INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

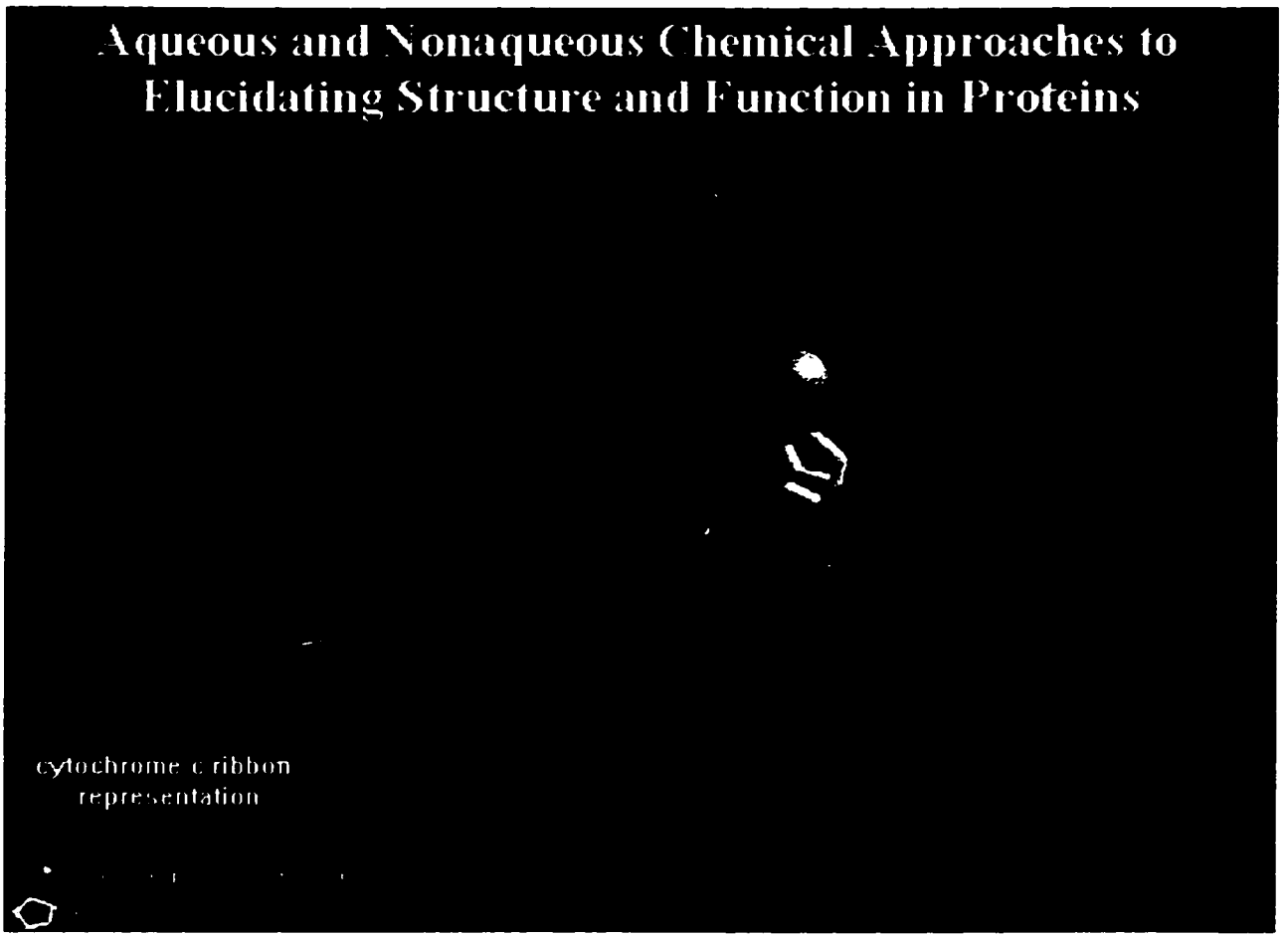
Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Université d'Ottawa · University of Ottawa

Aqueous and Nonaqueous Chemical Approaches to Elucidating Structure and Function in Proteins



cytochrome c ribbon representation

ABSTRACT

Aqueous and Nonaqueous Chemical Approaches to Elucidating Structure and Function in Proteins

Three **chemical approaches** were used to study proteins:

Chemical modification with iodomethane: A novel strategy was developed in which proteins were simultaneously reacted with [^{13}C] and [^{14}C]iodomethane to permit the identification of reactive groups by [^{13}C]NMR spectroscopy and the isolation of [^{14}C]labeled peptides containing individual methylated functional groups by autoradiography. Reaction with iodomethane N,N,N-trimethylated Nε-lysyl and N-terminal Nα-amino groups, N^1,N^3-dimethylated imidazole functions of histidyl residues, O-methylated phenolic hydroxyl functions of tyrosyl residues and S-methylated sulphide functions of methionyl residues. To our knowledge, none of the other derivatives, aside from the methionyl derivative, have been reported previously as *in vitro* chemical modifications of native proteins. The dimethylhistidyl and O-methyltyrosyl derivatives in particular, have not been reported as either *in vitro* or *in vivo* chemical modifications of protein. The identification of various classes of reacted functional groups and their corresponding amino acid residues in protein was facilitated by the observation that the ^{13}C-resonance positions of derivatives were relatively unaffected by the aqueous media and microenvironment. The methylation could be made selective for specific residues by the use of reversible blocking reagents and by adjusting reaction conditions.

Nonaqueous chemical modification techniques: Chemical modification of lyophilized proteins in nonaqueous environments was carried out by either dispersing protein in octane and reacting with dissolved reagent, or reacting protein *in vacuo* with a volatile reagent. The reaction of iodomethane with protein functional groups in nonaqueous environments and the derivatives formed paralleled those of aqueous reactions and were quantified by solution and solid state [^{13}C]NMR techniques. Reacting ethoxyformic anhydride or acetic anhydride with protein functional groups afforded acyl derivatives of Nα and Nε-amino groups and mixed anhydrides of side-chain carboxyl groups. These mixed anhydrides were stable in the absence of water and enabled the coupling of protein to various nucleophiles. The extent of derivatization of functional groups of lyophilized protein was directly related to the pK_a of the reactive group, pH of the solution from which the protein was lyophilized, and extent of surface exposure of functional groups under native conditions. The results suggest that the various physico-chemical factors which govern the reactivity of functional groups in nonaqueous environments depend on the protein solution structure prior to lyophilization. Protein modifications in a nonaqueous environment were carried out with no hydrolytic breakdown of protein, much greater economy of reagents, and permitted the derivatization of nanomole quantities of protein. Carboxyl groups of α-chymotrypsin, in particular, were amidated in high yields for the purpose of C-terminal analysis.

Enzyme kinetics: Five synthetic substrates containing different amino acid residues at the P_3 position (acetyl-X-Arg-Arg-AMC, where X is Gly, Glu, Arg, Val and Tyr and where AMC represents 7-amido-4-methylcoumarin) were used to investigate the S_3 subsite specificity of cathepsin B. At pH 6.0, the specificity constant, k_{cat}/K_m, for tripeptide substrate hydrolysis was observed to increase in the order Glu < Gly < Arg < Val < Tyr. Molecular modeling studies of substrates containing a P_3 Glu, Arg or Tyr covalently bound as the tetrahedral intermediate to cathepsin B suggest that enzyme specificity for a P_3 Tyr group is due to a favourable aromatic-aromatic interaction with Tyr^{75} on the enzyme as well as a possible hydrogen bond between the P_3 Tyr hydroxyl and the side-chain carboxyl of Asp^{69}.

To my family, past, present and future....

Acknowledgments

I would like to thank my supervisor Dr. Harvey Kaplan of the Department of Chemistry, University of Ottawa, for providing me with constant guidance and latitude during my studies. Dr. Kaplan was inspiring in his approach to problem solving and in his abilities to connect seemingly unrelated observations. The overall experience was very challenging, enjoyable and was carried out in an ideal working environment which nurtured independent thought and teamwork. Dr. Kaplan was actively involved and interested in the activities of every member of our research group. I always appreciated the words he would use to encourage me at those discouraging moments, his teachings of how to effectively play "the game of science", and the time he would take to share his insight about things that have nothing to do with chemistry at all. In a most sincere way, I feel extremely proud, appreciative and privileged to have worked in his laboratory, and I will always keep Dr. Kaplan's example as my indication of how a research supervisor should conduct himself. He has continued to show us in many ways that he always keeps his student's best interests close at heart.

I would like to thank my co-supervisor, Dr. Sadiq Hasnain, for giving me the opportunity to work at the National Research Council laboratories on an extremely interesting and multidisciplinary project, which comprised a portion of the material in this thesis. While I was his first graduate student, Dr. Hasnain did not hesitate to instill into me the optimism, enthusiasm, dedication and pride that propels this type of work. He always gave effortlessly of himself during our time together and after transferring into another position continued to assist me through various stages in the development of my program. Dr. Hasnain taught me how to see the positive in literally any situation. Like Dr. Kaplan, I feel extremely fortunate, proud and privileged to have worked with Dr. Hasnain. I must also mention the excellent help that Mrs. Tomoko Hirama provided and cannot understate the impact that it had on my progress.

My labmates were always there to support, humour and stimulate. I was fortunate to have learned important lessons of life from each of them. In particular, a very special thanks is necessary to my better half and sweetheart, Helen Vakos, for her smile, for making everything seem brighter in her presence, for her support, concern and help and her genuine love and loyalty. Our time together has been a wonderful adventure which I hope has been as fulfilling for her as it has been for me. From the bottom of my heart I wish to thank François Clairmont, Barry Lorbetskie, Nicolas Stewart, a most recent member of our group, Dr. Steve Lang, and a former member, Dr. Henri Bietlot. Their contributions in every manner were very much appreciated.

I am very grateful to Dr. Glenn Facey and Dr. Steve Lang for reviewing and explaining the principles of NMR theory and for taking it upon themselves to remind me on many occasions that I should write my thesis. Their support and humour made the writing process much easier. In particular, I wanted to thank Dr. Facey for giving me substantial latitude with use of the NMR instrument. I also wish to express my appreciation to Dr. R. R. Fraser for his concern and advice.

Last but not least, I cannot thank enough my mother and father, and my brother, Tarkan, for their incessant and considerable support throughout my Ph.D. program, and indeed, throughout my life. The impact of their nurturing and dedication becomes more apparent with each passing year and I truly feel blessed for having such a family.

TABLE OF CONTENTS

chymotrypsin ribbon representation

Chapter 1: Aqueous and nonaqueous chemical approaches to elucidating structure and function in proteins

1.1 Overview of protein chemistry.

Proteins comprise one facet of some very complicated systems in biological tissues. Their varing sizes and the precise three-dimensional arrangement of their groups are as interesting to study as their many functional roles. With infinite directions and topics to explore, it is not surprising that the protein chemist has diversified into many areas. Despite significant advances in techniques, routine investigations are often hampered by problems, attributed to the complexity of protein systems. Understanding relationships between structural features of proteins and their biological functions is improving, however, and the conduct of such comprehensive work has demanded that the protein chemist use a broad scope of approaches to address and help elucidate an equally broad range of interrelated topics.

Applications of protein chemistry are observed daily. All manner of protein products are readily available and used worldwide. It is even a fact that the merchandizing of protein products initiated what was perhaps the oldest form of practiced protein chemistry, namely, the tanning of leather by covalent chemical modification (Means and Feeney, 1971). Protein chemistry has been used in a myriad of other dyestuff and textile applications to date. Protein chemists are also very much involved with various research aspects of pharmaceutical and biological companies (Hashida *et al.*, 1994). Needless to say, the importance of protein chemistry and its applications cannot be understated.

The focus of protein chemistry is to promote the advancement and exploitation of the protein knowledge base for the benefit of mankind. As implied, protein chemistry is the use of physical and chemical techniques to explore protein systems (Means and Feeney, 1971; Darbre, 1986;

Hames *et al.*, 1990; Glasel *et al.*, 1995). An example of a pure structural study could be an X-ray structure analysis (Boyer, 1971; Wyckoff *et al.*, 1985a; Wyckoff *et al.*, 1985b; Darbre, 1986; Glasel *et al.*, 1995). Necessarily, the investigation provides only static information. An example of a pure functional study could be a bioassay, which provides insight into the function of a protein without necessarily providing structural information. Protein analyses are very often a weighted average of structural analyses and functional analyses, and there are many such examples of composite studies. In fact, the observation that structure affects function and function depends on structure serves as a constant reminder that one should consider both factors when carrying out protein related experiments. Inasmuch as composite studies are complicated, one advantage of pursuing such investigations is that any outcome tends to have more interesting repercussions, as it provides insight into both static and dynamic components. Understanding the properties of proteins and the relationship between structure and function is currently a subject of intensive study and is the key to promote and develop applied protein and enzyme technologies (Means and Feeney, 1971). Experiments at the heart of protein chemistry probe the interdependence of structure and function in proteins and are often referred to as structure-function studies (Boyer, 1970).

1.2 Historical aspects in the evolution of protein chemistry.

Modern protein chemistry did not develop according to a set plan but evolved, like many other disciplines, in accordance to the requirements of the times. In early investigations, little was known about proteins and experimental approaches tended to be customized out of necessity. Pioneering protein chemists such as Fischer investigated the properties of individual amino acids and the synthesis and properties of small peptides (Means and Feeney, 1971). In the 1920's it

was shown by Sumner that enzymes were in fact a special class of catalytic proteins and this prompted investigators to identify the protein residues responsible for catalysis. Enzyme kinetics became the established method to probe enzyme function (Cornish-Bowden and Wharton, 1988; Creighton, 1993; Purich, 1995). Apart from enzymes, whose function could readily be characterized by kinetics, initial investigations focused on protein primary structure. Gradually, it became apparent that higher orders of structure were equally important to study. The war effort of the 1940's opened the door to the development of protein chemical modification techniques where both nonspecific modifications and specific labeling procedures were employed (Olcott and Fraenkel-Conrat, 1947; Herriott, 1947; Balls and Jensen, 1952).

The diversification of protein chemical modifications was accelerated when existing physico-chemical methods such as X-ray diffraction and fluorescence techniques were adopted for the analysis of protein (Means and Feeney, 1971). Both methods required that certain chemical modifications of the protein be carried out to achieve success. In X-ray crystallography, heavy atom replacements of protein crystals were required to resolve the phase problem (Means and Feeney, 1971; Richards and Wyckoff, 1971; Carter and Sweet, 1997), while covalent fluorescent labeling techniques of protein were essential to improve protein fluorescence (Hirs, 1967; Means and Feeney, 1971). Improvements in physical methods of protein investigation, such as in the two above examples, often depended on the use of chemical approaches as in the case of chemical modification. In short, a physical approach often relied on a chemical technique to clinche the analysis, and vice versa. This complimentarity among chemical and physical methods of analysis is a recurring theme, and in this regard is similar to the interdependence observed of protein structure and function.

In yet another stage of advances, chemical and physical methods, and particularly the spectroscopies, were developed or made more sensitive. One analytical technique, namely amino acid sequence analysis, was eventually automated and represented one of the most significant improvements for the analysis of proteins in all fields of protein chemistry since it made readily available the primary sequence of proteins (Spackman *et al.*, 1958). At this stage of development, protein chemists were reaping the benefits of an extensive knowledge base and improved instrumentation, and could address more intricate details of structure and function of protein. While their investigations may have been more stringent than those of their predecessors, their existence was absolutely dependent on information provided by preliminary studies.

In the 1980's there were major advances in the field of molecular biology. It finally became practical to deduce a protein sequence indirectly, by sequencing the corresponding gene oligonucleotide sequence. This alternative represented a far more rapid protocol than conventional protein chemical methods. In addition to advances in deoxyribonucleic acid sequencing methods, chemical syntheses of deoxyribonucleic acid primers and the use of restriction enzymes was also introduced (Stryer, 1988). These three advances provided all the necessary tools for the development of site-directed mutagenesis, and molecular biologists could finally introduce conservative and specific changes to protein primary structure. The development of site-directed mutagenesis superceded other techniques of protein chemistry, and appeared to detract from the protein chemist's ability to make a significant contribution to structure-function studies using the wild-type protein as starting material. In fact, so great was the anticipated potential of molecular biology that the upcoming demise of various areas of protein chemistry was postulated by Malcolm (1978). In hindsight, the prediction was unfounded because of the failure

of gene nucleotide methods to account for proteolytic processing and other post-translational protein modifications, and because of the ever growing need for protein characterization in topics spanning all theoretical and practical fields.

1.3 The importance of structure-function studies

It is a well established fact that the properties of functional groups are influenced by their environment (Allinger *et al.*, 1976; Laidler and Meiser, 1982; March, 1988). The ionization of carboxylic acids and alcohols, for example, follows a dependency on the dielectric strength of the medium. There is a myriad of similar examples depicting the effect that environment has on intrinsic functional group properties. This observation is the driving rationale behind structure-function investigations in proteins, as no where is this effect more pronounced than in the observed properties of protein functional groups (Fersht, 1985; Creighton, 1993).

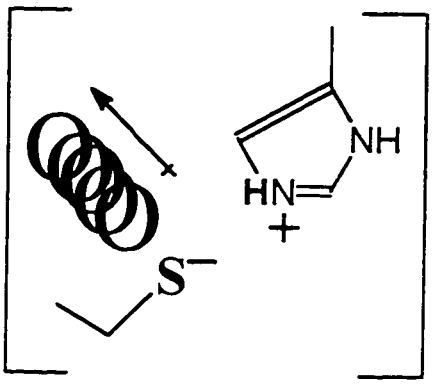


Figure 1.3.1: Schematic of papain active site thiol residue, proximal imidazole residue and α-helix. The positive charge of the imidazole and positive end of the dipole moment enhances the acidity of the thiol.

One particularly striking example (Figure 1.3.1) demonstrates that enzyme structure can significantly perturb the properties of a functional group by imposing an unusual local environment. The pK_a value of a thiol group is generally in the range of 8 to 10 (March, 1988). In the active site of papain, however, the reported pK_a value of the active site thiol of Cys^{25} is approximately 6 pK_a units less than expected, and represents a one million-fold enhancement of acidity (Creighton, 1993).

1.4 Tools of structure-function analyses.

It is very difficult to differentiate the relative importance of different techniques in protein chemistry as they each have their merits and are used collectively. Many volumes of Methods in Enzymology have been devoted to just this area. Protein chemists rely on information obtained by physical methods of analysis which are usually spectroscopic measurements (Wyckoff *et al.*, 1985; Hirs and Timasheff, 1986a,b; James and Oppenheimer, 1994; Sauer, 1995; Glaser, 1995), and chemical methods which are usually covalent modification studies (Means and Feeney, 1971; Lundblad and Noyes, 1984; Imoto and Yamada, 1989). Without attempting to generalize, physical methods tend to lend themselves well to structural analysis while chemical methods are better suited as a probe of structure and function (*vide infra*).

1.4.1 Examples of physical methods and their application to structure analyses. Modern experimentalists are often concerned with elucidating structure-function relationships, however, pure structural analyses of proteins have persisted. Proteins have up to four orders of structure, namely primary, secondary, tertiary and quaternary structures (Creighton, 1993). There are simply too many methods of protein analysis to list. Classical physical methods include X-ray diffraction techniques (Hirs, 1967; Boyer, 1970; Wyckoff *et al.*, 1985; Darbre, 1986; Glasel, 1995), infrared, ultraviolet and Raman spectroscopies, optical rotary dispersion and circular dichroism spectroscopies (Hirs and Timasheff, 1986a; Sauer, 1995), microcalorimetry (Hirs and Timasheff, 1986b), electrophoresis (Leggett Bailey, 1962; Hirs and Timasheff, 1972; Hames and Rickwood, 1990; Karger and Hancock, 1996a,b), hydrodynamic techniques (Hirs and Timasheff, 1985;

Glasel, 1995) and numerous chromatographies (Hirs, 1967; Hirs and Timasheff, 1972), to name a few. Modern techniques have added the use of neutron diffraction (Glasel, 1995), high field nuclear magnetic resonance for both solution (Wüthrich, 1986; Hirs and Timasheff, 1986b; James and Oppenheimer, 1994) and solid state analyses (Hirs and Timasheff, 1986b; Auger, 1995; Watts *et al.*, 1995; Palmer *et al.*, 1996), Mössbauer (Glasel, 1995; Sauer, 1995), electron pulse (Sauer, 1995) and specialized mass spectrometries (Darbre, 1986; McCloskey, 1990; Biemann, 1992; Glasel, 1995), and in depth computer modeling (Johnson and Brand, 1994) to the list of common approaches. While many techniques are available, the information provided by any one method requires complementation by other techniques. A handful of techniques and their applications are listed below.

Light spectroscopies. In each technique, light is perturbed when it is passed through a protein solution. The change in light property is quantified and interpreted as a measure of some aspect of the protein. Infrared, Raman, circular dichroism and optical rotary dispersion spectroscopies have been useful in elucidating the type and composition of secondary stucture elements of protein (Choma *et al.*, 1990), and ultraviolet spectroscopy has been used in many applications ranging from simple quantification, to probing the folding pathway of barnase (Serrano *et al.* 1992).

Column chromatography and electrophoresis. Some protocols invariably require the use of columns to purify protein and derivatives (Hirs and Timasheff, 1972; Darbre, 1986; Karger and Hancock, 1996a,b). Classically, resins have separated protein on the basis of size or charge

(Leggett Bailey, 1962). A more subtle separation technique uses the principle of isoelectric focusing, and purifies protein along a pH gradient (Hames, 1990). Affinity chromatography, another extremely powerful tool of protein purification (Burgess, 1987), separates proteins on the basis of specific binding interactions (Darbre, 1986). It has been used to provide protein for sequence analysis and other purposes and has been adopted for affinity electrophoresis (Hames, 1990).

Paper chromatography and electrophoresis. Classic paper electrophoresis and peptide mapping techniques have been used to detect changes in protein primary structure following covalent modification and fragmentation (Leggett Bailey, 1962; Darbre, 1986; Karger and Hancock, 1996a,b). Detection of the N→O acyl migration is just one of many examples which show the utility of these methods (Means and Feeney, 1971). Various modification reagents and enzyme strategies were necessarily part of the protocol (Boyer, 1970; Boyer, 1971; Darbre, 1986).

Microcalorimetry measurements. Microcalorimetry (Hirs and Timasheff, 1986b) has been used to measure unfolding enthalpies of protein. The technique was not available for the analysis of proteins until technical problems of downsizing were addressed. While simple in principle, the enthalpic profiles obtained are often rich with information. For example, in the case of the insecticidal protein derived from *Bacillus thuringiensis*, it was deduced that the toxic moiety of the protoxin underwent a conformational change upon activation to toxin (Choma *et al.*, 1991).

Common and miscellaneous structural measurements. X-ray diffraction (Wyckoff *et al.*, 1985a,b; Glasel, 1995) and solution NMR spectroscopy (Hirs and Timasheff, 1986b; Wüthrich, 1986; James and Oppenheimer, 1994; Glasel, 1995), in conjunction with computational protocols (Brooks *et al.*, 1983; Karplus and Petsko, 1990; Guntert and Wüthrich, 1991; Guntert *et al.*, 1991), are the established methods of three-dimensional crystal, and solution structure determination. X-ray crystallography is the major structural tool but requires substantial amounts of protein, labour and skill in order to obtain X-ray quality crystals. For these reasons, X-ray data of many proteins are not available. While X-ray data provides information of a static nature, solution NMR studies address the dynamic component of protein structure. NMR techniques are generally limited to smaller proteins because the experimental results are more easily interpreted and because very large proteins have broad lines due to long correlation times (Harris, 1983). Interestingly, the structural information provided by either method is remarkably similar. Protein folding and stability have been addressed using predictive computational methods (Dill and Stigter, 1995; Glaser, 1995; Lazaridis *et al.*, 1995) and kinetic unfolding experiments (Fersht *et al.*, 1992). NMR or ultraviolet spectroscopic measurements are often required to monitor the structural changes of this process (Roder, 1989; Englander and Mayne, 1992; Serrano *et al.*, 1992; Baldwin, 1993). Computer modeling packages have also been used to model protein-substrate interactions (Bohacek and McMartin, 1994; Bohacek and McMartin, 1995; McMartin and Bohacek, 1995). In other cases, NMR and Mössbauer spectroscopies have been used to probe local environments of metal ions bound to protein (Butler and Eckert, 1989; Aramini and Vogel, 1993; Aramini *et al.*, 1993; Aramini and Vogel, 1994; Aramini *et al.*, 1994a,b; Pikus *et al.*, 1996).

1.4.2 Examples of chemical methods of structure-function analysis. While protein function can be measured nonintrusively using physical approaches, as in the case of NMR monitored binding studies (Sykes *et al.*, 1970; Searle *et al.*, 1988; Paterson *et al.*, 1990; Jacobsen and Gerig, 1991; Mayne *et al.*, 1992; Spera and Bax, 1991; Tsang *et al.*, 1992; Takahashi *et al.*, 1992; Werner and Wemmer, 1992), it is more common practice to characterize protein function in response to a perturbation of structure. With due consideration to established physical methods, the availability of many modification reagents, enzymes and strategies, and the potential of chemical methods to contribute to elucidating structure-function relationships seem inherently greater. A multitude of chemical and enzymatic strategies have been used in structure-function studies. The power of chemical techniques of protein investigation have been complemented by improvements of various fields, such as peptide chemistry (Bodanszky and Bodanszky, 1994), and the development of site-directed mutagenesis (Stryer, 1988). There is a myriad of chemical approaches and reagents which have been reviewed for protein modification (Leggett Bailey, 1962; Hirs, 1967; Boyer, 1970; Boyer, 1971; Means and Feeney, 1971; Hirs and Timasheff, 1972; Duggleby and Kaplan, 1975a, Glazer *et al.*, 1976; Lundblad and Noyes, 1984; Imoto and Yamada, 1989; Young and Kaplan, 1989). As with physical methods, individual chemical methods provide only partial characterization and require combined studies. A very limited sample of chemical methods and their applications are outline below.

Titration studies. A very elegant method to obtain structure-function information was the classic protein titration assay (Hirs, 1967; Means and Feeney, 1971). Thermodynamic titration profiles were very useful for providing information on the nature and number of ionizable groups

in a protein and enabled the estimation of protein pI values. Titration hysteresis gave additional insight as to whether groups were surface exposed or buried.

Proteolytic processing. Tightly folded proteins, such as the proteinases, are generally not susceptible to proteolysis (Boyer, 1971) but can be acted upon by proteases as soon as they begin to unfold. In a particular protein folding experiment, the insecticidal protein from *Bacillus thuringiensis* was proteolyzed for a short time when placed under denaturing conditions (Choma and Kaplan, 1990). Several clean fragments were observed, which suggested the initial unfolding occurred only at certain key points. In effect, the enzyme had delineated the major structural motifs, which had not unfolded, from local structures between the motifs, which had unfolded and were proteolyzed.

Kinetic measurements. Kinetic methods, in particular enzyme kinetics, make use of the fact that enzymes have an easily measured activity, which in itself serves as a probe of structural integrity and the functioning of various residues (Cornish-Bowden and Wharton, 1988; Creighton, 1993; Purich, 1995). Clearly, unfolded or poisoned enzymes cannot have catalytic activity. Unlike physical methods though, kinetics are necessarily restricted to the study of enzymes, but the information obtained is in many cases extrapolatable to the better understanding of proteins in general. In the early development of protein chemistry, enzyme interactions with substrate, inhibitor and affinity label provided information which could not have been obtained by any other means. Enzyme kinetics were often complimented with other strategies that made use of isotopes and different dielectric media (Pulrich, 1980; Cleland, 1982). In this way kinetics could be used

to investigate functional groups essential to binding and/or catalysis, probe transition state energetics and reaction pathways, and differentiate between different classes of ionizable groups involved in catalysis.

Chemical modification of enzymes. Another strategy used to exploit the power of chemical methods has been the combined chemical modification and activity measurement approach. In a classic example, the amino groups of α–chymotrypsin and chymotrypsinogen were acetylated (Boyer, 1971). Covalent modification destroyed the activity of α–chymotrypsin. Interestingly, acetylated chymotrypsinogen could be treated with trypsin to generate active protease. This suggested that the proenzyme structure prevented modification of at least one residue important to catalysis. It was eventually shown that the positive charge of the N-terminal amino group of Ile-16 was essential for maintaining a catalytically essential salt bridge (McConn *et al.*, 1969). Following activation, acetylated chymotrypsinogen was shown to have an activity identical to that of wild-type α–chymotrypsin. While acetylation reduces the number of positive charges on a protein surface, the modification was inconsequential in this example. However, the consequences of charge replacements are protein dependent and difficult to anticipate and for this reason, other modifications have been attempted using guanidinating and amidinating reagents (Means and Feeney, 1971) which conserve the charge of amino groups. In another chemical modification study, dissolved cyanogen gas was used to crosslink the active site ion pair of carbonic anhydrase II. The single modification destroyed activity (Day *et al.*, 1989).

Competitive labeling. Normally, following a chemical derivatization or labeling procedure, some aspect of the modified protein is characterized. An alternate and extremely powerful chemical

approach to elucidating structure-function relationships is the competitive labeling approach of Duggleby and Kaplan (1975), where pK_a values and structure-induced perturbations of functional group reactivity can be quantified for each derivatizable group in a protein. Very small amounts of protein are amenable for analysis. The experimental design of this technique was distinct among other chemical approaches since it incorporated a label into protein, yet probed its native properties. In a particularly striking case, the method enabled the elucidation of individual pK_a values for every reactive amino group of a particular protein in a 50-protein ribosome complex, under native conditions (Hasnain *et al.*, 1977). The inability of some residues to incorporate label indicated that they were either buried or tightly associated to other proteins. It should be pointed out that no other method, past or present, could have provided information of this calibre.

Mutated proteins. Biosynthesis can be complimented with solution and solid phase synthetic techniques to produce modified proteins, in what has become accepted as protein semisynthetic methods (Offord, 1980). The approach is often used to compliment chemical modification and site-directed mutagenesis techniques in protein engineering studies. Semisynthesis of cytochrome c, for example, made it possible to incorporate bipyridyl-alanyl residues that could be used for fluorescence investigations of the heme group (Raphel and Gray, 1991; Imperiali *et al.*, 1993; Wuttke *et al.*, 1993). Notwithstanding, site-directed mutagenesis in itself remains an extremely powerful technique and has many applications. In a notable example, extensive site-directed mutagenesis was used to characterize the folding pathway of barnase (Serrano *et al.*, 1990; Serrano *et al.*, 1992). Using this extremely rigorous treatment, it was possible to calculate interaction energies between various residues and draw conclusions as to the folded state of specific structural motifs in barnase. Mutations in the active site which removed electrostatic

strain were found to increase stability towards the action of denaturants, however, it also decreased the catalytic efficiency. Mutations which added electrostatic strain into the active site region had the opposite effect (Meiering *et al.*, 1992). Furthermore, by substituting residues of barnase with residues of the related protein binase, a barnase mutant was engineered with high stability to unfolding (Serrano *et al.*, 1993). In another example, site-directed mutagenesis was used to remove a histidyl residue known to participate in the coordination of carbon dioxide by heme groups in myoglobin (DePillis *et al.*, 1994). The pocket originally occupied by the imidazole side chain could then be replaced with a number of aromatic heterocycles such as pyridine, and the effects on the performance of the heme group quantified.

Primary structure determination. Conventional acid hydrolysis is an extreme example of chemical degradation (Darbre, 1986). Despite this crude reaction, the approach allowed Sanger to identify the N-terminal amino acids of insulin using 2,4-fluorodinitrobenzene (Leggett Bailey, 1962; Stryer, 1988). A more elegant example, however, of extracting protein information by degradation is the established Edman procedure (Hirs, 1967; Hirs and Timasheff, 1972; Stryer, 1988). Primary structure confirmation of proteins suspected of being post-translationally processed and primary structure determinations of uncharacterized proteins still use the classic Edman N-terminal degradation procedure, which is a sequential stepwise chemical degradation of the N-terminal amino acid residue to produce a fluorescent derivative. With the exception of some refinements, the principle behind the technique is very much the same as the original method.

Analogous to N-terminal sequence determinations, it may be crucial to know the C-terminal sequence of a protein for characterization purposes, particularly those proteins who are suspected of undergoing proteolytic processing. However, in contrast to the Edman degradation procedure,

there are no general approaches to successfully sequence the troublesome C-terminal region of proteins (Fromageot *et al.*, 1950; Akabori *et al.*, 1952; Duggleby and Kaplan, 1975b; Darbre, 1986; Glaser, 1995). This seemingly simple problem was addressed with some success in the recent past using matrix assisted laser desorbtion ionization mass spectrometry in tandem with carboxypeptidase Y methods (Patterson *et al.*, 1995). Nevertheless, the chemistry of C-terminal degradation methods have not been perfected and require further investigation in order to provide reliable structure information.

1.5 An overview of the merit of physical and chemical methods

1.5.1 Physical methods. Physical methods are useful because they are usually nonintrusive with respect to protein structure. An excellent example of the potential power of physical methods of measurement is found in the NMR promoted elucidation of protein three-dimensional native structure. By combining homonuclear correlation spectroscopy measurements (COSY and NOESY) (Kessler *et al.*, 1988) it is possible to determine all amino acid residues in solution which are in close proximity. The premise is that if enough distance constraints are accumulated, there can only be one general three-dimensional conformation which satisfies all constraints. In this way, NMR spectroscopy and computational methods can be used to determine the dynamic solution structure of a protein without ever perturbing the native conformation (Wüthrich, 1986). NMR techniques are inherently insensitive compared to other methods, however, and an analysis normally requires protein concentrations of 1 to 5 mM, which may be impractical, due to solubility considerations or availability of protein.

1.5.2 Chemical methods: enzyme kinetics versus chemical modification. Unlike enzyme kinetics, chemical modification studies need not require that the protein possess an internal probe of structure and function since the incorporated reagent itself serves as the probe. For this reason, the structure-function study of proteins by chemical modification need not be restricted to the study of enzymes, and provides information about proteins that could not be readily obtainable by physical methods or kinetics. A drawback, however, is that chemical modification techniques are intrusive, in that they modify protein functional groups and possibly interactions which maintain protein structure. Depending on the nature of the modification and the focus of the investigation, there is always the risk that the modification could defeat the purpose of the investigation by altering the protein is some undesirable way, leading to spurious results. For example, amino acids in an enzyme may be guanidated and the effects of the guanidation quantified (Hirs and Timasheff, 1972). However, guanidation is typically carried out at pH values exceeding 10 and few proteins can tolerate these conditions without suffering adverse effects. It follows, therefore, that the results of guanidation on enzyme activity may be meaningless unless steps are taken to ensure structural integrity following modification. Physical methods such as circular dichroism have been invaluable for these reasons as a verification of post-reaction integrity of structure.

In rare cases, investigations that produce gross changes in structure may be tolerated, but this again is dependent on the aim of the experiment. Cyanogen bromide cleavage of protein followed by Edman degradation is such an example (Darbre, 1986). While both modifications are detrimental to protein structure, they do not defeat the purpose of the investigation, which is to elucidate the primary structure. As previously mentioned the technique of competitive labeling (Duggleby and Kaplan, 1975; Young and Kaplan, 1975) is also an exceptional case, since the

experimental design, the reaction conditions which are used, and the trace amounts of reagent employed preclude the requirement for structural verification.

1.5.3 Chemical methods: chemical modification versus site-directed mutagenesis.

Well before the advent of site-directed mutagenesis, chemical modification was being used to elucidate the roles of amino acid residues in protein. The potential of functional groups to be modified is determined by the ionization state, intrinsic nucleophilicity and accessibility to reagent (Means and Feeney, 1971). The latter factor has a marked dependency of the group's proximity to the surface of the protein. As far as the outcome of a reaction is concerned, the degree of surface exposure of a residue in protein constitutes what is an analogous but exaggerated example of the effect of steric hinderance in a typical organic reaction (March, 1988). While steric hinderance may be a nuissance in synthetic chemistry, the failure of a group to react in proteins can be just as informative as if it does react.

Unlike site-directed mutagenesis, chemical modification is not specific for particular residues at particular positions. Even different functional groups of different classes of amino acid residues can be, and often were, modified by the same reagent. There are, however, exceptions to the rule, as with site-directed reagents such as diisopropylfluorophosphate which apparently reacted only with the catalytic serine residue of the serine proteinases (Means and Feeney, 1971).

Despite the fact that the chemical modification is lacking specificity in comparison to site-directed mutagenesis, the lack of specificity need not be a disadvantage. Moreover, the list of chemical modification reagents is endless, while mutation studies are generally limited to substitution by one of the other nineteen naturally occurring amino acids. Another advantage is that chemical modification procedures can provide rapid and useful information on structurally

and functionally important regions of a protein, and often provide insights as to which specific amino acid residues have structural or functional roles. Thus, just as physical methods provide a structural foundation from which to carry out chemical modification studies, chemical modification studies can provide a structure-function foundation from which to carry out site-directed mutagenesis studies, where residue-specific modifications can be imposed on protein.

1.6 Limitations of physical and chemical methods and the demand for novel techniques to elucidate structure and function in protein.

Historically, the first milestone of protein chemistry was reached with advances in amino acid chemistry and the revelation that proteins were composed of amino acids. Scientists felt that once the primary structure of proteins was understood, the mystery of proteins would be clarified. It was soon confirmed using light spectroscopies that secondary structure elements existed. Eventually even X-ray crystallography advanced to a point where protein three-dimensional structure could be solved. New hopes for answers rested on the basis of these developments but still the biggest problems of the times were unresolved. When site-directed mutagenesis was developed, again there were hopes that unresolved issues could finally be put to rest. Still today, there are more questions raised by various techniques in protein chemistry than the answers they obtain.

To summarize, all types of physical and chemical methods are available, be they benign, disruptive, or even destructive protocols. Each protocol has different features, merits and drawbacks and its applicability is at times even protein dependent. In spite of the fact that there are many established techniques to investigate structure-function relationships and newer methods are in various stages of development, the complexity of proteins are such that novel approaches

are still required to obtain a better understanding of protein structure and function. The demand for additional methods of protein investigation is self-perpetuating, and as exemplified above, this claim continues to withstand the test of time. Scientists are currently struggling with the development of C-terminal methods of analysis (Darbre, 1986), protein folding studies (Ptitsyn, 1995) and numerous structure-function issues.

1.7 Novel approaches to elucidating structure and function in protein:

1.7.1 Enzyme catalyzed reactions in organic solvents

It was Klibanov who rigorously proved that enzymes were catalytically active in organic solvents (Zaks and Klibanov, 1988). Until his pioneering work was established, protein structure-function investigations in organic solvents were not attempted because protein structure was believed to be destroyed by organic media. Ironically, one of Klibanov's most important observations was that enzyme structure is more stable in apolar organic solvents than in water. The fact that structure-function studies could be carried out under nonaqueous conditions shed a new light into the ways proteins could be examined.

1.7.2 Overview of the present study

The work presented in this thesis demonstrates that exploiting novel approaches of protein investigation can contribute to (i) the understanding of protein structure and function and (ii) the development of novel procedures for the manipulation of protein. A description of the work carried out in each chapter follows, with some of the novel elements outlined.

Chapter 2: The use of iodomethane as a protein modifying reagent

While many reagents of organic synthesis are potentially applicable to protein systems, their chemistry with proteins has not been developed. To this end, the utility of iodomethane as a protein modifying reagent was investigated and it was indeed shown that from many viewpoints, iodomethane had appropriate properties that made it very ammenable for use in protein structure-function analyses. Reaction of proteins with iodomethane of various isotopic forms made it possible to carry out the following novel experiments: (i) Directly observe and investigate the reaction of functional groups in protein with iodomethane by [^{13}C]NMR spectroscopy. (ii) Use chemical and enzymatic degradation protocols, and high voltage paper electrophoresis (Leggett Bailey, 1962) to separate different methylated functional group residues in protein as their ^{14}C/^{13}C-labeled peptide fragments, and identify the reacted functional groups by NMR and mass spectral analyses. (iii) Directly observe an N-terminal sequence heterogeneity in a particular protein preparation by [^{13}C]NMR spectroscopy, and investigate the heterogeneous N-terminal trimethylated amino acid residue by the characteristic chemical shift resonance of its free trimethylated amino acid.

Chapter 3: Nonaqueous chemical modification of lyophilized proteins

Proteins in general, unlike enzymes with their catalytic properties, lack an inherent probe for the easy detection of changes in structure and function, and this, perhaps, is one reason why scientists have worked mainly with enzymes. While circular dichroism, ultraviolet, infrared and NMR spectroscopies have been successful in promoting the study of nonenzymatic proteins, each method has inherent limitations which detract from the generality of the procedure (Hirs and Timasheff, 1986a,b; Sauer, 1995; Glasel, 1995). Since the chemical labeling reagent itself serves

as a probe of structure and function, the use of reagents need not be limited to the study of enzymes. It seems unusual, however, that more chemical approaches, rather than reagents, have not been explored to full potential. While chemical reactions of protein are almost always carried out in water, it was never questioned that an aqueous environment may provide a relatively poor framework from which to characterize the outcome. Chemical reactions can become quite involved under the best of conditions, and as chemical modifications are less benign than physical methods, experimentalists who work with such complicated systems as proteins can encounter problems with structural integrity, hydrolytic breakdown and other phenomena which make the interpretation of results difficult or subject to question. These added complications need not occur, and to address these issues, a novel chemical modification procedure for proteins was developed, namely, the chemical modification of lyophilized proteins under nonaqueous conditions.

A nonaqueous chemical modification of lyophilized protein has two essential characteristics which can be taken advantage of: (i) water is absent and (ii) protein is structurally rigid (Zaks and Klibanov, 1988), but retains the essential elements of its native structure. By modifying proteins in the absence of water, specific facts of local structure and function were obtained, while the pitfalls of aqueous modification - namely, proteolysis, denaturation, and conformational dynamics - were avoided.

By implementing this approach, it was possible to (i) confirm the pH memory effect of proteins as postulated by Zaks and Klibanov (1988), (ii) derivatize protein amino groups, carboxyl groups, and tyrosyl and methionyl side-chains, (iii) radiolabel proteins with high specific radioactivity using tritium or carbon-14 enriched reagents (iv) carry out protein modification

reactions free of proteolysis, and (v) benefit from the technical aspects of nonaqueous chemical modifications of proteins - namely, cleaner reactions, improved cost-efficiency, facile workup and unparalleled convenience when monitoring the progression of a reaction. In short, with this approach, it was possible to interpret and carry out structure-function studies in ways which never could have been achieved under aqueous conditions.

Chapter 4: Microscale nonaqueous chemistry for the analysis of protein C-terminal sequences

As previously eluded, the direct sequencing of C-terminal regions of proteins has been an issue for some time. When applied to minute quantities of protein, there is apparently no protocol which is sufficiently general, sensitive and/or cost-effective enough to be used as the method of choice. Chemical methods are often unsuccessful, particularly in cases where C-termini are blocked (Darbre, 1986). Even in protein where the C-termini are free, microscaling the procedure for application to small amounts of proteins are rife with potential problems which are due, in part, to loss of protein, unavoidable contamination and interferents or long analysis times. Using the principles behind the method of Kaplan and Duggelby (1975b), an alternate general procedure was examined for the isolation of C-terminal sequences in protein. While other chemical approaches have drawbacks such as poor yield, product loss, non-generality and interferents during analysis, the potential problems in this nonaqueous method are inherently limited due to the novel experimental design. The procedure has been tested on minute quantities of protein and is potentially ammenable to the analysis of any trace protein having a free carboxyl terminus.

Chapter 5: Characterizing cathepsin B substrate specificity

The mechanism of cathepsin B activity is an area researched intensively by scientists of pharmaceutical companies. Since cathepsin B is implicated in many pathological conditions (Sloane, 1990), there is considerable interest in the development of specific potent inhibitors of cathepsin B activity. Thus far, however, the knowledge base required to achieve this goal is lacking and a better understanding of cathepsin B substrate specificity is required. To date, the full potential kinetic analysis of cathepsin B action has had to await several developments: (i) The production of a homogeneous nonglycosylated cathepsin B mutant. (ii) The elucidation of the three-dimensional crystal structure of cathepsin B. (iii) Modeling algorithms better suited to the study of protein interactions. Each criterion has been met recently, opening the door to the implementation of novel kinetic analyses. In this chapter, the S_3 subsite specificity of cathepsin B is finally characterized in a multicollaborative effort which required contributions from the National Research Council of Canada, the Shriners Hospital for Crippled Children, and Ciba-Geigy Corporation. Kinetic analyses of novel substrates, designed to probe S_3 subsite specificity, were carried out. Protein-substrate interactions were simulated by novel computational methods that used X-ray coordinates of a cathepsin B-inhibitor complex to impose positional constraints, and these results corroborated the outcome of the kinetic analyses. Another minor point is that the substrates were shown by the technique of high voltage paper electrophoresis to be cleaved only at the intended amide bond. This is one aspect of enzyme kinetics which is usually taken for granted by researchers in the field and rarely verified.

Chapter 6: [^{12}C]Iodomethane and the investigation of protein amino groups by [^{14}N]NMR spectroscopy

Amino group resonances of protein are normally very broad when observed by [^{14}N]NMR spectroscopy (Witanowski and Webb, 1972). Because of this feature, a structure-function study which uses this approach is inherently limited in the information it can provide. One of the useful features of iodomethane is that amino groups may be readily quaternized to form trimethylammonium derivatives of exceptional electronic symmetry. These derivatized groups are readily observed as very sharp resonance lines by [^{14}N]NMR spectroscopy. Therefore, the use of iodomethane finally provides a unique opportunity to directly observe and investigate the reaction, local environment and motion of amino groups in any protein by [^{14}N]NMR spectroscopy. The reaction of amino groups is presented herein.

1.8 References

Akabori, S., Ohno, K. and Narita, K. (1952) *Bull. Chem. Soc. Jap.* **25**, 214-218.

Allinger, N. L., Cava, M. P., de Jongh, D. C., Johnson, C. R., Lebel, N. A. and Stevens. C. L. (1976) *Organic Chemistry*, 2nd ed., Worth Publishers.

Aramini, J. M., German, M. K. and Vogel, H. J. (1993) *J. Am. Chem. Soc.* **115**, 9750-9753.

Aramini, J. M., Krygsman, P. H. and Vogel, H. J. (1994a) *Biochemistry* **33**, 3304-3311.

Aramini, J. M., McIntyre, D. D. and Vogel, H. J. (1994b) *J. Am. Chem. Soc.* **116**, 11506-11511.

Aramini, J. M. and Vogel, H. J. (1993) *J. Am. Chem. Soc.* **115**, 245-252.

Aramini, J. M. and Vogel, H. J. (1994) *J. Am. Chem. Soc.* **116**, 1988-1993.

Auger, M. (1995) *J. Chim. Phys* **92**, 1751-1760.

Baldwin, R. L. (1993) *Curr. Opin. Struct. Biol.* **3**, 84-91.

Balls, A. K. and Jensen, E. F. (1952) *Adv. Enzymol.* **13**, 321.

Biemann, K. (1992) In *Annu. Rev. Biochem.* **61**, 977-1010.

Bodanszky, M and Bodanszky, A. (1994) *The Practice of Peptide Synthesis*, 2nd ed., Springer-Verlag.

Bohacek, R. S. and McMartin, C. (1994) *J. Am. Chem. Soc.* **116**, 5560-5571.

Bohacek, R. S. and McMartin, C. (1995) *Nature Medicine* **1**, 177-178.

Boyer (1970) *The Enzymes*, 3rd ed. **1**, Academic Press.

Boyer (1971) *The Enzymes*, 3rd ed. **3**, Academic Press.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983) *J. Comput. Chem.* **4**, 187-217.

Burgess, R.R. (1987) In *Protein Engineering* (Oxender, D.L. and Fox, C.F. Eds.) pp.71-82, ARL.

Butler, A. and Eckert, H. (1989) *J. Am. Chem. Soc.* **111**, 2802-2809.

Carter, C. W., Jr. and Sweet, R. M. (1997) *Meth. Enzymol.* **276**, Academic Press.

Choma, C. T. and Kaplan, H. (1990) *Biochemistry* **29**, 10971-10977.

Choma, C. T., Surewicz, W. K., Carey, P. R., Pozsgay, M. and Kaplan, H. (1990) *J. Prot. Chem.* **9**, 87-94.

Choma, C. T., Surewicz, W. K. and Kaplan, H. (1991) *Biochem. Biophys. Res. Comm.* **179**, 993-938.

Cleland, W. W. (1982) *Crit. Rev. Biochem.* **13**, 385-428.

Cornish-Bowden, A. and Wharton, C. W. (1988) *Enzyme Kinetics* (Rickwood, D. Ed.) IRL Press.

Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2nd ed., W. H. Freeman and Company.

Darbre, A. (1986) *Practical Protein Chemistry-A Handbook*, John Wiley and Sons.

Day, R. A., Kirley, J., Tharp, R., Ficker, D., Strange, C. and Ghenbot, G. (1989) In *Techniques in Protein Chemistry* (Hugli, T. E. Ed.) pp. 517-525, Academic Press.

DePillis, G. D., Decatur, S. M., Barrick, D., Boxer, S. G. (1994) *J. Am. Chem. Soc.* **116**, 6981-6982.

Dill, K. A. and Stigter, D. (1995) *Adv. Prot. Chem.* **46**, 59-102.

Drago, R. S. (1992) *Physical Methods for Chemists*, 2nd ed., Saunders College Publishing.

Duggleby, R. G. and Kaplan, H. (1975a) *Biochemistry* **14**, 5168-5175.

Duggleby, R. G. and Kaplan, H. (1975b) *Anal. Biochem.* **65**, 346-354.

Englander, S. W. and Mayne, L. (1992) *Ann. Rev. Biophys. Biomol. Struct.* **21**, 243-265.

Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, W. H. Freeman and Company.

Fersht, A. R., Matouschek, A and Serrano, L. (1992) *J. Mol. Biol.* **224**, 771-782.

Fromageot, C. Jutisz, M., Meyer, D. and Penasse, L. (1950) *Biochim. Biophys. Acta* **6**, 283-289.

Glasel, J. A. and Deutscher, M. (1995) *Introduction to Biophysical Methods for Protein and Nucleic Acid Research*, Academic Press.

Glazer, A. N., Delange, R. J. and Sigman, D. S. (1976) In *Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T. S. and Work, E. Eds.) **4**, North-Holland Publishing Company.

Guntert, P, Braun, W. and Wüthrich, K. (1991) *J. Mol. Biol.* **217**, 517-530.

Guntert, P and Wüthrich, K. (1991) *J. Biomol. NMR* **1**, 447-456.

Hames, B. D. and Rickwood, D. (1990) *Gel Electrophoresis of Proteins: A Practical Approach*, 2nd ed., IRL Press.

Harris, R. K. (1983) *Nuclear Magnetic Resonance Spectroscopy: A Physicochemical View*, Pitman Publishing Inc.

Hashida, M., Nishikawa, M., Yamashita and Takakura, F. (1994) *Drug Dev. and Ind. Pharm.* **20**, 581-590.

Hasnain, S., Visentin, L. P. and Kaplan, H. (1977) *Eur. J. Biochem.* **80**, 35-41.

Herriott, R. M. (1947) *Adv. Prot. Chem.* **3**, 169.

Herschlag, D. (1988) *Bioorganic Chemistry* **16**, 62-96.

Hirs, C. H. W (1967) *Meth. Enzymol.* **11**, Academic Press.

Hirs, C. H. W. and Timasheff, S. N. (1972) *Meth. Enzymol.* **25B**, Academic Press.

Hirs, C. H. W. and Timasheff, S. N. (1985) *Meth. Enzymol.* **117**, Academic Press.

Hirs, C. H. W. and Timasheff, S. N. (1986a) *Meth. Enzymol.* **130**, Academic Press.

Hirs, C. H. W. and Timasheff, S. N. (1986b) *Meth. Enzymol.* **131**, Academic Press.

Imoto, T. and Yamada, H. (1989) In *Protein Function: A Practical Approach* (Creighton, T. E. Ed.) Chapter 10, IRL Press.

Imperiali, B., Prins, T. J. and Fisher, S. L. (1993) *J. Org. Chem.* **58**, 1613-1616.

Jacobsen, A. R. and Gerig, J. T. (1991) *J. Biomol. NMR* **1**, 131.

James, T. L. and Oppenheimer, N. J. (1994) *Meth. Enzymol.* **239**, Academic Press.

Johnson, M. L. and Brand, L. (1994) *Meth. Enzymol.* **240**, Academic Press.

Karger, B. L and Hancock, W. S. (1996a) *Meth. Enzymol.* **270**, Academic Press.

Karger, B. L and Hancock, W. S. (1996b) *Meth. Enzymol.* **271**, Academic Press.

Karplus, M and Petsko, G. A. (1990) *Nature* **347**, 631-639.

Kessler, H., Gehrke, M. and Griesinger, C. (1988) *Angew. Chem. Int. Ed. Engl.* **27**, 490-536.

Laidler, K. J. and Meiser, J. H. (1982) *Physical Chemistry*, Benjamin-Cummings Publishing Company.

Lazaridis, T., Archontis, G. and Karplus, M. (1995) *Adv. Prot. Chem.* **47**, 231-296.

Leggett Bailey, J. (1962) *Techniques in Protein Chemistry*, Elsevier Publishing Company.

Lundblad, R. L. and Noyes, C. M. (1984) *Chemical Reagents for Protein Modification*, **1** and **2**, CRC Press.

Malcolm, A. D. B. (1978) *Nature* **275**, 90-91.

March, J. (1988) *Advanced Organic Chemistry: Reactions, Mechanisms and Structure*, 3rd ed., John Wiley and Sons.

Mayne, L., Paterson, Y., Cerasoli, D. and Englander, S. W. (1992) *Biochemistry* **31**, 10678.

McCloskey, J. A. (1990) *Meth. Enzymol.* **193**, Academic Press.

McConn, J., Fasman, G. D. and Hess, G. P. (1969) *J. Mol. Biol.* **39**, 551-562.

McMartin, C. and Bohacek, R. S. (1995) *Journal of Computer-Aided Molecular Design* **9**, 237-250.

Means, G. E. and Feeney, R. E. (1971) *Chemical Modification of Proteins*, Holden-Day.

Meiering, E. M., Serrano, L. and Fersht, A. R. (1992) *J. Mol. Biol.* **225**, 585-589.

Offord, R. E. (1980) *Semisynthetic Proteins*, John Wiley and Sons.

Olcott, H. S. and Fraenkel-Conrat, H. (1947) *Chem. Rev.* **41**, 151.

Palmer, A. G., Williams, J. and McDermott, A. (1996) *J. Phys. Chem.* **100**, 13293-13310.

Paterson, Y., Englander, S. W. and Roder, H. (1990) *Science* **249**, 755.

Patterson, D. H., Tarr, G. E., Regnier, F. E. and Martin, S. A. (1995) *Anal. Chem.* **67**, 3971-3978.

Pikus, J. D., Studts, J. M., Achim, C., Kauffmann, K. E., Munck, E., Steffan, R. J., McClay, K. and Fox, B. G. (1996) *Biochemistry* **35**, 9106-9119.

Ptitsyn, O. B. (1995) *Adv. Prot. Chem.* **47**, 83-217.

Purich, D. L. (1980) *Meth. Enzymol.* **64**, Academic Press.

Purich, D. L. (1995) *Meth. Enzymol.* **249**, Academic Press.

Raphael, A. L. and Gray, H. B. (1991) *J. Am. Chem. Soc.* **113**, 1038-1040.

Richards, F. M. and Wyckoff, H. W. (1971) In The Enzymes, 3rd ed., (Boyer, P. D. Ed.) **3**, p. 683, Academic Press

Roder, H. (1989) *Method. Enzymol.* **176**, pp. 446-473, Academic Press.

Sauer, K. (1995) *Method. Enzymol.* **246**, Academic Press.

Searle, M. S., Forster, M. J., Birdsall, B., Roberts, G. C. K., Feeney, J., Cheung, H.T. A., Kompis, I and Geddes, A. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3787.

Serrano, L., Horovitz, A., Avron, B., Bycroft, M and Fersht, A. R. (1990) *Biochemistry* **29**, 9343-9352.

Serrano, L., Day, A. G. and Fersht, A. R. (1993) *J. Mol. Biol.* **233**, 305-312.

Serrano, L., Kellis, J. T., Cann, P., Jr., Matouschek, A. and Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 783-804.

Sloane, B. F., Moin, K., Krepela, E. and Rozhin, J. (1990) *Cancer Metastasis Rev.* **9**, 333-352.

Spackman, D. H., Stein, W. H. and Moore, S. (1958) *Anal. Chem.* **30**, 1190.

Spera, S., Ikura, M and Bax, A. (1991) *J. Biomol. NMR* **1**, 155.

Stryer, L. (1988) *Biochemistry*, 3rd ed., W. H. Freeman and Company.

Sykes, B. D., Schmidt, P. G. and Stark, G. R. (1970) *J. Biol. Chem.* **245**, 1180.

Takahashi, H. Suzuki, E. Shimada, I. and Arata, Y. (1992) *Biochemistry* **31**, 2464.

Tsang, P. Rance, R., Fieser, T. M., Ostresh, J. M., Houghten, R. A., Lerner, R. A. and Wright, P. E. (1992) *Biochemistry* **31**, 3862.

Watts, A., Ulrich, A. S. and Middleton, D. A. (1995) *Molecular Membrane Biology* **12**, 233-246.

Werner, M. H. and Wemmer, D. E. (1992) *J. Mol. Biol.* **225**, 873.

Witanowki, M. and Webb, G. A. (1975) In *Annual Reports on NMR Spectroscopy* (Mooney, E. F. Ed.) **5A**, pp. 395-457, Academic Press.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons.

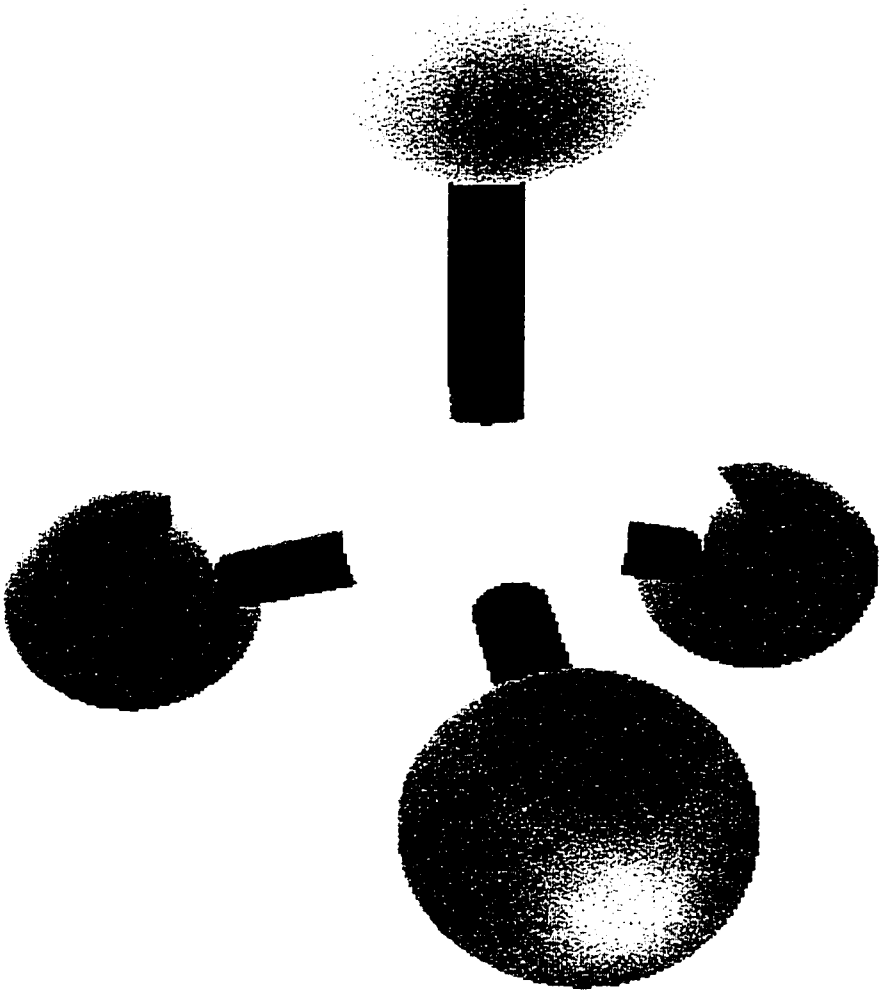
Wuttke, D. S., Gray, H. B., Fisher, S. L. and Imperiali, B. (1993) *J. Am. Chem. Soc.* **115**, 8455-8456.

Wyckoff, H. W., Hirs, C. H. W. and Timasheff, S. N. (1985a) *Meth. Enzymol.* **114**, Academic Press.

Wyckoff, H. W., Hirs, C. H. W. and Timasheff, S. N. (1985b) *Meth. Enzymol.* **115**, Academic Press.

Young, M. N. and Kaplan, H. (1989) In *Protein Function: A Practical Approach* (Creighton, T. E. Ed.) Chapter 9, IRL Press.

Zaks, A. and Klibanov, A. M. (1988) *J. Biol. Chem.* **263**, 3194-3201.



Chapter 2: Iodomethane as a protein modifying reagent

2.1 Introduction

Iodomethane is a common and effective methylating reagent of organic reactions (March, 1985a) and has gained popularity for the ease of which it methylates various functional groups such as amino, hydroxyl, thiol, and carboxyl groups. It has also been used for the ^{14}C-methylation and dimethylation of histidine imidazole nitrogens (Cowgill, 1957) and for the preparation of S-methylcysteine (Clark and Dijkstra, 1980). In sharp contrast to its application in organic synthesis, iodomethane has only been marginally used for protein derivatization. Of the known examples, the use of iodomethane is almost exclusively limited to the permethylation of proteins for mass spectral analyses (Dell, 1986; Knapp, 1990). In one example, it was used to alkylate methionine–29 in ribonuclease A (Link and Stark, 1968). In a second case, lysyl residues of the protein asialo-fetuim were reductively dimethylated with [^{14}C]formaldehyde and borohydride, and then trimethylated with iodomethane in aqueous methanol solution (Paik and Kim, 1980a; LaBadie *et al.*, 1976). In a last example, the single thiol of streptococcal proteinase under reducing conditions was alkylated using iodomethane (Liu and Elliot, 1971).

Iodomethane is detrimental to biological systems on several levels. Many reference manuals and catalogues (Merck Index, Beilstein, Aldrich, Fluka, Sigma, Materials Safety Data Sheets, etc.) have warnings to the effect that the toxicity of iodomethane is appreciable, however, the modes of action are poorly understood. It is reported to be toxic to central nervous system activity, severely irritating upon contact, and carcinogenic with chronic exposure (Kutob and Plaa, 1962; Amacher and Zelljadt, 1984; Amacher and Dunn, 1985; Lunn and Sansone, 1991). Iodomethane must presumably have some action on proteins, given that there are potentially

reactable functional groups in all proteins. The effects of iodomethane on protein activity nevertheless remain undefined.

While iodomethane may have detrimental effects on protein-based systems, paradoxically, the post-translational (*in vivo*) methylation of amino acid residues of various proteins appears to be essential in some cases to maintain biological function. Indeed, a wide variety of methylated proteins have been reported and investigated for some time (Table 2.1.1). 3-N-Methylhistidine, for example, is a product of muscle protein breakdown, but its biological significance is not well understood (Paik and Kim, 1980b). In certain pathological conditions, methylated amino acids can be found in the urine at abnormal concentrations (Paik and Kim, 1980b). Biological methylation reactions are typically carried out by highly specific S-adenosyl methionine methyl transferase enzymes (DiMaria *et al.*, 1982; Rowe *et al.*, 1986). The most common methylation reactions appear to take place at lysyl residues. Methylation can be specific to a particular lysyl residue (DeLange *et al.*, 1969; Schaeffer *et al.*, 1987) or extensive, to produce highly methylated proteins (Parish and Ada, 1969; Hiatt *et al.*, 1982). In histones, lysyl residues are often found as mono, di and trimethylated amino acid derivatives. These are thought to take part in the condensation of chromatin during mitosis (Paik and Kim, 1980c). Methylated lysyl residues have

Table 2.1.1: Examples of biologically methylated proteins and the derivatives formed

PROTEIN	METHYLATED AMINO ACID DERIVATIVES
HISTONES	Nε-MONO, DI & TRIMETHYLLYSINE, N^G-MONO & N^G,$N^{G'}$-DIMETHYL ARGININE, N^3-METHYLHISTIDINE
MYOSINS	Nε-MONO, DI & TRIMETHYLLYSINE, DIMETHYL ARGININE, N^3-METHYLHISTIDINE
ACTINS	Nε-MONO, DI & TRIMETHYLLYSINE, N^3-METHYLHISTIDINE
CYTOCHROME C	Nε-TRIMETHYLLYSINE
CYTOCHROME C557	Nα-DIMETHYLPROLINE
RIBOSOMAL PROTEINS	Nε-MONO, DI & TRIMETHYLLYSINE, DIMETHYL ARGININE, Nδ-METHYLGLUTAMINE, N-MONO & TRIMETHYLALANINE, N-METHYLMETHIONINE, METHYLESTER OF GLUTAMIC & ASPARTIC ACID
First discovered : 1950-1980	

data taken from Paik and Kim (1980d)

been observed in proteins of various functions, including myosin, actin, flagellin, ribosomes, cytochrome c, elongation factors and calmodulin, to name a few (DeLange *et al.*, 1969; Hardy and Perry, 1969; Kuehl and Adelstein, 1969; Parish and Ada, 1969; Weihing and Korn, 1970; Borun *et al.*, 1972; Chang *et al.*, 1978; Alix *et al.*, 1979; Hempel *et al.*, 1979; Arai *et al.*, 1980; Hiatt *et al.*, 1982; Klee *et al.*, 1982; Van Hemert *et al.*, 1984; Schaeffer *et al.*, 1987; Dever *et al.*, 1989; Paik *et al.*, 1989).

Given the void of information as far as protein methylation using iodomethane is concerned, and given the fact that reductive dimethylation of protein amino groups (Means and Feeney, 1968) has been very commonly used in structure-function studies, it seems unusual that the reaction of iodomethane with various functional groups of native proteins has not been characterized and has escaped attention as a possible alternative reagent for protein derivatization or a tool of structure-function studies. Its slight solubility in water is one factor that may have conveyed the idea that iodomethane would not be a suitable reagent for general use. Its low reactivity with respect to other reagents such as acetic anhydride is perhaps another factor. Ironically, these two properties of iodomethane could have been used to advantage since mild reagents are often preferred to carry out benign protein modifications. Other methylating reagents of varying specificity and potential are dimethylsulphate (Wissmann and Hillen, 1991), diazomethane (Herriott, 1947), acidic methanol (Wilcox, 1967), trimethyloxonium tetrafluoroborate or hexachloroantimonate and isoxazolium salts (Olah, 1973; Lundblad, 1995).

Advantageous features of protein methylation over other protein chemical modifications

Structure-function studies often require conservative modification of proteins. Reductive methylation is one method of choice and has been carried out with retention of bioactivity in many

cases (Means and Feeney, 1971a). The success of reductive methylation is due to the benign nature of the modification which preserves protein structure. In contrast, protein acetylation, one of the next smallest modifications, represents nevertheless a more dramatic change of steric bulk, and local and global electrostatic character of the protein. Charge interactions and steric requirements are among two of many protein structure-function determinants which should be considered before carrying out a chemical modification of polar residues:

(i) **Charge interactions**. Charge interactions of protein are partially responsible for the maintainance of structure, however, their relative importance is difficult to assess since charge interaction energies are very much dependent on the dielectric strength of the medium (Fersht, 1985). That being said, methylation of amino groups in protein occurs with conservation of charge over a wide pH range, whereas other modifications such as acetylation and 2,4-dinitrophenylation disturbs the original charge and may perturb the native structure.

(ii) **Steric considerations**. The manner in which protein residues pack together ultimately determines protein structure. One consequence of chemical derivatization of protein is the change of steric bulk in the area immediately surrounding the modified functional group. While it may appear that a change of steric bulk should always perturb structure and function, predicting the outcome of a specific derivatization is very often subject to interpretation. For example, a large group may not be considered bulky if there is nothing proximal to interact with it. This is consistent with the fact that substitutions involving changes of steric bulk in proteins are often more serious within the protein core where residues are closely associated (Serrano *et al.*, 1992). Nevertheless, replacements where a small group is replaced by a much larger group at the surface

of protein can still potentially have more serious consequences than a more isosteric replacement. Sterically speaking, the methyl group, which typically replaces a proton and some of its associated hydration shell, represents an extremely small change, and while there are enzymes which can differentiate such minimal replacements (DiMaria *et al.*, 1982; Rowe *et al.*, 1986), the general effect of the modification on proteins is presumably small.

The merit of iodomethane for protein modification

Just as protein methylation has merit, the use of iodomethane appears to be an attractive alternative to reductive methylation using formaldehyde and borohydride. As mentioned above, reductive methylation of amino groups with formaldehyde and borohydride has been extensively used and shown to produce conservative modifications with retention of biological activity in many cases (Means and Feeney, 1971a). It is expected that iodomethane should have the same desirable properties. Since iodomethane is small and uncharged, its reaction with proteins should be unaffected by electrostatic interactions and minimally by steric factors. Its reactivity towards distinct functional groups could be advantageous in effecting different modifications under appropriate conditions. A theoretical study reports that the interaction energy between ammonium and carboxylate groups decreases with each level of methylation of the amino group (Mavri and Vogel, 1994). While ionic interactions (i.e., salt bridges) do play a role in maintaining protein structure, a maximum decrease of 25% upon quaternization was calculated and this decrease would not appear to be a detriment to structure. In fact, the authors suggest that such a subtle change of interaction energy in lysyl residues may modulate physiological function.

Iodomethane is commercially available in various isotopic forms, which improves the scope of its applicablity in structure-function studies. In particular, by using [^{13}C]iodomethane, protein

modifications can be conveniently monitored by [^{13}C]NMR spectroscopy. In light of its known reactivity in organic syntheses (March, 1985a), iodomethane is a candidate for the modification of not just amino groups as is the case of reductive methylation, but potentially of imidazole, thiol, dialkylsulphide, phenol and carboxyl residues in protein. Also notable is that while reductive methylation only dimethylates amino groups in protein, thiol, amino and imidazole groups may be sequentially methylated to completion with iodomethane (March, 1985b; Cowgill, 1957). These derivatives, if formed, would in some cases represent changes that have not been previously reported as *in vitro* and/or *in vivo* modifications of native proteins. At any rate, by using iodomethane, the door to entirely novel structure-function studies could be opened.

Previous use of iodomethane as a structure-function tool

Proteins have complicated three-dimensional structures in which some amino acid residues are solvent exposed and others are buried. Functional groups have the highest probability of reaction if they are located at the surface of the protein where they are readily available to reagent (Means and Feeney, 1971b). The reaction in which iodomethane methylated methionine–29 of ribonuclease A is one example of the potential utility of this reagent in structure-function studies. While ribonuclease A has several methionine residues, methionine–29 is the most easily derivatized and this is consistent with the fact that in the X-ray structure, methionine–29 is solvent exposed, where it is readily accessible to reagent (Link and Stark,1968; Richards and Wyckoff, 1971). Thus far, the S-methyl approach was used to study apomyoglobin, basic myelin protein, basic pancreatic trypsin inhibitor, ribonuclease and α-chymotrypsin (Jones *et al.*, 1975; Jones *et al.*, 1976; Deber *et al.*, 1978; Harina *et al.*, 1978; Jaeck and Benz, 1979; Harina *et al.*, 1980; Smith *et al.*, 1980).

Focus of the present research

The S-methylation example of ribonuclease A (previous page) was among a handful of structure-function studies that used a protein specific protocol to characterize and rationalize a particular modification with iodomethane. A more general approach to observe protein systems has often involved the use of NMR spectroscopy. It followed that the next step in the development of iodomethane for use as a protein reagent rested on better characterizing the reaction of iodomethane with many protein functional groups and was felt that the technological advances of [^{13}C]NMR spectroscopy and the availability of high field instruments would provide a means of monitoring its reaction with proteins by using [^{13}C]iodomethane.

In the present research, we demonstrate that iodomethane is sufficiently reactive with proteins to be an extremely useful chemical modification reagent for investigating structure-function relationships of proteins. It should be possible using this protocol to examine group reactivities by NMR spectroscopy in order to obtain information concerning a protein's native structure, its folding pathway or any other structure-function relationship. The objective of the work was to define the conditions for reaction of iodomethane with proteins, characterize the types of derivatives formed, and determine the ^{13}C-chemical shifts for these derivatives. The information would then be available to facilitate the characterization of proteins modified with iodomethane or other methylating reagents for the purpose of carrying out structure-function studies.

Rationale of the present research

Following reaction of the insecticidal protein from *Bacillus thuringiensis* with iodomethane, biological assays are carried out in order to provide supporting evidence that proteins can tolerate

substantial reaction before structure is compromised. Subsequently, methylation of protein is carried out with [^{12}C], [^{13}C] or [^{14}C]iodomethane and the analysis of derivatives by (i) NMR spectroscopy of protein or (ii) NMR and mass spectroscopies of digested and purified peptide fragments. The model proteins used are bovine insulin, ribonuclease, serum albumin and α-chymotrypsin. Of these, most of the initial work is carried out on insulin because of its small size and relatively few reactable groups, which should simplify interpretation of the results (Figure 2.1.1). In addition, its three-dimensional structure is well characterized (Bordas *et al.*, 1983).

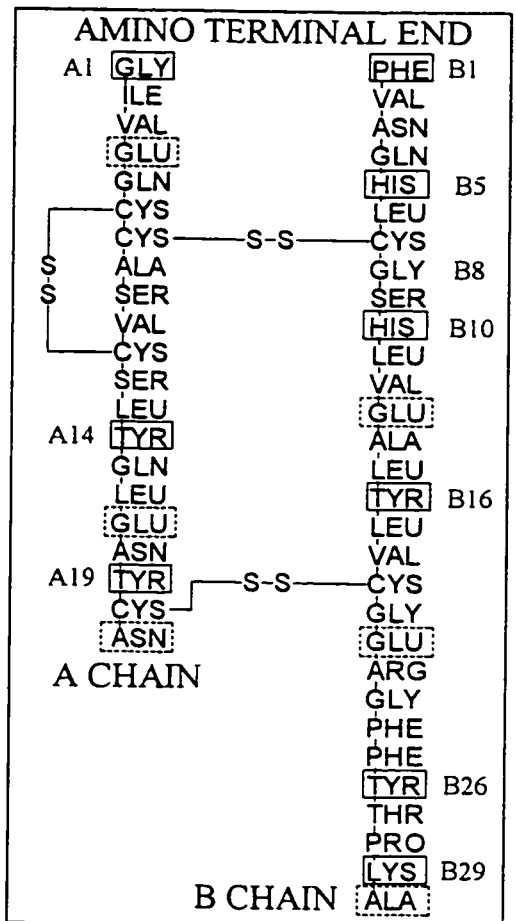


Figure 2.1.1: Primary sequence of insulin with potentially reactable groups boxed. (Solid boxes are confirmed reactions)

(Experiment A): Analysis of methylated insulin and its purified peptide fragments

The strategy adopted (Figure 2.1.2) is to use iodomethane of two isotopic forms, ^{13}C and ^{14}C, to label bovine insulin. [^{13}C]NMR spectroscopy will be used to identify derivatized functional groups in insulin by matching peak resonance positions to those of appropriate ^{13}C-methylated standard compounds. The purpose of the standards is to provide a rapid means of identifying the methylated residues of any protein by NMR spectroscopy. However, it may happen that protein structure (microenvironment) perturbs the normal environment of a methylated residue to the point where its ^{13}C-resonance no longer coincides with that of the standards, thus making identification impossible. It is therefore necessary to demonstrate that the

chemical shift of methylated residues are relatively independent of their microenvironment in the protein so that an identification can be made with standards. The [^{13}C]NMR spectra of methylated insulin is accordingly obtained in the presence and absence of denaturants to gain insight as to how the chemical shift varies with the microenvironment of the protein. The chemical shifts of standard ^{13}C-methylated amino acids are also acquired at two pH values and in the presence and absence of urea to see if recorded chemical shifts of reacted residues show any dependency to changes in the aqueous medium. This precautionary measure is taken so that any chemical shift effect attributable to general environmental factors and not the protein microenvironment, such as pH, urea and salt, can be corrected.

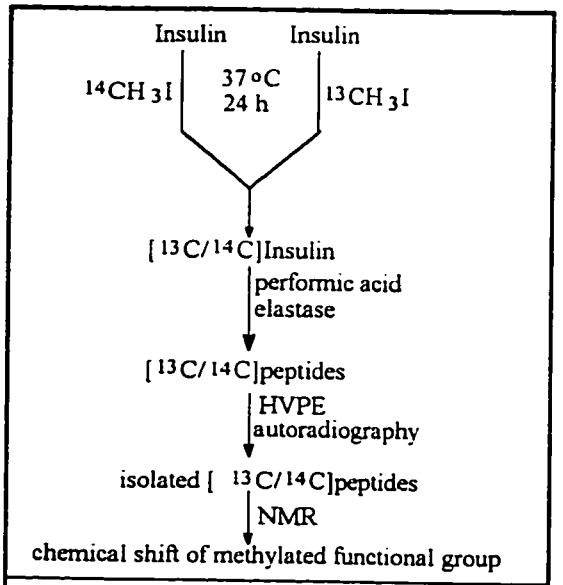


Figure 2.1.2: Strategy adopted to identify individual methylated residues of insulin.

Proteins contain several kinds of functional groups, so spectra of protein derivatives may be difficult to interpret. More specifically, one major problem in the interpretion of results by NMR spectroscopy may arise if a protein has many of the same type of reactive group. As an example, iodomethane could potentially methylate several lysyl residues which span the length of the protein sequence. If these ^{13}C-chemical shifts should superimpose, then the net resonance signal obtained is a sum of individually reacted residues at different sequence positions. While this may improve sensitivity,

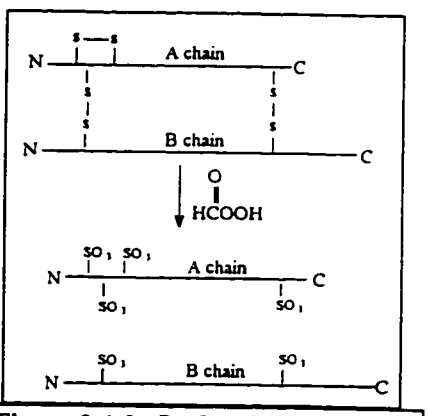


Figure 2.1.3: Performic acid oxidation of insulin A and B chains.

the fraction contributed by each lysyl residue to signal intensity, and hence, the extent of reaction of individual lysyl residues cannot be determined by NMR spectroscopy. With insulin, this problem could arise following methylation, should the two histidyl, four tyrosyl and six carboxyl residues have superimposing resonances. In anticipation of this problem, reacted functional groups of insulin will be separated into smaller peptides and reanalyzed by NMR spectroscopy so that individual groups can be observed and their reactivities quantified (Figure 2.1.2).

The degradation of methylated insulin makes use of chemical and enzymatic procedures. Following [^{13}C]NMR analysis of the methylated insulin, the disulphide bridges of the methylated protein are oxidatively cleaved (Figure 2.1.3) (Hirs, 1956) and the protein is subjected to an elastase digest (Wilkinson, 1986). The digest is carried out to ensure there will be at most one methylated residue in each peptide fragment. The resulting peptide fragments are separated by high voltage paper electrophoresis at two pH values in an apparatus described by Brown and Hartley (1966). The presence of trace ^{14}C-labeled peptides are used to advantage at this step to facilitate detection of electrophoretically separated peptide fragments on the paper by autoradiography. Once collected, (i) the [^{13}C]NMR spectra of individual methylated peptides are recorded, (ii) the reacted groups of bovine insulin are tentatively assigned on the basis of peptide fragment mobilities and chemical shift data and (iii) mass analysis of the peptides is used to corroborate the results of the investigation.

(Experiment B): Exploring the generality of protein reactions using [^{13}C]iodomethane with [^{13}C]NMR analysis

Bovine insulin should provide adequate results to determine if protein methylation using [^{13}C]iodomethane is feasible, however, additional work using other model proteins is required to

establish the generality of the procedure. To this end, bovine ribonuclease, serum albumin, and diisopropylphosphoryl-α–chymotrypsin are treated with [^{13}C]iodomethane in a similar fashion. The initial inhibition of α–chymotrypsin is carried out as a precautionary measure to remove any possibility of autolysis over the course of reaction. As was found to be the case for methylated insulin, protein derviatives are analyzed in urea solution to enhance solubility. The type and yield of methylated functional groups obtained in each protein is then identified using appropriate methylated standard compounds. These chemical shift values can then be compared amongst different proteins to determine if the results are consistent.

(Experiment C): The use of iodomethane as a criterion of protein purity.

In many cases, preproteins which are nonhomogeneously proteolyzed to the mature form may have N and C-terminal heterogeneities of primary structure. While such subtle differences are not observable by sodium dodecyl sulphate-polyacrylamide gel electrophoresis, the commonly accepted criterion of protein purity (Hames, 1990), a heterogeneity of the N-terminal amino acid should be observable by [^{13}C]NMR spectroscopy following reaction of the protein with [^{13}C]iodomethane. The unknown N-terminal heterogeneity can be presumed to be one of the common amino acid residues, however, its direct assignment in the intact protein following methylation is not a feasible strategy. This is so, because the value of the observed chemical shift partially depends on the identity of the adjacent amino acid residue in the protein (Taralp and Kaplan, 1997). In well characterized proteins, N-termini can be identified using an appropriate ^{13}C-trimethylated dipeptide or tripeptide standard. In unknown cases, however, it is difficult to predict what choice of standard peptide is appropriate to identify the N-terminus, as both the unknown N-terminal and adjacent residues contribute to the observed chemical shift value. A

different approach is therefore required. Unlike acetamide derivatives, advantage can be taken of the acid stability of trimethylated amino groups in protein. Following conventional acid hydrolysis (Darbre, 1986a), the derivatized N-terminal residues of proteins are liberated as ^{13}C-trimethylated amino acids. The heterogeneity is then identified by assignment of its chemical shift value to that of one of the twenty standard ^{13}C-trimethylated amino acids.

2.2 Materials and Methods

Proteins. Bovine insulin, α–chymotrypsin, ribonuclease and serum albumin were purchased from Sigma Chemical Company. Inactive diisopropylphosphoryl-α–chymotrypsin was prepared by incubation with diisopropylfluorophosphate (Jansen *et al.*, 1949; Oppenheimer *et al.*, 1966; Darbre 1986b). The insecticidal toxin from *Bacillus thuringiensis* was obtained and quantified according to the method of Bietlot *et al.* (1989). The protein was dialyzed at pH 4.5 against acetic acid and finally against water prior to methylation.

Standards. Histidinamide was purchased from Bachem. Other amino acid standards or precursors were purchased from Sigma Chemical Company.

Chemicals and Solvents. [^{13}C]Iodomethane, diisopropylfluorophosphate and citraconic anhydride were purchased from Sigma Chemical Company. [^{14}C]Iodomethane (specific activity = 56 mCi/mmole) was obtained from Amersham Canada Ltd. All other chemicals, reagents and solvents were high purity preparations obtained from commercial sources.

^{12}C-Methylation of the insecticidal toxin from *Bacillus thuringiensis* at pH 10.5. Toxin (14 mg), was placed in a screw-capped vial and dissolved in 200 mM sodium metaborate buffer (15 ml), pH 10.5. [^{12}C]Iodomethane (250 µl) was added and the vial was sealed tightly. The biphasic mixture was stirred at room temperature and aliquots (2.5 ml) were removed at time points up to 24 h. The reaction mixtures were transferred into a dialysis bag (3500 molecular weight cut-off) and dialyzed twice against 4L dH_2O and lyophilized. Quantification and CF-1 insect cell bioassays of methylated and control toxin were performed according to published procedures (Choma and Kaplan, 1990) by Dr. Ross E. Milne of the Forest Pest Management Institute, Forestry Canada. Lyophilized samples were incubated 18h in 0.1M 3-cyclohexylamino-1-propanesulphonic acid buffer (1 ml) pH 10.5 and centrifuged (10K, 10 minutes). An aliquot of the supernatant was quantified at 595 nm using the dye binding method of Bradford (1976) and the remainder was tested as a series of dilutions to examine the threshold response.

^{13}C-Methylation of bovine insulin, diisopropylphosphoryl-α–chymotrypsin, serum albumin and ribonuclease at pH 7.5 and pH 10. Protein (40 mg), was placed in a screw-capped vial and dissolved in 200 mM sodium phosphate buffer (10 ml) pH 7.5, or 200 mM sodium metaborate buffer (10 ml) pH 10. A 1:1 (v/v) solution (250 µl) of [^{13}C]iodomethane in acetonitrile was added below the surface of the solution. The vial was sealed tightly and the biphasic mixture shaken at 37°C for 24 h in a temperature controlled water bath. The reaction mixture was transferred to a dialysis bag (3500 molecular weight cut-off) and dialyzed twice against 4L dilute HCl (pH 2) and lyophilized.

^{13}C-Methylation of insulin at pH 7.5 in 9M urea. Insulin (50 mg) was placed in a screw-capped vial and dissolved in 9M urea, 200 mM phosphate buffer solution (10 ml). A 1:1 (v/v) solution of [^{13}C]iodomethane (500 µl) in acetonitrile was added below the surface of the solution. The vial was screw capped and sealed with Parafilm. The biphasic mixture was shaken at 37°C for 24 h. The reaction mixture was transferred to a dialysis bag (3500 molecular weight cut-off) and dialyzed twice against 4L dilute HCl (pH 2) and lyophilized.

^{13}C-Methylation of citraconylated insulin at pH 10. Insulin (50 mg) was placed in a screw-capped vial and dissolved in dH_2O (5 ml). The pH was adjusted to 9 with 5N NaOH. Four aliquots (25 µl) of citraconic anhydride were added with rapid stirring, and the pH was maintained with the addition of additional base. The reaction was terminated after 30 min when the pH remained constant. Sodium bicarbonate was added to the protein solution to a final concentration of 0.5 M and the pH was adjusted to 10 with 1 M NaOH. A 1:1 (v/v) solution (300µl) of [^{13}C]iodomethane in acetonitrile was added below the surface of the solution. The vial was sealed tightly and the biphasic mixture shaken at 37°C for 24 h. The reaction mixture was transferred to a 3500 molecular weight cut-off dialysis bag, dialyzed against 2 X 4 L dilute HCl at pH 2, and lyophilized.

^{14}C-Methylation of native insulin followed by characterization of individual reactive groups:

Methylation of insulin with [^{14}C]iodomethane. Bovine insulin (40 mg) was dissolved in 200 mM phosphate buffer, pH 7.5 (10 ml), in a glass scintillation vial. A 1:1 (v/v) solution of [^{14}C]iodomethane (500 µCi total activity) in acetonitrile was added below the surface of the

solution. [^{12}C]iodomethane (500 µl) was also added. The scintillation vial was screw capped and sealed with Parafilm. The biphasic mixture was reacted at 37°C with shaking for 24 h. The reaction mixture was transferred to a dialysis bag (3500 molecular weight cut-off), dialyzed against 2 X 4L dilute HCl (pH 2) and lyophilized.

Performic acid oxidation of methylated insulins. The procedure followed was that of Hirs for ribonuclease (1956). Half of the ^{13}C and ^{14}C-methylated insulins (2 X 20 mg) at pH 7.5 were dissolved in formic acid (2.5 ml) and combined to a total volume of 5 ml in a round bottom flask (40 mg total protein). Performic acid was made by adding 30% H_2O_2 (0.5ml) to formic acid (9.5ml) in a stoppered flask and allowed to stand at room temperature for 2 h. The performic acid and the formic acid solution of methylated insulins were cooled in an ice bath. The performic acid was then added to the protein solution and allowed to react at 4°C for 3 h. The reaction was stopped by adding cold dH_2O (350 ml) and the mixture lyophilized. Final traces of performic acid were removed by dissolving the material in dH_2O (20 ml) and relyophilizing.

Elastase digest. The procedure followed was that reported by Wilkinson (1986). Elastase typically cleaves C-terminal to glycine, alanine and serine residues (Hartley and Shotton, 1971). The above performic oxidized protein was dissolved in 1% NH_4HCO_3, pH 8.0 (20ml). To this solution was added porcine elastase (2 mg) dissolved in the same buffer (500 µl). The mixture was digested at 37°C for 4 h and lyophilized.

Separation of peptides by paper electrophoresis at pH 6.5 and tentative identification of the peptide fragments on the basis of mobility calculations. The elastase digested and performic

acid oxidized methylated insulins were applied to 3MM paper and subjected to high voltage paper electrophoresis in pH 6.5 buffer (acetic acid:pyridine:water, 3:100:900 by volume) at a voltage gradient of 40 volts/cm for 45 minutes. X-ray film was exposed at room temperature to the paper for 72 h and was developed to identify the radioactive bands. The mobility of each band with respect to dansyl sulphonic acid was calculated after correcting for electroosmosis by noting the position of a dansyl arginine marker. Bands were eluted using pH 6.5 buffer and purified further.

Purification of methylated peptides by paper electrophoresis at pH 2.1. The above eluted bands were dissolved in a minimum of pH 2.1 buffer (formic acid:acetic acid:water, 1:4:45), applied to 3MM paper and subjected to high voltage paper electrophoresis in pH 2.1 buffer at a voltage gradient of 60 volts/cm for 40 minutes. X-ray film was exposed to the paper for 72 h to identify the bands corresponding to radioactive peptides. The mobilities of these bands were calculated with respect to dansyl arginine after correcting for electroosmosis by noting the position of dansyl sulphonic acid. Each band was eluted from the paper with pH 2.1 buffer, the solution lyophilized, and the [^{13}C]NMR spectra recorded.

Acid hydrolysis of ^{13}C-methylated insulin and ^{12}C-methylated toxin. The protein derivative was suspended in 6N HCl, degassed, and sealed *in vacuo* in a pyrex hydrolysis tube. The hydrolysis was carried out for 72 h and 24 h respectively, at 110°C (Darbre, 1986a).

Synthesis of [^{13}C]NMR amino acid standards:

NH_2-His(Im^+Me_2): Acetyl histidine (40 mg) was dissolved in 200 mM metaborate buffer (10 ml), pH 10.5, and a 1:1 (v/v) solution (300µl) of iodomethane in acetonitrile was added

below the surface of the solution. The sealed reaction was shaken at 37°C for 48 h and the pH was monitored and reset periodically during this time. The entire solution was then made 6M in HCl by the addition of concentrated HCl and the compound was hydrolyzed *in vacuo* for 24 h at 110°C. The hydrolyzed compound was purified by ion exchange chromatography using Dowex 50X8-200 resin in the acid form. The compound was shown to be dimethyl histidine by [^{1}H] and [^{13}C]NMR spectroscopy.

Ac-NH-His($Im^{+}Me_2$)-NH_2: Histidine amide (40 mg) was dissolved in dH_2O (5 ml) and acetylated with acetic anhydride (5 X 20 µl), while maintaining the pH at 9 by the addition of 5N NaOH. Sodium metaborate was added to a final concentration of 200 mM and the pH was adjusted to 11 with the addition of 1M NaOH. The solution was transferred to a screw-capped vial and a 1:1 (v/v) solution (200 µl) of [^{13}C]iodomethane in acetonitrile was added. The capped biphasic mixture was shaken for 72 h at 37°C and the aqueous layer was separated. Acetate and other anions were removed prior to analysis by ion exchange using a Dowex-1 anion exchanger in the hydroxide form. The resulting solution was dried and the [^{13}C]NMR spectrum of the compound was recorded. An aliquot of the compound was then hydrolyzed in 6M HCl and purified by ion exchange chromatography using Dowex 50X8-200 resin. High voltage paper electrophoresis at pH 6.5 was performed and the mobility of the residue was identical to the dimethylhistidine produced by the alternate route above.

Ac-NH-Tyr(OMe)-NH_2 and NH_2-Tyr(OMe): N-Acetyl-L-tyrosine–amide (100 mg), was dissolved in 0.5 M sodium metaborate buffer (10 ml), pH 11, in a glass screw-capped vial. A 1:1 (v/v) solution of [^{13}C]iodomethane (300 µl) in acetonitrile was added. The biphasic mixture was

sealed and shaken at 37°C for 24 h. The solution was cooled in an ice bath to precipitate the methylated acetyl tyrosine amide. This was extracted with chloroform (3 X 10 ml) and the extracts were combined and dried under a stream of nitrogen to afford Ac-NH-Tyr(OMe)-NH_2 (20 mg). An aliquot was hydrolyzed in 6N HCl for 4 h in the presence of phenol. The hydrolysate was subjected to high voltage paper electrophoresis at pH 2.1 (20 min; 60 V/cm) alongside a commercially available ^{12}C-standard of O-methyltyrosine. The mobility of the hydrolysate and standard were shown to be identical using ninhydrin stain, lending confidence that the ^{13}C-derivatives were methylated on the phenolic hydroxyl group. Distortionless enhancement by polarization transfer analysis was performed on commercially available O-methyltyrosine at pH 10 and the position of the carbon resonance of the methyl group was verified. The ^{13}C-standards were used in place of commerically available ^{12}C-analogues because the poor solubility of these compounds at pH 7.8-8 made spectral acquisitions inconvenient without isotopic enrichment.

H_2N-Lys(ε–$^+NMe_3$), H_2N-Lys(ε–NMe_2) and H_2N-Lys(ε–NHMe): Polylysine hydrobromide (30 mg) was dissolved in 5mM sodium deuteroxide (3 ml) and a 1:1(v/v) solution (20 µl) of [^{13}C]iodomethane in acetonitrile was added. The pH was adjusted to 10.5 without buffering and the sealed vial was shaken at 37°C for 4.5 h. After NMR analysis, the polylysine derivative was hydrolyzed in 6N HCl for 18 h to liberate the Nε–trimethylated free amino acid. It was also of interest to identify the ^{13}C-chemical shift positions of the mono and dimethylated derivatives. To this end, the experiment was repeated with a trace amount of [^{13}C]iodomethane and sequentially forced to higher alkylation states with the addition of various amounts of [^{12}C]iodomethane (This method was successful in general for any amino acid Nα group). The progression of reaction, as monitored by NMR spectroscopy, indicated that the amino groups were increasingly reactive

towards iodomethane with each alkylation, as expected (March, 1985b). Commercially available Nε–monomethyllysine was used as an additional verification.

Me_3N^+-Gly-Leu, Me_3N^+-Gly-NH_2 and Me_3N^+-Phe–Gly-Gly: Glycylleucine, glycinamide or phenylalanylglycylglycine (13 mg) were placed in a glass vial and dissolved in 5 mM sodium deuteroxide (1 ml). A 1:1 (v/v) solution (20 ml) of [^{13}C]iodomethane in acetonitrile was added and the sealed reaction was shaken 9 h at 37°C in a water bath.

NH_2-His(Im-1-Me), NH_2-His(Im-3-Me), Ac-NH-His(Im-1-Me)-NH_2, Ac-NH-His(Im-3-Me)-NH_2: The monomethyl histidines were commerically available and were characterized by NMR spectroscopy without any further treatment. Preliminary attempts to generate the monomethyl acetylhistidinamides were hampered by purification problems, due in part to the similar chromatographic properties of the derivatives and due also to the propensity for imidazole groups to rapidly proceed to the dimethylated form. The following steps were carried out: i) Acetylate the amino terminus of histidinamide. ii) React the resulting acetylhistidinamide with a small amount of [^{13}C]iodomethane. iii) Characterize the resulting monomethyl acetylhistidinamides without further purification.

Histidinamide (23 mg) was acetylated in water (3 ml, pH 9) using acetic anhydride (25 µl) while maintaining the pH by addition of 5N NaOH. Complete acetylation was verified using a cadmium ninhydrin stain for amino groups. The pH of the solution was adjusted to 10.5 and a 1:1 (v/v) solution (100 µl) of [^{13}C]iodomethane in acetonitrile was added. The biphasic mixture was sealed and shaken at 37°C for 4 h. The pH was again readjusted and shaking was recommenced for another 2 h. The reaction was terminated and the mixture was characterized as follows:

i) An aliquot was removed and hydrolyzed *in vacuo* for 24 h in 6N HCl. The histidine derivatives were separated from other contaminants by ion exchange using a column packed with Dowex 50X8-200 resin and then co-eluted for analysis by high voltage paper electrophoresis. Standards of both monomethyl histidines and the dimethylhistidine standard (made and characterized above) were run alongside the eluted derivatives at pH 6.5 for 55 minutes using a voltage gradient of 60V/cm. The predominant compound corresponded to histidine, which indicated that histidine was always present and in excess of any monomethyl derivatives. Despite this precaution, the presence of trace amounts of dimethylhistidine was still apparent. More importantly, reasonable amounts of both monomethyl histidines were observed. There appeared to be a five fold excess of the 1-methyl derivative over the 3-methyl derivative on the basis of ninhydrin colour yield and an eight fold excess on the basis of proton NMR spectroscopy.

ii) An aliquot was added to D_2O (200 µl) and buffered at either pH 7.5 or 10. A thorough characterization of the mixture at both pH values was attempted by NMR spectroscopy on a Bruker 500MHz NMR spectrometer. The proton spectrum yielded a doublet of doublets and an apparent quartet of doublets in a region where the ^{13}C-enriched methyl groups were expected. These splittings were useful for identifying the monomethyl derivatives since the proton couplings of methyl protons in 1-methyl acetylhistidinamide and 3-methyl acetylhistidinamide are different (not shown). Integration of these peaks indicated that the 1-methyl derivative (doublet of doublets) was in eight fold molar excess of the 3-methyl derivative (essentially a doublet of doublet of doublets). This observation was corroborated by the difference in ninhydrin colour yield of the two derivatives. A heteronuclear multiquantum coherence correlation experiment was performed on each derivative. 1-Methyl acetylhistidinamide and 3-methyl acetylhistidinamide

derivatives gave corresponding carbon resonances of approximately 31.9 ppm and 33.9 ppm respectively. Although the cross peaks were grossly distorted and large, due to the inability of the algorithm to compensate for 100% ^{13}C-enrichment, the assignment was reliable since there were only two resonances present in the range where the methyl carbons were expected. Furthermore, the correct assignment of these resonances was corroborated by the closely matching carbon spectra of the monomethyl histidines which were obtained commercially.

Me_3N^+-Cys(NH_2)-S-S-Cys(NH_2)N^+Me_3: Cystine dimethylester dihydrochloride (10mg) was dissolved in dH_2O (4 ml) and the pH was adjusted and maintained at 9.5 by the addition of concentrated ammonia. After 45 min the solvent was evaporated under reduced pressure. The sample was suspended in D_2O (1 ml) and the supernatant containing the cystine diamide was reacted at 37°C for 8 h with [^{13}C]iodomethane (20 µl). The pH meter reading was maintained at 10 by the addition of 1 M sodium carbonate (100 µl).

H_2N-Met(S^+Me_2), H_2N-Cys(SMe) and H_2N-Cys(S^+Me_2): Distortionless enhancement by polarization transfer analysis of commercially available S-methyl methionine was performed to ascertain the ^{13}C-resonance of the S-methyl group. Acetyl cysteine (100 mg) was dissolved in D_2O (3ml) and a 1:1(v/v) solution (30 µl) of [^{13}C]iodomethane in acetonitrile was added. The reaction was shaken at 37°C (pH 10, 24 h) and aliquots were removed for NMR characterization. Only one product was observed. In order to characterize the dimethylated cysteine, a large excess of [^{12}C]iodomethane (150 µl) was added to the remaining reaction mixture and the biphasic reaction was sealed and shaken at 37°C for another 48 h. The spectrum showed two resonances

which were attributed to an R/S mixture of acetyl-NH-Cys($^{-}S^{12}Me^{13}Me$), but not a mixture of acetyl-NH-Cys($^{-}S^{12}Me^{13}Me$) and acetyl-NH-Cys($^{-}S^{13}Me^{13}Me$).

Attempted methylation of arginine, tryptophan, threonine and serine side-chains with iodomethane: Acetyl arginine (10 mg) was dissolved in 0.1 M NaOH and 25 µl of a 1:1 (v/v) solution of [^{13}C]iodomethane in acetonitrile was added and reacted 24 h with shaking in a sealed vial at 37°C. The resulting ^{13}C-spectrum was unfortunately to complicated to interpret. The reaction was repeated at pH 10 for 24 h in order to determine if methylations of arginine could occur under the reaction conditions relevant to proteins. This time it was shown that no reaction had occurred. In an analogous reaction, no ^{13}C-methyl group appeared for acetyl tryptophan, indicating that the indole ring was unreactive. Isopropanol and ethanol, which were model compounds for serine and threonine, produced ^{13}C-spectra that were complicated and inconclusive when reacted at very high pH values. Fortunately, with the exception of a large ^{13}C-MeOH peak, no resonances appeared which could be confused with methyl resonances of the other standards. The reactions were repeated at pH 10.5 for 36 h in order to determine if methylation of the hydroxyl group could occur using reaction conditions relevant to protein derivatizations. This time it was observed that no reaction had occurred.

Other standard chemical shift values: Other standards (Table 2.3.2) and some of those above were either obtained commercially [NH_2-Arg($N^{G}Me$), NH_2-Arg($N^{G,G}Me_2$), NH_2-Arg($N^{G,G'}Me_2$), NH_2-Trp(Ind-1-Me), NH_2-Thr(OMe) and NH_2-Ser(OMe)] or obtained by reaction with [^{13}C]iodomethane in the vapour phase, as described in chapter 3. Typically, 1 mg of starting material would be prepared as the free base and dried. Reaction with [^{13}C]iodomethane (4 µl)

would be carried out at 110°C for 3 days. The standard Me_3N^+-Ile–Val-NH_2, in particular, was prepared sequentially, by esterifying Me_3N^+-Ile (1 mg) in methanolic hydrogen chloride (0.1 N HCl in anhydrous methanol, 1 ml, 25°C, 12 h), removing solvent, and reacting (48 h, 25°C) in a 1:1 (v/v) solution (500 µl) of N,N-dimethylformamide and pyridine containing valinamide (4 mg).

NMR Spectra. [^{13}C]NMR spectra were obtained using a Gemini 200 MHz spectrometer. Methylated protein samples were analyzed in 9M urea (90% D_2O, 100 mM sodium phosphate) which gave a pH meter reading (p^{2}H) of 7.8-8.0. Insulin, reacted at pH 7.5 in particular, was recorded both in the presence and absence of 9M urea. Standards were adjusted to give a pH meter reading (p^{2}H) of 7.8-8 or 10 (100mM sodium carbonate buffer), both in the presence and absence of urea. Distortionless enhancement by polarization transfer experiments were performed to assign methyl groups of standards where necessary. Chemical shift values were relatively insensitive to the presence of salt so NMR spectroscopy of isolated peptides and hydrolyzed insulin were carried out in 0.1% ND_4DCO_3 solution to permit relyophilization. Acetonitrile (30 µl), with a methyl group ^{13}C-chemical shift of 1.70 ppm (Breitmaier and Voelter, 1987) was added to reference peak resonances. The nitrile carbon resonance appears at 119.93 ppm and urea is observed at 163.40 ppm. Unless otherwise indicated, spectral acquisitions of protein and peptides were terminated after 1024 scans, with 1s intervals between pulses.

2.3 Results and discussion

2.3.1 Supporting evidence that reaction with iodomethane is nondisruptive to protein structure.

The high pH-stable insecticidal protein produced from *Bacillus thuringiensis* was reacted at pH 10.5 with excess iodomethane. By 2 h elapsed time, most of the lysyl (70%) and 50% of the

histidyl residues were reacted, as indicated by amino acid analysis. Furthermore, the alkylation coincided with a dramatic decrease in solubility but essentially no loss of bioactivity (Table 2.3.1). These observations suggested that these charged groups were mainly surface exposed in the protein and not crucial for maintaining bioactivity. After 4 h of reaction, however, the protein lost all toxicity and it was proposed that loss of structure coincided with the loss of activity. Internal tyrosyl residues of proteins are generally protected from reaction (Means and Feeney, 1971c) but are in principle reactive at pH 10.5 when they deprotonate (Creighton, 1993). Furthermore, the contribution of tyrosyl residues in maintaining protein structure is a feature common to many proteins (Means and Feeney, 1971c). While the course of reaction of the tyrosyl residues was not monitored in the *Bacillus thuringiensis* toxin, it is believed that disruption of tyrosyl interactions by methylation of the phenolic hydroxyl functions was (i) either the cause of loss of structure and hence bioactivity, or (ii) a consequence of the loss of structure for other reasons, which resulted in subsequent methylation of exposed tyrosine groups. Wherever the truth lies, the results indicated that the protein could withstand substantial methylation before losing toxicity and structure.

Table 2.3.1: Protein recovery and results of *Bacillus thuringiensis* toxin bioassay following extensive methylation at pH 10.5

Reaction Time (Hours)	Protein Recovery (milligram/ml)	Threshold Response (nanogram)
24 (No iodomethane)	0.34	3.4
0	0.37	3.7
1	0.13	6.5
2	0.13	6.5
4	0.12	>120

2.3.2 Standards: NMR properties

Table 2.3.2 reports the ^{13}C resonances obtained for free, N-acetylated and C-amidated amino acid standards. The chemical shifts of free trimethylamino acid standards spanned a moderate

Table 2.3.2: ^{13}C-Chemical shift resonances of methylated amino acid and peptide standards				
Standard	**p^2H 7.8**	**p^2H 10**	**p^2H 7.8 9M Urea**	**p^2H 10 9M Urea**
Me_3N^+-Ala	52.04	-	52.06	-
Me_2N-Ala	41.52	-	-	-
MeNH-Ala	31.84	-	31.81	-
Me_3N^+-Ala–Ala	52.61	-	52.60	-
Me_2N-Ala–Ala	41.88	-	42.03	-
MeNH-Ala–Ala	32.02	-	32.07	-
Me_3N^+-Arg	52.59	-	52.60	-
NH_2-Arg(N^GMe)	28.17	-	28.35	-
NH_2-Arg($N^{G,G}Me_2$)	38.45	-	38.44	-
NH_2-Arg($N^{G,G'}Me_2$)	-	-	28.29	-
Me_3N^+-Asp	52.86	-	52.88	-
[Me_3N^+-Cys(NH_2)-S-]$_2$	53.39	-	53.42	-
Ac-NH-Cys(SMe)	15.45	15.43	-	-
Ac-NH-Cys(S^+Me_2)	26.67	26.70	-	-
	26.39	26.43	-	-
Me_3N^+-Cys(S^+\{carbaminomethyl\}\{Me\})	52.96	-	52.93	-
Me_3N^+-Gln	52.58	-	52.55	-
Me_3N^+-Glu	52.56	-	52.55	-
Me_3N^+-Gly	54.10	-	54.11	-
Me_3N^+-Gly-NH_2	54.90	54.88	54.91	54.89
Me_3N^+-Gly-Leu	55.03	55.02	55.06	55.06
Me_3N^+-His(Im^+Me_2)	53.00	-	53.01	-
Me_3N^+-His(Im^+Me_2)	36.55	-	36.59	-
	34.16	-	34.23	-
Ac-NH-His(Im^+Me_2)-NH_2	36.33	36.33	36.46	36.46
	33.93	33.92	34.04	34.04
NH_2-His(Im^+Me_2)	36.49	36.26	-	-
	34.11	33.94	-	-
Ac-NH-His(Im-3-Me)-NH_2	34.04	33.87	34.02	33.88
NH_2-His(Im-3-Me)	34.02	33.93	33.95	-

Standard	p^2H 7.5	p^2H 10	p^2H 7.5 9M Urea	p^2H 10 9M Urea
Ac-NH-His(Im-1-Me)-NH_2	32.28	32.88	32.17	31.90
NH_2-His(Im-1-Me)	32.24	31.98	32.19	-
Me_3N^+-Ile	52.29	-	52.33	-
Me_3N^+-Ile-Val-NH_2	53.07	-	-	-
Me_3N^+-Ile-NH_2	53.02	-	53.03	-
Me_2N-Ile-NH_2	42.19	--	42.32	-
MeNH-Ile-NH_2	33.10	-	33.24	-
Me_3N^+-Leu	52.45	-	52.46	--
Ac-NH-Lys(ε-$^+NMe_3$)	53.61	-	-	-
Me_3N^+-Lys(ε-$^+NMe_3$)	53.66	-	53.70	-
Me_3N^+-Lys(ε-$^+NMe_3$)	52.61	-	52.61	-
H_2N-Lys(ε-$^+NMe_3$)	53.61	53.57	-	-
H_2N-Lys(ε-NMe_2)	43.36	43.69	-	-
H_2N-Lys(ε-NHMe)	33.46	33.55	-	-
polyLys(ε-$^+NMe_3$)	53.73	53.67	53.72	-
polyLys(ε-NMe_2)	43.37	44.03	43.35	-
polyLys(ε-NHMe)	33.47	33.77	33.53	-
NH_2-Met(S^+Me_2)	25.41	25.24	-	-
Me_3N^+-Met(S^+Me_2)	53.08	-	53.07	-
Me_3N^+-Met after loss of (-S^+Me_3)	52.56	-	52.53	-
(S^+Me_3)	27.40	-	27.58	-
Me_3N^+-Phe	52.78	-	52.80	-
Me_2N-Phe	42.34	-	42.54	-
MeNH-Phe	32.92	-	32.94	-
Me_3N^+-Phe-Gly-Gly	53.41	53.41	53.41	--
Me_2N-Phe-Gly-Gly	42.18	42.06	42.42	--
MeNH-Phe-Gly-Gly	33.06	33.68	33.32	--
Me_3N^+-Pro	52.89	-	52.93	-
S^+Me_3 (Breitmaier and Voelter, 1987)	27.5	-	-	-
Me_3N^+-Ser	53.67	-	53.67	-
NH_2-Ser(OMe)	59.42	-	59.43	-
Me_3N^+-Thr	54.12	-	54.14	-

Standard	p^2H 7.8	p^2H 10	p^2H 7.8 9M Urea	p^2H 10 9M Urea
NH_2-Thr(OMe)	57.46	-	57.30	-
Me_3^+N-Trishydroxymethylaminomethane	53.72	-	-	-
Me_3^+N-Trishydroxymethylaminomethane	40.02	-	-	-
Me_3^+N-Trishydroxymethylaminomethane	27.35	-	-	-
Me_3N^+-Trp	52.81	-	52.80	-
NH_2-Trp(Ind-1-Me)	32.96	-	32.94	-
Me_3N^+-Tyr	52.78	-	-	-
Me_3N^+-Tyr(OMe)	52.79		52.78	-
Ac-NH-Tyr(OMe)	56.10	-	-	-
Me_3N^+-Tyr(OMe)	56.02	-	56.07	-
NH_2-Tyr(OMe)	56.17	56.13	-	-
Ac-NH-Tyr(OMe)-NH_2	56.15	56.15	56.12	56.13
Me_3N^+-Val	52.45	-	52.46	-
Me_4N^+	56.03	-	56.03	-

Standards were analyzed using a Gemini 200 MHz NMR spectrometer. Standards were adjusted to give a pH meter reading (p^2H) of 7.8-8.0 [700µl D_2O, 30µl CH_3CN, 50µl 1M sodium phosphate (pH 7.5), 5N NaOH] or 10 [700µl D_2O, 30µl CH_3CN, 50µl 1M sodium carbonate (pH 10), 5N NaOH]. If urea was added (25:26 (w/v)), the p^2H was reset (2N HCl). Reproducibility of chemical shift values were standard dependent and varied from 0.01-0.05 ppm. Acetonitrile was included as a reference peak.

range of approximately 2 ppm which allowed for the differentiation of individual amino acids by the value of the characteristic chemical shift. The variation of chemical shifts obtained is to a large degree due to the different environmental contributions of amino acid side-chains. The chemical shift dispersion was sufficient, for example, to carry out peak assignments in the case of mixtures of known trimethylamino acids (data not shown), but the ability to identify unknown amino acid derivatives was less defined. Methylated analytes could potentially be spiked with standards to facilitate assignment.

(i) Amidated versus free trimethylamino acid standards. It was initially hoped that free trimethylamino acid standards would provide a facile means to identify reacted N-termini of protein, however, amidated standards were prepared when it was found that free trimethylamino

acid standards inadequately mimicked the chemical shift of N-terminal derivatives of protein. Even in the case of dipeptidyl standards, the observed chemical shift values varied as different C-terminal amino acids were substituted. In table 2.3.2, trimethylated glycine, trimethylglycyl-leucine, and trimethylglycinamide provide an excellent example of the variability of chemical shift values. There are two possible explanations to rationalize the discrepancy (*vide infra*).

(a) Magnetic anisotropy effects. The amide bond of proteins and the carboxylic acid of free amino acids both possess double bond character but each differs in its NMR properties and orientation with respect to the Nα trimethyl groups. π-Electrons circulate in a magnetic field and induce an opposing magnetic field according to Lenz's law (Wolfson and Pasachoff, 1987). The net effect is to shield or deshield neighbouring nuclei, depending on whether they are oriented in the region where magnetic flux lines subtract or contribute to the external magnetic field (Drago, 1992). In simple olefins, for example, nuclei which are oriented at either end of the principle axis are deshielded and nuclei which are oriented side-on are shielded (Friebolin, 1991). The same induced magnetic field which causes a downfield shift of olefinic protons can presumably affect the chemical shift of an adjacent trimethyl N-terminus to varying degrees, depending on the magnetic anisotropy of the π-electron cloud. Magnetic anisotropy varies on the basis of whether the carboxyl group is free or amide bonded so related standards differing at the carboxylic acid region would be expected to show a difference in their chemical shift values.

(b) Stereoelectronic effects. The stereoelectronic effect (Deslongchamps, 1983) is another factor which may explain the change of chemical shift when free trimethylamino acid derivatives become amidated. Free N-trimethylamino acids differ from amidated analogues in their steric requirements, electrostatic interactions, inductive effects and resonance interactions. While the

net effect on the chemical shift value is certainly a weighted average of the above factors, it may be that the inductive and electrostatic effects are the more important factors. This argument is supported by the fact that Nα amino groups of zwitterionic amino acids typically have pK_a values above 9.0 (Leggett Bailey, 1962) while N-terminal amino groups of peptides have pK_a values of 6.8-8.0 (Creighton, 1993). Clearly, the removal of a negatively charged carboxylate has a pronounced effect on acidity, and by inference, other physical properties. This could, for example, form the basis of an NMR titration study of trimethylamino acid carboxyl groups, where the chemical shift position of zwitterionic trimethylamino acids would change as the pH is adjusted in the range where the carboxylate begins to titrate. If the contributions of magnetic anisotropy (explanation 1) are overlooked, the chemical shift of a trimethylamino acid at low pH would be expected to be closer to the amidated analogue than the zwitterionic N-trimethylated amino acid. In the example of trimethylglycine, the acidic oxygen has a partial negative charge and the carbonyl carbon has a partial positive charge. If one were to transmit the inductive effect through adjacent sigma bonds, one would find a partial negative charge placed on the ^{13}C-methyl carbons. The inductive contribution in trimethylglycyl peptides should differ from trimethylglycine since the carbonyl carbon is not as electropositive. This is because the peptide bonded nitrogen atom of the peptide bond donates electron density to the carbonyl carbon by a direct resonance interaction.

To sum up, amidated trimethylamino acid standards are required for investigations of intact protein. A careful choice of standard is still required since even the neighbouring amino acid, as in the case of $Me_3{}^{+}N$-Gly-NH_2 and $Me_3{}^{+}N$-Gly-Leu, can change the observed chemical shift value. The best scenario, for a protein of known structure, would be to use an N-trimethylamino dipeptidyl standard of the same composition as that found in the protein sequence.

(ii) Amino acid standards with derivatized side-chains. While the chemical shifts of the amidated standards were consistenty downfield (typically by 0.6-0.8 ppm) of their respective free trimethylamino acid standards (Table 2.3.2), the resonance position of the side-chains of amino acid derivatives appeared relatively insensitive to the presence or absence of free α–amino and α–carboxyl groups. Trimethylamino termini are separated from their carboxyl groups by one atom but side-chain residues have a much larger distance separation with respect to their α–amino and α–carboxyl groups. In the case of N^1,N^3-dimethylhistidine and O-methyltyrosine, the distance separating the side-chains was large enough that the observed chemical shift was essentially independent of the α-amino and α-carboxyl moieties. It appears that side-chain amino acid standards can be used without concern over the state of the Cα amino and carboxyl groups.

(iii) Amino acids with inert side-chains. Of all the functional groups which could potentially react with iodomethane in proteins, only the carboxyl groups of standards were apparently not esterified with iodomethane. Either the reaction did not proceed under the conditions employed, or the ester was quickly hydrolyzed before analysis. Esterification reactions using iodomethane typically require polar aprotic solvents and cesium carboxylates as starting material to be successful (March, 1985c). The side-chain residues of serine, threonine, arginine and tryptophan were inert to the action of iodomethane under the conditions employed. Methylated side-chain derivatives of these amino acids were obtained commercially and occupied positions in the NMR spectra which did not overlap with other functional group resonances.

(iv) Effect of aqueous medium on chemical shift values. The chemical shifts of all standards were relatively insensitive to the effects of pH and to the presence of 9M urea, a solubilizing denaturant necessary for many protein analyses. Deviations in chemical shift due to changes in

pH rarely exceeded 1 ppm over the pH range employed and these examples were restricted to methylated functional groups possessing a titratable proton such as Nε–methyl and dimethyllysine. As expected, trimethyllysine did not exhibit any pH dependency since it does not have a titratable proton (Zhang *et al.*, 1994).

The standards served to illustrate that reacted functional groups of amino acid residues in protein could potentially be identified by [^{13}C]NMR spectroscopy. Despite observations that chemical shift positions were relatively insensitive to the presence of urea, to changes of pH value and to a lesser extent, the adjacent amino acid residue, the preliminary nature of this work made it inappropriate to generalize since exceptions to the above were found as data was accumulated. For example, the methyl ether chemical shift of O-methyltyrosine differed significantly (0.15 ppm) from that of N-trimethyl-O-methyltyrosine and suggested that the N-terminal methyl groups affected the magnetic environment. This implied, however, that the O-methyl group might also affect the chemical shift value of the trimethylamino group, but such was not the case. In fact, the chemical shift values of N-trimethyltyrosine and N-trimethyl-O-methyltyrosine differed by only 0.01 ppm. It is not understood why these specific anomalies occurred, but a contribution of the aqueous medium cannot be overlooked. Furthermore, initial unfamiliarities with operating the NMR instrument and the theory of NMR operation are remote possibilities. It was observed later, for example, that improper shimming and phasing of the NMR spectrum and a poor signal to noise ratio could result in changes in chemical shift values of as much as 0.1 ppm, whereas the experimental accuracy of the NMR instrument is, in principle, much smaller for short acquisition times. In our experience, even the variation in chemical shift of carefully obtained measurements ranged as high as 0.05 ppm.

(v) Problems interpreting some of the spectra. There were potential problems in the synthesis of some standards (Table 2.3.2). The standards were made using starting materials without any prior purification, and the reaction temperatures employed were high in some cases. By and large, most of the NMR spectra obtained were of very high quality and easily interpreted. Others, such as standards used for serine, tyrosine, phenylalanine, cystine and methionine amino acids were more difficult to interpret, but achievable. The assignment of a chemical shift value to N-trimethyl-S-methyl-S-carbaminomethylcysteine was very difficult and is questionable. With the more problematic spectra, the appearance of multiple resonance peaks may have been a symptom of the reaction conditions employed (110°C, 4 d) or the use of substandard starting materials. In the case of methionine, methylation led to almost 100% elimination of trimethylsulphonium ion. This was suggested by the resonance peak at 27.40 ppm (Table 2.3.2) which was very close to the literature value of 27.5 ppm (Breitmaier and Voelter, 1987).

(vi) Inductive approach. The structure of each standard was not confirmed by [^{1}H]NMR spectroscopy, except in the case of N^1,N^3-dimethylhistidine and in cases where ^{12}C-standards were commercially available, such as O-methyltyrosine, Nε–methyllysine, and trimethylglycine. Since the focus of the investigation was not the synthesis and characterization of standards, few standards were verified directly. Each standard was characterized inductively, by reacting well characterized proteins with iodomethane and comparing the chemical shifts obtained for protein derivatives and standards. The argument which was adopted was to assume each standard was properly made and its chemical shift accurately recorded, unless shown otherwise. As indicated below, when applied to the characterization of methylated proteins, the protein derivatives and standards corroborated each other very effectively.

2.3.3 NMR analysis of insulin

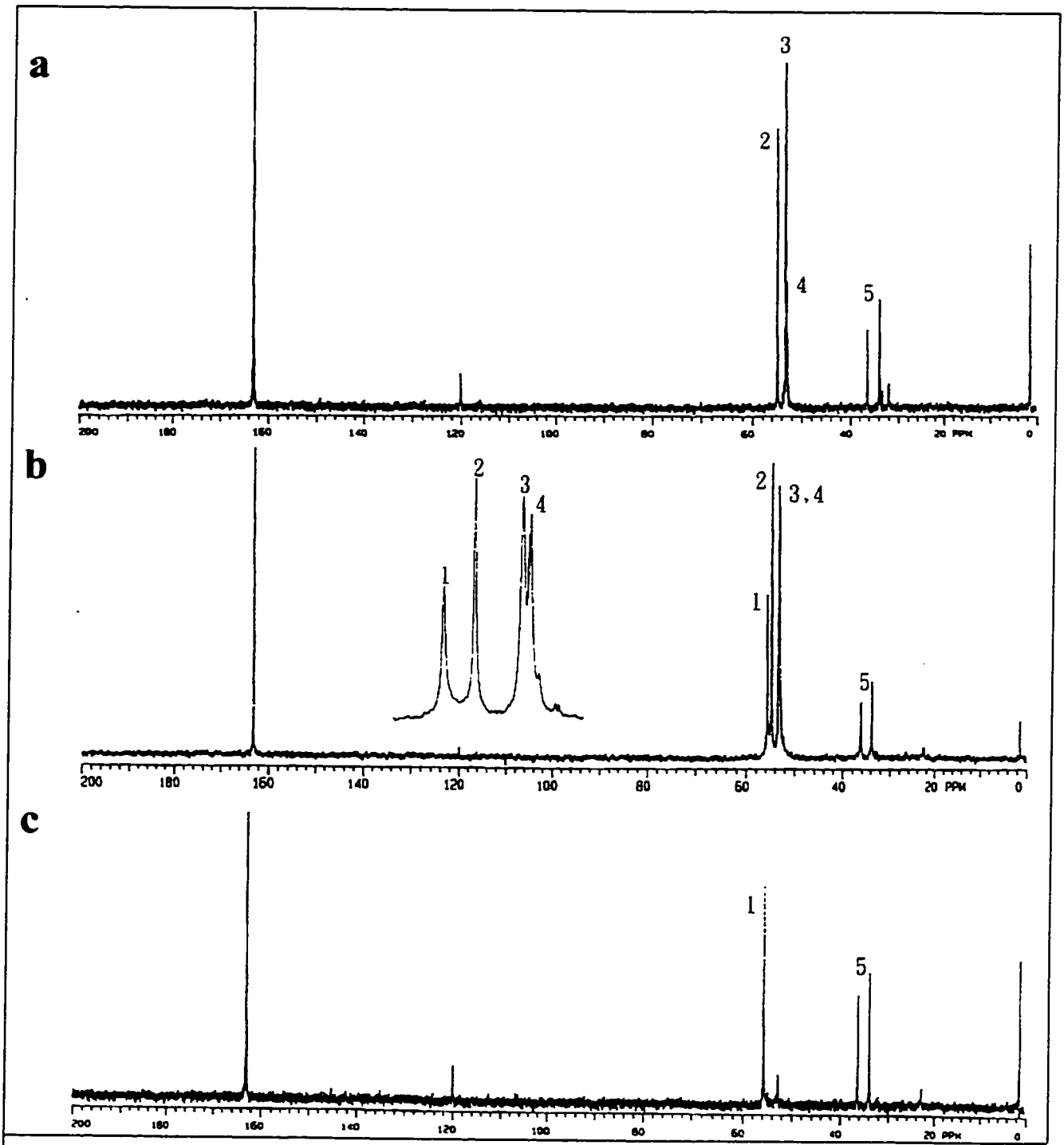


Figure 2.3.1: 50 MHz [^{13}C]NMR spectra obtained for reaction of insulin and citraconylated insulin with [^{13}C]iodomethane. (a)Insulin, pH 7.5; (b) Insulin, pH 10; (c) Citraconylated insulin, pH 10. Peak resonances correspond to the methyl groups in: 1-Tyr(O**Me**); 2-**Me**$_3$N$^+$-Gly; 3-Lys(ε–$^+$N**Me**$_3$); 4-**Me**$_3$N$^+$-Phe; 5-His(Im$^-$**Me**$_2$).

(i) Verifying the merit of standards. Insulin and citraconylated insulin were reacted with iodomethane (Figure 2.3.1). It was previously shown that within experimental accuracy, the chemical shifts of methyl groups were fairly insensitive to changes in the aqueous medium (Table 2.3.2). It was therefore presumed that the standards could serve as adequate probes to identify individual reacted groups in protein. By adopting this approach, tentative peak assignments were made to functional groups of insulin using appropriate standards. In so doing, it was verified that amidated standards better approximated the chemical shifts of amino termini in protein (*vide infra*), while side-chain methylated standards and corresponding Nα-acetylated, Cα-amidated side-chain standards could be used interchangeably to mimick the chemical shift of methylated side-chain residues of protein.

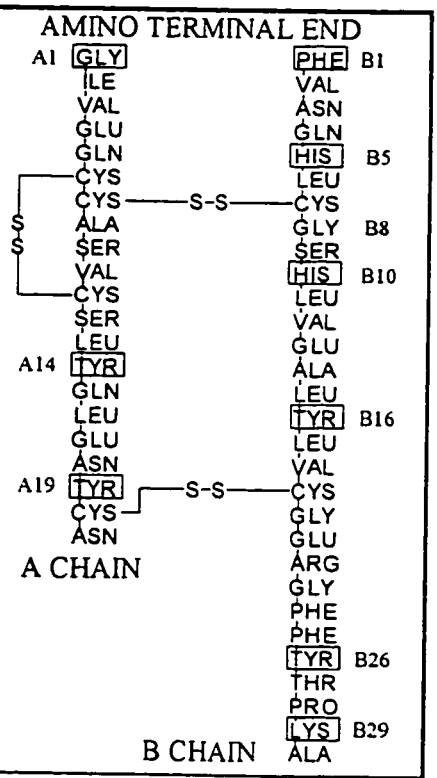


Figure 2.3.2: Primary structure of insulin with reactive groups indicated. Obtained from Legget Bailey (1962b).

(ii) Protein solubility following methylation. It was observed that insulin was less soluble in water following methylation. This loss of solubility was probably due to the loss of hydrogen bonding partners and not due to the loss of structure, as is the case of general denaturation (Means and Feeney, 1971d). For this reason, methylated insulin was necessarily dissolved in 9M urea solution in order to prepare a concentrated sample for NMR analysis and avoid long acquisition times. This was unfortunate in a sense, since the effects of microenvironment on chemical shift values would not be seen in urea solution. As indicated below, however, the influence of microenvironment was not significant.

(iii) Effect of microenvironment. A more theoretical reason for employing denaturant was to test for, and if necessary, account for or avoid any chemical shift perturbations which may originate from structural factors in the protein microenvironment. This concern over the effect of a microenvironment was unfounded since it was observed that the spectra of methylated insulin, when analyzed at p^2H 7.8 in the absence (not shown) and presence (Figure 2.3.1a,b) of urea, showed essentially identical chemical shifts within the accuracy of the NMR instrument.

(iv) Peak assignments. To reiterate, the best standards for assigning peak resonances of reacted amino termini in protein were the amidated amino acid standards, while the choice of standard for assignment of reacted side-chain groups was less consequential. Figure 2.3.1 illustrates the [^{13}C]NMR spectra obtained for the reaction of insulin at pH 7.5 and 10 and the reaction of citraconylated insulin at pH 10 with [^{13}C]iodomethane. The pH values of reaction were chosen arbitrarily and span the pK_a values of several different functional group classes. Accordingly, it was felt that the combination of these pH values would be highly informative. Several resonances appeared in the spectra at about 56 to 53 ppm and 36 to 34 ppm which can be attributed to the reaction of iodomethane with functional groups of insulin. At pH 10 (Figure 2.3.1b), the phenolic hydroxyl of up to four tyrosyl residues (namely A14, A19, B16 and/or B26; figure 2.3.2) reacted to give the most downfield chemical shift of 56.00 ppm, while the dimethylhistidyl moieties (B5 and/or B10) were the most upfield, with resonances at 36.8 and 34.10 ppm (Table 2.3.4). In between these extremes were the trimethyl ammonium derivatives of the B1 phenylalanyl and A1 glycyl N-termini and the B29 lysyl side-chain. It was assumed that at this pH value, most of each available functional group type should have reacted, however, as in the case of the standards, there were no unaccounted resonances observed which could be indicative of a methyl ester.

(v) Functional group reactivities of protein. The amino groups and histidyl side-chains of insulin reacted easily with iodomethane, as both are effective nucleophiles, solvent exposed and accessible to reagent. In fact, with excess iodomethane, methylation of these functional groups was driven essentially to completion. By first blocking the amino groups of insulin with citraconic anhydride, the only functional groups remaining for reaction with iodomethane were the tyrosyl and histidyl side-chains. Citraconylated insulin at pH 10 showed very little reaction of residual free amino groups and the predominant resonances corresponded to the side-chain hydroxyls of tyrosine and imidazoles of histidine (Figure 2.3.1c).

It was mentioned previously that in the case of insulin, protein structure did not significantly perturb the ^{13}C-chemical shifts of methyl groups, however, the effect of protein structure on functional group reactivity towards iodomethane was not addressed. Tyrosyl derivatives were only observed when insulin was reacted at pH 10, where protein is usually unfolded (Means and Feeney, 1971b). The observed inertness of tyrosyl residues towards iodomethane at pH 7.5 was potentially an excellent example of structure–induced protection, however, a better characterization of the reaction in protein is necessary to confirm this claim. Many reagents such as tetranitromethane have been similarly used to examine the distribution of tyrosyl residues in native proteins (Means and Feeney, 1971b). Tyrosyl residues are normally buried in native insulin and are inaccessible to reagent (Means and Feeney, 1971c). However, the structure of insulin exists as a weighted average of many conformations in dynamic equilibrium, and it is plausible that tyrosyl residues become surface exposed in minor conformations. Reaction with iodomethane can occur in these minor conformations at a greatly reduced apparent rate. Despite long reaction times, essentially no reaction was observed at pH 7.5 (Figure 2.3.1a). Even when in 9M

urea, tyrosyl residues were still unreacted (spectrum not shown). While it is possible that some folded structure was still protecting the tyrosyl residues, it is also likely that tyrosyl residues are simply not effective nucleophiles at pH 7.5, even when solvent exposed. This is consistent with the fact that lysyl and tyrosyl residues have the same pK_a values but the former is easily reacted at pH 7.5. The small fraction of lysyl residues which react as the free base at pH 7.5 attest to the effectiveness of nitrogen nucleophiles and the reaction is presumably driven to completion by LeChâtelier's Principle. All else equal, it is presumed that the weaker intrinsic nucleophilicity of phenolate ions places an inordinate limit on the extent of reaction. Ironically, attempts to directly quantify the relative rate of methylation of tyrosyl residues with respect to lysyl residues using reagent accessible standards were never conclusive enough at pH 7.5 to differentiate whether the rate of reaction in insulin was hampered by protection or simply too slow to be of any use. At pH 10 (Figure 2.3.1b), insulin is effectively unfolded and, more importantly, the majority of residues including tyrosine are in the reactive ionization state, leading to substantial reaction.

2.3.4 NMR analysis of insulin peptides

Following performic acid oxidation and elastase digestion of insulin methylated at pH 7.5, protein fragments were separated by high voltage paper electrophoresis as described earlier. Four fragments were observed (Figure 2.3.3), one which migrated as an acidic (A_1) peptide, one as a neutral (N) peptide and two which migrated as basic (B_1,B_2) peptides at pH 6.5. The peptides were collected separately and further purified by high voltage paper electrophoresis at pH 2.1. Figure 2.3.4 illustrates the position and relative intensity of each of the four methylated peptide bands after purification in the second dimension. The radioactive incorporation was such that extensive autoradiography (i.e., >72 h) was not necessary in order to locate individual bands.

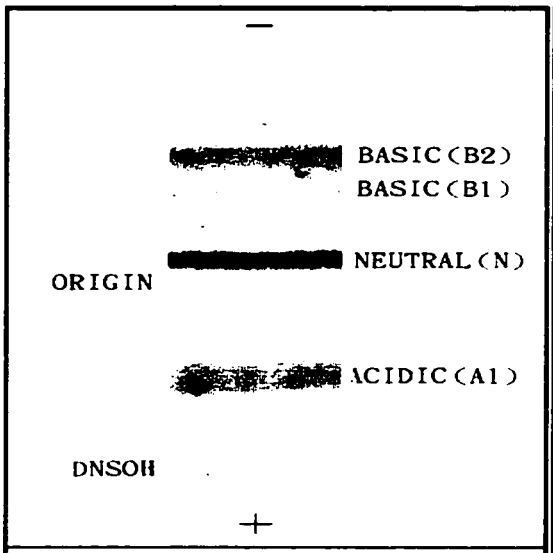


Figure 2.3.3: Illustration of the separation achieved of various methylated peptides in insulin following high voltage paper electrophoresis at pH 6.5

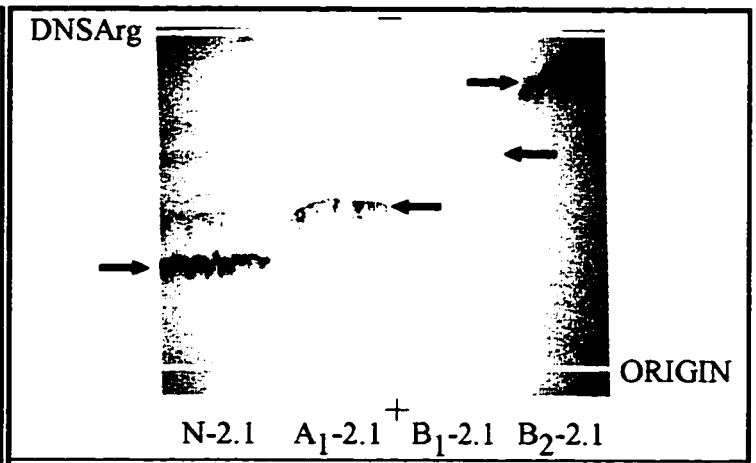


Figure 2.3.4: High voltage paper electrophoresis in the second dimension at pH 2.1. Resolution of 4 major bands (→): N-2.1(1), A_1-2.1(2), B_1-2.1(2) &B_2-2.1(1).

The resolution of several bands present in the neutral band at pH 6.5 is particularly striking. The relative mobility of the four bands of the first dimension, N, A_1, B_1 and B_2 , and the major methylated peptide bands of the second dimension, N-2.1(1), B_1-2.1(2), B_2-2.1(1), and A_1-2.1(2), (indicated by arrows) was calculated (Table 2.3.3) using mobility charts (Appendix 2) that were obtained empirically from model peptides of varying charge/mass ratio. Following two-dimensional electrophoresis, major peptide fragments appeared to be adequately resolved. These were eluted off the paper and examined by NMR spectroscopy. Figure 2.3.5

Table 2.3.3: Relative mobility of the major methylated peptides and their estimated sizes. Estimates were made assuming peptides had a net single or double charge and residues of an average molecular weight of 100g/mole.

High voltage paper electrophoresis at pH 6.5		
Band	**Mobility with respect to dansyl sulphonic acid**	**Estimated # of amino acid residues**
A_1	0.55	5 or 12
N	0	-
B_1	0.30	10 or 20
B_2	0.52	5 or 12
High voltage paper electrophoresis at pH 2.1		
	Mobility with respect to dansyl arginine	
A_1-2.1(2)	0.50	5 or 13
N-2.1(1)	0.32	13 or 37
B_1-2.1(2)	0.63	3 or 7
B_2-2.1(1)	0.90	2, 4 or 9*

*for a net charge of three

presents the recorded spectra of the major bands N-2.1(1), A_1-2.1(2), B_1-2.1(2) and B_2-2.1(1). Upon first inspection, the N-2.1(1) and B_1-2.1(2) bands (Figure 2.3.5a,c) contained both trimethylated amino and dimethylated histidyl derivatives, while the A_1-2.1(2) and B_2-2.1(1) bands (Figure 2.3.5b,d) contained only trimethylated amino groups. Since the primary structure of insulin (Leggett Bailey,1962b), the substrate specificity of elastase (Hartley and Shotton, 1971), the mobility of the peptide fragments and the type of methylated derivatives formed is known, some plausible primary structure(s) of each peptide was proposed (Table 2.3.4).

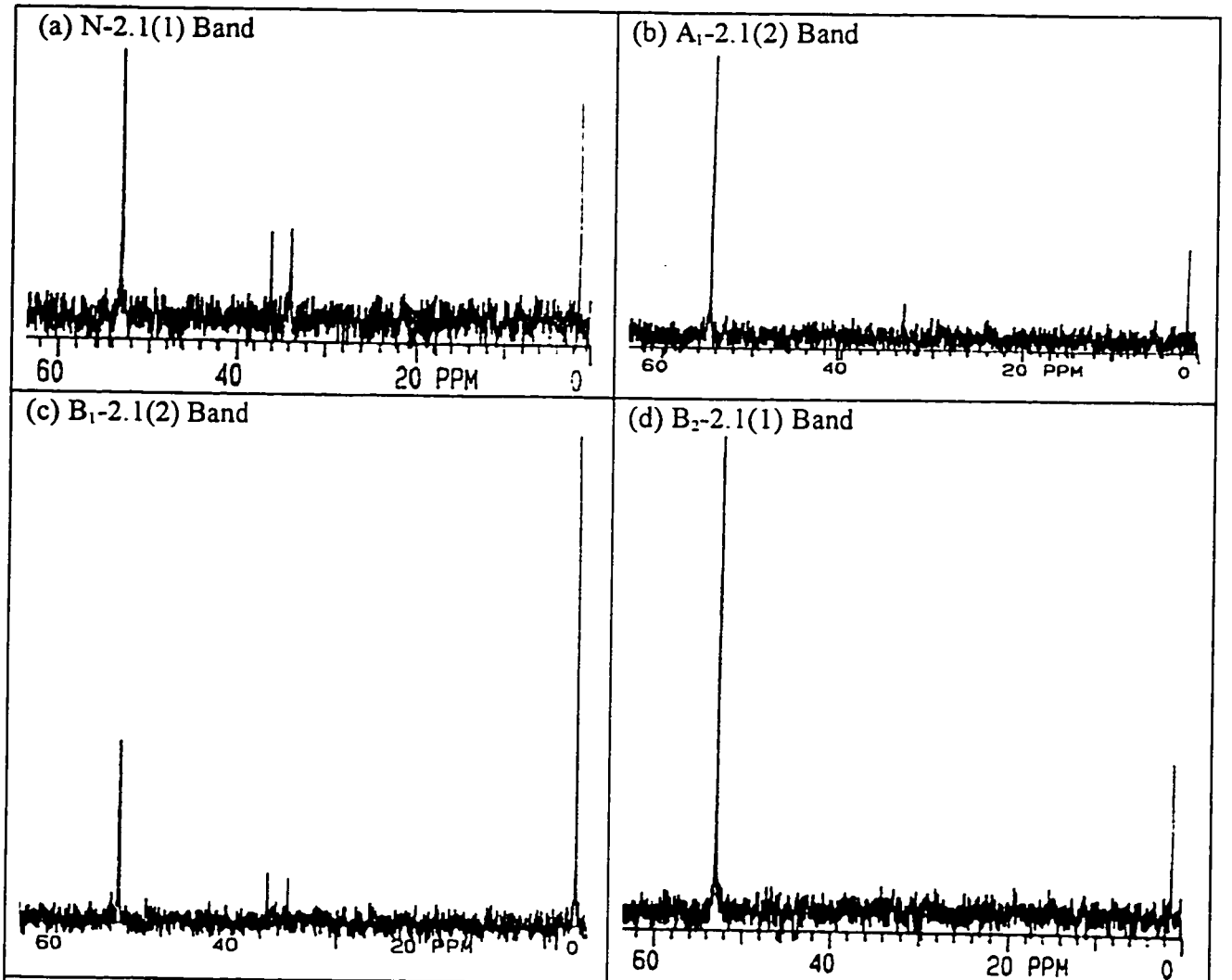


Figure 2.3.5: 50 MHz [^{13}C]NMR spectra of separated peptide derivatives following reaction of insulin with iodomethane (22700 scans acquired for B_1-2.1(2))

The spectrum (Figure 2.3.5a) of the N-2.1(1) fragment was consistent with the presence of dimethylhistidine and a trimethylated ammonium group. Since there was more than one derivatized functional group in the spectrum, the elastase digest must not have separated the two methylated derivatives into individual peptides. Perhaps the activity of that particular preparation of elastase was very low. Notwithstanding, the N-2.1(1) peptide could only have come from the insulin B chain. This was supported by the very close agreement between the observed chemical shift of the amino group (53.46 ppm) with that of the Me_3N^+Phe–Gly-Gly standard (53.41 ppm). More specifically, the fragment appeared to contain the aminoterminal phenylalanyl residue and at least the B5 histidine, but possibly the B10 histidine. The observed chemical shifts in the region for dimethylhistidyl deriva-

Table 2.3.4: ^{13}C-Chemical shifts (in ppm) of methylated peptides and methylated insulin

	D_2O p^{2}H 7.8	**9M urea p^{2}H 7.8**
Methylated Insulin		55.07
Me_3N^+-Gly-Leu	55.03	55.06
peptide A_1 **Me_3N^+**-Gly A1(Insulin)	55.07	
Methylated Insulin		53.40
Me_3N^+-Phe–Gly-Gly	53.41	53.41
peptide N and peptide B_1 **Me_3N^+**-Phe B1(Insulin)	53.46, 53.46	
Methylated Insulin		53.62
Lys(ε-**N^+Me_3**)	53.61	53.66
peptide B_2 B29 Lys ε–**N^+Me_3** (Insulin)	53.46	
Methylated Insulin		56.00
Ac-Tyr(**OMe**)NH_2	56.15	56.12
Methylated Insulin		34.10, 36.58
Ac-His(**Im^+Me_2**)NH_2	33.93, 36.33	34.04, 36.46
peptide N B5, B10 His **Im^+Me_2** (Insulin)	34.15, 36.48	
peptide B_1 B5 His **Im^+Me_2** (Insulin)	34.11, 36.46	

tives (36.48 and 34.15 ppm) were close to the acetylated and amidated dimethylhistidyl standards (36.33 and 33.93 ppm). The estimated size (Table 2.3.3) of this fragment (probably 13 residues, given a net +1 charge at pH 2.1) is contradictory, however, since a thirteen residue peptide would necessarily include the B10 histidine. The contribution of the second histidine should impart a doubly positive net charge to the peptide at pH 2.1, yet this is not consistent with the observed mobility. It has been previously observed, however, that deviations from ideality are common when interpolating the mobility chart for peptides containing sulphonic acid residues. The N-2.1(1) peptide was tentatively assigned the sequence Me_3N^+Phe–Val-Asn-Gln-His(Im^+Me_2)-Leu-Cys(SO_3H)-Gly, which coincided with the expected substrate specificity of elastase and observed net neutral charge at pH 6.5. With due consideration to possible deviations from ideal mobility, and as the integrity of the elastase preparation was questionable, a still larger peptide fragment with a net double charge and containing both histidines was still feasible since this peptide could also be neutrally charged at pH 6.5 due to the contribution of a B13 glutamate.

The signal in the spectrum of the A_1-2.1(2) fragment (Figure 2.3.5b) was consistent with a trimethylated ammonium group. The chemical shift (55.09 ppm) of this fragment was very close to that of the Me_3N^+Gly-Leu standard (55.06 ppm) and suggested that the fragment contained the N-terminal glycyl derivative from the insulin A chain. Furthermore, the peptide derivative was estimated to be composed of 5 amino acid residues (Table 2.3.3) on the basis of mobilities at pH 6.5 and pH 2.1. The proposed sequence for the A_1-2.1(2) fragment was therefore Me_3N^+Gly-Ile–Val-Glu-Gln. No other peptide could be consistent with the observed mobilities at pH 6.5 and 2.1, the peak resonance observed by NMR spectroscopy, and the structure of insulin. Although elastase is not known for cleaving C-terminal to glutaminyl residues, the adjacent A6 cysteic acid

may have confered some atypical behaviour. A more probable explanation, however, was that the elastase preparation was contaminated with α–chymotrypsin, which could readily cleave the bond. While more sophisticated analyses could have been used to verify the sequence of the peptide, its composition, namely Me_3N^+Gly-Ile–Val-Glu-Gln, was essentially established, and served to illustrate the power of a combined NMR and electrophoretic approach.

The NMR spectrum (Figure 2.3.5c) of the B_1-2.1(2) fragment was again indicative of the presence of dimethylhistidine and a trimethylated phenylalanyl N-terminus. The observed chemical shift of the amino group (53.46 ppm) was very close to that of the Me_3N^+Phe–Gly-Gly standard (53.41 ppm). It is noteworthy to point out that both the B_1-2.1(2) and N-2.1(1) trimethylphenylalanyl fragments have the same chemical shift value to two decimal places, and this is despite the fact that each peptide was recovered independently. The chemical shifts observed (36.46 and 34.11 ppm) in the region characteristic for dimethylhistidine were also diagnostic. The observed mobility and estimated number of residues (Table 2.3.3) of the B_1-2.1(2) fragment at pH 6.5 and pH 2.1 were 10 and 7, respectively. Considering these factors, the sequence finally predicted of this fragment was Me_3N^+Phe–Val-Asn-Gln-His(Im$^+Me_2$)-Leu, which would also be consistent with some proteolytic processing by contaminating chymotrypsin.

The resonance found in the spectrum (Figure 2.3.5d) of the B_2-2.1(1) fragment was consistent with the presence of a trimethylated ammonium group with a chemical shift value of 53.46 ppm. Although there are three candidates, the Me_3N^+Gly-Leu standard ruled out the glycyl N-terminus. The phenylalanyl N-terminus or another fragment containing the B29 lysine were the remaining choices. Each standard (53.41 ppm for Me_3N^+Phe–Gly-Gly and 53.61 ppm for NH_2-Lys(Nε$^+Me_3$)) was close to the observed chemical shift (53.46 ppm), but it was the

phenylalanyl standard that was the better match. The most probable structures of the B_2-2.1(1) fragment were proposed to be Me_3N^+Phe–Val (plus Asn or Asn-Gln), Glu-Arg-Gly-Phe–Phe–Tyr-Thr-Pro-Lys(Nε^+Me_3)-Ala or Thr-Pro-Lys(Nε^+Me_3)-Ala. The Me_3N^+Phe–Val hypothesis was ruled out because the peptide should not have migrated as a basic fragment at pH 6.5, which was the observation. Neither of the two remaining possiblities could be ruled out on the basis of size and mobility. On another note, the relatively large chemical shift difference between the lysyl standard and the B29 lysyl residue of the peptide was 0.15 ppm. This difference may be a consequence of the standard chosen or poor operation of the NMR instrument. In hindsight, neither explanation appears convincing. The original spectrum appeared to be adequately shimmed and phase corrected. As far as appropriate standards are concerned, the chemical shifts of many other standards indicated that side-chain resonances were relatively insensitive to the state of the α–amino and α–carboxyl residues, and the lysyl standard was already shown to be an adequate standard for the identification of reacted lysyl side-chains of insulin. It was for this reason, that the synthesis of an ideal standard for observing lysyl side-chain derivatives in peptides (e.g. Nα–acetyl-Lys(N^+Me_3)-NH_2) was never attempted, nor was it deemed necessary. While there may be other factors to consider when choosing a standard, the fact remains that the lysyl standard was a better standard for the identification of reacted lysine side-chains in insulin, than for the peptide fragment containing the B29 lysyl derivative. This observation suggests a very specific interaction may be occurring in the peptide fragment to cause a minor upfield shift of the NMR resonance. The nature of this interaction is unknown. A final recourse to remove all doubt as to the identity of the reacted residue in insulin would have been to acid hydrolyze the peptide bonds and obtain the NMR spectrum of the proposed B29 trimethyllysine derivative.

All other purified bands were too weak to be observed by overnight NMR acquisitions. There was thus far, no evidence to conclusively establish that the B10 histidine was reacted. The B10 histidine is known to coordinate to zinc and should have a decreased reactivity, however, this argument is only relevant to trace labeling studies where only the most reactive groups incorporate trace amounts of radiolabels. In this investigation, insulin was incubated with a very large excess of iodomethane for a prolonged time, so it would appear unusual if the B10 histidine was not reacted. It was hoped therefore, that some of the dimethylhistidyl signal of the N-2.1(2) spectrum came from the B10 histidyl derivative. To answer this question, mass spectral analyses of each peptide derivative was carried out. Unfortunately, the results were inconclusive because the computer algorithm employed for the elucidation of molecular weights could not cope with a 1:1 isotopic mixture of ^{12}C and ^{13}C-methyl groups in the peptide derivatives. The experiment has been rigorously repeated by Nicolas. A. S. Stewart of our laboratory using only [^{13}C]iodomethane. This time, mass analysis confirmed that the B10 histidyl residue was also among those reacted.

2.3.5 NMR analysis of other proteins

As part of the rationale of this investigation, it was important to establish the generality of the methylation approach. To achieve this, other model proteins, namely bovine ribonuclease, serum albumin, and diisopropylphosphoryl-α–chymotrypsin were methylated at pH 7.5 and 10. Their amino termini have been characterized (Leggett Bailey, 1962b; Kaplan, 1972; Peters, 1975). Spectral acquisitions of the methylated protein derivatives are indicated in figures 2.3.6, 2.3.7 and 2.3.8 and peak resonances are shown in table 2.3.5.

The proteins showed the same pattern of methylation as that observed of insulin. The methylation of all functional groups were substantial at pH 10 and substantial except for tyrosyl residues, which did not react at all, at pH 7.5. In addition to insulin, however, there were resonances observed at approximately 25.5 ppm, which was indicative of the methylation of methionine. Methionine is absent in insulin. Unlike some of the methionyl standards that were prepared at high temperature, the elimination of dimethylsulphide to produce trimethylsulphonium iodide was not observed under the reaction conditions used to derivatize protein. However, since the proteins were dialyzed following reaction, the possibility for this cannot be ruled out.

There was in general, a very good agreement of chemical shift resonances between methylated side-chain standards and the corresponding functional groups in each protein, which established the generality of this methylation technique for protein structure-function studies. Ribonuclease A, in particular, was soluble enough following methylation at pH 10 to carry out an NMR acquisition in the absence of urea (Figure 2.3.6; urea is absent from the spectrum). As presented in table 2.3.5, chemical shift values of ribonuclease A before and following addition of urea were essentially identical. This was the same finding as for insulin, with the exception that ribonuclease A is a larger and more specialized protein. It appeared again that the chemical shift was relatively insensitive to the protein microenvironment. While results suggest that the use of NMR spectroscopy for probing the microenvironment of a protein may be restricted to exceptional cases, the finding is also useful since it greatly simplifies the interpretation of spectra.

Both ribonuclease (Figure 2.3.6) and bovine serum albumin (Figure 2.3.7) are composed of a single chain containing a free N-terminus and many lysyl residues. It was therefore very encouraging that two peaks were obtained in the range corresponding to trimethylamino acid

derivatives, one of which originated from the lysyl groups and the other which could be tentatively assigned to the N-terminal trimethylamino acid derivative. Unfortunately, only the free

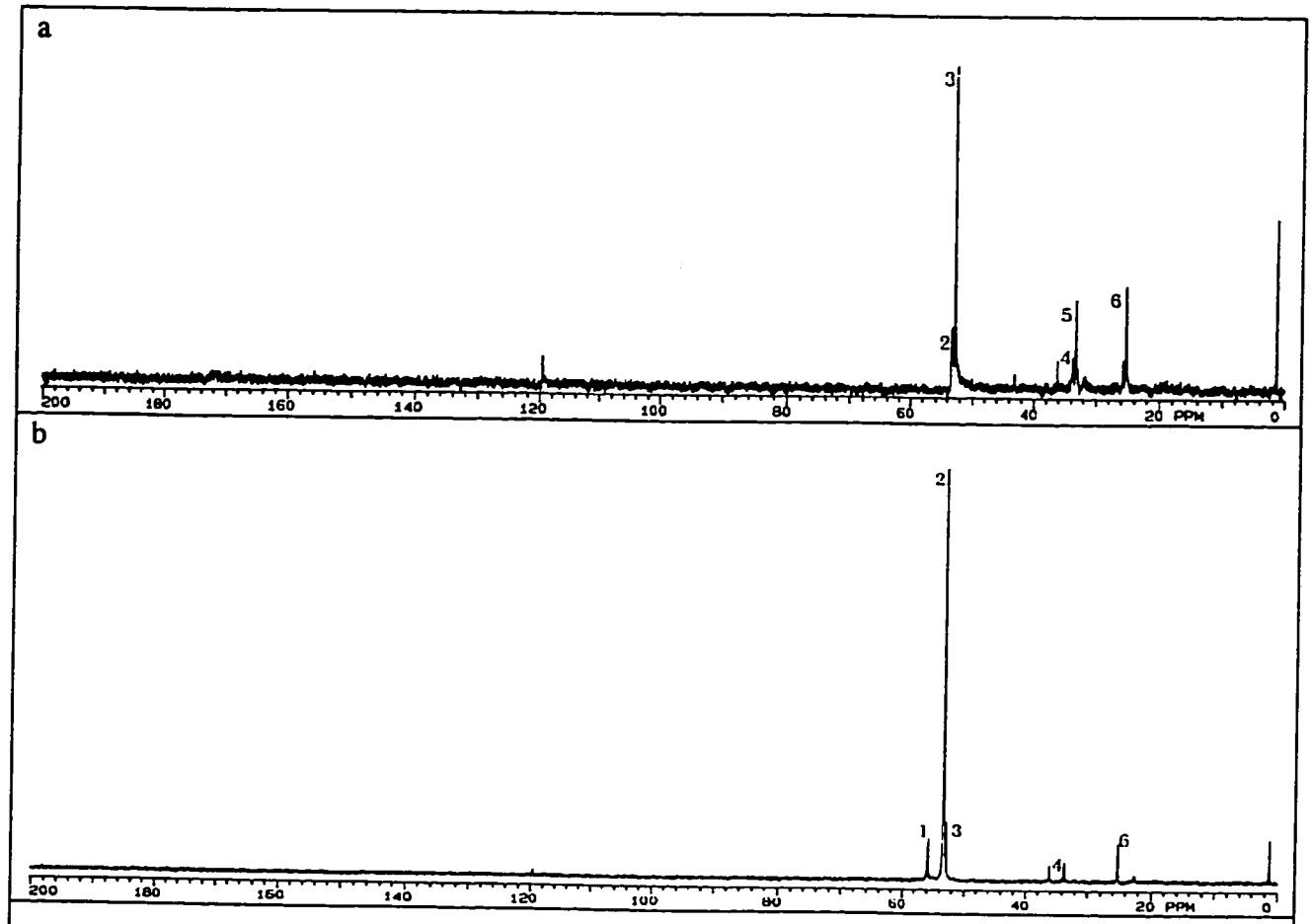


Figure 2.3.6. **50 MHz [^{13}C]NMR spectra obtained for reaction of ribonuclease A with [^{13}C]iodomethane.** (a) pH 7.5; (b) pH 10. Peak resonances correspond to the methyl groups in: 1-Tyr(OMe); 2- Lys(ε–$^{+}$NMe$_3$); 3- Me$_3$N^{+}Lys; 4- His(Im^{+}Me$_2$); 5- Lys(ε– NHMe); 6- Met($^{+}$SMe$_2$).

trimethyl amino acid standards were available, namely Me$_3$N^{+}Lys(ε–$^{+}$NMe$_3$) for ribonuclease and Me$_3$N^{+}Asp for bovine serum albumin, so that definitive peak assignments could not be made. It appears that in the case of ribonuclease at pH 7.5, methylation of lysyl side-chains is minimal.

Diisopropylphosphoryl-α–chymotrypsin is made up of three chains (Blow, 1971) and therefore four resonances are expected in the trimethylamino acid region, the fourth being that of lysine. Once again, this was the observation as shown in figure 2.3.8a.

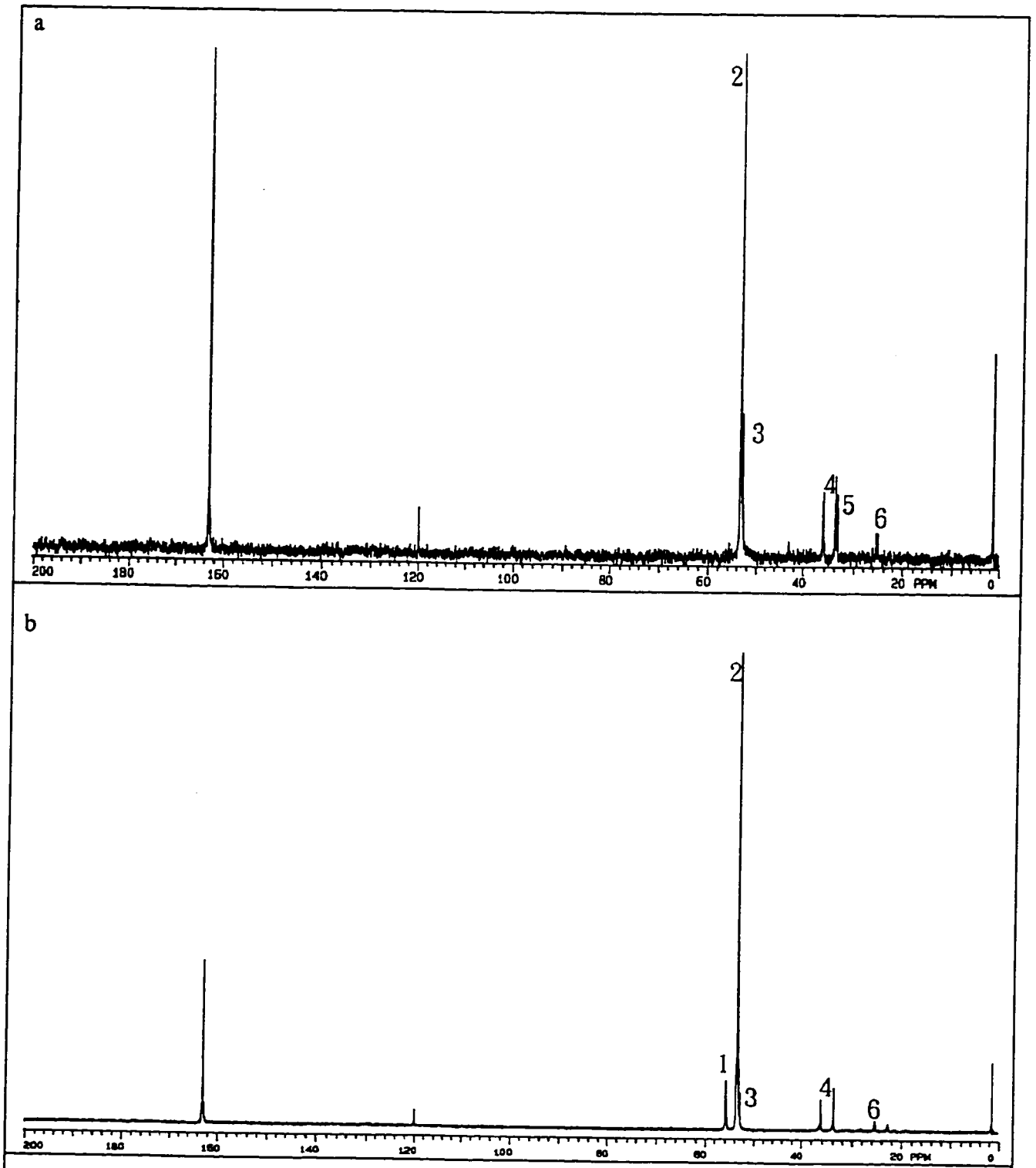


Figure 2.3.7. **50 MHz [^{13}C]NMR spectra obtained for reaction of bovine serum albumin with [^{13}C]iodomethane.** (a) pH 7.5; (b) pH 10. Peak resonances correspond to the methyl groups in: 1-Tyr(OMe); 2- Lys(ε-$^{+}$NMe$_3$); 3- Me$_3$N^{+}Asp; 4- His(Im^{+}Me$_2$); 5- Lys(ε- NHMe); 6- Met($^{+}$SMe$_2$).

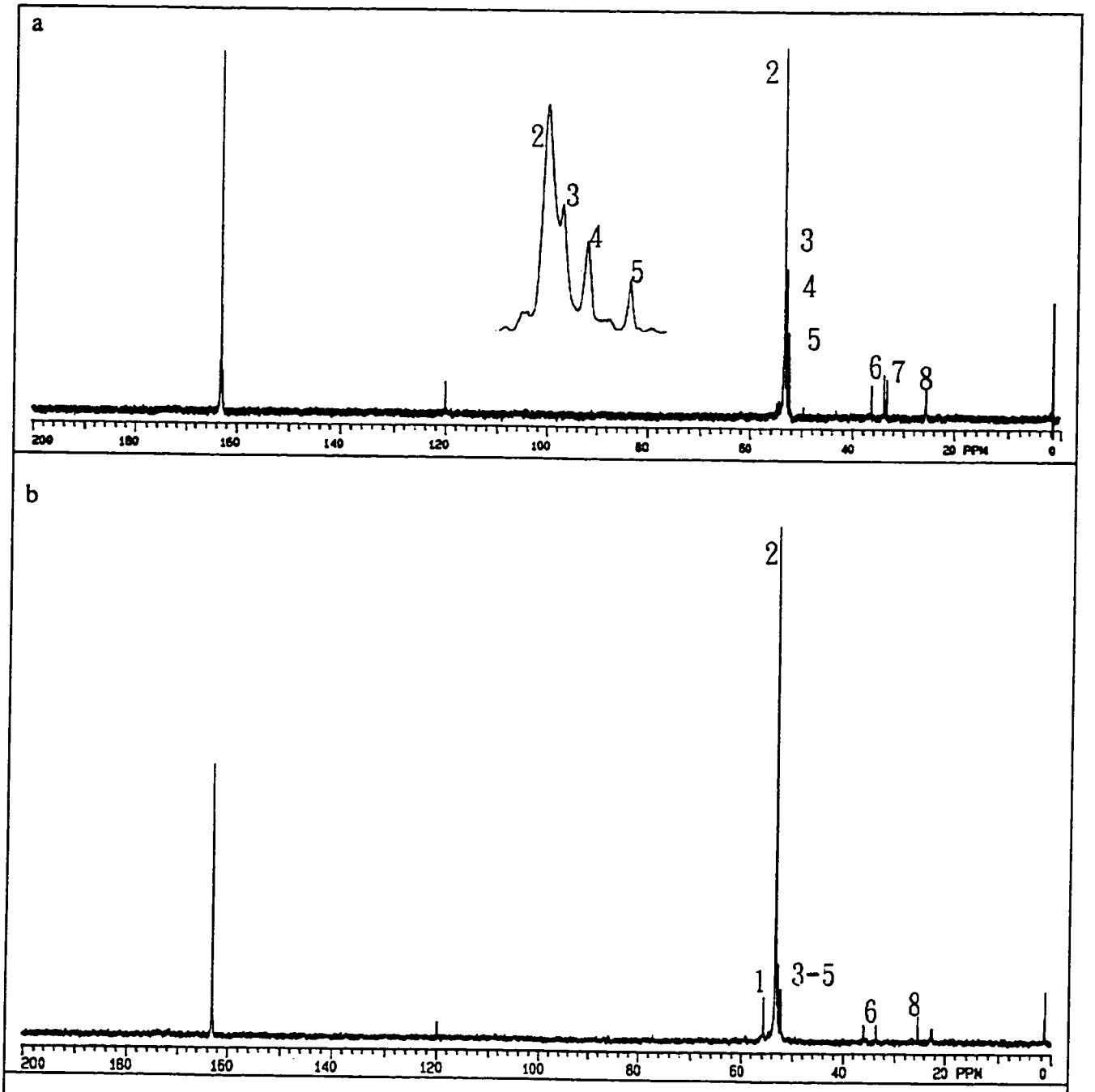


Figure 2.3.8. **50 MHz [^{13}C]NMR spectra obtained for reaction of diisopropylphosphoryl-α–chymotrypsin with [^{13}C]iodomethane.** (a) pH 7.5; (b) pH 10. Peak resonances correspond to methyl groups in: 1-Tyr(**OMe**); 2- Lys(ε–$^{+}$**NMe$_3$**); 3- **Me$_3$N^{+}**Cys; 4- **Me$_3$N^{+}**Ile; 5- **Me$_3$N^{+}**Ala; 6- His(Im^{+}**Me$_2$**); 7- Lys(ε–NH**Me**); 8- Met($^{+}$**SMe$_2$**).

Table 2.3.5: Chemical shifts in ppm for ^{13}C-methyl groups of amino acid standards, methylated ribonuclease A, methylated BSA and methylated diisopropylphosphoryl-α–chymotrypsin.

[^{13}C]Methyl Standards	Chemical Shift (ppm)							
	Standards	Ribonuclease A reacted† at			Bovine serum albumin reacted† at		α–Chymotrypsin reacted† at	
		pH 7.5	pH 10	pH 10	pH 7.5	pH 10	pH 7.5	pH 10
Ac-NH-Tyr(OMe)-NH_2	56.12	NO	56.16‡	56.15	NO	56.13	NO	56.18
H_2N-Lys(ε–$^+NMe_3$)	53.66	53.64‡	53.66‡	53.68	53.86	53.70	53.69	53.68
[Me_3N^+-Cys(NH_2)-S-]$_2$	53.42	-	-	-	-	-	53.50	NO
Me_3N^+Asp	52.88	-	-	-	53.22	53.20	-	-
Me_3N^+Lys(ε–$^+NMe_3$)	52.61	53.18‡	53.20‡	53.20	-	-	-	-
Me_3N^+-Ile–Val-NH_2	53.07‡	-	-	-	-	-	53.21	53.21
Me_3N^+-Ile–NH_2	53.03	-	-	-	-	-	53.21	53.21
Me_3N^+-Ala–Ala	52.60	-	-	-	-	-	52.71	52.72
H_2N-Lys(ε–NMe_2)	43.42	43.49‡	NO	NO	NO	NO	NO	NR
Ac-NH-His(Im^+Me_2)-NH_2	34.04, 36.46	34.06‡, 36.50‡	34.06‡, 36.49‡	34.08, 36.55	34.10, 36.58	34.11, 36.58	34.08, 36.59	34.09, 36.58
H_2N-Lys(ε–$^+NH_2Me$)	33.59	33.53‡	NO	NO	33.62	NO	33.59	NO
H_2N-Met(^+SMeMe)	25.41‡	25.59‡	25.57‡	25.64	25.66	25.62	25.65	25.66

†all proteins presented were reacted without urea
‡measured without urea
NO = not observed; NR = not resolved

2.3.6 Identifying unknown protein N-termini by NMR spectroscopy

Proteins are commonly analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis but the technique only provides information on the global integrity and purity of protein. Following a methylation reaction, however, NMR analysis of reacted protein can be used as an additional criterion of purity. More specifically, observing the number and types of N-terminal derivatives would be another way to investigate the homogeneity of a protein.

While the identity of the N-terminal residues of insulin were readily assigned by NMR spectroscopy, interpreting the spectra of the other model protein N-termini proved more challenging. As explained below, one problem remained. The spectra showed the expected number of N-termini, however, the ability to discriminate between N-termini of different proteins

was put to question. It was observed that the aspartic acid N-terminus of bovine serum albumin, the lysyl N-terminus of ribonuclease A and the isoleucyl N-terminus of diisopropylphosphoryl-α-chymotrypsin were all (presumably) trimethylated to give a resonance observable by NMR spectroscopy. The problem was that these chemical shift values were all 53.20±0.02 ppm, which seemed unusally coincidental (Table 2.3.5). Also, the isoleucyl N-terminal standard of diisopropylphosphoryl-α-chymotrypsin, with a chemical shift of 53.07 ppm, differed from the chemical shift observed in protein by a larger than expected value (53.07 versus 53.21 ppm). After methylating a different commercial preparation of insulin, an unaccountable peak was observed (Figure 2.3.9a), also with a chemical shift value of 53.20 ppm (Table 2.3.5). This peak could not be removed by extensive dialysis and appeared to be part of the protein. Three possibilities were surmised to explain this resonance, which was apparently common to four proteins following reaction. In the first case, it was proposed that an N-terminal impurity was indeed present in insulin and participated in the methylation reaction. Since the same N-terminal impurity is unlikely to be found in each of the three other proteins, it was subsequently proposed that many of the twenty possible trimethylamino acid residues of protein are in fact extremely close in chemical shift. If this is true, the merit of N-terminal identification in the intact protein may be compromised. The third explanation suggested that a common impurity or degradation product, which need not be an amino acid residue, was in fact present in all proteins during reaction. To resolve this matter, the derivatized insulin was hydrolyzed, and spectral analysis of the hydrosylate was carried out in hopes of characterizing the impurity.

Table 2.3.5 shows the measured chemical shifts of derivatized functional groups in methylated insulin before and after hydrolysis. The assignment of resonance positions in each

case, with exception to the unknown peak, was almost exact, and strongly suggested that the hydrolysis route was the method of choice for identifying N-termini. The spectrum (Figure 2.3.9b) of trimethyllysine, trimethylglycine, and trimethylphenylalanine were all 1:1:1 triplets with a coupling constant of approximately 12 Hz. It was observed that with careful shimming of the instrument and adequate scans, such spectra could be obtained for most trimethylamino acid standards. As discussed in chapter 6, the amino nitrogen, with a magnetic quantum number of one (Mann, 1978), normally benefits from efficient quadrupolar relaxation and cannot couple to methyl carbons in a manner which is observed on the NMR time scale. Quadrupolar relaxation, however, depends on the participation of a fluctuating electric field gradient, and in the special case of extreme tetrahedral symmetry, efficient relaxation is not possible (Brevard and Granger, 1981). Such is likely the case with the trimethylamino acid derivatives, as each methyl carbon resonance was accordingly split into three lines (Figure 2.3.9b). The extreme symmetry required to achieve this pattern was also noted. In the case of trimethylamino dipeptidyl standards, the overall symmetry is reduced, and the splitting

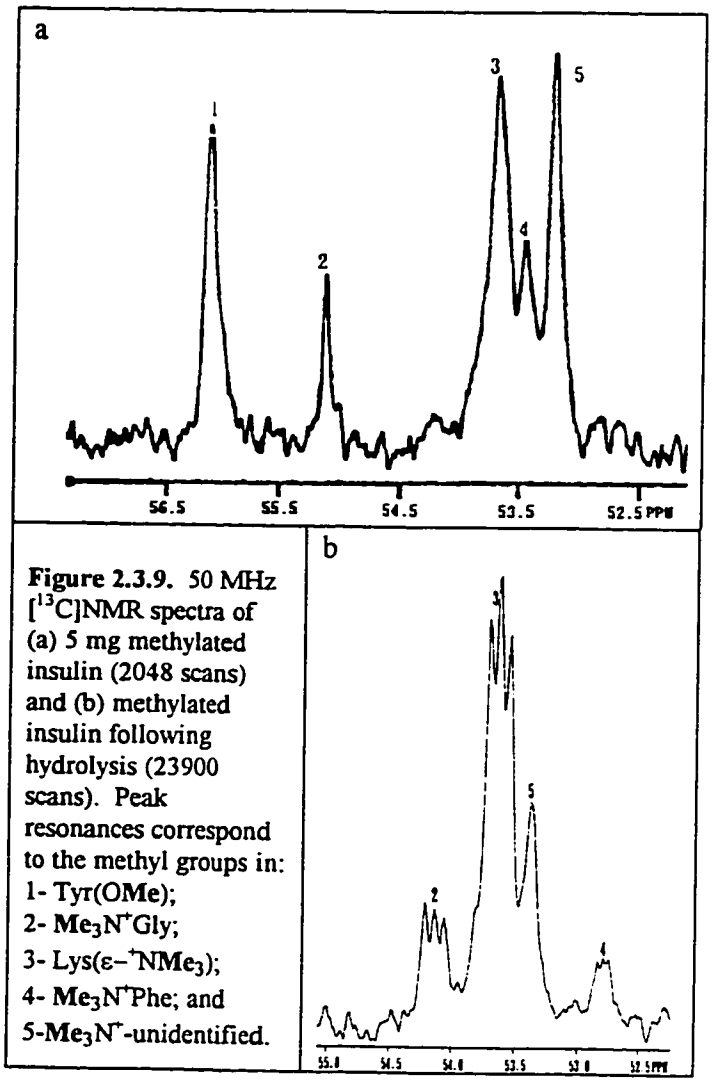


Figure 2.3.9. 50 MHz [^{13}C]NMR spectra of (a) 5 mg methylated insulin (2048 scans) and (b) methylated insulin following hydrolysis (23900 scans). Peak resonances correspond to the methyl groups in: 1- Tyr(OMe); 2- Me_3N^+Gly; 3- Lys(ε-$^+NMe_3$); 4- Me_3N^+Phe; and 5-Me_3N^+-unidentified.

pattern is lost, presumably as a result of more efficient relaxation (not shown). In the case of protein derivatives (Figure 2.3.9a), the splitting is never seen, and is likely due to the long correlation times of protein in solution (Harris, 1983). The unidentified resonance (Figure 2.3.9b) did not appear as a triplet and this suggested that it was not a free trimethylamino acid derivative. In support of this claim, its chemical shift did not match any values of the free trimethylamino acid

Table 2.3.5: Chemical shifts in ppm for ^{13}C-methyl groups of amino acid standards, methylinsulin and hydrolyzed methylinsulin

[^{13}C]Methyl Standards	Chemical Shift (ppm)			
	Standards in urea solution	Methylinsulin in urea solution	Standards in D_2O	Methylinsulin hydrolysate in D_2O
Ac-NH-Tyr(OMe)-NH_2	56.12	56.08	56.15	destroyed
Me_3N^+-Gly-Leu	55.06	55.11		
Me_3N^+-Gly			54.10	54.15
H_2N-Lys(ε-$^+NMe_3$)	53.66	53.66	53.61	53.65
Me_3N^+-Phe-Gly-Gly	53.41	53.43		
Me_3N^+-Phe			52.78	52.81
Me_3N^+-Ile-Val-NH_2		(53.20)	53.07	(53.40)

Proposed unknown N-terminus in parantheses

standards. Furthermore, the change in its chemical shift value following hydrolysis was small, and not in the direction consistent with that observed of N-terminal trimethylamino acid derivatives in protein. This suggested that it may be a dipeptide. The only dipeptide standard which was used in an attempt to identify the resonance peak was Me_3N^+Ile–Val-NH_2. This is so because it is present in the sequence of insulin as the A2 and A3 residues (Figure 2.3.2). Of any error which could be made in the conversion of proinsulin to insulin, it is quite plausible that the A1 glycyl residue is accidentally removed. Also consistent with the choice of standard is the fact that Ile–Val bonds are notorious for their slow acid hydrolysis, and it is quite possible, due to steric bulk, that the trimethylamino analogue is even more resistant and survives acid hydrolysis. Despite this

reasoning, the unidentified peak could not be characterized, as the chemical shift values of the unknown derivative and standard were not comparable. The resolution to this problem is awaiting further developments. One possible approach has been touched upon in chapter 6, where [^{14}N]NMR spectroscopy is used.

2.4 Conclusions

Iodomethane was an exceptionally useful reagent for protein modification. The results show that iodomethane can potentially probe structure-function relationships in proteins and serve as an analytical tool when used with NMR spectroscopy. Of the potentially reactable functional groups of protein, methylation proceeded with amino termini and side-chains of lysyl, histidyl, tyrosyl and methionyl residues. The tyrosyl residues were methylated at high pH values and the carboxyl residues did not appear to react. The reaction could be made specific for some functional groups by control of pH and use of reversible blocking reagents. With excess iodomethane, α–amino groups of N-termini, ε–amino groups of lysyl residues, imidazole groups of histidyl residues, phenolic hydroxyls of tyrosyl residues and sulphides of methionyl residues could be fully methylated. In particular, the trimethylammonium derivative of Nα-amino functional groups, the dimethyl imidazolium derivative of histidyl residues and O-methyl tyrosyl derivatives all represented novel *in vitro* chemical modifications of protein. The potential role of structural factors in perturbing reactivity was suggested by the failure of buried tyrosyl residues to react, while exposed amino, histidyl and methionyl groups did react. The utility of iodomethane was substantial since it was possible to rapidly identify methylated groups of protein by NMR spectroscopy following reaction with [^{13}C]iodomethane. Identification of various classes of methylated functional groups was facilitated by the observation that ^{13}C-resonances were

relatively unaffected by pH, urea and microenvironment. N-terminal amino acid residues of ^{13}C-methylated protein hydrosylates could be identified using NMR spectroscopy and appropriate standards. In the case of insulin, the use of [^{14}C]iodomethane further demonstrated the merit of this reagent by enabling the isolation of peptides containing individual methylated groups.

2.5 Acknowledgment

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and an NSERC scholarship to Alpay Taralp.

2.6 References

Alix, J.-H., Hayes, D.and Nierhaus, K. H. (1979) *J. Mol. Biol.* **127**, 375-395.

Amacher D. E. and Dunn E.M. (1985) *Environmental Mutagenesis* 7, 523-533.

Amacher D. E. and Zelljadt I. (1984) *Mutation Research* **136**, 137-145.

Arai, K., Clark, B. F. C. and Duffy, L. (1980) *Proc. Natl. Acad. Sci. U.S.A.* 77, 1326-1330.

Bietlot, H., Carey, P. R., Choma, C., Kaplan, H., Lessard, T. and Pozsgay, M. (1989) *Biochem. J.* **260**, 87-91.

Blow, D. M. (1971) In *The Enzymes* 3rd ed. **3** (Boyer, P. D. Ed.) pp. 185-212, Academic Press.

Bordas, J., *et al* (1983) *Proc. R. Soc. Lond. B* **219**, 21-39.

Borun, T. W., Pearson, D. W. and Paik, W. K. (1972) *J. Biol. Chem.* **247**, 4288-4298.

Bradford, M. M. (1976) *Anal. Biochem.* **72**, 255-260.

Breitmaier, E. and Voelter, W. (1987) *Carbon-13 NMR Spectroscopy: High Resolution Methods and Applications in Organic Chemistry and Biochemistry*, 3rd ed., p. 234, VCH Publishers.

Brevard, C. and Granger, P. (1981) *Handbook of High Resolution Multinuclear NMR*, pp. 18-19, John Wiley and Sons.

Brown, J. R. and Hartley, B. S. (1966) *Biochem. J.* **101**, 214.

Chang, F. N., Nairckas, I. J., Au, C. and Budzilowicz, C. (1978) *Biochim. Biophys. Acta* **58**, 89-94.

Choma, C. T. and Kaplan, H. (1990) *Biochemistry* **29**,10971-10977.

Clark, R.C. and Dijkstra, J. (1980) *Int. J. Biochem.* **11**, 577.

Cowgill, R.W.(1957) *J. Am. Chem. Soc.* **79**, 2249.

Creighton, T. E. (1993) Proteins: Structures and Molecular Properties, p. 6, W. H. Freeman.

Darbre, A. (1986a) In *Practical Protein Chemistry* (Darbre, A. Ed.) pp. 236-239, John Wiley and Sons.

Darbre, A. (1986b) In *Practical Protein Chemistry* (Darbre, A. Ed.) p. 127, John Wiley and Sons.

Deber, C. M., Moscareilo, M. A. and Wood, D. D. (1978) *Biochemistry* **17**, 898.

DeLange, R. J., Glazer, A. N. and Smith, E. L. (1969) *J. Biol. Chem.* **224**, 1385-1388.

Dell, A. (1986) In *Practical Protein Chemistry* (Darbre, A. Ed.) p. 532, John Willey and Sons.

Deslongchamps, P. (1983) *Stereoelectronic effects in organic chemistry*, 1st ed., Pergamon Press.

Dever, T. E., Costello, C. E., Owens, C. L., Rosenberry, T. L. and Merrick, W. C. (1989) *J. Biol. Chem.* **264**, 20518-20525.

DiMaria, P., Kim, S. and Paik, W. K. (1982) *Biochemistry* **21**, 1036-1044.

Fersht, A. (1985) *Enzyme Structure and Mechanism*, 2nd ed., p. 294, Freeman and Company.

Friebolin, H. (1991) *Basic One and Two Dimensional NMR Spectroscopy*, pp. 41-42, VCH Publishers.

Hames, D. (1990) In *Gel Electrophoresis of Proteins: A Practical Approach*, 2nd ed. (Hames, B. B. and Rickwood, D. Eds.) pp. 1-147, IRL Press.

Hardy, M. F. and Perry, S. V. (1969) *Nature* **223**, 300-302.

Harina, B. M., Dyckes, D. F., Willcott, M. R. and Jones Jr., W. C. (1978) *J. Am. Chem. Soc.* **100**, 4897.

Harina, B. M., Dyckes, D. F., Willcott, M. R. and Jones Jr., W. C. (1980) *J. Am. Chem. Soc.* **102**, 1120.

Harris R. K. (1983) *Nuclear Magnetic Resonance Spectroscopy: A Physicochemical View*, Pitman Publishing Inc.

Hartley, B.S. and Shotton, D.M.(1971) In *The Enzymes* 3rd ed. **3** (Boyer, P. D. Ed.) pp. 323-373, Academic Press.

Hempel, K., Thomas, G., Roos, G., Stocker, W. and Lange, H.-W. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 869-876.

Herriott, R. M. (1947) *Adv. Prot. Chem.* **3**, 169.

Hiatt, N. R., Garcia, R., Merrick, W. C. and Sypherd, P. S. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3433-3437.

Hirs, C.H.W.(1956) *J. Biol. Chem.* **219**, 611.

Jaeck, G. and Benz, F. W. (1979) *Biochem. Biophys. Res. Commun.* **86**, 885.

Jansen, E. F., Nutting, M.-D. F., Jang, R. and Balls, A. K. (1949) *J. Biol. Chem.* **179**, 189.

Jones Jr., W. C., Rothgeb, T. M. and Gurd, F. R. N. (1975) *J. Am. Chem. Soc.* **97**, 3876.

Jones Jr., W. C., Rothgeb, T. M. and Gurd, F. R. N. (1976) *J. Biol. Chem.* **251**, 7452.

Kaplan, H. (1972) *J. Mol. Biol.* **72**, 153-162.

Kaplan, H., Hefford, M., Chan, A., and Oda, G. (1984) *Biochem. J.* **217**, 135.

Klee, C. B. and Vanaman, T. C. (1982) *Adv. Prot. Chem.* **35**, 213-321.

Kuehl, W.M and Adelstein, R. S. (1969) *Biochem. Biophys. Res. Commun.* **37**, 59-65.

Kutob and Plaa (1962) *Toxicol. Appl. Pharmacol.* **4**, 354.

Knapp, D. R. (1990) In *Meth. Enzymol.* **193** (Mc Closkey, J.A. Ed.) pp.314-329, Academic Press.

LaBadie, J., Dunn, W. A. and Aronson, N. N., Jr. (1976) *Biochem J.* **160**, 85.

Leggett Bailey, J. (1962a) *Techniques in Protein Chemistry*, pp. 61-62, Elsevier Publishing Company.

Leggett Bailey, J. (1962b) *Techniques in Protein Chemistry*, p. 97, Elsevier Publishing Company.

Link, T.P. and Stark, G.R.(1968) *J. Biol. Chem.* **243**, 1082-1085.

Liu, T.-Y. and Elliott, S. D. (1971) In *The Enzymes* 3rd ed. **3** (Boyer, P. D. Ed.) p. 628, Academic Press.

Lundblad, R. L. (1995) *Techniques of Protein Modification*, Chapter 14, CRC Press.

Lunn G. and Sansone E. B. (1991) *American Industrial Hygiene Association Journal* **52**, 252-257.

Mann, B. E. (1978) In *NMR and the Periodic Table* (Harris, R. K. and Mann, B. E. Eds.) p. 96, Academic Press.

March, J. (1985a) *Advanced Organic Chemistry*, 3rd ed., p. 306, John Willey and Sons.

March, J. (1985b) *Advanced Organic Chemistry*, 3rd ed., p. 364, John Willey and Sons.

March, J. (1985c) *Advanced Organic Chemistry*, 3rd ed., Chapter 10, John Willey and Sons.

Marvi, J. and Vogel, H. J. (1994) *Proteins: Structure, Function and Genetics* **18**, 381-389.

Means, G. E. and Feeney, R. E. (1968) *Biochemistry* 7, 2192.

Means, G. E. and Feeney, R. E. (1971a) *Chemical Modification of Proteins*, pp. 130-134, Holden-Day.

Means, G. E. and Feeney, R. E. (1971b) *Chemical Modification of Proteins*, pp. 17-20, Holden-Day.

Means, G. E. and Feeney, R. E. (1971c) *Chemical Modification of Proteins*, pp. 8-9, Holden-Day.

Means, G. E. and Feeney, R. E. (1971d) *Chemical Modification of Proteins*, pp. 9-11, Holden-Day.

Olah, G. A. (1973) *Aldrichimica Acta* **6**, pp. 7-17.

Oppenheimer, H. L., Labouesse, B. and Hess, G. P. (1966) *J. Biol. Chem.* **241**, 2720.

Paik, W. K., Cho, Y. B., Forst, B. and Kim, S. (1989) *Biochem. Cell. Biol.* **67**, 602-611.

Paik, W.K. and Kim, S. (1980a) *Protein Methylation*: In *Biochemistry: A Series of Monographs* **1**, (Meister, A. Ed.) pp. 44-46, John Wiley and Sons.

Paik, W.K. and Kim, S. (1980b) *Protein Methylation*: In *Biochemistry: A Series of Monographs* **1**, (Meister, A. Ed.) pp. 240, 243-253, John Wiley and Sons.

Paik, W.K. and Kim, S. (1980c) *Protein Methylation*: In *Biochemistry: A Series of Monographs* **1**, (Meister, A. Ed.) pp.149-158, John Wiley and Sons.

Paik, W.K. and Kim, S. (1980d) *Protein Methylation*: In *Biochemistry: A Series of Monographs* **1**, (Meister, A. Ed.) pp. 12-17, John Wiley and Sons.

Parish, C. R. and Ada, G. L. (1969) *Biochem. J.* **113**, 489-499.

Peters, T., Jr. (1975) In *The Plasma Proteins: Structure, Function and Genetic Control*, 2nd ed., **1** (Putnam, F. W. Ed.) p. 142, Academic Press.

Richards, F. M. and Wyckoff, H. W. (1971) In *The Enzymes* 3rd ed. **3** (Boyer, P. D. Ed.) p. 682, Academic Press.

Rowe, P. M., Wright, L. S. and Siegel, F. L. (1986) *J. Biol. Chem.* **261**, 7060-7069.

Schaeffer, W. H., Lukas, T. J., Blair, I. A., Schultz, J. E. and Watterson, D. M. (1987) *J. Biol. Chem.* **262**, 1025-1029.

Serrano, L., Matouschek, A. and Fersht, A. R., (1992) *J. Mol. Biol.* **224**, 805-818.

Smith, K. M., Kehres, L.A. and Tabba, H. D. (1980) *J. Am. Chem. Soc.* **102**, 7151-7152.

Taralp, A. and Kaplan, H. (1997) *J. Protein. Chem.* **16**, 183-193.

Van Hemert, F. J., Amons, R. A., Pluymus, W. J. M., van Ormondt, H. and Moller, W. (1984) *Artemia. Embo. J.* **3**, 1109-1113.

Weihing, R. R. and Korn, E. D. (1970) *Nature* **227**,1263-1264.

Wilcox, P. E. (1967) In *Meth. Enzymol.* (Hirs, C. H. W. Ed.) **11**, p. 605.

Wilkinson (1986) In *Practical Protein Chemistry* (Darbre, A. Ed.) p. 144, John Wiley and Sons.

Wissmann, A. and Hillen, W. (1991) In *Meth. Enzymol.* (Sauer, R. T. Ed.) **208,** Chapter 18, Academic Press.

Wolfson, R. and Pasachoff, J. M. (1987) *Physics*, p. 767, Little, Brown and Company.

Zhang, M., Huque, E, and Vogel, H. J. (1994) *J. Biol. Chem.* **269**, 5099-5105.

Where is the water?

Chapter 3: Chemical modification of lyophilized proteins in nonaqueous environments

3.1 Introduction

Chemical modification of native proteins was one of the first methods employed to investigate structure-function relationships. Comprehensive descriptions of the techniques, reagents and strategies for the modification of native proteins are available in general reviews of the field (Means and Feeney, 1971; Glazer *et al.*, 1976; Lunblad and Noyes, 1984; Imoto and Yamada, 1989; Lundblad, 1995). The established solvent to carry out chemcal modifications has been water since proteins are generally presumed only soluble in aqueous media. In keeping with this fact, the reactions described specifically in chapter 2 were carried out under aqueous conditions. From a synthetic point of view, however, water has not been a choice solvent to carry out chemical transformations. Its seemingly obligatory use has restricted the choice and effectiveness of chemical modifying reagents of proteins because many of the chemicals used are insoluble in water, react rapidly with water and in the case of activating reagents, form water unstable derivatives with side-chain functional groups of protein. It would be advantageous, therefore, if chemical modifications of protein were carried out in the absence of water.

It is now well established that the catalytic properties of lyophilized enzymes remain intact in organic solvents (Klibanov, 1989; Chen and Sih, 1989; Wescott and Klibanov, 1994). This appears to be due to an associated layer of water which is not easily removed in the nonaqueous environment and is responsible for maintaining enzyme structure. Indeed, it has been observed that crystallized proteins have essentially the same structure when placed in organic solvent as in water (Fitzpatrick *et al.*, 1993; Fitzpatrick *et al.*, 1994). The findings imply that proteins, in general, should retain their native structure when lyophilized and dispersed in organic solvents,

however, physico-chemical studies to establish this have not been as clear as in the case of crystallized proteins. Some investigations indicate that the structure in the lyophilized state is the same as in solution (Careri *et al.*, 1980; Schinkel *et al.*, 1985; Rupley and Careri, 1991) while others indicate some limited but reversible conformational change (Poole and Finney, 1983; Prestrelski *et al.*, 1993; Desai *et al.*, 1994). There are likely to be differences among proteins in this regard but the results of extensive studies (Klibanov, 1989; Chen and Sih, 1989; Desai *et al.*, 1994; Wescott and Klibanov, 1994; Costantino *et al.*, 1995; Mishara *et al.*, 1996; Griebenov and Klibanov, 1997) provide strong evidence that most proteins in the lyophilized state at least retain the essential elements of their native structure in solution. If this is true, the reactivity of functional groups in lyophilized proteins should reflect the reactivity of functional groups of the protein in solution and provide information on the solution structure and its properties. The objective, therefore, of the present research was to demonstrate the feasibility and utility of nonaqueous chemical modification of proteins.

3.2 Materials and Methods

Proteins. Bovine insulin, α-chymotrypsin, ribonuclease and serum albumin were purchased from Sigma Chemical Company. Acetylated bovine serum albumin was prepared using acetic anhydride under aqueous conditions (Means and Feeney, 1971). Inactive diisopropylphosphoryl-α-chymotrypsin was prepared by incubation with diisopropylfluorophosphate (Darbre, 1986a).

Chemicals and Solvents. H_2N-Met($^+$SMe)$_2$ was purchased from Sigma Chemical Company. [^{13}C]Iodomethane was purchased from Sigma Chemical Company. [*Acetic*-1^{14}C]anhydride (250 μCi [tot. act.], 9.20 mCi/mmol [spec. act.]) was obtained from NEN Research Products and

[^{3}H]acetic anhydride (25 mCi, 6.94 Ci/mmol) and [^{14}C]iodomethane (1 mCi, 56mCi/mmole) were obtained from Amersham Canada Ltd. All other chemicals, reagents and solvents were high purity preparations obtained from commercial sources.

^{13}C-methylated amino acids were prepared as follows:

Ac-NH-Tyr(OMe)-NH_2: N-Acetyl-L-tyrosine-amide (100 mg) was dissolved in 0.5 M sodium metaborate buffer (10 ml), pH 11. A 1:1(v/v) solution (300 µl) of [^{13}C]iodomethane in acetonitrile was added, the reaction vessel sealed, and the biphasic mixture shaken in a water bath at 37°C for 24 h. The solution was cooled in an ice bath and the methylated acetyltyrosine amide was extracted with chloroform (3 X 10 ml). The extracts were dried under a stream of nitrogen.

H_2N-Lys(ε-$^{+}NMe_3$): Polylysine·HBr (30 mg) was dissolved in 5mM sodium deuteroxide (3 ml) and a 1:1(v/v) solution (20 µl) of [^{13}C]iodomethane in acetonitrile was added. The pH was adjusted to 10.5 without buffering and the sealed vial was shaken at 37°C for 4.5 h. The polylysine derivative was hydrolyzed (6N HCl, 18h) to liberate the Nε-methylated free amino acid.

Ac-NH-His(Im^{+}Me$_2$)-NH_2: Histidine amide (40 mg) was dissolved in dH_2O (5ml) and acetylated with acetic anhydride (5 X 20 µl) while maintaining the pH at 9 by the addition of 5N NaOH. Sodium metaborate was added to a final concentration of 200 mM and the pH was adjusted to 11 by the addition of 1M NaOH. The solution was transferred to a scintillation vial and a 1:1 (v/v) solution (200 µl) of [^{13}C]iodomethane in acetonitrile was added. The capped biphasic mixture was shaken for 72 h at 37°C and the aqueous layer was separated. Acetate and other anions were removed prior to analysis by ion exchange using a Dowex-1 anion exchanger.

Me_3N^+-Gly-Leu, Me_3N^+-Ile-NH_2, Me_3N^+-Ala-Ala and Me_3N^+-Phe-Gly-Gly: The un-methylated peptides (13 mg) were placed in a glass vial and dissolved in 5 mM sodium deuteroxide (1 ml). A 1:1 (v/v) solution (20 µl) of [^{13}C]iodomethane in acetonitrile was added and the sealed reaction was shaken at 37°C for 9 h in a water bath.

Me_3N^+-Cys(NH_2)-S-S-Cys(NH_2)N^+Me_3: Cystine dimethylester dihydrochloride (10mg) was dissolved in dH_2O (4 ml) and the pH was adjusted and maintained at 9.5 by the addition of concentrated ammonia. After 45 minutes the solvent was evaporated under reduced pressure. The sample was suspended in D_2O (1 ml) and the supernatant containing the cystine diamide was reacted at 37°C for 8 h with [^{13}C]iodomethane (20 µl). The p^{2}H was maintained at 10 by the addition of 1 M sodium carbonate (100 µl).

N-1-[5-dimethylamino-1-naphthalenesulfonyl]-2-hydroxy-1,3-diamino-propane: Dansyl chloride (500 mg) and 1,3-diamino-2-hydroxypropane (500 mg) were dissolved in pyridine (8 ml) and 1:1 (v/v) pyridine/N,N-dimethylformamide (10 ml), respectively. The diamine solution was stirred vigorously at room temperature and the solution of dansyl chloride was added slowly below the surface with a syringe. After 15 min, the reaction was diluted 1:1 (v/v) with dH_2O and the crude mixture stored at -20°C until required (yield ~90%). Aliquots of solution were subjected to high voltage paper electrophoresis in pH 6.5 buffer (Kaplan, 1972) for 25 minutes at a voltage gradient of 40 V/cm. The fluorescent amine ($\mu_{dansylsulfonic\ acid}$ = -0.8) was eluted from the paper with dH_2O, adsorbed to a Dowex 50X8-200 cation exchanger, and converted to the free amine by elution with 1 M ammonia. Fractions containing amine were collected and lyophilized.

Aqueous methylation of proteins at pH 7.5 and pH 10. Insulin (20 mg), α-chymotrypsin (20 mg), diisopropylphosphoryl-α-chymotrypsin (20 mg), and ribonuclease (20 mg) were placed in screw-capped vials and dissolved in 200 mM sodium phosphate buffer (10 ml) pH 7.5, or 200 mM sodium metaborate buffer (10 ml) pH 10. The buffer used was sufficient to titrate the protein without any measurable drop in the pH value. A 1:1 (v/v) solution (250 μl) of [^{13}C]iodomethane in acetonitrile was added. The vial was sealed and the biphasic mixture shaken at 37°C for 24 h.

Methylation of proteins at LpH 7.5 and LpH 10 in octane. Proteins were lyophilized directly in reaction vessels (typically threaded pyrex hydrolysis tubes). Only in the case of insulin at LpH 7.5 (i.e. *lyophilized* at pH 7.5) was it necessary to lyophilize protein from a large volume and transfer it to a reaction vessel. Insulin (20 mg) was lyophilized from a frozen solution of 1 mM sodium phosphate buffer (40 ml), pH 7.5 or 40 mM sodium metaborate buffer (1 ml), pH 10. Ribonuclease and α-chymotrypsin were lyophilized from a solution of 40 mM sodium phosphate buffer (1 ml), pH 7.5 or 40 mM sodium metaborate buffer (1 ml), pH 10. Prior to freezing, 10 μl (or 30 μl) 1N NaOH was added to readjust buffer pH to 7.5 (or pH 10). Anhydrous octane (2 ml) was added to the protein and the medium was sonicated until protein was finely dispersed. [^{13}C]Iodomethane (100 μl) was added. Vessels (and contents) were frozen, placed under vacuum and flame sealed to prevent loss of iodomethane. The protein dispersion was stirred 12 h at 75°C for the LpH 10 reactions and 24 h for the LpH 7.5 reactions. The tubes were opened, derivatized protein was centrifuged with two washes of octane and residual octane was removed *in vacuo*.

Preliminary octane and *in vacuo* methylation reactions of protein were monitored by 50 MHz [^{13}C]NMR spectroscopy. It was established that 12 h of reaction at LpH 10 and 24 h of reaction at LpH 7.5 was optimum to obtain maximum derivatization (Figures 3.2.1 and 3.2.2) while avoid-

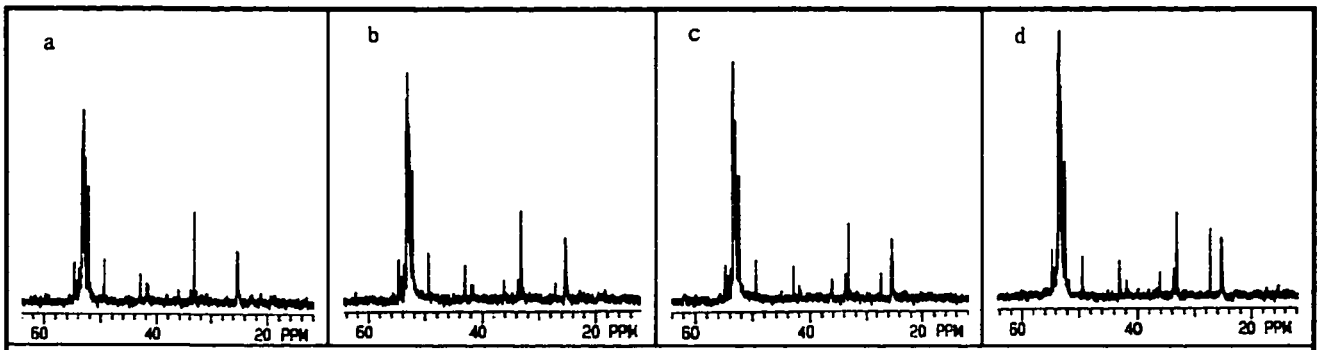


Figure 3.2.1: Typical reaction profile obtained for methylation of α-chymotrypsin, LpH 7.5, in octane (and *in vacuo*). The profiles are plotted to scale. Methylation commences rapidly but decreases notably thereafter. Reaction times are: a) 2 h; b) 6 h; c) 12 h; d) 24 h. Maximum reaction is essentially reached after 24 h.

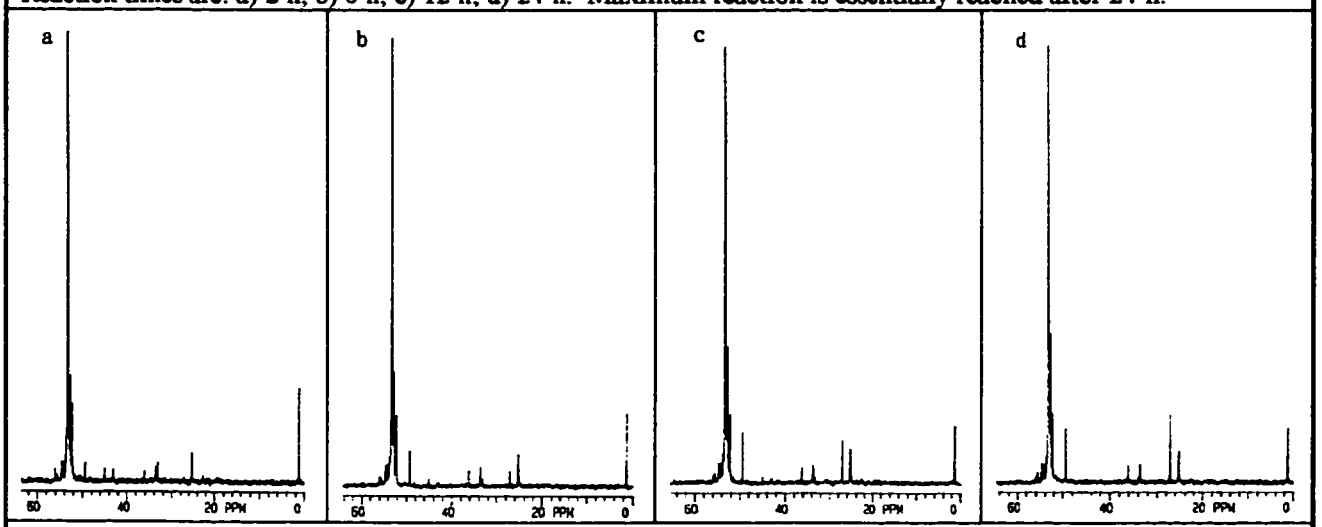


Figure 3.2.2: Typical reaction profile obtained for methylation of α-chymotrypsin, LpH 10.0, in octane (and *in vacuo*). Peak heights can be related to acetonitrile at 1.7 ppm. Methylation commences rapidly but decreases thereafter. Reaction times are: a) 2 h; b) 6 h; c) 12 h; d) 24 h. Maximum reaction is essentially reached after 12 h.

ing excessive crosslinking. Similarly, 75°C was found to be the optimum temperature to promote reaction and avoid crosslinking. All pH values were measured at room temperature. Possible changes of ionization states (at temperatures other than room temperature, e.g. 0°C) due to temperature dependent variations of pK_a values were overlooked in this preliminary investigation.

***In vacuo* methylation of proteins at LpH 7.5 and LpH 10.** The proteins were lyophilized using the same buffer system as described above for the octane reaction. A two-compartment reaction vessel was employed and protein was lyophilized directly in the protein compartment (Figure 3.2.3). The neck of the vessel was narrowed by flame and [^{13}C]iodomethane (10 μl) was transferred into the reagent chamber using a plastic tube attached to the end of a micropipette.

The top of the reaction vessel was closed with a screw cap and the reagent chamber was submersed in liquid nitrogen. The screw cap was removed, the end of the tube was fitted with a vacuum hose and the vessel was sealed *in vacuo*. The reaction vessel was incubated in an oven at 75°C for various times. At the end of the reaction, the reagent chamber was placed in liquid nitrogen, the vacuum seal was broken and the modified protein was removed. Alternatively, a single compartment reaction vessel, namely a pyrex hydrolysis tube fitted with a screw cap, could be employed (Figure 3.2.3). Protein was lyophilized and the tube was narrowed by flame at mid-height. Reagent was then added to the narrow portion of the vessel using a micropipette and the reagent was pulled down and frozen by immersing the bottom half of the vessel in liquid nitrogen. The condensation of trace moisture and oxygen inside the vessel was minimized by keeping the vessel capped whenever possible. The cap was eventually removed and the vessel sealed *in vacuo* as described above. The vessel could be incubated in either an oven or block heater. To terminate the reaction, the glass was scored and a septum with a hole bored through the rubber was fitted onto the top of the vessel. The septum served as a reservoir which could be filled with liquid nitrogen. After all iodomethane was frozen against the top portion of the glass, the vessel was cracked open and protein recovered. To open the vessel, a hot glass rod is pressed on the score line, ***away from frozen reagent***; ***Expanding*** *reagent vapour may* ***burst*** *vessel!*

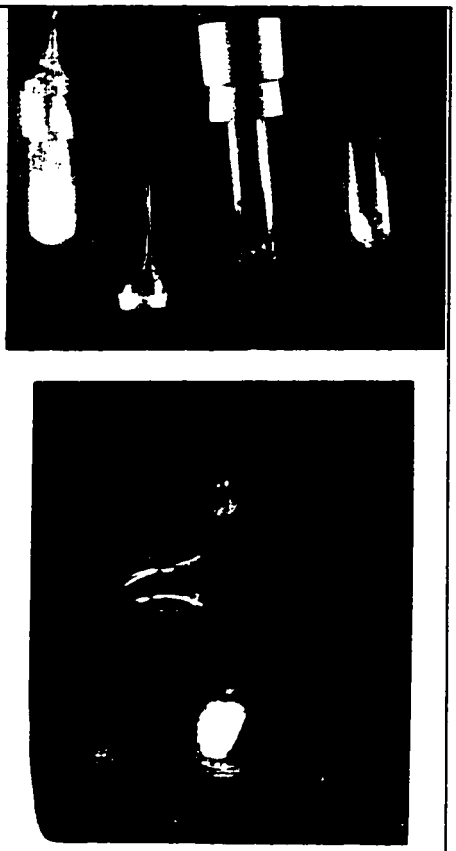


Figure 3.2.3: Reaction vessels for carrying out *in vacuo* reactions. (*top*) left to right: Sealed vessel, vessel reacted in block heater, vessel and attached rubber septum, opened reaction vessel; (*bottom*) left: reagent compartment, right: protein compartment. Attached glass rod (bottom) facilitates sealing the vessel.

***In vacuo* methylation of α-chymotrypsin at LpH 10 for observation by solid state NMR techniques.** α–Chymotrypsin (150 mg) was lyophilized out of 40mM sodium metaborate buffer (7 ml), at a final pH value of 10, and methylated 12 h in a single compartment reaction vessel containing [^{13}C]iodomethane (50 μl). The protein derivative was tightly packed into a standard NMR rotor and analyzed. Alternatively, the methylation could have been monitored directly in the NMR probe at various time points. A standard 5 mm NMR tube was cut and the shortened tube was packed with protein followed by a glass wool plug functioning as a spacer. The NMR tube was narrowed in the same manner as described for the pyrex reaction vessels, but required a very weak flame. A 1:1 (v/v) solution of [^{13}C]iodomethane (30 μl) in a high boiling point solvent such as octane or hexadecane was added and the tube was frozen and sealed *in vacuo* (Figure 3.2.4). The sample withstood incubation in an oven at 100°C for 10 minutes without exploding.

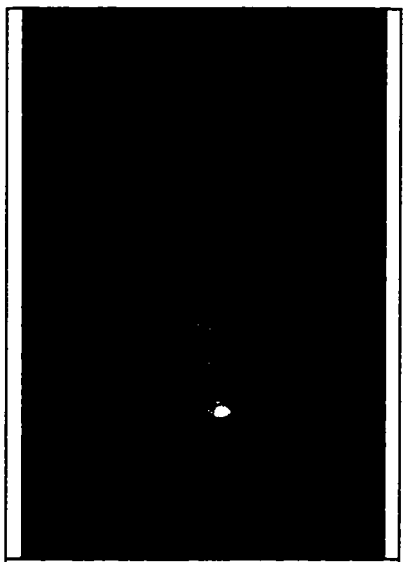


Figure 3.2.4: NMR tube serves as reaction vessel for solid state experiments. Protein (white) is tightly packed in the bottom half. The charred top contains glass wool as spacer.

***In vacuo* methylation of α-chymotrypsin at LpH 8.0 in presence of inhibitors.** α–Chymotrypsin (100 μg) was lyophilized directly in the reaction vessel from a solution of 5 mM sodium phosphate (1 ml), pH 8.0, containing 10 mM indole, 10 mM N-acetyl-L-tryptophan or no inhibitor. Iodomethane (25 μl) was added. The vessel was sealed *in vacuo* and incubated (75°C, 24 h).

Quantification of the activity of *in vacuo* methylated α-chymotrypsin. Protein was dissolved in 100 mM sodium formate buffer pH 4.0 (1ml). An aliquot containing 10 μg of enzyme was used

to determine activity. Rates were measured at pH 8.0 and 20°C on a pH stat apparatus using 10 mM N-acetyl-L-tyrosine ethyl ester in 0.1 M KCl and 5% acetonitrile, as substrate.

***In vacuo* acetylation of α-chymotrypsin at LpH 9.0 with [^{3}H]acetic anhydride.** A solution of α-chymotrypsin (2.5 mg/ml) was adjusted to pH 9.0 with 1N NaOH. Aliquots (1 ml) were lyophilized in the protein compartments of five reaction vessels. Diluted [^{3}H]acetic anhydride (1 X 10 µl [1.43 mCi/mmol: 25 mCi in 1 ml acetic anhydride]) was added to the reaction compartment of four reaction vessels. The fifth was used as a heat-control to which no reagent was presently added. Reaction vessels were sealed under vacuum as described above and incubated at 75°C. Reactions were terminated at various time points and protein isolated as described above. The heat-control protein was removed at the end along with the final reaction vessel. All proteins except the control were afterwards reacted with [^{12}C]acetic anhydride (2 X 20 µl) in 10 M urea (3 ml). The urea solution was made at pH 2.5 and adjusted to pH 9.5 prior to acetylation.

Quantification of ^{3}H incorporation into amino groups. The quantification procedure employed was that used in the competitive labeling technique (Young and Kaplan, 1989). Diluted[*Acetic*-1^{14}C]anhydride (3 X 100 µl [14.3 µCi/mmol: 250 µCi in 1 ml acetic anhydride]) was reacted with protein (90 mg) in 10 M urea (15 ml) to prepare fully ^{14}C-labeled protein for use as recovery standard. It was at this point that diluted [^{3}H]acetic anhydride (2 X 40 µl [1.43 mCi/mmol: 25 mCi in 1 ml acetic anhydride) was used to completely acetylate the heat-control protein (2.5 mg), also in 10 M urea (3 ml). The two protein derivatives enabled calculation of maximum ^{3}H/^{14}C ratios. All derivatized proteins were performic acid oxidized, in particular, to remove insolubilities observed in the nonaqueous reactions (this was not the case for protein

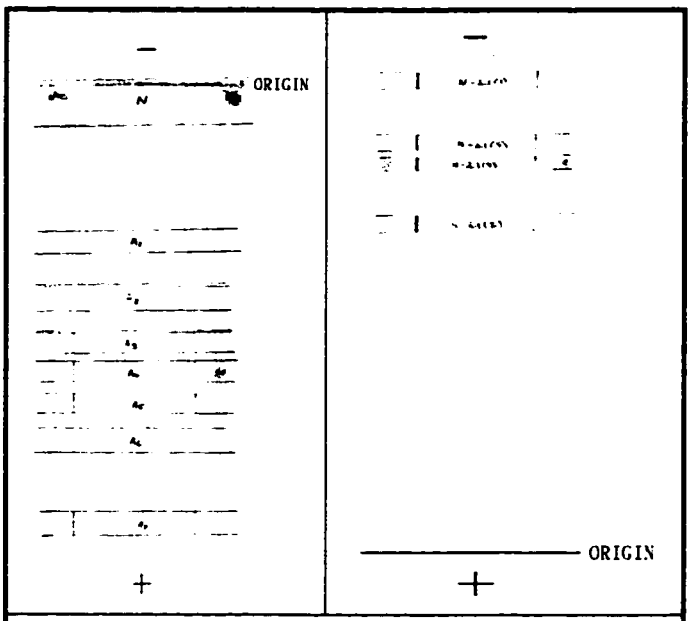


Figure 3.2.5: High voltage paper electrophoresis for separating acetylated peptides. (*left*) 1st dimension, pH 6.5; (*right*) 2nd dimension, pH 2.1. 2d autoradiography.

reacted at LpH 9 in the presence of 100 μmole sodium metaborate, but ^{3}H-incorporation was 20% less) and equal aliquots of the ^{14}C-labeled protein derivatives were transfered to each of the ^{3}H-labeled derivatives. A pepsin digestion completed the fragmentation protocol. Peptides containing the $^{3}H/^{14}C$-acetylated α-amino groups and ε-amino groups were separated by high voltage paper electrophoresis (Figure 3.2.5 and Table 3.2.1; A and N peptides) (Kaplan, 1972) using marker protein (10 mg) labeled in 10 M urea (5 ml) with [*Acetic*-1 ^{14}C]anhydride (100 μCi, 9.2 mCi/mmole) and then [^{12}C]acetic anhydride (2 X 25 μl). After separation in the first dimension, the acidic (A) peptides were counted individually to ensure there were no gross differences of reactivity of the different N-termini, and averaged. Aliquots of $^{3}H/^{14}C$-peptides were transferred into vials containing Aquasol-2 scintillation cocktail (5 ml). $^{3}H/^{14}C$-ratios were quantified on a

Table 3.2.1: Relative mobility of radiolabeled peptides following digestion of *in vacuo* [^{3}H]acetylated α-chymotrypsin (LpH 9.0, 75°C).

Dimension 1 (pH 6.5)		Dimension 2 (pH 2.1)	
Radioactive band	**Mobility with respect to dansylsulphonic acid**	**Radioactive band**	**Mobility with respect to dansylarginine**
A1	0.38	N-2.1 (2)	0.46
A2	0.50	N-2.1 (4)	0.56
A3	0.60	N-2.1 (5)	0.60
A4	0.66	N-2.1 (7)	0.71
A5	0.73		
A6	0.81		
A7	0.99		

LKB RackBeta liquid scintillation counter. Neutral (N-2.1) peptides were not quantified individually after separation in the second dimension. They were counted, instead, as a group of peptides, and a portion of the neutral band of the first dimension was eluted for this purpose.

Acetylation of α-chymotrypsin in octane at LpH 2.0, 9.0 and 10.5 with [^{14}C]acetic anhydride. A solution of α-chymotrypsin (10 mg/ml) was made and 3 aliquots (1 ml) were lyophilized in hydrolysis tubes. Each was then adjusted to pH 2.0, 9.0 or 10.5 using previously adjusted buffer systems (2 ml) of the composition 50 mM phosphate, borate and carbonate and relyophilized. Octane (2 ml) was added to each tube and diluted [*acetic*-1^{14}C]anhydride (1 x 12 µl [14.3 µCi/mmole: 250 µCi in 1 ml acetic anhydride]) was delivered. The tubes were stirred at room temperature for 24 h, the protein was spun down, washed twice with acetonitrile (2 ml) and collected. Each was suspended in 10 M urea (5 ml) and reacted with [^{12}C]acetic anhydride (2 X 25 µl). The quantification procedure was the same as described above for the *in vacuo* acetylation experiment, however, tritiated protein (50 mg) served as recovery standard and was prepared using diluted [^{3}H]acetic anhydride (2 X 25 µl [1.43 mCi/mmole: 25 mCi in 1 ml acetic anhydride]) in 10 M urea (10 ml). Completely labeled ^{14}C-protein (10 mg) was prepared in 10 M urea (5 ml) using [*acetic*-1^{14}C]anhydride (3 x 10 µl [14.3 µCi/mmole: 250 µCi in 1 ml acetic anhydride]). The counting ratio obtained for quantification purposes was ^{14}C/^{3}H.

Nonaqueous fluorescent labeling of carboxyl groups in acetylated bovine serum albumin. Acetylated bovine serum albumin (5 mg, LpH 7.0) was sealed in an evacuated reaction vessel with ethoxyformic anhydride (2 µl) and incubated in an oven at 65°C for 22 h. The reagent chamber was placed in liquid nitrogen to remove any unreacted reagent and the vacuum seal was broken.

The protein was suspended in an Eppendorf tube containing the fluorescent amine (100 μg) dissolved in N,N-dimethylformamide (100 μl). The coupling reaction was terminated after 1 h by repeated extraction of the reagent from the insoluble protein using N,N-dimethylformamide.

Solution ninhydrin and Pauly tests for methylated insulins (and proteins). Duplicate or triplicate aliquots of insulin (10-20 mg methylated under aqueous conditions and lyophilized, 2.5 mg under nonaqueous conditions) were used for either test. The solution Pauly test of insulin (or any protein) was preceded by an initial performic acid oxidation or trace pepsin digestion (chapter 2) to guarantee complete solubility during spectrophotometric analysis; it was previously observed that the Pauly chromophore was retained by protein insolubles. The Pauly test reference (Legget Bailey, 1962) contains a typographical error ("nitric acid" should read "nitrous acid") but we used a solution of sodium nitrite (7%). Sometimes the colour formation of *in vacuo* methylated protein was nonreproducible and unusually dark and it was believed to be due to interference by retained iodide. A 12 h delay before measurement sometimes ameliorated the problem. Alternatively, a dialysis prior to the Pauly test was useful. As a final check, methylated protein was subjected to acid hydrolysis (chapter 2), the amino acids were separated at pH 6.5 by high voltage paper electrophoresis (60V/cm, 30 minutes), and a Pauly spray test with a detection threshold of 0.1 μg histidine (Legget Bailey, 1962) confirmed the solution test. The solution ninhydrin test was a modification of the 1994 2nd Year University of Ottawa Biochemistry Laboratory Manual procedure. Protein insolubility was not a problem for the solution ninhydrin quantification since the chromophore leaches out of the protein matrix and into solution. Protein was dissolved/suspended in 3.8 M sodium acetate (0.5 ml, pH 5.3) containing 0.2 mM sodium cyanide. Cyanide was obtained from a fresh stock (1 M) in sodium acetate buffer, pH 5.3.

Methyl cellosolve (0.5 ml) containing 3% ninhydrin and ddH_2O (1 ml) were added and the sample was incubated with shaking at 75°C for 45 minutes. The reaction was terminated by adding 5 ml 1:1 (v/v) isopropanol/water and samples were agitated for 15 minutes and centrifuged to remove insoluble protein. The sensitivity of cadmium-ninhydrin spray reagent is reported at 0.5 nmole (Darbre, 198b) and a comparable sensitivity is anticipated in solution. Concentrated samples were diluted (1:1 isopropanol/water) before measurement (A_{570nm}). In both tests, controls/blanks were required and Beer's Law was obeyed to at least 0.8 absorbance units.

NMR Spectra. Standard [^{13}C]NMR spectra were obtained using a Gemini 200 MHz spectrometer with 1s between pulses. For the aqueous reactions derivatized proteins were dialyzed against 10 mM HCl and lyophilized. Methylated protein samples and standards were analyzed in 9M urea (chapter 2) which gave a pH meter reading of 8. Acetonitrile (30 µl) with a ^{13}C-chemical shift for its methyl group of 1.70 (Breitmaier and Voelter, 1987a) ppm was added to reference peak resonances. Methylated samples that were analyzed using a Bruker ASX-200 solid state NMR instrument required cross-polarization pulse sequences, with a 4s relaxation delay, 3.8µs 90° proton pulse, 2ms contact time, 50ms acquisition time and 18kHz proton spectral window (Fyfe, 1983a). Magic angle spinning (Fyfe, 1983b) of protein was carried out at 4 kHz except that of commercially available protein, which was spun at 5.5 kHz. Spectra were referenced to external hexamethylbenzene (16.9 ppm for the methyl carbons).

3.3 Results

Two experimental approaches for chemical modification of proteins in nonaqueous media were investigated. The first parallels the approach developed by Klibanov (Klibanov, 1984; Zaks

and Klibanov, 1988; Broos *et al.*, 1995) for enzymatic reactions in organic solvents. A protein solution was adjusted to the desired pH value, lyophilized and dispersed in octane. The modifying reagent was added to the protein dispersion and the reaction mixture was stirred in a temperature-controlled oven. The modified protein could be isolated simply by filtration or centrifugation, washed with octane and the residual organic solvent removed under vacuum. In a second approach, the reaction was carried out directly on the lyophilized protein. A reaction vessel with either one or two compartments was employed (Figure 3.2.3). Only acetylation reactions were carried out in the double compartment reaction vessel. In those cases, protein solution was lyophilized in one compartment and then modifying reagent was added to the other which was immersed in liquid nitrogen. The reaction vessel was sealed under vacuum and placed in a temperature-controlled oven. To terminate the reaction, the unreacted reagent was trapped out by placing the reagent compartment in liquid nitrogen and releasing the vacuum. The principle for employing the single chamber reaction vessels was the same. The modified protein from either procedure was dissolved in an aqueous medium for analysis by NMR spectroscopy or other analytical procedure. The corresponding aqueous reactions were carried out in the usual manner (Imoto and Yamada, 1989) by adding reagents to stirred protein solutions.

(i) Methylation of insulin. The reactions of insulin with iodomethane in water at pH 7.5 and pH 10 were compared with the nonaqueous reactions of insulin at LpH 7.5 (*lyophilized* at pH 7.5) and at LpH 10. The potential reactive functional groups of insulin are the α-amino termini of glycine and phenylalanine, the ε-amino group of lysine, four tyrosine phenolic hydroxyl groups and two histidine imidazole groups. Resonances corresponding to the methylated derivatives of all these groups were observed in NMR spectra of the ^{13}C-labeled proteins (Figure 3.3.1). Peak

resonances were assigned to their methylated derivatives by comparing their chemical shifts with those of [^{13}C]methylated standard compounds containing the expected functional groups (Table 3.3.1). The following similarities in the water, octane and *in vacuo* reactions were observed with

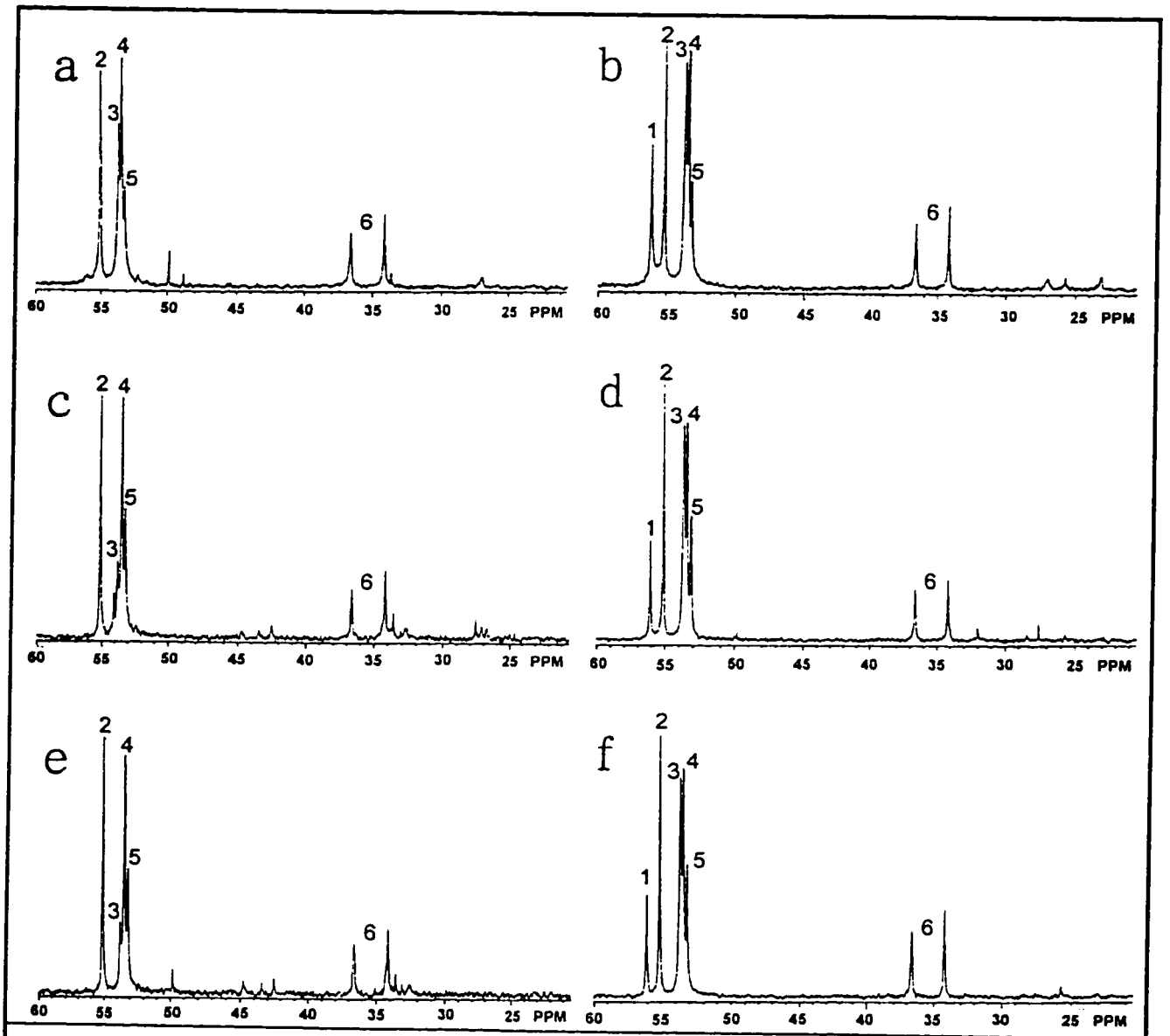


Figure 3.3.1. Expanded 50 MHz ^{13}C-NMR spectra (1024 transients) of 20 mg bovine insulin reacted with [^{13}C]iodomethane. Reaction conditions were: a) water pH 7.5, 24 h, 37°C; b) water pH 10, 24 h, 37°C; c) *in vacuo* LpH 7.5, 24 h, 75°C; d) *in vacuo* LpH 10, 12 h, 75°C; e) octane LpH 7.5, 24 h, 75°C; f) octane LpH 10, 12 h, 75°C. Peak resonances correspond to the methyl groups in: 1- Tyr(OMe); 2- $\mathbf{Me_3N^+}$-Gly; 3- Lys(ε-$^+$N$\mathbf{Me_3}$); 4- $\mathbf{Me_3N^+}$-Phe; 5- α-$^+$N$\mathbf{Me_3}$ unidentified; 6- His(Im$^+$$\mathbf{Me_2}$).

insulin: a) the same functional groups are modified in the water, octane and *in vacuo* reactions, b) the same derivatives of various functional groups are obtained, viz. amino groups are trimethylated to the quaternary state, histidine forms the dimethylimidazolium cation derivative and tyrosine forms the phenolic O-methyl derivative and c) the phenolic hydroxyl of tyrosine does not react at pH 7.5 or LpH 7.5 but reacts at pH 10 and LpH 10. Again, no methyl esters were found.

The octane and *in vacuo* reactions differed from the aqueous reaction in that the degree of methylation of the phenylalanyl α-amino, glycyl α-amino and most significantly the lysyl ε-amino group was considerably less at LpH 7.5 than at LpH 10. These differences are greater than is apparent in figure 3.3.1 because the vertical scales used in figures 3.3.1d and 3.3.1f are attenuated with respect to figures 3.3.1c and 3.3.1e. Insulin reacted under aqueous conditions (24 h, T = 37°C and pH = 10) or nonaqueous conditions *in vacuo* (24h, T=75°C and LpH =10) tested negative with Pauly's diazo reagent (Leggett Bailey, 1962; Darbre, 1986b) in both cases, weakly ninhydrin positive for the nonaqueous sample and ninhydrin negative for the aqueous sample. An extra resonance at 53.12 ppm in the chemical shift region for the α-amino termini is present in the spectrum of methylated insulin (Figure 3.3.1). Extra resonances are not always observed with other

Table 3.3.1: Chemical shifts in ppm for [^{13}C]Methyl Groups of Amino Acid Standards, Methylated Insulin and Methylated α-Chymotrypsin

[^{13}C]Standards	Chemical Shift (ppm)	Chemical Shift (ppm)	
		Insulin	α-Chymotrypsin
Ac-NH-Tyr(OMe)-NH_2	56.12	56.00	56.18
Me_3N^+-Gly-Leu	55.06	55.07	
H_2N-Lys(ε-$^+NMe_3$)	53.66	53.62	53.69
[Me_3N^+-Cys(NH_2)-S-]$_2$	53.42		53.50
Me_3N^+-Phe-Gly-Gly	53.41	53.48	
Me_3N^+-Ile-NH_2	53.03		53.21
Me_3N^+-Ala-Ala	52.60		52.71
H_2N-Lys(ε-$^+NHMe_2$)	43.42		43.37
Ac-NH-His(Im$^+Me_2$)-NH_2	34.04, 36.46	34.10, 36.58	34.08, 36.59
H_2N-Lys(ε-$^+NH_2Me$)	33.59		33.55
$^+SMe_3$ (Breitmaier and Voelter, 1987b)	27.5		27.54
H_2N-Met(^+SMeMe)	25.41		25.65

preparations of bovine insulin (refer to chapter 2). It is therefore believed to be due potentially to the presence of some insulin with an internal proteolytic cleavage in this commercial preparation.

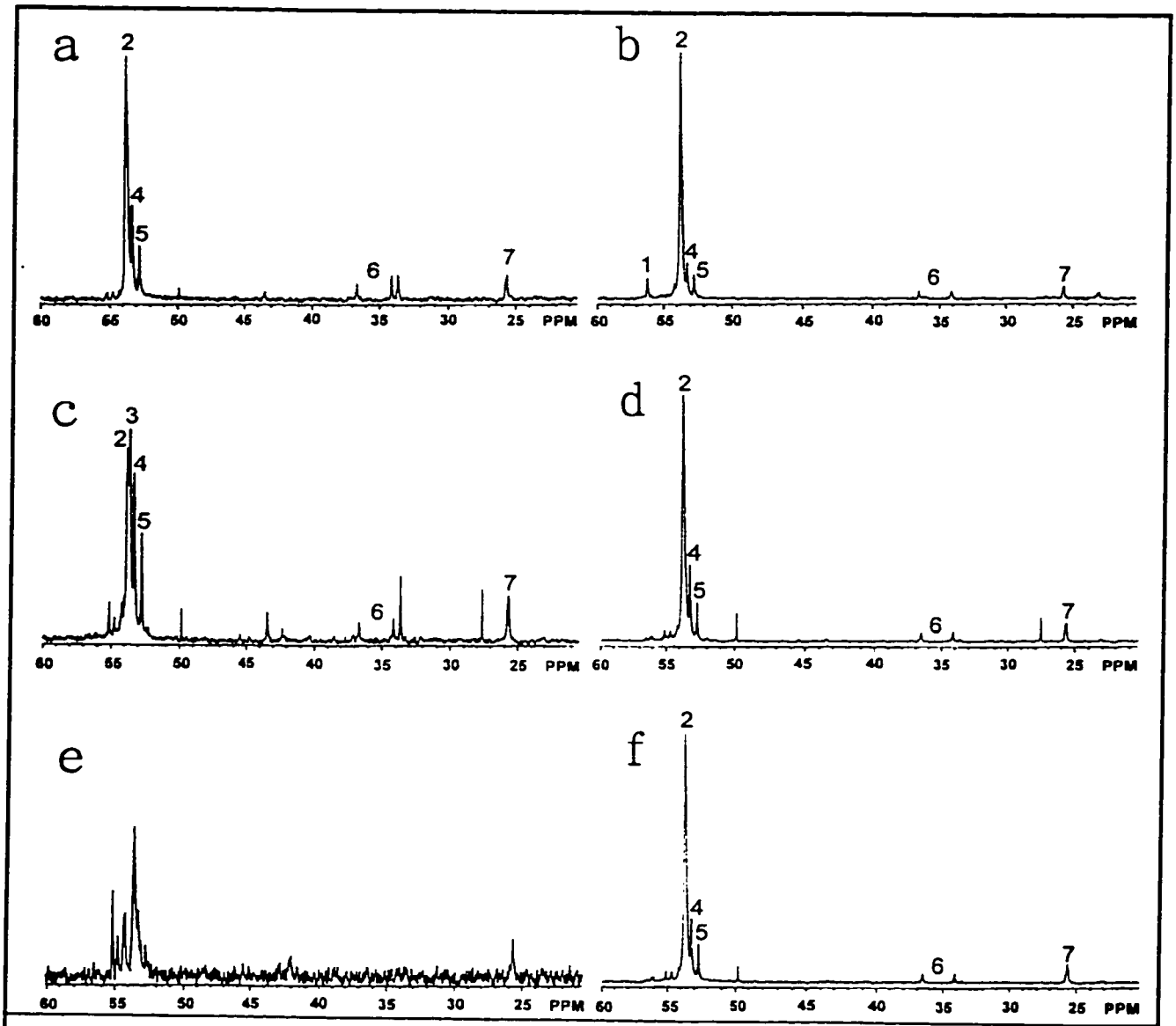


Figure 3.3.2. Expanded 50 MHz [^{13}C]NMR spectra (1024 transients) of 20 mg bovine α–chymotrypsin and diisopropylphosphoryl-α-chymotrypsin reacted with [^{13}C]iodo-methane. Reaction conditions were: a) diisopropylphosphoryl-α-chymotrypsin, water pH 7.5, 24 h, 37°C; b) diisopropylphosphoryl-α-chymotrypsin, water pH 10, 24 h, 37°C; c) α-chymotrypsin, *in vacuo* LpH 7.5, 24 h, 75°C; d) α-chymotrypsin, *in vacuo* LpH 10, 12 h, 75°C; e) α-chymotrypsin, water pH 7.5, 24 h, 37°C; f) α-chymotrypsin , octane LpH 10, 12 h, 75°C. Peak resonances correspond to the methyl groups in: 1- Tyr(O**Me**); 2- Lys(ε-$^{+}$N**Me**$_3$); 3- **Me**$_3$N^{+}-Cys; 4- **Me**$_3$N^{+}-Ile; 5- **Me**$_3$N^{+}-Ala; 6-His(Im^{+}**Me**$_2$); 7-Met($^{+}$S**Me**Me).

(ii) Methylation of α-chymotrypsin. The aqueous reaction of diisopropylphosphoryl-α-chymotrypsin at pH 7.5 and pH 10 with [^{13}C]iodomethane (Figures 3.3.2a and 3.3.2b) gave all the methylated derivatives observed with insulin. In addition, the dimethylsulphonium derivative of the methionine side-chain which is not present in insulin was observed. Nonaqueous reactions of α-chymotrypsin with [^{13}C]iodomethane differ from the aqueous reaction (Figure 3.3.2b) in that no O-methyltyrosine is observed at LpH 10 (Figures 3.3.2d and 3.3.2f). Methylation of ribonuclease paralleled that of chymotrypsin in that under aqueous conditions tyrosine was methylated at pH 10 but not at pH 7.5, and not methylated under nonaqueous conditions at LpH 10. α-Chymotrypsin has three amino termini and therefore three resonances are expected in the chemical shift region of trimethylated α-amino groups as observed in the nonaqueous reaction (Figure 3.3.2c). Peak 3 for the trimethylated cystine α-amino terminus is not visible in figures 3.3.2a, 3.3.2b, 3.3.2d and 3.3.2f because of the very intense neighbouring trimethylated lysine ε-amino resonance but is resolved in higher field spectrometers where finer shimming is possible. It was necessary to attenuate the vertical scale in figures 3.3.2b, 3.3.2d and 3.3.2f in order to accommodate the intense resonance at 53.69 ppm, due to the superimposition of the 14 lysine residues in α-chymotrypsin at pH 10 and LpH 10. For this reason the peak intensities for the other resonances at pH 10 and LpH 10 appear weaker than they do in the spectra for reactions at pH 7.5 and LpH 7.5 (Figures 3.3.2a and 3.3.2c) but the degree of methylation of these groups is at least as great or greater. When chymotrypsin was not inactivated with diisopropyl-fluorophosphate, more than three multiple peak resonances were observed in the aqueous reaction (Figure 3.3.2e) indicating that, unlike the nonaqueous reaction, autolysis had occurred generating additional α-amino groups. Also observed are resonances at 49.79 ppm ($^{13}CH_3OH$) probably due

to traces of water reacting with iodomethane, 43.37 ppm (^{13}C-ε-dimethylamino lysine) and 33.55 ppm (^{13}C-ε-monomethylamino lysine) due to incomplete methylation at LpH 7.5, and 27.54 ppm (trimethylsulphonium cation) possibly arising from breakdown of disulphide bridges. These are variable resonances arising from incomplete derivatization or minor side-reactions. For aqueous modifications, it is common practice to dialyze proteins following reaction. Consequently, any small breakdown products would not be observed. This is particularly evident in figure 3.3.2e where a large portion of the autolysis products have been removed by dialysis, requiring expansion of the vertical scale. For the nonaqueous reactions reported here the proteins need not be dialyzed so all low molecular weight degradation products are observed.

(iii) Protection studies. α-Chymotrypsin that was lyophilized at pH 8 in the presence of competitive inhibitors, indole or N-acetyl-L-tryptophan, was reacted *in vacuo* with iodomethane. In the absence of any inhibitor or with indole the enzyme was completely inactivated by reaction with iodomethane. In sharp contrast, even after 24 h reaction, substantial activity was retained when N-acetyl-L-tryptophan was present (Table 3.3.2).

Table 3.3.2: Activity Retained by Lyophilized α–Chymotrypsin in Presence and Absence of Competitive Inhibitors after Reaction with Iodomethane at 75°C for 24 h, LpH 8.0.

Inhibitor	%activity *
no inhibitor	0
indole	0
N-acetyl-L-tryptophan	68 ± 2

*Average of 2 trials relative to untreated α-chymotrypsin.

(iv) Acetylation of α-chymotrypsin. Modification of α-chymotrypsin was attempted under nonaqueous conditions using excess [^{14}C]acetic anhydride (Table 3.3.3). In a preliminary investigation, acetic anhydride was allowed to react with protein dispersed in octane at room temperature. The reaction was characterized by quantifying the ^{14}C/^{3}H ratio of incorporated acetyl

groups in peptides (Kaplan, 1972). Several lyophilization pH values were employed and the results, presented in table 3.3.3, indicate the radioactive incorporation increases in response to an increase of the LpH value. The extent of derivatization, however, was low and suggested that maximum reaction had not been attained in any of the examples. For this reason, further interpretation of the data was not attempted.

Table 3.3.3: Relative* incoporation of acetic anhydride into amino groups of α-chymotrypsin by nonaqueous derivatization in octane.

Amino group	LpH 2	LpH 9	LpH 10.5
α-amino	0.00534	0.0421	0.0864
ε-amino	0.0126	0.0529	0.120

*fraction relative to complete derivatization in water. T= 25°C, time = 24 h.

Table 3.3.4: Relative incorporation* of acetic anhydride into amino groups of α-chymotrypsin, LpH 9.0, by *in vacuo* nonaqueous derivatization

Amino group	Reaction time *in vacuo* (LpH 9.0)			
	1 h	8 h	24 h	60 h
α-amino	0.34	0.42	0.58	0.87
ε-amino	0.21	0.23	0.25	0.26

*fraction relative to complete derivatization in water. T = 75°C.

Modification with acetic anhydride was subsequently carried out on α-chymotrypsin *in vacuo* with protein at LpH 9 using excess [^{3}H]acetic anhydride. The reaction was monitored by quantifying the $^3H/^{14}C$ ratio, namely, the extent of incorporation of tritium into the acetylated α and ε-amino groups (Table 3.3.4) after proteolytic digestion and isolation of acetylated peptides (Kaplan, 1972). Complete acetylation of amino groups of proteins is readily achieved in water at pH 9 but in the nonaqueous reaction only 25% of the ε-amino groups and 90% of the α-amino groups were modified at prolonged reaction times. It also appears that acetylation of the available ε-amino groups reached completion much more rapidly than acetylation of the α–amino groups.

(v) Carboxyl group activation. Acetic anhydride forms a mixed anhydride with the side-chain carboxyl groups of proteins in water but these are rapidly hydrolyzed. However, in a nonaqueous medium hydrolysis cannot occur. Indirect evidence for the presence of anhydride was obtained

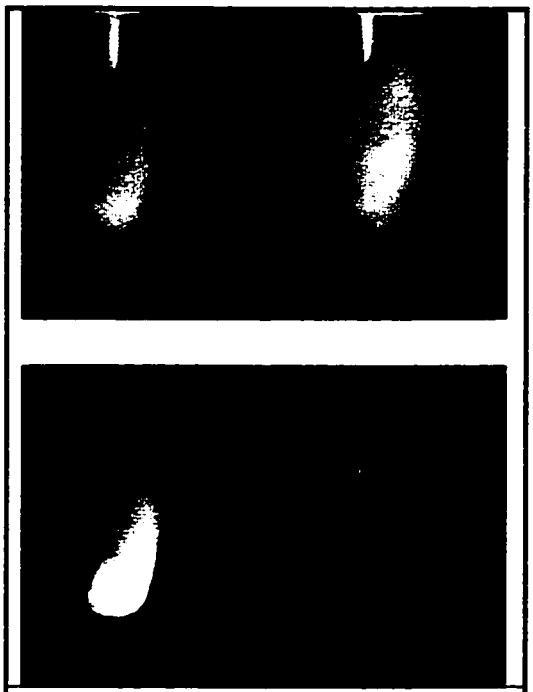


Figure 3.3.3. Chemical coupling of the carboxyl groups of acetylated bovine serum albumin with a fluorescent amine. The protein samples are shown under visible light (*top*) and broad band ultraviolet light (*bottom*). The protein on the left was reacted *in vacuo* with ethoxyformic anhydride and then coupled with the fluorescent amine. The protein on the right was treated the same way except no ethoxyformic anhydride was added.

from the observation that *in vacuo* ^{3}H-acetylated protein lost over 50% of its radioactivity after dialysis in 8M urea. Direct evidence for the formation of anhydrides under nonaqueous conditions was obtained by acetylating all the amino groups of bovine serum albumin under aqueous conditions, reacting the lyophilized acetylated protein *in vacuo* at LpH 7 with ethoxyformic anhydride and then incubating with a small amount of fluorescent amine (N-1-[5-dimethylamino-1-naphthalenesulfonyl]-2-hydroxy-1,3-diaminopropane) in N,N-dimethylformamide. An intense fluorescence was observed under ultraviolet light with the *in vacuo* modified protein but not in the control protein which was incubated *in vacuo* without ethoxyformic anhydride (Figure 3.3.3). Although amino groups do not necessarily have to be blocked, N-acetylated bovine serum albumin was used as a test protein to (i) ensure that carboxyl groups were the predominant nucleophile available in the protein for reaction with ethoxyformic anhydride, and (ii) prevent excessive crosslinking of protein (see discussion) due to intermolecular couplings of carboxyl and amino groups. Despite the precaution, the protein was insoluble after reaction.

(vi) Observation of derivatized protein functional groups by solid state NMR techniques. Solid state [^{13}C]NMR studies of protein derivatives have been carried out (Yan *et al.*, 1996). The *in vacuo* methylation of α-chymotrypsin at LpH 10 was observed by solid state NMR

spectroscopy using cross-polarization and magic angle spinning techniques (Figure 3.3.4). The resolution obtained was poorer than the solution spectra, however, it was obvious that three classes of functional groups of protein reacted, namely, amino groups, imidazole groups and sulphide groups. The appearance of peaks in the blank (underlying spectra) had not been observed before in solution spectra. When octane was added to the blank, the spectrum obtained (not shown) was unchanged. The spectrum of α-chymotrypsin taken directly from the bottle was also identical despite the fact that the protein appeared significantly more crystalline.

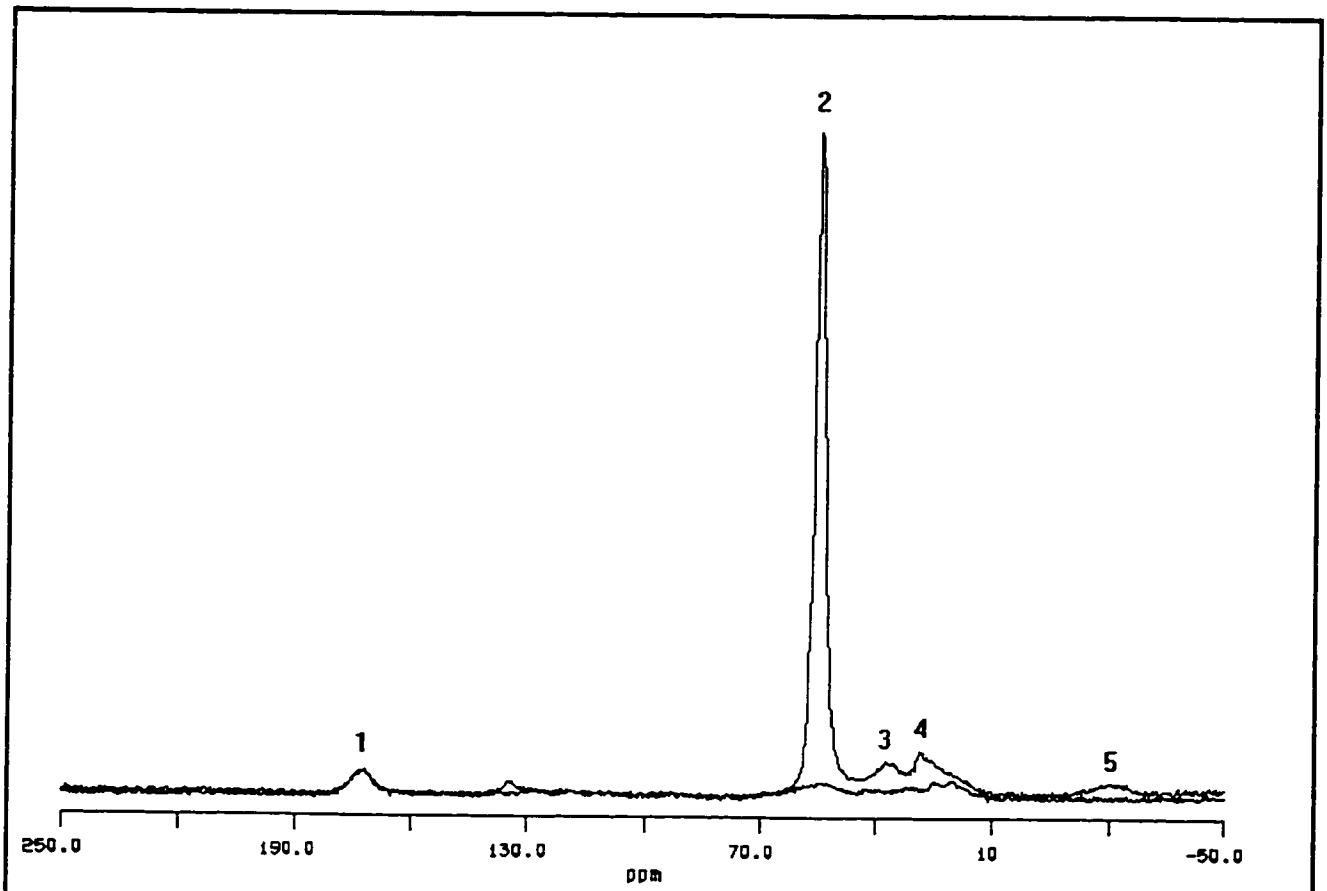


Figure 3.3.4: Superimposed 50 MHz solid state [^{13}C]NMR spectra of α-chymotrypsin, LpH 10.0, before (*underlying spectrum*, 17000 scans) and following (*overlaid spectrum*, 1000 scans) *in vacuo* methylation. Cross-polarization techniques and magic angle spinning were used. Carbon resonances of protein correspond to the following: 1) amide carbonyl carbons; 2) trimethylated ammonium groups; 3) dimethylated histidyl side-chains; 4) methylated methionyl side-chains and other sulphide derivatives; 5) motionally restricted [^{13}C]iodomethane.

3.4 Discussion

Klibanov and his co-workers (Klibanov, 1984; Zaks and Klibanov, 1988; Klibanov, 1989; Wescott and Klibanov, 1994; Broos *et al.*, 1995) observed that enzymes had certain properties in organic solvents which provided a rationale to carry out structure-function studies using chemical modification reagents under nonaqueous conditions. The most important feature is that proteins retain an active conformation which is at least very similar to the native solution structure. Proteins in organic solvents have a greatly enhanced thermostability which would permit the use of elevated temperatures to accelerate modification reactions without unfolding the native structure. For the *in vacuo* approach, it is expected that the thermal stability of the protein will be as great, if not greater, than in organic solvents. The thermal stability most likely arises from the rigidity of the protein structure in nonaqueous media (Zaks and Klibanov, 1988; Broos *et al.*, 1995). Reactivity data from nonaqueous reactions may therefore be more readily interpreted than in solution where it is complicated by the need to take into consideration the effect of all the dynamic conformational equilibria on the observed chemical properties (Young and Kaplan, 1989; Kaplan *et al.*,1971). These attributes and others suggest that there are significant potential advantages to carrying out chemical modifications under nonaqueous conditions:

(i) Use of volatile, organic soluble and water-labile reagents for protein modification. Reductive methylation of amino groups with formaldehyde and borohydride seldom results in disruption of the protein structure or loss of biological activity (Means and Feeney, 1971; Glazer *et al.*, 1976; Means, 1977; Lunblad and Noyes, 1984; Imoto and Yamada, 1989; Lundblad, 1995) and for this reason is widely used in protein structure-function studies (Zhang and Vogel, 1993; Zhang *et al.*, 1994). It is noteworthy to point out that formaldehyde and borohydride are quite

miscible in water. In contrast, iodomethane has been used so infrequently for protein methylation that it was not included in a catalogue of chemical reagents for protein modification (Lunblad and Noyes, 1984). As suggested in chapter 2, part of the reason for this is its sparing solubility in water, however, in organic solvents or under *in vacuo* conditions, the availability of potential modification reagents is greatly increased. While reductive methylation proceeds to give at most the dimethylamino derivative, iodomethane gave the quaternary trimethyl derivative. Although quaternization of amino groups is known to occur *in vivo* as a post-translational modification of proteins (Paik and Kim, 1975), this modification, to our knowledge, was never reported as an *in vitro* chemical modification of amino groups in native proteins and provides a means of placing a permanent positive charge on the α and ε-amino groups at all pH values. The formation of a dimethylimidazolium cation derivative with the side-chain of histidine was an additional bonus since this derivative has never been reported as either an *in vitro* chemical modification of a native protein or as an *in vivo* post-translational modification. Similarly, dimethylation of the imidazole function provides a means of placing a positive charge on the side-chain at all pH values.

The trimethyl ammonium and dimethylimidazolium derivatives are presumably formed via a series of deprotonation and alkylation steps. If this is true, it would appear that in both the octane and *in vacuo* reactions, hydroiodic acid must ultimately be transferred to and neutralized by lyophilized buffer or base for the methylation to proceed as observed. Perhaps the elevated temperature and the solubility of hydroiodic acid in octane or volatility of hydroiodic acid under evacuated conditions promoted such a transfer.

(ii) Use of the pH memory effect to achieve selective modification. It appeared from the results obtained in this initial investigation that most of the factors which affected the relative

reactivity of various functional groups of proteins in solution were retained by the lyophilized protein in the nonaqueous environment. Among the possible determinants, the most important were the intrinsic nucleophilicity of the group, the protein microenvironment and the ionization state of the group (Kaplan *et al.* 1971; Young and Kaplan, 1989). The reaction of ε-amino groups with acetic anhydride reached completion faster than the α-amino groups (Table 3.3.4) which was expected on the basis of their relative nucleophilicities. However, the degree of derivatization of the ε-amino group was much less than that of the α-amino groups. This would be expected on the basis of the pH memory effect (Zaks and Klibanov, 1988) where ionizable groups retain the ionization state they had in the solution from which they were lyophilized (Figure 3.4.1). In nonaqueous media there is no water present for a dynamic equilibrium to be established between the two ionization states, i.e. a group that is protonated will not deprotonate and its rate of reaction will be negligible relative to the deprotonated group. With pK_a values of approximately 7 to 8, more that 50% of the α-amino groups would be expected to be derivatized at LpH 9, whereas with pK_a values of approximately 10 to 11, less than 50% of the ε-amino groups would react (Creighton, 1993a). This pH memory effect is also evident in the reactions of the ε and α-amino groups and imidazole groups with iodomethane. At LpH 7.5, the degree of reaction of the ε-amino groups, relative to that of the α-amino groups, is clearly much less than at LpH 10 while the imidazole group with a pK_a value of approximately 6 to 7 (Creighton, 1993a) shows a much smaller difference (Figures 3.3.1 and 3.3.2). Since functional groups in proteins have pK_a values (Creighton, 1993a) varying from 3.5

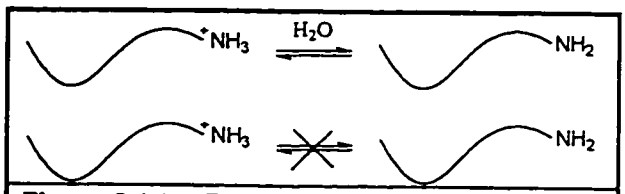


Figure 3.4.1: Properties of ionization states in aqueous and nonaqueous environments. (*top*) ionizable functional groups follow equilibrium dynamics in solution. (*bottom*) groups retain the ionization state of solution after removal of water.

to 12, this phenomenon of pH memory has the potential to be utilized as a means of achieving selective chemical modifications of ionizable functional groups by controlling the pH of lyophilization. For example, gaseous diazomethane could be used to preferentially form methyl esters of side-chain carboxyl groups over C-terminal carboxyl groups by lyophilizing protein at LpH 3.5-4 where the side-chain carboxyl groups are mainly protonated.

(iii) Enhanced ability over aqueous protocols to probe structure. The tyrosine phenolic hydroxyl function is readily methylated under aqueous conditions at pH 10 with model compounds and proteins used in this study. There was, however, a notable difference in the reactivity of this group between insulin and both α-chymotrypsin and ribonuclease in a nonaqueous environment in that only in insulin was this function methylated at LpH 10. This suggests that the phenolic side-chains are buried in the major conformational states of α-chymotrypsin and ribonuclease in solution at pH 10. The reason these side-chains react in an aqueous, but not in a nonaqueous environment, is due to the dynamic equilibrium that exists between the various conformational states in solution and, although the tyrosyl side-chains are exposed only in minor conformations, this can lead to substantial reaction over a period of time due to Le Châtelier's principle. In the nonaqueous state no such dynamic equilibria exist so that no substantial modification can occur. Insulin being a very small protein is unfolded to a large extent in solution at pH 10, and at LpH 10, exposing some or all of its tyrosyl side-chains.

(iv) Elimination of proteolysis during reaction. Proteolysis is a common problem in the study of proteins (Scopes, 1987; Volkin and Klibanov, 1989). Chemical modification of proteins often disrupts the native structure making the modified protein extremely sensitive to degradation by

trace amounts of proteolytic enzymes. It is particularly difficult to chemically modify proteases such as α-chymotrypsin under conditions where the enzyme is catalytically active without some autolysis. In the present study evidence for substantial hydrolytic breakdown was obtained in the aqueous reaction of α-chymotrypsin with iodomethane at pH 7.5 (Figure 3.3.2e). In contrast, the *in vacuo* reaction gave peak resonances only for the expected three methylated amino termini (Figure 3.3.2c) demonstrating that nonaqueous conditions can be used to perform the derivatization while eliminating hydrolytic breakdown of the protein. In the case of insulin, the presence of an extra α–amino resonance indicated a heterogeneity in the insulin preparation due to proteolysis. This ability to detect free α-amino termini demonstrates that the nonaqueous ^{13}C-methylation procedure with iodomethane can provide another criterion for the homogeneity of protein preparations or evidence for the presence of blocked amino termini.

(v) Novel protection studies. Ligands which bind to proteins have been shown to protect functional groups in the binding regions from chemical modification (Means and Feeney, 1971; Bosshard, 1979). However, this approach for identifying functional groups in the active sites of enzymes has found only limited application. Because most competitive inhibitors have relatively weak binding affinities, active site functional groups are still readily modified even in the presence of these inhibitors due to the dynamic binding equilibria that exist in aqueous media. In contrast, no such dynamic equilibria can exist under *in vacuo* conditions so if a nonvolatile ligand is bound, it cannot dissociate and the residues in that binding site will not be expected to react.

The results reported in table 3.3.2 show that lyophilized α-chymotrypsin is inactivated by methylation. As histidine-57 is essential for catalytic activity (Creighton, 1993b), methylation of this residue is expected to result in loss of activity. Indole which is a good competitive inhibitor

($K_I = 0.72$ mM at 25°C) (Barman, 1969) does not offer any protection whereas N-acetyl-L-tryptophan, which is a much weaker inhibitor, ($K_I = 17.5$ mM at 25°C) (Barman, 1969) gives substantial protection. α-Chymotrypsin was lyophilized from a solution containing these inhibitors at a concentration of 10 mM. It is therefore expected that virtually all the indole binding sites in the lyophilized enzyme are occupied whereas a substantial portion of the binding sites are free in the case of N-acetyl-L-tryptophan. The reason there is such a striking difference in protection is that the N-acetylproprionyl moiety of the latter example covers the catalytic site. If it can be assumed that 68% of the enzyme has the inhibitor bound and 32% is free, a dissociation constant (K_I) of 5 mM at 0°C is calculated, which is on the order of the experimental value reported at 25°C. Another interesting inference that can be made from the results is that dimethylation of the other histidine residue and trimethylation of the amino groups does not inactivate the enzyme. It is important to note that such information would be very difficult to obtain in an aqueous medium and provides yet another example of the potential utility of the chemical modification of proteins in nonaqueous environments. Detailed protection studies are being carried out in our laboratory by Nicolas A. S. Stewart.

(vi) Modification of carboxyl groups. Activated carboxyl groups are short-lived in water which require the use of high concentrations of activating reagents and nucleophiles to achieve extensive modification of carboxyl groups (Means and Feeney, 1971; Glazer *et al.*, 1976; Lunblad and Noyes, 1984; Imoto and Yamada, 1989; Lundblad, 1995). As demonstrated with acetylated bovine serum albumin, another advantage of nonaqueous conditions is that stable activated derivatives of carboxyl groups in proteins can be prepared with relatively small amounts of reagents. Coupling with nucleophiles to form water stable derivatives can then be carried out. In

the present study coupling was carried out with a fluorescent amine in N,N-dimethylformamide (Figure 3.3.3). However, there are a large number of procedures that could be used for such couplings and further attempts are required to determine the best strategies and their limitations. The need for additional work is best illustrated by the fact that most procedures attempted thusfar have resulted in unacceptable crosslinking of protein. Crosslinking was initially observed as a byproduct of excessive heating. When heated in the presence of alkylating reagents, the extent of crosslinking was found to decrease. The extent of crosslinking was protein dependent, but always less in methylated proteins than in heated blanks. In contrast, the problem was far worse when carboxyl group activating reagents (*vide* refs. cited in chapter 5) such as carbodiimides or anhydrides were employed. The problem was presumably due to intermolecular amide bond formation between amino and carboxyl groups of neighbouring proteins and this premise is supported by the fact that (i) very high molecular weight material is observed following gel electrophoresis of reacted protein, and (ii) proteins in solution do not crosslink intermolecularly to as great a degree. Even aqueous acetylation or trimethylation of amino groups prior to the use of carbodiimides could not prevent crosslinking, suggesting that intermolecular ester bond formation could also be playing a role. Apart from attempting to vary activating reagents, times and temperatures, lyophilization of protein at different pH values and in different matrixes were also investigated. Lyophilization at different pH values were inconclusive with respect to the best value to employ. Lyophilization in the presence of inorganic salts such as sodium chloride, zwitterionic species, or polyethyleneglycol did not improve the situation either. Of all initial investigations into this matter, the only factor which could be controlled, and which drastically affected the extent of crosslinking, at least apparently, was the lyophilization volume. Proteins which were lyophilized

in large volumes did not tend to crosslink to the same extent as those in smaller volumes. This was also consistent with the postulate that lyophilized proteins form intermolecular salt bridges. Presumably, as the lyophilization volume decreases, the proteins become more concentrated and form a higher percentage of intermolecular salt linkages which can be crosslinked. The number of salt linkages which actually participate in the crosslinking reaction is presumably low. This is supported by the observation that some arginine diacetate (*in vacuo*, 100°C, 24h) condensed to form N-acetyl arginine acetate (yield ≅ 0.1%). When the mixture was separated by high voltage paper electrophoresis at pH 6.5 and analyzed visually by spraying the chromatogram with the Sakaguchi reagent (*vide* chapter 5), a characteristic neutral band (N-acetyl arginine) was seen. The small extent of crosslinking is also suggested by the observation that α-chymotrypsin (LpH 9, data not shown) which was *in vacuo* labeled with [^{14}C]acetic anhydride (60 h, 75°C) incorporated only 20% additional radioactivity when compared to an identically treated sample which did not crosslink, apparently due to the inclusion of metaborate in the lyophilization matrix.

(vii) Convenience of characterization. Overseeing the progress of any chemical reaction is often desirable, however, some of the methods can be inconvenient and even wasteful. It is therefore useful to explore nonintrusive methods, and the use of solid state NMR spectroscopy is presented in this context. The *in vacuo* reaction of functional groups of α-chymotrypsin, LpH 10, with [^{13}C]iodomethane were observed using cross-polarization and magic angle spinning methods (Figure 3.3.4, overlying spectrum). The reacted protein lyophilisate was packed into an NMR rotor and analyzed. Had the reaction been incomplete, the protein could have been reacted further without any additional preparation. The protocol was even more convenient than it would have been in the case of solution NMR spectroscopy, since the sample would have to be separated

from urea and relyophilized before it could be reacted with additional iodomethane. The linewidths obtained (Figure 3.3.4) were sufficiently narrow so as to identify different classes of reacted nucleophiles, namely amino groups (peak 2), and histidyl (peak 3) and methionyl (peak 4) side-chains. The linewidths were not, however, sufficiently narrow as in solution to permit the differentiation of derivatized Nα and Nε amino groups. This was in spite of the fact that magic angle spinning was employed. Indeed, the linewidths of the protein lyophilisate were significantly broader than those observed of ideal crystalline materials (typically 15 Hz) and this finding suggested that carbon atoms in the lyophilisate were poorly ordered (Fyfe, 1983b,d). This suggested further that if the experiment was repeated on a more crystalline protein, the linewidths could be sharpened. Before attempts were made to produce protein crystals for methylation at high pH values, a sample of commericially available α-chymotrypsin was analyzed directly. The sample was visibly much more crystalline than the amorphous protein lyophilsate, however, the results (not shown) indicated that the improved crystallinity did not change the linewidths. It was therefore concluded that the apparent broad lines were a composite of many sharper carbon resonances possessing slightly different chemical shift values.

The reacted protein spectrum (Figure 3.3.4, overlaid spectrum) and unreacted protein spectrum (Figure 3.3.4, underlying spectrum) both contained natural abundence carbon resonances. NMR signals of naturally abundant carbons, particularly quaternary carbons, can be enhanced by using cross-polarization techniques (Fyfe, 1983a). This is emphasized by the fact that in normal pulse NMR experiments, natural carbon resonances of protein in solution were not observed. Very briefly, in the cross-polarization experiment, protons are irradiated and allowed to transfer their magnetization to carbons via through-space dipolar coupling. The transfer of spin

magnetization of the abundant proton spin reservoir to the dilute carbon nuclei is carried out coherently, while the Hartman-Hahn condition is satisfied in a spin-lock experiment (Fyfe, 1983a; Garces *et al.*, 1992). Carbon nuclei which are immediately adjacent to protons are amongst those most efficiently coupled, however longer range couplings to quaternary carbons are still adequate to boost signal intensities, provided there is enough contact time. As was shown in the present case, the contact time employed was sufficiently long (2 ms) to make possible the observation of amide carbonyl carbons. This in turn made it possible to superimpose the amide carbonyl carbon peaks (peak 1) and observe the true extent of reaction of histidyl and methionyl groups. The superimposition procedure was necessary for quantification purposes since naturally abundant carbon signals (Figure 3.3.4, underlying spectrum) also appeared in the chemical shift region spanning 20 - 40 ppm, and contributed to the total signal observed. While the appearance of naturally abundant carbon signals may be a nuisance, it illustrates that solid state experiments which use cross-polarization techniques are more quantitative for carbons with long relaxation times and can have certain advantageous features over solution NMR techniques (Fyfe, 1983c).

Methyl groups of solids undergo rapid rotation and do not benefit from as efficient a dipolar coupling. However, since methyl group resonances in proteins are clearly visible following reaction (Figure 3.3.4, overlaid spectrum), the contact time again appears to have been adequate. The fact remains that iodomethane was also observed (peak 5), despite repeated attempts to remove this residual reagent *in vacuo* and the presence of this resonance (peak 5) was rationalized on the basis that residual iodomethane is 100% ^{13}C-enriched and tightly adsorbed to protein.

Liquids, which by definition are in rapid motion, are not observable by cross-polarization techniques (Fyfe, 1983a,c). Such was the observation in these experiments. When octane or

hexadecane were added to protein samples, additional peak resonances were not seen. This was unusual in light of the fact that iodomethane appeared to be tightly associated to the matrix of protein. It was expected that with these higher boiling-point compounds, traces would be even more tightly associated than in the case of iodomethane. However, keeping in mind that iodomethane was ^{13}C-enriched, the failure to observe octane or hexadecane could simply have been a sensitivity problem.

The fact that octane and hexadecane were not observed led us to envisage a solid state, time-course NMR experiment in which protein, and a solution of reagent (in this case, [^{13}C]iodomethane) in octane or hexadecane would be allowed to react in a sealed NMR tube, such as the one illustrated in figure 3.2.2. While the iodomethane resonance would be observed in the spectrum, it would not interfere with the experiment since its chemical shift position is very much upfield of any other resonance (Figure 3.3.4, peak 5). Octane and hexadecane are necessary to reduce the vapour pressure of iodomethane to safer levels, as the volume available for expansion in an NMR tube is very limited in comparison to the volume of a typical reaction vessel. The NMR tube may then be incubated in the probe directly and examined periodically by NMR spectroscopy. In this manner, it should be possible to obtain several time-dependent spectra very conveniently without ever disturbing the sample. In our case, a tube containing iodomethane (15 μl) and hexadecane (15 μl) was placed in an oven at 100°C. The tube did not rupture, thereby supporting the applicability of this technique.

To summarize, two nonaqueous modification procedures were investigated in this chapter. Modification in organic solvent had the advantage that volatile and nonvolatile modifying reagents could be used. With the *in vacuo* procedure, stirring was not required to maintain the protein in a

dispersed state, the reaction temperature was not limited by the boiling point of the solvent, recovery of unreacted reagent was much simpler, and no further manipulation of the modified protein was required. Regardless of which nonaqueous procedure is used, significant advantages were observed over aqueous modifications: (1) Water is not present as a competing reactant so much smaller amounts of reagents are required which can be useful when expensive reagents are used. In the present study, for example, 10 µl of [^{13}C]iodomethane was used in the *in vacuo* reaction at an approximate cost of $1 whereas the equivalent aqueous reaction required 250 µl at a cost of $25. This is a substantial savings and would be even more substantial with costlier reagents such as radiolabeled compounds. Trace labeling studies, in particular, could make use of ^{14}C and ^{3}H-radiolabeled reagents of high specific activity, instead of resorting to other, more dangerous nuclei. (2) Unreacted reagents can be recovered which again can add to the cost-efficiency of the chemical modification. (3) Water insoluble reagents or gaseous reagents (e.g. isobutylene) can be employed at high concentrations or pressures to increase their effectiveness as modifying reagents. (4) Derivatized protein is easily isolatable and derivatizations can be conveniently monitored. (5) Derivatives which are unstable in water, such as activated carboxylic acids, can be prepared as reactive intermediates. (6) Hydrolytic degradation of protein can be prevented.

3.5 Conclusions: Chemical modification of lyophilized proteins in nonaqueous environments was feasible and practical. Reactions could be carried out by either dispersing protein in octane and reacting with dissolved reagent, or reacting protein *in vacuo* with a volatile reagent. Under aqueous and nonaqueous conditions, iodomethane N,N,N-trimethylated Nε-lysyl and Nα-terminal amino groups, N^{1},N^{3}-dimethylated histidyl imidazole functions, O-methylated tyrosyl hydroxyl functions and S-methylated methionyl sulphide groups in protein. The first two *in vitro* chemical

modifications were not reported for native proteins and the second and third modifications were not reported previously as any protein modification. Nonaqueous reactions were achievable with other reagents. Acid anhydrides acylated α and ε-amino groups and formed mixed anhydrides with side-chain carboxyl groups. All derivatives formed can potentially be used for novel structure-function studies. Under nonaqueous conditions, protein hydrolytic breakdown was prevented, smaller amounts of reagent could be used, unreacted reagent could be recovered, and activated but water-labile carboxyl intermediates could be prepared as long-lived intermediates. The extent of derivatization was directly related to the solution pH prior to lyophilization, and the pK_a and surface accessibility of the reacting functional group in the native protein. The physico-chemical factors governing the reactivity of protein functional groups in nonaqueous environments appeared to depend on the protein solution structure prior to lyophilization and opened a door to novel structure-function studies using the nonaqueous approach of protein modification.

3.6 Acknowledgment

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada, and an Ontario Graduate Scholarship to Alpay Taralp. Nicolas A. S. Stewart carried out the activity assays on methylated α-chymotrypsin.

3.7 References

Barman, T. E. (1969) *Enzyme Handbook II*, p. 621, Springer-Verlag.

Bosshard, H. R. (1979) *Methods Biochem. Anal.* **25**, 273-201.

Breitmaier, E. and Voelter, W. (1987a) *Carbon-13 NMR Spectroscopy: High-Resolution Methods and Applications in Organic Chemistry and Biochemistry*, 3rd ed., p. 109, VCH Publishers.

Breitmaier, E. and Voelter, W. (1987b) *Carbon-13 NMR Spectroscopy: High-Resolution Methods and Applications in Organic Chemistry and Biochemistry*, 3rd ed., p.234, VCH Publishers.

Broos, J., Visser, A. J. W. G., Engbersen, J. F. J., Verboom, W., van Hoek, A. and Reinhoudt, D. N. (1995) *J. Am. Chem. Soc.* **117**, 12657-12663.

Careri, G., Gratton, E., Yang. P. H. and Rupley, J. A. (1980) *Nature* **284**, 572-574.

Chen, C.-S. and Sih, C. J. (1989) *Agnew. Chem. Int. Ed. Engl.* **28**, 695-707.

Constantino, H. R., Griebenov, K., Mishra, P., Langer, R. and Klibanov, A. M. (1995) *Biochim. Biophys. Acta.* **1253**, 69.

Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, p. 6, W. H. Freeman.

Creighton, T.E. (1993) *Proteins:Structures and Molecular Properties*, pp.419-421,W.H. Freeman.

Darbre, A. (1986a) In *Practical Protein Chemistry* (Darbre, A. Ed.) p. 141, John Wiley and Sons.

Darbre, A.(1986b) In *Practical Protein Chemistry* (Darbre, A Ed.) pp.260-262, John Wiley and Sons.

Desai, U. R., Osterhout, J. J. and Klibanov, A. M. (1994) *J. Am. Chem. Soc.* **116**, 9420-9422.

Fitzpatrick, P. A., Ringe, D. and Klibanov, A. M. (1994) *Biochem. Biophys. Res. Commun.* **198**, 675-681.

Fitzpatrick, P. A., Steinmetz, A. C. U., Ringe, D. and Klibanov, A. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8653-8657.

Fyfe, C. A. (!983a) *Solid State NMR for Chemists*, pp. 183-187, C.F.C. Press.

Fyfe, C. A. (!983b) *Solid State NMR for Chemists*, pp. 140-146, C.F.C. Press.

Fyfe, C. A. (!983c) *Solid State NMR for Chemists*, pp. 282-289, C.F.C. Press.

Fyfe, C. A. (!983c) *Solid State NMR for Chemists*, p. 8, C.F.C. Press.

Garces, F. O., Rao, V. P., Garcia-Garibay, M. A. and Turro, N. J. (1992) *Supramolecular chemistry* **1**, 65-72.

Glazer, A. N., Delange, R. J. and Sigman, D. S. (1976) *Chemical Modification of Proteins*; In *Laboratory Techniques in Biochemistry and Molecular Biology* (Work, T.S. and Work, E. Eds.) **4**, pp. 3-205, North-Holland Publishing Company.

Griebenov, K. and Klibanov, A. M. (1997) *Biotechnol. Bioeng.* **53**, 351.

Imoto, T. and Yamada, H. (1989) *Chemical Modification*: In *Protein Function: A Practical Approach* (Creighton, T. E. Ed.) Chapter 10, pp. 247-277, IRL Press.

Kaplan, H. (1972) *J. Mol. Biol.* **72**, 153-162.

Kaplan, H., Stevenson, K. J., Hartley, B. S. (1971) *Biochem. J.* **124**, 289-299.

Klibanov, A. M. (1984) *Chemtech* **16**, 354-369.

Klibanov, A. M. (1989) *Trends Biochem. Sci.* **14**, 141-144.

Leggett Bailey, J. (1962) *Techniques in Protein Chemistry*, pp. 20-30, Elsevier Publishing Co.

Lundblad, R. L. (1995) *Techniques in Protein Modification*, CRC Press.

Lunblad, R. L. and Noyes C. M. (1984) *Chemical Reagents for Protein Modification*, 1 and 2, CRC Press.

Means, G. E. (1977) In *Meth. Enzymol.* (Hirs, C. H. W. and Timasheff, S. Eds.) **47**, 469-475.

Means, G. E. and Feeney, R. E. (1971) *Chemical Modificaton of Proteins*, Holden-Day.

Mishra, P., Griebenov, K and Klibanov, A. M. (1996) *Biotechnol. Bioeng.* **52**, 609.

Paik, W. K. and Kim, S. (1975) *Adv. Enzymol.* **42**, 227-286.

Poole, P. L. and Finney, J. L. (1983) *Biopolymers* **22**, 255-260.

Prestrelski, S. J., Arakawa, T. and Carpenter, J. F. (1993) *Arch. Biochem. Biophys.* **303**, 465-472.

Rupley, J. A. and Careri, G. (1991) *Adv. Protein Chem.* **41**, 37-52.

Schinkel, J. E., Downer, N. W. and Rupley, J. A. (1985) *Biochemistry* **24**, 352-357.

Scopes, R.K.(1987) *Protein Purification:Principles & Practice*, 2nd ed.,pp.250-252, Springer Verlag.

Volkin, D. B. and Klibanov, A. M. (1989) *Minimizing Protein Inactivation*: In *Protein Function: A Practical Approach* (Creighton, T. E. Ed.) Chapter 1, pp. 2-6, IRL Press.

Walsh, C. (1979) *Enzyme Reaction Mechanisms* (McCombs, L. W. and Bartlett, A. C. Eds.) pp. 74-78, W. H. Freeman and Company.

Wescott, C. R. and Klibanov, A. M. (1994) *Biochim. Biophys. Acta* **1206**, 1-9.

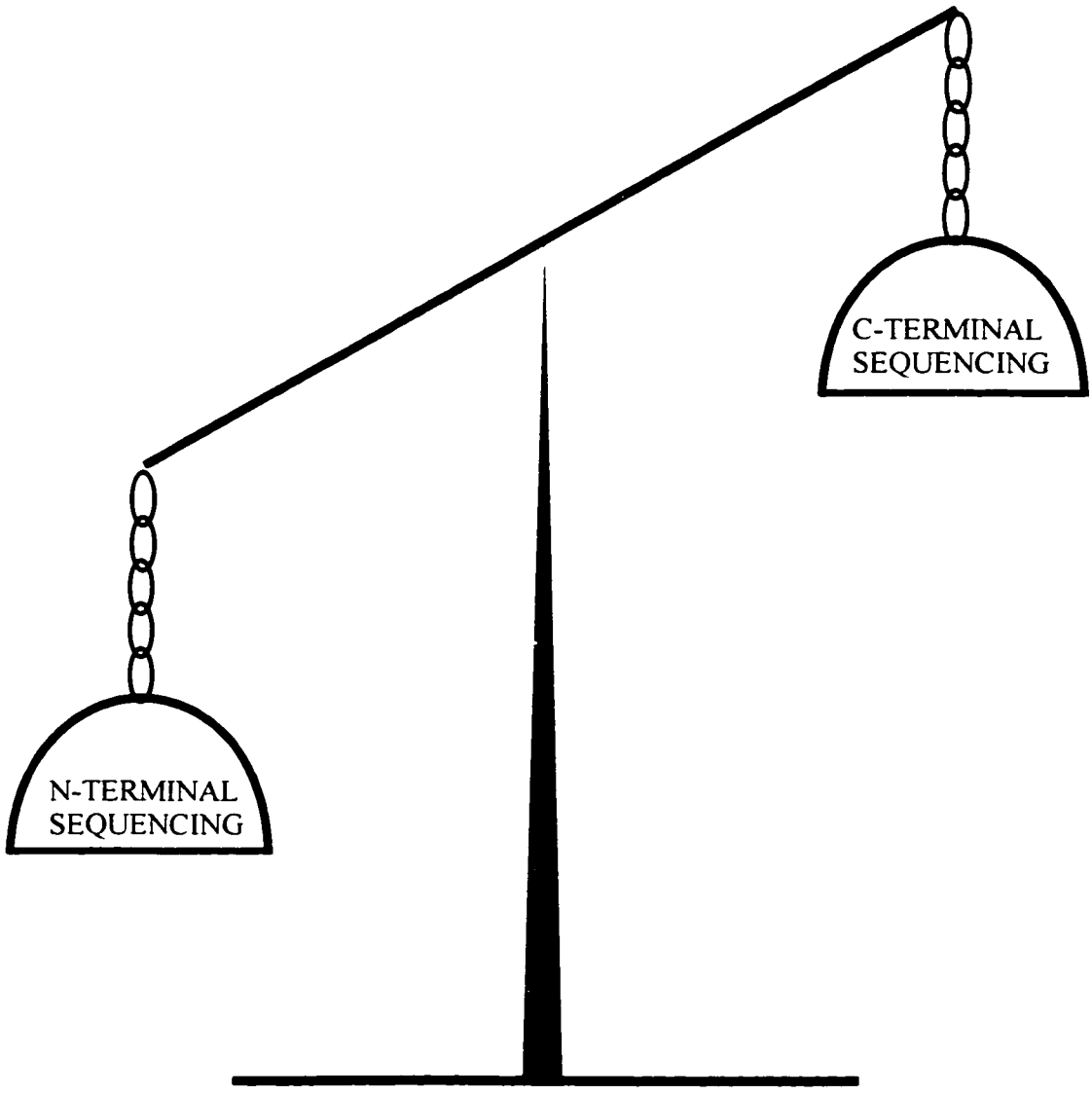
Yan, B., DeCaprio, A.P., Zhu, M. and Bank, S.(1996) *Chemico-Biological Interactions* **102**, 101-106.

Young, N .M. and Kaplan, H. (1989) *Chemical Characterization of Functional Groups in Proteins by Competitive Labelling*: In *Protein Function: A Practical Approach* (Creighton, T. E. Ed.) Chapter 8, IRL Press.

Zaks, A. and Klibanov, A. M. (1988) *J. Biol. Chem.* **263**, 3194-3201.

Zhang, M. and Vogel, H. J. (1993) *J. Biol. Chem.* **268**, 22420-22428.

Zhang, M., Thulin, E. and Vogel, H. J. (1994) *J. Protein Chem.* **13**, 527-535.


C-TERMINAL
SEQUENCING
N-TERMINAL
SEQUENCING

Chapter 4: Microscale nonaqueous chemistry for the analysis of protein C-terminal sequences

4.1 Introduction

4.1.1 Overview of C-terminal chemistry

Successfully identifying the C-terminal amino acid is an important criterion of protein characterization and continues to be investigated by protein chemists, using a variety of modern techniques. Protein primary structure is currently most often elucidated by inferring the amino acid sequence from the corresponding gene nucleotide sequence. While this approach is much simpler than direct protein sequencing, a major drawback is that it cannot account for changes in primary structure that may arise due to post-translational processing events which precede the formation of mature protein. Preproinsulin, for example, is enzymatically processed into a two-chain peptide of approximately half the molecular weight. Reliable, quick and direct amino acid sequence information is therefore often desired for (i) confirmation of protein primary structure, (ii) characterization of *in vivo* or post-translational chemical modifications such as acetylation, methylation, phosphorylation, sulphonation and glycosylation, (iii) characterization of proteolytic processing, and (iv) protein identification. Methods of sequence analysis use N-terminal Dansyl-Edman or phenylthiohydantoin chemistry, N and C-terminal enzymatic methods, various C-terminal chemical degradation methods and several mass spectral methods (Ambler, 1972; Callahan *et al.*, 1972; Gray, 1972; Konigsberg, 1972; Light, 1972; Richards and Lovins, 1972; Schroeder, 1972; Stark, 1972; Stryer, 1981; Winter, 1986a; Winter, 1986b; Waterfield *et al.*, 1986; Biemann and Scoble, 1987; Olivares *et al.*, 1987; Gibson *et al.* 1988; Biemann, 1992; Tsuji *et al.*, 1992; Stults *et al.*, 1993; Siudak, 1994; Caprioli and Suter, 1995). While N-terminal sequencing methods are far more successful than C-terminal methods, each has specific limitations

inherent to the procedure, making it necessary to use several strategies in order to obtain the complete sequence of a protein. The C-terminus remains a region that is often not analyzed because of the poor success of chemical and enzymatic methods, but the recent development of procedures using carboxypeptidase Y degradation in tandem with matrix-assisted laser desorption ionization mass spectrometry has narrowed this impasse (Patterson *et al.*, 1995). Among the chemical methods, extensive work was carried out using hydrazinolysis (Akabori *et al.*, 1952), reduction (Fromageot *et al.*, 1950), and tritiation (Ward, 1986).

A general chemical method for the determination of protein C-terminal sequences was developed by Duggleby and Kaplan (1975). Carboxyl groups of protein were labeled in aqueous media. Following incubation with proteinases, the C-terminal fragments could be isolated using a two-dimensional electrophoretic approach. This method has potential applications for the study of proteins in general, and practical uses in assessing protein C-terminal homogeneity. In fact, this procedure has all the desirable features to be the method of chioce, except that it lacks the sensitivity required to analyze nanomolar quantities of protein.

4.1.2 Aqueous C-terminal chemistry

The C-terminal problem was readdressed with variations of the diagonal electrophoresis approach of Duggleby and Kaplan (1975). Several aqueous methods were attempted, however, each had shortcomings. Of the variations attempted by Taralp and Kaplan, the most successful ones adopted a coupling of protein carboxyl groups to solubilizing molecules such as ethanolamine and ethylenediamine. Special mention could be made of the fluorescent tag N-1-[5-dimethylamino-1-naphthalenesulphonyl]-2-hydroxy-1,3-diaminopropane, which was easily

synthesized and imparted very desirable properties to the derivatized protein - namely, solubility, especially at low pH values, and a facile detection of the C-terminus by ultraviolet fluorescence. The potential sensitivity of fluorescence detection is illustrated in figure 4.1.1, which depicts 1 mg of protein after derivatization and extensive dialysis. Dialysis following derivatization is an integral part of these procedures since it is necessary to remove urea, excess amine, and water soluble carbodiimide, as well as other reaction byproducts (Means and Feeney, 1971; Duggelby and Kaplan, 1975; Darbre, 1986). Failure to do so leads to problems at the separation step, where paper electrophoresis is required. Even small scale reactions contain enough charge carriers to cause irratic electrophoretic separations and only with the smallest scale reactions can the need for dialysis be overlooked. Dialysis is also potentially detrimental in microscale reactions because the adsorbtion of protein to membrane may lead to unacceptable losses. While other strategies have been attempted to prevent protein loss, namely addition of polyethylene glycol or succinnylated protein as carrier, the implementation of these steps are not desirable since they can potentially lead to additional purification problems. Even in the best case scenario, the approach would still lose its elegance if it relied on these additional steps.

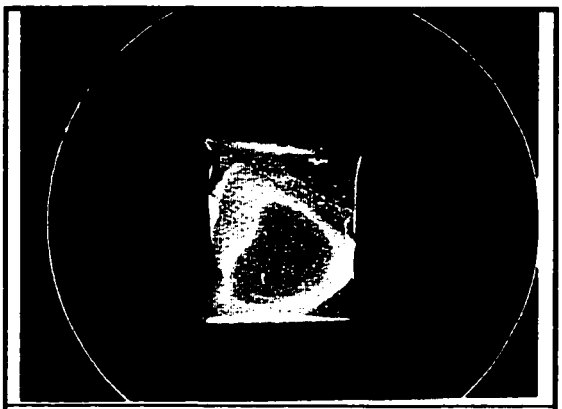


Figure 4.1.1: Fluorescence incorporated into the carboxyl groups of α-chymotrypsin (1 mg) following derivatization with the amine N-1-[5-dimethylamino-1-naphthalenesulphonyl]-2-hydroxy-1,3-diaminopropane. Protein was dialyzed extensively at room temperature against 100 mM NaCl & 10 mM HCl, and finally against HCl.

The potential success of aqueous microscale derivatizations for the separation of C-termini was hampered by low yields of carboxyl group derivatizations. In figure 4.1.2, α-chymotrypsin

(650μg) is derivatized with fluorescent amine. Following dialysis and incubation with trypsin and carboxypeptidase B (Darbre, 1986), the C-terminal peptides were separated by paper electrophoresis at pH 6.5. α-Chymotrypsin has three separate C-termini (Duggleby and Kaplan, 1975), and therefore at least three basic peptide bands were expected. In practice, for a protein of this complexity, approximately ten bands could have been expected. However, only two peptide bands (Figure 4.1.2; bands 2, 3) were seen, presumably due to incomplete coupling of fluorescent amine to the C-termini.

The poor efficiency of coupling was also corroborated by attempts to couple ethanolamine to activated carboxyl groups of protein (not shown) using reported methods (Means and Feeney, 1971; Darbre, 1986). Following derivatization and digestion, no acidic peptides should have been observed by paper electrophoresis at pH 6.5. However, the ninhydrin colour yield (Darbre, 1986) of peptides migrating to the anode was substantial and estimated at 40% of the control protein.

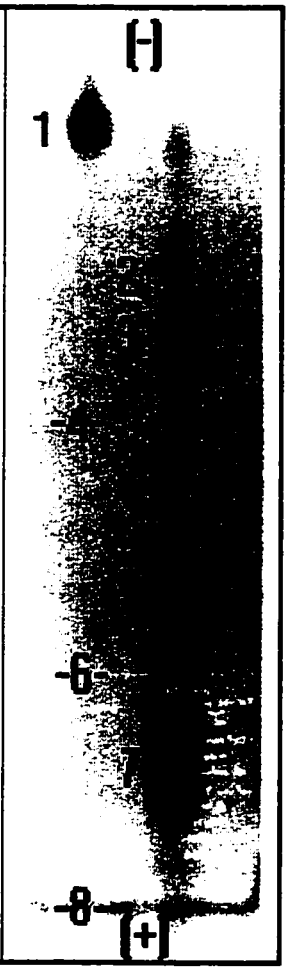


Figure 4.1.2: Fluorescently labeled α-chymotrypsin peptides separated by high voltage paper electrophoresis at pH 6.5, following incubation of protein with trypsin and carboxypeptidase B. The marked positions under ultraviolet light (302 nm) are: (1) N-1-[5-dimethyl-amino-1-naphthalenesul-phonyl]-2-hydroxy-1,3-diamino-propane; (2, 3) two distinct C-terminal peptides; (4) dansylsul-phonic acid starting line; (5) neutral band containg most fluorescent peptides; (6) origin; (7) fluorescent peptides with free aspartic and glutamic acid residues; (8) dansylsulphonic acid.

Another drawback of the aqueous approach was that the treatment of protein for C-terminal analysis became very different at the microscale (i.e. for < 50 μl total reaction volume). Technical problems were an issue.

Stirring was not possible and samples were agitated instead. The pH of reaction (pH 4.75) could no longer be maintained with a pH-stat and pyridine hydrochloride buffering was implemented, despite the fact that pyridine is known to catalyze the hydrolysis of activated acyl intermediates (Winter, 1986b). Some investigators have avoided this problem by trapping the activated carboxyl intermediate with N-hydroxysuccinimide (1 equivalent) followed by a secondary coupling with amine (Bodansky and Bodansky, 1994; Horn *et al.*, 1995). As mentioned previously, problems of dialysis is also a matter to consider.

A last problem which was discouraging was the need to develop a reliable detection and sequencing method that would not be limited by trace levels of contaminants and amine. It was observed, for example, that after paper electrophoresis, residual amine produced a fluorescent smear which could potentially frustrate detection and sequencing attempts. Even after dialysis, free amines were often observable. The example of figure 4.1.2 was perhaps one of the exceptional cases, where amine was removed effectively by dialysis and the fluorescent background was not prevelant. Even in this example, however, it is easy to see that residual amine could have been problematic if there had been a ten or one hundred fold decrease in the amount of protein present.

The factors discussed above give the impression that even if an aqueous C-terminal method should be developed for small amounts of protein, the approach would be limited and perhaps not very elegant to carry out. It was desired, therefore, to develop a technique for derivatizing and sequencing small amounts of protein that would be tolerant of impurities and could circumvent these limitations.

4.1.3 Focus of the present work

With the development of nonaqueous chemical modification procedures in our laboratory, (Kaplan and Taralp, 1997; Taralp and Kaplan, 1997) attention was shifted to the employment of nonaqueous media to address the C-terminal problem. With little direct experience to bias our judgement, we sought a nonaqueous procedure which would address the C-terminal problem and be specifically designed for use with minute quantities of protein. The derivatization protocols were chosen to limit protein loss, obtain high yields of coupling, prevent or limit the accumulation of salt or charge carriers while carrying out the procedure, and provide extreme sensitivity. It is hoped that the procedure will prove simple to carry out and be cost effective and pertinent for use with minute quantities of protein. The protease α-chymotrypsin served as the model protein because it has three C-termini which is ideal for testing the generality of the protocol (Duggleby and Kaplan, 1975).

Rationale

In this chemical approach, nonaqueous derivatization conditions are used to isolate and sequence the C-terminal amino acids of proteins. Following a sequential series of derivatizations with volatile reagents (Vath *et al.*, 1988; Knapp, 1990) and incubation with proteinase, peptide derivatives are separated by two-dimensional high voltage paper electrophoresis (Leggett Bailey, 1962; Duggleby and Kaplan, 1975), purified by paper chromatography, detected by autoradiography and sequenced by tandem mass spectroscopy. Unnecessary transfer of protein is avoided in order to limit losses. Reagents are removed from reaction by evaporation, rather than by other means. Manipulations such as dialysis and gel permeation chromatography are avoided.

Using volatile nonionized reagents places a restriction on the kinds of reagents which are available, but the scheme is nevertheless attractive since the reagents would preclude any need for dialysis. It should also be pointed out that the chemistry employed is designed to make protein derivatives particularly amenable for analysis by sensitive microspray mass spectral methods.

Figures 4.1.3 and 4.1.4 depict five variations (1 & 2 and 3, 4 & 5, respectively) of this theme. Derivatization steps I and II (Figure 4.1.3) are common to experimental variations 1 through 5. Variation 5 is identical to variation 4 but terminates following incubation with thermolysin.

Derivatization steps:

Lyophilization: step I. Protein is lyophilized out of aqueous solution in preparation for reaction in organic media. Since the solubility of protein appears to increase with the lyophilization volume, protein is lyophilized in the largest volume practical for a given amount of protein. The lyophilized protein (1) is expected to have substantial solubility in methanol (Means and Feeney, 1971).

Carboxyl group activation: step II. Protein is esterified in anhydrous methanol containing a small amount of hydrogen chloride. Methyl ester formation of protein (2) is usually quantitative after 24 h under the employed reaction conditions (Chibnall *et al.*, 1958; Means and Feeney, 1971). To convince ourselves, a time course esterification of model peptides was carried out using the same acid concentration. Peptide esters were subsequently amidated with diamines. The amidated peptides were analyzed by high voltage paper electrophoresis. It was deduced that phenylalanyl-glycyl-glycine must have been substantially esterified after only 30 min, and completely esterified after 9 h. Esterification did not proceed in the absence of acid.

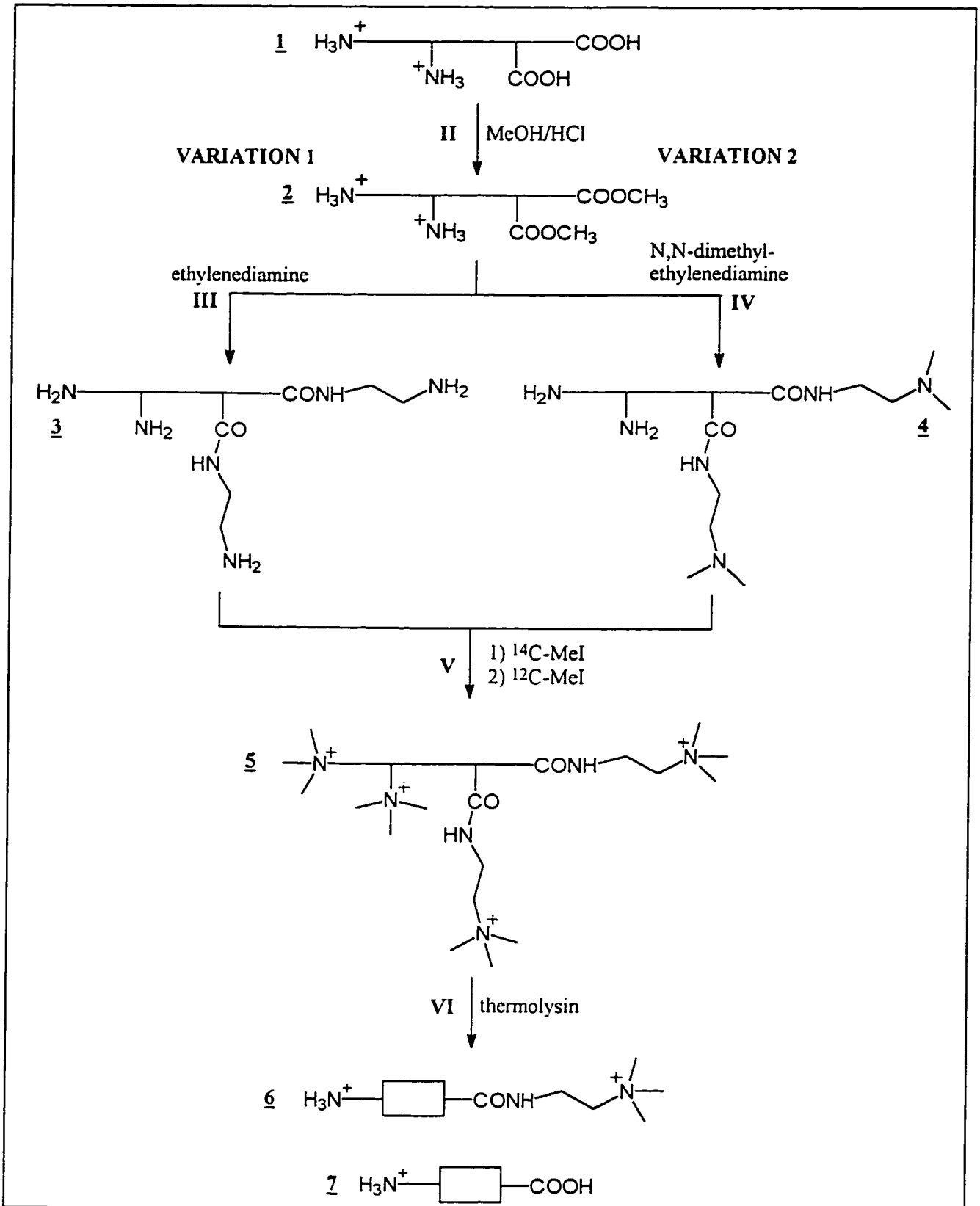


Figure 4.1.3: Nonaqueous C-terminal reaction scheme. Variations 1 and 2 are illustrated.

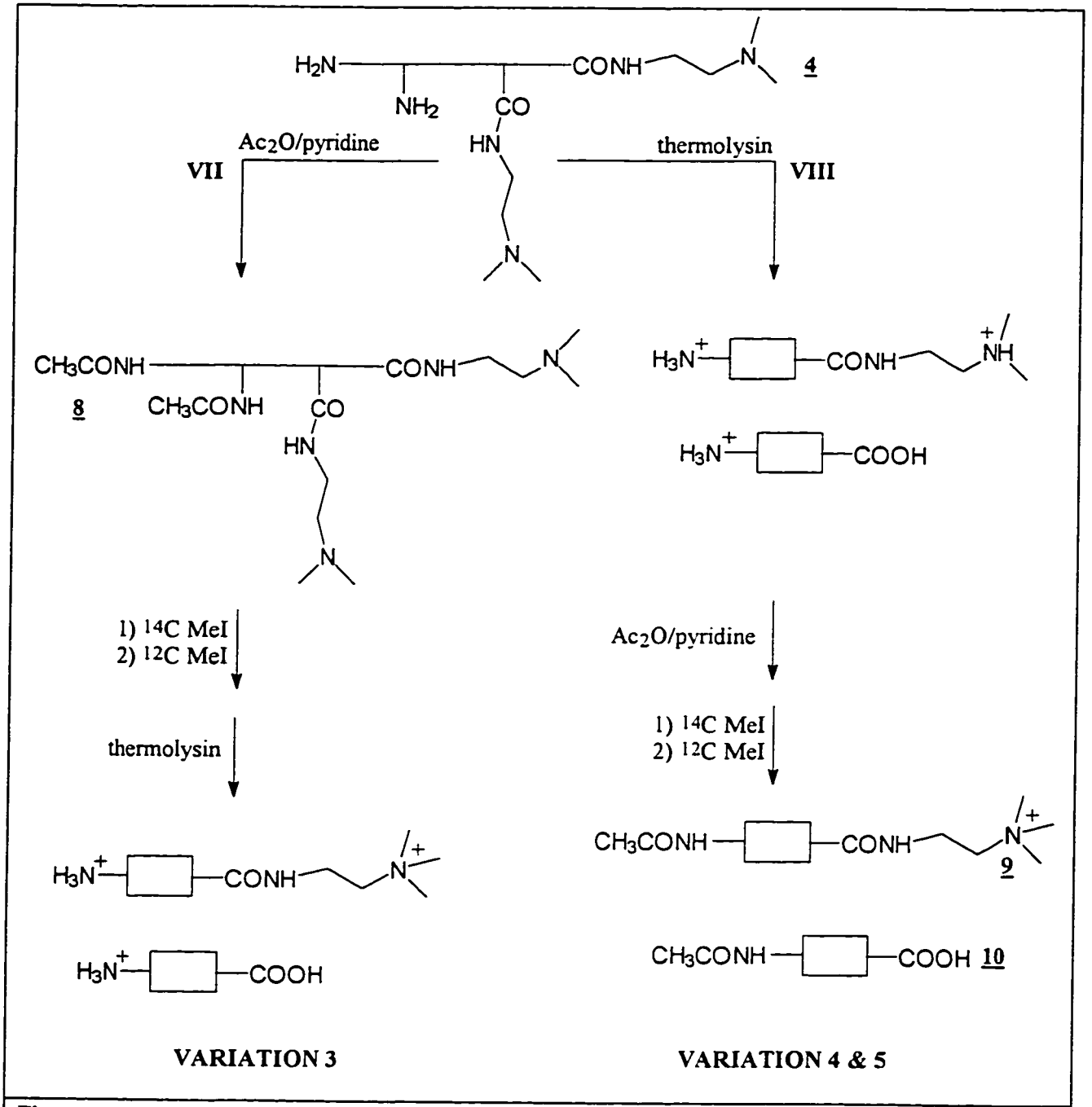


Figure 4.1.4: Nonaqueous C-terminal reaction scheme (continued). Variations 3, 4 and 5 are illustrated.

Esterification of protein was reported to occur with minor covalent modifications, one being the methanolysis of glutamine and asparagine side-chains and the other being the N→O acyl migration of serine or threonine residues of protein (Fraenkel-Conrat and Olcott, 1947; Phillips and Baltzly, 1947; Frieden, 1956; Chibnall *et al.*, 1958; Lenard and Hess, 1964; Iwai and Ando, 1967; Means and Feeney, 1971; Darbre, 1986). The methanolysis of glutamine and asparagine in acidic methanol is not expected to be problematic. In the N→O acyl migration reaction, however, the Nα-amide bond of serine or threonine residues of protein may be broken by nucleophilic attack of the serine or threonine hydroxyl group. The result is to transform the amide bond into an ester bond. Ester linkages of protein have been cleaved by the action of piperidine (Darbre, 1986) and in the context of this investigation, cleavage by diamine (steps III and IV, *vide infra*) would generate false C-termini and must be avoided. With due consideration to steps III and IV, the esterified protein is incubated (optional) under basic conditions, namely, in anhydrous pyridine, to promote an O→N reverse migration (Chibnall *et al.*, 1958; Means and Feeney, 1971; Darbre, 1986) which should restore the intergrity of the primary structure.

Coupling of diamine to protein carboxyl groups via amide bond formation: steps III or IV.

Normally, unactivated methyl ester groups under homogeneous reaction conditions require high concentrations of amine to effect displacement (March, 1985; Bodanszky and Bodanszky, 1994). It was observed that phenylalanyl-glycyl-glycyl-methyl ester was quantitatively amidated at room temperature after 10 h, when dissolved in N,N-dimethylethylenediamine or ethylenediamine. Amidation did not appear to damage the peptide bonds. In the context of

protein reactions, the amidation is carried out in neat ethylene diamine or N,N-dimethylethylenediamine to afford derivatives 3 and 4, respectively. The reaction is at least partially homogeneous since it was observed that both diamines dissolve many proteins in a relatively small volume. The ability of diamines to dissolve protein so effectively minimizes interprotein contact and crosslinking, and is one key point of the experimental strategy.

Variations 1 and 2; protein methylation: step V. The derivatized protein (3, 4) is trace labeled *in vacuo* with [^{14}C]iodomethane, followed by excess [^{12}C]iodomethane, to afford a chemically homogeneous protein derivative (5) in which amino, imidazole, thiol, sulphide and possibly phenolic hydroxyl groups are methylated to completion. The radiolabel need not be diluted with non-radioactive label as would be the case of aqueous reactions, so the specific radioactivity incorporated is expected to be very high (Kaplan and Taralp, 1997; Taralp and Kaplan, 1997).

Thermolysin digestion: step VI. Thermolysin is one of the few proteinases which cleave substrates N-terminal to the recognition site (Matsubara and Feder, 1971; Darbre, 1986). This is desirable since it precludes the accidental removal of the radioactive label at the C-terminus. The enzyme is also extremely stable and is active for extended periods of time. The digestion is carried out to produce two classes of peptides, 6 and 7. The C-terminal peptides (6) and the non C-terminal peptides (7) are then ready for separation by paper electrophoresis according to published procedures (Leggett Bailey, 1962; Duggleby and Kaplan, 1975).

Variations 3, 4 and 5: steps VII and VIII. It was observed that protein chains of bovine serum albumin are semi-specifically cleaved *in vacuo* at elevated temperatures by the action of ethanolamine or ethylenediamine. The occurrence is presumably a consequence of transamidation, in which the simple amine replaces the amino group moiety of the amide bond which is cleaved. In this context, variations 3, 4 and 5 are subsequent refinements which were introduced to address the unfortunate observation that protein amide bonds of esterified α-chymotrypsin are cleaved during the methylation step of variations 1 and 2, by the action of residual diamine.

The rationale of variations 3 and 4 is to eliminate the generation of false C-termini by preventing transamidation. The obvious route, short of removing all residual diamine, is to block the amino groups with another volatile reagent before methylation is carried out *in vacuo* at elevated temperature (variation 3). To achieve this, protein derivative (4) is acetylated in pyridine (step VII) and affords the desired protein derivative (8). Unfortunately, variation 3 (Figure 4.1.4) is not the answer because unblocked amine was observed after two-dimensional chromatography was performed. Either some N,N-dimethylethylenediamine is tightly adsorbed within the protein core and is not accessible for acetylation, or perhaps, intramolecular participation of the N,N-dimethylamino moiety and trace moisture is enough to unblock the adjacent N'-acetyl moiety at elevated temperature. To alleviate any potential problems that may arise from either possibility, acetylation is carried out only after digestion (Variation 4, step VIII) and methylation is carried out at room temperature. The rationale is that all N,N-dimethylethylenediamine should be readily accessible to acetic anhydride since peptides, unlike protein, should not harbour any diamine. The methylation reaction is carried out at room temperature to minimize the risk of unblocking the

acetylated diamine, and in N,N-dimethylformamide to promote reaction with iodomethane (March, 1985). One drawback of adopting variation 4 is that mass spectrometry becomes the only sequencing method available, since C-terminal peptides (9) are blocked at the N-terminus.

In variation 5, incubation with thermolysin and diagonal electrophoresis is carried out immediately after coupling dimethylethylenediamine to protein. This is by far the simplest derivatization procedure and can be carried out in one Eppendorf tube. The disadvantage of variation 5 is that peptides are not radioactive and will not expose X-ray film. The diagonal can be eluted using markers, but must be analyzed using a tandem mass spectrometer in series with a high performance liquid chromatography delivery system. This sophistication may not be readily available to researchers.

Diagonal paper electrophoresis for detection of C-terminal peptides. The protocol thusfar affords C-terminal peptides, which are amidated at the α-carboxyl group, and non C-terminal peptides, which have free α-carboxyl groups. Over the pH range of 2.1-4.4, C-terminal peptides have a constant charge, while non C-terminal peptides change their charge. Radioactive markers with either neutral or constant positive charge are added to the peptides. The peptides are then subjected to paper electrophoresis at pH 2.1, the paper is rotated 90°, and the peptides are resubjected to electrophoresis at pH 4.4. The peptides may be located by autoradiography (Duggleby and Kaplan, 1975). The constant charge marker, the neutral spot marker and the C-terminal peptides should describe a diagonal line (Figure 4.1.3). Non C-terminal peptides do not lie on this diagonal line. The diagonal line can be stitched onto another paper and the peptides purified by chromatography if necessary (Leggett Bailey, 1962).

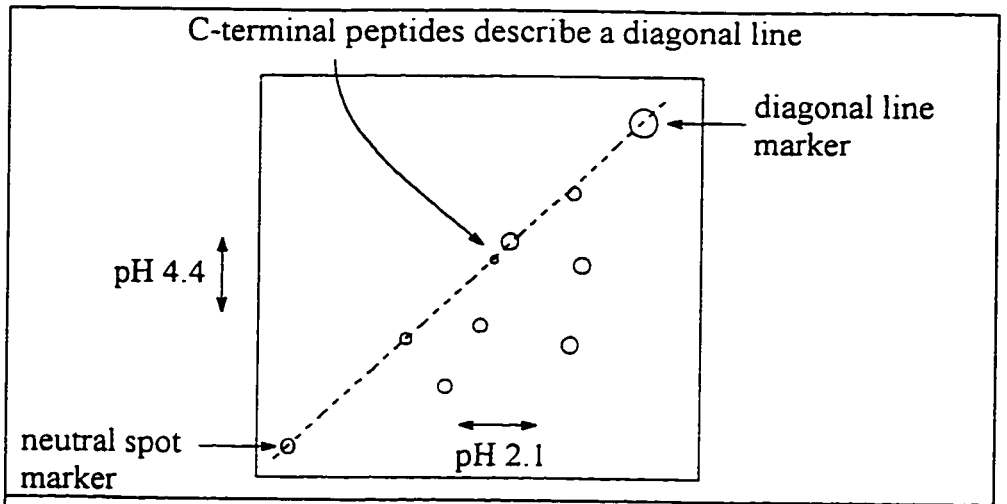


Figure 4.1.3: Two-dimensional high voltage paper electrophoresis for the isolation of C-terminal peptides.

Peptide detection using microspray volatilization and tandem mass spectral methods. C-Terminal peptide derivatives, much like betaine, contain a positively charged group, namely the trimethylammonium moiety. The liberation of trimethylamine from quaternary ammonium salts is reported to occur under basic aqueous conditions (March, 1986). Provided that trimethylamine is liberated under the conditions of mass analysis, C-terminal peptide derivatives are candidates for sensitive and selective detection by tandem microspray mass spectrometry (Biemann and Scoble, 1987; McCloskey, 1990; Anderegg *et al.*, 1988). The method can be extended to the analysis of peptides which liberate dimethylamine, as would be the case in variation 5. It is even conceivable that with further developments, microspray ionization at pH values >10 can take advantage of the trimethylammonium moiety to selectively analyze C-terminal peptides, since only these peptides would retain a net positive charge that would enable their selective ionization. In principle, this would even preclude the need for diagonal electrophoresis. Three modes of tandem (MS/MS) detection of C-terminal peptides are possible.

Theory of tandem microspray methods. Microspray mass spectrometry is in itself useful for the detection of minute amounts of peptides. It is at least an order of magnitude more sensitive than established electrospray methods (Bateman *et al.*, 1997; Kelly *et al.*, 1997). Microspray analysis is usually configured to detect positively charged residues but is amenable to negatively charged ions as well (Capriolli *et al.*, 1995). Tandem mass spectrometry methods have been extremely useful in the sequencing of peptides (Biemann and Scoble, 1987; McCloskey, 1990; Biemann, 1992). With the tandem approach, a greatly simplified spectrum can be obtained which is not as sensitive to the presence of contaminants at low analyte concentrations. The technique uses two stages of mass analysis, one to preselect an ion and the other to analyze fragments generated by collision-induced dissociation. With this technique, it is possible to select and analyze a specific mass ion species out of a complex mixture of ions. One such instrument, the triple quadrupole mass spectrometer, has been gaining favour due to its low cost, instrumental simplicity, and ease of computer control.

The heart of the instrument is composed of two quadrupole mass filters (Q_1 and Q_2) separated by an rf-only quadrupole (q), often termed the collision chamber (McCloskey, 1990; Capriolli, 1995). The quadrupole mass filters are composed of two pairs of parallel rods arranged so that the cross section roughly describes a square with concave sides. One pair of rods is fed a dc and rf voltage. The other pair is fed the opposite potential. The voltage applied to opposite pairs of rods is expressed by $U + V\cos\omega t$, where U is the dc voltage and $V\cos\omega t$ is the time-dependent rf voltage. At given values of U,V and

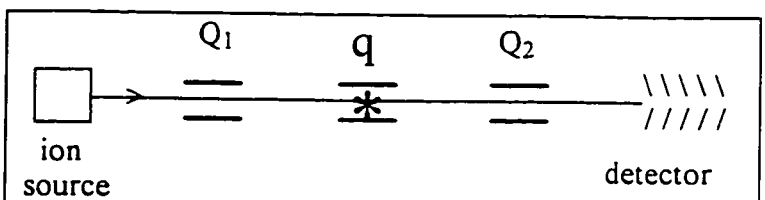


Figure 4.1.4. Triple quadrupole tandem mass spectrometer configuration.

to, only specific ions will have stable trajectories through the quadrupole. The mass/charge ratio of any ion which passes to the detector can be calculated, since there is a relationship between the applied potential and the ion mass/charge ratio. The remaining ions oscillate perpendicular to the flight path with ever increasing amplitude until finally colliding with one of the four rods.

In the collision chamber, incoming precursor ions are fragmented by inert gas. While fragmentation is desired, these ions are also scattered by the collision and must be redirected if they are ever to reach the detector. To this end, the rf-only source of the collision chamber refocuses the ions and promotes their efficient transfer to the third quadrupole. An off-axis ion detector follows Q_2 and transduces the arrival of an ion into a potential signal. The detector only functions to quantify the intensity of ions which pass Q_2, but it cannot assign mass/charge ratios.

MS/MS detection modes. There are three tandem modes of operation, depending on whether the mass/charge ratio of the precursor ion, product ion, or neutral fragment is fixed (McCloskey, 1990; Capriolli, 1995).

Mode 1. If the precursor ion filter (Q_1) is set to allow one precursor (parent) ion into the collision chamber, only fragments of that precursor ion are formed at q and are sequentially passed through to the product ion filter (Q_2). By sweeping the potential at Q_2, every fragment of varying mass/charge ratio which was formed at q can be made to sequentially pass through Q_2 to the detector. The mass/charge ratio of fragments are quantified at Q_2 and produce a product ion (daughter-ion) spectrum of the mass-selected precursor ion. This configuration of the instrument is the most common and is typically used to provide structural information of a specific ion, such as the primary sequence of a peptide ion. The method can be used in series with liquid

chromatography methods for the separation and analysis of complex mixtures (Lee *et al.*, 1989; Treston *et al.*, 1989).

Mode 2. Identifying a group of compounds who share a common moiety in a mixture of unrelated compounds may be a complicated problem, however, advantage can be taken of MS-MS techniques to greatly simplify the situation (McCloskey, 1990). If the desired analytes liberate a common fragment ion at q of known mass/charge ratio, Q_2 can be set to only allow that ion to pass. The parameters affecting the potential of Q_1 are varied instead. When the detector observes the fragment ion passing through Q_2, it is possible to calculate the mass/charge ratio of the corresponding parent ion which passed through Q_1, since the voltage parameters are known. In this configuration, a precursor ion (parent-ion) spectrum is obtained which only shows ions that liberated a certain fragment.

Microspray methods use low energy ionization techniques, and ions produced at the source are generally limited to multiply charged species of the parent ion. Mode 2 is normally used to determine the mass/charge ratio of a parent ion which liberates a certain moiety at q. Once this parent ion is known, it may be sequenced by mode 1. Peptides do occasionally fragment, however, before entering Q_1. This is a consequence of the internal energy acquired in the ionization process. In such a case, the peptide and its fragments enter Q_1, and the analysis will produce a profile analogous to a mixture of related compounds who share a common moiety. In the event that some initial fragmentation does occur, mode 2 can provide some sequence information.

Mode 3. In a last MS-MS configuration, Q_1 and Q_2 voltages can be scanned synchronously to allow for the identification of neutral fragments of a fixed mass. For example, if a compound

or group of compounds is known to release a neutral fragment at q, the mass filters Q_1 and Q_2 can be sweeped synchronously such that the mass/charge ratio of Q_2 is always less than Q_1 by an amount which corresponds to the mass of the neutral species. A constant neutral loss spectrum can be constructed from mass/charge data from either Q_1 or Q_2.

4.2 Materials and methods

Proteins. Thermolysin from *Bacillus thermoproteolyticus* and bovine pancreatic α-chymotrypsin were purchased from Sigma Chemical Company.

Chemicals and Solvents. Leucyl-leucyl-leucyl-tyrosyl-methyl ester was purchased from Bachem. [^{12}C]Iodomethane and di-*tert*-butyldicarbonate were purchased from Sigma Chemical Company. [^{14}C]iodomethane (1 mCi, 56mCi/mmole) was obtained from Amersham Canada Ltd. Anhydrous methanol, anhydrous N,N-dimethylformamide, 4M hydrogen chloride in 1,4-dioxane, acetic anhydride, ethylenediamine, N,N-dimethylethylenediamine and pyridine were obtained from Aldrich Chemical Company. Molecular sieves (4Å) were obtained from BDH and activated 5 d in an oven at 160°C. Pyridine was distilled once over ninhydrin and stored, and redistilled and dried over molecular sieves before use. Ethylenediamine and N,N-dimethylethylenediamine were distilled before use. All other chemicals, reagents and solvents were high purity preparations obtained from commercial sources.

Gel electrophoresis. Sodium dodecyl sulphate-polyacrylamide gels (15% crosslinking) were run on a Biorad electrophoresis apparatus. Gels were stained with Coomassie blue dye. Samples were prepared in buffer of the composition water/10%sodium dodecyl sulphate/glycerol

/0.5%bromophenol blue/6N HCl/urea (6.5:2:1:0.5:0.09:10 by weight). Samples in buffer were not heated before analysis.

Model peptides analyzed by conventional (MS) and tandem (MS/MS) methods were prepared as follows:

Leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$. Leucyl-leucyl-leucyl-tyrosyl-methyl ester (10 mg) was incubated 12 h at room temperature in neat N,N-dimethylethylenediamine (50 µl). The diamine was removed under reduced pressure and the peptide was subjected to high voltage paper electrophoresis at pH 2.1 at a voltage gradient of 60 V/cm for 20 min. The position of the peptide was located by staining the ends of the paper with ninhydrin. The band containing the peptide was eluted and the peptide was subjected to high voltage paper electrophoresis at pH 4.4. The purified peptide was eluted and analyzed by mass spectroscopy or served as starting material for other standard peptides.

Acetyl-leucyl-leucyl-leucyl-tyrosyl(O-acetyl)-$NHCH_2CH_2N^+(CH_3)_3$ and leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N^+(CH_3)_3$. Leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ (1 mg) was incubated 30 min in either a 1 % solution of acetic anhydride or di-*tert*-butyldicarbonate in pyridine (50 µl). The solutions were dried in pyrex hydrolysis tubes and the peptides sealed *in vacuo* with [^{12}C]iodomethane (20 µl). The tubes were incubated overnight at 100°C. The acetylated peptide was analyzed without further treatment. The tert-butyloxycarbonyl peptide was first incubated in trifluoroacetic acid (30 µl) for 30 min to liberate the amino terminus.

$(CH_3)_3N^+$-leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N^+(CH_3)_3$ and $(CH_3)_2N$-leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N^+(CH_3)_3$. Leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ (1 mg)

was transferred into a hydrolysis tube and sealed *in vacuo* with [^{12}C]iodomethane (20 µl). The peptide was reacted overnight at 100°C.

MS and MS/MS analyses. Microspray mass analyses were carried out by Dr. Pierre Thibault of the Institute for Biological Sciences, National Reasearch Council of Canada. A PE Sciex API-300 triple quadrupole mass spectrometer was used for conventional and tandem mass analyses. Samples were dissolved in 0.1 M formic acid. One hundred picomoles of standard was injected per analysis.

In the case of variation 5, analytes were first separated using a Zorbax RXC18 2.1mm X 150 mm reverse phase high performance liquid chromatography columnn. The components were eluted using a 5-95% acetonitrile gradient in an aqueous solution containing 0.1% trifluoroacetic acid. The effluent was split 1:15 (v/v) and samples were fed into the API-300 instrument. Tandem operation was configured to scan for parent ions losing the fragment ion, N,N-dimethylethylenediamine (89 m/z).

Preparation of α-chymotrypsin derivatives for microspray analysis:

Protein lyophilization. A 50 µg/ml solution of protein was made at 4°C. Without delay, a 500 µl aliquot (25 µg protein) was transferred into an Eppendorf tube, flash frozen in liquid nitrogen, and lyophilized. Protein was clearly visible as a very fine mesh after lyophilization. Some protein was esterified and then amidated while others were incubated immediately with diamine and served as control proteins.

Esterification. A 0.1 M solution (800 µl) of hydrogen chloride in anhydrous methanol (2.5% dioxane) is delivered to the protein. The Eppendorf tubes are briefly sonicated and shaken 36 h

at room temperature. The tubes are flash frozen, placed upright in a dessicator and the contents removed *in vacuo*. The tubes are kept under vacuum for several hours thereafter.

Pyridine treatment. Esterified protein (variation 1 and 3 only) was sometimes dissolved/suspended in anhydrous pyridine (500 µl) from 1 to 12 h. The tubes were flash frozen and the solvent evaporated in a dessicator.

Protein amidation. Esterified protein was dissolved/suspended in either ethylenediamine or N,N-dimethylethylenediamine (70 µl), briefly sonicated, and shaken 12 h at room temperature. In the case of variations 1 and 2, protein was transferred into small pyrex tubes. All samples were placed 24 h under vacuum. In the case of variation 4, the residual contents were suspended in n-butanol (2 X 1 ml), sonicated, and spun down. Residual diamine, in n-butanol, was decanted.

Methylation: Variations 1 and 2. Amidated protein was methylated *in vacuo* for 12 h at 75°C using neat [^{14}C]iodomethane (20 µCi) that was delivered as an octane solution (20 µl). The reaction tube was opened and without further manipulation, slipped into a large pyrex tube. *In vacuo* methylation was recommenced for 12 h using [^{12}C]iodomethane (25 µl). The reaction was terminated, and the inner tube containing sample was placed under vacuum.

Protein acetylation and methylation: Variation 3. Amidated protein was dissolved/suspended for 3 h in a 1% solution (300 µl) of acetic anhydride in pyridine. The reaction was terminated by addition of water (100µl). The entire contents were transferred into a small pyrex tube and the

solution was removed *in vacuo*. The protein was methylated in the same manner as variations 1 and 2.

Thermolysin digestion. Protein derivatives were dissolved/suspended and sonicated in 0.4% ammonium bicarbonate solution, 1 mM in calcium chloride, and 50 μg of thermolysin was added, dissolved in the same buffer solution (50 μl). The medium was adjusted to pH 7.8 if necessary with dilute acetic acid (1 %) and the digestion was carried out at 37°C for 48 h. The contents were lyophilized, water was added (200 μl) and the contents were relyophilized.

Peptide acetylation and methylation: Variation 4. Peptide derivatives were dissolved in 1% acetic anhydride solution in pyridine (300 μl), sonicated, and incubated 3 h. The reaction was terminated by addition of water (100μl) and the solution was removed *in vacuo*. The peptides were taken up in a 1:10 (v/v) octane:N,N-dimethylformamide solution (200 μl), that contained [^{14}C]iodomethane (20 μCi). The peptide derivatives were incubated 48 h at room temperature and then [^{12}C]iodomethane (20 μl) was introduced. The solution was left 8 h, removed *in vacuo* and disposed of cautiously.

Diagonal electrophoresis and paper chromatography. Peptides were spotted over 3 cm on Whatman 3 MM paper, wetted with pH 2.1 buffer of the composition formic acid/acetic acid/water (1:4:45 by volume) and subjected to high voltage electrophoresis at a voltage gradient of 20 V/cm for 90 min. The peptides were located using dansylsulphonic acid and dansylarginine as fluorescent markers. The strip containing the peptides was cut out, rotated 90° and stitched

onto another Whatman 3 MM paper. The paper underneath the strip was excised, the paper was wetted with pH 4.4 buffer of the composition pyridine/acetic acid/water (6:10:1200 by volume) and the peptides were subjected to electrophoresis at a voltage gradient of 30 V/cm for 60 min. In the case of variations 1-4, peptides were located by autoradiography. There were enough radioactive species in the sample to deliniate the neutral spot and diagonal positions without having to resort to markers. In some cases, the diagonal strip was cut out, restitched onto either 3 MM or 1 MM paper and chromatographed overnight with butanol/acetic acid/water/pyridine (15:3:12:10 by volume). In the case of variaton 5, dansylarginine was spotted onto the neutral band before the second dimension was run. At pH 4.4, dansyl arginine has a relative mobility of 2.38 with respect to the neutral band, as measured from the origin. The diagonal line was delineated using radioactive hexamethylethylenediamine obtained from another reaction. Peptides of the diagonal strip were eluted with pH 2.1 buffer and sent for analysis by high performance liquid chromatography and tandem mass spectral methods.

4.3 Results

4.3.1 Mass analyses of standard peptides.

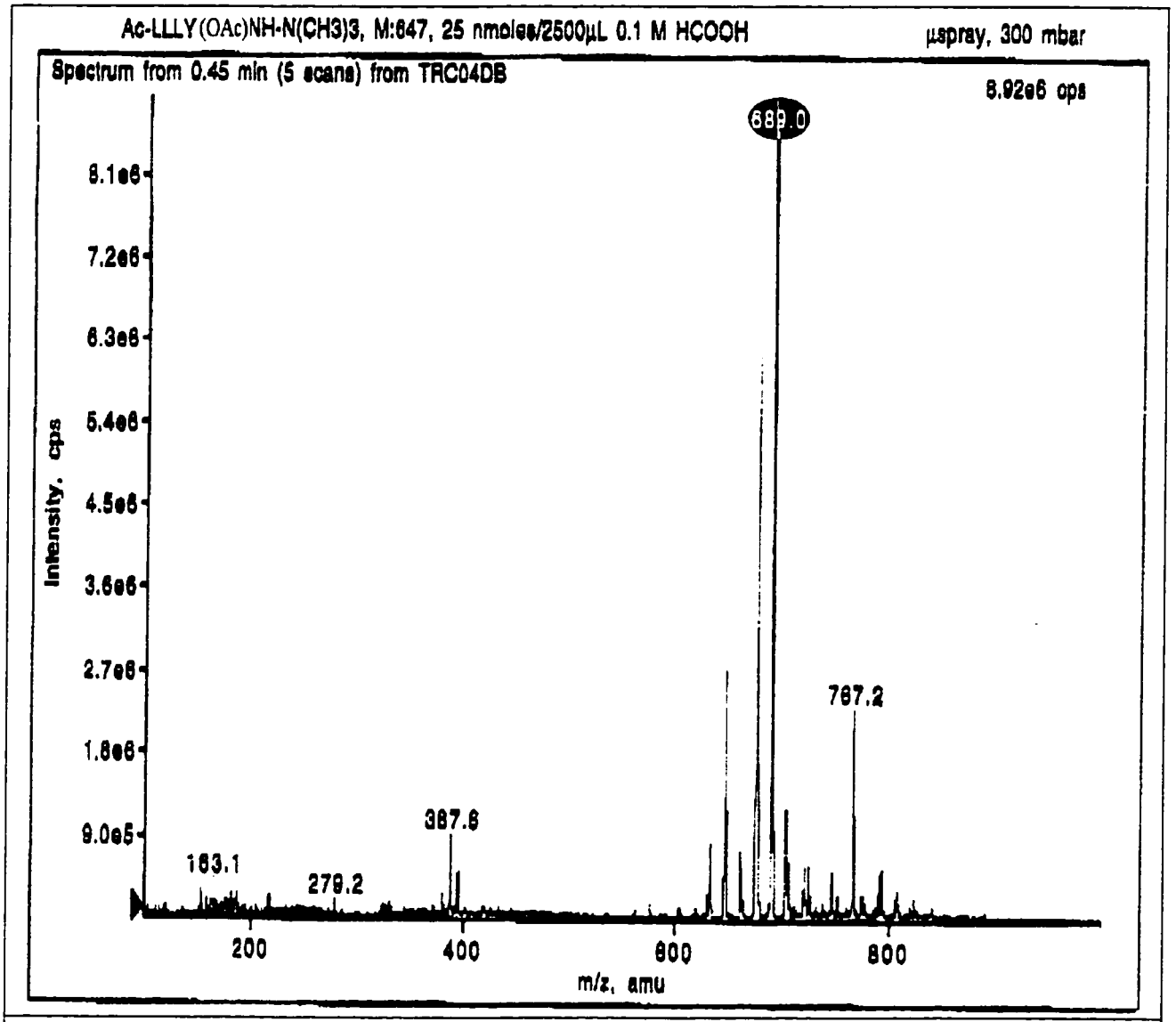


Figure 4.3.1. Microspray mass spectrum of acetyl-leucyl-leucyl-leucyl-tyrosyl(O-acetyl)-$NHCH_2CH_2N^+(CH_3)_3$. 100 pmole was injected into a triple quadrupole instrument (Q_1qQ_2) operated in single MS mode (qqQ). The 689.0 m/z ratio corresponds to the compound.

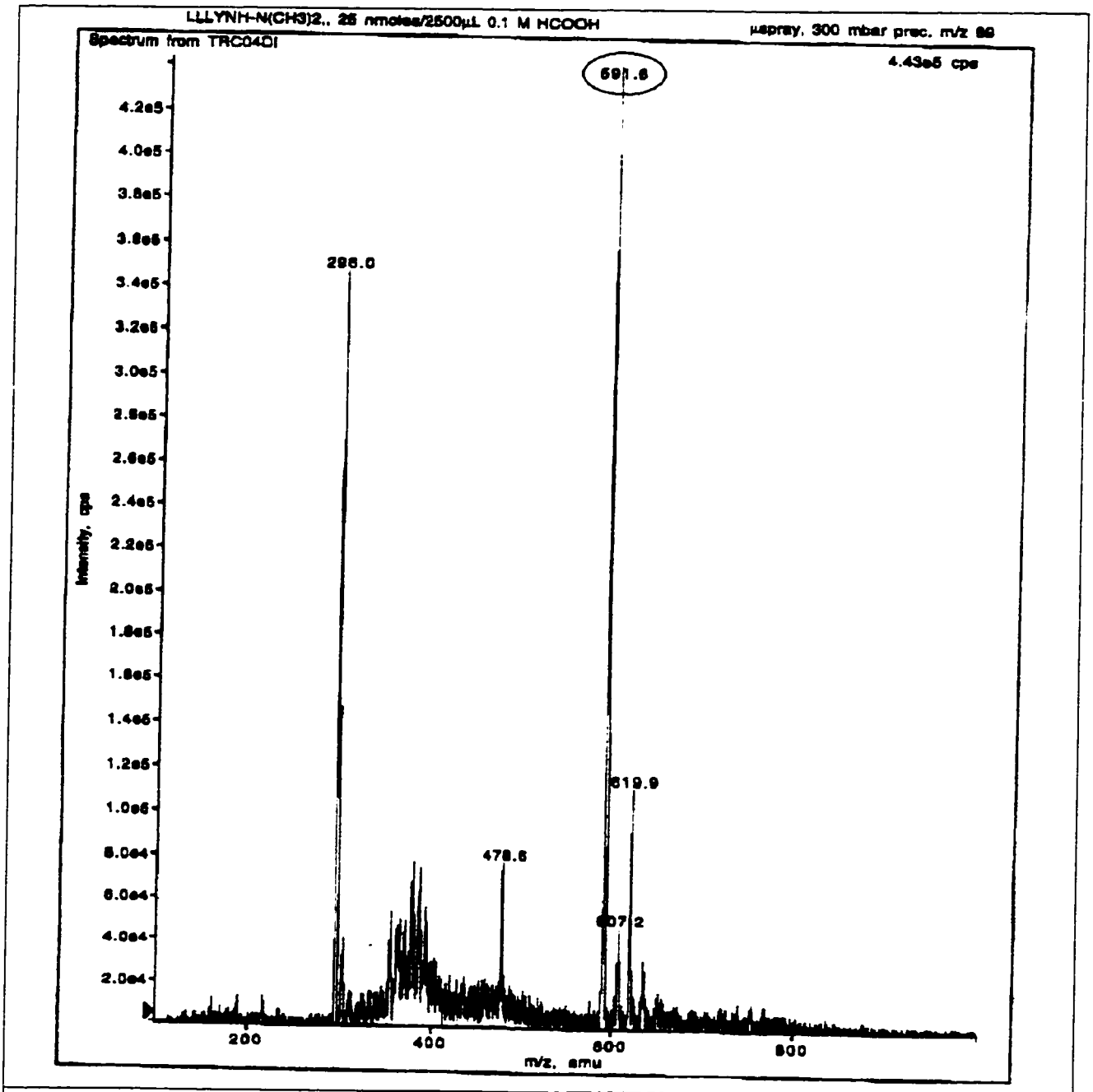


Figure 4.3.2. Precursor ion-spectrum of standard peptide leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ obtained by microspray MS/MS analysis. 100 pmole was injected into a triple quadrupole instrument (Q_1qQ_2). The product-ion mass filter (Q_2) was configured to pass ions with an 89 mass/charge ratio, corresponding to the N,N-dimethylethylenediamine fragment. Mass/charge ratios were measured at Q_1 for those ions which were detected. Mass/charge ratios correspond to the following: 591.8, leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$; 478.6, leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$; 296.6, doubly charged leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$.

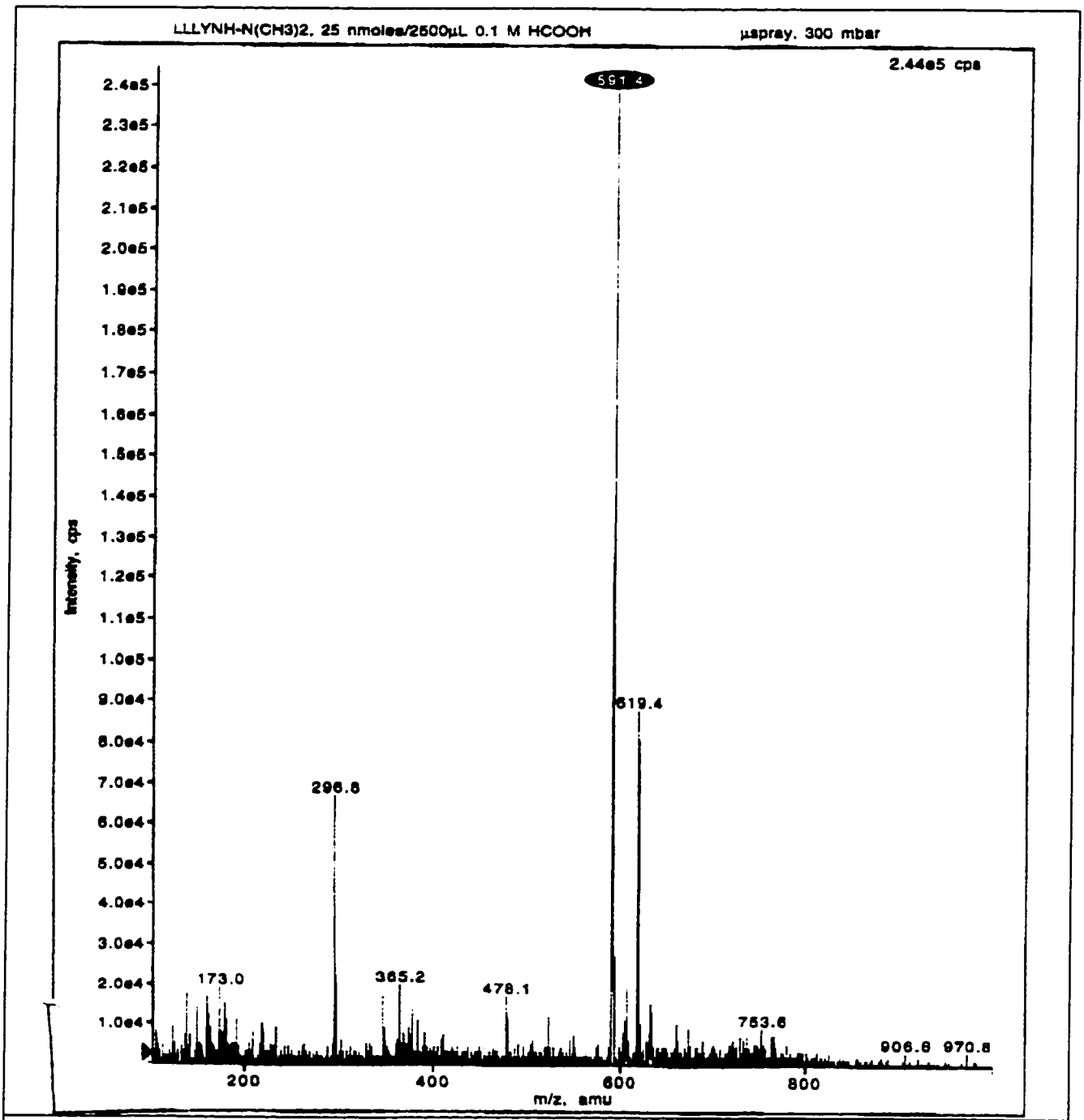


Figure 4.3.3. Constant neutral loss spectrum of standard peptide leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ obtained by microspray MS/MS analysis. 100 pmole was injected into a triple quadrupole instrument (Q_1qQ_2). The precursor and product-ion mass filters were sweeped such that a constant difference of 45 mass/charge units was maintained between Q_1 and Q_2, corresponding to the dimethylamine fragment. Mass/charge ratios were measured at Q_1 for those ions which were detected. Mass/charge ratios correspond to the following ion peaks: 591.8: leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$; 478.6, leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$; 365.2, leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$; 296.6, doubly charged leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$.

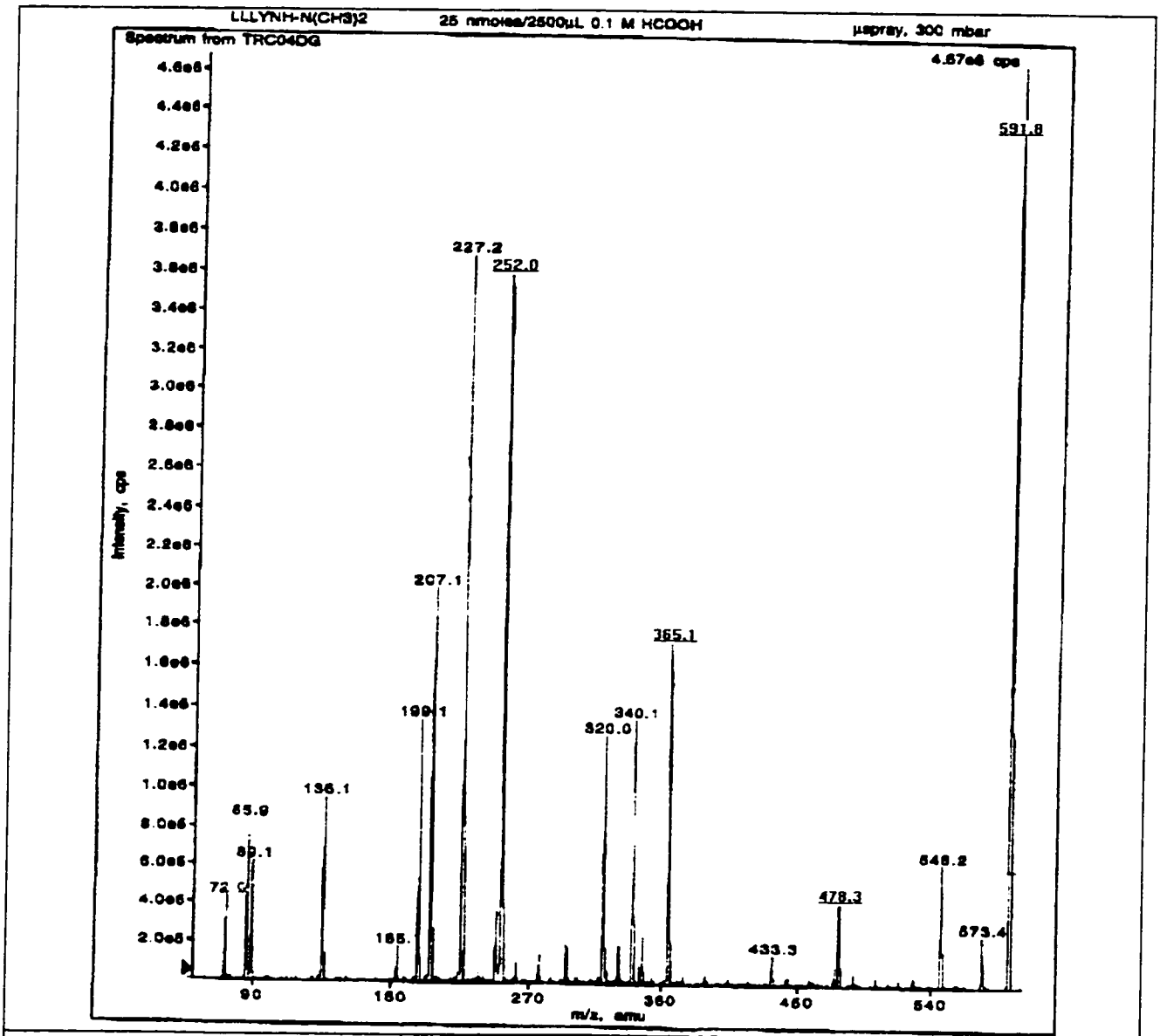


Figure 4.3.4. Product-ion spectrum of standard peptide leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ obtained by microspray MS/MS analysis. 100 pmole was injected into a triple quadrupole instrument (Q_1qQ_2). The precursor-ion mass filter Q_1 was configured to pass parent ions with a 591.8 mass/charge ratio, corresponding to the peptide. Mass/charge ratios were measured at Q_2 for parent and fragment ions which arrived at the detector. Mass/charge ratios and their respective ions are indicated in table 4.3.1.

Figures 4.3.1, 4.3.2, 4.3.3 and 4.3.4 illustrate the mass spectra of some peptide standards. In the case of acetyl-leucyl-leucyl-leucyl-tyrosyl(O-acetyl)-$NHCH_2CH_2N^+(CH_3)_3$ (Figure 4.3.1), the 689 m/z peak corresponded to the expected mass/charge ratio. Neither trimethylamine, nor a mass difference corresponding to trimethylamine was observed. Figure 4.3.2 depicts the precursor-ion spectrum of leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ with the instrument configured to search for an N,N-dimethylethylenediamine fragment ion (89 m/z) in the product mass filter. Peaks were found at 591.8 m/z, 478.6 m/z and 296.0 m/z. Figure 4.3.3 depicts the constant neutral loss spectrum of leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ with the instrument configured to search for the loss of a neutral dimethylamine fragment (45 m/z). Peaks were observed at 591.8 m/z, 478.6 m/z, 365.2 m/z and 296.0 m/z. Figure 4.3.4 and table 4.3.1 illustrates how successfully leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ can be sequenced using tandem analysis, with the instrument configured to analyze the fragmentation of a particular parent ion.

Table 4.3.1. Selected mass/charge ratios observed in the product ion spectrum of leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ by microspray MS/MS analysis. 100 pmole was injected into a triple quadrupole instrument (Q_1qQ_2). The precursor mass filter (Q_1) was set to pass the parent ion of 591.8 mass/charge ratio. The product mass filter (Q_2) was sweeped to allow detection of all fragment ions. A difference of 45 m/z among two species corresponds to the loss of neutral dimethylamine. All ions reported had strong signals in the spectrum.

mass/charge ratio	corresponding ion
591.8	leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$
546.2	leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ with loss of dimethylamine
478.3	leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$
433.3	leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ with loss of dimethylamine
365.1	leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$
340.1	leucyl-leucyl-leucyl-$C{\equiv}O^+$ acylium ion
320.0	leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ with loss of dimethylamine
252.0	tyrosyl-$NHCH_2CH_2N(CH_3)_2$
227.2	leucyl-leucyl-$C{\equiv}O^+$ acylium ion
207.1	tyrosyl-$NHCH_2CH_2N(CH_3)_2$ with loss of dimethylamine
89.1	N,N-dimethylethylenediamine

4.3.2 Preparation and mass analysis of protein.

Figures 4.3.5, 4.3.6 and 4.3.7 show the integrity of protein at different stages of derivatization. Suspending protein in methanol or acidified methanol did not appear to be problematic (Figure 4.3.5, lanes 1&2; figure 4.3.7, lanes 5&8). Incubation of protein and esterified protein with ethylenediamine (Figure 4.3.5, lanes 3&4) cleaved protein disulphide bonds. This was not observed when protein was incubated with N,N-dimethylethylenediamine (Figure 4.3.7, lanes 6&9). The molecular weight of amidated protein (Figure 4.3.5, lane 4; figure 4.3.7, lane 9) appears to increase with respect to control protein incubated in diamine (Figure 4.3.5, lane 3; figure 4.3.7, lane 8). *In vacuo* methylation of protein derivatized with ethylenediamine produced a smeared profile of radioactive material (Figure 4.3.6, lane 2). The two peptide chains observed in figure 4.3.5 (lane 4) were barely visible after methylation (Figure 4.3.6, lane 2). There was also radioactivity located at the separating and stacking gel interface and in the well.

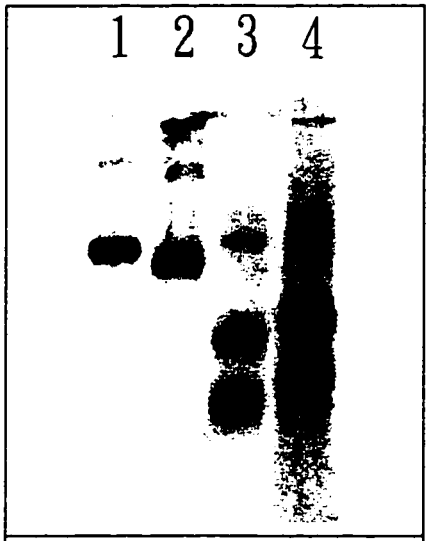


Figure 4.3.5. Esterification and amidation of α-chymotrypsin. A sodium dodecyl sulphate/15% polyacrylamide gel stained with Coomassie blue is shown. Lane 1, protein in methanol; lane 2, protein in HCl/methanol solution; lanes 3&4, protein incubated in pyridine (1 h), then 3 h in ethylenediamine.

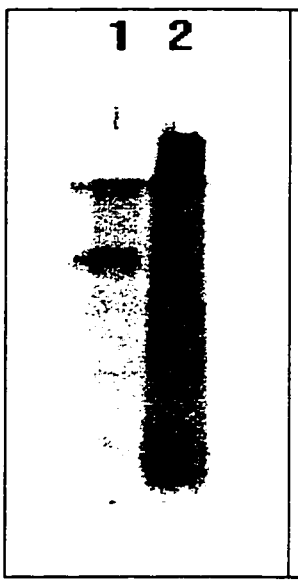


Figure 4.3.6. *In vacuo* [^{14}C]methylation of α-chymotrypsin derivatized with ethylenediamine. A sodium dodecyl sulphate/15% polyacrylamide gel was subjected 4 d to autoradiography. Lane 1, *in vacuo* [^{14}C]methylated α-chymotrypsin marker protein; lane 2, *in vacuo* [^{14}C] and [^{12}C]methylated α-chymotrypsin prederivatized with ethylenediamine.

Figures 4.3.8, 4.3.9, 4.3.10 and 4.3.11 show profiles of derivatized peptides that were separated by two-dimensional paper electrophoresis and paper chromatography, and observed by autoradiography. Variation 4 is not shown since there was nothing to

observe except for radioactive markers. The profile did show, however, that N,N-dimethylethylenediamine was extremely pure following redistillation. Figure 4.3.12 shows how the diagonal line of variation 5 was located.

The incorporation of radioactivity into protein was sufficient to enable easy detection, as exemplified by figure 4.3.8 (Variation 1). Strong spots were obtained after exposing X-ray film for only a short time. Diagonal line **1-2** (from 1 to 2) of each chromatogram (Figures 4.3.8 to 4.3.12) is presumed to contain the C-terminal peptides. The incorporation of radioactivity was such that other diagonal lines were easily observed. Diagonal lines **1-3** (from 1 to 3) and **1-4** (from 1 to 4) presumably deliniate a related series of non C-terminal peptides, and diagonal line **1-5** (from 1 to 5) clearly describes neutral species at pH 4.4. Paper chromatography of diagonal peptides indicated that a good separation was achievable for methylated proteins derivatized with ethylenediamine (Figure 4.3.9) and N,N-dimethylethylenediamine (not shown).

The intense spot produced by the relative abundence of hexamethylethylenediamine (Figure 4.3.8; top of diagonal line **1-2**) is clearly visible and shows that ethylenediamine is not removed quantitatively under vacuum and is present while methylation is carried out. The apparent smearing of protein (Figure 4.3.6) put to question if the C-terminal peptides described by line **1-2** was genuine or fabricated under the *in vacuo* methylation conditions of variation 1. Unfortunately, a nonesterified control protein was not available for comparative purposes. In each of variations 2 to 5, a control protein was included in the analysis. In each variation, the profiles obtained along diagonal lines **1-2** were identical for both the esterified and control protein (comparison not shown, except for variation 3; figure 4.3.11, compare left and right chromatograms).

It was hoped that N,N-dimethylethylenediamine would be more easily removed under

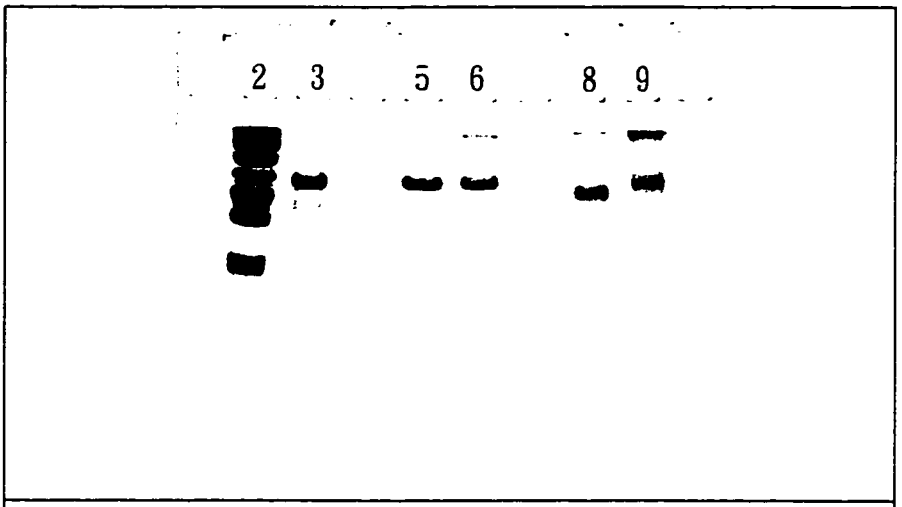


Figure 4.3.7. Esterification and amidation of α-chymotrypsin. A sodium dodecyl sulphate/15% polyacrylamide gel stained with Coomassie blue is shown. Lane 2, low molecular weight protein markers; lane 3, untreated protein; lane 5, protein in methanol; lane 6, protein in methanol, then incubated 12 h in N,N-dimethylethylenediamine; lane 8, protein treated in HCl/methanol solution; lane 9, protein treated in HCl/methanol solution, then incubated 12 h in N,N-dimethylethylenediamine.

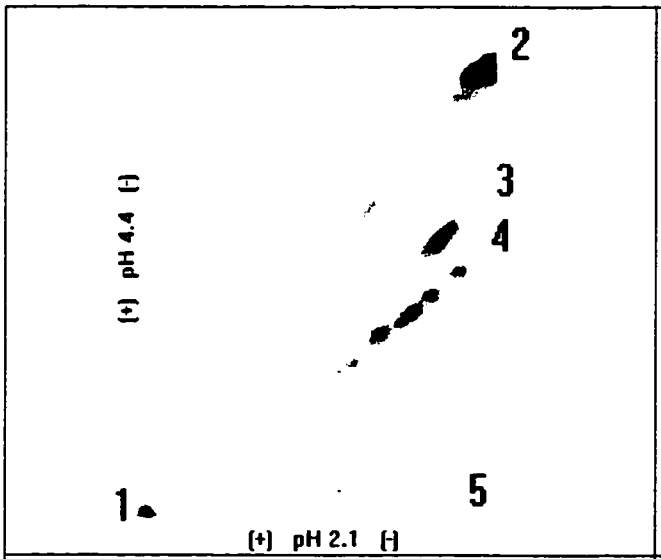


Figure 4.3.8. Two-dimensional high voltage paper electrophoresis of α-chymotrypsin derivatized according to variation 1. Protein was esterified and incubated with ethylenediamine, methylated *in vacuo* and digested. Autoradiography was 4 d. Numbers are included to assist discussion of the results in the text:1) neutral spot; 2) hexamethylethylenediamine and C-terminal peptides; 3&4) non C-terminal peptide series; 5) neutral peptides.

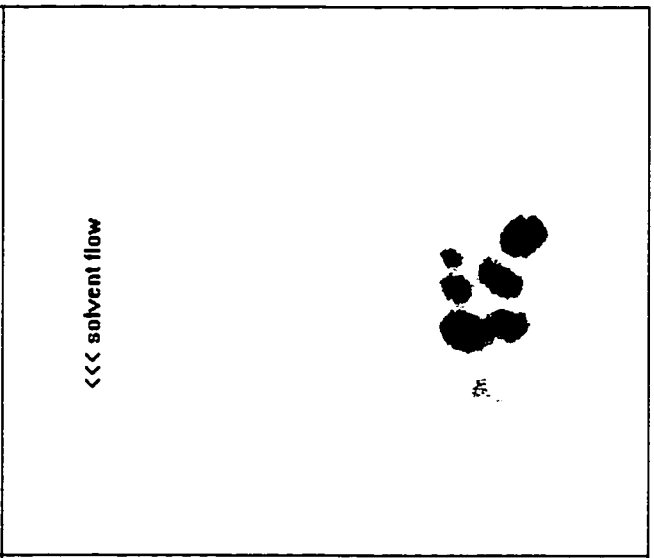


Figure 4.3.9. Paper chromatography to purify the diagonal peptides of variation 1. The diagonal line described by numbers 1 to 2 in figure 4.3.8 was stitched onto another Whatman 3 MM paper and chromatographed overnight. Autoradiography was carried out for 7 d. The spots to the right are different methylation states of ethylenediamine. Faint central spots were also observed.

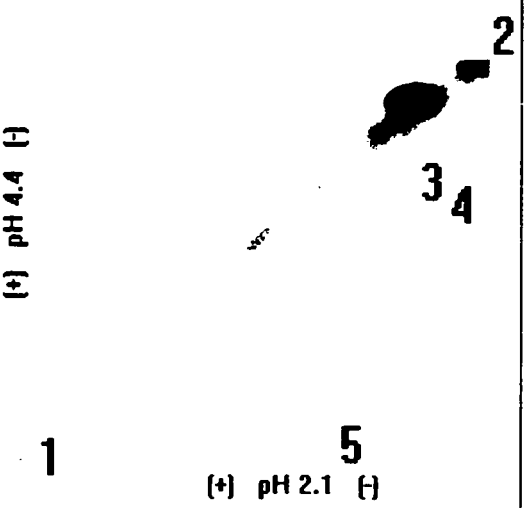


Figure 4.3.10. Two-dimensional high voltage paper electrophoresis of α-chymotrypsin derivatized according to variation 2. Protein was esterified and incubated with N,N-dimethylethylenediamine, methylated *in vacuo* and digested. Autoradiography was 4 d. Numbers are included to assist discussion of the results in the text: 1) neutral spot; 2) diagonal markers and C-terminal peptides; 3&4) series of non C-terminal peptides; 5) neutral peptides. The profile of nonesterified protein was identical.

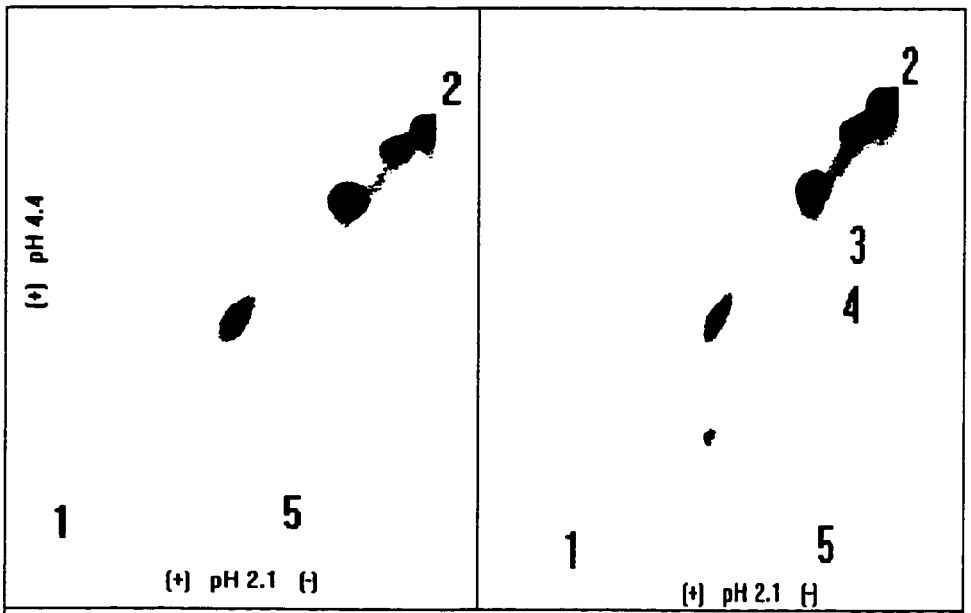


Figure 4.3.11. Two-dimensional high voltage paper electrophoresis of α-chymotrypsin derivatized according to variation 3. Protein was not esterified (*left*) or esterified (*right*), treated with pyridine, incubated in N,N-dimethylethylenediamine, acetylated, methylated *in vacuo* and digested. Autoradiography was 4 d. Numbers are included to assist discussion of the results in the text: 1) neutral spot; 2) diagonal markers and C-terminal peptides; 3&4) non C-terminal peptide series; 5) neutral peptides. Esterified and nonesterified proteins that were not treated with pyridine gave the same profiles as those that were treated.

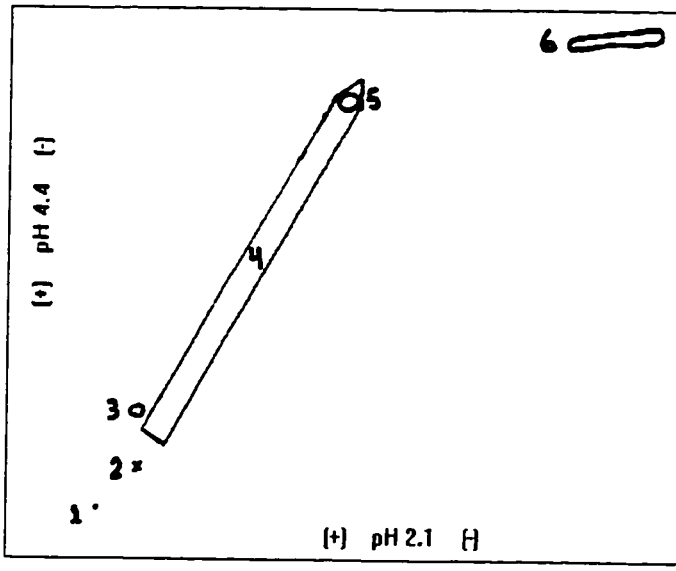


Figure 4.3.12. Two-dimensional high voltage paper electrophoresis of α-chymotrypsin derivatized according to variation 5. Protein was esterified, incubated in N,N-dimethyl-ethylenediamine, and digested. Numbering corresponds to: 1) origin; 2) neutral spot; 3) dansylarginine; 4) excised diagonal; 5) hexa-methylethylenediamine; 6) N,N-dimethylethyl-enediamine. Esterified and nonesterified proteins were treated identically and analyzed by mass spectrometry. Position 2 was calculated, compounds 3&6 were located under UV light, and 5 was located after 1 d autoradiography.

vacuum, however, this was not the observation. The profiles of variation 1 (Figure 4.3.9) and variation 2 (Figure 4.3.10) indicated that diamine was present. Figure 4.3.11 (Variation 3) shows that diagonal lines **1-3** and **1-4** are absent in the nonesterified control protein. The profiles were otherwise identical. The diagonal chromatogram of variation 5 (Figure 4.3.12) was unusual in that ethylenediamine (compound 6) and hexamethylethylenediamine (compound 5) did not lie on the same diagonal line.

Figures 4.3.13 and 4.3.14 show the chromatographic and mass spectral profiles obtained for standard peptide and protein analytes, respectively. Ultraviolet and conventional mass spectral detection profiles of leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ are very similar (Figure 4.3.13A and B), while the tandem mass spectral profile (Figure 4.3.13C) is missing one peak. Conventional mass analysis (not shown), and tandem analysis (Figure 4.3.13D) of a certain parent ion (Figure 4.3.13C; retention time = 14.43 min) indicated that the compound had a mass/charge

ratio of 592 m/z. The other peak (Figure 4.3.13C; retention time = 20.15 min) also corresponded to a mass/charge ratio of 591 m/z (not shown).

Unlike the profiles of figure 4.3.13, which were easily interpreted, profiles obtained of the diagonal peptides of derivatized α-chymotrypsin (Figure 4.3.14) were complicated. The signal to

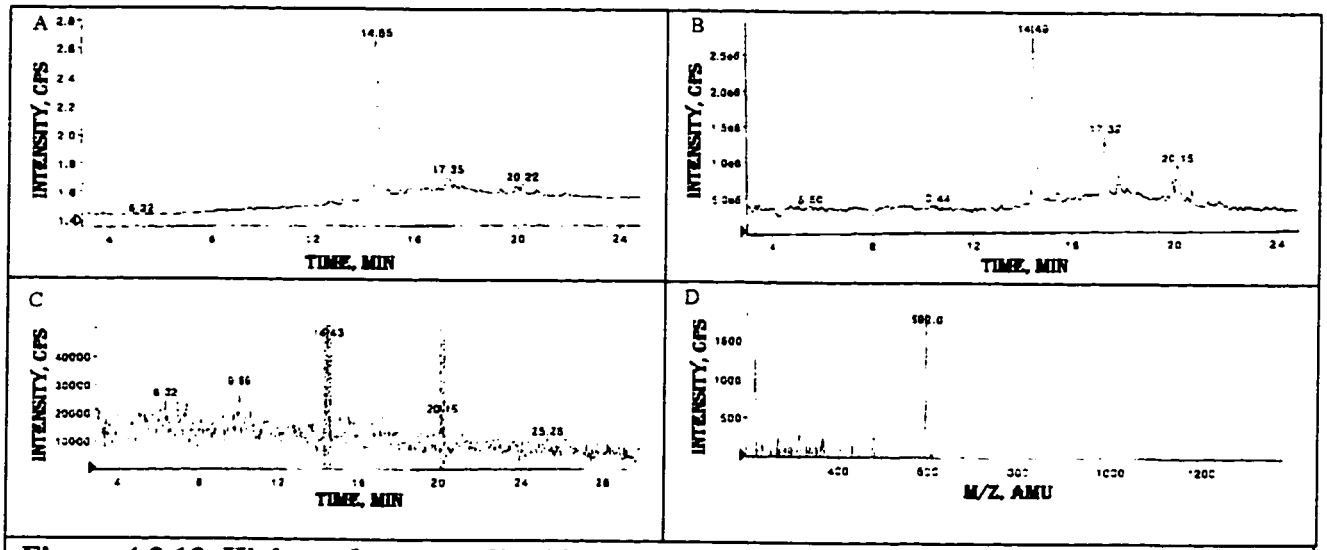


Figure 4.3.13. High performance liquid chromatography and mass analysis of the standard peptide leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$. 1 nmole was injected into a reverse phase column and chromatographed using a 5-95% acetonitrile gradient in an aqueous solution of 0.1% trifluroacetic acid. A one-fifteenth fraction of the flow was delivered to a quadrupole instrument for detection purposes. Detection modes used were (A) ultraviolet light (B) conventional mass spectroscopy (C) tandem mass spectroscopy. Precursor ion spectra were obtained for analytes losing N,N-dimethylethylenediamine (89 m/z). A precursor ion spectrum (D) is shown for the peak of profile C with retention time = 14.43 minutes.

noise ratio was very poor. More importantly, ultraviolet and conventional mass spectral profiles of nonesterified control protein were superimposable (not shown) to those of the amidated protein (Figure 4.3.14A and B). Even the profile using tandem mass spectral detection of control protein (not shown) and amidated protein (Figure 4.3.14C) suffered from weak signal intensity. While potential peptides were observed in conventional mass detection mode (Figure 4.3.14B) and tandem mass spectral mode (Figure 4.3.14C; retention time = 12.39 min), none of the peaks

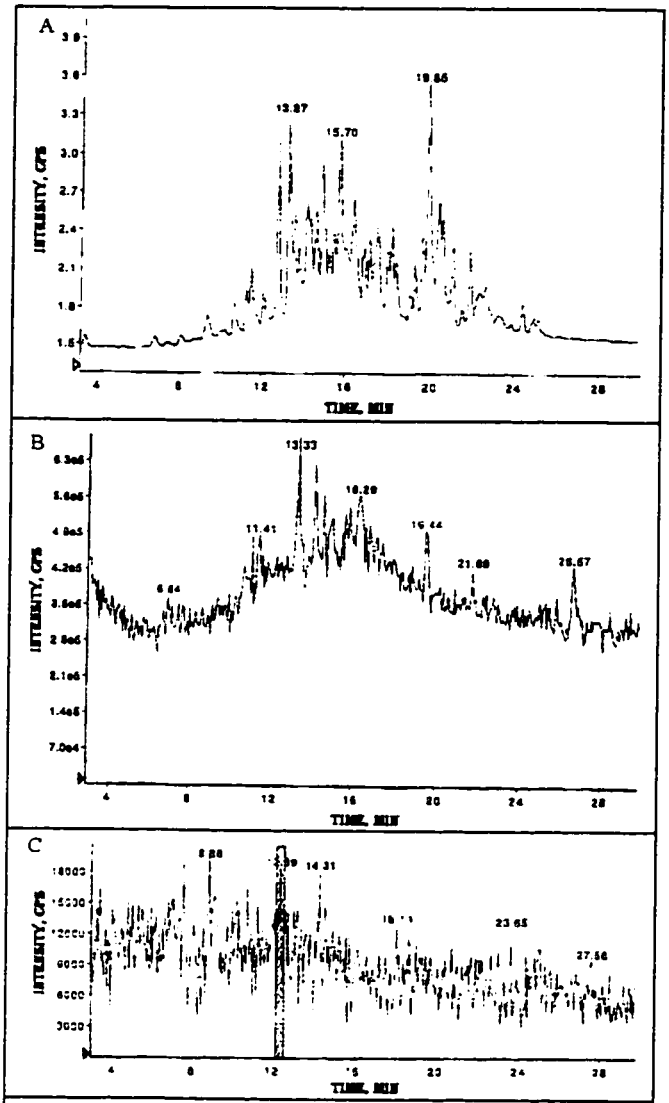


Figure 4.3.14. High performance liquid chromatography of eluted diagonal peptides of variation 5. 15% of total sample was injected into a reverse phase column and chromatographed using a 5-95% acetonitrile gradient in an aqueous solution of 0.1% trifluroacetic acid. A one-fifteenth fraction of the flow was delivered to a quadrupole instrument for detection purposes. Detection modes used were (A) ultraviolet light (B) conventional mass spectroscopy (C) tandem mass spectroscopy. A potential C-terminal peptide (C; retention time = 12.39 min) is indicated.

analyzed provided a qualifying mass/charge ratio characteristic of a C-terminal peptide of α-chymotrypsin (data not shown).

4.4 Discussion

An alternate nonaqueous approach for the analysis of protein C-termini has been presented for use with minute amounts of protein. Unlike other methods which sequentially degrade amino acid residues from the C-terminus, this method simply attempts to isolate the C-terminal peptide. While the merit of the technique was not realized in the given time, the derivatization of protein up to and including the amidation step appeared to be successful and quantitative. The approach can potentially be used to compliment existing methods once further refinements are made.

4.4.1 Mass spectral detection of peptides.

Triple quadrupole mass spectrometers have four operating modes and can produce conventional (Figure 4.3.1), precursor-ion (Figure

4.3.2), constant neutral loss (Figure 4.3.3), and product-ion (Figure 4.3.4) spectra. The predominant peak of figure 4.3.1 (689 m/z) corresponds to the molecular weight of acetyl-leucyl-leucyl-leucyl-tyrosyl(O-acetyl)-$NHCH_2CH_2N^+(CH_3)_3$. There is one minor peak which corresponds to unmethylated peptide (675.5 m/z), acetyl-leucyl-leucyl-leucyl-tyrosyl(O-acetyl)-$NHCH_2CH_2N(CH_3)_2$, and two very small peaks (661.4 m/z, 647.5 m/z) which represents the loss of the acetyl group in each peptide standard. As noted, with microspray and quadrupole methods, ionizations are relatively nondestructive and produce spectra which are predominantly composed of parent ions (McCloskey, 1990; Biemann, 1992). It was hoped that the model peptides examined would liberate fragments that could serve as qualifying indications of a C-terminal peptide. In this context, the most useful fragments that could be formed in the peptide acetyl-leucyl-leucyl-leucyl-tyrosyl(O-acetyl)-$NHCH_2CH_2N^+(CH_3)_3$ are trimethylamine and N,N,N-trimethylethylenediamine. Unfortunately, neither the fragments, nor loss of the fragments were observed by conventional mass spectral methods (Figure 4.3.1). Similarly, leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ did not appear to lose dimethylamine or N,N-dimethylethylene-diamine (data not shown). In contrast, leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ (Figures 4.3.2, 4.3.3, 4.3.4 and table 4.3.1) and another model peptide, $(CH_3)_2N$-leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N^+(CH_3)_3$ (data not shown), did lose dimethylamine, N,N-dimethyl-ethylenediamine and trimethylamine very easily under the conditions of tandem mass spectral analysis. The fact these fragments were not observed in the conventional operating mode was unexpected, and is presumed to be a consequence of the instrumental parameters chosen for acquiring the data. Nevertheless, the observed loss of such characteristic groups was an important finding and provided a potentially useful means to search for peptides of the

C-terminus, especially when analyzing mixtures of peptides in the presence of impurities. Potential problems associated with impurities is best illustrated in figure 4.3.1, where despite an involved purification, even a standard peptide is seen to contain impurities.

Figure 4.3.2 is the precursor-ion spectrum of the peptide leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$. In this case, the second mass filter, Q_2, is configured to allow only the N,N-dimethylethylenediamine ion to pass. The first mass filter, Q_1, was sweeped and the spectrum shows only ions which liberated a charged fragment of 89 mass/charge units. There is another fragment of 478.6 charge/mass units which corresponds to the peptide leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ and yet another fragment of 296.0 charge/mass units, which corresponds to leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$, but with a net double positive charge. The remaining background is presumed to come from the matrix used for microspray ionization. As suggested, the potential of tandem mass spectroscopy for detection of C-terminal peptides is significant. The peak at 619.9 mass/charge units corresponds to a mass increase of two methylene groups with respect to the analyte. The origin is uncertain.

Figure 4.3.3 is the constant neutral loss spectrum of the peptide leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$. The first and second mass filters are increased synchronously, such that a mass/charge ratio difference of 45 mass/charge units is maintained at all mass/charge values. In this operating mode, parent ions which liberate neutral dimethylamine in the collision chamber are observed. Once again, the predominant peak (591.4 m/z) corresponds to the peptide ion and another peak (296.8 m/z) is the doubly charged peptide ion. Two minor peaks (478.1 m/z and 365.2 m/z) correspond to leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ and leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$.

Product-ion spectra show characteristic fragmentation patterns of particular parent ions in a mixture of parent ions. These spectra are very often used to sequence individual peptides in a mixture, however, the molecular weight of the analyte must be known so that the instrument can be configured to allow the corresponding parent ion to pass through the first mass filter (Q_1). It was observed above that the precursor-ion and constant neutral loss spectra of figures 4.3.2 and 4.3.3 not only identified the mass/charge ratio of a model C-terminal peptide, but provided extensive sequence information. The sequence information provided was a bonus, and was presumed to be due to minor fragmentations which occurred during microspray ionization (McCloskey, 1990; Biemann, 1992; Capriolli *et al.*, 1995). It would have been sufficient if these spectra simply identified the mass/charge ratio of C-terminal peptides out of a mixture. With the parent ions identified, the product-ion spectrum of each potential peptide could be obtained by reanalyzing the sample. Figure 4.3.4 and table 4.3.1 shows the merit of this operating mode in sequencing the peptide leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$. The fragmentations reported are unmistakably characteristic of this peptide. For analytes which contain impurities and peptides of unknown molecular weight (McCloskey, 1990), precursor-ion and constant neutral loss spectra can first be acquired and used to determine the mass/charge ratio of potential C-terminal peptides in the mixture.

4.4.2. Protein derivatives.

The true test of the merit of nonaqueous derivatization of carboxyl groups is not just measured by product yield, but also in terms of convenience and simplicity. Performing the experiment requires a certain degree of skill, but once acquired, the protocols attempted are rather straightforward. While other reagents, namely carbodiimides, acid anhydrides and chlorides,

pyrocarbonates and alkylchloroformates, are more effective activators (Bodanszky and Bodanszky, 1994), there were problems associated with their use in nonaqueous media. First, the amount of residual salt and charge carriers remaining after reaction was unacceptable. In contrast, methanolic hydrogen chloride solution could be removed much more effectively *in vacuo* at room temperature. Second, the activated carboxyl groups were sufficiently reactive that proteins became crosslinked. It is believed that lyophilized proteins have interprotein salt bridges joining amino and carboxyl groups. Once activated, the carboxyl group may form an amide bond with the adjacent amino group, leading to crosslinked protein. The crosslinking phenomena can also be mediated by inter-protein ester bond formation, since it was observed that N-trimethylated proteins could still be made to crosslink using carboxyl group activating agents. In contrast, the poor leaving ability of methyl esters and the protonation state of amino groups in the esterified protein appeared to prevent crosslinking and improve shelf-life.

Figures 4.3.5, 4.3.6, and 4.3.7 show the integrity and yield of protein derivatizations as observed by sodium dodecyl sulphate polyacrylamide gel electrophoresis. The esterification (Figure 4.3.5, lane 2; figure 4.3.7, lane 8) proceeds without degradation of protein. The integrity of backbone amide bonds appears intact following reaction but the possibility of having N→O acyl migrations cannot be ruled out on this basis. The change of properties upon esterification is notable, for example, the control proteins in methanol consistently migrate at a slower rate (Figure 4.3.5, lane 1; figure 4.3.7, lane 5) compared to esterified protein. Since none of the samples were boiled or contained 2-mercaptoethanol, nothing can be said about the ratio of sodium dodecyl sulphate bound to protein or the expected mobility of the protein (Hames and Rickwood, 1990). The results at least suggest, however, that the structures are different. The

treatment of protein with ethylenediamine results in cleavage of the disulphide bonds (Figure 4.3.5, lanes 3&4). This has been reported to occur in basic aqueous solution by abstraction of one of the Cα protons of cystine (Darbre, 1986). It is suspected that the same mechanism operates in this example. The rupture of disulphide bonds had no bearing on whether the protein was previously esterified. While α-chymotrypsin has three chains, the smallest is only sixteen amino acid residues and is not observed. Also notable is the fact that the amidated protein (lane 4), which originally migrated ahead of the control (lane 1) as esterified protein (lane 2), now lags behind the corresponding control protein (lane 3) after treatment with ethylenediamine. The difference in migration rates between the control (lane 3) and reacted protein (lane 4) is significant enough to presume that amidation has occured. The fact that the amidated protein is observed as two distinct bands, and not smears, indicates that the yields of esterification and amidation must have been very good. As noted in the rationale section, treatment of the model peptide phenylalanyl-glycyl-glycine under identical conditions led to quantitative amidation.

The pretreatment of protein with pyridine remains optional as there is no evidence from initial investigations to suggest that protein amide bonds are being cleaved at serine or threonine ester bonds by diamine under the reaction conditions employed. Proteins appeared to be intact, with exception to the disulphide bonds, after treatment with ethylenediamine (Figure 4.3.5, lane 4). Either the N→O acyl migration is not occuring, or it is being reversed by the addition of pyridine or diamine without incident. This second possibility is not unreasonable since diamine, with an approximate concentration of 10 M, would be directly competing for the ester carbonyl carbon against the Nα-amino group of serine or threonine, the latter attempting a five-membered intramolecular ring cyclization reaction with an effective concentration many orders of magnitude

greater (Fersht, 1985; Creighton, 1993). Indeed, nonspecific transamidation and consequent degradation of protein is not observed (Figure 4.3.5, lanes 3&4; figure 4.3.7, lanes 6&9)

Using N,N-dimethylethylenediamine to amidate esterified protein gave different results than using ethylenediamine, however, they were no less encouraging. Most notable was the fact that N,N-dimethylethylenediamine does not cleave the disulphide bonds of α-chymotrypsin (Figure 4.3.7, lane 6&9). While it may appear as if the protein ester was never in contact with N,N-dimethylethylenediamine, the high molecular weight material trapped between the stacking and separating gels attests to the contrary and further suggests that minor crosslinking has occurred (lane 9). The amidated protein (lane 9) did not lag behind the control protein (lane 6) as was the case of protein derivatized with ethylenediamine (Figure 4.35, compare lanes 3&4). However, the gel appeared to be highly crosslinked (> 15%) and the migration of either α-chymotrypsin derivative was limited. It is difficult to conclude whether the amidated protein is laging behind the control, or even if it should be, given the circumstances. Compared to the esterified protein (Figure 4.3.7, lane 8) the amidated protein appears to have a different mobility. Once again, however, since the esterfied protein in lane 8 was not treated with N,N-dimethylethylenediamine, there may be a salt effect which perturbs the relative migrations of the protein in lanes 8&9. The only concrete evidence showing that carboxyl groups of protein were being amidated effectively with N,N-dimethylethylenediamine was from the diagonal experiment of variation 3, which showed that N-acetylated protein could not incorporate radioactivity into non C-terminal peptides (Figure 4.3.11; left) while N-acetylated and esterified protein did incorporate radioactivity (Figure 4.3.11; right).

The acrylamide gel (Figure 4.3.6) shows the profile of ethylenediamine derivatized protein following *in vacuo* ^{14}C-methylation. The results of the derivatization were difficult to interpret. At first glance, it appeared that the protein had been nonspecifically cleaved, generating false C-terminal peptides. This is supported by the fact that radioactivity is present along the entire length of the gel. The profile with Coomassie staining (not shown) was also identical. However, there is material trapped in the stacking gel and at the beginning of the separating gel which suggests that there are crosslinked protein fragments. If peptide bonds are being cleaved, insolubilities would not be expected. Futhermore, various amines and even methylated ethylenediamine are strongly stained with Coomasie blue and do not run off a gel as expected. Diamines were usually observed at the bottom one-third of a gel as a long smear (not shown). In some cases, amidated protein samples which were insufficiently evacuated and contained excessive diamine produced very blue backgrounds and showed evidence of salt effects. It was therefore necessary to look elsewhere in order to determine whether protein was being nonspecifically cleaved.

If false C-termini were being generated by heat promoted transamidation reactions, nonesterified control proteins, with free C-terminal carboxyl groups, should still have produced diagonal peptides along the C-terminal diagonal line (**1-2**). Unfortunately, a two-dimensional paper chromatogram had not been carried out for the control protein incubated with ethylenediamine (Variation 1), and it was therefore difficult to conclude with absolute certainty if any transamidation occured in the presence of residual ethylenediamine. While the success of the esterification and amidation steps cannot be denied, the outcome of the methylation step is less certain and it appears that residual diamine becomes problematic when protein is incubated at

elevated temperatures. In hindsight, more volatile diamine analogues such as 1,1-dimethyl-hydrazine and diaminomethane (Wang and Young, 1978) could potentially have been coupled cleanly to protein carboxyl groups. Dimethylhydrazine has a low boiling range of 62-64°C (Aldrich catalogue) and diaminomethane coexists in water with formaldehyde and 2 equivalents of volatile ammonia. Formaldehyde polymerizes to form paraformaldehyde upon dehydration but can still be removed using volatilizing conditions in which polymer and gaseous monomer coexist.

Peptide derivatives and two-dimensional paper electrophoresis.

Figures 4.3.8, 4.3.10, 4.3.11 and 4.3.12 show the diagonal chromatograms of variations 1 to 4, respectively. Diagonal line **1-2** of variation 1 (Figure 4.3.8) is undeniably strong and contains distinct peptides, however, as discussed previously, the peptides are likely to be false C-terminal peptides. More interesting are the strong peptides which lie off this C-terminal diagonal line, and which in themselves describe two other diagonal lines, **1-3** and **1-4**. Such non C-terminal diagonals were not observed before in aqueous derivatizations. The high specific activity which can be incorporated into peptides by nonaqueous methods offers a unique opportunity to examine other series of presumably related peptides. For example, diagonal line **1-3** may potentially represent a series of peptides which lose a net one positive charge in going from pH 2.1 to pH 4.4. If this is the case, the diagonal procedure would offer an incredibly powerful means of specifically isolating these peptides. Further, by using appropriate reagents and isotopes, and by controlling the LpH values of derivatization (chapter 3), it may be possible to investigate the reaction of just one type of amino acid of a protein. For example, methylation of protein at LpH 6 (lyophilized at pH 6) with excess [^{12}C]iodomethane, followed by trace methylation of the protein

at LpH 7.5 with [^{14}C]iodomethane may prove to be a way of specifically observing the reactivity of exposed N-terminal amino groups of protein.

Diagonals **1-2**, **1-3** and **1-4** were all subjected to paper chromatography. In each case, many of the peptides were resolvable by this method. Figure 4.3.9 shows the many methylation states of ethylenediamine (right side) and some faint spots in the more central areas. It is important to note that the experiment was begun with 1 nmole of starting material. Despite this, spots were observed, as in the case of figure 4.3.9, after only 1 week of autoradiography. This represents a substantial time-savings with respect to aqueous methods of derivatization, where radiolabeled reagents are normally diluted with regular reagent.

Variation 2 (Figure 4.3.10) was performed in the identical manner as variation 1, except that N,N-dimethylethylenediamine replaced ethylenediamine. It was hoped that this reagent would prove more easily removed under vacuum than its predecessor and pose less of a problem during the methylation step. Unfortunately, a nonesterified control protein was subjected to the remaining derivatization steps and produced a C-terminal diagonal line (**1-2**) that looked similar to that of the esterified protein. This indicated that false C-terminal peptides were generated. As observed, diamine was present during the methylation step. The very strong spot at one end of the diagonal corresponded to a methylation product of N,N-dimethylethylenediamine (Figure 4.3.10, top of diagonal line **1-2**). Non C-terminal diagonal lines (**1-3** and **1-4**) were also observed in the chromatogram of esterified protein (Figure 4.3.10) and nonesterified protein (not shown). Again, these other diagonal lines probably resulted from the reaction of iodomethane with any or all of the histidyl, tyrosyl, methionyl and lysyl side-chains of protein, and the potential applications of this finding should be further investigated.

Thus far, the isolation of a C-terminal peptide could not be confirmed by paper electrophoresis, due to the cleavage of protein chains by heat promoted transamidation reactions. Variations 3 and 4 were implimented to remove any risk of generating false C-termini. To achieve this, amino groups of the reaction mixture were acetylated prior to methylation. However, the results of variation 3 again showed that unblocked N,N-dimethylethylenediamine must have been present during the methylation step (Figure 4.3.11, diagonal line **1-2**). The three spots near the top are believed to represent various methylation states of N,N-dimethylethylenediamine, while the intense spot at the mid-point of the diagonal is believed to represent N,N,N-trimethyl-N'-acetylethylenediamine. The results suggest that either diamine is adsorbed and protected by the protein prior to acetylation, or acetylated diamine is unblocked under the conditions used for methylation. The profile of C-terminal diagonal lines (Figure 4.3.11; **1-2**) for esterified (right) and nonesterified (left) protein are similar, again indicating that C-terminal peptides are being falsely generated. Unlike variation 2, however, non C-terminal diagonal lines (**1-3** and **1-4**) of the control protein (left) were absent. This is consistent with the fact that protein lysyl residues should be acetylated and not capable of incorporating radioactivity. If this is true, diagonal lines **1-3** and **1-4** probably represent two different series of peptides containing amidated aspartic or glutamic acid residues. Further work is required to confirm this claim.

The methylation reaction of variation 4 was carried out at room temperature to prevent any possible generation of false C-terminal peptides. Furthermore, the protein was digested prior to methylation so that all residual N,N-dimethylethylenediamine would be acetylated. Unlike variations 1, 2 and 3, variation 4 had the advantage that the entire protocol could be conveniently

carried out in one Eppendorf tube. Unfortunately, the paper chromatogram of variation 4 (not shown) only appeared to contain a radioactive neutral spot and radioactive diagonal markers. These markers produced spots that were at least as intense as those observed in previous variations. With no apparent reason to explain the absence of radiolabeled peptides, it is believed that the methylated protein derivative was not proteolyzed.

Variation 5 (Figure 4.3.12) is the shortened version of variation 4 and was introduced when it was observed how easily leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ could be sequenced by tandem made spectrometry (Figure 4.3.4 and table 4.3.1). It was noted that the sensitivity of detection of this peptide was approximately the same as that of the quaternized ammonium analogue, leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N^+(CH_3)_3$. Furthermore, it was observed that the N,N-dimethylethylenediamine ion could easily be formed in the collision chamber and that dimethylamine was just as readily lost under the conditions of mass analysis as trimethylamine. These revelations suggested that it may not be necessary to derivatize the peptide with iodomethane. It was reasoned that variation 5 would avoid additional steps, allow the experiment to be carried out in a single Eppendorf tube, and most importantly, prevent the heat induced cleavage of protein by diamine. The disadvantage of this method was that individual peptides would no longer be seen by autoradiography. The success of variation 5 therefore required that the diagonal line be identified using markers, and the eluted peptides be separated by some other means. This made it necessary to use high performance liquid chromatography in series with tandem mass spectroscopy.

As observed, the chromatogram of variation 5 (Figure 4.3.12) had two apparent C-terminal diagonal lines (**1-5** and **1-6**). The markers, namely hexamethylethylenediamine (5) and

N,N-dimethylethylenediamine (6), did not describe the same line. Such anomalous mobility was observed previously with the compound acetyl-valyl-arginyl-arginyl-7-amido-4-methylcoumarin, of chapter 5 (not shown). In fact, no two of the three compounds described lay on a common diagonal line. The correct choice of diagonal marker therefore required some reasoning. The N,N-dimethylethylenediamine marker became very elipsoid when electrophoresed in the pH 4.4 direction. This immediately cast some doubt as to its reliability. Furthermore, it was shown that the peptides which lay along diagonal line **1-2** of variations 1, 2 and 3 were false C-terminal peptides, but peptides nonetheless. For this reason, it was safest and most consistent to choose the hexamethylethylenediamine marker to deliniate the C-terminal diagonal line of variation 5 (Figure 4.3.12; 4, excised diagonal). The diagonal strip was eluted to recover the peptides. Mass analyses of the contents of the diagonal strip (Figure 4.3.14) were inconclusive, in sharp contrast to the unambiguous results obtained for the peptide standard leucyl-leucyl-leucyl-tyrosyl-$NHCH_2CH_2N(CH_3)_2$ (Figure 4.3.13). The weak signals obtained (Figure 4.3.14) may be due to some or all of the following reasons, but only the first explanation is likely: (i) The amount of protein loaded was small in comparison to interferents normally coeluted from the Whatman paper. (ii) The wrong diagonal marker was chosen to recover the C-terminal peptides. (iii) Amidation and/or proteolysis of α-chymotrypsin was unsuccessful. With respect to points (i), (ii) and (iii), an internal standard containing a reactable carboxyl group could be introduced in the beginning stages to monitor the course of derivatization. This is especially important since verifying the presence of peptides on paper is not straightforward in variation 5, unlike variations 1 to 4, which take advantage of radioactivity.

Quite clearly, the problems associated when analyzing such small amounts of protein in the presence of trace interferents are still formidable. The success of microscale C-terminal analysis

may require the exclusion of two-dimensional paper electrophoresis, or the implementation of a post-chromatography purification protocol. This work has been continued by François R. Clairmont, who has modified the procedure for larger amounts of protein and has isolated the C-terminal peptide of the insecticidal protein from *Bacillus thuringiensis*, an achievement which attests to the merit of this procedure and encourages its continued development.

4.5. Conclusions

The investigation suggests that C-terminal peptides of protein can potentially be isolated by a nonaqueous chemical approach. Attempts to label carboxyl groups of protein by aqueous chemical methods had shortcomings when applied to very little amounts of material, however, nonaqueous derivatization of protein carboxyl groups was elegant, technically simple and applicable to minute amounts of protein. By avoiding chromatography or dialysis, product yields were clearly improved. The chemicals and reagents chosen made the protocol feasible for even small scale reactions. Esterification and subsequent amidation procedures were successfully carried out without observing significant protein degradation and crosslinking. The diamine used to label protein carboxyl groups fragmented predictably under the conditions of mass analysis, and potentially provides a means of characterizing individual C-terminal peptide derivatives by tandem mass spectral methods. Unfortunately, *in vacuo* methylation of protein could not be carried out without suffering degradation. Two last variations in protocol attempted to ameliorate the problem, however, the outcome was either inconclusive or indicated that more work was required. The use of diagonal electrophoresis can be extended to specifically isolate any derivatized protein functional group.

4.6 Acknowledgment

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) and an NSERC scholarship to Alpay Taralp.

4.7 References

Akabori, S., Ohno, K. and Narita, K. (1952) *Bull. Chem. Soc. Jap.* **25**, 214-218.

Ambler, R. P. (1972) In *Meth. Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 262-272, Academic Press.

Anderegg, R.J., Carr, S.A., Huang, I.Y., Hiipakka, R.A., Chang, C., and Liao, S. (1988) *Biochemistry* **27**, 4214-4221.

Bateman, K.P., White, R.L. and Thibault, P. (1997) *Rapid Communications in Mass Spectrometry* **11**, 307-315.

Biemann, K. (1992) *Annu. Rev. Biochem.* **61**, 977-1010.

Biemann, K. and Scoble, H.A. (1987) Science 237, 992-998.

Bodanszky, M. and Bodanszky, A. (1994) *The Practice of Peptide Synthesis*, 2nd ed., Springer-Verlag.

Callahan, P. X., McDonald, J. K. and Ellis, S. (1972) In *Meth. Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 282-298, Academic Press.

Capriolli, R.M. and Suter, M., J.-F. (1995) in *Introduction to Biophysical Methods for Protein and Nucleic Acid Research* (Glasel, J.A. and Deutscher, M.P. Eds.), Ch. 4, Academic Press.

Chibnall, A.C., Mangan, J.L., and Rees, M.W. (1958) *Biochem. J.* **68**, 114-118.

Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2nd ed., W. H. Freeman and Company.

Darbre, A. (1986) *Practical Protein Chemistry: A Handbook*, John Wiley and Sons.

Doscher, M.S. and Wilcox, P.E. (1961) *J. Biol. Chem.* **236**, 1328.

Duggleby, R. G. and Kaplan, H. (1975) *Anal. Biochem.* **65**, 346-354.

Fersht, A. (1985) *Enzyme Structure and Mechanism*, 2nd ed., W. H. Freeman and Company.

Fraenkel-Conrat, H. and Olcott, H.S. (1945) *J. Biol. Chem.* **161**, 259.

Frieden, E.H. (1956) *J. Am. Chem. Soc.* **78**, 1961.

Fromageot, C. Jutisz, M., Meyer, D. and Penasse, L. (1950) *Biochim. Biophys. Acta* **6**, 283-289.

Gibson, B.W., Yu. Z., Aberth, W., Burlingame, A.L. and Bass, N.M. (1988) *J. Biol. Chem.* **263**, 4182-4185.

Gray, W. R. (1972) In *Meth. in Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 333-344, Academic Press.

Hames, B. D. and Rickwood, D. (1990) *Gel Electrophoresis of Proteins: A Practical Approach*, 2nd ed., IRL Press.

Horn, M. J., Mayhew, J. W. and O'Dea, K. C. (1995) Abstract 531-M, *Protein Science, Ninth symposium abstracts* **4, suppl. 2**, 155.

Iwai, K. and Ando, T. (1967) In *Meth. Enzymol.* **11** (Hirs, C.H.W., Ed.) Academic Press.

Kaplan, H. and Taralp, A. (1997) In *Techniques in Protein Chemistry VIII* (Marshak, D. Ed.) pp. 219-230, Academic Press.

Kelly, J.F., Ramaley, L., and Thibault, P. (1997) *Anal. Chem.* **69**, 51-60.

Knapp, D. R. (1990) In *Meth. Enzymol.* **193** (McCloskey, J. A. Ed.) pp. 314-329, Academic Press.

Konigsberg, W. (1972) In *Meth. Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 326-332, Academic Press.

Lee, E. D., Henion, J. D. and Covey, T. R. (1989) *J. Microcolumn Sep.* **1**, 14.

Leggett Bailey, J. (1962) *Techniques in Protein Chemistry*, pp. 61-62, Elsevier Publishing Company.

Lenard, J. and Hess, G.P. (1964) *J. Biol. Chem.* **239**, 3275-3281.

Light, A. (1972) In *Meth. Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 253-262, Academic Press.

March, J. (1985) *Advanced Organic Chemistry: Reactions, Mechanisms and Structure*, 3rd ed., John Wiley and Sons.

Matsubara, H. and Feder, J. (1971) In *The Enzymes*, 3rd ed., **3** (Boyer, P. D. Ed.) pp. 721-791, Academic Press.

McCloskey, J. A. (1990) *Meth. Enzymol.* **193**, Academic Press.

Means, G. E.and Feeney, R. E. (1971) *Chemical Modification of Proteins*, Holden Day.

Olivares, J.A., Nguyen, N.T., Yonker, C.R., and Smith, R.D. (1987) *Anal. Chem.* **59**, 1232-1236.

Patterson, D. H., Tarr, G. E., Regnier, F. E. and Martin, S. A. (1995) *Anal. Chem.* **67**, 3971-3978.

Phillips, A.P. and Baltzly, R. (1947) *J. Am. Chem. Soc.* **69**, 200-204.

Richards, F. F. and Lovins, R. E. (1972) In *Meth. Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 314-326, Academic Press.

Schroeder, W. A. (1972) In *Meth. Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 298-313, Academic Press.

Siuzdak, G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11290-11297.

Smith, R.D., Wahl, J.H., Goodlett, D.R., and Hofstadier, S.A. (1993) *Anal. Chem.* **65**, 574-578.

Stark, G. R. (1972) In *Meth. Enzymol.* **25B** (Hirs, C. H. W. and Timasheff, S. N. Eds.) pp. 369-384, Academic Press.

Stryer, L. (1981) *Biochemistry*, 2nd ed., pp. 21-27, Freeman and Company.

Stults, J.T., Lai, J., McCune, S., and Wetzel, R. (1993) *Anal. Chem.* **65**, 1703-1708.

Taralp, A. and Kaplan, H. (1997) *J. Prot. Chem.* **16**, 183-193.

Treston, A., Kasprzyk, P., Cuttitta, F., Covey, T. and Mulshine, J. (1989) *Proc. 37th ASMS Conf. Mass Spectrom. Allied Topics*, p. 889.

Tsuji, K., Baczynskyj, L., and Bronson, G.E. (1992) *Anal. Chem.* **64**, 1864-1870.

Vath, J. E., Zollinger, M., and Biemann, K., (1988) *Fresenius Z. Anal. Chem.* **331**, 248-252.

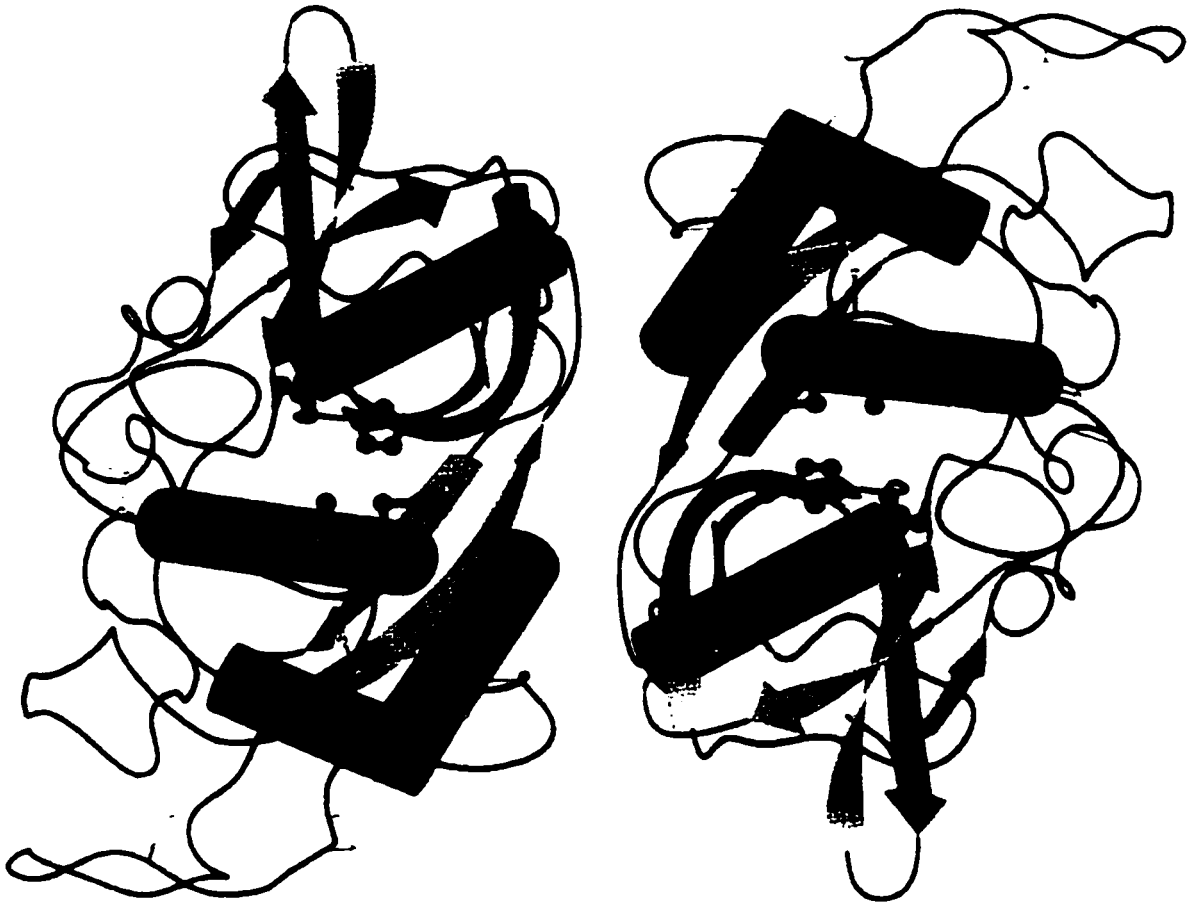
Wang, T.-T. and Young, M. N. (1978) *Anal. Biochem.* **91**, 696-699.

Ward, C. W. (1986) In *Practical Protein Chemistry: A Handbook* (Darbre, A. Ed.) pp. 491-525, John Wiley and Sons.

Waterfield, M. D., Scrace, G. and Totty, N. (1986) In *Practical Protein Chemistry, A Handbook* (Darbre, A. Ed.) pp. 411-426, John Wiley and Sons.

Winter, G. (1986a) In *Practical Protein Chemistry, A Handbook* (Darbre, A. Ed.) pp. 345-366, John Wiley and Sons.

Winter, G. (1986b) In *Practical Protein Chemistry, A Handbook* (Darbre, A. Ed.) pp. 367-374, John Wiley and Sons.



Cathepsin B dimer in unit cell

Chapter 5: Characterizing cathepsin B substrate specificity

5.1 Introduction

Lysosomal cathepsin B (EC 3.4.22.1), a cysteine proteinase normally active in protein turnover, has been implicated in several pathological conditions, including arthritis, muscular dystrophy and tumor metastasis (Mort *et al.*, 1984; Gopalan *et al.*, 1987; Katunuma and Kominami, 1987; Lah *et al.*, 1989; van Noorden *et al.*, 1989; Sloane, 1990; Sloane *et al.*, 1990; Moin *et al.*, 1992a; Moin *et al.*, 1992b; Buck *et al.*, 1992; Sinha *et al.*, 1993; Campo *et al.*, 1994; Cao *et al.*, 1994; Castiglioni *et al.*, 1994; Emmert-Buck *et al.*,1994; Honn *et al.*, 1994; Rempel *et al.*, 1994; Rozhin *et al.*, 1994; Sloane *et al.*, 1994a; Sloane *et al.*, 1994b; Visscher *et al.*, 1994; Berquin *et al.*, 1995; Calkins and Sloane, 1995; Mikkelsen *et al.*, 1995; Ryan *et al.*, 1995; Sinha *et al.*, 1995a; Sinha *et al.*, 1995b; Keppler and Sloane, 1996; Ren *et al.*, 1996). It is unique among the cysteine proteinases in that it has both endopeptidase and exopeptidase activities, can accept basic residues in a substrate at both the P_1 and P_2 positions and has complex pH dependencies (Barrett and Kirschke, 1981; Willenbrock and Brocklehurst, 1985; Pohl *et al.*, 1987; Khouri *et al.*, 1991). The detailed analysis of the pH dependence of cathepsin B catalyzed hydrolyses has revealed that at least seven groups which dissociate in the pH range of 3 to 9 can affect substrate binding and/or turnover (Hasnain *et al.*, 1992).

Kinetic analyses with N-terminal protected monopeptidyl and dipeptidyl substrates or various irreversible inhibitors demonstrated that the S_1 subsite prefers positively charged or straight chained aliphatic amino acids but exhibits a very poor affinity for negatively charged or bulky aromatic and branched aliphatic amino acids (Barrett and Kirschke, 1981; Crawford *et al.*, 1988; Wikstrom *et al.*, 1989; Krantz *et al.*, 1991; Pliura *et al.*, 1992). The S_2 subsite was shown to exhibit a preference for

phenylalanine and arginine (Figure 5.1.1, top), although the specificity constant, k_{cat}/K_m, is seven-fold higher for the former (Hasnain *et al.*, 1992). Interpretation of pH activity profile data, obtained with substrates containing an arginine at P_1 and either an arginine or phenylalanine at P_2, suggested the presence of carboxyl groups in the S_1 and S_2 subsites with pK_a values of 5.4 and 5.1 respectively (Hasnain *et al.*, 1992). The identity of the group with the pK_a of 5.1 was confirmed to be Glu^{245} by kinetic characterization of site-directed mutants, Glu^{245}-Gln^{245} and Glu^{245}-Ala^{245} (Hasnain *et al.*, 1993).

In contrast to the understanding of the S_1 and S_2 subsite specificities, S_3 specificity remains largely uncharacterized. Brömme *et al.* (1987) have shown that the binding of a series of N-terminal succinylated alanine-containing peptide substrates could be improved by increasing the length of the peptide chain. When comparing tripeptide and dipeptide alanine substrates, the former demonstrated a two-fold increase in the specificity constant k_{cat}/K_m. In addition, the S_3 subsite was shown to accept proline, unlike subsites S_1 or S_2 (Brömme *et al.*, 1987). Brömme *et al.* (1989) also examined S_3 and S_4 specificity using six tripeptide and tetrapeptide coumaryl substrates. For two of the substrates with valine and phenylalanine as the P_3 residue, they found no significant difference in binding energy. However, since the N-terminal blocking groups were not the same for the substrates and the P_3 residues selected did not have a sufficiently wide range of physico-chemical properties, the optimal specificity of the S_3 subsite remained unresolved. In a different approach, Koga *et. al.* (1991) digested soluble denatured proteins with cathepsin B. Fragments arising from endopeptidase activity were analyzed and interpretation of the results suggested a preference for the amino acids glycine, tryptophan, alanine, lysine, isoleucine and proline at the P_3 position of the substrate. Although only qualitative, the results, excluding the observation of lysine at P_3, supported the possibility of a hydrophobic pocket at the S_3 subsite.

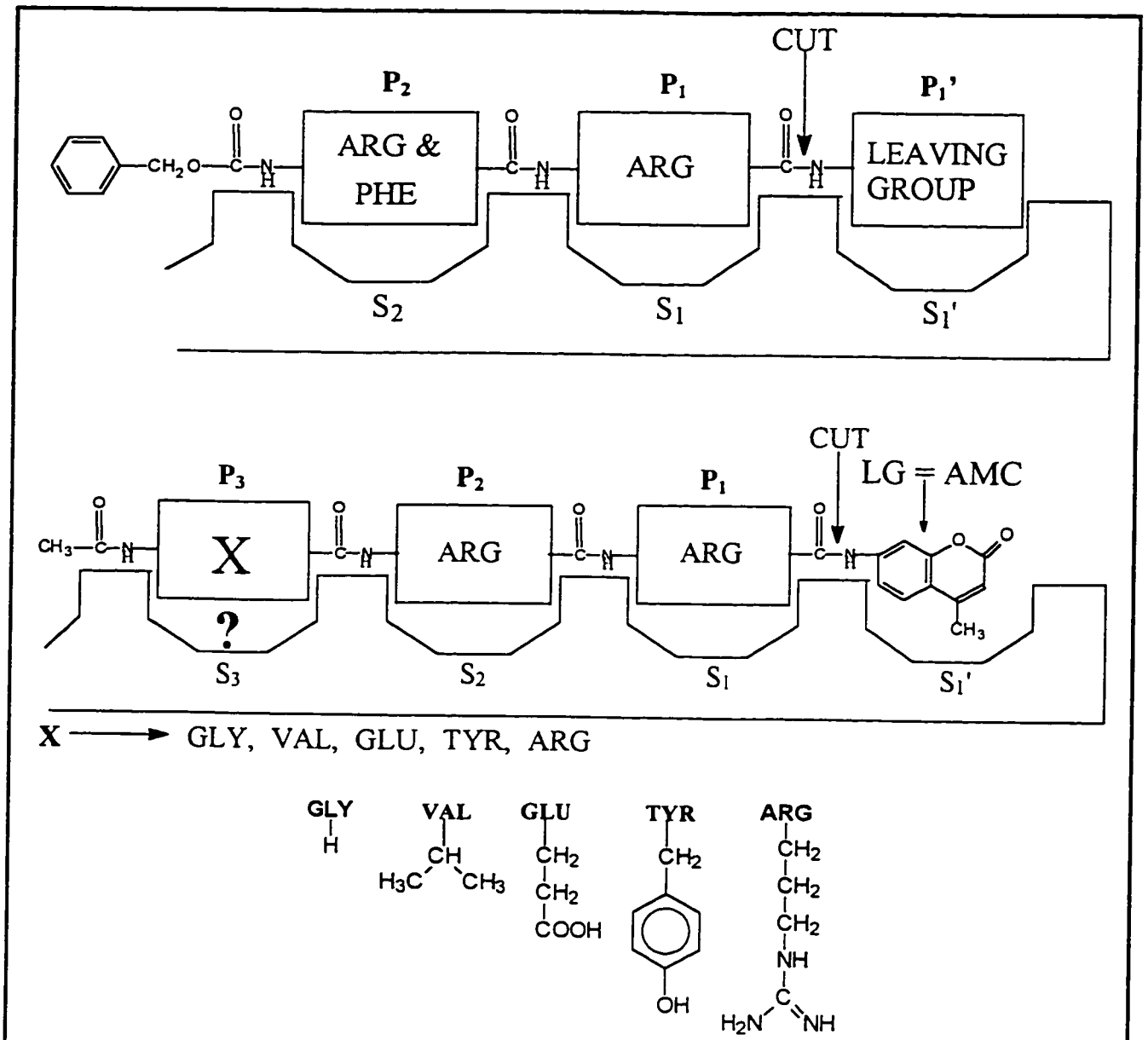


Figure 5.1.1: Schematic representation of the binding mode of substrates to cathepsin B. (*top*) Benzyloxycarbonyl dipeptidyl substrates with either arginine or phenylalanine bound to the S_2 pocket were previously shown to be good substrates of cathepsin B; (*bottom*) Acetyl tripeptidyl substrates with glycine, valine, glutamic acid, tyrosine or arginine were used to help define S_3 specificity. Leaving group (LG) = 7-amino-4-methylcoumarin (AMC)

Focus of the present research

In this chapter we describe the kinetic analysis of substrates of the type acetyl-X-Arg-Arg-AMC (where X is Gly, Glu, Arg, Val, or Tyr and where AMC represents 7-amido-4-methylcoumarin). The

side-chain residue of each substrate was chosen to help define specificity at the S_3 subsite (Figure 5.1.1, bottom). In addition, the difference in free energy change of binding ($\Delta\Delta G$ values) for these substrates was calculated to obtain the relative strengths of binding of the different P_3 side chains in the S_3 subsite. Finally, molecular modeling studies were carried out for tripeptide substrates containing Glu (in its protonated and deprotonated states), Arg and Tyr in the P_3 position in order to further clarify these enzyme-substrate interactions.

5.2 Materials and Methods

5.2.1 Enzyme expression and purification. The cDNA for rat cathepsin B was expressed in *Saccharomyces cerevisiae* as an α–factor fusion construct (Rowan *et al.*, 1992). Yeast starter cultures were grown in synthetic medium and then were grown in four litre shake flasks as reported before (Roowan *et al.*, 1992; Ernst, 1986). The recombinant enzyme was harvested and purified as previously described (Hasnain *et al.*, 1992).

5.2.2 Enzyme assays for the determination of kinetic constants. Prior to kinetic analysis, the enzyme concentration was determined by active site thiol titration (Figure 5.2.1) using the cysteine proteinase inhibitor E-64 (*N*-[*N*-(*L*-3-*trans*-carboxyoxiran-2-carbonyl)-*L*-leucyl]-agmatine) (Boehringer Mannheim Canada) according to the methodology of Barrett and Kirschke (1981).

Just before conducting experiments for the determination of kinetic constants, the enzyme, in 50 mM sodium phosphate buffer, pH 6.0, 1 mM ethylenediamine tetraacetic acid, was activated for one hour by the addition of dithiothreitol to a final concentration of 10 mM, and then kept on ice (Hasnain *et al.*, 1992). The active enzyme concentration was shown to be unchanged for the duration of the

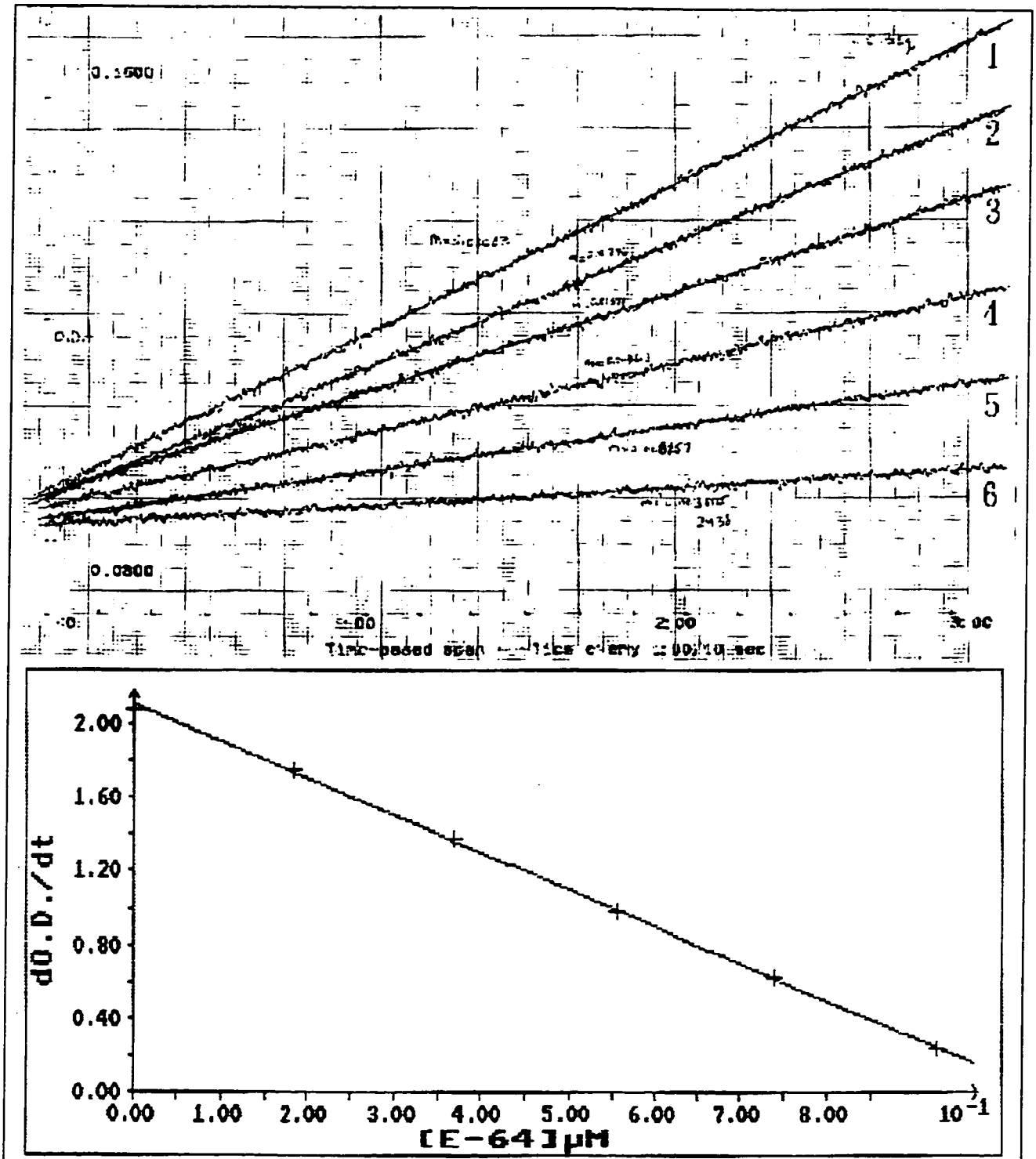


Figure 5.2.1: Obtaining the active enzyme concentration. (*top*) Spectrophotometric assay at 410 nm of cathepsin B activity in the presence of varing amounts of E-64: **1**, 0 μmole; **2**, 2 X 10^{-6} μmole; **3**, 4 X 10^{-6} μmole; **4**, 6 X 10^{-6} μmole; **5**, 8 X 10^{-6} μmole; **6**, 1 X 10^{-5} μmole. (*bottom*) Initial rate data of the hydrolysis of 20 mM ZRRpNA extrapolated to zero (dO.D./dt = 0) describes the concentration of E-64 which exactly titrates the active enzyme present.

experiments. Michaelis-Menten kinetic parameters for the coumaryl substrates were determined at 24°C by continuous fluorescence spectrophotometry (Figure 5.2.2), using an SLM Aminco DW2000 spectrophotometer equipped with the Total Fluorescence Accessory (Hasnain *et al.*, 1992). All data were fitted by the nonlinear regression data analysis program Enzfitter, of Leatherbarrow, supplied by Elsevier-Biosoft.

Assays were performed in 25 mM sodium phosphate buffer, 250 mM NaCl, 1 mM ethylenediamine tetraacetic acid, 3% dimethyl sulphoxide, pH 6.0 or 7.7, and in 25 mM sodium citrate buffer, 250 mM NaCl, 1 mM ethylenediamine tetraacetic acid, 3% dimethyl sulphoxide, pH 4.0. The pH activity profiles were determined for two substrates, acetyl-Val-Arg-Arg-AMC and acetyl-Arg-Arg-Arg-AMC. These pH activity profiles were determined over the pH range 3.2-8.4 by measuring k_{cat}/K_m at 0.1 pH unit intervals, using the relationship $k_{cat}/K_m=v/([E][S])$ when $[S]<<K_m$. Assay samples were monitored to ensure that pH was stable over the course of the reaction. Over the pH range 3.2-6.0, the reaction buffer consisted of 25 mM sodium citrate, 250 mM NaCl, 1 mM ethylenediamine tetraacetic acid, and 3% dimethyl sulphoxide. In the pH range 5.7-7.7, the reaction buffer was 25 mM sodium phosphate, 250 mM NaCl, 1 mM ethylenediamine tetraacetic acid and 3% dimethyl sulphoxide, and in the pH range of 7.7-8.4, the reaction buffer was 25 mM sodium borate, 250 mM NaCl, 1 mM ethylenediamine tetraacetic acid and 3% dimethyl sulphoxide.

The citrate buffer consistently afforded lower activity measurements when compared to the phosphate buffer through the pH range 5.7-6.0, and for this reason, there was a break in the continuity of the data of the pH activity profile. Similar results were observed by Hasnain *et al.* (1992) for the substrates benzyloxycarbonyl-Arg-Arg-X (where X is 7-amido-4-methylcoumarin or *p*-nitroanilide). The data in citrate buffer were corrected by a factor determined from measurements in phosphate

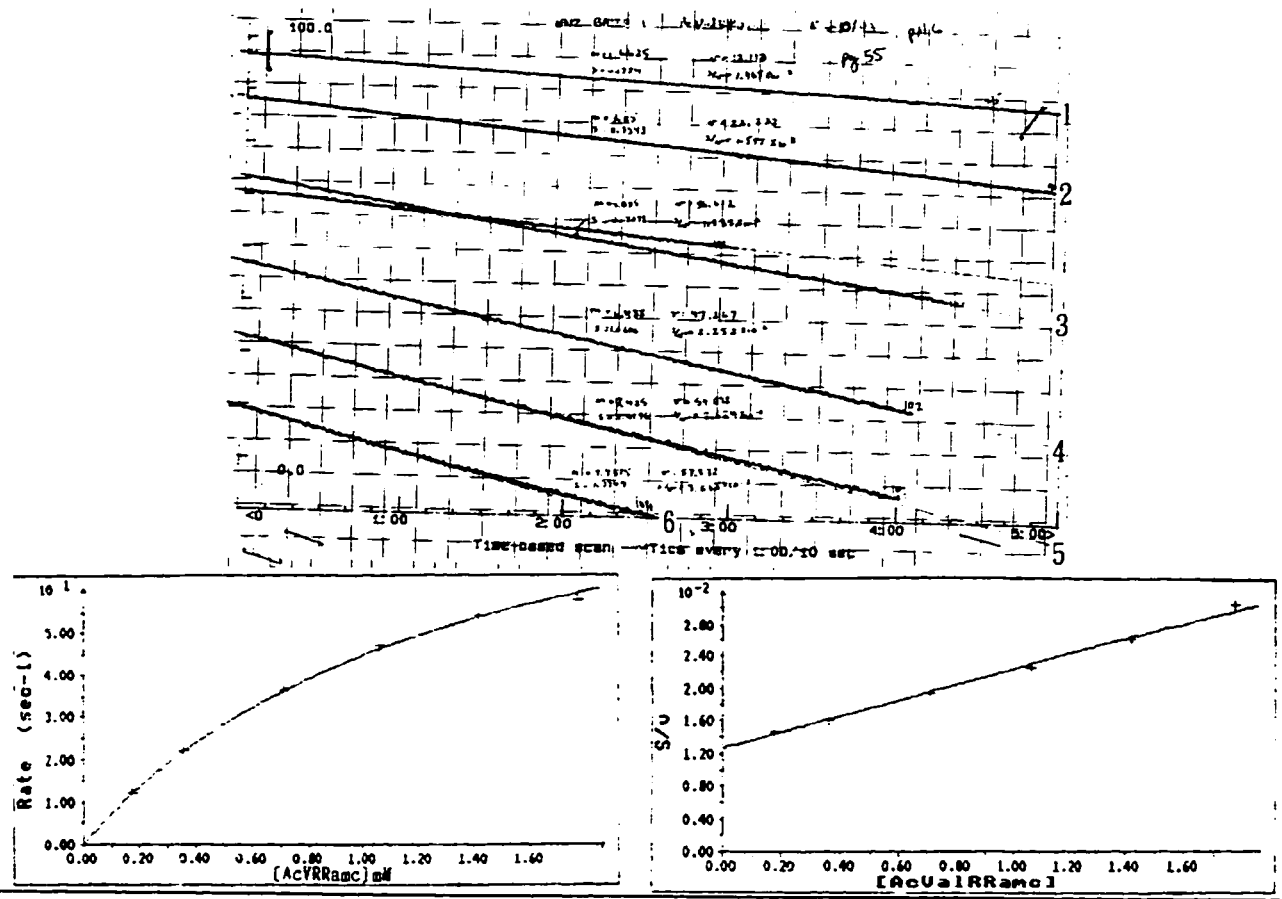


Figure 5.2.2: Sample enzyme assay using acetyl-Val-Arg-Arg-AMC as substrate at pH 6.0. (*top*) Initial rate data acquired using different substrate concentrations: (1) 0.177mM (2) 0.355mM (3) 0.710mM (4) 1.06mM (5) 1.42mM (6) 1.77mM. $\lambda_{excitation}$=370nm. (*bottom left*) Michaelis-Menten kinetics profile with rate (v) plotted against substrate concentration. (*bottom right*) Hanes-Woolfe plot for extraplotating the value of S/v.

buffer at four separate pH intervals (5.7-6.0) where citrate and phosphate buffers overlapped, as described by Hasnain *et al.* (1992). For the tripeptide substrates in this investigation, a similar effect was evident at pH 7.7, where borate and phosphate buffers overlapped, with greater activity in borate buffer. It was previously determined (Hasnain *et al.*, 1992) that pH 7.7 was the only value in the pH range where both phosphate and borate buffers could be made to overlap under the ionic strength conditions used in the experiment. Therefore, the following strategy was used to correct for this buffer effect. The enzyme activity measurements at pH 7.7 for the substrate acetyl-Val-Arg-Arg-AMC in each buffer system were carefully repeated in triplicate and were shown to be highly reproducible. By

averaging the enzyme activities in each buffer system at pH 7.7 and by calculating the ratio of these averaged activities for both buffer systems, a normalizing factor with small standard deviation was determined. Graphical analysis of the corrected data for two independent experiments using the substrate acetyl-Val-Arg-Arg-AMC and for one experiment using the substrate acetyl-Arg-Arg-Arg-AMC demonstrated that the profiles gave a best fit to the four proton ionization model of Hasnain *et al.* (1992), where dipeptide AMC substrates with an arginyl residue in the P_2 position were employed. The agreement with previous work supported the validity of this correction for the borate buffer.

5.2.2 Substrate synthesis. Substrates were provided by Isidoros Vlattas and coworkers of Ciba-Geigy Corporation. While many blocking groups and coupling reagents were available (March, 1985; Atherton and Sheppard, 1989; Bodanszky and Bodanszky, 1994) two approaches were decided upon. The peptides acetyl-Arg-Arg-Arg and acetyl-Val-Arg-Arg were synthesized on a Sasrin resin solid support by the employment of standard 9-fluorenylmethoxycarbonyl group methodology. The side chain functionality of *Na*-(9-fluorenylmethoxycarbonyl)-arginine was blocked with the 2,2,5,7,8-pentamethylchroman-6-sulphonyl group. Amino acids were coupled using 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate in *N*-methyl pyrrolidone. After being cleaved from the resin with 1% trifluoroacetic acid in dichloromethane, the peptides were coupled to AMC with 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride in N,N-dimethylformamide and dichloromethane, once again in the presence of 1-hydroxybenzotriazole. Substrates were deblocked by treating with 95% trifluoroacetic acid and 5% water.

The substrates acetyl-Arg-Arg-AMC, acetyl-Gly-Arg-Arg-AMC, acetyl-Glu-Arg-Arg-AMC and acetyl-Tyr-Arg-Arg-AMC were prepared in solution by the employment of *Na-tert*-butoxycarbonyl amino acids. In contrast to the solid support method, the initial syntheses of these substrates began

with the making of Arg-AMC and the coupling reaction was carried out in pyridine with dicyclohexylcarbodiimide. The side chain functionalities of Arg, Tyr and Glu were protected as *p*-toluenesulphonyl, 2-bromobenzyloxycarbonyl, and benzyl derivatives, respectively. Peptide coupling reactions with dicyclohexylcarbodiimide were generally achieved in a double solvent system. Typically, dichloromethane was used to dissolve the *tert*-butoxycarbonyl amino acid and coupling agent. To this was combined the growing AMC peptide, previously dissolved in *N,N*-dimethylformamide. Racemization was minimized by the employment of 1-hydroxybenzotriazole. Sequential removal of the *tert*-butoxycarbonyl group required treatment with 50% trifluoroacetic acid in dichloromethane at room temperature. Substrates were subsequently acetylated with acetic anhydride in pyridine. Deprotection of the side groups was effected by treatment with a 9:1 (v/v) mixture of hydrofluoric acid/anisole.

All substrates were purified by reverse phase high performance liquid chromatography using varied linear gradients between 0 and 60% acetonitrile, in 0.1% (v/v) trifluoroacetic acid (Figure 5.2.2). Substrate composition and purity were verified by amino acid, mass spectral and [^{1}H]NMR spectral analyses (Figures 5.2.3, 5.2.4 and 5.2.5).

5.2.3 Evidence supporting exclusive hydrolysis of substrates at the AMC peptide bond.

Exclusive hydrolysis of substrates at the AMC peptide bond was established by using high voltage paper electrophoresis to separate the products after incubation of substrates with cathepsin B (Figure 5.2.6). Since 7-amino-4-methylcoumarin is ninhydrin negative, the basic premise of this strategy was that cleavage at any site, other than the AMC peptide bond, would generate a free α-amino group which would yield a ninhydrin positive peptide. Substrates were incubated with cathepsin B in pH 6.5 buffer of the composition acetic acid:pyridine:water (3:100:1800 by volume). Reaction completion

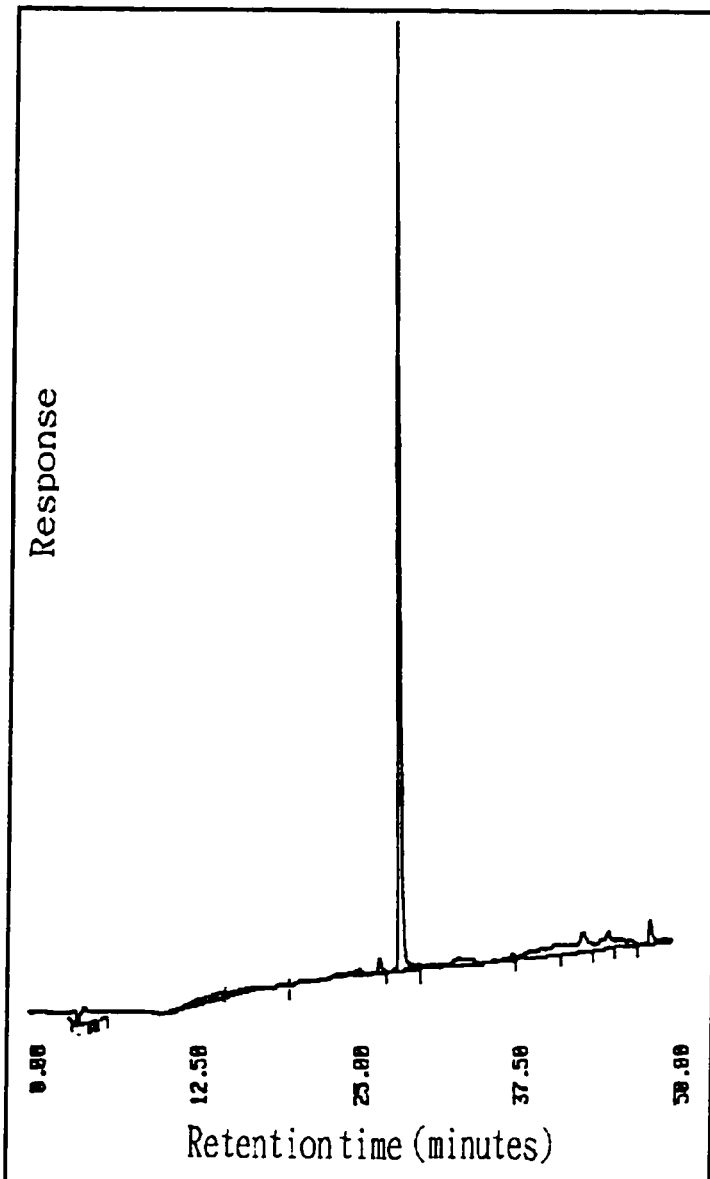


Figure 5.2.2: Final purification of the substrate acetyl-Val-Arg-Arg-AMC by reverse phase high performance liquid chromatography. The compound (retention time = 28.61 min) was collected and examined as indicated in the text.

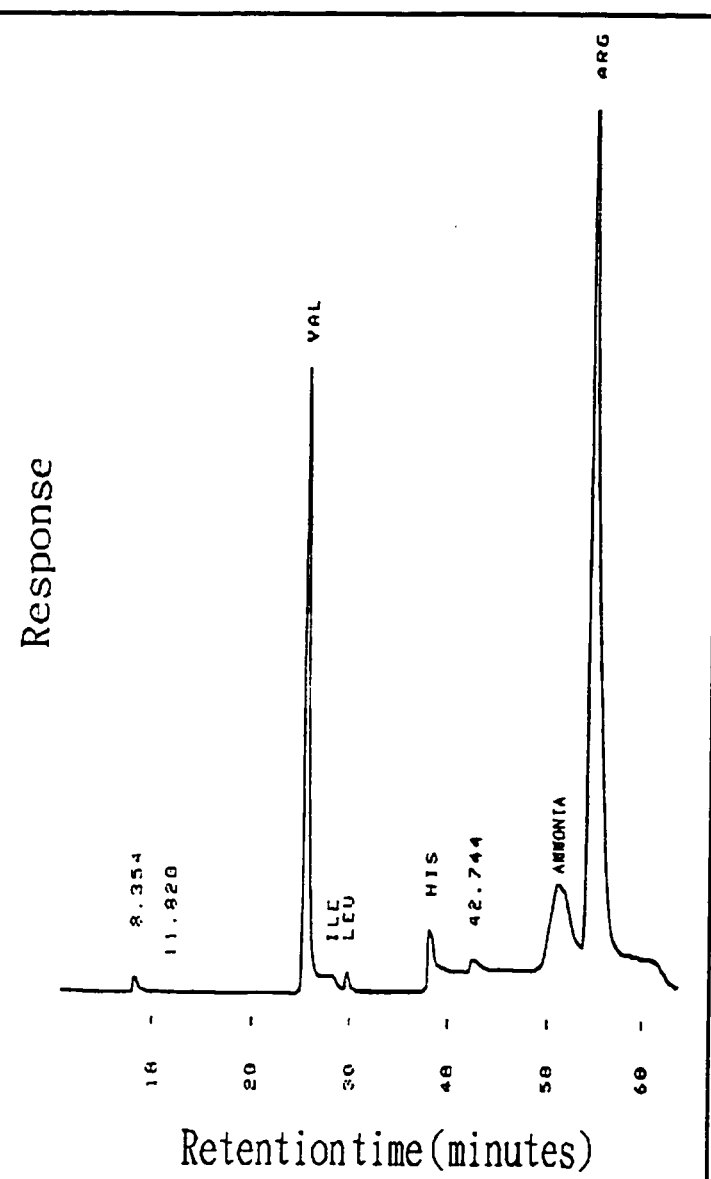
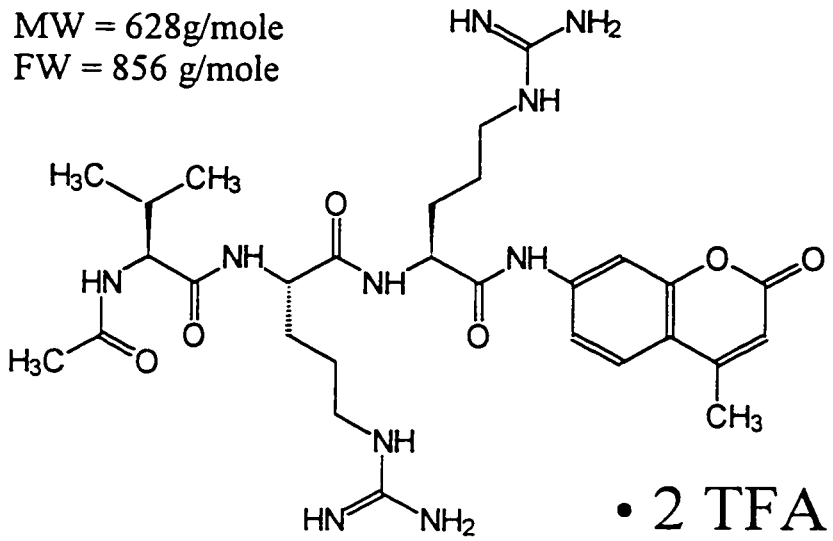


Figure 5.2.3: Amino acid analysis of the substrate acetyl-Val-Arg-Arg-AMC. An aliquot was dissolved in 6N HCl and hydrolyzed. Amino acid analysis was carried out with ninhydrin detection. Valine and arginine were observed in the ratio 1:2.



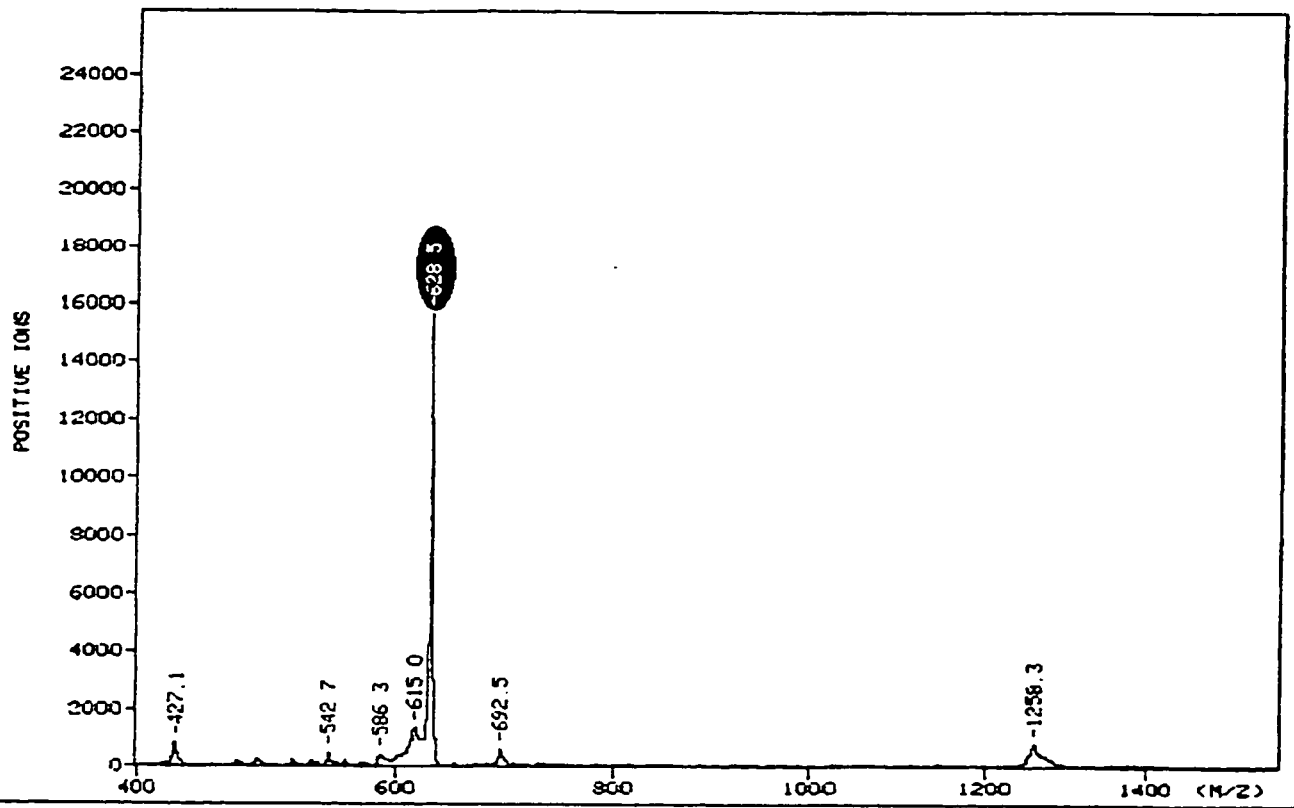


Figure 5.2.4: Plasma desorption mass spectral analysis of the substrate acetyl-Val-Arg-Arg-AMC using time of flight detection. The predominant $(M+H)^+$ ion at 628.5m/z corresponds to the substrate.

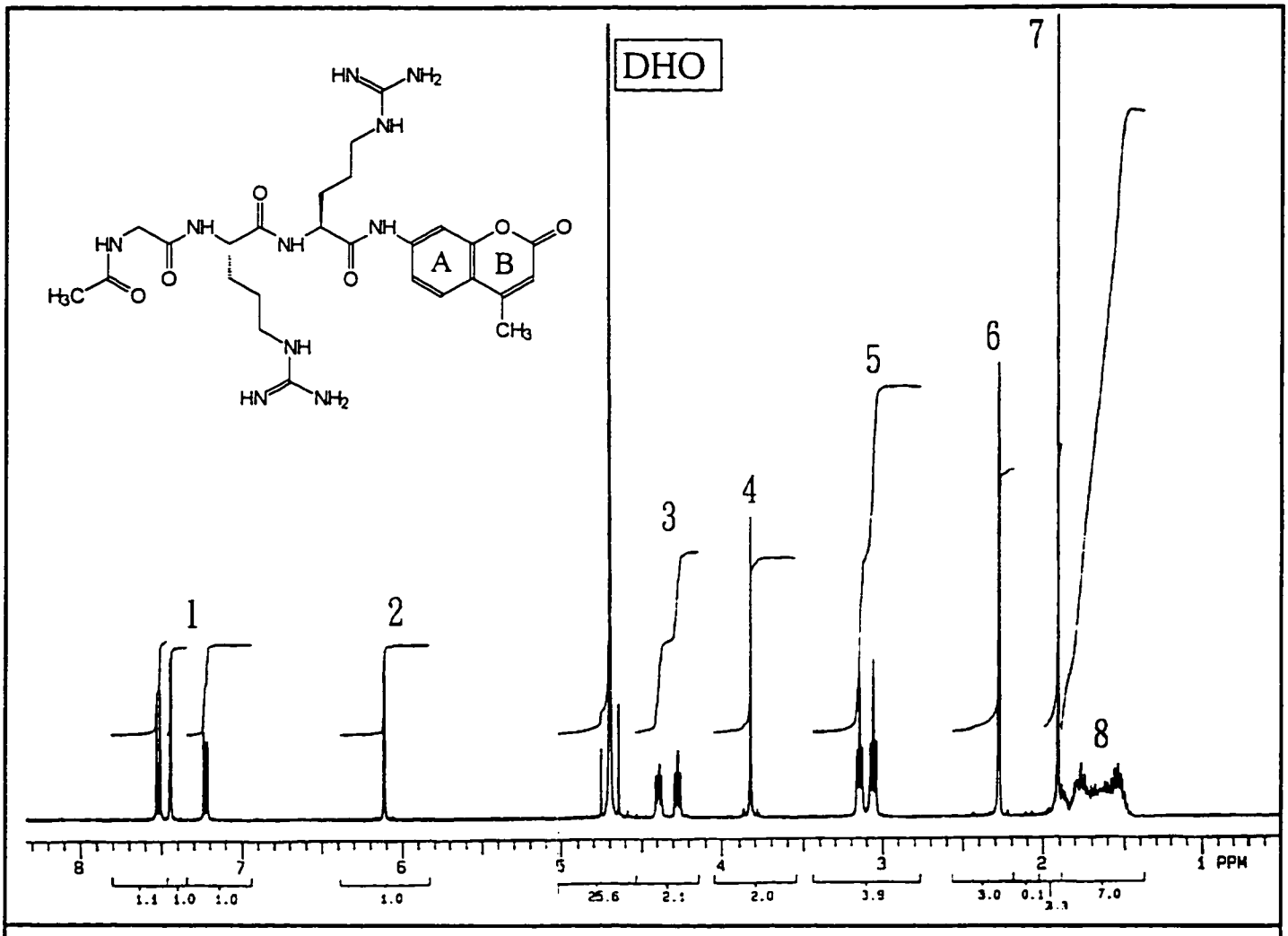


Figure 5.2.5: 400MHz [^{1}H]NMR spectrum of the substrate acetyl-Gly-Arg-Arg-AMC. Proton resonances correspond to: (1) Ring A aromatics (2) Ring B aromatic (3) Cα arginyls (4) Cα glycyl (5) Cδ arginyls (6) acetyl (7) Ring B methyl (8) Cβ and Cγ arginyls.

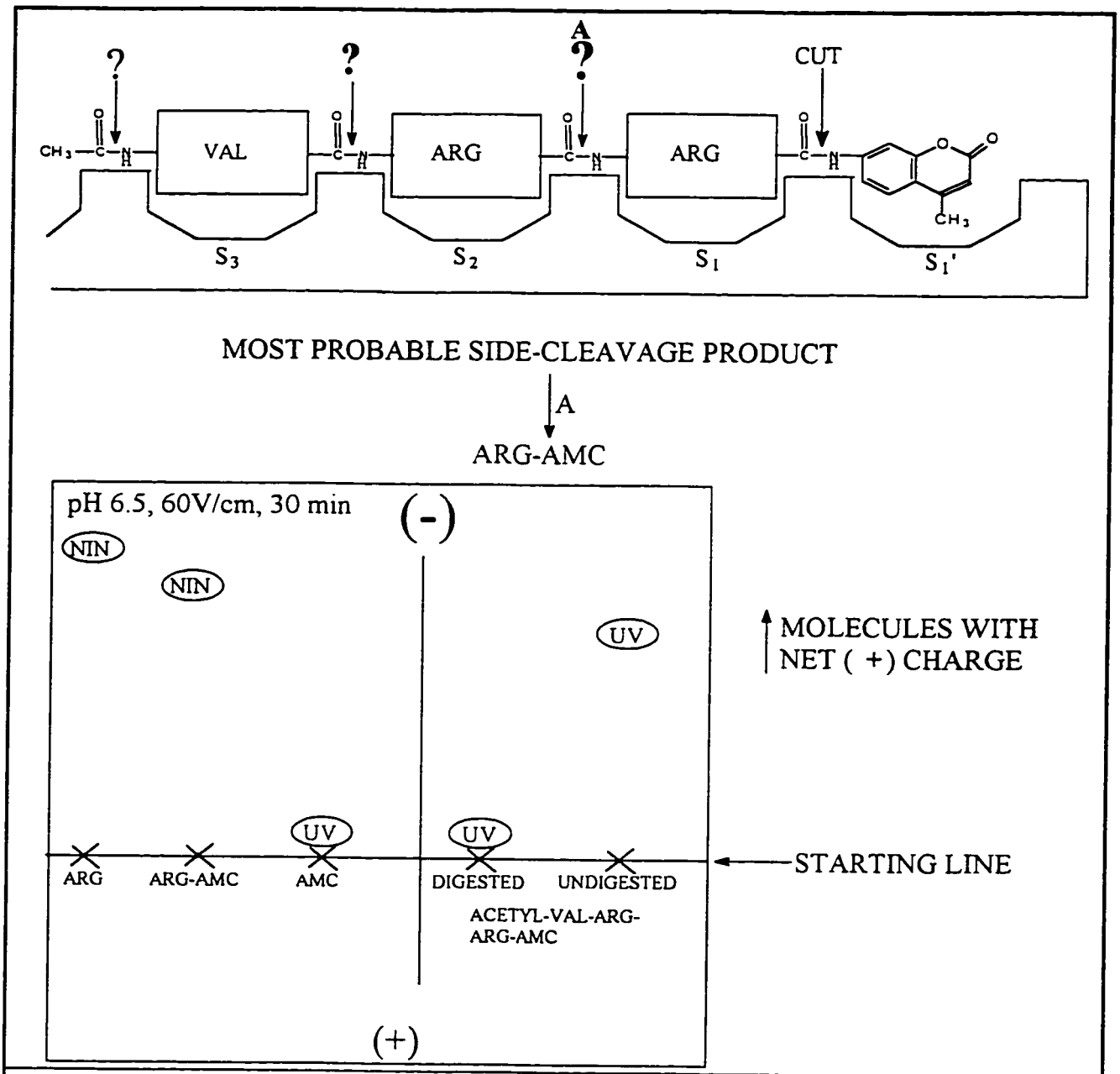


Figure 5.2.6: Schematic representation of the chromatogram obtained for acetyl-Val-Arg-Arg-AMC before and after incubation with cathepsin B. The coumaryl substrate, visible under ultraviolet light, migrates towards the cathode. In contrast, only one fluorescent neutral band is observed following incubation with catherpsin B, indicating that there is one coumaryl species. The migration of aminomethylcoumarin (AMC) is consistent with this band, indicating that cathepsin B has complelely cleaved the coumaryl bond. No other fluorescent or ninhydrin positive products were observed. In particular bond A, if cleaved,would be accompanied with concomitant production of the ninhydrin positive peptide Arg-AMC.

was verified spectrophotometrically, whereupon the digests were applied to 3MM Whatman paper and subjected to electrophoresis in pH 6.5 buffer (3:100:900 by volume) at a voltage gradient of 60 volts/cm for 30 minutes. All substrate digests yielded products that tested ninhydrin negative, demonstrating that cleavage occurs only at the AMC peptide bond. After spraying the chromatogram with the modified Sakaguchi reagent (Acher and Crocker, 1952), arginine residues were identified as a single red band, indicating there was only one product containing arginine. The final position of this fragment and its lack of fluorescence under ultraviolet light was consistent with the expected behaviour of each substrate after cleavage at the AMC peptide bond. The absence of other Sakaguchi positive material also ruled out other possible cleavage reactions. Standards of unwanted cleavage products were employed at predetermined concentrations and demonstrated that undesirable digest fragments would have been detected at the one mole percent level.

5.2.4 Molecular modeling of enzyme-substrate interactions. Modeling studies were carried out by Regine Bohacek at Ciba-Geigy Corporation for the tripeptide substrates containing a P_3 Tyr, Arg, or Glu, the latter in the protonated and deprotonated forms. Each substrate was constructed using, as a starting point, the X-ray structure of the inhibitor, Pro-Phe-Arg-chloromethylketone, bound to cathepsin B (Huber *et al.*, manuscript in prep). The MACROMODEL molecular modeling package (Mohamadi *et al.*, 1990) was used for this operation. The substrate was covalently attached to the enzyme simulating a tetrahedral intermediate by forming a covalent bond between the active site thiolate of Cys^{29} and the P_1 carbonyl carbon of the substrate.

To refine these enzyme/substrate complexes, low energy conformers were generated using simulating annealing followed by energy minimization. Each complex was subjected to 20 simulated

annealing experiments. In each experiment, the complex was subjected to a dynamics simulation during which the temperature (°C) varied from 300 degrees to 30 degrees over a period of 3 pico seconds. The time step was 3 femto seconds. These procedures were carried out using the QXP program developed at Ciba-Geigy (McMartin and Bohacek, 1995) with parameters from the AMBER force field (Weiner *et al.*, 1984). Comparison of the positions of the enzyme atoms from the X-ray structures of cathepsin B complexed with several different inhibitors showed that the positions of most of the enzyme atoms did not change significantly with binding of different inhibitors (Jia *et al.*, 1995; Huber, unpublished data). Therefore, most of the enzyme atoms were held stationary during energy minimization. Residues that were allowed to move were Gln^{23}, Cys^{29}, Asp^{69}, Glu^{122} and Glu^{245}.

5.3 Results and Discussion

5.3.1 Michaelis-Menten constants. Table 5.3.1 lists the substrates and the kinetic constants for their hydrolysis by cathepsin B. At pH 6.0, k_{cat}/K_m for the substrate acetyl-Arg-Arg-AMC was four-fold less than benzyloxycarbonyl-Arg-Arg-AMC. The difference appears largely due to an increase in K_m, which can be attributed to the substitution of the benzyloxycarbonyl group by the acetyl group. This is consistent with the fact that the X-ray structure of the inhibitor, benzyloxycarbonyl-Arg-Ser(O-benzyl)-chloromethylketone, in a covalent complex with cathepsin B, reveals a favourable interaction between the N-terminal benzyl group and the phenyl ring of Tyr^{75} (Jia *et al.*, 1995). This X-ray structure also suggests that, if the binding of the acetyl group were analogous to the position of the oxycarbonyl moiety in the inhibitor's benzyloxycarbonyl group, the N-terminal acetyl group, in acetyl-Arg-Arg-AMC, would not make any significant contacts with the enzyme. Due to the fact that the highest substrate concentration used in the

experiments with acetyl-Arg-Arg-AMC was almost ten times less than the reported K_m, a significant error can be expected in the value of the Michaelis constant. With due consideration of the limitated substrate concentration, the experimental data showed an increase in the K_m value for acetyl-Arg-Arg-AMC when compared to the value for benzyloxycarbonyl-Arg-Arg-AMC.

Table 5.3.1: Kinetic constants for recombinant cathepsin B expressed in yeast.
Enzyme was preactivated as described. The reaction buffers contained 3% dimethylsulphoxide and were 25 mM in sodium phosphate, 1 mM in ethylenediamine tetraacetic acid, pH 6.0 and 7.7, or 25 mM in sodium citrate, 1 mM in ethylenediamine tetraacetic acid, pH 4.0. The kinetic data and standard deviations were calculated from at least three separate determinations unless otherwise indicated. The range of substrate concentration used was 0.097 to 0.969 mM with acetyl-Arg-Arg-AMC, 0.086 to 0.860 mM with P_3=Glu, 0.080 to 0.800 mM with P_3=Gly, 0.081 to 0.810 mM with P_3=Arg, 0.090 to 0.900 mM with P_3=Val, and 0.062 to 0.624 mM with P_3=Tyr.

Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
pH 6.0			
benzyloxycarbonyl-Arg-Arg-AMC[a]	51.4±5.45	1.11±0.17	46400±3260
acetyl-Arg-Arg-AMC	98.9±14.6	8.67±1.63	11500±1080
acetyl-Glu-Arg-Arg-AMC	--	>10	8730±660
acetyl-Gly-Arg-Arg-AMC	69.2±4.84	4.60±0.75	15200±1500
acetyl-Arg-Arg-Arg-AMC	120±23.7	2.00±0.35	59600±4360
acetyl-Val-Arg-Arg-AMC	108±6.20	1.35±0.05	80300±4680
acetyl-Tyr-Arg-Arg-AMC	175±12.5	0.92±0.09	190000±8600
pH 4.0			
acetyl-Glu-Arg-Arg-AMC	15.0±2.20	2.15±0.21	7000±360
acetyl-Arg-Arg-Arg-AMC	3.94±2.45	1.24±0.83	3260±230
acetyl-Val-Arg-Arg-AMC	10.1±0.81	1.94±0.41	5290±730
acetyl-Tyr-Arg-Arg-AMC[b]	(17.7±3.80)	(1.56±0.40)	(11500±570)
pH 7.7			
acetyl-Arg-Arg-Arg-AMC[c]	123.0±14.1	1.65±0.09	74700±4500
acetyl-Val-Arg-Arg-AMC	86.8±5.00	0.69±0.07	126000±10400

[a]Data from Hasnain *et al.* (1992).
[b]Data from a single experiment, shown with standard errors.
[c]Data from a duplicate run, with the averages of the two runs reported.

For the tripeptide substrates at pH 6.0, specificity increased in the order Glu < Gly < Arg < Val < Tyr, with a 21-fold difference between the glutamate- and tyrosine-containing substrates. For the two substrates containing either arginine or valine at P_3, kinetic analysis at pH 7.7 showed that their second-order rate constants (k_{cat}/K_m) increased only slightly when compared to their

values at pH 6.0. However, at pH 4.0, the second-order rate constants were significantly lower than the respective values at pH 6.0. The k_{cat}/K_m values of the tripeptide substrates at pH 4.0 were generally fifteen to twenty-fold lower than at pH 6.0, the only exception being the result for acetyl-Glu-Arg-Arg-AMC. This major dependency of k_{cat}/K_m on pH has been observed previously for dipeptide substrates of cathepsin B which contained a P_2 arginine (Hasnain *et al.*, 1992; Hasnain *et al.*, 1993). The poor specificity observed for the substrate acetyl-Glu-Arg-Arg-AMC at pH 6.0 is best illustrated by a K_m value that is too large (estimated at greater than 10 mM) to measure. At pH 4.0, however, under which condition the side chain carboxyl group of the P_3 glutamyl could be more than 50 % protonated (assuming a pK_a of 4.5) (Fersht, 1985), there was a marked improvement in binding affinity as suggested by the drop in K_m compared to the value at pH 6.0. As a result k_{cat}/K_m values for this substrate were very similar at pH 4.0 and pH 6.0. For all other tripeptide substrates in this investigation (Table 5.3.1), as well as for the substrate benzyloxycarbonyl-Arg-Arg-AMC (Hasnain *et al.*, 1992), the approximately twenty-fold decrease in the value of k_{cat}/K_m at pH 4.0, when compared with pH 6.0, is, to a large extent, due to the protonation state of the Glu^{245} side chain in the S_2 subsite of cathepsin B (Hasnain *et al.*, 1993). It was concluded from the investigation of site-directed mutants that the side-chain carboxylate of the S_2 Glu^{245} forms a salt bridge with the guanidinium cation of a P_2 arginine in dipeptide substrates, such as benzyloxycarbonyl-Arg-Arg-*p*-nitroanilide. The strength of this interaction, which appears to stabilize the transition state complex, was shown to be weakened by 1.2 kcal/mol upon protonation of the of the Glu^{245} side-chain carboxylate (Hasnain *et al.*, 1993). This P_2 Arg-S_2 Glu^{245} interaction has recently been confirmed by the determination of the X-ray

structure of cathepsin B complexed with the inhibitor benzyloxycarbonyl-Arg-Ser(O-benzyl)-chloromethylketone (Jia *et al.*, 1995).

For the tripeptide substrates in this investigation, with the exception of acetyl-Glu-Arg-Arg-AMC, the observed drop in the value of k_{cat} and k_{cat}/K_m at pH 4.0, compared with pH 6.0, (Table 5.3.1) may similarly be explained by a weakening of the P_2-S_2 interaction upon the protonation of the Glu^{245} side-chain carboxylate. Any drop in k_{cat}/K_m that may have been expected for acetyl-Glu-Arg-Arg-AMC at pH 4.0 relative to pH 6.0, due to the decreased P_2-S_2 interaction, appears to have been compensated for by an improved P_3-S_3 interaction. Indeed, if the data are normalized to account for the twenty-fold drop normally observed at pH 4.0 relative to pH 6.0 for substrates with a P_2 Arg and a neutral P_3, there is an approximate twenty-fold increase in affinity for the protonated side-chain carboxyl of the P_3 Glu when compared with the deprotonated form.

Quantifying the constants K_m and k_{cat} (Table 5.3.1) was more difficult than k_{cat}/K_m, especially at pH 4.0, where product inhibition was generally strongest. Due to the limited amounts of substrate available, most assays were performed with substrate concentrations below the K_m value and at best equal to it. In this concentration range, precise data was required to obtain estimates of the constants k_{cat} and K_m. Furthermore, at pH 4.0, the determination of accurate initial rates of the two substrates acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC was hampered by strong product inhibition. Figure 5.3.1 illustrates a typical Hanes-Woolfe plot of assays with each of the tripeptide substrates. The small amount of scattering of the data points shows the precision of the data used to estimate values for K_m and k_{cat}.

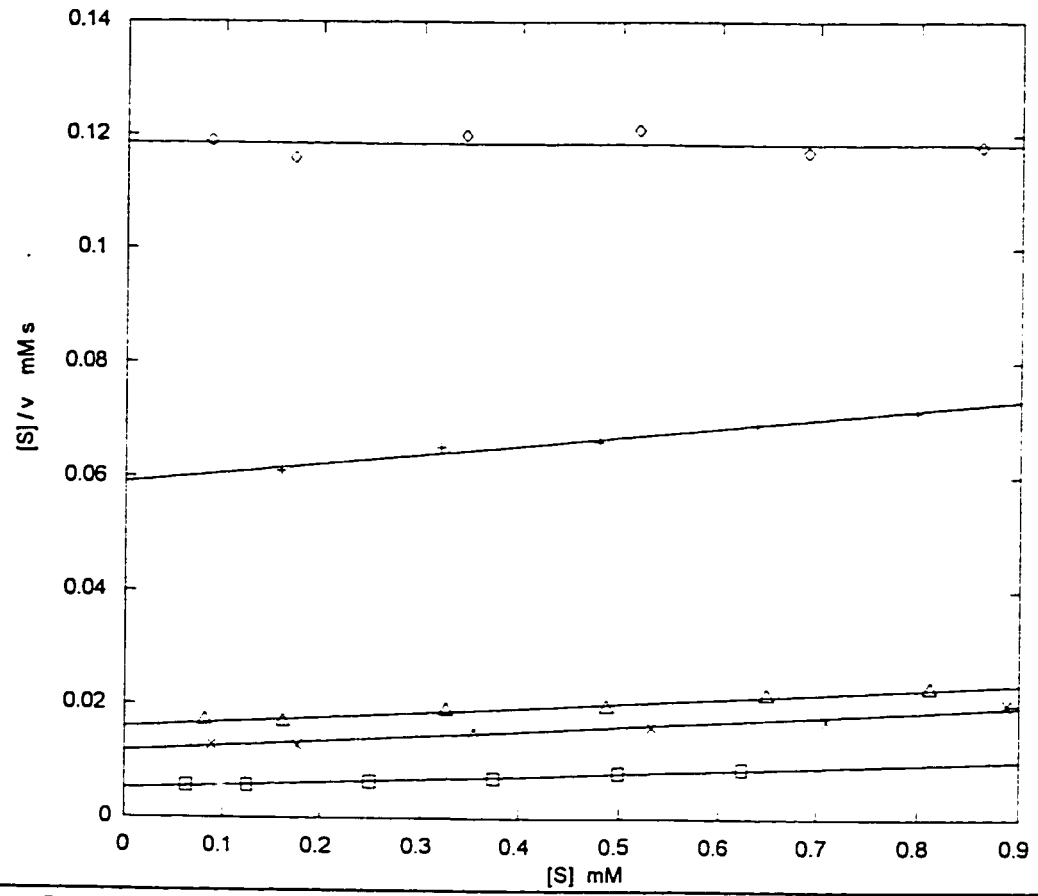
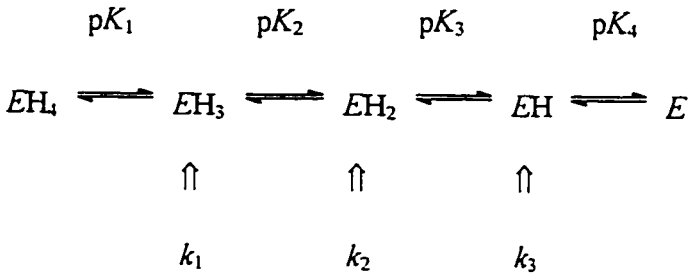


Figure 5.3.1: Hanes-Woolfe plots obtained by assaying recombinant cathepsin B at pH 6.0 with five novel tripeptide substrates. The substrates are acetyl-Tyr-Arg-Arg-AMC (□), acetyl-Val-Arg-Arg-AMC (x), acetyl-Arg-Arg-Arg-AMC (Δ), acetyl-Gly-Arg-Arg-AMC (+) and acetyl-Glu-Arg-Arg-AMC (◊). Robust and proportional weighting were used when fitting the data sets.

It is apparent from this data in Table 5.3.1 that the differences in specificity constants, k_{cat}/K_m, among these substrates is due to changes in K_m and k_{cat}. At pH 6.0, the trend in binding affinities ($1/K_m$) (obtained from the Michaelis constants of Table 5.3.1) for tripeptide substrates with various residues at P_3 can be ranked in the order Glu < Gly < Arg < Val < Tyr. This order parallels the trend in k_{cat}/K_m values and suggests that variations in specificity from subtrate to substrate are affected in part by changes in K_m. Changes in k_{cat} may also have a contributing effect on specificity. There appears to be an

increase in the value of k_{cat} (Table 5.3.1) as hydrophobic side-groups are introduced into the P_3 position of the substrate, with the same trend as that observed for specificity.

5.3.2 Analysis of pH activity profiles for the substrates acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC. The data for the pH activity profiles (Figure 5.3.2) were best fitted to a model describing two dissociation events in the ascending limb and two dissociation events in the descending limb as described before for the dipeptide substrate benzyloxycarbonyl-Arg-Arg-AMC (Hasnain *et al.*, 1992). EH_4 and E are inactive forms of the enzyme.



Data were fitted to this model according to the equation of Hasnain *et al.* (1992):

$$k_{obs}=k_{1\ limit}/([H^+]/K_1+1+K_2/[H^+]+K_2K_3/[H^+]^2+K_2K_3K_4/[H^+]^3)$$
$$+k_{2\ limit}/([H^+]^2/K_1K_2+[H^+]/K_2+1+K_3/[H^+]+K_3K_4/[H^+]^2)$$
$$+k_{3\ limit}/([H^+]^3/K_1K_2K_3+[H^+]^2/K_2K_3+[H^+]/K_3+1+K_4/[H^+])$$

where k_{obs} was the experimentally observed k_{cat}/K_m, and $k_{1\ limit}$, $k_{2\ limit}$ and $k_{3\ limit}$ were the highest theoretically obtainable k_{cat}/K_m values for each respective dissociation event.

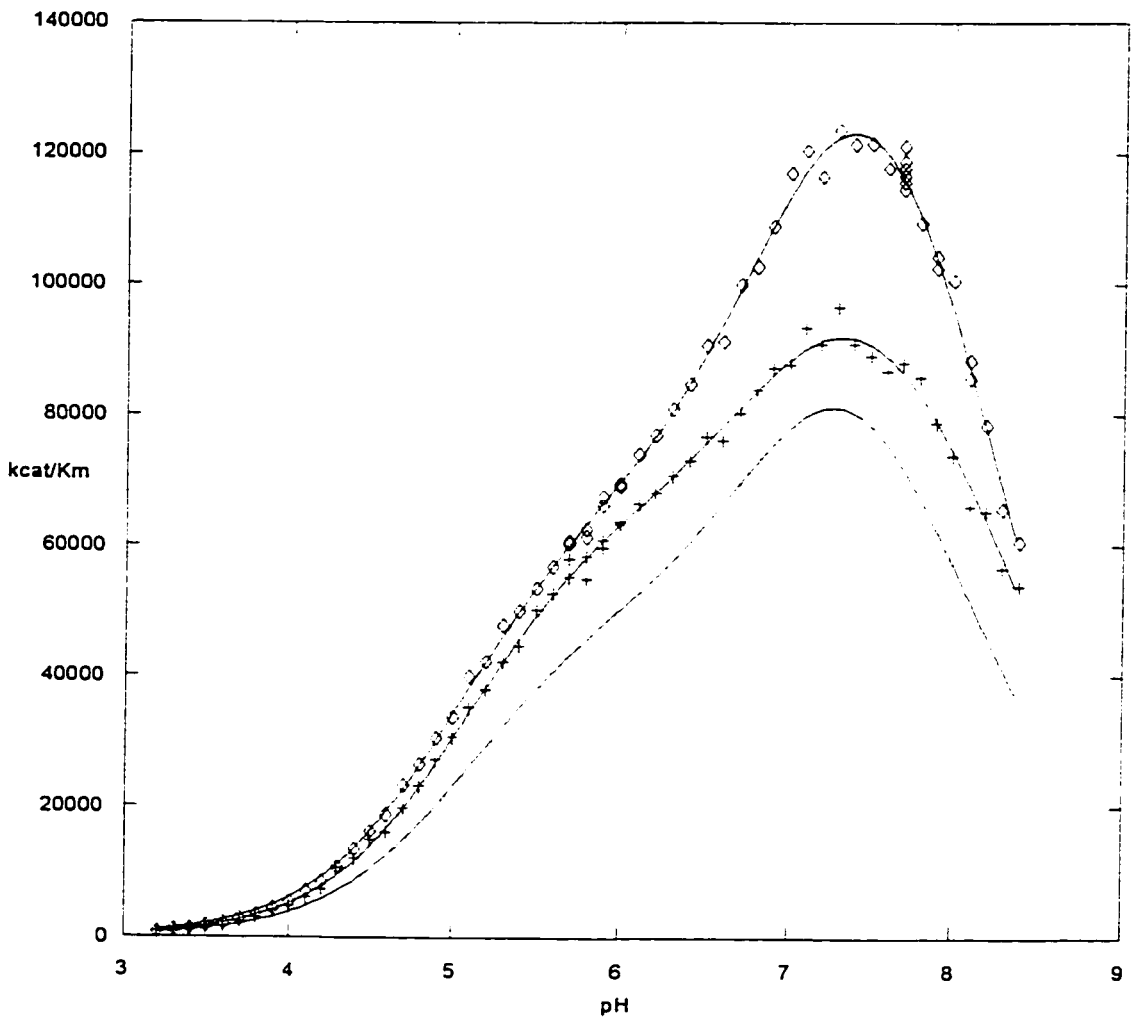


Figure 5.3.2: The pH activity profiles of recombinant cathepsin B. The substrates are acetyl-Arg-Arg-Arg-AMC (+) and acetyl-Val-Arg-Arg-AMC (◊). The lines through the data points represent the best fit using equations described in the text. A comparison to the substrate benzyloxycarbonyl-Arg-Arg-AMC (---) (Hasnain *et al.*, 1992) is shown.

The pH dependence of cathepsin B catalyzed hydrolysis of the substrates acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC (Figure 5.3.2) was very similar to that obtained previously for the substrate benzyloxycarbonyl-Arg-Arg-AMC (Hasnain *et al.*, 1992). The pK_a

values for the former tripeptide substrates were in close agreement with the values for the dipeptide substrate (Table 5.3.2). The greater uncertainty in the values for pK_a2 and pK_a3

Table 5.3.2: Ionization constants and k_{cat}/K_m limit values of dissociable groups that are important for substrate binding and catalysis obtained from pH activity profiles using the substrates acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC. One experiment for the substrate acetyl-Arg-Arg-Arg-AMC and two separate experiments for the substrate acetyl-Val-Arg-Arg-AMC were performed. The data for one of the determinations of the latter substrate are reported along with the standard errors. The pKa value 8.55±0.02 was previously determined with the substrate benzyloxycarbonyl-Phe-Arg-*p*-nitroanilide at 4°C, where kinetic analyses could be extended up to pH 9.4 (Hasnain *et al.*, 1992). This value represented the last dissociation event before all activity was lost. As this pK_a value affected catalysis, it was presumed that a similar final dissociation event should be observed in the kinetic profiles of coumaryl substrates of this investigation. The premise was supported, in that the profile of the substrate benzyloxycarbonyl-Arg-Arg-AMC could be fitted to a model which has a fourth dissociation event with a pK_a of 8.55 (Hasnain *et al.*, 1992). Since the present analyses were not carried through to sufficiently high pH values, the kinetic profiles could not be used to correctly define pK_a3 and pK_a4. To overcome this difficulty, the value of 8.55 was again used as a constant for pK_a4 in fitting the data for these two new substrates. The ease of which the last dissociation event could have been overlooked is best illustrated by the fact that the shape of the curve could be modeled incorrectly (not shown) with only one dissociation step in the decending limb in which the apparent pK_a3 value is actually a composite pK_a value that is intermediate to that of pK_a3 and pK_a4.

Substrate	benzyloxycarbonyl -Arg-Arg-AMC[a]	acetyl-Arg-Arg-Arg-AMC	acetyl-Val-Arg-AMC
pK_a1	5.07±0.03	5.06±0.02	4.98±0.02
pK_a2	6.95±0.09	6.90±0.10	6.93±0.04
pK_a3	7.71±0.12	7.69±0.17	7.98±0.08
pK_a4	(8.55±0.02)	(8.55±0.02)	(8.55±0.02)
$k_{1\ limit}$	49600±1350	64700±1270	65900±1100
$k_{2\ limit}$	111000±5730	111000±3970	160000±3600
$k_{3\ limit}$	46600±4790	78300±5010	63000±7500

[a]Data taken from Hasnain *et al.* (1992).

were due, in part, to the difficulty of obtaining accurate initial rates above pH 7.9, due to the pH instability of cathepsin B in that pH range. These data suggest that the interaction of the P_3 Arg in the S_3 subsite is not to any significant degree due to a charge-charge interaction, although a weak interaction cannot be ruled out.

5.3.3 Nature of the S_3-P_3 interaction. Table 5.3.3 reports the change in apparent binding energies among the dipeptide and tripeptide substrates at pH 6.0. Results are summarized as follows:

(i) Apparent binding energies for the dipeptide substrates acetyl-Arg-Arg-AMC and benzyloxycarbonyl-Arg-Arg-AMC. The change in apparent binding energy for the substrate acetyl-Arg-Arg-AMC relative to benzyloxycarbonyl-Arg-Arg-AMC was calculated to be 0.82 kcal/mol, suggesting that the benzyloxycarbonyl group interacts with the enzyme as a pseudo-P_3 residue. As mentioned previously, this finding is supported by the X-ray structure of a cathepsin B-inhibitor complex (Jia *et al.*, 1995). In this complex, the benzyl ring of the N-terminal benzyloxycarbonyl group of the inhibitor, benzyloxycarbonyl-Arg-Ser(O-Bzl)-CMK, makes a direct contact with the enzyme, forming an aromatic-aromatic interaction with Tyr^{75}, with the shortest distance of about 3.71 Å.

(ii) Binding energies of various P_3 side chains. The glycyl containing substrate served as the reference for the calculation of apparent binding energies of the P_3 side chains of the other tripeptide substrates (Table 5.3.3). At pH 6.0, the only P_3 side chain that demonstrated a weaker binding energy relative to glycyl was glutamyl, with a $\Delta\Delta G$ of +0.33 kcal/mol. The other three tripeptide substrates, with either Arg, Val or Tyr at P_3, exhibited an improved binding in that order. While the substrate containing a P_3 arginyl has an increased binding energy compared to glycyl, by about 0.81 kcal/mol, this enhancement cannot be to any large degree due to a charge-charge interaction at the S_3 subsite. First, a four-fold effect arising from an ionic interaction between the P_3 guanidinium group and a negatively charged group in the S_3 subsite should have been manifested in the pH activity profile. As shown in Figure 5.3.2 and Table 5.3.2, the pH activity profile data for the tripeptide, acetyl-Arg-Arg-Arg-AMC, could be fitted by the same equation used to fit the data for the dipeptide, benzyloxycarbonyl-Arg-Arg-AMC, and the substrate with a neutral P_3, acetyl-Val-Arg-Arg-AMC. Furthermore, the dissociation constants

derived from these data were virtually identical. In fact, the ratios of k_{cat}/K_m of acetyl-Arg-Arg-Arg-AMC and acetyl-Val-Arg-Arg-AMC are similar at pH 6.0 and 4.0. This is also true for the ratios of k_{cat}/K_m of acetyl-Arg-Arg-Arg-AMC and the tripeptide substrates with either glycyl or tyrosyl side chains at P_3, which suggests that the P_3 arginyl side-chain binding is not to any significant extent stabilized by an ionic interaction. The X-ray structure of the complex of cathepsin B and the inhibitor benzyloxycarbonyl-Arg-Ser(O-Bzl)-CMK suggests a relatively hydrophobic S_3 pocket defined by Tyr^{75}, Cα of Gly^{73}, Cβ of Asn^{72} and Cα of Asp^{69} (Jia *et al.*, 1995). Therefore one may suggest that the improved binding of arginine over glycine at P_3 may be largely due to hydrophobic interactions of the methylene groups in the arginine side chain with the relatively hydrophobic S_3 subsite. An interaction between the Asp^{69} side-chain carboxylate and the P_3 guanidinium cannot be entirely ruled out. However, if it occurs, this interaction does not contribute very significantly to substrate binding (see below for further discussion).

At pH 6.0, a comparison of the specificity constants for the substrates acetyl-Arg-Arg-AMC and acetyl-Gly-Arg-Arg-AMC (Table 5.3.1) shows virtually no difference. It appears, therefore, that the acetyl methyl and the atoms forming the adjacent amide bond in the tripeptide substrate make no significant contribution to binding. In addition, the X-ray structure of the cathepsin B-inhibitor complex reveals that the oxycarbonyl moiety between the pseudo-P_3 N-terminal benzyl group and the P_2 residue does not make any contact with the enzyme (Jia *et al.*, 1995). As such, the values for the change in free energy of binding (Table 5.3.3) for different tripeptide substrates also serve to define the incremental increase in binding energy of the P_3 residue when compared with the dipeptide substrate, acetyl-Arg-Arg-AMC.

Table 5.3.3: Change of free energy of binding (ΔΔG) between (i) the substrates benzyloxycarbonyl-Arg-Arg-AMC and acetyl-Arg-Arg-AMC and (ii) tripeptide substrates with amino acid substitutions at P_3 and the reference substrate acetyl-Gly-Arg-Arg-AMC. The ratios were averages calculated from k_{cat}/K_m values obtained from at least three determinations at pH 6.0. The substrate benzyloxycarbonyl-Arg-Arg-AMC (k_{cat}/K_m = 46400 $M^{-1}s^{-1}$), borrowed from Hasnain *et al.* (1992), served as a reference for the dipeptide acetyl-Arg-Arg-AMC, and the substrate acetyl-Gly-Arg-Arg-AMC (k_{cat}/K_m = 15200 $M^{-1}s^{-1}$) served as a reference for the tripeptides. $\Delta\Delta G$ was determined using the relationship $\Delta\Delta G = -RT\ln\{(k_{cat}/K_m^{[other]}) /(k_{cat}/K_m^{[reference]})\}$.

Substrate	$(k_{cat}/K_m^{[other]})/(k_{cat}/K_m^{[reference]})$	$\Delta\Delta G$ (kcal/mole)
acetyl-Arg-Arg-AMC	0.248	0.82
acetyl-Glu-Arg-Arg-AMC	0.574	0.33
acetyl-Arg-Arg-Arg-AMC	3.92	-0.81
acetyl-Val-Arg-Arg-AMC	5.28	-0.98
acetyl-Tyr-Arg-Arg-AMC	12.5	-1.50

(iii) Evidence supporting a predominantly hydrophobic pocket at S_3. From the specificity constants reported in Table 5.3.1 and changes in binding energy in Table 5.3.3, the nature of the S_3 subsite of cathepsin B can be deduced. The S_3 subsite shows a general preference for both aromatic and aliphatic groups. At pH 6.0, the substrate with the P_3 glutamyl is least favoured, while at pH 4.0 it is preferred over the arginyl and valyl side chains. Therefore, it appears that a negative charge at P_3 disrupts binding at the S_3 subsite. This may be due to a charge-charge repulsion involving Asp^{69} as well as unfavourable hydrophobic interactions.

If the Asp^{69} carboxylate is responsible for destabilizing the binding of a P_3 negative charge, it is surprising that the four-fold increase in specificity of the arginyl side chain compared to glycyl does not appear to result from a charge-charge attraction with Asp^{69}. A possible explanation for the four-fold increase in specificity of the Arg side chain compared to Gly is the possible van der Waals interaction of one or more of the arginyl side-chain methylene carbons with Tyr^{75}.

The two-fold increase in specificity of the P_3 tyrosyl side chain, relative to the P_3 valyl, and the four-fold increase in specificity of the P_3 tyrosyl, relative to the benzyl ring of the substrate

benzyloxycarbonyl-Arg-Arg-AMC, may result from an improved interaction of the phenyl ring of the substrate tyrosyl and Tyr^{75} in the S_3 subsite of the enzyme. In the X-ray structure of the cathepsin B-inhibitor complex, the benzyl ring of the benzyloxycarbonyl moiety is about 3.7 Å from the Tyr^{75} phenyl ring. The kinetic data suggest that both the valyl and tyrosyl side chains in the respective substrates make energetically more favourable contacts with Tyr^{75} than the benzyl ring of the benzyloxycarbonyl moiety.

5.3.4 Modeling tripeptide substrates containing Tyr, Glu or Arg at P_3. On the basis of supporting X-ray and kinetic evidence, residues Glu^{245}, Asp^{69}, Glu^{122}, Gln^{23} and Cys^{29} are believed to play either binding or catalytic roles in the hydrolysis of the tripeptide substrates. For this reason, Glu^{245}, Asp^{69}, Glu^{122}, Gln^{23} and the tetrahedral intermediate formed at Cys^{29} were allowed to move. As part of the study, it was important to investigate whether the final position of these residues were consistent with X-ray data of inhibitor complexes. In each case, the minimum energy positions of Glu^{245}, Glu^{122}, Gln^{23} and Cys^{29} corroborated X-ray structures and are not discussed further. The residue Asp^{69} was anticipated to interact with the guanidino group of the P_3 Arg substrate, however, this was not predicted by the modeling study. Instead, there was an unexpected repulsion shown between Asp^{69} and the P_3 Glu residue.

The modeling study with the substrate containing a P_3 Tyr (Figure 5.3.3) revealed that the phenyl ring of the substrate is partly stacked on the phenyl of Tyr^{75} of the enzyme, with the shortest distance between Cε of the substrate Tyr and Cε of Tyr^{75} of 3.4 Å. In the model, the

substrate tyrosyl hydroxyl forms a hydrogen bond with Asp^{69}. Another potential contact in the S_3 pocket is between the substrate Tyr Cβ and Cα of Gly^{73} (3.6 Å). The improved contact of the P_3 tyrosyl in the S_3 subsite, as suggested by the model, relative to the position of the pseudo-P_3 benzyl group of the dipeptide inhibitor observed in the X-ray structure (Jia *et al.*, 1995), is consistent with the kinetic data showing a four-fold increase in specificity for the P_3 Tyr compared to the N-terminal benzyloxycarbonyl group. A study by Serrano *et al.* (1991) revealed that aromatic-aromatic interactions in protein structures can contribute between 0.6 and 1.3 kcal/mol to protein stability. This effect is due to a quadrupole-quadrupole interaction between the aromatic rings, for which there is an associated potential energy that varies as $1/r^5$ (where r is the quadrupole separation distance) (Burley and Petsko, 1989). Therefore, any binding energy which is contributed by aromatic side chains of substrates or inhibitors interacting with aromatics on enzymes would be very sensitive to the distance between the groups.

The P_3 Arg modeling study (Figure 5.3.4) reveals that the methylene groups of the arginyl side chain are somewhat further from Gly^{73} (4.3 Å) and Tyr^{75} (4.4 Å) than the other modeled substrates, suggesting that a van der Waals interaction is unlikely. The guanidinium group of the P_3 Arg appears to form hydrogen bonds with the hydroxyl of Tyr^{75} (3.0 Å) and the carboxyl oxygen of Asp^{69} (2.7 Å) and, possibly, a weak hydrogen bond with the backbone carbonyl oxygen of Asn^{72} (3.4 Å). As discussed before, the kinetic data suggest that, if there is an interaction between the guanidinium moiety of the P_3 Arg and the carboxylate of Asp^{69}, there is no significant net gain in binding energy. Considering the fact that there would be a significant energy cost involved in desolvating the guanidinium cation and that this interaction would be completely solvent-exposed, it may be reasonable to expect that there may not be any significant gain in

binding energy for this charge-charge interaction. Therefore, the four-fold increase in specificity observed for a P_3 Arg, relative to Gly, may be due instead to the charge-neutral hydrogen bonds involving the substrate guanidinium and the hydroxyl of Tyr^{75} and possibly the carbonyl of Asn^{72}.

Figure 5.3.3: Possible binding mode of the substrate acetyl-Tyr-Arg-Arg-AMC. A possible binding mode of the substrate acetyl-Tyr-Arg-Arg-AMC to cathepsin B was determined by docking the substrate into a computer model of the active site obtained from the crystal structure of cathepsin B. The substrate was energy minimized as described in the text, using the Amber force field as implemented by the MACROMODEL software package.

Figure 5.3.4: Possible binding mode of the substrate acetyl-Arg-Arg-Arg-AMC. A possible binding mode of the substrate acetyl-Arg-Arg-Arg-AMC to cathepsin B was determined as described in Figure 5.3.3.

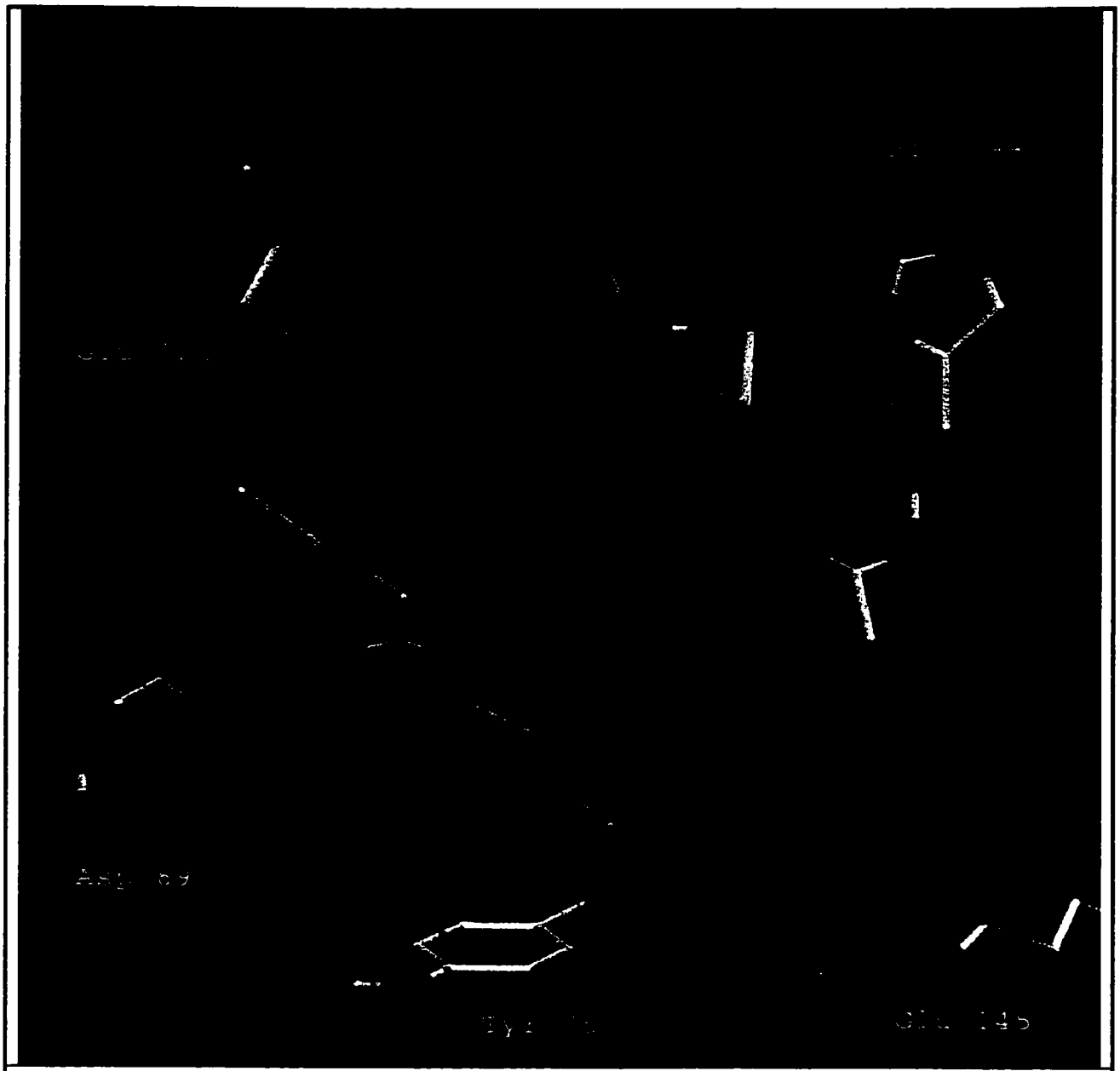


Figure 5.3.5: Possible binding mode of the substrate acetyl-Glu-Arg-Arg-AMC. A possible binding mode of the substrate acetyl-Glu-Arg-Arg-AMC to cathepsin B, where the Glu side-chain carboxyl is deprotonated, was determined as described in Figure 5.3.3.

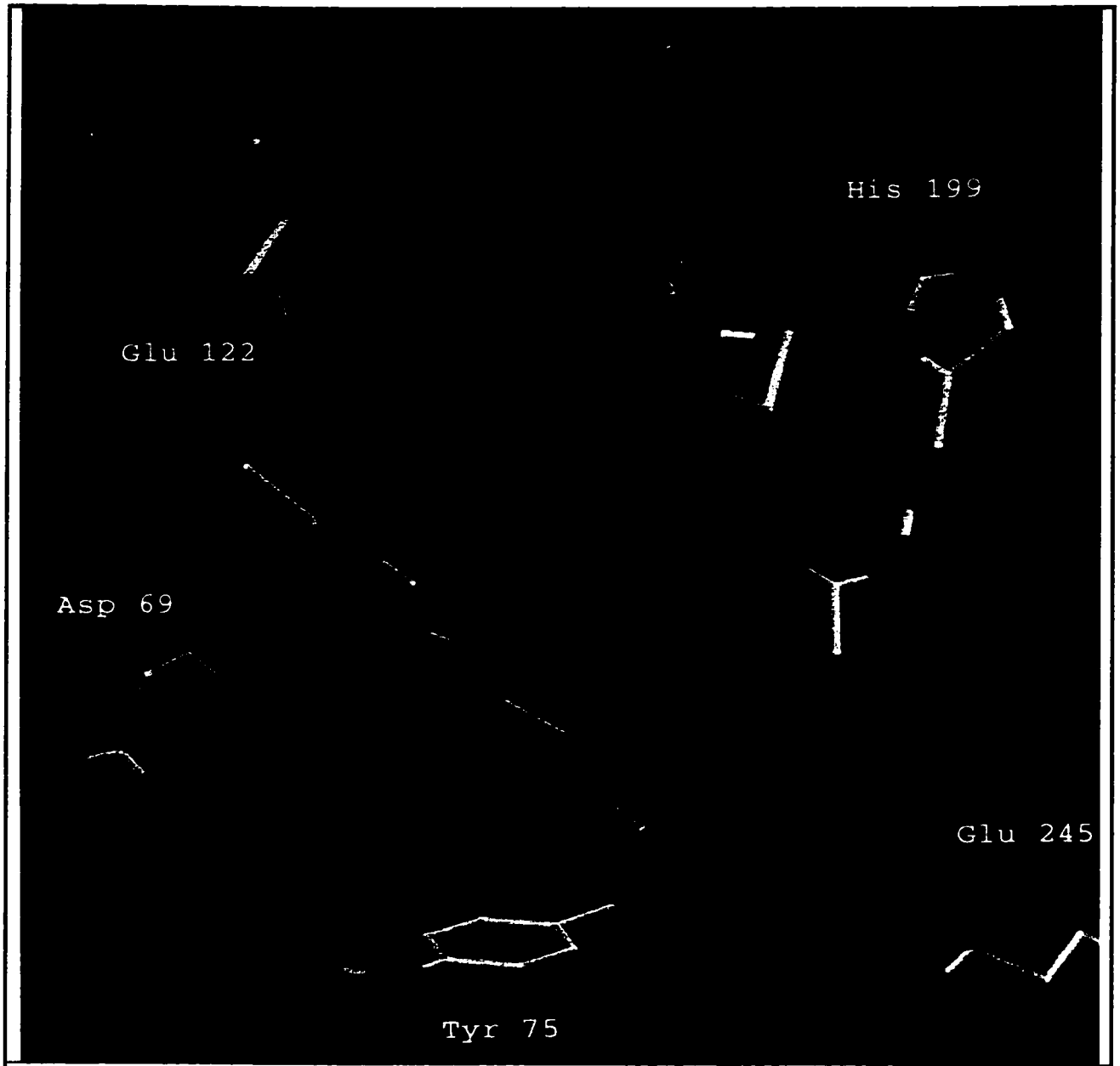


Figure 5.3.6: Possible binding mode of the substrate acetyl-Glu-Arg-Arg-AMC. A possible binding mode of the substrate acetyl-Glu-Arg-Arg-AMC to cathepsin B, where the Glu side-chain carboxyl is protonated, was determined as described in Figure 5.3.3.

The modeling study for the substrates containing either a protonated or deprotonated Glu (Figures 5.3.5 and 5.3.6) show that the side-chain methylene carbons of Glu make van der Waals contacts with the Cα of Gly^{73} (3.2 Å) and the phenyl ring of Tyr^{75} (3.6 Å). The kinetic data show

that k_{cat}/K_m drops 20-fold when the P_3 Glu deprotonates. One possible explanation of this significant destabilization of binding may be a charge-charge repulsion between Asp^{69} and the P_3 side-chain carboxylate, which are about 4.4 Å apart. Since the potential energy of charge-charge interactions varies as 1/r, where r is the charge separation, they can be manifested over relatively large distances. In fact, the modeling study with the deprotonated Glu shows the Asp^{69} side chain moving away from the P_3 carboxylate relative to the Asp^{69} side-chain position in the modeling studies with the other substrates (Unlike the long Arg side-chain, perhaps the P_3 Glu carboxylate is not as solvent accessible so that a significant charge interaction can be formed with Asp^{69}). Another possible explanation, however, may be that there is a repulsion between the π-electron cloud of Tyr^{75} and the negative charge on the substrate side chain (Burley and Petsko, 1989). To summarize, evidence consistent with a hydrophobic pocket at the S_3 subsite of cathepsin B is presented. The interactions which promote and favour the binding of one substrate over another are derived from steric factors at S_3 but may also involve some hydrogen bonding. The substrate binding in the S_3 subsite of cathepsin B is largely due to contacts with the phenyl ring of Tyr^{75} and to a lesser extent may involve the Cα of Gly^{73}. Additionally, hydrogen bonds involving the Tyr^{75} hydroxyl and the main-chain carbonyl oxygen of Asn^{72} may also make a contribution to substrate stabilization. Additional experiments are required to support or rule out the latter possibility. The best substrate among the tripeptides tested has tyrosine at P_3, which in comparison to the reference substrate containing glycine, has an increase in binding energy of 1.5 kcal/mol. This preference of cathepsin B for tyrosine at S_3 suggests a possible role for aromatic residues as specificity determinants for inhibitors. While these findings are significant for inhibitor design strategies, it may be useful to examine further the S_3 specificity of cathepsin B with substrates containing residues such as

tryptophan, phenylalanine, proline, methionine, leucine and isoleucine at the P_3 position. The study of these additional residues would provide a more complete understanding of the S_3 pocket and will contribute additional information for the design of more specific inhibitors.

5.4 Conclusions

Five synthetic substrates containing different amino acid residues at the P_3 position (namely, acetyl-X-Arg-Arg-AMC, where X is Gly, Glu, Arg, Val and Tyr) were used to better define the S_3 subsite specificity of cathepsin B. At pH 6.0, the specificity constant, k_{cat}/K_m, for tripeptide substrate hydrolysis was observed to increase in the order Glu < Gly < Arg < Val < Tyr. Molecular modeling studies of substrates containing a P_3 Glu, Arg or Tyr covalently bound as the tetrahedral intermediate to the enzyme suggested that the specificity for a P_3 Tyr is due to a favourable aromatic-aromatic interaction with Tyr^{75} on the enzyme as well as a possible hydrogen bond joining the P_3 Tyr hydroxyl function and the side-chain carboxyl group of Asp^{69}.

5.5 Acknowledgments

Alpay Taralp is indebted to Dr. Art Szabo for the use of his laboratory facilities and to the Natural Sciences and Engineering Research Council (NSERC) of Canada for a graduate scholarship. This work was supported in part by an NSERC operating grant to Dr. Harvey Kaplan and in part by the National Research Council of Canada. Further acknowledgments are due to Andrew Wigg of Ciba-Geighy Corporation, for assistance provided.

5.6 References

Acher, R. and Crocker, C. (1952) *Biochim. Biophys. Acta* **9**, 704-705.

Atherton, E. and Sheppard, R. C. (1989) *Solid phase peptide synthesis:A practical approach* (Rickwood, D. and Hames, B. D. Eds.) IRL Press.

Barrett, A. J. and Kirschke, H. (1981) *Meth. Enzymol.* (Lorand, L. Ed.) **80**, 535-581.

Berquin, I. M., Cao, L., Fong, D. and Sloane, B. F. (1995) *Gene (Amsterdam)* **159**, 143-149.

Bodanszky, M. and Bodanszky, A. (1994) *The practice of peptide synthesis*, 2nd ed., Springer-Verlag.

Brömme, D., Bescherer, K., Kirschke, H. and Fittkau, S. (1987) *Biochem. J.* **245**, 381-385.

Brömme, D., Steinert, A., Friebe, S., Fittkau, S., Wiederanders, B. and Kirschke, H. (1989) *Biochem. J.* **264**, 475-481

Buck, M. R., Karustis, D. G., Day, N. A., Honn, K. V. and Sloane, B. F. (1992) *Biochem. J.* **282**, 273-278.

Burley, S. K and Petsko, G. A. (1989) *TIBTECH* 7, 354-359.

Calkins, C. C. and Sloane, B. F. (1995) *Biological Chemistry Hoppe-Seyler* **376**, 71-80.

Campo, E., Munoz, J., Miquel, R., Palacin, A., Cardesa, A., Sloane, B. F. and Emmert-Buck M R. (1994) *American Journal of Pathology* **145**, 301-309.

Cao, L., Taggart, R. T., Berquin, I. M., Moin, K., Fong, D. and Sloane, B. F. (1994) *Gene (Amsterdam)* **139**, 163-169.

Castiglioni, T., Merino, M. J., Elsner, B., Lah, T. T., Sloane, B. F. and Emmert-Buck, M. R. (1994) *Human Pathology* **25**, 857-862.

Crawford, C., Mason, R. W., Wikstrom, P. and Shaw, E. (1988) *Biochem. J.* **253**, 751-758.

Emmert-Buck, M. R., Roth, M. J., Zhuang, Z., Campo, E., Rozhin, J., Sloane, B. F., Liotta, L. A. and Stetler Stevenson, W. G. (1994) *American Journal of Pathology* **145**, 1285-1290.

Ernst, J. F. (1986) *DNA (N.Y.)* **5**, 483-491.

Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, p. 156, W. H. Freeman and Company.

Gopalan, P., Dufresne, M. J. and Warner, A. H. (1987) *Can. J. Physiol. Pharmacol.* **65**, 124-129.

Hasnain, S., Hirama, T., Tam, A. and Mort, J. S. (1992) *J. Biol. Chem.* **267**, 4713-4721.

Hasnain, S., Hirama, T., Huber, C., Mason, P. and Mort, J. S. (1993) *J. Biol. Chem.* **268**, 235-240.

Honn, K. V. , Timar, J., Rozhin, J., Bazaz, R., Sameni, M., Ziegler, G. and Sloane, B. F. (1994) *Experimental Cell Research* **214**, 120-130.

Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J. S., To, R. and Huber, C. P. (1995) *J. Biol. Chem.*, **270**, 5527-5533.

Katunuma, N., and Kominami, E. (1987) *Rev. Physiol. Biochem. Pharmacol.* **108**, 2-20.

Keppler, D. and Sloane, B. F. (1996) *Enzyme and Protein* **49**, 94-105.

Khouri, H. E., Plouffe, C., Hasnain, S., Hirama, T., Storer, A. C. and Menard, R. (1991) *Biochem. J.* **275**, 751-757.

Koga, H., Yamada, H., Nishimura, Y., Kato, K. and Imoto, T. (1991) *J. Biochem.* **110**, 179-188.

Krantz, A., Copp, L. J., Coles, P. J., Smith, R. A. and Heard, S. B. (1991) *Biochemistry* **30**, 4678-4687.

Lah, T. T., Buck, M. R., Honn, K. V., Crissman, J. D., Rao, N. C., Liotta, L. A. and Sloane, B. F. (1989) *Clin. Exp. Metastasis* **7**, 461-468.

Lee, X., Ahmed, F. R., Hirama, T., Huber, C. P., Rose, D. R., To, R., Hasnain, S., Tam, A. and

March, J. (1985) *Advanced organic chemistry: Reactions, mechanisms and structure*, 3rd. ed., pp. 365-377 and refs. cited therein, John Wiley and Sons.

Mort, J. S. (1990) *J. Biol. Chem.* **265**, 5950-5951.

McMartin, C. and Bohacek, R. S. (1995) *Journal of Computer Aided Molecular Design* **9**, 237-250.

Mikkelsen, T., Yah, P. S., Ho, K. L., Sameni, M., Sloane, B. F. and Rosenblum, M. L. (1995) *Journal of Neurosurgery* **83**, 285-290

Mohamadi, F., Richards, N. G. J., Guida, W. C., Liskamp, R., Lipton, M. G. C., Chang, G., Hendrickson, T. and Still, W. C. (1990) *J. Comp. Chem.* **11**, 440-467.

Moin, K., Emmert, L. T. and Sloane, B. F. (1992a) *FEBS Letters* **309**, 279-282.

Moin, K., Day, N. A., Sameni, M., Hasnain, S., Hirama, T. and Sloane, B. F. (1992b) *Biochem. J.* **285**, 427-434.

Mort, J. S., Recklies, A. D. and Poole, A. R. (1984) *Arthritis Rheum.* **27**, 509-515.

van Noorden, C. J. F., Smith, R. E. and Rasnick, D. (1989) *J. Rheumatol.* **15**, 1525-1535.

Pliura, D. H., Bonaventura, B. J., Smith, R. A., Coles, P. J. and Krantz, A. (1992) *Biochem. J.* **288**, 759-762.

Pohl, J., Davinic, S., Blaha, I., Strop, P. and Kostka, V. (1987) *Analytical Biochemistry* **165**, 96-101.

Rempel, S. A., Rosenblum, M. L., Mikkelsen, T., Yan, P. S., Ellis, K. D., Golembieski, W. A., Sameni, M., Rozhin, J., Ziegler, G. and Sloane, B. F. (1994) *Cancer Research* **54**, 6027-6031.

Ren, W. P., Fridman, R., Zabrecky, J. R., Morris, L. D., Day, N. A. and Sloane, B. F. (1996) *Biochem. J.* **319**, 793-800.

Rowan, A. D., Mason, P., Mach, L. and Mort, J. S. (1992) *J. Biol. Chem.* **267**, 15993-15999.

Rozhin, J., Sameni, M., Ziegler, G. and Sloane, B. F. (1994) *Cancer Research* **54**, 6517-6525.

Ryan, R. E., Sloane, B. F., Sameni, M. and Wood, P. L. (1995) *Journal of Neurochemistry* **65**, 1035-1045.

Serrano, L., Bycroft, M. and Fersht, A. R. (1991) *J. Mol. Biol.* **218**, 465-475.

Sinha, A. A., Gleason, D. F., Deleon, O. F., Wilson, M. J. and Sloane, B. F. (1993) *Anatomical Record* **235**, 233-240.

Sinha, A. A., Gleason, D. F., Staley, N. A., Wilson, M. J., Sameni, M. and Sloane, B. F. (1995b) *Anatomical Record* **241**, 353-362.

Sinha, A. A., Wilson, M. J., Gleason, D. F., Reddy, P. K., Sameni, M. and Sloane, B. F. (1995a) *Prostate* **26**, 171-178.

Sloane, B. F. (1990) *Semin. Cancer Biol.* **1**, 137-152.

Sloane, B. F., Herman, C. J. and Padarathsingh, M. (1994a) *Cancer Research* **54**, 5241-5245.

Sloane, B. F., Moin, K., Sameni, M., Tait, L. R., Rozhin, J. and Ziegler, G. (1994b) *Journal of Cell Science* **107**, 373-384.

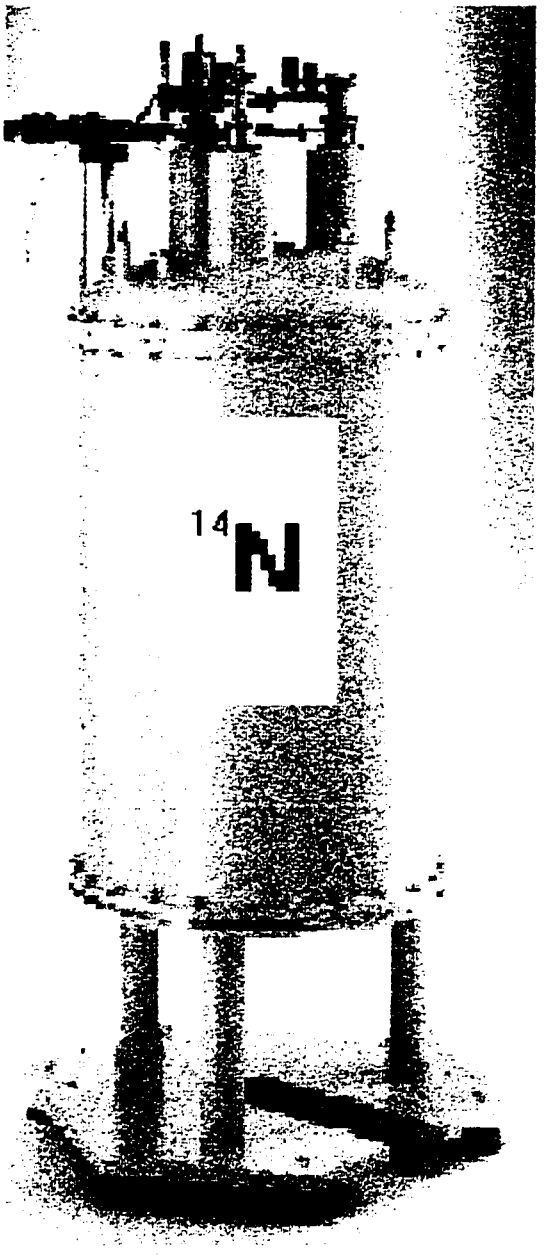
Sloane, B. F., Moin, K., Krepela, E. and Rozhin, J. (1990) *Cancer Metastasis Rev.* **9**, 333-352.

Visscher, D. W., Sloane, B. F., Sameni, M., Babiarz, J. W., Jacobson, J. and Crissman, J. D. (1994) *Modern Pathology* **7**, 76-81.

Weiner, S. J., Kollman, P. A. Case, D. A. Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. K. (1984) *J. Am. Chem. Soc.* **106**, 765-784.

Wikstrom, P., Kirschke, H., Stone, S. and Shaw, E. (1989) *Archives of Biochemistry and Biophysics* **270**, 286-293.

Willenbrock, F. and Brocklehurst, K. (1985) *Biochem. J.* **227**, 521-528.


14N

Chapter 6: [^{12}C]Iodomethane for the investigation of protein amino groups by [^{14}N]NMR

6.1 Introduction

Successfully adapting NMR techniques for use in the area of protein characterization is an ongoing theme in protein chemistry. The use of [^{14}N]NMR spectroscopy to observe Nα and Nε amino groups in proteins and amino acids is one potential extention of current protein NMR techniques. Despite the natural isotopic abundance of ^{14}N (99.6%) (Witanowski and Webb, 1972a) and the general wealth of amino groups in proteins, very little work has been carried out on protein systems using [^{14}N]NMR as a structure-function probe (*vide infra*). ^{14}N nuclei have average receptivity, low resonance frequencies and very short relaxation times compared to ^{15}N nuclei as a result of the large quadrupole moment (Witanowski and Webb, 1972a). Consequently, [^{14}N]NMR resonances are generally broad but much easier to observe than ^{15}N which has a very low natural abundance.

The broad lines obtained in ^{14}N-spectra is certainly one reason which may have removed any incentive to use [^{14}N]NMR for the study of biologically relevant compounds. Of those reported, studies generally focus on the variability of linewidths and not chemical shift measurements (Witanoski *et al.*, 1977a; Mann, B. E., 1978a). The ^{14}N-linewidths of several amino acids, related molecules, and amino groups of protein have been studied as a function of pH (Jenks, 1971; Tzalmona and Loewenthal, 1974; Cohen *et al.*, 1975; Richards and Thomas, 1975; Zhang *et al.*, 1994). Protonation states and proton exchange processes in aqueous solution were shown to affect substantially the ^{14}N lineshapes and chemical shift values (Cooper *et al.*, 1973; Tzalmona and Loewenthal, 1974; Cohen *et al.*, 1975; Blomberg *et al.*, 1976; Sheinblatt and Gutowsky,

1976). In the case of glycine, the transverse relaxation times and linewidths at low pH were affected by proton exchange, while at pH values greater than 6, the relaxation times and linewidths were a weighted average corresponding to the proportion of zwitterionic and anionic species at a given pH value (Tzalmona and Loewenthal, 1974). Linebroadening was greatest at either pH extreme. Chemical shift positions of quaternary nitrogens of bioactive peptides were studied in more isolated cases (Gerothanassis *et al.*, 1987; Karayannis *et al.*, 1990). Karayannis in particular, used [^{14}N]NMR spectroscopy to study the conformation of enkephalins. As NMR relaxation processes are sensitive to molecular motions (Berglund *et al.*, 1995), there is the potential of applying NMR spectroscopy to probe protein structure. Proteins tumble slowly in solution, but their side-chain residues are characterized by fast local motions (Breitmaier and Voelter, 1987a). Technical advances have made it possible to implement relaxation measurements which separate these dynamic components and allow for the study of internal dynamics of proteins and other large molecules (Berglund *et al.*, 1995). Motional studies have normally made use of ^{13}C, ^{15}N and ^{2}H nuclei (Fersht, 1985; Li *et al.*, 1996) but not ^{14}N nuclei, due to the problem of broad resonance lines. In one case, however, dynamic [^{14}N]NMR experiments were successfully carried out to probe the microenvironment of an Nε-trimethyllysyl residue in calmodulin (Zhang *et al.*, 1994). Galanter and Labotka (1991) also studied the binding of [^{14}N]nitrate to a transport protein by measuring the increase of linewidth upon association.

Of the above examples, Galanter and Labotka (1991), and Zhang *et al.* (1994) successfully applied [^{14}N]NMR techniques to study proteins. With respect to this chapter, however, the work of Zhang *et al.* (1994) was particularly relevant. The proteins they investigated contained many amino groups which were observable as broad lines. Extensive broadening of nitrogen resonances

of amino groups has been attributed to the effects of deprotonation (Gerothanassis *et al.*, 1987; Tzalmona and Loewenthal, 1974; Cohen *et al.*, 1975). As expected, these lines broadened in increasingly alkaline solutions. Broad lines would normally have limited the scope of the investigation, however, the proteins they used, namely cytochrome c and calmodulin, were exceptional in that both had one Nε-trimethylated lysyl residue present as a post-translational modification. As expected, and unlike the other nitrogen resonances, the linewidth corresponding to the trimethyllysyl resonance did not show any pH dependency. Similar results were obtained with the [^{14}N]NMR analysis of free amino acid analogues, namely lysine, Nε-methyllysine, Nε-dimethyllysine, and Nε-trimethyllysine. More importantly, the trimethyllysyl residues of calmodulin and cytochrome c were observed as narrow resonance lines which made motional NMR studies of the region surrounding the lysyl residue possible. While Zhang and coworkers demonstrated that trimethylammonium groups and [^{14}N]NMR spectroscopy could be used to study protein dynamics, the protocol was not generally applicable since it required that proteins be post-biosynthetically trimethylated.

In chapter 3 a nonaqueous chemical modification procedure was presented for trimethylating amino groups in proteins with iodomethane. As in the case of reductive methylation, trimethylation of amino groups proceeds with retention of structure and bioactivity under appropriate reaction conditions. Nonaqueous protein methylations were carried out for up to 12 h at 75°C (Taralp and Kaplan, 1997) and even higher temperatures, with no indication of peptide bond cleavage by [^{13}C]NMR analysis. In fact, *in vacuo* methylated α-chymotrypsin was fully active when steps were taken to protect the active site prior to reaction with iodomethane (Taralp and Kaplan, 1997). Previous evidence also supports that enzymes retain structure and

activity in nonaqueous environments, at temperatures as high as 100°C (Zaks and Klibanov, 1984; Zaks and Klibanov, 1988). By retention of structure, it is implied that once protein is derivatized, the trimethylammonium groups which are formed have the potential to faithfully report on the microenvironment of the corresponding amino groups from which they were prepared. This is significant since the study of trimethylammonium groups by NMR spectroscopy need not be limited to chemical shift measurements of sharp lines but opens the door to motional studies such as those conducted by Zhang (1994) or studies of local environment such as those conducted by Galanter and Labotka (1991), and can in principle be applied to any protein system. It follows, therefore, that advantage can be taken of iodomethane to: (i) monitor the reaction of amino functional groups in proteins using [^{14}N]NMR spectroscopy; (ii) characterize derivatized Nα and Nε amino acid residues by [^{14}N]NMR spectroscopy and (iii) obtain information on the local dynamics and microenvironment of the amino group from which the trimethyl ammonium derivative was originally prepared, using conventional NMR techniques.

Focus of the present study

From the purely analytical viewpoint, trimethylated ammonium derivatives should enable the identification of Nα and Nε amino groups by a very facile procedure. The methylation approach is therefore likely to demonstrate some merit, and while [^{13}C]NMR spectroscopy has been similarly applied (Taralp and Kaplan, 1997) using costlier [^{13}C]iodomethane, the much larger chemical shift dispersion but similar linewidth of nitrogen nuclei (Mann, 1978a) compared to carbon nuclei in trimethylamino acids suggests that an easier resolution should be possible using [^{14}N]NMR spectroscopy. The potential use of this procedure is best illustrated by virtue of example. In a

particular case of chapter 2, an N-terminal trimethylammonium ^{13}C-resonance, apparently common to four different methylated proteins, could not be identified by [^{13}C]NMR spectroscopy. Quite clearly then, the ability to exploit two different nuclei to identify amino acids by NMR spectroscopy has a better chance of success.

The immediate focus of this initial investigation was to study the reaction of iodomethane with amino acids and amino groups of protein by [^{14}N]NMR spectroscopy, and identify the reacted amino groups. The test proteins presented are bovine insulin, α-chymotrypsin, chicken lysozyme and horse cytochrome c.

Rationale

Proteins can be reacted with [^{12}C]iodomethane *in vacuo* to form distinct acid stable trimethyl ammonium derivatives of Nα and Nε amino groups of amino acid residues. As previously shown by Zhang *et al.* (1994), these derivatized amino acid residues should be observable by [^{14}N]NMR spectroscopy as sharp resonance lines. As the focus of the investigation is to identify various amino acids of protein by the [^{14}N]NMR chemical shift of their trimethylated amino groups, very sharp peak resonances are desired to maximize the chances of success. The linewidths obtained for *in vacuo* derivatized protein are expected to be comparable to that measured by Zhang and coworkers for calmodulin (23 Hz). While this may be narrow enough to carry out the experiment successfully, Zhang *et al.* (1994) had also noted that the free Nε-trimethylated amino acid derivative of lysine had even sharper lines (3 Hz). Since it is so crucial that different peak resonances do not superimpose for this approach to succeed, an acid hydrolysis of the protein following methylation will be undertaken to liberate the derivatized amino acid residues. The free

trimethylated amino acids are then analyzed by [^{14}N]NMR spectroscopy and assigned by the aid of trimethylated standard amino acids. A last requirement is that the NMR spectra be acquired at high pH values. This is necessary since other amino group resonances, which are broad and present in large excess following hydrolysis, may complicate the spectrum. At high pH values, however, these additional resonances become too broad to be observed and disappear into the noise, leaving only the sharp resonance lines of the trimethylated amino acids.

[^{14}N]NMR background information

The ^{14}N resonance position distinguishing the environment of NH_4^+ from NO_3^- in the compound NH_4NO_3 was one of the first chemical shift differences to be reported (Proctor and Yu, 1950). Since then, ^{14}N nuclei have generally been used in chemical shift studies on natural abundance samples or quadrupolar relaxation studies to obtain structural information (Witanowski *et al.*, 1977a) and information on molecular dynamics (Mann, 1978a). The NMR properties of the ^{14}N nucleus have been reviewed (Mann, 1978a; Witanowski and Webb, 1972a; Mooney and Winson, 1969). With a natural abundance of 99.63 % and a relative gyromagnetic ratio of 7.22×10^{-2} with respect to the proton, ^{14}N nuclei have a relative sensitivity of 10^{-3} with respect to the proton and 5.69 with respect to naturally abundent ^{13}C nuclei (Witanowski and Webb, 1972a). Although ^{14}N signals are easily detected using modern pulse NMR techniques, the observed linewidths are often large.

(i) Electric field gradients. All spin >1/2 nuclei have (i) an electric quadrupole moment and associated electric field gradient at the nuclear site, which originates from electrical asymmetries

in the molecule, and (ii) a magnetic dipole moment which is found in spin 1/2 nuclei (Reisse, 1982; Harris, 1983a; Drago, 1992d). Consequently, nuclear spins of quadrupolar nuclei interact with magnetic fields to give rise to quadrupolar and Zeeman energy terms, respectively (Harris, 1983a). Electric field gradients vary in magnitude as the inverse cube of the distance from a charge and therefore, those gradients which impinge on atomic nuclei must originate from charge density variations that are near the nucleus (Akitt, 1972). For the reason that electric field gradients fall off rapidly with distance, the only two factors which should normally influence the electric field gradient about a particular nucleus are the arrangement of neighbouring bonds and the electron density associated with each bond near the nucleus. A case in point is the tetramethylammonium ion versus benzyltrimethylammonium ion example, where both ions have similar electronic distributions and electric field gradients near the nitrogen center, as evidenced by the reported linewidths of 7 and 12 Hz, respectively (Akitt, 1972)

(ii) Quadrupolar relaxation and linewidths. The quadrupole moment interacts with the electric field gradient to provide a torque on the nucleus, and this has an associated energy term. Due to molecular motion, the electric field gradient fluctuates as a function of time and this induces energy transitions within the nucleus, leading to efficient relaxation (Reisse, 1982). ^{14}N nuclei have a spin quantum number (I) of 1 and a fairly large quadrupole moment compared to deuterium (Witanowski and Webb, 1972a). Consequently, relaxation is rapid and ^{14}N linewidths ($\Delta\nu_{1/2}$) are generally on the order of 100 Hz to 1000 Hz, depending on the electronic symmetry about the nitrogen nucleus. The chemical shift range of a given class of nitrogen compounds is

usually only on the order of 50 ppm, and as such, broad lines in close proximity are often unresolvable (Mann, 1978a).

A simplified expression (1) that incorporates the extreme narrowing condition (i.e. $\tau << 1/w_0$) can be used to describe the expected linewidth (where $\Delta\nu_{1/2} = 1/T_{2Q}$) of quadrupolar nuclei of relatively small molecules in nonviscous liquids and in the absence of line-broadening contributions due to chemical exchange phenomena (Harris, 1983a). The linewidth ($\Delta\nu_{1/2}$ in radians), spin-spin relaxation time (T_{2Q}), spin-lattice relaxation time (T_{1Q}), statistically averaged molecular correlation time (τ_c), electric field gradient (eq) and its deviation from axial symmetry (η), and nuclear quadrupole moment (eQ) are all related as shown below (Witanowski and Webb, 1972b; Gerothanassis *et al.*, 1987).

$$1/T_{1Q} = \boxed{\Delta\nu_{1/2} = 1/T_{2Q} = 3\pi/2(1 + \eta^2/3)\,(eqeQ/h)^2\,\tau_c \qquad \mathbf{1}}$$

The linewidth ($\Delta\nu_{1/2}$) is a function of the nuclear quadrupole moment, electric field gradient, and correlation time. The rotational contribution (i.e. τ_c) to linewidth is normally the least important consideration since ^{14}N nuclear relaxation is usually dominated by the quadrupolar coupling constant term (eqeQ/h) (Drago, 1992a). In the case of tetraalkylammonium salts with narrow lines, such is not the case.

(iii) Narrow lines. ^{14}N relaxation times are typically no greater than 100 ms and this explains why the linewidth, being inversely related to T_{2Q}, is often quite broad (Mann, 1978b). However, there are known examples of extremely narrow lines. It is apparent from expression **1** that if the electric field gradient (eq) is very small, as in the case of extreme electronic symmetry around a

nucleus, relaxation times become extremely long and linewidths extremely short. Some examples include trigonal planar (nitromethane and NO_3^-) (Witanowski and Webb, 1972c) and tetrahedral (R_4N^+) (Mann, 1978b; Witanowski and Webb, 1972d) nitrogen compounds. While nitromethane and nitrate do not appear to have three dimensional symmetry, they must have electronic symmetry in order to give rise to such narrow lines upon observation. The ease of which electronic symmetry can be distorted is evident by the fact that ion pairing of tetraalkyl ammonium salts is enough to broaden lines (Harris, 1983a). This sensitivity to environment is the basis for the nitrate binding experiments which were carried out by Galanter and Labotka (1991).

(iv) Other relaxation mechanisms. In the absence of a strong electric field gradient, quadrupolar relaxation is not efficient and participates very little in the overall relaxation process. Other relaxation processes, which are less efficient and normally associated with spin one-half nuclei, become the principle factors which define the linewidth (Witanowski *et al.*, 1977b). As with spin 1/2 nuclei, dipole-dipole interactions are therefore another possible mechanism of relaxation. Every nuclear spin generates a local magnetic field upon molecular motion. The interaction of another local magnetic field with the magnetic dipole of ^{14}N has an associated energy term. Molecular motion fluctuates the local magnetic field and this promotes nuclear energy transitions and leads to the relaxation of the nucleus (Breitmaier and Voelter, 1987b). This process is most efficient when the fluctuations occur at w_o, the Larmor frequency (in radians). The relaxation of $ND_4^+Cl^-$ below 220 K, for example, was dominated by intramolecular N-D dipole-dipole interactions. The example serves to illustrate that even two quadrupolar nuclei can have dipolar coupling interactions. In aqueous solutions, intermolecular dipole-dipole interactions

relax nitrogen nuclei of ammonium salts by the reorientation of water dipoles and their associated local magnetic fields (Witanowski *et al.*, 1977b). Both inter and intramolecular dipolar mechanisms can contribute to the relaxation of tetraalkylammonium salts.

(v) Correlation times and line broadening in proteins. The correlation time (τ_c) is used to model the bulk loss of rotational correlation (i.e. the randomization of the ensemble) due to Brownian motion, as a function of time (Abragam, 1983). Analytically, τ_c is a constant which is used to model the auto-correlation and spectral density functions, as defined by a statistical treatment of the loss of correlation by the ensemble of nuclear spins (Harris, 1983b). These functions have been used to calculate Zeeman level transition probabilities of nuclei and determine the dependence of various relaxation processes on correlation time (τ_c) (and hence, molecular tumbling) (Reisse, 1982; Drago, 1992b). All relaxation processes have a dependency on correlation time. In addition, the efficiency of longitudinal relaxation (the T_1 process) is affected by the external magnetic field strength (Reisse, 1982; Harris, 1983b). While not apparent from equation **1**, T_1 is most efficient when $\tau_c = 1/w_0$. The longitudinal relaxation rate in the rotating frame ($T_{1\rho}$) also passes through a maximum at a larger correlation time (τ_c) than the T_1 process. In contrast, transverse relaxation (T_2), which affects the linewidth, becomes increasingly efficient as correlation times (τ_c) (and hence, molecular weights) increase so that large molecules tend to give broader lines. For example, in the case of free trimethylysine (Zhang *et al.*, 1994), the average tumbling rate was too fast to promote efficient relaxation, and the linewidths were narrow. In contrast, the same trimethyllysyl derivative in much larger and slower tumbling

macromolecules such as cytochrome c and calmodulin gave linewidths that were an order of magnitude greater. These biomolecules had much longer correlation times.

6.2 Materials and Methods

Proteins. Bovine insulin, α-chymotrypsin, chicken lysozyme and horse cytochrome c were purchased from Sigma Chemical Company.

Chemicals and solvents. [^{12}C]Iodomethane was purchased from Sigma Chemical Company. All other chemicals, reagents and solvents were high purity preparations obtained from commercial sources.

Trimethyl amino acid standards. Trimethylglycine (betaine) was purchased from BDH chemicals. Precursors of amino acid standards were purchased from Sigma Chemical Company. Phenylalanine, polylysine hydrobromide, isoleucine, alanine, cystine, lysine, threonine and valine were dissolved (50mg) in distilled water (3 ml) and the pH was adjusted to 10-11 with 5N NaOH. [^{12}C]Iodomethane (50 ml) was added and the tightly capped tube was shaken at 37°C for 48 h. Base was added periodically to maintain the pH value in the range of 10-11. In the case of polylysine, methylation was followed by dialysis. The polylysyl derivative was dissolved in 6N HCl and subjected to acid hydrolysis at 110°C for 24 h.

***In vacuo* methylation of proteins.** Lysozyme, insulin or cytochrome c (20 mg) were dissolved in 40 mM sodium metaborate (1ml) and adjusted to pH 9.2, 9.2 and10, respectively, with 5N

NaOH. Chymotrypsin (100 mg) was dissolved in water (3 ml), cooled, and the pH quickly adjusted to 10.8 with 5N NaOH. Each solution was made up in a single-compartment pyrex reaction vessel, followed by flash freezing and lyophilization. The tubes were narrowed by flame at the mid-point and [^{12}C]iodomethane (25 ml; 100 ml for chymotrypsin) was transferred into the tube. The top of the reaction vessel was closed with a screw cap and the bottom half was submersed in liquid nitrogen. The screw cap was removed, the end of the tube was fitted with a vacuum hose and the vessel was sealed *in vacuo*. The reaction vessel was incubated in a block heater at 75°C for 24h. At the end of the reaction, the tubes were scored. Prior to opening the vessel, a rubber septum with a hole cut through the rubber was fitted onto the top of the vessel and the unreacted iodomethane was frozen onto the glass at the top by pouring liquid nitrogen into the cavity (chapter 3). The vacuum seal was broken and the modified protein was removed.

Acid hydrolysis of proteins. The protein derivatives were transferred into another pyrex hydrolysis tube and 6N HCl (1 ml) was added. The resultant suspension was frozen, degassed and sealed *in vacuo* by flame. The tubes were incubated at 110°C for 24 h, opened, and the contents removed under vacuum in a dessicator containing sodium hydroxide pellets.

NMR spectra. [^{14}N]NMR spectra were obtained using a Varian XL-300 MHz spectrometer. A nonmagnetic, 200 picofarad capacitor was inserted to lower the tuning range of the probe to accomodate the ^{14}N-resonance condition (21.7 MHz). Methylated amino acid samples and protein hydrolysates were analyzed in D_2O with sufficient 5N NaOH added to give a pH meter reading of 11. Unless otherwise stated, 512 scans were acquired for the analysis of standards. A

1 Hz exponential linebroadening function was applied to the free induction decay prior to Fourier transformation. A 0.8 s acquisition time, 90° pulse (36 μs) and 0 s relaxation delay were used. Observed linewidths ($\Delta\nu_{1/2obs} = \Delta\nu_{1/2} + 1$ Hz) were measured directly from the spectra. The spectral window was 12000 Hz. Only proton decoupled spectra were obtained. Tetramethylammonium chloride (3 mg) served as internal standard and was positioned at 0 ppm. Tetramethylammonium chloride is 333.5 ppm upfield of neat nitromethane, the most accepted standard (Witanowski and Webb, 1972d). One spectrum of methylated and hydrolyzed lysozyme was acquired at p^2H 1 in the absence of tetramethylammonium chloride. The spectrum of ethylenediamine was obtained at p^2H 2.

6.3 Results.

The linewidths of trimethylammonium derivatives at p^2H 11 (Table 6.3.1) were an order of magnitude narrower than the measured linewidth of ethylenediamine at p^2H 2. Still narrower, was the linewidth of tetramethylammonium chloride. Of the handful of amino acid derivatives, linewidths were narrowest for those with small alkyl side-chain residues. The measured linewidths were greater for amino acid derivatives with bulkier alkyl side-chains such as isoleucine and valine, and greater still for side-chains with more complicated functional groups such as threonine and phenylalanine. Cystine, with two amino groups, had the broadest lines. It should be pointed out that in every case, the linewidth of tetramethylammonium chloride served to verify that shimming among samples was uniform. Of the protein hydrolysates analyzed by NMR spectroscopy (Table 6.3.1 and all figures), each spectrum appeared to have a chemical shift value characteristic of Nε-trimethyllysine. Apart from the reference signal, there were no other

Table 6.3.1: [^{14}N]chemical shifts (in ppm) and linewidths (in Hertz) of trimethylammonium derivatives of amino acid standards, tetramethylammonium chloride and ethylenediamine; Chemical shifts of methylated cytochrome c, insulin, lysozyme and α-chymotrypsin.

[^{14}N] Standard compounds	Chemical shifts (left) and linewidths (right) of standard trimethylamino acids, tetra-methylammonium chloride and ethylenediamine		Chemical shifts observed in			
	δ (ppm)	$\Delta\nu_{1/2}$ (Hertz)	Cytochrome c reacted at LpH 10	Insulin reacted at L pH 9.2	Lysozyme reacted at LpH 9.2	α-Chymotrypsin reacted at LpH 10.8
$[Me_3N^+$-Cys-S-$]_2$	(14.70, 9.98)	15.2, 8.0	-	-	-	NO
Me_3N^+-Val	11.41	4.3	-	-	-	-
Me_3N^+-Ile	11.27	4.2	-	-	-	11.28
Me_3N^+-Phe	(9.70)	8.0	-	(10.24)	-	-
Me_3N^+-Ala	9.12	3.0	-	-	-	9.145
Me_3N^+-Thr	9.09	(7.6)	-	-	-	-
H_2N-Lys(ε-$^+NMe_3$)	5.77	2.7	5.77	5.75	5.77	5.76
Me_3N^+-Lys(ε-$^+NMe_3$)	9.25	(3)	-	-	NO	-
Me_3N^+-Lys(ε-$^+NMe_3$)	5.75	3.2	-	-	NR	-
Me_3N^+-Gly	4.62	2.8	-	4.61	-	4.577
Me_4N^+	0.00	1.8	0	0	0	0
$H_3N^+CH_2CH_2^+NH_3$	-	42[‡]				

[‡]measured at p^{2}H 2; NO = expected but not observed; NR = not resolved.
linewidths were measured directly from each spectrum without removing the 1 Hz linebroadening function.
parenthesis indicate values may be inaccurate; hyphens indicate the measurement is not applicable.

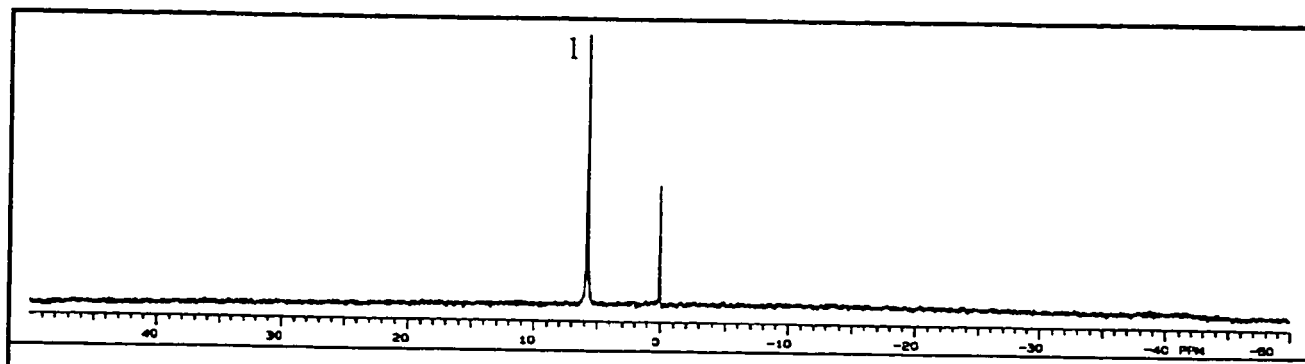


Figure 6.3.1: 21.7 MHz [^{14}N]NMR spectrum of cytochrome c following methylation at LpH 10 and acid hydrolysis. 10000 scans were acquired. Peak resonance 1 corresponds to H_2N-Lys(ε-$^+NMe_3$).

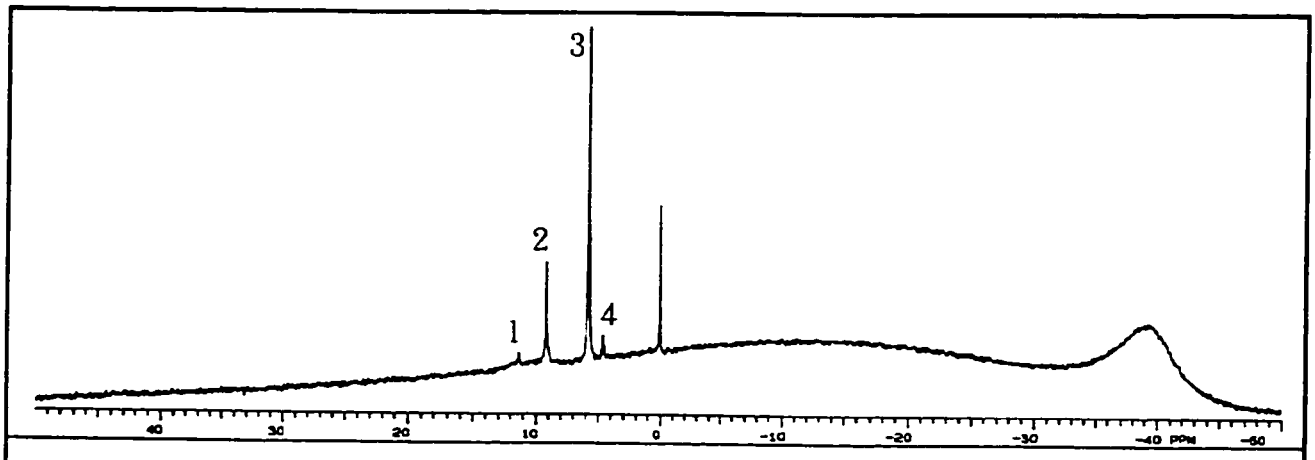


Figure 6.3.2: 21.7 MHz [^{14}N]NMR spectrum of α-chymotrypsin following methylation at LpH 10.8 and acid hydrolysis. 62000 scans were acquired. Peak resonances tentatively correspond to the following methylated amino acids: 1 - Me_3N^+-Ile; 2 - Me_3N^+-Ala; 3 - H_2N-Lys(ε-$^+NMe_3$); 4 - Me_3N^+-Gly.

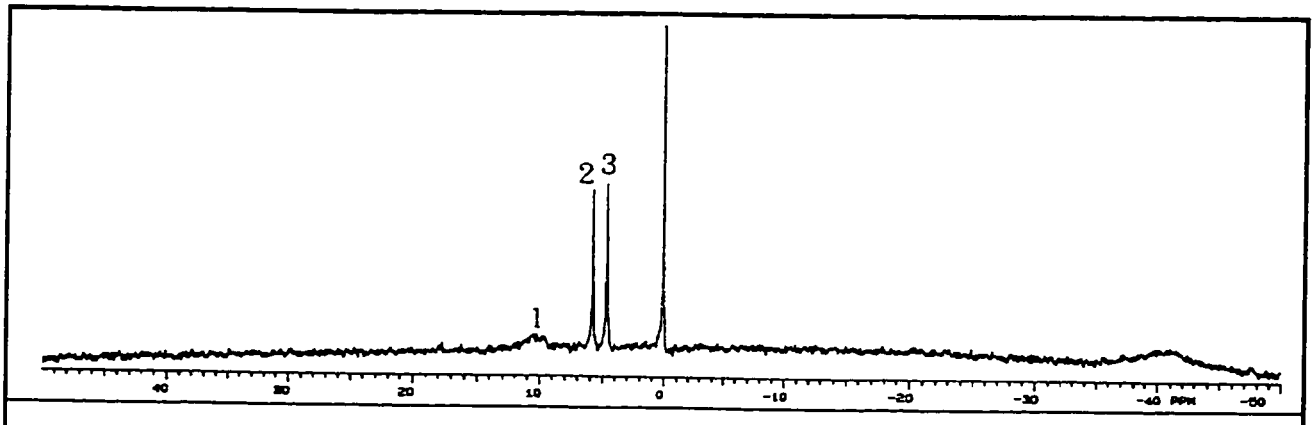


Figure 6.3.3: 21.7 MHz [^{14}N]NMR spectrum of insulin following methylation at LpH 9.2 and acid hydrolysis. 45000 scans were acquired. Peak resonances tentatively correspond to the following methylated amino acids: 1 - Me_3N^+-Phe; 2 - H_2N-Lys(ε-$^+NMe_3$); 3 - Me_3N^+-Gly.

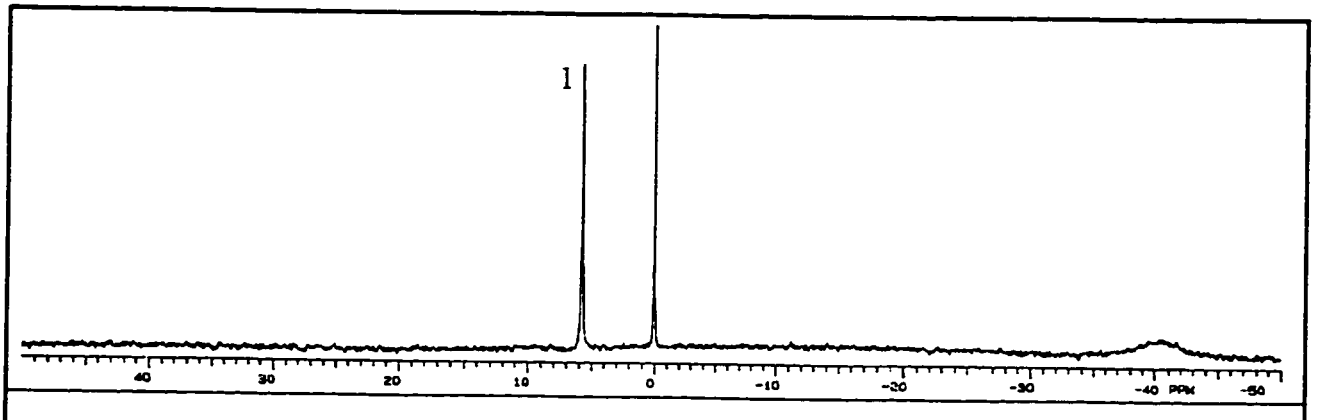


Figure 6.3.4: 21.7 MHz [^{14}N]NMR spectrum of lysozyme following methylation at LpH 9.2 and acid hydrolysis. 43500 scans were acquired. Peak resonance 1 corresponds to H_2N-Lys(ε-$^+NMe_3$).

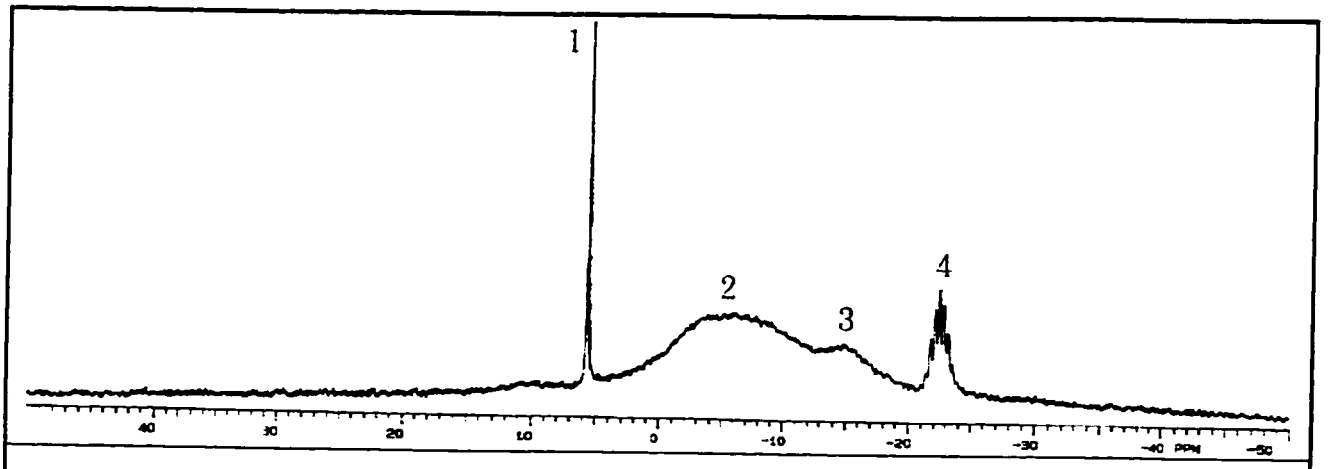


Figure 6.3.5: 21.7 MHz [^{14}N]NMR spectrum of lysozyme following methylation at LpH 9.2 and acid hydrolysis. 43500 scans were acquired at p^{2}H 1. Peak resonances tentatively correspond to the following: 1 - H_3N^+- Lys(ε-$^+$NMe_3); 2,3 - D_3N^+-amino acid resonances; 4 - $D_4N^+Cl^-$.

resonance peaks observed for the hydrosylate of either cytochrome c (Figure 6.3.1) or lysozyme (Figure 6.3.4). The hydrosylate of α-chymotrypsin (Figure 6.3.2) gave four peak resonances while insulin (Figure 6.3.3) gave two distinct peak resonances and one broad resonance in the region presumably occupied by trimethylamino acids (peak 1). Broad underlying resonances were observed, particularly in the case of α-chymotrypsin (Figure 6.3.2), but the distortion of the spectrum was not problematic. As shown in Figure 6.3.5, the spectrum of lysozyme, when deliberately acquired under acidic conditions, was significantly more difficult to interpret.

6.4 Discussion.

6.4.1 Standards

The ammonium ion (H_4N^+, not shown) is one exceptional example of a nitrogen nucleus with sharp lines, typically on the order of 20 Hz at neutral pH values (Zhang *et al.*, 1994). This is possible because protonation forces a near perfect symmetrical electronic geometry about the

nitrogen nucleus and greatly reduces the efficiency of the quadrupolar relaxation mechanism. The principle is supported by [^{1}H]NMR spectroscopy measurements at low pH values (not shown) where amino protons, which are not normally coupled to nitrogen, appear as triplets with intensity 1:1:1. Another reason the nitrogen resonance is so sharp is because the ammonium ion is very small, tumbles rapidly, and has an extremely short correlation time (see theory). The predominant mechanism of relaxation in this example is unclear. Normally the quadrupolar mechanism is the only determinant of linewidth, however, in the case of narrow nitrogen resonances the linewidths are probably determined by minor contributions from several factors. Proton exchange (Drago, 1992c) may cause minor distortions of tetrahedral symmetry. With the temporary loss of symmetry, an electric field gradient could be established and this would allow for some quadrupolar relaxation to take place (Reisse, 1982). An intramolecular dipole-dipole relaxation mechanism is probably another major process, where a locally fluctuating magnetic field of a bound proton directly interacts with the nuclear magnetic moment of nitrogen to promote relaxation (Breitmaier and Voelter, 1987b). Last is the intermolecular dipole-dipole mechanism which involves the reorientation of water dipoles.

The trimethylation procedure made possible the observation of exceptionally sharp [^{14}N]NMR resonances. Ethylene diamine (Table 6.3.1) was chosen as a comparative model to emphasize the significance of quaternization of amino groups for obtaining sharp lines. Ethylenediamine is close to the average molecular weight of amino acids, and it is therefore expected that there will be a similar contribution of the molecular correlation time (τ_c) to the linewidths in both ethylenediamine and the amino acids. When acquired at p^{2}H 2 where amines exist almost exclusively as the ammonium ion, the measured linewidth was 42 Hz. At higher p^{2}H

values, where the resonance is a composite and weighted average of acidic and basic species, the linewidth would be expected to be much greater. In the case of the trimethylamino acids, however, the nitrogen atom was quaternized using nondissociable methyl groups, thus eliminating the possibility of deuterium ion exchange at all p^2H values, and the loss of tetrahedral symmetry at high p^2H values. The latter characteristic is important since it makes possible the observation of sharp lines at p^2H values where all other amino groups are in the base form. The loss of tetrahedral symmetry which accompanies the removal of the deuterium ion from nonquaternary alkylamino groups promotes very efficient quadrupolar relaxation (see theory) so these lines are generally too broad to be observed.

Looking to Table 6.3.1, in which spectra were acquired at p^2H 11, the sharpest line corresponded to the resonance of tetramethylammonium chloride, which, due to its near perfect symmetry, relative small size, and nondissociating groups had a linewidth of 1.8 Hz. The predominant relaxation mechanism is presumably of a dipole-dipole nature. The interaction of a local fluctuating magnetic field with the nitrogen magnetic moment can be intramolecular, namely, between proton and nitrogen, or intermolecular, between water dipoles and nitrogen. Trimethylglycine, trimethyllysine and trimethylalanine also produced very sharp lines of 2.8, 2.7 and 3.0 Hz, respectively. The two lines for hexamethyllysine were sharp as well. These lines are broader than that observed of tetramethylammonium chloride (1.8 Hz) since the perfect electronic symmetry is distorted by the adjacent alkyl moiety. The next broadest lines corresponded to trimethylisoleucine and trimethylvaline with linewidths of 4.2 and 4.3 Hz. The presence of a bulkier side-chain appears to disrupt the electronic symmetry more effectively, however it is difficult to comment on whether the observed difference in linewidth is significant. Still broader

lines corresponded to trimethylthreonine and trimethylphenylalanine, with linewidths of 7.8 and 8.0 Hz, respectively. It is noteworthy to point out that while trimethylthreonine is isosteric with trimethylvaline, the former linewidth is double. The largest linewidths corresponded to the methylated cystine standard. Two resonances were observed in the spectrum. Without additional evidence, it was difficult to be certain, however, it is believed that one of the resonance peaks corresponds to N-trimethylcystine while the other corresponds to N,N'-hexamethylcystine. This interpretation is consistent with the observation that the resonances were too narrow at high p^2H values to be anything other than trimethylammonium groups, and is in keeping with the tendency of nitrogen nucleophiles to quaternize by reaction with iodomethane (March, 1985).

While it is difficult to comment on the reasons for the observed differences in linewidth, the line width in each case is ultimately a reflection of the degree of electronic symmetry about the nitrogen nucleus. The linewidth dependence may also have something to do with the local correlation time for rotation of the trimethylammonium moiety. The line width differences are probably not due to (i) differences in the overall correlation time for molecular tumbling, since the compounds have very similar molecular weights, or (ii) conformational changes in the molecule which give rise to multiple ^{14}N lines that happen to exchange slowly with one another. Ionizable groups, such as the hydroxyl group of threonine, are also not likely to explain the results since at pH 11, the equilibrium would highly favour one of the species and the other species would not give rise to a resonance. In the case of cystine, conformational exchange may be a factor which gives rise to broad lines (Harris, 1983c). If rotation of two cysteine units about the cystine disulphide bond is moderately fast on the NMR time scale, the observed peak will be a composite of at least two semi-resolved conformations. In principle, this would place the exchange

phenomena in the near-fast exchange region (Drago, 1992f). While this hypothesis was not verified, it could have been easily tested by varying the acquisition temperature or magnetic field strength. An increase of temperature would be expected to sharpen the lines since the exchange rate would increase, while the separation of different conformations, in Hertz, would be unchanged. Conversely, an increase of magnetic field strength would increase the separation (in Hertz) between different conformations, but would not affect the exchange rate. The exchange process would become even less averaged, and the peak would be expected to broaden (Breitmaier and Voelter, 1987c).

The chemical shift values (Table 6.3.1) of each of the trimethylamino acid standards were distinct, and with few exceptions, quite dispersed. It is interesting to note that while tetramethylammonium chloride is most upfield with respect to the trimethylamino acids, the trend is reversed in ^{13}C-spectra (chapter 2). The wider ^{14}N-chemical shift dispersion of trimethylamino acids with respect to ^{13}C analogues appears to be due to the inherently greater sensitivity of the nitrogen nucleus to environmental effects. The fact that nitrogen is one atom closer than its methyl groups to the side-chain residues may play a role, but the relationship is probably not straightforward. Nevertheless, the results suggest that when combined with the use of [^{13}C]NMR spectroscopy (chapter 2), it is almost certain that iodomethane can be used for analytical purposes. It was noted above that the linewidths of various trimethylamino acid derivatives were variable. This is significant since it suggests that amino acids can be identified on the basis of both chemical shift and linewidth measurements. This rationale was applied to various test proteins (*vide infra*), and although an acid hydrolysis was necessarily carried out in order to obtain sharp resonance lines, the procedure was not any more laborious with respect to the ^{13}C-approach, as

acid hydrolysis was eventually incorporated into the ^{13}C-protocol to promote the analysis of unknown amino termini of insulin (chapter 2). For all proteins tested (*vide infra*), peak assignments were only tentative and require confirmation. The list of trimethylamino acid standards is incomplete so there is always the danger that other trimethylamino acid derivatives of protein, or for that matter, ammonium derivatives in general, could be misassigned as one of the trimethylamino acid standards currently available.

6.4.2 Model protein hydrosylates analyzed by [^{14}N]NMR spectroscopy

The chemical shift resonances of methylated and hydrolyzed proteins correlated very well, in the majority of cases, to trimethylamino acid standards representing the expected functional groups (Table 6.3.1). Cytochrome c (Figure 6.3.1) has a blocked Nα-terminus and therefore only one resonance, namely the superimposed peaks of many Nε-trimethyllysine derivatives, was expected (peak 1). While sensitivity of the method may have prevented the observation of a less abundent trimethylammonium derivative, such as a reacted N-terminus, there were, nevertheless, no other resonances observed. The poorer sensitivity of the [^{14}N]NMR protocol with respect to the [^{13}C]NMR protocol was obvious from the 10000 scans required to obtain a good spectrum. Indeed, when an equal amount of cytochrome c was analyzed by [^{13}C]NMR spectroscopy, there were, in fact, minor impurities observed in the Nα amino group region of the spectrum. This was readily seen after only 1024 scans.

In the case of α-chymotrypsin, 100 mg of protein was derivatized to assure that N-termini would be observed. The protein was lyophilized without buffering in order to minimize the amount of salt present, since it was observed that other samples became quite warm during the

NMR acquisition. While additional buffer has previously led to shimming problems, it was observed following analysis (Figure 6.3.2) that in its absence, the reaction did not proceed to completion as evidenced by the weak signals obtained for the amount of protein used. Another observation which suggested that the reaction was incomplete was the very large relative abundance of broad nitrogen resonances, which produced a distortion of the baseline at p^2H 11. Such a pronounced distortion should not have been observed had the trimethylamino acid signals been more intense. α-Chymotrypsin has three N-termini which commence with alanine, isoleucine and a half cystine residue. Four trimethylamino acid resonances were observed. The standards for the isoleucyl (peak 1), alanyl (peak 2) and lysyl (peak 3) derivatives were all very close to observed peaks of the protein sample (Table 6.3.1). The last resonance (peak 4) appeared to have a weaker correlation to the glycyl standard. While glycine is not found at the N-terminus of α-chymotrypsin, contamination or proteolytic processing cannot be ruled out. This explanation is reasonable since the protein was not dialyzed after methylation. Furthermore, since the cystine standards were broad, it is likely that the resonance peak corresponding to it in the protein sample was too broad to be observed. It is weakly possible, however, that the cystine standard was improperly interpreted, and that the observed resonance (peak 4) at a chemical shift value close to that of trimethylglycine was in fact the cystine chemical shift. This would be consistent with the fact that chemical shift values of sulphur containing standards were often difficult to interpret by [^{13}C]NMR spectroscopy (chapter 2).

The results obtained for insulin (Table 6.3.1 and figure 6.3.3) were quite conclusive since the spectrum obtained was consistent with the expected chemical shifts and relative linewidths of trimethylphenylalanine (peak 1), Nε-trimethyllysine (peak 2) and trimethylglycine (peak 3). The

correlation between the trimethylphenylalanine standard and the broad peak (peak 1) whose midpoint was at 10.24 ppm was not as strong as in the other cases and may be due to the weakness of the signal, which made it difficult to observe the resonance. The unidentified resonance of insulin in chapter 2 was not apparent here, but as stated before, it could be due to a sensitivity problem. Alternatively, it could be the downfield portion of peak 1.

The last protein analyzed by [^{14}N]NMR spectroscopy was lysozyme (Figure 6.3.4). Lysozyme has a free N-terminal lysyl residue but the Nα resonance was not observed, again, due to the sensitivity problem. The spectrum only contained one peak which was presumably made up of many superimposed Nε-trimethyllysine resonances (peak 1). Since the resonance position of the Nε-trimethyl moiety of hexamethyllysine is so close to Nε-trimethyllysine (Table 6.3.1), it was not possible to comment on whether the Nε moiety of the N-terminal lysyl group of lysozyme was reacted. The effectiveness of the high p^{2}H strategy for simplifying spectra is easily observed when comparing figures 6.3.4 and 6.3.5. In figure 6.3.5, the spectrum of hydrolyzed lysozyme was acquired at p^{2}H 1. Peaks 2 and 3 correspond to amino groups of liberated amino acids (Witanowski *et al.*, 1981). These resonances would normally be too broad to observe at high p^{2}H values. The septet (Figure 6.3.5, peak 4) is actually a nonet, characteristic of deuterated ammonium chloride (Witanowski and Web, 1972d), and can be removed altogether at p^{2}H 11 (Figure 6.3.4). The ability to carry out acquisitions at different p^{2}H values provides another subtle way to obtain information. In this case, the observation of a substantial ammonium ion resonance (peak 4) was diagnostic of amino acid decomposition during acid hydrolysis.

To summarize, the results suggest that trimethylation can be used for analytical purposes. The labeling of protein N-termini by this procedure is being pursued by Helen Vakos of our laboratory as an application of the work carried out in this chapter. While the use of ^{15}N enriched amino acids and proteins would remove any need to use this technique, the method presented is easily applicable, cost and labour efficient, and retains the native structure of protein. The use of ^{15}N amino acids in growth media for the production of labeled proteins is particulary notorious for being expensive and labour intensive. Furthermore, trimethylated ammonium derivatives are still sensitive to microenvironment (Table 6.3.1) and molecular motions (Zhang *et al.*, 1994), so linewidth measurements and dynamic NMR experiments can also form the basis of some structure-function studies.

With respect to the [^{13}C]iodomethane labeling procedure of chapter 2, this method is less costly and can be used more effectively to probe structure. The observed chemical shift dispersion of different nitrogen nuclei is also greater. The [^{14}N]NMR approach is, however, significantly less sensitive with respect to protein labeled with 100% [^{13}C]iodomethane, so that in cases where hydrolysis is not carried out, practical matters such as acqusition time and protein solubility may limit the technique to the observation of superimposed Nε-trimethyllysyl residues. At any rate, large amounts of protein are required and the availability of protein may therefore be another factor to consider. This may necessitate the use of 10 mm NMR tubes to contain more protein. Another drawback (or advantage) is that only trimethylamino acid derivatives are observable by [^{14}N]NMR spectroscopy as sharp lines at high p^2H values. With [^{13}C]NMR spectroscopy, tyrosyl, methionyl, cystyl and histidyl derivatives are also observed.

6.5 Conclusions

Methylated functional groups of protein could be observed by [^{14}N]NMR as very sharp resonance lines following hydrolysis. For each trimethylamino acid synthesized, the observed chemical shift and linewidth measurements were, in combination, distinct enough to provide a means of identification using standards. The principle of simplifying the spectra, by acquiring data at high p^{2}H values, was also very useful. Methylation of protein for [^{14}N]NMR analysis can be potentially used to study local motions and environments of amino groups. When used in combination with [^{13}C]NMR, the methylation protocol is an even more useful and reliable analytical tool.

6.6 Acknowledgment

This work was supported by a Natural Sciences and Engineering Research Council grant. We are grateful to the American Technical Ceramics Corporation for donating nonmagnetic capacitors.

6.7 References

Abragam, A. (1983) *Principles of Nuclear Magnetism: International Series of Monographs on Physics* (Marshall, W. C. and Wilkinson, D. H. Eds.) pp. 270-272, pp. 297-300, Oxford Science Publications.

Akitt, J. W (1972) In *Annual Reports on NMR Spectroscopy* **5A** (Mooney, E.F. Ed.) pp. 466-467, Academic Press.

Berglund, H., Baumann, H., Knapp, S., Landenstein, R. and Hard, T. (1995) *J. Am. Chem. Soc.* **117**, 12883-12884 and references cited therein.

Blomberg, F., Maurer, W. and Ruterjans, H. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1409-1413.

Breitmaier, E. and Voelter, W. (1987a) *Carbon-13 NMR Spectroscopy: High-Resolution Methods and Applications in Organic Chemistry and Biochemistry*, 3rd. ed., p. 177, VCH Publishers.

Breitmaier, E. and Voelter, W. (1987b) *Carbon-13 NMR Spectroscopy: High-Resolution Methods and Applications in Organic Chemistry and Biochemistry*, 3rd. ed., p. 164, VCH Publishers.

Breitmaier, E. and Voelter, W. (1987c) *Carbon-13 NMR Spectroscopy: High-Resolution Methods and Applications in Organic Chemistry and Biochemistry*, 3rd. ed., pp. 128-129, VCH Publishers.

Brevard, C. and Granger, P. (1981) *Handbook of High Resolution Multinuclear NMR*, p. 104, John Wiley and Sons.

Cohen, E.A., Shiller, A.M., Chan, S.I. and Manatt, S.L. (1975) *Organic Magnetic Resonance* 7, 605-609.

Drago, R. S. (1992a) *Physical Methods for Chemists*, 2nd ed., p. 319, Saunders College Publishing.

Drago, R. S. (1992b) *Physical Methods for Chemists*, 2nd ed., pp. 326-328, Saunders College Publishing.

Drago, R. S. (1992c) *Physical Methods for Chemists*, 2nd ed., pp. 257-258, p. 290, Saunders College Publishing.

Drago, R. S. (1992d) *Physical Methods for Chemists*, 2nd ed., p. 211, Saunders College Publishing.

Drago, R. S. (1992e) *Physical Methods for Chemists*, 2nd ed., p. 241, Saunders College Publishing.

Drago, R. S. (1992f) *Physical Methods for Chemists*, 2nd ed., pp. 290-295, Saunders College Publishing.

Fersht, A. (1985) *Enzyme Structure and Mechanism*, 2nd ed., p. 33-36, W. H. Freeman and Company.

Galanter, W.L. and Labotka, R.J. (1991) *Biochimica et Biophysica Acta* **1079**, 146-151.

Gerothanassis, I.P., Karayannis, T., Sakarellos-Daitsiotis, M., Sakarellos, C. and Marraud, M. (1987) *Journal of Magnetic Resonance*, **75**: 513-516.

Harris, R. K (1983a) *Nuclear Magnetic Resonance Spectroscopy: A Physicochemical View*, pp. 130-137, Pitman Publishing Inc.

Harris, R. K (1983b) *Nuclear Magnetic Resonance Spectroscopy: A Physicochemical View*, pp. 67-68, pp. 85-93, Pitman Publishing Inc.

Harris, R. K (1983c) *Nuclear Magnetic Resonance Spectroscopy: A Physicochemical View*, pp. 118-130, Pitman Publishing Inc.

Jenks, G.J. **(1971)** *J. Chem. Phys.* **54**, 1558.

Karayannis, T., Gerothanassis, I.P., Sakarellos-Daitsiotis, M., Sakarellos, C. and Marraud, M. (1990) *Biopolymers* **29**, 423-439.

Richards, R.E. and Thomas, N.A. (1975) *J. Chem. Soc., Perkin Trans.* **2**, 368.

Mann, B.E. (1978a) In *NMR and the Periodic Table* (Harris, R.K. and Mann, B.E. Eds.) p. 96, Academic Press.

Mann, B.E. (1978b) In *NMR and the Periodic Table* (Harris, R.K. and Mann, B.E. Eds.) p. 98, Academic Press.

March, J. (1985) *Advanced Organic Chemistry: Reactions, Mechanisms, and Structure* 3rd ed., p. 364, John Wiley & Sons,.

Mooney, E.F. and Winson, P.H. (1969) In *Annual Reports on NMR Spectroscopy* **2** (Mooney, E.F. Ed.) Academic Press.

Proctor, W.G. and Yu, F.C. (1950) *Physical Reviews* 77, 716-717.

Reisse, J. (1982) In *The Multinuclear Approach to NMR Spectroscopy* (Lambert, J. B. and Riddell, F. G. Eds.) chapter 4, D. Reidel Publishing Company.

Sheinblatt, M. and Gutowsky, H.S. (1976) *J. Am. Chem. Soc.* **86**, 4814.

Taralp, A. and Kaplan, H. (1997) *Journal of Protein Chemistry* **16**, 183-193.

Tzalmona, A. and Loewenthal, E. (1974) *The Journal of Chemical Physics* **61**, 2637-2639.

Witanowski, M., Stefaniak, L. and Webb, G.A. (1977a) In *Annual Reports on NMR Spectroscopy* 7 (Webb, G.A. Ed.) p. 118, Academic Press.

Witanowski, M., Stefaniak, L. and Webb, G.A. (1977b) In *Annual Reports on NMR Spectroscopy* 7 (Webb, G.A. Ed.) p. 234, Academic Press.

Witanowski, M., Stefaniak, L. and Webb, G.A. (1981) In *Annual Reports on NMR Spectroscopy* **11B** (Webb, G.A. Ed.) pp. 257-261, Academic Press.

Witanowski, M. and Webb, G.A. (1972a) In *Annual Reports on NMR Spectroscopy* **5A** (Mooney, E.F. Ed.) p. 396, Academic Press.

Witanowski, M. and Webb, G.A. (1972b) In *Annual Reports on NMR Spectroscopy* **5A** (Mooney, E.F. Ed.) p. 409, Academic Press.

Witanowski, M. and Webb, G.A. (1972c) In *Annual Reports on NMR Spectroscopy* **5A** (Mooney, E.F. Ed.) p. 426, Academic Press.

Witanowski, M. and Webb, G.A. (1972d) In *Annual Reports on NMR Spectroscopy* **5A** (Mooney, E.F. Ed.) p. 435, Academic Press.

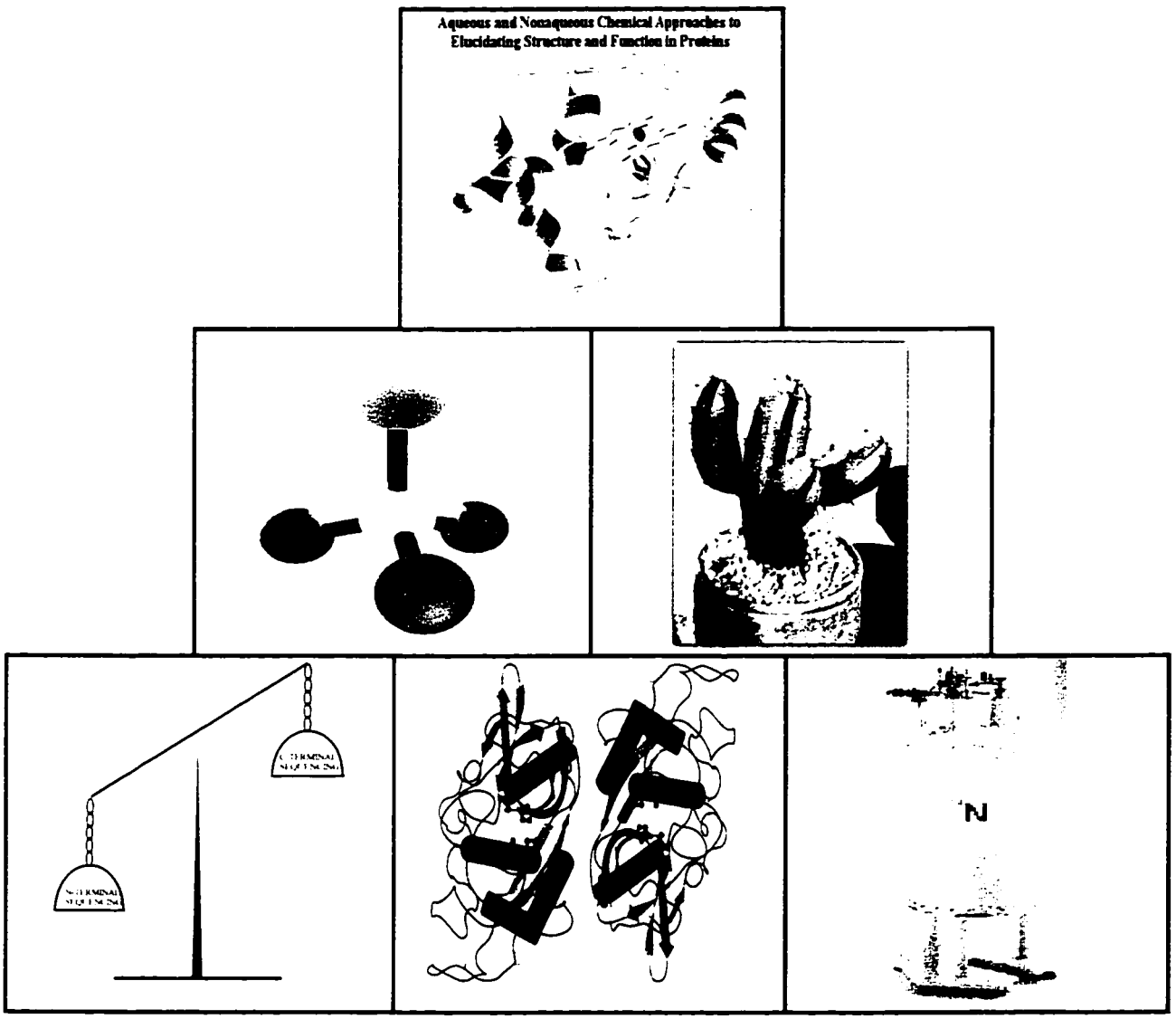
Wolfson, R. and Pasachoff, J. M. (1987) *Physics*, Little, Brown and Company.

Zaks, A. and Klibanov, A.M. (1984) *Science* **224**, 1249-1251.

Zaks, A. and Klibanov, A.M. (1988) *The Journal of Biological Chemistry* **263**, 3194-3201.

Zhang, M., Huque, E. and Vogel, H.J. (1994) *The Journal of Biological Chemistry* **269**, 5099-5105.

Zhigang, L., Raychaudhuri, S., and Wand, J. A. (1996) *Protein Science* **5**, 2647-2650.


Aqueous and Nonaqueous Chemical Approaches to
Elucidating Structure and Function in Proteins
N

General conclusions

The theme of the present work was to use chemical approaches to elucidate protein structure and function. Both aqueous and nonaqueous methods were developed and investigated. Some were quite novel. While chapters 2 through 6 examined different facets of protein chemistry, the work carried out in each chapter had one common element, namely, that a chemical approach was employed to probe structure and function, and a physical technique served to quantify the effects of the perturbation. It is noteworthy to point out that despite the fact that chemical modification and enzyme kinetics are among the oldest structure-function tools, it was still possible to carry out novel work in either area. Some of the distinctions among the chapters are discussed below.

The novelty of chapter 2 was largely based on the fact that iodomethane was not previously investigated as a suitable reagent for protein modification. It turned out that iodomethane was an exceptionally useful reagent, and made possible the synthesis of new protein derivatives for carrying out structure-function studies. The research carried out in chapter 3 and chapter 6 exploited the misconception that protein structure-function studies in organic solvents are invalid. Once Klibanov challenged and dispelled what was professed to be sound wisdom, this area of protein investigation was defined. It was clearly shown from the results of chapter 3 that a nonaqueous environment provides the necessary framework to carry out extremely useful chemical modifications, and would enable the development of numerous theoretical studies and practical applications. An environment devoid of water permitted the derivatization of small amounts of protein using nonconventional modification reagents. This feature was exploited particularly in chapter 4, where it was applied with some success to the investigation of the C-terminal problem of proteins. Chapter 5 reports the S_3 subsite specificity of cathepsin B. The enzyme was successfully characterized by acquiring kinetic rate data of the hydrolysis of substrates specifically designed to define S_3 specificity.

While established chemical approaches such as enzyme kinetics were used to investigate proteins, the work presented shows that novel chemical approaches of protein investigation can be developed and exploited to contribute to the understanding of protein structure and function.

Appendix 1: Claims to original research and publications arising from this thesis

Claims to original research

1. First demonstration that iodomethane can be used as a general probe of protein structure and function, and as an analytical tool of protein characterization.

2. First use of [^{13}C] and [^{14}N]NMR spectroscopy to characterize the reaction of protein functional groups with iodomethane.

3. First reported *in vitro* N,N,N-trimethylation of amino groups of native protein.

4. First reported *in vitro* or *in vivo* N^1,N^3-dimethylation of histidyl imidazole groups and O-methylation of tyrosyl hydroxyl groups of protein.

5. First demonstration that the nonaqueous chemical modification of lyophilized protein with reagents in the vapour phase or dissolved in organic solvents is feasible to carry out structure-function studies.

6. First direct chemical evidence that the pH memory effect postulated by Klibanov is due to the retention of ionization state of individual functional groups of protein.

7. First use of lyophilization pH values and conformational factors to achieve good selectivity for reactions involving functional groups of protein.

8. First use of nonaqueous conditions to eliminate the risk of proteolysis during the course of reaction.

9. First use of nonaqueous conditions to prevent the hydrolysis of labile reaction intermediates of protein.

10. First characterization of the S_3 subsite specificity of cathepsin B.

Refereed publications arising from this thesis:

(1) Nicolas A. S. Stewart, **Alpay Taralp** and Harvey Kaplan (1997) "Imprinting of Lyophilized α-Chymotrypsin Affects the Reactivity of the Active-Site Imidazole," *Biochemical and Biophysical Research Communications*, accepted for publication.

(2) Harvey Kaplan and **Alpay Taralp** (1997) "Nonaqueous Chemical Modification of Lyophilized Proteins," in *Techniques in Protein Chemistry VIII* (Marshak, D. Ed.) pp. 219-230, Academic Press.

(3) **Alpay Taralp** and Harvey Kaplan (1997) "Chemical Modification of Lyophilized Proteins in Nonaqueous Environments," *Journal of Protein Chemistry* **16**, 183-193.

(4) **Alpay Taralp**, Harvey Kaplan, Iou-Iou Sytwu, Isidorus Vlattas, Regine Bohacek, Anna. K. Knap, Tomoko Hirama, Carol P. Huber and Sadiq Hasnain (1995) "Characterization of the S_3 Subsite Specificity of Cathepsin B," *Journal of Biological Chemistry* **270**, 18036-18043.

Scientific presentations:

(1) Nicolas A. S. Stewart, **Alpay Taralp** and Harvey Kaplan "Ligand-Induced Changes in the Active-Site of Lyophilized Chymotrypsin," The Eleventh Symposium of the Protein Society, conference proceedings, *Protein Science* **6** (Suppl. 2), 568-S, Boston, MA, July 1997.

(2) Harvey Kaplan and **Alpay Taralp** "Chemical Modification of Lyophilized Proteins in Nonaqueous Environments," The Tenth Symposium of the Protein Society, conference proceedings, *Protein Science* **5** (Suppl. 1), 471-S, San Jose, CA, August 1996.

(3) **Alpay Taralp** and Harvey Kaplan "Chemical Modification of Lyophilized Proteins in Nonaqueous Environments," The 6th Ottawa-Carleton Chemistry Institute Minisymposium in Bioorganic Chemistry, Ottawa, May 1996.

(4) **Alpay Taralp** and Harvey Kaplan "Iodomethane as an NMR Active Protein Modification Reagent," The VIII MOOT NMR Minisymposium, Ottawa, September 1995.

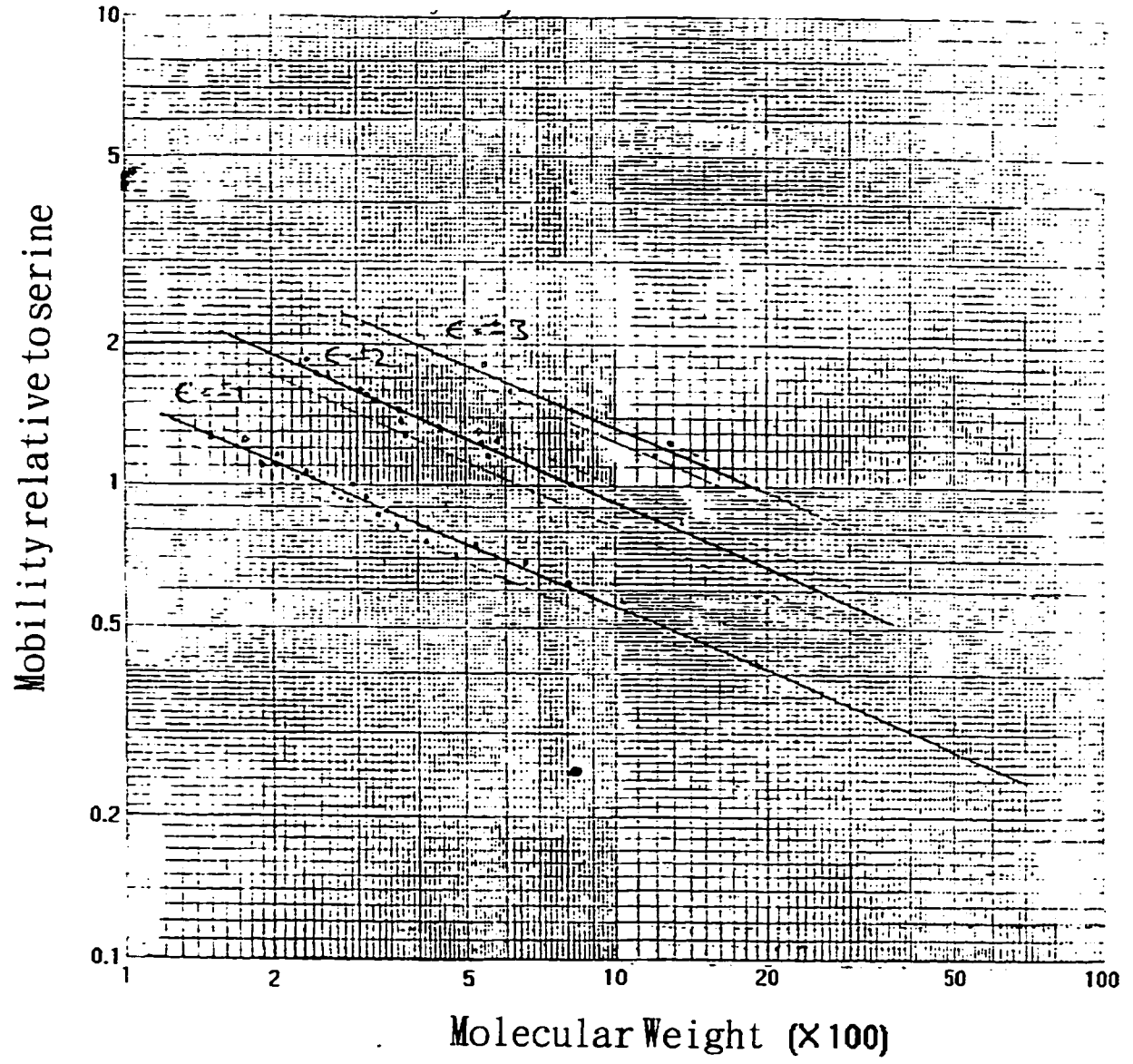
(5) **Alpay Taralp**, Nicolas A. S. Stewart and Harvey Kaplan "Iodomethane as an NMR Active Protein Modification Reagent," The Ninth Symposium of the Protein Society, conference proceedings, *Protein Science* **4** (Suppl. 2), 503-T, Boston, MA, July 1995.

(6) **Alpay Taralp**, Nicolas A. S. Stewart and Harvey Kaplan "Iodomethane as an NMR Active Protein Modification Reagent," The 5th Ottawa-Carleton Chemistry Institute Minisymposium in Bioorganic Chemistry, Ottawa, May 1995.

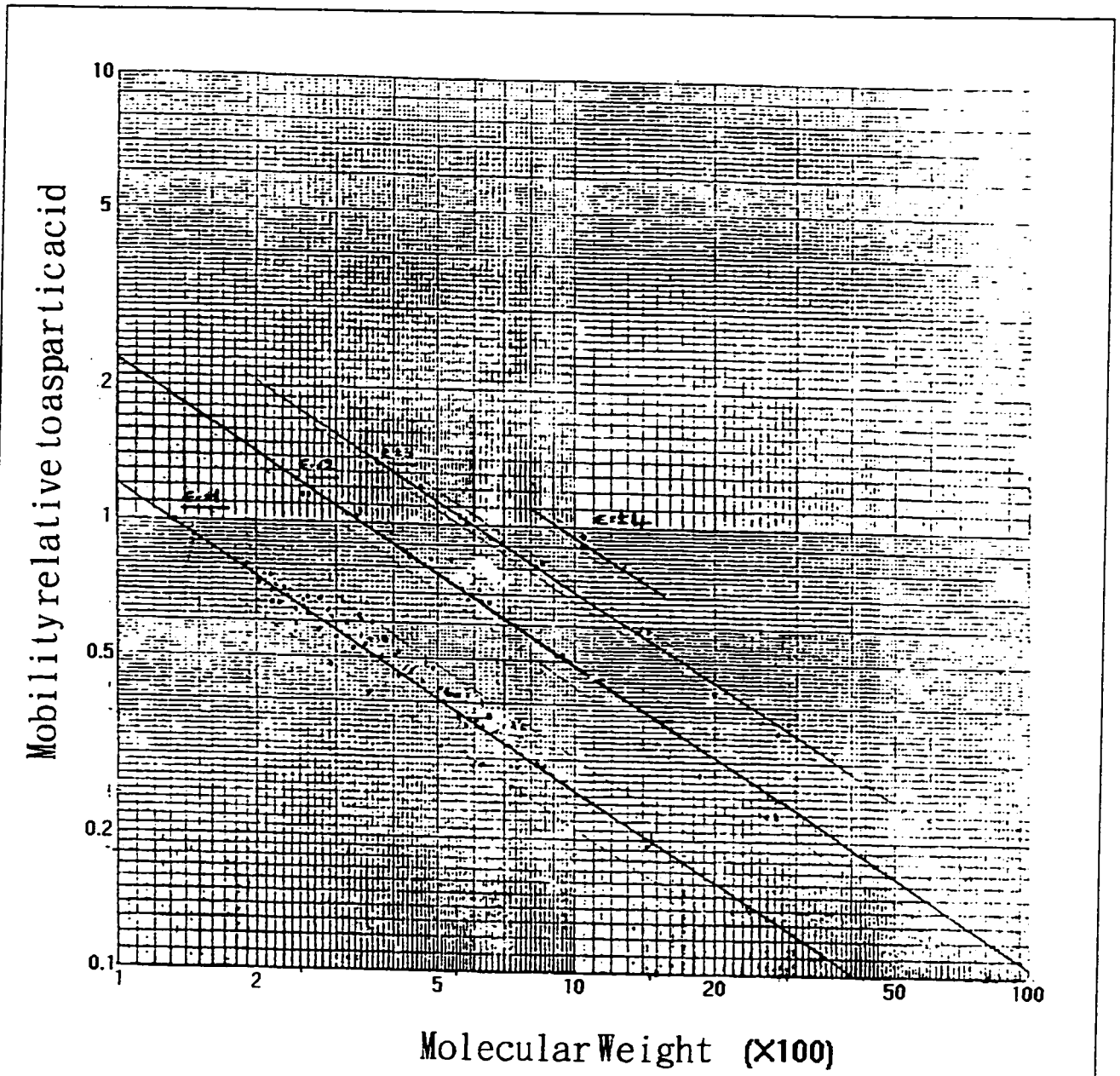
(7) **Alpay Taralp**, Harvey Kaplan, Iou-Iou Sytwu, Isidoros Vlattas, Regine Bohacek, Anna K. Knap, Tomoko Hirama, Carol P. Huber and Sadiq Hasnain "Characterization of the S_3 Subsite Specificity of Cathepsin B: The Importance of P_3 on Substrate Binding," The Eight Symposium of the Protein Society, conference proceedings, *Protein Science* **3** (Suppl. 1), 128-M, San Diego, CA, July 1994 and The 4th Ottawa-Carleton Chemistry Institute Minisymposium in Bioorganic Chemistry, Ottawa, May 1994.

(8) Harvey Kaplan, **Alpay Taralp**, Vicky L. Knowles and R. P. Oomen "Selective Methylation of Tyrosine and Histidine Residues in Proteins with Methyl Iodide: Methylation of Insulin," The Seventh Symposium of the Protein Society, conference proceedings, *Protein Science* **2** (Suppl. 1), 455-T, San Diego, CA, July 1993 and The Fourth Ontario-Quebec Minisymposium in Synthetic and Bioorganic Chemistry, Ottawa, October 1993.

Appendix 2: High voltage paper electrophoresis mobility charts for use at pH 2.1 and pH 6.5



Mobility chart for noncysteic acid peptides migrating at pH 2.1. The mobility of dansyl arginine is 1.45 with respect to serine as measured from the neutral spot (dansyl sulphonic acid).



Mobility chart for nonhistidyl peptides migrating at pH 6.5. The mobility of dansylsulphonic acid is 0.69 with respect to aspartic acid as measured from the neutral spot (dansylarginine).