
133
445
THS

THESIS

MICHIGAN STATE LIBRARIES
3 1293 01766 5112

This is to certify that the

thesis entitled

A Study of Cyanylation Reagents for the Assignment of Cysteine Residues and Disulfide Bond Pairings in Proteins by MALDI-TOF Mass Spectrometry

presented by

Qing Li

has been accepted towards fulfillment
of the requirements for

Master degree in Chemistry

Major professor

Date 10 July 1998

O-7639

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY
Michigan State
University

PLACE IN RETURN BOX
to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
FEB 2 0 2000		

1/98 c/CIRC/DateDue.p65-p.14

A STUDY OF CYANYLATION REAGENTS FOR THE ASSIGNMENT OF CYSTEINE RESIDUES AND DISULFIDE BOND PAIRINGS IN PROTEINS BY MALDI-TOF MASS SPECTROMETRY

By

Qing Li

A THESIS

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

MASTER OF SCIENCE

Department of Chemistry

1998

ABSTRACT

A STUDY OF CYANYLATION REAGENTS FOR THE ASSIGNMENT OF CYSTEINE RESIDUES AND DISULFIDE BOND PAIRINGS IN PROTEINS BY MALDI-TOF MASS SPECTROMETRY

By

Qing Li

A novel methodology to locate cysteines and disulfide bond pairings in peptides and proteins has been developed by the Watson group. This new approach employs a specific reaction between free sulfhydryls and a cyanylation reagent such as 2-nitro-5-thiocyanobenzoic acid or 1-cyano-4-dimethylamino-pyridium tetrafluoroborate to cyanylate free sulfhydryl groups. The N-terminal peptide bond of the cyanylated cysteine residue can be cleaved under alkaline conditions to form an amino-terminal peptide and a series of 2-iminothiazolidine-4-carboxyl (*ITC*) peptides which can be mapped to the sequence by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry.

This thesis focuses on two new cyanylation reagents: 2-thiocyanopyridine (PYRSCN) and 4-(dimethylamino) phenyl thiocyanate. These two reagents were found to be similar in reactivity to NTCB when reacting with free sulfhydryl groups; a comparison is also made with CDAP. The two new reagents tend to form both the cyanylated product and a mixed disulfide product. Structural features of the cyanylation reagent and the pH of the reaction medium were studied using model peptides. Based on a plausible reaction pathway for the cyanylation process, a new series of reagents is proposed which may avoid the formation of the mixed disulfide during the reaction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

Chapter I

Introduction

I. Introduction

Since its discovery in 1912, mass spectrometry (MS) has provided key insight for broad and diverse disciplines. However, many compounds of biological interest are difficult to volatilize directly and are usually derivatized in order to increase volatility for ion production by electron ionization in the vapor phase. This frequently leads to thermal degradation at the same time to produce a complicated mass spectrum which is difficult to interpret (1). Advances in chemical technology have been the engine powering the biotechnology industry. Analytical chemists have added fresh impetus to bioresearch with two new mass spectrometry ionization tools: matrix-assisted laser desorption/ionization (MALDI) (2) and electrospray (ESI) (3). Commercial availability of these instruments has made routine the analysis of high-mass compounds including proteins, peptides, carbohydrates, oligonucleotides, natural products, and drug metabolites with picomole to femtomole sensitivity (4). Today, MS has been revolutionary in spurring research in protein biochemistry, glycobiology, and biotechnology, and most recently in DNA sequencing (5).

Protein analytical techniques are essential for determining a protein's primary structure as well as detecting and defining posttranslational modifications of proteins so that the complete structure of a protein can be obtained (6, 7). Nucleotide sequences of genes embody the information required to deduce the primary structure (amino acid sequence) of proteins, but they do not reveal whether the side chains of amino acids are

chemically modified *in vivo* after translation of genes or the extent of any modifications. Defining posttranslational modifications is essential in the case of expressed proteins that are defined for use as pharmaceuticals. This is required to ensure that they have the same chemical structure as their natural counterparts or an acceptable alternate structure. Unacceptable structures could be detrimental to recipients due to unwanted pharmaceutical or immunological activities.

Defining the locations and structures of posttranslational modifications in proteins is one of the most important contributions that mass spectrometry has made and continues to make in the field of protein biochemistry (8). The unique potential of MS is most evident when it is used in conjunction with conventional methodologies like automated protein sequence and amino acid analysis. In contrast to the latter two approaches in which identification is based entirely on chromatographic retention relative to one of the 20 commonly occurring amino acids or a derivative thereof, MS relies on measurement of molecular mass, an intrinsic physical property. Mass spectrometry is clearly the method of choice for characterization of posttranslationally modified proteins, since in most cases well-developed, accurate, and sensitive chemical or biochemical approaches are not available.

Disulfide bond formation after gene expression is one of the most important posttranslational modifications frequently encountered in protein characterization (9, 10). Mass spectrometry, combined with other chemical/biochemical techniques, has been extensively applied to the recognition of both free sulfhydryls and disulfide bonds in proteins.

In this thesis, novel methodologies for recognizing free sulfhydryl groups and disulfide bonds will be discussed in Chapter 2, and a study of reagents for the cyanylation reaction will be described in detail in Chapter 3. The objective of this chapter is to introduce basic aspects of MALDI and current methodologies for characterizing cysteine sulfhydryl status in proteins.

II. Characterization of Cysteine and Disulfide Bond Status in Proteins

Cysteine contributes to protein biological functions by using its free sulfhydryl group (-SH) as the active site for enzyme catalysis (11). Sulfhydryl groups can also play an important role in a variety of physiological and biochemical processes (11). Disulfide bonds, on the other hand, comprise the major covalent cross-linkage in proteins and maybe *intra-* or *inter*-molecular in nature. *Intra*-chain disulfide bonds serve to confer conformational stability on the folded polypeptide chain. While *inter*-chain disulfide bonds are functional in maintaining the quaternary structure of multi-chain proteins, serving as the only linkage between subunits or providing covalent stability to structures otherwise maintained by non-covalent forces. Both free cysteines and disulfides, therefore, play a unique role in the activity of proteins (9).

A Localization of Sulfhydryl Groups

1. Classical Approach

The classical approach for locating free sulfhydryl groups in proteins involves several steps. The first step is to modify free sulfhydryl groups, usually by an irreversible

reaction such as alkylation, under conditions that can prevent or minimize sulfhydryl/disulfide exchange. Second, the protein is cleaved by enzymes or chemical reagents between cysteine residues. Third, the digestion mixture is separated by chromatography. Finally, the derivatized sulfhydryl-containing peptides are identified, mapped to sequence by the Edman technique and/or mass spectrometry, and related to specific segments of the protein.

A key to the success of this approach is to choose an appropriate derivatization reaction and 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 bond 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 it should easily attach a radioactive element to facilitate the identification 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.

There are many reagents that had been used in the modification of cysteine groups. Alkylation of the free sulfhydryl group is one of the most traditional ways. This is usually achieved by reaction with compounds containing an alkyl halide group such as iodoacetate, iodoacetamide, bromoacetate, etc., among which iodoacetamide shows the highest reactivity and specificity to SH, alkylation can be accomplished by converting the SH to S-carboxymethylcysteine (Figure 1.1) (11-13). The iodide liberated can be

determined by an electrochemical method (14) or by absorbance at 226nm (15). To avoid the formation of iodine by photooxidation of iodide ions, the carboxymethylation is performed in the dark.

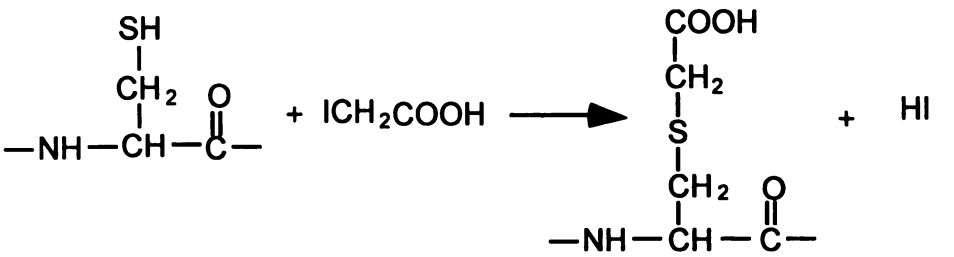


Figure 1.1 The modification of cysteine with iodoacetic acid to form S-carboxymethylcysteine

This commonly used cysteine alkylation method does not yield material which is easily amenable to Edman sequence analysis. This is either because of solubility problems with the alkylated proteins and peptides or because the 3-phenyl-2-thiohydantoin (PTH) derivatives (formed by Edman degradation) of the alkylated cysteine residues cannot be readily identified by standard procedures such as gas-liquid chrmatography. Because of these reasons, some other alkylation reagents, such as 4-vinyl pyridine (4-VP), which yields the pyridylethylcysteine (PECys) derivative were developed. Alkylation with 4-VP provides several advantages. For instance, the PECys residues are stable to hydrolysis in 6N HCl, and they increase the solubility of peptides at neutral and lower pH, etc. (16). Although this provides another choice of alkylation reagent, iodoacetate is still the most popular derivatization reagent for modifying sulfhydryl groups because of its simplicity and specificity (17-18).

Two derivatives of 7-fluoro-2,1,3-benzoxadazole (19), 7-fluoro-2,1,2-benzoxadiazole 4-sulfonate (SBD-F) and 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD-F), have proven useful for the assay of sulfhydryls by reason of their formation of fluorescent products (Figure 1.2).

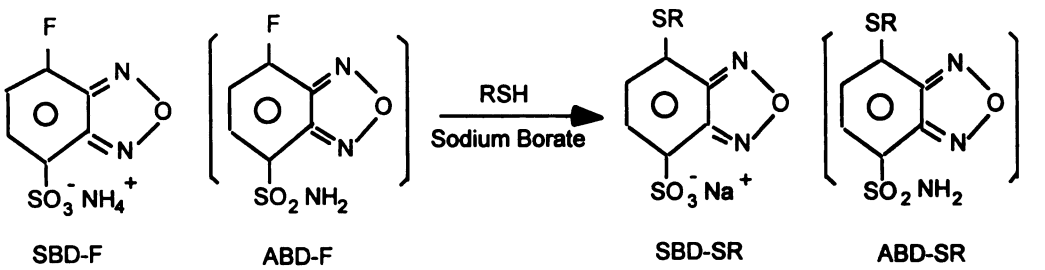


Figure 1.2 Derivatization of the cysteine group by SBD-F or ABD-F

The adducts exhibit fluorescence at a long wavelength (~515nm) and show a spectral shift corresponding to environmental hydrophobicity (20, 21). The reaction is performed under borate buffer, pH 9.5, containing 1mM EDTA to prevent metal-catalyzed oxidation of the sulfhydryls. For the derivatization of sulfhydryls in large proteins, the proteins must be denatured to ensure that every SH can be attacked by the reagents (22, 23).

Organomercurial compounds are high affinity and specific reagents for SH groups. In contrast to alkylation, the mercaptidation reaction is reversible. Among various organomercurial reagents, monofunctional organomercurial compounds have been widely exploited. Spectrophotometric titration with *p*-mercuribenzoate (Figure 1.3) proposed by Boyer is the most preferred because of its simplicity, high sensitivity, selectivity and precision (24).

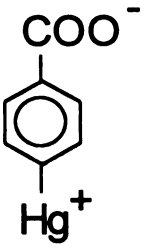


Figure 1.3 *p*-mercuribenzoate

Another method to modify free SH groups is based on sulfhydryl/disulfide exchange. Among oxidants of SH groups, disulfides occupies a special position since their reactions with sulfhydryls are absolutely specific. This is referred to as sulfhydryl/disulfide exchange. Among the various disulfides proposed, Ellman's reagent 5,5'-dithio-*bis*-(2-nitrobenzoic acid), DTNB (Figure 1.4), has received the widest use (25). Grassetti and Murray (26, 27) demonstrated that 2,2'-dithiodipyridine or 4,4'-dithiodipyridine were good substituents for Ellman's reagent. These compounds showed strong reactivity to SH groups in a wide pH range (pH 1-8) (28). The particular value of pyridyl disulfides as reactivity probes and labelling reagents is that they are two-protonic-state electrophiles, i.e., they markedly increased their electrophilicity when protonated on a pyridyl nitrogen atom.

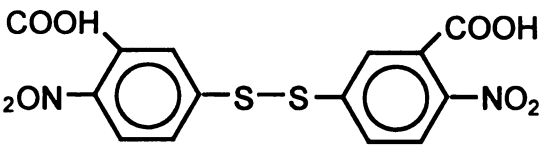


Figure 1.4 Ellman's Reagent: 5,5'-dithio-bis-(2-nitrobenzoic acid)

The other method used in the modification of –SH groups is oxidation. Sun and Smith used performic acid to modify certain residues and thereby cause a characteristic change in the peptide molecular weight. This change in molecular weight is determined by FABMS (fast-atom-bombardment mass spectrometry) and used to help identify peptides. Results for a series of small peptides demonstrate that Cys, Met, and Trp are the only residues that undergo a change in molecular weight under the conditions they used. The molecular weight of a peptide increases by 48 Da for each cysteinyl-containing peptide (29). When treating with perfomic acid, destruction of tryptophan and partial destruction of tyrosine will occur, thus the practical application of this approach is limited.

Chin and Wold (30) employed the combination of tributylphosphine (Bu_3P) to reduce disulfide bonds and of 4-(aminosulfonyl)-7-fluoro-2,1,3-benzoxadiazole (ABD-F) to block free sulfhydryls to characterize both free sulfhydryls and disulfide bonds in a number of proteins. Since the two reagents apparently do not react with each other, their combination offers a convenient and quite general method for the complete characterization of free Cys (-SH) and cross-linked Cys (S-S) in proteins. In their paper, they demonstrated their strategy on model compounds (e.g., Cys and glutathione), proteins such as enolase from yeast and rabbit muscle, containing only free Cys, and a protein, superoxide dismutase, containing both free and crosslinked Cys. Using a similar procedure, Kirley (31, 32) determined the location of three disulfide bonds and one free sulfhydryl group in the β-subunit of (Na^+, K^+)-ATPase.

Another reagent, 5-[2-((iodoacetyl)amino)-ethyl]naphthalene-1-sulfonic acid (IAEDANS) (Figure 1.5) has been proven powerful for sulfhydryl derivatization (33-35).

Recently, Sturrock (36) *et al* used this reagent to label a free cysteine residue in human testicular angiotensin-converting enzyme (tACE). After isolating the fluorescent peptide from enzymatic digests by HPLC, the sequence of the fluorescent peptides was mapped by MALDI-PSD (matrix assisted laser desorption/ionization - post source decay). The sequence data established that Cys496 is a free thiol in tACE.

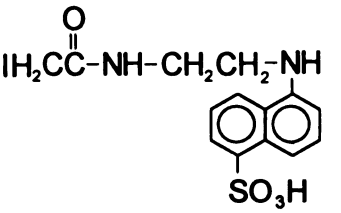
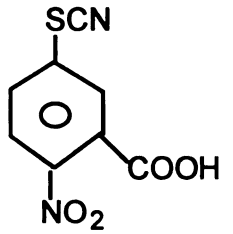


Figure 1.5 5-[2-((iodoacetyl)amino)-ethyl] napthalene-1-sulfonic acid (IAEDANS)

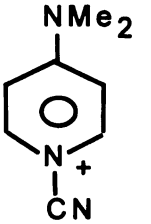
2. Cleavage at Cysteine Residues

This method utilizes a cleavage reaction at cysteine residues. The conversion of an SH group into an SCN group was first achieved in two stages by successive treatment of a protein with Ellman's reagent and then with cyanide (37, 38). Catsimpoolas (37) showed that the cyanide cleavage of the disulfide bond yields a sulfhydryl and a thiocyano group. In pH ranges below 8, the latter cyclizes to produce an acyliminothiazolidine moiety on the peptide chain. This product is unstable and hydrolyzes spontaneously to effect cleavage of the peptide link on the amino group. Stark (39) showed that 2-nitro-5-thiocyanobenzoic acid (NTCB) (Figure 1.6) specifically cyanylates cysteine sulfhydryls, which subsequently cleave at the N-terminal side of the cyanylated cysteinyl residues under mildly alkaline conditions to form an amino-terminal peptide and series of 2-iminothiazolidine-4-carboxyl (ITC) peptides. If a protein contains

n cysteine residues, the cleavage reaction results in the formation of **n+1** peptide fragments, mass assignment of which indicates the number and location of cysteine residues. While potentially a very useful method, it has seldom been used for sequence determination because all but the N-terminal peptide is blocked by the ITC group, which is not amenable to Edman degradation (40, 41) . Papayannopoulos and Biemann (42) first used CID tandem mass spectrometry to sequence the NTCB cleavage reaction



NTCB



CDAP

Figure 1.6 Structures of 2-nitro-5-thiocyanobenzoic acid (NTCB) and 1-cyano-4-dimethylamino-pyridium salt (CDAP)

products of a protease inhibitor isolated from *Sarcophaga bullata*. Mass spectrometry provides much more accuracy than SDS page for mass assignment and can be applied to the sequencing of ITC blocked peptides. To date, the NTCB approach has primarily been employed to locate total cysteines in various proteins although selective identification of free cysteine in the presence of disulfide bonds was also reported (42-47).

Another cyanylation reagent that had been used is 1-cyano-4-dimethylamino-pyridinium salt (CDAP) (Figure 1.6). Reported by Wakselman (48) this reagent was

synthesized from 4-dimethylaminopyridine. CDAP was found to be reactive toward protein sulfhydryl groups in neutral or acidic medium, then the cyanylated product can be cleaved under alkaline conditions (37). It was also reported that this reaction is highly specific; papain when reacting with CDAP at pH 3.6 for 11 minutes, loses more than 98% of its catalytic activity. The cyanylated product is isolable and this reaction has no oxidative side reactions.

In this thesis, more cyanylation reagents will be investigated, and the chemistry and mechanism of the cyanylation reactions will be discussed in detail in Chapter 3.

B. Localization of Disulfide Bonds

Although the amino acid sequences of proteins are readily deduced from the corresponding cDNA, modifications occuring to the protein after translation cannot be predicted accurately. Although there are good methods for quantifying the number of disulfide bonds in proteins, the unambiguous determination of the locations of disulfide bonds remains a difficult task.

Two most often used terms in cysteine chemistry are sulfhydryl/disulfide exchange and disulfide bond scrambling (or disulfide bond interchange). Although sulfhydryl/disulfide exchange (such as Ellman's reagent) is the most specific reaction, both sulfhydryl/disulfide exchange and disulfide bond scrambling in proteins occur in mildly alkaline media when structures permit and result in mismatched nonnative disulfide bond structures (Figure 1.7). While under acidic conditions from pH 2 to pH 6.5, this sulfhydryl/disulfide exchange and disulfide bond scrambling reactions can be minimized (49, 50).

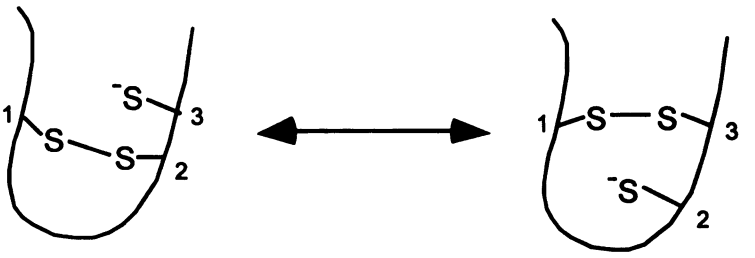


Figure 1.7 Illustration of sulfhydryl/disulfide exchange in proteins

Classical Strategy

As illustrated in Figure 1.8, a classical approach (51, 52) in locating the disulfide bond linkage includes three steps. First, a protein is cleaved by enzymes or chemical reagents between half-cystinyl residues under conditions that can avoid or minimize disulfide scrambling. Second, the digestion mixture is separated by reversed-phase HPLC. Third, the amino acid sequence or molecular masses of these peptides are determined using Edman degradation or mass spectrometry. Finally, the disulfide linkages are deduced, but this method is usually time-consuming. The other traditional method is enzymatic digestion. The method usually requires an enzyme to be specific so that it can produce well-defined fragments that can simplify the identification of the disulfide bond. It is very likely that some proteins cannot be cleaved by specific cleavage reagents; thus, they must be cleaved by non-specific reagents. However, when using nonspecific reagents, it is difficult to relate the disulfide-bridged peptides to specific segments of the protein simply by determining their molecular masses by MS.

Recently, more methods have been developed for the identification of disulfide-containing peptides, especially various mass spectrometric methods. A new methodology developed in Watson's research group will be discussed in detail in Chapter 2.

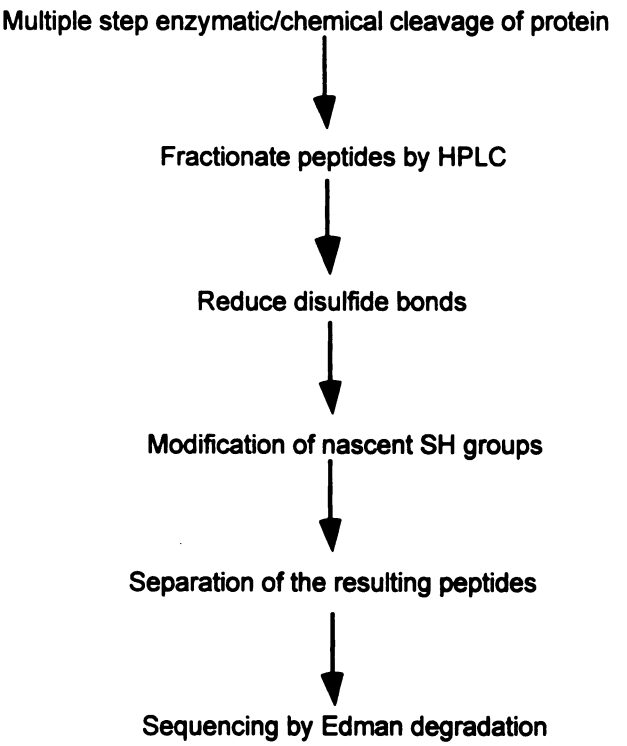


Figure 1.8 Classical strategy for disulfide bond assignment of a protein

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

Lasers originally found applications in mass spectrometry in the early 1970s for desorption/ionization of small organic compounds that contained appropriate chromophores (53). Abrupt energy absorption initiates a phase change in a microvolume of the absorbing sample from a solid to a gas while also inducing ionization. In the late 1980s, research into the desorption/ionization of nonvolatile materials for mass spectrometry has fostered the development of matrix-assisted laser desorption ionization, MALDI (54-56). Although the first attempt to generate ions of organic molecules by direct laser desorption/ionization dates back to early 1970s (57). However, the size of the analytes which can be desorbed and ionized was limited to ~1000 daltons. When larger molecules are analyzed, more energy is required to desorb them. However, such a high energy flux leads to pyrolytic or photochemical decomposition of analytes. The breakthrough came in the late 1980s when Hillenkamp and Karas successfully experimented with the use of a matrix, nicotinic acid (58, 59). It was discovered that desorption/ionization of large, nonvolatile molecules such as proteins could be similarly obtained when such samples were irradiated by a laser after being codeposited with a large molar excess of an energy-absorbing "matirx" material, even though the analyte did not strongly absorb the wavelength of the laser radiation. Initial experiments with this technique were conducted with frequency-quadrupled Nd-YAG lasers that emit at 266nm. The ejected ions were analyzed by a time-of-flight (TOF) mass spectrometer. After that MALDI has quickly gained notoriety as an analytical tool that can accurately determined the molecular weight of proteins up to (and exceeding) 100kDa from low-

picomole amounts of sample (60). Despite the lack of an established ionization mechanism, the dependence on the matrix material in the desorption/ionization process gave rise to the terminology matrix-assisted laser desorption/ionization (MALDI). Today MALDI has developed into an extremely powerful technique for accurate and sensitive analysis of molecular masses extending to greater than 200,000 Da. With the introduction of two new techniques, post-source decay (PSD) and delayed extraction (DE), MALDI-TOF MS is also becoming an effective tool for structure elucidation of biopolymers.

All the work that will be discussed later was performed on the MALDI instrument (PerSeptive Biosystem Voyager Elite) available in MSU/NIH Mass Spectrometry Facility. This instrument can be operated in the reflectron mode to acquire PSD spectra for structural analysis, and it also has the delayed extraction function which improves its resolution.

A. The Matrix

For the most commonly employed laser wavelengths (337 or 355nm), several hundred different organic compounds, selected for their light absorbing properties, have been investigated for use as MALDI matrices (61, 62). However, only a few of these are widely applicable, e.g., sinapinic acid (SA), 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnamic acid (α-CHCA) (63, 64). This is largely because features other than absorption of the appropriate wavelength of laser radiation are necessary for a compound to be a useful MALDI matrix. Other important matrix characteristics include formation of a single phase with the analyte in the solid phase, solubility in the same

solvents required for dissolution of the analyte, vacuum compatibility (low vapor pressure), a chemical composition that promotes ionization (matrix substituents that can donate protons to the analyte), nonreactivity with the analyte, and other physical properties such as low heat of sublimation and a capacity to crystallize readily (65). Some of these criteria may explain the variable success in analyzing different classes of macromolecules (i.e., small peptides, proteins, glycoproteins, polysaccharides, or oligonucleotides) by MALDI using different matrices.

The physicochemical events leading to the transfer of proteins to the gas phase and their ionization in MALDI have not yet been fully elucidated (66). The matrix is believed to serve to minimize sample fragmentation from the laser beam by absorbing the incident laser energy, resulting in the sample and matrix molecules being ejected into the gas phase. One model for the mechanism assumes that the uppermost layers of matrix are induced to undergo a phase transition from the solid to the gas phase. The subsequent expansion of these matrix molecules into the vacuum drags the matrix-isolated protein molecules into the gas phase. During the transfer to the gas phase, the protein undergoes ionization through proton transfer reactions with the matrix by a reaction process that remains to be explained.

B. Sample Preparation

Proper sample preparation is critical for successful analysis by MALDI-MS (60). Matrix solutions are prepared in water, water-acetonitrile, or water-alcohol mixtures at a concentration of about 5-10ug/ul, depending on the solubility properties of the matrix. The analyte is prepared at a concentration of about 0.1ug/ul in a solvent that is miscible

with the matrix solution (for peptides or proteins, aqueous 0.1% trifluoroacetic acid (TFA) is frequently used). The matrix and analyte solutions are mixed to give a final matrix : analyte molar ratio of approximately 10,000:1 and a final volume of 0.5-2ul. The mixture is applied to a stainless steel MALDI-MS probe tip, and allowed to dry by either ambient evaporation, heating with a stream of warm air, or under vacuum. During the drying process, the matrix codeposits from solution with the analyte. The nature of the matrix-analyte interaction has not yet been defined, but it may involve analyte molecules that have become embedded in the matrix crystal lattice or on the surface of rapidly forming matrix crystals. Studies based on optical microscopy have shown that matrix crystal formation and size vary, depending on the matrix and solvent, but seem to be independent of the analyte. The decrease in the "quality" of matrix crystal morphology (where the matrix-analyte deposit appears as a powder, film, or noncrystalline glassy surface) leads to a lower quality MALDI mass spectrum. These conditions can result from the presence of high levels of contaminating salts, buffers, or surfactants.

Several factors affect the sensitivity of a MALDI-MS experiment. In many cases, analysis of smaller amounts of sample yields a better MALDI signal. This may simply be the result of approaching a more optimum molar ratio between the matrix and the analyte, especially with macromolecules whose masses exceeds 10kDa. Alternatively, dilution of a contaminant which can interfere with analyte signal acquisition may also lead to improved spectra.

In this thesis, the experiment is performed on the Voyager Elite system which uses a plate holder that can accommodate 100 samples and permit a much higher sample

throughput. During the drying process, the matrix codeposits from solution with the analytes. The sample is inserted into the vacuum chamber by a probe and a vacuum lock in the VT2000, and by an "automatic mechanical hand" in the Voyager Elite MALDI-TOF mass spectrometer, and irradiated with short duration(1 to 10ns) pulses of an UV laser beam.

To date, the chemistry of "matrix-assistance" remains incompletely understood; the choice of matrix and method of application is still empirical. Sample preparation by different matrices and solvents (69), matrix additives (70), and evaporation rate (71,72) affects the resolution and sensitivity of MALDI. Optimal results require parallel analyses under different conditions.

C. Instrumentation

Time-of-flight (TOF) mass spectrometers (60) are instruments noted for their nearly unlimited mass range and high sensitivity. In these mass analyzers, ions are formed by a pulsed ionization process (pulsed laser radiation for MALDI) in a short (~5cm) source region containing an electric field as shown in Figure 1.9. The electric field accelerates the ions into a long (~1m) field-free drift region. The time required for ions to traverse the drift region (TOF) is dependent on their mass and is described by the relationship: $TOF = L/v = L(m/2zeV)^{1/2}$, where L is the length of the flight tube, v is the ion velocity, m is the mass of the ion, z is the charge state of the ion, and V is the acceleration potential. Thus, low-mass ions have a shorter flight time than heavier ions. The clock used to measure the TOF of the MALDI ions is triggered by the laser pulse. All the ions that are formed from a single laser pulse and that are accelerated by the electric field give rise to a

transient TOF signal from the detector at the end of the flight tube. To improve ion counting statistics (60), multiple transient TOF mass spectra can be summed from 10 to several hundred laser pulses. Time-to-mass conversion is achieved by incorporating the tof of ions of known mass (from internal or external standards) into a calibration algorithm from which mass values can be computed for unknown ions with experimentally determined flight times.

Additional virtues of the time-of-flight instrument are simplicity and low cost. However, because 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 mass resolution is disappointingly low (73). The poor resolving power (defined as the capacity to distinguish ions differing slightly in mass) is a result of peak broadening which increases with the mass of the ionized macromolecule. This limits the capacity to detect certain protein modifications and/or protein sequence variations, especially at higher masses. In spite of this, it is possible to determine the mass of MALDI-TOF ions with an accuracy up to +/- 0.01% for proteins with molecular masses between 1 and 40 kDa, and with somewhat poorer accuracy for proteins above 40 kDa (60).

D. Novel Techniques in MALDI-TOF MS

MALDI has been a superior method for molecular mass determination of proteins and peptides, and other biological materials, but lacks the capacity for structural analysis, because it is a soft ionization technique. This deficiency has been overcome since the introduction of post-source decay (PSD) analysis of MALDI-generated ions. The PSD

(74,75) technique is based on mass analysis of product ions from unimolecular or collision-induced decay taking place in the field-free region between the ion source and the reflectron. In contrast to a linear instrument, reflection instruments can be used as an as an energy analyzer to differentiate ions that are otherwise detected as species with the same velocity in the linear mode. Ions formed as a result of metastable decomposition of fully accelerated precursors are detected at the same arrival time (same apparent mass) as their precursor in the linear mode of TOF because they have the same velocity. However, because the product ions have lower kinetic energy, they can often be resolved in the reflectron mode of the TOF instrument by lowering the potential of the reflector. Therefore, PSD, like MS/MS techniques, can provide full or partial sequence information of medium sized peptides (<2500 Da). Since its commercialization about seven years ago, MALDI-PSD has quickly evolved into a powerful technique for sequence determination of peptide at the low picomole scale.

E. Applications of MALDI-TOF MS

Over the past few years, MALDI has become among the most powerful methods yet available for macromolecular characterization of living systems. A wide range of achievements in proteins, nucleotides, and glycobiology have been attained in a short time.

MALDI has been extensively used for the determination of primary covalent structures of proteins derived from both natural and recombinant sources (76). MALDI

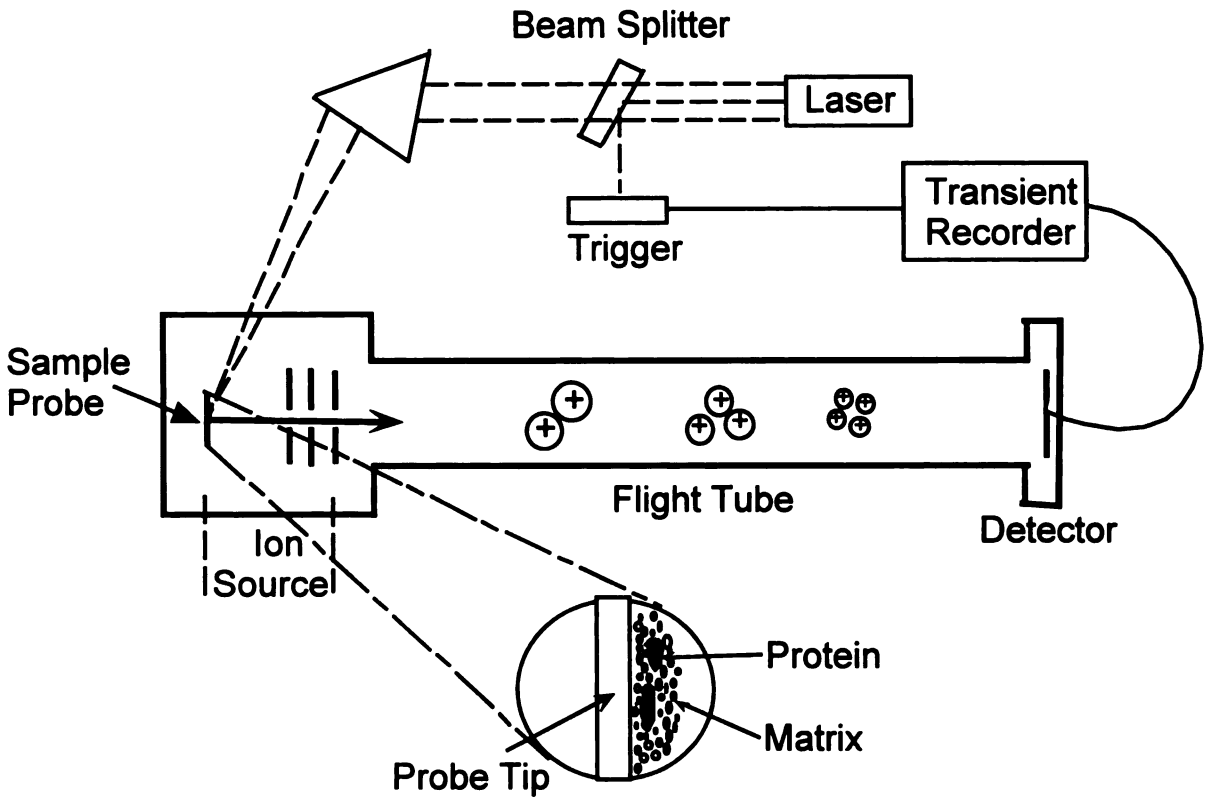


Figure 1.9 Scheme of matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF MS)

provides routine and reliable means to analyzing tryptic digests and glycoproteins which typically have a high degree of heterogeneity (77). The high tolerance of MALDI to extraneous compounds has been used to study the degradation of hemoglobin by a malaria parasite. This remarkable achievement shows direct analysis of cell contents is now within the reach of mass spectrometry (78).

MALDI-MS is an excellent method when dealing with protein posttranslational modifications. Covalently modified protein N- and C- termini (79), disulfide bonds (80, 81), phosphorylation (82), glycosylation (83), lipidation (84) and protein DNA interactions (85) have all been studied by MALDI.

In addition to the use of MALDI as a sequencing tool, its ability toward the analysis of combinatorial libraries has also been explored.

The capacity for mass spectrometric analysis of oligonucleotides has generally lagged behind that of oligopeptides. However, this position has changed markedly in the last few years (86, 87), especially with the development of new matrices. Investigation of a 50-mer is now somewhat routine (88, 89) and there are individual reports of the analysis of much larger RNAs (90).

It is believed that MALDI-MS will play more and more important roles in biological studies. As is often true in a new and rapidly developing field, the most exciting breakthroughs may, in the end, occur in areas not even anticipated at the present time.

III. References:

1. Thorgersen, D. F., Skowronski, R. P., and Macfarlane, R. D., *Biochem. Biophys. Res. Commun.*, 60, 616 (1974)
2. Karas, M., and Hillenkamp, F., *Anal. Chem.*, 60, 2299-2301 (1988)
3. Meng, C. K., Mann, M., and Fenn, J. B., Proc. 36th ASMS Conference on Mass Spectrometry and Allied Topics, San Francisco, CA, (1988), p771.
4. Siuzdak, G., *Proc. Natl. Acd. Sci.*, 91, 11290 (1994)
5. Burlingame, A. L., Boyd, R. K., and Gaskell, S. J., *Anal. Chem.*, 68, 599R-651R (1996)
6. Carr, S. A., Roberts, G. D., and Hemling, M. E., in *Mass Spectrometry of Biological Materials*, C. N. McEwen and B. S. Larsen (Editors), Marcel Dekker, New York, (1990), Ch. 3, p87.
7. Yan, S. C. B., Grinnell, B. W., and Wold, F., *Trends Biochem. Sci.*, 14, 264 (1989).
8. Carr, S. A., and Biemann, K., Methods Enzymol., 106, 29-58 (1984)
9. Creighton, T. E., in *Mechanism of Protein Folding*, Edited by Pain, R. H., IRL Press, (1992)
10. Hirayama, K., and Akashi, S., in *Biological Mass Spectrometry: Present and Future*, Edited by Matsuo, T., Caprioli, R. M., Gross, M. L., and Seyama, Y., John Wiley & Son Ltd, (1994)
11. Torchinsky, Y. M., in *Sulfur in Protein*, (Metzler, D., Translation Ed.), Pergamon, Oxford, (1981)
12. Jocelyn, P. C., *Biochemistry of the SH Group*, Academic Press, London, (1974)

13. Jocelyn, P. C., in Methods in Enzymol., Vol. 143, Academic Press, edited by Jakoby, W. B., Griffith, O. W. Inc., (1987)

14. Stark, G. R., and Stein, W. H., *J. Biol. Chem.*, 239, 3755 (1964)

15. Swoboda, G., and Hasselbach, W., *Hoppe-Seyler's Z. Physiol. Chem.*, 354, 1611(1973)

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

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

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

19. Toyo'oka, T., and Imai, K., *Anal. Chem.*, 56, 2461-2464 (1984)

20. Toyo'oka, T., and Imai, K., *Analyst*, 109, 1003 (1984)

21. Sueyoshi, T., Miyata, T., Iwanaga, S., Toyo'oka, T., and Imai, K., *J. Biochem. (Tokyo)*, 97, 1811 (1985)

22. Toyo'oka, T., and Imai, K., *Anal. Chem.*, 57, 1931 (1985)

23. Kobayashi, T., Kudo, I., Karasawa, K., Mizushima, H., Inoue, K., and Nojima, S., *J. Biochem. (Tokyo)*, 98, 1017 (1985)

24. Boyer, P. D., *J. Am. Chem. Soc.*, 76, 4331 (1954)

25. Ellman, G. L., *Arch. Biochem. Biophys.*, 82, 70 (1959)

26. Grassetti, D. R., and Murray, J. F., *Arch. Biochem. Biophys.*, 119, 41 (1967)

27. Grassetti, D. R., and Murray, J. F., *J. Chromatogr.* 41, 121 (1969)

28. Stuchbury, T., Shipton, M., Norris, R., Malthouse, J. P. G., Brocklehurst, K., Herbert, J. A. L., and Suschitzky, H., *Biochem. J.*, 151, 417 (1975)

29. Sun, Y., Smith, D. L., *Anal. Biochem.*, 172, 130-138 (1988)

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

31. Kirley, T. L., *J. Biol. Chem.*, 264, 7185-7192 (1989)

32. Kirley, T. L., *Anal. Biochem.*, 180, 231-236 (1989)

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

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

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

36. Sturrock, E. D., Yu, X. C., Wu, Z., Biemann, K., and Riordan, J. F., *Biochemistry*, 35, 9560-9566 (1996)

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

38. Degani, Y., and Patchornik, A., *Biochemistry*, 13, 1 (1974)

39. Stark, G. R., *Methods in Enzymology*, Vol. 47, 129-132 (1977)

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

41. Schaffer, M. H., and Stark, G. R., *Biochem. Biophys. Res. Commun.*, 71, 1040-1047 (1976)

42. Papayannopoulos, I. A., and Biemann, K., *Protein Sci.* 1, 278-288 (1992)

43. Mahboub, S., Richard, C., Delacourte, A., and Han, K., *Anal. Biochem.*, 154, 171-182 (1986)

44. Altamirano, M. M., Plumbridge, J. A., and Calcagno, M. L., *Biochem.*, 31, 1153-1158 (1992)

45. Tang, C., Yuksel, K. U., Jacobson, T. M., and Gracy, R. W., Arch. Biochem. Phys., 283, 12-19, (1990)

46. Nefsky, B., and Bretscher, A., *Proc. Natl. Acad. Sci.* USA, 86, 3549-3553 (1989)

47. Minard, P., Desmadril, M., Ballery, N., Perahia, D., and Mouaward, L., *Eur. J. Biochem.*, 185, 419-423 (1989)

48. Wakeselman, M., and Guibe-Jampel, E., *J. C. S. Chem. Comm.*, 21, 1976

49. Lumper, L., and Zahn, H., *Adv. Enzymol.*, 27, 199 (1965)

50. Spackman, D. H., Stein, W. H., and Moore, S., *J. Biol. Chem.*, 235, 648-659 (1960)

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

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

53. Cotter, R. J., *Anal. Chem.*, 56, 485 (1984)

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

55. Hillenkamp, F., Karas, M., *Methods in Enzymology*, Academic Press, New York, Vol. 193, p280

56. Karas, M., Bahr, U., and Giessman, U., *Mass Spectrom. Rev.*, 10, 335 (1991)

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

58. Karas, M., and Hillenkamp, F., *Anal. Chem.*, 60, 2299-2301 (1988)

59. Karas, M., Bachmann, D., Bahr, U., and Hillenkamp, F., *Int. J. Mass Spectrom. Ion Processes*, 78, 53 (1987)

60. Zaluzec, E., Gage, D., Watson, J. T., *Protein Expression and Purification*, 6, 109 (1995)

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

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

63. Beavis, R. C., and Chait, B. T., *Rapid Commun. Mass Spectrom.*, 3, 432 (1989)

64. Beavis, R. C., Chaudhary, T., and Chait, B. T., *Org. Mass Spectrom.*, 27, 156 (1992)

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

66. Chait, B. T., and Kent, S. B. H., *Science*, 257, 1885-1894 (1992)

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

68. Busch, K. L., *J. Mass Spectrom.*, 30, 233 (1995)

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

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

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

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

73. Cotter, R. J., *Biomed. Environ. Mass Spectrom.*, 18, 513 (1989)

74. Kaufmann, R., Kirsch, D., and Spengler, B., *Int. J. Mass Spectrom. Ion Proc.*, 131, 355 (1994)

75. Kaufmann, R., Spengler, B., *Rapid Commun. Mass Spectrom.*, 7, 902 (1993)

76. Carr, S. A., in *Mass Spectrometry in the Biological Sciences*, Burlingame, A. L., Carr, S. A., Eds.; Humana Press: Totowa, NJ, (1996)

77. Billeci, T. M., and Stults, J. T., *Anal. Chem.*, 65, 1709 (1993)

78. Goldberg, D. E., Slater, A. F. G., Beavis, B. T., Chait, B. T., Cerami, A., and Henderson, G. B., *J. Exp. Med.*, 173, 961 (1991)

79. Specht, B., Oudenampsen-Kruger, E., Ingendoh, A., Hillenkamp, F., Leaius, A. G., and Spener, F., *J. Biotechnol.*, 33, 259 (1994)

80. Robertson, J. G., Adams, G. W., Medzihradszky, K. F., Burlingame, A. L., and Villafranca, J. J., *Biochemistry*, 33, 11563 (1994)

81. Chang, J. Y., Schindler, P., and Chatrenet, B., *J. Biol. Chem.*, 270, 11992 (1995)

82. Liao, P-C., Leykam, J., Andrew, P.C., Gage, D., and Allison, J., *Anal. Biochem.*, 219, 9 (1994)

83. Burlingame, A. L., *Curr. Opin. Biotechnol.*, 7, 4 (1996)

84. Wilcox, M. D., Schey, K. L. Busman, M., and Hildebrandt, J. D., *Biochem. Biophys. Res. Commun.*, 212, 367 (1995)

85. Zhao, Y., and Chait, B. T., *Anal. Chem.*, 66, 3723 (1994)

86. Nordhoff, E., Kirpekar, F., and Roepstorff, P., *Mass Spectrom. Rev.*, 15, 67 (1996)

87. Limbach, P. A., Crain, P. F., and McCloskey, J. A., *Curr. Opin. Biotechnol.*, 6, 96 (1995)

88. Nordhoff, E., Kirpekar, F., Karas, M., Cramer, R., Hahner, S., Hillenkamp, F., Kristiansen, K., Roepstorff, P., Lezius, A., *Nucleic Acids Res.*, 22, 2460 (1994)

89. Nordhoff, E., Karas, M., Cramer, R., Hahner, S., Hillenkamp, F., Kirpekar, F., Lezius, A., Muth, J., Meier, C., and Engles, J. W., *J. Mass Spectrom.*, 30, 99 (1995)

90. Kirpekar, F., Nordhoff, E., Kristiansen, K., Roepstorff, P., Lezius, A., Hahner, S., Karas, M., and Hillenkamp, F., *Nucleic Acids Res.*, 22, 3866 (1994)

Chapter II

Novel methodologies for the assignment of both cysteine residues and disulfide bond pairings in proteins

I. A strategy to locate cysteine residues in proteins by specific chemical cleavages followed by mass mapping by MALDI-TOF MS

A. Introduction

Cysteine contributes to protein biological functions by using its free sulfhydryl (SH) group in the active site for enzyme catalysis, as the chelating site for metal ions, and as the active site of disulfide reshuffling enzymes. Pinpointing the number and location of cysteine residues thus becomes an important strategy for determining protein structure. While it is true that the large majority of new protein sequences are now deduced from nucleotide sequences of clones cDNA (or, in the case of prokaryotic proteins, from cloned genomic DNA fragments), differentiating between free cysteine residues (sulfhydryl groups) and cystines (disulfide bonds) is still cumbersome and tedious; furthermore, the cDNA technique, like other conventional techniques, cannot identify posttranslational modifications.

Classical approaches for localizing protein free sulfhydryl groups usually involve modification of free sulfhydryls before and after reduction of a protein, degradation of the derivatized protein chemically or enzymatically into smaller peptide fragments, followed by HPLC fractionation of the peptides and Edman degradation of the derivatized peptides

(1,2). Recently, peptide mapping by mass spectrometry of proteolytic digests before and after HPLC fractionation also has been reported (3-5). However, this procedure requires the sulfhydryl derivatization under a carefully controlled conditions so that sulfhydryl-disulfide bond exchange is not an issue. Secondly, this approach requires that a protein be cleaved at both sides of cysteine residues to give peptides that contain only one cysteine. Multiple-step enzymatic or chemical digestions usually are performed in this case which is tedious and requires a sample size on the order of nanomoles. Thirdly, derivatizing reagents must be chosen to facilitate HPLC separation with UV detection and to avoid the retention time overlap of derivatized cysteine and other amino acids on an Edman sequencer. Although other approaches such as methods based on affinity chromatography of sulfhydryl groups have also been proposed, their feasibility remains to be tested.

B. Cyanylation of Cysteine Residues

One of the most widely used techniques for examination of structure-function relationships in proteins is that of chemical modification of amino acid side chains. The role of thiol groups in proteins has been particularly well studied, mainly because of the availability of a number of specific thiol-modifying reagents such as organo-mercurials (Boyer, 1954) and 5,5'-dithio-*bis*-(2-nitrobenzoic acid), DTNB (Ellman, 1959). Jacobson et al. (6) showed that 2-nitro-5-thiocyanobenzoic acid (NTCB) selectively cyanylates cysteine's free sulfhydryl group under mild conditions. Here in this methodology, we chose NTCB as the cyanylation reagent. Since the reactivity of NTCB toward a thiol group is related to the easy displacement by the sulfur nucleophile of the good leaving

group *p*-nitrothiophenolate. The cyanylation is selective for free sulfhydryl groups, disulfide bonds are unreactive to NTCB. The selective cyanylation and subsequent cleavage of free sulfhydryls can thus be achieved in the presence of disulfide bonds by carrying out the reaction without prior reduction of the protein (6). Other reagents for cyanylation of SH groups have been prepared by Wakselman (7) and by Brocklehurst et al (8, 9). These reagents may offer some advantages over NTCB, but their utility needs to be further examined. Further research on cyanylation reagents was done in our laboratory as discussed in Chapter 3.

C. Chemical Cleavages at Cysteine Residues

NTCB cleavage of peptide chains at the N-peptide bonds of cysteine residues was first observed by Catsimpoolas and Wood (10). However, because of the several 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. Further research on NTCB by Jacobson (6) showed that after cyanylation reaction by NTCB, 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-iminothiazolidine-4-carboxylyl (ITC) peptides (Figure 2.1). If a protein contains **n** cysteine residues, the cleavage reaction results in the formation of **n+1** peptide fragments, and mass analysis of the fragments indicates the number and location of cysteine residues.

Although the original paper claimed that the cleavage reaction can come to completion with little side reactions for most of the peptides and proteins tested, Degani

and Patchornik (11) found that a β-elimination reaction also occurred depending on the structural properties of the respective peptides.

Jacobson et al (6) proposed a mechanism for the cleavage reaction. This mechanism implicates that hydroxide ion catalyzes the cleavage, as shown by the influence of rising pH on the rate of cleavage reaction. A possible explanation of the role of OH^- is shown in Figure 2.2. The amide nitrogen is apparently too weak a base to carry out the nucleophilic attack (the pKa of a protonated amide is 4.0 or less), and the reaction probably proceeds only after attack of OH^- on the carbonyl carbon of the amide. This generates a much more basic nitrogen, which then can participate in the concerted cyclization and cleavage, without formation of an acyliminothiazolidine intermediate.

Since both cleavage and β-elimination occur under mildly alkaline conditions, the two reactions are competitive, sometimes resulting in only a moderate degree of cleavage of a peptide bond. Another inherent problem of the cleavage reaction is that it takes too long to accomplish. Various side reactions may rise after long hours of exposure to alkaline solution, which brings about some uncertainty into the analytical results. From a systematic study done by Watson's group (17) on the cleavage reaction as a function of pH, solvents and amino acid structures for polypeptides containing different amino acids adjacent to N-terminus of cyanylated cysteines, we know that higher pH can greatly promote both cleavage and β-elimination reactions, but the extent of β-elimination does not increase significantly even at a pH as high as 12. Ammonia, a stronger nucleophile than hydroxyl anions, greatly accelerates the cleavage of peptide chains and minimizes side reactions related to the prolonged incubation. The optimized conditions were applied to the recognition of free sulfhydryl groups in a variety of peptides.

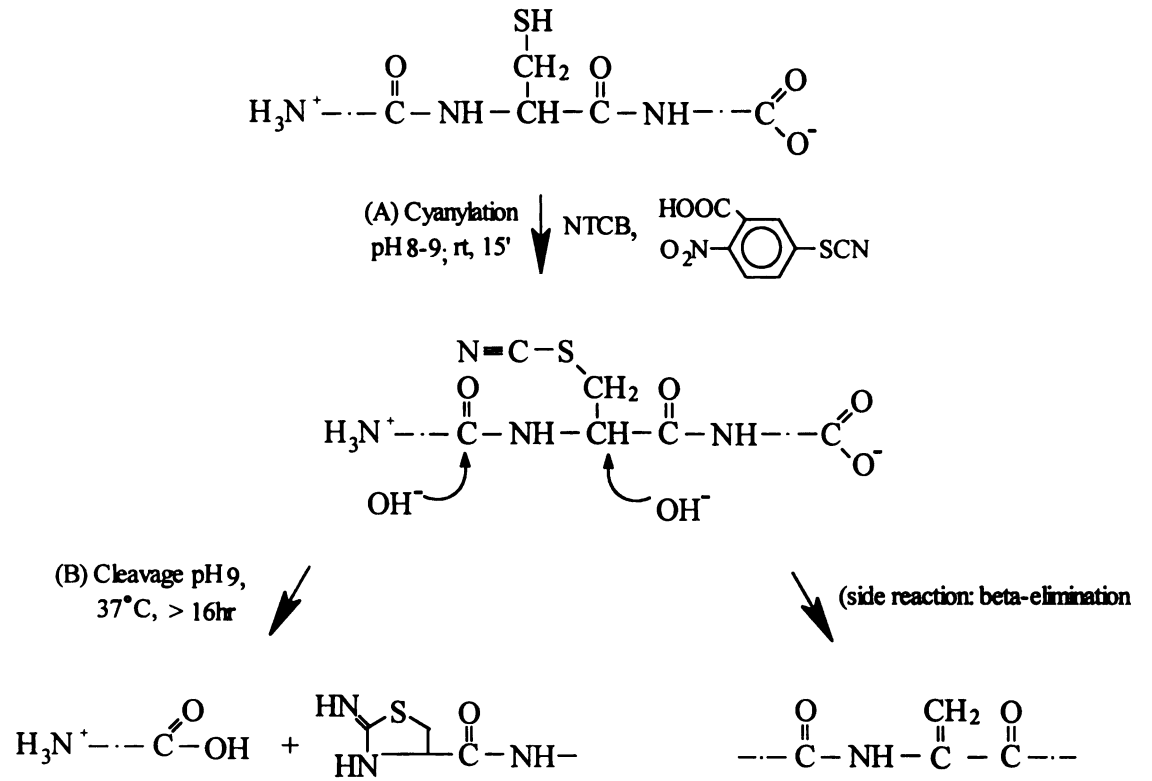


Figure 2.1 Reaction between cysteine residue and 2-nitro-5-thiocyanobenzoic acid (NTCB), (A) Cyanylation and (B) Cleavage Reaction. β-elimination competes with cleavage reaction under alkaline conditions.

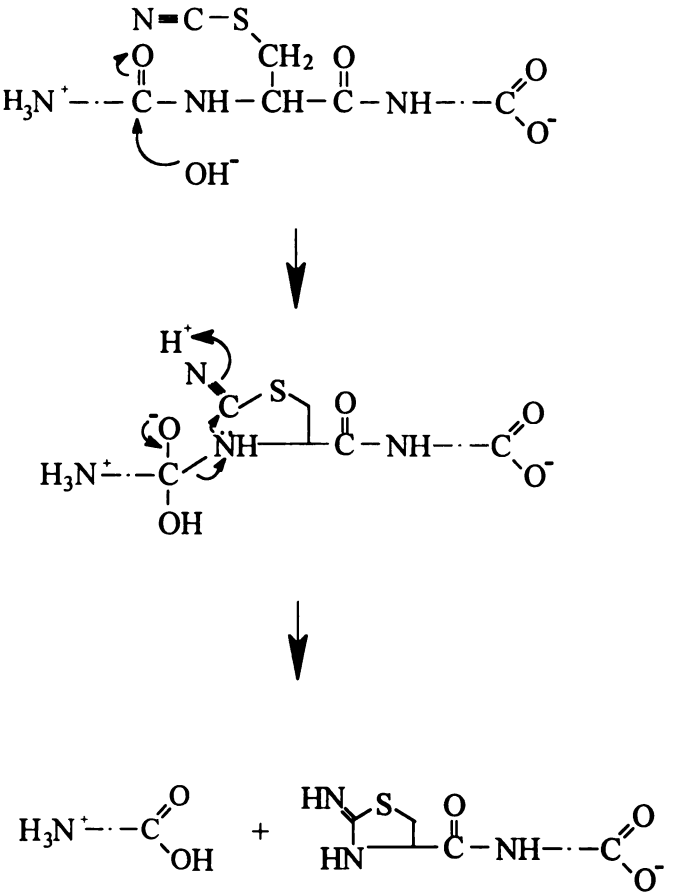


Figure 2.2 A mechanism for the base-catalyzed cleavage reaction

II. A Strategy for Assignment of Disulfide Bond Pairings in Proteins

A. Introduction

Many biological and therapeutically important proteins (peptides) contain disulfide bonds. Because the disulfide bond is an important element of protein structure, it is necessary to know the locations of disulfide bonds in order to understand more fully any unique contributions they may make to protein (peptide) structure and function. 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 challenge protein chemists.

Current methodology (3, 12) for assignment of disulfide bonds in proteins involves cleavage of protein chains between half-cystinyl residues with specific cleavage reagents, such as cyanogen bromide or proteases, such as trypsin, to obtain peptides that contain only one disulfide bond. The resulting mixture of peptides is separated, and the amino acid compositions, sequences or molecular masses of the peptides are determined by Edman degradation or mass spectrometry or both. Assignment of these peptides to specific segments of the protein leads to the recognition of disulfide crosslinkages.

Although this approach is well established and has been used with much success, it is limited in many aspects. First, disulfide-containing peptides can be identified by their amino acid composition or sequence only if they are purified to homogeneity. This requirement may not be achieved in the case of large proteins, or when the quantity of protein is very small. Second, the above methodology requires the protein chain to be cleaved between every half-cystinyl residue so that the resulting peptides contain no more than one disulfide pair. However, even multi-step enzymatic or chemical digestions

rarely yield a full set of diagnostic fragments because proteins are often resistant to some cleavage reagents. It will be even more difficult or impossible to find an enzyme which is likely to cleave between two closely spaced or adjacent cysteines (3). It is likewise important to note that the conditions that are frequently used to cleave a protein chain are also the conditions that frequently lead to disulfide scrambling (13). As a result, the location of disulfide bonds determined by this approach may not be the same as the location of disulfide bonds in the native protein. The problem of disulfide exchange plagues all attempts to locate disulfide bonds in proteins, irrespective of the general method currently used. Although supplementary methodologies such as non-specific fragmentation by partial acid hydrolysis (14) have been proposed to avoid disulfide bond scrambling, it is difficult to deal with the data obtained by these non-specific techniques.

Gray (15, 16) has described an approach for analyzing disulfide linkage patterns in highly bridged small peptides with close or adjacent cysteine residues. In his experiments, peptides were partially reduced under controlled conditions, the isomers of the partially reduced protein separated by HPLC, the nascent free thiols alkylated, and the positions of alkylated cysteines recognized from the results of sequence analysis and related to the disulfide bond pair that had been reduced and cyanylated. However, it is obviously tedious, if not impractical, to sequence an alkylated high-mass peptide or protein using this approach.

B. A Novel Strategy for Disulfide Assignment

Watson's group has developed a simple methodology for recognizing the location of free cysteine groups in peptides and proteins (17). This approach employs a specific

chemical reaction between sulfhydryls and 2-nitro-5-thiocyanobenzoic acid (NTCB) to selectively cyanylate cysteine thiols (18). The N-terminal peptide bond of the modified cysteinyl residue can then be cleaved under alkaline conditions to form an amino-terminal peptide and a series of 2-iminothiazolidine-4-carboxyl (ITC) peptides which can be mass mapped to the sequence of the original molecule by MALDI-MS. Disulfide bonds do not react with NTCB and therefore do not interfere with the determination of sulfhydryl groups. The optimized cleavage conditions greatly facilitated the cleavage and make this methodology much more attractive in terms of speed of the analysis and interpretation of the data.

Another novel strategy (outlined in Figure 2.3) developed by Watson's group for the assignment of disulfide bond pairings in proteins using the above specific chemical cleavage of partially reduced and cyanylated protein isomers with mass mapping of the resulting peptides by MALDI-TOF MS (19). In this methodology, as shown in Figure 2.4 for a simple protein containing only two disulfide bonds, the denatured protein is partially reduced by tris(2-carboxyethyl)phosphine (TCEP) in a buffer solution at pH 3.0 to produce a mixture of residual intact protein and isomers of partially reduced species. Conditions can be optimized so that the predominant products are isomers in which only a single disulfide bond has been reduced . Nascent sulfhydryls are immediately cyanylated by 1-cyano-4-dimethylamino-pyridinium tetrafluoroborate (CDAP) (20) under the same buffer conditions as used for reduction. The partially reduced and cyanylated protein isomers are then separated by reversed-phase HPLC, followed by analysis of HPLC fractions by MALDI-MS to determine which isomers are singly reduced/cyanylated. Those shifted by +52 Da correspond to a singly reduced/cyanylated

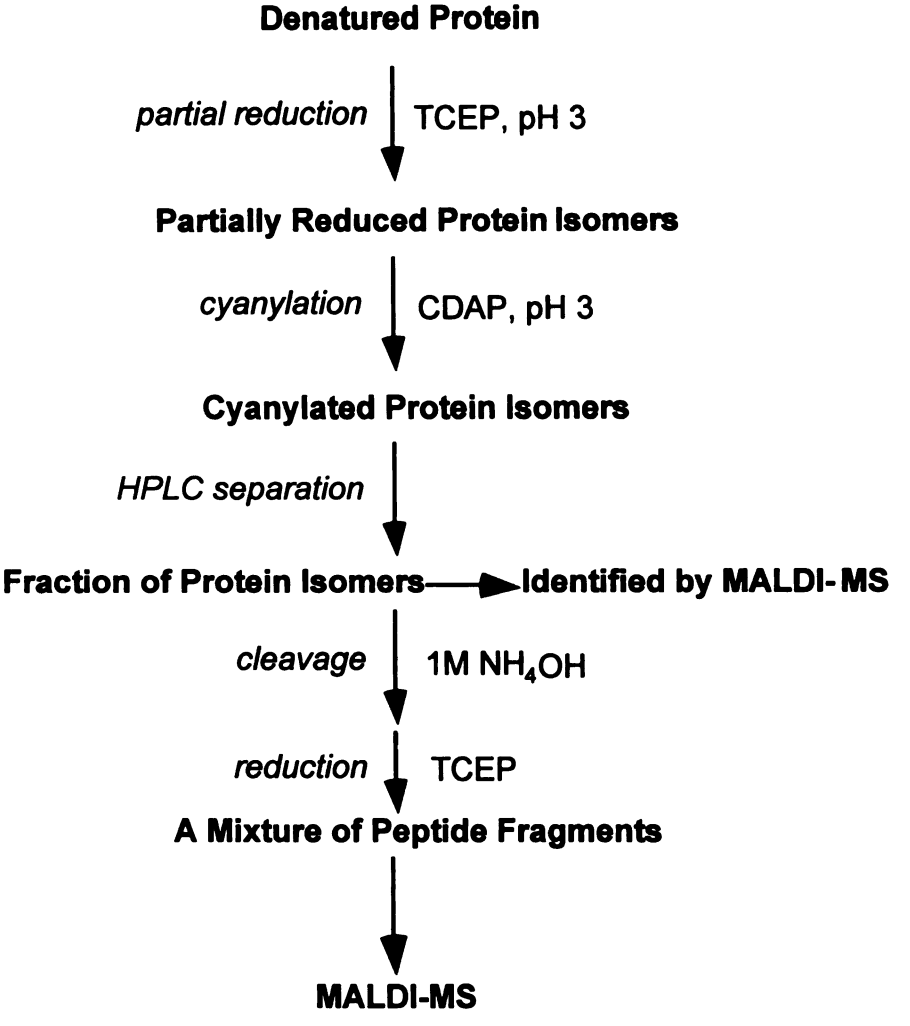


Figure 2.3 Descriptive overview of the cyanylation/cleavage methodology for assignment of disulfide bond pairings in proteins.

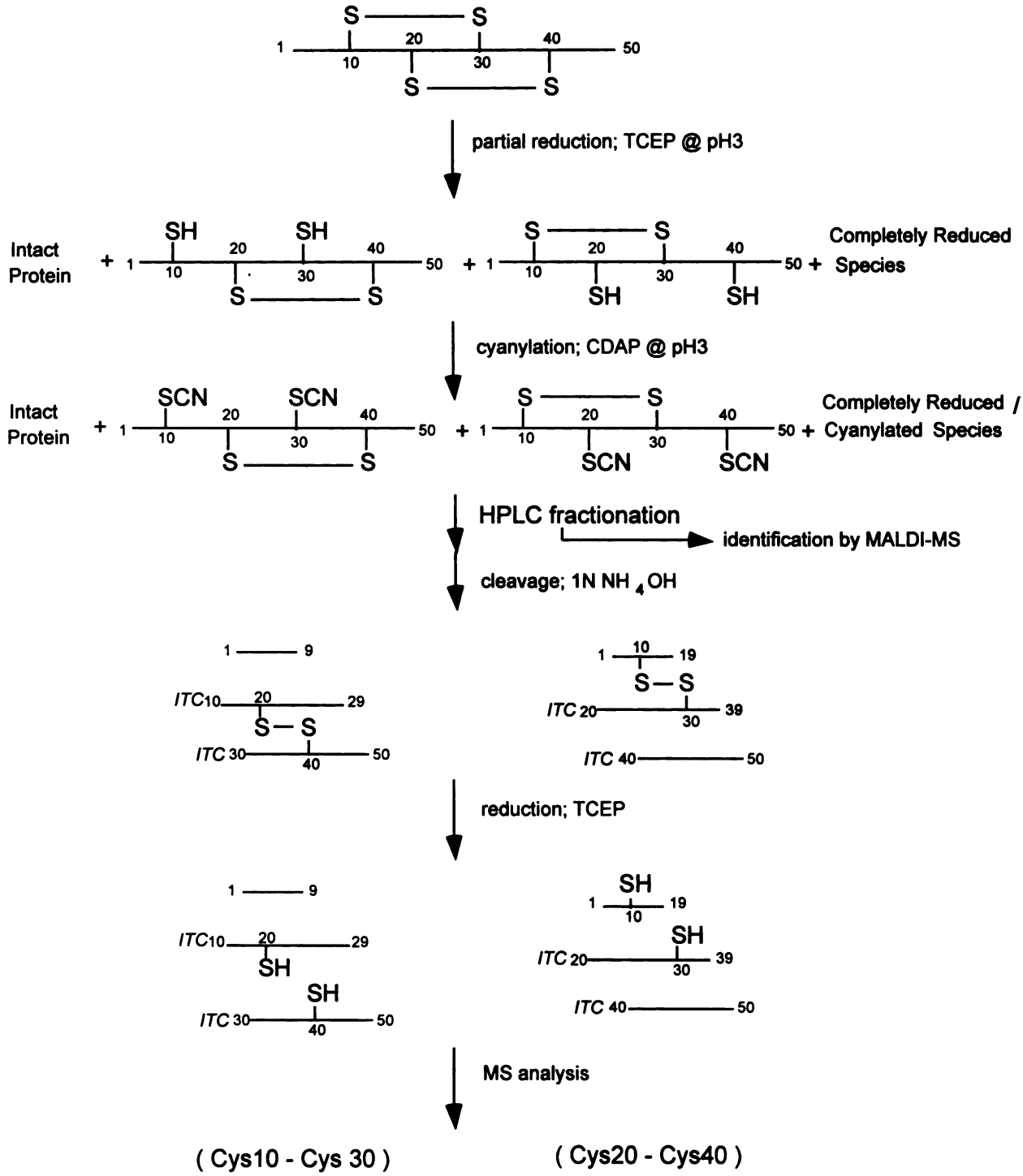


Figure 2.4 Chemical overview of the cyanylation/cleavage methodology. Partial reduction means that the protein of interest is reduced under controlled conditions into a mixture of isoforms, each of which corresponds to reduction of only one of its disulfide bonds; for a protein containing **n** cystines, **n** isoforms of the singly-reduced protein will result. ITC stands for iminothiazolidinyl carboxyl residue at the amino terminus.

species; a shift of +104 Da corresponds to a doubly reduced/cyanylated 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 MALDI-MS. The masses of the resulting peptide fragments are related to the location of the paired cysteines that had undergone reduction, cyanylation, and cleavage.

A primary advantage of this approach is its underlying simplicity. The data obtained from this approach are straightforward and easily interpreted because only a few protein isomers are produced, and each is relevant. Typically, only **n-1** intermediates are needed to define an **n**-bridge system. Redundant information is frequently obtained to confirm an assignment. Secondly, we describe a practical way to circumvent problems related to disulfide bond scrambling because both reduction and cyanylation are performed in an acidic media. Finally, our approach can be used for the assignment of adjacent or closely spaced cysteines for which conventional methodology fails. Therefore, two problems associated with current methodologies are solved using this new methodology.

In Chapter 3, the cyanylation reagent, CDAP, is compared with other reagents; also, the results of studies of cyanylation reactions are summarized.

III. References

1. Dabre, A., *Practical Protein Chemistry: A Handbook*, John Wiley & Sons, (1985) p.149.
2. Chin, C. C. Q., and Wold, F., *Anal Biochem.*, **214**, 128 (1993)
3. Hirayama, K., and Akashi, S., in *Biological Mass Spectrometry: Present and Future* (Matsuo, T., Caprioli, R. M., Gross, M. L., and Seyama, Y., Eds.), John Wiley & Sons Ltd., (1994). P. 299-312
4. Rpbertson, J. G., Adams, G. W., Medzihradszky, K. F., Burlinggame, A. L., and Villafranca, J. J., *Biochemistry*, **33**, 11563 (1994)
5. Glocker, M. O., Arbogast, B., Milley, R., Cowgill, C., and Deinzer, M. L., *Proc. Natl. Acad. Sci. USA.*, **91**, 5868 (1994)
6. Jacobson, G. R., Schaffer, M. H., Stark, G. R., and Vanaman, T. C., *J. Biol. Chem.*, **248**, 6583 (1973)
7. Wakeselman, M., Guibe-Jampel, E., Raoult, A., and Busse, W. D., *J. Chem. Soc. Chem. Commun.*, **1**, 21 (1976)
8. Brocklehurst, 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)
9. Brocklehurst, K., Malthouse, J. P. G., *Biochem. J.*, **175**, 761 (1978)
10. Catsimpoolas, N., and Wood, J. L., *J. Biol. Chem.*, **241**, 1790 (1966)
11. Degani, Y., and Patchornik, A., *Biochemistry*, **13**, 1 (1974)
12. Smith, D. L., and Zhou, Z., *Methods Enzymol*, **193**, 374 (1990)
13. Ryle, A. P., and Sanger, F., *Biochem. J.*, **60**, 535 (1955)
14. Zhou, Z., and Smith, D. L., *J. Protein Chem.*, **9**, 523 (1990)

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

16. Gray, W. R., *Protein Sci.*, **2**, 1749 (1993)

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

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

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

Chapter III Study on Cyanylation Reagents

I. Introduction

Among the 20 amino acids that compose proteins, cysteine has unique properties. For a new protein with unknown sequence, the quantitative determination of cysteine content is very beneficial to further characterization of the protein.

The cysteine sulfhydryl group is distinguished from other side chains of amino acids by its high reactivity and by the exceptionally diverse chemical reactions in which it participates, such as alkylation, acylation, oxidation, sulfhydryl-disulfide bond exchange, charge-transfer complexes, reactions with organic heavy metal compounds (1). So there are many ways for the modification of free sulfhydryl group. In the strategy developed by Watson's group as discussed in Chapter 2, modification of cysteine's free sulfhydryl group is the first step. Research on sulfhydryl modification can be dated back to the 1950s' (2) and several reagents were involved in the modification of the cysteine group.

A. Organic Mercurial Reagents:

Organic mercurial compounds are high affinity, specific reagents for the protein SH group. They have been widely used in studies of the biochemical function of sulfhydryl groups since the pioneering research of Hellerman (3) with *p*-mercuribenzoate (*p*-MB). Spectrophotometric measurement of the extent and rate of reaction of *p*-mercuribenzoate with sulfhydryl groups proposed by Boyer (2) stands in first place because of its sensitivity, simplicity, selectivity, and precision. This classical method is

selective because although combination of *p*-mercuribenzoate with other groups or compounds in the protein preparations may occur, only the combination with the -SH groups results in a large increase in the molecular absorptivity at 250nm. Spectral measurements may thus serve to distinguish between combinations of *p*-mercuribenzoate with -SH groups and other groups (2) (Figure 3.1). Interestingly, the rate and the extent of the reaction of SH groups of many proteins with *p*-MB is higher at pH 4.6 than at pH 7-8 because of the high affinity of *p*-MB for hydroxide ions (4). As a result, the reactivity of the reagent increases as the pH of the solution is decreased. Other reagents, such as 1-(4-chloromercuriphenylzao)-napthol-2 (CMPN; mercury-orange), and *p*-(hydroxymercuri)-benzoic acid (5), were also proposed to improve the spectrophotometric detection of SH groups in lysine monoxygenase (6) and in carboxyhemoglobin (7).

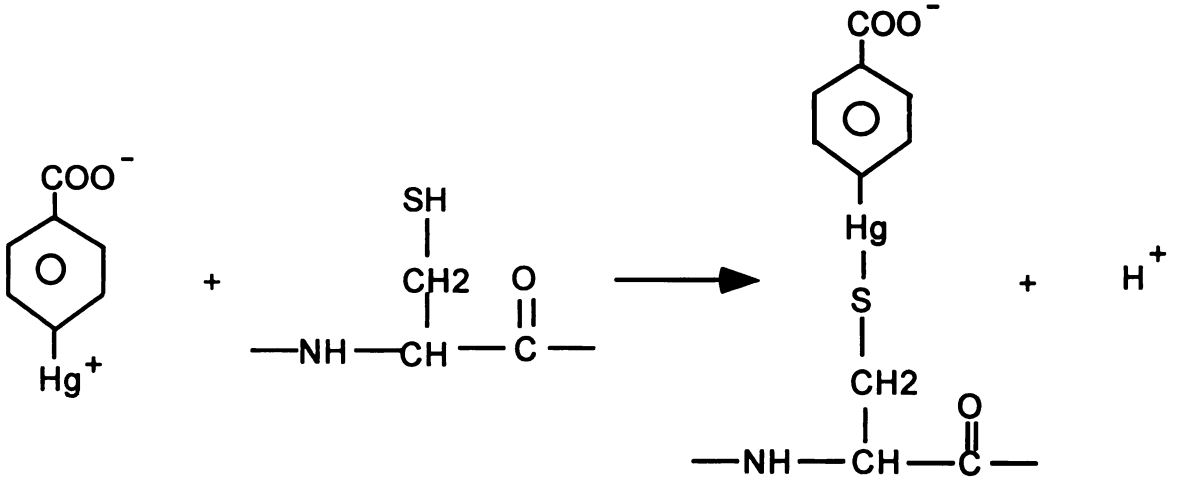


Figure 3.1 Chemical reaction between *p*-mercuribenzoate and sulfhydryl group

B. Sulfhydryl / Disulfide Exchange:

Among oxidants of SH groups, disulfides occupy a special position since their reactions with sulfhydryls are absolutely specific. This reaction is referred to as sulfhydryl/disulfide exchange. As can be seen from the below equations, this reaction consists of two steps of nucleophilic substitution with the formation of a mixed disulfide at the intermediate stage.

$$R_1SSR_1 + RSH \text{---- } R_1SH + R_1SSR \qquad \text{Equation 3.1}$$

$$R_1SSR + RSH \text{ ---- } R_1SH + RSSR \qquad \text{Equation 3.2}$$

Among the various disulfides proposed, Ellman's reagent, 5,5'-dithio-*bis*-(2-nitrobenzoic acid), DTNB, had received the widest application (8). The reaction of thiolate anion with excess Ellman's reagent at pH 7-8 is favored toward the stoichiometric formation of TNB thiolate (5-thio-2-nitrobenzoate, TNB^-) and a mixed disulfide (RS-TNB) (Figure 3.2).

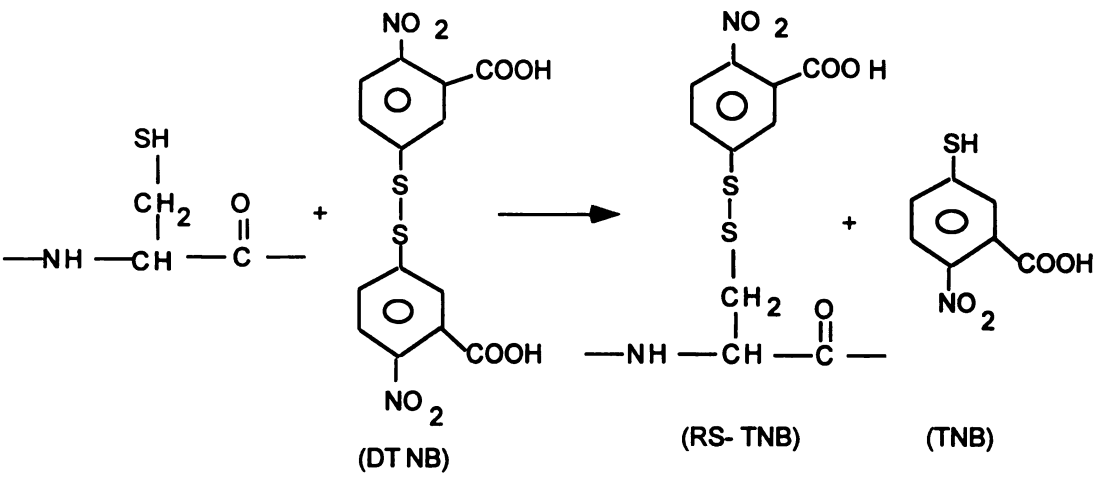
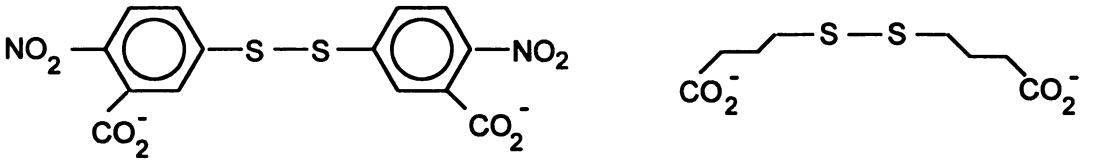


Figure 3.2 The reaction of Ellman's reagent, 5,5'-dithio-*bis*-(2-nitrobenzoic acid), with cysteinyl residues in proteins

More work was done by Vanaman and Stark (9) on this reagent. They proposed the use of enzyme-thionitrobenzoate mixed disulfides as intermediates in the preparation of specific derivatives of enzyme -SH groups and the potential of the use of enzyme thiocyanates in specific chemical cleavage of polypeptide chains at cysteine and cystine residues. They found that the reaction of the sulfhydryl group with DTNB and other symmetrical disulfides are different. (Figure 3.3).



DTNB

3-carboxypropyldisulfide

Figure 3.3 Structure of DTNB and 3-carboxypropyldisulfide

Even the interatomic distances and bond angles in the above two compounds are slightly different, DTNB and 3-carboxypropyldisulfide can achieve similar orientations. So the failure of 3-carboxypropyldisulfide to react with sulfhydryl group reflects the large intrinsic difference between the reactivity of an aliphatic disulfide and DTNB (9). The reaction of DTNB with free -SH in proteins can be summarized as follows: it forms a mixed disulfide with concomitant release of 1 mole of TNB per mole of -SH as shown in Figure 3.2 (8). The position of equilibrium for this disulfide interchange reaction is very much in the direction of the mixed disulfide, since the aromatic mercaptide anion has the

low pKa of 5.1 (8). The use of enzyme-TNB mixed disulfide to achieve specific and reversible chemical modification of -SH groups in proteins seems to be very promising.

Reaction of disulfides with cyanide has been used by Wood and Catsimpoolas (10) as a route to specific chemical cleavage of peptides which contain cystine. The unsymmetrical disulfide enzyme-TNB is cleaved by cyanide in only one way, i.e., to release TNB. For specific cleavage at cystine and cysteine residues, one might (a) reduce the protein disulfides with reducing agent, (b) allow all of the -SH groups to react with excess DTNB, and (c) add excess cyanide at pH 7. The product is an N-terminal iminothiazolidine ring.

Later on, another reagent 2,2'-dipyridyl disulfide (2-Py-S-S-2-Py) (Compound I) is more often used than 5,5'-dithio-*bis*-(2-nitrobenzoic acid) because its reactions with thiols are essentially stoichiometric over a wide range of pH (Grassetti & Murray, 1967; Brocklehurst & Little, 1973) (11,12,13). The particular value of pyridyl disulfides as reactivity probes and labelling reagents is that they are two-protonic-state electrophiles, i.e., they markedly increase their electrophilicity when protonated on a pyridyl nitrogen atom (e.g.,compound II). Mixed disulfides containing the 2-pyridyl moiety (compound III) have found application in the purification of thiol enzymes by covalent chromatography. The mixed disulfides retain the ability to act as two-protonic-state electrophiles and thus permit selective isolation or labelling of low-pKa thiols in acidic media. The range of such electrophilic reagents is being extended to provide variation in solubility, ionization properties and spectroscopic characteristics and in the nature of the electrophilic center such as the cyanylating reagent 2-thiocyanopyridine (Compound IV). This reagent provides thiol derivatives that are sterically relatively undemanding so that it

may possess advantages relative to 2-nitro-5-thiocyanatobenzoic acid (NTCB) analogous to those that compound I possess relative to 5,5'-dithio-*bis*-(2-nitrobenzoic acid) (DTNB) (11,12).

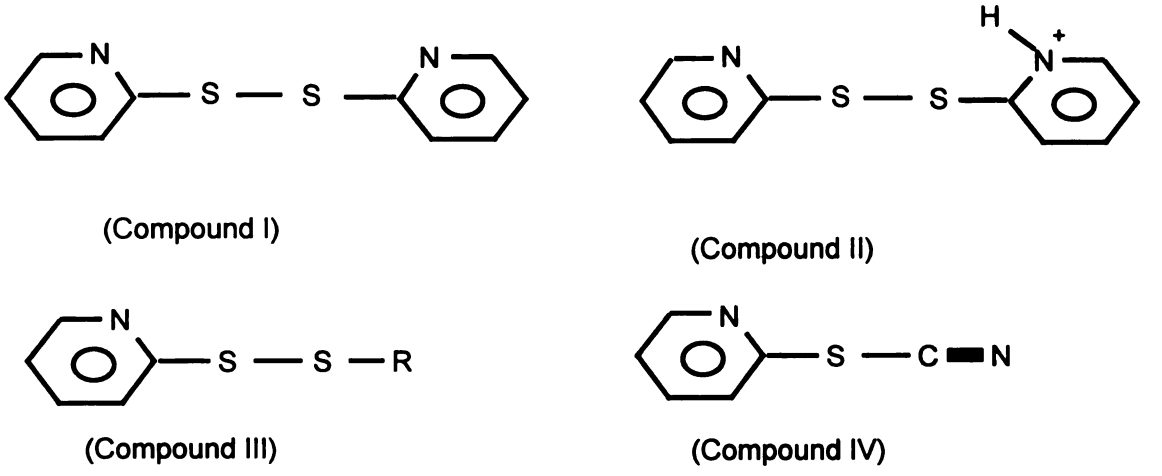


Figure 3.4 Cysteine Modification Reagents

C. NTCB (2-nitro-5-thiocyanobenzoic acid) as a reagent:

The 2-nitro-5-thiocyanobenzoic acid mentioned above is a widely used reagent for quantitative cyanylation of the cysteine residues (14). Its reaction is shown in Figure 3.5. NTCB is shown to be a potential tool for the selective chemical cleavage of peptide chains at the N-terminal of the cyanylated cysteinyl residues. Earlier reports (Wood and Catsimpoolas, 1963; Catsimpoolas and Wood, 1966) (10,15) showed that β-thiocyanoalanine residues when generated by excess cyanide treatment of cystine disulfide bonds, underwent cyclization to form 2-iminothiazolidine rings, with consequent cleavage of N-peptide bond. (For a review, see Spande et al., 1970) (16).

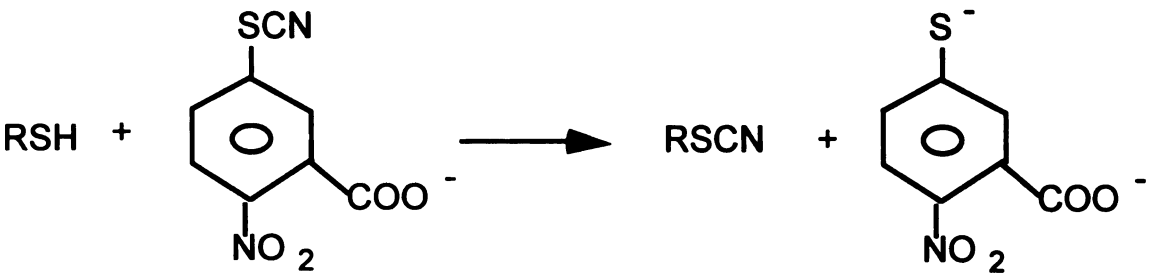
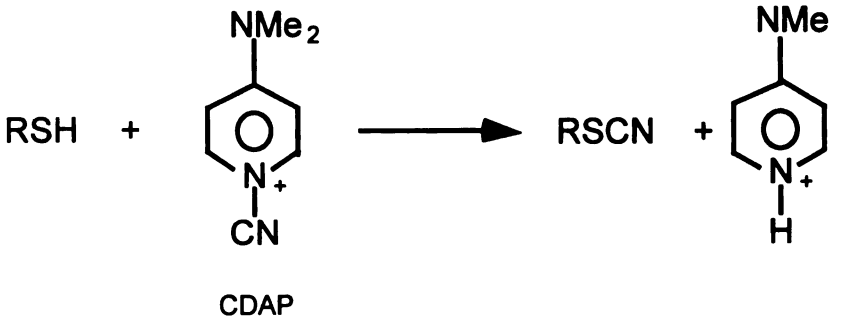


Figure 3.5 Reaction of NTCB with SH

D. CDAP (1-cyano-4-dimethylamino-pyridium salt) as a reagent:

Another reagent called 1-cyano-4-dimethylamino-pyridinium (CDAP) perchlorate or fluoroborate is also used as a cyanylation reagent (Wakselman, 1976) (17). Its reaction is shown in Figure 3.6. According to Wakselman, CDAP reacts rapidly with thiols in neutral and acidic medium: in 11 minutes at pH 3.6, 98% of the catalytic activity of papain is inhibited; Cys 7 and Cys 19 of the reduced B-chain of bovine insulin are quantitatively cyanylated at pH 3.5, then the N-peptide bonds are selectively cleaved at pH 9.5. This water soluble reagent was also found to react with the protein sulfhydryl group in a wide pH range, from pH 2-7, and it can lead to isolable thiocyanates without oxidative side reactions. Then the protein chain may be selectively cleaved in alkali.



CDAP

Figure 3.6 Reaction of CDAP with a generic free sulfhydryl group(RSH)

This chapter will focus more on other cyanylation reagents such as 2-thiocyanopyridine and 4-(dimethylamino) phenyl thiocyanate and compare them with the traditional reagents such as NTCB and CDAP. Although other reagents mentioned above can modify the free sulfhydryl group in proteins and peptides, they either can only be used under alkaline conditions which may cause disulfide bond scrambling or there is no specific cleavage reaction for the modified cysteine residues in order for us to use mass spectrometry to analyze them. Our laboratory uses both NTCB (18) and CDAP (19) to modify cysteine residues, but there are limitations of these two reagents. NTCB can only be used at pH 8 to 9, so that disulfide scrambling is very likely to happen. CDAP seems to be a good choice since it can be used at pH 3, and the reaction is fast and specific. However, for some peptides and proteins which have very tight structures, their cysteine groups may be difficult to cyanylate because of their poor accessibility. In other cases, we might want to increase the pH of the reaction or raise the reaction temperature so that poorly accessible cysteine groups can be cyanylated. CDAP is very unstable above pH 4, having a very short half life; for this reason, it is necessary to make a fresh CDAP solution for the cyanylation reaction. Unlike CDAP, PYRSCN(2-thiocyanopyridine) and

4-(dimethylamino) phenyl thiocyanate are very stable and they have similar structures to NTCB. We expect PYRSCN and 4-(dimethylamino) phenyl thiocyanate can be used in a wide pH range. More research was done on these two NTCB series of reagents.

II. Materials and Methods:

A. Chemicals:

Guanidine hydrochloride was a product of Boehringer-Mannheim Biochemicals (Indianapolis, IN), citric acid, sodium citrate, 1-cyano-4-dimethylamino-pyridinium tetrafluoroborate (CDAP) were purchased from Sigma and used without further purification. Acetonitrile and TFA were of HPLC grade. 2-nitro-5-thiocyanobenzoic acid (NTCB) and Trizma buffer were purchased from Sigma and used without further purification. The 0.01M solution of NTCB was prepared in either 0.1M Tris-HCl or the buffer made in 6M guanidine-HCl solution and the pH of the solution was adjusted to pH 8.0. The 0.10M CDAP solution in pH 3.0, 0.1M citrate buffer-4M guanidine-HCl was freshly prepared prior to use. Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride was purchased from Pierce Chemical co. (Rockford, IL). The TCEP solution in 0.1M citrate buffer at pH 3.0 was prepared as 0.01M stock solution and stored under N2 at –20C for weeks with little deterioration. Rabbit muscle creatine phosphokinase was purchased from Sigma and used without further purification.

Peptide YEKPLQNFTLCFR (MW 1657.1), peptide RGPCRAFI (MW 919.1) and peptide RYVVLPRPVCFEKGMNYTVR (MW 2427.9) were purchased from Bachem

California Inc.. The 1nmol/ml protein and peptide solutions were prepared in 0.1M citrate buffer, pH 3.0, containing 6 M guanidine-HCl.

B. Synthesis of 2-thiocyanopyridine (PYRSCN)

2 grams of 2-mercaptopyridine and 2.9-3.6 grams of NaCN was suspended in 30-40 ml methanol saturated with KBr (20). This mixture was cooled to –5C and stirred while a solution of 1.32 ml Br_2 in 7.4 ml methanol was added dropwise. The reaction mixture was then poured into ice-water and the product was extracted with ethyl ether. The upper layer of the extract was collected in a 50 ml flask, and a rotovaporator was used to vaporize the ether. The residual solution in the flask is the PYRSCN reagent; the color is dark brown.

Purification of PYRSCN was done by first adding some anhydrous sodium sulfate to absorb water that was left in the solution and removal by filtration. Then 10ml of methylene dichloride was added to dissolve it. A column was prepared, with some glass wool on the bottom, then 0.3cm anhydrous sodium sulfate, then 5 cm silica gel was added on the top. 10 ml of methylene dichloride was used to wet the column, and then PYRSCN solution was poured onto the column. The first 10 ml of the eluent was discarded; the following eluent was collected. The impurities elute slower than the pure compound, so they were left on the top part of the column. Use nitrogen stream to evaporate the methylene dichloride from the solution. After purification, the color of PYRSCN is light yellow. It is not soluble in water, but soluble in 2:1 actonitrile:water.

C. Detection of PYRSCN by GC/MS:

GC/MS was done on a Varian Saturn 2000. GC conditions were: 35C for 3 minutes, then ramp to 260C at 15C/min. Solvent delay was set for 5 minutes. An aliquot of 0.5ul PYRSCN reagent was dissolved in 24ml of chloroform, and a 1ul sample was injected. A molecular ion peak and a fragment peak corresponding to loss of SCN were both detected, identifying the PYRSCN reagent.

D. Mass Spectrometry:

MALDI mass spectra were obtained on a Voyager Elite time-of-flight (TOF) mass spectrometer (PerSeptive Biosystems Inc., Framingham, MA) equipped with a model VSL-337ND nitrogen laser (Laser Science, Newton, MA). The accelerating voltage in the ion source was set to 25 KV. Data were acquired in the positive or negative linear mode of operation. Time-to-mass conversion was achieved by external calibration using standards of bradykinin (m/z 1061.2), bovine pancreatic insulin (m/z 5734.5) 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 0.1% 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.

E. Cyanylation of Nascent Sulfhydryls:

For CDAP: to the partially reduced protein mixture was added a 10-fold molar excess of CDAP solution over the total cysteine content. Cyanylation of the nascent sulfhydryl groups was accomplished by incubation at room temperature for another 10-15 minutes.

For NTCB: to the partially reduced protein mixture was added a 10-fold molar excess of NTCB solution over the total cysteine content. Cyanylation of the nascent sulfhydryl groups was accomplished by incubation at 37C for 30 minutes.

For PYRSCN and 4-(dimethylamino) phenyl thiocyanate: to the partially reduced protein mixture was added a 50-fold molar excess of PYRSCN solution or 4-(dimethylamino) phenyl thiocyanate solution over the total cysteine content. Cyanylation of the nascent sulfhydryl groups was accomplished by incubation at room temperature for 30 minutes.

F. HPLC Separation of Partially Reduced and Cyanylated Protein Isomers:

Partially reduced and cyanylated species were separated by reversed-phase HPLC with linear gradient elution using Waters model 6000 pumps controlled by a PC computer. UV detection was at 215nm. Mobile phase are acetonitrile and water. The columns were a Vydac C18 (#214TP54, 10-um particle size, 300-A pore, 4.6 x 250mm) . The major HPLC fractions were collected manually and the masses of the collected protein isomers were determined by MALDI-MS. Appropriate fractions were then dried for further use. The sample size of 10 nanomoles was used for convenient detection from

conventional HPLC columns; the use of microbore columns should allow the use of much smaller sample sizes.

G. Cleavages of singly reduced and cyanylated protein isomers

To the dried HPLC fraction was added 1 M NH_4OH. Cleavage of the peptide chain was performed at room temperature at pH 11 for one hour. Excess ammonia was removed in a vacuum system.

H. Complete reduction of remaining disulfide bonds

Truncated peptides, still linked by residual disulfide bonds, were completely reduced by reacting with 2ul of 0.1M TCEP solution at 37C for 30 minutes at pH 3-5 to minimize the possibility of reoxidation. Samples were diluted with 100ul of 50% (v/v) acetonitrile/0.1% TFA solution prior to analysis by MALDI-MS.

III. Results and Discussion:

A. Locate the free sulfhydryl groups in rabbit muscle creatine phosphokinase using PYRSCN as the cyanylation reagent:

Creatine Phosphokinase is a protein containing 380 amino acid residues and 4 free sulfhydryl groups at positions 73, 145, 253 and 282.

The cyanylation reaction was conducted on this protein using a 50-fold of excess PYRSCN reagent at pH 7, room temperature for 30 minutes. The modified protein was purified by RP-HPLC using gradient elution to remove excess reagent and solvent, then cleavage reactions were conducted on HPLC fractions 1-3 using 1M ammonia at pH 11

for one hour. Sample was then dried and pH was adjusted to pH 3 using citrate buffer, and then take this to MALDI for detection.

The HPLC chromatogram showed 3 peaks (Figure 3.7). The first peak gave no signal above m/z 1000 in MALDI mass spectrum, so it might be due to some interference from the PYRSCN reagent. The second and the third peak gave complementary data. MALDI data are shown below (Figure 3.8) for the cleavage products of the cyanylated material represented by peak 3 in the HPLC chromatogram. From the mass spectrum of the cleavage products of peak 2 the protein with only its cysteine 73 and 145 cyanylated, there are three peaks: m/z 8236.02 is the mass of residues 1 to 72, m/z 8165.21 is the mass of the iminothiazolidine (ITC) derivative of residue 73 to 144, and m/z 4117.95 is the doubly charged ion of m/z 8236.02. This indicates that the second HPLC peak is the cyanylation product with only cysteine 73 and 145 cyanylated. The mass spectrum of the third HPLC peak gave a full set of data, that is, this cyanylation peak is the peak with all four cysteines cyanylated. On one hand this tells us that the accessibility of each of the four free sulfhydryl groups in creatine phosphokinase is different, on the other hand since CDAP cyanylation of creatine phosphokinase only gave us a single cyanylation peak on HPLC which suggests that CDAP is a stronger cyanylation reagent than PYRSCN.

From the data shown below (Table 3.1) we can successfully deduce the free sulfhydryls' locations in creatine phosphokinase. Although PYRSCN is not as reactive as CDAP toward free sulfhydryl groups, it still cyanylates the protein as indicated by the MALDI spectrum in Figure 3.8. PYRSCN seems to be an effective cyanylating reagent around pH 7.

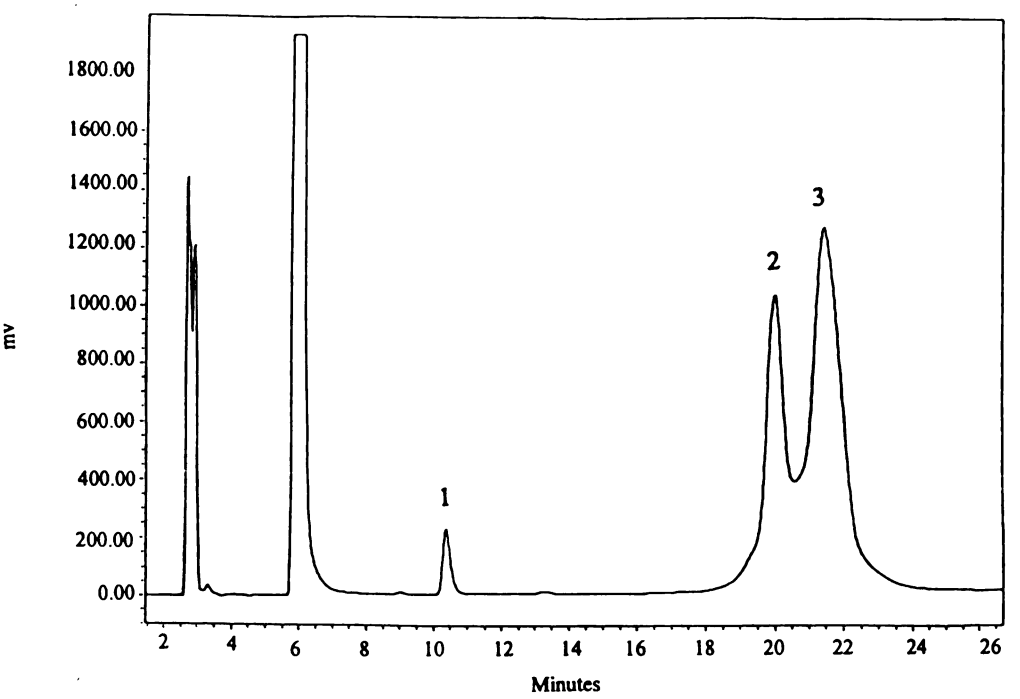


Figure 3.7 HPLC chromatogram of cyanylated creatine phosphokinase

73 145 253 282
1---GCV---HCS---FCV---TCP---380

Fragment	Calculated m/z [M+H]+	Observed m/z [M+H]+
1-72	8234.2	8228.03
73-144	8161.0	8160.52
145-252	12625	12614
253-281	3414.0	3411.76
282-380	10719	10716.5
253-380	14055	13377

Table 3.1 Calculated and observed masses for cleavage products of creatine phosphokinase

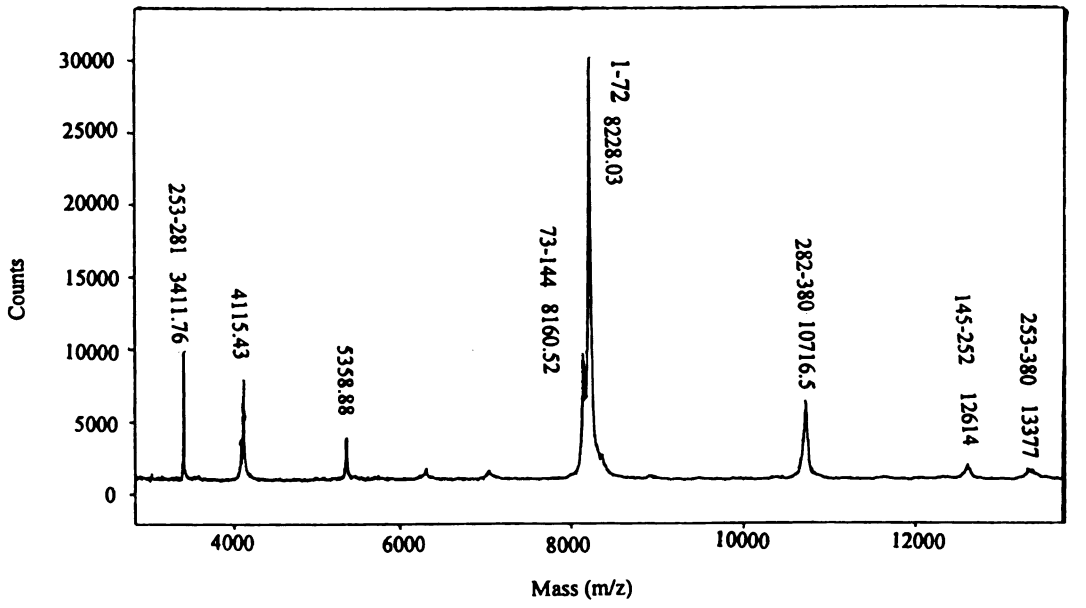


Figure 3.8 MALDI spectrum of cleavage product of creatine phosphokinase HPLC peak 3 in Figure 3.7

B. pH effect on cyanylation reaction:

Cyanylation reactions were done on three model peptides : peptide 919, peptide 1657 and peptide 2428, according to molecular weight, calculated and observed values are shown in Table 3.2. Experiments were conducted by taking aliquots of a 1 nmol/μl peptide solution and adding to it a 50-fold molar excess of the cyanylation reagent solution. React under room temperature for 30 minutes. Both pH 7 and pH 8 were tried on the three model peptides. Then HPLC was used to separate the products. In each cyanylation mixture; three HPLC peaks were observed, each peak was integrated and percentage of each one was given below in Table 3.3, Table 3.4 and Table 3.5.

Table 3.2 Calculated and observed masses for cyanylation product of peptide 919, 1657 and 2427.9

		Cyanylated Product $[M+H]^+$		Mixed Disulfide $[M+H]^+$		Dimer $[M+H]^+$	
Peptide Sequence	M.W.	Cal.	Observed	Cal.	Observed	Cal.	Observe
RGPCRAFI	919.3	945.3	945.53	1029.3	1029.42	1837.6	1836.2
YEKPLQNFTLCFR	1657.1	1683.1	1683.18	1767.1	1767.33	3313.2	3312.24
RYVVLPRPVCFEKGMNYTVR	2427.9	2453.9	2454.43	2537.9	2539.7	4854.8	4857.3

For all the peptides reacting with PYRSCN shown in Figure 3.9, Figure 3.11 and Figure 3.13: HPLC peak 1 is the cyanylated product, peak 2 is the mixed disulfide product and peak 3 is the dimer peak.

1. **Results of Study on Peptide 2428:**

Table 3.3 Product ratios of peptide 2428 after reacting with PYRSCN under both pH 7 and pH 8.

pH	Percent of Cyanylated Peptide	Percent of Mixed Product	Percent of Dimer
pH7	68%	14%	12%
pH8	50%	7%	43%

HPLC chromatograms of cyanylation products of peptide 2428.

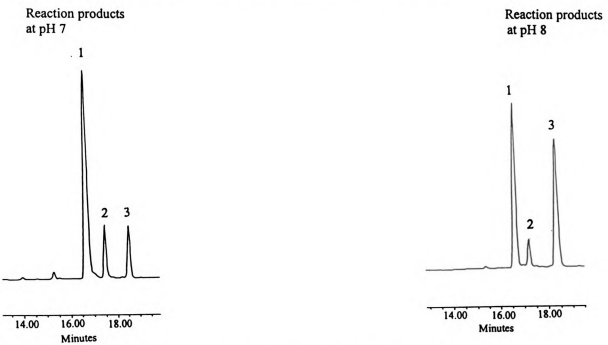


Figure 3.9 HPLC chromatogram of cyanylated peptide 2428, at pH 7.0 & 8.0, room temperature, 50-fold PYRSCN for 30 minutes

The following figures show the MALDI spectrum obtained from the compounds represented by each of the three peaks in Figure 3.9

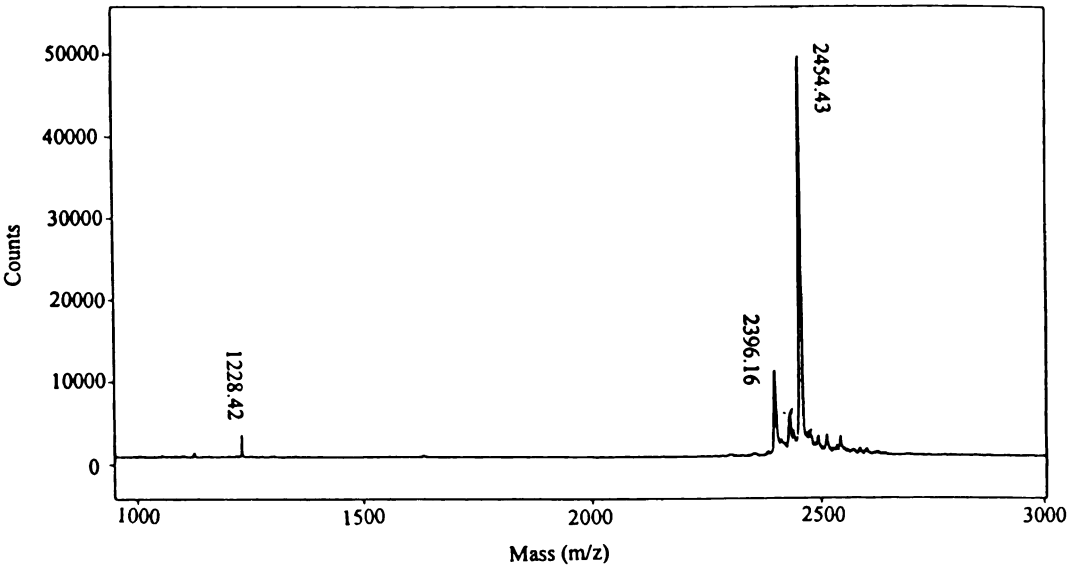


Figure 3.10(a): MALDI spectrum of peptide 2428 HPLC peak 1

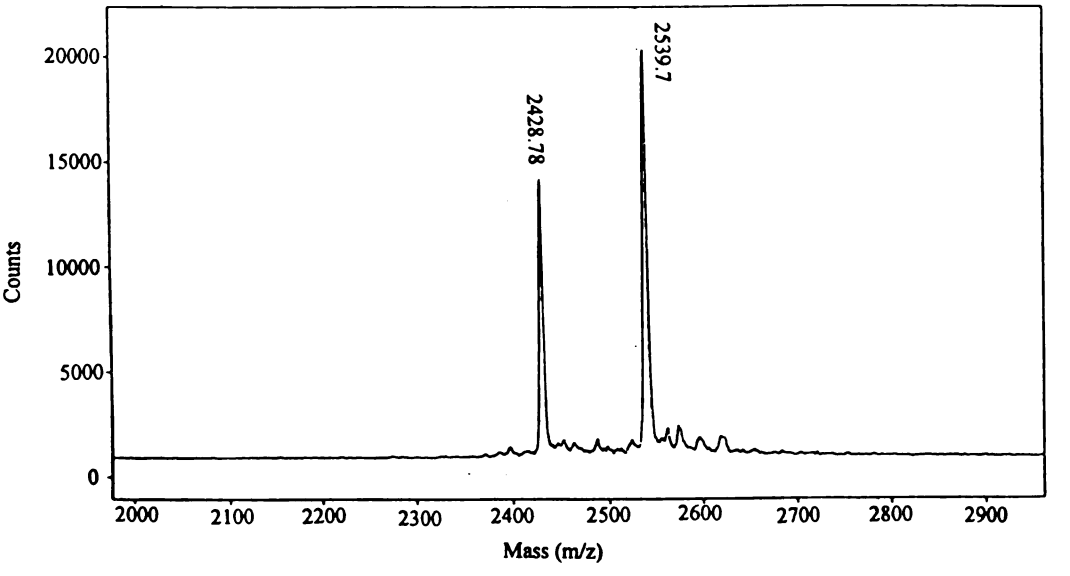


Figure 3.10 (b) MALDI spectrum of peptide 2428 HPLC peak 2.

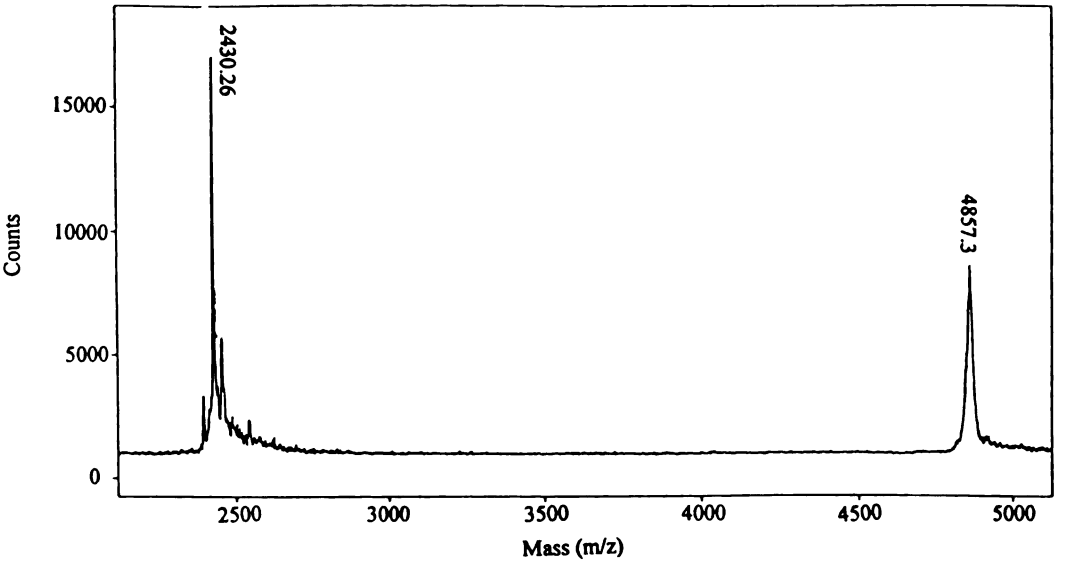


Figure 3.10 (c) MALDI spectrum of peptide 2428 HPLC peak 3.

2. Results of Study on Peptide 1657:

Table 3.4 Product ratios of peptide 1657 after reacting with PYRSCN under pH 7 and pH 8.

pH	Percent of Cyanylated Peptide	Percent of Mixed Product	Percent of Dimer
pH7	60%	25%	15%
pH8	63%	12%	25%

HPLC chromatograms of cyanylation products of peptide 1657:

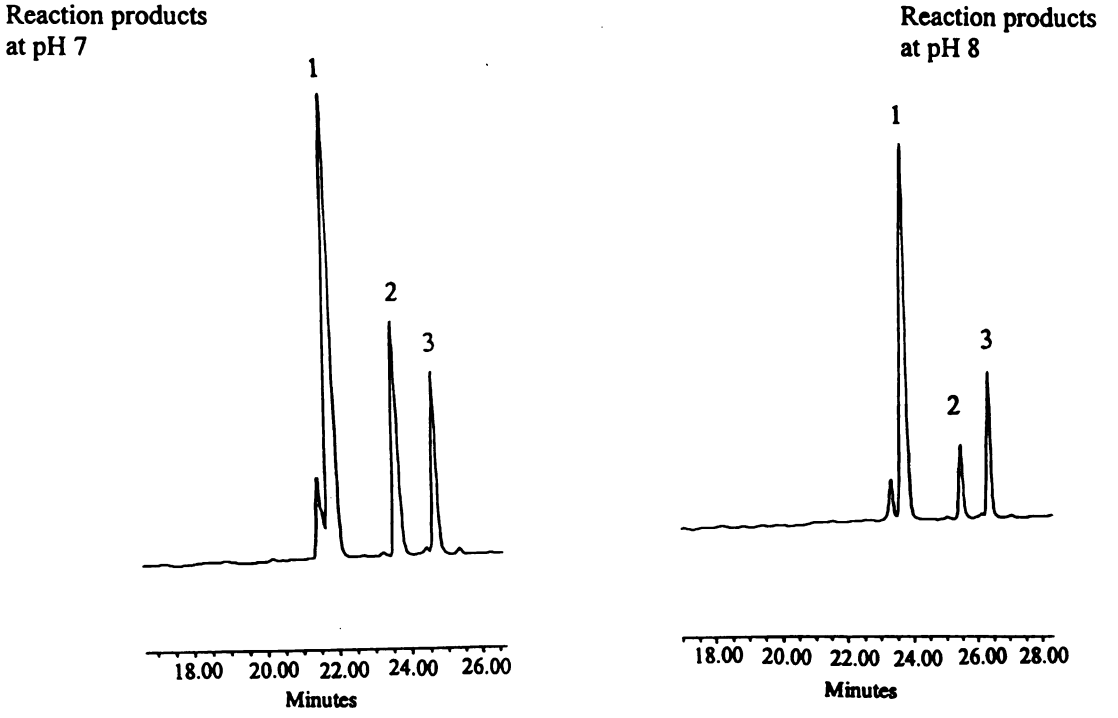


Figure 3.11 HPLC of cyanylated peptide 1657, at pH 7.0 & 8.0, room temperature, 50-fold PYRSCN for 30 minutes.

The following figures show the MALDI spectrum obtained from the compounds represented by each of the three peaks in Figure 3.11.

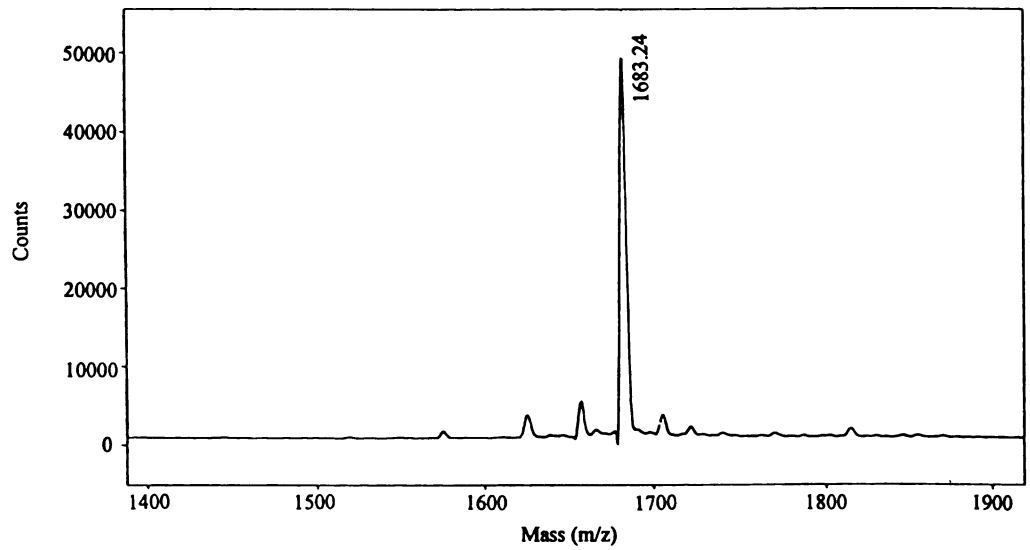


Figure 3.12 (a) MALDI spectrum of the above peptide 1657 HPLC peak 1

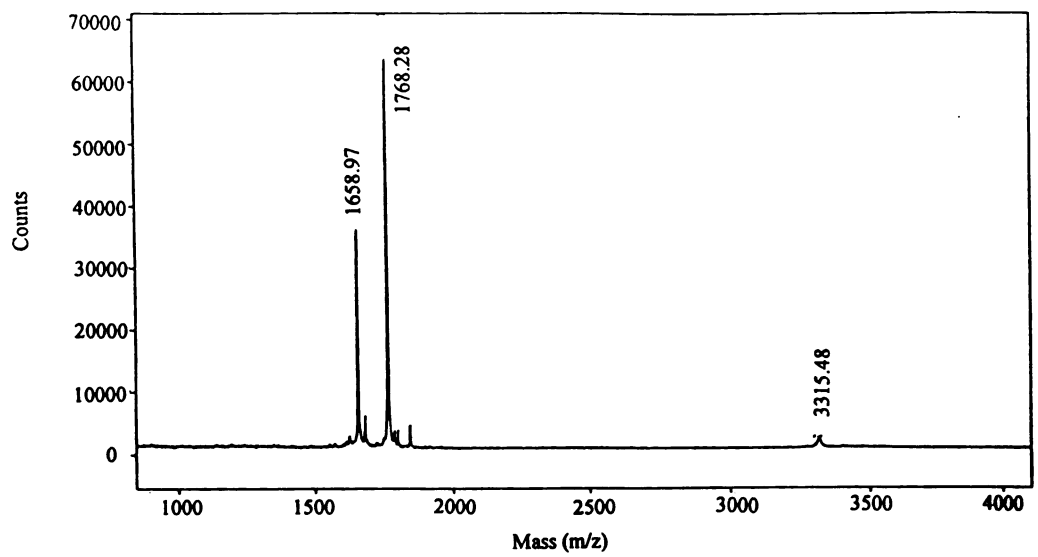


Figure 3.12 (b): MALDI spectrum of the above peptide 1657 HPLC peak 2.

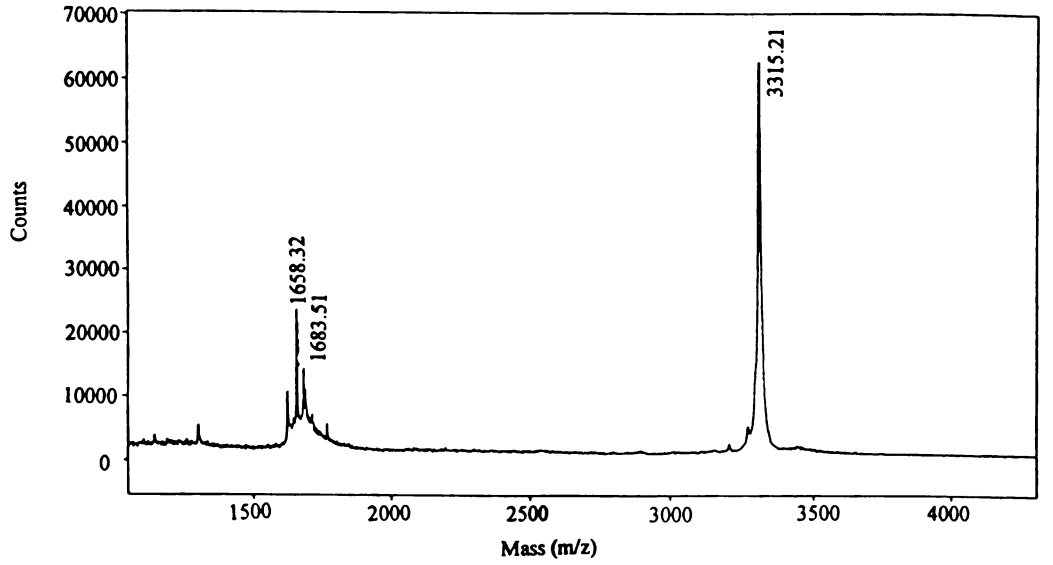


Figure 3.12 (c): MALDI spectrum of the above peptide 1657 HPLC peak 3.

3. Results of Study on Peptide 919:

Table 3.5 Product ratios of peptide 919 after reacting with PYRSCN under pH 7 and pH 8.

pH	Percent of Cyanylated Peptide	Percent of Mixed Product	Percent of Dimer
pH7	50%	33%	17%
pH8	58%	27%	15%

HPLC chromatograms of cyanylation products of peptide 919:

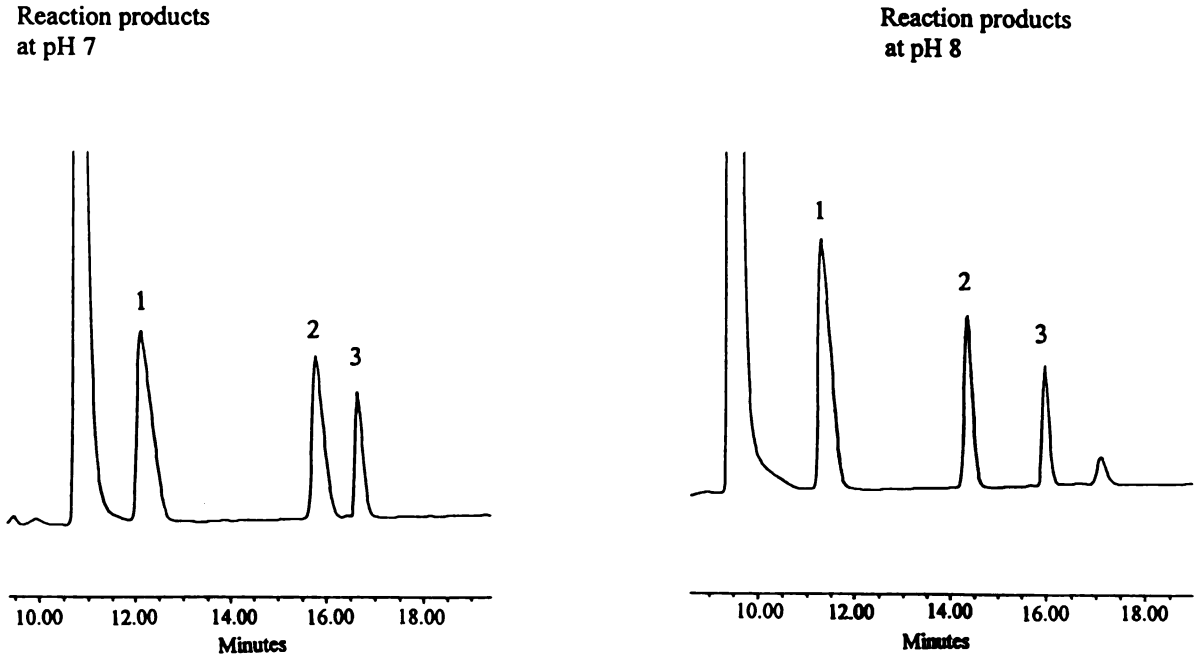


Figure 3.13 HPLC chromatograms of cyanylated peptide 919, at pH 7.0 & 8.0, room temperature, 50-fold PYRSCN for 30 minutes.

The following figures show the MALDI spectrum obtained from the compounds represented by each of the three peaks in Figure 3.13.

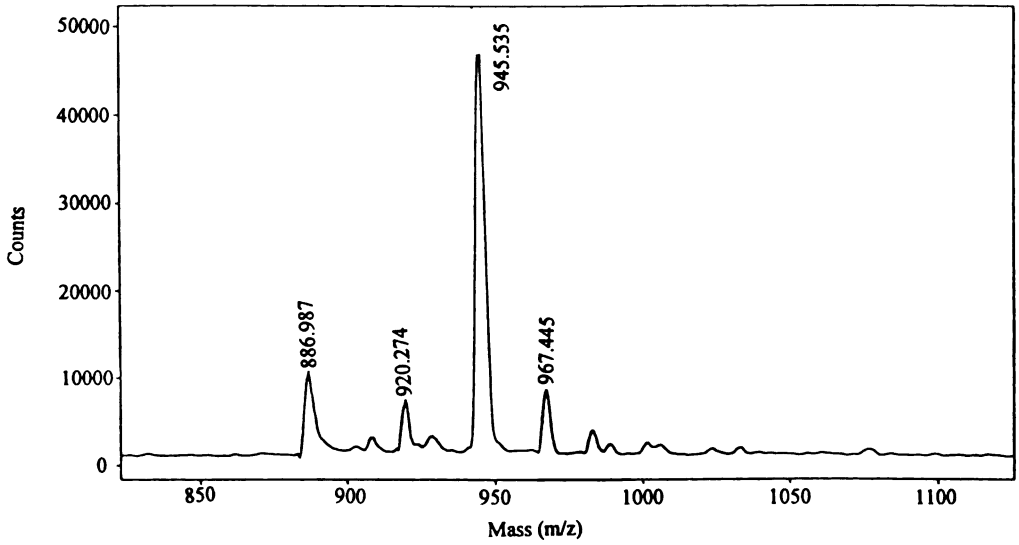


Figure 3.14(a) MALDI spectrum of the above peptide 919 HPLC peak 1

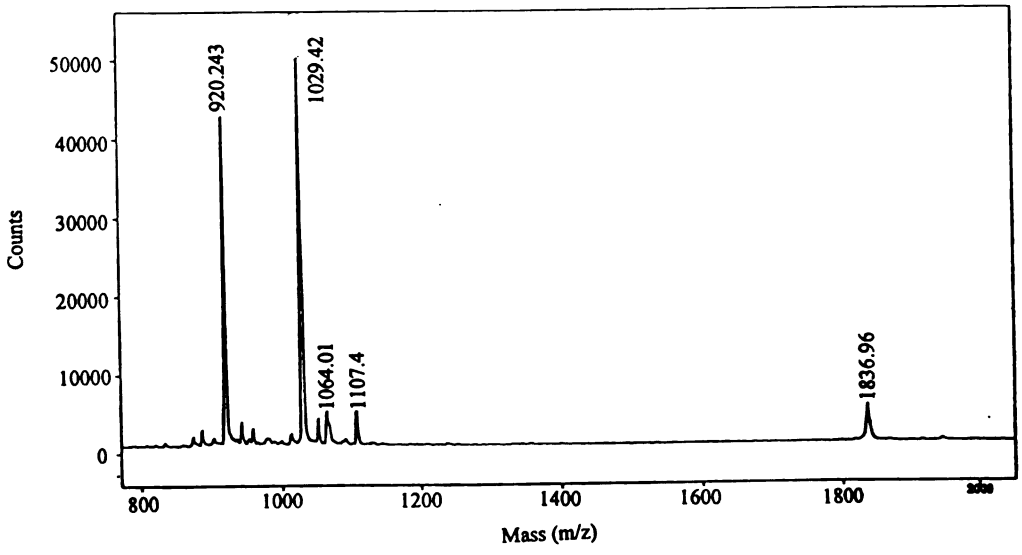


Figure 3.14 (b): MALDI spectrum of the above peptide 919 HPLC peak 2.

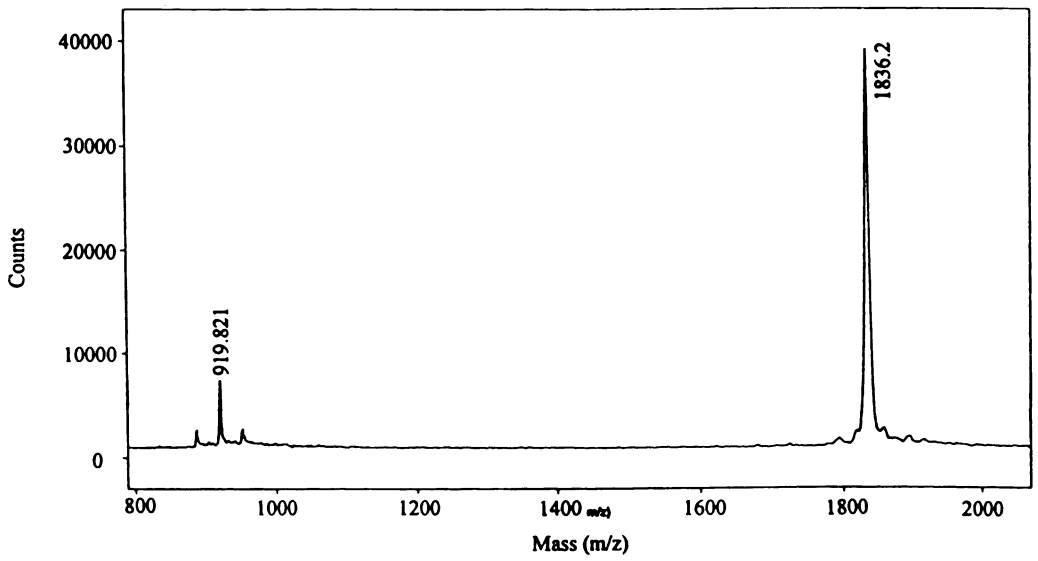


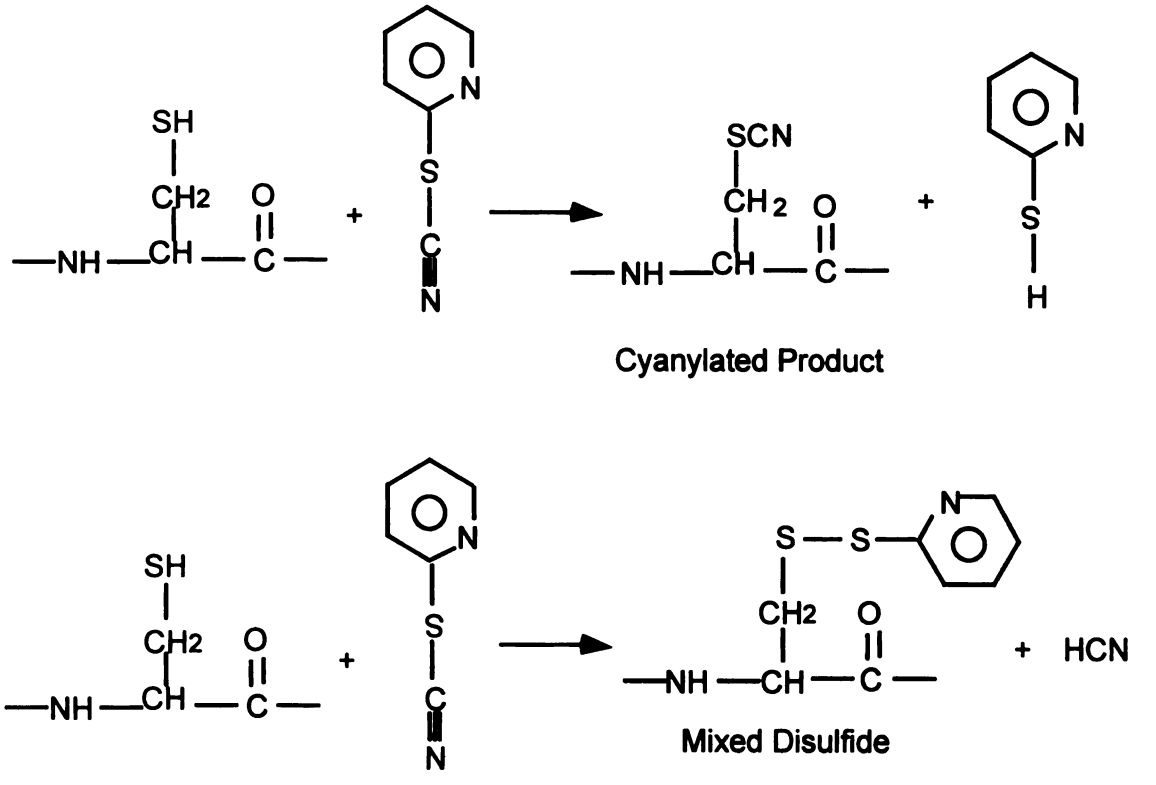
Figure 3.14 (c): MALDI spectrum of the above peptide 919 HPLC peak 3.

From the above data, we can see that for the PYRSCN reagent, as the pH goes from 7 to 8, the mixed disulfide products decrease dramatically for the three peptides studied.

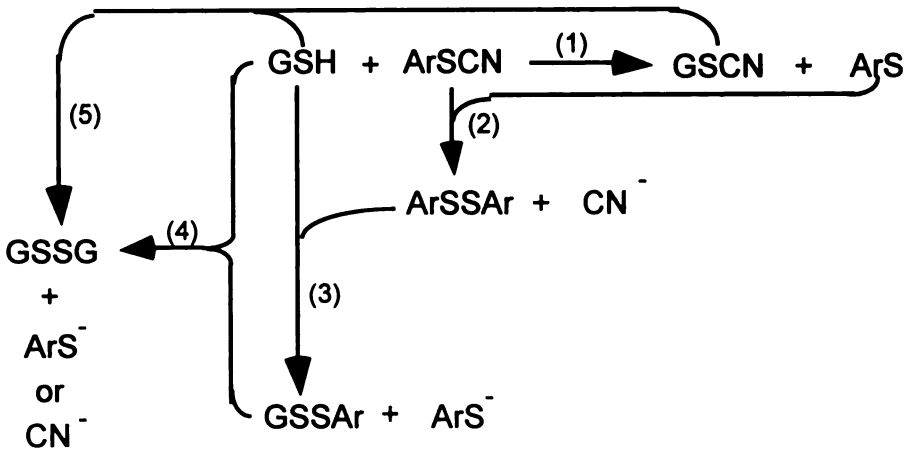
The desired cyanylation product results from the cyanide group on the cyanylation reagent replacing the hydrogen on the sulfhydryl group (from -SH to –SCN), so there is a mass shift of +25 Da for the cyanylated product. The mixed disulfide results from the other side of the cyanylation reagent replacing the hydrogen atom on the sulfhydryl group shown in scheme 3.1, so for our PYRSCN reagent, the mass shift for the mixed disulfide is +109 Da.

Generally after applying the cyanylation reagent on peptides and proteins, three reactions take place, the cyanylation reaction which has a mass shift of +25 Da for any cyanylation reagent, and a side reaction which is called the mixed disulfide reaction as

was discussed above, for PYRSCN reagent it gives a mass shift of + 109 Da, for the 4-dimethylamino phenyl thiocyanate reagent it gives a mass shift of + 151 Da. Another side reaction is the formation of a dimer of the original peptide. The dimer is usually formed under more basic conditions, the thiolate ion reacts with itself and forms a dimer. A dimer will not have much effect on the following cleavage reactions because the only reaction that might take place is the dissociation of the dimer to the original peptide which has no negative effect on our methodology in locating the sulfhydryl or the disulfide bond positions. Hence, this reaction is not described further at the present time.



Scheme 3.1 Formation of cyanylated and mixed product using PYRSCN reagent



Scheme 3.2 Reaction proposed to occur in GSH and NTCB reaction by Degani and Patchornik

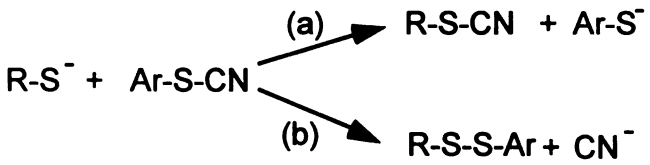
C. Introduction to the reaction pathway:

Some research has been done on the NTCB (2-nitro-5-thiocyanobenzoic acid) reagent by Degani (21). Since the structures of our two reagents are similar to NTCB in that they all have a SCN group attached to a benzene ring, we observed the same mixed disulfide formation reactions with NTCB and the other two reagents. The proposed reaction pathway of NTCB reacting with thiol groups could very possibly be used to explain the cyanylation reactions of PYRSCN and 4-(Dimethylamino)phenyl thiocyanate.

Both Degani (21), and Price (22) observed the mixed disulfide reaction with the NTCB reagent. They found that when reacting NTCB with GSH (glutathione), formation of GSTNB was seen. A proposed sequence of reactions, leading to the formation of both GSTNB (i.e., GSSAr) and GSSG (oxidized glutathione), is shown in Scheme 3.2. According to this scheme of consecutive competitive reactions, ArS^- formed by primary reaction 1, displaces CN^- from still unreacted ArSCN to form ArSSAr (reaction 2). The

aromatic disulfide then reacts with free GSH to form GSSAr (reaction 3) which in turn reacts in reaction 4 with still unreacted GSH to form GSSG. The latter can also arise from GSCN and GSH through reaction 5 (23). All the secondary reactions 2-5 exist independently.

From Price's work in 1976, he also observed the GSSAr formation when reacting NTCB with a free sulfhydryl group, but he only suggested that reaction of 2-nitro-5-thiocyanatobenzoic acid (NTCB) with thiol groups can proceed via two possible pathways, as shown in the following Scheme 3.3. The reacting thiol group is shown in its ionized form.



Scheme 3.3 Two proposed pathways for the reaction of NTCB with GSH by Price

Pathways (a) and (b) represent two competitive reactions; for some peptides and proteins reaction (a) is dominant , and for others reaction (b) is dominant. This means different peptides and proteins will follow different cyanylation reaction pathways.

Degani (21) found that pH had an effect on the reaction rates and hence the percent of the mixed disulfide that is formed except reaction 2, which is pH independent

in this range. This is because all the nucleophiles can act as a leaving group and vise versa and all of the nucleophiles in the system react primarily in their anionic form; their effective concentrations at any given pH will therefore depend on the pKa's of their respective conjugate acids. The pKa's of GSH and HCN are 8.56 (24) and 9.32 (25), respectively, whereas that of ArSH is only about 5.1 (26). At pH 7 and above, ArSH is therefore essentially fully ionized, and hence in this region, the rate of reaction 2 does not depend on the pH.

From our data, we found that pH does have an effect on the reaction of PYRSCN with all the three peptides that we studied. The general trend is as the pH goes higher, we see less side reaction which means less mixed disulfide was formed. The reason could be explained by the above theory. As the pH goes higher, more peptides are in their ionized forms, thus more SH groups are in the S^- form. Since it is the thiolate that attacks the cyanylation reagent as a nucleophile, the more thiloate ion, the faster reaction 1 goes. Under these conditions, less ArSCN is unreacted, so less ArSSAr is formed, and then less mixed disulfide GSSAr is formed. While at pH 7, there is more peptide in its protonated form, so less is reacted with the cyanylation reagent, and thus more ArSCN is unreacted, so more mixed disulfide products are formed. So from our results we can see the better pH for cyanylation of SH group is pH 8, but disulfide scrambling at pH 8 is unaviodable. We used pH 7 for the following experiments.

We can also see from the above data that under same reaction conditions, using the same reagent, a different peptide gives a different proportion of mixed and cyanylated products. Price 's work in 1976 found the same thing. This can only mean that the reaction also depends strongly on the structure of the peptides.

D. Effect of reagent structure on the cyanylation reaction:

The cyanylation reaction was conducted using both reagents PYRSCN and 4-(dimethylamino) phenyl thiocyanate at pH 7. The percentage of each peak is compared in Table 3.6.

Results of Study on Peptide 2428:

Table 3.6 Product rations of peptide 2428 after reacting with PYRSCN and 4-(dimethylamino) phenyl thiocyanate at pH 7.

pH 7	% of Cyanylated Peptide	% of Mixed Product	% of Dimer
PYRSCN	50%	43%	7%
Dimethyl reagent	66%	23%	11%

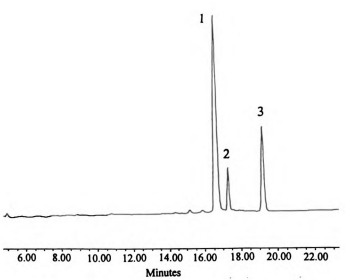


Figure 3.15 HPLC chromatogram of the cyanylation product of peptide 2428 reacted with 50-fold molar excess of 4-(dimethylamino) phenyl thiocyanate at pH 7, room temperature for 30 minutes.

Unlike the cyanylation reaction with PYRSCN reagent (Figure 3.14), the mixed disulfide for 4-(dimethylamino) phenyl thiocyanate is eluted later than the dimer, so now peak 2 is the dimer peak, while peak 3 is the mixed disulfide peak.

From this experiment we found that a different reagent structure (PYRSCN and 4-dimethylamino phenyl thiocyanate) also has an effect on the cyanylation reaction and it can be explained by the reaction chain theory.

The reaction chain has some noteworthy properties (21) as described above that all nucleophiles can act as leaving groups and visa versa. The order of decreasing nucleophilicity is GSH > CN^- > ArS^-, because the alkyl group is an electron donating group, while an aromatic ring can stabilize the electrons. So when a free sulfhydryl is attached to an alkyl group, the thiolate is a better nucleophile than when it is attached to an aromatic ring. Electrons are better stabilized in the ring, so the GSH is a stronger nucleophile than ArSH. So the order of increasing leaving groups is just the opposite: ArS^- > CN^- > GSH. From this we can also see that CN^- is thus both a moderate nucleophile and a moderate leaving group. We can deduce that the reaction rate for any reaction in the system depends on the specific combination of a nucleophile and a leaving group, those reactions involving extreme combinations being the fastest ones. Thus a sulfur substituent on a *p*-nitro-phenyl ring is more electrophilic than the sulfur atom of the aliphatic side chain; k3 is therefore larger than k4 although both reaction 3 and 4 involve the same nucleophile (GSH) and the same leaving group (ArS^-). A similar difference in electrophilicity (or thiophilicity) between carbon and sulfur may account for the still lesser reactivity of ArSCN compared to that of either ArSSAr or GSSAr. Also, slight conjugation effects within the aromatic thiocyanate molecule (27,28) would

tend to somewhat stabilize it. On the other hand, the symmetry of ArSSAr may further contribute to the reactivity of this disulfide, which is the highest in the system.

We assessed two cyanylation reagents with model peptide 2428. Both experiments were conducted under the same reaction conditions at pH 7. The 4-(dimethylamino) phenyl thiocyanate reagent has much less mixed disulfide formed than the PYRSCN reagent. Since the reaction conditions are exactly the same, so it must be the different structure of the reagent that makes a difference. When reacted, 4-(dimethylamino) phenyl thiocyanate has less electrophilicity than the PYRSCN reagent. Because dimethyl amino is an electron donating group, the sulfur atom attached to the dimethyl amino phenyl ring is more nucleophilic, thus less electrophilic. On the other hand, the sulfur atom of the PYRSCN reagent is attached to a pyridine ring; the nitrogen atom on the pyridine ring withdraws electron from the sulfur atom, thus it is electron poor, so it has greater electrophilicity, and less nucleophilicity. Thus, when a thiol group reacts with these two cyanylation reagents, since the thiolate ion is a stronger nucleophile than the ArSCN, so it will attack the cyanylation reagent, also because ArS^- is a better leaving group than CN^- as was described above, so GSCN will be formed as indicated in Scheme 3.2 reaction 1. For the mixed disulfide reaction, because $PYRS^-$ has a greater electrophilicity than 4-(dmethylamino) phenyl thiloate ion, so reaction 3 in Scheme 3.2 with PYRSCN reagent tend to be easier , thus more mixed disulfide is formed when using PYRSCN reagent. So the 4-(dimethylamino) phenyl thiocyanate reagent is a better cyanylation reagent than PYRSCN when reacting with peptide 2428.

From the above discussion, an electron withdrawing group on the benzene ring of the cyanylation reagent seems to cause more mixed disulfide to be formed, but the

function of different substituents on cyanylation reaction 1 which is the main reaction still needs further investigation. Since this is a reaction chain, each reaction affects the other, so the interaction among all these reactions must be considered.

IV. Summary:

All four reagents can give the cyanylation reactions as we expected, but in different pH ranges. Experiments were done by conducting the cyanylation reaction under specific pH conditions. Samples of the reaction mixture were dried and pH was adjusted to acidic conditions using citrate buffer, then analyzed by MALDI-MS.

For comparison, a 50-fold molar excess of NTCB was used at pH 8-9, 37C for 30 minutes, and a single M+25 peak was observed for peptide 919, peptide 1657 and peptide 2428, while on peptide 2428 we saw some cleavage products. For all three peptides, we saw a trace of β-elimination products as was observed by others (21).

The CDAP reagent can be used at pH 3-4. Only 10-fold molar excess of the reagent at pH 3 for 15 minutes is enough to give a complete cyanylated product for all the peptides tried. Since the reaction is completed at very acidic pH, disulfide bond scrambling should be surpressed, and no β-elimination and cleavage reaction can take place.

PYRSCN was first tried at pH 3, because from its stable structure we expected it to be reactive in a wide pH range from acidic to basic. Later on, we found under very acidic pH there is little cyanylation reaction; however as pH was increased, the cyanylation reactions took place. At pH 6, there is already some cyanylated product,

while at pH 7 more is cyanylated. Since under acidic pH disulfide scrambling could be largely prevented, we use pH 6.5-7 as our reaction condition; a 50-fold molar excess of reagent was applied at room temperature for 30 minutes. The cyanylation reaction was observed on all three peptides studied, together with some β-elimination and a little bit of cleavage.

4-(dimethylamino) phenyl thiocyanate was also first tried at acidic pH, but not much cyanylation reaction was observed; however, at increasing pH (up to 8), more and more cyanylated products were observed. At pH 7, there is already a significant amount of cyanylated product formed. When using 4-(dimethylamino) phenyl thiocyanate reagent to cyanylate cysteine, we also observed mixed disulfide and dimer peaks as we did in the PYRSCN and NTCB cyanylation reactions. We found that the pH of the reaction, peptide structure and substituents on the benzene ring of the cyanylation reagent all have an effect on the proportion of cyanylated or mixed products that were formed. Further research needs to be done to investigate the effect of substituents on the cyanylation reaction.

From the above we can conclude that NTCB, PYRSCN, and 4-(dimethylamino) phenyl thiocyanate have similar structures, all have SCN groups and benzene rings; these are designated a series of SCN reagents. These SCN reagents can only be effective for cyanylation under slightly acidic to basic pH. PYRSCN and 4-(dimethylamino) phenyl thiocyanate have advantages over NTCB in that they can be used from pH 6-7, while NTCB can only be used above pH 8, so disulfide scrambling can be prevented to some degree when using PYRSCN and 4-(dimethylamino) phenyl thiocyanate. Among SCN

reagents, mixed disulfide reactions were reported for all of them (21); the effect of this mixed disulfide on the following cleavage reaction still needs more study. CDAP is the only reagent so far that can be used under low pH; reaction of CDAP with –SH is fast and specific, but CDAP is very unstable, so for some difficult peptide or proteins with tight structures that can only be cyanylated under more basic pH, we might need to try the other two reagents. More work still needs to be done to test the potency of the two new reagents.

So, for peptides and proteins that need to be cyanylated under relatively high pH, such as pH 6-pH 7, then the two new reagents proposed have the advantages over both CDAP and NTCB.

V. Future work

So far, it seems that CDAP is still one of the best choices for the cyanylation reaction. Looking at its structure (Figure 3.16), we can see it is a salt, and the positive charge is on the nitrogen atom directly attached to the CN, and since it only has a CN group, so there is no mixed disulfide reaction, thus we can avoid this side reaction which seems to exist in all SCN reagents. Also because the positive charge is close to the CN group, it makes the CN group have greater electrophilicity toward the thiolate ion than the CN group on the SCN reagents. From the reaction conditions for both the CDAP and NTCB, we can see it only takes 10 to 15 minutes for the CDAP to react, while it takes 30 minutes for most of the SCN reagents to react. So a reagent with only a CN group, but not a SCN group can avoid the mixed disulfide reaction , and also it might react much faster than the SCN reagents.

The new reagents that we proposed are a series of imides such as N-cyanoimide or N-cyanophthalimide (Figure 3.16) (29).

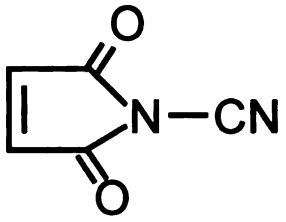


Figure 3.16 Structure of N-cyanoimide salt

These reagents are like CDAP; the CN group is attached to a nitrogen atom and at the ortho position on the ring there are electron withdrawing groups which will make the CN group more electropositive, so the cyanylation reaction might take place easily. Unfortunately, these reagents are not commercially available now, so more research needs to be done in this area. Also, the potency of the PYRSCN reagent still needs to be investigated on more proteins and peptides.

VI. References:

1. Torchinsky, Y. M., in *Sulfur in Protein*; (Metzler, D., Translation Ed.), Pergamon, Oxford, (1981)
2. Boyer, P. D., *J. Am. Chem. Soc.*, **76**, 4331 (1954)
3. Hellerman, L., *Physiol. Revs.*, **17**, 454 (1937)
4. Kochkina, V. M., and Torchinsky Y. M., *Biochem. Biophys. Res. Communs.*, **63**, 393 (1975)
5. Sakai, H., *Anal. Biochem.*, **26**, 387 (1968)
6. Yamuchi, T., Yamamoto, S., and Hayaishi, O., *J. Biol. Chem.*, **250**, 7127 (1975)
7. Gibson, Q. H., *J. Biol. Chem.*, **248**, 1281 (1973)
8. Ellman, G. L., *Arch. Biochem. Biophys.*, **82**, 70 (1959)
9. Vanaman, T. C., and Stark, G. R., *J. Biol. Chem.*, **245**, 3565 (1970)
10. Catsimpoolas, N., and Wood, J. L., *J. Biol. Chem.*, **241**, 1790 (1966)
11. Grassetti, D. R., and Murray, J. F., *Arch. Biochem. Biophys.*, **119**, 41 (1967)
12. Grassetti, D. R., and Murray, J. F., J. Chromatogr. **41**, 121 (1969)
13. Stuchbury, T., Shipton, M., Norris, R., Malthouse, J. P. G., Brocklehurst, K., Herbert, J. A. L., and Suschitzky, H., *Biochem. J.*, **151**, 417 (1975)
14. Jacobson, G. R., Schaffer, M. H., Stark, G. R., and Vanaman, T. C., *J. Biol. Chem.*, **248**, 6583 (1973)
15. Catsimpoolas, N., and Wood, J. L., *J. Biol. Chem.*, **241**, 1790 (1966)
16. Spande, T. F., Witkop, B., Degani, Y., and Patchornik, A., *Advan. Protein Chem.*, **24**, 98 (1970)

17. Wakselman, M., Guibe-Jampel, E., Raoult, A., and Busse, W. D., *J. Chem. Soc. Chem. Commun.*, **1**, 21 (1976)

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

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

20. Kottke, K., Friedrich, F., and Pohloudek-Fabini, R., *Arch. Pharm.*, **300**(7), 583 (1967)

21. Degani, Y., and Patchornik, A., *Biochemistry*, **13**, 1 (1974)

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

23. Aldrige, W. N., *Biochem. J.*, **48**, 271 (1951)

24. Friedman, M., Cavins, J. F., and Wall, J. S., *J. Am. Chem. Soc.*, **87**, 3672 (1965)

25. Britton, H. T. S., and Robinson, R. A., *Trans. Faraday Soc.*, **28**, 531 (1948)

26. Ellman, G. L., *Arch. Biochem. Biophys.*, **74**, 443 (1958)

27. Bordwell, F. G., and Boutan, P. J., *J. Am. Chem. Soc.*, **78**, 851 (1956)

28. Campbell, T. W., and Rogers, M. T., *J. Am. Chem. Soc.*, **70**, 1029 (1948)

29. Stephens, R. W., and Domeier, L., *Synthetic Commun.*, **21**, 2025 (1991)

MICHIGAN STATE UNIV. LIBRARIES
31293017665112