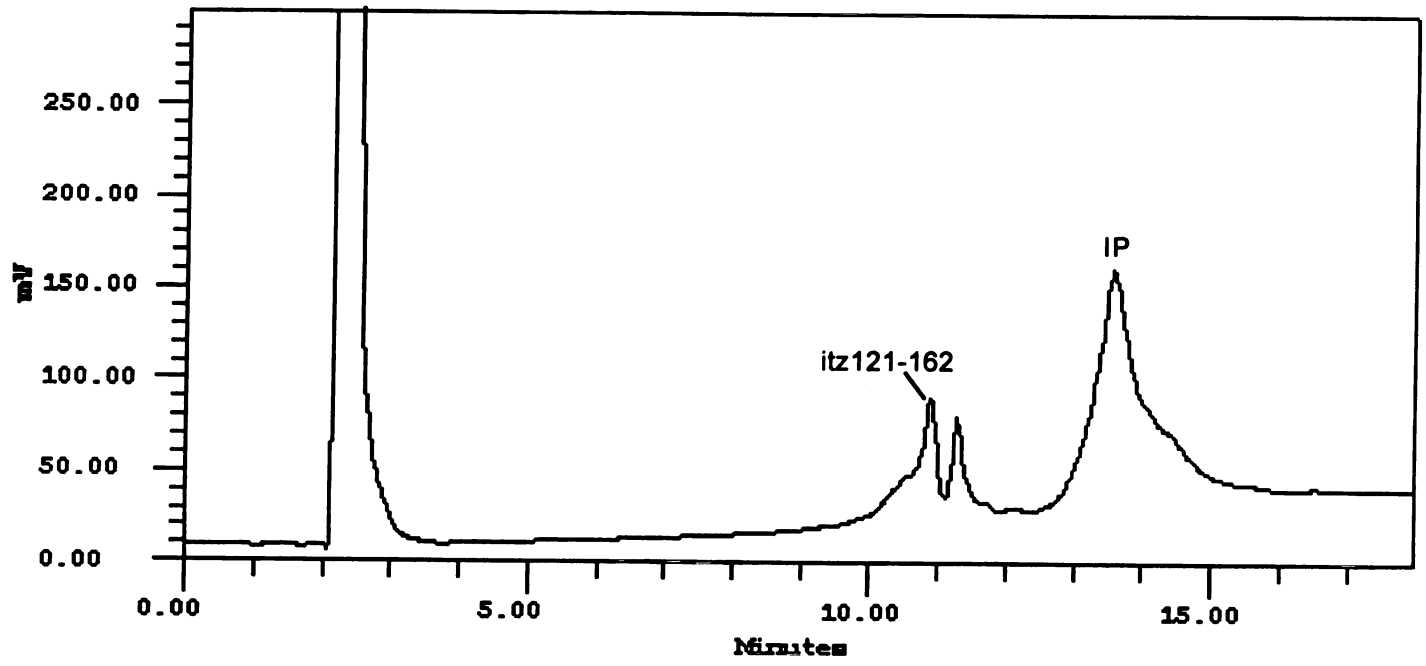


Figure 2.6. HPLC chromatogram of cleavage produts following procedure CDAP/TCEP.

HPLC conditions are the same as described in Figure 2.4.

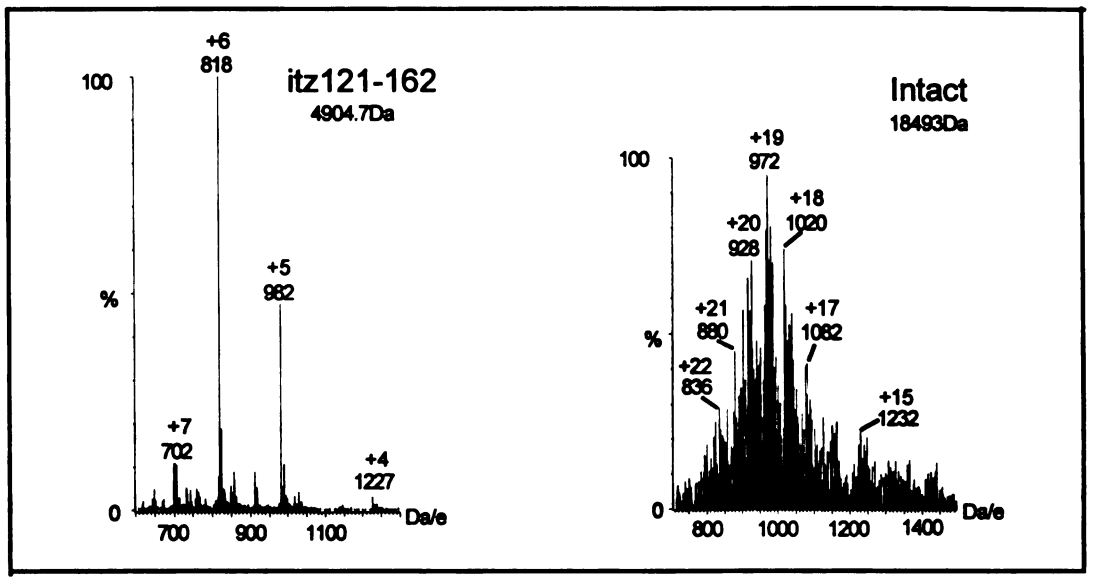


Figure 2.7. ESI spectra of cleavage products from procedure CDAP/TCEP of β-lactoglobulin A

1). Every peak in the HPLC chromatogram was identified by ESI-MS and assigned to a specific fragment (Table 1).

2). The peaks for fragments itz160-162(m/z 397) and 1-65(m/z 7243.4) that cannot be identified by MALDI-MS were successfully determined by ESI-MS.

3). HPLC and ESI-MS showed a large β-elimination peak *1-105(m/z 11823.0 Da) and itz*121-162 (m/z 4870.7 Da), but those peaks were totally suppressed in MALDI-MS.

4). MALDI-MS identified a large peak for fragment itz66-105(m/z 4653.5) and no peak for fragment 1-65(m/z 7247.4), but HPLC peak belonging to fragment itz66-105 is smaller than that of fragment 1-65, which clearly indicate the signal suppression phenomenon of MALDI.

5). The results of the analysis of cysteine residue status in β-lactoglobulin A showed a promise for the method of HPLC separation/ESI-MS, and this method can also be further used by on-line LC/MS.

B. Recognizing the Location of Cysteines and Cystines in Ovalbumin

Ovalbumin is a large protein containing 385 amino acids and having a molecular weight of 42699Da. It contains one disulfide bond and four free cysteines (Figure 2.8). The expected masses of cleavage fragments with the two procedures of reduction/cyanylation and cyanylation/reduction are shown in Table 2.2.

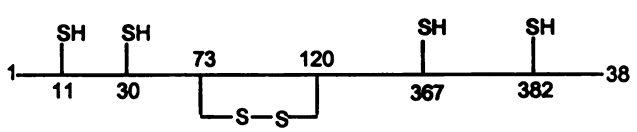


Figure 2.8. Structure of Ovalbumin

Table 2.2. Expected and Observed Masses of Cleavage Product of Ovalbumin

Reduction/Cyanylation

Fragments	Expected Mass(Da)	ESI(Da)	MALDI(Da)
1-10	1011.1	1011.0	no
itz11-29	2378.7	2377.4	2377.4
itz30-72	4762.5	4760.2	4761.4
itz73-119	5427.1	5426.9	5426.1
itz120-366	27581.1	27581.0	no
itz367-381	1715.1	1713.8	1714.8
itz382-385	429.5	430.0	no
*1-29	3312.8	3311.2	3311.2
itz*30-119	10112.6	10108.3	no
itz*367-385	2067.6	2066.5	2068.1
itz*120-381	29219.2	29198.7	no

Cyanylation/Reduction

Fragments	Expected Mass(Da)	ESI(Da)	MALDI(Da)
1-10	1011.1	1011.0	no
itz11-29	2378.7	2377.1	2378.5
itz30-366	37684.7	37684.0	no
itz367-381	1715.1	1714.0	1715.2
itz382-385	429.5	430.0	no

* β-elimination

itz = iminothiazolidine derivative.

The spectra of MALDI-MS with the two procedures of CDAP/TCEP and TCEP/CDAP are presented in Figure 2.9A and 2.9B, respectively. MALDI-MS identified peaks of itz11-29 (m/z 2377.4), itz30-72 (m/z 4761.4), itz73-119 (m/z 5426.1), itz367-381 (m/z 1714.8) and β-elimination peaks of *1-29 (m/z 3311.2), itz*367-385 (m/z 2068.1) resulting from procedure TCEP/CDAP (Figure 2.9B). The expected peaks for fragments 1-10 at m/z 1011.1, itz382-385 at m/z 429.5 and itz120-366 at m/z 27581.1 in procedure TCEP/CDAP were undetected. Instead of five expected fragment peaks resulting from procedure CDAP/TCEP, only three peaks were identified by MALDI-MS, the expected peaks at m/z 1011.1 and m/z 37684.7 were undetected. Therefore, MALDI-MS gave an incomplete mass-mapping for protein ovalbumin. The location of cysteines and cystines in ovalbumin cannot be recognized unambiguously.

The HPLC chromatogram and corresponding ESI spectra of cleavage mixture resulting from procedure TCEP/NTCB are shown in Figure 2.10 and Figure 2.11 respectively. Fragments of 1-10, itz11-29, itz30-72, itz73-119, itz120-366, itz367-381, an itz382-385 have been all identified. Therefore, total numbers and locations of cysteine residues in ovalbumin can be recognized unambiguously by HPLC/ESI-MS.

The HPLC chromatogram and corresponding ESI-MS spectra of cleavage products resulting from procedure of NTCB/TCEP are presented in Figure 2.12 and Figure 2.13, respectively. The fragments of 1-10, itz11-29, itz30-366, itz367-381, and itz382-385 were identified, indicating the location of free cysteine residues in ovalbumin at Cys 11, Cys 30, Cys 367, and Cys 382. By comparing the results of procedures of TCEP/NTCB and NTCB/TCEP, the location of cystine in ovalbumin should be at Cys 73 and Cys 120.

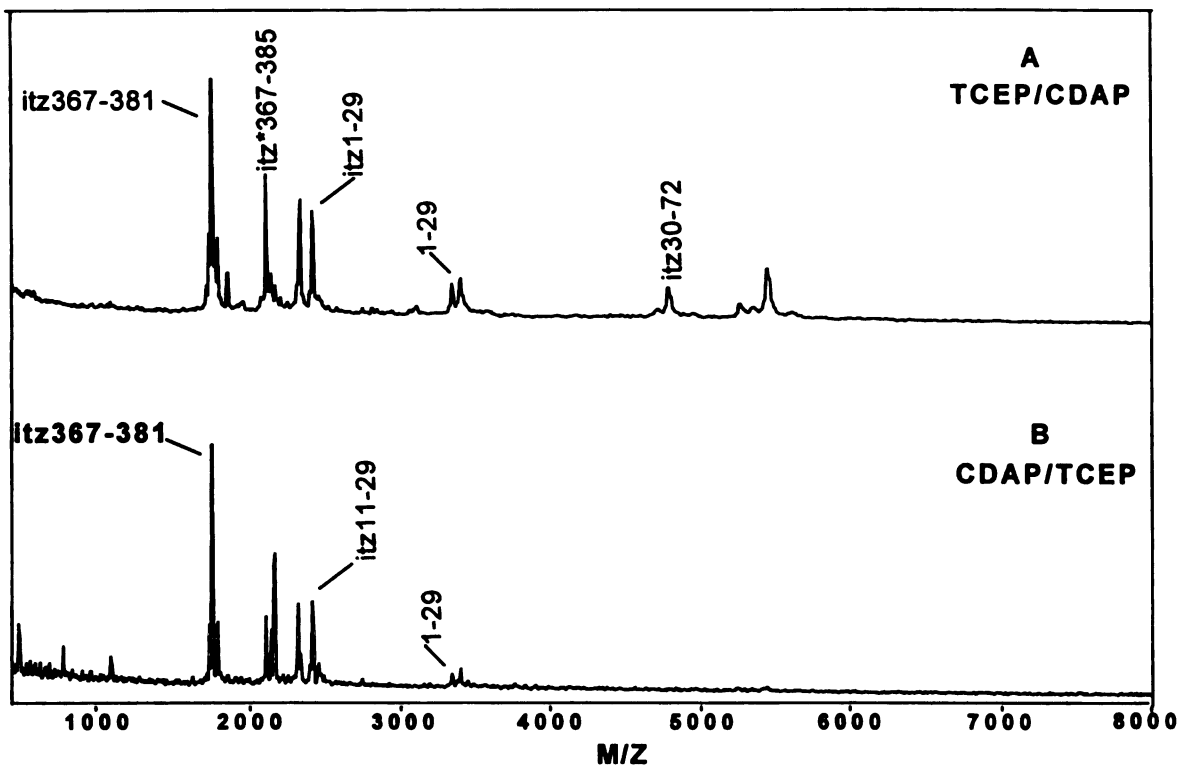


Figure 2.9. MALDI spectra of cleavage products of ovalbumin following reduction/cyanylation/cleavage (panel A) or cyanylation/cleavage/reduction (panel B).
itz = iminothiazolidine derivative.

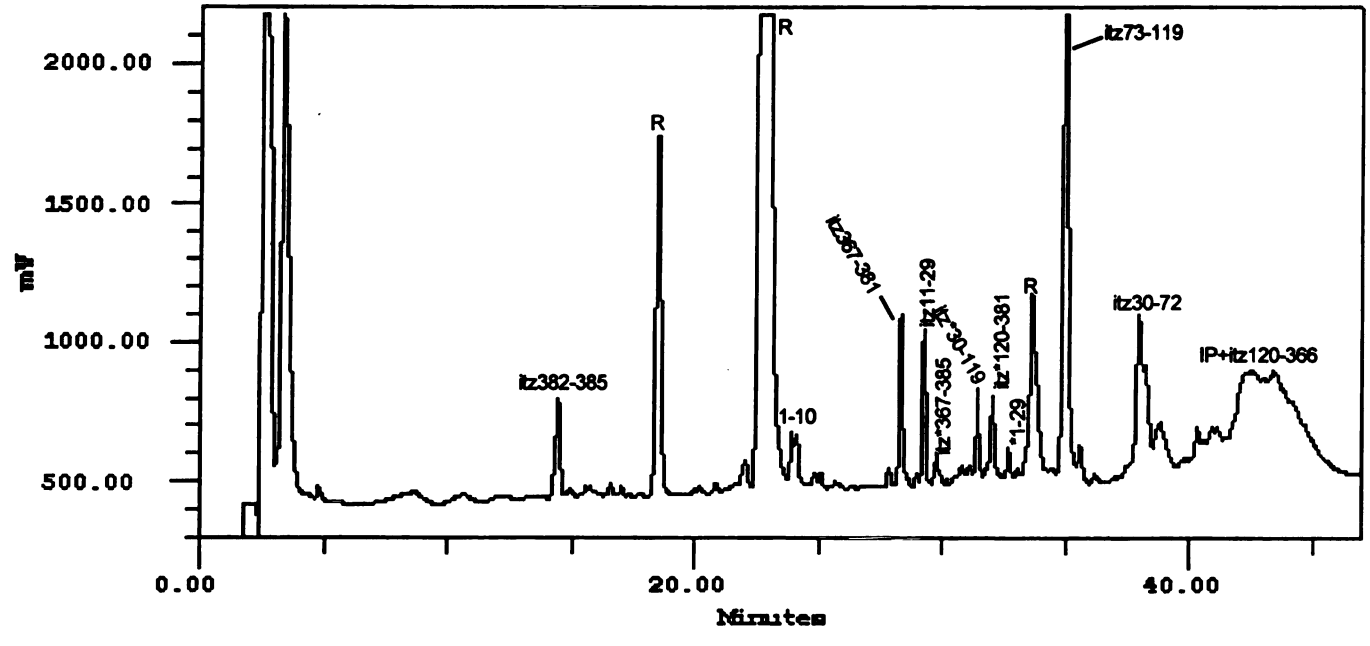


Figure 2.10. HPLC chromatogram of mixtures of cleavage product following procedure TCEP/NTCB of ovalbumin.

Solvent A: 0.1% TFA; solvent B: 90% acetonitrile containing 0.1% TFA.

Gradient: 0%-70% B linear in 50 min.

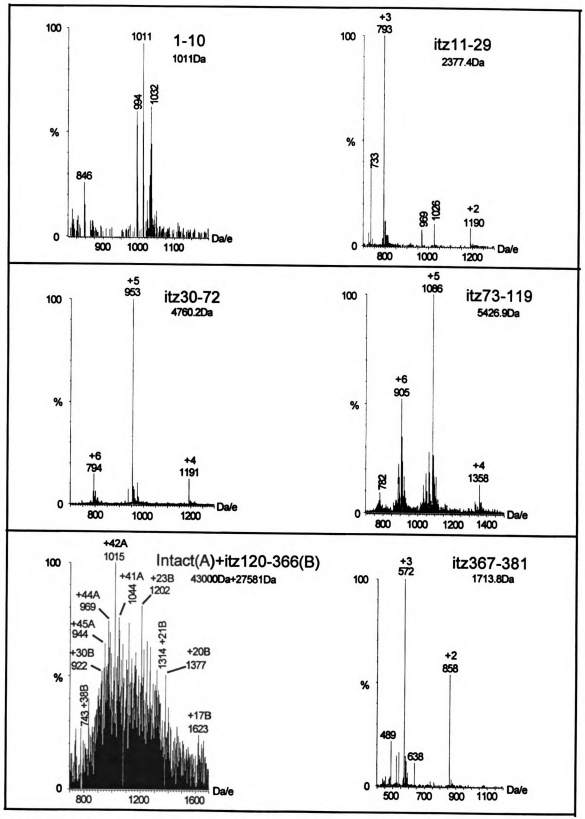


Figure 2.11. ESI spectra of cleavage products following procedure TCEP/NTCB of ovalbumin

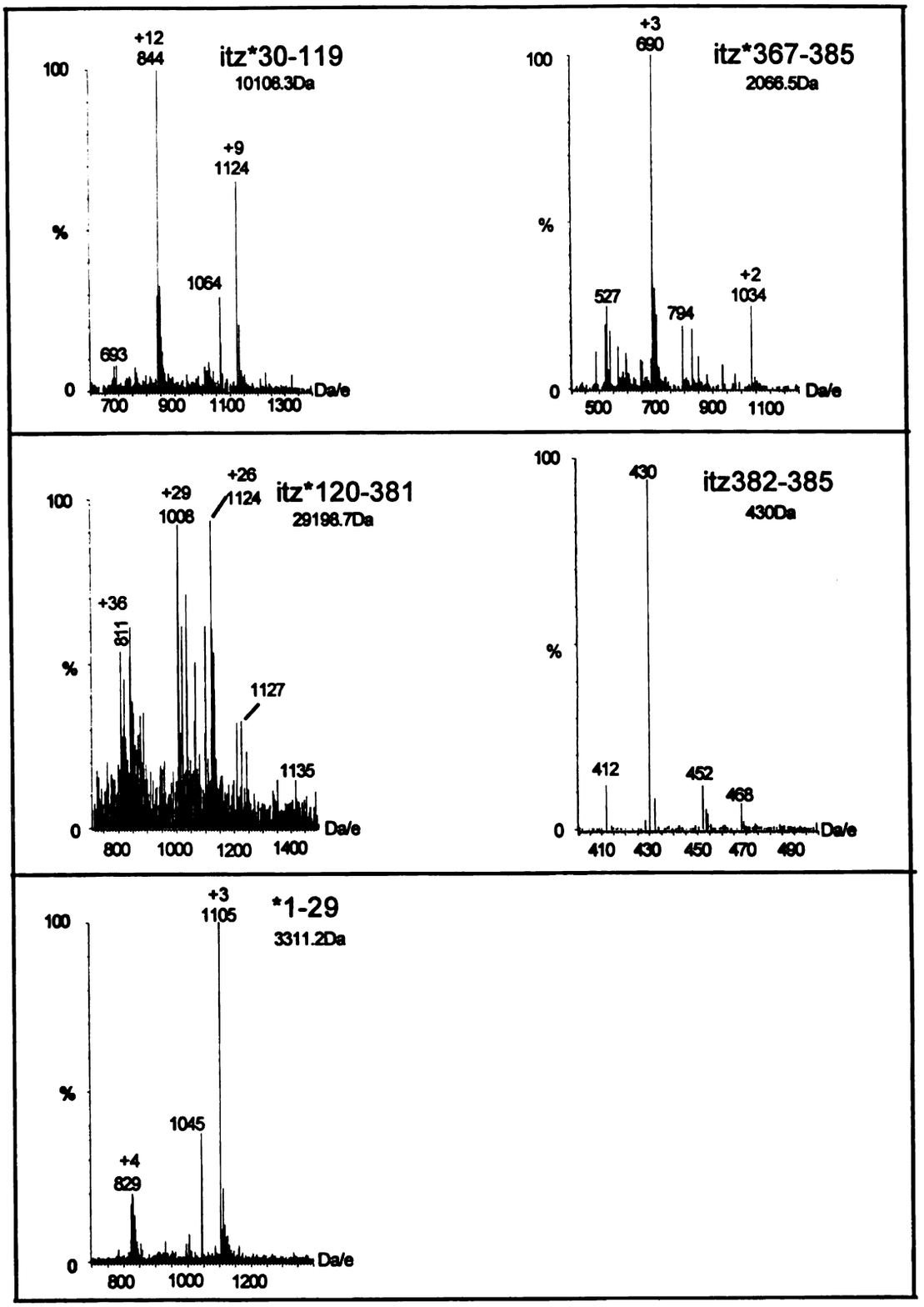


Figure 2.11. Continue

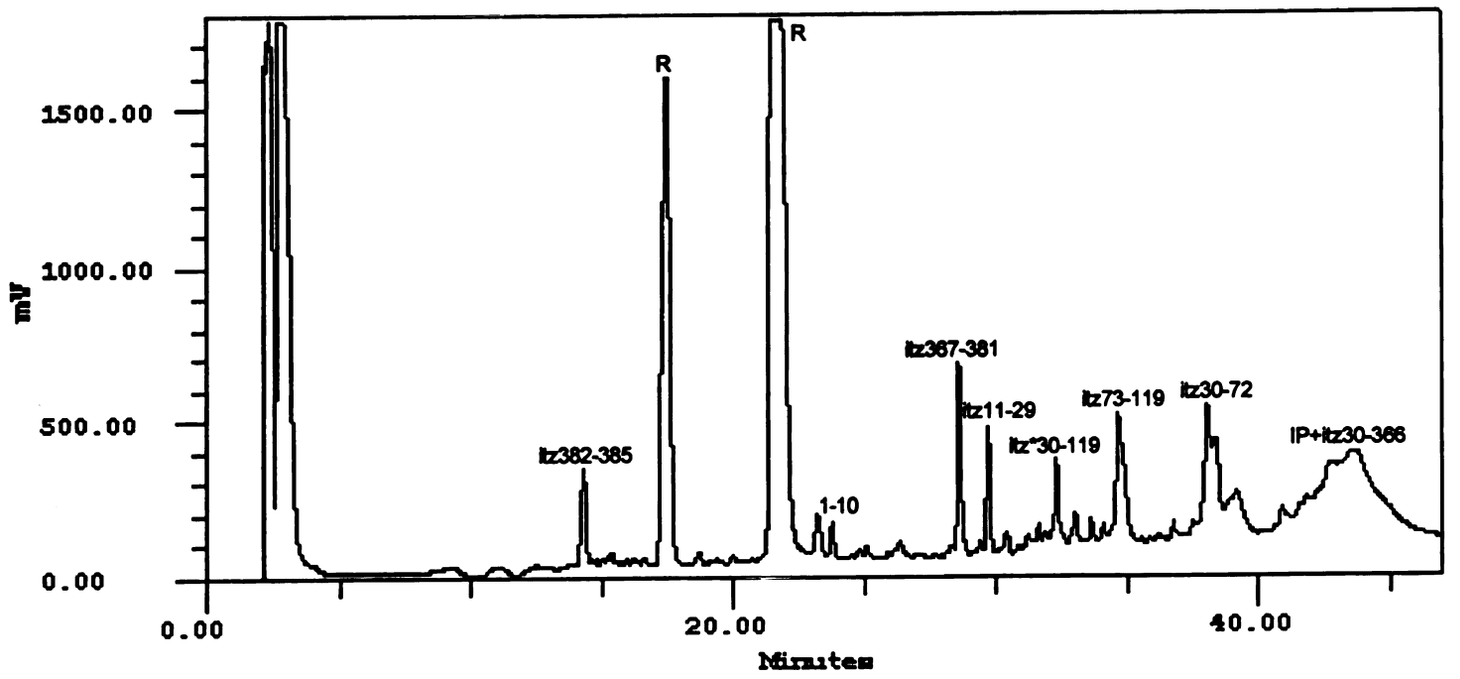


Figure 2.12. HPLC chromatogram of mixture ofcleavage products following the procedure NTCB/TCEP of ovalbumin.

Solvent and gradient conditions are the same as described in Figure 2.10.

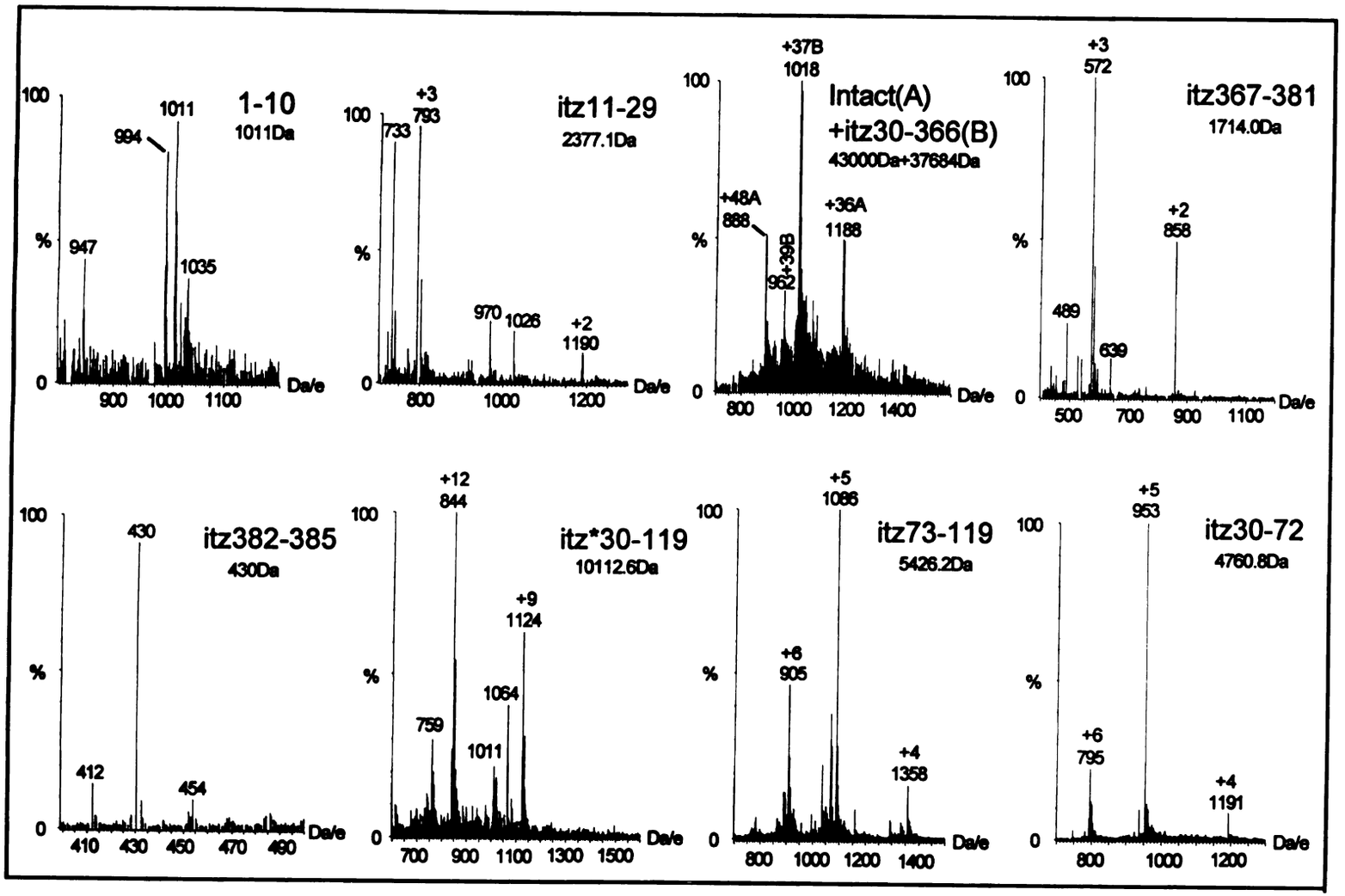


Figure 2.13. ESI spectra of cleavage products following the procedure NTCB/TCEP of ovalbumin

Because ovalbumin only contains one disulfide bond, the linkage of disulfide bond must be between Cys 73 and Cys 120.

Comparing the results of MALDI and ESI, several aspects could be concluded:

1). All the undetected peaks by MALDI-MS were identified by HPLC/ESI-MS (summarized in Table 2.2). Therefore, HPLC / ESI-MS gave the complete mass mapping for ovalbumin.

2). Peak of ESI at m/z 37684.7 in procedure NTCB/TCEP and peak of ESI at m/z 27581 in procedure TCEP/NTCB are very noisy which are due to the broad HPLC peaks that include the intact ovalbumin.

3). Compare the two HPLC chromatograms for procedure TCEP/NTCB and NTCB/TCEP. The procedure NTCB/TCEP also forms the peaks for fragments itz30-72 and itz73-119 at m/z 4760.8 and m/z 5426.2 which should not appear in the procedure of NTCB/TCEP, although they have lower intensity. Therefore, the cyanylation reaction with NTCB at pH 8.7 induced the thiol-disulfide exchange of ovalbumin.

C. Evaluation of CDAP and NTCB

Based on the results of ovalbumin in Section B, NTCB induced the thiol/disulfide exchange of ovalbumin in procedure of cyanylation/reduction. In order to evaluate the performance of NTCB and CDAP. We used CDAP to perform the same procedure of CDAP/TCEP as that of NTCB described in section B. The HPLC chromatograms and corresponding ESI spectra of the cleavage product from procedure CDAP/TCEP of ovalbumin are illustrated in Figure 2.14 and Figure 2.15, respectively.

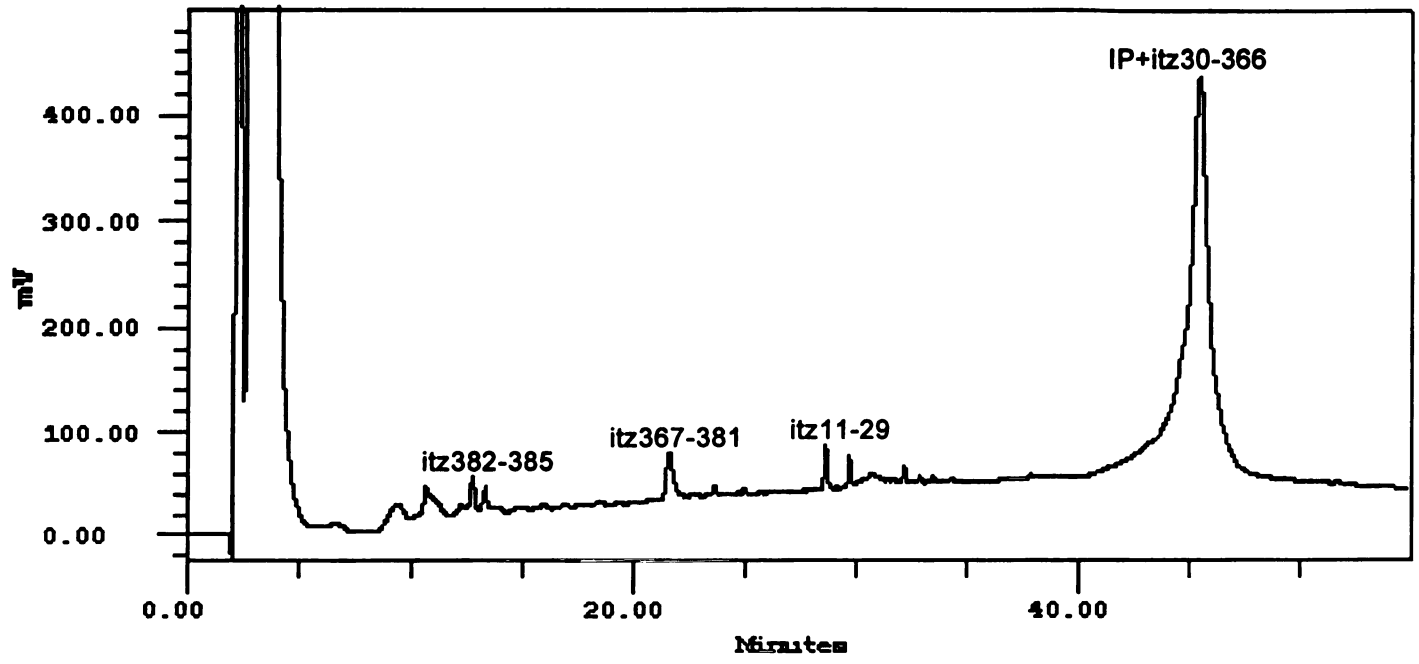


Figure 2.14. HPLC chromatogram of mixture of cleavage product from procedure CDAP/TCEP of ovalbumin

Solvent and gradient conditions were the same as described in Figure 2.10.

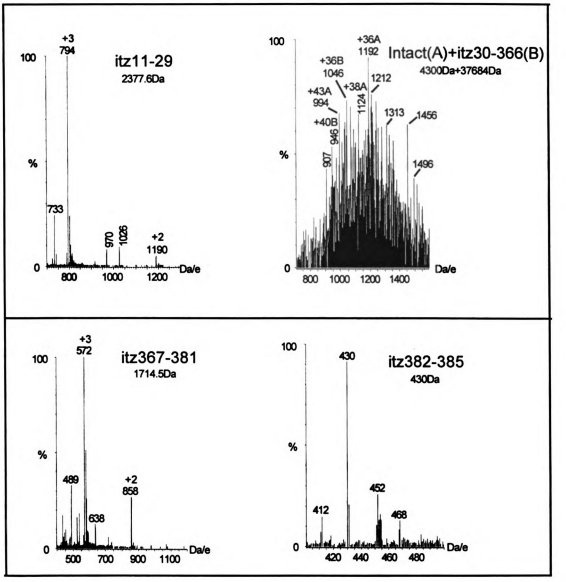


Figure 2.15. ESI spectra of cleavage product from procedure CDAP/TCEP of ovalbumin

The results are summarized in Table 2.3. Several aspects can be concluded:

1). No peaks for fragments itz30-72 and itz73-119 at m/z 4762.5 and m/z 5427.1 resulting from the procedure CDAP/TCEP were identified. Therefore, CDAP reaction did not induce the thiol-disulfide exchange.

2). CDAP has the lower capacity to cyanylate free cysteines. The fragment 1-10 of ovalbumin cannot be detected by HPLC by using 20-fold excess of CDAP.

3). By comparing the results from NTCB and CDAP, we can obtain clear assignment for cysteines and cystines in ovalbumin.

C. Conclusions

- The use of NTCB or CDAP is unique in that it specifically targets the site being analyzed.
- Since CDAP reaction is performed at pH 3, it eliminates the thiol-disulfide exchange in the analysis.
- HPLC separation / ESI-MS overcomes the signal suppression phenomenon of MALDI-MS; it can become a complementary method to MALDI-MS. The ESI method can be further used with on-line LC/MS.

Table 2.3. Comparison of CDAP and NTCB

Fragments	CDAP/TCEP(Mass: Da)	NTCB/TCEP(Mass: Da)
1-10	no	1011.0
itz11-29	2377.6	2377.1
itz30-366	37684.0	37684.0
itz367-381	1714.5	1714.0
itz382-385	430.0	430.0
#itz*30-119	no	10112.6
#itz73-119	no	5426.2
#itz30-72	no	4760.8

#: fragments due to thiol-disulfide exchange. *: β-elimination

itz = iminothiazolidine derivative.

References:

1. Catsimpoolas, N., and Wood, J. L., *J. Biol. Chem.*, 241, 1790 (1966).
2. Jacobson, G. R., Schaffer, M. H., Stark, G. R., and Vanaman, T. C., *J. Biol. Chem.*, 248, 6583 (1973).
3. Degani, Y., and Patchornik, A., *Biochemistry*, 13, 1 (1974).
4. Wu, J., and Watson, J. T., *to appear in Anal. Biochem.* (1998).
5. Wakeselman, M., Guibe-Jampel, E., Raoult, A., and Busse, W. D., *J. Chem. Soc. Chem. Commun.*, 1, 21(1976).
6. Brockleehurst, K., Malthouse, J. P. G., Baines, B. S., Blenkinsop, R. D., Churcher, J. A., Mushiri, M. S., and Ormerod, F., *Biochem. Soc. Transactions*, 6, 261 (1978).
7. Brocklehurst, K., Malthouse, J. P. G., *Biochem. J.*, 175, 761 (1978).
8. Walkselman, M., and Guibe-Jampel, E., J. C. S. Chem. Comm., 21 (1976).
9. Casey, R., and Lang, A., *Biochem. J.*, 145, 251 (1975).
10. Price, N. C., *Biochem. J.*, 159, 177 (1976).
11. Schaffer, M. H., and Stark, G. R., *Biochem. Biophys. Res. Comm.*, 71, 1040 (1976).
12. Lu, H. S., and Gracy, R. W., *Arch. Biochem. Biophys.*, 212 (2), 347 (1981).
13. Thomas, M. L., Janatova, J., Gray, W. R., and Tack, B. F., *Proc. Natl. Acad. Sci. USA*, 79, 1054 (1982).
14. Roberts, D. D., and Goldstein, I. J., *J. Biol. Chem.*, 259, 909 (1984).
15. Matsudaira, P., Jakes, R., Cameron, L., and Atherton, E., *Proc. Natl. Acad. Sci. USA*, 82, 6788 (1985).
16. Mahboub, S., Richard, C., Delacourte, A., and Han, K., *Anal. Biochem.*, 154, 171 (1986).

17. Bahler, M., Benfenati, F., Valtorta, F., Czernik, A. J., and Greengard, P. *J., Cell Biol.*, 108, 1841 (1989).

18. Minard, P., Desmadril, M., Ballery, N., Perahia, D., and Mouawad, L., *Eur. J. Biochem.*, 185, 419 (1989).

19. Nefsky, B., and Bretscher, A., *Proc. Natl. Acad. Sci. USA*, 86, 3549 (1989).

20. Eshdat, Y., Chapot, M. P., and Strosberg, A. D., *FEBS Letters*, 1-2, 166 (1989).

21. Katayama, E., *J. Biochem.*, 106, 988 (1989).

22. Sutherland, C., and Walsh, M. P., *J. Biol. Chem.*, 264, 578 (1989).

23. Tang, C., Yuksel, K. U., Jacobson, T. M., and Gracy, R. W., *Arch. Biochem. Phys.*, 283(1), 12 (1990).

24. May, J. M., Buchs, A., and Carter-Su, C., *Biochem. J.*, 29, 10393 (1990).

25. Wines, B. D., and Easterbrook-Smith, S. B., *Molecular Immunology*, 28 (8), 855 (1991).

26. Som, S., and Friedman, S., *J. Biol. Chem.*, 266, 2937 (1991).

27. Altamirano, M. M., Plumbridge, J. A., and Calcagno, M. L., *Biochem. J.*, 1153 (1992).

28. Goold, R., and Baines, A. J., *Eur. J. Biochem.*, 224, 229 (1994).

29. Papayannopoulos, I. A., and Biemann, K., *Protein Sci.*, 1, 278 (1992).

30. Wu, J., Gage, D. A., Watson, J. T., *Anal. Biochem.*, 235, 161 (1996).

31. Karas, M., and Hillenkamp, F., *Anal. Chem.*, 60, 2299 (1988).

32. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M., *Science*, 246, 64 (1989).

33. Muddiman, D. C., Bakhtiar, R., Hofstadler, S., and Smith, R. D., *J. Chem. Edu.*, 74, 1288 (1997).

34. Stark, G. R., *Methods in Enzymology*, vol. 47, 129 (1977).

35. Denslow, N. D., and Nguyen, H. P., in *Techniques in Protein Chemistry* VII, Academic Press, Inc., (1996), p.241-248.

36. Price, N. C., *Biochem. J.*, 159, 177 (1976).

37. Nakagawa, S., Tamakashi, Y., Hamana, T., Kawase, M., Taketomi, S., Ishibashi, Y., Nishimura, O., Fukuda, T., *J. Am. Chem. Soc.*, 116, 5513 (1994).

38. Wu, J., and Watson, J. T., *Protein Sci.*, 6, 391 (1997).

CHAPTER 3

DISULFIDE MASS MAPPING IN PROTEINS CONTAINING ADJACENT CYSTEINES WITH CYANYLATION/CLEAVAGE METHODOLOGY

I. Introduction

Disulfide bond formation between cysteine residues is a posttranslational event that commonly occurs in proteins synthesized in the endoplasmic reticulum. In many cases, disulfide bond formation contributes to the stability of the tertiary structure of the folded protein molecule (1). A full description of the covalent structure of proteins demands that the connectivity of the bridged cysteines be analyzed. Although there are good methods for quantifying the number of disulfide bonds in proteins, the unambiguous determination of the location or pairing of disulfide bonds continues to be a challenging task for the protein chemists.

The general strategy for locating disulfide bonds in proteins by conventional methods involves several steps (2, 3). First, a protein is cleaved by enzymes or chemical reagents between half-cystinyl residues to obtain peptides that contain only one disulfide bond. Second, the amino acid composition, amino terminal sequence or molecular masses of these peptides are determined using an amino acid analyzer, manual or automated Edman degradation, and/or mass spectrometry. Finally, peptides identified using these data are related to specific segments of the protein (3). There are two major problems are

related to the methodology. First, if a protein contains a pair of adjacent or closely spaced cysteines in its primary structure, it is impossible to obtain peptides containing a single disulfide bond (3). Second, the conditions most frequently used in specific enzymatic digestions may promote disulfide bond scrambling (4). Although supplementary methodologies such as non-specific fragmentation by partial acid hydrolysis (2) have been proposed to avoid disulfide bond scrambling, it is difficult to deal with the data obtained by these non-specific techniques.

Different approaches have been made to assign the disulfide linkages of proteins containing adjacent or closely spaced cysteine residues (5-17). Among them, the methodology proposed by Gray (16, 17) has the capacity to analyze proteins containing adjacent or closely spaced cysteine residues. In his experiments, peptides were partially reduced, the nascent free thiols alkylated, and their positions recognized from the results of sequence analysis. However, it is obviously tedious, if not impractical, to sequence an alkylated high-mass peptide or protein using this approach.

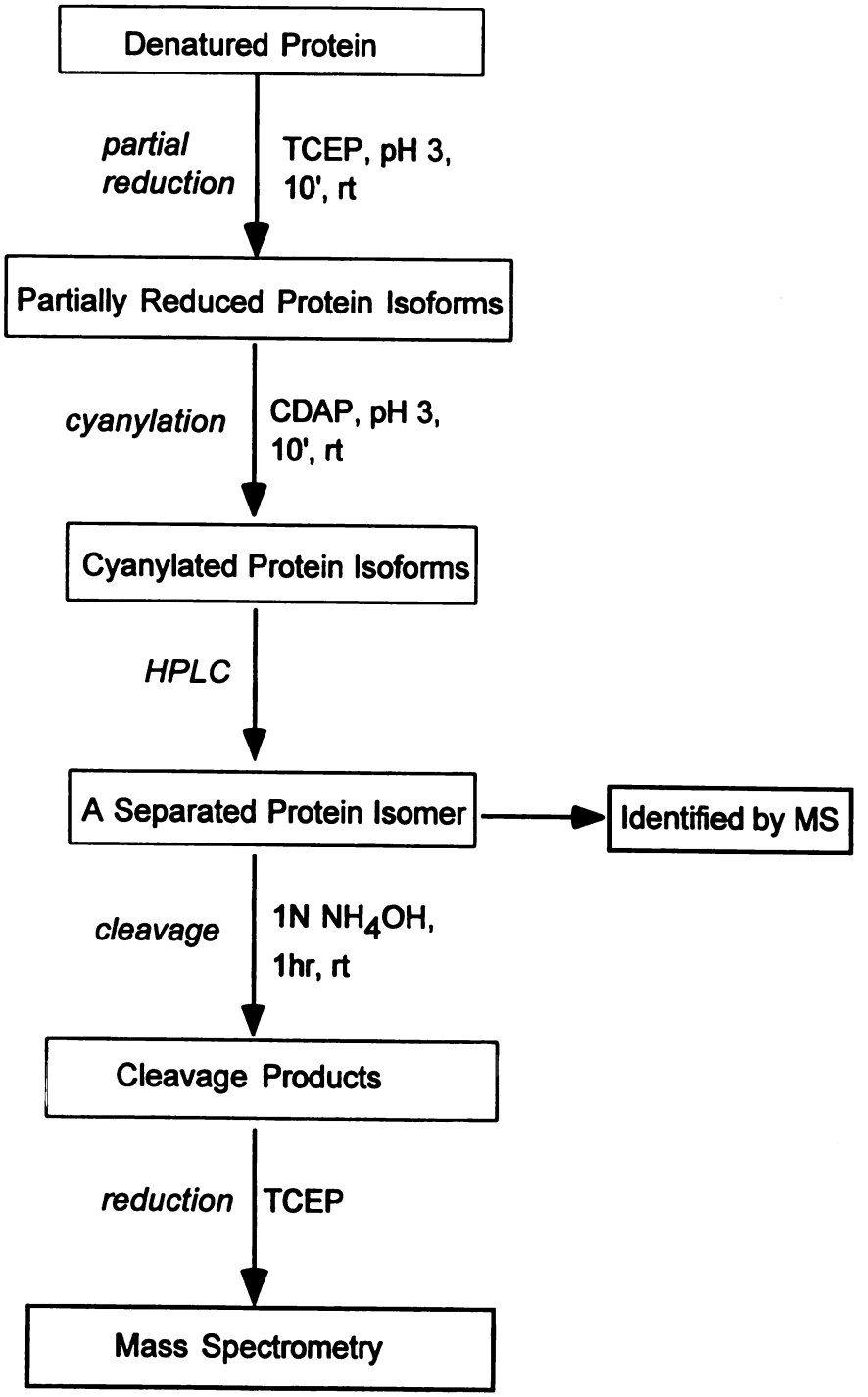
Recently, Wu and Watson described a novel approach for the assignment of disulfide bonds in proteins of known sequence (18). In this approach, the denatured protein was subjected to limited reduction by tris(2-carboxyethyl)phosphine (TCEP) in pH 3.0 citrate buffer to produce a mixture of partially reduced protein isomers; the nascent sulfhydryls were immediately cyanylated by 1-cyano-4-dimethylamino-pyridinium tetrafluoroborate (CDAP) under the same buffered conditions. The cyanylated protein isomers, separated by and collected from reversed-phase HPLC, were subjected to cleavage of the peptide bonds on the N-terminal side of cyanylated cysteines in aqueous ammonia to form truncated peptides that were still linked by residual disulfide

bonds. The remaining disulfide bonds were then completely reduced to give a mixture of peptides that can be mass mapped by MALDI-MS. The masses of the resulting peptide fragments were related to the location of the paired cysteines that had undergone reduction, cyanylation, and cleavage. This strategy minimizes the disulfide bond scrambling and is simple, fast, and sensitive. Furthermore, it is possible to assign the disulfide linkages of proteins containing adjacent cysteines by this new methodology.

In this chapter, the applicability of partial reduction, cyanylation, chemical cleavage, and mass mapping approach to the disulfide structure analysis of proteins containing adjacent cysteine residues will be demonstrated. Two model proteins, long R^3 insulin-like growth factor-I and insulin-like growth factor-I, will be used to show the capability of the new approach for the determination of disulfide structure involved in adjacent cysteine residues.

II. Methodology for Assignment of Disulfide Linkages

The methodology utilized to determine disulfide linkages of proteins containing adjacent cysteines are based on the strategy proposed by Wu & Watson (18). The approach is illustrated in scheme 3.1. The disulfide bonds of a denatured protein are partially reduced with TCEP by controlled TCEP concentration and reducing time in a pH 3.0 buffer. Then the resulting nascent sulfhydryls will be cyanylated by CDAP in the same buffer solution. After the separation of partially reduced/cyanylated protein isomers, HPLC fractions can be analyzed by MALDI-MS or ESI-MS to determine which isomers are singly reduced and cyanylated. Those shifted by +52 Da correspond to a singly reduced/cyanylated species; +104 Da correspond to doubly reduced/cyanylated



Scheme 3.1. Conceptual overview of the methodology for assignment of disulfide bond pairings in proteins

species, etc. Those isomers with a 52-Da mass shift from the mass of the intact protein are dried and subjected to specific chemical cleavage in aqueous ammonia. The cleaved peptides, which may be linked by residual disulfide bonds, are then completely reduced to give a mixture of peptides that can be mass mapped by either MALDI-MS or LC/ESI-MS. The masses of the resulting peptide fragments are related to the location of the paired cysteines that had undergone reduction, cyanylation, and cleavage.

III. Experimental Section

MALDI-TOF MS

MALDI mass spectra were obtained on a Voyager Elite time-of-flight (TOF) mass spectrometer (PerSeptive Biosystems Inc., Framingham, MA) equipped with delayed extraction and a model VSL-337ND nitrogen laser (Laser Science, Newton, MA). The accelerating voltage in the ion source was set to 20 kV. Grid voltage and guide wire voltages were 93.6% and 0.2% of the accelerating voltage, respectively. Data were acquired in the positive linear DE mode of operation. Time-to-mass conversion was achieved by external and/or internal calibration using standards of bradykinin (m/z 1061.2), bovine pancreatic insulin (m/z 5734.5), and horse skeletal myoglobin (m/z 16952) obtained from Sigma Chemical Co. (St. Louis, MO). All experiments were performed using α-cyano-4-hydroxycinnamic acid (Aldrich Chemical Co., Milwaukee, WI) as the matrix. Saturated matrix solutions were prepared in a 50% (v/v) solution of acetonitrile/aqueous 1% TFA, and mixed in equal volumes with peptide or protein

samples, and applied to a stainless-steel sample plate. The mixture was allowed to air dry before being introduced into the mass spectrometer.

LC/ESI-MS

The LC system used for LC/ESI-MS analysis was Perkin-Elmer API Applied Biosystems 173. The separation was carried out on a C18 capillary column. Solvent A was 0.085% TFA in water and solvent B was 0.085% TFA in CH_3CN. The linear gradient was 5% B to 65% B in 135 minutes at a flow rate of 7 µl/min. ESI mass spectra were obtained on-line on a Micromass Platform-LC mass spectrometer. This instrument was equipped with a single quadrupole mass analyzer. The capillary voltage was set at 3.25KV, cone voltage at 30V, and source temperature at 80^0C.

HPLC

The HPLC is equipped with Waters model 6000 pumps controlled by a personal computer. UV detection was at 215nm. The column was a Vydac C18 (#218TP54, 10µm particle size, 300-Å pore, 4.6×250mm).

Chemicals

Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Pierce, Rockford, IL. Guanidine hydrochloride was a product of Boehringer-Mannheim Biochemicals (Indianapolis, IN). Human recombinant long R^3 insulin-like growth factor I (LR^3IGF-I), 2-nitro-5-thiocyanobenzoic acid (NTCB), 1-cyano-4-dimethylamino-pyridinium tetrafluoroborate (CDAP), were purchased from Sigma. Human recombinant

insulin-like growth factor I was obtained from Austral Biologicals (San Ramon, CA). The proteins were purified before using. Acetonitrile and TFA were HPLC grade. The TCEP solution in 0.1 M citrate buffer at pH 3 was prepared as 0.1 M stock solution and stored under N_2 at -20^0C for weeks with little deterioration. The 0.1 M CDAP solution in pH 3.0, 0.1M citrate buffer was freshly prepared prior to use.

Partial Reduction of Proteins

10nmol Protein samples (IGF-I and LR^3IGF-I) were solubilized in 40µl of 0.1M citrate buffer (pH 3.0) containing 6 M guanidine-HCl. Partial reduction of disulfide bonds were carried out by adding 10 molar fold of TCEP for the cystine content of the proteins (300nmol of TCEP is reacted with 10nmol of IGF-I or LR^3IGF-I as 1nmol of IGF-I or LR^3IGF-I contains 3nmol of cystine). The reduction was conducted at room temperature for 10 minutes.

Cyanylation of Nascent sulfhydryls

To the partially reduced protein mixture was added a 30-fold excess of CDAP over the cysteine content of IGF-I or LR^3IGF-I. Cyanylation of the nascent sulfhydryl groups was accomplished by incubation at room temperature for another 10 minutes.

HPLC Separation of Partially Reduced and Cyanylated Protein Isomers

Partially reduced and cyanylated species were separated by reversed-phase HPLC with linear gradient of 30% B to 50% B in 45 minutes, where solvent A was 10% CH_3CN containing 0.1% TFA and solvent B was 90% CH_3CN containing 0.1% TFA. The

predominant HPLC fractions were collected manually and the masses of the collected protein isomers were determined by MALDI-MS or ESI-MS. Appropriate fractions were then dried for further use.

Cleavage of Singly Reduced and Cyanylated Protein Isomers

To the dried HPLC fractions was added 2μl of 6M guanidine-HCl in 1M NH_4OH aqueous solution to dissolve the protein residue and then 5 μl of 1M NH_4OH. Cleavage of the peptide chain was performed at room temperature for one hour. Excess ammonia was removed in a vacuum system.

Complete Reduction of Remaining Disulfide Bonds

Truncated peptides, still linked by residual disulfide bonds, were completely reduced by reacting with 2 μl of 0.1M TCEP solution at pH 3.0, 0.1M citrate buffer at 37^0C for 30 minutes. Samples were diluted with 100 μl of a 50% (v/v) CH_3CN/1% TFA for MALDI-MS analysis or 100 μl of 5% (v/v) CH_3CN/0.085% TFA for LC/ESI-MS analysis.

IV. Results and Discussion

A. Disulfide Mapping of IGF-I

Insulin-like growth factor I (IGF-I) (Mr = 7648.6) is a single-chain polypeptide of 70 amino acid residues containing three intramolecular disulfide bonds, two of which involve adjacent cysteines (19) (Figure 3.1). IGF-I is postulated to be the mediator of

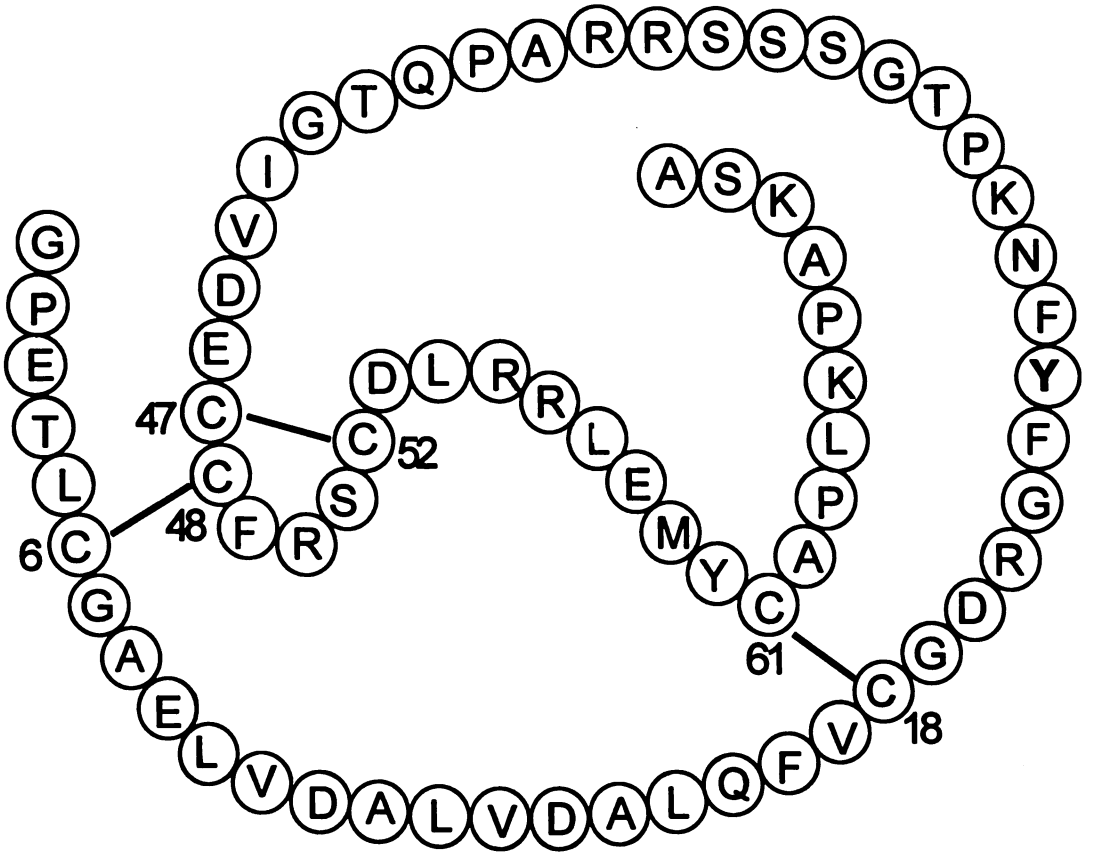


Figure 3.1. Amino acid sequence and disulfide structure of IGF-I

growth hormone action on skeletal tissue as well as mitogenic activity on several cell types (20, 21). While the complete amino acid sequence of IGF-I was determined in 1978 (19), very little experimental data is available describing the secondary and tertiary structure of the molecule. A three-dimensional model of human IGF-I has been suggested first by Blundell *et al* in 1978 (22). This model is based on the proposed tertiary structure of porcine insulin as determined by X-ray crystallography (23). From this model, a disulfide bond arrangement between Cys6-Cys48, Cys47-Cys52 and Cys18-Cys61 was predicted for human IGF-I. However, since IGF-I contains three disulfide bonds, several isomeric forms can theoretically be formed, especially in the region with the two adjacent cysteine residues, i.e., Cys47 and Cys48.

Several approaches have been used to identify the disulfide linkages of IGF-I (24-27). While the conventional method of enzymatic digestion did not completely assign the disulfide linkages because of the two adjacent cysteines (24), chemical synthesis and the combination of enzymatic digestion and fast-bombardment mass spectrometry have been utilized to try to assign the linkages. The method based on the chemical synthesis determined disulfide linkages of IGF-I by comparing the retention time during HPLC chromatography of the enzymatic digested peptide fragments with these of chemically synthesized peptides. Two types of peptides related to two adjacent cysteines, Type I with Cys6-Cys47 and Cys48-Cys52 as well as Type II with Cys6-Cys48 and Cys47-Cys52, have been synthesized (25). The disulfide bond system of IGF-I was determined to be the Type II form. This method is tedious and only provided indirect experimental evidences. The other approaches utilized a combination of multiple steps of enzymatic digestion, Edman degradation, and mass spectrometry (26, 27). Raschdorf *et al.* assigned

the disulfide linkages of IGF-I by a three-step mass spectrometric analysis (26). Firstly, the correct molecular weight of the intact protein was determined by fast atom bombardment (FAB) mass spectrometric analysis. Secondly, two-fold enzymatic degradation (chymotrypsin followed by V_8 protease, FAB mapping of the cleavage products) was employed to assign the disulfide linkage of Cys18-Cys61. Thirdly, FAB tandem mass spectrometry and manual Edman degradation were further used to analyze the peptide fragment, which contained two other disulfide bonds and two adjacent cysteines. Axelsson *et al.* (27) utilized the combination of enzymatic digestion, reversed-phase HPLC, FAB mass spectrometry, and one-cycle Edman degradation. They first digested the protein with V_8 protease, then analyzed it by FAB-MS to determine the disulfide linkage unrelated to adjacent cysteine residues. The peptide fragment involved in adjacent cysteine residues was further degraded by one-cycle Edman degradation, followed by trypsin digestion and FAB-MS analysis. The methodology consisting of multiple steps of enzymatic digestion, Edman degradation, and mass spectrometry is tedious and time consuming.

The disulfide linkages of IGF-I can be determined unambiguously by the new approach based on the partial reduction, cyanylation, chemical cleavage, and mass spectrometry (18). The overview of the methodology used to assign the disulfide linkages of IGF-I is illustrated in Figure 3.2.

Partial Reduction and Cyanylation of IGF-I

The partial reduction of IGF-I was carried out in citrate buffer (pH 3.0) containing 6 M Guanidine-HCl. Guanidine-HCl was utilized to completely denature the protein.

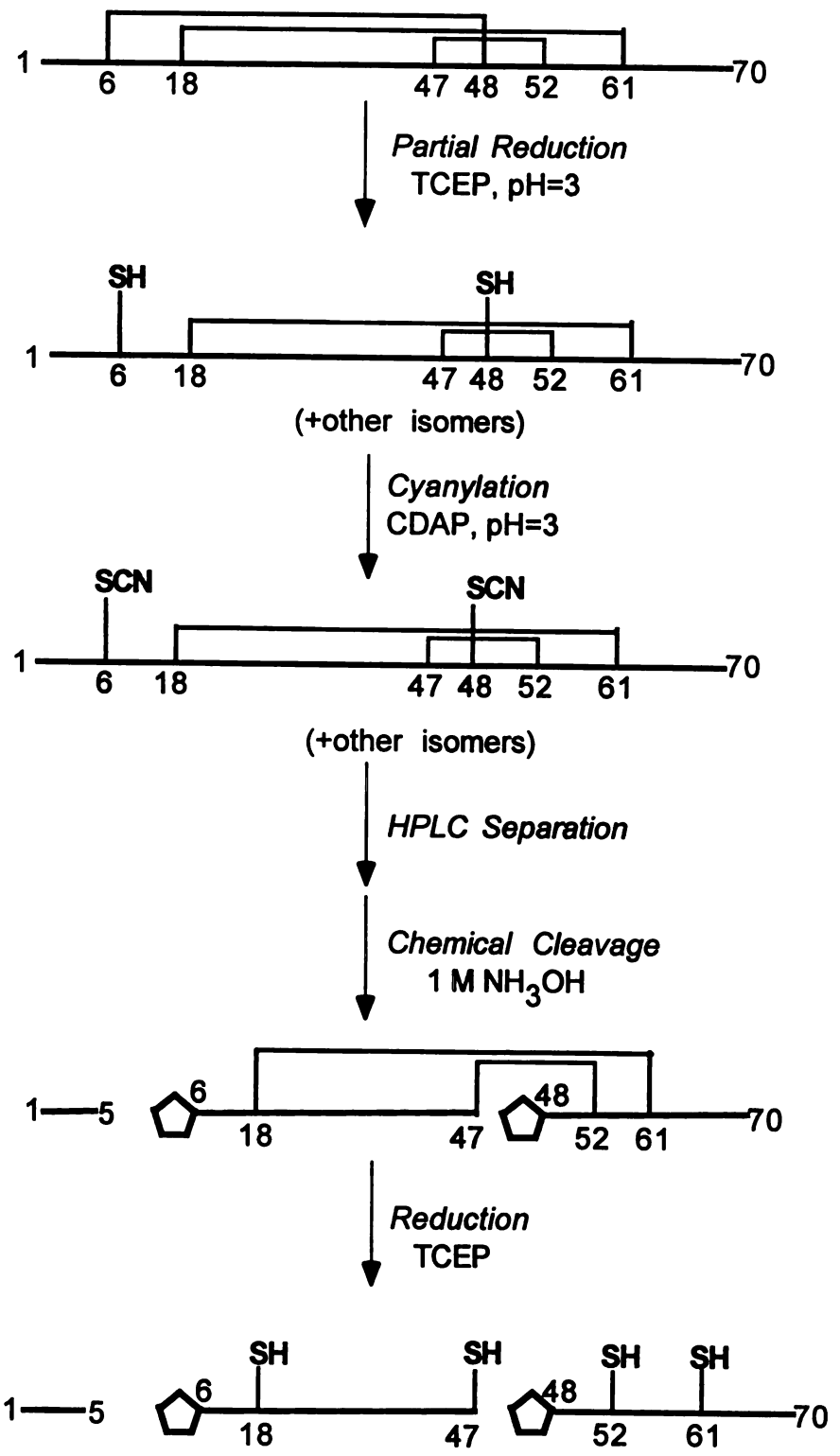


Figure 3.2. Overview of the methodology used to assign the disulfide linkages in IGF-I

Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was used as a reducing reagent since it has proved to be an excellent reducing agent for disulfide bonds (12, 17, 18, 28). Reduction by TCEP can be carried out at pH 3.0 to suppress disulfide bond scrambling. Furthermore, at pH 3.0, the reduction of disulfide bonds is kinetically controlled which makes partial reduction possible (17).

After the partial reduction of the protein, the nascent sulfhydryls can be cyanylated by CDAP in the same citrate buffer (pH 3.0). There are several advantages of using CDAP as the cyanylation reagent (18, 28). Firstly, cyanylation by CDAP can be carried out at pH 3.0, which is compatible with the partial reduction by TCEP. It is therefore unnecessary to remove excess TCEP, change buffer, or readjust the pH prior to using CDAP. Secondly, CDAP reacts instantly with TCEP even at pH 3.0, and the excess of TCEP in the reaction system will be consumed upon addition of CDAP, and no further reduction during cyanylation is possible. This is definitely an advantage for the control of reduction in an unknown protein. While a larger amount of CDAP has to be applied in case more TCEP is required for partial reduction of proteins with a tight structure, a large excess of CDAP (>50 fold molar excess over peptide sulfhydryls) and excessive incubation time (> 2 hours at room temperature) will result in some side reactions. Finally, CDAP is water soluble, which is compatible to the reaction system.

Because IGF-I consists of three disulfide bonds, use of the partial reduction technique should result in the production of three isoforms, each of which consists of a singly reduced isomer emanating from a different disulfide bond. However, as shown in the chromatogram (Figure 3.3) obtained after partial reduction and cyanylation, besides a major peak for the intact unreduced protein there are only two discernible peaks (peak 1

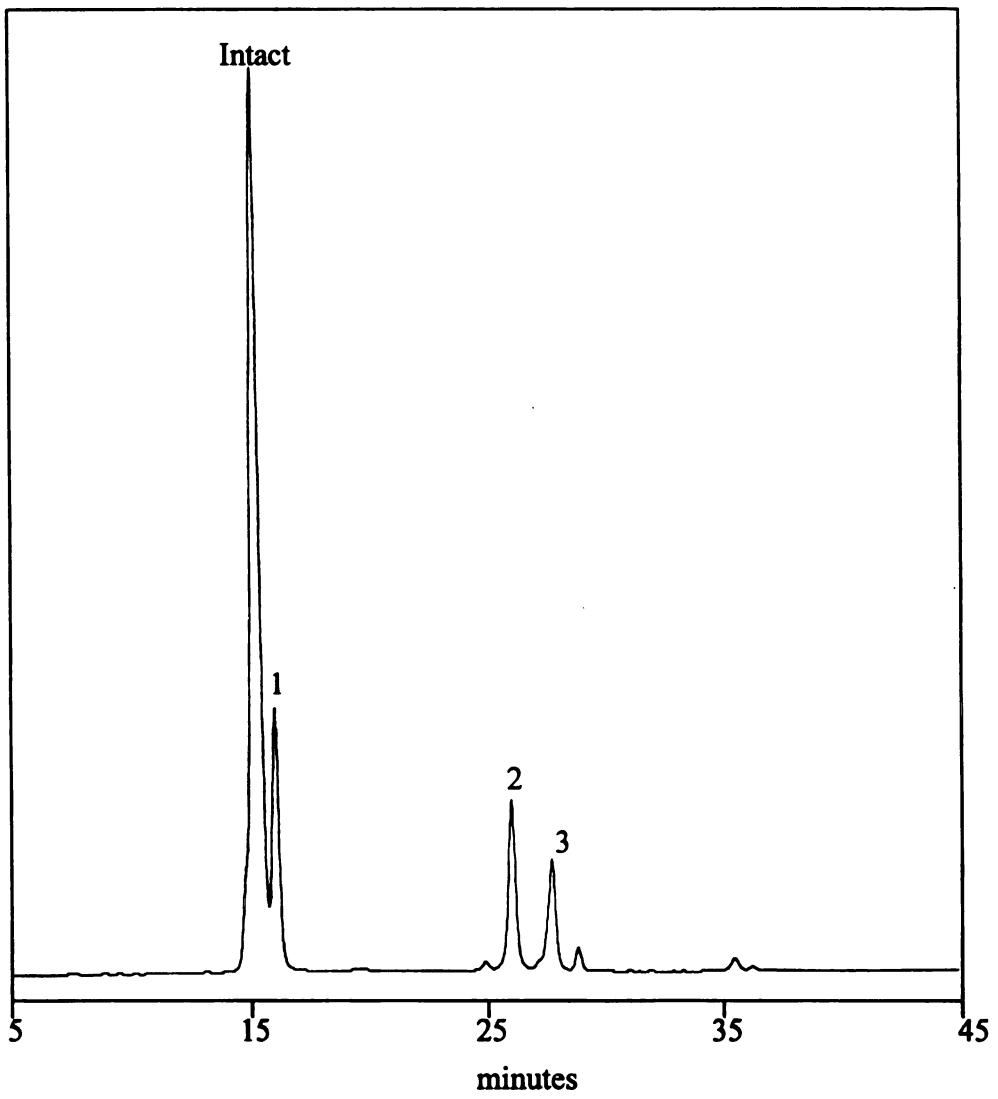


Figure 3.3. HPLC separation of denatured IGF-I and its partially reduced/cyanylated isomers. Separation was carried out on a Vydac C18 column at a flow rate of 1.0 ml/min with a linear gradient 30-50% B in 45 minutes, where A = 0.1% TFA in water and B = 0.1% TFA in CH_3CN. Peaks 1 and 2 represent singly reduced/cyanylated species. Peak 3 represent doubly reduced/cyanylated species, as determined by MALDI-TOF analysis.

and peak 2) each of which represents a singly reduced isoform as determined by mass analysis showing that the molecular weight of each species shifted by 52 Da from that of the intact protein (a 2-Da shift for conversion of a cystine to two cysteines followed by a 50-Da shift for the replacement of a hydrogen on each of two free sulfhydryls with a cyano group). Peak 3 resulted from the reduction of two disulfide bonds as indicated by a mass shift of 104 Da from the intact protein. This result is not unexpected, as the some disulfide bonds are much more stable than others as indicated in a recent study of the stability of the disulfide bond of Cys18 to Cys61 in IGF-I (29). This Cys18-Cys61 disulfide bond is preferentially formed in the folding process, a feature that may explain its resistance to reduction under our reaction conditions. In any case, in dealing with a protein known to contain three disulfide bonds, the determination of the pairing of two of the disulfide bonds allows the remaining one to be determined by default.

Cleavage of Peptide Chains

After the reversed-phase HPLC separation of partial reduced/cyanylated protein isomers, the peptide bonds at N-terminal side of cyanylated cysteine residues can be cleaved in 1N NH_4OH at room temperature for 1 hour in the presence of guanidine-HCl (18). These modified cleavage conditions have the advantage over the conventional cleavage method by using the mildly alkaline conditions (pH~9) (18, 30). Ammonia can be easily removed from the reaction system and the cleavage reaction is fast (in one hour).

After cleavage at cyanylated cysteinyl residues, the truncated peptide chains, still linked by the remaining disulfide bonds, can be easily reduced by excess TCEP. Finally,

the peptide mixture is diluted to minimize the adverse effect of guanidine on the subsequent analysis by MALDI-MS.

Interpretation of MALDI Data

Because cleavage of the peptide chain takes place only at cyanylated cysteinyl sites, there will be three fragments for each singly reduced/cyanylated protein isomer (and sometimes two overlapped fragments corresponding to β-elimination at one cysteine site). The mass of each fragment is related to the position of the two-cyanylated cysteinyl residues, which in turn can be used to deduce the disulfide bond linkage. β-Elimination, an alternative to peptide chain cleavage, provides mass spectral data corresponding to overlapped peptides (that otherwise would have cleaved) and serves as a confirmation for the disulfide bond pairing assignment.

Table 3.1 lists the calculated m/z values for possible fragments due to cleavage of the peptide chains at different sites depending on which disulfide bond was reduced and cyanylated. Figure 3.4A and B are two MALDI spectra of peptide mixtures resulting from cleavage of isomers of singly reduced/cyanylated IGF-I corresponding to HPLC peaks 1-2 in Figure 3.3, respectively. HPLC peak 3 is due to the reduction of disulfide bonds of Cys47-Cys52 and Cys6-Cys48 according to the MALDI spectra of the cleavage mixture of HPLC peak 3.

The mass spectrum in Figure 3.4A corresponds to the cleavage products represented by HPLC peak 1 in Figure 3.3. Two peaks at m/z 4910.4 and m/z 2192.1 are due to fragments 1-46 and itz52-70, respectively (expected m/z 4909.3 and 2189.6). From these data, one can deduce that peptide chain cleavages occur at Cys47 and Cys52. The

Table 3.1. Calculated and observed m/z values for possible fragments resulting from the cleavage reaction of IGF-I chains at sites of designated cysteine pairs

Reduction of Disulfide	Fragment	Calculated m/z	Observed m/z
Cys47-Cys52	1-46	4909.3	4910.4
	47-51	638.7	nd
	52-70	2189.6	2192.1
Cys6-Cys48	1-5	516.5	nd
	6-47	4539.0	4539.2
	48-70	2682.1	2683.8
	*1-47	4976.5	4977.4
Cys15-cys61	Unreduced		

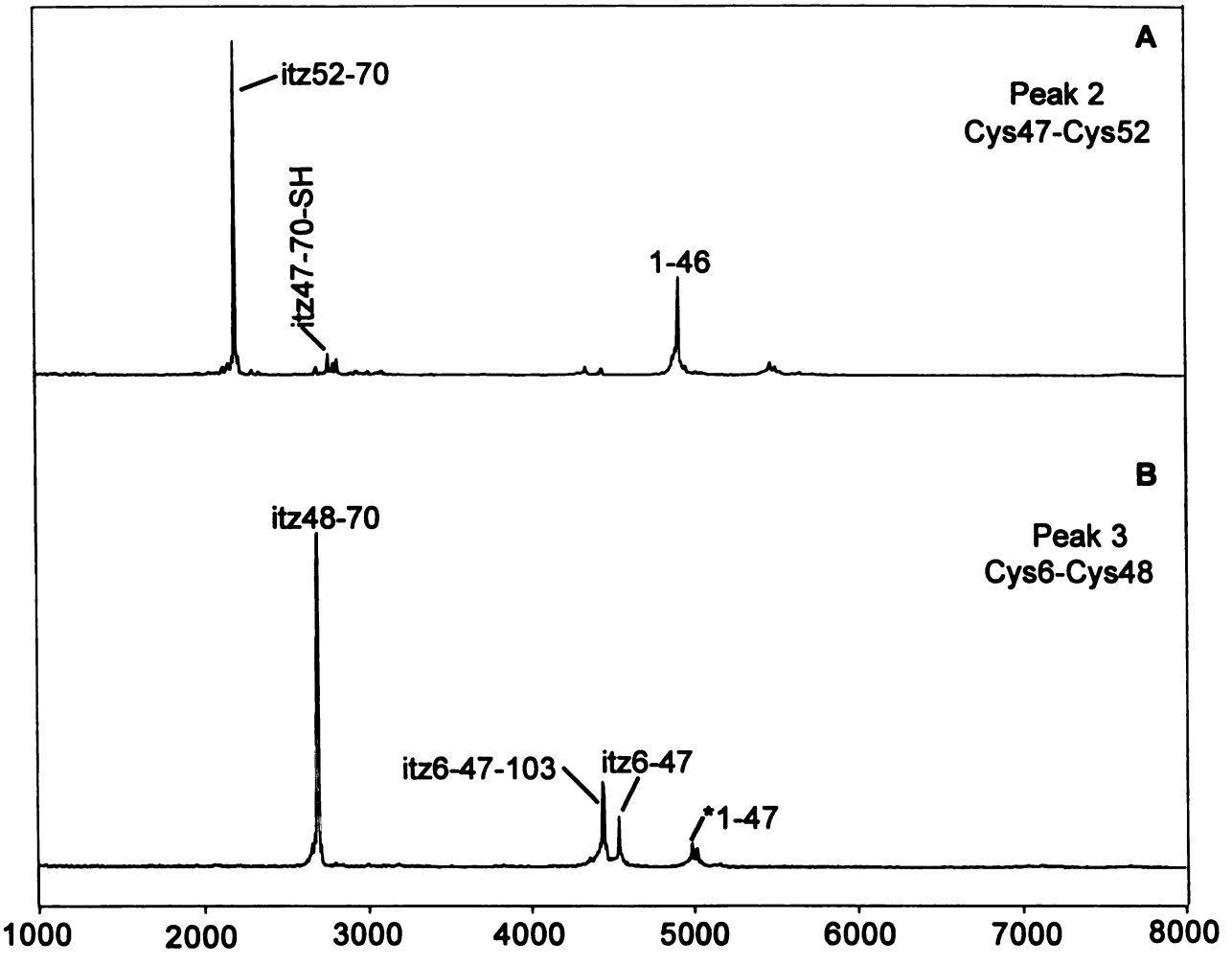


Figure 3.4. The MALDI mass spectra of peptide mixtures resulting from the cleavage of the two singly reduced/cyanylated IGF-I isomers, corresponding to the HPLC peaks 2 and 3 in Figure 3.2, respectively. The symbols "itz'" and * represent the iminothiazolidine derivatives and protonated β-elimination products, respectively.

MALDI peak marked itz47-70-SH in Figure 3.4A indicates the cleavage of the peptide bond at Cys47, and the cleavage of SH group from Cys52. This further confirms the cleavage site at Cys47. The fragment itz47-51 (m/z 638.7) was not determined by MALDI, which is due to the signal suppression at low-mass range of MALDI. Overall, a disulfide bond linkage between Cys47-Cys52 can be unambiguously deduced.

With similar strategy, the other disulfide-bond linkage, Cys6-Cys48 also can be recognized from Figure 3.4B. The MALDI spectrum in Figure 3.4B, corresponding to the cleavage product represented by HPLC peak 2 (Figure 3.3), shows peaks at m/z 2683.8 and m/z 4539.2, corresponding to fragment itz48-70 and itz6-47, respectively. Although another expected fragment 1-5 (m/z 516.5) is missing, it is still possible to deduce Cys6-Cys48 from these two fragments because no other combination gives such masses. One minor β-elimination product, residues 1-47 (m/z 4977.4), is particularly informative for confirmation of the assignment in this case. It should be pointed out that the question-marked peak in Figure 3.4B corresponds arithmetically to itz6-47 minus 103 Da; we have no rational explanation for its origin at this time. Whether this peak corresponds to the expulsion of a cysteine residue from itz6-47 awaits further study with other peptides or proteins containing adjacent cysteines.

After the two disulfide linkages of Cys47-Cys52 and Cys6-Cys48 are assigned unambiguously, the third one can be reasonably deduced as Cys18-Cys61 because there are three disulfide bonds in IGF-I, even though the singly reduced/cyanylated isomer could not obtained for this disulfide bond.

B. Disulfide Mapping of LR^3IGF-I

Recombinant LR^3IGF-I is a variant of human IGF-I that contains arginine replacing glutamate-3; it also has an amino terminal extension of 13 amino acids (30). LR^3IGF-I shows higher biological activities than its analog, IGF-I. Like IGF-I, LR^3IGF-I contains adjacent cysteines at positions 60 and 61. The disulfide-bonding scheme, assigned on the basis of homology to the insulin (or IGF-I) sequence and shown in Figure 3.5, has never been verified experimentally. In preparing authentic LR^3IGF-I from recombinant sources, it is important to confirm that the disulfide bond linkage is the same as for IGF-I isolated from natural sources, since a mismatching of disulfide bonds could have a major influence on any biological activity, as has been observed with disulfide-bonded isomers of insulin (31). Figure 3.6 shows the HPLC separation of LR^3IGF-I and isomers of its partially reduced/cyanylated species. The chromatogram shows similar results as those of obtained for IGF-I. There are only two discernible peaks, each of which represents a singly reduced isoform as determined by mass analysis. Stronger reducing conditions (increase the ratio of reducing reagent, apply higher temperature, and prolong the reduction time) result in the formation of doubly reduced/cyanylated protein isomers, but the third disulfide bond still refused to reduce, indicating the third disulfide bond is very stable.

Both MALDI-MS and on-line LC/ESI-MS were used to analyze the cleavage products of singly reduced/cyanylated isomers. The results are shown in Table 3.2.

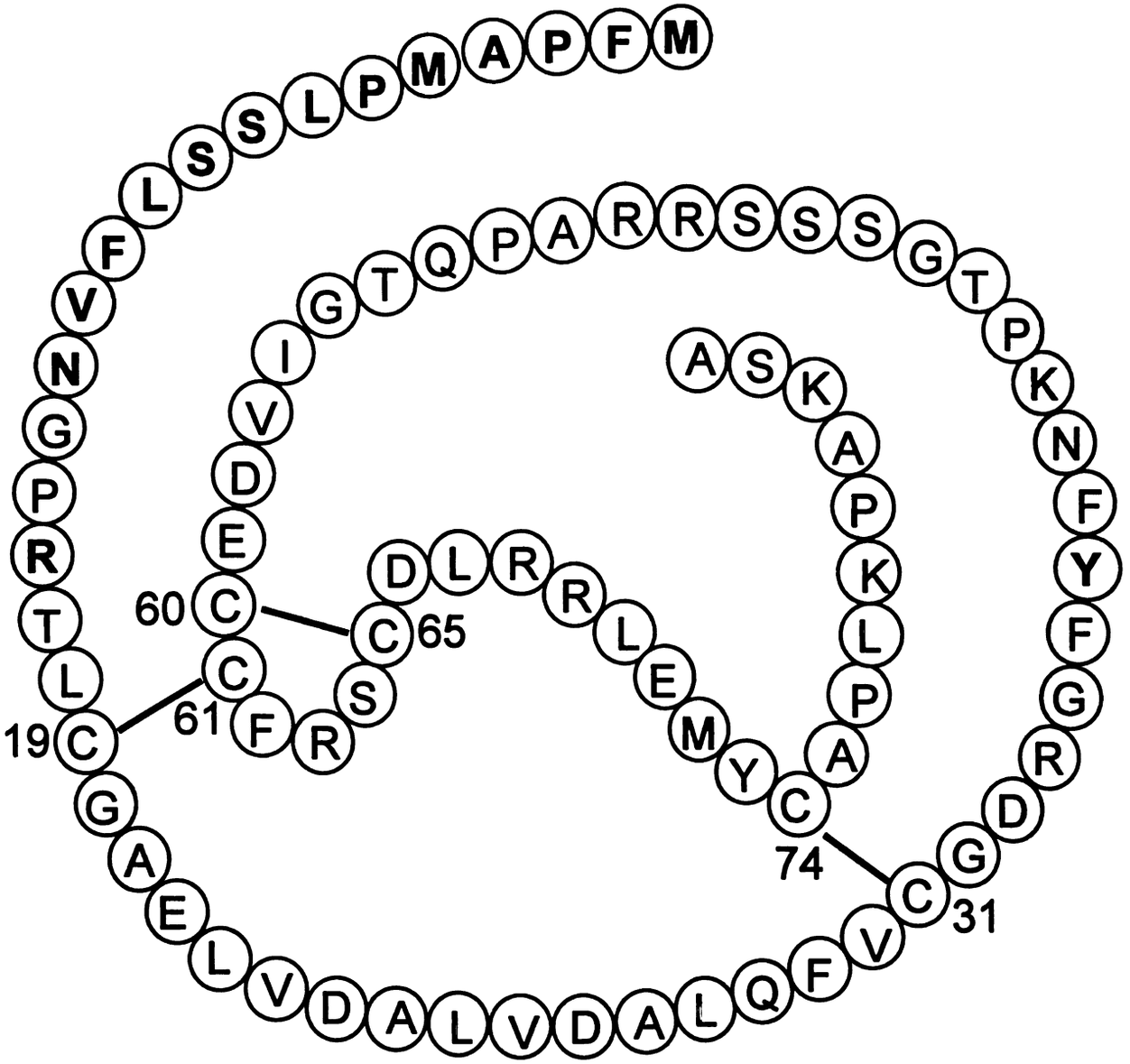


Figure 3.5. Amino acid sequence and disulfide structure of $LR^3IGF\text{-}I$
The bold letters at N-terminal site indicate the 13 amino acid extension.
The bold R represents the replacement of E^3 of IGF-I.

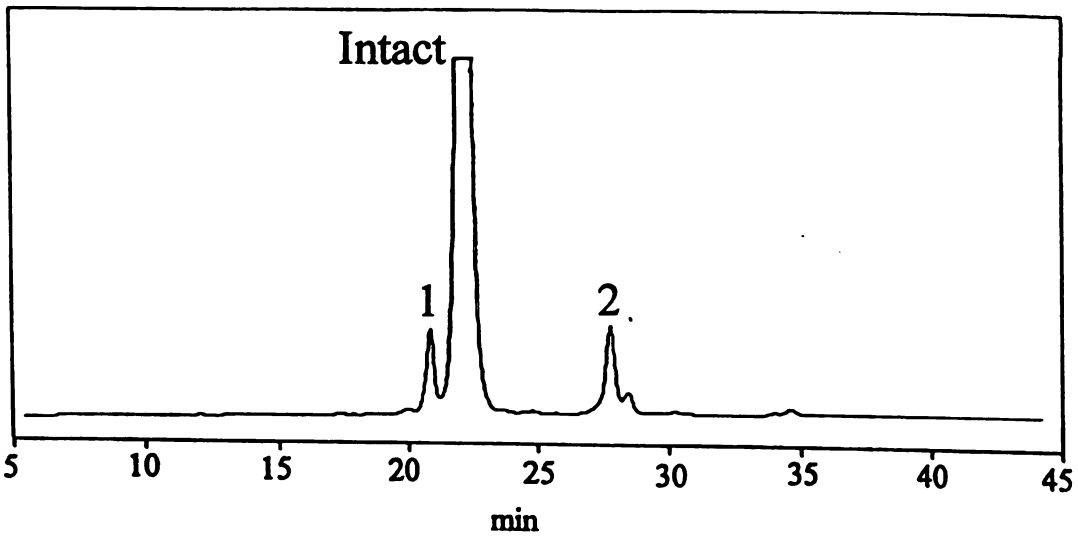


Figure 3.6. HPLC separation of denatured LR^{3}IGF-I and its partially reduced/cyanylated isomers. Separation was carried out on a Vydac C18 column at a flow rate of 1 ml/min with a linear gradient 30-50% B in 45 minutes, where A = 0.1% TFA in water and B = 0.1% TFA in CH_3CN. Peaks 1 and 2 represent singly reduced/cyanylated species, as determined by MALDI-TOF analysis.

Table 3.2. Calculated and observed m/z values for possible fragments resulting from the cleavage reaction of LR^3IGF-I chains at sites of designated cysteine pairs

Reduction of Disulfide	Fragment	Calculated m/z	Observed m/z by MALDI	Observed m/z by LC/ESI-MS
Cys60-Cys65	1-59	6371.2	6373.3	6372.7
	60-64	637.7	nd	638.0
	65-83	2188.6	2191.6	2189.7
	β(60-83)	2749.3	2755.1	nd
Cys19-Cys61	1-18	1978.4	1979.8	1977.8
	19-60	4538.0	4543.8	4540.0
	61-83	2681.1	2686.3	2683.9

Interpretation of MALDI Data

The two HPLC fractions (HPLC peak 1 and 2 in Figure 3.6) were subjected to cleavage under the described conditions. The corresponding MALDI mass spectra are shown in Figure 3.7A and B. The peaks at m/z 2191.6 and 6373.3 are due to fragments itz65-83 and 1-59, respectively. From these data, one can deduce that the cleavage sites are at positions 60 and 65. The fragment itz60-64 was missing from the MALDI spectrum due to the low-mass range signal suppression of MALDI-MS. However, the peaks at m/z 2755.1, representing overlapped peptide fragment itz60-83, where β-elimination occurs at Cys60, confirm that position 60 is the cleavage site. Therefore, disulfide linkage between Cys60 and Cys65 can be recognized.

The cleavage products from HPLC peak 2 show m/z values attributable to fragments1-18, itz19-60, and itz61-83, respectively, suggesting Cys19 and Cys61 are cleavage sites (Figure 3.7B). The peak with a single question mark showed a mass increase of 103 Da from the fragment 1-18, suggesting a cysteine attached to the fragment, whereas the peak with double question marks showed a mass decrease of 103 Da from the fragment 19-60, suggesting a cysteine removed from the fragment as occurrence in IGF-I. Like IGF-I, further study should be conducted to address the problem. On the other hand, the third disulfide bond in LR^3IGF-I can be easily assigned to Cys31-Cys74 based on the linkages of other two disulfide bonds. Therefore, unambiguous assignment of the disulfide structure is evident even if such undesired fragments are present, because the fragments from specific cleavage at the native disulfide bonds are sufficient to make a positive conclusion.

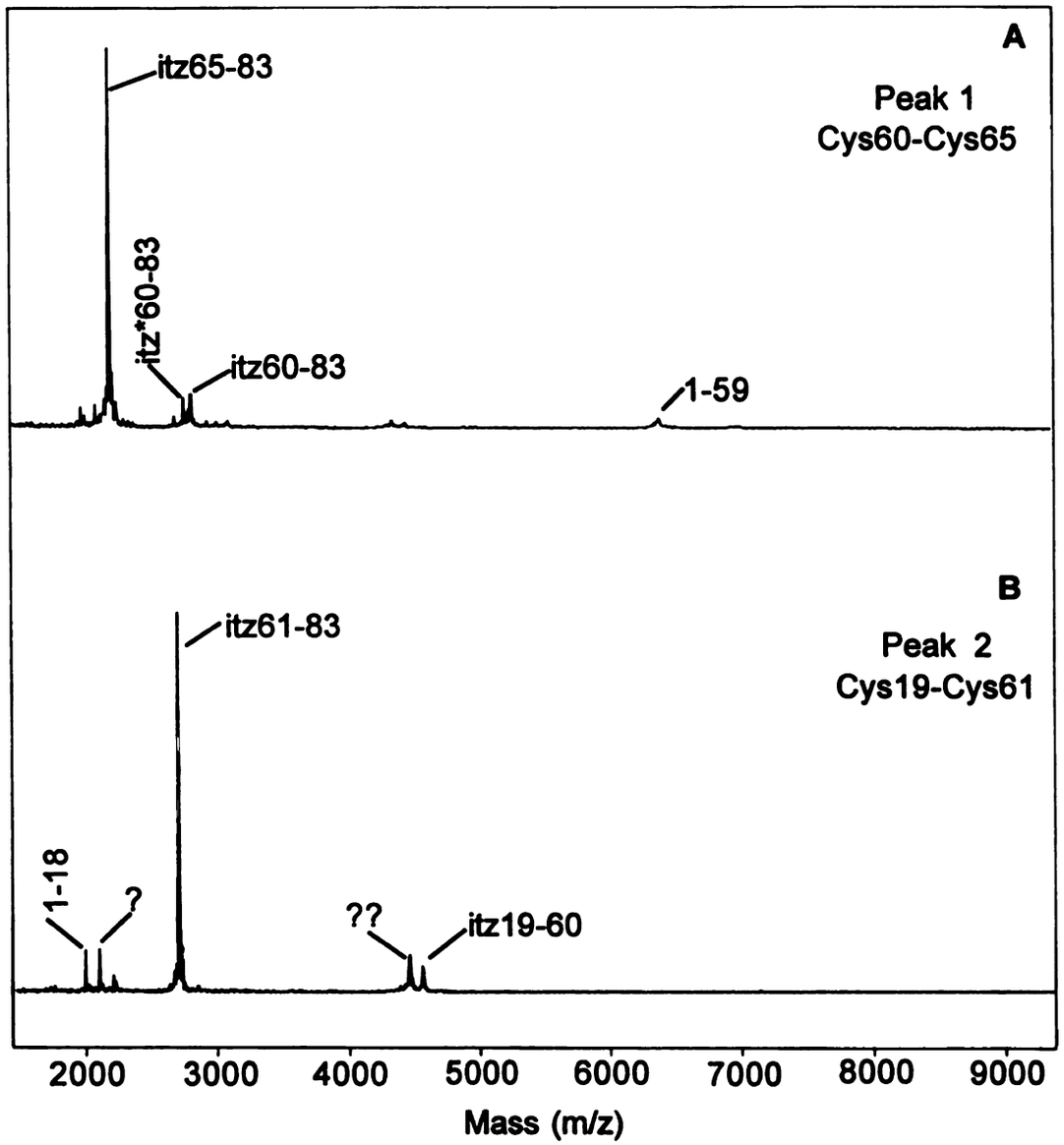


Figure 3.7. The MALDI mass spectra of peptide mixtures resulting from the cleavage of the two singly reduced/cyanylated LR^3IGF-I isomers, corresponding to the HPLC peaks 1 and 2 in Figure 3.4, respectively. The symbol "itz" and * represent iminothiazolidine derivatives and β-elimination products, respectively. The peaks with question marks are discussed in the text.

Interpretation of LC/ESI-MS Data

In order to confirm the disulfide structure of LR^3IGF-I assigned by MALDI-MS, especially for the low-mass range fragments and question marked peaks, LC/ESI-MS was utilized to analyze the cleavage products of singly reduced/cyanylated isomers. Figure 3.8A and B as well as Figure 3.9A and B show the results.

The first isomer (peak 1 in Figure 3.6) was analyzed by LC/ESI-MS to give the reconstructed total ion current (RTIC) chromatogram in Figure 3.8A and the corresponding mass spectra in Figure 3.8B. The peaks in the RTIC in Figure 3.8A can be assigned to specific fragments 1-59, itz60-64, itz65-83, itz65-83 minus the mass of H_2O, and the uncleaved protein (IP in Figure 3.8A) according to the corresponding masses from ESI spectra (Figure 3.8B). The ESI spectra in Figure 3.8B give masses 6372.7Da, 639 Da (protonated peptide), and 2189.7 Da which are due to fragments 1-59, itz60-64, and itz65-83, respectively. The missing peak (m/z 638.0 corresponding to fragment itz60-64) in MALDI spectrum was identified by LC/ESI-MS. From these data, the cleavage sites can be easily deduced, which are at positions 60 and 65. Therefore, the disulfide linkage should be between Cys60 and Cys65.

The peaks in the RTIC in Figure 3.9A resulting from analysis of cleavage products of the second singly reduced/cyanylated isomer (peak 2 in Figure 3.6) can be assigned to fragments 1-18 (1977.8 Da), itz19-60 (4540.98 Da), and itz61-83 (2684.93 Da) according to corresponding ESI spectra in Figure 3.9B, which indicate Cys19 was linked to Cys 61. While the peak corresponding to 1-18 plus 103 Da that was identified by MALDI-MS was not detected by LC/ESI-MS, the 'redundant' peak B in the ESI spectrum in the middle panel of Figure 3.9B that corresponds arithmetically to itz19-60 minus 103 Da

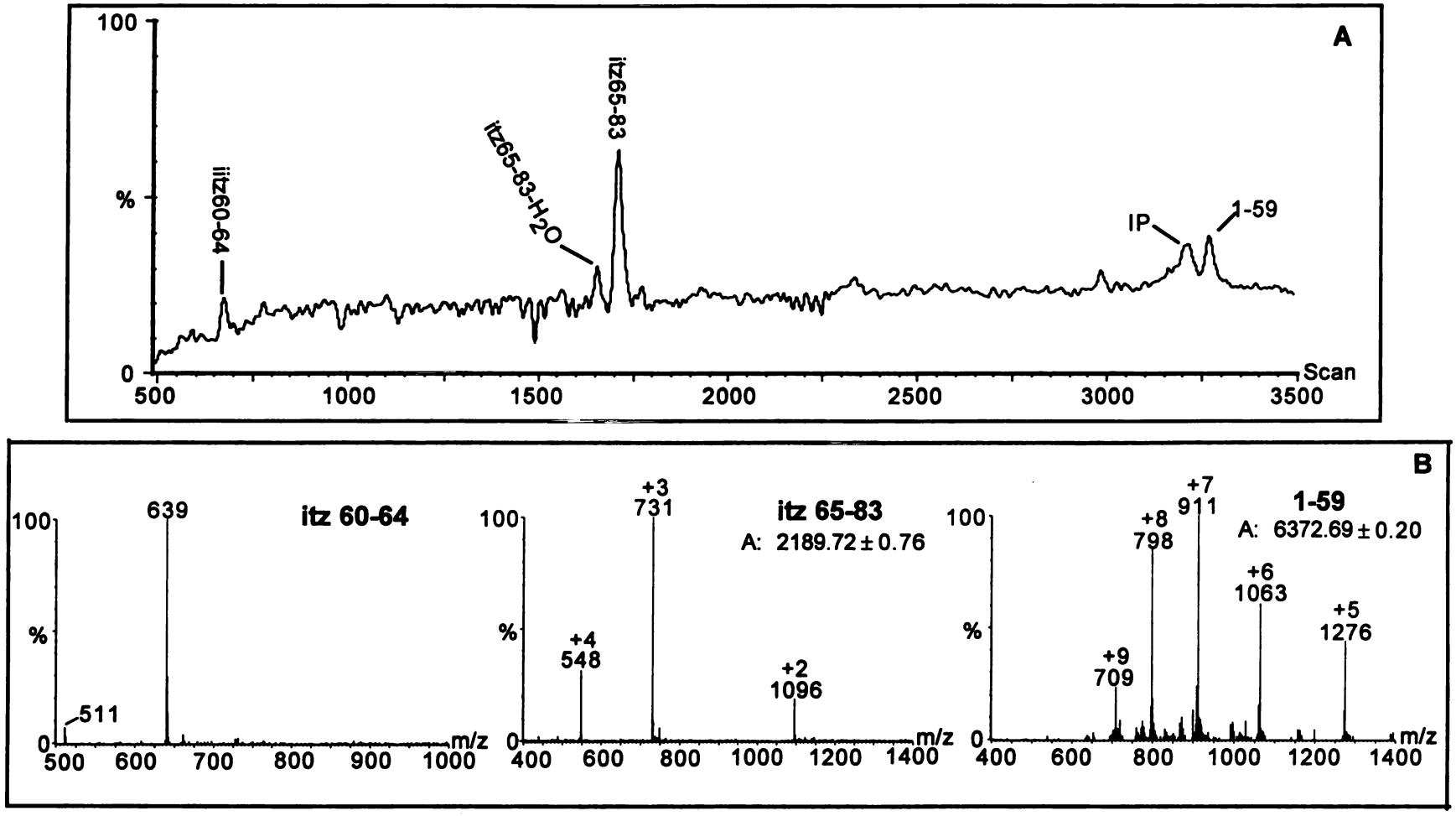


Figure 3.8. LC/ESI-MS results of peptide mixture from the cleavage of a singly reduced/cyanylated LR^{3}IGF-I isomer, corresponding to HPLC peak 1 in Figure 3.6. Figure 3.8A is RTIC of the cleavage products. Figure 3.8B is the ESI mass spectra of the peptides corresponding to the TIC peaks in Figure 3.8A. The symbol 'itz' represent iminothiazolidine derivatives.

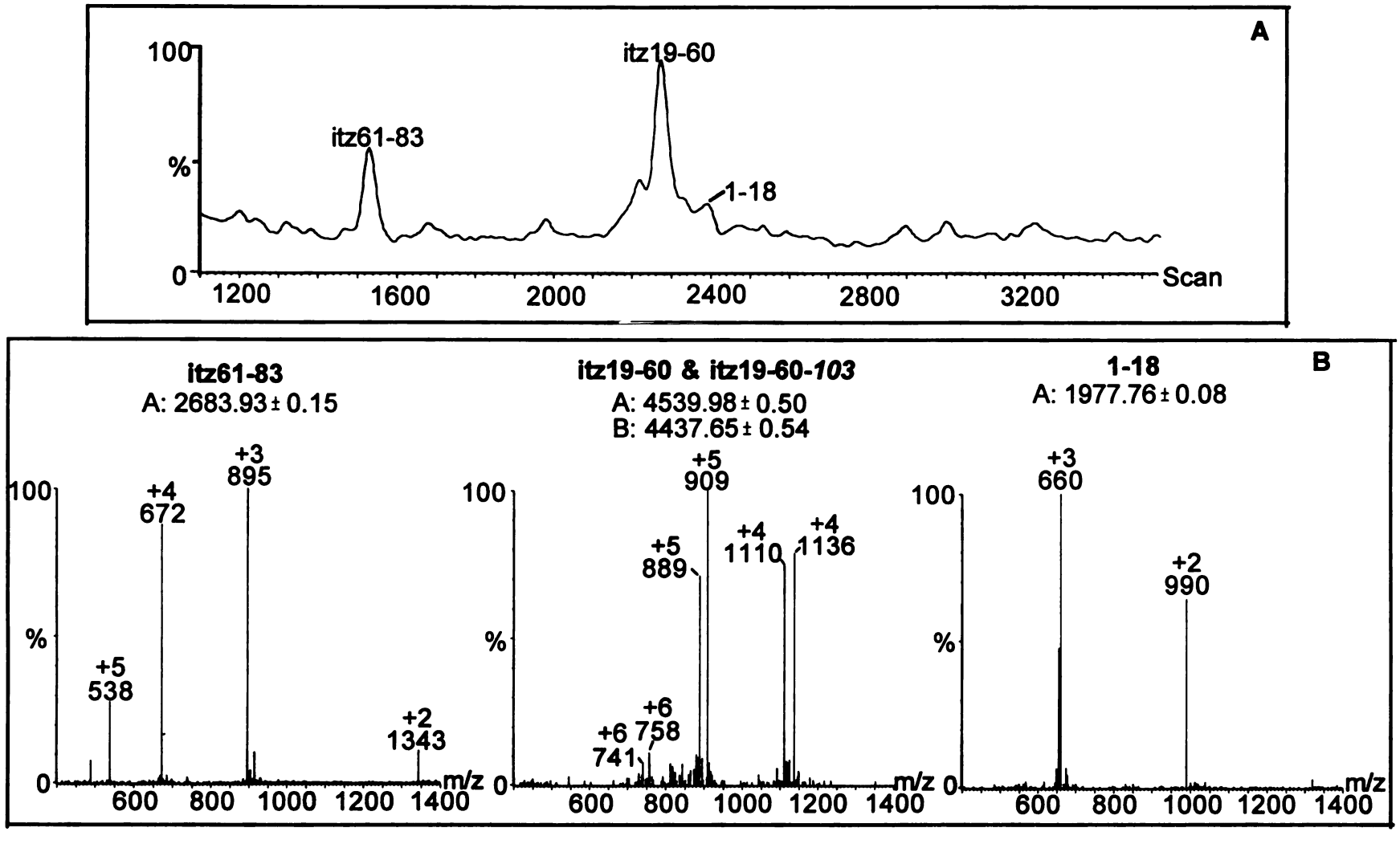


Figure 3.9. LC/ESI-MS results of peptide mixture from the cleavage of a singly reduced/cyanylated LR^{3}IGF-I isomer, corresponding to HPLC peak 2 in Figure 3.6. Figure 3.9A is RTIC of the cleavage products. Figure 3.9B is the ESI mass spectra of the peptides corresponding to the TIC peaks in Figure 3.9A.

was still present. Overall, the disulfide bond structure of LR^3IGF-I can be easily assigned by the combination of MALDI and ESI-MS. These results provide direct evidence for the disulfide-bonding pattern in LR^3IGF-I that has hitherto only been presumed by homology to that in IGF-I (30).

V. Conclusion

In this chapter, the disulfide linkages of IGF-I and LR^3IGF-I have been determined by the methodology of partial reduction, cyanylation, chemical cleavage, and mass mapping. Both proteins contain three disulfide bonds and two adjacent cysteines. We have demonstrated that the methodology provides a simple and effective method for the assignment of disulfide-bond pairing in proteins containing adjacent cysteine residues, a challenging problem for the conventional approach. Furthermore, the new approach offers an important advantage of minimizing disulfide bond scrambling. The combination of MALDI and LC/ESI-MS are very powerful tools to analyze a variety of cleavage products for disulfide mapping.

VI. Future Work

The approach used in this chapter to assign the disulfide linkages in proteins involved in partial reduction, cyanylation, HPLC separation, chemical cleavage, and mass mapping. Theoretically, it is not necessary to separate the mixture of singly reduced/cyanylated protein isomers if only the singly reduced isomers are produced during the partial reduction process. Controlling the partial reduction conditions (ratio of the reducing reagent to disulfide content, reduction time and temperature) can satisfy this

condition. Omitting the step of chromatographic separation of partial reduced/cyanylated protein isoforms could make the analysis faster and less sample consumption. An overview of a proposed new approach is illustrated in Figure 3.10 based on a protein containing two disulfide bonds.

In the case of two disulfide bonds in a protein, there are three possible disulfide linkage patterns (Case I, II, III in Figure 3.10). The partial reduction can be conducted at pH 3.0 in the controlling conditions to produce only singly reduced protein isoforms, two isoforms (I, II, or III) for each case shown in the Figure 3.10. Cyanylation can be performed in the same buffer (pH 3.0). Right after the cyanylation, the pH of the solution can be adjusted to pH 12.0 by ammonia, followed by incubation for 1 hour at room temperature. After evaporating ammonia from the solution, the remaining disulfide bonds can be reduced by TCEP. MALDI-MS or LC/ESI-MS can be used to analyze the cleavage mixture. Because the sequence of a protein is known as well as the numbers and location of cystines can be determined by the method described in Chapter 2, the disulfide linkages of a protein can be determined based on the combination of different peptide fragments for each case (Figure 3.10). The masses of peptide fragments for each case can be predicted according to its disulfide linkages. By matching the predicted and observed masses, the disulfide-bond-pairing pattern can be determined.

It should be pointed out that the numbers of peptide fragments in the cleavage mixture would be increased with the increasing number of disulfide bonds in the protein. MALDI-MS sometimes has the signal suppression phenomenon as described in Chapter 2, which will result in the incomplete identification of the peptide fragments in the cleavage products, providing uncertainty in the disulfide linkage determination. While

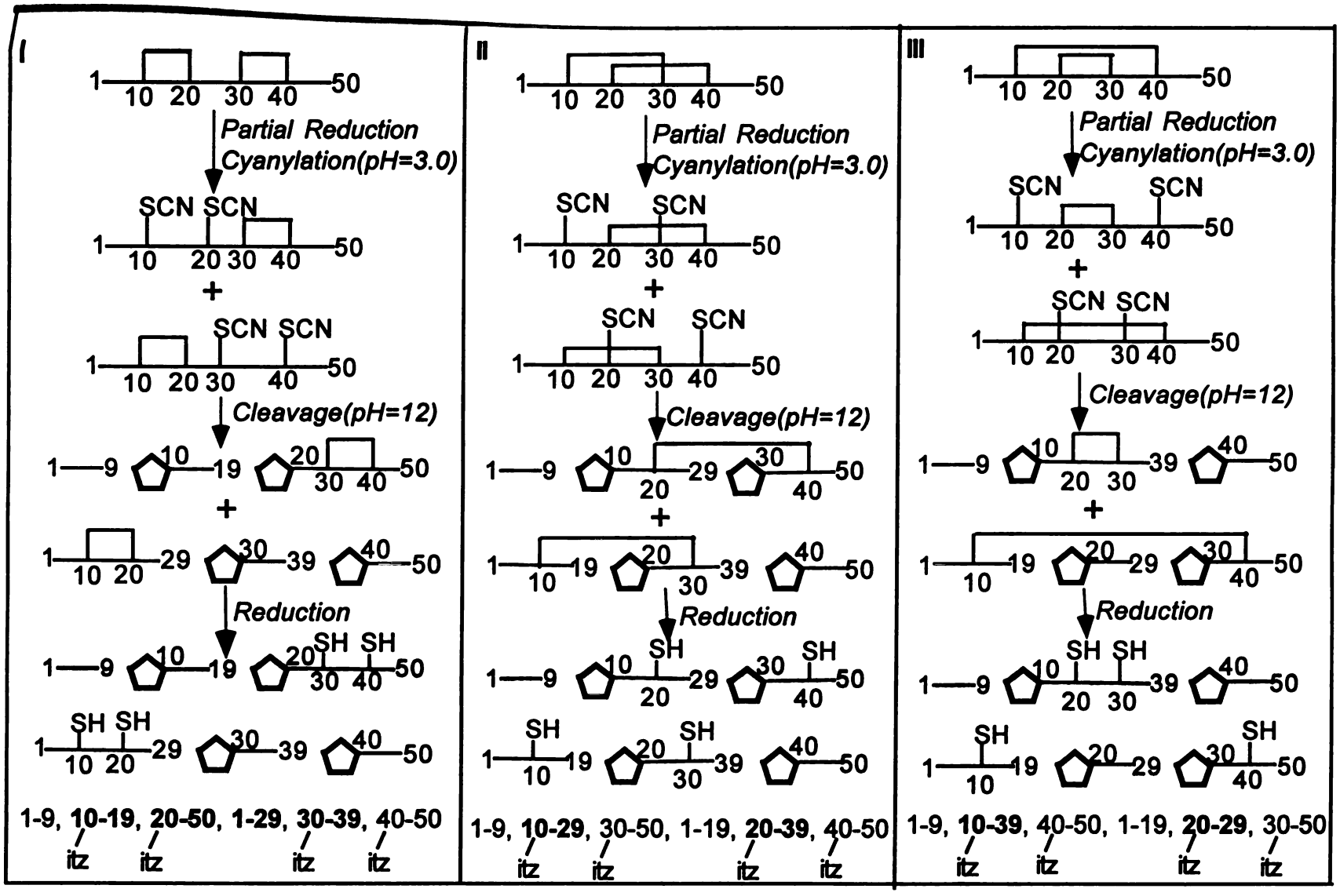


Figure 3.10. The overview of proposed non-chromatographic approach

LC/ESI-MS might be used to solve the problem, disulfide linkages may still be identified even though the peptide fragments cannot be completely determined by mass spectrometry. The peptide fragments indicated in bold numbers in Figure 3.10 for each case can be used to identify the disulfide linkages because they have unique masses from the set of peptide fragments that result from the other cases. Alternatively, the combination of the masses of several not complete peptide fragments can also be used to determine the disulfide linkages. For example, the combination of fragments 1-9, 10-19, and 20-50 in case I of Figure 3.10 can be used to determine the disulfide linkage pattern of Cys10-Cys20 and Cys30-Cys40 because the same combination cannot be obtained for the other two cases (II and III). An effective way to do the calculation and matching is by the computer software. The combination of partial reduction, cyanylation, mass spectrometry and computer software matching would provide a powerful technique to assign the disulfide linkages in proteins.

VII. References

1. Creighton, T. E., *Bioessays*, 8, 57 (1988).
2. Smith, D. L., Zhou, Z., *Methods Enzymol.*, 193, 374 (1990).
3. Hirayama, K., Akashi, S., in *Biological Mass Spectrometry: Present and Future*, Matsuo T., Caprioli, R. M., Gross, M. L., and Seyama, Y., eds., pp 299 (1994).
4. Ryle, A. P., Sanger, F., *Biochem. J.*, 60, 535 (1955).
5. Stults, J. T., Bourell, J. H., Canova-Davis, E., Ling, V. T., Laramee, G. R., Winslow, J. W., Griffin, P. R., Rinderknecht, E., and Vandlen, R. L., *Biomed. Environ. Mass Spectrom.*, 19, 655 (1990).
6. Fox, J. W., Elzinga, M., and Tu, A. T., *Biochemistry*, 18, 678 (1979).
7. Serensen, H. H., Thomsen, J., and Bayne, S., *Biomed. Environ. Mass Spectrom.* 19, 713 (1990).
8. Nokihara, K., Morita, N., Yamaguchi, M., and Watanabe, M., *Anal. Lett.*, 25, 513 (1992).
9. Zhang, D., and Liang, S., *J. Protein Chem.*, 12, 735 (1993).
10. Krishnamurthy, T., Hauer, C. R., Prabhakaran, M., Freedy, J. G., and Hayashi, K., *Biol. Mass Spectrom.*, 23, 719 (1994).
11. Tang, Y., and Selsted, M. E., *J. Biol. Chem.*, 268, 6649 (1993).
12. Gray, W. R., *Protein Science*, 2, 1749 (1993).
13. Haniu, M., Hsieh, P., Rohde, M. F., and Kenney, W. C., *J. Biochem. Biophys.*, 310, 433 (1994).
14. Yamashita, H., Nakatsuka, T., and Hirose, M., *J. Biol. Chem.*, 270, 29806 (1995).
15. Fairlie, W. D., Stanton, P. G., and Hearn, M. T. W., *Biochem. J.*, 314, 449 (1996).

16. Gray, W. R., Luque, F. A., Galyean, R., Atherton, E., Sheppard, R. C., Stone, B. L., Reyes, A., Alford, J., Mcintosh, M., Olivera, B. M., Cruz, L. J., and rivier, *J., Biochemistry*, 23, 2796 (1984).

17. Gray, W. R., *Protein Sci.*, 2, 1732 (1993).

18. Wu, J., and Watson, J. T., *Protein Science*, 6, 391 (1997).

19. Rinderknetcht, E., and Humbel, R. E., *J. Biol. Chem.*, 253, 2769 (1978).

20. Froesch, E. R., Burgi, H., Ramseier, E. B., Bally, P., and Labhart, A., *J. Clin. Invest.*, 42, 1816 (1963).

21. Rinderknecht, E., and Humbel, R. E., *Proc. Natl. Acad. Sci., USA*, 73, 2365 (1976).

22. Blundell, T. L., Bedarkar, S., Rinderknecht, E., and Humbel, R. E., *Proc. Natl. Acad. Sci. USA*, 75, 180 (1978).

23. Blundell, T. L., Dodson, G. G., Hodgkin, D. C., and Mercola, D. A., *Adv. Protein Chem.*, 26, 279 (1972).

24. Forsberg, G, Palm, G., Ekebacke, A., Josephson, S., and Hartmanis, M., *Biochem. J.*, 271, 357 (1990).

25. Iwai, M., Kobayashi, M., Tamura, K., Ishii, Y., Yamada, H., and Niwa, M., *J. Biochem.*, 106, 949 (1989).

26. Raschdorf, F., Dahinden, R., Maerki, W., and Richter, W. J., *Biomed. Environm. Mass Spectrom.*, 16, 3(1988).

27. Axelsson, K., Johansson, S., Eketorp, G., Zazzi, H., Hemmendorf, B., and Gellerfors, P., *Eur. J. Biochem.*, 206, 987 (1992).

28. Wu, J., Gage, D. A., and Watson, J. T., *Anal. Biochem.*, 235, 161 (1996).

29. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., Nilsson, B., *Biochemistry*, 31, 1749 (1992).

30. Tomas, F. M., Knowles, S. E., Chandler, C. S., Francis, G. L., Owen, P. C., and Ballard, F. J., *J. Endocrinology*, 137, 413 (1993).

31. Rinderknetcht, E., Humbel, R. E., *FEBS Lettt.*, 89, 283 (1978).

CHAPTER 4

TRAPPING AND IDENTIFICATION OF INTERMEDIATES DURING THE REFOLDING OF LR^3IGF-I AND IGF-I

I. Introduction

Considerable insight to the folding and unfolding pathways of a protein can be obtained from trapping and characterizing intermediate structures involved in the dynamic process (1-4). A particularly difficult aspect in the study of protein folding is the fact that intermediates may be short-lived and therefore difficult to isolate and analyze structurally and functionally. Disulfide-containing proteins provide an opportunity to capture sulfhydryl-containing intermediates by chemical reaction during the time course of folding or unfolding (5-8). The folding pathways of several proteins (9-11) have been studied in this way; among them, bovine pancreatic trypsin inhibitor (BPTI) (6, 8, 12, 13), and ribonuclease A (14-18) have been extensively characterized.

In order to isolate and characterize the intermediates that are involved in folding or unfolding of proteins containing disulfide bonds, it is necessary to stop both the inter- and intramolecular thiol/disulfide exchange reactions that convert one intermediate to another. Two traditional trapping methods have been used. One method traps all sulfhydryls irreversibly by alkylation with iodoacetate or iodoacetamide. However, rearrangement of intermediates during trapping with iodoacetate has been observed for both BPTI (8) and ribonuclease A (19), since quenching by 0.2M iodoacetate is comparable to or, in some cases, slower than the rate of some of the intramolecular rearrangements that interconvert intermediates of the same oxidation state (20). Another

method quenches thiol/disulfide exchange by lowering the pH to < 2 by adding acid. An advantage of acid trapping is its reversibility; intermediates trapped at low pH can be transferred to high pH to allow further rearrangement or folding in experiments designed to more completely characterize particular pathways (8, 10). While quenching by acidification is rapid and occurs at the diffusion-controlled rate, it does not completely stop thiol/disulfide exchange (21). Therefore, structural characterization of a given trapped intermediate must be done promptly.

We have developed methodology to trap and identify folding intermediates based on cyanylation methodology and mass spectrometry (22). Our approach has two advantages. Firstly, acidification quenches the thiol/disulfide exchange rapidly, and cyanylation of free sufhydryls precludes further exchange because of irreversible chemical modification. Secondly, trapping of sulfhydryls is already part of our procedure for structural elucidation of the intermediates.

After trapping and isolation of folding or unfolding intermediates, the disulfide structure must be determined. Conventional methodology for determining disulfide structures of folding intermediates involves cleavage between every cysteine residue in the intermediates by enzymatic/chemical digestion, which is tedious, cumbersome, and may risk artifact formation due to disulfide exchange. Furthermore, conventional approaches may have difficulty in analyzing intermediates containing adjacent cysteines since proteolytic and/or chemical degradation may not achieve cleavage between cysteine residues (23).

Our methodology involves partial reduction of disulfide bonds in a protein, cyanylation of sufhydryl groups, HPLC separation of partially reduced/cyanylated protein

isoforms, chemical cleavage of the peptide bonds at the N-terminal side of cyanylated cysteine residues, and mass analysis of cleavage products by mass spectrometry (24). This new developed methodology has the advantages of preventing disulfide scrambling and cleaving the peptide backbone between adjacent cysteines; furthermore, it is fast, simple, and sensitive (25).

Insulin-like growth factor-I (IGF-I) is a single-chain polypeptide of 70 amino acid residues containing three intramolecular disulfide bonds, two of which involve adjacent cysteines (26) (Figure 4.1). IGF-I is postulated to be the mediator of growth hormone action on skeletal tissue as well as mitogenic activity on several cell types (27, 28). Several research groups have studied refolding pathway of IGF-I (29-33). Hober, *et al* (29, 30) trapped the folding intermediates of IGF-I by pyridylethylation at pH 8.7. Five major forms of IGF-I were detected, which includes a one-disulfide intermediate, two two-disulfide intermediates, and a mismatched three-disulfide intermediate (29). A different folding pattern was obtained by trapping the folding intermediates with acidification (31-33). Instead of five intermediates detected by pyridylethylation, six intermediates were identified by trapping with acidification, which include a mixed disulfide intermediate (32, 33). A nonnative two-disulfide intermediate was identified with acidification, but not identified with pyridylethylation. Furthermore, while a one-disulfide intermediate (Cys18-Cys16) was identified as a major form in the folding process of IGF-I by trapping with pyridylethylation (29), it only appeared as a minor component when trapped by acidification (32, 33). Therefore, a further study of the folding intermediates of IGF-I needed to be conducted.

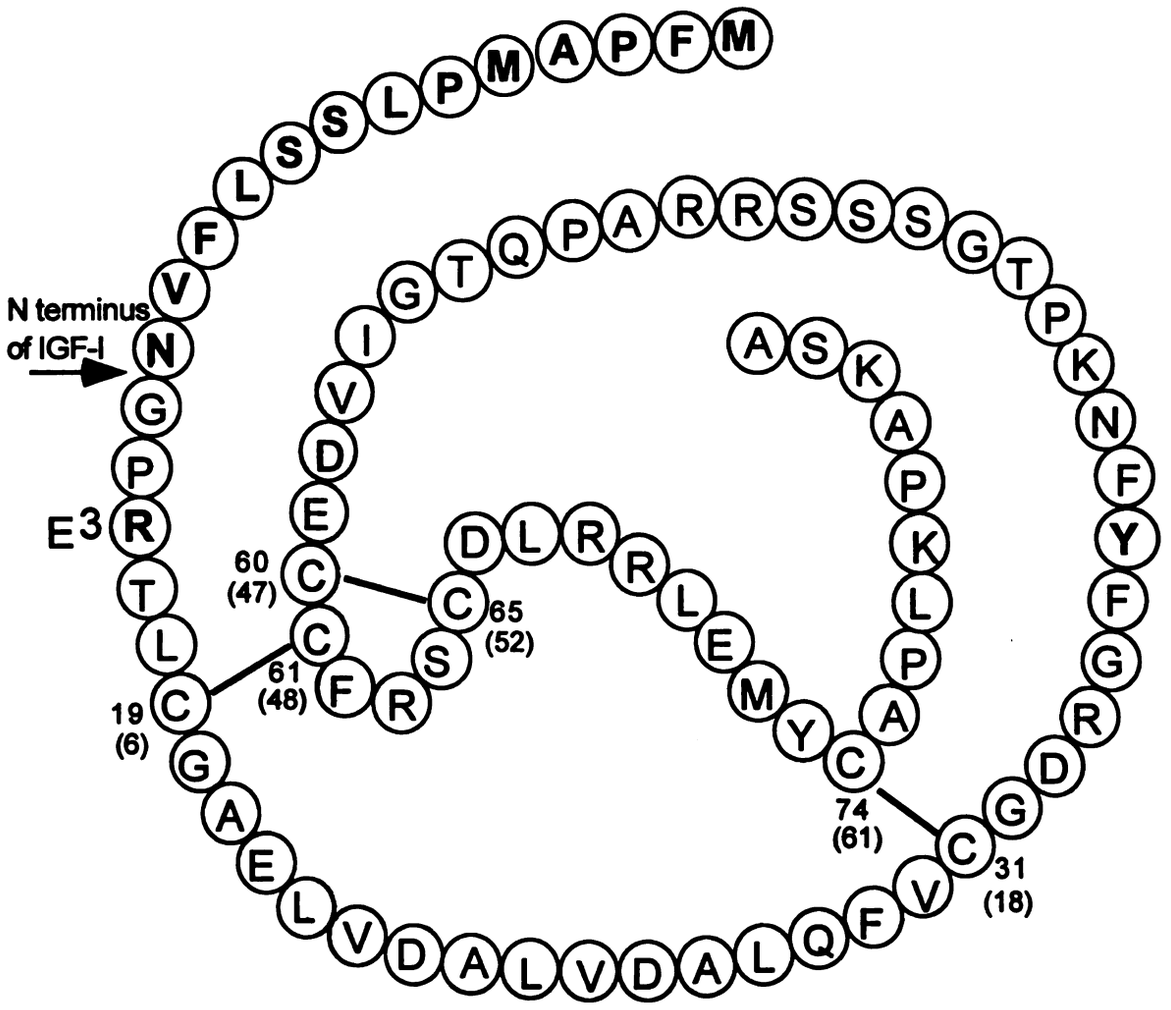


Figure 4.1. Amino acid sequences and disulfide structures of $LR^3IGF\text{-}I$ and IGF-I. The numbers in the parentheses indicate the location of cysteine residues of IGF-I. The 13 bold letters at the N-terminus indicate the 13-amino acid extension in $LR^3IGF\text{-}I$. The bold R represents the replacement of E^3 of IGF-I.

Recombinant human long R^3 insulin-like growth factor-I (LR^3IGF-I) is a variant of human insulin-like growth factor-I (IGF-I) in which glutamate 3 is replaced by arginine and a 13-residue extension appears at the N-terminus (Fig. 4.1). It contains three disulfide bonds and two adjacent cysteines. The disulfide linkages have been determined by partial reduction/cyanylation/mass mapping (25). LR^3IGF-I is substantially more potent than IGF-I in affecting carbohydrate metabolism and in stimulating the growth of fetal tissue in animals (34). The results of a folding study of LR^3IGF-I (31) were significantly different from those published for IGF-I refolding (29, 30, 32-33). However, the disulfide structure of the folding intermediates of LR^3IGF-I has not been studied by the authors (31).

In this chapter, we integrate the different structural features of IGF-I and LR^3IGF-I into the observed folding and refolding dynamics by using the same methodology to study the behavior of both proteins. Furthermore, our methodology permits direct determination of the disulfide pattern in each of the intermediates, which were trapped without artifact formation. We have trapped the intermediates of refolding of IGF-I and LR^3IGF-I and determined disulfide structures of the intermediates by cyanylation methodology and mass spectrometry. The disulfide structure of all the refolding intermediates of both IGF-I and LR^3IGF-I were identified by our methodology. It is the first direct experimental evidence for the disulfide structure of the nonnative three-disulfide intermediate, which was only predicted previously by the structural constraints of disulfide bonds in the three-disulfide intermediates (29). The refolding pattern of IGF-I obtained by our methodology was similar to that of obtained by trapping with

acidification. Furthermore, the results of our investigation provide integrated information for the folding of insulin-like growth factor families.

II. The Strategy for Trapping Folding Intermediates

The procedures for trapping and analyzing refolding and reductive unfolding intermediates of proteins are shown in Figure 4.2. Reduced-unfolded protein was refolded at pH 8.7 in the presence of redox buffer consisting of oxidized glutathione (GSSG) and reduced glutathione (GSH). The folding intermediates were trapped by cyanylating the sulfhydryl groups of the intermediates with 1-cyano-4-dimethylamino-pyridinium (CDAP) at pH 2~3. The trapped cyanylated intermediates were separated by reversed-phase HPLC. The numbers of disulfide bonds were determined by MALDI-MS based on the mass shift in MW between native protein and that of intermediates. The mass shift should be 104 Da between native protein and the one-disulfide intermediate, 52 Da between native protein and the two-disulfide intermediates, and no mass shift between the intact protein and the three-disulfide intermediates. The 52-Da mass shift between intact protein and the two-disulfide intermediate arises from 2 Da for reduction of one disulfide bond and 25 Da for replacement of a hydrogen atom on each of the sulfhydryl groups with CN, etc. After HPLC separation and purification, the disulfide structure of the intermediates was determined based on the methodology of partial reduction, cyanylation, chemical cleavage, and mass mapping (24).

The reverse direction of refolding involving disulfide formation in a protein is reductive unfolding involving the reduction of disulfides by a free-thiol reagent such as cysteine. A study of reductive unfolding can provide insight to protein refolding since

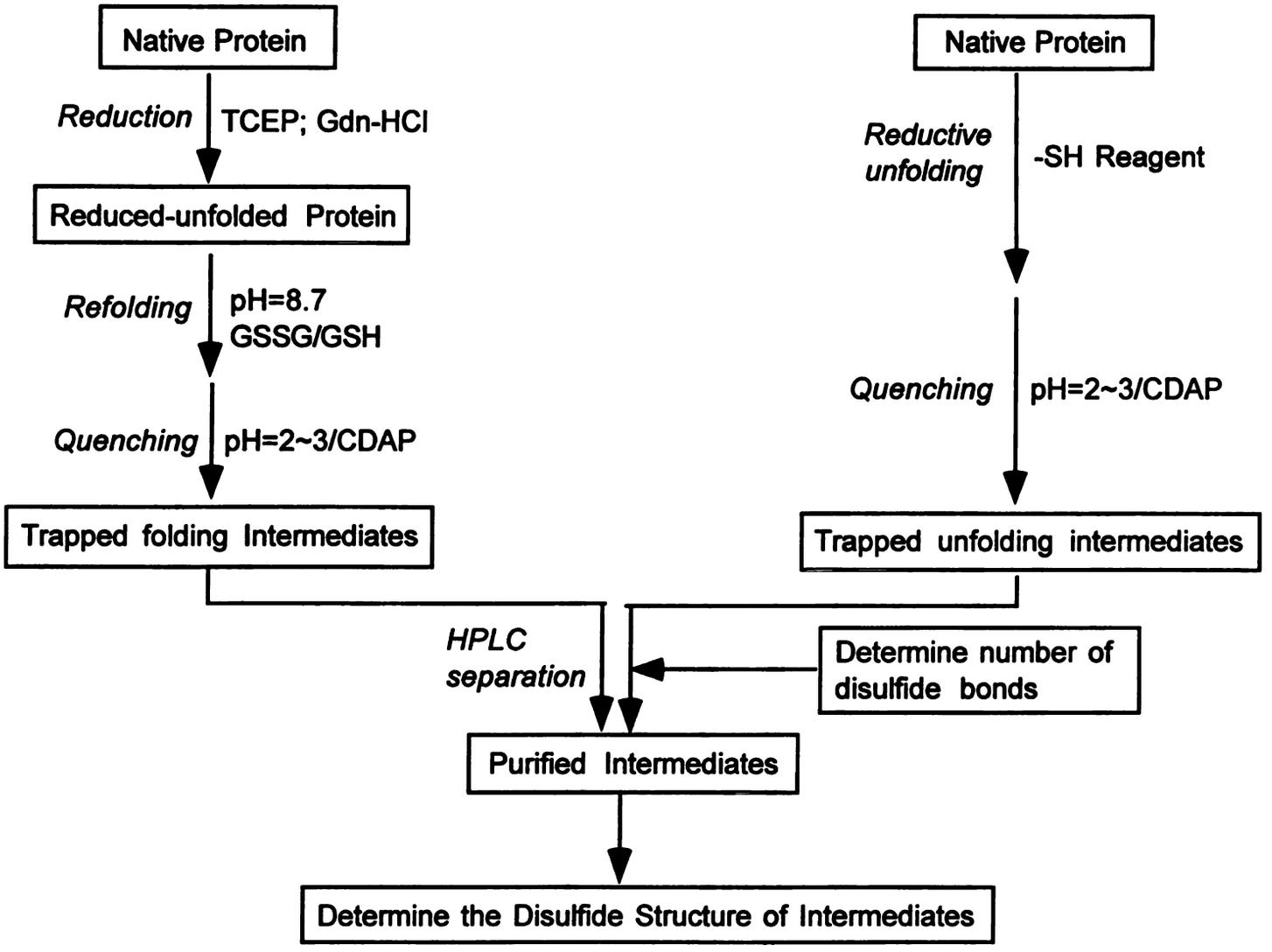


Figure 4.2. The overview of the processes of refolding and reductive unfolding of proteins

reductive unfolding is the reverse of refolding. The disulfide bonds of a native/folded protein are reduced in the presence of a thiol reagent, cysteine. The unfolding intermediates are trapped, separated, and analyzed by the same methods as those used in the study of folding intermediates.

Compared to the conventional approach, our methodology provides several advantages. First, the trapping reaction, occurring in acidic solution, greatly reduces the risk of sulfhydryl/disulfide exchange that bothers alkylation trapping. Secondly, the CDAP quantitatively cyanylates free sulfhydryl groups to stop further folding. Unlike acid quench, which is totally reversible and requires further chemical modification in order to determine the disulfide structure of the folding intermediates, the cyanylated intermediates can be directly subjected to characterization of disulfide structure by the approach described in chapter 3. Thirdly, the disulfide-mapping technique itself provides unique advantages over conventional techniques, as demonstrated in previous chapter.

III. The Experimental Section

MALDI-TOF-MS

MALDI mass spectra were obtained on a Voyager Elite time-of-flight (TOF) mass spectrometer (PerSeptive Biosystems Inc., Framingham, MA) equipped with delayed extraction and a model VSL-337ND nitrogen laser (Laser Science, Newton, MA). The accelerating voltage in the ion source was set to 20 KV. Grid voltage and guide wire voltages were 93.6% and 0.2% of the accelerating voltage, respectively. Data were acquired in the positive linear DE mode of operation. Time-to-mass conversion was

achieved by external and/or internal calibration using standards of bradykinin (m/z 1061.2), bovine pancreatic insulin (m/z 5734.5), and horse skeletal myoglobin (m/z 16952) obtained from Sigma Chemical Co. (St. Louis, MO). All experiments were performed using α-cyano-4-hydroxycinnamic acid (Aldrich Chemical Co., Milwaukee, WI) as the matrix. Saturated matrix solutions were prepared in a 50% (v/v) solution of acetonitrile/aqueous 1% TFA, and mixed in equal volumes with peptide or protein samples, and applied to a stainless-steel sample plate. The mixture was allowed to air dry before being introduced into the mass spectrometer.

Materials

Recombinant human long insulin-like growth factor-I (LR^3IGF-I) was purchased from Sigma Chemical Co. (St. Louis, MO). Insulin-like growth factor-I was obtained from Austral Biologicals (San Ramon, CA). The proteins were purified before use. Oxidized glutathione, reduced glutathione, and cysteine were obtained from Sigma Chemicals (St. Louis, MO). Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride was purchased from Pierce Chemical Co. (Rockford, IL.). Guanidine hydrochloride was obtained from Boehringer-Mannheim Biochemicals (Indianapolis, IL.). Citric acid, sodium citrate, hydrochloric acid, and 1-cyano-4-dimethylamino-pyridinium tetrafluoroborate (CDAP) were purchased from Sigma and used without further purification. Acetonitrile and trifluoroacetic acid (TFA) were of HPLC grade. TCEP solution in 0.1M citrate buffer at pH 3.0 was prepared as 0.10 M stock solution and stored under N_2 at -20^{0}C for weeks with little deterioration. The 0.10M CDAP solution in 0.1 M citrate buffer at pH 3.0 was freshly prepared prior to use.

Refolding Experiments

LR^3IGF-I or IGF-I (0.1mg) was dissolved in 0.5ml citrate buffer, pH 3.0, containing 6 M guanidine-HCl and 0.1M TCEP reducing agent. The denaturation and reduction of the proteins were carried out at 37^0C for 2 hours. HPLC was used to separate the reaction mixture and the fraction corresponding to the reduced LR^3IGF-I or IGF-I was collected, dried and stored at -80^0C.

The refolding of reduced and denatured protein was initiated by diluting the reduced/unfolded protein sample with 0.10 M Tris-HCl buffer (pH 8.71), containing 1mM oxidized glutathion (GSSG), 10mM reduced glutathion, 0.2M KCl, 1 mM EDTA, to a final protein concentration of 0.1mg/ml. Folding intermediates were trapped in a time course manner by removing aliquots (0.1ml) of protein solution and mixing them with 1.0M HCl solution containing freshly prepared 0.2 M CDAP to pH 2~3. Cyanylation of free sulfhydryl groups by CDAP proceeded at room temperature for 10 minutes. The trapped intermediates were immediately separated by HPLC. The fractions were collected manually and analyzed by MALDI-TOF MS. Those with 0-Da, 52-Da, 104-Da, 156-Da increases over the mass of the intact protein corresponding to the three-disulfide (non-native or native N), 2-disulfide, 1-disulfide, and 0-disulfide (R) species, respectively. The collected intermediates were dried in a speedvac and stored in a -80^0C freezer.

Reductive Unfolding Experiments

Native LR^3IGF-I and IGF-I (0.1mg/ml) was reduced/unfolded in 0.10 M Tris-HCl buffer (pH 8.71), containing 0.25mM cysteine, 0.20 M KCl, 1 mM EDTA. The reduced-

unfolded intermediates were trapped, separated, analyzed with the same method as described for folding intermediates.

After unfolding reached equilibrium, 0.1ml of unfolding solution was removed and cysteine was added to a final concentration of 100 mM. After incubate at different time points, the intermediates were trapped, separated, and analyzed by the method describe above.

Disulfide Mapping of Refolding and Unfolding Intermediates

One-disulfide intermediate (I): After capture by CDAP, purification by HPLC, and drying by speedvac, the one-disulfide intermediate (I) was subjected to chemical cleavage by adding 2 µl of 1 M aqueous NH_4OH containing 6 M guanidine-HCl to dissolve the protein residue in the bottom of a 500-µl Eppendorf tube. Then an additional 5 µl of 1 M NH_4OH were added, and cleavage of the peptide chain was performed at room temperature for one hour. Excess ammonia was removed in the speedvac. Truncated peptides, which may still be linked by residual disulfide bonds, were completely reduced by reacting with 2 µl of 0.1 M TCEP solution at 37^0C for 30 minutes at pH 3-5. Samples were diluted with 20 µl of a 50% (v/v) acetonitrile/1% TFA solution prior to analysis by MALDI-MS.

Two-disulfide intermediate: Two experiments were applied for recognizing disulfide pairing in the two-disulfide intermediate. One portion of the purified cyanylated intermediate was subjected to the same treatment as the one-disulfide intermediate described above. This approach was applied to define the location of the two remaining sulfhydryl groups (that had not yet formed a covalent disulfide bond) by mass mapping of

the resulting cleaved peptides. Another aliquat of the same cyanylated intermediate was subjected to partial reduction/further cyanylation/chemical cleavage/mass mapping to determine the disulfide bond structure (24, 25). Partial reduction of the disulfide bonds in the two-disulfide intermediate was performed by adding a 20-molar excess of TCEP which was estimated from the ratio of the HPLC peak area for two-disulfide intermediate to the total of all peak areas representing various forms of the protein; the total protein in the mixture was fixed. The reduction was performed at room temperature for 10 minutes. After cyanylating the sulfhydryl groups of the partially reduced intermediate, HPLC was used to separate the partially reduced/cyanylated protein isoforms. The fractions corresponding to the singly reduced/cyanylated isomers were subjected to chemical cleavage, reduction of the remaining disulfide bond, and mass analysis of the cleavage products.

Three-disulfide intermediate: the three-disulfide intermediate contains three disulfide bonds based on no observed mass shift from the MW of the intact protein. The assignment of disulfide bonds in three-disulfide intermediate was performed by partial reduction/cyanylation/cleavage/complete reduction according to the procedure described above of the assignment of disulfides in two-disulfide intermediate.

HPLC Separation

The purification and separation of reduced LR^3IGF-I and IGF-I, the trapped (cyanylated) intermediates, and the partial reduction/cyanylation products were performed by reversed-phase HPLC with linear gradient elution using Waters model 6000 pumps controlled by a PC. UV detection was at 215 nm. The column was a Vydac

C18 (#218TP54, 10-μm particle size, 300-Å pore, 4.6×250 mm). Solvent A was water containing 0.1% TFA. Solvent B was acetonitrile/water (9:1, v/v) containing 0.1% TFA. The gradient was 30%-50% solvent B linear in 45 minutes. The flow rate was 1ml/min. The HPLC fractions were collected manually and the contents were then dried for further use.

IV. Results and Discussion

A. Refolding and Reductive Unfolding of LR^3IGF-I

Trapping and HPLC Separation of Folding Intermediates of LR^3IGF-I

The distribution of intermediates at any time during the refolding process can be visualized by a chromatogram of cyanylated and other (i.e., three-disulfide species) intermediates. Figure 4.3A shows chromatograms of an array of intermediates trapped by CDAP at acidic condition at various times showing the appearance and disappearance of certain species during the refolding process. In addition to native LR^3IGF-I (N) and completely reduced LR^3IGF-I (R), a total of 3 peaks were identified at different time points. The structural information of intermediates was determined by MALDI-MS based on the mass shift between the MW of intermediate and that of intact protein. Peak 1 contains two species, a three-disulfide intermediate and a mixed two-disulfide intermediate, according to the mass shift between the MW of intermediates and that of intact protein. While no mass shift was observed for the three-disulfide intermediate, a 612-Da mass shift was determined for mixed two-disulfide-intermediate. Peak 2 is a two-disulfide intermediate because a 52-Da mass shift was determined. Peak 3 represents a

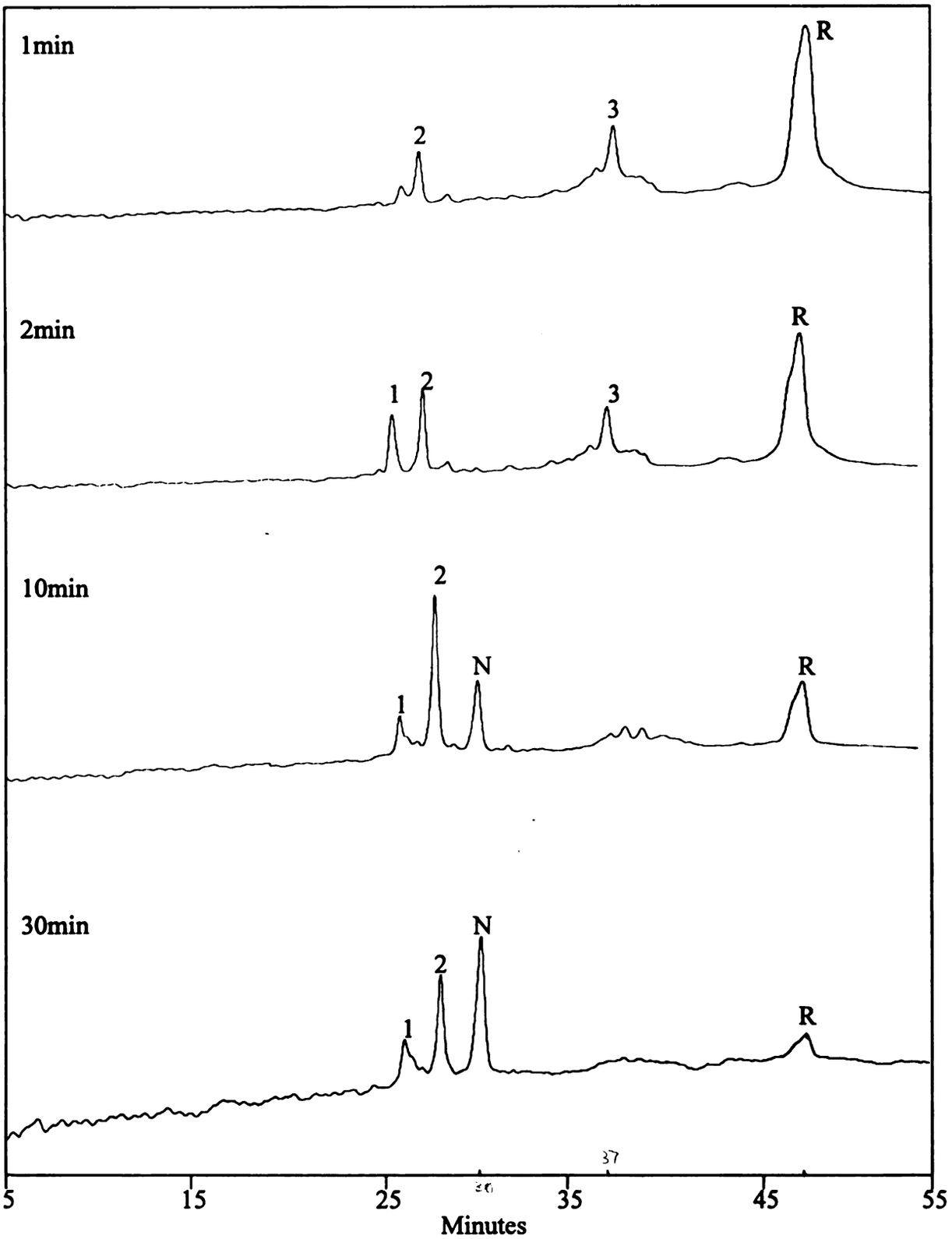


Figure 4.3A. HPLC separation of CDAP-trapped intermediates during the time course of refolding of LR^{3}IGF-I

one-disulfide intermediate based on a mass shift of 104 Da between the intermediate and the intact protein.

Reduced/unfolded LR^3IGF-I is first converted to a one-disulfide intermediate (peak 3) and a two-disulfide intermediate (peak 2) after 1 minute in the presence of 10mM GSH and 1mM GSSG. A three-disulfide intermediate and a mixed two-disulfide intermediate (coeluted in peak 1) appeared by 2 minutes. Native protein (N) was formed by 10 minutes. The ratio of the components changed after 10 minutes. However, the ratio of the components after several hours was almost the same as that at 30 minutes, indicating that the refolding of LR^3IGF-I reached equilibrium by 30 minutes. Therefore, the folding intermediates of LR^3IGF-I appeared in the order of one-disulfide intermediate, two-disulfide intermediate, three-disulfide intermediate, and native protein in the folding process of LR^3IGF-I.

After equilibrium was reached, if the concentration of GSSG was increased to 100mM and the mixture incubated for another 30 minutes, almost all of the two-disulfide and three-disulfide intermediates were converted to native/folded protein (Figure 4.3B). This result implies that the nonnative three-disulfide intermediate can convert to the native protein by the pathway(s) involving the two-disulfide intermediate.

Trapping and HPLC Separation of Unfolding Intermediates of LR^3IGF-I

Figure 4.4 shows the time-dependant distribution of the intermediates of reductive unfolding of LR^3IGF-I in the form of several HPLC chromatograms.

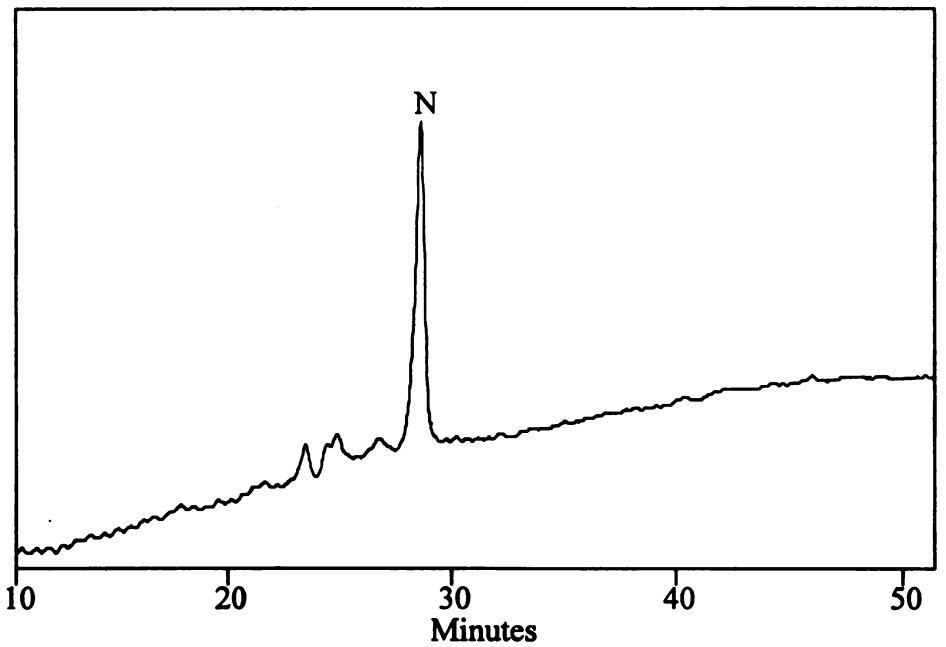


Figure 4.3B. HPLC separation of CDAP-trapped folding products after folding reached to the equilibrium, increased the concentration of GSSG to 100 mM and incubated for another 30 minutes.

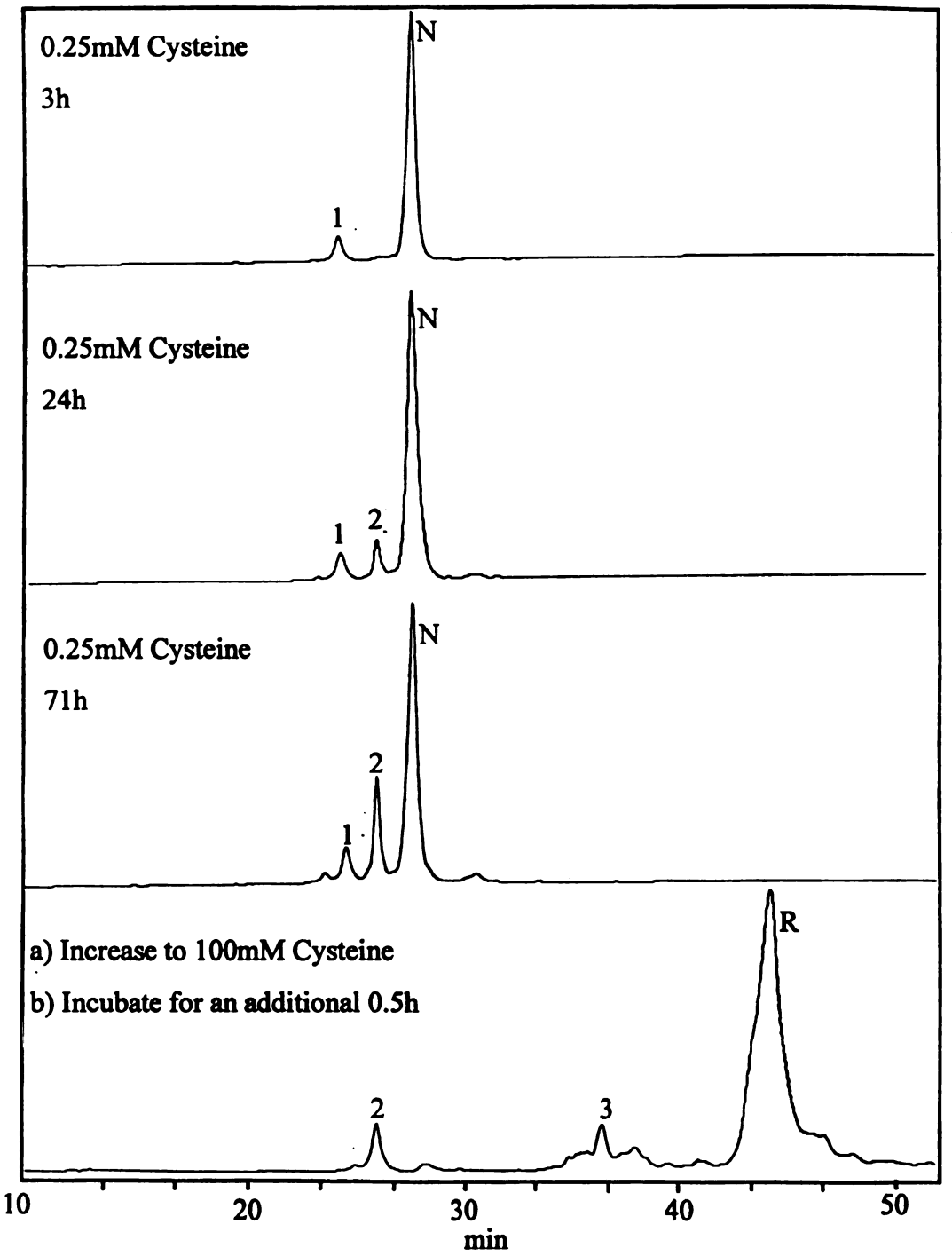


Figure 4.4. HPLC separation of CDAP-trapped intermediates in the time course of reductive unfolding of LR^{3}IGF-I

In order to avoid the formation of the mixed-disulfide intermediate, cysteine is used as a thiol reagent in the reductive unfolding. The pattern of the intermediates of reductive unfolding was similar to that of the refolding of LR^3IGF-I. A total of three intermediates was identified in addition to native (N) and reduced (R) forms. Peak 1 is the three-disulfide intermediate, peak 2 is the two-disulfide intermediate, and peak 3 is the one-disulfide intermediate. Reductive unfolding of LR^3IGF-I reached equilibrium at approximately 71 hours. However, if the concentration of cysteine was increased to 100 mM and incubated for another 30 minutes after reaching the 71-hour equilibrium point under the initial conditions, the one-disulfide intermediate (peak 3) and a large amount of reduced/unfolded protein (R) were formed. The intermediates of reductive unfolding of LR^3IGF-I appeared in the order of three-disulfide intermediate (peak1), two-disulfide intermediate (peak 2), one-disulfide intermediate (peak 3), and finally reduced/unfolded protein, which is the reverse of the folding process of LR^3IGF-I. It should be pointed out that some of the two-disulfide intermediate(s) should exist in the pathway of the conversion of native protein to the three-disulfide intermediate. However, no two-disulfide intermediates were trapped and identified even at early time of unfolding, indicating that the two-disulfide intermediates had too short a life time for our methodology.

Disulfide Structure of One-Disulfide Intermediate (peak 3) of LR^3IGF-I

The one-disulfide Intermediate (peak 1) contains only one disulfide bond, thus four sulfhydryl groups should be available (free) in the intermediate I. After intermediate I is trapped by CDAP at acidic condition, four sulfhydryl groups should be cyanylated. The

chemical cleavage reaction can be conducted in 1M aqueous NH_4OH right after the separation of the intermediates by HPLC, and MALDI-MS can be used to analyze the masses of the cleavage products. The location of the sulfhydryls can be determined by identifying the cleavage sites of the intermediate because only free sulfhydryl groups will be cyanylayed and the peptide bond at the N-terminal side of the cyanylated sulfhydryls will be cleaved.

Figure 4.5 shows the MALDI spectrum of the cleavage products of the cyanylated one-disulfide intermediate corresponding to HPLC peak 3 in Figure 4.3A. The expected and observed masses of the cleavage products of the one-disulfide intermediate are shown in Table 4.1. Fragments 1-18, itz19-59, itz61-83, and itz65-83 were detected. Ideally, five cleavage products were expected; however, because the adjacent cysteines were free, cleavage between them yields a single residue (as itz) and it is too low in mass to be detected (interference from matrix ions). Observing a cleavage product terminating at residue 59 implies that Cys60 was cyanylated. As a whole, these data indicate that peptide bonds at the N-terminal side of cyanylated cysteine residues at positions 19, 60, 61, and 65 were cleaved. Thus four free cysteine residues were located at positions 19, 60, 61, and 65. Obviously, no cleavage occurs at Cys31 and Cys74; otherwise, fragments related to cleavage at cyanylated Cys31 and Cys74 should be identified by MALDI-MS. Therefore, the Cys31-Cys 74 linkage can be deduced. It is notable that this intermediate has a native disulfide structure.

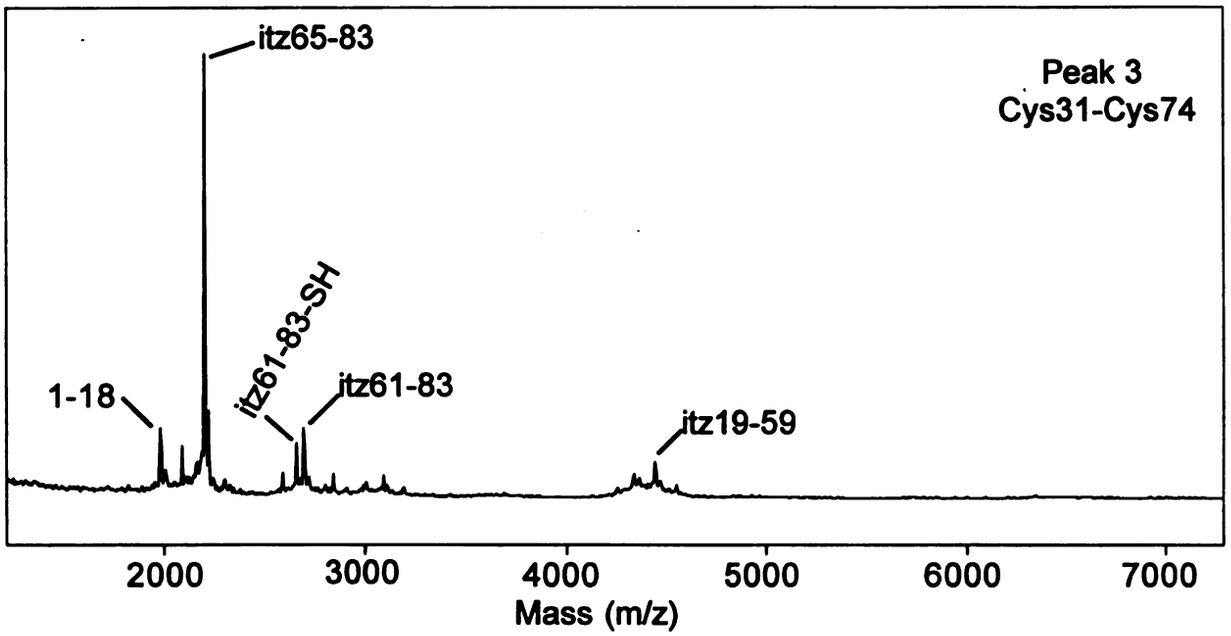


Figure 4.5. MALDI spectrum of peptide mixture resulting from the cleavage of the purified and cyanylated intermediate corresponding to peak 3 in Figure 3.4. The symbol "itz" represents iminothiazolidine derivatives.

Table 4.1. The expected and observed masses of cleavage products of one-disulfide intermediate of $LR^3IGF\text{-}I$

Fragments	Expected masses (Da)	Observed masses (Da)
1-18	1978.4	1978.5
itz19-59	4435.8	4436.7
itz60	145.2	nd
itz61-64	535.5	nd
itz65-83	2188.6	2190.7
itz61-83	2681.1	2684.0

Disulfide Structure of Two-Disulfide Intermediate (peak 2)

The intermediate corresponding to HPLC peak 2 in Figure 4.3A contains two disulfide bonds. The disulfide structure of the intermediate can be determined by the methodology of partial reduction/ cyanylation/chemical cleavage described in Chapter 3. The isolated fraction of HPLC peak 2 (Figure 4.3A) was subjected to partial reduction and cyanylation. Figure 4.6 shows the chromatogram of the intact intermediate and its partially reduced/cyanylated products. Two one-disulfide bonded species are expected from partial reduction of the intermediate (a two-disulfide species). However, only one one-disulfide species arises from the partial reduction and cyanylation. That is, the HPLC peak 1 (Figure 4.6) has one disulfide opened that has the retention time and MW of the intact intermediate, and peak 2 has two disulfides opened (its MW is 52 Da greater than that of the intact intermediate). Only one of the two expected one-disulfide isoforms is detected after cyanylation (peak 2 in Figure 4.6) because of the high stability of one of the two disulfide bonds in the intermediate.

Although the disulfide structure of the intermediate can be directly determined from the cleavage products of the residual cyanylated intermediate and partially reduced cyanylated isoform of the intermediate, it is much easier to identify the location of the two free sulfhydryls first. Table 4.2 lists the expected and observed masses of the cleavage products of HPLC peak 2 in Figure 4.3A. The MALDI spectrum (Figure 4.7A) of the cleavage products of the species corresponding to the HPLC peak 2 in Figure 4.3A shows fragments 1-59, itz65-83, b-elimination product itz60-(74*)-83, and uncleaved product itz60-(74)-83, indicating that cleavage occurred at cysteine residues 60 and 65. Therefore, Cys60 and Cys65 are the two free sufhydryls in the intermediate.

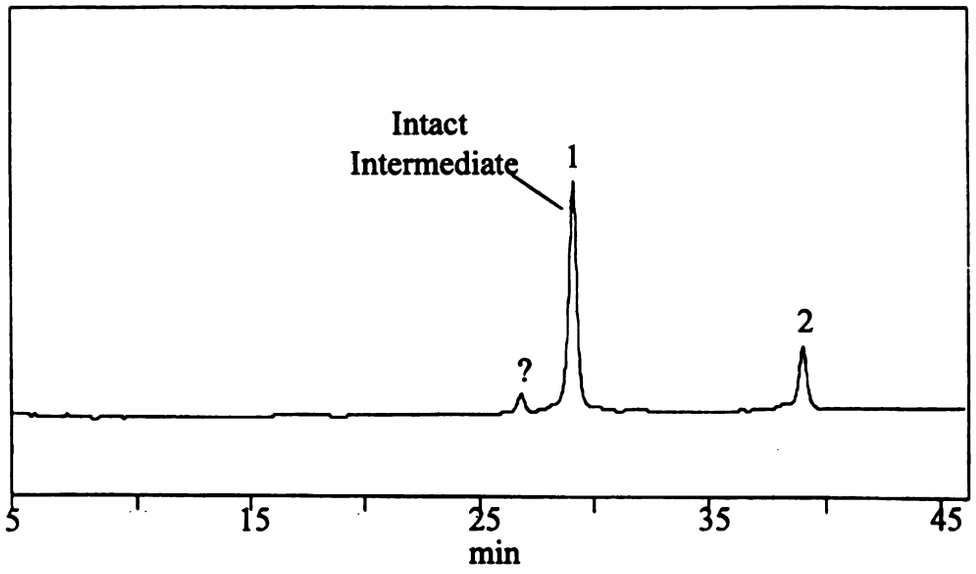


Figure 4.6. HPLC separation of intact intermediate (peak 1 here) corresponding to HPLC peak 2 in Figure 4.3 and one of its expected partially reduced/cyanylated isomers (peak 2). The peak with a question mark did not correspond to any specific cleavage fragments to the disulfide structures of the intermediates as analyzed by MALDI.

Table 4.2. The expected and observed masses of cleavage products of the purified/cyanylated two-disulfide intermediate of $LR^3IGF\text{-}I$

Fragments	Expected masses (Da)	Observed masses (Da)
1-59	6371.2	6373.3
itz60-64	637.7	nd
itz65-83	2188.6	2191.6
itz60-(65*)-83	2748.3	2755.1

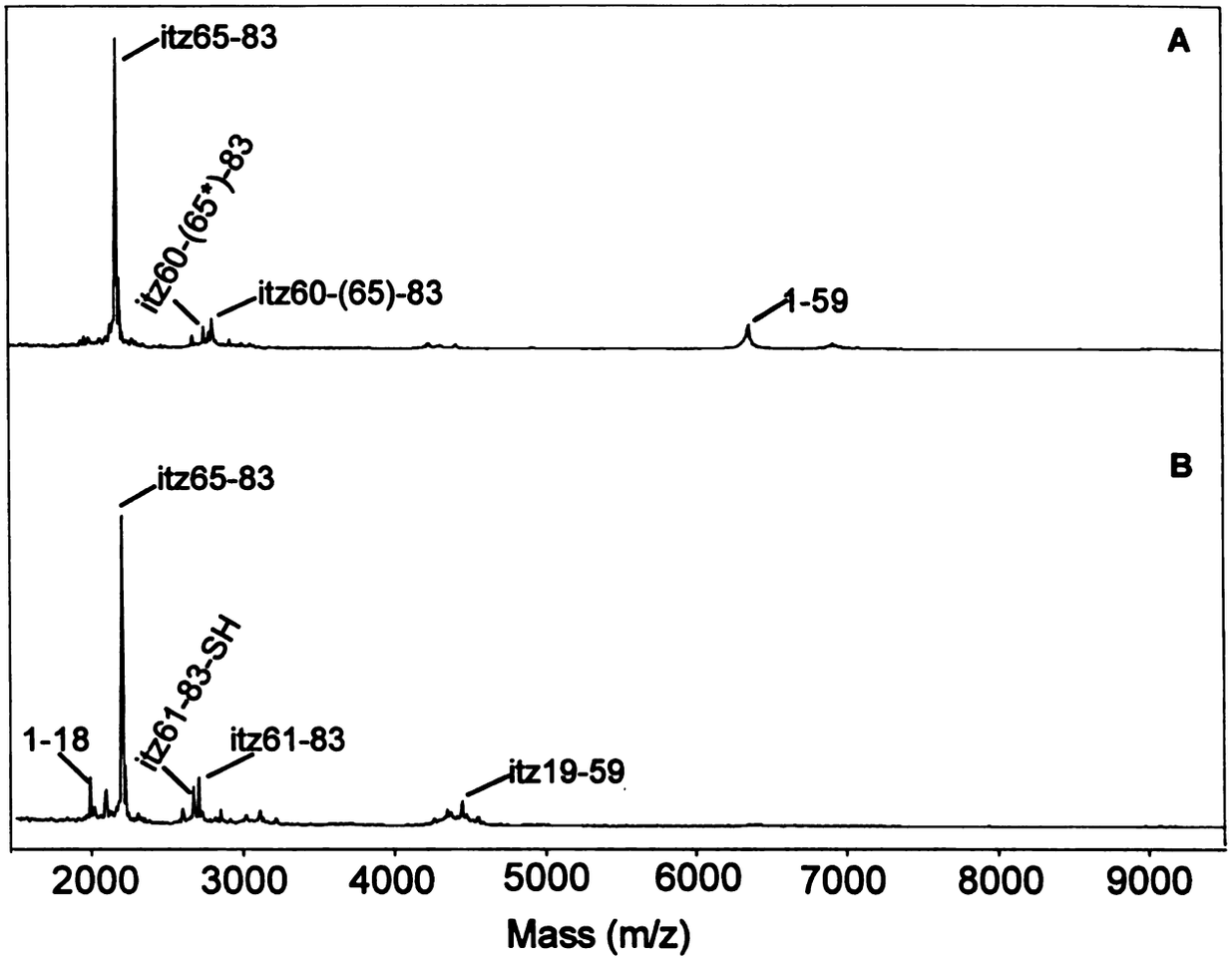


Figure 4.7. MALDI spectra of peptide mixtures resulting from the cleavage of the purified/cyanylated intermediate corresponding to HPLC peak 2 in Figure 4.3 and peak 1 in figure 4.6 (A) and one of its partially reduced/cyanylated isomers (peak 2 in Figure 4.6) (B). The symbol "itz" and * represent iminothiazolidine derivatives and β-elimination species, respectively. Itz60-(65)-83 represents that the residue in parenthesis (residue 65) is cyanylated but remains uncleaved. Itz60-(65*)-83 represents that the β-elimination occurs at residue 65.

Figure 4.7 B shows the MALDI spectrum of cleavage products of the cyanylated species represented by HPLC peak 2 in Figure 4.6. Fragment 1-18, itz19-59, itz61-83, and itz65-83 are identified, implying that another disulfide pair, Cys19-Cys61, must have been reduced, cyanylated and cleaved. Therefore, cysteine 19 is connected to cysteine 61 as a disulfide bond in the two-disulfide intermediate, and the second disulfide bond must be linked between the two remaining cysteine residues, 31 and 74.

In summary, the two-disulfide intermediate corresponding to HPLC peak 2 in Figure 4.3A contains a native structure, Cys19-Cys61 and Cys31-Cys74.

Disulfide Structure of Three-Disulfide Intermediate (peak 1)

Using the similar methodology of partial reduction/cyanylation described above one can recognize the disulfide structure of the three-disulfide intermediate. The HPLC chromatogram of the residual intact three-disulfide intermediate and its partially reduced/cyanylated isoforms is shown in Figure 4.8. Mass analysis showed that peaks 2 and 3 represent singly reduced/cyanylated species. Three singly reduced isoforms were expected from partial reduction, but the HPLC chromatogram shows only two singly reduced/cyanylated isoforms, indicating that one of three disulfide bonds is sufficiently stable to resist to reduction. Figure 4.9A and B show MALDI spectra of cleavage products of the species represented by peak 2 and peak 3, respectively. Table 4.3 lists the observed and expected masses of cleavage products of the species represented by peaks 2 and 3 in Figure 4.8. The MALDI spectrum of peak 2 (Figure 4.9A) identified fragment peaks consisting of residues 1-60, itz61-83, and itz65-83, indicating that Cys61 is linked to Cys65 as a disulfide bond.

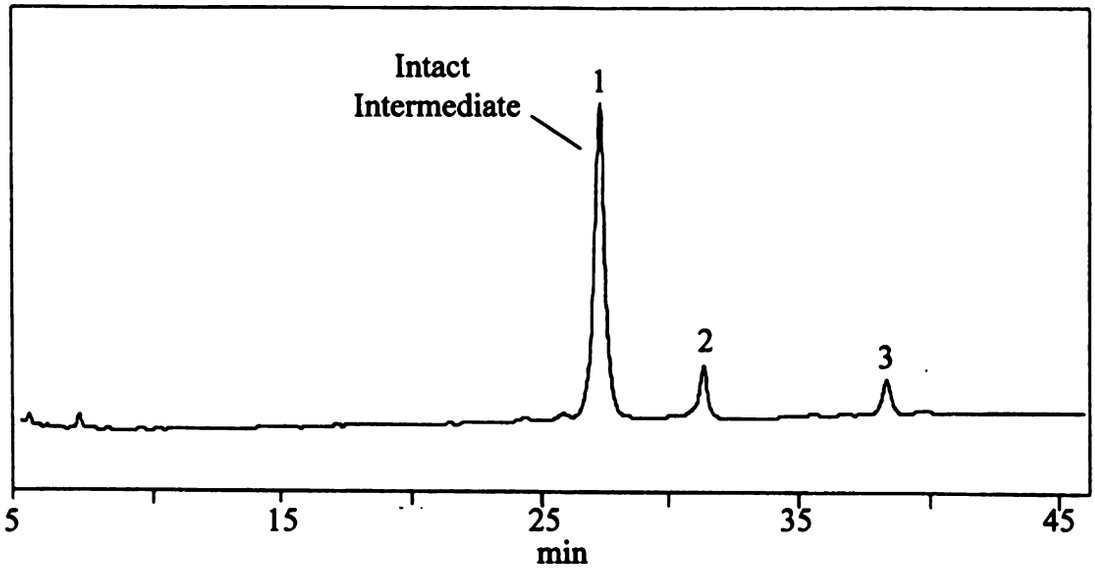


Figure 4.8. HPLC separation of residual intact intermediate corresponding to HPLC peak 1 in Figure 4.3 and its partially reduced/cyanylated isomers.

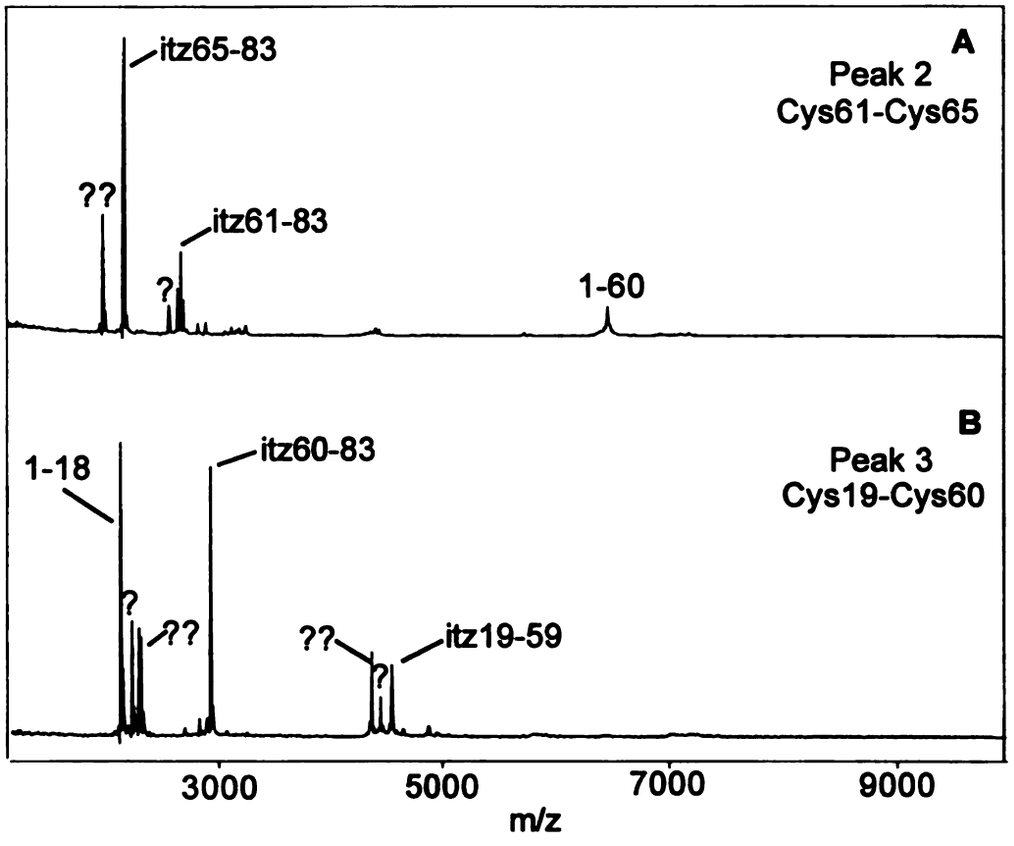


Figure 4.9. MALDI spectra of peptide mixtures resulting from cleavage of partially reduced/cyanylated isomers corresponding to HPLC peaks 2 and 3 in figure 4.8 of the three-disulfide intermediate (peak 1 in Figure 4.3). The symbol "itz" indicates iminothiazolidine derivatives. The question marked peaks are discussed in the text.

Table 4.3. The expected and observed masses of cleavage products of singly reduced/cyanylated isomers of the three-disulfide intermediate of $LR^3IGF\text{-}I$

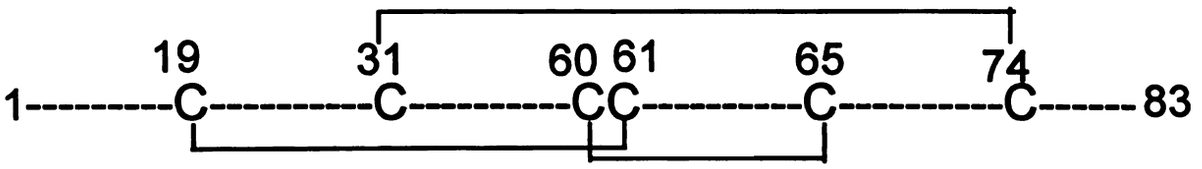
Fragments	Expected masses (Da)	Observed masses (Da)
1-60	6473.4	6476.3
itz61-83	2681.1	2684.9
itz65-83	2188.6	2191.4
1-18	1978.4	1978.6
itz19-59	4435.8	4437.4
itz60-83	2783.3	2788.1

The MALDI spectrum of the cleavage products of the species represented by peak 3 (Figure 4.9B) shows fragment peaks corresponding to residues 1-18, itz19-59, and itz60-83, indicating that Cys19 is connected to Cys60 as a disulfide bond. The third disulfide bond must be linked between the two remaining cysteines residues, Cys31 and Cys74. It should be pointed out that in addition to the mass spectral peaks corresponding to the specific cleavages related to reduced disulfides, some of the peaks corresponded to a mass of the MW of the expected fragment ±103 Da or ±185 Da were identified. How these ions were produced is still not clear. It may result from impurities of recombinant LR^3IGF-I or some side reactions. Further work should be done to obtain a reasonable explanation. A MS/MS experiment might be used to analyze the ±103Da and ±185Da species to provide sequence information. In spite of the existence of the ±103Da and ±185Da peaks, assignment of disulfide linkages is still possible, since the specific cleavage related to the cyanylated nascent sulfhydryls from reduced disulfides provides enough information.

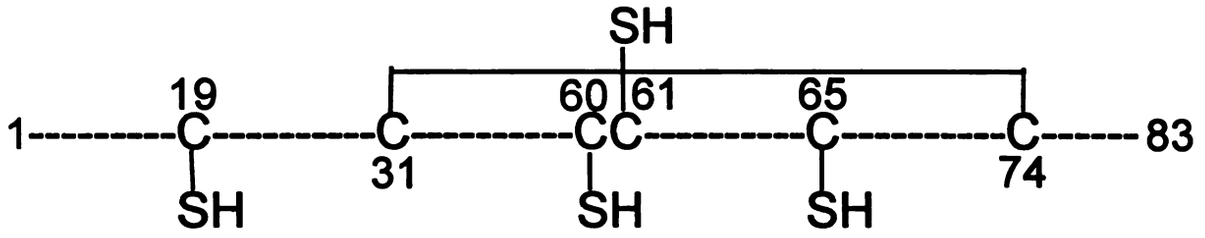
To summarize, three intermediates in the refolding of LR^3IGF-I have been captured and characterized, which include a one-disulfide intermediate (I), a two-disulfide intermediate (II), and a three-disulfide intermediate (III) (Figure 4.10).

B. Refolding and Reductive Unfolding of IGF-I

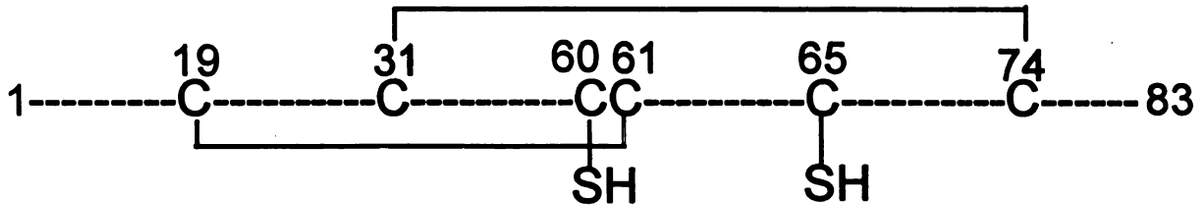
LR^3IGF-I is an analog of IGF-I with an N-terminal extension of 13 amino acids and replacement of glutamate 3 of IGF-I with arginine. The effect of structural features on the folding of LR^3IGF-I and IGF-I can be obtained by comparing their refolding and unfolding processes.



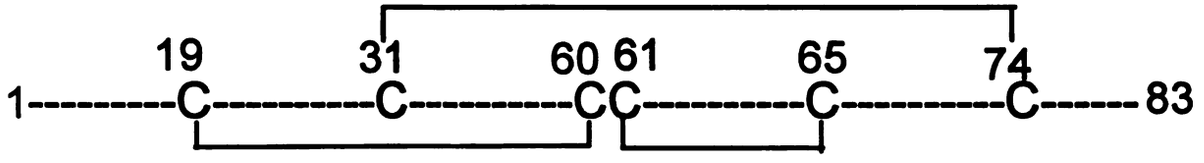
Native $LR^3IGF\text{-}I$ (N)



One-disulfide intermediate (I)



Two-disulfide intermediate (II)



Nonnative three-disulfide intermediate (III)

Figure 4.10. The disulfide structure of folding intermediates of $LR^3IGF\text{-}I$

Trapping and HPLC Separation of Folding Intermediates of IGF-I

Folding intermediates of IGF-I trapped at 10 seconds, 1 minute, 2.5 minutes, and 30 minutes, respectively, were separated by HPLC (Figure 4.11). The number of disulfide bonds was determined from mass shift between the MW of intermediates and intact IGF-I based on analysis by MALDI-MS. A total of 6 peaks were identified in addition to the native (N) and reduced (R) IGF-I. Peak 1 represents a three-disulfide species. Peaks 3, 4, 5 represent two-disulfide species. Peak 6 represents a one-disulfide specie. Peak 2 represents a mixed disulfide species according to the mass shift between the MW of the intermediate and that of the intact protein.

One of the two-disulfide intermediates (peak 3) and the one-disulfide intermediate appeared after 10 seconds. The other two two-disulfide intermediates (peaks 4 and 5) and the mixed-disulfide intermediate (peak 2) showed up after 1 minute. After 2.5 minutes the three-disulfide intermediate (peak 1) and native IGF-I formed. The folding between 1 minute and 2.5 minute was fast. The pattern of the folding intermediates of IGF-I at 30 minutes was similar to that at 2.5 minutes. The only difference is a small change in the ratio of the intermediates; the peak corresponding to native IGF-I was the largest one at 30 minutes, while the two-disulfide intermediate (peak 3) was the largest one at 2.5 minutes. Therefore, the appearance of the folding intermediates of IGF-I is in the order of a one-disulfide intermediate, the two-disulfide intermediates, and the three-disulfide species including native IGF-I.

The distribution of the folding intermediates of IGF-I trapped by cyanylation at acidic condition is similar to the published results that were obtained by acidic trapping (32, 33). Six intermediates, including a one-disulfide intermediate, three two-disulfide

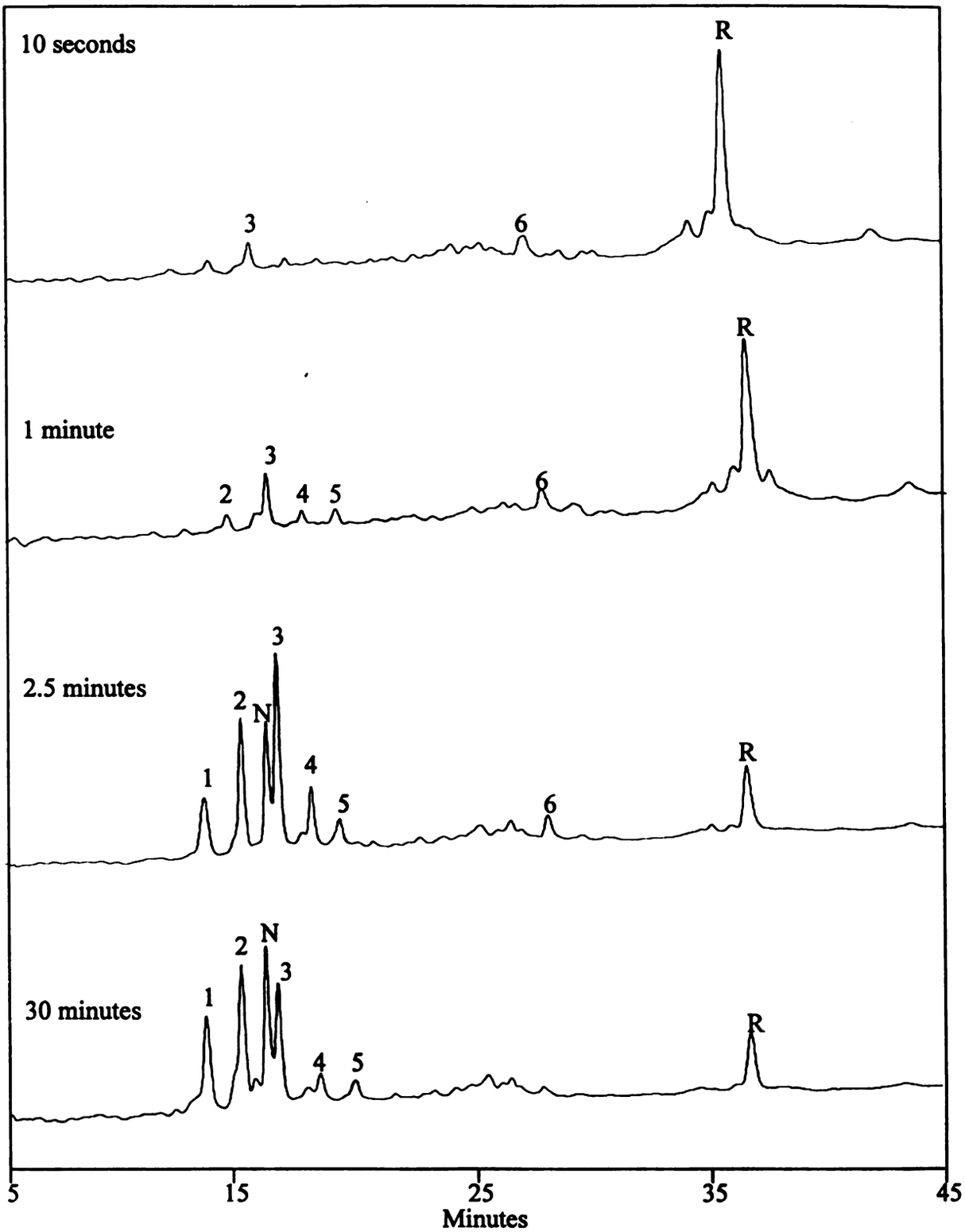


Figure 4.11. HPLC separation of CDAP-trapped intermediates during the time course of refolding of IGF-I.

intermediates, a three-disulfide intermediate, and a mixed disulfide intermediate, were captured here by our methodology and as well as by acidic trapping by others (32, 33), while only four intermediates were trapped by pyridylethylation under basic condition (29). Furthermore, the one-disulfide intermediate was the major form captured by pyridylethylation (29), while it was the minor form in the distribution of the folding intermediates trapped by our cyanylation methodology as described herein and as well as by acidification as reported elsewhere (32, 33).

Trapping and HPLC Separation of Unfolding Intermediates of IGF-I

Figure 4.12A shows the time-dependent distribution of the intermediates of reductive unfolding of IGF-I in the form of several HPLC chromatograms. Nonnative three-disulfide intermediate III (peak 1) appeared after 30 minutes in the presence of 0.25 mM cysteine. The ratio of the three-disulfide intermediate III and native IGF-I (N, peak 2) was almost the same at the time points of 2, 20, and 71 hours. This indicates that an equilibrium was reached at 2 hours. Two different features can be revealed between the unfolding processes of LR^3IGF-I (Figure 4.4) and IGF-I (Figure 4.12A). Firstly, The conversion of native IGF-I to the nonnative three-disulfide intermediate of IGF-I is faster than that of native LR^3IGF-I to the nonnative three-disulfide intermediate of LR^3IGF-I. This implies that the nonnative three-disulfide intermediate of IGF-I is more stable than the nonnative three-disulfide intermediate of LR^3IGF-I. This also can be seen from the refolding processes of IGF-I and LR^3IGF-I, in which the nonnative three-disulfide intermediate (peak 1 in Figure 4.11) of IGF-I has a higher yield (to total amount of IGF-I)

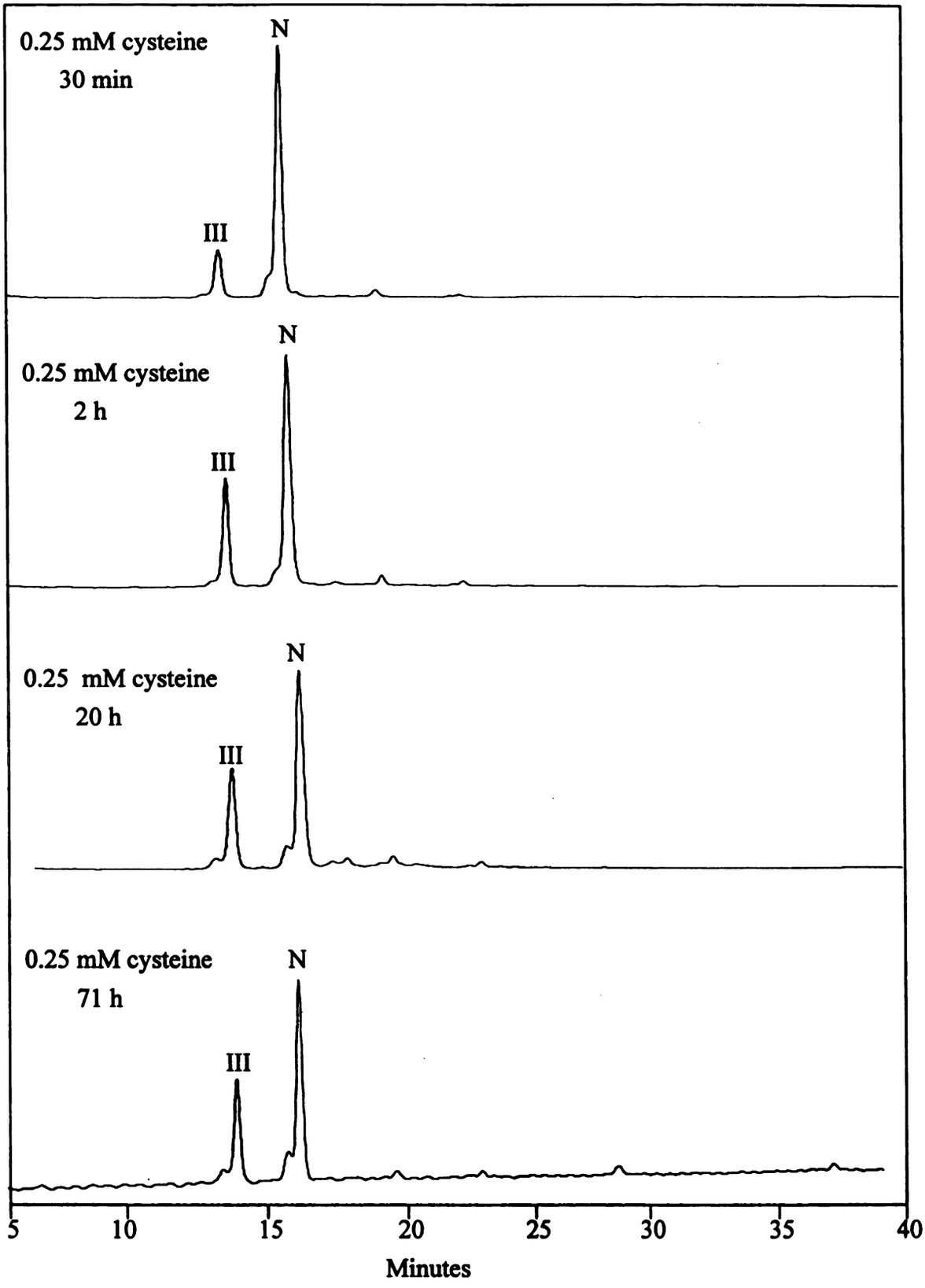


Figure 4.12A. HPLC seperation of CDAP-trapped intermediates in the time course of reductive unfolding of IGF-I

than that of the nonnative three-disulfide intermediate (peak 1 in Figure 4.3A) of LR^3IGF-I. Secondly, while there was no detectable two-disulfide intermediate II_A (see disulfide structure of II_A in Figure 4.19) in the unfolding process of IGF-I in the presence of 0.25 mM cysteine until even 71 hours, the two-disulfide intermediate (peak 2 in Figure 4.4) was appeared at 24 hours. This indicates that disulfide bond Cys47-Cys52 in IGF-I is more stable than that of corresponding disulfide bond Cys60-Cys65 in LR^3IGF-I.

Using the similar process to the unfolding of LR^3IGF-I, the concentration of cysteine was increased to 100 mM and incubated for another 1 and 30 minutes after the equilibrium of the reductive unfolding of IGF-I. Figure 4.12B shows the chromatograms of the unfolding products of IGF-I at 1 and 30 minutes after the equilibrium. The one-disulfide intermediate (I) and the two-disulfide intermediate (II_A) were formed at 1 minute. After 30 minutes, two-disulfide intermediate II_A was disappeared and one-disulfide intermediate I almost completely converted to reduced-unfolded IGF-I (R). The second-stage unfolding (after the equilibrium) of IGF-I is faster than that of LR^3IGF-I (Figure 4.4), which indicates that the disulfide bond Cys6-Cys48 in native IGF-I is reduced rapidly after the reduction of disulfide bond Cys47-Cys52. This result may be due the cooperative effect between these two disulfide bonds.

Disulfide Structure of Folding and Unfolding Intermediates of IGF-I

The disulfide structure of the folding intermediates of IGF-I were determined by using the same methodology as described earlier for identifying the disulfide structure of the folding and unfolding intermediates of LR^3IGF-I. Table 4.4 lists the expected and

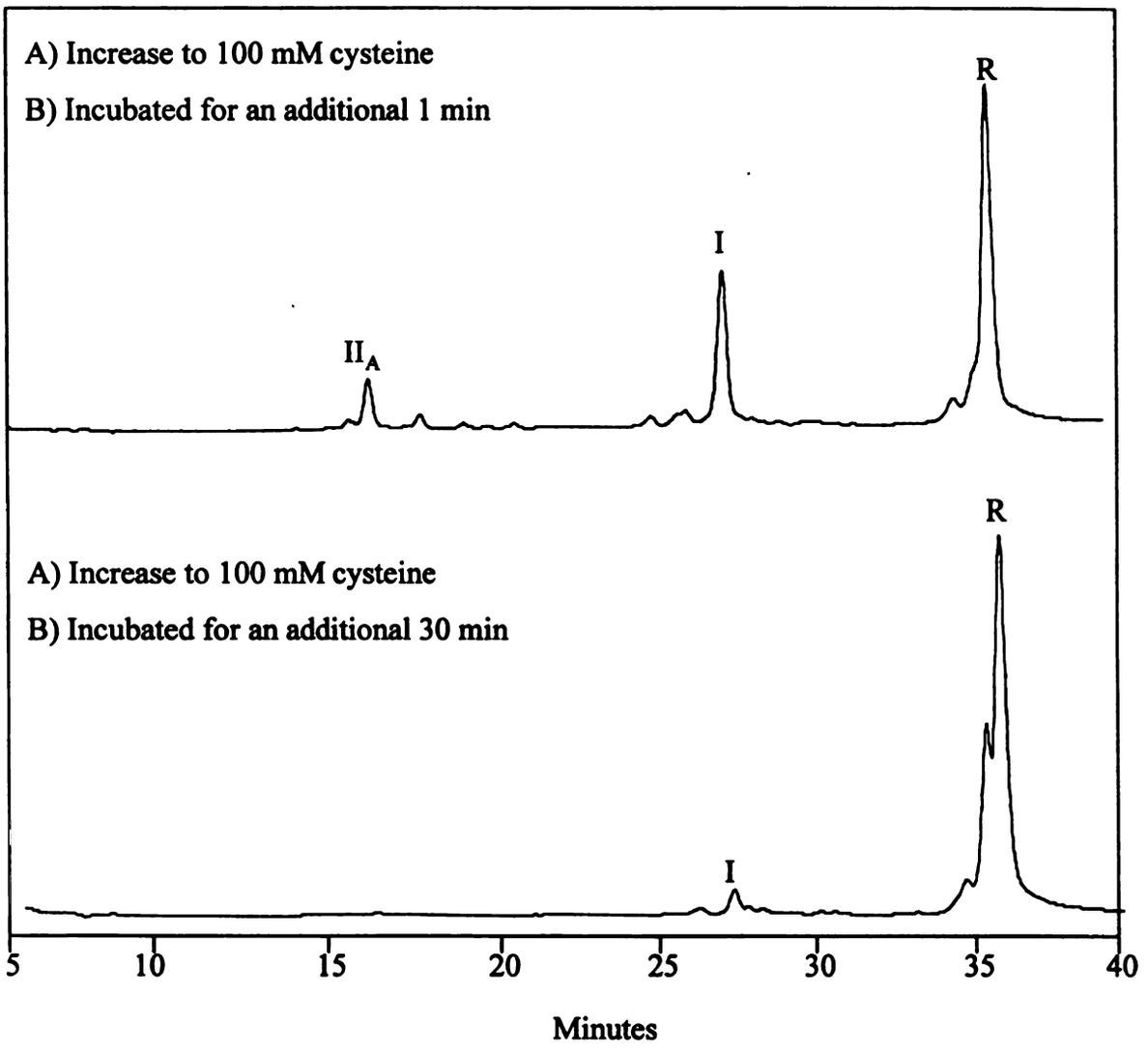


Figure 4.12B. HPLC separation of CDAP-trapped intermediates in the unfolding after the equilibrium.

Table 4.4. The expected and observed masses of the fragments related to the disulfide assigment of the intermediates of IGF-I

Fragments	Expected masses (Da)	Observed masses (Da)
1-5	515.54	nd
1-47	5010.5	5012.2
1-46	4908.5	4909.1
itz6-46	4435.8	4433.9
itz47	145.2	nd
itz47-70	2784.1	2786.7
itz47-51	535.5	nd
itz48-51	637.7	nd
itz48-70	2681.1	2684.4
itz52-70	2188.6	2191.0

observed masses of the fragments related to the disulfide assignment of the intermediates of IGF-I.

Figure 4.13 represents the chromatogram of partially reduced/cyanylated isomers of the three-disulfide intermediate (peak 1 in Figure 4.11). Peaks 2 and 3 indicate the singly reduced/cyanylated isomers of the intermediate based on the mass shift between the MW of isomers and intact intermediate. The fractions of peaks 2 and 3 were collected and incubated in aqueous ammonia. From analysis of the cleavage products of the cyanylated isomer represented by peak 2, fragments consisting of residues 1-47, itz48-70, itz52-70, and itz48-70 minus the mass of an SH group were identified (Figure 4.14A), indicating cleavage sites at residues 48 and 52. Therefore, the connection between Cys48 and Cys52 can be deduced. The MALDI spectrum of the cleavage products of peak 3 (Figure 4.14B) shows the fragments itz6-46 and itz47-70, indicating cleavage sites at residues 6 and 47. Thus, Cys6 is connected to Cys47 as a disulfide bond in the intermediate. Although a singly reduced/cyanylated isomer was not obtained for the remaining disulfide bond due to its high stability, its linkage between Cys18 and Cys61 can be easily deduced by default. Thus, the three-disulfide intermediate (peak 1 in figure 4.11) has disulfide bonds between residues 48-52, 5-47, and 18-61.

The intermediate corresponding to HPLC peak 3 in Figure 4.11 is a two-disulfide intermediate. The chromatogram of its partial reduction/cyanylation reaction mixture is shown in Figure 4.15. Instead of two partially reduced/cyanylated isomers, only one isomer was identified (peak 2) apparently because of the stability of one of the disulfide bonds. The MALDI spectrum (Figure 4.16A) of the cleavage reaction mixture for the species represented by peak 1 in Figure 4.16 (also peak 3 in Figure 4.11) shows

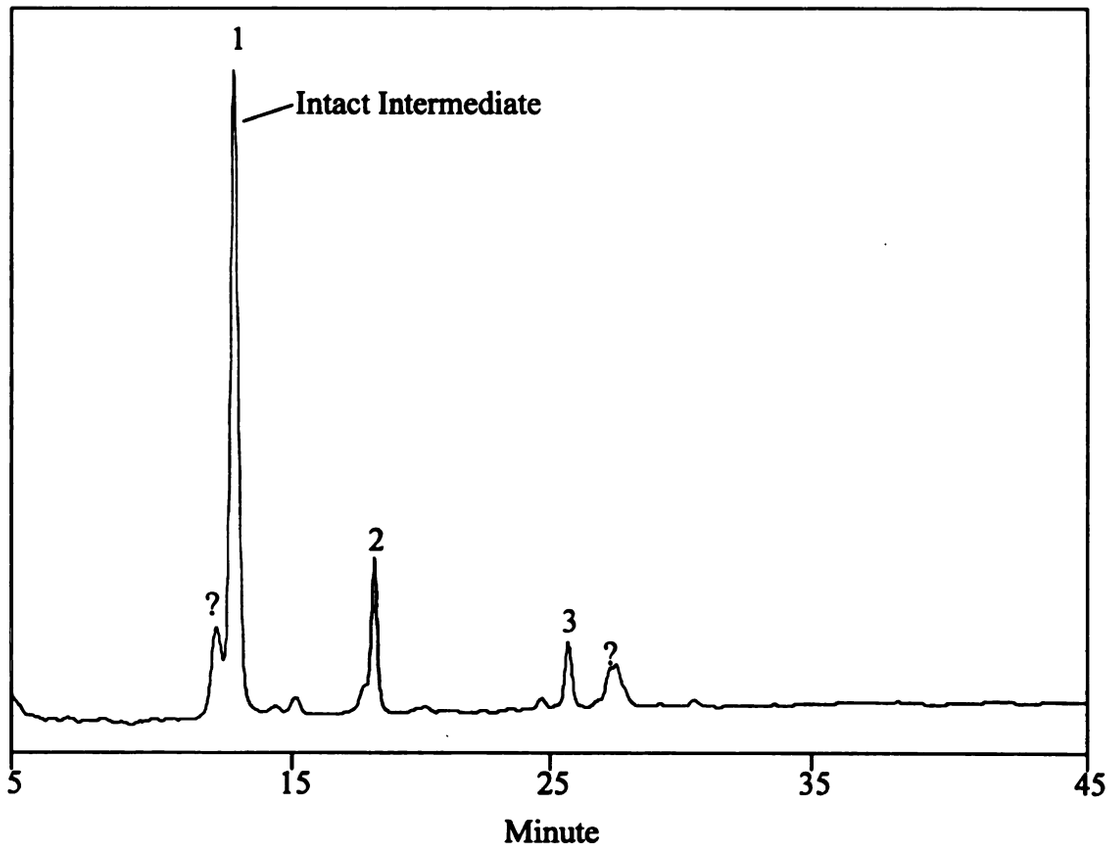
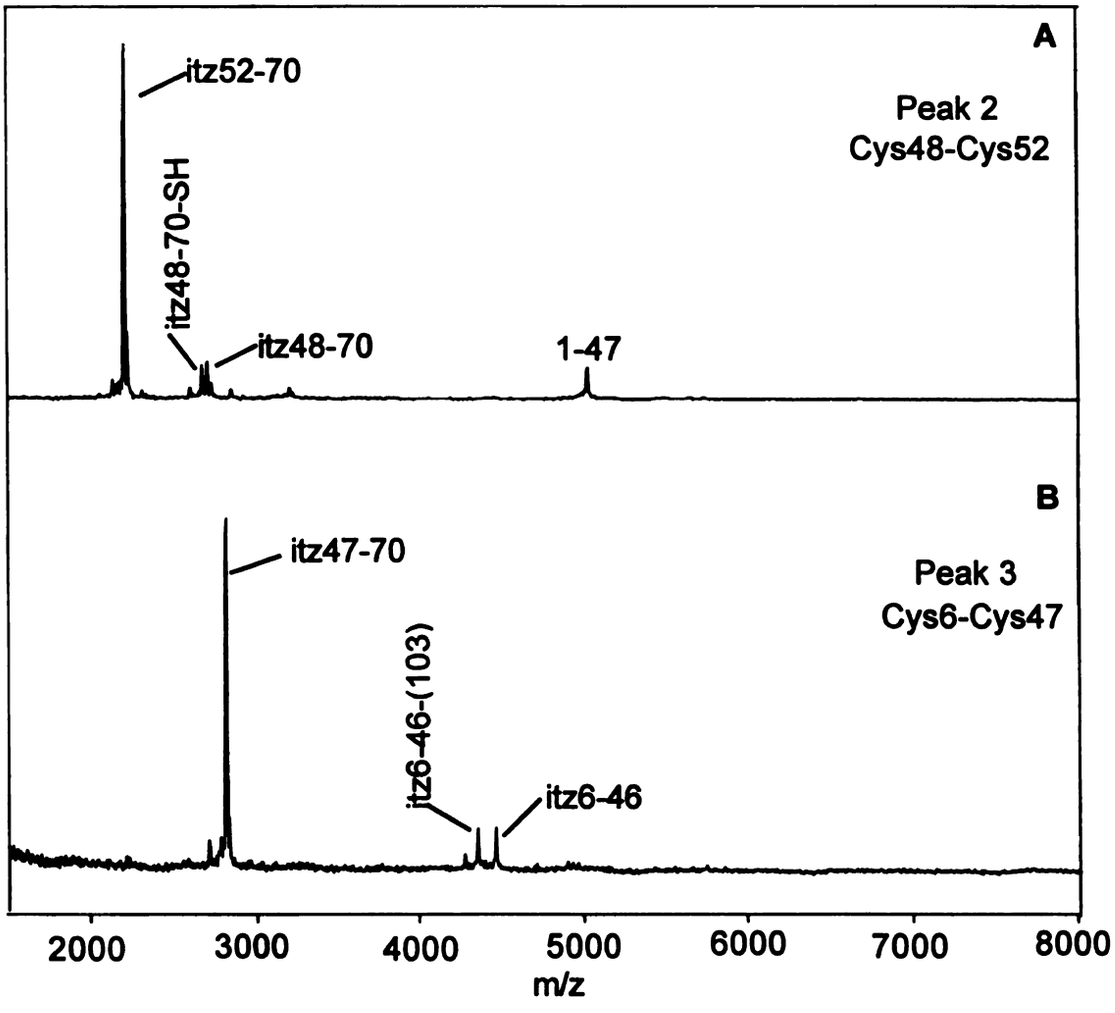


Figure 4.13. HPLC separation of residual intact three-disulfide intermediate corresponding to HPLC peak 1 in Figure 4.11and its partially reduced/cyanylated isomers. Peaks 2 and 3 each represent two-disulfide species. The question marked peaks are discussed in the text.



Figur 4.14. MALDI spectra of peptide mixtures resulting from the cleavage of the two singly reduced/cyanylated isomers, corresponding to HPLC peaks 2 and 3 in Figure 4.13, respectively. The symbol "itz" represents iminothiazolidine derivatives.

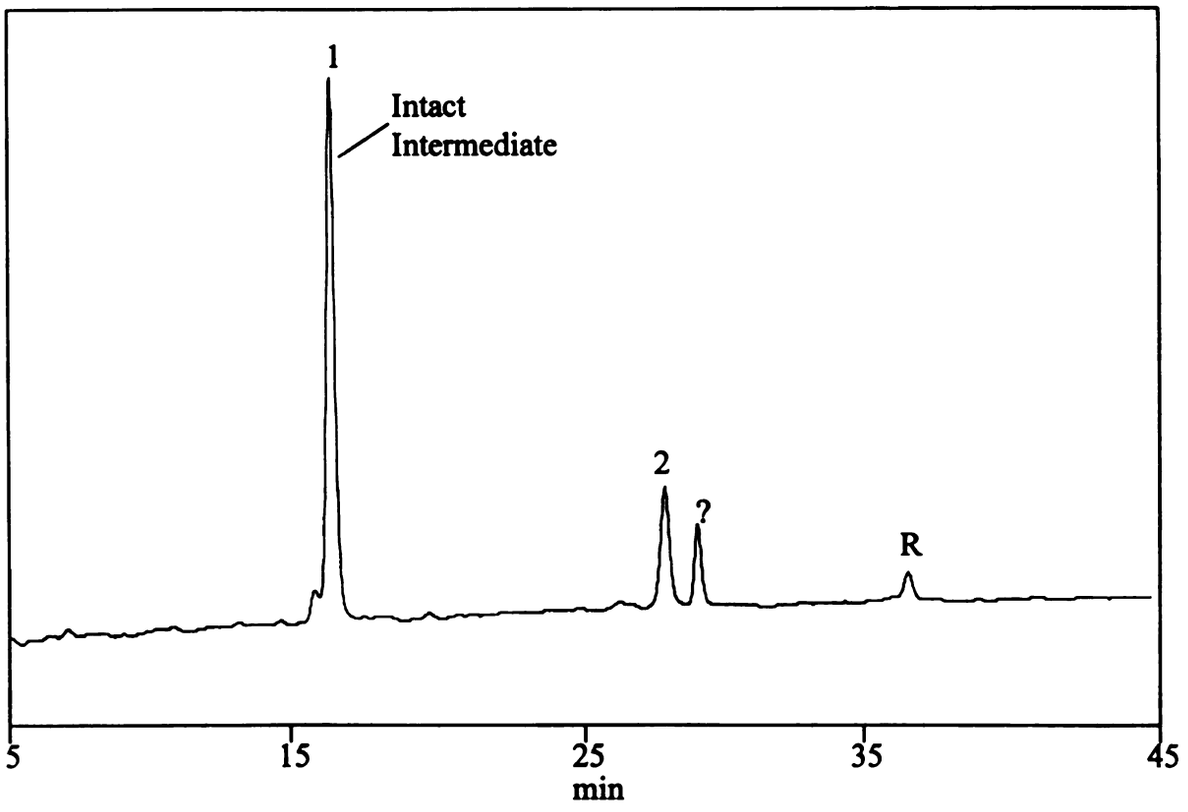


Figure 4.15. HPLC chromatogram of partial reduction/cyanylation mixture showing residual intact intermediate (two-disulfide species) corresponding to HPLC peak 3 in Figure 4.11and one of the expected partially reduced/cyanylated species (peak 2). The question marked peak is discussed in the text.

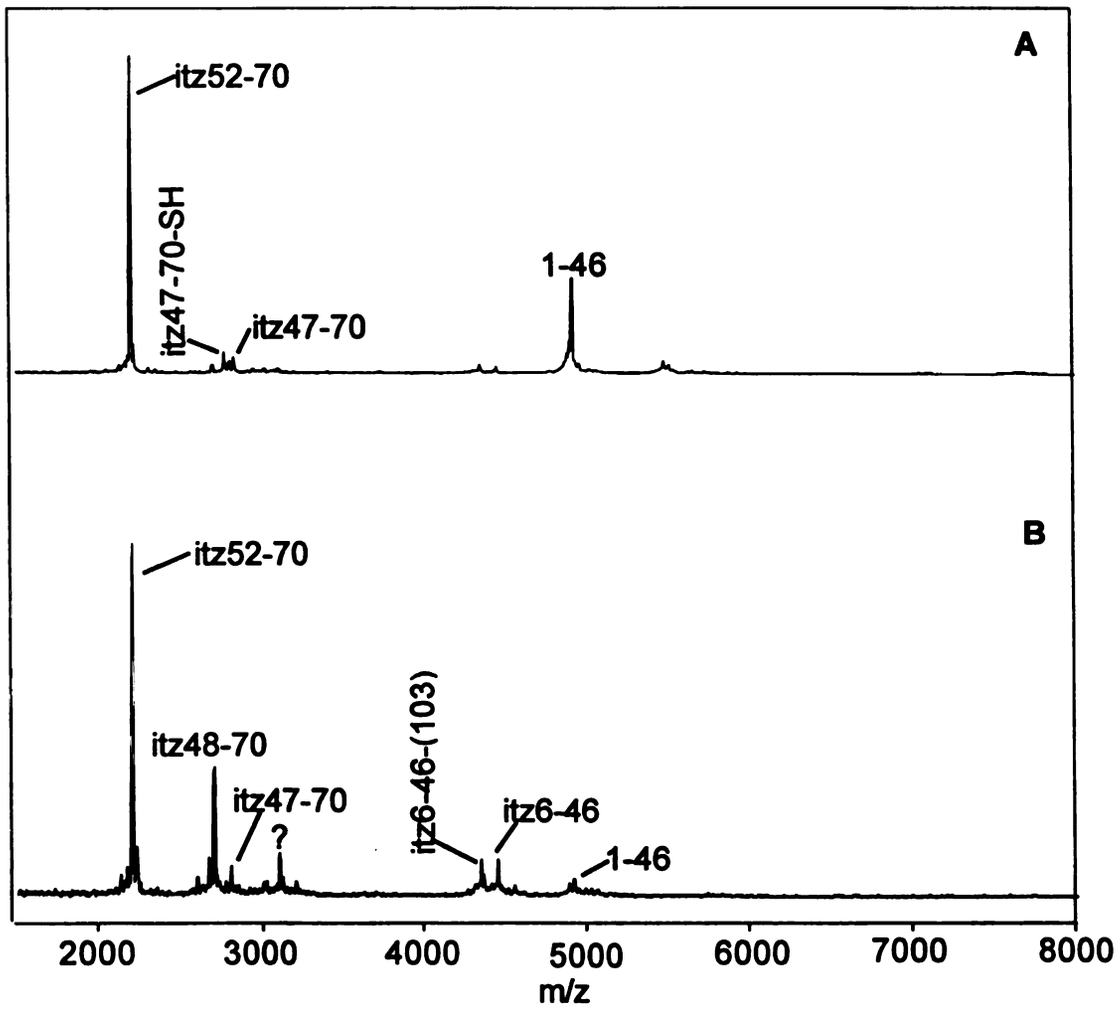


Figure 4.16. MALDI spectra of peptide mixtures resulting from the cleavage of the purified and cyanylated intermediate corresponding to peak 1 in Figure 4.14 (also to peak 3 in Figure 4.11) (A) and the partially reduced/cyanylated isomer corresponding to HPLC peak 2 in Figure 4.14 (B). The symbol "itz" and * represent iminothiazolidine derivatives and β-elimination peak, respectively. The peak with a question mark cannot be assigned to any specific cleavage fragment of the intermediate.

fragments consisting of residues 1-46, itz47-70, itz52-70, and itz47-70 minus the mass of SH group, indicating that the cleavage sites occur at cysteine residues 47 and 52, which represent two free sulfhydryl groups in the two-disulfide intermediate. The MALDI spectrum (Figure 4.16B) of the cleavage products represented by peak 2 in Figure 4.15 shows fragments consisting of residues itz6-46, 1-46, itz47-70, itz48-70, and itz52-70, indicating that another disulfide bond, Cys6-Cys48, must have been reduced, cyanylated and cleaved. Therefore, Cys6 is connected to Cys48 as a disulfide bond in the two-disulfide intermediate. The remaining two cysteine residues, Cys18 and Cys61, must be linked together as a disulfide bond as deduced by default. Thus, the two-disulfide intermediate corresponding to peak 3 in Figure 4.11 contains a native disulfide structure, Cys6-Cys48 and Cys18-Cys61.

It should be pointed out that the question marked peaks in the chromatograms in Figure 4.13 and Figure 4.15 did not correspond to any specific cleavage fragments related to the disulfide structures of the intermediates as analyzed by MALDI. We thought at first that the peaks corresponded to species resulting from a side reaction of CDAP with the intermediates. However, these species also showed up in the chromatogram when only partial reduction with TCEP was performed but without cyanylation with CDAP. Therefore, the question-marked peaks apparently result from the process of partial reduction, which may be due to some impurity that coeluted with the intermediates during the HPLC separation of the folding products.

Figure 4.17A and B show the MALDI spectra of cleavage products of the cyanylated two-disulfide intermediates corresponding to peaks 4 and 5 in Figure 4.11, respectively. The fragments 1-47, itz48-70, and itz52-70 were identified (Figure 4.17A)

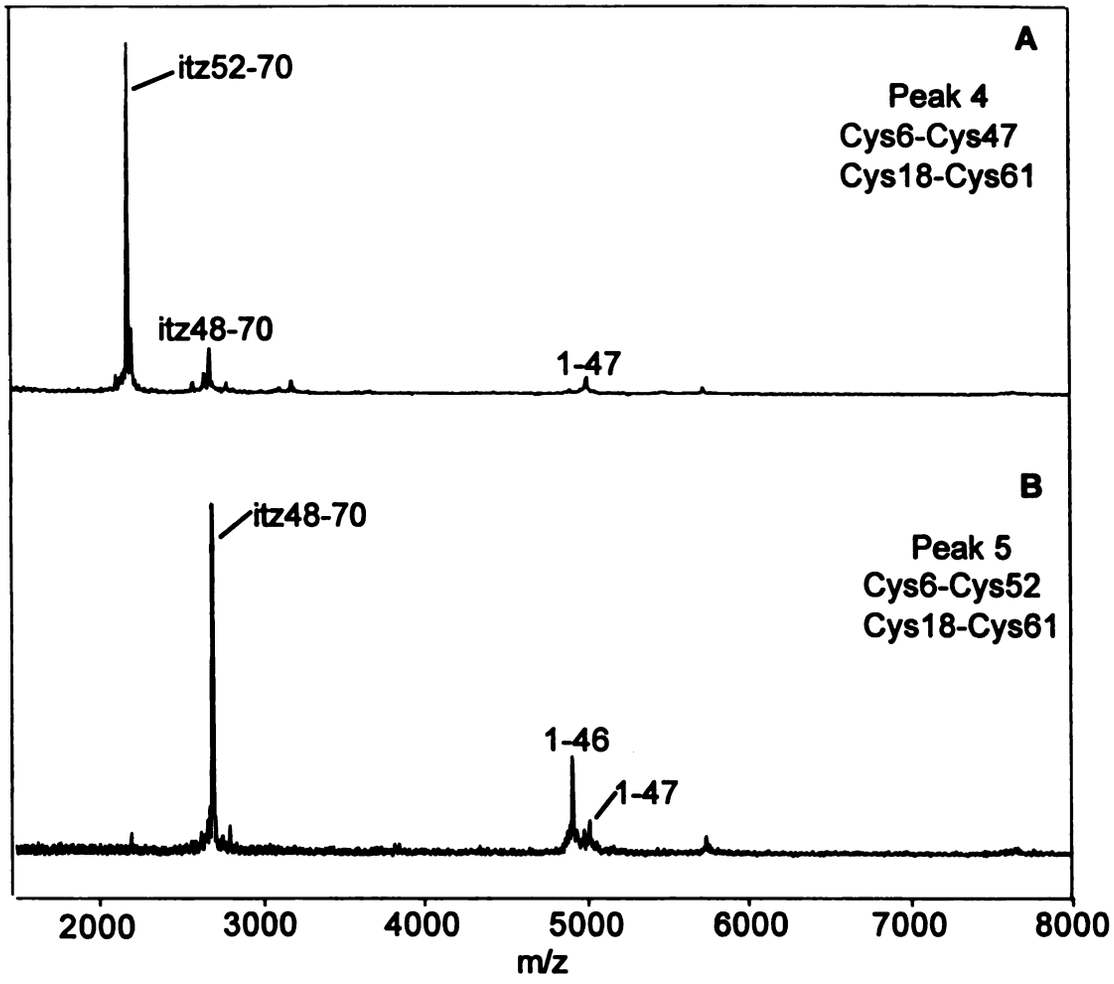


Figure 4.17. MALDI spectra of peptide mixtures resulting from the cleavage of purified and cyanylated two-disulfide intermediates corresponding to peaks 4 and 5 in Figure 4.11, respectively. The symbol "itz" represents iminothiazolidine derivatives.

from the cleavage products of the intermediate represented by peak 4, indicating that Cys48 and Cys52 are two free sulfhydryl groups; thus Cys6, Cys18, Cys47, and Cys61 must be linked together as two disulfide bonds. The disulfide linkages can be reasonably deduced as Cys6-Cys47 and Cys18-Cys61 based on the known high stability of the disulfide bond between Cys18 and Cys61.

Fragments 1-46, 1-47, and itz48-70 were identified in the MALDI spectrum (Figure 4.17 B) of the cleavage products of peak 5 in Figure 4.11, indicating cleavage sites at residues 47 and 48. Because residues 47 and 48 are free sulfhydryls in this intermediate, the linkages of the two-disulfide bonds should be between Cys6-Cys52 and Cys18-Cys61 considering the known high stability of the disulfide bond between Cys18-Cys61. Therefore, the two-disulfide intermediates corresponding to HPLC peak 4 in Figure 4.11 contains the nonnative disulfide structure, Cys6-Cys47 and Cys18-Cys61; the intermediate corresponding to HPLC peak 5 in Figure 4.11 has the nonnative disulfide structure of Cys6-Cys52 and Cys18-Cys61.

The intermediate corresponding to HPLC peak 6 in Figure 4.11 contains one disulfide bond. Figure 4.18 shows the MALDI spectrum of the cleavage products of this purified and cyanylated one-disulfide intermediate. The fragments itz6-46, 1-46, itz48-70, itz47-70, and itz52-70 were identified, indicating that four free sulfhydryl groups are located at positions 6, 47, 48, and 52. Therefore, Cys18 is linked to Cys61 as a disulfide bond.

To summarize, six folding intermediates were captured and characterized by our methodology (Figure 4.19). The one-disulfide intermediate (I) contains disulfide bond between Cys18-Cys61. Three two-disulfide intermediates were identified. One two-

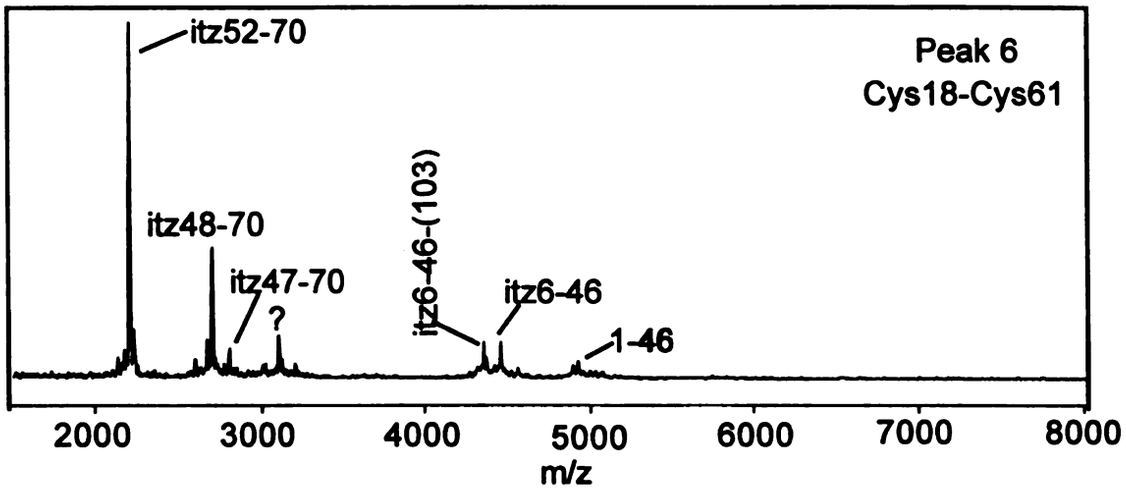


Figure 4.18. The MALDI spectrum of cleavage products of purified/cyanylated intermediate corresponding to HPLC peak 6 in Figure 4.11. The symbol "itz" represents iminothiazolidine derivatives.

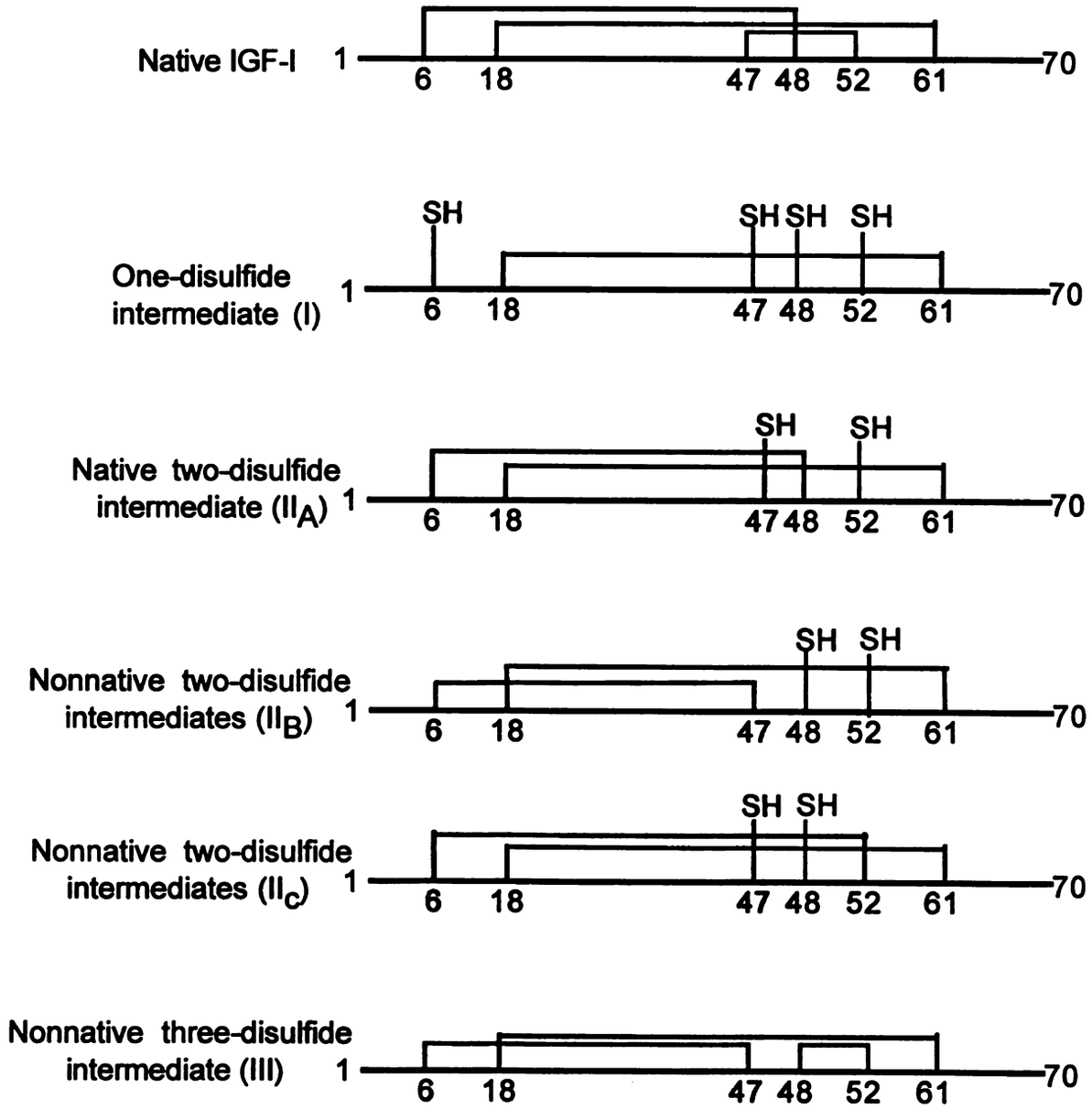


Figure 4.19. The disulfide structure of refolding intermediates of IGF-I.

disulfide intermediate (II_A) contains native disulfide structure Cys6-Cys48 and Cys18-Cys61. The other two two-disulfide intermediates have nonnative disulfide structures, Cys6-Cys47 and Cys18-Cys61 (II_B); Cys6-Cys52 and Cys18-Cys61 (II_C). The nonnative three-disulfide intermediate (III) contains disulfides Cys6-Cys47, Cys48-Cys52, and Cys18-Cys61. A mixed disulfide intermediate was also identified.

V. Isolation and Refolding of the Folding Intermediates of LR^3IGF-I

Additional detailed information on the refolding of a protein can be obtained from a stop/go experiment. The folding intermediates are isolated and purified at pH 3, and then subjected to a second-stage refolding by increasing the pH to 8.7 in the presence of GSSG/GSH.

Figure 4.20 shows the distribution of the intermediates after 30-minutes of continued refolding starting from the purified two-disulfide intermediate (II) or the purified three-disulfide intermediate (III) of LR^3IGF-I. The three-disulfide intermediate (III) can convert to the two-disulfide intermediate (II) and native LR^3IGF-I. The distribution of species after 30 minutes of second-stage refolding is similar to the equilibrium distribution shown in Figure 4.3A, especially for the continued refolding of the purified two-disulfide intermediate (II) as shown in Figure 4.20 (top panel). Continued refolding of the purified two-disulfide intermediate (II) gave a similar result to that of the purified three-disulfide intermediate (III) as shown in Figure 4.20. These results imply that the three species (II, III, and N) can interconvert to each other by breakage/reformation of disulfide bonds.

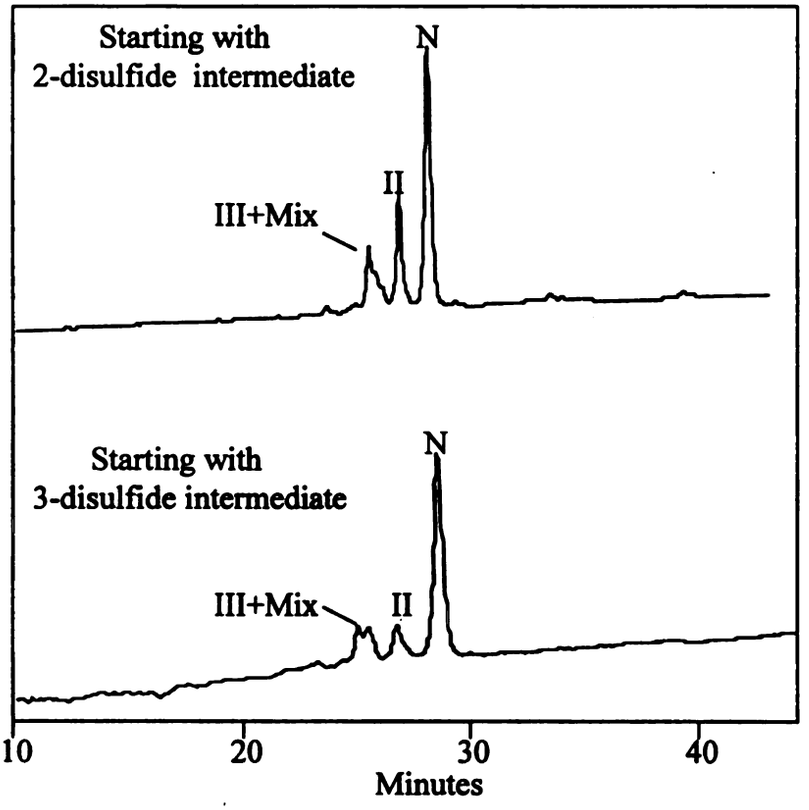


Figure 4.20. HPLC separation of CDAP-trapped intermediates after 30-minutes of refolding starting from the purified 2-disulfide intermediate (II) or 3-disulfide intermediate (III) of LR^3IGF-I in the same refolding buffer as used in the refolding of reduced/unfolded LR^3IGF-I.

VI. Comparison of the Folding Process of LR^3IGF-I and IGF-I

LR^3IGF-I is an analog of IGF-I with the differences at N-terminal side. There are several different aspects between the folding processes of IGF-I and LR^3IGF-I.

1). Fewer intermediates in the folding pathway of LR^3IGF-I were identified than in the folding pathway of IGF-I. The extension of 13 amino acids at N-terminus of LR^3IGF-I might result in the reducing the options of forming different disulfide bonds due to the lower flexibility in LR^3IGF-I.

2). In the folding process of IGF-I, two nonnative two-disulfide intermediates were detected. However, there were no detectable nonnative two-disulfide intermediates in the folding or reductive unfolding process of LR^3IGF-I.

3). The disulfide structures of the nonnative three-disulfide intermediate, the native two-disulfide intermediates, and the one-disulfide intermediate of LR^3IGF-I and IGF-I have similar disulfide-linkage patterns.

4). The folding process of IGF-I is much faster than that of LR^3IGF-I in the early folding stages, which may be due to the lower flexibility of LR^3IGF-I since the protein is a more extended random coil at the early folding stage.

VII. Conclusions

The folding and reductive unfolding intermediates of LR^3IGF-I were captured and characterized by cyanylation methodology, which include a native one-disulfide intermediate, a native two-disulfide intermediate, and a nonnative three-disulfide intermediate. The pathways of refolding and reductive unfolding of LR^3IGF-I can be proposed on the basis of the time courses of the refolding and unfolding (Figure 4.3A,

4.3B and 4.4) and on the structure of disulfide bonds of the intermediates. The proposed pathways are illustrated in Figure 4.21.

In the pathway of the refolding of LR^3IGF-I, the reduced/unfolded LR^3IGF-I was converted to one-disulfide intermediate (I) with the formation of the first disulfide bond, 31-74. Then intermediate I was converted to two-disulfide intermediate (II) with the formation of the second disulfide bond, 19-61. Three-disulfide intermediate (III) can be formed from the two-disulfide intermediate II based on the results of the stop/go experiment starting with the two-disulfide intermediate II (Figure 4.20). Some transient nonnative two-disulfide intermediates should exist in the pathway of two-disulfide intermediate II to three-disulfide intermediate III since two-disulfide intermediate II cannot convert to the three-disulfide intermediate III without the breakage/reformation of disulfide bonds. However, no nonnative two-disulfide intermediates were trapped even at an early stage of the unfolding because of the apparent short lifetime of the intermediates. Alternatively, it is possible that the three-disulfide intermediate III is formed by the pathway of the one-disulfide intermediate (I) to nonnative two-disulfide intermediates, and then to the three-disulfide intermediate (Figure 4.21).

Further experiments should be conducted to answer whether only one pathway (two-disulfide intermediate to three-disulfide intermediate) or two pathways (either the two-disulfide intermediate (II) or the one-disulfide intermediate (I) to the three-disulfide intermediate (III)) exist. Unfortunately, these two pathways are unlikely to be distinguished by using our folding methodology (stop/go) since the one-disulfide intermediate (I) is the precursor of the two-disulfide intermediate (II). Therefore, we cannot determine whether the three-disulfide intermediate (III) is formed from the

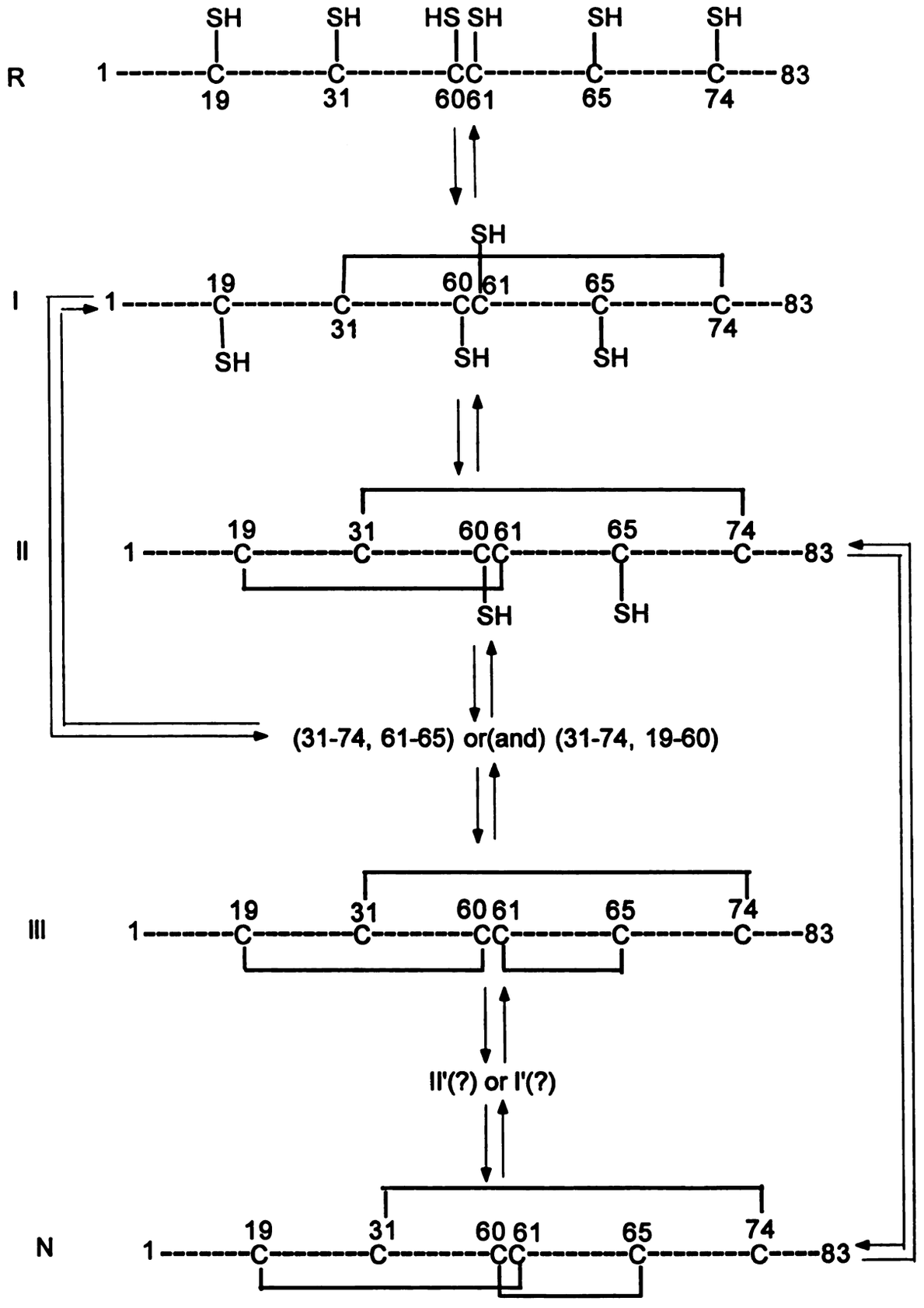


Figure 4.21. The predicted folding pathway of LR^{3}IGF-I

pathway of the two-disulfide intermediate (II) to the nonnative two-disulfide intermediate or from one-disulfide intermediate (I) to the nonnative two-disulfide intermediate. The technique of mutating Cys60 and Cys65 (such as mutate the cysteines to alanines) may be used to solve the problem because the two-disulfide intermediate (II) will not formed from the one-disulfide intermediate (I) after the mutation, and the only precursor of the three-disulfide intermediate (III) would be the one-disulfide intermediate (I).

The folded/native LR^3IGF-I can finally be formed by two pathways (Figure 4.21). One pathway between the three-disulfide intermediate (III) and native LR^3IGF-I can be reasonably suggested according to the time course of the refolding and unfolding of LR^3IGF-I. In this pathway, some transient two- or one-disulfide intermediates (I' and/or II") chould be involved through the breakage/reformation of disulfide bonds. It is also possible that an alternative pathway between the two-disulfide intermediate (II) and native LR^3IGF-I may exist based on two reasons. Firstly, the two-disulfide intermediate (II) already contains two native disulfide bonds, one can reasonably suggest that intermediate II could convert to native protein with the formation of the third disulfide bond, 61-65. Secondly, increasing the oxidized glutathione concentration in the equilibrium mixture of refolding species (previously established with lower GSSG concentration) resulted in all the intermediates (including intermediate II and intermediate III) being converted to native protein (Figure 4.3B) also suggests the existence of this pathway. The possible conversion of two-disulfide intermediate II to native protein will increase at higher concentration of GSSG so that the equilibrium will move to the formation of native protein from the pathway of the three-disulfide intermediate III to two-disulfide intermediate II, finally to native protein.

By comparison, the folding intermediates of IGF-I were also identified, which include a native one-disulfide intermediate (I), a native two-disulfide intermediate (II_A), two nonnative two-disulfide intermediates (II_B and II_C), and a nonnative three-disulfide intermediate (III). The disulfide structures of the native intermediates contain the similar linkage patterns in LR^3IGF-I and IGF-I. Two nonnative two-disulfide intermediates were identified in the folding process of IGF-I, but not in the refolding of LR^3IGF-I. The folding of IGF-I was much faster than that of LR^3IGF-I in the early folding stage.

A folding pathway of IGF-I can be suggested based on the time course of the refolding and the disulfide structure of the intermediates (Figure 4.22). The one-disulfide intermediate is formed from the reduced/unfolded IGF-I by the formation of the disulfide bond between 18 and 61. The two-disulfide intermediate II_A is formed from the one-disulfide intermediate by the formation of the second disulfide bond between 6 and 48. The two nonnative two-disulfide intermediates (II_B and II_C) can be formed from the one-disulfide intermediate by the formation of the disulfide bond between 6 and 47 (as in II_B) or between 6 and 52 (as in II_C). Alternatively, it is possible that the two nonnative two-disulfide intermediates could be formed from the two-disulfide intermediate (II_A) by breakage/reformation of disulfide bonds.

The nonnative three-disulfide intermediate could be formed from the nonnative two-disulfide intermediate (II_B) by the formation of the disulfide bond between 48 and 52. Finally, native IGF-I could be formed either from the native two-disulfide intermediate (II_A) by the formation of the third disulfide bond between 47 and 52 or from the nonnative three-disulfide intermediate (III) by the breakage/reformation of disulfide bonds.

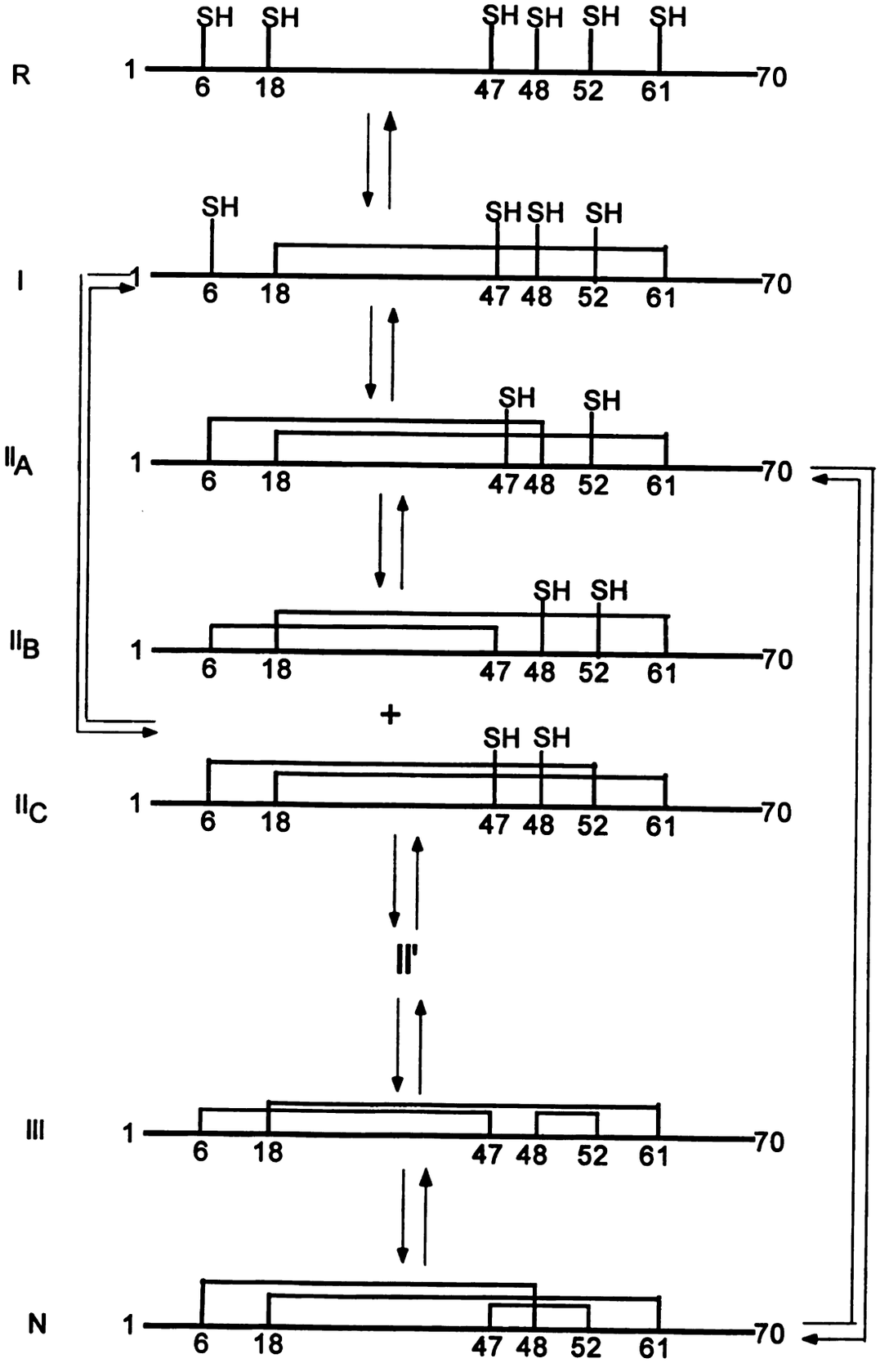


Figure 4.22. The predicted folding pathway of IGF-I.

It should be pointed out that the folding pathways of LR^3IGF-I and IGF-I described in Figure 4.21 and Figure 4.22 are predicted only based on the time dependent appearance and disulfide structure of the folding intermediates. Two problems must be solved before the actual pathway can be defined. Firstly, the well-populated species are not necessarily the productive species that account for the flow of intermediates. One way to identify the productive intermediates is to perform stop/go folding experiments of all species of the intermediates, both major and minor, that present at the same stage of equilibrium. The dilemma faced by this approach is that productive intermediates may exist as minor species. Finding all these minor species will be a daunting task. The predicament can be further complicated by the argument that an undetected species is not necessarily a non-existing species. Secondly, protein folding may occur through multiple pathways and these pathways may not be distinguishable. The other techniques such as mutation of the cysteine residues might solve the problem.

The trapping methodology, based on the cyanylation of free thiol groups under acidic conditions, showed similar results to those of acid trapping (31-33), but different from those of pyridylethylation trapping at alkaline condition (29) in terms of the distribution of intermediates during the folding processes of LR^3IGF-I and IGF-I. The methodology for the determination of disulfide structure by chemical cleavage and mass mapping of the fragments can be used to determine the disulfide linkages of the folding intermediates containing adjacent cysteines. The direct experimental evidences for the disulfide linkages of the nonnative three-disulfide intermediates of LR^3IGF-I and IGF-I were obtained for the first time by our methodology, which cannot be solved by the

conventional approach. Furthermore, the cyanylation methodology is much faster and sensitive.

VIII. References

1. Kim P. S., Baldwin R. L. (1990) *Annu. Rev. Biochem.* 59, 631-660.
2. Matthews C.R. (1993) *Annu. Rev. Biochem.* 62, 653-683.
3. Ptitsyn O.B. (1995) *Curr. Opin. Struct. Biol.* 5, 74-78.
4. Li Y.J., Rothwarf D.M., Scheraga H.A. (1995) *Nature Struc. Biol.* 2, 489-494.
5. Creighton, T.E. (1974). *J. Mol. Biol.* 87, 563-577.
6. Creighton, T.E., Goldenberg, D.P. (1984) *J. Mol. Biol.* 179, 497-526.
7. Scheraga, H.A., Konishi, Y., Ooi, T. (1984) *Adv. Biophys.* 18, 21-41.
8. Weissman, J.S., Kim, P.S. (1991) *Science* 253, 1386-1393.
9. Creighton, T.E., Ewbank J.J. (1994) *Biochemistry* 33, 1534-1538.
10. Chang, J.Y. (1996) *Biochemistry* 35, 11702-11709.
11. Glocker M.O., Arbogast B., Milley R., Cowgill C., Deinzer M.L. (1994) *Proc. Natl. Acad. Sci. USA* 91, 5868-5872.
12. Darby, N.J., Norin P.E., Talbo, G., Creighton, T.E. (1995) *J. Mol. Biol.* 249, 463-477.
13. Weissman, J.S., Kim, P.S. (1992) *Proc. Natl. Acad. Sci. USA* 89, 9900-9904.
14. Xu, X., Rothwarf, D.M., Scheraga, H.A. (1996) *Biochemistry* 35, 6406-6417.
15. Rothwarf, D.M., Scheraga, H.A. (1993a) *Biochemistry* 32, 2671-2679.
16. Rothwarf, D.M., Scheraga, H.A. (1993b) *Biochemistry* 32, 2680-2689.
17. Rothwarf, D.M., Scheraga, H.A. (1993c) *Biochemistry* 32, 2690-2697.
18. Rothwarf, D.M., Scheraga, H.A. (1993d) *Biochemistry* 32, 2698-2703.
19. Rothwarf, D. M., Scheraga, H. A. (1991) *J. Am. Chem. Soc.* 113, 6293-6294.
20. Creighton, T.E. (1992) *Bioessays* 14, 195.

21. Pain, R.H. (1994). In *mechanisms of protein folding. Oxford University Press*, New York, p112.

22. Wu, J., Yang, Y., Watson, J.T. (1998) *Accepted by Protein Science.*

23. Smith, D.L., Zhou, Z.R. (1990) *Methods Enzymol.* 193, 374-389.

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

25. Yang, Y., Wu, J., Watson, J.T. (1998) *Accepted by J. Am. Chem. Soc..*

26. Rinderknetcht, E., and Humbel, R. E., *J. Biol. Chem.*, 253, 2769 (1978).

27. Froesch, E. R., Burgi, H., Ramseier, E. B., Bally, P., and Labhart, A., *J. Clin. Invest.*, 42, 1816 (1963).

28. Rinderknecht, E., and Humbel, R. E., (1976) *Proc. Natl. Acad. Sci., USA*, 73, 2365.

29. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992) *Biochemistry*, 31, 1749-1756.

30. Hober, S., Forsberg, G., Palm, G., Hartmanis, M., and Nilsson, B. (1992) *Biochemistry*, 31, 1749-1756.

31. Milner, S.J., Francis, G.L., Wallace, J.C., Magee, B.A., Ballard, F.J. (1995) *Biochem. J.* 308, 865-871.

32. Hua, Q.X., Narhi, L., Jia W., Arakawa, T., Rosenfeld, R., Hawkins, N, Miller, J.A., Weiss, M. (1996) *J. Mol. Biol.* 259, 297-313.

33. Miller, J.A., Narhi, L.O., Hua, Q.X, Rosenfeld, R., Arakawa, T., Rohde, S.P., Lauren, S., Stoney, K.S., Tsai, L., Weiss, M.A. (1993) *Biochemistry* 32, 5203-5213.

34. Tomas, F.M., Knowles, S.E., Chandler, C.S., Francis, G.L., Owens, P.C., Ballard, F.J. (1995) *J. Endocrinology* 137, 413-421.