NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

STRUCTURAL CHARACTERIZATION OF THE INSECTICIDAL PROTEIN FROM *BACILLUS THURINGIENSIS*

CHRISTIN TERESA CHOMA

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biochemistry

University of Ottawa

ISBN 0-315-62327-6

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

T. T. T.

Put up in a place
where it's easy to see
the cryptic admonishment
T. T. T.

When you feel how depressingly
slowly you climb,
it's well to remember that
Things Take Time.

-Piet Hein

ABSTRACT

STRUCTURAL CHARACTERIZATION OF THE INSECTICIDAL PROTEIN FROM *BACILLUS THURINGIENSIS*

During sporulation, *Bacillus thuringiensis* subsp. *kurstaki* produces a crystalline inclusion body which is toxic upon ingestion by susceptible Lepidopteran larvae. The major component of crystals from Lepidopteran-specific subspecies of *B. thuringiensis* is a 130-kDa protein, protoxin. Following ingestion by susceptible larvae, protoxin is proteolyzed to yield a 58-70 kDa toxic fragment, toxin. In the present study, a simplified procedure was used for isolating and purifying toxin generated by the tryptic digestion of protoxin from *B. thuringiensis* subsp. *kurstaki* HD-73.

Characterization of this toxin showed that it is derived from the N-terminal half of the protoxin molecule. The toxin is insoluble at neutral pH values but is moderately soluble at alkaline values above pH 9. Application of several spectroscopic and theoretical procedures to the purified toxin showed that the protein is composed of approximately equal amounts of α-helix, ß-sheet and random coil structures. The tertiary structure of toxin was shown to be comprised of two primary domains; these domains correspond to the toxic and specificity (or binding) domains predicted from analysis of protoxin gene nucleotide sequences. Evidence was obtained that at least one additional domain is present as a structural component of the C-terminal specificity domain.

Both the toxic moiety within the protoxin molecule and free toxin were found to be unusually resistant to unfolding by chemical denaturants and to proteolysis. In contrast, the C-terminal half of protoxin could be readily unfolded and was extremely susceptible to proteolytic digestion. The unfolded protoxin and unfolded toxin were shown to refold rapidly into their native and biologically active conformations. Evidence was obtained that the conformation of the toxic moiety of protoxin is very similar to the conformation of toxin.

Chemical modification of the cysteine and lysine residues in the protoxin did not affect the biological activity of the protein. However, the introduction of positive, negative or neutral groups onto these residues had a large effect on the solubility of the protein. These results, along with the results obtained from the unfolding/folding studies, strongly indicate that the primary function of the C-terminal half of the protoxin molecule is to promote the formation of a stable crystal.

To my parents,
who always encouraged me not to settle for
the lowest common denominator...

ACKNOWLEDGEMENTS

I wish to take this opportunity to express my sincere appreciation to Dr. H. Kaplan of the Department of Chemistry, University of Ottawa, for supervising my doctoral studies. Prior to embarking on these studies, I had no knowledge whatsoever of protein chemistry; this deficiency was rapidly rectified largely due to Dr. Kaplan's innate talent for transforming intricate concepts into easily comprehensible ideas, and his inexhaustible patience. I am appreciative of all that he has taught me, but even more so of the spirit in which he viewed my training. He has always treated me as a peer rather than a student, and has encouraged me to explore fully all the research avenues open to me, even when my interests were somewhat unorthodox.

Most of the work presented in this thesis was conducted at the National Research Council of Canada. In my three years of graduate studies I have had the privilege and benefit of working with a large number of scientists who have generously permitted me the use of their facilities and assisted me with experimental design and data analysis. In particular, I wish to thank Dr. W. Surewicz for help with infrared spectroscopy and calorimetry, Dr. M. Yaguchi for help with amino acid analyses, Dr. P. Carey for assisstance with Raman spectroscopy, Dr. A. Szabo for the use of his fluorescence facility, and Drs. M. Young and S. Hasnain for cheerfully fielding the many questions posed to them.

Research is, by its nature, an often solitary and sometimes disheartening pursuit, and I thank the technicians at the N. R. C. and the graduate students in Dr. Kaplan's laboratory for helping make my graduate studies pleasant and productive. Special thanks are due to Mr. T. Lessard for introducing me to the peculiarities of working with the *Bacillus thuringiensis* protein, and to Ms. T. Hirama and I. Rasquinha for the many times they offered me their assistance. The camaraderie of E. Armstrong, J. Atkins, H. Bietlot and R. Milne has been much appreciated.

Prior to the present studies, my employment provided me the pleasure of making the acquaintance of many scientists and engineers in the Space Division of N. R. C. Although this thesis represents a significant diversion from space-related research, numerous opportunites have been made available to me over the past three years to continue my active involvement in the N. R. C. microgravity programme. I wish to offer my sincerest thanks to Drs. P. Kumar and Z. Saghir, Mr. G. Campbell and Mr. R. Wilkinson for encouraging my interest in this field and for making my continuing participation possible. I also wish to thank Mr. J. Aitken and Mr. M. Pygas of the National Aeronautical Establishment for their interest in the microgravity project I undertook during the course of the present studies, and for helping make those experiments so enjoyable.

The poems which preface the chapters in this thesis are by Piet Hein, a Danish scientist and poet; his witty and creative poems effortlessly bridge the gap between science and humanism.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	aminoethyl
B.t.	*Bacillus thuringiensis*
CAM	carbaminomethyl
CAPS	3-(cyclohexylamino)-1-propane-sulphonic acid
CD	circular dichroism
CF	Chou-Fasman
CM	carboxymethyl
deg	degree
dmol	decimole
DNFB	2,4-dinitrofluorobenzene
E64	trans-epoxysuccinyl-L-leucylamido-6-(4-guanidino)-butane
EDTA	ethylenediaminetetraacetic acid
GOR	Garnier-Osguthorpe-Robson
GuHCl	guanidinium chloride
HPLC	high performance liquid chromatography
i.d.	inside diameter
IR	infrared
kDa	kilodalton
LD_{50}	dose resulting in 50% mortality
MOPS	3-(N-morpholino)propanesulphonic acid
MPD	methylpentanediol
mW	milliwatts
nm	nanometer
nmol	nanomole
N.R.C.	National Research Council of Canada
o.d.	outside diameter
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
SCAM	succinyl-carbaminomethyl
SCM	succinyl-carboxymethyl
SDS	sodium dodecyl sulphate
TCA	trichloroacetic acid
TPCK	N-tosyl-L-phenylalanine chloromethyl ketone
TRIS	tris(hydroxymethyl)aminomethane
UV	ultraviolet

Chapter 1.

INTRODUCTION

Problems

Problems worthy

of attack

prove their worth

by hitting back.

- Piet Hein

Chapter 1.

INTRODUCTION

1.1 OVERVIEW OF BACILLUS THURINGIENSIS

1.1.1 Survey of bioinsecticides

The control of insect pests in both the agricultural and forestry industries is largely achieved by the liberal application of chemical insecticides. Worldwide, the annual expenditure for chemical pesticides is estimated to be well over four billion dollars (Wilcox *et al.*, 1986). Although it has been known since the early part of this century that certain bacteria, fungi and viruses possess insecticidal actions, microbials have until recently remained a virtually untapped pesticide resource. Over one thousand micro-organisms or microbial products are known to be effective against insects, and yet as of 1986 only thirteen microbial organisms were approved by the American Environmental Protection Agency for field use (Wilcox *et al.*, 1986). However, the combined impact of the increasing incidence of insect resistance to chemical insecticides, and heightened public awareness of the potential human toxicity and carcinogenicity of these chemicals, has served to renew interest in alternate forms of insect control such as microbial pesticides. Since the advent of genetic engineering, the promise of generating new recombinant biological insecticides with novel host specificities has further served to invigorate interest in this field.

A fundamental advantage of bioinsecticides over chemical pesticides is that the former are only active against a narrow species range and have little or no activity against non-target species. Although there are thousands of insect pests, only a few hundred are so destructive as to warrant active control. Microbial products can be formulated to target a specific insect pest, leaving the plants, animals and other insects of the ecosystem unharmed.

The most successful and widely used bioinsecticide products are those based on the bacterium *Bacillus thuringiensis*. Insecticides formulated with *B. thuringiensis* are highly active against a narrow range of insects, making these products an ideal component of integrated pest management policies. Commercial production of *B. thuringiensis*

insecticides began in the United States in 1958, but their use did not become widespread until the early 1980's at which time they were also deployed in Canada. In both Canada and the United States, *B. thuringiensis* is extensively used against food crop insects. In recent years, the bioinsecticide has also been widely used in Canada against forest pests, being employed in spruce budworm control operations in Ontario and Quebec, in the control of jackpine budworm and western spruce budworm in the Western provinces, and against hemlock looper and gypsy moth in the Atlantic provinces (Statistics Canada, 1987).

B. thuringiensis products are produced by a number of companies worldwide. Aside from the environmental safety of this bioinsecticide, other advantages of *B. thuringiensis* products are that they act rapidly, and no evidence of insect resistance has been reported arising from normal field applications. However, *B. thuringiensis* formulations represent only one percent of total worldwide pesticide sales. Use of this bioinsecticide has been limited by problems encountered with cost-effective field application of the formulations, removal of the insecticide from foliage by rain, and sunlight inactivation of the insecticidal activity. As these practical problems are presently being addressed in many laboratories, it can be anticipated that the use of *Bacillus thuringiensis*-based insecticides will become increasingly widespread during the coming years.

1.1.2 History of *B. thuringiensis*

B. thuringiensis was first isolated from diseased silk moth (*Bombyx mori*) larvae by Ishiwata in 1901, and again ten years later by Berliner from larvae of the flour moth (*Anagasta kuhniella*). It was Berliner who first described the presence of an inclusion body within the sporulated cells, and who termed the organism *Bacillus thuringiensis*, after the German province of Thuringia. Attempts were made in the following decades to utilize *B. thuringiensis* as an insecticide, but so little was known about the source and specificity of the insecticidal action that the organism could not be efficiently exploited. It was not until the early 1950's that preliminary information regarding the basis of the insecticidal action

became available. The presence of diamond-shaped crystals in sporulating *B. thuringiensis* cells was described by Steinhaus in 1951. Several years later the crystals were shown to be composed largely of protein, and it was demonstrated that these crystals were responsible for the insecticidal properties of *B. thuringiensis* (Angus, 1954; Hannay and Fitz-James, 1955). By 1956 it was clear that different strains of *B. thuringiensis* were toxic to different insect species (Angus, 1956). The efforts of Howard Dulmage to systematically catalogue the various *B. thuringiensis* species lead to considerable information on strain specificity and to a comprehensive set of *Bacillus thuringiensis* strains, the HD collection (Dulmage *et al.*, 1981). With the establishment of the cabbage looper (*Trichoplusia ni*) as the industry-wide test organism, and agreement on *B. thuringiensis* subsp. *kurstaki* HD-1 as the industrial bioinsecticide standard (Dulmage, 1970), information on the specificity and toxicity of *B. thuringiensis* could be evaluated in a more systematic manner. In the last decade, renewed interest in the properties of *B. thuringiensis* has lead to significant insights into the mode of action of this bioinsecticide, but only modest advances have been made towards understanding the structure of the insecticidal component. The future will see advances in the development of improved bioinsecticides through genetic engineering, as the relationship between structure and function of the *B. thuringiensis* insecticide is elucidated.

1.1.3 Biological properties of *Bacillus thuringiensis*

B. thuringiensis is a Gram-positive, aerobic, sporulating soil bacterium which forms a parasporal crystal. Except for the presence of the entomocidal crystal, *B. thuringiensis* is indistinguishable from *Bacillus cereus* or *Bacillus anthracis*. Over twenty subspecies of *B. thuringiensis* have been classified based on their flagellar antigens (Dean, 1984). The subspecies also differ in the shape of the crystal, the number of toxic proteins in the crystal, and their insect host specificities. Until the late 1970's, it was believed that *B. thuringiensis* toxicity was limited to Lepidopteran larvae. Since then, *B. thuringiensis* subspecies have been identified which are lethal to mosquitoes (Yamamoto and

McLaughlin, 1981), black flies (Thomas and Ellar, 1983a) and beetles (Krieg *et al.*, 1983). Over 130 species of Lepidoptera, Diptera and Coleoptera are susceptible to at least one subspecies of *B. thuringiensis* (Dean, 1984).

The biosynthesis of the entomocidal crystal takes place concomitant with sporulation (Huber and Lüthy, 1981). The crystal grows to approximately the same size as the spore (1μ in length) and constitutes 30% of the cellular protein (Dean, 1984). In the last phase of sporulation, the crystal, which is generally bipyramidal in shape, is released into the environment together with the spore when the cell lyses. The prominence of the crystal within *B. thuringiensis* cells suggests that the crystal confers substantial survival value to individual cells. It is known that spores of *B. thuringiensis* germinate slowly, and the natural concentration of the bacterium in the soil is low (Aronson *et al.*, 1986). An insect host provides an excellent nutritional environment for bacterial proliferation and sporulation, and therefore permits the bacteria to reach the high concentrations required for cell mating and exchange of genetic material (Gonzalez *et al.*, 1982).

1.1.4 Insect pathology of *B. thuringiensis*

B. thuringiensis is pathogenic primarily to insects in the larval stages of development. In susceptible Lepidopteran insect larvae, the midgut pH is usually highly alkaline (between pH 9 and 10), while the hemolymph pH is approximately neutral. It is believed that the alkaline pH of the gut aids the insect in the breakdown of tannins and other plant material which remain insoluble at lower pH values (Milne *et al.*, 1990). When a susceptible insect ingests a parasporal crystal from *B. thuringiensis*, the high midgut pH helps solubilize the crystal, and gut proteases cleave the crystal protein (protoxin) to form the active toxin (Fast, 1981). Within minutes of ingestion, the insect gut and mouth become paralyzed, the pH of the midgut drops and the pH of the hemolymph rises (Heimpel and Angus, 1959). Within one hour of crystal ingestion, the microvilli in the midgut have swollen, and within four hours the midgut cells have separated from the basement membrane and are totally destroyed (Greigo *et al.*, 1980; Oron *et al.*, 1985).

Investigations into the mode of action of the crystal protein suggest that the mechanism of action of the toxin involves the inhibition of active potassium transport across the midgut epithelium (Griego *et al.*, 1979; Harvey and Wolfersberger, 1979). Alternatively, it has been proposed that the toxin binds to specific cell-surface receptors in the midgut and forms pores in the epithelium; ions and small molecules then equilibrate across the membrane, leading to cell lysis (Knowles and Ellar, 1987). Changes in the endoplasmic reticulum and mitochondria, disruption of ion and glucose transport and oxygen uptake, and loss of adenosine triphosphate from midgut cells, all contribute towards total body paralysis and death (Aronson *et al.*, 1986).

In some insects, the crystal protein appears to act alone in causing death, while in other insect species the spore apparently plays a synergistic role. These observations lead Heimpel and Angus (1959) to classify Lepidopteran larvae into three groups. Type I insects are killed solely by the protein crystal, and the spores of the bacterium do not enhance toxicity. In type II insects, the spore enhances the toxic action of the crystal, while type III insects can only be killed by a mixture of spores and crystals. It is believed that most type III insects do not possess highly alkaline guts and thus cannot efficiently dissolve the *B. thuringiensis* crystal. As a pH close to neutrality favours spore germination, type III insects are killed primarily by the proliferation of *B. thuringiensis* within their midgut (Aronson *et al.*, 1986). Possible explanations for the observed range in host specificities is that crystal proteins from various subspecies of *B. thuringiensis* are processed with varying efficiencies by gut proteases, and that the receptors in different insect guts exhibit various affinities for binding the activated toxin (Whiteley and Schnepf, 1986). In addition, the production of more than one type of toxic protein by some strains of *B. thuringiensis* raises the possibility that synergistic or antagonistic interactions between the toxic molecules and potential receptors in the midgut epithelial cells give rise to the observed variation in toxicity of *B. thuringiensis* crystals towards different insects (Wilcox *et al.*, 1986).

1.2 THE CRYSTAL PROTEIN FROM *BACILLUS THURINGIENSIS*

1.2.1 Structure of the entomocidal crystal

Crystals from various subspecies of *B. thuringiensis* differ in the number of protein components, but in most crystal types the predominant protein is the protoxin. Other proteins associated with the parasporal inclusion body include, for example, a mosquitocidal protein found in *B. thuringiensis* subsp. *kurstaki* HD-1 (Donovan *et al.*, 1988). Unlike most Lepidopteran-specific subspecies of *B. thuringiensis*, the crystals from subspecies *israelensis* are composed of several proteins, some of which are toxic towards Dipteran larvae, and others which are lytic towards mammalian cells (Thomas and Ellar, 1983b). However, it is the 130-kDa protoxin protein found in Lepidopteran-specific crystals which has been most extensively studied, and which is the subject of the present investigation.

The bipyramidal crystals are composed of protoxin molecules packed into a crystalline array held together by disulphide bonds between the protein molecules (Huber *et al.*, 1981). Although it was reported by Holmes and Monroe (1965) and Huber *et al.* (1981) that the repeating subunit in crystals from a number of *B. thuringiensis* isolates was a dimer of the protoxin molecule, more recent work questions the validity of this conclusion. A study into the solubilization of the crystal by Nagamatsu *et al.* (1984) provided strong evidence that the crystal from *B. thuringiensis* subsp. *dendrolimus* is composed of a single subunit of molecular mass 145 kDa. In addition, although the parasporal crystals are too small to yield X-ray diffraction patterns which would permit determination of the structure of the protoxin molecule, small angle X-ray powder patterns have been obtained from crystal isolates of *B. thuringiensis* subsp. *tenebrionis*. These patterns showed that each unit cell of the crystal contains only one protoxin molecule (Li *et al.*, 1988). To date, none of the Lepidoptran-specific protoxins have been re-crystallized to yield X-ray diffraction quality crystals.

1.2.2 Chemistry of the crystal protein

The entomocidal crystal protein is insoluble at neutral pH, but is fairly soluble at pH 9 and above. *In vitro*, the rate of solubilization of the crystal can be greatly enhanced by the addition of thiol reagents such as ß-mercaptoethanol (Bulla *et al.*, 1977). Various amounts of carbohydrate have been reported to be associated with different crystal types, but there is little agreement amongst the published reports (Whiteley and Schnepf, 1986). Amino acid analyses of different crystal preparations show that all the protoxin proteins have similar amino acid compositions (Fast, 1981). These compositions are not remarkable and do not reflect the unusual properties of the protoxin molecule.

1.2.3 Activation of protoxin

Study of the *B. thuringiensis* crystals during the 1960's and 1970's lead to a range of estimates for the molecular mass of the principal crystal protein component. The situation was complicated by the presence of contaminating bacterial proteases which acted on the crystal protein and lead to the accumulation of protein fragments. By the early 1980's it was established that the primary component of Lepidopteran-specific crystals was a protein with a molecular mass of approximately 130 kDa, and that a minor fraction of protein from each crystal type had a molecular mass of approximately 68 kDa (Tyrell *et al.*, 1981). The larger protein, protoxin, was found to be non-toxic towards cultured cells of the spruce budworm (*Choristeneura fufimerana*), whereas the smaller protein was very toxic (Huber and Lüthy, 1981). Protoxin could be converted into the smaller protein by the action of either purified mammalian proteases or insect gut juice (Lilley *et al.*, 1980). From these observations, Bulla *et al.* (1981) concluded that the crystals of *B. thuringiensis* are composed of protoxin molecules which are proteolyzed into smaller, toxin molecules within the insect gut. As the toxin has a molecular mass which is half that of the protoxin, there was some controversy in the literature as to whether protoxin was a dimer of the toxin molecule. However, it was established by Andrews *et al.* (1985) that protoxin is stoichiometrically converted to toxin by the action of proteases. The molecular mass of the

protease-resistant, Lepidopteran-specific toxin molecule varies from approximately 58 to 70 kDa, depending on the subspecies origin of the crystal and the proteases used to accomplish the activation of protoxin to toxin (Höfte and Whiteley, 1989).

1.2.4 Structure of the protoxin genes

To date, fourteen distinct crystal protein genes in *B. thuringiensis* have been identified. Thirteen of these genes code for a family of related insecticidal proteins, called the Cry, or crystal, proteins. These thirteen *cry* genes have been divided into four major classes based on gene homology, protein structural similarities and insect specificities (Höfte and Whitely, 1989). CryI proteins are specific to Lepidoptera, CryII proteins are toxic to both Lepidoptera and Diptera, CryIII proteins are effective against Coleoptera, and CryIV proteins are specific to Diptera. The proteins characterized in the present study, the protoxin and toxin from *B. thuringiensis* subsp. *kurstaki* HD-73, belong to the CryI class.

B. thuringiensis bacteria harbour between two and seventeen plasmids, depending upon the subspecies (Lereclus *et al.*, 1982). Curing bacteria of their plasmids results in the loss of crystal production, indicating that the *cry* genes are located on plasmids (Gonzalez and Carlton, 1980). When plasmid DNA from the industrial standard strain *B. thuringiensis* subsp. *kurstaki* HD-1 was digested with the restriction enzyme HindIII, three different DNA fragments with lengths of 4.5, 5.3 and 6.6 kilobases were found to contain protoxin genes (Kronstad and Whiteley, 1986). These genes have recently been reclassified as the *cryIA(a)*, *cryIA(b)* and *cryIA(c)* genes, and code for protoxin proteins of 1176, 1155 and 1179 amino acids, respectively (Höfte and Whiteley, 1989). Only the *cryIA(c)* gene is found in *B. thuringiensis* subsp. *kurstaki* HD-73 (Adang *et al.*, 1985), but multiple heterogeneous copies of the protoxin gene have been identified in many *B. thuringiensis* subspecies. As the toxicities of products from a given class of genes are very similar, the advantages of gene diversity remain somewhat obscure. However, it is possible that the heterogeneous copies of the protoxin gene are regulated differently within the bacterial cell, or that the various gene products of a particular subspecies act in a

synergistic manner and thus contribute to the insect host spectrum of a *B. thuringiensis* subspecies.

When the *cryIA* gene sequences are aligned and gaps are introduced to maximize homology, an 85% overall homology between the three sequences becomes evident (Adang *et al.*, 1987). The differences that do exist among the gene types are not randomly distributed, but rather are clustered in several distinct regions (Adang *et al.*, 1987). The N-terminal 280 amino acids of all three gene types are 98% conserved. The following 350 amino acids delineate a section of each sequence, called the hypervariable region, where homology falls to approximately 60%. The hypervariable region can be further divided into two regions. In the case of the *cryIA(c)* gene type, the sequence between amino acid residues 280-460 shows some variance with the *cryIA(a)* and *cryIA(b)* gene types, while the region between residues 460 and 640 shows pronounced heterogeneity (Geiser *et al.*, 1986; Andrews *et al.*, 1987). The coding for the C-terminal half of the protoxin molecule by the three genes again shows strong conservation of the amino acid sequence, although the preservation of the sequence is less striking than that observed for the first 280 N-terminal residues.

Elucidation of the *B. thuringiensis cry* gene sequences has lead to the identification of two distinct functional regions in the protoxin molecule. Deletion studies of protoxin genes from *B. thuringiensis* subsp. *berliner* revealed that only the amino-terminal segment encoding the first 630 amino acids was required to produce a protein with insecticidal activity comparable to that of the native protoxin (Lilley *et al.*, 1980). The protein encoded by this gene fragment had a molecular mass of 68 kDa, which corresponds to the size of protease-activated toxin. Further deletion studies on other protoxin genes confirmed that in all cases, the toxin is derived from the amino terminal half of the protoxin molecule and is encoded by both the highly conserved and the hypervariable regions of the gene. In contrast, the function of the conserved C-terminal half of the protoxin molecule remains unclear, but sequencing of the *cry* genes revealed that this region has an unusual

distribution of cysteine and lysine residues. In *B. thuringiensis* subsp. *kurstaki* HD-73 for example, the C-terminal half contains 14 of the 16 cysteine residues, and 31 of the 34 lysine residues (Adang *et al.*, 1985). Although no role has yet been proposed for the lysine residues, the abundance of cysteine residues suggests that the C-terminal region plays a structural role in crystal stabilization through the formation of disulphide bonds between protoxin molecules (Nickerson, 1980; Huber *et al.*, 1981; Bietlot *et al.*, 1990). As the N-terminal region alone is sufficient for toxic activity, it has been postulated that the C-terminal region, in addition to its role in crystal formation, serves to protect the toxic moiety or mediates binding to insect gut epithelial cell receptors (Aronson *et al.*, 1986). However, experimental support for these postulated functions is presently not available.

1.3 STRUCTURE OF THE PROTOXIN AND TOXIN MOLECULES: SYNOPSIS OF THE PRESENT STUDY

It has been generally accepted that the structure of a protein is closely related to its biological function. The elucidation of protoxin primary structures from the *cry* gene sequences has revealed several unusual structural features of the protein, as noted above. Deletion and point mutation studies are serving to unravel the functional roles played by individual regions of the molecule, and of single amino acids within the protein. However, despite the economic and environmental importance of the *B. thuringiensis* insecticidal proteins, there have been few direct structural studies on the protoxin and toxin *proteins*, as opposed to structural information deduced from gene sequences. Perhaps the primary reason for the lack of direct research has been the large size of the protoxin molecule and its low solubility, which have made the protein difficult to study. It has also been difficult to obtain reasonable quantities of protoxin free of contaminating spores and proteases. Recent developments in the separation of protoxin crystals from spores have largely overcome this barrier (Carey *et al.*, 1986), opening the way for detailed structural studies on the insecticidal protein.

The focus of the present work is the characterization of the structure of toxin and protoxin from *B. thuringiensis* subsp. *kurstaki* HD-73. This strain of *B. thuringiensis* was chosen as it carries only a single *cryIA(c)* gene (Adang *et al.*, 1985; Appendix D), and thus synthesizes a homogeneous population of protoxin molecules. The intent of the present work is to contribute to the overall understanding of the structural properties of the toxin, and to relate this data to the unique biological activity of the insecticidal protein. In Chapter 2, the chemical properties of the toxin are characterized in order to provide a foundation for detailed biophysical and biochemical studies. A comprehensive investigation into the secondary structure of the toxin molecule is presented in Chapter 3. The deduced secondary structure elements are compared where possible to structures predicted by gene sequence alignment studies. In Chapter 4, a low resolution map of the tertiary structure of the toxin molecule is deduced using the methods of limited proteolysis and differential scanning calorimetry. Roles for the observed structural features are tentatively assigned, based on information in the literature regarding the molecular-level function of the toxin.

The work presented in Chapter 5 arose from the observation that toxin is far more resistant to chemical denaturation than most other proteins. This unusual resistance is clearly demonstrated in the very slow unfolding of the toxin in denaturants, and the protein's ability to readily refold into its native, active conformation upon removal of the denaturant. In contrast, protoxin appears to be far more susceptible to denaturation. The results are interpreted with a view to providing insights into possible conformational changes in the protoxin toxic moiety when the C-terminal region is cleaved and the molecule becomes functionally active.

In Chapter 6, chemical modification studies offer further insights into the tertiary structure of the protoxin molecule, and suggest a means of broadening the insect host specificity of this bioinsecticide. The main body of the thesis concludes with Chapter 7, where the information acquired during the course of this study on the structure of the

protoxin and toxin molecules is drawn together and discussed in the context of the inter-relationship between protein structure and function.

Although Appendix A presents a study quite separate from the main body of this dissertation, it addresses an issue relevant to the question of protein structure. The difficulty of obtaining large crystals of a protein for detailed structure analysis has greatly hampered the utility of the X-ray diffraction technique; true to most proteins, efforts to obtain diffraction quality crystals of the *B. thuringiensis* insecticidal protein have met with little success. An approach which may help overcome the difficulty of crystallizing proteins is reviewed, and a preliminary study addressing several practical aspects of this strategy is presented.

1.4 REFERENCES

Adang, M. J., M. J. Staver, T. A. Rocheleau, J. Leighton, R. F. Barker and D. V. Thompson (1985). Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. *Gene* **36**, 289-300.

Adang, M. J., K. F. Idler and T. A. Rocheleau (1987). Structural and antigenic relationships among three insecticidal crystal proteins of *Bacillus thuringiensis* subsp. *kurstaki*. (in) *Biotechnology Advances in Invertebrate Pathology and Cell Culture* (K. Maramorosch, ed.) Academic Press, New York, pp. 85-99.

Andrews, R. E., M. M. Bilbos and L. A. Bulla (1985). Protease activation of the entomocidal protoxin of *Bacillus thuringiensis* subsp. *kurstaki*. *Appl. Environ. Microbiol.* **50**, 737-740.

Angus, T. A. (1954). A bacterial toxin paralysing silkworm larvae. *Nature* **173**, 545-546.

Angus, T. A. (1956). Extraction, purification and properties of *Bacillus sotto* toxin. *Can. J. Microbiol.* **2**, 416-426.

Aronson, A. I., W. Beckman and P. Dunn (1986). *Bacillus thuringiensis* and related insect pathogens. *Microbiol. Rev.* **50**, 1-24.

Bietlot, H., I. Vishnubhatla, P. R. Carey, M. Pozsgay and H. Kaplan (1990). Characterization of the cysteine residues and disulphide linkages in the protein crystal of *Bacillus thuringiensis* subsp. *kurstaki* and *entomocidus*. *Biochem. J.* **267**, 309-315.

Bulla, L. A., K. J. Kramer and L. I. Davidson (1977). Characterization of the entomocidal parasporal crystal of *Bacillus thuringiensis*. *J. Bacteriol.* **130**, 375-383.

Bulla, L. A., K. J. Kramer, D. J. Cox, B. L. Jones, L. I. Davidson and G. L. Lockhart (1981). Purification and characterization of the entomocidal protein of *Bacillus thuringiensis*. *J. Biol. Chem.* **256**, 3000-3004.

Carey, P. R., P. Fast, H. Kaplan and M. Pozsgay (1986). Molecular structure of the protein crystal from *Bacillus thuringiensis*: a Raman spectroscopic study. *Biochim. Biophys. Acta* **872**, 169-176.

Dean, D. H. (1984). Biochemical genetics of the bacterial insect-control agent *Bacillus thuringiensis*: Basic principles and prospects for genetic engineering. *Biotechnol. Genet. Eng. Rev.* **2**, 341-363.

Donovan, W., C. Dankogski, M. Gilbert, M. Gawron-Burke, R. Groat and B. Carleton (1988). Amino acid sequence and entomocidal activity of the P2 crystal protein. *J. Biol. Chem.* **263**, 651-657.

Dulmage, H. T. (1970). Insecticidal activity of HD-1, a new isolate of *Bacillus thuringiensis* var. *alesti*. *J. Invert. Pathol.* **15**, 232-239.

Dulmage, H. T. and Co-operators (1981). Insecticidal activity of isolates of *Bacillus thuringiensis* and their potential for pest control. (in) *Microbial Control of Pests and Plant Diseases* (H. D. Burges, ed.) Academic Press, London, pp. 193-222.

Fast, P. G. (1981). The crystal toxin of *Bacillus thuringiensis*. (in) *Microbial control of pests and plant diseases* (H. D. Burgess, ed.) Academic Press, London, pp. 223-248.

Geiser, M., S. Schweitzer and C. Grimm (1986). The hypervariable region in the genes coding for entomopathogenic crystal proteins of *Bacillus thuringiensis*: nucleotide sequence of the *kurhd1* gene of subsp. *kurstaki* HD-1. *Gene* **48**, 109-118.

Gonzalez, J. M. and B. C. Carlton (1980). Patterns of plasmid DNA in crystalliferous and acrystalliferous strains of *Bacillus thuringiensis*. *Plasmid* **3**, 92-98.

Gonzalez, J. M., B. S. Brown and B. C. Carleton (1982). Transfer of *Bacillus thuringiensis* plasmids coding for delta-endotoxin among strains of *B. thuringiensis* and *B. cereus*. *Proc. Natl. Acad. Sci. USA* **79**, 6951-6955.

Griego, V. M., D. Moffett and K. D. Spence (1979). Inhibition of active K+ transport in the tobacco hornworm (*Manduca sexta*) midgut after ingestion of *Bacillus thuringiensis* endotoxin. *J. Insect Physiol.* **25**, 283-288.

Griego, V. M., L. J. Fancher and K. D. Spence (1980). Scanning electron microscopy of the disruption of tobacco hornworm, *Manduca sexta*, midgut by *Bacillus thuringiensis* endotoxin. *J. Invert. Pathol.* **35**, 186-194.

Hannay, C. L. and P. C. Fitz-James (1955). The protein crystals of *Bacillus thuringiensis* Berliner. *Can. J. Microbiol.* **1**, 694-698.

Harvey, W. R. and M. G. Wolfersberger (1979). Mechanism of inhibition of active potassium transport in isolated midgut of *Manduca sexta* by *Bacillus thuringiensis* endotoxin. *J. Exp. Biol.* **83**, 293-304.

Heimpel, A. M. and T. A. Angus (1959). The site of action of crystalliferous bacteria in *Lepidoptera* larvae. *J. Insect Pathol.* **1**, 152-170.

Höfte, H. and H. R. Whiteley (19889). Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Holmes, K. C. and R. E. Monroe (1965). Studies on the structure of parasporal inclusions from *Bacillus thuringiensis*. *J. Mol. Biol.* **14**, 572-581.

Huber, H. E. and P. Lüthy (1981). *Bacillus thuringiensis* delta-endotoxin: Composition and activation. (in) *Pathogenesis of invertebrate microbial diseases* (E. W. Davidson, ed.) Allanheld, Osmum and Co., Totowa, N.J., pp. 209-234.

Huber, H. E., P. Lüthy, H. R. Ebersol and J. L. Cordier (1981). The subunits of the parasporal crystal of *Bacillus thuringiensis*: size, linkage and toxicity. *Arch. Microbiol.* **129**, 14-18.

Knowles, B. and D. Ellar (1987). Colloid-osmotic lysis is a general feature of the mechanism of action of *Bacillus thuringiensis* delta-endotoxins with different insect specificities. *Biochim. Biophys. Acta* **924**, 509-518.

Krieg, A., A. Huger, G. Langenbruch and W. Schnetter (1983). *Bacillus thuringiensis* var. *tenebrionis*: a new pathotype effective against larvae of Coleoptera. *J. Appl. Entomol.* **96**, 500-508.

Kronstad, J. W. and H. R. Whiteley (1986). Three classes of homologous *Bacillus thuringiensis* crystal protein genes. *Gene* **43**, 29-40.

Lereclus, D., M. M. Lecadet, J. Ribier and R. Dedonder (1982). Molecular relationships among plasmids of *Bacillus thuringiensis*: conserved sequences through 11 crystalliferous strains. *Mol. Gen. Genet.* **186**, 391-398.

Li, J., R. Henderson, J. Carroll and D. Ellar (1988). X-ray analysis of the crystalline parasporal inclusion in *Bacillus thuringiensis* var. *tenebrionis*. *J. Mol. Biol.* **199**, 543-544.

Lilley, M., R. N. Ruffell and H. J. Somerville (1980). Purification of the insecticidal toxin in crystals of *Bacillus thuringiensis*. *J. Gen. Microbiol.* **118**, 1-11.

Milne, R., A. Z. Ge, D. Rivers and D. H. Dean (1990). Specificity of insecticidal crystal proteins: implications for industrial standardization. (in) *Analytical chemistry of Bacillus thuringiensis* (C. Hickle and W. Fitch, eds.) ACS Symposium series no. 432, pp. 22-35.

Nickerson, K. W. (1980). Structure and function of the *Bacillus thuringiensis* protein crystal. *Biotech. Bioeng.* **22**, 1305-1333.

Oron, U., M. Sokolover, A. Yawetz, M. Broza, B. Sneh and A. Honigman (1985). Ultrastructural changes in the larval midgut epithelium of *Spodoptera littoralis* following ingestion of delta-endotoxin of *Bacillus thuringiensis* var. *entomocidus*. *J. Invert. Pathol.* **45**, 353-369.

Statistics Canada (1987). Selected Forestry Statistics Canada: Information Report Ex40.

Steinhaus, E. A. (1951). Possible use of *Bacillus thuringiensis* Berliner as an aid in the biological control of the alfalfa looper. *Hilgardia* **20**, 259-265.

Thomas, W. E. and D. J. Ellar (1983a). Mechanism of action of *Bacillus thuringiensis* var. *israelensis* insecticidal delta-endotoxin. *FEBS Lett.* **154**, 362-368.

Thomas, W. E. and D. J. Ellar (1983b). *Bacillus thuringiensis* var. *israelensis* crystal delta-endotoxin: effects on insect and mammalian cells *in vitro* and *in vivo*. *J. Cell Sci.* **60**, 181-197.

Tyrell, D. J., L. A. Bulla, R. E. Andrews, K. J. Kramer, L. I. Davidson and P. Nordin (1981). Comparative biochemistry of entomocidal parasporal crystals of selected *Bacillus thuringiensis* strains. *J. Bacteriol.* **145**, 1052-1062.

Wilcox, D. R. *et al.* (1986). Genetic Engineering of Bioinsecticides. (in) *Protein Engineering: Applications in Science, Medicine and Industry*, Academic Press, New York, pp. 395-413.

Yamamoto, T. and R. E. McLaughlin (1981). Isolation of a protein from the parasporal crystal of *Bacillus thuringiensis* var. *kurstaki* toxic to the mosquito larva of *Aedes taeniorhynchus*. *Biochem. Biophys. Res. Commun.* **103**, 414-421.

Chapter 2.

CHARACTERIZATION OF *B. THURINGIENSIS* SUBSP. *KURSTAKI* HD-73 TOXIN

Getting down to fundamentals

It will steadily shrink,
our earthly abode,
until antipode stands
upon antipode.

Then, soles together,
the planet gone,
we'll know the ground
that we rest upon.

- Piet Hein

Chapter 2.

CHARACTERIZATION OF *B. THURINGIENSIS* SUBSP. *KURSTAKI* HD-73 TOXIN

2.1 INTRODUCTION

A prerequisite to the comprehensive study of a compound is the ability to generate pure, characterized preparations of the material. In the case of biological macromolecules, this requirement often imposes strict limitations on the nature and scope of the investigations which can be undertaken. The elucidation of the structural characteristics of a protein requires nanomolar to micromolar quantities of pure material. For many studies, this prerequisite can only be fulfilled by the engineering of cloned gene expression systems, and long and laborious purification protocols. Fortunately, the protoxin protein is abundantly produced in *Bacillus thuringiensis*, and is conveniently packaged as a crystal within the bacterium. However, it is only very recently that a protocol for isolating large quantities of pure crystals has been developed. Efforts towards this end have been hampered by the fact that the bacterium synthesizes the protoxin crystal at the same time that the spore is produced. Since the crystal and spore are very similar in size, density and surface characteristics (Andrews *et al.*, 1987), difficulties in separating crystals from spores typically resulted in low recovery yields of crystals from *B. thuringiensis* cells (Andrews *et al.*, 1987).

When the activated toxin rather than the protoxin is the subject of study, some published purification procedures advocate bypassing the tedium of purifying crystals. As the toxin is quite resistant to proteases, treatment with mammalian trypsin of a spore/crystal mixture from lysed *B. thuringiensis* cells results in the isolation of only one large polypeptide, the toxin (Aronson and Arvidson, 1987). However, following the subsequent steps used to purify the toxin, less than 10% of the theoretical yield of toxin was recovered (Aronson and Arvidson, 1987). Other methods for producing pure toxin vary in their approach and the purification steps used (Chestukhina *et al.*, 1982; Andrews *et al.*, 1985), but all are lengthy protocols which produce only poor yields of the purified protein. Furthermore, the toxic fragments produced from CryI protoxins by these procedures differ significantly in their size, isoelectric point, carbohydrate content and

amino acid composition (Chestukhina *et al.*, 1982; Andrews *et al.*, 1985; Aronson and Arvidson, 1987; Andrews *et al.*, 1987).

The generation of toxin from purified crystals is straightforward. Only two simple purification steps are required, since few contaminants are generated during the activation process. Protoxin and toxin purified using the protocols established at the National Research Council of Canada and the University of Ottawa (Carey *et al.*, 1986; Bietlot *et al.*, 1989) were used in all the work presented in this thesis. In this chapter, the characterization of the toxin produced by bovine trypsin digestion of *B. thuringiensis* subsp. *kurstaki* HD-73 protoxin crystals is described.

2.2 MATERIALS AND METHODS

2.2.1 Materials

B. thuringiensis subsp. *kurstaki* HD-73 stock culture cells were obtained from the Forestry Pest Management Institute, Sault Ste. Marie, Ontario. Renografin was purchased from Squibb Canada Ltd., Montreal, Quebec. All proteases used (bovine pancreatic TPCK [N-tosyl-L-phenylalanine chloromethyl ketone] trypsin, pronase type XIV, α-chymotrypsin type II, carboxypeptidase B, carboxypeptidase Y, papain type III and elastase type III) were purchased from Sigma Chemical Co., St. Louis, Missouri. Dialysis tubing (50 kDa cutoff) was obtained from Spectrum Medical Industries, Los Angeles, California. Precast gels, buffer strips and marker proteins for use with the Pharmacia Phast gel electrophoresis system were purchased from Pharmacia Canada Ltd., Dorval, Quebec, as was the 'Mono-Q' column used for high performance liquid chromatography (HPLC). Polyvinylidene difluoride membranes were obtained from Applied Biosystems, Foster City, California. Disposable cellulose acetate syringe filters (0.45 μm pore size) were purchased from Cole-Parmer, Chicago, Illinois. CAPS (3-(cyclohexylamino)-1-propanesulphonic acid) buffer, the detergent Triton X-100, L-norleucine and the papain inhibitor trans-epoxysuccinyl-L-leucylamido-6-(4-guanidino)-butane (E64) were obtained from Sigma Chemical Co., St. Louis, Missouri. Ultra high purity hydrochloric acid was

obtained from Pierce Chemical Co., Rockford, Illinois. Reagents used for HPLC were purchased from Waters Canada Ltd., Mississauga, Ontario. All other chemicals were reagent-grade; all solutions were prepared using reverse-osmosis quality water purified by the Milli-Q water system, Millipore Canada Ltd., Mississauga, Ontario.

2.2.2 Purification of Protoxin

B. thuringiensis subsp. *kurstaki* HD-73 was grown and harvested by Mr. P. Fleming, Division of Biological Sciences, N.R.C. Cells were grown with aeration in a 28 litre fermentor at 30°C in medium composed of 0.5% yeast extract, 0.5% Trypticase, 0.1% glucose and 0.08% K_2HPO_4. Cells were harvested by centrifugation after approximately 40 hours of culture, when cell proliferation had stopped and crystal formation within the bacteria was complete (Krieg and Miltenburger, 1984).

Protoxin crystals were isolated from the cells by Mr. T. Lessard of the Division of Biological Sciences, N.R.C. The procedure used was a modification of an earlier protocol (Carey *et al.*, 1986) and permitted the isolation of between 1-2 grams of pure protoxin crystals from the harvested cells. The cells were lysed in reverse-osmosis quality water and the preparation was washed twice in 1M NaCl containing 0.01% Triton X-100. The pellet was then suspended in water and mixed with Renografin to give a final concentration of 64% (v/v) Renografin. The suspension was sonicated and then spun at 25,000x*g* for one hour at 4°C in a swinging bucket rotor. Non-germinated spores were found to pellet at the bottom of the centrifuge tube, while germinated spores and crystals remained in suspension or formed a layer at the top of the preparation. The fractions containing crystals were suspended in 50 ml of water, sonicated, and spun at 45,000x*g* for 30 minutes at 4°C. The preparation was found to sediment into three layers. Pure crystals pelleted at the bottom of the tube, and were overlayed with a layer of germinated spores. The top layer was composed of cell debris and lipid. Careful removal of the contaminating layers yielded protoxin crystals which, upon examination by phase contrast microscopy, were found to be between 99% to 99.9% pure (free of spores). In practice, it was often necessary to repeat

each of the above purification steps several times in order to achieve this level of purity. Solubilization of these crystals in 0.1M pH 10.5 CAPS buffer containing 0.1% ß-mercaptoethanol yielded only a single protein band of molecular mass 130 kDa upon sodium dodecyl sulphate (SDS) gel electrophoresis (Bietlot *et al.*, 1989). Amino acid analysis of the crystal protein (protoxin) yielded a composition (Carey *et al.*, 1986) which corresponded to that predicted from the *B. thuringiensis* subsp. *kurstaki* HD-73 gene sequence (Adang *et al.*, 1985).

2.2.3 Preparation of toxin

To a suspension of 100 mg of crystal in 5 ml of 0.1M CAPS buffer, pH 10.5, 5 mg of TPCK-trypsin was added and the suspension was stirred for 20 hours at 20°C. The suspension was centrifuged at 45,000x*g* at 4°C for 10 minutes to remove any undissolved crystals. The supernatant was then passed through a cellulose acetate filter in order to remove any remaining particulates. To obtain toxin free of residual trypsin and peptides, ammonium sulphate was added in aliquots to a final concentration of 40% (w/v). After each addition of $(NH_4)_2SO_4$, the preparation was stirred for 15 minutes at 4°C, the precipitate was removed, the next aliquot of $(NH_4)_2SO_4$ was added to the supernatant and the process was repeated. To remove the salt, each precipitate fraction was dialyzed separately against water at 4°C overnight using 3.5 kDa cutoff dialysis tubing. The toxin fractions were then analyzed by sodium dodecyl sulphate (SDS) gel electrophoresis to identify samples contaminated by trypsin; these samples were subsequently subjected to further ammonium sulphate fractionation or to gel permeation chromatography. During the present investigation, this procedure was modified so that only a single aliquot of 40% (w/v) $(NH_4)_2SO_4$ was added to the digest. The sample was stirred for 15 minutes at 4°C, and then dialyzed against three changes of water at 4°C over 24 hours using 50 kDa cutoff dialysis tubing. The precipitated toxin was analyzed for purity using SDS gel electrophoresis, isoelectric focusing and high performance liquid chromatography.

2.2.4 Characterization of the toxin

2.2.4.1 Polyacrylamide gel electrophoresis (PAGE)

PAGE was used both to assess the purity of the toxin preparations, and to estimate the isoelectric point of the protein. The SDS/PAGE method used is based on that described by Laemmli (1970). Toxin samples were boiled for 2 minutes in an SDS buffer consisting of 8M urea, 2.5% SDS, 5% ß-mercaptoethanol and 10mM Tris-HCl, adjusted to pH 8.3. After cooling, 1 µl aliquots of the samples were applied to preformed 10-15% gradient gels. The constituent proteins of each sample were separated using a Pharmacia Phast gel electrophoresis apparatus, employing the voltage and current settings suggested by the manufacturer. Gels were stained for 10 minutes at 50°C in 0.1% Coomassie blue stain containing 30% methanol and 10% acetic acid, and were destained in 20 minutes using three changes of destaining solution (30% methanol and 10% acetic acid) at 50°C. The molecular mass of the toxin was estimated using a standard curve of log (molecular mass) *vs.* relative mobility, determined with standard molecular mass markers. The markers used were: phosphorylase (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20 kDa) and α–lactalbumin (14 kDa).

Isoelectric focusing, originally described by Svensson (1962), served as a further test of the homogeneity of the toxin sample, in addition to estimating the pI of the protein. Toxin was dissolved in 0.05M CAPS buffer, pH 10.5, and 1 µl of sample was applied to the preformed gel. The samples were focused using a Pharmacia Phast gel electrophoresis apparatus, employing the procedure suggested by the manufacturer. Ampholytes within the gel migrated to their respective pI and formed a gradient of pH 3 to 9 between the anode and cathode. Marker proteins of known pI were applied as standards; the isoelectric point of the toxin was determined from a calibration curve of the marker proteins (pI *vs.* distance migrated from cathode by protein). Staining and destaining procedures for the gels were as outlined above.

2.2.4.2 High performance liquid chromatography (HPLC)

HPLC of toxin samples served as another indicator of the purity of the preparations. A Gilson HPLC and model 302 pump system, fitted with a 'Mono-Q' anion-exchange column, was used. Approximately 0.05 mg of toxin dissolved in 0.02M CAPS, pH 10.0, was injected into the column. The protein was eluted with a smooth salt gradient, which at 0-time consisted of 0.02M CAPS, pH 10.0, and at the conclusion of the 80 minute run consisted of 0.02M CAPS plus 0.5M NaCl. Eluted proteins were detected by absorption at 280 nm.

2.2.4.3 Absorption coefficient

The absorption coefficient of the toxin at 280 nm was determined by quantifying protein samples of known absorbance by amino acid analysis. To 4.0 ml of toxin solution of known absorbance at 280 nm, 250.0 nmol of norleucine was added. A 500.0 µl aliquot of this sample was hydrolyzed in 6N HCl *in vacuo* at 110^{o}C for 24 hours. The amino acids in the sample were quantified using a Technicon TSM amino acid analyzer fitted with a ninhydrin detection system. By comparing the areas of each amino acid peak to the area of the internal norleucine standard, the amount of each amino acid within the sample could be quantified. Three duplicate toxin solutions were analyzed in this manner. From the amino acid composition of toxin deduced from the gene sequence, the amount of toxin in each sample could be calculated. Beer's law states that:

$$A = \varepsilon l C$$

where A is the absorbance of the sample (here at 280 nm), ε is the absorption coefficient, l is the pathlength of the light through the sample (in cm) and C is the concentration of the sample (in mol liter^{-1}). Since A is dimensionless, ε has the units liters mol^{-1} cm^{-1}. When the concentration of the sample is expressed in mg/ml, the units of ε change accordingly.

2.2.4.4 pH-dependence of toxin solubility

Approximately 8 mg of toxin was added to 0.5 ml of buffer solution containing 0.1M of each of the following: boric acid, potassium chloride, citric acid and disodium phosphate. The solubility of the toxin at pH values ranging from pH 2-12 was determined by adjusting the pH of the sample, stirring vigorously for 30 minutes at room temperature, and then measuring the pH of the suspension. Undissolved protein was removed by centrifugation and the absorbance of the supernatant at 280 nm was measured. The concentration of toxin in solution was calculated from the absorbance of the sample at 280 nm and the absorption coefficient, determined above. The supernatant was returned to the pellet of undissolved toxin, the sample was vortexed, the pH adjusted by an increment of approximately 0.5 pH unit, and the procedure was repeated.

2.2.4.5 N-terminal sequence determination

Automated gas-phase sequencing was performed on an Applied Biosystems 474A protein sequencing system, using a modification of the method originally described by Edman and Begg (1967). Toxin (approximately 0.5 nmol) from an SDS/polyacrylamide gel was electroblotted onto a polyvinylidene difluoride membrane (Matsudira, 1987). Using the protocol suggested by the manufacturer, the first twenty N-terminal amino acids were sequenced by Mr. D. Watson, Division of Biological Sciences, N.R.C.

2.2.4.6 Proteolysis of purified toxin

The sensitivity of toxin to proteolysis was determined by incubating approximately equimolar amounts of toxin and protease in 0.1M CAPS buffer, pH 10.5, at room temperature for either 5 minutes or one hour. The samples were then boiled in SDS sample buffer and analyzed by SDS/PAGE. The proteases tested were pronase type XIV, α–chymotrypsin type II, carboxypeptidase B, carboxypeptidase Y, papain type III and elastase type III. The papain inhibitor E64 was added to samples containing papain immediately before boiling, thus inhibiting any further proteolysis of the toxin by the

enzyme in the SDS sample buffer. This precaution was taken as papain is unusually stable and retains residual activity even under highly denaturing conditions (Sluyterman, 1967; Glazer and Smith, 1971).

2.2.4.7 Carbohydrate analysis

The carbohydrate content of protoxin and toxin from *B. thuringiensis* subsp. *kurstaki* HD-73 was determined by the phenol/sulphuric acid method of Dubois *et al.* (1956). Due to the low solubility of protoxin and toxin at neutral pH values, samples (3 mg) were initially solubilized in 0.1M KOH (pH 13). Phenol and sulphuric acid were then added directly to the solubilized samples. A standard curve of the absorbance at 480 nm of 1 µg/ml to 20 µg/ml glucose was prepared in a similar manner, and was used to convert the colour intensity of the protein samples into a measure of their sugar content. Triplicate samples of protoxin and toxin were analyzed.

2.3 RESULTS

2.3.1 Homogeneity of the toxin preparations

The characterization of the protoxin crystals from which toxin is generated is described by Carey *et al.* (1986). Overnight incubation of protoxin crystals at pH 10.5 with bovine trypsin yielded a viscous, colloidal suspension of toxin which readily precipitated upon addition of ammonium sulphate. Dialysis of the precipitate against water using 50 kDa cutoff dialysis tubing was effective in removing salt, residual trypsin and peptides. The SDS gel electrophoresis pattern of toxin samples dialyzed in either 3.5 kDa or 50 kDa cutoff membrane tubing is compared in Fig. 2.1. The use of 50 kDa cutoff dialysis tubing consistently yielded toxin preparations which were electrophoretically pure.

Isoelectric focusing also served as an indication of the homogeneity of the toxin preparations. The very low solubility of toxin at neutral pH values often resulted in the protein precipitating during focusing. This problem could generally be overcome by

applying the toxin sample to the centre of the gel immediately prior to focusing, rather than at the cathode end as suggested in the Pharmacia manual. When focusing was successful, only a single band was obtained (Fig. 2.2). The absence of multiple bands upon isoelectric focusing provides further evidence of the electrophoretic homogeneity of the toxin preparations.

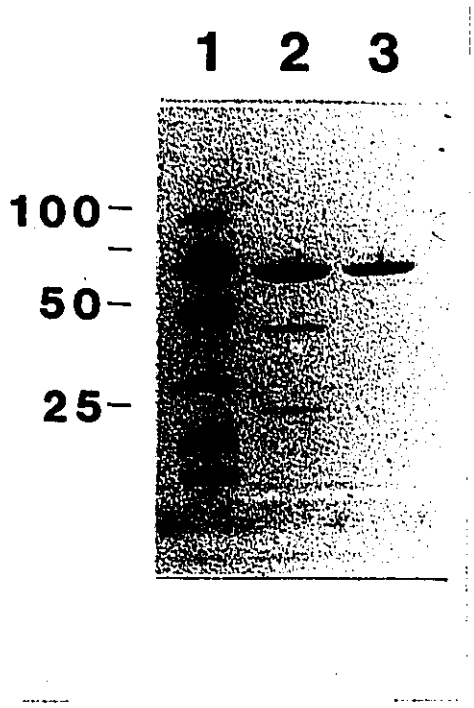


Figure 2.1. Purification of toxin by dialysis against water using 50 kDa cutoff dialysis tubing.
Shown are the SDS gel electrophoresis patterns of toxin generated by tryptic digestion of protoxin crystals followed by extensive dialysis against water in either 3.5 kDa cutoff dialysis tubing (lane 2) or 50 kDa cutoff dialysis tubing (lane 3). Lane 1 shows the molecular mass markers: phosphorylase (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20 kDa) and α–lactalbumin (14 kDa).

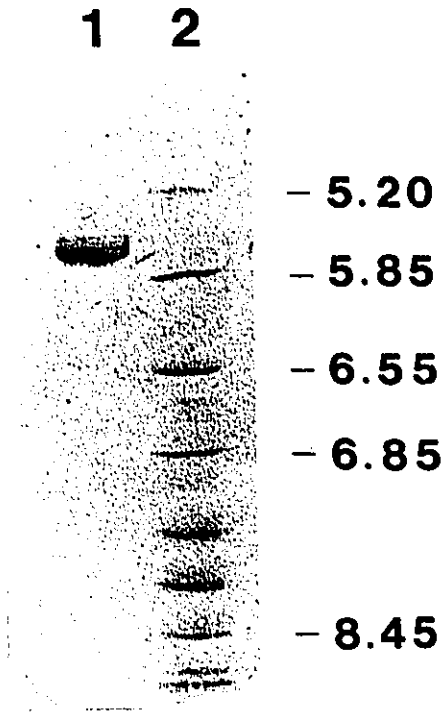


Figure 2.2. Determination of the pI of toxin.
Toxin was focused on a pH 3-9 gradient gel. Lane 1 shows the focused toxin band. Lane 2 shows the following pI marker proteins: ß-lactoglobulin A (pI = 5.20); bovine carbonic anhydrase B (5.85); human carbonic anhydrase B (6.55); horse myoglobin acidic band (6.85); horse myoglobin basic band (7.35); lentil lectin acidic band (8.15); lentil lectin middle band (8.45); lentil lectin basic band (8.65); trypsinogen (9.3); cytochrome C (10.25).

Toxin samples subjected to analysis by HPLC gave rise to only a single major peak; the few contaminants observed were present in very small quantities (Fig.2.3). When this peak was collected, concentrated and analyzed by SDS gel electrophoresis, only a single band with a molecular mass identical to that of toxin was obtained.

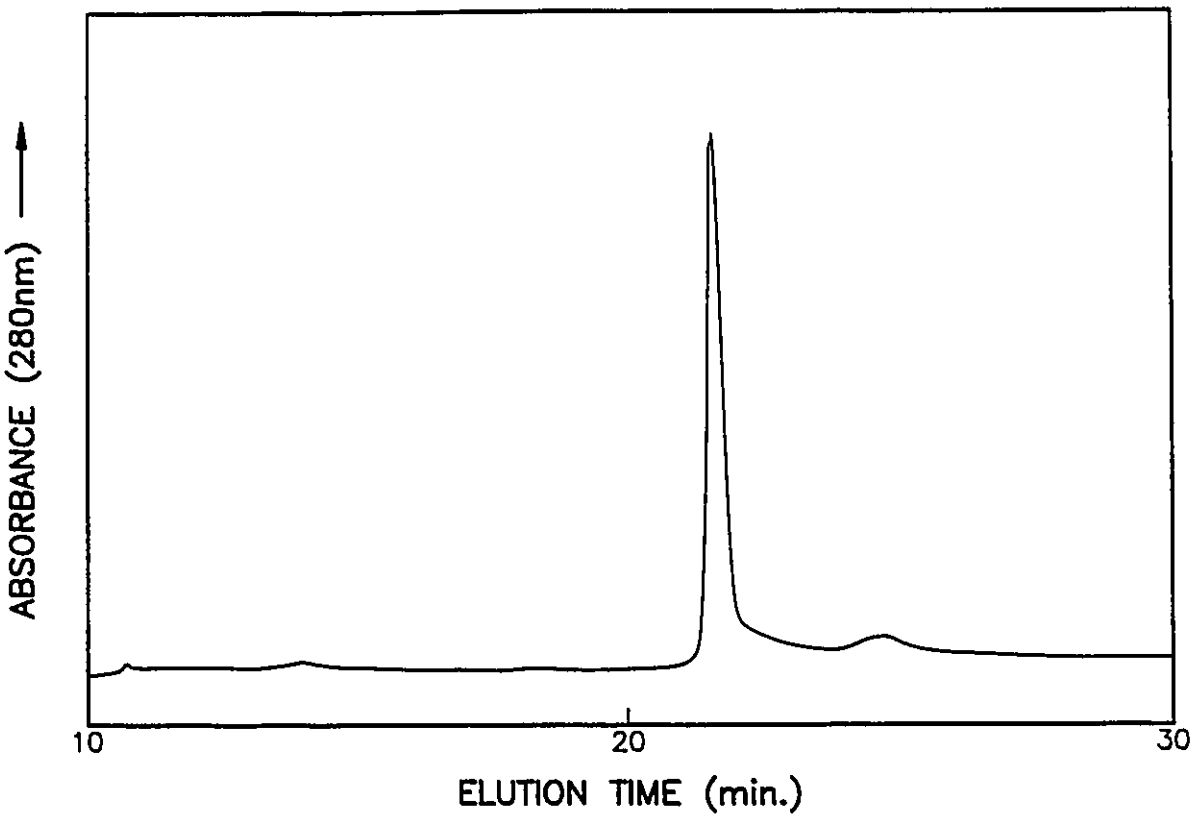


Figure 2.3. HPLC of toxin.
Toxin, generated by tryptic digestion of protoxin crystals and extensively dialyzed against water using 50 kDa cutoff dialysis tubing, was subjected to high pressure liquid chromatography. The trace shows the components detected at 280 nm during elution by a salt gradient.

2.3.2 Characterization of the toxin

From Fig. 2.1, it can be estimated that the molecular mass of the trypsin-resistant toxin is approximately 67 kDa. Isoelectric focusing indicated that the pI of the protein is 5.5 ± 0.3 (Fig. 2.2). Both these results were consistent for all toxin preparations tested, indicating that the protocol used for generating toxin from protoxin crystals gives highly reproducible toxin preparations.

The absorption coefficient of the toxin at 280 nm was obtained by quantitative amino acid analysis of toxin samples of known absorbance. Duplicate analyses of three toxin preparations indicated that a toxin solution at pH 10.5 and at a concentration of 1 mg/ml would give an absorbance of 1.61 ± 0.12 for a 1 cm pathlength at 280 nm. This absorption coefficient was used in calculating the solubility of toxin at various pH values in the range 2-12. It was shown independently that changes in pH have a negligible effect on this absorption coefficient (H. Kaplan, personal communication). It is evident from Fig. 2.4 that toxin is soluble only under very acid or alkaline conditions. At alkaline pH values of 10.5 or greater, toxin is soluble to about 8 mg/ml, while below pH 3.5, a solubility of approximately 5 mg/ml is observed.

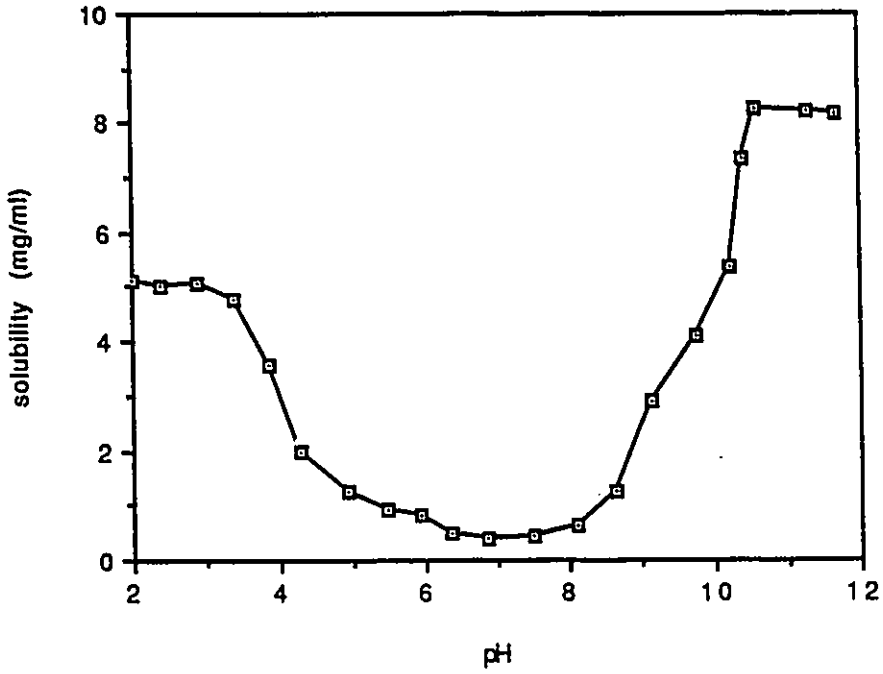


Figure 2.4. The pH-dependant solubility of toxin.
The solubility of toxin (mg/ml) over the pH range 2-12 was determined.

Sequencing of the first 20 N-terminal amino acids of the trypsin-generated toxin gave the sequence IETGYTPIDISLSLTQFLLS. This corresponds to amino acid residues 29-49 as deduced from the protoxin gene nucleotide sequence (Adang *et al.*, 1985). No secondary sequence was detected, indicating that the trypsin cleavage site preceding isoleucine-29 in protoxin is by far the most predominant N-terminal cleavage, and that toxin molecules generated by trypsin digestion have homogeneous, not ragged, N-termini.

The incubation of toxin with a number of proteases showed that the toxin is remarkably stable not only towards further proteolysis by trypsin, but to other proteases as well. It is evident from Fig. 2.5 that of the six proteases tested, only pronase could significantly proteolyze toxin. The action of pronase was rapid, with little toxin remaining

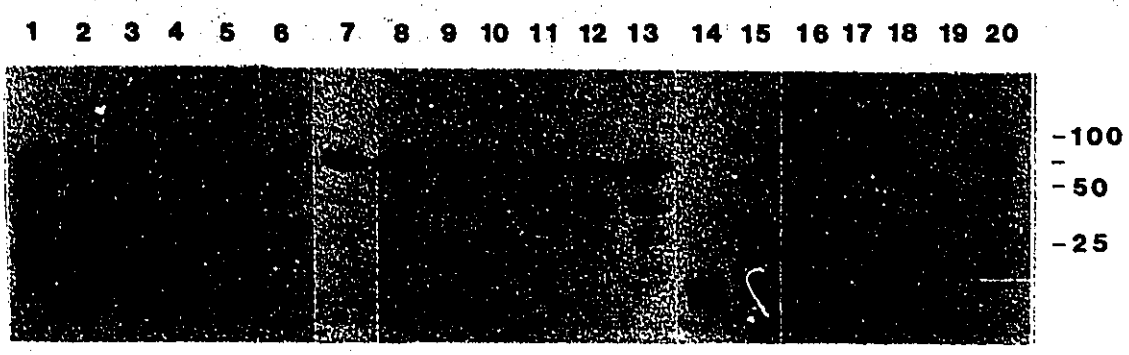


Figure 2.5. Protease resistance of toxin.
SDS gel electrophoresis patterns are shown for toxin incubated at pH 10.5 for either five minutes or one hour with each of the following proteases: lane 1, with pronase, 5 min.; lane 2, with pronase, 60 min.; lane 3, with α–chymotrypsin, 5 min.; lane 4, with α–chymotrypsin, 60 min.; lane 5, toxin alone (no protease added); lane 6, with carboxypeptidase B, 5 min.; lane 7, with carboxypeptidase B, 60 min.; lane 8, with carboxypeptidase Y, 5 min.; lane 9, with carboxypeptidase Y, 60 min.; lane 10, with papain, 5 min.; lane 11, with papain, 60 min.; lane 12, with elastase, 5 min.; lane 13, with elastase, 60 min.; lane 14, pronase marker; lane 15, α–chymotrypsin marker; lane 16, carboxypeptidase B marker; lane 17, carboxypeptidase Y marker; lane 18, papain marker; lane 19, elastase marker; lane 20, molecular mass markers.

after 5 minutes incubation; prolonged incubation for one hour did not result in significantly more proteolysis. The absence of a time effect may be due to pH inactivation of the enzyme; however, in light of the known ability of papain to proteolyze proteins in boiling SDS sample buffer (Chapter 4), it is possible that the proteolysis observed with pronase also occurred upon boiling the sample in the denaturing buffer. As the toxin unfolded under these conditions, it would have been cleaved by pronase before the protease itself was denatured. Of the other proteases tested, only α–chymotrypsin had any effect on the toxin, and the amount of proteolysis was very small even after one hour of incubation.

Analysis of solubilized protoxin and toxin samples for carbohydrate content showed insignificant amounts (<0.5% w/w) of sugar associated with either protein. Although this background level could not be eliminated by extensive washing of the protoxin crystals, it cannot be ruled out that the carbohydrate detected was non-covalently bound to the protein, and was merely a contaminant from the fermentation broth.

2.4 DISCUSSION

Tryptic digestion of protoxin crystals which are free of spores and contaminating endogenous proteases results in a toxin preparation which can be purified to electrophoretic homogeneity by a rapid and straightforward procedure. Precipitation of the toxin with ammonium sulphate, followed by extensive dialysis against water using 50 kDa cutoff dialysis tubing, resulted in a toxin preparation which was free of contaminating peptides and residual trypsin. The recovery of toxin prepared by this procedure was shown to be 60-70% of the theoretical maximum, and the protein was highly toxic towards spruce budworm larvae (Bietlot *et al.*, 1989). In the present study, toxin preparations were shown to be homogeneous by a number of fractionation techniques. On both SDS and isoelectric focusing gels, toxin samples consistently yielded only a single band. HPLC fractionation resulted in a single large peak being separated from several other components which were present in trace amounts.

The molecular mass of toxin estimated from SDS gels is 67 kDa. This approximation was verified by gel permeation chromatography using Sephadex G-200 (Bietlot *et al.*, 1989) and is in general agreement with the size of protease-stable toxins generated from protoxin isolated from other strains of *B. thuringiensis*: 63-kDa toxin from *kurstaki* HD-263 (Aronson and Arvidson, 1987), 58-kDa toxin from *dendrolimus* (Nagamatsu *et al.*, 1984), 68 kDa from *kurstaki* LB1 (Andrews *et al.*, 1985), and 65-80 kDa from several other subspecies of *B. thuringiensis* (Chestukhina *et al.*, 1982).

Together with the information that the N-terminal trypsin cleavage site is between arginine-28 and isoleucine-29 in protoxin, it can be estimated from the gene-deduced sequence that the *B. thuringiensis* subsp. *kurstaki* HD-73 toxin molecule spans residues 29 to approximately 625 of the protoxin molecule. The C-terminal trypsin cleavage site was determined to be between lysine-623 and alanine-624; quantitative amino acid analysis was used to verify that toxin consists of amino acid residues 29-623 of the protoxin molecule (Bietlot *et al.*, 1989).

The identification of arginine-28 as the N-terminal cleavage site is in agreement with Nagamatsu *et al.* (1984), who determined the same cleavage site for toxin generated by trypsin digestion of protoxin from *B. thuringiensis* subsp. *dendrolimus*. Similarly, Aronson and Arvidson (1987) identified isoleucine as the N-terminal amino acid in toxin generated from *B. thuringiensis* subsp. *kurstaki* HD-263; in conjunction with their amino acid analysis of the toxin, this N-terminus would be consistent with a cleavage site between residues 28 and 29. In contrast to these findings, Chestukhina *et al.* (1982) reported that no N-terminal processing occurred during the generation of toxin from protoxin isolated from four species of *B. thuringiensis*, and that in each case only the C-terminal half of the protoxin molecule was removed by trypsin. In all these investigations, including the present study, the N-terminal half of the protoxin molecule (less the first 28 amino acid residues) was found to be resistant to trypsin digestion. It was also shown in each instance that this protease-resistant fragment was toxic to insect larvae (Chestukhina *et al.*, 1982;

Nagamatsu *et al.*, 1984; Aronson and Arvidson, 1987; Bietlot *et al.*, 1989). These results together provide conclusive evidence that the toxin is derived from the N-terminal half of the protoxin molecule.

It is interesting to note that in studies reporting processing of the N-terminus (Nagamatsu *et al.*, 1984; Aronson and Arvidson, 1987), the resultant toxin was found to contain no cysteine residues and very few lysine residues. In the present study for example, toxin generated from *B. thuringiensis* subsp. *kurstaki* HD-73 contains only three of the original 34 lysine residues, and none of the 16 cysteine residues found in the protoxin (Bietlot *et al.*, 1989). It should further be noted that toxin encompasses the regions encoded by the conserved, variable and hypervariable regions of the protoxin gene sequence, discussed in Chapter 1. The unusual distribution of amino acid residues, and the spanning of the three distinct homology regions by the toxin molecule, both figure prominently in relating the structural features elucidated in this thesis to the toxin's function.

Once it was established which portion of the protoxin molecule is encompassed by the toxin, it was possible to estimate the absorption coefficient of toxin from the number of amino acids which absorb strongly at 280 nm. Assuming that tryptophan and tyrosine have absorption coefficients of 5700 M^{-1} cm^{-1} and 1300 M^{-1} cm^{-1}, respectively (Cantor and Schimmel, 1980), a solution of toxin at a concentration of 1 mg/ml would give an absorbance of 1.45 for a 1 cm pathlength at 280 nm. This estimate agrees favourably with that determined by amino acid analysis, which indicates that for a 1 cm pathlength, an absorbance of 1.61 at 280 nm corresponds to 1 mg/ml of toxin.

The pI of toxin was estimated from isoelectric focusing gels to be 5.5 ± 0.3. In contrast, the pI of a 58-kDa toxin prepared by trypsin digestion of protoxin from *B. thuringiensis* subsp. *dendrolimus* was reported to be 8.9 (Nagamatsu *et al.*, 1984). Using gels containing 6M urea, Aronson and Arvidson (1987) measured the pI of a 63-kDa toxin from *B. thuringiensis* subsp. *kurstaki* HD-263 to be 6.5. Presumably, the urea was

present to prevent the toxin from precipitating during focusing. When *B. thuringiensis* subsp. *kurstaki* HD-73 toxin prepared in the present study was similarly subjected to isoelectric focusing on urea gels, a single band with a pI of 6.0 ± 0.3 was obtained. In view of the unusual properties of toxin in chemical denaturants (see Chapter 5), it is probable that little toxin unfolding occurred during focusing in urea gels. A slightly acidic pI can be predicted simply from the amino acid composition of the toxin, since the toxin contains 152 acidic groups (Asp, Glu) and 129 basic groups (Lys, Arg and His). Indeed, prediction of the pI using the Pustell sequence analysis programme, which assumes that all amino acids are on the surface of the protein, gave a pI estimate of 6.3. However, given the intricate secondary and tertiary structure composition of toxin (Chapters 3 and 4), it is probable that some of the charged groups are buried within the protein and thus have altered ionization constants.

A pI close to neutrality is in agreement with the very low solubility observed for the toxin between pH 5-8. The low solubility of *B. thuringiensis* toxins in general has been noted by Andrews *et al.* (1987). Aronson and Arvidson (1987) indicate that solubility can be significantly improved at moderately alkaline pH values by adding 2M KSCN to the toxin sample. Since KSCN weakens hydrophobic interactions, it is possible that the observed low solubility of toxin is in part caused by intermolecular hydrophobic interactions. The buffer generally used in the present study to solubilize toxin, CAPS, contains a hydrophobic moiety and permits somewhat higher solubilities to be achieved than is possible with inorganic buffer salts. Interactions between hydrophobic areas on the toxin's surface and the hydrophobic region of the CAPS molecule would discourage association between toxin molecules. In all buffer systems tested, toxin was always most soluble at pH values of 10.5 or greater. The increased solubility may be due to deprotonation of tyrosine and lysine residues, converting the former residues to negatively charged species, and the latter to neutral groups. The molecule would therefore take on an overall negative charge which would help compensate for the general hydrophobicity of the

toxin. Given the highly basic nature of the Lepidopteran larval gut, good solubility and the retention of biological activity under the adverse environmental conditions of the insect gut are necessary if the toxin is to be physiologically active.

The resistance of toxin towards further proteolysis by larval gut enzymes, and by proteases in general, is one of the distinguishing features of this protein. Indeed, several protocols for the generation of toxin have advocated incubating protoxin crystals with insect gut juice in order to generate the protease-resistant toxin (Lecadet and Dedonder, 1967; Lilley *et al.*, 1980). However, the homogeneity of toxin prepared using a mixture of uncharacterized proteases must be questioned. In the present study, advantage is taken of the resistance of toxin to digestion by pancreatic trypsin to generate pure preparations of the protein. The observed stability towards proteolysis by a variety of proteases suggests that toxin has a tightly folded tertiary structure; this observation is substantiated in Chapter 4, which examines the domain structure of the toxin molecule.

Although protoxin and toxin have been widely reported to be glycoproteins (Fast, 1981; Bulla *et al.*, 1982; Andrews *et al.*, 1987), the reports are often contradictory, with values ranging from 12% carbohydrate (Bateson and Stainsby, 1970) to no carbohydrate (Huber and Lüthy, 1981). In the present study, less than 0.5% carbohydrate was found to be associated with protoxin or toxin from *B. thuringiensis* subsp. *kurstaki* HD-73. It has not been conclusively shown in any report that the observed sugar is covalently bound to the protein, although it has been demonstrated that the carbohydrate cannot be removed by treating the protein with 8M urea (Nickerson, 1980) or other denaturants (Muthukumar and Nickerson, 1987). Despite speculation that the carbohydrate may play a role in recognizing the insect gut binding receptor (Muthukumar and Nickerson, 1987; Haider and Ellar, 1987), no correlation was found between sugar content and toxicity (Nickerson, 1980). The evidence for the presence and role of carbohydrate in protoxin and toxin therefore remains ambiguous.

2.5 CONCLUSIONS

Toxin generated by trypsin digestion of purified *B. thuringiensis* protoxin crystals was shown by polyacrylamide gel electrophoresis and HPLC to be a homogeneous preparation free of contaminating peptides. The toxin is highly toxic towards spruce budworm larvae (Bietlot *et al.*, 1989). The protein spans amino acids 29-623 of the protoxin molecule, and therefore encompasses the regions encoded by the conserved, variable and hypervariable regions of the protoxin gene sequence. The evolution of these three distinct gene regions to code for the amino acid sequence of the toxin hints at the unusual biological function of the molecule. In contrast, few of the toxin's chemical characteristics reflect the unique *in vivo* action of this protein. One of the few distinguishing characteristics of the toxin is that it attains its maximum solubility only at pH values of 10.5 or higher. This coincides with the approximate gut pH of susceptible Lepidopteran larvae, and thus represents the environment in which the toxin must be soluble and stable if it is to be biologically active. The demonstrated resistance of the toxin towards degradation by a number of proteases illustrates the suitability of the toxin for withstanding the hostile environment of the insect gut. The results of structural studies on the toxin, which constitute the remainder of this thesis, provide insights into some of the novel functional aspects of this unique protein and its protoxin precursor.

2.6 REFERENCES

Adang, M. J., M. J. Staver, T. A. Rocheleau, J. Leighton, R. F. Barker and D. V. Thompson (1985). Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. *Gene* **36**, 289-300.

Andrews, R. E., M. M. Bilbos and L. A. Bulla (1985). Protease activation of the entomocidal protoxin of *Bacillus thuringiensis* subsp. *kurstaki*. *Appl. Environ. Microbiol.* **50**, 737-742.

Andrews, R. E., R. M. Faust, H. Wabiko, K. C. Raymond and L. A. Bulla (1987). The biotechnology of *Bacillus thuringiensis*. *CRC Crit. Rev. Biotechnol.* **6**, 163-232.

Aronson, J. N. and H. C. Arvidson (1987). Toxic tryptic digest fragment from the *Bacillus thuringiensis* parasporal protein. *Appl. Environ. Microbiol.* **53**, 416-421.

Bietlot, H., P. R. Carey, C. T. Choma, H. Kaplan, T. Lessard and M. Pozsgay (1989). Facile preparation and characterization of the toxin from *Bacillus thuringiensis* var. *kurstaki*. *Biochem. J.* **260**, 87-91.

Bulla, L. A., D. B. Bechel, K. J. Kramer, Y. I. Shethna, A. I. Aronson and P. C. Fitz-James (1982). Ultrastructure, physiology and biochemistry of *Bacillus thuringiensis*. *CRC Crit. Rev. Microbiol.* **8**, 147-189.

Cantor, C. R. and P. R. Schimmel (1980). *Biophysical chemistry, part II: Techniques for the study of biological structure and function.* W. H. Freeman and Co., San Francisco, pp. 380-381.

Carey, P. R., P. Fast, H. Kaplan and M. Pozsgay (1986). Molecular structure of the protein crystal from *Bacillus thuringiensis*: a Raman spectroscopic study. *Biochim. Biophys. Acta* **872**, 169-176.

Chestukhina, G. G., L. I. Kostina, A. L. Mikhailova, S. A. Tyurin, F. S. Klepikova and V. M. Stepanov (1982). The main features of *Bacillus thuringiensis* delta-endotoxin molecular structure. *Arch. Microbiol.* **132**, 159-162.

Dubois, M., K. A. Gilles, J. K. Hamilton, P. A. Rebers and F. Smith (1956). Colorimetric method for determination of sugars and related substances. *Anal. Chem.* **28**, 350-356.

Edman, P. and G. Begg (1967). A protein sequenator. *Eur. J. Biochem.* **1**, 80-91.

Fast, P. G. (1981). The crystal toxin of *Bacillus thuringiensis*. (in) *Microbial control of pests and plant diseases* (H. D. Burgess, ed.) Academic Press, London, pp. 223-248.

Glazer, A. N. and E. L. Smith (1971). Papain and other plant sulfhydryl proteolytic enzymes. (in) *The Enzymes*, vol. III (P. D. Boyer, ed.) Academic Press, New York, pp. 501-537.

Haider, M. Z. and D. J. Ellar (1987). Analysis of the molecular basis of insecticidal specificity of *Bacillus thuringiensis* crystal delta-endotoxin. *Biochem. J.* **248**, 197-201.

Huber, H. E. and P. Lüthy (1981). *Bacillus thuringiensis* delta-endotoxin: Composition and activation. (in) *Pathogenesis of invertebrate microbial diseases* (E. W. Davison, ed.) Allanheld, Osmum and Co., Totowa, N.J., pp. 209-234.

Krieg, A. and H. G. Miltenburger (1984). Bioinsecticides: I. *Bacillus thuringiensis*. (in) *Advances in Biotechnological Processes*, vol. 3, Alan R. Liss Inc., New York, pp. 273-290.

Laemmli, U. K. (1979). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lecadet, M. M. and R. Dedonder (1967). Enzymatic hydrolysis of the crystals of *Bacillus thuringiensis* by the proteases of *Pieris brassicae*. I. Preparation and fractionation of the lysates. *J. Invertebr. Pathol.* **9**, 310-321.

Lilley, M., R. N. Ruffell and H. J. Somerville (1980). Purification of the insecticidal toxin in crystals of *Bacillus thuringiensis*. *J. Gen. Microbiol.* **1**, 1-8.

Matsudaira, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**, 10035-10038.

Muthukumar, G. and K. W. Nickerson (1987). The glycoprotein toxin of *Bacillus thuringiensis* subsp. *israelensis* indicates a lectinlike receptor in the larval mosquito gut. *Appl. Environ. Microbiol.* **53**, 2650-2655.

Nagamatsu, Y., Y. Itai, C. Hatanaka, G. Funatsu and K. Hayashi (1984). A toxic fragment from the entomocidal crystal protein of *Bacillus thuringiensis*. *Agric. Biol. Chem.* **48**, 611-619.

Nickerson, K. W. (1980). Structure and function of the *Bacillus thuringiensis* protein crystal. *Biotech. Bioeng.* **22**, 1305-1333.

Sluyterman, L. A. (1967). The effect of methanol, urea and other solutes on the action of papain. *Biochim. Biophys. Acta* **139**, 418-429.

Svensson, H. (1962). Isoelectric fractionation, analysis and characterization of ampholytes in natural pH gradients. III. Description of apparatus for electrolysis in columns stabilized by density gradients and direct determination of isoelectric points. *Arch. Biochem. Biophys. Suppl. I*, 132-136.

Chapter 3.

THE SECONDARY STRUCTURE OF THE INSECTICIDAL TOXIN

Technique

The height of
technical felicity
is to combine
sublime simplicity
with just sufficient
ingenuities
to show how difficult
to do it is.

- Piet Hein

Chapter 3.

THE SECONDARY STRUCTURE OF THE INSECTICIDAL TOXIN

3.1 INTRODUCTION

When protein fibres such as silk were first subjected to X-ray analysis, it became evident that the constituent molecules of the fibre contained local, ordered structures (Astbury, 1938). The spatial distances of the repeating units determined from such fibres were instrumental in the elucidation of three ordered arrangements found in proteins: the α–helix, and the parallel and antiparallel ß-pleated sheet (Pauling *et al.*, 1951). The X-ray structures which were solved in the 1960's established that helices, sheets and other structural features are universal in proteins. Together with the protein refolding experiments of Anfinsen and Haber (1961), these structures confirmed the hypothesis of Linderström-Lang (1959) that the protein polypeptide chain is organized into a multi-level hierarchy, with the lower level elements determining the structures of the higher elements. Thus, the amino acid sequence (or primary structure) contains all the information required for the protein to fold into its unique, functional, three dimensional shape or conformation. Repeating, regular arrangements of the polypeptide backbone form the principal structural units from which the protein is constructed; these units are the secondary structure elements of helices, sheets and turns. Regions of non-repetitive structure, called random coils, are also classified as secondary structure. The complex interactions of secondary structure assemblies with each other give rise to the higher architectural levels of supersecondary structures, domains, globular proteins and aggregates (Creighton, 1984; Jaenicke, 1987), culminating in the highly ordered, functional protein.

The studies presented in this thesis were undertaken with a view to determining the overall architecture of the protoxin and toxin from *B. thuringiensis* subsp. *kurstaki* HD-73. Since it is the active insecticidal protein, emphasis is placed on elucidating the underlying structure of toxin. Prior to this study, no experimental inquiries into the secondary structure of any *B. thuringiensis* toxin had been reported, although predictive methods had been applied to protoxin gene-deduced amino acid sequences (Schnepf *et al.*, 1985; Shibano *et al.*, 1985), and Raman spectroscopic studies on crystals from three strains of *B.*

thuringiensis had been conducted (Carey *et al.*, 1986; Pozsgay *et al.*, 1987). The purpose of the present comprehensive investigation into the secondary structure of toxin was to further characterize the protein, and to determine if any unusual features in the secondary structure contribute to the unique properties of this insecticidal protein.

3.1.1 Techniques for elucidating protein secondary structure

Secondary structures are non-superimposable, three-dimensionsional arrangements of atoms arising from the rotation of bonds in the polypeptide backbone; the overall spatial arrangement of atoms in each secondary structure element is often referred to as its conformation. The assembly of secondary structure units to form the functional protein in turn gives rise to the protein's tertiary structure. The basis for the experimental estimation of protein secondary structure is the interaction of light with the different structural elements in the folded polypeptide backbone. In principle, the spectral signature of each secondary structure element is unique and should provide an unambiguous, quantitative measure of the individual structural units within the protein. In practice, the interaction of the protein environment with the polypeptide chain and associated chromophores gives rise to a confusing range of signals, making it difficult to extract structural information from the conformation-sensitive regions of the spectrum. Although these difficulties could be avoided by employing X-ray crystallography, the obvious practical limitations of crystallographic analysis often preclude the use of this technique. In the course of the present study for example, efforts to crystallize the *B. thuringiensis* subsp. *kurstaki* HD-73 toxin resulted only in the production of microcrystallites. Since relatively few proteins can be successfully crystallized, mathematical methods have been developed for deciphering conformational information from the spectra of proteins in solution.

Three spectroscopic techniques and two predictive methods were used to elucidate the secondary structure of toxin. A variety of approaches were utilized as each technique has particular strengths and weaknesses; thus, a consensus estimate should provide the most reliable assessment of the toxin's secondary structure. The following sections

summarize the advantages and disadvantages of each technique, and outline the steps employed to extract conformational information from the spectra.

3.1.1.1 Raman Spectroscopy

When a population of molecules is illuminated with monochromatic light of a frequency not absorbed by the sample, most of the photons are transmitted without being affected. However, some of the incident light is scattered, and while most of this scattered light has the same wavelength as the incident radiation, a small proportion of the photons transfer energy to molecules in the sample, causing those molecules to enter a higher vibrational state. These photons are then scattered from the sample at frequencies lower than that of the incident beam (Van Wart and Scheraga, 1978). As a result of these frequency shifts, a collection of spectral lines can be observed at right angles with respect to the incident beam; these lines constitute the Raman spectrum of the sample.

The Raman spectrum of a protein consists of contributions from the peptide backbone and various amino acid sidechains. The -CO-NH- peptide bonds have a number of vibrational degrees of freedom (or vibrational modes) which give rise to several spectral bands. Of these bands, the frequencies of the amide I band (due primarily to C=O stretching) and the amide III band (arising largely from N-H bending) are sensitive to the secondary structure of the protein (Yu, 1977). The characteristic frequencies of the various secondary structure elements have been determined from spectra of polypeptide models and of proteins whose X-ray structures are known (Carey, 1982). In practice, the assignment of spectral features arising from secondary structure components relies heavily on data gathered from model polypeptides, as assignments from proteins of known structure are complicated by overlapping bands in the amide III region arising from amino acid side chains.

Several methods have been put forward in the literature for quantifying the various amounts of secondary structure in a protein from its Raman spectrum (Lippert *et al.*, 1976; Williams, 1986). In the present study, the method of Lippert *et al.* was applied to Raman

spectra of toxin. This method was chosen for its ease of application to non-digitized data (Raman spectra were obtained on a strip-chart recorder) and because of the method's good success rate in estimating the secondary structure content of proteins (Tu, 1986).

Depending on the pH of its environment, poly(L-lysine) can assume either an α–helix, ß-sheet or random coil conformation (Chen and Lord, 1974). Lippert *et al.* (1976) used poly(L-lysine) at various pH values in water and ^2H_2O to determine which frequencies of the Raman spectrum were most responsive to the conformations of the model polypeptide. Data from the amide I and amide III regions in ^2H_2O (designated amide I' and amide III') were required as the amide I band in aqueous samples is obscured by the strong water bending mode centered at 1645 cm^{-1} (Harada and Takeuchi, 1986). It was determined that the Raman intensity at three frequencies could describe the overall secondary structure of a protein sample. In aqueous samples, the amide III band intensity at 1240 cm^{-1} is due to random coil and ß-sheet structure. In ^2H_2O, the signal at 1632 cm^{-1} in the amide I' band is due to α–helix content, while random coil and ß-sheet give rise to the amide I' band intensity at 1660 cm^{-1}. For the purpose of analysis, the Raman amide modes are assumed to be a summation of linear superpositions of bands due to the component conformations. Thus, if the peak heights of a protein spectrum at 1240 cm^{-1} in H_2O and 1632 and 1660 cm^{-1} in ^2H_2O are measured relative to a band whose intensity is not affected by conformation or deuteration, such as the CH_2 bending mode at 1448 cm^{-1}, then these relative spectral intensities will be additive functions of the amount of α–helix, ß-sheet and random coil in the protein. The relationship between conformational content of a protein and the above relative spectral intensities obtained from the protein's Raman spectra is given by four simultaneous equations which are solved to give the relative proportions of each structure (Lippert *et al.*, 1976).

Both Lippert *et al.* (1976) and Williams (1986) claim good agreement between estimates of structure content obtained from Raman spectra and results derived using other techniques. However, neither method has been rigorously tested for accuracy in estimating

protein secondary structure. The method of Lippert *et al.* has been criticized because its basis is poly(L-lysine) conformations, rather than proteins of known X-ray structure. However, estimation methods developed using one X-ray structure database do not necessarily give accurate estimates of structure when tested against a different X-ray structure database (Dev, 1987). In common with all the techniques available for estimating protein secondary structure content, estimates obtained from Raman spectra must be treated with caution. In addition, acquiring quality Raman spectra of proteins is often difficult due to intense fluorescence of chromophores within the sample (Carey, 1982). Nonetheless, Raman spectroscopy has the advantage over infrared spectroscopy and circular dichroism in that it can provide information on the environment experienced by a number of side chains, particularly tyrosine and tryptophan (Tu, 1986).

3.1.1.2 Infrared spectroscopy

Like Raman spectroscopy, infrared (IR) spectroscopy probes the vibrational energy levels of bonds within a molecule. However, in IR spectroscopy only those transitions between vibrational levels of the molecule's ground state which result from the absorption of light in the IR region of the electromagnetic spectrum (10^3 to 10^5 nm) are detected. These vibrational levels are generated by the characteristic stretching and bending motions of various bonds within the molecule. The amide I band (1620-1700 cm^{-1}) is primarily a stretching vibration of the C=O bond and is the band most sensitive to polypeptide backbone conformation (Krimm, 1962). IR studies of model polypeptides have lead to the identification of amide I frequencies arising from α–helix, ß-sheet, turn and random coil conformations (Susi and Byler, 1983). However, in proteins the amide I band is a composite of these overlapping bands. This has made it difficult to determine the number and relative proportions of the secondary structure components contributing to the amide I band profile of a protein.

In order to overcome this obstacle, computational procedures have been developed for decomposing the complex amide I band into its underlying components (Susi and

Byler, 1983). The most effective procedure for narrowing IR bands is Fourier self-deconvolution (Kauppinen *et al.*, 1981). In this method, the cosine Fourier transform of the original broad infrared band $A(\nu)$ (Fig. 3.1) yields the exponentially decaying function $I(x)$. The wider the original absorbance band, the more rapidly its Fourier transform decays. Therefore, if the rate of decay of this exponential term were decreased, the width of the bands in the infrared spectrum would be reduced. This can be achieved by multiplying $I(x)$ with an exponentially increasing function to yield a new function, $I'(x)$. By performing the reverse Fourier transform on $I'(x)$, a new absorbance band $A'(\nu)$ is obtained which is narrower than $A(\nu)$. Deconvolution of the original amide I band by this method permits the determination of the number of band subcomponents and their frequencies. The secondary structure of a protein is then estimated by band-fitting analysis of the original amide I band, using the information obtained from the deconvoluted spectrum as input. Curve fitting is performed by iterative adjustment of the heights and half-widths of the component bands until the best fit of the original contour is obtained. The fractional areas of the fitted component bands are directly related to the relative populations of the conformational structures represented by these components (Byler and Susi, 1986).

The general validity of this approach has been tested; good estimates were obtained from the IR spectra of proteins whose X-ray structures are known (Susi and Byler, 1986). However, deconvolution and band-fitting cannot be used indiscriminately. Random noise and water vapour in the original spectrum may appear in the deconvolved spectrum as artifacts, indistinguishable from real amide bands, and curve-fitting involves a certain degree of subjectivity (Surewicz and Mantsch, 1988). Nonetheless, IR spectroscopy is a valuable method for providing estimates of the secondary structure of a protein either in solution or as a solid.

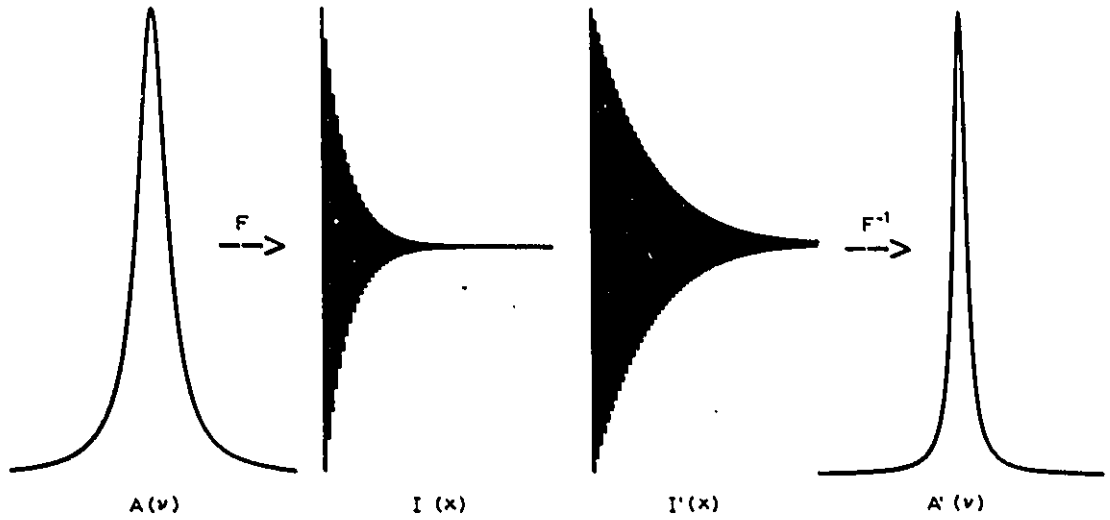


Figure 3.1 Illustration of the process of band-narrowing by Fourier self-deconvolution.
The cosine Fourier transform of the original infrared band A(ν) yields the function I(x). Multiplication of I(x) with an increasing exponential function gives I'(x) which, after reverse Fourier transformation, leads to the new, narrower infrared band A'(ν) (Surewicz and Mantsch, 1988).

3.1.1.3 Circular dichroism

Circular dichroism (CD) differs fundamentally from the two techniques described above. The CD spectrum of a protein is not generated by the vibrations of bonds in the sample molecules, but rather by their differential absorption of left and right circularly polarized light. The differential absorption results primarily from the spatial asymmetry of the constituent amino acids of the polypeptide backbone, and thus provides information on the secondary structure of the protein.

The analysis of CD spectra for secondary structure content was initiated by Greenfield and Fasman (1969) using poly(L-lysine) at various pH values. However, unlike model polypeptides, proteins do not contain long stretches of any single structure. This approach was therefore found to have limited application in estimating protein

secondary structure, although it was valuable for identifying the spectral characteristics of the four principal secondary structure elements.

In the present study, two more recent approaches for quantifying structure were applied to the CD spectrum of toxin. In the first method, a linear summation of a set of reference spectra characteristic of a particular conformational class are fitted to the sample spectrum. These standard spectra are deduced from proteins of known X-ray structures (Chen *et al.*, 1972; Chang *et al.*, 1978). However, as CD can only be conducted on proteins in solution, it must be assumed that the secondary structure of each solubilized reference protein is identical to that of the crystallized protein. In addition, it is unlikely that a reference spectrum can accurately represent the variations observed amongst proteins for a given secondary structure element (*e.g.*, the length of a ß-sheet affects its CD contribution). Another drawback is that the number of spectra in the reference set usually exceed the information content of the data for the protein of unknown structure. As a result, more than one solution can generally be found to fit the data equally well (Johnson, 1988). In an effort to constrain the instability of the solution for a sample spectrum and to overcome the problem of defining reference spectra, Provencher and Glöckner (1981) proposed a method in which the sample CD spectrum is analyzed directly as a linear combination of the digitized CD spectra of sixteen proteins of known structure. The CONTIN programme by Provencher conducts the necessary calculations. Using this method, the estimates of secondary structure content derived for various proteins of known X-ray structure were usually good, although some estimates of ß-sheets and turns were very inaccurate (Provencher and Glöckner, 1981). In general, CD can be relied upon to provide good estimates of α–helix content since (i) the spectrum arising from this structural element occurs at wavelengths (>190 nm) far from the absorption band of water vapour and (ii) the shape of the α–helix spectrum is distinct from the spectra of the other secondary structure elements.

3.1.1.4 Predictive methods

Since instructions for the specific conformation of a protein are believed to reside in the amino acid sequence, it should in principle be possible to predict conformation solely from the protein's primary structure. However, there is uncertainty as to how protein folding occurs, and what drives one conformation to be strongly preferred over a host of other possible conformations. The methods presently available for predicting protein secondary structure from the amino acid sequence therefore have limited accuracy. Nonetheless, they provide a useful tool for estimating the location of individual predicted secondary structure elements within the polypeptide chain. Predictive methods therefore compliment experimental estimates of protein secondary structure.

Each predictive method is based on one of several approaches (probabilistic, physicochemical or stereochemical) which are used to define in which particular secondary structure conformation a given amino acid residue will participate. The two predictive methods applied in the present study to the toxin amino acid sequence, that of Chou and Fasman (1978) and Garnier *et al.* (1978), belong to the probabilistic class. In this approach, the frequencies with which the amino acid residue types participate in each of three conformational states (α–helix, ß-sheet and random coil) are determined from a data base of known protein structures. These frequencies are then taken as propensities for a given residue type to occur in the respective conformational state. The propensities can be used to predict structural sequence regions if a sequential cluster of four or more amino acids with preferences for a particular secondary structure greater than the defined threshold value are found. In the case of the Chou-Fasman method, each amino acid residue is inspected in turn. In the method of Garnier *et al.* (1978), the influence on a specific residue of other amino acids up to ten positions away on the polypeptide chain is taken into consideration before assigning the residue to a structural class. The method of Garnier *et al.* has a 57% probability of assigning a particular amino acid to the correct class, as compared to 47% for the Chou-Fasman method (Kabsch and Sander, 1983). Both these

values are better than chance alone, which 33% of the time would correctly assign a given residue to one of three structural classes. However, the problem of ambiguous criteria for some aspects of secondary structure (such as the ends of α–helices and ß-sheets) decreases the accuracy of predictive methods (Argos and MohanaRao, 1986). In addition, the rules governing the assignment of a residue to a structural class are often vague and as a result the same prediction method applied to the same protein can yield different results in the hands of different investigators (Schulz, 1988). Finally, prediction programmes and experimental techniques for estimating secondary structure have all evolved from analysis of soluble, globular protein structures, but preliminary neutron scattering studies indicate that toxin has an elongated, flat shape (personal communication, X. Lee, Division of Biological Sciences, N.R.C.). It is therefore particularly important that predictive methods and experimental techniques be used in conjunction with each other in order to determine if a concensus estimate of the secondary structure of toxin can be obtained.

3.2 MATERIALS AND METHODS

3.2.1 Materials

Toxin was prepared from protoxin crystals and characterized for homogeneity as described in Sections 2.2.3 and 2.2.4. Deuterium oxide and *d*-10-camphorsulphonic acid were obtained from Aldrich Chemical Co., Milwaukee, Wisconsin. CAPS buffer salt was purchased from Sigma Chemical Co., St. Louis, Missouri. Solutions prepared with H_2O used reverse-osmosis quality water purified by the Milli-Q water system, Millipore Canada Ltd., Mississauga, Ontario.

3.2.2 Raman spectroscopy

The sample cell for hydrated solid toxin samples consisted of a quartz capillary tube approximately 2 mm i.d., sealed at one end. A toxin pellet was obtained by spinning a suspension of toxin in H_2O for five minutes in a clinical centrifuge. Deuterated samples

were obtained by pelleting and resuspending the toxin sample three times in fresh ^2H_2O over a 72 hour period, and then pelleting the protein in the quartz capillary. The Raman spectrum of the pelleted protein was obtained by orienting the incident laser beam, pellet and spectrophotometer optics slit such that only those photons leaving the sample 90^o to the incident beam entered the spectrophotometer (Carey, 1982). The samples were illuminated with 150 - 180 mW of the 514.5 nm line from an argon laser, and the Raman photons were detected using a Jarrell-Ash 0.75 m double monochromator. Spectra were collected on a strip-chart recorder and calibrated using the lines from an argon lamp. Thus, peak positions reported here are accurate to within 2 cm^{-1}.

Spectral measurements on air-dried toxin were obtained by packing a pin hole in a stainless steel rod with dry protein, and orienting the pin hole in the focussed laser beam so that the beam grazed the surface of the sample. The laser was operated at 50 mW, and the scattered light was examined in the usual 90^o excitation-detection geometry (Carey, 1982). An average of six spectra were collected for each data set. Spectra could not be obtained for toxin in solution since very high protein concentrations, beyond the solubility limit of toxin, are required for Raman spectroscopy.

3.2.3 Infrared spectroscopy

The vibrational mode most useful for the analysis of protein secondary structure is the amide I band between approximately 1600 and 1700 cm^{-1}. As water gives a strong absorption band centered at 1640 cm^{-1}, toxin samples were deuterated by pelleting and resuspending the toxin three times in ^2H_2O over a 72 hour period. The protein was then dissolved in ^2H_2O containing 0.1M CAPS made alkaline to p^{2}H 10.5 by the addition of NaOH in ^2H_2O, and then incubated for 24 hours at 4^oC prior to scanning.

The direct comparison of spectra from solubilized and solid samples is not straightforward due to the much slower rate of $^1H \rightarrow ^2H$ exchange in solids. In order to compare spectra of solid and solubilized toxin deuterated to the same degree, part of a

solubilized toxin sample in alkaline ^2H_2O was used to obtain the solution spectrum. The other part was neutralized with HCl diluted in ^2H_2O, and the spectrum of the resulting precipitate was immediately measured.

Samples were assembled between calcium fluoride windows separated by a 50 μm teflon spacer. Infrared spectra were recorded at 22°C with a computerized Digilab FTS-60 instrument. For each spectrum, 250 interferograms were co-added and Fourier-transformed to give a resolution of 2 cm^{-1}. Spectra were corrected for the weak absorbance of the CAPS/^2H_2O buffer. In order to eliminate the spectral contributions of atmospheric water vapour, the sample compartment was continuously flushed with dry nitrogen. The technique of Fourier deconvolution was used to resolve overlapping infrared bands (Kauppinen *et al.*, 1981; Surewicz and Mantsch, 1988), and curve fitting was performed using least squares procedures (Fraser and Suzuki, 1966).

3.2.4 Circular Dichroism.

As two CD spectropolarimeters were available, spectra of toxin were obtained in duplicate on each, and the data were analyzed using two different programmes for estimating secondary structure content. Spectra were acquired on both an Aviv Model 61DS solid state CD spectropolarimeter interfaced with an AT&T personal computer, and on a Jasco J600 instrument interfaced with an IBM microcomputer. Both spectropolarimeters were calibrated with 0.1% (w/v in water) *d*-10-camphorsulphonic acid (Chen and Yang, 1977). Spectra were analyzed for secondary structure content by the nonlinear, least squares curve-fitting procedure of Chang *et al.* (1978), and independently by the CONTIN programme (obtained from the Max Planck Institute for Biophysical Chemistry, Heidelberg), based on the procedure of Provencher and Glöckner (1981).

Toxin was dissolved in 25mM CAPS buffer, pH 10.5, to a concentration of 0.5 to 1 mg/ml. The concentration of protein in each sample was determined by amino acid analysis as described in Section 2.2.4.3. Protein solutions were placed in a 0.2 mm pathlength cylindrical quartz cell and spectra were recorded in the far ultra-violet region

between 184 to 240 nm on the Aviv instrument, and between 190 to 240 nm on the Jasco spectropolarimeter. Typically, eight spectra were averaged and the buffer baseline subtracted before the data were analyzed for secondary structure content.

Data analysis required that each digitized spectrum be converted to units of mean residue molar ellipticity ($[\theta]_\lambda$). This value was calculated at increments of 0.2 nm from the following relationship:

$$[\theta]_\lambda \text{ (deg cm}^2 \text{ dmol}^{-1}) = (66{,}700/593)\theta_\lambda /100[P]l$$

where θ_λ is the observed ellipticity in degrees, l is the pathlength in cm, [P] is the protein concentration in mol dm^{-3}, 66,700 is the molecular weight of toxin, and 593 is the number of peptide bonds per protein molecule.

3.2.5 Predictive methods

The amount of α–helix, ß-sheet and random coil were predicted from the gene-deduced amino acid sequence of toxin using the methods of Chou and Fasman (1978) and Garnier *et al.* (1978). Both methods were accessed through the *PeptideStructure* programme, obtained from the Max von Pettenkofer Institute, University of Munich. By incorporating the refinements of Nishikawa (1983), the programme resolves many ambiguous assignments in the original Chou-Fasman method. The location of the predicted secondary structure elements along the polypeptide chain were displayed graphically using the companion programme *PlotStructure*. The *PepPlot* programme, obtained from the McArdle Research Laboratory, University of Wisconsin, permitted the 'weight' of each amino acid assignment to be judged from the graphical display, since the greater the intensity of the peak, the more satisfied the criteria for the prediction of that structural element. Among other capabilities, *PepPlot* displays the arrangement of acidic, basic, hydrophobic and hydrophilic residues in the sequence, permitting the general character of segments of the protein to be assessed. This information was used to calculate

hydropathy curves for the toxin and protoxin using the method of Kyte and Doolittle (1982).

3.3 RESULTS

3.3.1 Raman Spectroscopy

Raman data have previously been obtained for *B. thuringiensis* subsp. *kurstaki* HD-73 protoxin, both as air-dried crystals and as an aqueous solid pellet (Carey *et al.*, 1986). Table 3.1 compares the Raman frequencies and relative peak intensities of the protoxin spectra with corresponding data collected during the present study for toxin. Data from a deuterated solid toxin pellet is included in the table, but there is no corresponding data in the literature for protoxin. For each toxin spectrum, a smooth baseline was drawn to intersect the trace. In the case of the air-dried pellet for example (Fig. 3.2), the baseline converges with the trace at 700, 800, 915, 1160, 1520 and 1780 cm^{-1}. By convention, the Raman peak near 1453 cm^{-1} is used as an internal standard against which all other peak heights are measured, since this feature arises primarily from the CH_2 scissoring mode of methylene side chain groups and thus is not influenced by protein conformation. Therefore, the 1453 cm^{-1} feature was arbitrarily assigned an intensity of 10.0, and the intensities of all other peaks in the same spectrum were expressed relative to this internal standard. The sixth column in Table 3.1 lists the amino acid moieties making major contributions to each observed Raman feature. Peak assignments were taken from the literature (Thomas and Prescott, 1983; Parker, 1983); assignments for the principal Raman peaks obtained from an air-dried toxin pellet are shown in Fig. 3.2.

The amide I band from the toxin pellet in H_2O appears as a strong feature with an intensity maximum at 1664 cm^{-1} (Table 3.1 and Fig. 3.3). This frequency is in the middle of the range expected for random coil structures (1660-1670 cm^{-1}) and at the lower end of the range for polypeptides in ß-sheet structures (1665-1680 cm^{-1}) (Parker, 1983). The amide I band shifts to 1669 cm^{-1} when the protein is air-dried, and to 1658 cm^{-1} when the

Table 3.1 Raman Frequencies (cm^{-1}), Relative Intensities,[a] and Assignments for Toxin from *Bacillus thuringiensis* Subsp. *kurstaki* HD-73

Protoxin (H_2O)[b]	Protoxin (Dry)	Toxin (H_2O)	Toxin (^2H_2O)	Toxin (Dry)	Assignment[c]
770 (1.8)	762 (2.6)	759 (2.5)	763 (1.3)	760 (2.3)	W
830[d]	831 (3.5)	830 (5.1)	831 (3.8)	830 (2.6)	Y
854	853 (3.7)	850 (4.3)	850 (3.4)	852 (3.0)	Y
		879 (1.0)			G
		902 (0.7)			A
				926 (1.5)	T
		936 (1.7)	935 (br)		V, L
	968 (1.4)	960 (1.0)	957 (br)	960 (1.6)	V, L
					(Amide III in ^2H_2O)
1005 (7.4)	1005 (6.8)	1005 (11.8)	1005 (10.3)	1005 (7.5)	F
1034 (3.6)	1032 (2.6)	1036 (3.6)	1032 (2.9)	1031 (2.5)	F, G
		1053 (2.0)			K, E, S
		1075–1105 (br)			E, T, A
	1104 (1.6)			1102 (1.8)	A
	1126 (1.6)	1127 (2.6)	1130 (1.3)	1128 (1.8)	V, L, I, E, D, CC_{α}
			1162 (sh)		V, L
1167 (sh)	1180 (2.1)	1174 (2.3)	1180 (4.0)	1180 (1.7)	V, L
1212 (sh)	1211 (sh)	1206 (5.7)	1210 (7.7)	1207 (4.6)	F, Y
1245 (7.8)	1245 (7.1)	1246 (7.8)	1253 (4.8)	1256 (6.2)	Amide III
1318 (sh)	1310 (sh)	1317 (sh)	1320 (sh)	1312 (6.2)	CH_2 twist/wag
1338 (8.5)	1337 (8.0)	1338 (7.1)	1335 (5.8)	1338 (7.4)	CH_2 twist/wag
	1358 (sh)	1359 (sh)		1358 (sh)	W
		1405 (sh)	1413 (br, sh)		$-COO^-$
		1425 (sh)			CH_2 scissor
1452 (10.0)	1453 (10.0)	1453 (10.0)	1453 (10.0)	1453 (10.0)	CH_2 scissor
1557 (3.8)	1554 (3.7)	1557 (4.2)	1554 (3.1)	1557 (2.4)	W
1586 (sh)	1585 (sh)	1587 (sh)	1584 (sh)	1585 (sh)	W
1612 (8.9)	1614 (6.2)	1616 (8.3)	1610 (5.2)	1614 (5.8)	Y
1664 (13.1)	1669 (11.1)	1664 (13.7)	1658 (9.2)	1669 (9.5)	Amide I

[a] Relative intensities in parenthesis.
[b] Data for Btk HD-73 protoxin from Carey *et al.* (1986).
[c] Assignments based on those in the literature (Thomas *et al.*, 1983; Parker, 1983).
[d] $I_{830} > I_{853}$ (Carey *et al.*, 1986).
Abbreviations: br, broad; sh, shoulder; str, stretch.

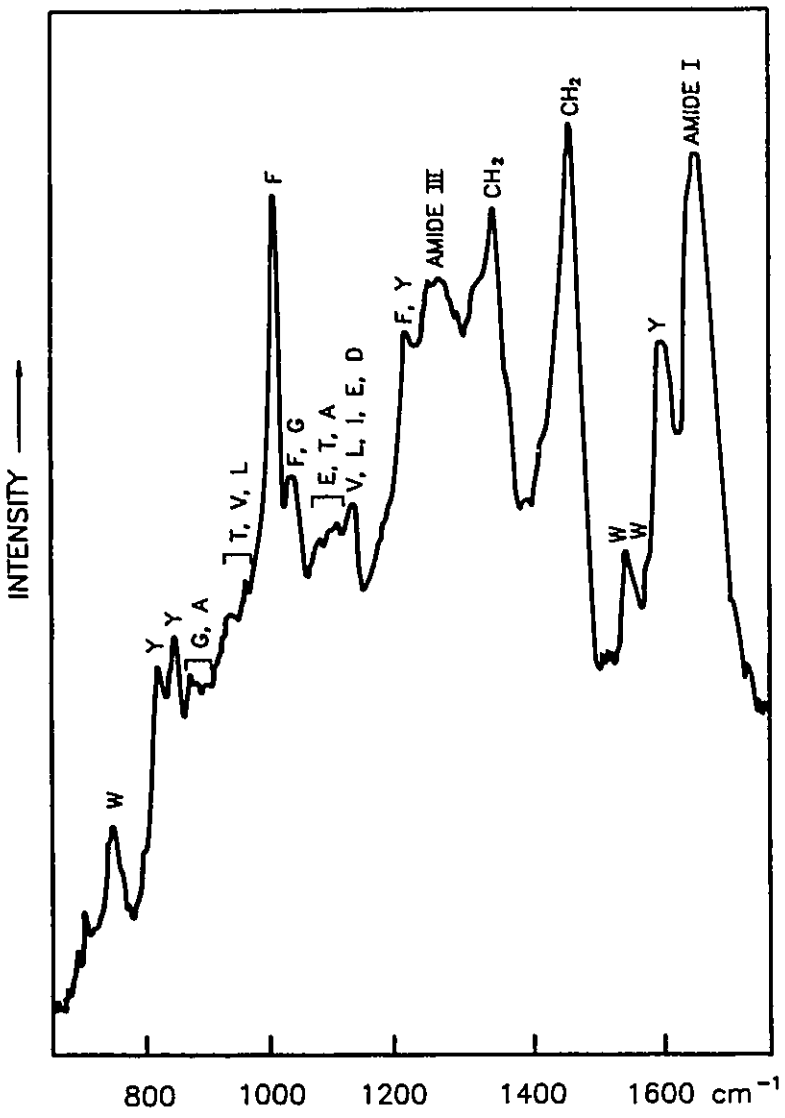


Figure 3.2 Raman spectrum of air-dried toxin.
The more prominent spectral features are labelled; single letters indicate the amino acid side-chains giving rise to the feature. The argon laser 514.5 nm wavelength was used at a power of approximately 50 mW.

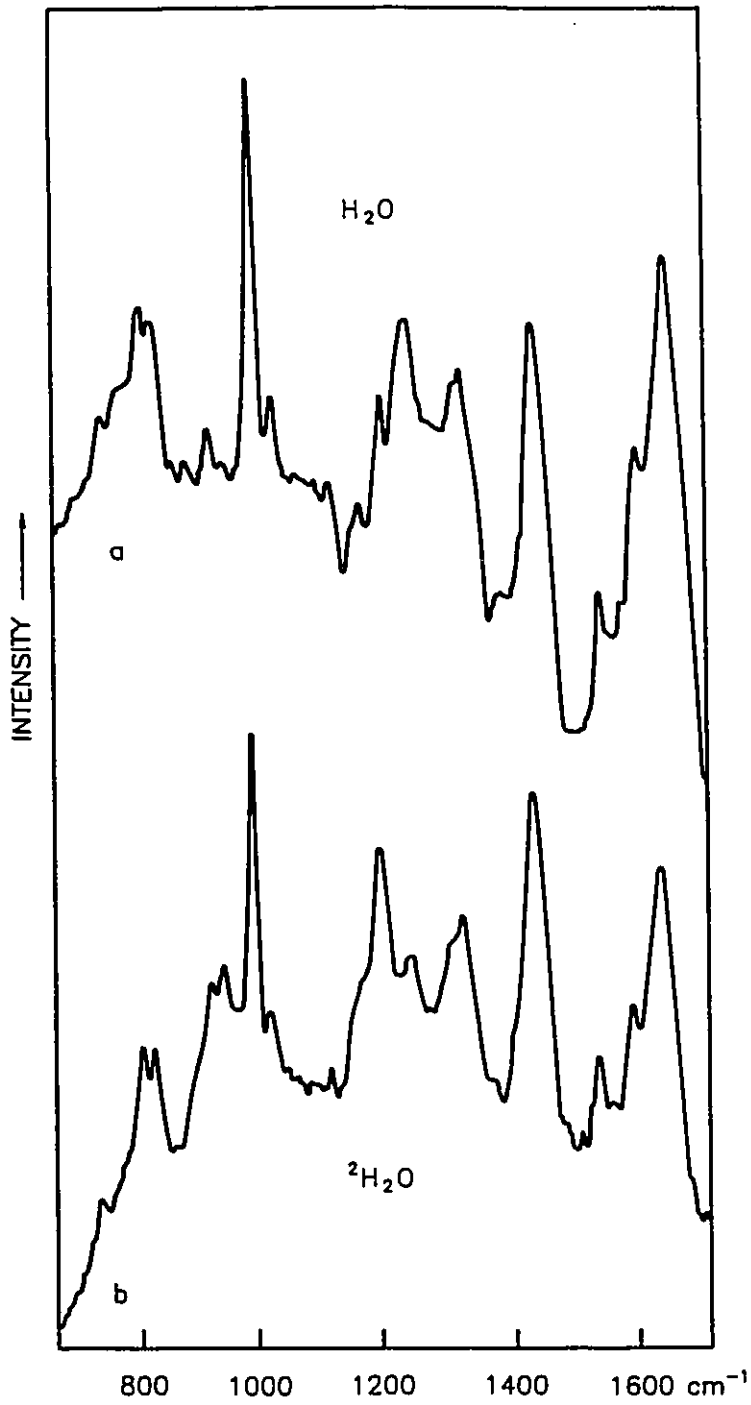


Figure 3.3. Raman spectra of hydrated and deuterated solid toxin.

Comparison of the Raman spectra obtained from toxin as an aqueous pellet in (a) H_2O and (b) ^2H_2O. The argon laser 514.5 nm wavelength was used at a power of approximately 160 mW.

protein pellet is deuterated (Table 3.1). The former shift indicates that a small conformational change probably occurs upon drying the protein, while the latter shift is a consequence of hydrogen-deuterium exchange in the peptide linkages (Thomas and Prescott, 1983).

The amide III region, which spans 1220-1310 cm^{-1}, is also diagnostic of polypeptide structure. For ß-sheet structures the amide III signature occurs as an intense band between 1220 and 1240 cm^{-1}; for α–helix structures, a band of much lower relative intensity occurs in the 1270-1310 cm^{-1} range. The amide III features due to random coil structures are intermediate between those for helical and sheet structures. The spectrum from the hydrated toxin pellet shows a broad amide III peak in the 1225-1275 cm^{-1} region, with a maximum at 1246 cm^{-1} (Table 3.1), suggesting the presence of significant amounts of random coil structure. A small change in the amide III profile of the dried toxin again indicates a minor conformational change. A much larger spectral shift occurs when the protein is deuterated: the amide III peak shifts to the amide III' region near 960 cm^{-1}, leaving a much reduced spectral intensity in the 1220-1300 cm^{-1} region. The intensity which remains is due to NH amide groups which have not undergone $^1H \rightarrow ^2H$ exchange, and to various amino acid side-chains which make a contribution to this region of the spectrum (Carey, 1982).

The amount of α–helix, ß-sheet and random coil structure in hydrated solid toxin was estimated using the method of Lippert *et al.* (1976). The relative spectral intensities at 1240 cm^{-1} in H_2O and at 1632 and 1660 cm^{-1} in ^2H_2O were obtained from Fig. 3.3. Solving the four simultaneous equations relating conformational content and these intensities yielded an estimate of 20% α–helix, 35% ß-sheet and 45% random coil, with the estimate for ß-turn divided between the helix and random coil terms (Carey, 1982).

3.3.2 Infrared Spectroscopy

The presence of both α–helix and ß-sheet structures in toxin was corroborated by Fourier-transform infrared spectroscopy. The conformationally-sensitive amide I region

from toxin dissolved in ^2H_2O is shown in Fig. 3.4(A). The spectrum shows a maximum at 1644 cm^{-1} and a shoulder at approximately 1630 cm^{-1}. As in other proteins, the observed amide I band contour is a composite of overlapping components which represent different secondary structure elements. These components can be better resolved by Fourier

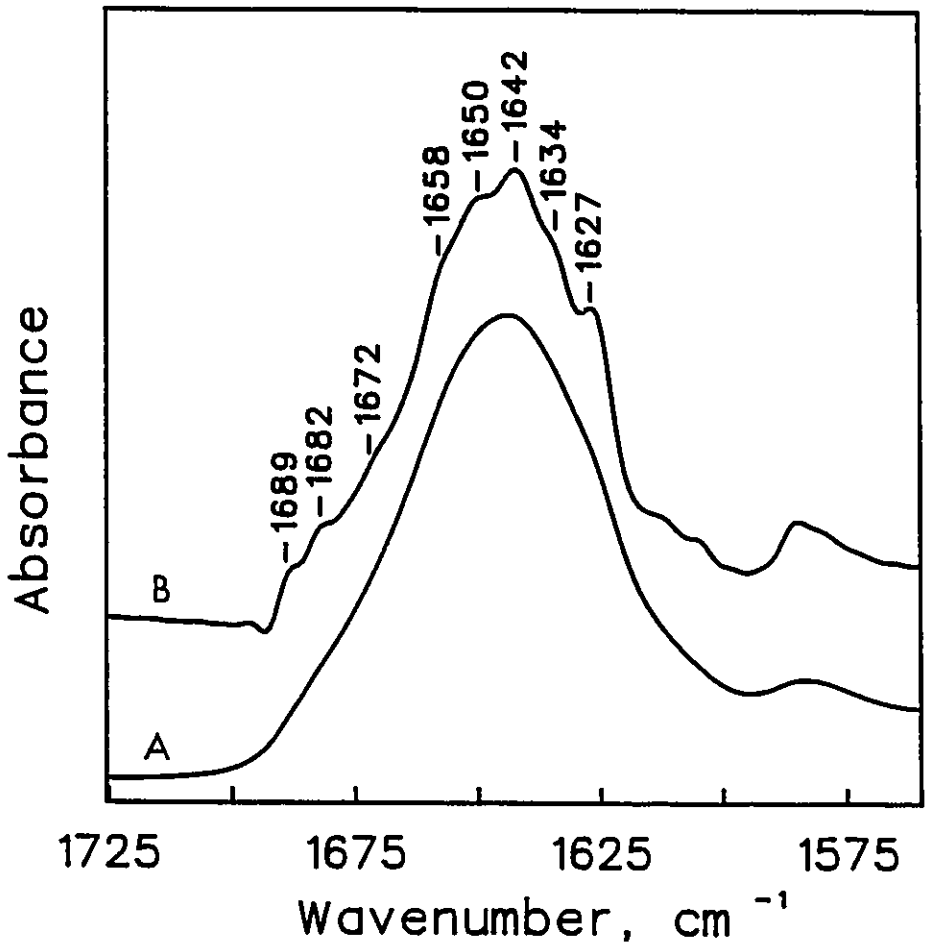


Figure 3.4. Infrared spectrum of the amide I band of toxin solubilized in ^2H_2O, p^{2}H 10.5.
The spectrum is shown (A) before and (B) after band narrowing by Fourier self-deconvolution. A Lorentzian line with a half-width of 14 cm^{-1} and a resolution enhancement factor of 2.3 (Kauppinen *et al.*, 1981) was used to achieve the resolution enhancement.

self-deconvolution analysis of the spectrum (Surewicz and Mantsch, 1988). The deconvolved spectrum of solubilized toxin, shown in Fig. 3.4(B), reveals eight component bands in the amide I region. Based on previously established correlations between spectral features and protein secondary structure (Byler and Susi, 1986), the bands at 1627 and 1634 cm^{-1} can be assigned to ß-sheet structures. The presence of more than one 'ß-component' in the spectral region between 1620 and 1640 cm^{-1} likely represents ß-type segments with slightly different patterns of hydrogen bonding. However, the assignment of these different bands to specific classes of ß-structure, such as parallel or antiparallel sheet, is presently not possible (Surewicz and Mantsch, 1988).

The band at 1642 cm^{-1} in Figure 3.4(B) represents largely random coil segments in which C=O and NH moieties are hydrogen-bonded to the solvent. Bands at 1650 and 1658 cm^{-1} are in the spectral region which is highly characteristic of amide groups in α–helices. The remaining component bands at 1672 and 1689 cm^{-1} are believed to represent turns, although one of these, most likely the band at 1672 cm^{-1}, may also originate from the 'in phase' vibrations of amide groups involved in ß-sheet structures (Surewicz and Mantsch, 1988).

Using the frequencies of these component bands as input parameters, curve-fitting analysis was conducted on the spectrum in Fig. 3.4(A). The fitted bands and their assignments are shown in Fig. 3.5. The fractional areas of the fitted component bands assigned to each secondary structure were added and expressed as a fraction of the total amide I band area. This analysis indicated that the two bands assigned to α–helices (Fig. 3.5) account for approximately 40% of the total area of the amide I band contour. Bands arising from ß-sheet structures account for 22-32% of the area, depending on whether the 1672 cm^{-1} band is assigned to ß-sheet or to turn structures. The estimate for turns ranges from 8-18%, depending on the above assignment, and the random coil structure constitutes approximately 20% of the total amide I band area.

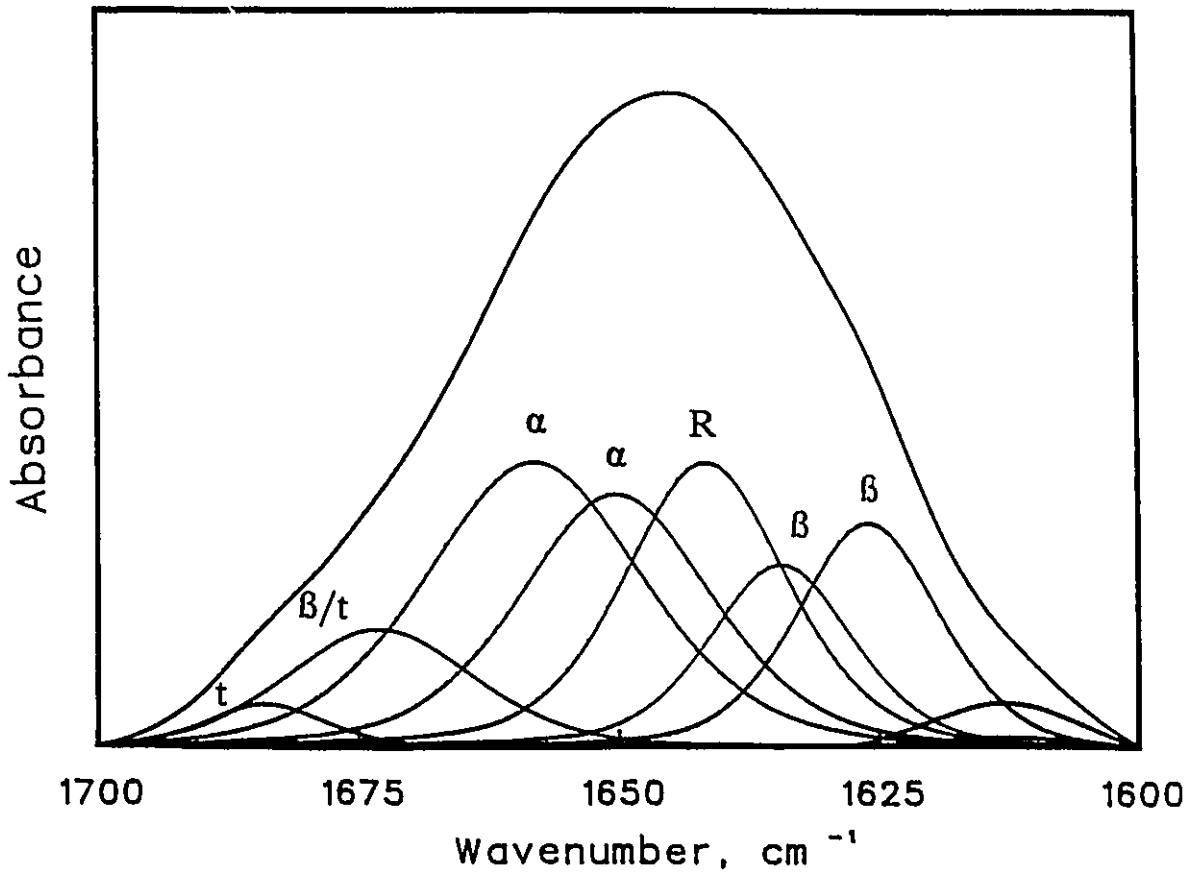


Figure 3.5. Band-fitting of the IR spectrum of solubilized toxin.
The frequencies of the component bands identified in the Fourier-deconvolved spectrum (Fig 3.4) were used as input parameters for curve-fitting the original amide I band. The best fit was obtained by iterative adjustment of the heights and half-widths of the component bands. Assignments for the bands are taken from the literature (Surewicz and Mantsch, 1988).

An advantage of infrared spectroscopy over the other techniques employed during the course of this study is that spectra can be measured both on solubilized toxin (<10 mg/ml) under alkaline conditions, and on hydrated pellets under neutral conditions. When solubilized toxin was precipitated from CAPS/^2H_2O by the addition of HCl, the spectrum

obtained was very similar to that from the solubilized protein (Fig. 3.6) but was shifted to slightly larger wavenumbers. Deconvolution and band fitting analysis of this contour yielded secondary structure estimates which were essentially identical to that obtained for the solubilized protein.

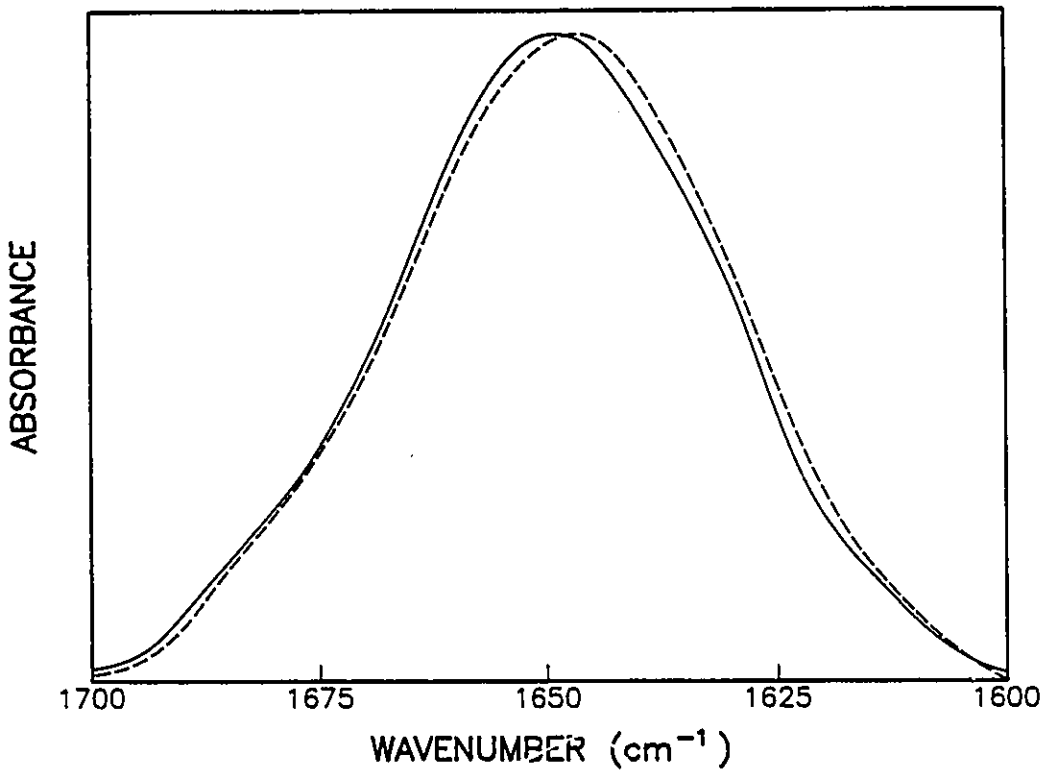


Figure 3.6. IR spectra of the amide I band of solubilized and solid toxin. Toxin solubilized in ^2H_2O, p^{2}H 10.5, was used to obtain the solution IR spectrum (dashed line). The sample was then neutralized and the spectrum of the precipitated toxin was measured (solid line).

3.3.3 Circular Dichroism

Spectra of toxin were obtained on two spectropolarimeter models and the data were analyzed for secondary structure content using two independent fitting programmes. Once normalized for differences in concentration, the spectra obtained from the two instruments were virtually superimposable. Analysis of the data (Fig. 3.7) by the CONTIN programme estimated the content of α–helix, ß-sheet and random coil to be 33%, 32% and 35%, respectively, with ß-turns included in the random coil term (Chen *et al.*, 1972). The procedure of Chang *et al.* (1978) gave an essentially identical estimate of 33% α–helix, 33% ß-sheet, 11% turns and 23% random coil structure.

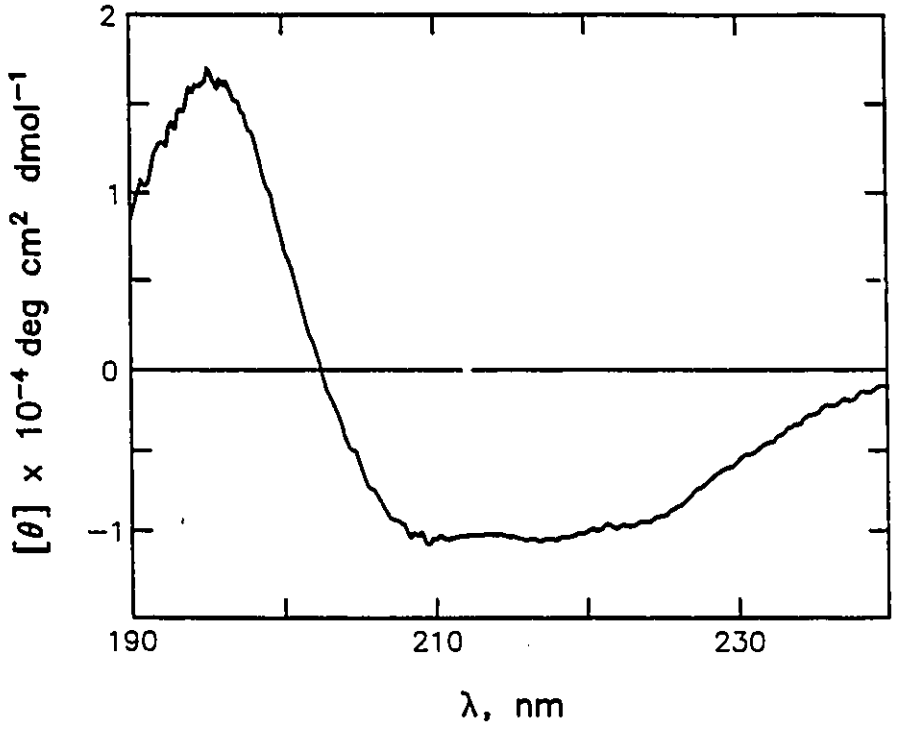


Figure 3.7. Far ultra-violet circular dichroism spectrum of toxin.
The protein, solubilized at pH 10.5 in CAPS buffer, was scanned at 10 nm/minute. Eight spectra were averaged and corrected for buffer absorption.

3.3.4 Predictive Methods

The prime advantage of predictive methods over experimentally-derived estimates of secondary structure is that the predicted structural elements can be located in the amino acid sequence. Using *PlotStructure*, the positions of α–helices and ß-sheets in toxin predicted by the methods of Chou and Fasman (1978) and Garnier *et al.* (1978) were identified (Fig. 3.8). Residues not assigned to either of these structures were classified as random coil. The strength of the Chou-Fasman assignment of individual residues to α–helix or ß-sheet conformation is depicted graphically in Fig. 3.9; also shown is a complimentary display of residues which form or break α–helices or ß-sheets. In addition,

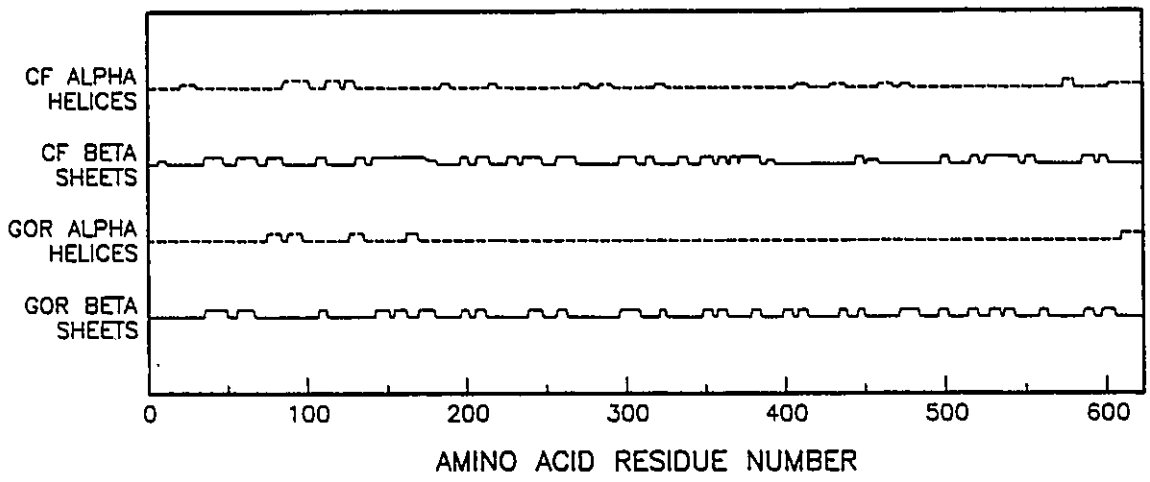


Figure 3.8. Location of α–helices and ß-sheets in the toxin polypeptide chain.
The position and relative length of each structural element was estimated using the methods of Chou-Fasman (CF) and Garnier-Osguthorpe-Robson (GOR). The strength of the prediction (strong or weak) of an individual element can be judged from the height of the rise representing that element. The following criteria were used in appointing a structural element: (i) at least 4 out of 7 adjacent residues must be predicted to exist in the same conformational state; (ii) a minimum of 4 residues is required for an α–helix and (iii) a minimum length of 3 residues is needed for a ß-sheet structure to be predicted. The predicted positions of random coil structures is not shown.

a representation of the charge or hydropathic property of each amino acid in the toxin sequence is displayed in Fig. 3.9, together with a hydropathy measure of the sequence as determined by the method of Kyte and Doolittle (1982).

Table 3.2 summarizes the secondary structure estimates for toxin obtained by experimental techniques, and compares these results with the predictions obtained by the methods of Chou and Fasman (1978) and Garnier *et al.* (1978). Given the limited accuracy of prediction programmes and the problems associated with their implementation and interpretation (Argos and MahanaRao, 1986), there is overall good agreement between secondary structure estimates obtained by spectroscopic means and by predictive algorithms.

TABLE 3.2 Secondary structure content of toxin as estimated by spectroscopic and predictive methods.

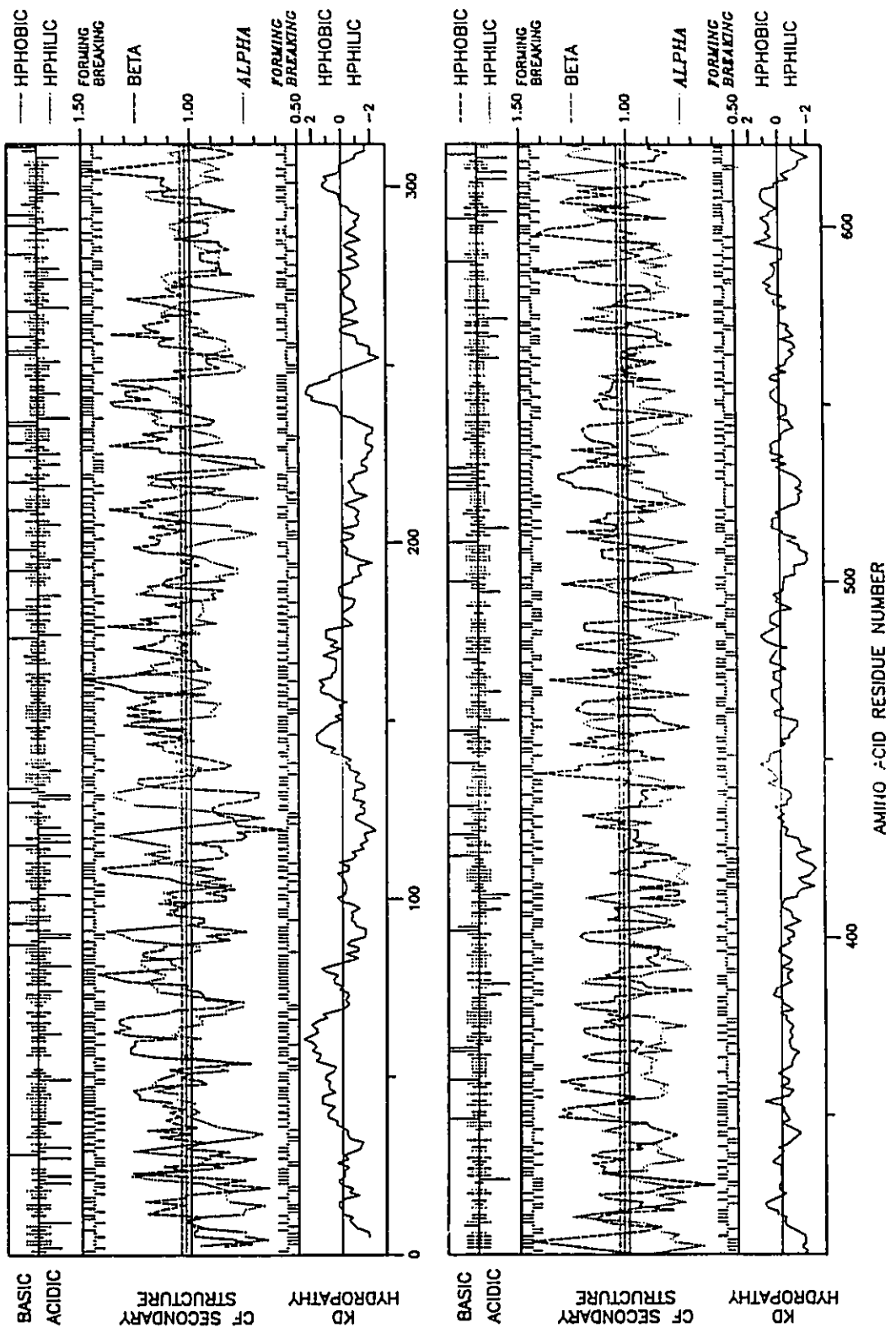
	Raman	Infrared	Circular dichroism[a]		Predictive methods[a]	
			PG	CWY	CF	GOR
% α–helix	20	40	33	33	23	12
% ß-sheet	35	22-32	32	33	47	41
% turn	-[b]	8-18	-[c]	11	-	-
% random coil	45	20	35	23	30	47

a: PG: Provencher and Glockner, (1981); CWY: Chang *et al.*, (1978); CF: Chou and Fasman, (1979); GOR: Garnier *et al.*, (1978)
b: Estimate for turn divided between α–helix and unordered terms.
c: Estimate for turns included in unordered structure term.

Figure 3.9. Secondary structure and hydropathy predictions for toxin. The gene-deduced toxin sequence was analyzed by the *PepPlot* programme. The N-terminal half of the sequence is depicted in the upper plot, and the C-terminal half is portrayed in the lower plot. Each plot is divided into three parts. In the first part, each residue in the sequence is represented by a line at the position where it occurs in the polypeptide chain. Hydrophilic charged residues are indicated by solid lines; those which rise above the baseline are basic, and those which fall below the baseline are acidic. Hydrophilic, uncharged residues are indicated by dotted lines, and hydrophobic residues are shown as dashed lines.

The second part of the plot shows the Chou-Fasman (CF) propensity measures for α–helices (dotted lines) and ß-sheets (dashed lines). Above the curves is a display of the residues which are ß-sheet forming or breaking as defined by Chou and Fasman (1978). Below the curves is a complimentary display of residues which are α–helix forming or breaking. As the curve rises above the threshold for its line type (dotted or dashed), it satisfies one criterion for the propagation of an α–helix or ß-sheet structure. To nucleate an α–helix, there must be four or more α–forming residues and not more than one breaking residue within a span of six residues. For a ß-sheet to nucleate, there should be at least three ß-forming residues and not more than one breaking residue within a window of five. If the curves for α–helix or ß-sheet propagation drop below the solid line threshold, and if there is at least one breaking residue in four, then the structure may terminate.

The third part of the plot displays the Kyte and Doolittle (KD) (1982) hydropathy measure. The curve is the average of a residue-specific hydropathy index over a window of nine residues. When the line is in the upper half of the frame it indicates a hydrophobic region, and when it is in the lower half, a hydrophilic region is indicated.



Chapter 3. Secondary structure of Bt toxin **Fig. 3.9**

3.4 DISCUSSION

The consensus reached from the experimental and predictive methods is that toxin contains roughly equal amounts (approximately 30-35%) of α–helix and ß-sheet structure. The amount of ß-turn and random coil could only be differentiated by two of the techniques, but both these estimates were similar, suggesting approximately 15% turn and 20% random coil content. The overall amount of secondary structure in toxin is therefore not unusual and suggests that the structural basis for the entomocidal action of toxin resides in either the distribution of the secondary structure elements, or else at a higher level in the structural hierarchy, perhaps in the interactions of secondary structure units which give rise to the tertiary structure of the protein.

As there have been no previous experimental investigations into the secondary structure of toxin, a comparison of the present data to results reported in the literature must necessarily focus on the findings from prediction studies. From Fig. 3.8 it is evident that the methods of Chou and Fasman (1978) and Garnier *et al.* (1978) both predict that toxin is composed primarily of ß-sheet structure. The Chou-Fasman algorithm suggests that the helices are dispersed throughout the length of the toxin sequence, with the strongest predictions occurring between residues 90-130. Analysis by the method of Garnier *et al.* indicates that, except for a helix at the extreme C-terminus of the molecule, the α–helix content is contained exclusively between amino acid residues 80-180. The suggestion of helices in this region by two independent methods adds weight to the prediction; in addition, it is interesting to note that many of the ß-sheet elements are predicted by both programmes to be in identical locations.

The region encompassed by the first 200 amino acids, and particularly the segment between residues 30-80, is predicted to be strongly hydrophobic (Fig. 3.9). A similar observation has been made for the corresponding region of the sequence deduced from the *cryIA(a)* gene (Schnepf *et al.*, 1985). These authors compared the hydropathy of the most extensive hydrophobic region in their sequence (residues 31-66) to the hydropathies of the

transmembrane domains of bacteriorhodopsin and of the B fragment of diphtheria toxin. They concluded that residues 31-66 could constitute a transmembrane segment, even though this region was predicted to be in the ß-sheet conformation. On the basis of alignment studies and hydropathy plots, Höfte and Whiteley (1989) have suggested that the proteins encoded by each of the *cryIA* genes contain a transmembrane sequence in the first 100 amino acid residues. A similar suggestion was made following analysis of several Lepidopteran-specific protoxin genes by Hefford *et al.* (1987), and by Thorne *et al.* (1986) for the Diptera-specific protoxin gene from *B. thuringiensis* subsp. *israelensis*. The conservation amongst the three cryIA gene products of the predominantly hydrophobic sequence spanning residues 36-68 (Schnepf *et al.*, 1985) indicates that these amino acids may play a critical role in the function of the protein. However, there is presently no other experimental evidence in the literature to support the transmembrane sequence hypothesis.

The amount of predicted helical content in the toxin (especially by the method of Garnier *et al.*, 1978) is somewhat lower than that measured spectroscopically (Table 3.2). As pointed out earlier in this chapter, estimates of secondary structure content determined experimentally are based on a number of approximations and assumptions. However, all the experimentally-determined estimates for the secondary structure content of toxin are in good agreement, indicating that the predictive methods underestimate the amount of α–helix in this protein.

A comparison of Raman spectral data obtained from toxin with data found in the literature for protoxin (Carey *et al.*, 1986) indicates that the band positions and intensities from hydrated pellets of both proteins are very similar (columns 1 and 3, Table 3.1). This comparison suggests that the spectra from the two proteins are essentially superimposable. In particular, the amide I (1664 cm^{-1}) and amide III (1246 cm^{-1}) band positions and intensities for aqueous pellets of protoxin and toxin are almost identical. Since these bands reflect protein conformation, their similarity suggests that the overall secondary structure content of the two proteins do not differ greatly. Data from a deuterated *B. thuringiensis*

subsp. *kurstaki* HD-73 protoxin pellet is not available from the literature, so it is not possible to calculate the secondary structure content of this protoxin. However, Carey *et al.* (1986) estimated that the crystal protein from *B. thuringiensis* subsp. *kurstaki* HD-1 has a secondary structure content of 25% α–helix, 21% ß-sheet and 54% random coil. Unfortunately, this subspecies of *B. thuringiensis* carries three different protoxin genes (Kronstad and Whiteley, 1986) so it is not possible to determine what proportion of each spectral contribution arises from each of the gene products. Nonetheless, this estimate of secondary structure content is similar to that calculated in the present study for toxin. This is a somewhat surprising result given that the C-terminal half of the protoxin molecule contains 14 cysteine residues which are engaged in interchain disulphide bonds (Bietlot *et al.*, 1990), whereas toxin contains no cysteine residues. Furthermore, toxin and the C-terminal half of protoxin are composed of amino acids with decidedly different hydropathic properties. This is evident from Fig. 3.10, which shows the Kyte-Doolittle hydropathy plot for protoxin. It is clear that the N-terminal half of the molecule from which the toxin is derived is hydrophobic, whereas the C-terminal half is largely hydrophilic. The distinctly different physicochemical characters of the two halves of the protoxin molecule contribute to the interesting tertiary structure of this protein (Chapters 4 and 5).

As with any technique, there are uncertainties inherent in the estimate of secondary structure content by Raman spectroscopy. This is due to ambiguity in the position of the spectral baselines (and hence peak heights), and also to the difficulty of averaging irregularities in the individual, non-computerized spectra. Nonetheless, Raman spectroscopy indicates that the relative amounts of α–helix, ß-sheet and random coil content in protoxin and toxin are similar, although the technique provides no information on the distributions of these elements in the molecules. Based on a Chou-Fasman prediction analysis of the sequence deduced from the *cryIA(a)* gene, Schnepf *et al.* (1985) reported that the N-terminal half of the protoxin molecule is primarily in the ß-sheet conformation, while the C-terminal half contains a larger amount of α–helix structure.

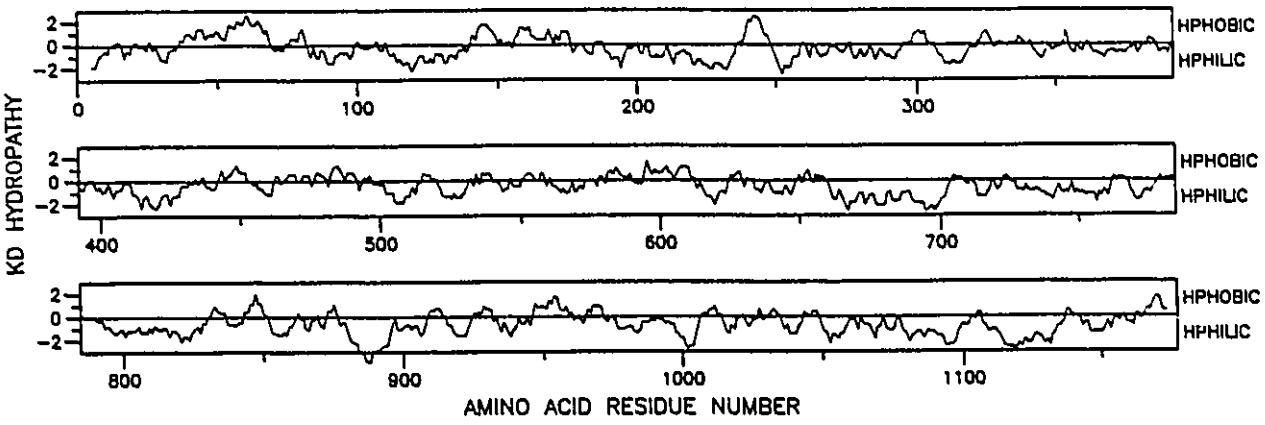


Figure 3.10. Hydropathy plot for protoxin.
The curve is the average of a residue-specific hydropathy index (Kyte and Doolittle, 1982) over a window of nine residues. When the line is in the upper half of the frame it indicates a hydrophobic region, and when it is in the lower half, a hydrophilic region is indicated.

An advantage of Raman spectroscopy over infrared and CD is the insight the spectra provide into the overall environment experienced by certain classes of amino acids within a protein. The relative peak intensity of the tyrosine doublet at 830 and 850 cm^{-1} is an indication of the average environment experienced by the tyrosine side-chains and the degree of hydrogen bonding at the tyrosine -OH group (Tu, 1986; Harada and Takeuchi, 1986). A ratio of approximately one suggests that the majority of the tyrosine residues are involved in weak-to-medium hydrogen bonds. Ratios previously reported (Carey *et al.*, 1986; Pozsgay *et al.*, 1987) for hydrated crystal pellets from *B. thuringiensis* subsp. *kurstaki* HD-1 and NRD-12 indicate that the average environment experienced by the approximately 60 tyrosine side-chains involve fairly weak hydrogen donor and/or acceptor bonds to water and other residues. A similar situation was observed in the present study for hydrated pellets of *B. thuringiensis* subsp. *kurstaki* HD-73 toxin. The ratio of I_{850}/I_{830} (column 3, Table 3.1) for the aqueous pellet is approximately 0.85. This

suggests that the majority of the 26 tyrosines in toxin are weakly hydrogen-bonded to water and/or other residues. In air-dried toxin (column 5, Table 3.1) the ratio increases only marginally to 1.2, indicating that the tyrosines are still predominantly involved in hydrogen bonding.

Vibrations of the tryptophan indole ring give rise to characteristic Raman bands at 760, 880, 1360, 1555, and 1585 cm^{-1} (Lord and Yu, 1970). All these features are observed in the aqueous toxin pellet spectrum and are identified in Table 3.1. The intensity of the 1360 cm^{-1} band is particularly influenced by the environment (Tu, 1986). This peak is therefore a useful indicator of whether most tryptophan side-chains are exposed or buried, since a sharp, intense band indicates predominantly buried side-chains. The 1360 cm^{-1} feature appears as a shoulder in all three spectra of toxin (Fig. 3.2 and 3.3), indicating that the majority of the ten tryptophan residues in toxin are not buried in a hydrophobic environment and may be accessible to water. Given that these highly hydrophobic tryptophans are found in a hydrophobic protein (Fig. 3.10), it is rather surprising that these residues are not buried. That they are not buried was corroborated in an independent manner during the course of the present study using fluorescence emission spectroscopy. Unfortunately, many difficulties are encountered in interpreting fluorescence data obtained from multi-tryptophan proteins (Weber, 1972). Therefore, a study of the accessibility of the tryptophans in toxin to the external environment using the double-quenching technique of Somogyi *et al.* (1985) was terminated after several preliminary experiments. Due to the incomplete nature of the results, they will not be presented in this thesis. However, the data clearly showed that at least 50% of the fluorescence intensity of toxin could be selectively quenched by ionic quenching agents, indicating that the tryptophans giving rise to this fluorescence are located at the surface of the protein.

3.5 CONCLUSIONS

The spectroscopic data reported in the present study is the first experimental information on the secondary structure of any *B. thuringiensis* toxin. As discussed recently (Dev, 1987; Surewicz and Mantsch, 1988; Schultz, 1988), experimental estimates and predictions of secondary structure content are based on a number of approximations and assumptions, and the data obtained from any one physical technique must be interpreted with caution. However, when estimates from several techniques are in agreement, a much greater degree of confidence can be attached to these results. In the present study, all the experimental approaches used lead to a consensus that toxin, either as a solid or in solution, is a highly folded protein with approximately 70% of the amino acids involved in forming α–helices and ß-sheet structures. A highly folded conformation is in agreement with results presented in the following chapter, where evidence is presented that the tertiary structure of toxin is composed of at least three domains.

3.6 REFERENCES

Andrews, R. E. Jr., R. M. Faust, H. Wabiko, K. C. Raymond and L. A. Bulla (1987). The biotechnology of *Bacillus thuringiensis*. *CRC Crit. Rev. Biotechnol.* **6**, 163-230.

Anfinsen, C. B. and E. Haber (1961). Studies on the reduction and re-formation of protein disulphide bonds. *J. Biol. Chem.* **236**, 1361-1363.

Argos, P. and J. K. MohanaRao (1986). Predictions of protein structure. *Methods Enzymol.* **130**, 185-207.

Astbury, W. T. (1938). X-ray adventures among the proteins. *Trans. Faraday Soc.* **34**, 378-388.

Bietlot, H., I. Vishnubhatla, P. R. Carey, M. Pozsgay and H. Kaplan (1990). Characterization of the cysteine residues and disulfide linkages in the protein crystal of *Bacillus thuringiensis* subsp. *kurstaki* and *entomocidus*. *Biochem. J.* **267**, 309-315.

Brousseau, R. and L. Masson (1988). *Bacillus thuringiensis* insecticidal crystal toxins: gene structure and mode of action. *Biotech. Adv.* **6**, 697-724.

Byler, D. M. and H. Susi (1986). Estimation of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25**, 469-487.

Carey, P. R. (1982). *Biochemical Applications of Raman and Resonance Raman Spectroscopies*, Academic Press, New York.

Carey, P. R., P. Fast, H. Kaplan and M. Pozsgay (1986). Molecular structure of the protein crystal from *Bacillus thuringiensis*: a Raman spectroscopic study. *Biochim. Biophys. Acta* **872**, 169-176.

Chang, C. T., C. S. Wu and J. T. Yang (1978). Circular dichroic analysis of protein conformation: inclusion of the ß-turns. *Anal. Biochem.* **91**, 13-31.

Chen, M. C. and R. C. Lord (1974). Laser-excited Raman spectroscopy of biomolecules. IV. Some polypeptides as conformational models. *J. Am. Chem. Soc.* **96**, 4750-4752.

Chen, G. C. and J. T. Yang (1977). Two point calibration of circular dichrometer with d-10-camphorsulphonic acid. *Anal. Lett.* **10**, 1195-1207.

Chen, Y. H., J. T. Yang and H. M. Martinez (1972). Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* **11**, 4120-4131.

Chou, P. Y. and G. D. Fasman (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* **47**, 45-148.

Crighton, T. E. (1984). *Proteins*. W. H. Freeman and Co., New York, pp. 221-264.

Dev, S. B. (1987). Quantitative prediction of protein secondary structure - where is the Lacuna? *J. Biol. Phys.* **15**, 57-61.

Fraser, R. D. and E. Suzuki (1966). Resolution of overlapping absorption bands by least squares procedures. *Anal. Chem.* **38**, 1770-1773.

Garnier, J., D. J. Osguthorpe and B. Robson (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **120**, 97-120.

Greenfield, N. and G. Fasman (1969). Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108-4115.

Harada, I. and H. Takeuchi (1986). Raman and ultraviolet resonance Raman spectra of proteins and related compounds. (in) *Spectroscopy of Biological Systems* (R. J. Clark and R. E. Hester, eds.) John Wiley and Sons, New York, pp. 113-175.

Hefford, M. A., R. Brousseau, G. Prefontaine, Z. Hanna, J. A. Condie and P. C. Lau (1987). Sequence of a lepidopteran toxin gene of *Bacillus thuringiensis* subsp. *kurstaki* NRD-12. *J. Biotech.* **6**, 307-322.

Höfte, H. and H. R. Whiteley (1989). Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Jaenicke, R. (1987). Folding and association of proteins. *Prog. Biophys. Molec. Biol.* **49**, 117-237.

Johnson, W. C. (1988). Secondary structure of proteins through circular dichroism spectroscopy. *Ann. Rev. Biohys. Biophys. Chem.* **17**, 145-166.

Kabsch, W. and C. Sander (1983). How good are predictions of protein secondary structure? *FEBS Lett.* **155**, 179-182.

Kauppinen, J. K., D. J. Moffatt, H. H. Mantsch and D. C. Cameron (1981). Fourier self-deconvolution: a method for resolving intrinsically overlapped bands. *Appl. Spectrosc.* **35**, 271-277.

Krimm, S (1962). Infrared spectra and chain conformation of proteins. *J. Mol. Biol.* **4**, 528-540.

Kronstad, J. W. and H. R. Whiteley (1986). Three classes of homologous *Bacillus thuringiensis* crystal protein genes. *Gene* **43**, 29-40.

Kyte, J. and R. F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Linderström-Lang, K. U. and J. A. Schellman (1959). Protein structure and enzyme activity. (in) *The Enzymes*, vol. I, second edition (P. D. Boyer, ed.) Academic Press, New York, pp. 443-510.

Lippert, J. L., D. Tyminski and P. J. Desmeules (1976). Determination of the secondary structure of proteins by laser Raman spectroscopy. *J. Am. Chem. Soc.* **98**, 7075-7080.

Lord, R. C. and N. T. Yu (1970). Laser-enhanced Raman spectroscopy of biomolecules. I. Native lysozyme and its constituent amino acids. *J. Mol. Biol.* **50**, 509-524.

Nagamatsu, Y., Y. Itai, C. Hatanaka, G. Funatsu and K. Hayashi (1984). A toxic fragment from the entomocidal crystal protein of *Bacillus thuringiensis*. *Agric. Biol. Chem.* **48**, 611-619.

Nishikawa, K. (1983). Assessment of secondary structure prediction of proteins: comparison of computerized Chou-Fasman method with others. *Biochim. Biophys. Acta* **748**, 285-299.

Parker, F. S. (1983). *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*, Chapter 3, Plenum Press, New York, pp. 83-155.

Pauling, L., R. B. Corey and H. R. Branson (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205-211.

Pozsgay, M., P. Fast, H. Kaplan and P. R. Carey (1987). The effect of sunlight on the protein crystals from *Bacillus thuringiensis* var. *kurstaki* HD1 and NRD12: a Raman spectroscopic study. *J. Invertebr. Pathol.* **50**, 246-253.

Provencher, S. W. and J. Glöckner (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33-37.

Schnepf, H. E., H. C. Wong and H. R. Whiteley (1985). The amino acid sequence of a crystal protein from *Bacillus thuringiensis* deduced from the DNA base sequence. *J. Biol. Chem.* **260**, 6264-6272.

Schulz, G. E. (1988). A critical evaluation of methods for prediction of protein secondary structures. *Ann. Rev. Biophys. Biophys. Chem.* **17**, 1-21.

Shibano, Y., A. Yamagate, N. Nakamura, T. Iizuka, H. Sugisaki and M. Takanami (1985). Nucleotide sequence coding for the insecticidal fragment of the *Bacillus thuringiensis* crystal protein. *Gene* **34**, 243-251.

Somogyi, B., A. Papp, A. Rosenberg, I. Seres, J. Matko, G. Welch and P. Nagy (1985). A double-quenching method for studying protein dynamics: separation of the fluorescence quenching parameters characteristic of solvent-exposed and solvent-masked fluorophors. *Biochemistry* **24**, 6674-6679.

Surewicz, W. K. and H. H. Mantsch (1988). New insight into protein secondary structure from resolution-enhanced infrared spectra. *Biochim. Biophys. Acta* **952**, 115-130.

Susi, H. and D. M. Byler (1983). Protein structure by Fourier transform infrared spectroscopy: second derivative spectra. *Biochem. Biophys. Res. Commun.* **115**, 391-397.

Susi, H. and D. M. Byler (1986). Resolution-enhanced Fourier transform infrared spectroscopy of enzymes. *Methods Enzymol.* **130**, 290-311.

Thomas, G. J. Jr. and B. Prescott (1983). Structure similarity, difference and variability in the filamentous viruses fd, If1, IKe, Pf1 and Xf: investigation by laser Raman spectroscopy. *J. Mol. Biol.* **165**, 321-356.

Thorne, L., F. Garduno, T. Thompson, D. Deckler, M. Zounes, M. Wild, A. Walfield and T. Pollock (1986). Structural similarity between the lepidoptera- and diptera-specific insecticidal endotoxin genes of *Bacillus thuringiensis* subsp. *kurstaki* and *israelensis*. *J. Bacteriol.* **166**, 801-811.

Tu, A. T. (1986). Peptide backbone conformation and microenvironment of protein side chains. (in) *Spectroscopy of Biological Systems* (R. J. Clark and R. E. Hester, eds.), John Wiley and Sons, New York, pp. 47-112.

Van Wart, H. E. and H. A. Scheraga (1978). Raman and resonance Raman spectroscopy. *Methods Enzymol.* **49**, 67-148.

Weber, G. (1972). Use of fluorescence in biophysics: some recent developments. *Ann. Rev. Biophys. Bioeng.* **1**, 553-570.

Williams, R. W. (1986). Protein secondary structure analysis using Raman amide I and amide III spectra. *Methods Enzymol.* **130**, 311-350.

Yang, J. T., C. S. Wu and H. M. Martinez (1986). Calculation of protein conformation from circular dichroism. *Methods Enzymol.* **130**, 208-269.

Yu, N. T. (1977). Raman spectroscopy: A conformational probe in biochemistry. *CRC Crit. Rev. Biochem.* **4**, 229-249.

Chapter 4.

THE DOMAIN STRUCTURE OF TOXIN

On problems

Our choicest plans
have fallen through,
our airiest castles
tumbled over,
because of lines
we neatly drew
and later neatly
stumbled over.

- Piet Hein

Chapter 4.

THE DOMAIN STRUCTURE OF TOXIN

4.1 INTRODUCTION

In the preceding chapter, the secondary structure content of toxin was determined experimentally, and sites of local folding of the polypeptide chain into individual secondary structure elements were predicted. In the hierarchical model of Linderstrom-Lang (1959), the stability of an element higher in the hierarchy depends on the stability and strength of association of elements lower in the hierarchy. Thus, the association of secondary structure elements into ordered, discrete structural and functional entities gives rise to the tertiary-level structure of the protein. The tertiary structures of many proteins consist of two or more tightly folded units called domains. Individual domains within a protein molecule generally interact with each other, but these associations are far less extensive than the interactions between the secondary structure elements which give rise to each domain (Jaenicke, 1987).

The criteria used to classify these structural regions are often rather subjective. Perhaps the most useful definition is that given by Schulz and Schirmer (1979), who classify domains into two categories: functional and structural. Functional domains are regions of the polypeptide chain which are autonomous in the sense that they possess many of the characteristics associated with an intact globular protein. When these domains are excised from the parent protein, they retain their structure and function. On the other hand, structural domains are geometrically separate entities which can generally only be identified upon inspection of a three-dimensional protein structure. These domains usually have molecular masses of 20 kDa or less, while functional domains are often considerably larger. Therefore, a functional domain may consist of one or more structural domains. Structural domains, also sometimes termed 'subdomains', do not remain folded when excised from the parent protein.

The importance of domains is evident in many aspects of protein function, as the active sites of enzymes often reside at domain interfaces, and various substrates or ligands may bind to different domains performing different functions (Jaenicke, 1987). In

addition, since functional domains fold independently of each other (Kirschner and Bisswanger, 1976; Privalov, 1982), domains have been postulated to be major intermediates in the folding pathways of multidomain proteins (Janin and Wodak, 1983). Structurally, domains are usually linked to their neighbours by a single stretch of polypeptide chain (Creighton, 1984). The tightly-folded conformation of a functional domain is relatively resistant to proteolytic enzymes, but the polypeptide segments linking the domains are often flexible and susceptible to cleavage (Ghelis and Yon, 1982). Therefore, the method of limited proteolysis of a protein provides a diagnostic test for the presence of functional domains, and may permit their separation and isolation. This approach was first used to elucidate the domain structure of ß-D-galactosidase from *Escherichia coli* (Goldberg, 1969) and of immunoglobulin G (Edelman *et al.*, 1969), and has subsequently been employed in the identification of the autonomous domains of many other proteins (Kirschner and Bisswanger, 1976). In the present study, the techniques of limited proteolysis, differential scanning calorimetry and infrared spectroscopy were used to characterize the domain structure of toxin.

4.1.1 Differential scanning calorimetry

Differential scanning calorimetry is a versatile technique which can be used to study events of biochemical significance such as metabolic reactions, ligand binding or the unfolding of biological macromolecules (Langerman and Biltonen, 1979). In calorimetry, the amount of heat required to raise the temperature of a system by a specific amount is measured; at constant pressure, this quantity is the heat capacity (C_p) of the system. In biochemical applications, the system consists of an aqueous solution of the macromolecule of interest. As heat is applied to the system, changes occur in the sample causing it to be transformed from one state to another. An example of such a transformation is a protein in its native conformation being transformed by the absorption of heat into an unfolded polypeptide. In order to minimize interactions which could initiate aggregation during the heating process, the concentration of the protein in the sample is kept as low as possible.

The instrument must therefore be capable of resolving the contribution of the protein, which is generally less than 1% of the total solution heat capacity (Privalov and Potekhin, 1986). Thus to study the unfolding of a protein by calorimetry it is essential to employ a differential scheme of measurement so that the small quantity of heat absorbed or released by the protein itself can be detected (Sturtevant, 1972). This is accomplished by placing equal volumes of the dissolved sample and the solvent in two closely matched cells, and continuously measuring the heat capacities of the two cells while they are heated at a constant rate. The raw data output from a differential scanning calorimeter is typically in the form of an *x*-*y* plot, showing the difference in heat capacity between the two cells (*y* axis) with temperature (*x* axis).

Calorimetric data can be used to directly evaluate enthalpy changes in the sample since the heat capacity determined at constant pressure is a temperature derivative of the enthalpy function:

$$C_p = (\partial H/\partial T)_p$$

Integration of the heat capacity over the temperature range of the thermal unfolding of the protein yields the calorimetric enthalpy (ΔH_{cal}) of the transition.

An estimate of enthalpy changes in the sample can be conducted if the reaction is reversible (*i.e.*, if the thermally-unfolded protein refolds after cooling to yield an identical unfolding profile upon reheating in the calorimeter). The enthalpy change for a two state process in which the protein can be considered to be either folded or unfolded, with no significant population of intermediates, is given by the van't Hoff equation:

$$(\partial \ln k/\partial T)_p = \Delta H_{vH}/RT^2$$

where k is the equilibrium constant of the process and ΔH_{vH}, usually referred to as the van't Hoff enthalpy, is the enthalpy change of the process. The change in van't Hoff

enthalpy (ΔH_{vH}) for the unfolding of a protein can be estimated from calorimetric data by the following expression:

$$\Delta H_{vH} = 4RT_m^2(C_{ex(max)}/\Delta H_{cal})$$

where $C_{ex(max)}$ is the maximal value of the heat capacity of the transition and T_m is the temperature at which C_{ex} reaches its maximal value (Takahashi *et al.*, 1981).

Comparison of the calorimetric and van't Hoff enthalpies for the thermal unfolding of a protein provides information on the co-operativity of the unfolding process. The value of $\Delta H_{vH}/\Delta H_{cal}$ for many small proteins is approximately one, indicating that the unfolding of the protein is fully co-operative and closely approximates a two-state transition (Pfeil and Privalov, 1976). For these proteins, the population of all intermediate states between the native and denatured states is very small, and so denaturation can be considered to be an all-or-none process. For many proteins however, unfolding is not a simple two-state process: the ratio of $\Delta H_{vH}/\Delta H_{cal}$ is significantly less than one, indicating that disruption of the native structure gives rise to substantial concentrations of several unfolding intermediates (Privalov, 1982). A ratio of less than one is thus often observed for large, multidomain proteins in which the domains unfold independently of each other (Privalov and Potekhin, 1986). Differential scanning calorimetry therefore provides a means of estimating the number of independent unfolding domains in a protein. The technique was employed in the present study to corroborate the number of domains in toxin observed by limited proteolysis.

4.1.2 Infrared spectroscopy

Infrared spectroscopy can also be used to monitor the thermal unfolding of a protein. The amide I band of the protein is scanned while the sample is heated; as the temperature is increased, the melting of the various secondary structure elements leads to changes in the band contour. This method does not have the resolution required to identify

the independent thermal events leading to the denaturation of individual domains. However, melting curves can reveal gross structural changes resulting from major denaturation processes (Surewicz and Mantsch, 1988). This information can then be correlated with other data obtained independently regarding the protein's domain structure.

4.2 MATERIALS AND METHODS

4.2.1 Materials

B. thuringiensis subsp. *kurstaki* HD-73 toxin was prepared from protoxin crystals and characterized as described in Sections 2.2.3 and 2.2.4. Deuterium oxide was obtained from Aldrich Chemical Co., Milwaukee, Wisconsin. Precast gels and buffer strips for the Pharmacia Phast gel electrophoresis system were purchased from Pharmacia Canada Ltd., Dorval, Quebec. TPCK-trypsin, elastase and papain were obtained from Sigma Chemical Corp., St. Louis, Missouri, as were the papain inhibitor E64, ultra-pure urea and CAPS buffer salt. All other chemicals were high purity preparations obtained from commercial sources. Solutions were prepared using reverse-osmosis quality water.

4.2.2 Limited proteolysis of toxin

4.2.2.1 Papain:

Toxin (0.1-1 mg) was dissolved in 200 µl of 0.1M CAPS buffer, pH 10.5. Papain (0.02-1 mg) was activated by the addition of 0.1% ß-mercaptoethanol, and was then added to the toxin. The digest was incubated at room temperature for between 1-30 minutes, and was then added to boiling SDS sample buffer (8M urea, 2.5% SDS, 5% ß-mercaptoethanol and 10mM Tris-HCl, adjusted to pH 8.3). Residual papain activity in the boiling SDS sample buffer could be inhibited by adding a hundredfold excess (inhibitor:papain) of E64 to the digest prior to boiling.

4.2.2.2 Trypsin and Elastase:

Toxin (0.1-1.0 mg) was dissolved at 20°C in 200 µl of 8M urea buffered at pH 9.5 with 0.1M CAPS. After one hour, trypsin or elastase (0.05-3 mg) was added and the digest was incubated at 20°C for 5-15 seconds. The digest was then added to SDS sample buffer at room temperature and boiled.

4.2.3 Gel electrophoresis

Polyacrylamide SDS-gels (10-15% gradient) were run on a Pharmacia Phast electrophoresis system and were stained with Coomassie blue as described in Section 2.2.4.1. The molecular masses of the toxin fragments were estimated using a standard curve of log(molecular mass) *vs.* relative mobility, determined with molecular mass markers. The markers used were: phosphorylase (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20 kDa) and α–lactalbumin (14 kDa).

4.2.4 N-terminal sequence analysis

Toxin fragments from SDS-gels were electroblotted onto polyvinylidene difluoride membranes and the first 20 amino acids were sequenced using a model 470A Applied Biosystems gas-phase sequencer, as described in Section 2.2.4.5.

4.2.5 Differential scanning calorimetry

Calorimetry was performed using a Microcal MC-2 scanning microcalorimeter interfaced with an IBM personal computer. Toxin was dissolved to a concentration of approximately 5 mg/ml in 0.1M CAPS buffer, pH 10.5 containing 0.15M NaCl. The presence of salt in the sample helped overcome protein aggregation and precipitation at elevated temperatures. After degassing under vacuum, 1.2 ml of toxin sample, and an equal volume of 0.1M CAPS/0.15M NaCl buffer, were loaded into the sample and reference cells, respectively. The cells were pressurized with N_2 to two atmospheres to reduce the formation of gas bubbles within the cells during heating. The cells were

equilibrated at 15°C for 30 minutes, and then the temperature of the two compartments was increased at a constant rate of 1°C/min. The differential heat capacity between the sample and reference cells was measured continuously during the heating process. Scans were typically terminated at 95°C. Following each experiment, the sample and reference cells were each loaded with 0.1M CAPS/0.15M NaCl buffer, pH 10.5, and the equilibration and scanning process was repeated in order to obtain a baseline. The concentration of protein in each sample was determined by amino acid analysis as described in Section 2.2.4.3.

Each data set was normalized for the scan rate and number of moles of toxin in the sample, thus converting the data into units of calorie/degree/mole. The data set was then baseline-corrected by either (i): subtracting the stored buffer-buffer baseline or (ii): calculating the baseline using the splines subroutine in the data analysis package supplied by Microcal Inc., Amherst, Massachusetts. In this method, a reference temperature is chosen directly below the temperature where the protein denaturation transition begins, and another reference is selected immediately above the end of the transition. The data is then processed to determine the smooth baseline which best describes all data points except those lying between the two reference temperatures.

Although simple calorimetric curves can be analyzed for enthalpies by direct evaluation of the area under the unfolding transition, the calorimetric and van't Hoff enthalpies from complex denaturation curves are obtained by fitting the experimental data to a theoretical curve. The calorimetric data from toxin was subjected to curve-fitting analysis based on non-linear least squares minimization, using the fitting subroutine in the Microcal data analysis package. It was assumed that ΔH_{vH} and ΔH_{cal} are not equal, as this permitted all initialization parameters to be chosen manually. The number and midpoint temperatures of the minimal number of transitions necessary to describe the calorimetric curve were chosen and subtracted from the experimental curve. Their positions and shapes were optimized by iterative minimization until their sum approximated the shape and area of the experimental curve.

4.2.6 Infrared spectroscopy

Toxin was solubilized in CAPS/^2H_2O buffer and the amide I band was scanned with a Digilab FTS-60 instrument as described in Section 3.3.3. A circulating thermostated water bath was used to raise the temperature of the sample compartment at a rate of 0.25^oC/minute. The sample was scanned at approximately 5^oC intervals.

4.3 RESULTS

4.3.1 Limited proteolysis

The fragmentation of toxin by papain was first observed during the initial characterization of this protein (Chapter 2) when the protease stability of the toxin was determined. In that study, a given protease was incubated with toxin for either five minutes or one hour and the sample was then analyzed for digestion products by SDS/PAGE. It was found that if the papain inhibitor E64 was added to toxin/papain mixtures prior to boiling in the sample buffer, the toxin was not proteolyzed (Section 2.2.4.6). If E64 was not present when the sample was boiled, proteolysis did occur. Under these conditions, it was observed that the quantity of proteolysis products present on SDS gels was not dependent on the length of incubation of papain with the toxin at room temperature (Fig. 4.1). However, the degree of proteolysis was affected by the amount of papain present. While the addition of very large amounts of papain (toxin:papain; 1:10 by mass) resulted in extensive proteolysis of the toxin into small peptides in the boiling sample buffer, the presence of smaller amounts of papain (toxin:papain; 5:1 by mass) resulted in the generation of only two fragments. These fragments had apparent molecular masses of approximately 36 and 33 kDa (Fig. 4.1). If roughly equal amounts of papain and toxin were present in the sample, it was observed that the intensity of the first (36 kDa) band decreased, and a third band with an apparent molecular mass of 26 kDa appeared on the gel (Fig. 4.1).

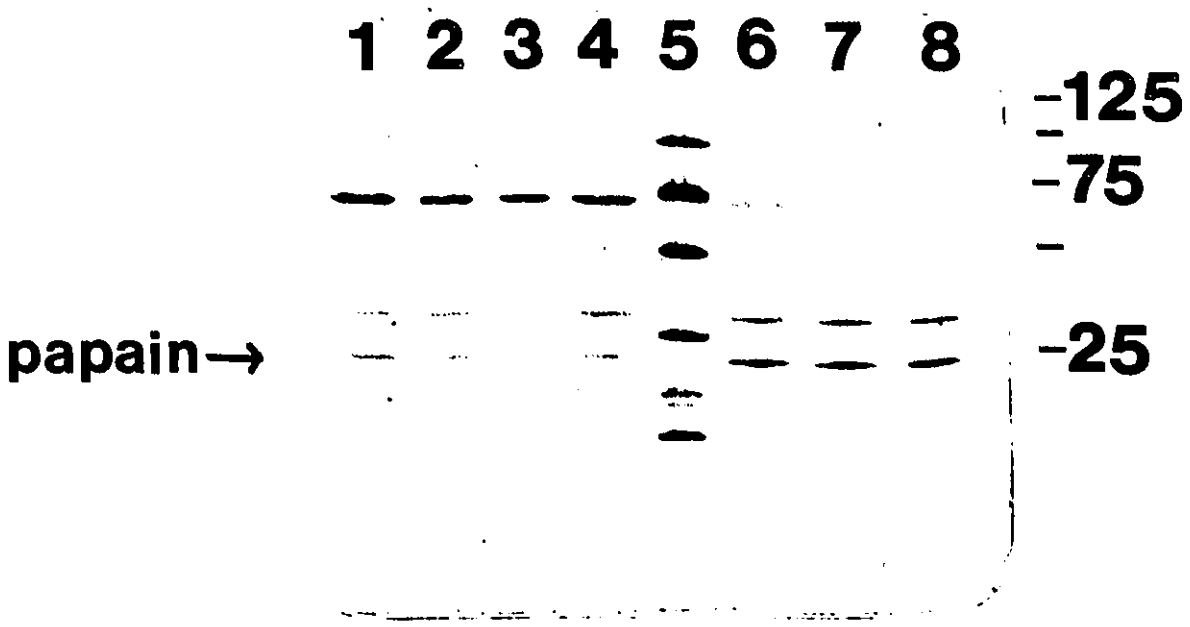


Figure 4.1. Limited proteolysis of toxin by papain under denaturing conditions.
Proteolysis was carried out as described in Section 4.2.2 with the following protein ratios and incubation times: Lanes 1-4, toxin:papain 5:1 by mass; lane 1, 1 min.; lane 2, 5 min.; lane 3, 10 min.; lane 4, 30 min. Lanes 6-8, toxin:papain 1:1 by mass; lane 6, 1 min.; lane 7, 5 min.; lane 8, 10 min. Lane 5 contains the molecular mass standards. The markers used were: phosphorylase (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20 kDa) and α–lactalbumin (14 kDa).

The toxin fragments were electroblotted onto polyvinylidene difluoride membranes and the first twenty amino acids were sequenced. The fragments with apparent molecular masses of 36 and 26 kDa each gave the sequence FSGPEFTFPLYGTMGNAAPQ, while the fragment with the apparent molecular mass of 33 kDa gave the sequence IETGYTPIDISLSLTQFLLS. The former sequence corresponds to that following glycine 327 of the gene-deduced protoxin sequence (Adang *et al.*, 1985), and the latter corresponds

to the sequence of the N-terminus of the toxin generated by tryptic cleavage of the arginine at position 28 (Section 2.3.2). From the gene-deduced sequence (Adang *et al.*, 1985), the calculated molecular mass of the N-terminal fragment (spanning residues 29-327) is 34.4 kDa. Similarly, since the C-terminus of toxin is residue 623 (Bietlot *et al.*, 1989), then the molecular mass of the large C-terminal fragment (residues 328-623) is calculated to be 32.3 kDa. It is not possible to calculate the true molecular mass of the small C-terminal fragment as the cleavage site giving rise to this polypeptide is unknown; the N-terminus could not be sequenced since the fragment does not appear on the gels. The absence of the small C-terminal fragment indicates that it is proteolyzed by papain in hot SDS buffer.

A comparison of the apparent and calculated molecular masses (36 *vs.* 32.3 kDa) for the large C-terminal fragment shows that this polypeptide runs considerably slower than expected on the gradient Phast SDS/PAGE gels. There is no obvious reason for this behaviour, but there are many examples in the literature of proteins which run anomalously on SDS/PAGE (Andrew, 1986). On a homogeneous slab SDS/PAGE gel run on a Biorad minigel system, the large C-terminal and the N-terminal fragments migrated very close together, accurately reflecting their similar molecular masses. It therefore appears that the anomalous relative mobility of the C-terminal toxin fragment is due to some property of the precast Phast gels.

It was found that incubation of toxin with elastase or trypsin in CAPS buffer at pH 9.5 for periods of up to two hours at room temperature did not result in proteolysis of the toxin. However, limited proteolysis could be induced by the addition of very large amounts of trypsin (toxin:trypsin; 1:10 by mass) or elastase (toxin:elastase; 1:2 by mass) to toxin dissolved in 8M urea, followed immediately by addition of SDS sample buffer at 20^{o}C and subsequent boiling (Fig. 4.2). Prolonged incubation of the toxin/protease samples in 8M urea at room temperature prior to the addition of SDS sample buffer and subsequent boiling did not increase the extent of toxin proteolysis, indicating that proteolysis was occurring under the strongly denaturing conditions of the hot SDS buffer.

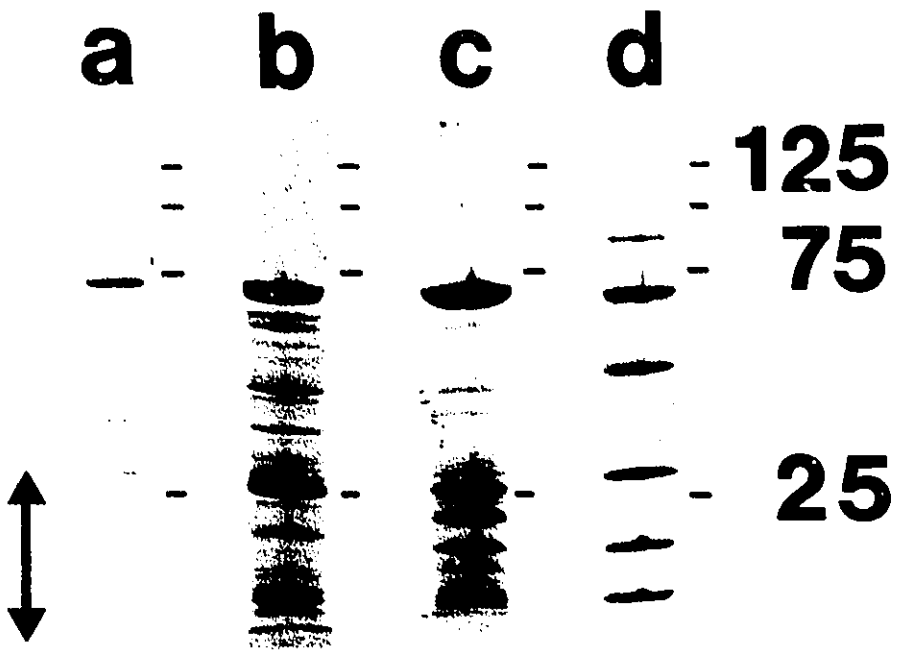


Figure 4.2. Limited proteolysis of toxin by papain, elastase and trypsin under denaturing conditions.
The proteolytic products generated by limited proteolysis of toxin with papain (lane a), elastase (lane b) and trypsin (lane c) are compared. Lane (d) contains the molecular mass markers. The double-headed arrow indicates the positions of bands arising from the proteases themselves. Elastase and trypsin autodigest to give multiple bands, whereas papain does not.

Elastase was found to generate two major toxin fragments which had apparent molecular masses of 38 and 33 kDa on Phast SDS gradient gels. Under similar conditions, trypsin gave rise to two major fragments with apparent molecular masses of 38 and 35 kDa (Fig. 4.2).

4.3.2 Differential scanning calorimetry

Differential scanning calorimetry performed on two preparations of the toxin at pH 10.5 gave a reproducible double-peaked endotherm, with maxima in the excess heat capacity occurring at 62°C and 69°C (Fig. 4.3). A shoulder on the high temperature peak is evident, and deconvolution of the data revealed a third distinct unfolding transition with an excess heat capacity maximum at 74°C (Fig. 4.4). The total calorimetric enthalpy calculated for the unfolding of the toxin is approximately 240 kcal/mole. The averaged contributions of the individual transitions obtained from the two scans are shown in Table 4.1, as are the calculated ΔH_{vH} values. The total ΔH_{vH} value for the unfolding of toxin is approximately 230 kcal/mol.

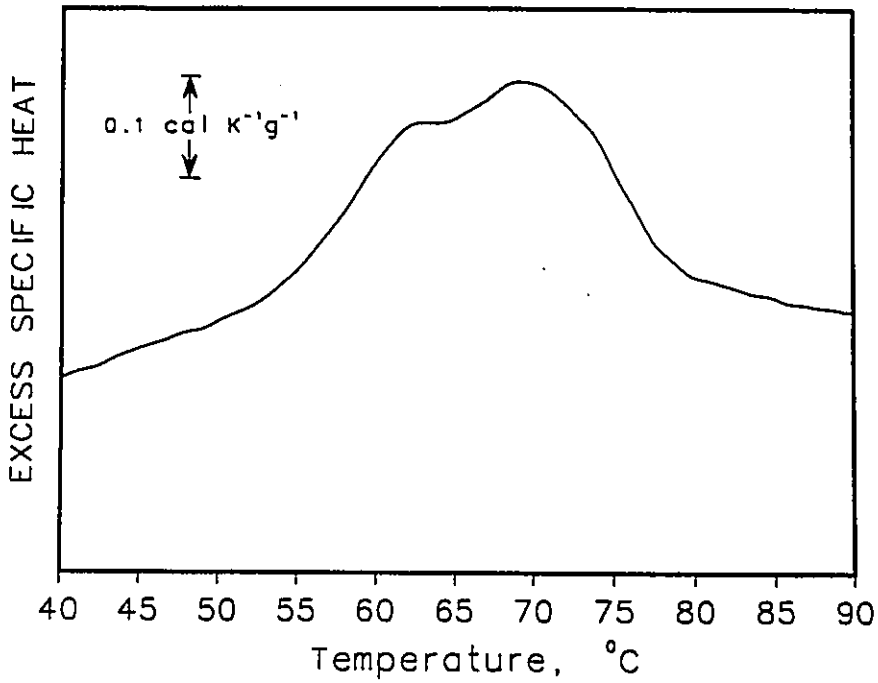


Figure 4.3. A plot of the excess specific heat as a function of temperature obtained from differential scanning calorimetry of toxin.
The concentration of protein was 4.7 mg/ml in 0.1M CAPS buffer, pH 10.5, containing 0.15M NaCl. The baseline was not subtracted.

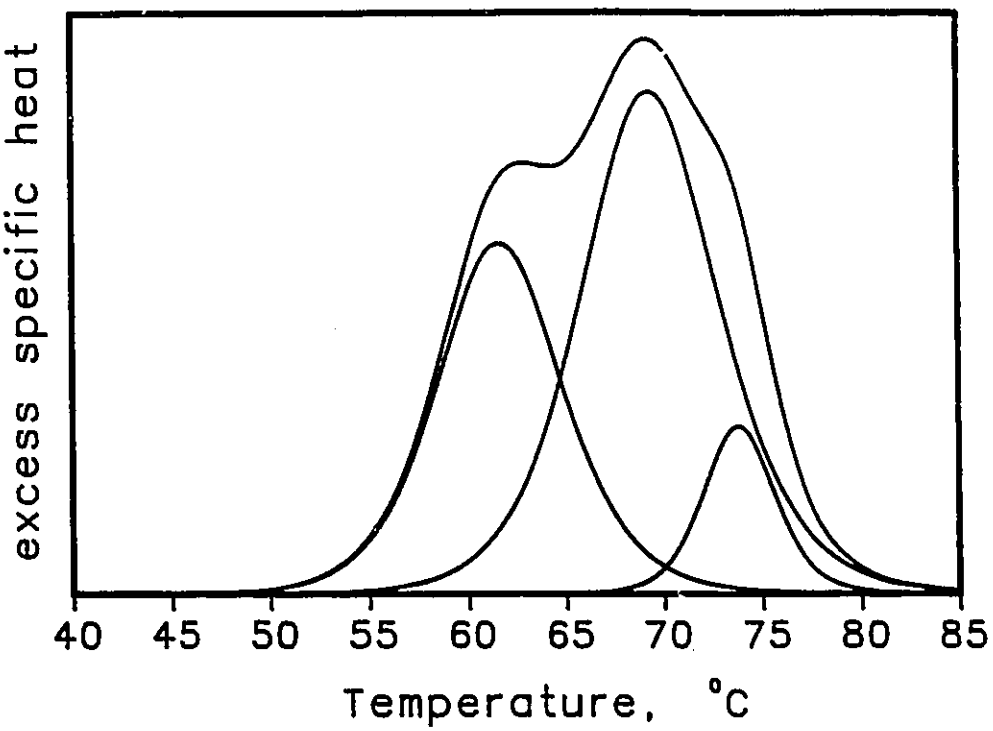


Figure 4.4. Curve-fitting of toxin calorimetric data.
The scan shown in Fig. 4.3 was baseline-corrected using the splines method. The curve was fitted with the minimal number of transitions required to approximate the shape of the experimental curve, and the transition parameters were then adjusted until the sum of the transitions gave the best fit of the experimental data. The theoretical excess heat capacity curve shown is superimposable on the experimental curve.

Table 4.1 Calorimetric and van't Hoff enthalpies of the transitions comprising the thermal unfolding of toxin.

T_m (°C)	ΔH_{cal} (kcal/mol)[a]	ΔH_{vH} (kcal/mol)[a]	$\Delta H_{vH}/\Delta H_{cal}$[b]
62	97, 91; 94	90, 83; 87	0.93
69	122, 115; 119	102, 120; 111	0.93
74	24, 27; 26	39, 27; 33	1.3

(a) the first two numbers in each entry are estimated from each of two scans; the third number is the average of these two values.

(b) obtained from the average of each ΔH_{vH} and ΔH_{cal} value.

To verify that the two major transitions arise from the thermal denaturation of independently unfolding structures, toxin was heated in the calorimeter at 1°C/minute to 64°C (Fig. 4.5A), which corresponds to the temperature midpoint between the two major transitions. Following rapid cooling and re-equilibration at 15°C, the sample was reheated in the calorimeter to 95°C (Fig. 4.5B). Although there was now no indication of the first major transition peak, the second major transition still occurred at 69°C, and had a ΔH_{cal} of approximately 150 kcal/mole.

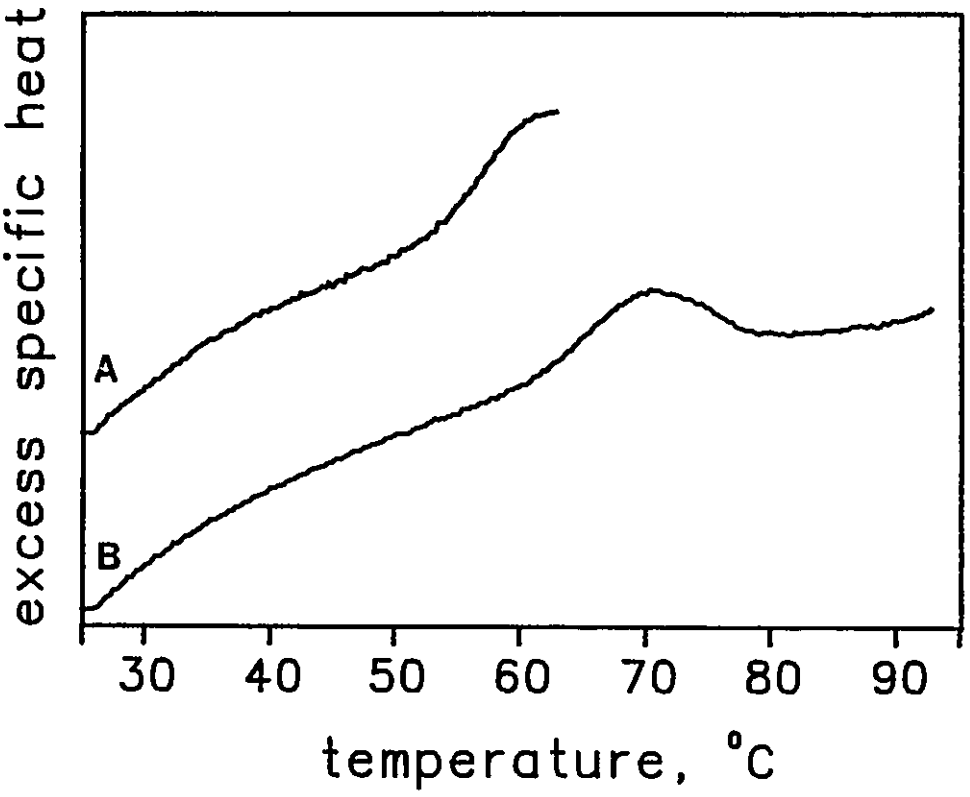


Figure 4.5. Separate calorimetry scans of the two principal unfolding domains in toxin.
The toxin was heated to 64°C, which is the temperature corresponding to the midpoint between the two major thermal events (curve A). Following cooling and re-equilibration, the same sample was heated to 95°C (curve B). The baselines were not subtracted.

4.3.3 Infrared spectroscopy

The change in the contour of the toxin amide I band with increasing temperature over the range 25-82°C is shown in Fig. 4.6. It can be observed that the band slowly broadens as the temperature is increased, and that broadening becomes noticeably more pronounced as the temperature increases over 60°C. There is no sudden change in the shape of the band which could be interpreted to reflect a major disruption of protein conformation (*e.g.*, the unfolding of one or more of the toxin domains).

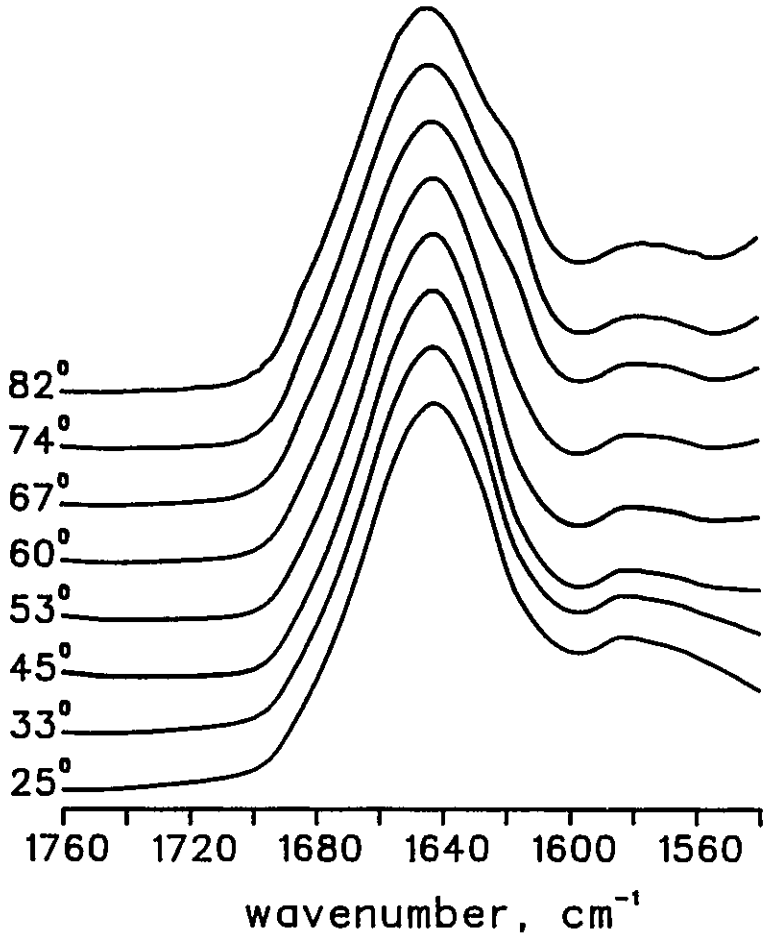


Figure 4.6. Change in the toxin infrared amide I band contour between 25°C and 82°C.

4.4 DISCUSSION

It was shown in Section 2.3.2 that the toxin molecule is resistant to proteolysis by a number of proteases, including papain. The results obtained here indicate that toxin is proteolyzed by papain immediately upon addition of boiling SDS sample buffer, before the papain is inactivated. Relatively large amounts of papain must be present for a significant amount of proteolysis to occur. Although it is somewhat surprising to see proteolysis taking place under such harsh denaturing conditions, papain is known to be unusually resistant to chemical denaturation (Sluterman, 1967). Proteases isolated from spruce budworm larvae have also been observed to proteolyze toxin and other proteins in boiling SDS sample buffer (personal communication, R. Milne, Forestry Canada).

The proteolysis of toxin by papain results in a discrete and highly reproducible fragmentation pattern. The reproducibility of this pattern indicates that the proteolytic cleavages are occurring at specific sites in the toxin molecule. At a papain:toxin ratio of approximately 1:5 by mass, two major fragments are generated (Fig. 4.1). One fragment was found to have the same N-terminus as toxin and had a calculated molecular mass of approximately 34.4 kDa. The N-terminus of the second fragment was identified as phenylalanine 328; this fragment had a calculated molecular mass of 32.3 kDa, although its apparent molecular mass, estimated from the SDS gel, was 36 kDa. When the papain:toxin ratio was higher (approximately 1:1 by mass), a second cleavage occurred in the C-terminal half of the molecule, giving rise to a fragment with an N-terminus at phenylalanine 328 and an apparent molecular mass of 26 kDa (Fig. 4.1). From the apparent molecular masses of the two C-terminal fragments on the gel (36 and 26 kDa), it can be estimated that the cleavage site which generates the second C-terminal fragment is approximately 100 residues in from the C-terminus of the toxin. However, a fourth fragment with a mass corresponding to 100 amino acids is not observed on the gel.

A two band fragmentation pattern from toxin in 8M urea can be generated with trypsin or elastase if the toxin/protease sample is added to SDS buffer at 20^{o}C and then

raised to boiling. The molecular masses of the fragments generated by these two proteases are very similar to those obtained with the lower concentration of papain (papain:toxin; 1:5 by mass). However, compared to the amount of proteolysis observed with papain, much less toxin is proteolyzed by trypsin or elastase (Fig. 4.2), and no proteolysis is observed unless the toxin is initially pretreated at 20°C with 8M urea. As will be shown in the following chapter, the extent of toxin unfolding after one hour in 8M urea is small, so it is probable that most of the observed proteolysis occurs as the toxin/protease sample is heated in the SDS sample buffer. Nonetheless, even a slight loosening of the toxin structure by pretreatment in urea at 20°C could cause a significant increase in the rate and extent of toxin unfolding upon addition of the strongly denaturing SDS buffer. Under these conditions, the toxin would be proteolyzed upon addition of trypsin or elastase, before the protease itself became denatured. When the toxin was not pretreated with 8M urea, the proteases became inactivated before any unfolding of the toxin occurred. The suggestion that proteolysis occurs in the SDS sample buffer is supported by the observation that the length of incubation of urea-treated toxin with protease at 20°C had no positive effect on the extent of proteolysis; indeed, the longer the incubation, the less the amount of proteolysis, indicating that the enzyme was denaturing even before the SDS sample buffer was added.

It has been observed previously that proteins treated with SDS can be cleaved at specific sites with proteases (Cleveland *et al.*, 1977). The consistency with which the toxin is proteolyzed into two fragments of approximately equal size by three different proteases demonstrates that these cleavages are not fortuitous, but rather reflect the fundamental architecture of the toxin. The results suggest that as the toxin unfolds in the hot SDS buffer, the interdomain region connecting the N- and C-terminal domains of the toxin is exposed first and is cleaved before the protease is inactivated. A somewhat slower event is the unfolding of the polypeptide chain connecting the small (approximately 100 amino acid) sub-domain to the major C-terminal domain. If relatively high concentrations of papain are initially added to the sample, sufficient proteolytic activity remains after the initial cleavage

to cleave this second interdomain region as it is exposed. Since no band corresponding to a mass of 100 amino acids is observed on the gels, this subdomain probably unfolds rapidly in the hot SDS buffer, and is proteolyzed to small peptides.

Differential scanning calorimetry provides independent support for the above conclusions regarding the toxin's domain structure. Multiple-peaked transitions have been reported previously for a number of proteins, and are believed to reflect the existence within the protein molecule of two or more domains which unfold independently (Takahashi *et al.*, 1981; Tischenko *et al.*, 1982; Privalov, 1982). The two major events observed during the heat denaturation of toxin (Fig. 4.3) were shown experimentally to occur independently of each other (Fig. 4.5). Deconvolution of the data resolved a shoulder on the higher temperature transition peak and showed it to be a distinct thermal event (Fig. 4.4).

The thermal unfolding of toxin occurs over a range of approximately $35^{o}C$, producing a broad denaturation profile. High concentrations of toxin (approximately 5 mg/ml) were therefore required in order to obtain scans in which data were readily discernable from instrument noise. However, at such high concentrations aggregation and precipitation of the protein at elevated temperatures was a recurrent problem: of nine attempts to obtain calorimetric data on toxin, only three scans (including the discontinuous two-stage scan shown in Fig. 4.5) were successfully completed. Qualitatively, the unfolding profiles from these scans were very similar, and showed that at least three transitions are involved in the unfolding of the toxin molecule. Quantitatively, similar ΔH_{cal} estimates were obtained from the two continuous scans for both the overall unfolding of the toxin and for the individual transitions; however, the estimates of ΔH_{vH} obtained by curve-fitting of the two data sets differed significantly (Table 4.1). Attempts to resolve these differences by baseline manipulation or by increasing the number of adjustable parameters in the curve-fitting process did not alleviate the discrepancies.

The thermodynamic analysis of the unfolding of toxin is also hampered by the fact that the calculation of ΔH_{vH} is strictly only valid for a reversible process, whereas thermally unfolded toxin is not observed to refold under the experimental conditions used. However, the thermal denaturation of many proteins is essentially irreversible, but reversible thermodynamics are still applied to the study of their unfolding behaviour (Takahashi *et al.*, 1981; Privalov, 1982; Privalov and Potekhin, 1986; Sturtevant, 1987). Despite the limitations of the data, the ΔH_{cal} and ΔH_{vH} values were estimated for the three unfolding transitions (Table 4.1). It is obvious that the two major transitions occur independently of each other (Fig. 4.5) in the manner of two independently unfolding domains. This conclusion is supported by the values of $\Delta H_{vH}/\Delta H_{cal}$ for each of these transitions, which in both cases gave a value of approximately one. This is interpreted to mean that the unfolding events giving rise to these transitions occur in an all-or-none fashion, with one process being unaffected by the other (Privalov and Potekhin, 1986). In contrast, the value of $\Delta H_{vH}/\Delta H_{cal}$ for the third transition is greater than one, suggesting that this transition may not occur independently (Sturtevant, 1987). The ratio of $\Delta H_{vH}/\Delta H_{cal}$ for the total protein would be expected to be less than one if the molecule exhibits several discrete intermediate states during unfolding (Sturtevant, 1987). The present data provides a value marginally less than one, which suggests that at least some of the domains in toxin (probably those giving rise to the first and second thermal transitions) are independently unfolding structures. However, these conclusions are based on only two calorimetric scans of the toxin, and so the interpretations must be treated with caution. A more comprehensive calorimetric study is required before a true thermodynamic analysis of the thermal unfolding of toxin can be undertaken.

Results from infrared spectroscopy of the amide I band of toxin corroborate that thermal denaturation begins at approximately 55^{o}C (Fig. 4.6); however, no distinct event which can be attributed to the unfolding of a domain is observed. Therefore, the two major

thermal events observed by differential scanning calorimetry are not sufficiently discrete to be resolved by infrared spectroscopy.

Since functional domains are independently-folding, stable structures, it should be possible to isolate and characterize the domains which constitute a protein. Examples of characterized domains which retain their functionality when removed from the parent protein include a domain from aspartokinase I:homoserine dehydrogenase I from *Escherichia coli* and an N-terminal domain from *E. coli* DNA polymerase I (Kirschner and Bisswanger, 1976). In the present study, an attempt was made to take advantage of the different thermal stabilities of the two major domains of toxin. Preliminary proteolytic experiments were conducted at elevated temperatures with a view to isolating the relatively thermostabile domain so that its functional role in the protein could be determined, and also so that conditions leading to its refolding after thermal denaturation could be determined. It was found that heating the toxin to between 55-64^{o}C in pH 10.5 CAPS buffer and then treating with protease did not result in proteolysis of the toxin. Incubating toxin in various concentrations of either SDS or urea at a variety of elevated temperatures, followed by the addition of protease, either left the toxin untouched or caused it to be completely proteolyzed, depending on the temperature of the pretreatment and the enzyme used. These preliminary experiments suggest that either there is only a narrow range of physical and chemical conditions which destabilize one domain and leave the other unaffected, or that the destabilization of one domain causes the entire protein to destabilize and become susceptible to proteolysis. The stabilizing effects of interactions at domain interfaces are believed to be of critical importance in the overall stability and integrity of many multidomain proteins (Janin and Wodak, 1983).

The proteolytic and calorimetric data presented above show that toxin is composed of an N-terminal and a C-terminal domain of approximately equal mass, and these two major domains appear to unfold independently in the manner of functional domains (Takahashi *et al.*, 1981). The proteolytic data indicate that the C-terminal domain contains

a subdomain, while the calorimetric data show that the high temperature transition is composed of two thermal events. It is therefore postulated that the low temperature transition is due to the thermal unfolding of the N-terminal domain, while the high temperature transition is due to the unfolding of the C-terminal domain and its subdomain.

While the above results were in press (Choma *et al.*, 1990), Convents *et al.* (1990) reported very similar conclusions regarding the domain structure of the toxin from *B. thuringiensis* subsp. *berliner*. Using the technique of limited proteolysis, Convents *et al.* (1990) concluded that the toxin is composed of N- and C-terminal domains of approximately equal mass; however, the cleavage site giving rise to these two fragments was not determined. An interesting observation made in this paper is that the N-terminal domain of toxin from *B. thuringiensis* subsp. *berliner* is less stable in chemical denaturants at pH 11 than is the C-terminal domain. If the same is true of toxin from *B. thuringiensis* subsp. *kurstaki* HD-73, then this observation by Convents *et al.* (1990) lends support to the postulation that the first thermal transition observed by differential scanning calorimetry of toxin is due to the unfolding of the N-terminal domain.

The domain structures of various bacterial toxins such as *Pseudomonas* exotoxin A and diphtheria toxin have been extensively studied, and it has been determined that these toxins are composed of several functional domains which mediate operations such as receptor binding and membrane translocation (Neville and Hudson, 1986; Zalman and Wisnieski, 1985). As discussed in Section 1.2.4, gene nucleotide sequence comparisons of the various protoxin genes have shown that the *B. thuringiensis* subsp. *kurstaki* HD-73 toxin molecule, like the other cryIA gene products, can be divided into three distinct regions: the conserved (residues 1 to approximately 280), variable (residues 280-460) and hypervariable (residues 460-623) regions (Andrews *et al.*, 1986). It may be assumed that the difference in insect-target specificities of the various *B. thuringiensis* toxins is due to amino acid differences in the proteins, and it has been proposed that target specificity resides in either the variable or hypervariable regions of the molecule (Geiser *et al.*, 1986).

In the cryIA(a) gene product, the specificity domain for toxicity towards *Bombyx mori* was shown by Ge *et al.* (1989) to reside between amino acid residues 332-450. This was determined by exchanging this region of the *cryIA(c)* gene (from *B. thuringiensis* subsp. *kurstaki* HD-73) with the excised region of the *cryIA(a)* gene coding for residues 332-450. The cryIA(c) gene product, normally toxic towards *Manduca sexta*, was now fully toxic towards *Bombyx mori*. If, as is generally accepted (Andrews *et al.*, 1986; Brousseau and Masson, 1988; Höfte and Whiteley, 1989), the toxin must bind to receptors in the insect midgut in order to elicit a toxic response, then the results of Ge *et al.* (1989) suggest that the receptor binding domain is encompassed by amino acid residues 332-450 in the cryIA(a) gene product. Given the similarities between the protoxin genes, it is probable that the receptor binding domain in all the gene products lies in this general vicinity. It has been further postulated (Brousseau and Masson, 1988; Höfte and Whiteley, 1989) that the conserved region of the protoxin gene codes for the toxicity domain. This proposal is largely based on the prediction, noted in Section 3.4, that the first 100 amino acids may constitute a transmembrane sequence (Schnepf *et al.*, 1985). A structure capable of membrane penetration would be required if the toxin acts by forming pores in the insect midgut cell membranes (Section 1.1.4; Knowles and Ellar, 1987). In light of this evidence for the functional roles of the conserved and variable regions of the toxin molecule, the present observation that proteases preferentially cleave the toxin molecule in the vicinity of amino acid residue 327 lends weight to the conclusion that this is the interdomain region between the toxic and binding domains.

Deletion studies have shown that the extreme C-terminus of the toxin from *B. thuringiensis* subsp. *kurstaki* HD-73 plays a vital role in the structure and function of the molecule. A gene truncated to encode the first 611 amino acids was found to produce a protein fully toxic to *Manduca sexta*, whereas truncating the gene further to encode only the first 596 amino acids produced no stable protein product (Aronson *et al.*, 1986). Adang *et al.* (1985) noted that a *cryIA(c)* gene encoding the first 490 amino acids produced a stable,

but non-toxic, protein. It is not clear why a product encompassing residues 1-490 is stable whereas a protein composed of residues 1-596 is not; nonetheless, the fact that the former protein is not toxic indicates that the extreme C-terminus is necessary for the structural integrity of the *B. thuringiensis* subsp. *kurstaki* HD-73 toxin. It is interesting to note that in the gene fragment exchange experiments of Ge *et al.* (1989), exchange of genetic material in the region of the gene coding for approximately residues 540-620 resulted in gene products which were not toxic, even though the region encoding the specificity domain remained intact. While no specific function has been assigned to this region of the toxin molecule (termed the 'toxic boundary domain' by Andrews *et al.*, 1986), the papain proteolytic data and calorimetric results presented here provide experimental evidence that this region folds into a discrete subdomain. The schematic representation of the domain structure of toxin presented in Fig. 4.7 is based on the experimental data presented in this chapter, and on the gene alignment and site-directed mutagenesis studies reported in the literature.

Concurrent with the present experiments on the domain structure of toxin, an interesting observation regarding the domain structure of the protoxin molecule was made. As noted in Section 3.4, the protoxin molecule can be divided into two approximately equal parts: the N-terminal half which is predominantly hydrophobic, and the C-terminal half which is largely hydrophilic. The former section of the molecule is believed to be composed mainly of ß-sheet structures, while the latter appears to contain a higher proportion of α–helices. These observations indicate that the two halves of the protoxin molecule are structurally distinct, and that they may have very different biological functions (Section 1.2.4). Using the technique of limited proteolysis, the domain structure of the C-terminal half of the protoxin molecule was shown to be highly unusual (Choma *et al.*, 1990). Incubation of trypsin in pH 10.5 CAPS buffer with solubilized protoxin from a variety of *B. thuringiensis* strains resulted in six fragments with masses of 8-10 kDa being cleaved from the C-terminus. The last cleavage resulted in the production of the protease-

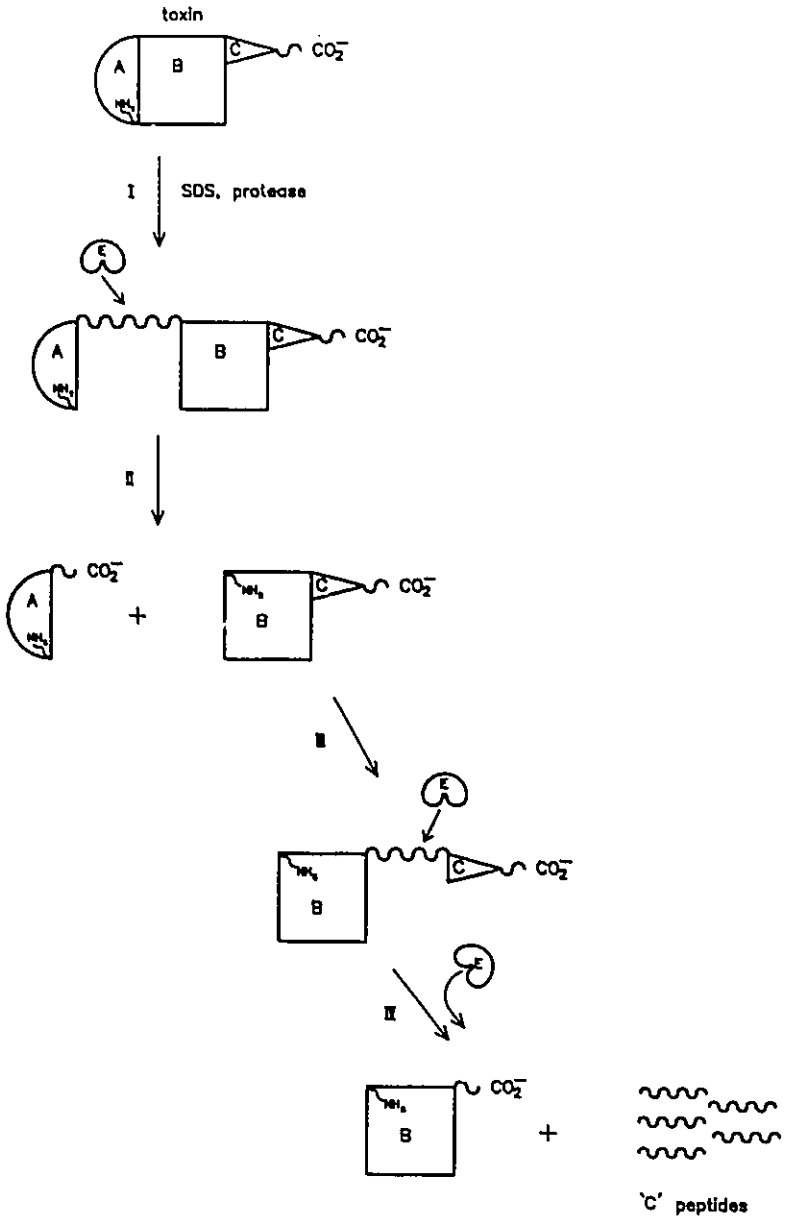


Figure 4.7. Schematic diagramme of the domain structure of toxin as elucidated by limited proteolysis.
Toxin added to hot SDS in the presence of papain (step I) unfolds to expose the interdomain region between the N-terminal toxic domain (A) and the binding and toxic boundary domains (B and C, respectively) in the C-terminal half of the molecule. This exposed stretch of polypeptide is cleaved by papain at residue 327 (step II) to yield the N-terminal domain, which remains folded and thus protease resistant, and the two connected C-terminal domains. Upon longer incubation in hot SDS, the interdomain region between the binding and toxic boundary domains is exposed (step III) and cleaved by the protease. The binding domain remains intact, but the toxic boundary domain unfolds and is proteolyzed (step IV) into peptides.

stable toxin. These fragments were not cleaved in a random fashion, but rather in a sequential manner, with the first cleavage occurring approximately 80 amino acids in from the C-terminus, and then every 80-90 amino acids along the polypeptide chain in the direction of the N-terminus. There is no precedent in the literature for a multidomain structure giving rise to the large number of observed sequential proteolytic cleavages. One possible explanation for the sequential degradation is that the small domains (or subdomains) interact with each other such that proteases are sterically hindered from approaching a potential cleavage site until the adjacent domain is removed. Another possibility is that successive conformational changes take place along the polypeptide chain with each successive cleavage. These results show that the conversion of protoxin to toxin proceeds through a series of obligatory intermediates. It is therefore likely that a number of toxic fragments with different potencies are generated by the proteolytic enzymes in the insect gut. Variation among larvae in rates or specificities of proteolysis will alter this distribution and may account, at least in part, for the differences in the host spectrum of protoxins from various *B. thuringiensis* strains. In addition, although little is known concerning the functional role of the C-terminal half of the protoxin molecule, the fact that its sequence is highly conserved (Section 1.2.4; Andrews *et al.*, 1986) indicates that this region has an important structural role. Höfte and Whitely (1989) have suggested that the C-terminal region is conserved because it is essential for crystal formation; therefore, the domain structure which gives rise to the sequential proteolysis may be a consequence of structural constraints imposed by the requirement for the formation of a regular crystal lattice involving disulphide bridges.

4.5 CONCLUSIONS

Calorimetric and limited proteolysis data provide independent evidence that there are at least three domain structures in the toxin molecule. The observed proteolytic cleavage site in toxin at glycine 327 is in the interdomain region between the constant (toxic domain)

and variable (binding domain) regions observed by gene fragment exchange experiments (Ge *et al.*, 1989) and predicted by comparison of gene nucleotide sequences (Brousseau and Masson, 1988; Höfte and Whiteley, 1989). The third, smaller structure is the toxic boundary domain. Three domains have been predicted from hydropathic and gene sequence analyses of the toxin (Andrews *et al.*, 1987): an N-terminal domain of approximately 300 amino acids, followed by a variable sequence domain of approximately 200 residues and a C-terminal domain of approximately 150 residues. The proteolytic cleavage patterns and calorimetric data presented here correlate well with these predictions.

4.6 REFERENCES

Adang, M. J., M. J. Staver, T. A. Rocheleau, J. Leighton, R. F. Barker and D. V. Thompson (1985). Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. *Gene* **36**, 289-300.

Andrew, A. T. (1986). *Electrophoresis: theory, techniques and biochemical and clinical applications* (second edition). Clarendon Press, Oxford, pp. 117-123.

Andrews, R. E. Jr., R. M. Faust, H. Wabiko and K. C. Raymond (1987). The biotechnology of *Bacillus thuringiensis*. *CRC Crit. Rev. Biotechnol.* **6**, 163-230.

Aronson, A. I., W. Beckman and P. Dunn (1986). *Bacillus thuringiensis* and related insect pathogens. *Microbiol. Rev.* **50**, 1-24.

Bietlot, H., P. R. Carey, C. T. Choma, H. Kaplan, T. Lessard and M. Pozsgay (1989). Facile preparation and characterization of the toxin from *Bacillus thuringiensis* var. *kurstaki*. *Biochem. J.* **260**, 87-91.

Brousseau, R. and L. Masson (1988). *Bacillus thuringiensis* insecticidal crystal toxins: gene structure and mode of action. *Biotech. Adv.* **6**, 697-724.

Choma, C. T., W. K. Surewicz, P. R. Carey, M. Pozsgay, T. Raynor and H. Kaplan (1990). Unusual proteolysis of the protoxin and toxin from *Bacillus thuringiensis*: structural implications. *Eur. J. Biochem.* **189**, 523-527.

Cleveland, D. W., S. G. Fischer, M. W. Kirschner and K. U. Laemmli (1977). Peptide mapping by limited proteolysis in sodium dodecyl sulfate and analysis by gel electrophoresis. *J. Biol. Chem.* **252**, 1102-1106.

Convents, D., C. Houssier, I. Lasters and M. Lauwereys (1990). The *Bacillus thuringiensis* delta-endotoxin: evidence for a two domain structure of the minimal toxic fragment. *J. Biol. Chem.* **265**, 1369-1375.

Creighton, T. E. (1984). *Proteins*. W. H. Freeman and Co., New York, pp. 224-225.

Edelman, G. M., B. A. Cunningham, W. E. Gall, P. D. Gottlieb, U. Rutishauser and M. J. Wexdal (1969). The covalent structure of an entire gamma-G immunoglobulin molecule. *Proc. Nat. Acad. Sci. U.S.A.* **63**, 78-85.

Ge, A. Z., N. I. Shivarova and D. H. Dean (1989). Location of the *Bombyx mori* specifictiy domain on a *Bacillus thuringiensis* delta-endotoxin. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4037-4041.

Geiser, M., S. Schweitzer and C. Grimm (1986). The hypervariable region in the genes coding for entomopathogenic crystal proteins of *Bacillus thuringiensis*: nucleotide sequence of the *kurhl* gene of subsp. *kurstaki* HD1. *Gene* **48**, 109-118.

Ghelis, C. and J. Yon (1982). *Protein folding*. Academic Press, New York, pp. 410-413.

Goldberg, M. E. (1969). Tertiary structure of *Escherichia coli* ß-D-galactosidase. *J. Mol. Biol.* **46**, 441-446.

Höfte, H. and H. R. Whitely (1989). Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Jaenicke, R. (1987). Folding and association of proteins. *Prog. Biophys. Molec. Biol.* **49**, 117-237.

Janin, J. and S. J. Wodak (1983). Structural domains in proteins and their role in the dynamics of protein function. *Prog. Biophys. Molec. Biol.* **42**, 21-78.

Kirschner, K. and H. Bisswanger (1976). Multifunctional proteins. *Ann. Rev. Biochem.* **45**, 143-166.

Knowles, B. and D. Ellar (1987). Colloid-osmotic lysis is a general feature of the mechanism of action of *Bacillus thuringiensis* delta-endotoxins with different insect specificities. *Biochim. Biophys. Acta* **924**, 509-518.

Langerman, N. and R. L. Biltonen (1979). Microcalorimeters for biological chemistry: applications, instrumentation and experimental design. *Methods Enzymol.* **61**, 261-318.

Linderstrom-Lang, K. U. and J. A. Schellman (1959). Protein structure and enzyme activity. (in) *The Enzymes*, vol. I, second edition (P. D. Boyer, ed.) Academic Press, New York, pp. 443-510.

Neville, D. M. and T. H. Hudson (1986). Transmembrane transport of diphtheria toxin, related toxins, and colicins. *Ann. Rev. Biochem.* **55**, 195-224.

Pfiel, W. and P. L. Privalov (1976). Thermodynamic investigations of proteins. *Biophys. Chem.* **4**, 23-50.

Privalov, P. L. (1982). Stability of proteins. *Adv. Protein Chem.* **35**, 1-104.

Privalov, P. L. and S. A. Potekhin (1986). Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol.* **131**, 4-51.

Schnepf, H. E., H. C. Wong and H. R. Whiteley (1985). The amino acid sequence of a crystal protein from *Bacillus thuringiensis* deduced from the DNA base sequence. *J. Biol. Chem.* **260**, 6264-6272.

Schulz, G. E. and R. H. Schirmer (1979). *Principles of protein structure.* Springer-Verlag, New York, pp. 46-47.

Sluterman, L. A. (1967). The effect of methanol, urea and other solutes on the action of papain. *Biochim. Biophys. Acta.* **139**, 418-429.

Sturtevant, J. M. (1972). Calorimetry. *Methods Enzymol.* **26**, 227-253.

Sturtevant, J. M. (1987). Biochemical applications of differential scanning calorimetry. *Ann. Rev. Phys. Chem.* **38**, 463-488.

Surewicz, W. K. and H. H. Mantsch (1988). New insight into protein secondary structure from resolution-enhanced infrared spectroscopy. *Biochim. Biophys. Acta* **952**, 115-130.

Takahashi, K., J. L. Casey and J. M. Sturtevant (1981). Thermodynamics of the binding of D-glucose to yeast hexokinase. *Biochemistry* **20**, 4693-4697.

Tischenko, V. M., V. P. Zav'yalow, G. A. Medgyesi, S. A. Potekhin, and P. L. Privalov (1982). A thermodynamic study of cooperative structures in rabbit immunoglobulin G. *Eur. J. Biochem.* **126**, 517-521.

Zalman, L. S. and B. J. Wisnieski (1985). Characterization of the insertion of *Pseudomonas* exotoxin A into membranes. *Infect. Immun.* **50**, 630-635.

Chapter 5.

STRUCTURAL IMPLICATIONS OF THE UNFOLDING AND FOLDING OF PROTOXIN

The Optimist's Obelisk

I go past a stone
by the road twice a day
in my regular
toing and froing.
The sight of it tells me
I'm over half way-
whether I'm
coming or going.

- Piet Hein

Chapter 5.

STRUCTURAL IMPLICATIONS OF THE UNFOLDING AND FOLDING OF PROTOXIN

5.1 INTRODUCTION

The preceding three chapters have dealt with the characterization of structural features in the activated toxin molecule. Where possible, the novel information has been related to known functional aspects of this entomocidal protein. In the second half of this thesis, the scope of the investigation is broadened to include protoxin, with the specific objective of determining the structural and functional significance of those portions of the protoxin molecule which are excised during the activation process: the C-terminal half of the molecule, and the first 28 amino acids of the N-terminus.

As noted earlier (Sections 1.2, 3.4 and 4.4), the protoxin molecule appears to be composed of two distinct structural regions. The C-terminal half is hydrophilic, contains 14 of the 16 cysteine residues and 31 of the 34 lysine residues, and is composed of approximately six subdomains. The N-terminal half from which the toxin is derived is hydrophobic, resistant to proteolysis, and is composed of at least three domains which appear to have specific functional roles (Sections 4.3 and 4.4). The primary role of the C-terminal half of the protoxin molecule is believed to be in promoting crystal formation (Höfte and Whiteley, 1989). It has also been postulated that the carboxy-terminal half may somehow protect the toxic moiety from degradation within the *B. thuringiensis* bacterium, or that perhaps it functions in the attachment of the molecule to the insect gut epithelial receptors (Aronson *et al.*, 1986). However, the fact that the sequence of the C-terminal half is highly conserved among *cryI* protoxin genes suggests that this region has a crucial function. Although it may be that this sequence conservation merely reflects an essential structural requirement for crystal formation, it has recently been postulated (Andrews *et al.*, 1986; Brousseau and Masson, 1988) that the C-terminal half of protoxin affects the conformation and protease sensitivity of the N-terminal, toxic moiety. This hypothesis is in agreement with results described by Aronson *et al.* (1986; see Section 4.4), which showed that truncation of the *B. thuringiensis* subsp. *kurstaki* HD-73 gene to encode the first 611 amino acids lead to a fully toxic protein being expressed, but further truncation to

the 596th codon lead to no stable protein being produced. Truncation of the gene to the 603rd codon resulted in the generation of a relatively stable but non-toxic protein (Schnepf and Whiteley, 1985). Similarly, it was noted by Andrews *et al.* (1985) that deletions in the 3' end of the gene lead to the accumulation of proteins smaller than that predicted by the size of the gene. It has also been reported that the removal of several amino acids from the C-terminus of a dual-specificity toxin from *B. thuringiensis* subsp. *aizawai* is sufficient to abolish Lepidopteran specificity without affecting toxicity to Dipteran larvae (Haider *et al.*, 1989).

In light of this evidence that the C-terminus in some way confers specificity and/or stability to the N-terminal toxic moiety, the present study was undertaken to investigate whether the C-terminal half of the protoxin molecule affects the folding, and hence conformation, of the N-terminal half. Specifically, the overall goal of the work presented here was to establish whether the C-terminal half of protoxin is required for proper folding of the N-terminal toxic moiety, and whether a conformational change occurs during the proteolytic processing of the protoxin to toxin.

5.1.1 The unfolding and folding of proteins

Studies on ribonuclease by Anfinsen and Haber (1961) provided evidence that the amino acid sequence contains all the necessary information for a protein to fold into its native, active conformation. Recently there has been renewed interest in the problem of protein folding, kindled by the need to predict the conformation of engineered proteins; however, many fundamental questions regarding the phenomenon of protein folding have remained unanswered. The complexity of the problem is compounded by the lack of understanding of whether insights obtained from *in vitro* renaturing experiments are applicable to understanding the *in vivo* folding processes of the nascent polypeptide chain.

The terms 'denatured protein' and 'unfolded protein' remain poorly defined, largely because each protein reacts differently in a given denaturing environment. *In vitro*, denaturing conditions can be achieved using a large number of organic and inorganic

compounds, with urea and guanidinium chloride (GuHCl) being the two most common denaturants. GuHCl is the more potent of the two denaturants, and appears to interact only with residues on the surface of the protein. It acts by making the aqueous environment a better solvent for nonpolar residues, thus diminishing the hydrophobic interactions within the protein molecule and removing the driving force for these groups to remain buried in the protein interior (Hibbard and Tulinsky, 1978). Since hydrophobic interactions are believed to be the major factor stabilizing the conformations of proteins (Creighton, 1978), this is sufficient to cause denaturation. Urea also denatures proteins by making nonpolar groups more soluble. As the stability of different regions of a protein will be dependent to varying degrees on hydrophobic interactions, it is probable that these denaturants produce a wide range of effects in a single protein molecule, depending upon the local geometry of the interacting groups (Creighton, 1984). For the purpose of most protein folding studies, this resulting heterogeneous assembly of denatured proteins is simplified conceptually: the unfolded protein is represented by a population of random coil conformations in which each bond in the polypeptide backbone can freely rotate independently of the rotation of all other groups along the chain (Jaenicke, 1987).

Determining the sequence of events which leads to the regeneration of the native protein structure from the unfolded polypeptide has proven to be an extremely complex problem. The complete description of the folding pathway for a given protein implies the characterization of the denatured state and the final native state, together with all the intermediates along the transition. In general, it has not been possible to characterize discrete intermediates, primarily because these species are thermodynamically unstable and thus only have a transient existence (Tanford, 1968). In addition, the refolding process as a whole is very rapid; for example, the estimated half-time for refolding of staphylococcal nuclease is 0.1-0.2 seconds (Ghelis and Yon, 1982). It has nonetheless been established for a number of proteins (e.g., bovine pancreatic trypsin inhibitor, ribonuclease and penicillinase) that a defined pathway is followed during the regeneration of native tertiary

structure from the unfolded protein (Jaenicke, 1987). However, until a better understanding of the folding pathways for proteins emerges, the folding of most proteins will continue to be approximated by the two-state model, in which the polypeptide is assumed to be either completely denatured or totally reconstituted into its native state, and alternates between these two states depending on the environment (Creighton, 1978).

Given that the entire length of the denatured polypeptide chain will not be in a totally random conformation at every moment in time, it has been suggested that the areas of residual structure act as nuclei for the refolding process (Wetlaufer, 1973; Creighton, 1978; Jaenicke, 1987). A 'consensus pathway' for sequential protein folding (Tanford, 1970) proposes that folding starts independently and more or less simultaneously in many regions along the polypeptide chain by the formation of short stretches of secondary structure. These short-lived structures involve only next-neighbour interactions in an area of several residues, and appear and disappear in a matter of microseconds (Jaenicke, 1987). Eventually in the folding process, more stable or longer structures appear which interact with other structures in the polypeptide chain, and form 'micro-domains'. In multidomain proteins, these 'micro-domains' subsequently merge into true domain structures. At this stage, these intermediate conformations of the polypeptide have some stability due to the extensive interactions of the secondary structure elements within the domains. The native tertiary structure is regenerated when the domains come together and interact with each other in the manner of true folded proteins. Postulating such a role for domain structures in protein folding helps explain how an unfolded polypeptide chain manages to find its distinct energy minimum within an extremely short period of time, compared with the time predicted for a random search mechanism (Ghelis and Yon, 1982).

In the present study, the extent and rate of unfolding of toxin, and of the toxic moiety in protoxin, was determined by monitoring the increasing protease sensitivity of the proteins upon prolonged incubation in chemical denaturants. Fluorescence emission spectroscopy was also used to compare the time dependence and extent of unfolding of

toxin and of protoxin. Using tryptophan as an intrinsic probe, changes in the fluorescence spectrum of toxin or protoxin upon incubation in GuHCl or urea revealed changes in the environment of the tryptophans, and thus in the conformations of the proteins. The refolding of the denatured proteins into their native conformations was followed using the same techniques; in addition, *in vitro* toxicity assays were used to verify that the proteins had not only refolded into what appeared to be their native conformations, but also into their biologically active conformations. The resulting data were analyzed with a view to answering the following questions: (i) Is the C-terminal half of the protoxin molecule required for the folding of the N-terminal toxic moiety? (ii) Is the activation process accompanied by a conformational change in the toxic moiety? (iii) Does the crystal protein (protoxin) fold independently of its incorporation into the crystal lattice?

5.2 MATERIALS AND METHODS

5.2.1 Materials

B. thuringiensis subsp. *kurstaki* HD-73 protoxin and toxin were prepared and characterized as described in Sections 2.2.2 and 2.2.3. Both proteins were quantified by their absorbance at 280 nm, as described in Section 2.2.4.3. TPCK-trypsin, elastase, porcine pepsin, papain and the papain inhibitor E64 were all obtained from Sigma Chemical Corp., St. Louis, Missouri. Sequenal grade urea and guanidinium chloride were obtained from Pierce Chemical Co., Rockford, Illinois. Precast gels and buffer strips for the Pharmacia Phast electrophoresis system were purchased from Pharmacia Canada Ltd., Dorval, Quebec. Dialysis tubing (3.5 kDa cutoff) was purchased from Spectrum Medical Industries, Los Angeles, California. All other chemicals were high purity reagents obtained from commercial sources. Solutions were prepared with reverse-osmosis quality water purified by the Milli-Q water system.

5.2.2 Designations for denatured and renatured protein.

The protocol used to investigate the effect of urea and GuHCl on protoxin, and their effect on toxin generated by trypsin from protoxin both in the presence and absence of denaturants, is summarized in Fig. 5.1. A similar protocol, shown in Fig. 5.2, was used to generate toxin by papain digestion of protoxin. Bold-face type designates proteins which were assayed for biological activity.

5.2.3 Generation of toxin by papain proteolysis

Protoxin crystals (50 mg) were suspended in 3 ml of 0.1M CAPS buffer, pH 10.5, and 1 mg papain (activated by the addition of 0.1% ß-mercaptoethanol) was added. The sample was stirred for two hours at 20°C, then a hundredfold excess (papain:inhibitor) of E64 was added. The sample was then centrifuged and filtered, and the supernatant was treated in a manner identical to that described for trypsin-generated toxin (Section 2.2.3).

5.2.4 N-Terminal sequence analysis.

The N-terminal sequence of toxin generated by the action of papain on protoxin was determined using a model 470A Applied Biosystems gas-phase sequencer, as described in Section 2.2.4.5.

5.2.5 Proteolysis of protoxin and toxin in denaturants.

Protoxin (8 mg/ml), or toxin (4 mg/ml) generated from protoxin by the action of either trypsin (T_t; Fig. 5.1) or papain (T_p; Fig. 5.2), were incubated in 8M urea. The urea solution was buffered at either pH 8.0 (0.1M Tris buffer), pH 9.0 (0.1M carbonate/ bicarbonate buffer) or pH 10.0 (0.1M CAPS buffer). Samples were incubated at 20°C for periods ranging from 15 minutes to 20 hours. ß-mercaptoethanol (0.01%) was added to the samples to ensure that the protoxin molecules were in a reduced state. After a given incubation time, 20 µl of reaction mixture was removed and 5 µl of either activated papain (15 mg/ml), trypsin (25 mg/ml) or elastase (25 mg/ml) was added. After 5 minutes, the

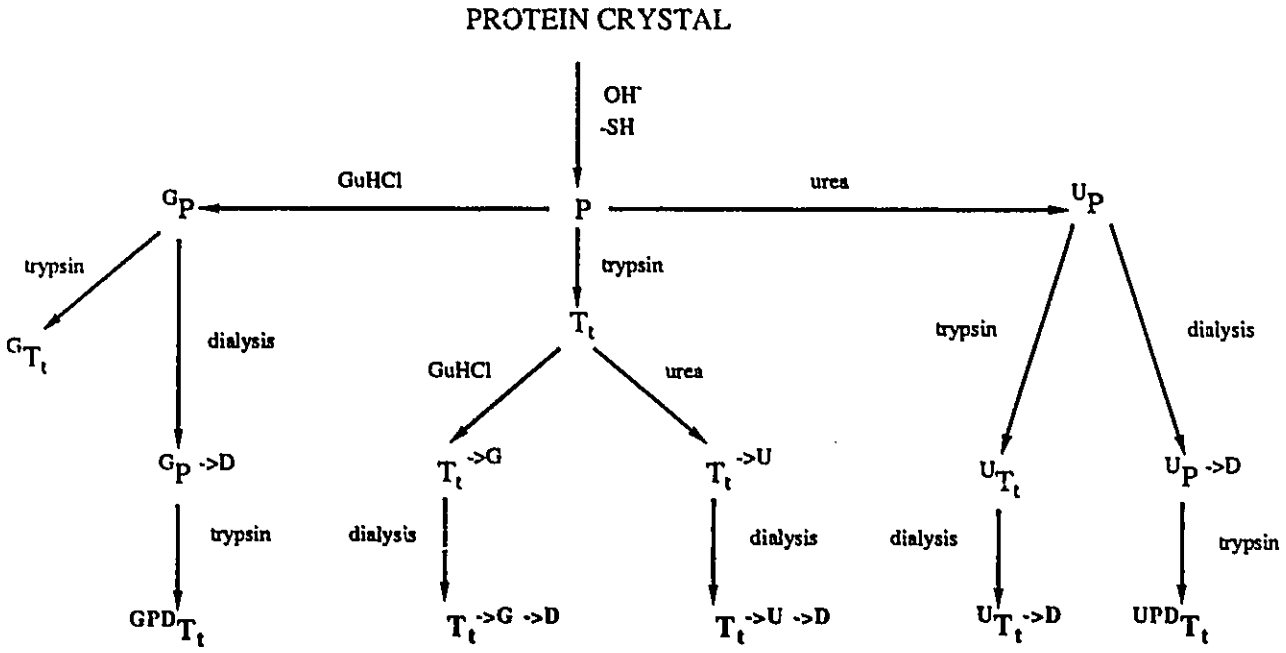


Figure 5.1. Generation of toxins from the protein crystal by trypsin: Protocol for unfolding and refolding of protoxin and toxin from urea and GuHCl.

P: protoxin; ^{G}**P**: protoxin incubated in GuHCl; ^{G}**T$_t$**: toxin generated from protoxin in GuHCl by trypsin; ^{G}**P**$\rightarrow$**D**: dialyzed ^{G}P; **GPDT$_t$**: toxin generated by trypsin from ^{G}P$\rightarrow$D.

T$_t$: native trypsin-generated toxin; **T$_t$**$\rightarrow$**G**: T$_t$ incubated in GuHCl; **T$_t$**$\rightarrow$**G**$\rightarrow$**D**: dialyzed T$_t$$\rightarrow$G; **T$_t$**$\rightarrow$**U**: T$_t$ incubated in urea; **T$_t$**$\rightarrow$**U**$\rightarrow$**D**: dialyzed T$_t$$\rightarrow$U.

^{U}**P**: protoxin incubated in urea; ^{U}**T$_t$**: toxin generated by trypsin from protoxin in urea; ^{U}**T$_t$**$\rightarrow$**D**: dialyzed ^{U}T$_t$; ^{U}**P**$\rightarrow$**D**: dialyzed ^{U}P; **UPDT$_t$**: toxin generated by trypsin from ^{U}P$\rightarrow$D.

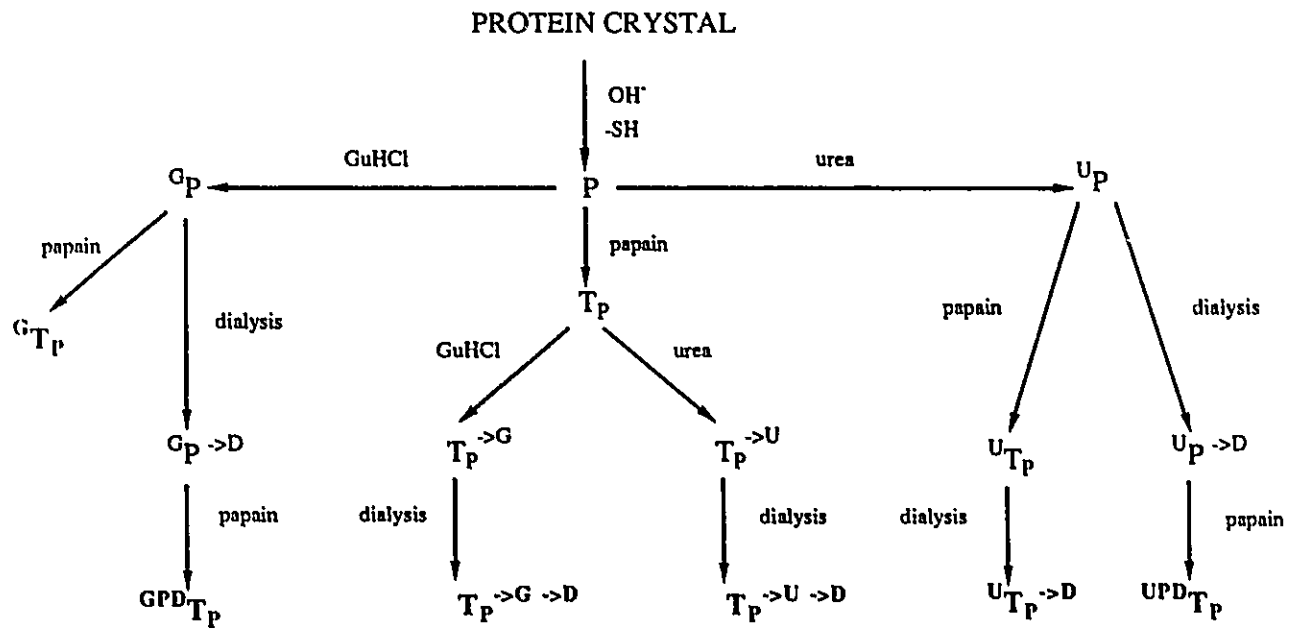


Figure 5.2. Generation of toxins from the protein crystal by papain: Protocol for unfolding and refolding of protoxin and toxin from urea and GuHCl.

P: protoxin; ^{G}P: protoxin incubated in GuHCl; $^{G}T_P$: toxin generated from protoxin in GuHCl by papain; $^{G}P \rightarrow D$: dialyzed ^{G}P; $GPDT_P$: toxin generated by papain from $^{G}P \rightarrow D$.

T_P: native papain-generated toxin; $T_P \rightarrow G$: T_P incubated in GuHCl; $T_P \rightarrow G \rightarrow D$: dialyzed $T_P \rightarrow G$; $T_P \rightarrow U$: T_P incubated in urea; $T_P \rightarrow U \rightarrow D$: dialyzed $T_P \rightarrow U$.

^{U}P: protoxin incubated in urea; $^{U}T_P$: toxin generated by papain from protoxin in urea; $^{U}T_P \rightarrow D$: dialyzed $^{U}T_P$; $^{U}P \rightarrow D$: dialyzed ^{U}P; $UPDT_P$: toxin generated by papain from $^{U}P \rightarrow D$.

inhibitor E64 was added in hundredfold excess (inhibitor:papain) to the papain-containing samples, and the samples were boiled for 2 minutes in SDS sample buffer (2.5% SDS, 5% ß-mercaptoethanol, 1 mM EDTA, 10 mM Tris-HCl, pH 8.3). As noted in Section 2.2.4.6, addition of inhibitor was necessary as papain retains residual activity even under strongly denaturing conditions (Sluterman, 1967). In samples containing trypsin or elastase, proteolysis was stopped after the 5 minute incubation period by adding boiling SDS sample buffer directly to these samples. After 2 minutes at 100°C, the samples were cooled and analyzed by SDS/PAGE.

Proteolysis of protoxin and toxin (T_p) in GuHCl, pH 9.0, was conducted only with papain, as preliminary experiments showed that trypsin and elastase were so rapidly inactivated in GuHCl that very large quantities were required to achieve proteolysis. The procedure used for proteolysis in GuHCl was as above, except that after the 5 minute incubation period with papain, the proteins were precipitated from GuHCl by addition of 200 µl 20% (w/v) TCA. The precipitate was washed three times with water, and then solubilized by boiling in SDS sample buffer. This procedure, which was shown to recover proteins quantitatively from GuHCl, eliminated salt-induced streaking of the proteins on the SDS/PAGE gels.

5.2.6 Gel Electrophoresis.

Polyacrylamide SDS-gels (10-15% gradient) were run on a Pharmacia Phast electrophoresis system and were stained with Coomassie blue as described in Section 2.2.4.1. Densitometric measurements of the toxin band were made using an LKB-2222-010 Ultro Scan XL Laser Densitometer.

5.2.7 Insect cell bioassays.

The *in vitro* biological activities of native and renatured toxins were assessed by means of a lawn assay using the CF-1 cell line, which is derived from larvae of the spruce budworm (Gringorten *et al.*, 1990). The assays were performed in collaboration with Dr.

Luc Peloquin at the Biotechnology Research Institute, Montreal. Briefly, CF-1 cells were suspended in low-temperature melting agar and poured into plates over a layer of high-temperature melting agar. After incubating the plates overnight at 25°C, 1 µl aliquots of twofold serial dilutions of each toxin sample were applied to the surface of the assay plates. The concentration of protein in the first sample in each dilution series was 20 ng/µl, and all sample dilutions were prepared using 0.1M CAPS buffer, pH 10.5. After application of the samples, the plates were incubated for one hour at room temperature, and were then flooded for ten minutes with 0.2% trypan blue in 0.9% MOPS (3-[N-morpholino]propane-sulphonic acid) buffer, pH 7.3. The plates were washed with 0.9% MOPS, and then destained overnight in 0.34% KCl. Relative toxicities of the samples were determined from the threshold levels of toxin required to damage the CF-1 cells. These levels were estimated to lie between the last visible spot coloured by trypan blue, indicating injured cells, and the following unaffected sample spot. All samples were tested in duplicate. The limits of sensitivity of this assay are ± 1 dilution sample (Gringorten *et al.*, 1990).

5.2.8 Proteolysis of the protoxin C-terminal region.

The sequential degradation of the C-terminal half of the protoxin molecule was achieved as described in Choma *et al.* (1990). Protoxin crystals were treated with either 8M urea or 6M GuHCl (containing 0.01% ß-mercaptoethanol and 0.1M carbonate/ bicarbonate buffer, pH 9.0) for periods of 1-6 hours. The samples were then extensively dialyzed in 3.5 kDa cutoff dialysis tubing against water at 4°C for 24 hours. The precipitated samples, or fresh protoxin crystals, were dissolved in 0.5 ml CAPS buffer, pH 10.5, containing 0.01% ß-mercaptoethanol. In each sample, the final concentration of protein was 4-8 mg/ml. Trypsin (0.25 µg), dissolved in 50 µl 0.1M CAPS pH 10.5, was added to each sample. Aliquots (50 µl) were removed at timed intervals, added to boiling SDS sample buffer and analyzed by SDS/PAGE.

5.2.9 Limited proteolysis by papain in SDS.

Protoxin or toxin (0.5 - 1.0 mg) were dissolved in 200 µl of either 0.1M CAPS buffer pH 10.0, or 8M urea containing 0.1M CAPS at pH 10.0; samples in addition contained 0.01% ß-mercaptoethanol. After one hour, activated papain (0.3 mg) was added, followed by SDS sample buffer and subsequent boiling. No papain inhibitor was added prior to boiling the samples; thus, residual active papain proteolyzed the protoxin or toxin as the protein unfolded under these strongly denaturing conditions (see Section 4.3.1).

5.2.10 Fluorescence.

Emission spectra of toxin and protoxin under native and denaturing conditions were collected with a SLM-800C spectrofluorimeter. Samples, contained in 3 ml quartz cuvettes with 1 cm pathlengths, were maintained at 22°C. The samples were excited at 295 nm and emission was monitored at 340 nm. Emitted light was passed through an emission polarizer oriented at 35.3° to the vertical in order to minimize distortions due to Brownian motion (Badeau and Brand, 1979). Corrections were made for the signal from the appropriate blank and the wavelength dependence of the instrument response (Ghiggino *et al.*, 1983). Samples were maintained at pH 9.0 with 0.1M carbonate/bicarbonate buffer; for all fluorescence measurements, the absorbance of the sample at the excitation wavelength was less than 0.1.

To ensure that the cysteine residues in protoxin were in a reduced state, the samples contained 0.01% ß-mercaptoethanol. For both protoxin and toxin, the fluorescence spectrum of the completely unfolded state was estimated from the emission spectrum of a peptic digest of each protein. Toxin or protoxin were each suspended in 10% formic acid to a concentration of 2.5 mg/ml, and porcine pepsin (pepsin:protein, 1:100) was added. After 18 hours at 37°C, the samples were freeze dried, and then dissolved in pH 9.0 0.1M carbonate/bicarbonate buffer for fluorescence measurements.

The effect of rapid dilution on the fluorescence spectra of denatured toxin and protoxin was observed by adding 15 µl of 10 mg/ml protein, incubated in either 6M GuHCl or 8M urea containing 0.01% ß-mercaptoethanol, to 1.5 ml 0.1M carbonate/ bicarbonate buffer, pH 9.0. The buffer was rapidly stirred during sample addition. Fluorescence spectra of the diluted samples were collected immediately, and on an hourly basis thereafter.

5.3 RESULTS

5.3.1 Generation of toxin by papain digestion

A protocol very similar to that followed to generate toxin from protoxin using trypsin can also be used to generate toxin using papain. Papain-generated toxin (T_p; Fig. 5.2) is indistinguishable from trypsin-generated toxin (T_t; Fig. 5.1) on SDS/PAGE gels, and has very similar toxicity against CF-1 cells when tested using the lawn assay (Table 5.1). Amino acid sequencing of T_p gave the N-terminal sequence Gly-Glu-Arg-Ile. This corresponds to residues 26-29 of the gene-deduced amino acid sequence for *B. thuringiensis* subsp. *kurstaki* HD-73 (Adang *et al.*, 1985), and shows that the activation of protoxin by papain removes the first 25 amino acids, as compared to the first 28 which are removed during bovine trypsin activation (Section 2.3.2). The C-terminal cleavage site was not determined, but as T_p has the same apparent molecular mass as T_t, this cleavage site, like the N-terminal site, must also lie very close to that produced by trypsin.

5.3.2 Protease sensitivity of toxin and protoxin in denaturants

Accessibility of peptide bonds to proteolytic enzymes is a useful approach for monitoring unfolding/folding processes in proteins (Ghelis and Yon, 1982). Since toxin is very resistant to further proteolysis by most proteases, including papain (Section 2.3.2), susceptibility to proteolysis should provide a convenient method for following the unfolding of the toxin. The experimental protocol adopted in the present study involved

Table 5.1. Toxicity of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 toxins[a] to CF-1 cells.

Effect of six hour incubation in denaturant on the toxicity threshold level.

Toxin	Threshold Level(ng/μl)[b]
T_t	0.32 - 0.16
$T_t \rightarrow U \rightarrow D$	0.32 - 0.16
$U_{T_t} \rightarrow D$	0.32 - 0.16
UPD_{T_t}	0.32 - 0.16
$T_t \rightarrow G \rightarrow D$	0.64 - 0.32
GPD_{T_t}	0.64 - 0.32
T_p	0.64 - 0.32
$T_p \rightarrow U \rightarrow D$	0.64 - 0.32
$U_{T_p} \rightarrow D$	0.64 - 0.32
UPD_{T_p}	0.64 - 0.32
$T_p \rightarrow G \rightarrow D$	0.64 - 0.32
GDP_{T_p}	0.64 - 0.32

(a) Preparation protocol for each toxin is given in Figs 5.1 and 5.2.

(b) Threshold level lies between the given values.

incubating protoxin or toxin in urea or GuHCl for a specified period of time, then adding the protease for five minutes, followed by boiling SDS sample buffer. Although most proteases are rapidly inactivated in the presence of denaturants, if relatively large quantities (ca. 1:1 mol/mol) are added, activity remains for a sufficient length of time to proteolyze any susceptible peptide bonds (Section 4.3.1). In the case of papain, even boiling SDS buffer is not sufficient to rapidly inactivate the enzyme, and so it was necessary to add the inhibitor E64 to the digest prior to boiling (Sections 2.2.4.6 and 4.3.1). Following incubation in urea or GuHCl and treatment with protease, the amount of toxin remaining was quantified by SDS gel electrophoresis and laser densitometry. Since papain has unusual stability in denaturants, it was the most suitable enzyme to use for quantifying the proteolytic susceptibility of toxin and protoxin under various denaturing conditions. Trypsin and elastase were also used to verify that the results obtained were not due to some peculiar property of papain.

Papain-generated toxin in 8M urea ($T_p \rightarrow U$) was found to be surprisingly resistant to proteolysis by papain. The extent of proteolysis increased at higher pH values and with incubation time (Fig. 5.3; Figs. 5.4B, 5.5B, 5.6B). After one hour of incubation at pH 8.0, most of the toxin remained resistant to proteolysis, and even after 20 hours a significant amount of toxin remained intact after treatment with protease (Figs. 5.3 and 5.4B). While a pH effect was evident, it was not dramatic; after 20 hours incubation in 8M urea at pH 9.0 and even at pH 10.0, some toxin still remained following treatment with protease (Figs. 5.3, 5.5B and 5.6B).

Trypsin-generated toxin in urea ($T_t \rightarrow U$) showed similar resistance towards proteolysis by trypsin (Figs. 5.4D, 5.5D and 5.6D) at pH 8.0, 9.0 and 10.0. Toxin incubated in 8M urea ($T_t \rightarrow U$) at these three pH values, followed by treatment with elastase (Figs. 5.4F, 5.5F and 5.6F) also showed that significant proteolysis is observed only after long incubation in 8M urea, although susceptibility increases as the incubation pH increases.

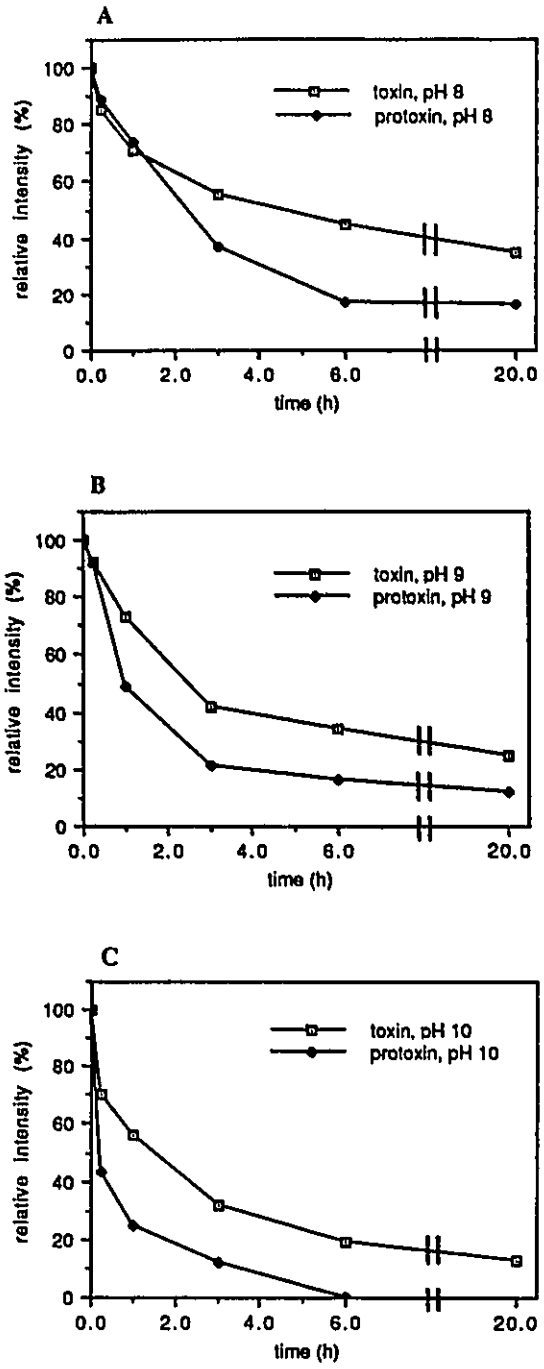


Figure 5.3. The sensitivity to digestion by papain of toxin incubated in 8M urea ($T_p \rightarrow U$) (□), and toxin generated in 8M urea from protoxin ($^{U}T_p$) (◆).

The amount of toxin at each time point was determined by quantifying the Coomassie blue staining intensity of the 67-kDa band. The zero-time sample point for protoxin was determined by quantifying the amount of toxin generated in the absence of denaturant. A): proteolysis conducted at pH 8.0; B): proteolysis at pH 9.0; C): proteolysis at pH 10.0.

Figure 5.4. Protease sensitivity at pH 8.0 of toxin incubated in 8M urea ($T_p \rightarrow U$ and $T_t \rightarrow U$), and of toxin generated from protoxin in 8M urea ($^{U}T_t$ and $^{U}T_p$).

Samples of protoxin or toxin were incubated in 8M urea at pH 8.0 for various time intervals (15 minutes to 20 hours) before addition of protease. The samples were then subjected to SDS-gel electrophoresis followed by staining with Coomassie blue. Proteolysis of protoxin by papain (gel A), trypsin (gel C) and elastase (gel E) is shown. Lane 1 shows the protoxin protein; lane 8 shows the molecular mass markers, with molecular masses indicated in kDa. Lane 2 indicates the amount of toxin generated by the protease from the protoxin sample in the absence of urea. Lanes 3-7 are 15 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations, respectively, of protoxin in 8M urea prior to addition of protease.

Gel (B) shows the corresponding incubations of toxin (T_p) with papain, while gels (D) and (F) show the results of incubating toxin (T_t) in 8M urea followed by addition of trypsin (gel D) or elastase (gel F). In each of the three gels, lane 1 shows the protease alone, and lane 2 is the 0-time toxin sample (no protease added). Lanes 3-7 are 15 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations of toxin in 8M urea prior to addition of protease. Lane 8 shows the molecular mass markers. Note that trypsin and, to a lesser extent elastase, autodigest under the denaturing conditions used.

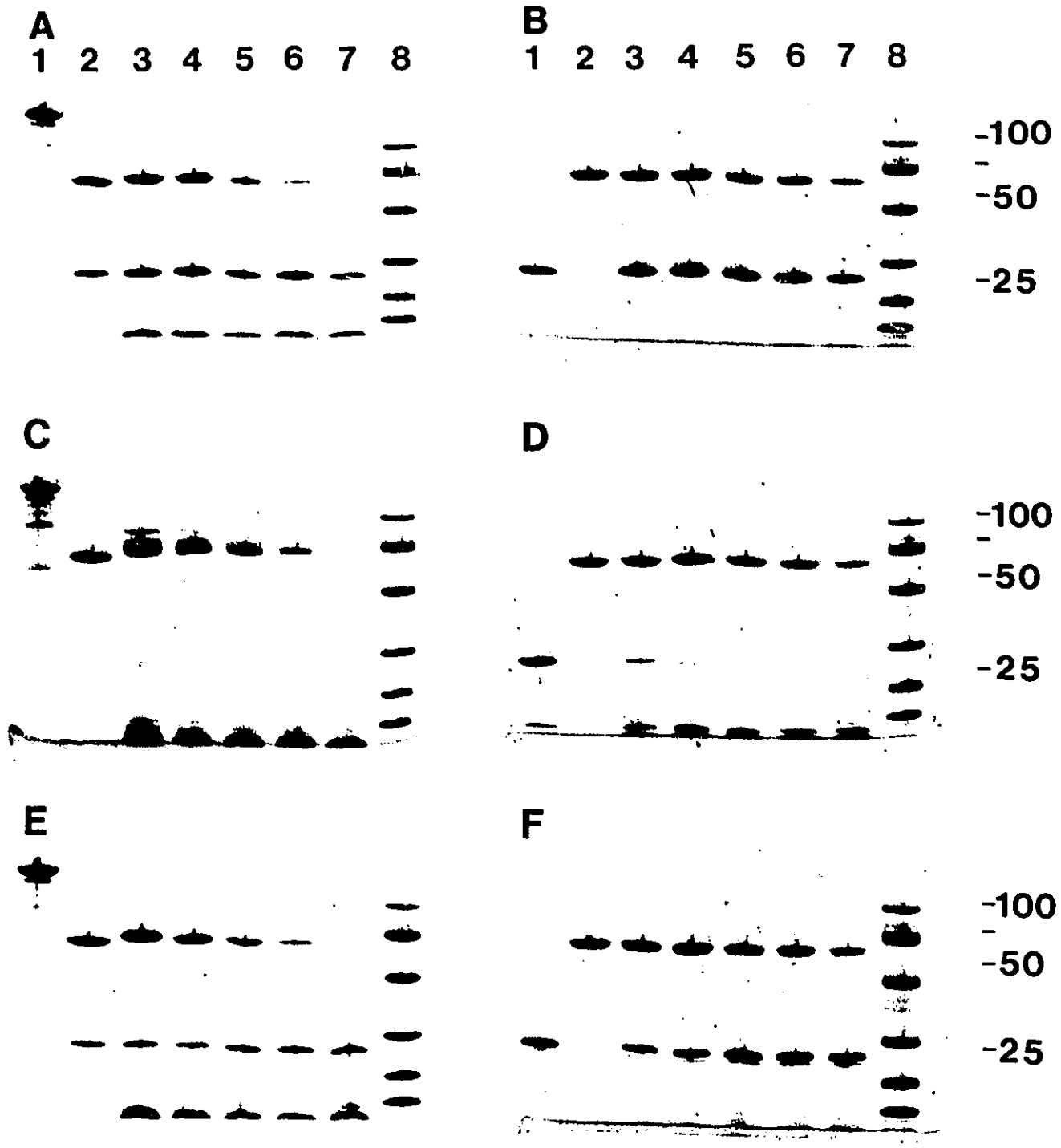
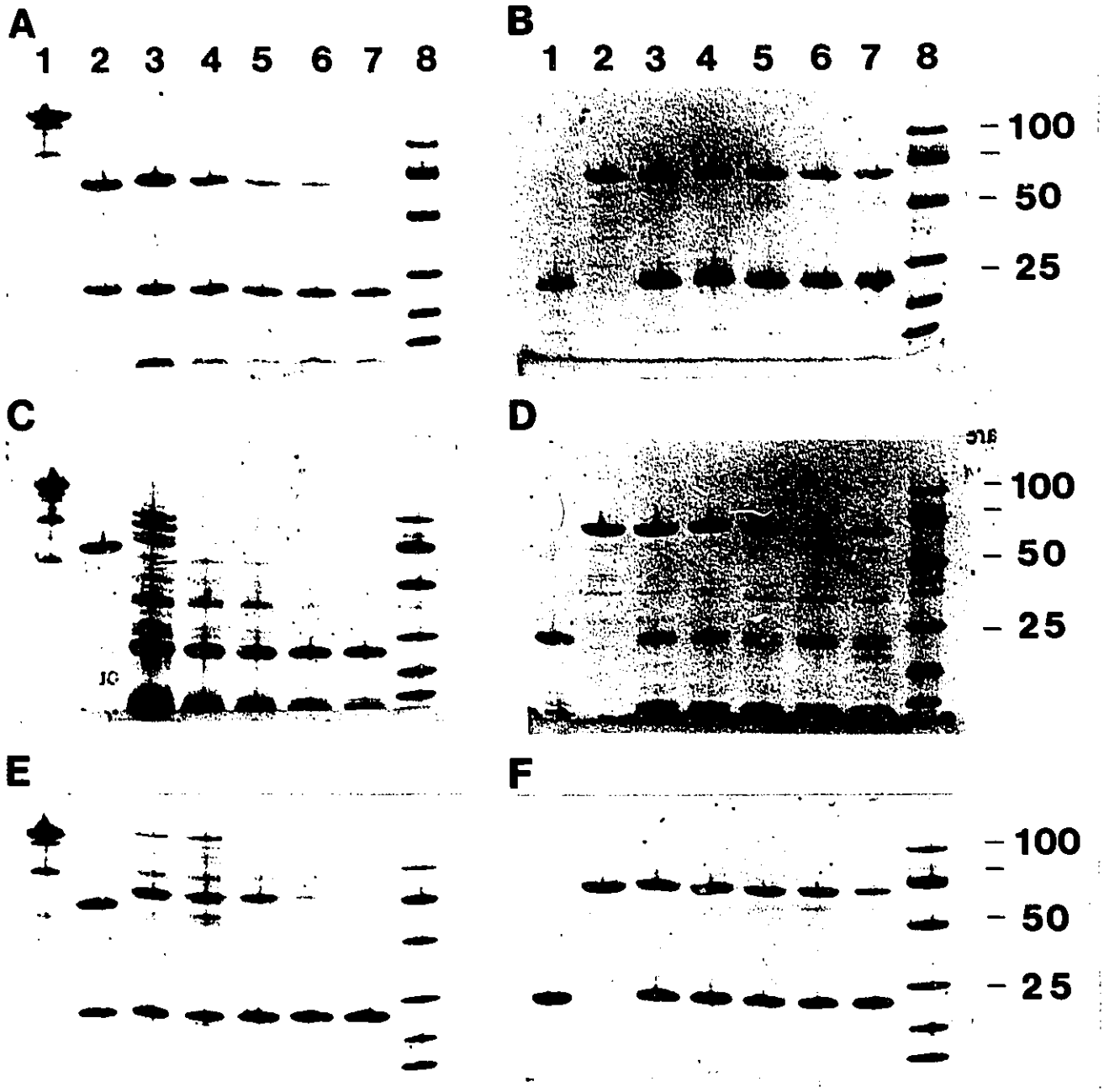

A
1 2 3 4 5 6 7 8
B
1 2 3 4 5 6 7 8
-100
-50
-25
C
D
-100
-50
-25
E
F
-100
-50
-25

Figure 5.5. Protease sensitivity at pH 9.0 of toxin incubated in 8M urea ($T_p \rightarrow U$ and $T_t \rightarrow U$), and of toxin generated from protoxin in 8M urea ($^{U}T_t$ and $^{U}T_p$).

Except for the pH of incubation (pH 9.0 rather than pH 8.0), the incubation conditions and samples used were exactly as for Fig. 5.4. Gels (A), (C) and (E) show protoxin proteolyzed by papain, trypsin and elastase, respectively. Lane 1 shows the protoxin protein, lane 2 indicates the amount of toxin generated by the protease from the protoxin sample in the absence of urea, and lane 8 shows the molecular mass markers. Lanes 3-7 are 15 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations, respectively, of protoxin in 8M urea prior to addition of protease.

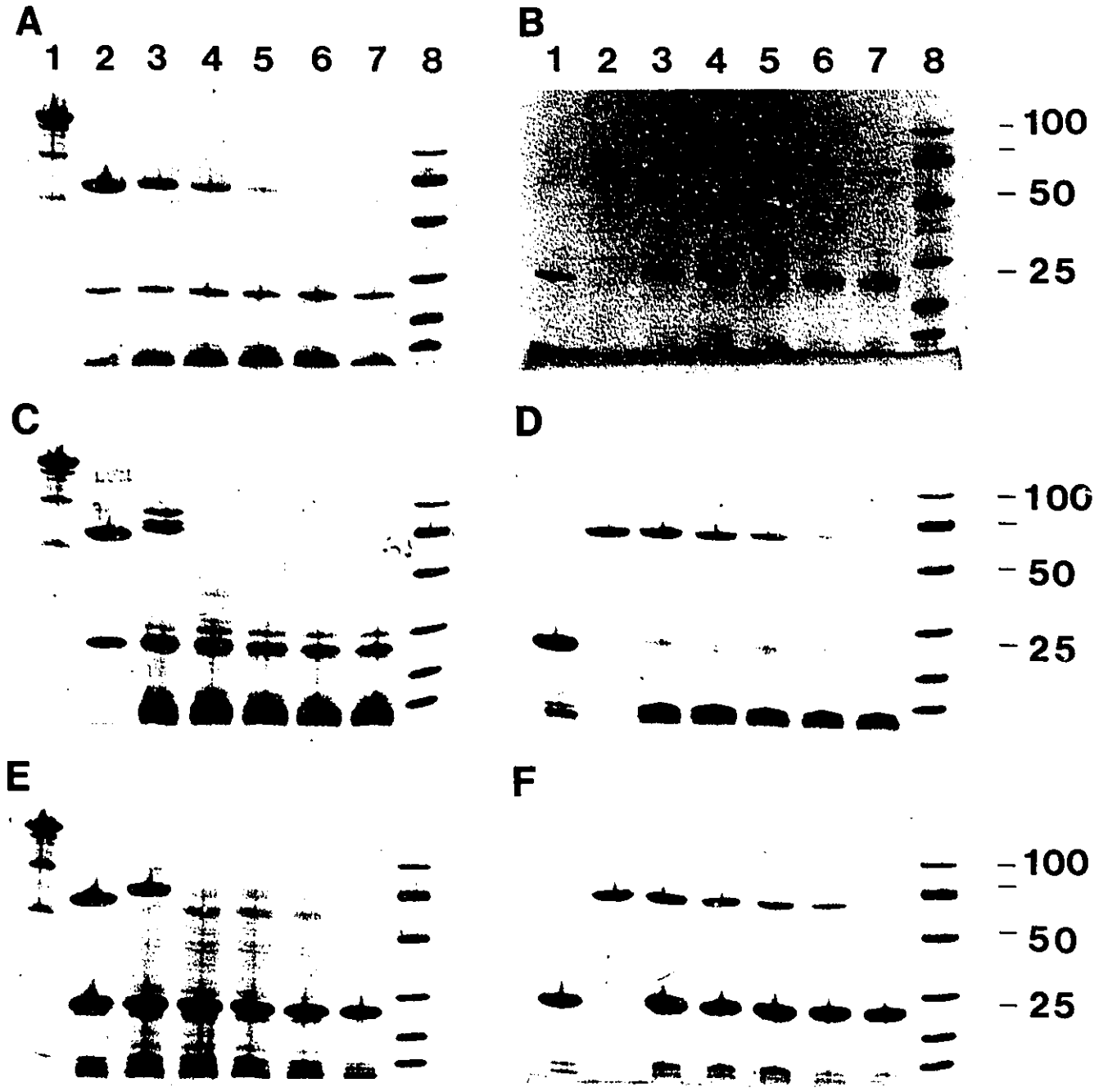
Gel (B) shows the corresponding incubations of toxin (T_p) with papain, while gels (D) and (F) show the results of incubating toxin (T_t) in 8M urea followed by addition of trypsin (gel D) or elastase (gel F). In each of the three gels, lane 1 shows the protease alone, and lane 2 is the 0-time toxin sample (no protease added). Lanes 3-7 are 15 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations of toxin in 8M urea prior to addition of protease. Lane 8 shows the molecular mass markers.



Chapter 5. Folding/unfolding of protoxin **Fig. 5.5**

Figure 5.6. Protease sensitivity at pH 10.0 of toxin incubated in 8M urea ($T_p \rightarrow U$ and $T_t \rightarrow U$), and of toxin generated from protoxin in 8M urea ($^{U}T_t$ and $^{U}T_p$).

Samples and conditions were exactly as for Figs 5.4 and 5.5, except that the 8M urea solution was buffered at pH 10.0 with CAPS. Gels (A), (C) and (E) show the digestion of protoxin in urea with papain, trypsin and elastase, respectively. The remaining gels show the result of proteolysis of T_p with papain (gel B), and of T_t with trypsin and elastase (gels D and F, respectively).



Chapter 5. Folding/unfolding of protoxin **Fig. 5.6**

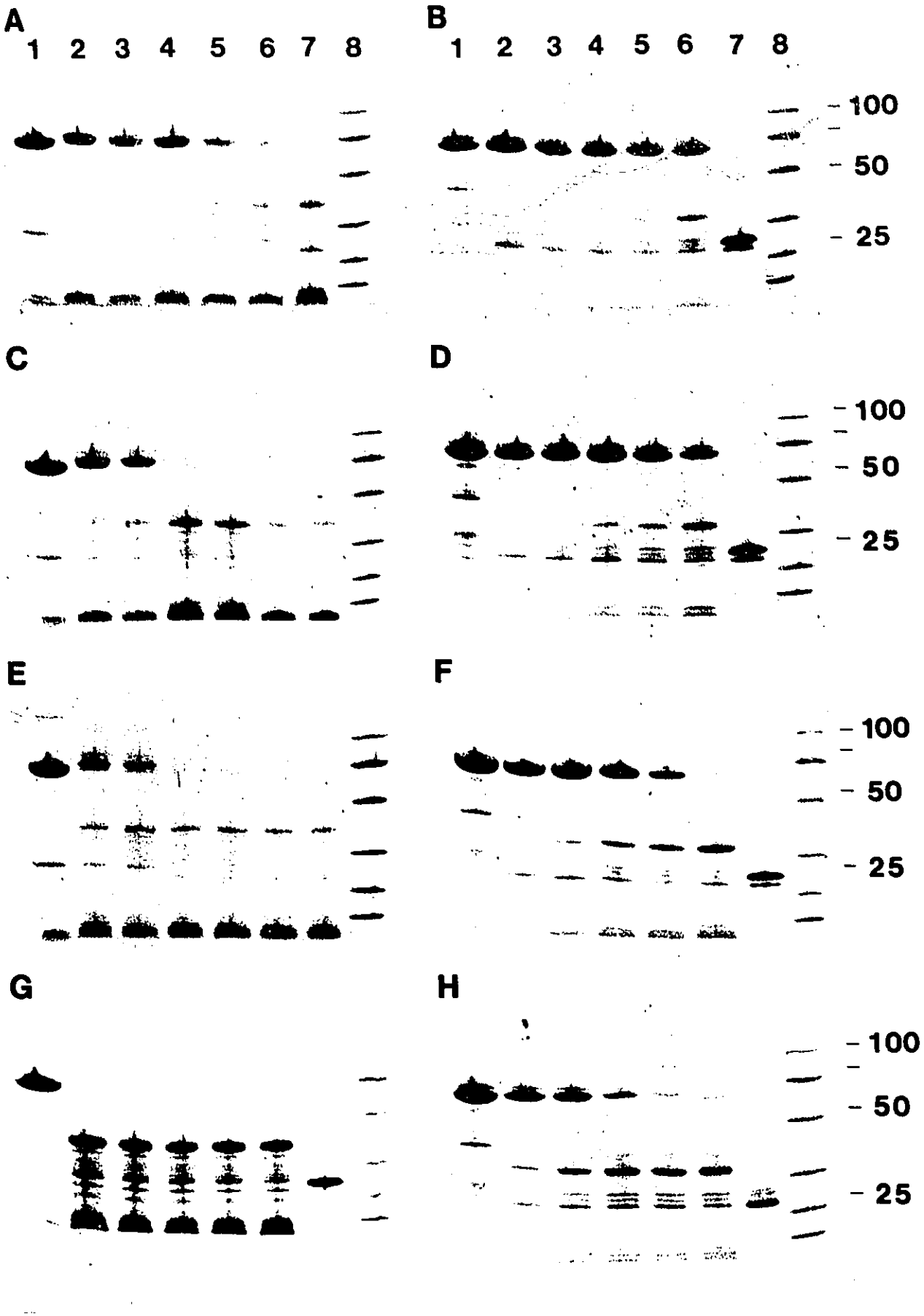
When papain is added to protoxin in 8M urea (^{U}P), a fragment ($^{U}T_p$) is rapidly generated. The remainder of the protoxin molecule is digested by the protease into small peptides. The amount of $^{U}T_p$ generated decreases with increasing pH and incubation time in 8M urea (Fig. 5.3; Figs. 5.4A, 5.5A and 5.6A). Although $^{U}T_p$ has the same apparent molecular mass as T_p (compare lane 3, gel A with lane 2, gel B, Fig. 5.4), its rate of proteolysis by papain is significantly greater than that observed for toxin ($T_p \rightarrow U$) incubated for the same length of time in 8M urea at the same pH (Fig. 5.3; Fig. 5.4A *vs.* Fig. 5.4B; Fig. 5.5A *vs.* Fig. 5.5B; Fig. 5.6A *vs.* Fig. 5.6B). The decreasing amount of $^{U}T_p$ observed with increasing incubation time in 8M urea, and with increasing pH, is not unique to digestion by papain. A similar increased sensitivity to proteolysis is observed when protoxin in 8M urea is digested with trypsin: $^{U}T_t$ is proteolyzed much faster relative to $T_t \rightarrow U$ (Fig. 5.4C *vs.* Fig. 5.4D; Fig. 5.5C *vs.* Fig. 5.5D; Fig. 5.6C *vs.* Fig. 5.6D). The fragment generated by elastase from protoxin in urea likewise shows the same susceptibility to proteolysis (Fig. 5.4E *vs.* Fig. 5.4F; Fig. 5.5E *vs.* Fig. 5.5F; Fig. 5.6E *vs.* Fig. 5.6F).

Increased protease sensitivity is also observed with fragments generated in GuHCl. Toxin-like fragments generated in GuHCl by proteolysis with papain ($^{G}T_p$), and native toxin incubated in GuHCl ($T_p \rightarrow G$), were tested for their susceptibility to proteolysis upon incubation in several concentrations of GuHCl, pH 9.0. The difference in protease sensitivity observed between $^{G}T_p$ and $T_p \rightarrow G$ is much greater than the corresponding difference observed in urea between $^{U}T_p$ and $T_p \rightarrow U$. The pronounced sensitivity of $^{G}T_p$ over $T_p \rightarrow G$ at all concentrations of GuHCl tested is clearly evident from Fig. 5.7. Laser densitometric measurements of the toxin bands shown in Fig. 5.7 permitted the rate of decrease in $^{G}T_p$ and $T_p \rightarrow G$ to be quantified (Fig. 5.8). Native toxin ($T_p \rightarrow G$) appears to be relatively stable to proteolysis in 4.5M GuHCl, and even in 5.0M GuHCl a significant amount remains protease-resistant after 20 hours incubation (Fig. 5.7, gels B and D; Fig. 5.8, A and B). However, proteolysis is more rapid in 5.5M GuHCl, and $T_p \rightarrow G$ is totally

Figure 5.7. Protease sensitivity at pH 9.0 of toxin incubated in GuHCl ($T_p \rightarrow^{G}$) and of toxin generated from protoxin in GuHCl ($^{G}T_p$).

Samples of protoxin or toxin were incubated in GuHCl for various time intervals before the addition of papain. The samples were then subjected to SDS/PAGE, followed by staining with Coomassie blue. Proteolysis of protoxin in 4.5M GuHCl (gel A), 5.0M GuHCl (gel C), 5.5M GuHCl (gel E), and 6.0M GuHCl (gel G) is shown. Lane 1 shows the amount of toxin (T_p) generated by the protease from the protoxin sample in the absence of GuHCl. Lanes 2-7 are 15 minute, 30 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations, respectively, of protoxin in GuHCl prior to the addition of papain. Lane 8 shows the molecular mass markers, with the molecular masses indicated in kDa.

Proteolysis of toxin in 4.5M GuHCl (gel B), 5.0M GuHCl (gel D), 5.5M GuHCl (gel F), and 6.0M GuHCl (gel H) is also shown. Lanes 1-6 are 0 minute, 30 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations, respectively, of toxin in GuHCl ($T_p \rightarrow^{G}$) prior to the addition of papain. Lane 7 shows papain alone, and lane 8 shows the molecular mass markers.



Chapter 5. Folding/unfolding of protoxin Fig. 5.7

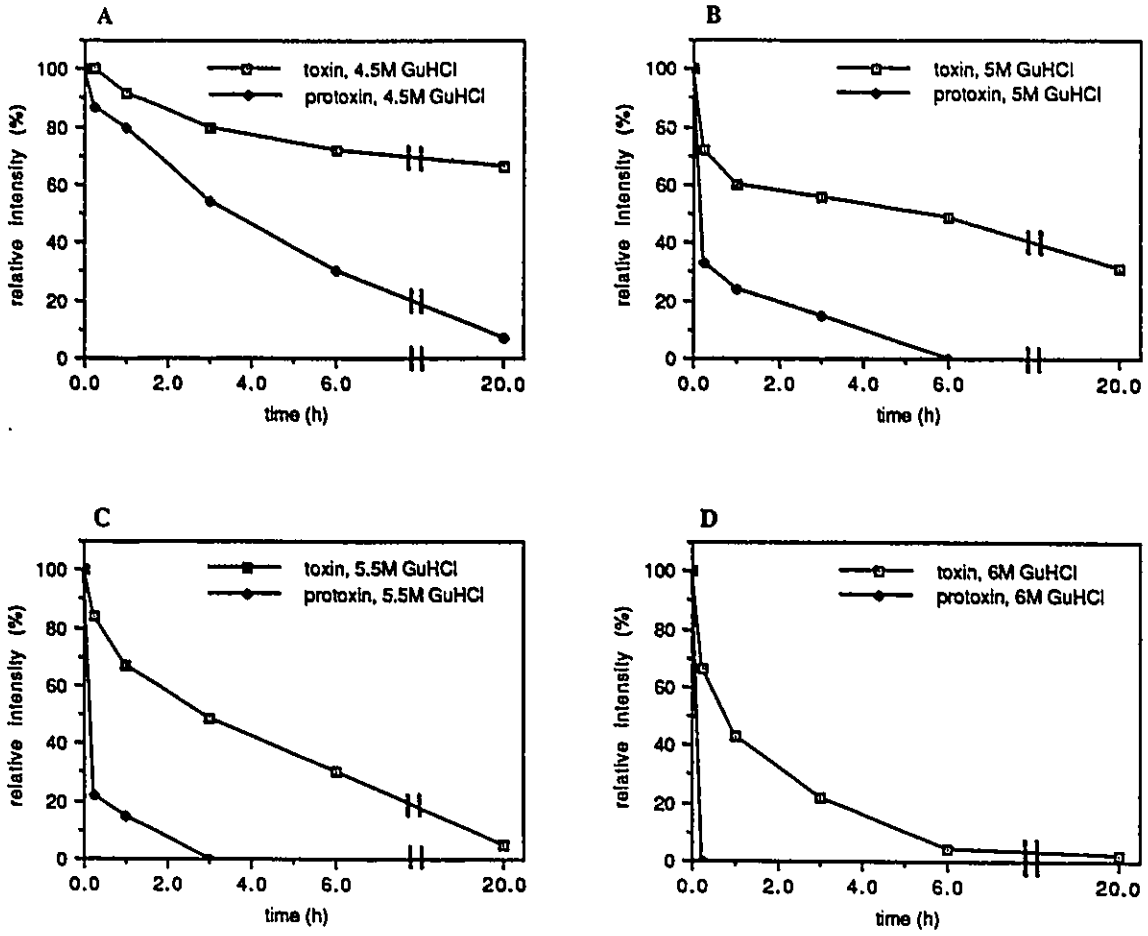


Figure 5.8. Sensitivity to papain digestion at pH 9.0 of toxin incubated in GuHCl ($T_p \rightarrow G$) (□), and toxin generated in GuHCl (^GT_p) (◆).

The amount of toxin at each time point was determined by quantifying the Coomassie blue staining intensity of the 67-kDa band on the gels shown in Fig. 5.7.

proteolyzed after 6 hours in 6M GuHCl (Fig. 5.7, gels F and H; Fig. 5.8, C and D). In marked contrast, the toxin-like fragment generated in GuHCl ($^{G}T_p$) is quite rapidly proteolyzed even in 4.5M GuHCl (Fig. 5.7, gel A; Fig. 5.8A), and is much more rapidly digested as the concentration of GuHCl increases. The difference in protease sensitivity of $^{G}T_p$ and $T_p \rightarrow G$ is clearly illustrated in Figs 5.7 and 5.8; in particular, $^{G}T_p$ is so rapidly digested in 6M GuHCl that none could be visualized on SDS/PAGE gels.

When the protease-sensitive toxins generated in denaturant ($^{U}T_t$, $^{U}T_p$ and $^{G}T_p$) were dialyzed extensively against 0.1M carbonate/bicarbonate buffer, pH 9.0, they become as resistant to proteolysis as the native toxins (T_p and T_t). Their protease resistance was indistinguishable from that of native toxins upon re-addition of denaturant (Figs. 5.9 and 5.10). Similarly, if native toxins are incubated in denaturant and then dialyzed, their original protease resistance is restored (compare Fig. 5.3B and Fig. 5.10A).

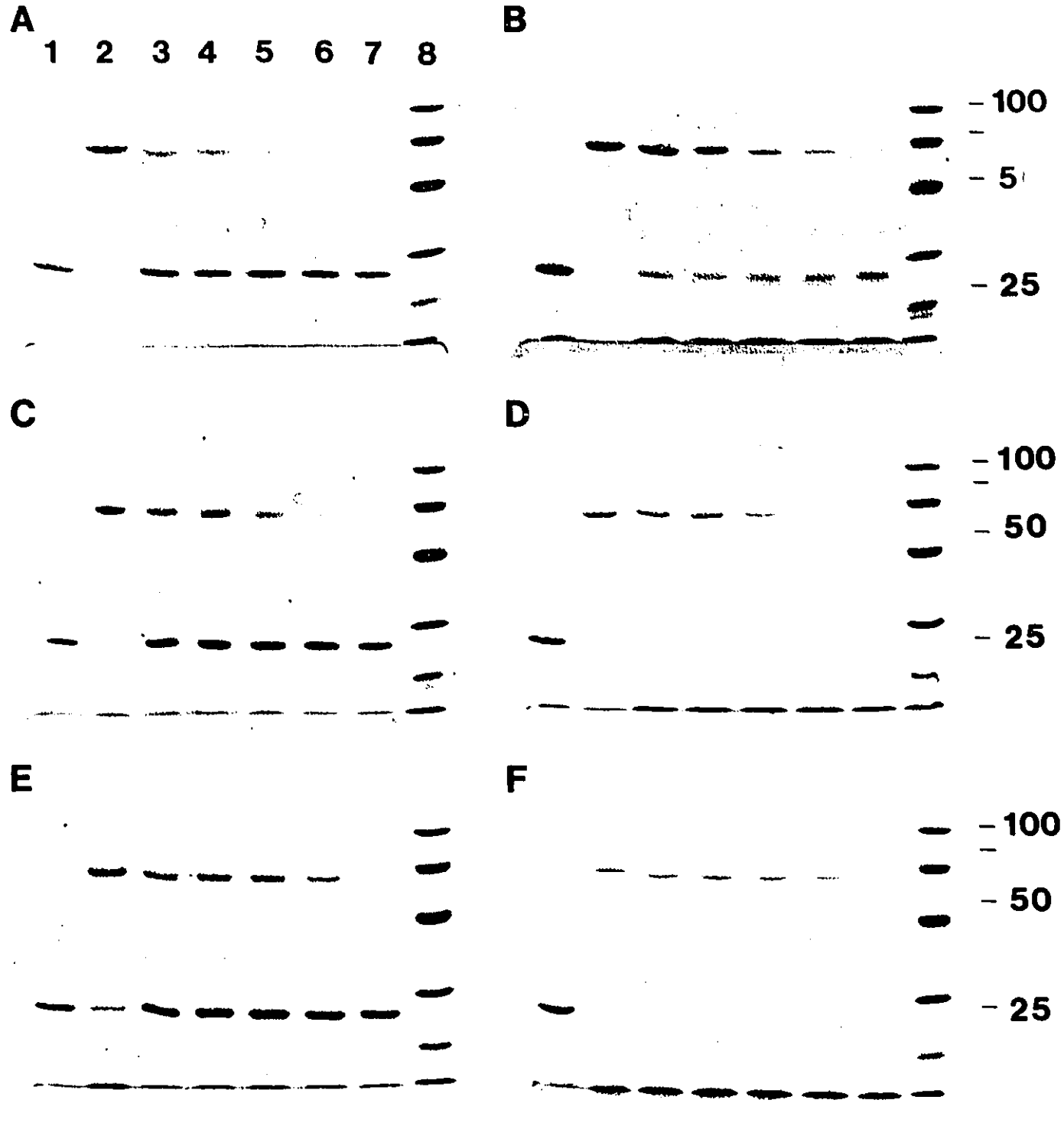
5.3.3 Toxicity assays of denaturant-treated toxins

Native toxins (T_t, T_p), toxins following treatment with denaturant ($T_t \rightarrow U \rightarrow D$, $T_p \rightarrow U \rightarrow D$, $T_t \rightarrow G \rightarrow D$, $T_p \rightarrow G \rightarrow D$), and toxins generated in denaturant ($^{UPD}T_t$, $^{UPD}T_p$, $^{GPD}T_t$ and $^{GPD}T_p$), were thoroughly dialyzed and tested for biological activity. A representative assay plate showing results from the CF-1 cell assay is presented in Fig. 5.11. The threshold levels of the various toxins required to injure CF-1 insect cells were quantified, and are presented in Table 5.1. Native toxins exposed to either 8M urea or 6M GuHCl for six hours, followed by dialysis ($T_t \rightarrow U \rightarrow D$, $T_p \rightarrow U \rightarrow D$, $T_t \rightarrow G \rightarrow D$, $T_p \rightarrow G \rightarrow D$), retained full toxicity. If the corresponding toxins were first generated in the presence of urea ($^{U}T_t$ and $^{U}T_p$) and then dialyzed ($^{U}T_t \rightarrow D$ and $^{U}T_p \rightarrow D$), they also had the same toxicity as native toxin. Similarly, fully toxic 67-kDa proteins ($^{UPD}T_t$, $^{UPD}T_p$, $^{GPD}T_t$ and $^{GPD}T_p$) could be generated from protoxin incubated in 8M urea or 6M GuHCl for six hours, followed by dialysis and addition of papain or trypsin. Within the limits of the sensitivity of this assay (± 1 dilution unit; Gringorten *et al.*, 1990), these results demonstrate that there is no difference in toxicity between any of the toxins tested.

Figure 5.9. Sensitivity to proteolysis of native toxin, and of toxin generated in urea, following dialysis and re-addition of 8M urea.

Toxin was generated by papain (T_p) and extensively dialyzed against 0.1M carbonate/bicarbonate buffer, pH 9.0 (sample I). Toxin was also generated in 8M urea ($^{U}T_p$) after one hour in denaturant, and was then dialyzed against pH 9.0 buffer ($^{U}T_p \rightarrow D$, sample II). In addition, toxin was generated after six hours in 8M urea ($^{U}T_p$), and then dialyzed (sample III). The three toxin samples were then re-dissolved in 8M urea, pH 9.0, and a proteolytic time course was conducted using papain. Gels (A), (C) and (E) show the proteolysis of samples I, II and III, respectively. Lane 1 shows papain alone, lane 2 shows the toxin sample without the further addition of papain (0-time), and lanes 3-7 show 15 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations of the sample in urea prior to the addition of papain. Lane 8 shows the molecular mass markers.

An identical protocol was followed for toxin proteolyzed by trypsin. Toxin was generated by trypsin (T_t) and then dialyzed (sample IV). Toxin was also generated in urea by trypsin after one hour incubation followed by dialysis ($^{U}T_t \rightarrow D$, sample V), or after six hours incubation followed by dialysis (sample VI). These three samples were then re-dissolved in 8M urea, pH 9.0, and a proteolytic time course was conducted using trypsin. Gels (B), (D) and (F) show the proteolysis of samples IV, V and VI, respectively. Lane 1 shows the trypsin marker, lane 2 shows the 0-time sample, and lanes 3-7 show 15 minute, 1 hour, 3 hour, 6 hour and 20 hour incubations of the sample in urea prior to the addition of trypsin. Lane 8 shows the molecular mass markers, with molecular masses indicated in kDa.



Chapter 5. Folding/unfolding of protoxin

Fig. 5.9

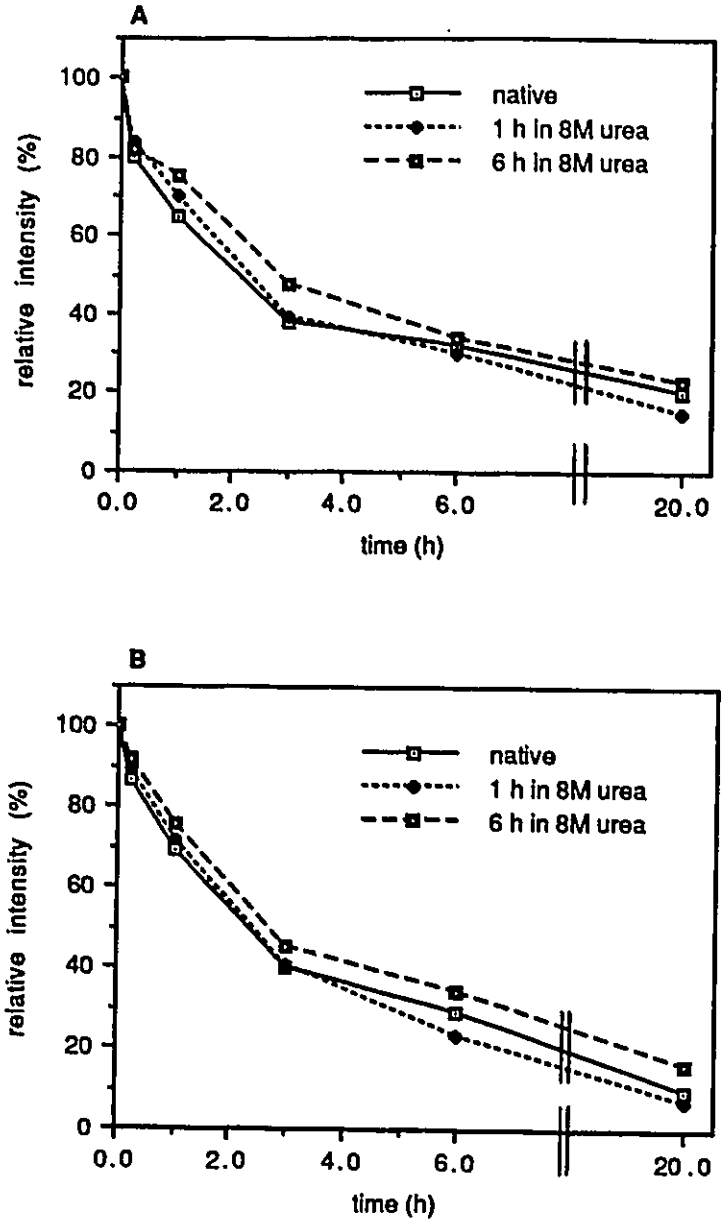


Figure 5.10. Sensitivity to proteolysis at pH 9.0 of native toxin, and of toxin generated from protoxin during incubation in 8M urea.

A: Sensitivity to papain: (i) Native papain-generated toxin, treated in 8M urea for six hours, was dialyzed and then resuspended in 8M urea (⧈); (ii) toxin generated by papain digestion of protoxin after one hour incubation in 8M urea, pH 9.0, then dialyzed ($^{U}T_{p}\rightarrow^{D}$) and re-incubated in 8M urea (◆), and (iii) toxin generated by papain digestion of protoxin after six hours in 8M urea, pH 9.0, then dialyzed ($^{U}T_{p}\rightarrow^{D}$) and re-incubated in 8M urea (□). **B:** Sensitivity to trypsin: procedure as above, except that trypsin was used to generate the native toxin (T_t) and the urea-generated toxin ($^{U}T_{t}\rightarrow^{D}$). The stability of these proteins towards trypsin upon re-addition of 8M urea is shown.

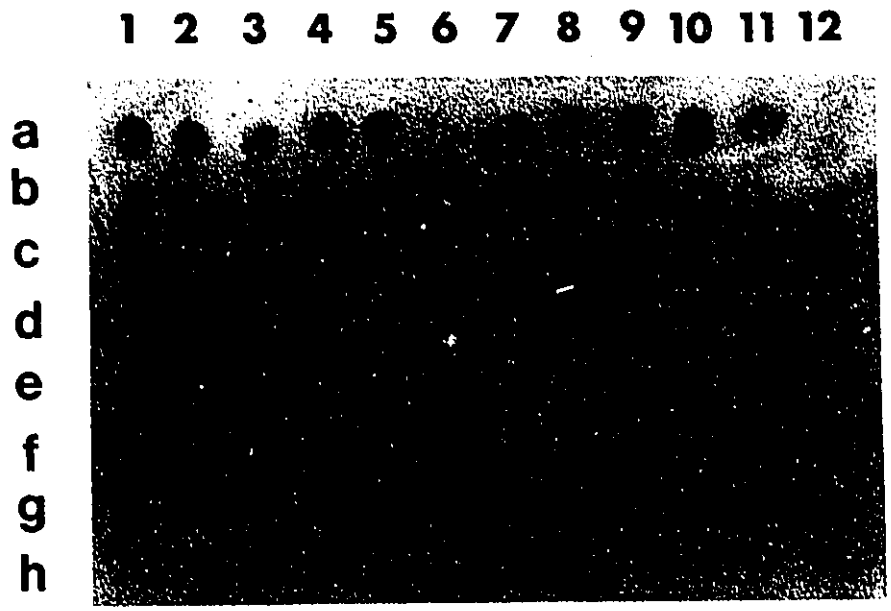


Figure 5.11. The biological activity of toxins treated with denaturants, or generated in denaturants, as assessed using the CF-1 cell lawn assay.
The assay was conducted as described in Section 5.2.7. Eight, two-fold serial dilutions of each sample (1-12) were applied to the plate, with protein concentrations decreasing in the direction a→h. The concentration of protein in each initial sample was 20 ng/μl. The samples assayed on this plate were: (1) T_p; (2) $T_p \rightarrow U \rightarrow D$; (3) $U T_p \rightarrow D$; (4) $UPD T_p$; (5) $GDP T_p$; (6) CAPS buffer blank; (7) T_t; (8) $T_t \rightarrow U \rightarrow D$; (9) $U T_t \rightarrow D$; (10) $UPD T_t$; (11) $GDP T_t$; (12) CAPS buffer blank.

5.3.4 Unfolding and Refolding of Protoxin and Toxin.

5.3.4.1 Limited proteolysis of protoxin and toxin.

The generation of toxin from protoxin has been shown to occur by an unusual proteolytic process in which fragments of approximately 10-kDa are removed in sequence from the C-terminus until the protease-resistant toxin is released (Section 4.4; Choma *et al.*, 1990). Protoxin incubated in 8M urea for 6 hours or in 6M GuHCl for 1 hour, followed by dialysis ($U_P \rightarrow D$ and $G_P \rightarrow D$), exhibited the same sequential proteolysis by trypsin as native crystal protein (Fig. 5.12). The 67-kDa fragments generated ($UPD T_t$ and $GPD T_t$)

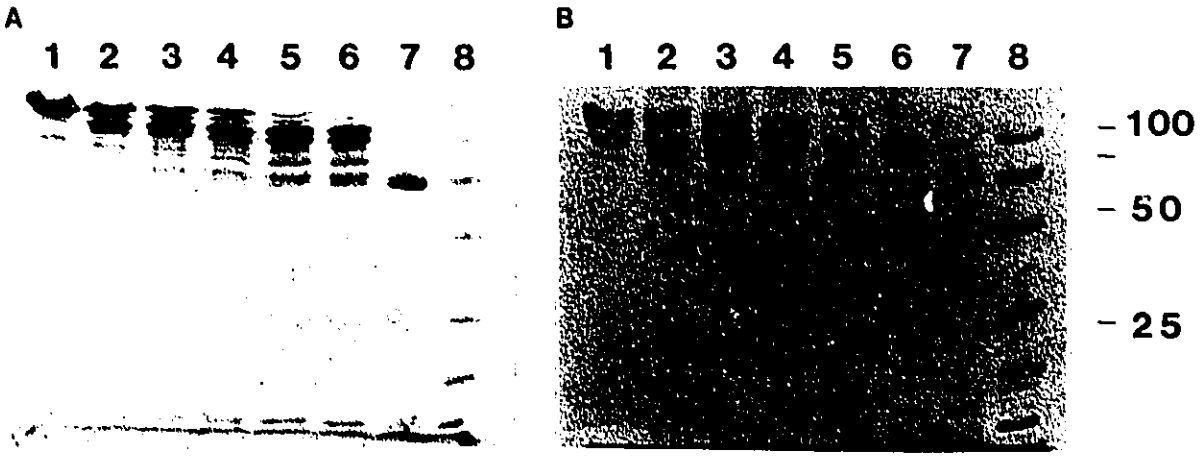


Figure 5.12. Sequential degradation of the C-terminal half of urea-treated protoxin, and of native protoxin.
Protoxin was incubated for six hours in 8M urea, pH 9.0, and was then dialyzed ($^{U}P \rightarrow D$). $^{U}P \rightarrow D$ (gel A) and native protoxin crystal (gel B) were then dissolved in 0.1M CAPS buffer, pH 10.5, and were digested with trypsin as described in Section 5.2.8. Samples were incubated with trypsin for the following lengths of time before being boiled in SDS sample buffer: lane 1, 0 minutes; lane 2, 6 minutes; lane 3, 10 minutes; lane 4, 20 minutes; lane 5, 40 minutes; lane 6, 60 minutes; lane 7, 120 minutes. Lane 8 shows the molecular mass markers, with molecular masses indicated in kDa. Identical results were obtained from protoxin incubated in 6M GuHCl for 1 hour, followed by dialysis ($^{G}P \rightarrow D$).

did not degrade to smaller fragments on the gel, and they were fully toxic against CF-1 cells (Table 5.1). When the protoxin was incubated in 6M GuHCl for 6 hours, the sequential proteolysis was not observed after dialysis, although a protease-resistant toxic fragment was still obtained.

Under controlled denaturing conditions, limited proteolysis of native toxin with papain cleaves peptide bonds in the interdomain regions yielding three fragments (Section 4.3.1). If toxin is incubated in 8M urea, pH 10.0 for three hours ($T_t \rightarrow U$) prior to the addition of papain and SDS buffer, the same three fragments are generated (Fig 5.13).

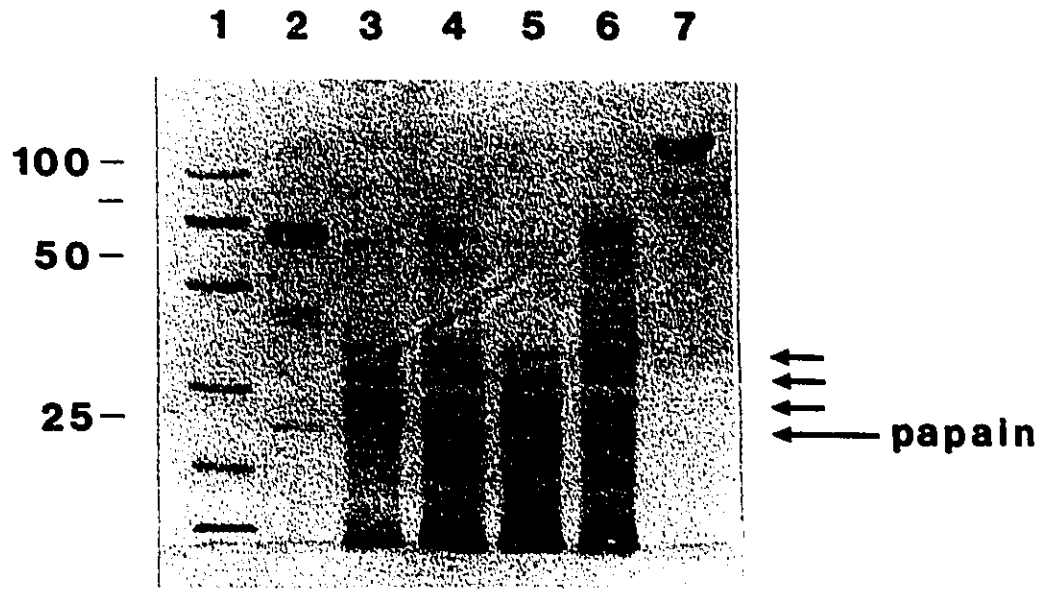


Figure 5.13. Limited proteolysis of toxin and protoxin.
Toxin or protoxin were incubated either in pH 10.0 buffer or in 8M urea, pH 10.0, for three hours. Limited proteolysis of the samples was then conducted as described in Section 5.2.9. Lane 1 shows the molecular mass markers, with molecular masses indicated in kDa; lane 2 is the toxin; lane 3 shows toxin treated in urea for three hours prior to limited proteolysis by papain; lane 4 shows the result of incubating protoxin in urea for 3 hours prior to limited proteolysis; lane 5 is the result of incubating toxin in 0.1M CAPS buffer for three hours before limited proteolysis, lane 6 shows the effect of limited proteolysis on protoxin incubated in CAPS, and lane 7 shows the protoxin alone. The three fragments arising from samples subjected to limited proteolysis are indicated by arrows, and the band corresponding to papain is marked.

When protoxin is incubated either in pH 10.0 buffer, or in 8M urea pH 10.0 (^{U}P) for three hours followed by papain and boiling SDS buffer, the same three fragments are also observed on SDS/PAGE gels (Fig 5.13).

5.3.4.2 Fluorescence emission of toxin.

Figure 5.14A shows the effect of 8M urea on the fluorescence emission spectrum of toxin. On addition of urea, the emission spectrum shifts to longer wavelengths. If toxin incubated for 6 hours in 8M urea ($T_t \rightarrow U$) is diluted rapidly in buffer, its fluorescence spectrum is superimposable on that of the native toxin (spectrum #1, Figure 5.14A). A larger red shift in the emission spectrum is observed when toxin is exposed to 6M GuHCl ($T_t \rightarrow G$; Figure 5.14B). As observed with urea, the emission spectrum of native toxin could be regenerated upon rapid dilution of toxin which had been incubated for 6 hours in 6M GuHCl (spectrum #1, Figure 5.14B).

5.3.4.3 Fluorescence emission of protoxin.

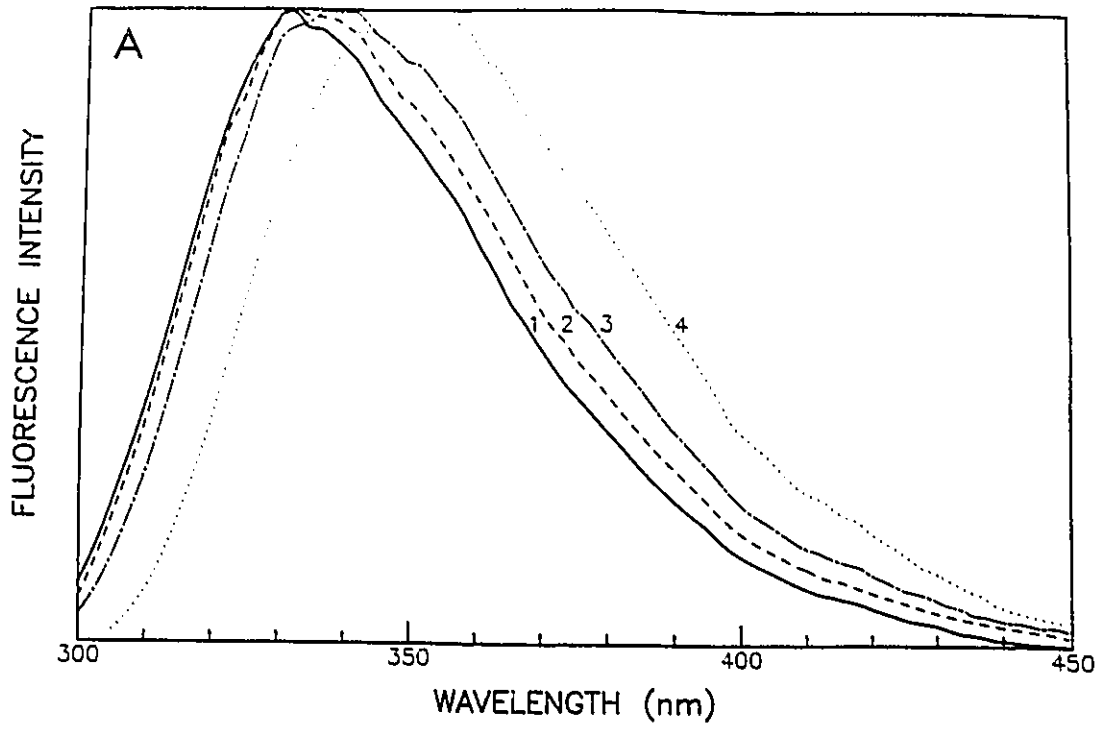
Native protoxin and toxin have very similar fluorescence emission spectra (spectrum #1, Figures 5.14 and 5.15). In the presence of either urea or GuHCl, the emission spectrum of protoxin is shifted towards longer wavelengths than that observed for toxin. As in the case of the toxin, there is a much larger red shift with GuHCl (spectrum #2, Fig. 5.15A *vs.* spectrum #3, Fig. 5.15B). Upon dilution of protoxin incubated in 8M urea for 6 hours, the emission spectrum of native protoxin was rapidly regenerated and was superimposable on that of the native protoxin (spectrum #1, Figure 5.15A). The emission spectrum of native protoxin could be regenerated from protoxin incubated for 1 hour in 6M GuHCl, but not after 6 hours incubation (Figure 5.15B). In the latter case, the emission spectrum remained slightly shifted to longer wavelengths (spectrum #2, Figure 5.15B).

5.3.4.4 Fluorescence maxima of toxin and protoxin in denaturants.

It is clear from Figs. 5.14 and 5.15 that the emission spectra of toxin and protoxin shift to longer wavelengths with increasing time of incubation in denaturant, and that the shift is much more pronounced in GuHCl than in urea. The emission spectra of the two proteins in 8M urea and 6M GuHCl were measured after 15 minutes incubation in the denaturant, and on an hourly basis thereafter. The shift in the emission maxima of the toxin

Figure 5.14. Fluorescence emission spectra of trypsin-generated toxin.

A: 1) native toxin (T_t); 2) toxin incubated in 8M urea for 1 hour ($T_t \rightarrow U$); 3) toxin incubated in 8M urea for 6 hours ($T_t \rightarrow U$); 4) peptic digest of toxin in 8M urea. **B:** 1) native toxin (T_t); 2) toxin incubated in 6M GuHCl for 1 hour ($T_t \rightarrow G$); 3) toxin incubated in 6M GuHCl for 6 hours ($T_t \rightarrow G$); 4) peptic digest of toxin in 6M GuHCl. All samples were at pH 9.0. For both Figs. 5.14A and 5.14B, rapid dilution into pH 9.0 buffer of the samples which gave spectra #2 and #3 resulted in the generation of spectra which were superimposable onto spectrum #1.


A
FLUORESCENCE INTENSITY
1
2
3
4
300
350
400
450
WAVELENGTH (nm)

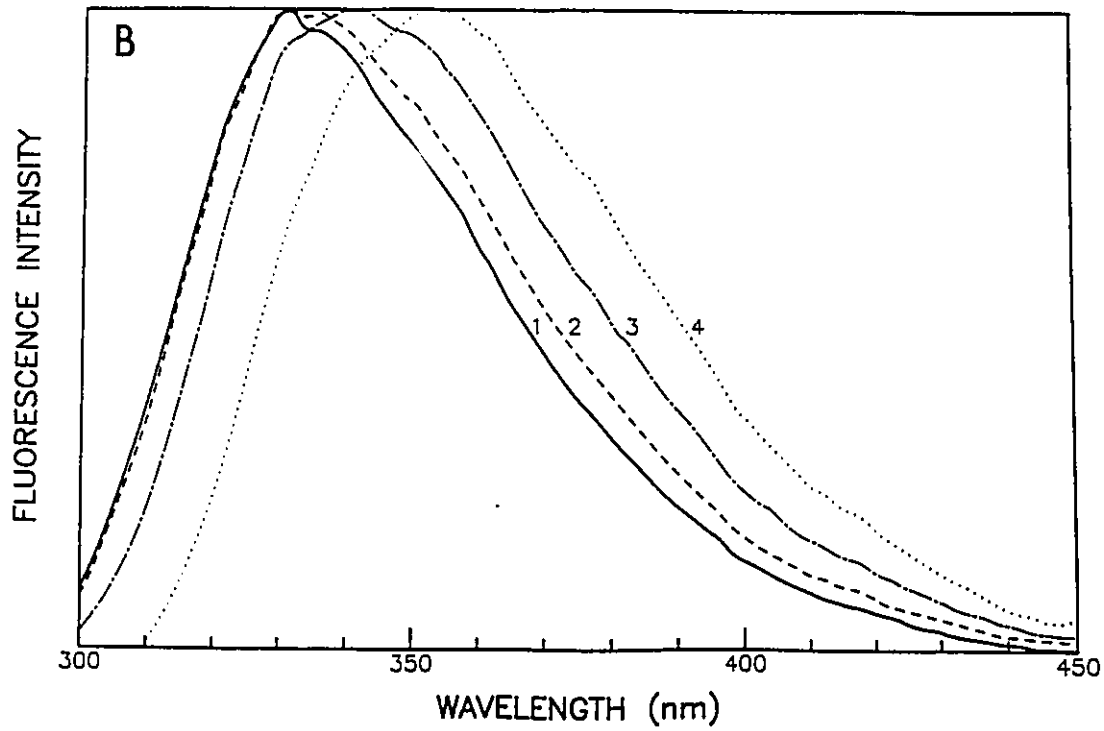
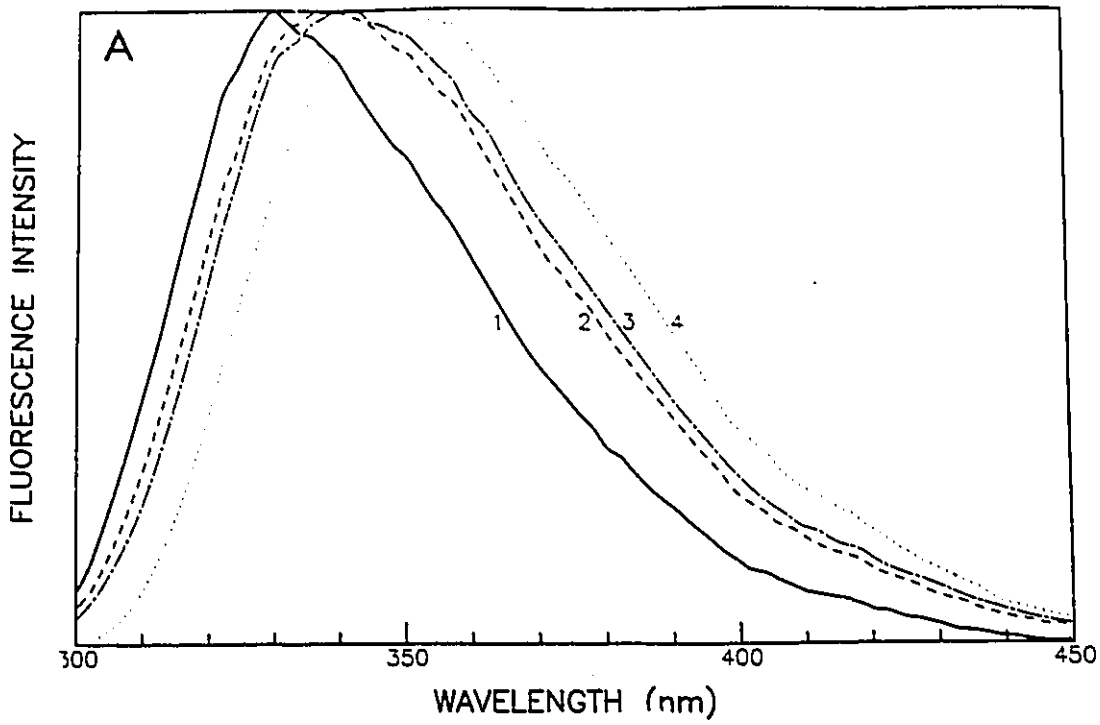
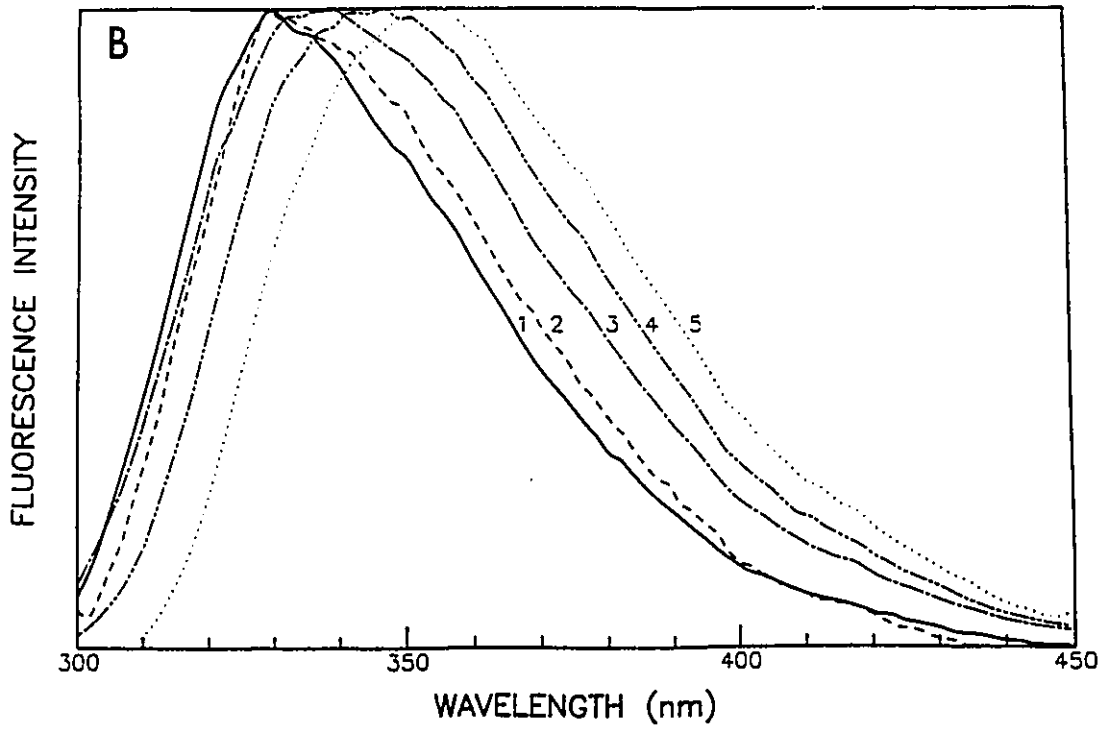

B
FLUORESCENCE INTENSITY
1
2
3
4
300
350
400
450
WAVELENGTH (nm)

Figure 5.15. Fluorescence emission spectra of protoxin.

A: 1) native protoxin (P); 2) protoxin incubated in 8M urea for 1 hour (^{U}P); 3) protoxin incubated in 8M urea for 6 hours (^{U}P); 4) peptic digest of protoxin in 8M urea. Rapid dilution into pH 9.0 buffer of the samples which gave spectra #2 and #3 resulted in the generation of spectra which were superimposable onto spectrum #1. **B:** 1) native protoxin (P); 2) dilution of the sample which gave spectrum #4; 3) protoxin incubated in 6M GuHCl for 1 hour (^{G}P); 4) protoxin incubated in 6M GuHCl for 6 hours (^{G}P); 5) peptic digest of protoxin in 6M GuHCl. Rapid dilution of the sample which gave spectrum #3 resulted in the regeneration of spectrum #1. All samples were at pH 9.0.


A
FLUORESCENCE INTENSITY
1
2
3
4
300
350
400
450
WAVELENGTH (nm)


B
FLUORESCENCE INTENSITY
1
2
3
4
5
300
350
400
450
WAVELENGTH (nm)

spectra with time of incubation in either urea or GuHCl is shown in Fig. 5.16A; the corresponding shifts observed with protoxin under identical conditions are shown in Fig. 5.16B.

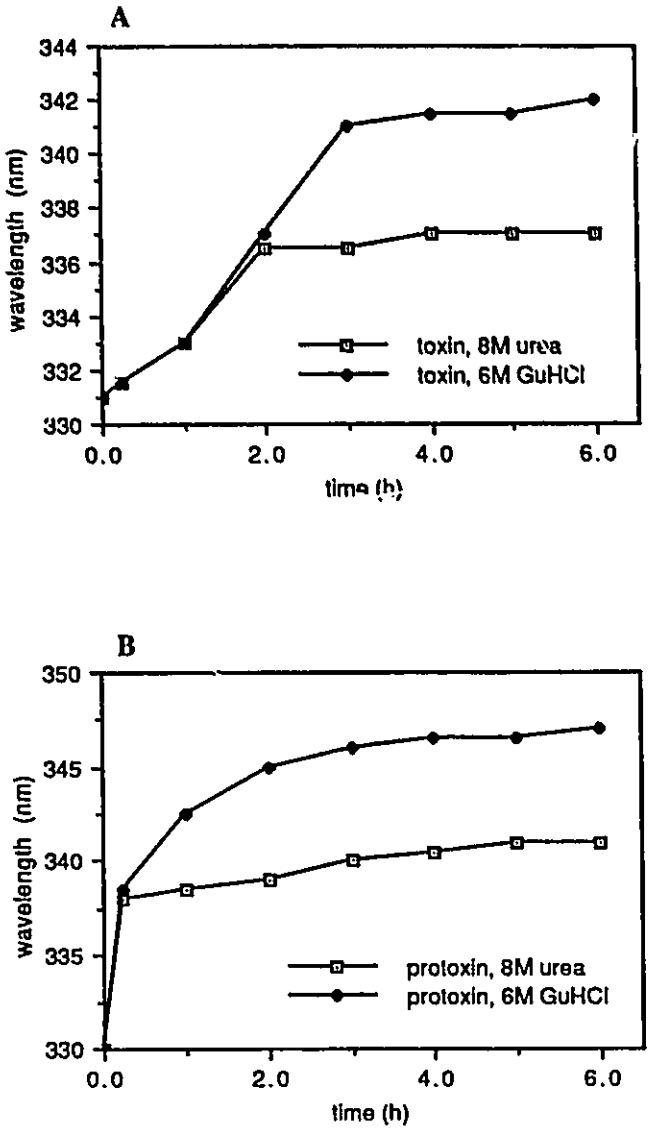


Figure 5.16. Fluorescence emission maxima of toxin and protoxin as a function of incubation time in denaturants.
A: Trypsin-generated toxin in 8M urea ($T_t \rightarrow {}^{U}$) and trypsin-generated toxin in 6M GuHCl ($T_t \rightarrow {}^{G}$). **B:** Protoxin in 8M urea (^{U}P) and protoxin in 6M GuHCl (^{G}P). All samples were at pH 9.0.

5.4 DISCUSSION

The striking asymmetry between the N- and C-terminal halves of the protoxin molecule in terms of amino acid composition and physical properties (Sections 1.2, 3.4 and 4.4) raises the question as to whether these two regions behave independently, or interact in a co-operative manner. Of particular interest is whether the C-terminal half interacts with the N-terminal region and affects the conformation of the toxic moiety. If the two halves of the protoxin molecule unfold and fold independently of each other, this would suggest that the two regions do not interact in a co-operative manner. In addition, if activated toxin cannot unfold and then refold to give a biologically active protein, this would suggest that a primary function of the C-terminal half of protoxin is to ensure the correct folding of the toxic moiety; however, if toxin *can* refold into its active conformation, this would indicate that the C-terminus functions in some other capacity. Unfolding/folding studies therefore provide a direct approach for addressing these questions.

The results shown in Figs. 5.3, 5.4, 5.5 and 5.6 indicate that both papain-generated toxin (T_p) and trypsin-generated toxin (T_t) unfold very slowly in 8M urea. In contrast, the toxin-like fragments ($^{U}T_t$ and $^{U}T_p$) which arise when protoxin is incubated in denaturant and then treated with protease are significantly less stable to proteolysis, indicating that these fragments are more unfolded after a given incubation time in urea than are $T_t \rightarrow U$ and $T_p \rightarrow U$. Similar results are obtained in GuHCl, but the difference in stability between $^{G}T_p$ and $T_p \rightarrow G$ is much greater (Figs. 5.7 and 5.8) than that observed in urea. The increased susceptibility to proteolysis of $^{U}T_t$, $^{U}T_p$ and $^{G}T_p$ indicates that they have a more unfolded conformation in the presence of denaturant than do native toxins incubated in urea or GuHCl ($T_t \rightarrow U$, $T_p \rightarrow U$ and $T_p \rightarrow G$). However, as shown in Fig. 5.9, when the denaturant is removed ($^{U}T_t \rightarrow D$ and $^{U}T_p \rightarrow D$) these toxins become as resistant to proteolysis upon readdition of the denaturant as native toxins ($T_t \rightarrow U$ and $T_p \rightarrow U$). Furthermore, if native toxins are incubated in denaturant and then dialyzed, their

protease resistance is not altered (compare 'native' curve, Fig. 5.10A and 'toxin' curve, Fig. 5.3B).

Native toxins exposed to either 8M urea or 6M GuHCl for six hours, followed by dialysis ($T_t \rightarrow U \rightarrow D$, $T_p \rightarrow U \rightarrow D$, $T_t \rightarrow G \rightarrow D$, $T_p \rightarrow G \rightarrow D$), were found to retain full toxicity (Table 5.1). If the corresponding toxins were first generated in the presence of urea or GuHCl and then dialyzed ($^{U}T_t \rightarrow D$ and $^{U}T_p \rightarrow D$), they also had the same toxicity as native toxin. Similarly, toxins generated from protoxin incubated in 8M urea or 6M GuHCl for six hours ($^{UPD}T_t$, $^{UPD}T_p$, $^{GPD}T_t$ and $^{GPD}T_p$) had the same toxicity. Thus, within the sensitivity limits of the CF-1 cell bioassay (± 1 dilution unit; Gringorten *et al.*, 1990), these results demonstrate that biological activity is regenerated upon removal of denaturant, and that both the toxic moiety in the protoxin molecule and free toxin refold into their active conformations.

Protoxin incubated in 8M urea for six hours or in 6M GuHCl for one hour, followed by dialysis, was found to give the same distinctive proteolytic cleavage pattern of the C-terminal half of the molecule as that obtained from fresh crystal protein (Fig. 5.12). This indicates that, following denaturation, the C-terminal half of the protoxin molecule refolds into its native conformation. However, protoxin incubated for six hours in GuHCl did not yield this characteristic fragmentation pattern upon addition of trypsin. This suggests that prolonged exposure to GuHCl causes some irreversible changes to the C-terminal half of the molecule. These changes do not occur in the N-terminal half, since active toxin can still be generated. It is possible that upon prolonged incubation in GuHCl, some random oxidation of the cysteine residues occurs, thus preventing the regeneration of the native conformation of the C-terminal region of the protoxin.

Evidence that incubation in urea has only a minor effect on the overall conformation of toxin and the toxic moiety in protoxin was shown by limited proteolysis under controlled denaturing conditions. When protoxin or toxin were incubated in pH 10.0 buffer, or in 8M urea pH 10.0 (P, T_t, ^{U}P, $T_t \rightarrow U$) for three hours, followed by addition of papain and SDS

sample buffer, three major fragments were obtained from each sample (Fig. 5.13). As discussed in Section 4.4, these fragments are believed to arise from cleavage of exposed peptide bonds in the interdomain regions of the toxin. Since these same fragments are obtained from toxin after three hours incubation in 8M urea, pH 10.0, this result indicates that while the toxin is made somewhat protease sensitive by treatment with urea at high pH (Fig. 5.3), its integral domain structure remains unaffected. The same three-fragment pattern arises from limited proteolysis of native protoxin, and of protoxin exposed to urea for three hours. This observation indicates that the conformation of the toxic moiety of the protoxin is very similar, if not identical, to that of activated toxin, and that this integral structure remains unaffected by exposure to urea.

The results of fluorescence emission spectroscopy show that toxin unfolds to a much greater extent in GuHCl than in urea (Fig. 5.14), but that it refolds into its native conformation upon removal of denaturant, even after six hours incubation in 6M GuHCl (Fig. 5.14B). If it is assumed that the emission spectrum of a peptic digest of toxin reflects the completely unfolded state, it appears that toxin does not totally unfold in 6M GuHCl (Fig. 5.14B). Protoxin appears to unfold more readily in urea than does toxin, but the native spectrum can be regenerated even after six hours incubation (Fig. 5.15A). The difference in the response of toxin and protoxin to GuHCl is more pronounced, and after prolonged incubation in GuHCl the spectrum of protoxin approaches that of a peptic digest of the protein (Fig. 5.15B). Nonetheless, much of the native structure can apparently be restored to this extensively unfolded protein simply by diluting the denaturant (spectrum #2, Fig. 5.15B). As protoxin exposed to GuHCl for six hours no longer yields the characteristic C-terminal fragmentation pattern, but does produce a toxic protein (Table 5.1), it would appear that the toxic moiety in protoxin refolds to its native conformation, whereas the C-terminal region does not. Therefore, the native fluorescence spectrum is not completely regenerated, but remains slightly shifted to longer wavelengths.

Further evidence obtained from fluorescence spectroscopy supports the conclusion that the C-terminal region of protoxin is more readily denatured than is the toxic moiety or free toxin. A comparison of the shift in wavelength of the fluorescence maximum of toxin and protoxin spectra (Fig. 5.16) reveals an interesting feature of the unfolding process in protoxin, namely, that it appears to consist of two distinct events: a rapid step which is complete within 15 minutes, followed by a slow unfolding over the next 2-3 hours (Fig. 5.16B). In both urea and GuHCl, the fast step results in a similar red shift in the fluorescence maximum. Furthermore, the rate of the slower process in protoxin is very similar to the overall rate of toxin unfolding (Fig. 5.16A). These results suggest that the toxic moiety within the N-terminal half of protoxin unfolds at the same rate as free toxin, and that the initial rapid shift in the fluorescence spectrum of protoxin is due to the unfolding of the C-terminal half of the molecule.

The results presented here show that neither protoxin or toxin are irreversibly inactivated by treatment with denaturants, contrary to earlier reports (Calabrese *et al.*, 1980; Fast, 1981; Huber and Lüthy, 1981; Tyrell *et al.*, 1981). Protoxin readily refolds in dilute solution, suggesting that folding of the molecule precedes crystal formation. The results further show that the C-terminal half of the protoxin molecule (and the extreme N-terminus, which is also removed during activation) are not required for proper folding of the N-terminal toxic moiety. Indeed, the toxic moiety and the C-terminal region appear to fold and unfold independently, behaving as if they were two autonomous proteins fused together. This observation is consistent with the view that the two halves of the molecule have independent functions, and that the primary role of the C-terminal region is for crystal formation.

Prior to the present study, there was little evidence to indicate whether the toxic moiety in the protoxin molecule undergoes a conformational change upon activation. If, as shown here, the C-terminal region apparently serves only to stabilize the crystal and therefore functions quite independently of the N-terminal half, then it is quite possible that

the toxic moiety is present in an active conformation. Protoxin does display some toxic activity toward CF-1 insect cells, but the threshold level is approximately two orders of magnitude higher than it is for toxin (Milne, 1989). The low toxicity of the protoxin molecule may be due to steric hindrance, with the bulky C-terminal half of the molecule preventing interactions between the protein's binding domain and the insect cell binding site.

Although toxin generated in urea is more unfolded than native toxin the amount of unfolding, while significant, is still quite small. This observation may reflect the fact that the activation process is accompanied by a minor conformational change which is prevented under denaturing conditions. However, there is another possible explanation. The present results show that, in the presence of denaturants, the C-terminal half of protoxin is largely or entirely unfolded. Since this half of the molecule is still linked to the C-terminal end of the toxic moiety, it is likely that a partial unfolding of the proximal portion of the N-terminal toxic moiety occurs, decreasing the region's structural stability. While it does not necessarily follow that this would destabilize the entire toxic moiety, the evidence cited in Section 4.1 shows that toxins truncated at the C-terminal end are much less stable molecules (Schnepf and Whiteley, 1985; Andrews *et al.*, 1985; Aronson *et al.*, 1986). The results obtained in this study indicate that no major conformational change occurs on activation of protoxin to toxin. A similar conclusion was reached by Huber-Lukac *et al.* (1983), who found that monoclonal antibodies raised to toxin also bound to protoxin. However, the possibility that some minor conformational change occurs upon activation cannot be ruled out.

The sixteen cysteine residues in protoxin have been shown to form disulphide bridges (Nickerson, 1980; Bietlot *et al.*, 1990). Although most denatured and reduced proteins containing disulphide bridges refold into their native conformations, the conditions required for refolding are generally quite stringent. In a comparison of ten denatured and reduced enzymes, only ribonuclease recovered 100% activity, with most enzymes

averaging 50% recovered activity (Ghelis and Yon, 1982). Considering its molecular mass (130 kDa) and large number of cysteines, it is surprising that denatured protoxin refolds so readily into what appears to be its native conformation. The most likely explanation for the ease with which protoxin refolds is that the sulfhydryl groups reside on the surface of the protein molecule and do not form any intrachain disulphide bridges. The present results therefore provide additional experimental evidence that all sixteen of the cysteine residues of native protoxin are on the surface and form only interchain disulphide bonds (Bietlot *et al.*, 1990).

Two other observations regarding the present results warrant mention. Firstly, there are examples in the literature of precursor proteins which cannot correctly refold once an activation peptide has been removed (Ghelis and Yon, 1982). It is therefore interesting to note that the correct refolding of free toxin does not require information contained in the amino acid sequence of either the extreme N-terminus of the polypeptide, or in the C-terminal half. Secondly, renaturation of protoxin and toxin occurred only when dilute concentrations (0.1 mg/ml or less) of protein were used. At high concentrations, the unfolded proteins tended to aggregate and precipitate from solution in a denatured state. The increased tendency of unfolded proteins to aggregate and the need to use low protein concentrations during renaturation is well documented (Creighton, 1978).

5.5 CONCLUSIONS

The primary structure of the C-terminal portion of the protoxin molecule is highly conserved even though it is not required for toxic activity. Höfte and Whiteley (1989) have suggested that the sequence is conserved because the C-terminal region serves an essential structural role in crystal formation. The results obtained here support this hypothesis. Although it has been postulated that the C-terminus affects the conformation and protease sensitivity of the N-terminal, toxic moiety (Andrews *et al.*, 1986; Brousseau and Masson, 1988), the present results show that the C-terminal region is not required for the correct

folding of the N-terminal region, and that in fact the N-terminal and C-terminal halves of the protoxin molecule fold and unfold independently. It cannot be ruled out that the C-terminal region has a very minor effect on the conformation of the toxic moiety, but all the experimental evidence obtained in this study indicates that the toxic moiety is present in the protoxin molecule in an active conformation.

5.6 REFERENCES

Adang, M. J., M. J. Staver, T. A. Rocheleau, J. Leighton, R. F. Barker and D. V. Thompson (1985). Characterized full-length and truncated plasmid clones of the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73 and their toxicity to *Manduca sexta*. *Gene* **36**, 289-300.

Andrews, R. E. Jr., M. M. Bilbos and L. A. Bulla (1985). Protease activation of the entomocidal protoxin of *Bacillus thuringiensis* subsp. *kurstaki*. *Appli. Environ. Microbiol.* **50**, 737-742.

Andrews, R. E. Jr., R. M. Faust, H. Wabiko and K. C. Raymond (1987). The biotechnology of *Bacillus thuringiensis*. *CRC Crit. Rev. Biotechnol.* **6**, 163-230.

Anfinsen, C. B. and E. Haber (1961). Studies on the reduction and re-formation of protein disulphide bonds. *J. Biol. Chem.* **236**, 1361-1363.

Aronson A. I., W. Beckman and P. Dunn (1986). *Bacillus thuringiensis* and related insect pathogens. *Microbiol. Rev.* **50**, 1-24.

Badeau, M. G. and L. Brand (1979). Time-resolved fluorescence measurements. *Methods Enzymol.* **61**, 378-425.

Bietlot, H., I. Vishnubhatla, P. R. Carey, M. Pozsgay and H. Kaplan (1990). Characterization of the cysteine residues and disulphide linkages in the protein crystal of *Bacillus thuringiensis* subsp. *kurstaki* and *entomocidus*. *Biochem. J.* **267**, 309-315.

Brousseau, R. and L. Masson (1988). *Bacillus thuringiensis* insecticidal crystal toxins: gene structure and mode of action. *Biotech. Adv.* **6**, 697-724.

Calabrese, D. M., K. W. Nickerson and L. C. Lane (1980). A comparison of protein crystal subunit sizes in *Bacillus thuringiensis*. *Can. J. Microbiol.* **26**, 1006-1009.

Choma, C. T., W. K. Surewicz, P. R. Carey, M. Pozsgay, T. Raynor and H. Kaplan (1990). Unusual proteolysis of the protoxin and toxin from *Bacillus thuringiensis*: structural implications. *Eur. J. Biochem.* **189**, 523-527.

Creighton, T. E. (1978). Experimental studies of protein folding and unfolding. *Prog. Biophys. Molec. Biol.* **33**, 231-297.

Creighton, T. E. (1984). *Proteins*. W. H. Freeman and Co., New York, pp. 265-334.

Fast, P. G. (1981). The crystal toxin of *Bacillus thuringiensis*. (in) *Microbial Control of Pests and Plant Diseases* (H. D. Burgess, ed.), Academic Press, London, pp. 223-248.

Ghelis, C. and J. Yon (1982). *Protein Folding*. Academic Press, New York, pp. 562.

Ghiggino, K. P., P. F. Skitton and P. J. Thistlethwaite (1983). Beta-carboline as a fluorescence standard. *J. Photochem.* **31**, 113-121.

Gringorten, J. L., D. P. Witt, R. E. Milne, K. van Frankenhuyzen, P. G. Fast and S. S. Sohi (1990). An *in vitro* system for testing *Bacillus thuringiensis* toxins: the lawn assay. *J. Invert. Path.* **56**, 237-242.

Haider, M. Z., G. P Smith and D. J. Ellar (1989). Delineation of the toxin coding fragments and an insect-specificity region of a dual toxicity *Bacillus thuringiensis* crystal protein gene. *FEMS Microbiol. Lett.* **58**, 157-164.

Hibbard, L. S. and A. Tulinsky (1978). Expression of functionality of alpha-chymotrypsin. Effects of guanidine hydrochloride and urea in the onset of denaturation. *Biochemistry* **17**, 5460-5468.

Höfte, H. and H. R. Whiteley (1989). Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Huber, H. E. and P. Lüthy (1981). *Bacillus thuringiensis* delta-endotoxin: Composition and activation. (in) *Pathogenesis of invertebrate microbial diseases* (E. W. Davidson, ed.) Allenheld, Osmum and Co., Totowa, N.J., pp. 209-234.

Huber-Lukac, M., P. Lüthy and D. G. Braun (1983). Specificities of monoclonal antibodies against the activated delta-endotoxin of *Bacillus thuringiensis* var. *thuringiensis*. *Infect. Immun.* **40**, 608-612.

Jaenicke, R. (1987). Folding and association of proteins. *Prog. Biophys. Molec. Biol.* **49**, 117-237.

Milne, R. (1989). Chemical activation of the delta-endotoxin from *Bacillus thuringiensis* sub. *sotto*. M.Sc. thesis. Laurentian University, Sudbury, Ontario.

Nickerson, W. K. (1980). Structure and function of the *Bacillus thuringiensis* protein crystal. *Biotechnol. Bioeng.* **12**, 1305-1335.

Schnepf, H. E. and H. R. Whiteley (1985). Delineation of a toxin-encoding segment of a *Bacillus thuringiensis* crystal protein gene. *J. Biol. Chem.* **260**, 6273-6280.

Sluterman, L. A. A. (1967). The effect of methanol, urea and other solutes on the action of papain. *Biochim. Biophys. Acta.* **139**, 418-429.

Tanford, C. (1968). Protein denaturation, Parts A and B. *Adv. Protein Chem.* **23**, pp. 121-282.

Tanford, C. (1970). Protein denaturation, Part C. *Adv. Protein Chem.* **24**, pp. 2-97.

Tyrell, D. J., L. A. Bulla, R. E. Andrews, K. J. Kramer, L. I. Davidson and P. Nordin (1981). Comparative biochemistry of entomocidal parasporal crystals of selected Bacillus thuringiensis strains. J. Bacteriol. **145**, 1052-1057.

Wetlaufer, D. B. (1973). Nucleation, rapid folding and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 697-701.

Chapter 6.

CHEMICAL MODIFICATION OF PROTOXIN

Originality

Original thought

is a straightforward process.

It's easy enough

when you know what to do.

You simply combine

in appropriate doses

the blatantly false

and the patently true.

- Piet Hein

Chapter 6.

CHEMICAL MODIFICATION OF PROTOXIN

6.1 INTRODUCTION

A recurrent theme in the preceding chapters is that the N- and C-terminal regions of the protoxin molecule differ significantly in biological function and chemical structure. Functionally, only the N-terminal toxic moiety is necessary for toxicity while the C-terminal region is apparently required solely for crystal formation; structurally, the N- and C-terminal halves are markedly different in secondary structure, hydropathic properties, domain structure and amino acid composition. The present chapter focuses on the unusual distribution of the sulfhydryl and amino groups in protoxin. The effects of chemical modification of these residues on the physical and biological properties of the protein are examined as an approach for elucidating their structural and functional roles in the molecule.

As was discussed in Chapter 1, remarkably little is known about the molecular basis for the action of the *B. thuringiensis* insecticidal protein. Comparison of gene nucleotide sequences together with deletion and truncation studies have provided evidence for the location of the specificity and binding domains in toxin (Section 4.4; Ge *et al.*, 1989; Höfte and Whiteley, 1989). However, there is presently no information on which residues are directly involved in mediating the biological function of the protein. This lack of basic understanding of the structure/function relationship at the amino acid sequence level has precluded the rational application of site-directed mutagenesis as a means of engineering novel proteins with broadened insect host specificities and enhanced field stabilities.

Prior to the advent of site-directed mutagenesis, chemical modification was widely used as an approach for elucidating the roles of amino acid residues in proteins. The technique is not usually specific for residues at predetermined sites in the polypeptide chain: whether a residue is modified or not is determined by its inherent nucleophilicity, its ionization state, and its proximity to the surface of the protein and thus its accessibility to the reagent. All residues of a particular amino acid, or even of a class of functional groups,

may be modified by a reagent. Despite the lack of specificity of this approach compared to site-directed mutagenesis, chemical modification can rapidly provide useful information on potential active sites and structurally important regions of the molecule, and often provides insights into specific amino acid residues which have significant structural or functional roles. The information provided by chemical modification studies can then form the basis of a site-directed mutagenesis strategy for introducing residue-specific modifications into the protein.

The 16 cysteine residues in the protoxin molecule from *B. thuringiensis* subsp. *kurstaki* HD-73 form symmetrical interchain bonds, with each cysteine residue forming a bridge to the corresponding cysteine in an adjoining protoxin molecule (Bietlot *et al.*, 1990). Under the alkaline conditions of the larval gut (Section 1.1.4; Andrews *et al.*, 1987), these bridges are broken and the protein molecule is solubilized. All the cysteine residues and 31 of the 34 lysine residues are removed during the proteolysis of protoxin to toxin (Section 2.4; Bietlot *et al.*, 1989). The unusual distribution of cysteine and lysine residues suggests that they have some functional and/or structural role in the protoxin molecule. Although the sulfhydryl groups apparently play a crucial role in crystal formation (Nickerson, 1980; Höfte and Whiteley, 1989), the role of the lysine residues is less clear. The effects of chemical modification of the sulfhydryl and amino groups in protoxin on the physical and biological properties of the protein are investigated in this chapter.

6.2 MATERIALS AND METHODS

6.2.1 Materials

Aminoethyl-8 reagent and guanidinium chloride were purchased from Pierce Chemical Co., Rockford, Illinois. Deuterium oxide and 2,4-dinitrofluorobenzene (DNFB) were obtained from Aldrich Chemical Co., Milwaukee, Wisconsin. TPCK-trypsin, CAPS buffer salt and L-norleucine were purchased from Sigma Chemical Corp., St. Louis,

Missouri. Precast gels and buffer strips for the Pharmacia Phast electrophoresis system were purchased from Pharmacia Canada Ltd., Dorval, Quebec. Dialysis tubing was purchased from Spectrum Medical Industries, Los Angeles, California. All other chemicals used were high purity reagents obtained from commercial sources. Solutions were prepared with reverse-osmosis quality water purified by the Milli-Q water system.

6.2.2 Preparation of protoxin and toxin

B. thuringiensis subsp. *kurstaki* HD-73 protoxin and toxin were prepared and characterized as described in Sections 2.2.2 and 2.2.3. Chemically modified protoxins were digested with trypsin to yield toxin using the procedure described in Section 2.2.3. Protoxin and toxin were quantified by their absorbance at 280 nm, as outlined in Section 2.2.4.3.

6.2.3 Preparation of protoxin derivatives

6.2.3.1 Carboxymethyl cysteine (CM) and Carbaminomethyl cysteine (CAM) protoxin

The protocols used are slight modifications of the procedures outlined by Means and Feeney (1971); the principal change is that the derivatizations were conducted in the absence of urea. Protoxin (10 mg) was dissolved at 20°C in 10 ml of 0.1M bicarbonate buffer containing 0.1% ß-mercaptoethanol. To help solubilize the protoxin, 2M NaOH was added to bring the solution to pH 10.0. After 5 minutes of rapid stirring, HCl was added to lower the pH to 9.0. The solution was thereafter maintained at pH 9.0 by the addition of base as required.

Carboxymethylation of the cysteine residues was achieved by adding 270 mg of iodoacetic acid, dissolved in 1 ml of 1M NaOH, to the stirring protoxin solution. To carbaminomethylate the sulfhydryl groups, 270 mg of dry iodoacetamide was added to the solubilized protein. Each reaction was stopped after 5 minutes by the addition of 100 µl of ß-mercaptoethanol. The samples were extensively dialyzed against water at 4°C, and were stored in water at 4°C.

6.2.3.2 Succinyl-Carboxymethyl cysteine (SCM) and Succinyl-Carbaminomethyl cysteine (SCAM) protoxin

The protocols used were modifications of procedures given by Means and Feeney (1971); the principal change is that the derivatizations were conducted under non-denaturing conditions. CM or CAM protoxin (10 mg) were dissolved in 10 ml 0.1M carbonate/bicarbonate buffer, pH 9.0. At 1 minute intervals, succinic anhydride was added in 4, 25 mg aliquots to the rapidly stirring sample, and the pH was maintained at 9.0 by the addition of base. The samples were dialyzed at 4°C against three changes of water, and were stored at 4°C in water.

6.2.3.3 Aminoethyl cysteine (AE) protoxin

Aminoethylation of native protoxin was achieved by modifying the procedure of Schwartz *et al.* (1980). Again, the principal modification was that the reaction was conducted under non-denaturing conditions. Protoxin (10 mg) was dissolved in 10 ml of 0.1M CAPS buffer, pH 10.0 containing 0.15% ß-mercaptoethanol. The temperature of the sample was raised to 40°C and maintained there for the duration of the reaction. The sample was stirred rapidly and kept under nitrogen to prevent oxidation of the cysteine residues.

N-(ß-iodoethyl)trifluoroactamide (also called aminoethyl-8 reagent) was added in a 25-fold molar excess over the concentration of -SH in the sample. The reagent (1.1 g) was dissolved in 1 ml of methanol and was added to the stirring sample in two aliquots over a 30 minute period. The sample was maintained at pH 10.0 by the addition of base. The pH of the reaction mixture stabilized after approximately 2 hours, and the sample was then thoroughly dialyzed against water at 4°C.

6.2.4 Quantification of succinylated lysine

6.2.4.1 Rationale

DNFB (2,4-dinitrofluorobenzene) reacts with free amino groups in proteins to form acid-stable derivatives. If a succinylated protein is treated with DNFB and then hydrolyzed in acid, only those lysine residues which were originally succinylated will appear as free lysine upon amino acid analysis. Therefore, the amount of lysine detected will be a direct measure of the extent of succinylation in the original protein.

6.2.4.2 Method

Succinylated protein (1 mg) was dissolved in 1 ml of 6M guanidinium chloride. The sample was saturated with sodium bicarbonate and 25 µl of DNFB (50% in acetonitrile) was added. The sample was protected from light and shaken at 20^{o}C for 4 hours, then a further 25 µl of DNFB solution was added and the sample was shaken for 16 hours. One drop of octanol was then added to reduce foaming and the sample was acidified to pH 2 by the addition of concentrated HCl. Unreacted DNFB was extracted with ether and residual salt was removed by washing the protein pellet several times with water. The sample was dried, hydrolyzed, and the succinyl-lysine content was quantified by amino acid analysis.

6.2.5 Amino acid analysis

Amino acid analysis was conducted as described in Section 2.2.4.3. Carbaminomethyl cysteine is not stable to acid hydrolysis and is converted to carboxymethyl cysteine. Carboxymethyl cysteine was observed to elute immediately before aspartic acid, and aminoethyl cysteine eluted in front of the lysine peak.

6.2.6 Circular dichroism and infrared spectroscopy

Circular dichroism spectra of the derivatives were obtained with a Jasco J600 spectropolarimeter using the procedure described in Section 3.2.4. Proteins were

dissolved in 25mM CAPS buffer, pH 10.5 to a concentration of approximately 0.1 mg/ml; spectra were corrected for differences in sample concentrations. Due to its strong absorbance in the ultra-violet region, no ß-mercaptoethanol was used to help dissolve the native protoxin crystal sample. Infrared spectra of the amide I band of native and derivatized protoxin pellets were recorded with a Digilab FTS-60 instrument. The procedure described in Section 3.2.3 for deuterating and scanning the samples was followed.

6.2.7 pH-dependent solubility of derivatives

The procedure outlined in Section 2.2.4.4 was followed; the only modification was that the buffer for the native protoxin sample contained 0.01% ß-mercaptoethanol. The absorption coefficients of the derivatives were determined by amino acid analysis to be very similar to that of native protoxin: at a concentration of 1 mg/ml, protoxin gives an absorbance of 1.37 for a 1 cm pathlength at 280 nm. This value was used to calculate the concentration of protoxin (both native and derivatives) in the pH range 5-11.

6.2.8 Proteolysis of protoxin

The sequential degradation of the C-terminal half of both native and chemically-modified protoxins was observed using the procedure described previously (Section 5.2.8; Choma *et al.*, 1990).

6.2.9 Gel electrophoresis

SDS polyacrylamide gels (10-15% gradient) were run on a Pharmacia Phast electrophoresis system and were stained with Coomassie blue as described in Section 2.2.4.1.

6.2.10 Cell bioassays

The biological activities of toxins generated by tryptic digestion of the chemically modified protoxins were quantified using CF-1 insect cells in a lawn assay, as described

previously (Section 5.2.7; Gringorten *et al.*, 1990). Samples were diluted to 50 ng/µl in 0.1M CAPS buffer, pH 10.5, and twofold serial dilutions were applied to the assay plates. Threshold levels were estimated to lie between the last visible spot coloured by trypan blue, indicating injured cells, and the following unaffected sample spot.

6.2.11 Insect bioassays

The toxicities of the protoxin derivatives towards eastern spruce budworm (*Cholistoneura fuminiferana*) were determined by insect bioassays conducted by Ms. C. Badau of the Forestry Pest Management Institute, Sault Ste. Marie, Ontario. Day-1 sixth instar larvae were force-fed with 2 µl of either native crystal protoxin or derivatized protoxin dissolved in 0.2M CAPS, pH 10.5. Ten-fold dilutions of protoxin, starting at 1000 µg/µl were prepared; the concentration of protein in each original sample was determined by absorbance at 280 nm, and the samples were diluted accordingly. The toxicity of each dilution was determined by force-feeding 15 larvae. After treatment, the larvae were held at 25°C on a 16 hour photoperiod and mortality was scored after 10 days. Once the approximate concentration (within an order of magnitude) of each protoxin derivative required to kill spruce budworm larvae had been determined by this method, the procedure was repeated using two-fold serial dilutions spanning the appropriate concentration range. On average, six two-fold serial dilutions were each fed to 35 larvae and mortality (LD_{50}) was determined after 10 days. The statistical reliability of the data was analyzed by the *Polo-PC* probit analysis programme, obtained from LeOra Software, Berkeley, California.

6.3 RESULTS

6.3.1 Extent of modification of sulfhydryl and amino groups

The extent of chemical modification of cysteine residues in the CM, CAM and AE protoxin derivatives, and of the lysine residues in the SCM and SCAM derivatives, is

summarized in Table 6.1. Amino acid analysis was conducted on at least three preparations of each derivative; the mean value for the extent of derivatization is reported. It is clear from Table 6.1 that, within the limits of experimental error for amino acid analysis, all the cysteine residues are derivatized in the CM, CAM and AE protoxins. In contrast, only about 35% of the lysine residues are succinylated in the SCM and SCAM protoxins.

Table 6.1 Extent of chemical modification of cysteine and lysine residues in protoxin derivatives.

Derivative	# modified residues*	percent modified residues
CM	15.7 ± 1.2 cysteines	98
CAM	16.4 ± 0.8 cysteines	102
AE	15.7 ± 1.1 cysteines	98
SCM	11.6 ± 1.4 lysines	34
SCAM	12.6 ± 1.7 lysines	37

* average amino acid analysis from three preparations.

6.3.2 Solubility of derivatized protoxins

The effect of derivatization of cysteine and lysine groups on the solubility of protoxin was determined for the pH range 5-11 (Fig. 6.1). All the derivatives except AE protoxin were found to be more soluble than native protoxin below pH 8, whereas above pH 9 native protoxin (in the presence of 0.01% ß-mercaptoethanol) was slightly more soluble than any of the derivatized protoxins.

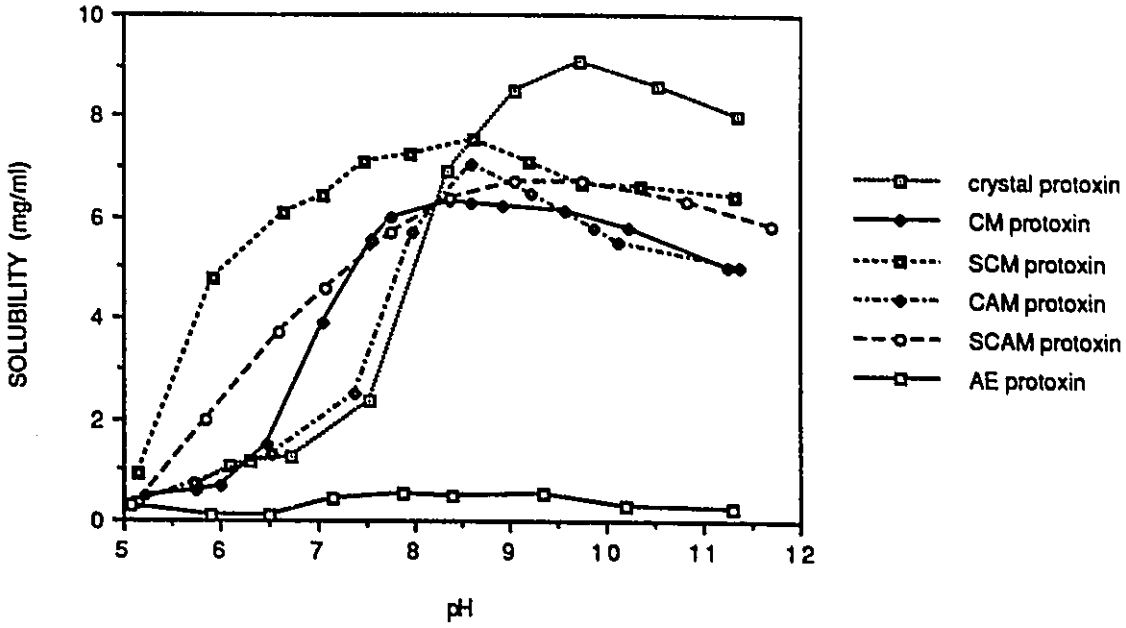


Figure. 6.1 The solubilities of native crystal protoxin and derivatized protoxins as a function of pH.

The solubilities (mg/ml) of native crystal protoxin and the protoxin derivatives over the pH range 5-11 was determined. The native protoxin sample contained 0.01% ß-mercaptoethanol.

6.3.3 Circular dichroism and infrared spectroscopy of protoxin derivatives

Circular dichroism and infrared spectroscopy were conducted on the native and derivatized protoxins in order to determine if the chemical modifications caused gross changes in overall structure. Circular dichroism spectra of the derivatives were similar to the spectrum for native protoxin, but some variations in the 200-230 nm region were observed (Fig. 6.2). Although the shape of the spectra were very similar, the intensity of the curves obtained from the derivatives differed from that of the native crystal protein. Conformational changes in the derivatized protoxins are also suggested by differences in shape and intensity maxima of the infrared spectra of the modified proteins (Fig. 6.3).

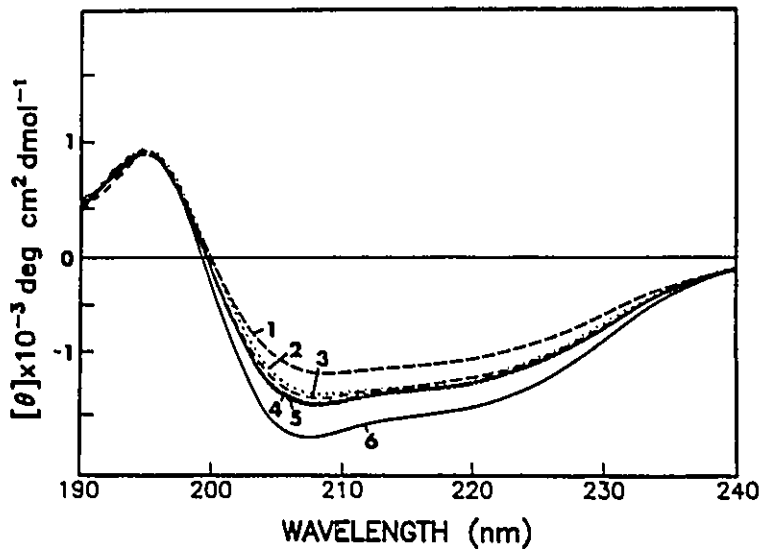


Figure. 6.2 Far ultra-violet circular dichroism spectra of native and derivatized protoxins.
Spectra were normalized for differences in concentration. #1: AE protoxin; #2: SCM protoxin; #3: SCAM protoxin; #4: CAM protoxin; #5: CM protoxin; #6: native protoxin.

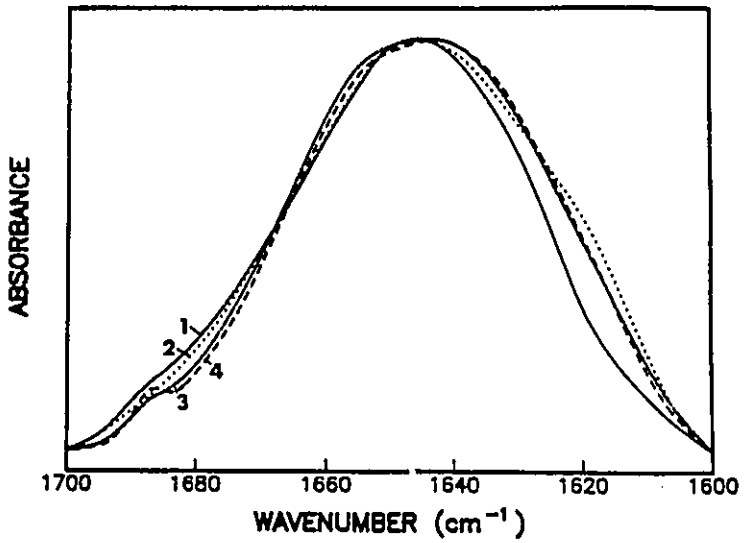


Figure. 6.3 Infrared spectra of the amide I bands from native and derivatized protoxins.
Protein pellets were deuterated for 72 hours prior to scanning. The spectra shown correspond to the following proteins: #1: native crystal protoxin; #2: AE protoxin; #3: CAM protoxin; #4: CM protoxin. The spectrum from SCAM protoxin was superimposable on the spectrum obtained from CAM protoxin (spectrum #3), and the amide I band from SCM protoxin was identical to that obtained from CM protoxin (spectrum #4).

6.3.4 C-terminal degradation of protoxin derivatives

As previously discussed (Section 4.4; Choma *et al.*, 1990), fragments of approximately 10 kDa are removed by trypsin in a sequential manner from the C-terminal half of protoxin; the process ends with the 67-kDa protease-resistant toxin being produced. Upon exposure to trypsin, CM, CAM, AE, SCM and SCAM protoxins were each found to give rise to a sequential fragmentation pattern similar to that obtained from native protoxin (Fig. 6.4). Proteolysis of all the derivatives ended in the production of a protease-resistant 67-kDa fragment.

6.3.5 Toxicities of the chemically modified protoxins

The results of CF-1 cell toxicity lawn assays (Section 5.2.7; Gringorten *et al.*, 1990) on the protease-resistant 67-kDa fragment from each of the protoxin derivatives are presented in Table 6.2. These results show that toxins generated from all the derivatives except AE protoxin are as toxic to CF-1 cells as is native toxin. The results of force-feeding

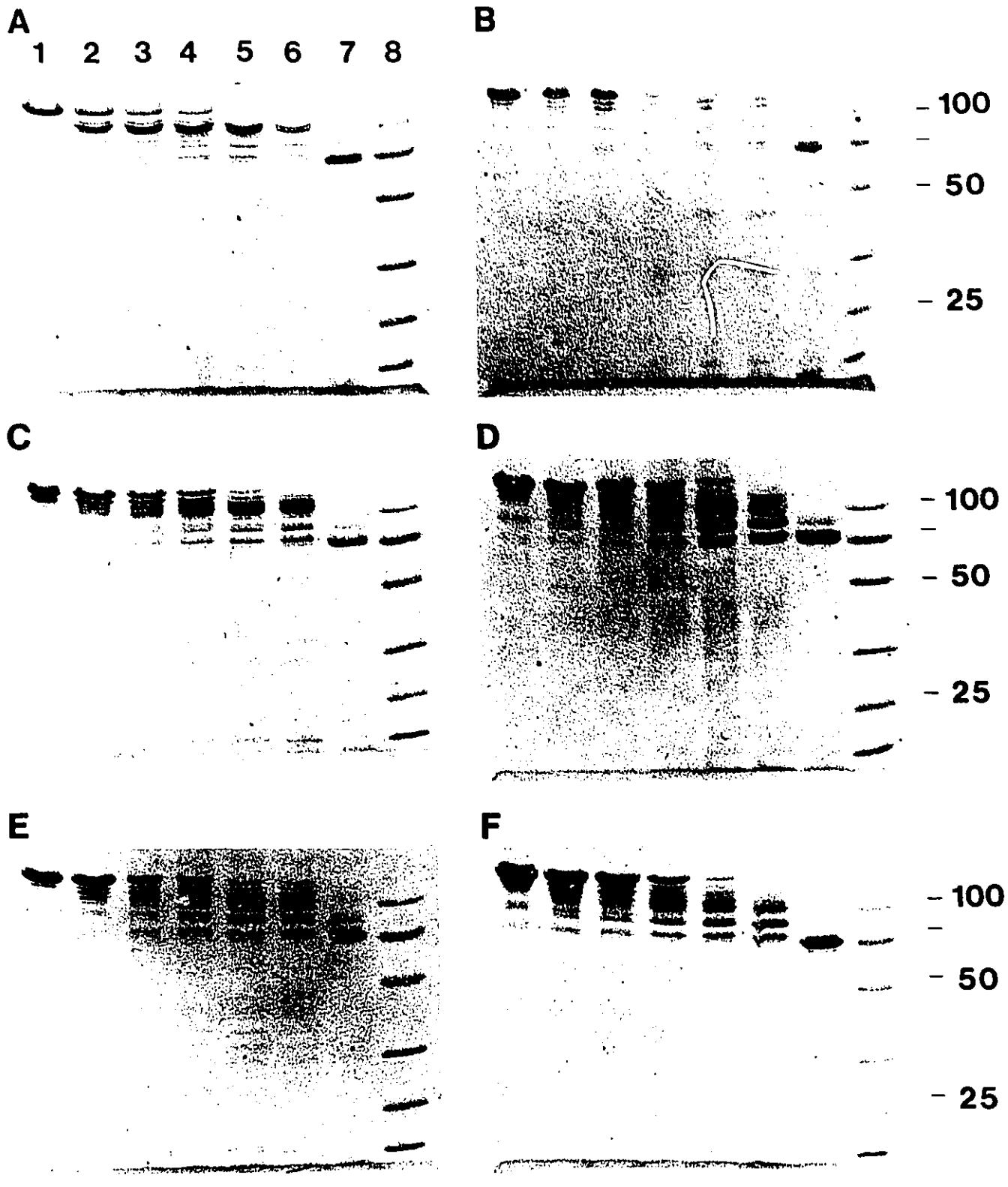
Table 6.2 Toxicity to CF-1 cells of toxins prepared from chemically modified protoxins.

Parent derivative	Threshold level (ng/μl)*
native protoxin	0.38-0.19
CM protoxin	0.38-0.19
CAM protoxin	0.38-0.19
AE protoxin	2.14-1.07
SCM protoxin	0.38-0.19
SCAM protoxin	0.38-0.19

* Threshold level lies between the given values

Figure. 6.4 Sequential tryptic proteolysis of native crystal and derivatized protoxins.

Protoxins were dissolved in pH 10.5 CAPS buffer containing 0.01% ß-mercaptoethanol and sequential degradation of the C-terminal half of the proteins was conducted as described in Section 5.2.8. The protoxin derivatives were: (A) native crystal protoxin; (B) AE protoxin; (C) CM protoxin; (D) SCM protoxin; (E) CAM protoxin; (F) SCAM protoxin. Incubation times of the protoxin with trypsin prior to boiling in SDS sample buffer are designated by the numbered lanes as follows: lane 1, 0 minutes; lane 2, 6 minutes; lane 3, 10 minutes; lane 4, 20 minutes; lane 5, 40 minutes; lane 6, 60 minutes; lane 7, 120 minutes. For all samples, the protease-resistant toxin is the principal band after 120 minutes of proteolysis. Lane 8 shows the molecular mass markers, with molecular masses indicated in kDa.



Chapter 6. Chemical modification of protoxin Fig. 6.4

derivatized protoxins directly to spruce budworm (*Choristeneura fufimerana*) larvae indicate that chemically-modified protoxins have the same level of insecticidal activity as does native crystal protoxin (Table 6.3).

Table 6.3 Toxicity of derivatized protoxins to spruce budworm larvae.

Derivative	LD_{50} (μg/larva)	95% Confidence Limits	
		Lower	Upper
crystal protein	31	25	39
CM protoxin	51	41	65
CAM protoxin	26	16	42
AE protoxin	51	31	88
SCM protoxin	76	81	125
SCAM protoxin	47	31	72

6.4 DISCUSSION

Treatment of solubilized native protoxin with either iodoacetic acid, iodoacetamide or N-(ß-iodoethyl)trifluoroactamide resulted in total conversion of the cysteine residues to carboxymethyl, carbaminomethyl or aminoethyl cysteine, respectively (Table 6.1). In CM protoxin, the cysteines are converted into cationic groups, whereas in CAM and AE protoxin, the modified residues are uncharged. Amino acid analysis of these derivatives gave no indication that residues other than the cysteines had reacted with the modifying reagents. The ease with which the sulfhydryl groups in protoxin can be modified under non-denaturing conditions supports the conclusion of Bietlot *et al.* (1990) that all the cysteine residues are on the surface of the protoxin molecule, since a surface location for

these residues is the most straightforward explanation for their ready accessibility to modifying reagents.

In contrast to the ease of derivatizing the cysteine residues in protoxin, only 35% of the lysine residues in CM and CAM protoxin could be modified with succinic anhydride (Table 6.1). Since succinylation converts cationic amino groups to anionic groups, the relative difficulty of modifying the lysine residues may be attributed in part to electrostatic and steric effects. This view is supported by the observation that 70% of the lysines could be modified to neutral residues upon addition of acetic anhydride to either CM or CAM protoxin. However, the inability to modify all the amino groups with either of these reagents may also reflect the inaccessibility of some of the lysine residues within the protoxin molecular structure.

Modifications of the sulfhydryl and amino groups in protoxin significantly alter the solubility of the protein (Fig. 6.1). The introduction of negatively charged groups onto the cysteine residues in CM protoxin renders the protein more soluble at neutral pH than the native protoxin, in which the sulfhydryl groups are uncharged below pH 8.5. CM protoxin is also more soluble near neutral pH values than its electrically neutral counterpart, CAM protoxin. The addition of succinyl groups to the lysines of the CM derivative significantly increase its solubility; a similar effect is observed with SCAM protoxin. The insolubility of AE protoxin at all pH values is surprising, but may be due to the overall electrical neutrality of the molecule, since the introduction of 16 extra basic groups into the protein effectively balances the number of acidic residues (146 basic groups *vs.* 152 acidic groups).

Both the circular dichroism and infrared spectra of the derivatives show that chemical modification of the cysteine and lysine residues in protoxin have some effect on the secondary structure of the protein. The observed changes were not quantified, but a qualitative assessment of the spectra showed no gross changes in the conformations of the derivatives. All the modified protoxins show decreased intensity in their circular dichroism

spectra in the range 200-230 nm relative to that observed for native protoxin (Fig. 6.2), although the overall shapes of the spectra remain unaltered. In addition, the CD spectra from all the modified protoxins (except AE protoxin) are virtually superimposable. This suggests that the structural changes giving rise to the altered spectra are due to the modification of the cysteine groups, since succinylation of lysine residues appears to have little added effect on altering the CD spectra. Similar conclusions can be drawn from the infrared spectra of the derivatives (Fig. 6.3). The shapes and intensity maxima of the spectra from CM, CAM and AE protoxins all differ slightly from the spectrum for the native protein, but succinylation of CM or CAM protoxin has no added effect on the amide I band.

Digestion of the derivatized protoxins with trypsin followed by SDS/PAGE resulted in fragmentation patterns which were similar to that obtained from native protoxin (Fig. 6.4). Difficulties were encountered in obtaining a distinct banding pattern from AE protoxin, probably due to the virtual insolubility of the protein. However, the other derivatives clearly show the sequential appearance of bands with increasing time of incubation with trypsin. The fragmentation patterns of SCM and SCAM appear more complex and the bands less distinct than those obtained from the non-succinylated derivatives, suggesting that the introduction of more negative charges into the molecule may result in the exposure of additional cleavage sites. The retention of the unusual sequential proteolytic pattern by all the protoxin derivatives indicates that no gross change in the tertiary structure of the C-terminal half of the protoxin molecule results from the chemical modification of the cysteine and lysine residues.

Toxicity assays on the protease-resistant 67-kDa fragments from each of the derivatives (Table 6.2) and on the intact modified protoxins (Table 6.3) show that, except for the AE derivative, modification of the cysteine and lysine residues in protoxin has no effect on the biological activity of the toxin molecule derived from the N-terminal half of protoxin. Toxin from AE protoxin exhibited lower toxicity only in the CF-1 cell assay; the

LD_{50} measured by the spruce budworm feeding assay was the same as those obtained for the other protoxins (Table 6.3). It is not clear why different results for the AE derivative were obtained by the two assay methods. Nonetheless, the bioassay results suggest that the structure of the N-terminal toxic half of the protoxin molecule remains essentially unaffected by chemical modification of the C-terminal half. The retention of toxicity in these modified proteins further supports the view that the toxic region of protoxin acts in a manner quite autonomous from the remainder of the molecule (Chapter 5; Choma and Kaplan, 1990). It cannot be ruled out that derivatizations in the C-terminal region do cause structural changes in the N-terminal toxic moiety of the intact protoxin molecule, but if such is the case, the N-terminal half rearranges into its native, fully toxic conformation following digestion of the C-terminal region.

The results obtained here conflict with the observations of Schesser *et al.* (1977) and Bulla *et al.* (1977), who reported that treatment of protoxin from *B. thuringiensis* subsp. *kurstaki* HD-1 with iodoacetic acid resulted in a total loss of toxicity towards the tobacco hornworm, *Manduca sexta*. Schesser *et al.* (1977) did not report the reaction conditions used for the carboxymethylation of protoxin, nor was the derivative characterized. Bulla *et al.* (1977) treated non-reduced protoxin with a large excess of iodoacetic acid for 12 hours at pH 9, and yet reported that only four of the cysteines were derivatized, and that the protoxin was biologically inactive.

Arvidson *et al.* (1989) noted an effect on the host specificity of crystal protein from *B. thuringiensis* subsp. *kurstaki* HD-73, depending on whether the cysteine residues of the protoxin were oxidized or reduced. Reduction of the sulfhydryl groups was reported to significantly decrease the toxicity of protoxin towards *Heliothis virescens* larvae; toxicity was restored upon reoxidation. It is unclear why the oxidation state of these residues should affect host specificity, since the toxic moiety of the protoxin molecule would apparently remain unaltered. However, this observation by Arvidson *et al.* (1989) suggests that the oxidation state of the sulfhydryl groups may affect processing of the

protoxin in the insect gut. The present results show that extensive chemical modification of the sulfhydryl and amino groups of the protoxin from *Bacillus thuringiensis* subsp. *kurstaki* HD-73 leaves the toxicity properties of the molecule apparently unaffected and has no major effect on the overall structure of the molecule, although the solubility properties of the protein are greatly affected. Since native crystal protoxin is essentially insoluble at neutral pH, and as there is evidence that inefficient solubilization of the crystal is responsible for the low susceptibility of at least some insects to the protein (Tojo and Aizawa, 1983; Jaquet *et al.*, 1987), the possibility arises that more soluble and potent protoxins can be developed by site-directed mutagenesis of the cysteine and lysine residues.

6.5 CONCLUSIONS

The 16 cysteine residues in native protoxin can be quantitatively reacted with iodoacetic acid, iodoacetamide or N-(ß-iodoethyl)trifluoroacetamide. The carboxymethyl cysteine derivative was found to be significantly more soluble at neutral pH values, where both native protoxin and the carbaminomethyl cysteine derivative exhibit low solubilities. The aminoethyl derivative had an extremely low solubility at all pH values. Succinic anhydride was found to react with only 35% of the lysine residues in both the carboxymethyl and carbaminomethyl protoxin derivatives, but these modified protoxins nonetheless exhibited significantly increased solubilities at neutral pH values. All the derivatives were found to retain full insecticidal activity toward spruce budworm larvae, although toxin derived from the aminoethyl derivative showed decreased toxicity in the *in vitro* insect cell assay. It is probable that the cysteine residues and modified lysine residues are on the surface of the protein, and that derivatization does not significantly alter the conformation of the protoxin molecule.

6.6 REFERENCES

Andrews, R. E., R. M. Faust, H. Wabiko, K. C. Raymond and L. A. Bulla (1987). The biotechnology of *Bacillus thuringiensis*. *CRC Crit. Rev. Biotechnol.* **6**, 163-232.

Arvidson, H., P. E. Dunn, S. Strnad and A. I. Aronson (1989). Specificity of *Bacillus thuringiensis* for lepidopteran larvae: factors involved *in vivo* and in the structure of a purified protoxin. *Molec. Microbiol.* **3**, 1533-1543.

Bietlot, H., P. R. Carey, C. T. Choma, H. Kaplan, T. Lessard and M. Pozsgay (1989). Facile preparation and characterization of the toxin from *Bacillus thuringiensis* var. *kurstaki*. *Biochem. J.* **260**, 87-91.

Bietlot, H., I. Vishnubhatla, P. R. Carey, M. Pozsgay and H. Kaplan (1990). Characterization of the cysteine residues and disulphide linkages in the protein crystal of *Bacillus thuringiensis* subsp. *kurstaki* and *entomocidus*. *Biochem. J.* **267**, 309-315.

Bulla, L. A., K. L. Kramer and L. I. Davidson (1977). Characterization of the enotomocidal parasporal crystal of *Bacillus thuringiensis*. *J. Bacteriol.* **130**, 375-383.

Choma, C. T., W. K. Surewicz, P. R. Carey, M. Pozsgay, T. Raynor and H. Kaplan (1990). Unusual proteolysis of the protoxin and toxin from *Bacillus thuringiensis*: structural implications. *Eur. J. Biochem.* **189**, 523-527.

Choma, C. T. and H. Kaplan (1990). Folding and unfolding of the protoxin from *Bacillus thuringiensis*: evidence that the toxic moiety is present in an active conformation. *Biochemistry* (in press).

Ge, A. Z., N. I. Shivarova and D. H. Dean (1989). Location of the *Bombyx mori* specificity domain on a *Bacillus thuringiensis* delta-endotoxin. *Proc. Natl Acad. Sci. U.S.A.* **86**, 4037-4041.

Gringorten, J. L., D. P. Witt, R. E. Milne, K. van Frankenhuyzen, P. G. Fast and S. S. Sohi (1990). An *in vitro* system for testing *Bacillus thuringiensis* toxins: the lawn assay. *J. Insect. Path.* **56**, 237-242.

Höfte, H. and H. R. Whiteley (1989). Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Jaquet, F., R. Hutter and P. Luthy (1987). Specificity of *Bacillus thuringiensis* delta-endotoxin. *Appl. Environ. Microbiol.* **53**, 500-504.

Means, G. E. and R. E. Feeney (1971). *Chemical modification of proteins*. Holden Day, Inc., San Fransisco, pp. 254.

Nickerson, K. W. (1980). Structure and function of the *Bacillus thuringiensis* protein crystal. *Biotechnol. Bioeng.* **12**, 1305-1335.

Schesser, J. H., K. J. Kramer and L. A. Bulla (1977). Bioassay for homogeneous parasporal crystal of *Bacillus thuringiensis* using the tobacco hornworm *Manduca sexta*. *Appl. Environ. Microbiol.* **33**, 878-880.

Schwartz, W. E., P. K. Smith and G. P. Royer (1980). N-(ß-iodoethyl)trifluoroacetamide: a new reagent for the aminoethylation of thiol groups in proteins. *Anal. Biochem.* **106**, 43-48.

Tojo, A. and D. Aizawa (1983). Dissolution and degradation of *Bacillus thuringiensis* delta-endotoxin by gut juice protease of the silkworm *Bombyx mori*. *App. Environ. Microbiol.* **45**, 576-580.

Chapter 7.

CONCLUSIONS AND FUTURE EXPERIMENTS

The Gift

A gift of great
utility
is common
copability:
The knack of getting
each thing done
before the grind
has spoiled the fun.

-Piet Hein

Chapter 7.

CONCLUSIONS AND FUTURE EXPERIMENTS

7.1 CONCLUSIONS

The theme of the work presented in this thesis has been the characterization of the structure of the Lepidopteran-specific insecticidal protein from *Bacillus thuringiensis* subsp. *kurstaki* HD-73. Although it has been known for many years that the insecticidal properties of *B. thuringiensis* reside in the protein crystal (Angus, 1954), very little is known about either its structure or mechanism of action. Given the unique biological activity and commercial importance of the crystal protein, it is surprising that so few studies have focused on elucidating the structural and chemical basis of the protein's function. As pointed out in Chapter 1, direct studies on the protoxin have been hampered by difficulties in isolating substantial quantities of crystals free of contaminating spores and proteases, and by the large size and low solubility of the protein. Studies into the insecticidal action of the protoxin and toxin have suffered from the use of poorly defined preparations, resulting in a body of literature which is inconclusive and often contradictory (Fast, 1982; Andrews *et al.*, 1987). The lack of well-characterized protein has also resulted in most information on the structure of the insecticidal protein being inferred from comparison of protoxin gene nucleotide sequences. Recent developments in the separation of protoxin crystals from spores (Section 2.2.2; Carey *et al.*, 1986) have overcome the first of these difficulties, permitting the isolation of large quantities of protoxin and toxin with reproducible properties. The ability to isolate protoxin crystals rapidly and to generate homogeneous preparations of toxin was a prerequisite to the structural studies which comprise this thesis. Since toxin is the active insecticidal agent, the present studies focused on elucidating its structural characteristics; studies on protoxin were aimed primarily at determining the effect of the C-terminal half of the molecule on the structure of the N-terminal toxic moiety.

Characterization of toxin generated by tryptic digestion of protoxin crystals showed that it is (i) derived from the N-terminal half of the protoxin molecule; (ii) is homogeneous as judged by several criteria; (iii) is only soluble at pH values above 9 and (iv) is resistant to digestion by many proteases. The latter two characteristics indicate that toxin is well

suited to functioning in the high pH, protease-rich environment of the larval Lepidopteran gut. The comprehensive study of the secondary structure of the toxin presented in Chapter 3 employed a variety of physical and computational procedures. Circular dichroism, Raman and infrared spectroscopy independently estimated that toxin consists of approximately equal amounts of α-helix, ß-sheet and random coil structures. Predictive methods suggest that the α-helices are clustered in the N-terminal region and are uncommon elsewhere in the molecule. A high proportion of hydrophobic residues occur near the predicted helices; it has been suggested that this region constitutes a transmembrane sequence which penetrates the insect gut epithelium, forming pores in the membrane and thereby severely disrupting the ionic balance of the cells (Schnepf *et al.*, 1985).

The locations of the toxicity and binding domains in toxin have been predicted from alignment studies of the known protoxin gene sequences (Andrews *et al.*, 1987; Höfte and Whiteley, 1989). Using the method of limited proteolysis, it was shown here that toxin is proteolyzed at two locations. The primary site is located in the interdomain region between the predicted toxic and binding domains (Geiser *et al.*, 1986). The second cleavage site corresponds to the interdomain region between the predicted binding and toxic boundary domains (Andrews *et al.*, 1987). The presence of a minimum of three domains in toxin was corroborated by differential scanning calorimetry: two major thermal events are observed, and a third transition is evident upon deconvolution of the data. At least two of the domains unfold independently of each other in a manner consistent with structural domains; however, the data could not discern whether the third domain unfolded in a co-operative or independent manner. Combining the proteolytic and thermal denaturation data, it is proposed that the observed low temperature transition arises from the unfolding of the toxicity domain, while the middle and high temperature events arise from the thermal denaturation of the binding and toxic boundary domains, respectively.

The toxic moiety within the protoxin molecule was observed to generate the same distinctive proteolytic fragmentation pattern as that obtained upon limited proteolysis of

activated toxin. This pattern is not altered by prolonged incubation of either protein in urea. The observed resistance of the toxin and toxic moiety in protoxin towards chemical denaturation lead to an investigation of the structural role of the C-terminal half of the protoxin molecule. Proteolysis studies showed that toxin could be generated from protoxin in the presence of urea or guanidinium chloride, but that this protein was less protease-stable than toxin activated in the absence of denaturants. Native-like protease resistance could be attained by simply removing the denaturant. Similarly, the native fluorescence spectrum of both protoxin and toxin could be readily regenerated from samples which had been extensively treated with denaturants. Toxins generated in, or otherwise exposed to denaturing agents, were found upon the restoration of non-denaturing conditions to be fully toxic towards insect cells. In conjunction with fluorescence data on the time dependence of the unfolding of protoxin and toxin, it was concluded that the C-terminal half of the protoxin molecule behaves autonomously from the N-terminal toxic moiety. The two regions do not interact in a co-operative manner, but unfold and fold independently of each other. The C-terminal region has no apparent effect on the conformation of the N-terminal region, suggesting that the toxic moiety exists in the protoxin molecule in its active conformation. The results presented in Chapter 5 further suggest that the major functional role of the C-terminal half of protoxin is in crystal formation.

Additional evidence that the C-terminal region of protoxin does not interact with the N-terminal toxic moiety was obtained by chemical modification of the 16 cysteine residues which are removed during proteolysis of protoxin to toxin. Complete derivatization of these residues to either cationic or uncharged groups had no effect on the toxicity of the preparations to spruce budworm larvae, suggesting that the conformation of the toxic moiety remained unaffected. Thirty one of the 34 lysine residues in protoxin are also removed during the activation process; modification of 35% of the lysine residues had no effect on the activity of the protoxin. The retention of biological activity and the increased

solubilities of these derivatized protoxins raises the possibility of expanding the host specificity range of the protein by modifying these apparently non-critical residues.

7.2 FUTURE DIRECTIONS

7.2.1 Justification for additional research

This thesis represents the first comprehensive study at the protein level into the inter-relationship between structure and function in the *B. thuringiensis* insecticidal protein. As a consequence, some of the results presented here are necessarily of a preliminary nature. The information obtainable from any single experimental technique is generally limited and the data may be insufficient for unambiguous interpretation. Where possible, a number of approaches and methods were applied to the study of each problem undertaken. However, in the absence of a three dimensional structure, it is not possible to discriminate at the molecular level the exact nature and position of the structural features which collectively give rise to the unique properties of any protein. The techniques applied here could only provide information on the gross structural organization of the insecticidal protein. Nevertheless, the results obtained revealed several unusual features of both the protoxin and toxin molecules. Foremost amongst these is the independent behaviour of the N- and C-terminal halves of the protoxin. The two regions behave as though they were autonomous proteins fused together, suggesting that each region may have originally followed a separate evolutionary pathway but that the two proteins fused in order to confir upon the insecticidal protein the unique ability to form crystals within the bacterium. Given its large size and substantial number of cysteine bridges, the protoxin is also unique in the readiness with which it regains its native conformation following treatment with denaturants. The rapid unfolding and ease of proteolysis of the C-terminal region contrasts with the resistance of the toxin and toxic moiety towards denaturation and proteolytic degradation; this asymmetry further extends to distinct differences in hydropathy, domain structure and amino acid composition. Taken in conjunction with the truly novel biological

functions of the toxic moiety and C-terminal region, the insecticidal protein from *Bacillus thuringiensis* is deserving of a concerted effort by crystallographers, protein chemists, molecular biologists and entomologists to elucidate the structural basis and molecular mechanism of action of this unique protein. If the properties of the protein are found to arise from novel structural features, then the protoxin and toxin will constitute a valuable model system contributing to the elucidation of the principles which govern the structural organization of proteins in general.

7.2.2 Future experiments

As so little is currently known concerning the molecular mechanism of action of the protoxin and toxin, it is evident that the scope for future work to unravel the inter-relationship between structure and function is extensive. The suggestions presented here are limited to experiments which follow directly from the work presented in this thesis.

The most obvious and direct manner by which to gain insight into the underlying structural basis of the protein's insecticidal action is by analyzing, comparing and contrasting its X-ray crystallographic structure with those of other known toxins. Although the protoxin is naturally produced in a crystalline form, efforts to grow protoxin or toxin crystals to a size usable for X-ray structural studies have thus far been unsuccessful. However, microcrystallites of toxin have been obtained, suggesting that a detailed investigation into conditions whereby a toxin solution at high pH is slowly lowered to neutral pH may succeed in growing crystals suitable for X-ray analysis.

Building upon the results obtained in the present investigation, detailed chemical modification studies will further indicate which residues are crucial to the biological functioning of the protoxin and toxin. For example, loss of toxicity following modification of histidine, tyrosine or tryptophan residues will suggest that some of these residues, which are often located in active site regions, are crucial to the structure of the toxicity or binding domains of the protein. Residues which can be modified without detrimental

effects on activity will further indicate which amino acids can be altered by site-directed mutagenesis to diversify the properties of the insecticide.

The calorimetric results showing that the toxicity and binding domains unfold independently of each other suggest that cloning of the toxin fragments corresponding to these domains may result in stable products which retain their original structures and functions. The cloned domains could then be used to elucidate the events occurring at the molecular level which lead to the binding and toxic action of the toxin. Changes in specificity or binding constant as compared to the intact protein would imply that interactions between the two domains, perhaps giving rise to an active site cleft, are central to the functioning of the toxin.

The results presented here indicate that the domain structure of the toxic moiety in protoxin is similar to that in activated toxin. A calorimetric study of the unfolding of protoxin would provide insights into the effect of the C-terminal region on the tertiary structure of the toxic moiety, particularly regarding the number of domains and the co-operativity of their unfolding. The sequential proteolytic degradation of the C-terminal region has been shown to be part of the molecular mechanism by which toxin is activated from protoxin. Calorimetric studies on the protoxin could provide insights into the domain structures which give rise to this novel phenomenon.

7.3 REFERENCES

Andrews, R. E., R. M. Faust, H. Wabiko, K. C. Raymond and L. A. Bulla (1987). The biotechnology of *Bacillus thuringiensis*. *CRC Crit. Rev. Biotechnol.* **6**, 163-232.

Angus, T. A. (1954). A bacterial toxin paralysing silkworm larvae. *Nature* **173**, 545-546.

Carey, P. R., P. Fast, H. Kaplan and M. Pozsgay (1986). Molecular structure of the protein crystal from *Bacillus thuringiensis*: a Raman spectroscopic study. *Biochim. Biophys. Acta* **872**, 169-176.

Fast, P. G. (1981). The crystal toxin of *Bacillus thuringiensis*. (in) *Microbial control of pests and plant diseases* (H. D. Burgess, ed.) Academic Press, London, pp. 223-248.

Geiser, M., S. Schweitzer and C. Grimm (1986). The hypervariable region in the genes coding for entomopathogenic crystal proteins of *Bacillus thuringiensis*: nucleotide sequence of the *kurhl* gene of subsp. *kurstaki* HD1. *Gene* **48**, 109-118.

Höfte, H. and H. R. Whiteley (1989). Insecticidal crystal proteins of *Bacillus thuringiensis*. *Microbiol. Rev.* **53**, 242-255.

Schnepf, H. E., H. C. Wong and H. R. Whiteley (1985). The amino acid sequence of a crystal protein from *Bacillus thuringiensis* deduced from the DNA base sequence. *J. Biol. Chem.* **260**, 6264-6272.

Appendix A

MIXING OF AQUEOUS SOLUTIONS IN MICROGRAVITY

Last things first

Solutions to problems
are easy to find:
the problem's a great
contribution.
What is truly an art
is to wring from your mind
a problem to fit
a solution.

- Piet Hein

Appendix A

MIXING OF AQUEOUS SOLUTIONS IN MICROGRAVITY

A.1 INTRODUCTION

This thesis has dealt with characterizing the *B. thuringiensis* insecticidal protein using a variety of chemical and physical techniques, and represents the most comprehensive study to date on the structure of this protein. Nonetheless, it is obvious that details of the structure/function inter-relationship which govern the unique characteristics of this protein are not understood in depth. The recommendation was made in Chapter 7 that efforts be directed towards obtaining protoxin or toxin crystals suitable for X-ray diffraction studies, as this would provide the most direct information on the underlying structural basis of the protein's molecular mode of action. Efforts during the present study to crystallize *B. thuringiensis* subsp. *kurstaki* HD-73 toxin resulted in the growth of microcrystallites. While the crystallization of a 64-kDa toxin from an unidentified subspecies of *B. thuringiensis* has been reported (Garfield and Stout, 1988), no preliminary results regarding the structure of this toxin appeared, suggesting that the crystals obtained may be too disordered or are otherwise not suitable for X-ray analysis.

The difficulty of obtaining quality crystals in general has limited the application of X-ray diffraction as a means of elucidating protein structures. Proteins have small ordering energies, and when deprived of their hydration requirements (*e.g.*, by the addition of salt or organic solvent), tend to precipitate out of solution as an amorphous solid rather than form crystals. When the addition of a precipitant agent to a protein solution does result in the formation of crystals, often only minute crystals, or large crystals with a high degree of internal disorder, are produced. The kinetics and thermodynamics associated with the incorporation of soluble macromolecules into highly ordered crystalline structures are being elucidated for a number of model proteins (Tiller, 1986; Atka and Asai, 1988) and confirm that the process is generally thermodynamically unfavourable. Thermodynamics thus underlies the usual need to test a large number of parameters in order to find conditions

conducive to the crystallization of any single protein. Protocols have been developed to shorten this trial-and-error process (Carter and Carter, 1979). However, substantial effort and protein must still be committed to finding controlled precipitation conditions which lower the protein's solubility, causing it to come out of solution in a uniform manner so that individual molecules are precisely positioned for incorporation into the growing crystal lattice structure. As only several hundred protein structures have been solved to date, it is evident that for most proteins these conditions remain unknown. Thus, the lack of success in growing toxin crystals during the present study exemplifies the rule rather than the exception in protein crystal growth. The realization that obtaining novel protein crystals remains a largely empirical process with only a low probability of success, coupled with previous experience and continuing interest in the field of materials processing in space, lead to the auxiliary research project briefly described in this Appendix.

During the last decade, several research groups have studied the effect of the microgravity (or 'zero gravity') environment of space on protein crystal growth (Littke and John, 1984; Bugg, 1986; Drenth *et al.*, 1987). These experiments were based on the rationale that the absence of density-driven convective flow may benefit macromolecular crystal formation. During conventional crystal growth, protein is taken up from solution and incorporated into the lattice, thus depleting the supply of soluble protein molecules near the surface of the crystal and causing density changes at the crystal interface. In a gravitational field, the low density solution surrounding the crystal rises, creating convective flow patterns. This forces the solution to flow by the crystal at such a rate that steady-state diffusion of protein molecules from the bulk solution to the crystal face ceases to be a rate-limiting step in growth; in addition, non-uniform conditions are experienced at various parts of the crystal surface. As protein crystals are extremely fragile, it is believed that the depletion layer and convective flows are sufficient to cause crystal defects (Pusey *et al.*, 1988; Broom *et al.*, 1988). In microgravity, the virtual elimination of these effects permits the role of convection on crystal growth to be examined directly. Microgravity also

serves to minimize sedimentation of the growing crystal which can further interfere with uniform crystal growth. In addition, it should be possible to grow the crystals in drops which have only a small area of contact with a solid surface, thus minimizing unwanted nucleation at the solid/liquid interface. The results of eleven different protein crystal growth experiments conducted on U.S. shuttle flights have been critically examined (DeLucas *et al.*, 1989). This evaluation showed that for many (but not all) proteins tested, microgravity resulted in the growth of larger crystals with more uniform morphologies yielding diffraction data to significantly higher resolutions than the best crystals of these proteins which can be grown in Earth (1-g) gravity.

The vapour diffusion or 'hanging drop' method has been the technique used with most success in both 1-g and μ-g crystallization experiments (DeLucas *et al.*, 1986). In this method, the protein and precipitant solutions are extruded together from an injector to form a 'hanging' or 'pendent' droplet with a volume ranging from 10 μl (in 1-g) to 80 μl (in μ-g). The drop is enclosed in a sealed chamber containing a 1-2 ml reservoir of precipitant solution. As the concentration of precipitant in the reservoir is higher than it is in the drop, the system is driven towards equilibrium by the vapour diffusion of water from the drop to the reservoir. This effectively raises the concentration of precipitant in the drop and forces the protein from solution.

Several unique problems arise when the pendent drop method is used to grow crystals in microgravity. On the ground, it is typical for a crystal to require weeks or months to grow to a useable size. Due to suppressed convection effects, establishing equilibrium between the drop and the reservoir may take longer in μ-g than in 1-g, slowing the crystallization process (Fowlis *et al.*, 1988). However, only short space flights of approximately one week are generally available to Western researchers. A second problem is that the protein and precipitant solutions must be mixed together in μg, as premixing of the solutions prior to launch may lead to premature nucleation and crystal growth. Due to the empirical approach to protein crystal growth and the economics of space research,

several hundred protein/precipitant pairs may be flown in a single protein crystal growth apparatus. Presently, the in-flight mixing of each sample pair is achieved by repeatedly withdrawing and extruding the two solutions through a single syringe from which the 'pendent' drop is grown (DeLucas *et al.*, 1989). Although this approach is mechanically simple, air bubbles are often introduced into the drop, increasing the likelihood of multiple nucleation and the growth of microcrystallites rather than large single crystals. A third problem involves selecting the material from which the protein/precipitant solution is extruded and attached to during the crystal growth period. The chosen material for the injector must be non-wettable by a large number of protein and precipitant solutions exhibiting a wide range of physical properties, and should be amenable to being cast from a mold. Creeping of the protein solution along the outer face of the injector would encourage nucleation at the large solid/liquid interface, thus largely defeating the purpose of using the μ-g environment; cut surfaces, even when polished, are sufficiently irregular to offer multiple nucleation sites. Injectors made of polypropylene or glass have been used with some success to crystallize proteins in μ-g (Schoen and Seifert, 1987; Giege *et al.*, 1988), but problems with fluid creeping and unwanted nucleation at the solid/liquid interface persist.

The purpose of the present study was to address specific aspects of the second and third problems outlined above. Firstly, what material should the injector be constructed of so that the 'pendent' drops formed are stable, even to perturbations in μ-g levels? Secondly, is it necessary to actively premix the protein and precipitant solutions in μ-g prior to droplet extrusion, or is the mixing which results when two fluids are injected into each other sufficient to yield a homogeneous solution? These questions were addressed by a series of simple experiments conducted on aircraft flying parabolic trajectories.

A.2 MATERIALS AND METHODS

A.2.1 Materials

Bovine haemoglobin (2X crystallized and lyophilized) was purchased from Sigma Chemical Co., St. Louis, Missouri. Methylpentanediol (MPD) and polyethylene glycol (PEG; molecular weight 4,000) were purchased from Aldrich Chemical Co., Milwaukee, Wisconsin. All other chemicals were high purity reagents obtained from commercial sources. Solutions were prepared with reverse-osmosis quality water purified by the Milli-Q water system.

The following materials were tested for their suitability to support stable 'pendent' droplets of protein and precipitant solutions: stainless steel, Teflon, silicone rubber, polyethylene, polypropylene, polycarbonate, polystyrene, polysulfone, polyurethane and borosilicate glass. The plastic resins were obtained from Cole-Parmer Co., Chicago, Illinois and Canus Plastics Ltd., Ottawa. All materials were in the form of small bore tubing with dimensions of 1.6 mm i.d. and 3.2 mm o.d.

A.2.2 Fluids

One protein solution in combination with three precipitant solutions was chosen to represent the spectrum of physical properties typified by a 'general' protein crystallization drop. Precipitant agents are usually salts, organic solvents or polymers; one example from each of these classes was tested here. In order to adequately assess injector materials and solution mixing behaviour, fluids exhibiting widely different physical properties were required. Thus, fairly high concentrations of each precipitant were employed, representing the extremes of concentrations used to actually induce protein crystallization (Carter and Carter, 1979). Haemoglobin was chosen as the model protein for these studies because of its ready availability and its colour, which aided visual observation of the mixing of protein and precipitant solutions. The solutions used to assess injector materials for their suitability to support stable 'pendent' drops in microgravity consisted of 1.5 mg/ml haemoglobin in

0.05M Tris-HCl, 0.1M sodium citrate, pH 7.0, containing either: (a) 1M ammonium sulphate; (b) 15% (v/v) MPD or (c) 15% (w/v) PEG 4000.

The mixing behaviour of a protein solution with a precipitant solution was studied using 3.0 mg/ml haemoglobin in 0.05M Tris-HCl, 0.1M sodium citrate, pH 7.0. This protein solution was extruded simultaneously with one of the following precipitant solutions: (a) 2M ammonium sulphate; (b) 30% PEG or (c) 30% MPD. All solutions were buffered as described above. The mixing behaviour of two identical solutions (*e.g.*, 30% PEG with 30% PEG) was observed by mixing a drop of food colouring in the reservoir of one solution.

Viscosities of all solutions were measured using a Ubbelohde viscometer. Surface tension and interfacial tension measurements were made using a Fisher Scientific Model 21 surface tensiomat. All experiments and measurements were conducted at 20°C.

A.2.3 Droplet generation

The apparatus shown in Fig. A.1 was designed to repeatedly extrude a fixed volume from up to 10, 1 ml syringes simultaneously. The syringe plungers were ganged together such that a 2 kg downward force on the top plate caused all the plungers to travel exactly 5.0 mm, resulting in the extrusion of 80 µl from each syringe. The plungers were depressed in a single smooth motion so that approximately 3 seconds were required to extrude the drops. Drops with a final volume of 80 µl were used to observe drop stability as this approximates the largest volume required for the adequate growth of protein crystals, and also represents a drop size (approximately 5 mm diameter) easily observed by eye.

A.2.4 Injector design

The stability of 'pendent' drops on the various injector materials was determined by inserting a straight, blunt 16 gauge stainless steel needle into a 1.5 cm length of tubing, such that the tubing protruded 5 mm beyond the end of the needle. One needle/tubing

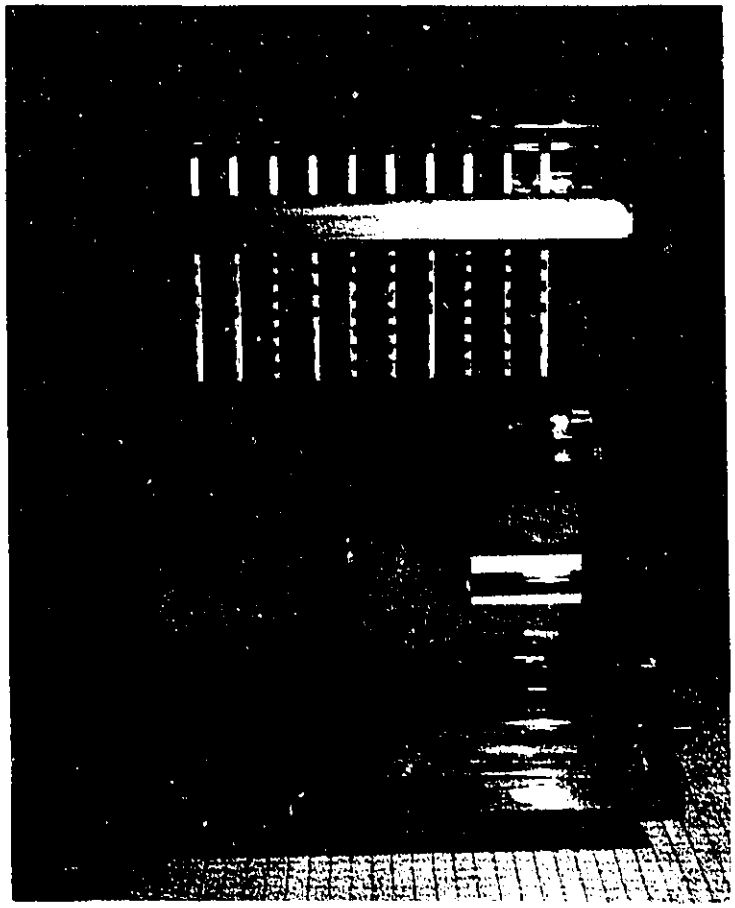


Figure. A.1. Apparatus used to simultaneously generate fixed volume droplets from multiple injectors.

injector was attached to the end of each syringe. Syringes, needles and tubing were discarded after each experiment to avoid cross-contamination of fluids.

The mixing of protein and precipitant solutions in μ-g was studied using injectors in various configurations, as seen in Fig. A.1. Each injector was composed of a blunt 16 gauge needle sheathed with either Teflon or silicone rubber. Protein and precipitant solutions were brought together in one of three ways. (i) Two injectors were bent at 45^{o} so that their tips touched. Eighty microlitres of protein solution was extruded through one injector, and an equal volume of precipitant through the other. (ii) Precipitant solution was extruded through a straight injector, and protein solution through an injector bent at 90^{o}.

The bent injector was attached to a spring mechanism which permitted smooth transverse motion of the injector tip. This allowed two pregrown drops to be brought into contact with each other. (iii) Precipitant solution was extruded through a straight injector. Protein was extruded through an injector bent into a 'U' shape, which permitted the injector tip to be positioned directly beneath the straight injector. The vertical separation between the two tips was 6 mm; thus, during extrusion the two drops were partly grown before contacting each other.

A.2.5 Data collection

One drop was extruded from the individual syringes at the onset of each stable microgravity period (generally 4-5 seconds into the parabola). A Super VHS camera positioned in front of the droplet generator provided a video record of the experimental results. A 1/10 second timer in the camera's field of view was synchronized with the initiation of droplet extrusion. Since three-axis accelerometer traces for each μ-g period were available post-flight, this allowed observations made during the course of each experiment to be correlated to the μ-g level at that moment in time. Still photographs were also collected, using a 35 mm camera fitted with a motor drive. This permitted photographs to be collected at 3 frames/second. The mounting of the droplet generator and cameras in the KC-135 cabin is shown in Fig. A.2.

A.2.6 Microgravity requirements

The apparatus was flown on both the National Aeronautical Establishment's T-33 and NASA's KC-135 aircraft for a total of 120 parabolas. A typical parabolic maneuver is depicted graphically in Fig. A.3. Approximately 30 seconds of 10^{-2} to 10^{-4}g is experienced at the top of the parabola, during which the aircraft is in a state of free-fall. This state is achieved by reducing the lift on the wings to zero while matching the thrust generated by the engines to the drag caused by the airflow striking the aircraft. Since all objects within the aircraft are accelerating towards the earth at essentially the same rate, they

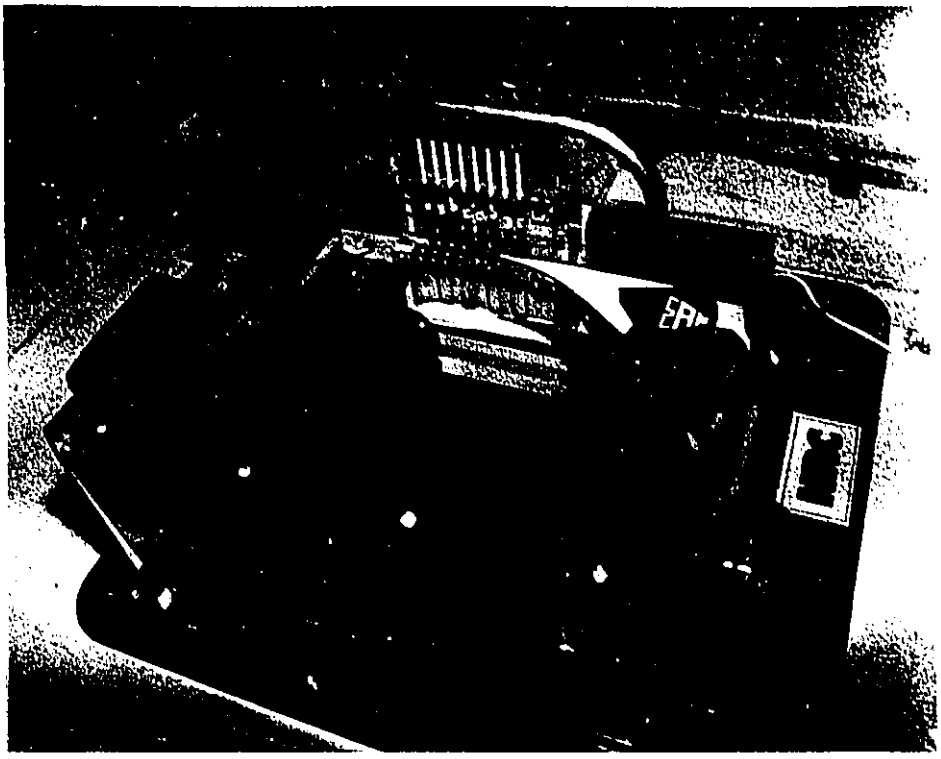


Figure. A.2 Configuration of droplet generator and cameras in the KC-135 cabin.

have no tendency to move with respect to each other. Therefore, floating objects continue to float until the end of the maneuver, at which time the aircraft is pulled up for the start of the next parabola. On the KC-135, a typical flight consists of 40 consecutive parabolas. The present experiment was manually controlled, and thus for ease of operation was bolted to the floor of the aircraft. Due to vibrations and perturbations in the control of the aircraft, an average of 10^{-2}g was experienced by the experiment; remote-controlled, free-floating experimental packages experience approximately 10^{-4}g. The term 'microgravity' is therefore not strictly correct but is nonetheless widely used to describe the environment produced during the free-fall portion of the parabola.

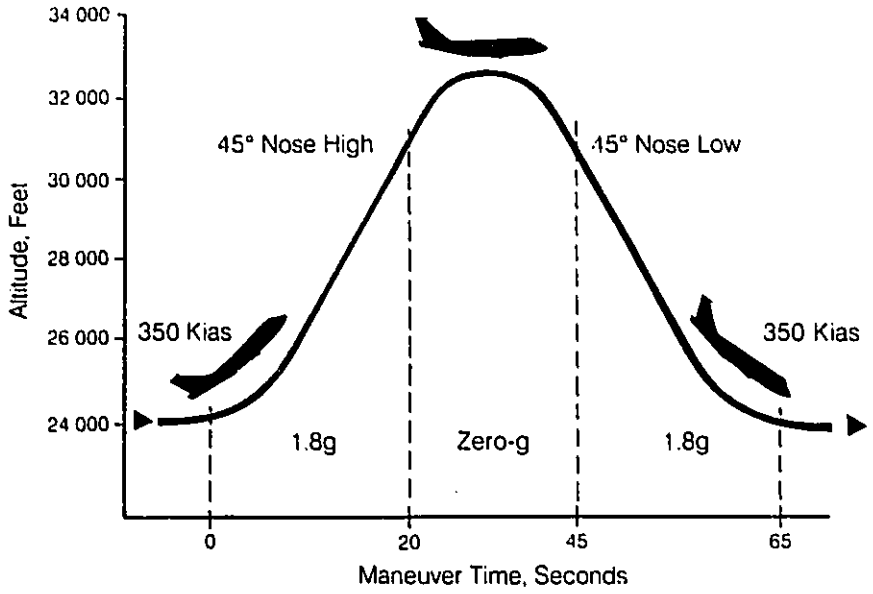


Figure. A.3. Trajectory of the T-33 and KC-135 aircraft during parabolic maneuvers.

A.3 RESULTS

A.3.1 Selection of injector material

The physical properties of the three protein/precipitant solutions used to assess the suitability of injector materials to produce stable 'pendent' drops in microgravity are given in Table A.1. Eighty microlitre drops of each solution were extruded a total of 10 times from each injector. The injectors tested were composed of: stainless steel, Teflon, silicone rubber, polypropylene, polyethylene, polycarbonate, polystyrene, polysulfone, polyurethane or glass. Despite the range of physical properties exhibited by the protein/ precipitant solutions, a given injector tended to produce similar results with all three fluids. None of the solutions formed 'pendent' drops from the stainless steel or glass injectors; the fluids immediately crept along the outer surface of these materials. About 20% of the drops

TABLE A.1. Physical properties of protein/precipitant solutions used to assess injector materials.

Solution*	Density (g/ml)	Viscosity (millipoise)	Surface tension (dynes/cm)
1M $(NH_4)_2SO_4$	1.119	16.0	36.9
15% MPD	1.010	20.5	36.3
15% PEG	1.029	44.4	43.9

* All solutions contained 1.5 mg/ml haemoglobin buffered to pH 7.0 in addition to the indicated precipitant.

formed from polyethylene and polypropylene injectors were 'pendent', while with polycarbonate, polyurethane and polystyrene no creeping was observed in about 25% of the trials. Half the drops remained 'hanging' to the polysulfone injector, whereas Teflon and silicone rubber consistently supported 'pendent' drops of all three solutions in at least 9 out of 10 trials. From the video record of each experiment and the accelerometer trace of each parabola, the moment at which these hanging drops separated from the injector tip could be correlated to the g-level experienced by the drops. It was estimated that for injectors made of Teflon and silicone rubber, the 80 µl drops remained attached and pendent in gravitational fields which fluctuated between -0.1 and +0.4 g's parallel to the axis of the injector.

In order to determine whether one material (Teflon or silicone rubber) is superior over the other in supporting stable 'hanging' drops, the injectors were bent 90^{o} so that the drops were 'pendent' perpendicular to the principal fluctuations in g-level (along the *z*-axis). Both materials performed similarly under these adverse conditions. It can therefore

be concluded that, of the materials tested, Teflon and silicone rubber are the materials of choice for supporting stable, 'pendent' protein drops in microgravity.

A.3.2 Mixing of protein and precipitant solutions

The physical properties of the solutions used to observe mixing between protein and precipitant in microgravity are given in Table A.2. The concentration of solute (protein or precipitant) is double that used in the above experiments, as each solute was halved when protein and precipitant drops were brought into contact with each other. The interfacial tensions between each precipitant solution and the haemoglobin solution were measured. A reproducible value of 59 dynes/cm was obtained for the ammonium sulphate/protein interface and 79 dynes/cm for the PEG/protein interface, but no stable interface was obtained between MPD/haemoglobin in 1-g.

TABLE A.2. Physical properties of protein and precipitant solutions used to observe fluid mixing in microgravity.

Solution*	Density (g/ml)	Viscosity (millipoise)	Surface tension (dynes/cm)
3.0 mg/ml haemoglobin	1.024	11.9	46.1
2M $(NH_4)_2SO_4$	1.142	17.0	72.8
30% MPD	1.011	42.0	33.0
30% PEG	1.056	112.3	55.3

* All solutions were buffered to pH 7.0

The first mixing experiment consisted of pregrowing an 80 microlitre drop of protein on an injector bent 90^{o} in relation to a second injector holding an equal volume drop of precipitant. The injector holding the 'pendent' protein drop was slowly and smoothly moved transversely until it contacted the precipitant drop. As some of the solutions possess fairly high surface tensions (Table A.2) it was anticipated that the drops might recoil rather than coalesce. In fact, the haemoglobin drop was observed to fuse with drops of all three precipitant solutions. However, although fusion was accompanied by intense surface oscillations (which in all cases were damped within 1 second), mixing of protein and precipitant did not necessarily follow. An example of this is seen in the left-hand sequence in Fig. A.4(A), which clearly shows two distinct layers in a fused drop of haemoglobin and ammonium sulphate solutions. The layers remain separate and show no sign of mixing throughout the 25 second microgravity period. (The sealed tube at the bottom of each frame in Fig. A.4(A) contains coloured ethanol and air. With the virtual elimination of density differences between the two phases, surface tension effects dominate, and ethanol and air partition so as to minimize the surface area between the two fluids.)

Very similar results were obtained when pregrown drops of protein and PEG were brought together, except that the separation between the solutions was even more distinct than that observed with protein/ammonium sulphate. With protein and MPD, however, partial mixing of the fluids did accompany fusion of the drops, as seen in the left-hand sequence in Fig. A.4(B).

Little mixing occurred when protein and precipitant solutions were co-extruded simultaneously into each other, even when mixing was encouraged by oscillating gravitational fields. The sequence shown in Fig. A.4(A) was obtained during a flight in which g-levels oscillated from approximately +0.3g to -0.2g. The centre drop is the product of the PEG solution, flowing through the left injector, being extruded simultaneously with an equal volume of haemoglobin solution, flowing through the right

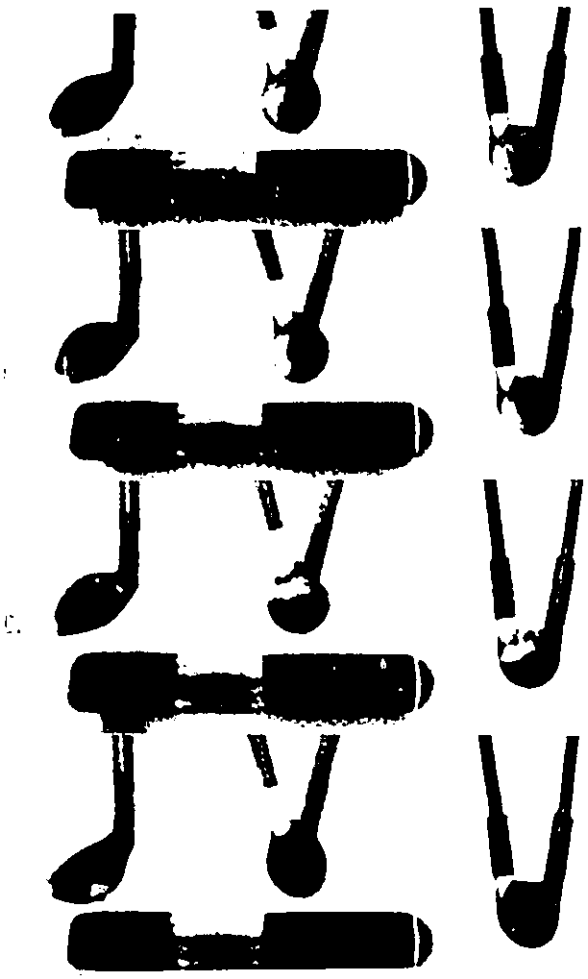


Figure. A.4(A). Mixing of haemoglobin with ammonium sulphate and PEG solutions.
The sequence (top to bottom) was photographed at 4 second intervals.
Left: Eighty µl drops of 2M ammonium sulphate (left injector) and 3.0 mg/ml haemoglobin (right injector) were grown separately, then the injector holding the ammonium sulphate drop was moved horizontally until the two fluids made contact.
Centre: Eighty µl of 3.0 mg/ml haemoglobin (right injector) and 80 µl of 30% PEG (left injector) were simultaneously extruded. Since the injectors were touching, the two fluids were effectively injected into each other.
Right: Eighty µl of 2M ammonium sulphate (right injector) were extruded simultaneously with 80 µl of haemoglobin (left injector). The distance between the two injector tips was 4 mm; thus, the drops were partly grown before they contacted each other.

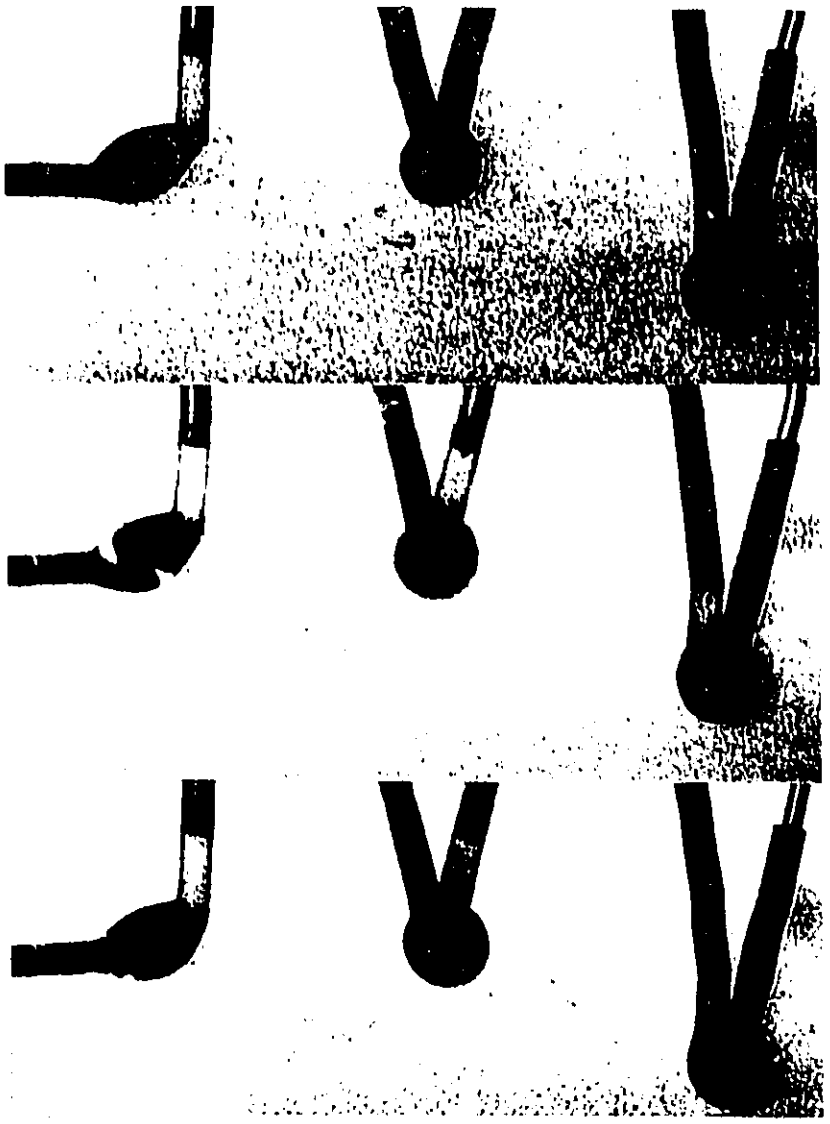


Figure. A.4(B). Mixing of haemoglobin with ammonium sulphate and PEG solutions.
The sequence (top to bottom) was photographed at 2 second intervals.
Left: Eighty µl drops of 3.0 mg/ml haemoglobin (left injector) and 30% MPD (right injector) were grown separately, then the injector holding the protein drop was slowly moved horizontally until the two fluids made contact.
Centre: Eighty µl of 3.0 mg/ml haemoglobin (left injector) and 80 µl of 30% PEG (left injector) were simultaneously extruded. Since the injectors were touching, the two fluids were injected into each other.
Right: Eighty µl of water (right injector) was extruded simultaneously into 80 µl coloured water (left injector).

injector. The interface between the protein and PEG remains sharp throughout the sequence, despite the severe oscillations in microgravity which cause the droplet mass to rotate in the direction of the fluctuating gravitational field. The right-hand pair of injectors in Fig. A.4(A) show a similar situation, where ammonium sulphate is the uncoloured fluid. The co-ordinated movement of the three drops in Fig. A.4(A) in response to the unstable 'microgravity' conditions is clearly depicted in the photographic sequence. It can also be observed that the interface between haemoglobin and ammonium sulphate is less stable than the interface between the protein solution and PEG. Nonetheless, little mixing occurs, whether the salt solution is ejected simultaneously into the extruding protein stream, or whether the two drops are pregrown before being brought gently into contact.

Figure A.4(B) contrasts the mixing properties of three pairs of fluids. The left hand pair of injectors shows the partial mixing which occurs when pregrown drops of MPD and protein are brought together, as mentioned above. The centre pair illustrates the lack of mixing between the PEG and haemoglobin solutions when the two fluids are simultaneously extruded into each other. This contrasts sharply with the right hand pair of injectors, which shows the result of co-extruding coloured and uncoloured water; in this case, a virtually homogeneous drop is obtained. Surface tension measurements of water both with and without the dye showed that the addition of food colouring had no effect on the surface tension.

The centre pair of injectors in Fig. A.4(B) also illustrate an interesting observation made during periods of stable microgravity. A high viscosity fluid such as 30% PEG, pushed through a tube with a given force, will flow with a lower velocity than a less viscous fluid such as haemoglobin extruded through an identical tube with an equal force. The haemoglobin solution will therefore emerge from its injector first. The difference in flow rate of the two solutions causes the fluid inside the drop to rotate in the direction of less viscous to more viscous liquid, and thus a counter-clockwise rotation of the centre droplet in Fig. A4(B) is observed.

The reluctance of protein and precipitant solution to mix in microgravity was also seen when droplets of the two fluids were partly extruded before contacting each other, as was the case when the two injectors were positioned one above the other, their tips 6 mm apart (Fig. A.5). As the two solutions were not actively injected into each other, mixing would result primarily from the turbulence caused by the fusion of the two fluid masses, and on a longer time scale, by surface-tension driven convection and diffusion. Only the

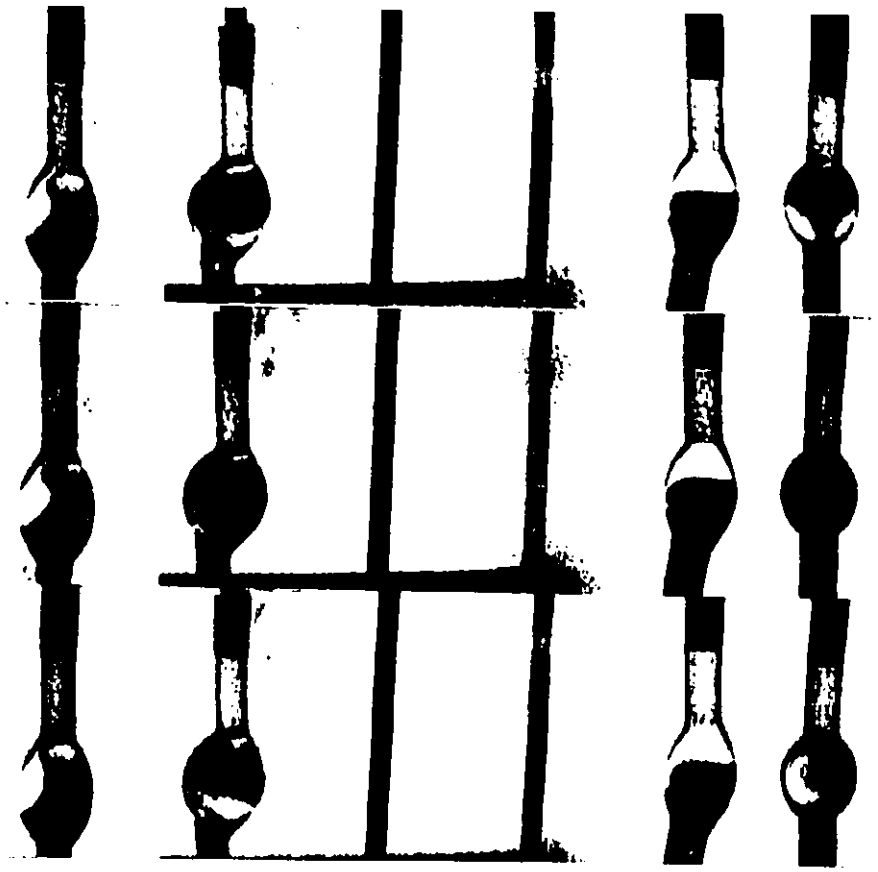


Figure. A.5. Passive mixing of haemoglobin with precipitant solutions. Photographed at 5 second intervals: far left; ammonium sulphate/haemoglobin, centre left; MPD/haemoglobin, centre right; PEG/haemoglobin, far right; water/water. Note that protein solution was extruded through the bottom injector and precipitant solution through the top injector in each pair.

combination of a water drop contacting coloured water showed significant mixing after 20 seconds. In a similar experiment where an uncoloured precipitant drop contacted a drop of the same solution containing dye (Fig. A.6), no mixing was observed except in the case of the water/water control. Attempts to measure the interfacial tensions in 1-g between a precipitant solution containing food colouring and the same solution without dye were unsuccessful, as the process of layering one solution on the other was sufficient to bring about what appeared to be complete mixing.

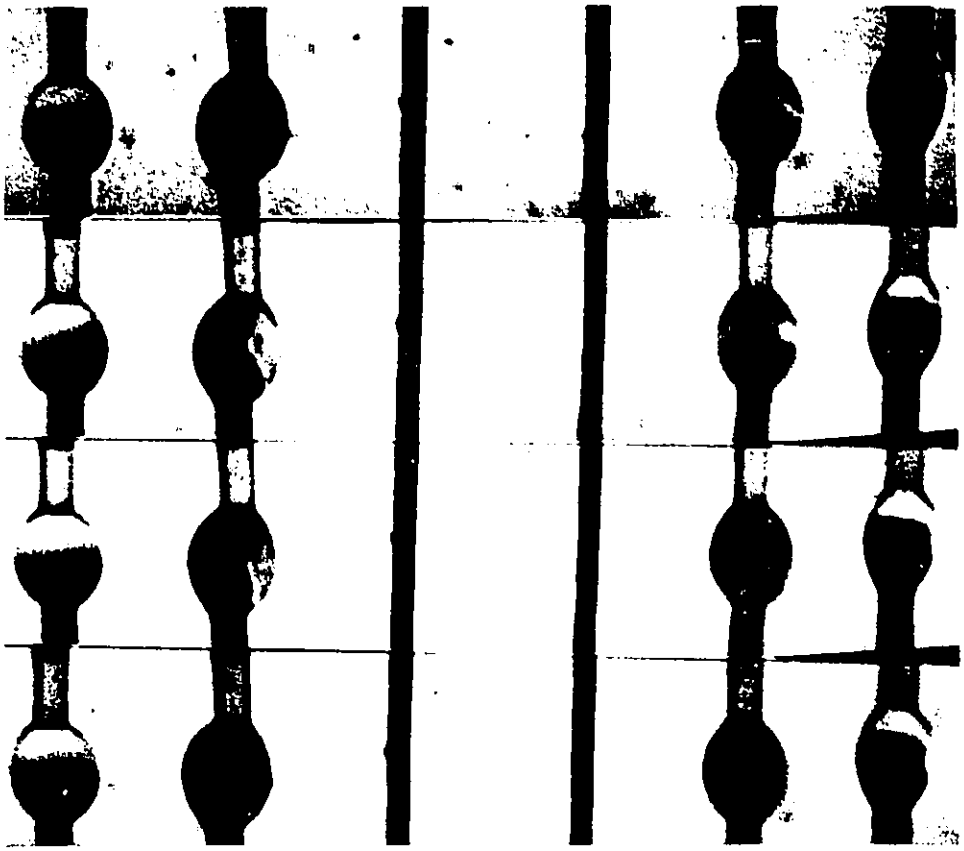


Figure. A.6. Passive mixing of like precipitant solutions.
Far left; 30% PEG/PEG, centre left; 2M ammonium sulphate/ammonium sulphate, centre right; water/water, far right; 30% MPD/MPD. Elapsed time between photographs: 4 seconds.

A.4 DISCUSSION AND CONCLUSIONS

The suitability of ten materials for producing 'pendent' protein/precipitant drops was tested with the intent of identifying the best material for supporting protein crystallization drops in microgravity. The materials were in the form of small bore tubing with the dimensions 1.6 mm i.d. and 3.2 mm o.d. All the protein/precipitant drops had a volume of 80 µl. It is not possible to grow pendent drops of this size in 1-g, and in the fluctuating µ-g environment produced by the parabolic flight of aircraft, the drops were unstable on many of the injector materials tested. Only silicone rubber and Teflon proved suitable for supporting drops of all three protein/precipitant solutions tested. Although the potential injector materials tested here by no means represent an exhaustive list, it is possible that silicone rubber or Teflon are indeed the ideal materials for this application, and it is surprising that neither has been used in microgravity protein crystallization experiments. Both materials are non-wettable by aqueous solutions, Teflon is easily machined, and silicone rubber is amenable to molding (Bowers and Zisman, 1969). As mentioned earlier, injectors cast from a mold may prove superior to machined injectors, as molded surfaces tend to be smoother than cut faces and might thus deter unwanted nucleation at the injector/protein solution interface.

The mixing behaviour of a model protein solution with three precipitant solutions representing a range of physical properties clearly showed the need to premix protein and precipitant solutions prior to extrusion and formation of a crystallization droplet; this need is more pronounced in microgravity than it is in 1-g. In the absence of density-driven convection, the momentum of the two fluids as they are injected into each other is insufficient to bring about mixing. The observation period for each individual mixing event was limited to 20-25 seconds, and therefore little additional mixing due to surface-tension driven convection effects or diffusion could be observed. During an actual crystallization experiment, a period of several days would be available for the drop to achieve homogeneity and produce protein crystals. However, in view of the cost of the short space

flights presently available and given that establishing equilibrium between the drop and precipitant reservoir is slower in μ-g than in 1-g (Fowlis *et al.*, 1988), passive achievement of homogeneity in the 'pendent' drop by surface-tension effects and diffusion would be unacceptably slow (Rosenberger, 1986).

The observed lack of mixing was especially evident with fluids exhibiting large interfacial tensions, which would tend to stabilize the boundary between the two solutions. However, even between solutions with interfacial tensions so low that an interface could not be obtained in 1-g (*e.g.*, between the MPD and haemoglobin solutions), little mixing was observed in microgravity. In 1-g it is not possible to grow pendent drops larger than approximately 20 μl, and therefore ground control experiments were necessarily conducted using much smaller drops than those used during μ-g tests. In 1-g, the fusion of a protein drop with a drop of any of the three precipitant solutions rapidly resulted in apparent homogeneity, suggesting that density-driven convection plays a major role in passive mixing. The very simple data collection system employed in the present study precludes the extraction of true quantitative information on the behaviour of these fluids in microgravity, but it is evident that in a space protein crystallization facility, up to a hundred or more protein/precipitant solution pairs must be actively premixed prior to the formation of the 'pendent' crystallization droplets. The approach presently used (DeLucas *et al.*, 1989) involves repeated extrusion and withdrawal of the two solutions through the injector, but this action is liable to introduce air bubbles into the crystallization drop. A different approach would be to co-extrude the protein and precipitant solutions through a miniature pre-mixing chamber immediately upstream of the injector. Mixing in the chamber would be achieved by a series of small static baffles. However, a potential limitation of this approach would be the cost of machining several hundred such mixing chambers. A second alternative might be to install a tuneable piezoelectric transducer at the tip of each injector. Each transducer would vibrate at a frequency which would effectively ultrasonicate the drops: the molecules at the protein/precipitant interface would be excited, thus making the

boundary unstable and thereby promoting mixing. The transducers would need to be individually tuned, as a different frequency may be required to destabilize each protein/ precipitant interface without destabilizing the outer miniscus, which could fragment the drop.

A.5 REFERENCES

Ataka, M. and M. Asai (1988). Systematic studies on the crystallization of lysozyme: determination and use of phase diagrams. *J. Cryst. Growth* **90**, 86-93.

Bowers, R. and W. Zisman (1969). Surface Properties. (in) *Engineering design for plastics*. (E. Baer, ed.), Reinhold Publishing Co., London, pp. 691-705.

Broom, M. B., W. Wintherow, R. S. Snyder and D. C. Carter (1988). Preliminary observations of the effect of solutal convection on crystal morphology. *J. Cryst. Growth* **90**, 130-135.

Bugg, C. E. (1986). The future of protein crystal growth. *J. Cryst. Growth* **76**, 535-544.

Carter, C. W., Jr. and C. W. Carter (1979). Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.* **254**, 12219-12223.

DeLucas, L. *et al.* (1986). Preliminary investigations of protein crystal growth using the space shuttle. *J. Cryst. Growth* **76**, 681-694.

DeLucas, L. *et al.* (1989). Protein Crystal Growth in Microgravity. *Science* **246**, 651-654

Fowlis, W. W., L. DeLucas, P. Twigg, S. Howard, E. Meehan and J. Baird (1988). Experimental and theoretical analysis of the rate of solvent equilibration in the hanging drop method of protein crystal growth. *J. Cryst. Growth* **90**, 117-129.

Garfield, J. and C. Stout (1988). Crystallization and preliminary X-ray diffraction studies of a toxin crystal protein from a subspecies of *Bacillus thuringiensis*. *J. Biol. Chem.* **263**, 11800-11801.

Giege, R., B. Lorber, V. Mikol, D. Moras, M. Ruff, A. Theobald and J. Thierry (1988). Crystal growth of biological macromolecules and crystallization of proteins in weightlessness. *Bull. Inst. Pasteur*, **86**, 9-20.

Littke, W. and C. John (1984). Protein single crystal growth under microgravity. *Science* **225**, 203-204

Pusey, M., W. Witherow and R. Naumann (1988). Preliminary investigations into solutal flow about growing tetrgonal lysozyme crystals. *J. Cryst. Growth* **90**, 105-111.

Rosenberger, F. (1986). Inorganic and protein crystal growth - similarities and differences. *J. Cryst. Growth* **76**, 618-636.

Schoen, E. and F. Seifert (1987). The protein crystallization facility for EURECA. Proceedings of the 38th Congress of the International Astronautical Federation, Brighton, England.

Tiller, W. A. (1986). Thermodynamic and kinetic considerations for crystal growth of complex molecules from solution. *J. Cryst. Growth* **76**, 554-561.

Appendix B.

Claims to original research

1. Characterization of the papain- and trypsin-generated toxins from the crystal protein of *Bacillus thuringiensis* subsp. *kurstaki* HD-73.

2. Determination of the secondary structure of toxin from *Bacillus thuringiensis* subsp. *kurstaki* HD-73.

3. Elucidation of the domain structure of the toxin from *Bacillus thuringiensis* subsp. *kurstaki* HD-73.

4. Demonstration that the unfolded protoxin and unfolded toxin refold to their native and biologically active conformations following treatment with chemical denaturants. To the best of my knowledge, the protoxin is the largest single polypeptide chain shown to regain its native conformation after being unfolded in denaturants.

5. Demonstration that the N-terminal and C-terminal halves of the protoxin molecule unfold and fold independently of each other.

6. First evidence that the toxic moiety in protoxin is present in an active conformation.

7. First demonstration that the cysteine sulfhydryl groups of the protoxin can be modified without loss of biological activity.

8. First demonstration that the lysine amino groups can be modified without loss of biological activity.

9. Development of chemically modified protoxins with greatly improved solubility properties.

10. Demonstration that Teflon and silicone rubber are ideal materials for producing 'pendent' protein drops in microgravity.

Appendix C.

PUBLICATIONS ARISING FROM THIS THESIS

Refereed Publications:

Choma, C. T., W. K. Surewicz and H. Kaplan. The effect of salt on the thermal stability of the toxin and protoxin from *Bacillus thuringiensis*: a calorimetric study (in preparation).

Choma, C. T. and H. Kaplan. *Bacillus thuringiensis* crystal protein: effect of chemical modification of the cysteine and lysine residues (submitted).

Choma, C. T. and H. Kaplan (1990). Folding and unfolding of the protoxin from *Bacillus thuringiensis*: evidence that the toxic moiety is present in an active conformation. *Biochemistry* (in press).

Choma, C. T., W. K. Surewicz, P. R. Carey, M. Pozsgay, T. Raynor and H. Kaplan (1990). Unusual proteolysis of the protoxin and toxin from *Bacillus thuringiensis*: structural implications. *Eur. J. Biochem.* **189**, 523-527.

Choma, C. T., W. K. Surewicz, P. R. Carey, M. Pozsgay and H. Kaplan (1990). Secondary structure of the entomocidal toxin from *Bacillus thuringiensis* var. *kurstaki* HD-73. *J. Prot. Chem.* **9**, 87-94.

Bietlot, H., P. R. Carey, **C. T. Choma**, H. Kaplan, T. Lessard and M. Pozsgay (1989). Facile preparation and characterization of the toxin from *Bacillus thuringiensis* var. *kurstaki*. *Biochem. J.* **260**, 87-91.

Scientific Presentations:

Choma, C. T. and H. Kaplan. Folding/unfolding of the protoxin and toxin from *Bacillus thuringiensis* subsp. *kurstaki*. Fourth annual meeting of the Protein Society, San Diego, California, August 1990.

Choma, C. T. Solution behaviour in microgravity: considerations for protein crystal growth. NRCC second workshop on microgravity experiments, Ottawa, May 1990.

Choma, C. T., W. K. Surewicz, T. Raynor, P. R. Carey, M. Pozsgay and H. Kaplan. Domain structure of the protoxin and toxin of *Bacillus thuringiensis* var. *kurstaki*. Third annual meeting of the Protein Society, Seattle, Washington, August 1989.

Rucker, C., W. Kung and **C. T. Choma.** Design for a protein crystallization facility for Space Station. NRCC 'Spacebound' 89: R and D Opportunities On Board the Space Station.' Ottawa, May 1989.

Pozsgay, M., P. R. Carey, **C. T. Choma** and T. Lessard. Structure-function studies of the insecticidal protein from *Bacillus thuringiensis*. The enzyme catalysis process: Energetics, mechanism and dynamics. NATO Advanced Studies Institute Conference, Barga, Italy, July 1988.

Patent Application:

Choma, C. T. and H. Kaplan. Improved bioinsecticides from *Bacillus thuringiensis*. Canadian patent application submitted to Kirby, Eades, Gale, Baker and Potvin, Patent and Trademark Agents, Ottawa (8 August, 1990).

Appendix D.

GENE NUCLEOTIDE SEQUENCE AND AMINO ACID ANALYSES OF PROTOXIN AND TOXIN FROM *B.T.* SUBSP. *KURSTAKI* HD-73

Table D.1 Amino acid analysis of *B.t.* subsp. *kurstaki* HD-73 protoxin*

amino acid	# residues
Ala	66
Arg	73
Asp	61
Asn	85
Cys	16
Glu	91
Gln	45
Gly	82
His	22
Ile	74
Leu	94
Lys	34
Met	10
Phe	54
Pro	53
Ser	91
Thr	67
Trp	19
Tyr	57
Val	84

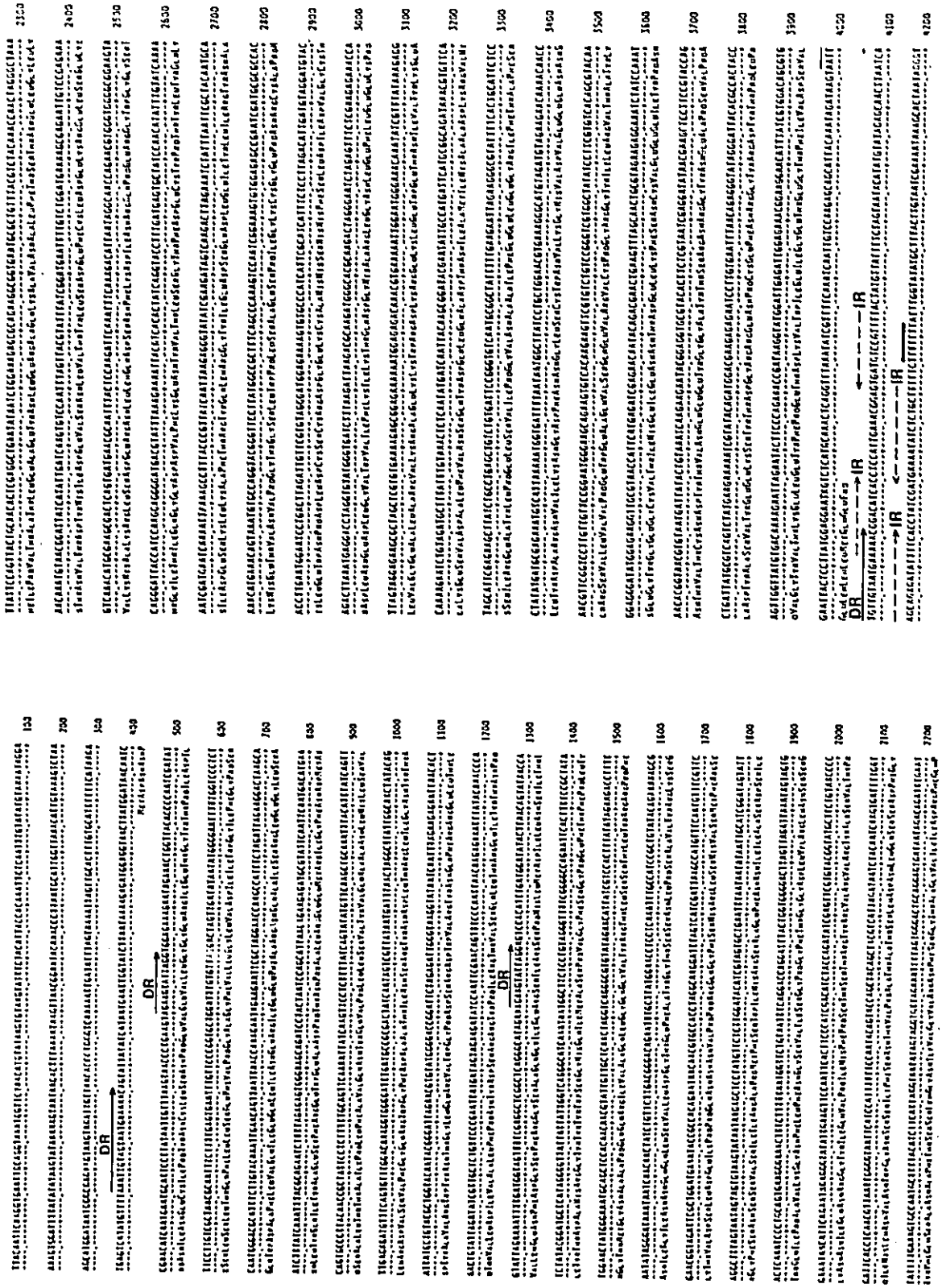
* Deduced from gene nucleotide sequence (Adang *et al.*, 1985, *Gene* **36**, 289-300.)

Table D.2 Amino acid analysis of *B.t.* subsp. *kurstaki* HD-73 toxin*

amino acid	# residues
Ala	37
Arg	43
Asp+Asn	67
Cys	0
Glu+Gln	54
Gly	43
His	9
Ile	45
Leu	48
Lys	3
Met	7
Phe	36
Pro	30
Ser	60
Thr	36
Trp	10
Tyr	27
Val	39

* From Bietlot *et al.*, 1989, *Biochem. J.* **260**, 87-91

Fig. D1 Gene nucleotide sequence of the protoxin gene from *B.t.* subsp. *kurstaki* HD-73 (from Adang *et al.*, 1985, *Gene* 36, 289-300).



If you know what I mean

A poet should be of the
old-fashioned meaningless brand:
obscure, esoteric, symbolic,-
the critics demand it;
so if there's a poem of mine
that you do understand
I'll gladly explain what it means
till you don't understand it.

-Piet Hein