Title
Engineering of bovine pancreatic trypsin inhibitor for NMR and protein folding

Author
Altman, John David

Publication Date
1991

Peer reviewed|Thesis/dissertation

Engineering of Bovine Pancreatic Trypsin Inhibitor

for NMR and Protein Folding

by

John David Altman

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

Date **University Librarian**

Degree Conferred: 6/16/91

To Leslie

Acknowledgements

I tend to regard the debts that are incurred during graduate education as a form of reverse Social Security: we take from the system when we are young and inexperienced and give back later in life. The analogy breaks down on at least two accounts. We never stop learning and "taking" from others throughout our scientific careers. In choosing a career in science, we begin by admiting our ignorance and we express a continuing desire to learn, through our own endeavors and through the work and ideas of others. Secondly, we do not write and receive "checks" to and from a faceless bureaucracy; rather the debts we owe—and will hopefully honor—are happily very personal. I cannot begin to pay back directly all those who have aided me during the course of my graduate education. I do promise, however, to attempt to aid others as I have been aided. And of course I can acknowledge in a personal way those whose contributions to my education have been particularly important.

I begin by noting that my graduate career was unique—hopefully all facets of all scientific careers are unique—but I believe that my experience was well outside the norm, at least for the Department of Pharmaceutical Chemistry at UCSF. For several years, I carried out most of my work on the expression of BPTI on the campus of Genentech, ten miles south and east of UCSF, in their departments of Biomolecular Chemistry and later Protein Engineering. For this, I am in debt to many, first for the opportunity to have this experience, and second for what I learned while working there. At the top of the list are my research adviser, Tack Kuntz, for arranging this collaboration and for having the confidence in me to be productive and independent while away from UCSF; Stephen Anderson, in whose lab I began my work at Genentech, and who is now a professor at Rutgers University; and finally Tony Kossiakoff, the director of the department of Protein Engineering at Genentech, who permitted me to finish my work at Genentech, even after Steve left for Rutgers. Steve and Tony did much more than provide me with facilities and

lab space; they have been, and continue to be, important collaborators and teachers whose opinions I value immensely.

I owe Jim Wells more than my gratitude for encouraging me to use Dennis Henner's pDH108 plasmid expression vector and for being a model of enthusiasm, productivity, creativity, and intelligence—I still owe him three bottles of Anderson Valley Boont Amber Ale for winning the bet that the pDH108 based expression system for BPTI (which became pHAZY in my hands) would prove to be better than the original secretion vectors. I have not forgotten, Jim. I am grateful to Dennis Henner for providing me with the pDH108 vector and for always being available and helpful during the preparation of my **Protein Engineering** manuscript.

At various times during my stay at Genentech, I shared lab space with Cara Berman-Marks, Mark Zoller, John Gill, Paul Carter, and Bob Kelley. Much of what I know about molecular biology I learned while working in the trenches with all of them. I am also indebted to Steve Bass and Brian Cunningham of the Wells lab for their expertise and enthusiasm, and to Mike Randal and Charlie Eigenbrot of the Kossiakoff lab for sharing results from BPTI crystal structures.

I conclude the Genentech portion of the acknowledgements with two scientists from Sweden who were my closest colleagues and friends all through the time I spent at Genentech, Björn Nilsson and Lars Abrahmsén. Björn was a post-doc in the Anderson lab and taught me how to get bacteria to make proteins from cows. I also shared with him many discussions of protein folding and secretion, the Giants, and Muddy Waters.

Lars arrived in the Wells lab from the RIT at almost exactly the same time as Björn's departure. Although I did not directly collaborate with Lars as I had with Björn, Lars and I shared many scientific and personal interests. I have rarely encountered someone as easy to talk with and as personally generous as Lars.

I have shared my interest in BPTI with many in the Kuntz group at UCSF, and I have learned from all of them. Among this group are Mark Hurle, who also worked with me in the Anderson lab at Genentech, Charlie Eads, and Phyllis Kosen.

My work could not have been carried out without a first rate NMR facility. For this, I am grateful to Tack Kuntz, Tom James (also a member of my thesis committee), and Vladimir Basus, and to Mark Day and Don Kneller for their heroic efforts at software development and computer system administration. I also thank Norm Oppenheimer, the third member of my thesis committee, especially for encouraging me to pursue the random fractional deuteration project. I am also grateful to Norm for chairing my orals committee, and for helping me over that hurdle.

I have already mentioned my research advisor, Tack Kuntz, in several contexts. I am most grateful to him for his role as a teacher and for the breadth of his intellectual support. I look forward to continuing this relationship as my career progresses.

My personal debts are in many ways more important than my scientific ones. I cannot begin to enumerate the enriching experiences I have shared with Brian Shoichet, also a member of the Kuntz lab. Brian will be the best man at my wedding, which will occur scarcely a week after this thesis is filed. Brian's influence runs unseen throughout this document, but I know that it is there.

My parents, Sandy and Larry Altman, have frequently struggled to understand what my graduate studies were all about, but they have always been supportive. Though they are not scientists, they have instilled in me values which led to my choice of science as a career: the importance of worldly curiousity and learning and the value of truth and integrity. These values guide me in all aspects of my life. I cannot begin to repay them for all they have given me. My accomplishments are a tribute to them. They are also a tribute to my grandparents Rose and Sydney Bowman, whose confidence in me has never wavered.

Finally, I dedicate this thesis to my fiancée, Leslie Taylor. Leslie has never failed to stand by me with support , encouragement, and patience. The filing of this thesis accidentally coincides with our wedding. A doctoral thesis represents the end of one phase of a scientific career and the beginning of another, just as a wedding represents the end of one phase of a relationship and the beginning of the next. I now look forward to both exciting changes in my life, and our lives together.

Engineering of Bovine Pancreatic Trypsin Inhibitor for NMR and Protein Folding

John David Altman

Protein structure determination by nuclear magnetic resonance spectroscopy is now well established. However, the best structures derived from NMR data do not define the atomic coordinates of proteins with the same precision as is possible using the older technique of x-ray crystallography. Many recently introduced methods for increasing the precision of protein structures determined by NMR spectroscopy require the production of isotope labeled proteins. Most often, these proteins are produced by a heterologous expression system in the bacterium Escherichia coli, grown in a defined minimal medium. Here, a previously described secretion expression system for bovine pancreatic trypsin inhibitor (BPTI) (Marks, C. B., *et al.* (1986). *J. Biol. Chem.*, **261**, 7115-7118.) was found to be inadequate for the production of BPTI labeled with deuterium or ^{15}N. A new intracellular expression system for BPTI was constructed, leading to the production of BPTI fusion proteins that were found in insoluble inclusion bodies. The BPTI fusion proteins were refolded prior to cleavage and affinity purification by chymotrypsin using a novel single column procedure. Mutants of BPTI containing a single free thiol were constructed for specific modification of the protein with a nitroxide spin label. Distances from the paramagnetic center to specific protons in spin labeled BPTI were measured by the paramagnetic contribution to the line width. Data from two dimensional ^{1}H NMR experiments were not consistent with a single site model for the position of the spin label. Further improvements in the spin labeling methodology and NMR experiments for the extraction of long range distance information are discussed.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction: The problem of the precision of protein structures determined by NMR spectroscopy

1.1. Why Do We Determine Protein Structures?

Most important biological functions are mediated by linear chains of amino acids called proteins, and the function of a specific protein is dependent on the three-dimensional fold of the linear chain. I can think of no better example that illustrates the relationship between protein structure and function than the human class I histocompatibility antigen HLA-A2, a membrane bound protein responsible for "presenting" a wide range of peptides to other actors in the immune system. Exactly how so many peptides are recognized by a single protein was a mystery until the solution of the three-dimensional structure of the soluble domains of the HLA-A2 protein, an event which transformed the field of molecular immunology (Bjorkman *et al.*, 1987).

The HLA-A2 structure was solved by a technique known as x-ray crystallography. Until recently, x-ray crystallography was the only technique available for solving protein structures. However in 1983, the structure of micelle-bound glucagon was solved using data from a technique known as nuclear magnetic resonance spectroscopy (NMR) (Braun *et al.*, 1983); this was followed in 1985 by the first report of the structure of a globular protein in solution, bull seminal proteinase inhibitor IIA (BUSI IIA) (Williamson *et al.*, 1985). Since then, the rate at which solution structures of proteins have been solved by NMR has increased rapidly. X-ray crystallography and NMR spectroscopy remain the only methods capable of determining the structures of proteins to atomic resolution.

The resolution of a structure determination is related to the precision with which the atomic coordinates are known. Unfortunately, differences in methodology make a direct comparison of the resolution of structures determined by x-ray crystallography and NMR impossible; in fact, most researchers seem to avoid the comparison altogether. Nevertheless, early NMR structures were said to be equivalent to 3 Å crystal structures, while some more recent structures are claimed to be equivalent to 2.0-2.5 Å x-ray crystal structures (Wüthrich, 1989; Clore and Gronenborn, 1991b). X-ray crystal structures are

commonly better than this; for example, the structure of the C30A/C51A mutant of bovine pancreatic trypsin inhibitor (BPTI) was recently solved to 1.6 Å (Eigenbrot *et al.*, 1990).

Frequently, a "medium" resolution structure is all that is required to reveal the secrets behind a protein's biological function. For example, the mechanism of antigen presentation by the HLA-A2 protein was quite clear in the initial 3.5 Å crystal structure (Bjorkman *et al.*, 1987). However, many applications, including iterative design of new protein catalytic functions (Carter and Wells, 1987) and of enzyme inhibitors (DesJarlais *et al.*, 1990), benefit from structures which are determined to the highest resolution possible. Clearly, there is a need for methods which allow more precise definitions of protein structure by NMR.

The major theme of this thesis is the development of methods to increase the precision of protein structure determination by nuclear magnetic resonance spectroscopy. In particular, I have chosen to focus on the extraction of a set of constraints that have not previously been applied to a full solution structure determination, namely the long range distance constraints available from NMR experiments with spin labeled proteins (Schmidt and Kuntz, 1984). A model experimental system was required for these studies, and for this I chose the protein BPTI. Unfortunately, the path from choice of a model system to the production of proteins suitable for spin label analysis was not a straight one, and I was forced to develop a new heterologous expression system for BPTI. The development of this expression system therefore became a co-equal major theme of this thesis, underscoring the important contribution that developments in biotechnology have made to the field of biophysics. Other themes include the development of mutants suitable for high yield, specific modification by spin label reagents and the possible utility of some of these mutants for protein folding studies.

Since I began the work of this thesis, a large number of important methods for increasing the precision of structures determined by NMR have been described. In this first chapter, I will review a number of the most successful and/or promising methods and at-

tempt to place spin labeling in the context of these other developments, highlighting areas of overlap. I will begin, however, with some historical notes, an outline of the complete procedure for solution structure determination, and a more detailed explanation of what is meant by the precision of a protein structure.

1.2. Protein NMR in a historical context

The field of protein structure determination by protein NMR has experienced two revolutions in a time span of less than a decade. The hallmark of the first revolution was the first complete sequence specific assignment for a protein—BPTI—published in 1982 (Wagner and Wüthrich, 1982b). Developments leading up to this achievement began as early as 1976 with the first practical demonstration of two-dimensional NMR (Aue *et al.*, 1976), but it is still the sequence specific resonance assignment which allows the description of protein structure to atomic detail. Under the regime of the first revolution, which I shall call the homonuclear era, chemical shift assignments could routinely be obtained for proteins with less than approximately 100 residues, although some larger proteins were assigned using only two-dimensional NMR methods. A two-year-old table of specifically assigned proteins is given in a review by Markley (1989) (Markley, 1989); the application of two-dimensional NMR to protein structure is thoroughly reviewed in the book by Wüthrich (1986)(Wüthrich, 1986).)

The hallmark of the second revolution was the introduction of heteronuclear three-dimensional NMR (Fesik and Zuiderweg, 1988; Marion *et al.*, 1989b) (recently reviewed in (Fesik and Zuiderweg, 1990)). Further advances in the period following the second revolution, which I shall call the heteronuclear era, have allowed the chemical shift assignment and structure determination of proteins up to 17 kiloDaltons (kDa) in molecular weight (Clore *et al.*, 1991b), with similar data for proteins as large as 30-40 kDa expected in the future. The second revolution did not make obsolete the methods introduced during

the homonuclear era; indeed, quite excellent structures of small proteins are still determined using only homonuclear two-dimensional ^{1}H NMR (Omichinski *et al.*, 1990), although application of new heteronuclear techniques can lead to significantly better structures for small proteins (Zuiderweg and Fesik, 1989; Güntert *et al.*, 1991b). In summary, increasing the size of proteins amenable to analysis by NMR, together with increasing the precision of the structure determination, constitute the two primary goals of method development in protein NMR spectroscopy (Wider *et al.*, 1989; Clore and Gronenborn, 1991b).

1.3. Protein structure determination by NMR: A coarse outline

The earliest complete outline of the procedures necessary to determine a protein structure in a solution environment were presented by Wüthrich in 1982 (Wüthrich *et al.*, 1982). After the acquisition of spectral data, there are three basic steps—and of course many often laborious substeps—in an NMR structure determination: (1) sequential assignment of the resonances from each proton in the molecule; (2) extraction of specific structural constraints, including short interatomic distances and torsion angles; and (3) construction of model structures compatible with the input constraints using a mathematical algorithm. These three steps apply equally in the homo- and heteronuclear eras, with changes only in the details.

In the homonuclear era, isolated proton spin systems—bounded by peptide bonds and therefore exactly corresponding to amino acids—were identified in experiments which correlated protons through bonds (*e.g.*, the COSY experiment (Aue *et al.*, 1976)) and were then linked in sequence by spatial connections observed in NOESY spectra (Jeener *et al.*, 1979; Kumar *et al.*, 1980a). Sequential assignments in the heteronuclear era are based on more complicated experiments (Ikura *et al.*, 1990; Kay *et al.*, 1990b), but the assignments derived from them are more robust because they do not rely on through space interactions,

which are interpreted based on a statistical argument (Billeter *et al.*, 1982). Assignments are made for all ^{13}C and ^{15}N atoms, primarily as an aid for assignment of the proton spectrum. As I shall discuss below, the bulk of the structural information is derived from proton-proton correlations.

The second step in the structure determination procedure is the identification of proton pairs which are close in space. Protons which are close in space may be correlated through a magnetization transfer process mediated by the nuclear Overhauser effect; these correlations are observed in NOESY spectra. In the absence of chemical shift degeneracy, NOESY cross-peaks may be assigned by inspection of a table of assigned chemical shifts. In the presence of resonance overlap, which is quite severe in large proteins, one chooses among several degenerate resonance assignments by reference to a structural model or by correlation of one or both protons of the pair to the chemical shift of an attached heteronucleus. Based on the intensity of a NOESY peak, the through space correlation is translated to one of three or four distance ranges. The lower limit is usually the sum of the van der Waals radii, while the upper limit can be as much as 5 Å (Wüthrich, 1986); these numbers are entered into the lower and upper bounds limit matrices, respectively. NOEs between protons may be observed in two-, three-, or four-dimensional NMR spectra. NOEs between protons and heteronuclei are not useful for two reasons: (1) the magnitude of the NOE is a function of the square of the gyromagnetic ratio of each nucleus in the pair (Noggle and Schirmer, 1971), and the gyromagnetic ratios for ^{13}C and ^{15}N are approximately 4 and 10 times smaller than protons, respectively, and (2) the predominant NOE to a heteronucleus is from an attached proton, and this carries no conformational information. Structural information can also be derived through measurement of vicinal coupling constants which place constraints on the intervening torsion angle via an empirical Karplus relation (Karplus, 1963).

The third step in the structure determination is the construction of structural models that are consistent with all of the input constraints using a mathematical algorithm such as

distance geometry (Kuntz *et al.*, 1979; Kuntz *et al.*, 1989) or the pseudo-energy based approach of restrained molecular dynamics (Kaptein *et al.*, 1985; Scheek *et al.*, 1989) or simulated annealing (Nilges *et al.*, 1988a; Nilges *et al.*, 1988). The basic elements of these procedures have also been implemented in a combined approach (Nilges *et al.*, 1988b).

In the metric matrix distance geometry algorithm, the lower and upper bounds distance matrices are first smoothed using procedures such as the triangle inequality, and a distance for each pair of protons between the smoothed bounds range is randomly chosen, creating a single, symmetric distance matrix. This procedure is repeated many times—usually between 10 and 100—creating a number of distance matrices. Each of these matrices are then translated into individual three dimensional structures using a procedure called embedding; the embedded structures are further minimized with respect to constraint violations. At the end of the procedure, one is left with a number of structures, all of which are stereochemically correct and compatible with the input constraints within some level of tolerance. The notion of the precision of an NMR structure arises from a consideration of the deviation of these structures from an average structure.

The basis of the molecular dynamics-based approaches is the introduction of a pseudo potential that is meant to simulate the observed NOE distance constraint. In the molecular dynamics approaches, one takes snapshots from late time points—ensuring that the conformations are near some local minimum—from single or multiple runs. Many different conformations will be represented for two reasons: (1) the pseudo potentials corresponding to the observed distance range are flat over that range, and (2) conformational sampling is an inherent feature of molecular dynamics.

1.4. Constraints and the precision of protein structures

One exact constraint per degree of freedom is necessary to describe the conformation of a molecule. In cartesian space, this corresponds to 3n constraints, where n is the

number of atoms in the molecule; the number is substantially less when using internal coordinates. The position of atoms in a protein chain are also correlated, due to the presence of relatively fixed bond lengths and angles. These represent conformation independent constraints on the protein structure.

The number of inexact constraints—such as might be available from the measurement of an NOE between two protons—necessary to describe a protein structure is a fuzzy concept. The intuitive notion that the more constraints, the better the structure, is certainly correct (Havel and Wüthrich, 1985), but no precise mathematical relationship may be formulated. NMR spectroscopists often talk about the number of interresidue NOEs observed per amino acid residue (Wagner, 1990; Forman-Kay *et al.*, 1991). This number varies widely within a protein sequence, with values of up to 40 found in the interior of a protein, while 2-3—mostly sequential—are found for surface loops. Similar numbers are also found in model calculations starting with crystal structure data, indicating that it is not necessary to invoke motional models to explain the variation in the number of NOEs per residue; a simple consideration of proton density may be all that is necessary.

I shall further illustrate the relationship between constraints and the precision of a structure determination, focusing on the distance geometry algorithm. A large number of constraints per residue effectively limits the difference between the upper and lower bounds in the smoothed bounds matrices for many atoms in the vicinity. This reduces the randomly chosen distances within this range to a small region of distance space, and when structures are finally embedded, a smaller region of real space is compatible with the input constraints.

The precision of an NMR structure determination is commonly expressed as the root mean square difference (RMSD) in the atomic coordinates for a collection of a large number of structures—all of which are compatible with the available constraints—from the average coordinate position for all structures (Havel and Wüthrich, 1985). Typically, the contribution from the comparatively well-determined backbone is expressed separately

from the RMSD for all atoms; with experimental NMR data, a range of 1 to 2 Å is found for the backbone atoms, while a range of 1.5 to 3 Å is observed for all atoms (Clore and Gronenborn, 1989a).

X-ray crystallographers do not normally speak of the precision of the structure determination. Rather, they generate a single three dimensional structure and talk about its resolution—directly related to the minimum interatomic spacing that gives rise to reflections they can measure from a single crystal (Creighton, 1984b)—and its accuracy, expressed in the form of the well known crystallographic R factor which is a precise, if one-dimensional, description of how well the structural model fits the diffraction data. The development of methods for describing the accuracy of an NMR structure determination, the NMR R factor, is both young (Borgias and James, 1988; Borgias and James, 1989; Gonzalez *et al.*, 1991) and controversial (Clore and Gronenborn, 1989b), but I expect that a widely accepted standard will emerge. The precision of an NMR structure determination is analogous to the crystallographic resolution, but it will never be as mathematically simple. The concepts found in the phrases, "the crystal structure of JDase was determined to a resolution of 2.2 Å" and "the all-atom RMSD of the 35 NMR structures calculated for JDase was 2.5 Å" will probably never merge.

1.5. Obtaining more NMR constraints

It is well known that improvement in the resolution of structures obtained from x-ray diffraction from single crystals requires collection of additional data at higher diffraction angles. Due to the three dimensional nature of the problem, doubling the resolution requires the collection of 8 times the number of data points. Improving the resolution of solution structures from NMR data also requires the collection of additional data. These data will come from the following sources, not necessarily in order of importance, which will be discussed in the sections to follow: (1) higher magnetic fields, (2) determination of

additional dihedral angle constraints, using the relationship between vicinal coupling constants, ^3J, and the dihedral angle, (3) stereospecific assignments of prochiral atoms, (4) extraction of additional NOEs using isotope editing techniques and higher dimensional spectroscopy to remove ambiguity problems due to spectral overlap, (5) extraction of more quantitative information from previously assigned NOEs, (6) determination of additional hydrogen bond constraints, and (7) determination of long range distance information using the powerful relaxation properties of paramagnetic spin labels. Of course, not all of the members of this list are completely independent of the other members; *e.g.* the determination of stereospecific assignments for diastereotopic β protons requires the measurement of at least one dihedral angle and several individual coupling constants. I shall discuss all of these methods and their interrelationships in the sections to follow. Finally, I will present my motivation for choosing to focus on data obtained from spin labeled proteins.

1.5.1. Increased Magnetic Field Strength

The initial NMR spectra of proteins (Saunders *et al.*, 1957) clearly show conformation dependent features (Kowalsky, 1962; Kowalksy, 1965; McDonald and Philips, 1967), but in general the resonances could not be resolved and assigned to specific residues due to spectral overlap. Chemical shift dispersion is a function of magnetic field, and advances in superconducting technology have allowed the construction of magnets with much higher magnetic field strengths. The field strength standard in the early 1980's was equivalent to a proton resonance frequency of 500 MHz; in the late '80's, magnets with proton resonance frequencies of 600 MHz were introduced and have since become increasingly common. The greater chemical shift dispersion at 600 MHz will allow more confident assignment of NOE cross peaks which were overlapped or otherwise chemically shift degenerate in 2D or 3D spectra at 500 MHz. Higher field strengths also increase the sensitivity of NMR experiments, allowing the measurement of a greater number of weak NOEs and/or the use of more dilute solutions—the current limit is not far from 1 mM, even at 600

MHz. This will be important for experiments on higher molecular weight proteins where solubility may become a problem (Forman-Kay *et al.*, 1989).

1.5.2. Dihedral Angle Constraints

1.5.2.1. Measurement of $^3J_{HN\alpha}$

The relationship between the vicinal coupling constant ^3J and the intervening dihedral angle θ was first described by Karplus (Karplus, 1959); the Karplus equation is quadratic in $\cos\theta$ and must be calibrated for particular chemical systems by fitting the coupling constant data to independently determined conformational data (Bystrov, 1976). The most recent calibration for the angle ϕ in proteins—describing rotation about the N–C^α bond ($\theta = |\phi - 60\ ^\circ|$) and related to the coupling constant $^3J_{HN\alpha}$—utilized data from BPTI (Pardi *et al.*, 1984). In applying this relationship to proteins of unknown structure, ambiguities may arise because the angle ϕ is not a single valued function of $^3J_{HN\alpha}$. However, only a restricted range of ϕ angles is normally found in proteins (Richardson, 1981; Wüthrich, 1986), removing some of the ambiguity.

In all but the smallest proteins, it will not be possible to measure $^3J_{NH\alpha}$ in one dimensional experiments, even if they are resolution enhanced. Instead, the coupling constant is measured by examination of the antiphase cross peak fine structure in phase sensitive COSY or double quantum filtered COSY (DQF-COSY) experiments (Marion and Wüthrich, 1983; Rance *et al.*, 1983). In samples where the linewidths approach the value of the coupling constants, the observed separation between the relative extrema in the antiphase cross peak will not reflect the true coupling constant due to a phenomenon known as antiphase cancellation (Neuhaus *et al.*, 1985). However, the true coupling constant can sometimes be determined using spectral simulations which vary the linewidth and relevant coupling constants (Widmer and Wüthrich, 1986; Widmer and Wüthrich, 1987; Smith *et al.*, 1991).

Several other techniques have been introduced for the measurement of $^3J_{HN\alpha}$ in ^{15}N labeled proteins. Montelione and Wagner have recently extended their technique for measurement of long range heteronuclear coupling constants (discussed below) to a technique that is capable of accurately measuring small homonuclear coupling constants (Montelione and Wagner, 1989). This technique is quite successful with labeled peptides but has not yet been applied to ^{15}N labeled proteins; its utility with large proteins may be limited because of its reliance on coherence transfer from $^{13}C_\alpha$ to ^{15}N with a $^1J_{C(\alpha)N} \approx 11$ Hz (Montelione and Wagner, 1989; Sørensen, 1990). Wüthrich's group has introduced a method for determination of $^3J_{HN\alpha}$ within ranges necessary for assignment of the ϕ angle to common secondary structures—less than 6.0 Hz for α-helices, greater than 8.0 for β-sheets, or between the two for conformationally averaged structures (Neri *et al.*, 1990). Finally, a new experiment named HMQC-J (heteronuclear multiple quantum coherence - J) was introduced by the NIH group (Kay *et al.*, 1989; Kay and Bax, 1990). Even though the experiment gives significantly narrower multiple quantum line widths in the F_1 dimension compared to normal homonuclear experiments, spectral simulations and nonlinear optimizations were required to relate the apparent coupling constants to the true values (Forman-Kay *et al.*, 1990). Nevertheless, Forman-Kay *et al.* claim that "the accuracy far exceeds that of the apparent coupling constants measured directly from homonuclear COSY-type experiments without correction for the broad homonuclear NH line widths.... The lower limit for measurable couplings in the HMQC-J experiment is well below that of the COSY experiments due to the intrinsically narrower multiple quantum line widths."

1.5.2.2. Measurement of $^3J(H^\alpha, H^{\beta x})$

The E.COSY (Griesinger *et al.*, 1985) or P.E.COSY (Mueller, 1987) experiments can be used to measure accurately homonuclear coupling constants between two of three mutually coupled nuclei—such as the α and β protons in an AMX side chain spin system—without the complication of antiphase cancellation; the experiments are limited only

by the digital resolution of the spectrum and resonance overlap. Together with qualitative intra- and interresidue NOE information, these measurements can provide information on the χ_1 dihedral angle. This will be discussed in the next section on stereospecific assignments. The α-amide coupling constants in glycine can also be measured with this technique.

1.5.3. Stereospecific Assignments

1.5.3.1. Effect of stereospecific assignments on the precision of calculated structures

In the absence of stereospecific assignments for diastereotopic protons, it is necessary to use pseudoatoms (Wüthrich *et al.*, 1983; Wüthrich, 1986) located at the average position of the two atoms when calculating structures using any of the common algorithms such as metric matrix distance geometry (Kuntz *et al.*, 1979; Kuntz *et al.*, 1989). A correction to the NOE distance constraint, equal to the maximum possible error, must be added. Similar pseudo structures must be employed for the two terminal methyl groups in valine and leucine, with even larger corrections. A growing body of evidence suggests that the stereospecific assignment of β-methylene protons and the terminal methyl groups of valine and leucine significantly improves the precision of the structure determination (Driscoll *et al.*, 1989a; Driscoll *et al.*, 1989b; Folkers *et al.*, 1989; Holak *et al.*, 1989a; Holak *et al.*, 1989b; Kraulis *et al.*, 1989; Widmer *et al.*, 1989). Two systematic studies convincingly demonstrate the increase in precision using actual (Driscoll *et al.*, 1989c) and synthetic (Güntert *et al.*, 1989) NMR data respectively.

1.5.3.2. Homonuclear methods for stereospecific assignment

Several methods have appeared for obtaining stereospecific assignments; all methods use at least the $^3J_{\alpha\beta 2}$, $^3J_{\alpha\beta 3}$, and a quantitative interpretation of the intensity of intraresidue NOEs between the amide, α, and β protons (Hyberts *et al.*, 1987). Similar

methods have been described for stereospecific assignment of valine methyl groups (Zuiderweg *et al.*, 1985; Härd *et al.*, 1990). The dependence of the NOEs on the ϕ angle may not be adequately accounted for with either procedure (Basus, 1989; Wagner, 1990). Subsequent work showed that in favorable cases, the sequential NOEs between the β protons and the NH_{i+1} proton can help provide the correct stereospecific assignment (Arseniev *et al.*, 1988). Two computer programs have been described which obtain the stereospecific assignments by a conformational database search using the experimental coupling constants, intraresidue NOEs, and sequential NOEs as input (Güntert *et al.*, 1989; Nilges *et al.*, 1990). In principle, these programs account for a full range of ϕ, ψ, and χ_1 angles and should provide more accurate and a greater number of stereospecific assignments. Finally, stereospecific assignments have been made using global NOE information after initial three dimensional structures have been constructed, using either distance geometry (Weber *et al.*, 1988; Güntert *et al.*, 1991a) or dynamical simulated annealing (Habazettl *et al.*, 1990). These methods are capable of obtaining stereospecific assignments for glycine α protons, relatively immobile γ and δ methylene groups of long side chains such as arginines, and the terminal methyl groups of valine and leucine, in addition to the β methylene groups that have been assigned using only local information.

1.5.3.3. Selective isotope labeling for stereospecific assignments

Several methods have appeared which allow stereospecific assignment based on selective isotope labeling and a knowledge of the biosynthetic pathways in *E. coli* (LeMaster, 1987; Neri *et al.*, 1989) . These methods will most probably be superseded by global heteronuclear labeling methods, described below, which are more in the spirit of multidimensional spectroscopy.

1.5.3.4. Heteronuclear coupling constants, torsion angles and stereospecific assignments

Bystrov demonstrated a number of years ago that the determination of heteronuclear coupling constants between the amide ^{15}N and β protons would be valuable for the characterization of the χ_1 dihedral angle (Bystrov, 1976). Recent two dimensional homonuclear NMR experiments developed by Montelione and Wagner allow the measurement of these small coupling constants in ^{15}N labeled proteins (Montelione *et al.*, 1989); three dimensional analogs of this experiment have been successfully applied by Wüthrich's group (Wider *et al.*, 1989). Together with knowledge of the $\alpha\beta$ homonuclear coupling constants, these experiments not only allow the determination of the χ_1 angle, but as Wagner points out, they make "the stereospecific assignments of β-methylene protons trivial" (Wagner, 1990). Stereospecific assignments based on the heteronuclear coupling constants are more rigorous than any of the approaches which rely on measurement of NOEs, and I expect that the method will be widely applied in the future. Long range heteronuclear couplings between ^{13}C and ^{1}H nuclei can also be observed and will yield more reliable determinations of the dihedral angle ϕ (Wagner, 1990). An extension of this concept to a three-dimensional NMR experiment should substantially remove the difficulties in interpretation that arise from overlap with the two dimensional technique (Montelione *et al.*, 1989; Edison *et al.*, 1991).

1.5.4. Extraction of more NOEs

1.5.4.1. The homonuclear case

Following completion of the ^{1}H assignment procedure, long range NOEs can often be assigned by a lookup procedure in which the cross peak position is compared to a list of assignments (Eads and Kuntz, 1989). Frequently, more than one proton shares a single chemical shift (within a specified tolerance), and the spectroscopist must resort to other

techniques for making the NOE assignment. Sometimes this can be done by a consideration of line shapes shared by previously assigned cross peaks (Denk *et al.*, 1986). In the case of NOEs involving amide protons, removal of the chemical shift degeneracy can frequently be accomplished by acquiring NOE data at a second temperature (Basus, 1989). Finally, NOE assignments can be made by reference to a structural model, available from any of several sources, including the following: (1) an x-ray crystal structure, of the identical protein or a homolog, (2) a solution structure of a closely related mutant, or (3) a low resolution structure built from a positively assigned subset of the total available data (Clore and Gronenborn, 1989a). With the latter procedure, more precise solution structures can be built in an iterative manner.

1.5.4.2. Isotope labeling techniques

The chemical shift degeneracy problem can also be removed by suitable isotope labeling schemes. These may involve selective biosynthetic isotope labeling of specific amino acids, either with a heteronucleus such as ^{15}N or ^{13}C (see (Muchmore *et al.*, 1989) and references therein), or specific deuteration (*e.g.*, see (Anglister, 1990)). Global labeling of proteins with ^{15}N, ^{13}C, or random fractional deuteration have also been successfully applied to proteins which show severe proton chemical shift degeneracy (Gronenborn *et al.*, 1989; Marion *et al.*, 1989a; Torchia *et al.*, 1989; Forman-Kay *et al.*, 1990); efficient *E. coli* expression systems have been essential in most of these studies. Several new multinuclear NMR pulse sequences have been developed to take advantage of the newly available proteins. Some of these will be discussed below.

1.5.4.2.1. Heteronuclear labeling

Spectral editing by selective isotope labeling by amino acid residue type has a long history in NMR studies of proteins (Markley *et al.*, 1968); however, advances in heterologous protein expression technology have revolutionized the application of this technique, particularly in conjunction with heteronuclear NMR, to studies of protein structure

(McIntosh and Dahlquist, 1990). Beginning in the middle of the 1980's, several groups used selective ^{15}N labeling by amino acid residue type to obtain resonance assignments and NOE information (LeMaster and Richards, 1985; Torchia *et al.*, 1989; McIntosh *et al.*, 1990); studies of this type require the preparation of many isotopomers, frequently in a number of different auxotrophic bacterial strains (McIntosh and Dahlquist, 1990), and can be tedious. In most cases, they will probably be supplanted by methods utilizing global heteronuclear labeling.

More recently, heteronuclear 2D or 3D experiments have been devised to remove chemical shift degeneracies by the introduction of an ^{15}N or ^{13}C chemical shift axis. Though these experiments can sometimes be performed at natural abundance (Wagner and Bruhwiler, 1986), they are most useful with isotope enriched samples. A combination of these experiments can be used to determine the ^{1}H, ^{15}N, and ^{13}C chemical shift assignments and subsequently, the mid- and long-range NOE assignments which carry structural information (Kay *et al.*, 1989; Fesik and Zuiderweg, 1990). The two dimensional ^{15}N–^{1}H HMQC-NOESY carries essentially the same information as the amide region of a homonuclear NOESY experiment, with the amide proton chemical shift replaced by that of the ^{15}N nucleus (Gronenborn *et al.*, 1989; Forman-Kay *et al.*, 1990). The amide nitrogen *and* proton chemical shifts involved in a particular NOE will only be degenerate in rare cases; in this manner, the experiment compliments a homonuclear NOESY experiment much like acquiring an additional ^{1}H–^{1}H NOESY spectrum at an alternate temperature. Two dimensional HMQC-COSY and HMQC-J spectra are useful for assignment and coupling constant determinations, respectively (Forman-Kay *et al.*, 1990).

Three- and four-dimensional NMR spectra further remove the problems of chemical shift degeneracy by the addition of ^{15}N and ^{13}C frequency labeling of the proton NOE cross peaks (Fesik and Zuiderweg, 1988; Marion *et al.*, 1989b; Kay *et al.*, 1990a; Clore *et al.*, 1991a). Heteronuclear three-dimensional NMR was recently reviewed by Fesik and Zuiderweg (Fesik and Zuiderweg, 1990). The finest demonstration of the promise of

multi-dimensional spectroscopy is the recent paper by (Clore *et al.*, 1991b) which describes the structure determination of interleukin 1β in solution, "based on 3146 experimental restraints comprising 2780 distance and 366 torsion angle (ϕ, ψ, and χ_1) restraints". Four-dimensional NMR experiments (Kay *et al.*, 1990a; Clore *et al.*, 1991a) were necessary to assign unambiguously a large enough number of NOEs to permit a high-resolution structure determination, compared to the previous low-resolution structure (Clore *et al.*, 1990c).

1.5.4.2.2. Random Fractional Deuteration

The earliest applications of deuterium labeling to ^1H NMR spectroscopy of proteins were designed to reduce the overall number of proton resonances, permitting the observation of signals from a limited number of residues that were present at natural abundance (Crespi *et al.*, 1968; Markley *et al.*, 1968; Crespi *et al.*, 1970). These are frequently termed "selective protonation" experiments (LeMaster, 1989). Increased spectral resolution due to the decrease in the linewidth of amide protons in perdeuterated proteins dissolved in ^1H_2O was also noted in early studies (Crespi *et al.*, 1973; Crespi *et al.*, 1974). This effect is due to the decrease in the number of ^1H–^1H dipolar interactions, the primary relaxation mechanism for most protons in most proteins. These themes were further pursued in studies utilizing fully protonated ligands bound to fully deuterated proteins, such as Markley and Jardetzky's experiments with thymidine bound to staphylococcal nuclease (SNase) (Markley and Jardetzky, 1970), Crespi and Katz's work with ^1H-flavin mononucleotide (FMN) bound to a fully deuterated flavoprotein (Crespi *et al.*, 1972), and a more recent study of ^1H-mellitin bound to ^2H-chicken calmodulin (Seeholzer *et al.*, 1986).

The expectation of significantly narrowed line widths in samples with lower than natural abundance proton densities led Kalbitzer *et al.* (Kalbitzer *et al.*, 1985) and LeMaster and Richards (LeMaster, 1988; LeMaster and Richards, 1988) to prepare protein samples with 50-90% ^2H at all non-exchangeable positions, an approach LeMaster calls "random fractional deuteration" (RFD). Several reviews of the RFD approach have since appeared

(LeMaster, 1989; LeMaster, 1990a; LeMaster, 1990b). The Kalbitzer *et al.* paper is a very preliminary one-dimensional NMR study of the 43 kDa protein EF-Tu, but it does provide the theoretical foundation for the technique. In particular, the equation for the expected reduction in line width of the resonance of a residual proton, assuming all surrounding protons are substituted by deuterons is given: $(\gamma_1{}^2 I_1(I_1 + 1))/(\gamma_2{}^2 I_2(I_2 + 1))$. With the correct values for the gyromagnetic ratios and nuclear spins, a limiting reduction factor of approximately 16 is expected. In practice, the linewidths of the histidyl C-2H resonances were decreased from 11 Hz to 2 Hz, a factor of 5.5. The difference between the expected value of 16 in the limiting case and the observed value of 5.5 can be ascribed to nearby proton dipoles and to the presence of relaxation mechanisms other than dipole-dipole interactions.

The LeMaster and Richards experiments are far more complete and are the first to really demonstrate the utility of the random fractional deuteration technique. The sequential assignment procedure for the 108-residue protein *Escherichia coli* thioredoxin was significantly aided through the use of 75% uniformly deuterated samples along with samples deuterated at the αβ positions of selected residues (LeMaster and Richards, 1988). Due to the reduction in proton populations, cross peak volumes in two-dimensional NMR experiments with the 75% deuterated samples must be decreased by the square of the isotopic dilution factor relative to the non-deuterated sample. However, the sensitivity of the experiments is a function of the peak height and not the peak volume. Line narrowing—the result of the combination of fewer relaxation pathways due to the dilution of proton density and the suppression of so-called passive coupling to spins not directly involved in creation of the cross peak—largely compensated for the decrease in peak volume. The approximate decrease in peak height for cross peaks between αβ protons in the 75% deuterated sample was only 2-3 fold, while the sensitivity for cross peaks between carbon bound and amide protons (dissolved in 95% H_2O, with all of the amide deuterons are back-exchanged for protons) was approximately equal to that of the unlabeled sample. The decrease in proton

density in the 75% deuterated samples significantly reduces spin diffusion (Kalk and Berendsen, 1975), allowing longer mixing times to be used in the NOESY experiments, providing an additional gain in sensitivity. This proves to be particularly valuable in the amide-amide region of ^{1}H NOESY spectra, where the sensitivity of the deuterated sample exceeds that of the unlabeled sample (LeMaster and Richards, 1988). Torchia *et al.* have used this effect to identify helical regions in larger proteins such as staphylococcal nuclease (SNase) (Torchia *et al.*, 1988). Finally, 50% deuterated samples were shown to be suitable, perhaps even advantageous, for use in relay experiments such as HOHAHA (LeMaster, 1988). In these experiments, the relay peaks decrease in intensity in a predictable manner as the magnetization is transferred to more remote protons of the spin system and as the probability of the spin system being uninterrupted by deuterons decreases. This effect is helpful in discriminating between single and double relays which are sometimes ambiguous in HOHAHA spectra

The reduction of spin diffusion pathways may also improve the accuracy of quantitative distances extracted from the NOE data using the isolated spin pair approximation (ISPA) (Gronenborn and Clore, 1985). The use of longer mixing times may permit the observation of cross peaks with a higher signal-to-noise ratio, allowing more accurate measurement of buildup curves (Kumar *et al.*, 1981; Gronenborn and Clore, 1985). However, it will not be possible to apply relaxation matrix analysis (Borgias and James, 1988) (see below) to the NOESY spectra from deuterated proteins, because in the randomly labeled sample every protein molecule will have a different set of relaxation pathways. Should these methods become part of the standard procedure for structure refinement using NOE data (Yip and Case, 1989) (this point is controversial (Clore and Gronenborn, 1989b)), then the attractiveness of the RFD approach for distance determination will be diminished.

1.5.4.3. Alternate processing techniques: S/N criteria for inclusion

Cross peaks in NOESY spectra have inherently low signal-to-noise ratios which affects the accuracy of intensity measurements. Additional spectral artifacts—*e.g.*, t_1 noise, overlap of zero-quantum cross peaks with true NOESY peaks, and baseline distortions—might be removed with modifications to either the method of data acquisition (Otting *et al.*, 1990; Wider, 1990) or processing (Denk *et al.*, 1985; Otting *et al.*, 1986). Further developments in spectral processing alternatives to the FFT (Hoch, 1989; Hoffman *et al.*, 1989; Hoffman and Levy, 1989) may also provide more accurate estimates for the intensity of low signal-to-noise peaks. These techniques, and others (Denk *et al.*, 1986; Gochin *et al.*, 1990), might also be used for extracting accurate intensities from overlapped peaks.

1.5.5. Obtaining more information from existing NOEs

The NMR constraints that are used for structural model building are *"imprecise but completely correct"* (Wüthrich, 1986) (p. 189, Wüthrich's italics). Historically, NOE cross peaks have been placed in three categories, namely strong, medium and weak, representing the ranges 1.8–2.7, 1.8–3.3, and 1.8–5.0 Å respectively; the lower limit corresponds to the sum of the van der Waals radii of two protons (Clore and Gronenborn, 1989a). As I have shown, all of the various algorithms used to calculate three-dimensional structures from NOE data treat these constraints in a more or less random fashion within these ranges.

The observed intensity of the NOE for particular pairs of protons in a protein is a complicated function of the distance between the protons, the presence of neighboring magnetic dipoles, and the correlation time describing the reorientation of the vector between the relevant nuclei. The simplest method for extracting the internuclear distances is to use the isolated spin pair approximation (ISPA) and intensity data from the initial portion of an NOE buildup curve (Wagner, 1990). A single isotropic correlation time is assumed, and

all buildup rates are compared to that observed for two protons at a fixed reference distance, such as the distance between a pair of methylene protons or neighboring protons in an aromatic ring. With this procedure, the distances are systematically underestimated, even for short mixing times (Borgias and James, 1988). Conservative estimates of the relative distance ranges are therefore used to compensate for the systematic errors (Clore and Gronenborn, 1989a). More accurate estimates of the distances may be available from a Complete Relaxation Matrix Analysis (CORMA) approach. The use of tighter (and more accurate) distance ranges improves the precision of the structure determination in model calculations (Thomas *et al.*, 1991).

1.5.5.1. CORMA and NMR refinement approaches

The matrix equations describing the relaxation behavior of the protons of any molecule—taking into account only pairwise dipolar interactions and neglecting higher order interactions—are relatively simple (Macura and Ernst, 1980; Borgias and James, 1988). Using these equations, it is possible in principle to calculate proton-proton distances directly from the experimental intensities (Olejniczak *et al.*, 1986); however, this method requires accurate measurement of the intensities of diagonal peaks which are not available in larger systems, such as even small proteins. The alternative approach is to back calculate the NOE spectrum from a suitable structural model, compare the calculated with the actual spectrum, and then adjust the model in an iterative procedure until the best fit of the calculated to the observed spectrum is obtained (Boelens *et al.*, 1989; Borgias and James, 1990). Yip and Case have recently developed a program which calculates the derivatives of the NOE intensities with respect to the nuclear coordinates and incorporates these expressions into an additional force in a molecular dynamics force field to automatically refine structures based on the experimental data (Yip and Case, 1989) . All of these methods currently assume a single isotropic correlation time for the entire system, an approach which has been criticized by Clore and Gronenborn (Clore and Gronenborn,

1989a; Clore and Gronenborn, 1989b). The algorithms do allow treatment of individual correlation times to be fit as additional variables, but it remains to be seen if the NOESY spectra contain enough information to justify this approach.

Finally, it has long been recognized that NOESY spectra reflect a time-averaged picture of protein structure, and that any single static model may not even be compatible with all of the available NOE constraints (Kim and Prestegard, 1989). A restrained molecular dynamics-based approach for correctly modeling the effect of time averaged constraints was recently introduced (Torda *et al.*, 1990; Pearlman and Kollman, 1991). In the near future, these methods may be combined with the methods which refine structures based on the NOE intensities.

1.5.6. Hydrogen Bond Constraints

Even though protein secondary structures fall out of the sequential assignment process and often represent the best defined portions of a solution structure, the hydrogen bonds which are the hallmark of the classic secondary motifs (Pauling and Corey, 1951; Pauling *et al.*, 1951; Schulz and Schirmer, 1979) are not well defined (Wagner, 1990). The presence of a hydrogen bond can usually be detected by slow exchange of the amide proton in D_2O solution (Wüthrich, 1986), but the assignment of the hydrogen bond acceptor, often a carbonyl oxygen atom, can only be determined by a network of NOEs to other protons in the secondary structure (Wagner *et al.*, 1987). Hydrogen bonds can also be assigned after a set of initial structures is available, leading to an iterative approach to structure refinement. Correct identification of a hydrogen bond constitutes a valuable and quite tight structural constraint, but only a limited number are expected to be available for any particular protein. For example, in the recent high-resolution structure determination of interleukin 1β, an all β-pleated sheet protein containing 153 residues, 57 hydrogen bonds were identified, 55 of which involved NH···CO backbone interactions. These represent

quite restrictive constraints on the three dimensional structure of the protein; hydrogen bonds in α-helices provide almost no constraints on the tertiary structure of proteins.

1.5.7. Long Range Distance Information

NOE, dihedral angle, and hydrogen bond constraints are all limited to short range interactions, on the order of 5 Å or less. Distance geometry simulations suggest that the inclusion of long range distance information, such as that available from spin labeled proteins, significantly improves the precision of the structure determination (I. D. Kuntz and J. Thomason, personal communication). The theory which relates NMR relaxation rates (and therefore linewidths) and the distance of the observed nucleus from a paramagnetic center was developed by Soloman and Bloembergen (Solomon and Bloembergen, 1956) and its application to proteins was most recently reviewed by Kosen (Kosen, 1989). Because these equations are dependent on the square of the gyromagnetic ratio of the observed nucleus, the proton is by far the most sensitive nucleus in proteins for spin label experiments.

Experimental attempts to extract this kind of information have been rare; a list of examples is given in a review by Kosen (1989) (Kosen, 1989)). More often, specifically tailored spin probes have been used to characterize protein binding pockets, such as in the elegant work of Frey, Anglister, and McConnell on anti-dinitrophenyl-spin-label antibodies that has been recently reviewed (Anglister, 1990). Spin labels are covalently attached to proteins; spin probes are not, and their use requires modeling of binding kinetics and thermodynamics. I shall only be concerned with spin labels.

1.5.7.1. Spin Labeling in the Kuntz Lab

Based on earlier work of McConnell and coworkers (Wien *et al.*, 1972), Schmidt and Kuntz used 300 and 500 MHz ^{1}H NMR spectra of spin labeled hen egg lysozyme in D_2O solution to extract several long range distance constraints (Schmidt and Kuntz, 1984). A major problem with these studies was the paucity of resonance assignments and the lack

of resolution in the one-dimensional spectra of lysozyme. For these reasons, a spin label study of derivatives of BPTI was undertaken (Kosen *et al.*, 1986). The primary advantages of BPTI for these studies were the availability of complete resonance assignments from the pioneering work of Wüthrich and colleagues (Wagner and Wüthrich, 1982b) and the increase in spectral resolution afforded by two-dimensional NMR spectroscopy (Ernst *et al.*, 1987). Following the laborious development of purification procedures to obtain pure preparations of mono-labeled derivatives, a qualitative analysis of the results from two-dimensional absolute value homonuclear COSY spectra (Aue *et al.*, 1976; Kumar *et al.*, 1980b) showed overall agreement between the observed broadening of assigned resonances in the fingerprint region of the protein and the x-ray crystal structure (Deisenhofer and Steigmann, 1975). Quantitative analysis of the two-dimensional data sets has been hampered by the inability to deconvolute the contribution to the decrease in COSY cross peak intensity from each of the active coupling partners, *e.g.*, the amide and alpha protons which yield cross peaks in the "fingerprint" region (Wagner and Wüthrich, 1982b) of protein spectra (Kuntz, personal communication).

1.5.7.2. Important issues pertaining to spin labeling

1.5.7.2.1. Problems with the single structure assumption

The extraction of distance information from spin label data require the assumption that a single structure exists for the protein in solution. This assumption can break down under two conditions. The first problematic situation arises if the spin label can reside in two distinct coordinate positions relavtive to the protein. This may result when the two positions represent approximately the same conformational energy. It may also result from the more trivial case in which the spin label used to modify the protein actually consists of two stereoisomers. The second condition is related to the first and arises from conformational exchange between multiple species which will seriously complicate the analysis of the paramagnetic relaxation behavior.

1.5.7.2.2. Concentration limitations

The low sensitivity of NMR requires the use of proteins at relatively high concentrations; generally a concentration of at least 2 mM is needed for adequate signal-to-noise for two dimensional homonuclear ^1H NMR experiments in a reasonable period of time at a field strength of 500 MHz. At this concentration, each protein molecule, on average, occupies a box that is 94 Å on a side:

$$\left(\frac{1\ l}{2 \times 10^{-3}\ mol} \times \frac{10^{27}\ Å^3}{1\ l} \times \frac{1\ mol}{6.022 \times 10^{23}\ molecules}\right)^{\frac{1}{3}}$$

This is equivalent to a sphere of radius 58.3 Å. The maximum dimensions of BPTI are about 19 × 29 Å (Creighton, 1984b) (p. 223). Even though BPTI has only a very weak tendency to self associate at these concentrations and pH values used in the NMR experiments (pH 2.0 or 4.6) (Gallagher and Woodward, 1989), care must be taken to account for possible intermolecular interactions that might be present in dynamic exchange processes. This can be done by taking measurements at several different protein concentrations, but problems with signal-to-noise place a rather severe practical lower limit to the protein concentration.

1.5.7.2.3. Choice of T_1 or T_2

The modified Solomon-Bloembergen equations describe the effect of the paramagnetic spin on both T_{1p} and T_{2p}. The relevant equations are (Kosen, 1989):

$$\frac{1}{T_{1P}} = \frac{2.46 \times 10^{-32}\ cm^6\ sec^{-2}}{r^6}\left(\frac{3\tau_c}{1 + \omega_I^2\tau_c^2}\right) \qquad (1.1)$$

$$\frac{1}{T_{2P}} = \frac{1.23 \times 10^{-32}\ cm^6\ sec^{-2}}{r^6}\left(4\tau_c + \frac{3\tau_c}{1 + \omega_I^2\tau_c^2}\right) \qquad (1.2)$$

Assuming a correlation time of approximately 2 nsec and a spectrometer frequency of 500 MHz ($\omega_I = 2\pi \times 500 \times 10^6$ rad/sec), $1/T_{2P}$ is 27.5 fold greater than $1/T_{1P}$ at a given

distance. However, the effect of the paramagnetic contribution to proton relaxation must be compared to the relevant relaxation rate in the absence of the spin label (James, 1975):

$$\left(\frac{1}{T_x}\right)_{obs} = \left(\frac{1}{T_x}\right)_o + \frac{1}{T_{xP}} \qquad (1.3)$$

Assuming an average linewidth of for BPTI of 10 Hz, the value of $1/T_2$ is approximately 32 sec^{-1}; with an estimated T_1 value of 0.1 sec, $1/T_1$ equals 10 sec^{-1}. Therefore, measurements of $1/T_{2P}$ are likely to be 10 fold more sensitive than measurements of $1/T_{1P}$.

Until recently, it has not been possible to determine directly proton T_2's in coupled systems using a Carr-Purcell-Meiboom-Gill type experiment (Meiboom and Gill, 1958). However, inspired by the work of Wagner and coworkers on the measurement of heteronuclear relaxation parameters (Nirmala and Wagner, 1988), Gochin *et al.* have developed an experiment to measure proton T_2's by effectively decoupling other protons in the same spin system by their different interaction with an attached heteronucleus (Gochin, 1991). This experiment was available for neither the previous lysozyme (Schmidt and Kuntz, 1984) and BPTI (Kosen *et al.*, 1986) experiments nor the work of this thesis. In each of these latter cases, the value of T_2 was estimated by an indirect and approximate measurement of the linewidth, using the relation

$$T_2 = \frac{1}{\pi W_{1/2}} \qquad (1.4)$$

1.5.7.2.4. Contribution to linewidth: 1D

Due to a lack of resolution in one-dimensional NMR spectra, Schmidt and Kuntz used difference spectroscopy (Campbell *et al.*, 1975) and a one-dimensional convolution difference method (Campbell *et al.*, 1973) to estimate the contribution of an attached nitroxide spin label to the linewidths of hen egg lysozyme in D_2O solution. Unspecified methods were used to account for spin multiplets. They were able to detect a change in peak amplitude of down to 5%, corresponding to a distance of approximately 25 Å. However, they

were only able to measure the effect for 21 assigned resonances; the remaining peaks in the difference spectra were either severely overlapped or unassigned.

1.5.7.2.5. Use of 2D experiments

Two-dimensional experiments provide vastly improved spectral resolution compared to their one-dimensional counterparts. The advantages and disadvantages of the use of a number of homonuclear experiments were described by (Kosen, 1989); for clarity, Table I of her review is reproduced here.

Table 1.1‡

Comparison of advantages and disadvantages of two-dimensional NMR spectra for purposes of distance measurements

Two-dimensional experiment	Advantages	Disadvantages
Absolute-value COSY	High signal-to-noise ratio Spectral simulation is simple Integrated intensity is strongly dependent on linewidth when the change in linewidth is approximately equal to the coupling constant	W_{P1} and W_{P2} cannot be measured separately
Double-quantum filtered COSY	High signal-to-noise ratio Spectral simulation is simple Moderate dependence of peak height on linewidth W_{P1} and W_{P2} can be measured independently	Integrated intensity is zero
Phase-sensitive NOESY	Simulation is not necessary W_{P1} and W_{P2} can be measured independently Many cross-peaks Direct measurement of T_{1P} is possible	Low signal-to-noise ratio $T_{1P} < T_{2P}$ Uncertainty in distance measurement increased due to greater separation of nuclei, if W_{P1} and W_{P2} are not evaluated separately
Phase-sensitive HOHAHA	High signal-to-noise Moderate dependence of peak height on linewidth W_{P1} and W_{P2} can be measured independently Many cross-peaks	Spectral simulation is not possible

‡Reproduced from Kosen, P.A. (1989). "Spin labeling of proteins." *Methods Enzymol.* **177**, 86-121.

Not noted in her table is a disadvantage common to all of the homonuclear two-dimensional experiments; the wide variation in the signal-to-noise ratio of the cross peaks in the absence of the spin label—due to the spread of coupling constants or internuclear distances which give rise to two dimensional peaks—creates a wide range in the precision and accuracy of the measurement of the effect of the spin label. This is in contrast to two-dimensional heteronuclear experiments, where conformation-independent and therefore more uniform ^1J coupling constants, give rise to direct cross peaks with a much narrower range of the signal-to-noise ratio. Kosen concluded that the best two-dimensional homonuclear experiment to use is the absolute value COSY.

In a study of *Megasphaera elsdenii* flavodoxin, Moonen *et al.* used two-dimensional NOESY difference spectra to obtain a qualitative analysis of those residues close to the riboflavin, a cofactor which is diamagnetic in the oxidized state and paramagnetic in the reduced, semiquinone state (Moonen *et al.*, 1984). Several attempts were made to apply the 2D difference spectroscopy technique to spin labeled BPTI, but all were hampered by a slight shift in the chemical shifts of many cross peaks, most likely due to slight changes in pH upon addition of the ascorbate reducing reagent (I. D. Kuntz, personal communication). Two-dimensional difference spectra taken on samples which must be removed from the probe are difficult due to slight but significant changes in sample volumes and shim settings, magnetic field and electronic instabilities, and baseline artifacts (Neuhaus and Williamson, 1989). Despite these difficulties, successful two-dimensional difference spectroscopy experiments have been carried out by Anglister and McConnell (Anglister, 1990; Zilber *et al.*, 1990) and by the group at Abbott (Fesik and Zuiderweg, 1989). These experiments are most commonly performed on selectively deuterated samples in D_2O solutions.

1.5.7.2.5.1. Heteronuclear Experiments for measurement of relaxation parameters

Two-dimensional heteronuclear experiments have recently been devised for the measurement of the relaxation properties of the heteronucleus (Nirmala and Wagner, 1988). Since the relaxation of the heteronucleus is dominated by the attached proton at a known fixed distance, these experiments are most useful for the characterization of atomic motions (Wagner and Nirmala, 1989; Clore *et al.*, 1990b). Unfortunately, these experiments cannot provide a sensitive measure of the effects of paramagnetic relaxation. The effect of a spin label on the nuclear T_2 is proportional to the square of the gyromagnetic ratio (Solomon and Bloembergen, 1956). Hence $1/T_{2P}$'s for protons are 16 times greater than for ^{13}C and 97 times greater than for ^{15}N. Obviously, a heteronuclear experiment for measuring the proton T_2 would provide a much more sensitive measure of the distance of a nucleus from the spin label, and such an experiment has recently been developed by Gochin in our lab (Gochin, 1991). A further advantage of the experiment is that there is no need to separate the contributions to the change in peak intensity from each nucleus giving rise to the cross peak, since the paramagnetic interaction with the proton is dominant. Gochin's experiment is just now being applied in our lab to measure proton T_2's of spin labeled BPTI in a ^{1}H-^{13}C experiment. Extension of the technique to a ^{1}H-^{15}N experiment is trivial. The ^{15}N experiment will require biosynthetic labeling to achieve adequate signal-to-noise, but the ^{13}C experiment has been demonstrated on natural abundance samples. Enrichment of the protein with ^{13}C would provide a valuable boost in signal-to-noise, but uniform labeling must be kept at or below 10% to avoid unwanted spin exchange mediated by one bond ^{13}C-^{13}C couplings. This level of enrichment corresponds to a 10 fold increase in signal-to-noise.

1.5.7.3. Combining isotope labeling with spin labeling

When I began the work described in this thesis, three-dimensional experiments and stereospecific assignments were still 2-3 years away. The spin labeling approach had yielded promising preliminary results (Schmidt and Kuntz, 1984; Kosen *et al.*, 1986) but had not been applied to an actual structural determination. What seemed to be missing was both more data and more accurate data. I concluded that the most accurate data from spin label experiments was that obtained from singlet peaks in one-dimensional difference spectra (Schmidt and Kuntz, 1984), but that these techniques were limited by excessive spectral overlap. Therefore, for my thesis work I decided to pursue a combination of the random fractional deuteration technique of LeMaster with the spin labeling approach being pursued in our lab. Whether the resolution enhancement and multiplet reduction available from the RFD approach would prove sufficient for an accurate accounting of one dimensional difference spectra was not known *a priori*; nevertheless, the deuterated sample would also prove valuable for the detailed characterization of the reduction of spin diffusion and for the possibly more accurate measurement of distances derived from proton NOEs.

1.6. Choice of BPTI as the Model System

The requirements of a system for the application of the combined spin label/RFD approach to protein structure are as follows: (1) suitability for NMR analysis, preferably with previously assigned resonances, (2) a cloned gene with an efficient expression system that allows biosynthetic labeling under stringent conditions, preferably in *E. coli*, and (3) well characterized protein modification chemistry. As has so often been the case in the past, bovine pancreatic trypsin inhibitor (BPTI) appeared to serve as an almost ideal system for these studies. The complete ^{1}H NMR spectrum of BPTI and several derivatives were assigned by Wüthrich's group at the ETH (Wagner and Wüthrich, 1982b; Stassinopoulou *et al.*, 1984; Wagner *et al.*, 1987; Wagner *et al.*, 1987) with a few additions by Woodward

and coworkers (Tüchsen and Woodward, 1987). A clone and expression system were available from Dr. Stephen Anderson of Genentech, already a close collaborator with the Kuntz group. Anderson's system was capable of producing a large number of mutants of BPTI, a number of which might be suitable for the RFD/spin label work. Finally, Phyllis Kosen in the Kuntz lab had developed a purification scheme for 3 spin labeled derivatives of BPTI obtained from a non-selective reaction with an amino group specific reagent (Kosen *et al.*, 1986). The yields from this procedure are only on the order of 5-10%, but only a few milligrams of protein is needed for one dimensional experiments. Experiments designed to spin label histidine mutants of BPTI were in progress (Kosen, personal communication).

It seemed as if the technology required for a spin label study of RFD-BPTI was in place. The next chapter reveals that the experiments I proposed placed greater demands on existing technologies—particularly the *E. coli* expression system—than they were capable of handling. The bulk of this thesis describes developments in a new *E. coli* expression system for BPTI that allow coupling of isotope labeling with spin labeling. The new expression system also permits the production of a number of classes of BPTI mutants that were desired for protein folding experiments but were not available with the existing expression system. These developments are described in Chapter 2, which is an expansion of a paper recently published in **Protein Engineering** (Altman *et al.*, 1991). A new class of mutants was constructed which allow high yield modification reactions with spin label reagents, and these are described in Chapter 3. The actual spin labeling experiments on deuterated BPTI were never carried out, so this thesis represents a progress report rather than a piece of work which validates (or invalidates) the spin label approach for increasing the precision of solution structures. However, Chapter 4 presents the current results from proton NMR experiments on one of the mutants described in Chapter 3. The efforts of my collaborators in Anthony Kossiakoff's lab at Genentech and Stephen Anderson's lab at the

Center for Advanced Biotechnology and Medicine at Rutgers, which utilize mutants I describe in Chapter 2, are briefly reviewed in Chapter 5.

The work that I describe in this thesis, together with the dramatic advances in the field of protein NMR in the heteronuclear era, highlight the increasing reliance of protein NMR research on the fruits of biotechnology. Even starting with a successful heterologous expression system, the stringent demands of isotope labeling in *E. coli* can make it a non-trivial procedure. Accordingly, the development of new methodologies for the expression of BPTI required most of my attention in the course of this thesis work. Nevertheless, I am confident that the methods I have developed open the doors to a wide variety of NMR and protein folding experiments with BPTI that were previously barred by a limited expression system. In light of the prevalent use of BPTI in model studies, the ability to produce an unlimited spectrum of mutants of BPTI represents a significant advance.

Chapter 2

Intracellular Expression of BPTI

2. Introduction

The bulk of this chapter constitutes a manuscript describing the development of a new intracellular expression system for derivatives of BPTI that was recently published in the journal **Protein Engineering** (Altman *et al.*, 1991). The version of the manuscript included here is identical to that which appeared in the journal. Before I present the manuscript, however, I include a number of prefatory remarks that further describe the motivation behind the work. Several appendices are included after the manuscript and are intended as a guide to researchers who wish to pursue studies using the tools I have built. Some of these appendices provide details of the expression system that are not appropriate for a widely circulated journal; others are merely my suggestions for designs which improve upon the tools I have built.

2.1. BPTI expression: a historical perspective

At the conclusion of the previous chapter, I presented arguments in favor of the choice of BPTI as a model system for the application of a combined random fractional deuteration with spin labeling approach for the extraction of long range distances in proteins. The crucial prerequisite for the success of the experiments was an ability to express BPTI—or any of several mutants—in an extensively deuterated media. A secretion expression system for BPTI was developed in the laboratory of Stephen Anderson; *in vitro* refolding and peptide bond cleavage procedures were not required (Marks *et al.*, 1986; Marks *et al.*, 1987a; Marks *et al.*, 1987b). However, not all mutants could be expressed in their system, and the yields of protein were highly variable (Nilsson *et al.*, 1991).

It is instructive to consider the choices and decisions that were made by Anderson during the development of his expression system. The most important decision Anderson made was to express BPTI with a secretion vector rather than an intracellular one. Many of the general arguments in favor of secretion are presented in a recent review (Stader and

Silhavy, 1990). These include simplification of purification, correct processing of "pre" sequences by signal peptidases, correct formation of disulfide bonds, and localization of the heterologous protein to a compartment with relatively few harmful proteases. More specifically, Anderson was aware of "rumors" that others had attempted intracellular expression of BPTI and failed (S. Anderson, personal communication); these rumors seemed to be substantiated by the subsequent appearance of a paper describing the successful intracellular expression of BPTI, but only after 14 copies of the gene were cloned into a single plasmid (von Wilcken-Bergmann *et al.*, 1986). In addition, successful secretion systems were in wide use at Genentech (Gray *et al.*, 1985).

The choices of a promoter and signal sequence were relatively minor compared to the decision to use a secretion vector. To a large degree, Anderson's decision to use the *phoA* promoter and the ST:II signal sequence (Marks *et al.*, 1987b) were based on the prevailing systems at Genentech. The *phoA* promoter, although undependable in shake flask experiments, was optimal for industrial scale fermentations, and the ST:II signal sequence gave better yields than the alkaline phosphatase signal sequence for several secreted proteins (Chang *et al.*, 1987). I must reemphasize that both choices were based on an optimization of protein production in rich media in ten liter industrial fermentors, where extremely high cell densities are available.

The Anderson-Kuntz long term project to dissect the oxidative refolding pathway of BPTI by the removal of individual native disulfide bonds was hindered by the inability to express mutants lacking the 5-55 disulfide bond, *e.g.* C5A/C55A, in the *phoA*/ ST:II expression system. Björn Nilsson attempted to express this mutant in a protein A secretion expression vector that he brought to the lab from his previous lab in Sweden (Nilsson *et al.*, 1987; Nilsson and Abrahmsen, 1990; Nilsson *et al.*, 1991). Although the protein A vector provided 2-4 times better yields than the *phoA*/ST:II vector for several mutants (Nilsson *et al.*, 1991), the yields of C5A/C55A were only fractions of a milligram per liter (B. Nilsson, personal communication). In all secretion expression systems, the yields of

the C14A/C38A mutant were 3-10 times greater than wild type BPTI. The protein A expression system did allow Nilsson to characterize the competition between intracellular folding, leading to blockage of secretion, and "facile secretion of unfolded material" (Nilsson *et al.*, 1991).

2.2. Attempts to adapt the BPTI secretion systems to the production of deuterated protein

I spent considerable effort attempting to express 80% randomly deuterated C14A/C38A BPTI from any of the available secretion expression vectors using minimal media conditions—based on succinate and alanine as the sole carbon sources—similar to those developed by LeMaster (LeMaster and Richards, 1988). A variation of the MOPS minimal media (Neidhardt *et al.*, 1974) was used for expression with the *phoA* promoter, while the standard M9 or M63 media (Miller, 1972) were used for work with the protein A vectors. In preliminary experiments with an equivalent "all-protio" media, the protein A vectors gave higher yields than the *phoA* vectors. However, subsequent experiments in deuterated media utilizing the protein A vector were beset by one significant problem: titering of dilutions of the cultures on agar plates with and without ampicilin indicated that only 0.1% of the cells still harbored the pEZ:475 expression plasmid. The loss of antibiotic selection due to destruction of ampicillin by β-lactamase is a common problem (Studier *et al.*, 1990); the problem is in this case exacerbated by the induction of a general leaky phenotype associated with the expression of secreted protein A fusion proteins (Abrahmsén *et al.*, 1986; L. Abrahmsén, personal communication).

My final effort to express a secreted BPTI in a deuterated media was to move the gene for the protein A-BPTI fusion protein into a plasmid vector carrying a gene for kanamycin resistance instead of the ampicillin resistance gene of the pEZ vectors. Kanamycin inhibits protein synthesis by inhibiting the translocation of the ribosome along

the mRNA; the product of the kanamycin resistance gene I used was an aminoglycoside 3'-phosphotransferase which inactivates the kanamycin (Oka *et al.*, 1981; Ausubel *et al.*, 1987). My construction was based on the plasmid pK3ZZIGF-1, a gift from Björn Nilsson. With the kanamycin vector, I was able to prepare enough 80% deuterated Gly-BPTI(C14A/C38A) for a single 1.5 mM 0.4 ml NMR sample, but the signal to noise of the two dimensional spectra from this sample was not sufficient to observe even one half of the fingerprint peaks in a double quantum filtered COSY of the protein dissolved in 90% H_2O/10% D_2O.

2.3. Intracellular expression of protein A fusion proteins

As I was finishing the purification of the deuterated protein expressed from the kanamycin vector, I was encouraged by Jim Wells to try expression of BPTI in a new vector designed by Dennis Henner for the intracellular expression of protein A fusion proteins. The vector was used by Henner for the expression of IGF-1, and was also used by Paul Carter and Bob Kelley, with whom I shared a lab at Genentech, for the expression of other small, disulfide containing proteins. Based on previously observed difficulties with intracellular expression of BPTI (von Wilcken-Bergmann *et al.*, 1986), I was skeptical regarding the chance of success of the approach. On the other hand, apparently no one had tried to express BPTI as an intracellular fusion protein. The manuscript that follows addresses these issues.

Protein Engineering vol.4 no.5 pp.593–600, 1991

Intracellular expression of BPTI fusion proteins and single column cleavage/affinity purification by chymotrypsin

John D.Altman, Dennis Henner[1], Björn Nilsson[2,3], Sephen Anderson[4] and Irwin D.Kuntz[5,6]

Department of Pharmaceutical Chemistry S-926, University of California, San Francisco, San Francisco, CA 94143, [1]Genentech, Inc. 460 Pt. San Bruno Blvd, South San Francisco, CA 94080, USA, [2]Royal Institute of Technology, Department of Biochemistry and Biotechnology, S-100 44 Stockholm, Sweden, [4]Center for Advanced Biotechnology and Medicine, 675 Hoes Lane, Piscataway, NJ 08854 and [5]Department of Pharmaceutical Chemistry S-926, University of California, San Francisco, San Francisco, CA 94143, USA

[1]Present address: Kabi AB, Department of Biochemistry, S-112 87 Stockholm, Sweden

[6]To whom correspondence should be addressed

The novel and efficient expression system described here produces formerly poorly expressed, proteolytically unstable mutants of bovine pancreatic trypsin inhibitor (BPTI). A new, single column method for the cleavage of a recombinant fusion to BPTI and affinity purification of the BPTI moiety by immobilized chymotrypsin is an integral part of the system. Wild-type and mutant BPTI molecules are expressed in *Escherichia coli* as fusion proteins forming intracellular inclusion bodies. Transcription initiation is under the control of the *E.coli trp* promoter. The expressed protein is tripartite fusion comprising (i) a portion of the TrpLE leader peptide, (ii) a synthetic IgG binding domain derived from protein A and (iii) the BPTI variant. Solubilization of the inclusion bodies and refolding of the fusion proteins in a thiol-disulfide shuffling system yields correctly folded inhibitor molecules. In the single column purification and cleavage procedure, immobilized chymotrypsin cleaves the refolded fusion protein and releases affinity purified active BPTI mutants with correct N-termini. Mutant BPTI molecules which do not fold into active inhibitors are also stably expressed in inclusion bodies but cannot be purified by this method. Unlike previously described secretion systems for the production of BPTI, expression levels in this system appear to be independent of both the mutation in the BPTI gene and the activity of the expressed protein. Mutants poorly expressed in secretion systems can now be produced in sufficient quantities for protein folding studies and structural analysis using X-ray crystallography and NMR spectroscopy.

Key words: BPTI/mutagenesis/protein folding/site-specific protein cleavage/TrpLE-protein A fusion proteins

Introduction

The introduction of a proposed mechanism for the oxidative refolding of reduced and denatured bovine pancreatic trypsin inhibitor (BPTI) (Creighton, 1977; Creighton and Goldenberg, 1984) has provided an impetus for the production of mutants to test predictions based upon the model. Two different themes emerge from recent studies. One line of investigation has focused on the effects of the removal of one of the three native disulfides in BPTI; double mutants at cysteines 14/38 and 30/51 have been described (Marks *et al.*, 1986, 1987a,b; Goldenberg, 1988). Other studies have examined the role of individual core residues at each step of the folding pathway (Goldenberg *et al.*, 1989; Coplen *et al.*, 1990). The ability to produce mutants of BPTI has also permitted the experimental testing of theoretical predictions (Levitt, 1981) of the role of specific proline isomerizations in the refolding of guanadinium hydrochloride (GdnHCl) denatured protein (Hurle *et al.*, 1990, 1991). Until now, however, limitations in our expression systems (among several described below) inhibited our ability to study several mutants, most notably mutants lacking the 5-55 disulfide bond (Δ5-55).

Six different expression systems for the production of recombinant BPTI in *Escherichia coli* have been described (Table I). Four of these rely on secretion of the protein into the periplasmic space of *E.coli* and direct recovery of native protein which has folded *in vivo*. These systems yield either correctly processed proteins with their signal peptides removed (Marks *et al.*, 1986; Goldenberg, 1988) or mature fusion proteins (Nilsson *et al.*, 1990) which may be specifically cleaved *in vitro* to release a glycine-extended BPTI. Two methods for the intracellular production of BPTI in inclusion bodies have been described. The first, labor-intensive method relies on stepwise multiple duplications of the BPTI gene until 14 copies of the gene are under the control of a single promoter (von Wilcken-Bergmann *et al.*, 1986); this system yields BPTI molecules with an additional *N*-formyl methionine at the N-terminus. A second method (Auerswald *et al.*, 1987) utilizes a fusion of the BPTI(M52E) mutant coding sequence to the 3′ end of a complete *lacZ* gene; the BPTI moiety of the expressed fusion protein was released by cyanogen bromide cleavage. All previously described methods other than direct secretion require modification of the BPTI amino acid sequence, either by N-terminal extensions or mutations to remove methodology-specific cleavage sites such as Met52.

Levels of BPTI expression in secretion systems vary by as much as two orders of magnitude (Nilsson *et al.*, 1990, 1991) as a function of the mutation in the BPTI gene. In a limited number of cases, variation in expression levels correlates with thermal stability of the mutant protein (Coplen *et al.*, 1990; Nilsson *et al.*, 1990) or inversely correlates with *in vitro* folding rates of reduced and denatured protein (Nilsson *et al.*, 1990). For some mutants, it has been practically impossible to obtain the quantities of protein needed to carry out folding studies or structural analyses by NMR or crystallography (Eigenbrot *et al.*, 1990).Our inability to express large amounts of several BPTI mutants [e.g. mutants missing the 5-55 disulfide (Δ5-55) and the N43L mutant] or to label the wild-type protein with ^{15}N or deuterium in secretion-based expression systems led us to reconsider intracellular expression.

Here, we report a new, simple and efficient intracellular expression system for the production of BPTI variants as fusion protein inclusion bodies that can be refolded without prior purification or cleavage of the fusion protein. Furthermore, we

Table I. BPTI expression systems[a]

Mode	Plasmid	Promoter	Signal	Fusion	Reference
Secretion	p393	*phoA*	PhoA	–	Marks *et al.* (1986)
	p657	*phoA*	STII	–	Marks *et al.* (1987a,b)
	pTI103	*lpp/lac*	OmpA	–	Goldenberg (1988)
	pEZZ:BPTI	*spa*	SpA	ZZ	Nilsson *et al.* (1990)
Intracellular	pRK48.1.1	*lac*	–	β-gal	Auerswald *et al.* (1987)
	piWiT10wL1	*T5 p25*	–	–	von Wilcken-Bergmann *et al.* (1986)
	pHAZY:BPTI	*trp*	–	TrpLE-Z	This report

[a]Six previously described BPTI expression systems, plus the vector described in this report.
Abbreviations are as follows: PhoA, the *E.coli* alkaline phosphatase promoter or signal sequence; STII, heat stable enterotoxin II signal sequence; *lpp/lac*, a hybrid *E.coli* lipoprotein:lac protomer; OmpA, *E.coli* outer membrane protein A; *T5 p25*, a synthetic promoter corresponding to the bacteriophage T5 P25 promoter; *trp*, *E.coli trp* promoter; Z, single synthetic domain of staphylococcal protein A; *spa*, staphylococcal protein A; TrpLE, *trp* operon leader peptide; β-gal, β-galactosidase.

present the first method for the cleavage of BPTI fusion proteins that does not require modification of the BPTI amino acid sequence (Auerswald *et al.*, 1987) or leave N-terminal extensions (Nilsson *et al.*, 1990) that may interfere with a salt bridge between the N- and C-termini of the wild-type protein (Brown *et al.*, 1978; Goldberg and Creighton, 1984; Vincent *et al.*, 1971). This procedure, which is generally useful for all mutants which retain a high affinity for chymotrypsin, also provides a powerful affinity purification for proteins which are properly folded. It represents the first description of a single column fusion protein cleavage/affinity purification procedure and may prove useful for the production of other protein inhibitors of proteases.

Materials and methods

Construction of pDH108

Plasmid pDH108, a general vector for the intracellular expression of staphylococcal protein A (SpA) fusion proteins, was constructed by standard techniques. The *E.coli* promoter, derived from pHGH207-2 (de Boer *et al.*, 1982) is used for transcription of the fusion protein. The first seven amino acids of the translated protein are encoded by synthetic DNA identical to the *trpLE* gene (Miozzari and Yanofsky, 1978). A ZZ moiety consists of two synthetic SpA domains, as described (Nilsson *et al.*, 1987). A polylinker site follows the ZZ moiety to allow simple construction of in-frame fusion proteins. Following the expression cassette is a synthetic transcriptional terminator based on the sequence of the lambda t_0 terminator (Scholtissek and Grosse, 1987). The vector is pBR322 (Sutcliffe, 1978) from the *Hin*dIII site to the *Eco*RI site. Details of the DNA sequence will be provided upon request. Schematically, the sequence contains the following fragments.

Fragment 1. The *E.coli trp* promoter from an *Eco*RI site to an *Xba*I site, as described for HGH207-2 (de Boer *et al.*, 1982). The *Eco*RI site was filled in with the Klenow fragment of DNA *Pol*I and four nucleotide triphosphates (dNTPs). Subsequent ligation of this blunt end destroys the site.

Fragment 2. Synthetic DNA encoding the spacer between the ribosome binding site, the initiation codon and several codons of the *trpLE* gene (Miozzari and Yanofsky, 1978). The sequence of the synthetic oligonucleotide is:

MetLysAlaIlePheValLeuAsn
CTAGAATTATGAAAGCAATTTTCGTACTGAATGC

Fragment 3. *Fsp*I–*Bam*HI fragment of pEZZ318, formerly termed pEZZ18 (Löwenadler *et al.*, 1987), encoding the mature portion of the synthetic protein A (termed ZZ), through part of the polylinker.

Fragment 4. Synthetic DNA encoding the lambda t_0 terminator (Scholtissek and Grosse, 1987), and the pBR322 tet promoter region, up to the pBR322 *Hin*dIII site. The sequence of the synthetic region is:

5'GATCTAGGCCTAACGCTCGGTTGCCGCCGGGCGTT-
TTTTATTGTTAACTCATGTTTGACAGCTTATCATCGA-
TAAGCTT3'

Fragment 5. pBR322 from the *Hin*dIII site to the *Eco*RI site. The *Eco*RI site was filled in with the Klenow fragment of DNA *Pol*I and dNTPs; ligation destroys the site.

Construction of pDH:BPTI mutants

The construction of the phagemids encoding secreted Z:BPTI fusion proteins in the vector pEZZ318 has been described (Nilsson *et al.*, 1990). The corresponding plasmids for the expression of intracellular Z:BPTI fusion proteins were constructed from 336 bp fragments from the *Bgl*II site to the *Eco*RI site of pEZ:BPTI and the 4786 bp vector fragment of pDH108, cut with the restriction enzymes *Bgl*II and *Stu*I. The *Eco*RI site of pEZ:BPTI was filled in with large (Klenow) fragment *E.coli* DNA polymerase I and dNTPs. Ligation of the filled-in *Eco*RI site to the blunt end of the *Stu*I–*Bgl*II fragment of pDH108 regenerates the *Eco*RI site.

A chymotrypsin-sensitive site was introduced between the synthetic protein A domain and the BPTI domain in the phagemid pEZ:475, coding for the C14A/C38A double mutant of BPTI, using oligodirected site-specific mutagenesis by the α-phosphorothioate method (Nakamaye and Eckstein, 1986) as supplied by Amersham. Single-strand DNA from pEZ:475 was prepared with the helper phage M13K07 (Vieira and Messing, 1987). The oligonucleotide:

*** *
GAAGTCAGGCCTATAAGTCATGTCTACT

corresponding to the non-coding strand of the mutant gene, was used to introduce the mutations Asn-Gly → Thr-Tyr. Mismatches are indicated with asterisks. The mutation was verified by dideoxy sequencing of single-stranded DNA (Sanger *et al.*, 1977) and moved into pDH108 as described in the previous paragraph, to create the plasmid pHAZY:475. Additional BPTI mutants, originally made in the STII-based vector (C.Marks and S.Anderson, unpublished data) or in the pEZ:BPTI vector (B.Nilsson and S.Anderson, unpublished data) using the *mutL* method (Carter, 1991) were cloned into the pHAZY vector as *Stu*I–*Eco*RI fragments.

Expression of secreted proteins from pEZ:BPTI and ST:II based vectors

Expression of proteins from the pEZ:BTI and ST:II based vectors was as described (Nilsson and Abrahmsén, 1990).

Expression of the fusion proteins

Escherichia coli strain MM294 (Meselson and Yuna, 1968) was transformed with plasmids carrying the relevant BPTI mutant according to the PEG/DMSO procedure (Chung and Miller, 1988). Transformants were selected on LB agar plates with 15 μg/ml tetracycline. Saturated overnight cultures of bacteria in 2YT media (Miller, 1972) were diluted 25-fold into M9 media supplemented with 0.5% glucose, 0.5% casamino acids, 1 mM $MgSO_4$, 0.01% thiamine hydrochloride (B1) and 15 μg tetracycline and grown at 37°C. Expression of recombinant fusion protein was induced at mid-log phase (OD_{580} = 1) by the addition of 25 μg/ml β-indole acrylic acid. Cells were harvested in late log phase by centrifugation at 5000 *g* for 5 min.

For small-scale experiments (<25 ml culture volume), cells were resuspended in 1/10 culture volume in a buffer containing 25 mM Tris−HCl, 5 mM EDTA, 1 mM phenyl methyl sulfonyl fluoride (PMSF) pH 8.0 (buffer A) and lysed by sonication. The TrpLE-Z-BPTI fusion protein was found exclusively in the pellet upon centrifugation for 10 min in a microcentrifuge. The pellet was washed with 1/10 culture volume 0.5% Triton X-100, 1 mM EDTA and finally resuspended by sonication in 1/10 culture volume buffer A supplemented with 7 M GdnHCl.

Samples were prepared for SDS−PAGE on 16% gels as follows: 10 μl of GdnHCl solubilized inclusion bodies, representing 100 μl of bacterial culture, was diluted with 100 μl water and protein was precipitated by the addition of 1 ml of acetone and incubation on dry ice for 30 min. Protein was pelleted by centrifugation for 10 min in a microcentrifuge and the supernatant was aspirated. Pellets further were dried *in vacuo*. Protein was redissolved in 50 μl SDS sample buffer, and 25 μl was applied to the SDS gels (Laemmli, 1970). Gels were stained with Coomassie Brilliant Blue R-250 (BioRad) and scanned on a LKB 2202 Ultroscan Laser Densitometer interfaced to a Hewlett Packard 3390A integrator.

Refolding of intact fusion proteins

The GdnHCl-solubilized fusion proteins were diluted 10- to 20-fold into a refolding buffer containing 10 mM reduced glutathione, 1 mM oxidized glutathione, 100 mM Tris−HCl, 200 mM KCl and 1 mM EDTA, pH 8.7 (final concentrations; buffer B). For most mutants, the refolding reactions were incubated at 37°C for 3 h. For potentially less stable mutants (e.g. N43L), refolding reactions were incubated at 4°C overnight. Refolding buffer was exchanged for 0.2 M triethanolamine, pH 7.8 by gel filtration on PD-10 columns (Pharmacia) and the refolded proteins were assayed for trypsin inhibition activity as previously described (Marks *et al.*, 1986).

Preparative scale isolation of refolded protein from inclusion bodies

The preparation of large quantities (>10 mg) of protein essentially followed published procedures (Nagai and Thøgersen, 1987) with minor modifications. Sonication was used to aid in the disruption of the cells and shearing of the DNA following the addition of the detergent buffer. Sonication was also used to resuspend the inclusion bodies in Triton X-100 wash buffer. Following two washes with 200 ml Triton X-100 buffer, the inclusion bodies were further washed with 200 ml 10 mM Tris−HCl, 1 mM EDTA, pH 8.0 to remove the residual detergent.

The washed inclusion bodies were dissolved in 200 ml buffer A supplemented with 7 M GdnHCl at 4°C. Following sonication, the suspension was stirred for 2 h at 4°C. The solubilized inclusion bodies were refolded by 10-fold dilution into BPTI refolding buffer B, and the pH was readjusted to 8.7. No attempt was made to exclude oxygen. Refolding reactions were stirred at room temperature overnight. The glutathione and residual GdnHCl were removed from the refolded protein by dialysis (SpectraPor3 MWCO 3500 membranes) against 20 vol of 50 mM Tris−HCl, pH 7.5 (three changes). The dialyzed protein solution was centrifuged for 10 min at 40 000 *g* and filtered through a 0.45 μm filter to remove precipitated material.

Cleavage and affinity purification

The filtrate was loaded onto a column of chymotrypsin immobilized on Affi-Gel 10 (BioRad) as previously described (Marks *et al.*, 1986). Linear flow rates did not exceed 0.4 cm/min. Following loading, the column was washed with 5 column volumes of a buffer containing 0.1 M triethanolamine hydrochloride, 0.3 M NaCl and 10 mM $CaCl_2$, pH 7.8 (buffer C). An additional 0.2 column volumes of fresh chymotrypsin resin and 0.2 column volumes wash buffer were then added to the column, which was sealed and incubated on a gentle rocker at 4°C for at least 12 h. The column was then washed with 5 column volumes of buffer C and BPTI was eluted with a solution of 250 mM KCl pH 1.7, as previously described. BPTI mutant proteins were further purified on a Mono S FPLC column (Pharmacia) as previously described (Hurle *et al.*, 1990). Purity was verified by reverse-phase HPLC chromatography on a C18 column (Vydac), SDS−PAGE (16% gels) (Laemmli, 1970) and native gel electrophoresis (Goldenberg, 1989). The site of cleavage by chymotrypsin was determined by automated Edman degradation on an Applied Biosystems 470A protein sequencer equipped with a 120A PTH analyzer. Sequence interpretation was performed on a VAX 8650 (Henzel *et al.*, 1987). Further confirmation of the cleavage site was demonstrated by a molecular ion peak in the Liquid Secondary Ion Mass Spectrum taken on a JEOL-HX110HF Tandem Mass Spectrometer.

Results

Expression

In Figure 1, the expression cassette from the pHAZY:BPTI plasmid described in this study is compared with the previously described secretion vector, pEZ:BPTI (Nilsson *et al.*, 1990). The pHAZY:BPTI plasmid was designed to complement the pEZ:BPTI vector; cloning of mutant BPTI genes between the two vectors is facilitated by the *Eco*RI site at the 3′-end of the BPTI gene and either the *Bgl*II site in the Z domain or the *Stu*I site which spans the dipeptide sequence Arg-Pro at the N-terminus of BPTI. The *spa* promoter and signal sequence of the pEZ:BPTI vector are replaced in the pHAZY:BPTI vector by the *trp* promoter and parts of the *trpLE* gene and the lambda t_0 transcription terminator is added. The details of the peptide cleavage site in the linker region between the Z and BPTI domains in the two vectors are also shown. The hydroxylamine sensitive Asn-Gly sequence of the pEZ:BPTI vector is replaced by the potentially chymotrypsin-sensitive sequence ThrTyr. These replacements immediately before the Arg-Pro sequence at the N-terminus of BPTI do not interfere with the *Stu*I site.

To demonstrate the robust nature of the TrpLE:Z:BPTI expression system, we examined the expression levels of six classes of BPTI mutants which affect or are expected to affect the nature of intermediates seen in the oxidative folding pathway (Creighton and Goldenberg, 1984). These mutants are also expected to have a profound effect on the stability of the native state of the protein. In Table II, we have classified the mutants according to either the replacement of a pair of cysteine residues

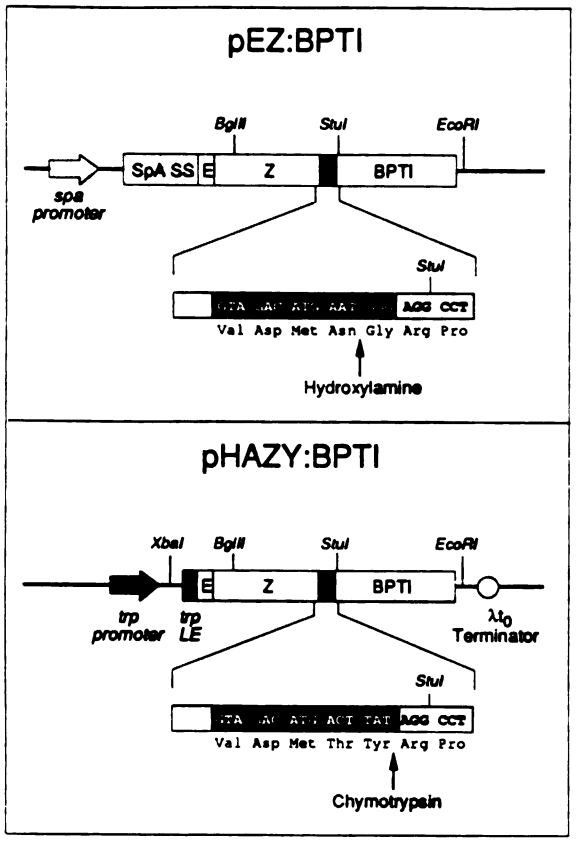


Fig. 1. Comparison of the pEZ:BPTI (Nilsson *et al.*, 1990) and pHAZY:BPTI expression cassettes. The constructions are identical from the beginning of the E region through the C-terminus of BPTI, except for the peptide cleavage site in the linker reigon, shown in black. pEZ:BPTI does not contain a transcription terminator. The restriction sites that are used in shuttling between the pEZ:BPTI (secretion) vector and the pHAZY:BPTI vector are highlighted. The linker regions between the Z domain and the N-terminus of pEZ:BPTI and pHAZY:BPTI contain sites which are specifically cleaved by hydroxylamine or chymotrypsin, respectively. These sites are shown with an arrow.

Table II. Six classes of BPTI mutants and relative yields in the pHAZY expression system[a]

Class	Mutant	Relative expression
Wild type	Wild type	1.00
Δ14-38	C14A/C38A	0.70
Δ30-51	C30V/C51A	1.48
Δ5-55	C5A/C55A	1.19
	C5V/C55A	1.24
X(43)	N43L	1.03
	N43S	0.68
Non-native intermediate	C14A/C55A	0.66

[a]Six classes of mutants and their relative yield when expressed from the pHAZY vectors. Densities are from scanning laser densitometry of the SDS gel shown in Figure 2 and are normalized to the band of wild-type BPTI. Fusion proteins were expressed in 25 ml shake flask experiments. Inclusion bodies were isolated, solubilized and applied to the gel as described in Materials and methods.

by amino acids with small aliphatic side chains (alanine or valine), or to replacements at the conserved Asn43 residue (Creighton and Charles, 1987). The side chain of Asn43 forms major tertiary interactions in the core of BPTI via hydrogen bonds between its side chain and main chain atoms of tyrosine 23 (two hydrogen bonds) and glutamic acid 7 (one hydrogen bond). The six classes of mutants are: (i) wild-type BPTI, (ii) mutants missing the 14/38 (Δ14-38) disulfide, (iii) mutants missing the 30/51 (Δ30-51) disulfide, (iv) mutants missing the 5/55 (Δ5-55) disulfide, (v) replacement of the conserved asparagine residue at position 43, and (vi) C14A/C55A, in which two cysteine residues which are not paired in the native structure are replaced by alanines. The mutant in class six is designed to mimic one of the non-native intermediates seen in the folding pathway of BPTI.

Phase contrast microscopy at ×1000 magnification clearly showed the presence of dense refractile bodies in >90% of the *E.coli* fron late log phase cultures of bacteria harboring the pHAZY plasmids (data not shown). Cells harboring refractile bodies were elongated and were frequently observed to have difficulty septating. Initial experiments demonstrated that all of the detectable heterologously expressed protein was found in an insoluble fraction following cell lysis by sonication (data not shown). Comparison of samples run under reducing and non-reducing conditions indicated that no detectable intermolecular disulfides were formed between proteins in inclusion bodies.

The gel shown in Figure 2 measures the expression of the mutant proteins present in inclusion bodies prior to refolding. Quantitation of the relative expression levels based on scanning laser densitometry of the gel in Figure 2 shows that differences in expression levels are <3-fold (Table II). These differences are probably not significant.

Refolding and activity

Upon 10-fold dilution into a refolding buffer containing oxidized and reduced glutathione, the GdnHCl-solubilized proteins

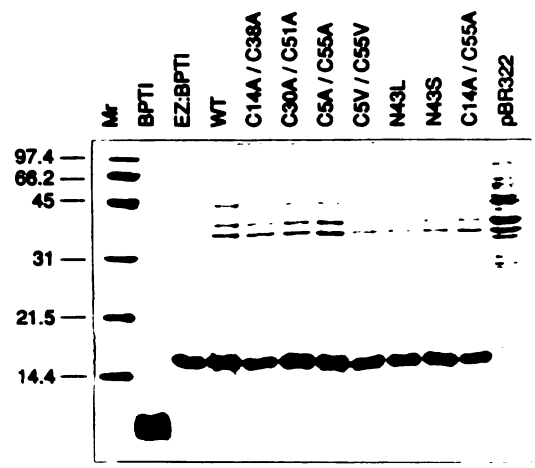


Fig. 2. 16% SDS–PAGE gel of the GdnHCl-solubilized fusion proteins prior to refolding. **Lane 1.** BioRad Lo molecular weight markers: lysozyme (14.4 kd), soybean trypsin inhibitor (21.5 kd), carbonic anhydrase (31.0 kd), ovalbumin (45.0 kd), bovine serum albumin (66.2 kd), and phosphorylase B (92.5 kd). **Lane 2.** wild-type BPTI standard (Mobay Corp.). **Lane 3.** purified, secreted EZ:BPTI (C14A/C38A) standard. **Lanes 4–11.** BPTI variants expressed in the pHAZY:BPTI vector. **Lane 12.** GdnHCl-solubilized pellet from cells not harboring a pHAZY:BPTI plasmid.

remained soluble and formed only intramolecular disulfide bonds. A small amount of precipitate was observed in the folding reactions, but very little of this material was detectable by Coomassie staining when it was resolubilized in boiling SDS and analyzed by SDS−PAGE (data not shown).

Mutants in five of the six classes described above are known inhibitors of trypsin (Marks *et al.*, 1987a,b; Goldenberg *et al.*, 1989; Nilsson *et al.*, 1990). The non-native two-dilsulfide intermediate we have attempted to mimic in the C14A/C55A mutant (class 6) is not expected to be an active inhibitor. Our results on the activity of the refolded proteins are completely consistent with these observations. Refolded proteins from inclusion bodies produced in shake flask experiments were assayed for their ability to inhibit trypsin in stoichiometric inhibition assays. The results are shown in Figure 3 where the activities are expressed as milligrams of inhibitor protein per liter of culture, relative to the molecular weight of wild-type BPTI. For comparison, expression levels for several of these mutants in two secretion systems are included (Nilsson *et al.*, 1990). The high yields of the Δ5-55 and Asn43 mutants are of particular importance to us as we have not previously been able to produce significant quantities of these proteins. The relative range of the activity numbers is slightly greater than the expression levels obtained by densitometry of the gel in Figure 2 and shows no correlation with those numbers. This discrepancy will be addressed below.

Chymotrypsin cleavage and affinity purification

Next, we explored the possibility of cleaving the refolded TrpLE:Z:BPTI fusion proteins and affinity purifying the released BPTI product in a single column. As an example, ~6.6 μmol BPTI activity from large-scale refolding reactions of the C5V/C55A fusion protein, corresponding to 100 mg of correctly processed BPTI, was bound to 50 ml chymotrypsin columns in two separate experiments. In the first experiment, the column was simply washed with buffer C followed by elution of the BPTI variant at low pH. The second experiment was designed to test for possible increased accessibility of the engineered cleavage site in a batch mode procedure, and was performed as described in Materials and methods. The results, shown in Figure 4, indicate that both procedures efficiently cleave the fusion proteins and release affinity-purified molecules with a molecular weight indistinguishable from authentic BPTI by SDS−PAGE. In this experiment and others (data not shown), the batch mode procedure is seen to be slightly more effective at processing the fusion proteins. The yield of trypsin inhibition activity eluted from the column in each case was 70%.

Proteins eluted from the chymotrypsin column were further purified on Mono S cation exchange FPLC columns (Hurle *et al.*, 1990). Integration of the Mono S chromatogram and subsequent identification of the major peak by N-terminal protein sequencing and mass spectrometry showed that ~90% of the protein eluted from the chymotrypsin column after the batch mode procedure was cleaved to yield protein with an N-terminal sequence identical to authentic wild-type BPTI. Proteins in all five classes of mutants which yielded active inhibitors (Table II and Figure 3) could be cleaved in this manner (data not shown). Cleavage within the BPTI protein was not observed even for mutants without the 5/55 disulfide bond which might have shown chymotryptic sensitivity at Phe4. No attempt was made to purify the C14A/C55A double mutant with this procedure.

Discussion

Nilsson *et al.* (1990) described the production of secreted mutants of BPTI fused to two synthetic domains of protein A (the Z domain). They observed proteolytic degradation of a fraction of the secreted protein in the C-terminal BPTI domain; for a limited number of mutants, the yield of full-length protein was inversely correlated with the *in vitro* oxidative folding rate of the isolated BPTI domain. To account for these observations, they proposed

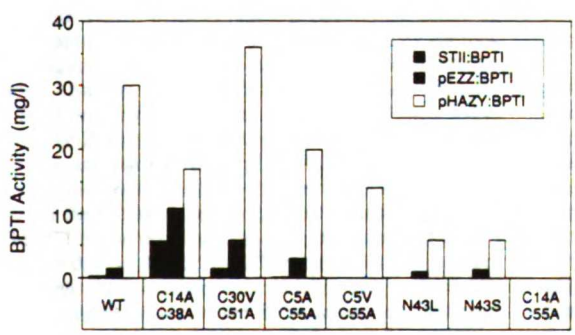


Fig. 3. Histogram comparing expression levels for three expression systems: STII:BPTI, pEZ:BPTI and pHAZY:BPTI. Yields for the STII:BPTI and pEZ:BPTI expression systems were determined as described (B.Nilsson, unpublished data). Expression levels are expressed as milligrams of BPTI activity per liter of culture in shake flask experiments. Masses are relative to the molecular weight of wild-type BPTI.

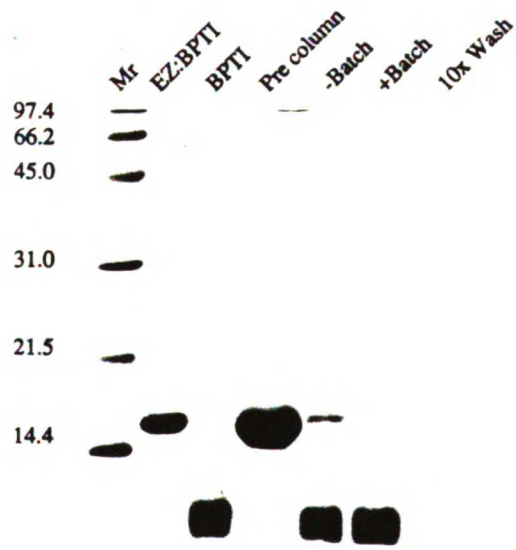


Fig. 4. 16% SDS−PAGE showing the proteins eluted off the chymotrypsin columns. **Lane 1.** BioRad Lo molecular weight markers as in Figure 2. **Lane 2.** EZ-BPTI(C14A/C38A) standard, purified on an IgG column. **Lane 3.** BPTI standard (Traysylo, Mobay Corp.). **Lane 4.** *trpLE*-Z-BPTI(C5V/C55A) from large-scale refolding reaction. **Lane 5.** *trpLE*-Z-BPTI (C5A/C55A) protein eluted off chymotrypsin column without batch mode mixing of the affinity resin. **Lane 6.** *trpLE*-Z-BPTI(C5V/C55A) protein eluted off chymotrypsin column following overnight batch mode mixing of affinity resin. **Lane 7.** Ten-fold concentrated sample from the wash of the chymotrypsin column with buffer C.

a model in which secretion of full-length 'fast folding' mutants is blocked by intracellular folding (Nilsson *et al.*, 1990; Randall *et al.*, 1987). The yield of several BPTI mutants was severely restricted in all secretion systems.

Here, we have constructed a complementary intracellular expression system for BPTI that is not subject to the complications imposed by secretion and is therefore more robust. Three features were crucial in the design of our new expression system. First, we desired high level expression without gene duplications, as described in the system of von Wilcken-Bergman *et al.* (1986). This requirement was met by the use of the strong and inducible *trp* promoter for the control of the initiation of transcription and fusion to part of the *trpLE* gene for efficient translation initiation (Yansura, 1990). Second, we required a method for stabilizing and purifying BPTI mutants that were not properly folded. Production of proteins in inclusion bodies typically provides enough stabilization and purification to proceed with the isolation of misfolded heterologous proteins but it is not possible to predict *a priori* that an expression system will lead to the formation of inclusion bodies. Accordingly, this requirement was met by sandwiching a synthetic domain of staphylococcal protein A between the TrpLE and BPTI domains. The synthetic Z domain forms an extremely stable independent folding unit (B.Nilsson, unpublished data) and its IgG-binding properties may be used to affinity purify fusions to unfolded proteins. Third, we desired a method for the cleavage of the fusion proteins that did not require the introduction of amino acid substitutions in the BPTI sequence (e.g. the replacement of Met52 for cyanogen bromide cleavages) or leave N-terminal extensions (e.g. hydroxylamine, which yields Gly-BPTI (Nilsson *et al.*, 1990), or direct production of fMet-BPTI (von Wilcken-Bergmann *et al.*, 1986). A mild enzymatic method was preferred.

All of the TrpLE-Z-BPTI mutants we have examined are expressed in inclusion bodies. Because proteins in inclusion bodies may be isolated in a nearly pure form, it is not necessary to purify the intracellularly expressed fusion proteins on IgG columns prior to refolding. Darby and Creighton (1990) have recently published a note describing a system in which the BPTI mutant M52R is produced as a fusion protein in inclusion bodies. In their system, it was necessary to purify the unfolded fusion proteins on a cation exchange resin at an extreme pH (pH2) prior to refolding. Our results indicate that purification of solubilized protein from inclusion bodies prior to refolding is not generally necessary, even for systems restricted to BPTI. Occasionally, a fraction of the TrpLE-Z-BPTI fusion protein in the refolding reactions appeared degraded, but this was minimized by washing the insoluble aggregates with Triton X-100 prior to solubilization and refolding (Babbitt *et al.*, 1990). Our ability to refold the TrpLE-Z-BPTI proteins directly from solubilized inclusion bodies suggests that the Z domain may not be required for stabilization and purification and that a simple TrpLE-BPTI fusion would meet the design requirements listed above. We chose not to pursue this course for the following reasons: (i) the cellular localization of TrpLE-BPTI fusion proteins cannot be predicted; (ii) despite the observation that the TrpLE-Z-BPTI proteins form aggregates *in vivo* in *E.coli*, the Z domain is extremely stable and soluble, and may facilitate refolding of the BPTI moiety of the fusion proteins by increasing their solubility; (iii) the Z domain provides a versatile shuttle between a secretion vector and an intracellular vector via compatible cloning sites; (iv) because of the procedures we have developed for further processing of the fusion proteins, the presence of the Z domain presents no disadvantages.

Based upon Coomassie-stained SDS gels, expression levels of TrpLE-Z-BPTI fusion proteins are not a function of mutations in the BPTI gene. Expression levels based upon the activity of refolded protein appear to be more sensitive to mutations in the BPTI gene, but the interpretation of these results requires more detailed descriptions of the folding kinetics and thermodynamic stabilities of the mutants described here. These studies, which will be the subject of a future paper, should allow the design of optimal conditions under which the fusion proteins may be refolded and the yield of trypsin inhibitor activity subsequently optimized. Despite the qualifications we have introduced when comparing yields based on activities of refolded protein, the significance of this work lies more in the absolute yield of previously poorly expressed mutants, e.g. C5V/C55A and N43L, than in the differences among mutants; indeed, we have recently purified >200 mg of C5V/C55A from 100 g of wet *E.coli* cell paste.

Cleavage of fusion proteins

In the course of attempts to purify intact secreted Z-BPTI fusion proteins on chymotrypsin columns, we discovered several chymotrypsin-sensitive sites in the Z domain. This led us to test the seemingly paradoxical idea that the fusion protein could be precisely cleaved at an engineered site by a protease that is inhibited by one of the domains of the fusion protein. Our results clearly demonstrate the success of this approach. However, the detailed mechanism by which the fusion proteins are cleaved is still a matter of speculation. We present three possibilities. Some fraction of the proteins may be cleaved at the designed site before binding to the chymotrypsin in the inhibitory mode. Alternatively, the fusion proteins may have an appreciable off rate from their complex with chymotrypsin (Lazdunski *et al.*, 1974) which will allow them to be cleaved after an initial binding step. In a more likely mechanism, the fusion proteins are first immobilized through binding to immobilized chymotrypsin and are then cleaved by a second molecule of chymotrypsin. The accessibility of the engineered site in the chymotrypsin–BPTI complex may be inferred from a structural model. In the crystal structure of the trypsin–BPTI complex (Deisenhofer and Steigmann, 1975) the N-terminus of BPTI is well exposed to solvent and is at the opposite end of the inhibitor molecule from Lys15, which fits into the primary specificity pocket of trypsin. We have assumed a similar structural model for the BPTI–chymotrypsin complex. This places the linker region of the immobilized fusion protein as far as possible from the chymotrypsin molecule to which it is complexed, where it is effectively poised for cleavage by a second molecule of chymotrypsin. In support of this hypothesis, we have observed chymotryptic activity towards the substrate *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanilide in aliquots of column washes following BPTI binding, indicating that chymotrypsin slowly leaches off the column, even after repeated cycles of BPTI binding at pH 7.8 and release at low pH. The release of chymotrypsin from the column is not so rapid as to release significant quantities of chymotrypsin–BPTI complexes, as the recovery of BPTI activity is quite high. However, once the Z-BPTI fusion proteins have bound to the column, only enzymatic quantities of free chymotrypsin are necessary to cleave the proper peptide bond in the linker region between the Z and BPTI domains.

The use of chymotrypsin to cleave the Z-BPTI fusion proteins provides the additional benefit of a powerful purification in the same step. Fusion proteins which are not properly refolded are either cleaved by the chymotrypsin or simply pass through the column; this assures that all proteins which make it through the

procedure are active protease inhibitors. The procedure also removes the cleaved protein A domain, which is simply washed off the affinity column following cleavage. Some downstream purification is necessary to remove traces of leached chymotrypsin or inhibitor molecules which may have ragged ends, but these represent only a small fraction of the protein eluted from the column.

Our main interest has been in BPTI mutants which affect the folding pathway or stability of the protein and not the trypsin inhibitor activity. However, we have produced a mutant, TrpLE-Z-BPTI(C38A), which when refolded as described above is predominantly in the mixed disulfide form with glutathione and which does not bind to trypsin or chymotrypsin in our stoichiometric inhibition assays. Chymotrypsin in solution cleaves this fusion protein to yield a correct N-terminus at molar ratios of 40:1, fusion protein:chymotrypsin, at 25°C, pH 7.8 with yields >90% (data not shown). This modification extends the enzymatic cleavage methodology to other mutants which are no longer inhibitors of chymotrypsin but which still retain a protease-resistant fold.

General utility of the pDH:108 expression vector

Several other small proteins, including IGF-1, lung surfactant SPC, atrial natriuretic factor (ANF) (D.Henner, unpublished data) and kistrin (P.Carter, M.Dennis and R.Lazurus, personal communication) have been expressed as TrpLE-Z fusion proteins. Inclusion body formation was observed in some, but not all of these systems, in a growth-condition-dependent manner. When protein was expressed in a soluble form, it could be purified, prior to refolding when necessary, from the mixture of other cytoplasmic *E.coli* proteins by IgG affinity chromatography (M.Dennis and P.Carter, personal communication). Therefore, the expression system has potential broad utility. The chymotrypsin cleavage scheme may not be appropriate for most TrpLE-Z fusion proteins, but other mild enzymatic cleavage methods may be adapted for individual needs (Carter, 1990).

Future work

The work described here provides the basis for a large variety of other studies of mutants that were previously unavailable in high yields. The structure of the C5V/C55A mutant has been solved recently to a resolution of 1.3 Å from a crystal with a space group identical to wild-type crystal form II (M.Randal, personal communication); folding and stability studies of a number of Δ5-55 mutants are also in progress. Studies which examine the structural consequences of amino acid replacements at the buried Asn43 are of considerable interest (Goldenberg *et al.*, 1989) and are currently under way. We are also interested in a number of NMR experiments that utilize biosynthetically isotope-labeled protein. We have begun a series of model experiments which utilize randomly deuterated protein (LeMaster and Richards, 1988) and nitroxide spin labeling (Kosen *et al.*, 1986) for the measurement of many long-range distances in BPTI. The highly efficient expression system described here is indispensible to these studies.

Finally we speculate that the combined fusion protein cleave/affinity purification procedure we have developed may be useful for processing other recombinantly produced fusions to protein inhibitors of broad specificity proteases. The advantages of the method are that it is extremely simple to apply, it does not require mutation of the amino acid sequence to remove sites which may be sensitive to cleavage by chemical reagents, and it will not introduce impurities due to side reactions caused by harsh reaction conditions.

Acknowledgements

J.D.A. is grateful to Dr Anthony Kossiakoff and Genentech, Inc. for providing lab space and support. We thank the following additional people at Genentech: Mike Covarrubias for fermentations, Byron Nevins and Andrea Louie for protein sequencing, James Bourell and Karl Clauser for MS data and Carol Morita for assistance with the preparation of figures. We thank Dr Phyllis Kosen for advice on protein folding, and we are grateful to Dr Lars Abrahmsen and Dr Paul Carter for providing critical readings of the manuscript. We wish to acknowledge the following grant support during the course of this work: NSF grant DMB-860691 to S.A. and NIH grant RO1 GM19267 to I.D.K.

References

Auerswald,E.-A., Schroder,W. and Kotick,M. (1987) *Biol. Chem. Hoppe-Seyler*, **368**, 1413–1425.
Babbitt,P.C., West,B.L., Buechter,D.D., Kuntz,I.D. and Kenyon,G.L. (1990) *Bio/Tech.*, **8**, 945–949.
Brown,L.R., De Marco,A., Richarz,R., Wagner,G. and Wüthrich,K. (1978) *Eur. J. Biochem.*, **88**, 87–95.
Carter,P. (1990) In Ladisch,M.R., Wilson,R.C., Painton,Chih-duen,C. and Builder,S.E. (eds), *Protein Purification: from Molecular Mechanisms to Large-Scale Processes*. American Chemical Society Symposium Series no. 427, pp. 181–193.
Carter,P. (1991) *Mutagenesis: a Practical Approach*. IRL Press, Oxford, UK, in press.
Chung,C.T. and Miller,R.H. (1988) *Nucleic Acids Res.*, **16**, 3580.
Coplen,L.J., Frieden,R.W. and Goldenberg,D.P. (1990) *Proteins. Struct. Funct. Genet.*, **7**, 16–31.
Creighton,T.E. (1977) *J. Mol. Biol.*, **113**, 275–293.
Creighton,T.E. and Charles,I.G. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 511–519.
Creighton,T.E. and Goldenberg,D.P. (1984) *J. Mol. Biol.*, **179**, 497–526.
Darby,N.J. and Creighton,T.E. (1990) *Nature*, **344**, 715–716.
de Boer,H.A., Comstock,L.J., Yansura,D. and Heyneker,H. (1982) In Rodriguez,R.L. and Chamberlain,M.J. (eds), *Promoters: Structure and Function*. Praeger, New York, pp. 462–481.
Deisenhofer,J. and Steigmann,W. (1975) *Acta Crystallogr., B*, **31**, 238–240.
Eigenbrot,C., Randal,M. and Kossiakoff,A.A. (1990) *Protein Engng*, **3**, 591–598.
Goldenberg,D.P. (1988) *Biochemistry*, **27**, 2481–2489.
Goldenberg,D.P. (1989) In Creighton,T.E. (ed.), *Protein Structure: a Practical Approach*. IRL Press, Oxford, pp. 225–250.
Goldenberg,D.P. and Creighton,T.E. (1984) *J. Mol. Biol.*, **179**, 527–545.
Goldenberg,D.P., Frieden,R.W., Haack,J.A. and Morrison,T.B. (1989) *Nature*, **38**, 127–132.
Henzel,W.J., Rodriguez,H. and Watanabe,C. (1987) *J. Chromatogr.*, **404**, 41–52.
Hurle,M.R., Marks,C.B., Kosen,P.A., Anderson,S. and Kuntz,I.D. (1990) *Biochemistry*, **29**, 4410–4419.
Hurle,M.R., Anderson,S. and Kuntz,I.D. (1991) *Protein Engng*, **4**, 451–455.
Kosen,P.A., Scheek,R.M., Naderi,H., Basus,V.J., Manogaran,S., Schmidt,P.G., Oppenheimer,N.J. and Kuntz,I.D. (1986) *Biochemistry*, **25**, 2356–2364.
Laemmli,U. (1970) *Nature*, **227**, 680–685.
Lazdunski,M., Vincent,J.-P., Schweitz,H., Person-Renner,M. and Pudles,J. (1974) In Fritz,H., Tschesche,H., Greene,L.J. and Truscheit,E. (eds), *Proteinase Inhibitors—Bayer Symposium V*. Springer Verlag, Berlin, pp. 420–431.
LeMaster,D.M. and Richards,F.M. (1988) *Biochemistry*, **27**, 142–150.
Levitt,M. (1981) *J. Mol. Biol.*, **145**, 251–263.
Lowenadler,B., Jansson,B., Paleus,S., Holmgren,E., Nilsson,B., Mocks,T., Palm,G., Josephson,S., Philipson,L. and Uhlen,M. (1987) *Gene*, **58**, 87–97.
Marks,C.B., Vasser,M., Ng,P., Henzel,W. and Anderson,S. (1986) *J. Biol. Chem.*, **261**, 7115–7118.
Marks,C.B., Naderi,H., Kosen,P.A., Kuntz,I.D. and Anderson,S. (1987a) In Oxender,D.L. (ed.), *Protein Structure, Folding and Design*. Alan R. Liss, New York, pp. 335–340.
Marks,C.B., Naderi,H., Kosen,P.A., Kuntz,I.D. and Anderson,S. (1987b) *Science*, **235**, 1370–1373.
Meselson,M. and Yuan,R. (1968) *Nature*, **217**, 1110–1114.
Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Miozzari,G.F. and Yanofsky,C. (1978) *J. Bacteriol.*, **133**, 1457–1466.
Nagai,K. and Thøgersen,H.C. (1987) *Methods Enzymol.*, **153**, 461–481.
Nakamaye,K.L. and Eckstein,F. (1986) *Nucleic Acids Res.*, **14**, 9679–9698.
Nilsson,B. and Abrahmsén,L. (1990) *Methods Enzymol.*, **185**, 144–161.
Nilsson,B., Moks,T., Jansson,B., Abrahmsén,L., Elmblad,A., Holmgren,E., Henrichson,C., Jones,T.A. and Uhlén,M. (1987) *Protein Engng*, **1**, 107–113.

Nilsson.B., Kuntz.I.D. and Anderson.S. (1990) In King.J. and Gierasch.L. (eds), *Protein Folding. Deciphering the Second Half of the Genetic Code*. American Association for the Advancement of Science, Washington, DC, pp. 117–122.
Nilsson.B., Berman-Marks.C., Kuntz.I.D. and Anderson.S. (1991) *J. Biol. Chem.*, **266**, 2970–2977.
Randall.L.L., Hardy.S.J.S. and Thom.J.R. (1987) *Annu. Rev. Microbiol.*, **141**, 507–541.
Sanger.F., Nicklen.S. and Coulsen.A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
Scholtissek.S. and Grosse.F. (1987) *Nucleic Acids Res.*, **15**, 3185.
Sutcliffe.J.G. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 77–90.
Vieira.J. and Messing.J. (1987) *Methods Enzymol.*, **153**, 3–11.
Vincent.J.-P., Chicheportiche.R. and Lazdunski.M. (1971) *Eur. J. Biochem.*, **23**, 401–411.
von Wilcken-Bergmann.B., Tils.D., Sartorius.J., Auerswald.E.A., Schroeder.W. and Müller-Hill.B. (1986) *EMBO J.*, **5**, 3219–3225.
Yansura.D.G. (1990) *Methods Enzymol.*, **185**, 161–166.

Received on November 11, 1990; accepted February 19, 1991

Note added in proof

It has recently come to our attention that two additional expression systems have been described. The references are as follows.

Auerswald.E.A., Horlein.D., Reinherdt.G., Schroder.W. and Schnabel.E. (1988) *Biol. Chem. Hoppe Seyler*, **369**, 27–35.
Norris.K., Norris.F., Bjorn.S.E., Diers.I. and Peterson.L.C. (1990) *Biol. Chem. Hoppe Seyler*, **371**, 37–42.

Appendix 2A

2A. Comments on the workup of BPTI from the pHAZY expression system

The procedures for refolding the guanidinium hydrochloride solubilized TrpLE:Z:BPTI fusion proteins and the subsequent purification steps that I presented in the **Protein Engineering** manuscript are sketched in the flow chart in figure 2A.1. Many of these procedures are unwieldy due to the large scale of the workup. In this appendix, I present several suggestions for simplifying or otherwise improving these procedures.

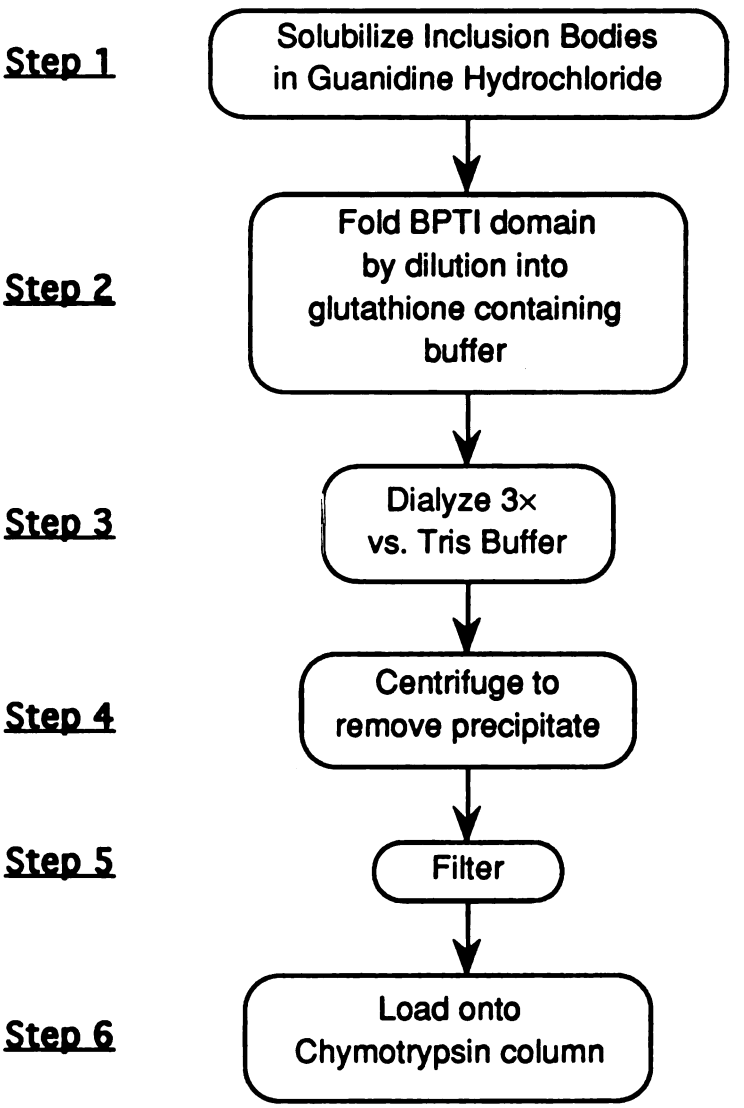


Figure 2A.1: Flow chart of procedures for the workup of isolated TrpLE:Z:BPTI fusion protein containing inclusion bodies.

2A.1. The effect of oxygen during on refolding (Figure 2A.1, Step 2)

The large scale refolding reactions are carried out in one or two liter vessels, usually plastic beakers. Creighton-type folding kinetics experiments with BPTI are usually performed on a small scale in glass vessels (Creighton, 1984a), but early on I seemed to obtain higher yields in plastic vessels (in uncontrolled experiments), so I have retained the practice. Accurate kinetic measurements of the steps in refolding require anaerobic reactions conditions (Creighton, 1984a), but I made no attempts to exclude oxygen from the large scale preparative refolding reactions. For convenience, these reactions are carried out overnight at room temperature. At the end of the reaction, the solutions yield a negative test for free thiols by the Ellman assay (Ellman, 1959; Creighton, 1989), indicating a depletion of reduced glutathione by air oxidation. Higher yields of correctly refolded BPTI might therefore be obtained by the exclusion of oxygen from this reaction, but this is not a certainty. In addition, it is extremely inconvenient to carry out anaerobic reactions at this scale in most protein chemistry labs.

2A.2. Comments on dialysis and clarification steps required after refolding (Figure 2A.1, Steps 3-5)

I found it necessary to change the buffer between the refolding reaction and the application of the protein to the chymotrypsin cleavage/affinity column. Binding of the inhibitor to the column may have been prevented by the presence of 0.7 M guanidine hydrochloride or residual glutathione which may inactivate chymotrypsin (though this is not likely, considering the negative Ellman assay mentioned in the previous section). The buffer exchange was accomplished by dialysis, a relatively labor free process. In this case, however, the large volume of the reaction solution required three dialysis steps against at least 50 liters of Tris buffer, which is sometimes inconvenient on the lab scale. Furthermore, a flocculent precipitate developed in both the refolding reactions and during the dialysis that was impossible to remove by filtration through 0.45 μm membranes (even

with prefilters or prefiltration through Whatmann paper) or by centrifugation in large bottles (due to the limited *g* force available in rotors such as the Sorval GS-3 or the GS-A). My solution to this was to centrifuge the dialyzed solutions in 45 ml batches in a Sorval SW-34 rotor at high RPMs, followed by filtration through 0.45 μm membranes. This is an inconvenient process, requiring a couple of hours of "feeding" the centrifuge. I have attempted to use high surface area 0.45 μm filters such as the MilliPak60 from Millipore, omitting the centrifugation step, but these also tended to clog about half way through the filtration. A more convenient procedure is obviously desirable.

I attempted to reduce the volume of the protein solution prior to dialysis by carrying out the centrifugation first, followed by an ultrafiltration step on an Amicon YM-1 or YM-2 membrane. This is also not optimal, due to the slow speed of ultrafiltration on these membranes (even at room temperature; concentration at 4 °C was not possible) and due to the formation of an additional precipitate during the concentration step which completely clogged the membranes.

There are a number of alternatives which I have not yet investigated. These include the use of filter aids such as Celite. It may also be possible to use Fast Flow Q Sepharose (Pharmacia), a relatively cheap anion exchange resin—it will not bind the fusion protein at an appropriate pH—as a filter (P. Carter and R. Kelley, personal communication).

Concentration by ion exchange should also be considered. It will be necessary to adjust the pH to the range of 4-5 and to further dilute the refolded protein by two to three volumes to lower the concentration of guanidineadine hydrochloride. This brings the total volume to at least six liters, but this can be easily loaded onto the Fast Flow S Sepharose resin (Pharmacia) in a reasonable period of time, even in the presence of the precipitate. This procedure has the advantage of removing the guanidine and glutathione by washing the column following loading, allowing omision of the dialysis steps. The protein can be eluted in a relatively small volume by a step gradient to a high salt (NaCl) concentration. The pH of the eluate can be adjusted to approximately 8 (with the addition of an appropriate

buffer such as Tris) prior to loading the chymotrypsin column; if the high salt concentration appears to be a problem during the chymotrypsin step, the solution can be diluted and the volumes will still be manageable. Finally, the substitution of urea for guanidine hydrochloride in the solubilization of the inclusion bodies might be considered, thus facilitating the ion exchange step. I have chosen to use guanidine to avoid the possibility of protein carbamylation by cyanates which form over time in concentrated urea solutions. The use of freshly prepared solutions of Ultra-Pure urea will minimize this problem.

Along similar lines, protein concentration and removal of guanidine and glutathione by hydrophobic interaction chromatography can also be considered. This requires addition of ammonium sulfate to 2 M to the refolding reaction, and no dilution, prior to column loading. On the other hand, the Fast Flow Phenyl Sepharose resins are rather expensive, and recovery is not as good as with the ion exchange resins.

Precipitation proteins by either ammonium sulfate or organic solvents such as cold acetone is a common method of fractionation and/or concentration. It was not considered here because the protein solutions are already rather dilute—around 0.2 mg/ml—limiting the recovery of precipitated protein.

Finally, concentration by "dialysis" against dry polyethylene glycol (PEG) could be considered (Pohl, 1990). However, this is a slow process which may require several changes of PEG. In addition, a further dialysis step will be required to remove the remaining guanidine and glutathione.

In summary, the most attractive procedure appears to be the combined concentration and buffer exchange afforded by cation exchange on a Fast Flow S Sepharose resin. The method is both fast and relatively inexpensive and includes a minimum number of high recovery steps prior to loading the chymotrypsin column.

Appendix 2B

2B. Guide to cloning sites and restriction maps

This appendix is intended as a guide to the use of the complementary BPTI cloning and expression vectors as they currently exist, including the pHAZY vectors for the intracellular expression of TrpLE-Z-BPTI fusion proteins, the pEZ vectors for expression of secreted Z-BPTI fusion proteins, and the ST:II based vectors for the direct expression of BPTI. Protein expression from the pHAZY vectors is certainly the most robust (*i.e.*, it is possible to obtain high yields of nearly any mutant desired), but it is not possible to make mutants in them with methods that rely on the production of single stranded DNA. Quite reasonable yields may be obtained from the pEZ or ST:II vectors for many mutants, but some mutants will not express at all. On the other hand, both sets of vectors contain single strand origins, allowing the construction of mutants by any number of methods based on mismatch primer directed mutagenesis. These mutants may then be subcloned into the pHAZY vectors using a small number of appropriate restriction sites. This appendix will present a consideration of the most useful sites.

Before I begin, I must make a few comments concerning cassette mutagenesis. From a naive point of view, the small BPTI gene would seem an ideal system for cassette mutagenesis, which would allow the construction of mutants in pHAZY directly without the need for single stranded DNA. However, reverse translation of the BPTI protein sequence into ambiguous nucleotides reveals relatively few sites where silent mutations can be made in order to introduce new, unique restriction sites. The sites which do exist in the BPTI gene are not spaced optimally for large scale mutagenesis projects using cassette mutagenesis. New, solid phase mutagenesis methods may be applied in this system (Hultman *et al.*, 1990) (B. Nilsson, personal communication), but for now the oligo-directed mismatch mutagenesis methods utilizing single stranded DNA remain standard.

2B.1. A practical note on the use of ssDNA from the pEZ and pST:II series

The phage F1 single strand origin in the pEZ series directs the packaging of the *sense* DNA strand, relative to the BPTI gene. Therefore, sequencing and mutagenic oligonucleotides must be synthesized to match the antisense strand. The single strand origin in pST:II directs the packaging of the antisense strand, requiring the use of oligonucleotides corresponding to the sense strand.

2B.2. Restriction sites useful for subcloning

Sites within the BPTI gene - There are three useful restriction sites in some, but not all BPTI mutants in each of the three vectors. A *Sty* I site spans residues Ala(25)-Lys(26)-Ala(27) and is present in all BPTI vectors we have constructed. The *Pst* I site spanning residues Gly(37)-Cys(38)-Arg(39) is removed upon mutation of Cys-38, most frequently to alanine. Finally, there is a *Stu* I site spanning residues Arg(1)-Pro(2) in all of the pEZ and pHAZY vectors* . This site was introduced by Björn Nilsson during the construction of the pEZ vectors; most of the ST:II series precede the pEZ series and therefore lack the *Stu* I site. The *Sty* I and *Pst* I sites are unique only in the pEZ series.

Sites within the protein A "Z" domain - The only site within the Z domain that is frequently used is the *Bgl* II site.

Sites flanking the coding regions - Sites 5' to the coding region are typically located in promoter regions and are not often used to move mutants from one vector to another. The *Not* I octanucleotide site in the pEZ series defines the 5' end of the more general protein A

---

* The only exception to this is the pEZ:His3:wt plasmid which contains the Ala-Ala-Pro-Phe linker immediately before Arg-1 and has Pro-2 mutated to alanine. This mutant was designed solely for the purpose of the cleavage of the fusion protein by Ala-64 subtilisin, which was prevented by the proline in the P2' position.

secretion vector expression cassette. It is therefore useful for subcloning mutants into yet another series of plasmids, designated pK3, which contain a gene for kanamycin resistance, an intracellular marker. The pHAZY series contains an *Xba* I site between the Shine-Dalgarno sequence (Shine and Dalgarno, 1975) and the initiation codon; the transcription start site is -29 to the translation initiation codon. The St:II series contains the identical sequence between the trascription start site through the *Xba* I site, but otherwise the 5' promoter regions are very different, destroying the utility of the *Xba* I sites.

The most useful site at the 3' end of the gene is the *Eco*R I site. This site is unique in the pEZ and pHAZY series, but there is an additional *Eco*R I site in the ST:II series. There are *Cla* I and *Hin*d III sites in the 3' region of all the vectors, but these should be used with caution. The *Cla* I site is not unique in pEZ. In addition, the *Hin*d III site is in the promoter region of the tetracycline resistance gene in both pHAZY and ST:II. Finally, the sequences between the *Eco*R I and *Hin*d III sites differ in pHAZY from those of pEZ and ST:II; the λ t_0 transcription terminator is cloned in this position in pHAZY.

The wild type BPTI gene in the pEZ series differs from all other mutants in the same series due to its lack of a 3' *Eco*R I site, an artifact of an early subcloning procedure (B. Nilsson, personal communication). This fact is a little annoying, particularly when one wishes to make point mutations in the wild type gene in pEZ and then subclone them into pHAZY. The 3' sequences of pEZ:wt should be made identical to the rest of the series.

Outside of the 5' promoter and translated leader peptide sequences, there is another important difference between pEZ and pHAZY. The linker region between the Z domain in pEZ contains a hydroxylamine sensitive Asn-Gly site, while pHAZY replaces this with a chymotrypsin sensitive Thr-Tyr site. This difference makes subcloning between the vectors using the *Bgl* II site in the Z domain less useful, which is unfortunate. Instead, *Stu* I becomes the most practical 5' site for subcloning bewteen pEZ and pHAZY. Use of this site gives generally small *Stu* I-*Eco*R I fragments (186 bp) with a 5' blunt end that are a lit-

tle more difficult to subclone than are *Bgl* II-*Eco*R I fragments (332 bp), but both fragments have been subcloned reproducibly.

In summary, most mutagenesis and expression work can be carried out using only pEZ and pHAZY. Mutant DNA sequences are generally made in pEZ and subcloned into pHAZY on *Stu* I-*Eco*R I fragments. Deviations from the subcloning procedure are currently required when making mutants starting from pEZ:wt, but this should be corrected in the future. The only way to move genes from ST:II into pHAZY is on very small *Sty* I-*Eco*R I (113 bp) or *Pst* I-*Eco*R I (76 bp) fragments, but both procedures require inconvenient partial digestions of the pHAZY vector plasmids. It will certainly be useful to prepare a silent mutant of the ST:II wild type gene with the *Stu* I site spanning the Arg(1)-Pro(2) sequence.

Complete restriction and circular plasmid maps for the pHAZY, pEZ, and pST:II plasmids are given in the pages to follow.

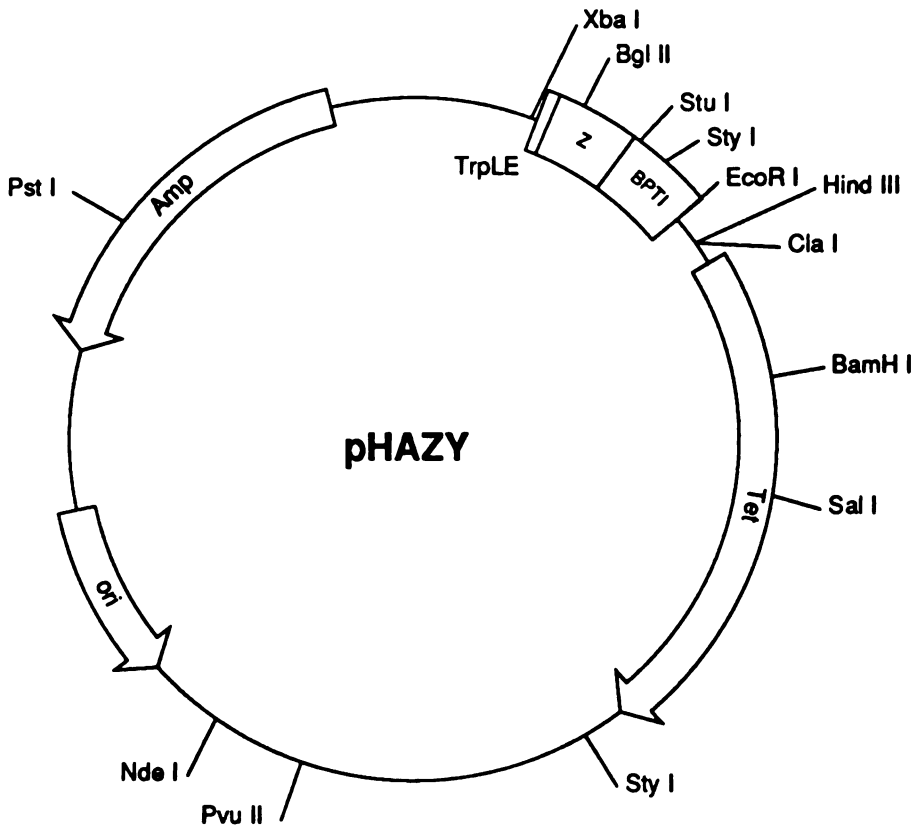


Figure 2B.1 The pHAZY expression plasmid

MAP restriction endonuclease sites:
Cut positions given are the last base of the fragment on the 5'-side (left side) of the cut.
The TrpLE and linker protein sequences are in normal font. *The Z domain is in italics.* BPTI is in underlined.

The following enzymes were examined:

AATI	AATII	ACCIII	AFLII	ALWNI	AOSI
APAI	APALI	ASEI	ASP700I	ASP718I	ASUII
AVIII	AVRII	AXYI	BALI	BAMHI	BBEI
BBUI	BCLI	BGLI	BGLII	BSMI	BSPHI
BSPMI	BSPMII	BSSHII	BSTBI	BSTEII	BSTI
BSTXI	BSU36I	CLAI	CSP45I	CVNI	DRAI
DRAIII	DRDI	EAGI	EARI	ECO47III	ECO52I
ECO81I	ECONI	ECO065I	ECORI	ECORV	ECOT22I
ESPI	FSPI	HINDIII	HPAI	KPNI	KSP632I
MLUI	MROI	NAEI	NARI	NCOI	NDEI
NHEI	NOTI	NRUI	NSII	NUNII	PAER7I
PFLMI	PSTI	PVUI	PVUII	RSPXI	RSRII
SACI	SACII	SALI	SAUI	SCAI	SFII
SMAI	SNABI	SPEI	SPHI	SPLI	SPOI
SSPI	SSTI	SSTII	STUI	STYI	TTH111I
XBAI	XHOI	XMAI	XMAIII	XMNI	XORII

C1 (1f): |>u 1>+++++ ss (5118 bases)+++++>u 5118>|

```
        10          20          30          40          50          60
GAATTAATTC ATGCTGTGGT GTCATGGTCG GTGATCGCCA GGGTGCCGAC GCGCATCTCG
    ASP700I
    XMNI
     ASEI

        70          80          90         100         110         120
ACTGCACGGT GCACCAATGC TTCTGGCGTC AGGCAGCCAA TCGGAAGCTG TGGTATGGCT
           APALI                                                BSPMI

       130         140         150         160         170         180
GTGCAGGTCG TAAATCACTG CATAATTCGT GTCGCTCAAG GCGCACTCCC GTTCTGGATA

       190         200         210         220         230         240
ATGTTTTTTG CGCCGACATC ATAACGGTTC TGGCAAATAT TCTGAAATGA GCTGTTGACA
                                   SSPI

       250         260         270         280         290         300
                                                                   Met
ATTAATCATC GAACTAGTTA ACTAGTACGC AAGTTCACGT AAAAAGGGTA TCTAGAATTA
  ASEI          SPEI        SPEI                                XBAI
                 HPAI

       310         320         330         340         350         360
 LysAlaIl ePheValLeu AsnAlaGlnH isAspGluAl a*ValAspAsn* *LysPheAsnLys*
TGAAAGCAAT TTTCGTACTG AATGCGCAAC ACGATGAAGC CGTAGACAAC AAATTCAACA
                         AVIII
                         FSPI
                         AOSI
                          BSMI

       370         380         390         400         410         420
 *GluGlnGl nAsnAlaPhe TyrGluIleL euHisLeuPr oAsnLeuAsn GluGluGlnArg*
AAGAACAACA AAACGCGTTC TATGAGATCT TACATTTACC TAACTTAAAC GAAGAACAAC
             MLUI          BGLII

       430         440         450         460         470         480
 *AsnAlaPh eIleGlnSer LeuLysAspA spProSerGl nSerAlaAsn LeuLeuAlaGlu*
GAAACGCCTT CATCCAAAGT TTAAAAGATG ACCCAAGCCA AAGCGCTAAC CTTTTAGCAG
                      DRAI                    ECO47III

       490         500         510         520         530         540
 *AlaLysLy sLeuAsnAsp AlaGlnAlaP roLys*ValAs pMetThrTyr ArgProAspPhe
AAGCTAAAAA GCTAAATGAT GCTCAGGCGC CGAAAGTAGA CATGACTTAT AGGCCTGACT
                          NUNII                       STUI
                          NARI                        AATI
                           BBEI

       550         560         570         580         590         600
 CysLeuGl uProProTyr ThrGlyProA laLysAlaAr gIleIleArg TyrPheTyrAsn
TCTGCCTAGA GCCTCCATAT ACGGGTCCCG CTAAGGCCAG AATTATCAGA TACTTCTACA

       610         620         630         640         650         660
 AlaLysAl aGlyLeuCys GlnThrPheV alTyrGlyGl yAlaArgAla LysArgAsnAsn
ACGCCAAGGC TGGGCTCTGC CAGACCTTTG TATATGGTGG CGCTAGAGCT AAGAGAAACA
    STYI

       670         680         690         700         710         720
 PheLysSe rAlaGluAsp CysMetArgT hrCysGlyGl yAla***
ATTTCAAGAG CGCAGAGGAC TGCATGAGGA CCTGTGGTGG TGCTTAAGGG CCCCGGGAAT
                                                AFLII APAI
                                                      SMAI
                                                       XMAI

       730         740         750         760         770         780
TCCTAACGCT CGGTTGCCGC CGGGCGTTTT TTATTGTTAA CTCATGTTTG ACAGCTTATC
ECORI                                HPAI
```

```
      790         800         810         820         830         840
ATCGATAAGC TTTAATGCGG TAGTTTATCA CAGTTAAATT GCTAACGCAG TCAGGCACCG
   CLAI    HINDIII

      850         860         870         880         890         900
TGTATGAAAT CTAACAATGC GCTCATCGTC ATCCTCGGCA CCGTCACCCT GGATGCTGTA

      910         920         930         940         950         960
GGCATAGGCT TGGTTATGCC GGTACTGCCG GGCCTCTTGC GGGATATCGT CCATTCCGAC
                                                ECORV

      970         980         990        1000        1010        1020
AGCATCGCCA GTCACTATGG CGTGCTGCTA GCGCTATATG CGTTGATGCA ATTTCTATGC
                                   NHEI                          AVIII
                                    ECO47III                     FSPI
                                                                 AOSI

     1030        1040        1050        1060        1070        1080
GCACCCGTTC TCGGAGCACT GTCCGACCGC TTTGGCCGCC GCCCAGTCCT GCTCGCTTCG

     1090        1100        1110        1120        1130        1140
CTACTTGGAG CCACTATCGA CTACGCGATC ATGGCGACCA CACCCGTCCT GTGGATCCTC
                                                              BAMHI
                                                              BSTI

     1150        1160        1170        1180        1190        1200
TACGCCGGAC GCATCGTGGC CGGCATCACC GGCGCCACAG GTGCGGTTGC TGGCGCCTAT
                     NAEI          NUNII                     NUNII
                                   NARI                      NARI
                                    BBEI                      BBEI

     1210        1220        1230        1240        1250        1260
ATCGCCGACA TCACCGATGG GGAAGATCGG GCTCGCCACT TCGGGCTCAT GAGCGCTTGT
                                                    RSPXI
                                                    BSPHI
                                                       ECO47III

     1270        1280        1290        1300        1310        1320
TTCGGCGTGG GTATGGTGGC AGGCCCCGTG GCCGGGGGAC TGTTGGGCGC CATCTCCTTG
                                                  NUNII
                                                  NARI
                                                   BBEI

     1330        1340        1350        1360        1370        1380
CATGCACCAT TCCTTGCGGC GGCGGTGCTC AACGGCCTCA ACCTACTACT GGGCTGCTTC
   SPHI
   BBUI

     1390        1400        1410        1420        1430        1440
CTAATGCAGG AGTCGCATAA GGGAGAGCGT CGACCGATGC CCTTGAGAGC CTTCAACCCA
    ECONI                          SALI

     1450        1460        1470        1480        1490        1500
GTCAGCTCCT TCCGGTGGGC GCGGGGCATG ACTATCGTCG CCGCACTTAT GACTGTCTTC

     1510        1520        1530        1540        1550        1560
TTTATCATGC AACTCGTAGG ACAGGTGCCG GCAGCGCTCT GGGTCATTTT CGGCGAGGAC
                           NAEI   ECO47III

     1570        1580        1590        1600        1610        1620
CGCTTTCGCT GGAGCGCGAC GATGATCGGC CTGTCGCTTG CGGTATTCGG AATCTTGCAC

     1630        1640        1650        1660        1670        1680
GCCCTCGCTC AAGCCTTCGT CACTGGTCCC GCCACCAAAC GTTTCGGCGA GAAGCAGGCC

     1690        1700        1710        1720        1730        1740
ATTATCGCCG GCATGGCGGC CGACGCGCTG GGCTACGTCT TGCTGGCGTT CGCGACGCGA
        NAEI BGLI      XMAIII                          SPOI
                       ECO52I                          NRUI
                       EAGI
```

```
      1750        1760        1770        1780        1790        1800
GGCTGGATGG CCTTCCCCAT TATGATTCTT CTCGCTTCCG GCGGCATCGG GATGCCCGCG

      1810        1820        1830        1840        1850        1860
TTGCAGGCCA TGCTGTCCAG GCAGGTAGAT GACGACCATC AGGGACAGCT TCAAGGATCG
                BSPMI

      1870        1880        1890        1900        1910        1920
CTCGCGGCTC TTACCAGCCT AACTTCGATC ACTGGACCGC TGATCGTCAC GGCGATTTAT

      1930        1940        1950        1960        1970        1980
GCCGCCTCGG CGAGCACATG GAACGGGTTG GCATGGATTG TAGGCGCCGC CCTATACCTT
      BGLI                                              NUNII
                                                        NARI
                                                         BBEI

      1990        2000        2010        2020        2030        2040
GTCTGCCTCC CCGCGTTGCG TCGCGGTGCA TGGAGCCGGG CCACCTCGAC CTGAATGGAA

      2050        2060        2070        2080        2090        2100
GCCGGCGGCA CCTCGCTAAC GGATTCACCA CTCCAAGAAT TGGAGCCAAT CAATTCTTGC
  NAEI                                      PFLMI

      2110        2120        2130        2140        2150        2160
GGAGAACTGT GAATGCGCAA ACCAACCCTT GGCAGAACAT ATCCATCGCG TCCGCCATCT
                AVIII       PFLMI
                FSPI        STYI
                AOSI
                 BSMI

      2170        2180        2190        2200        2210        2220
CCAGCAGCCG CACGCGGCGC ATCTCGGGCA GCGTTGGGTC CTGGCCACGG GTGCGCATGA
                                                  BALI        AVIII
                                                              FSPI
                                                              AOSI

      2230        2240        2250        2260        2270        2280
TCGTGCTCCT GTCGTTGAGG ACCCGGCTAG GCTGGCGGGG TTGCCTTACT GGTTAGCAGA

      2290        2300        2310        2320        2330        2340
ATGAATCACC GATACGCGAG CGAACGTGAA GCGACTGCTG CTGCAAAACG TCTGCGACCT

      2350        2360        2370        2380        2390        2400
GAGCAACAAC ATGAATGGTC TTCGGTTTCC GTGTTTCGTA AAGTCTGGAA ACGCGGAAGT

      2410        2420        2430        2440        2450        2460
CAGCGCCCTG CACCATTATG TTCCGGATCT GCATCGCAGG ATGCTGCTGG CTACCCTGTG
                          BSPMII
                          MROI
                          ACCIII

      2470        2480        2490        2500        2510        2520
GAACACCTAC ATCTGTATTA ACGAAGCGCT GGCATTGACC CTGAGTGATT TTTCTCTGGT
                          ECO47III

      2530        2540        2550        2560        2570        2580
CCCGCCGCAT CCATACCGCC AGTTGTTTAC CCTCACAACG TTCCAGTAAC CGGGCATGTT

      2590        2600        2610        2620        2630        2640
CATCATCAGT AACCCGTATC GTGAGCATCC TCTCTCGTTT CATCGGTATC ATTACCCCCA

      2650        2660        2670        2680        2690        2700
TGAACAGAAA TTCCCCCTTA CACGGAGGCA TCAAGTGACC AAACAGGAAA AAACCGCCCT

      2710        2720        2730        2740        2750        2760
TAACATGGCC CGCTTTATCA GAAGCCAGAC ATTAACGCTT CTGGAGAAAC TCAACGAGCT

      2770        2780        2790        2800        2810        2820
GGACGCGGAT GAACAGGCAG ACATCTGTGA ATCGCTTCAC GACCACGCTG ATGAGCTTTA
                                    ASP700I
                                    XMNI
```

```
      2830        2840        2850        2860        2870        2880
CCGCAGCTGC CTCGCGCGTT TCGGTGATGA CGGTGAAAAC CTCTGACACA TGCAGCTCCC
     PVUII

      2890        2900        2910        2920        2930        2940
GGAGACGGTC ACAGCTTGTC TGTAAGCGGA TGCCGGGAGC AGACAAGCCC GTCAGGGCGC
                                                 DRDI

      2950        2960        2970        2980        2990        3000
GTCAGCGGGT GTTGGCGGGT GTCGGGGCGC AGCCATGACC CAGTCACGTA GCGATAGCGG
                                             TTH111I

      3010        3020        3030        3040        3050        3060
AGTGTATACT GGCTTAACTA TGCGGCATCA GAGCAGATTG TACTGAGAGT GCACCATATG
                                                   APALI
                                                         NDEI

      3070        3080        3090        3100        3110        3120
CGGTGTGAAA TACCGCACAG ATGCGTAAGG AGAAAATACC GCATCAGGCG CTCTTCCGCT
                                                              KSP632I
                                                              EARI

      3130        3140        3150        3160        3170        3180
TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GGCGAGCGGT ATCAGCTCAC

      3190        3200        3210        3220        3230        3240
TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA ACGCAGGAAA GAACATGTGA

      3250        3260        3270        3280        3290        3300
GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT

      3310        3320        3330        3340        3350        3360
AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC
                                             DRDI

      3370        3380        3390        3400        3410        3420
CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT GCGCTCTCCT

      3430        3440        3450        3460        3470        3480
GTTCCGACCC TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG

      3490        3500        3510        3520        3530        3540
CTTTCTCATA GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG

      3550        3560        3570        3580        3590        3600
GGCTGTGTGC ACGAACCCCC CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT
           APALI

      3610        3620        3630        3640        3650        3660
CTTGAGTCCA ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG
                                                 ALWNI

      3670        3680        3690        3700        3710        3720
ATTAGCAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC

      3730        3740        3750        3760        3770        3780
GGCTACACTA GAAGGACAGT ATTTGGTATC TGCGCTCTGC TGAAGCCAGT TACCTTCGGA

      3790        3800        3810        3820        3830        3840
AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGGTTTTTTT

      3850        3860        3870        3880        3890        3900
GTTTGCAAGC AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT

      3910        3920        3930        3940        3950        3960
TCTACGGGGT CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCATGAGA
                                                         RSPXI
                                                         BSPHI

      3970        3980        3990        4000        4010        4020
TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT TAAATCAATC
                                  DRAI                  DRAI
```

```
      4030        4040        4050        4060        4070        4080
TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT GCTTAATCAG TGAGGCACCT

      4090        4100        4110        4120        4130        4140
ATCTCAGCGA TCTGTCTATT TCGTTCATCC ATAGTTGCCT GACTCCCCGT CGTGTAGATA

      4150        4160        4170        4180        4190        4200
ACTACGATAC GGGAGGGCTT ACCATCTGGC CCCAGTGCTG CAATGATACC GCGAGACCCA

      4210        4220        4230        4240        4250        4260
CGCTCACCGG CTCCAGATTT ATCAGCAATA AACCAGCCAG CCGGAAGGGC CGAGCGCAGA
                                                  BGLI

      4270        4280        4290        4300        4310        4320
AGTGGTCCTG CAACTTTATC CGCCTCCATC CAGTCTATTA ATTGTTGCCG GGAAGCTAGA
                                            ASEI

      4330        4340        4350        4360        4370        4380
GTAAGTAGTT CGCCAGTTAA TAGTTTGCGC AACGTTGTTG CCATTGCTGC AGGCATCGTG
                          AVIII                       PSTI
                          FSPI
                          AOSI

      4390        4400        4410        4420        4430        4440
GTGTCACGCT CGTCGTTTGG TATGGCTTCA TTCAGCTCCG GTTCCCAACG ATCAAGGCGA

      4450        4460        4470        4480        4490        4500
GTTACATGAT CCCCCATGTT GTGCAAAAAA GCGGTTAGCT CCTTCGGTCC TCCGATCGTT
                                                          XORII
                                                          PVUI

      4510        4520        4530        4540        4550        4560
GTCAGAAGTA AGTTGGCCGC AGTGTTATCA CTCATGGTTA TGGCAGCACT GCATAATTCT

      4570        4580        4590        4600        4610        4620
CTTACTGTCA TGCCATCCGT AAGATGCTTT TCTGTGACTG GTGAGTACTC AACCAAGTCA
                                              SCAI

      4630        4640        4650        4660        4670        4680
TTCTGAGAAT AGTGTATGCG GCGACCGAGT TGCTCTTGCC CGGCGTCAAC ACGGGATAAT

      4690        4700        4710        4720        4730        4740
ACCGCGCCAC ATAGCAGAAC TTTAAAAGTG CTCATCATTG GAAAACGTTC TTCGGGGCGA
                      DRAI                    ASP700I
                                              XMNI

      4750        4760        4770        4780        4790        4800
AAACTCTCAA GGATCTTACC GCTGTTGAGA TCCAGTTCGA TGTAACCCAC TCGTGCACCC
                                                              APALI

      4810        4820        4830        4840        4850        4860
AACTGATCTT CAGCATCTTT TACTTTCACC AGCGTTTCTG GGTGAGCAAA AACAGGAAGG

      4870        4880        4890        4900        4910        4920
CAAAATGCCG CAAAAAAGGG AATAAGGGCG ACACGGAAAT GTTGAATACT CATACTCTTC

      4930        4940        4950        4960        4970        4980
CTTTTTCAAT ATTATTGAAG CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT
   KSP632I                                            RSPXI
   EARI  SSPI                                         BSPHI

      4990        5000        5010        5020        5030        5040
GAATGTATTT AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG AAAAGTGCCA

      5050        5060        5070        5080        5090        5100
CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG
       AATII                    RSPXI
                                BSPHI

      5110
AGGCCCTTTC GTCTTCAA
```

aatII(GACGTC): 5044
accIII(TCCGGA): 2422
aflII(CTTAAG): 703
ahaIII(TTTAAA): 440 3990 4009 4701
alwNI(CAGNNNCTG): 3644
apaI(GGGCCC): 708
apaLI(GTGCAC): 69 3049 3547 4793
aseI(ATTAAT): 3 241 4297
balI(TGGCCA): 2202
bamHI(GGATCC): 1133
bglI(GCCNNNNNGGC): 1687 1921 4240
bglII(AGATCT): 385
bsmI(GAATGC): 320 2111
bspHI(TCATGA): 1247 3953 4961 5066
bspMI(ACCTGC): 123 1821
bspMII(TCCGGA): 2422
claI(ATCGAT): 781
draI(TTTAAA): 440 3990 4009 4701
eagI(CGGCCG): 1697
earI(CTCTTC): 3111 4915
ecoNI(CCTNNNNNAGG): 1380
ecoRI(GAATTC): 717
ecoRV(GATATC): 943
fspI(TGCGCA): 323 1018 2114 2212 4346
hindIII(AAGCTT): 787
hpaI(GTTAAC): 257 756
mluI(ACGCGT): 373
naeI(GCCGGC): 1159 1527 1687 2041
narI(GGCGCC): 506 1171 1192 1306 1963
ndeI(CATATG): 3055
nheI(GCTAGC): 987
nruI(TCGCGA): 1730
pflMI(CCANNNNNTGG): 2073 2122
pstI(CTGCAG): 4367
pvuI(CGATCG): 4493
pvuII(CAGCTG): 2824
salI(GTCGAC): 1409
scaI(AGTACT): 4604
smaI(CCCGGG): 712
speI(ACTAGT): 253 261
sphI(GCATGC): 1320
sspI(AATATT): 216 4928
stuI(AGGCCT): 531
styI(CCWWGG): 604 2127
tth111I(GACNNNGTC): 2977
xbaI(TCTAGA): 291
xmaI(CCCGGG): 712
xmaIII(CGGCCG): 1697
xmnI(GAANNNNTTC): 1 2789 4721

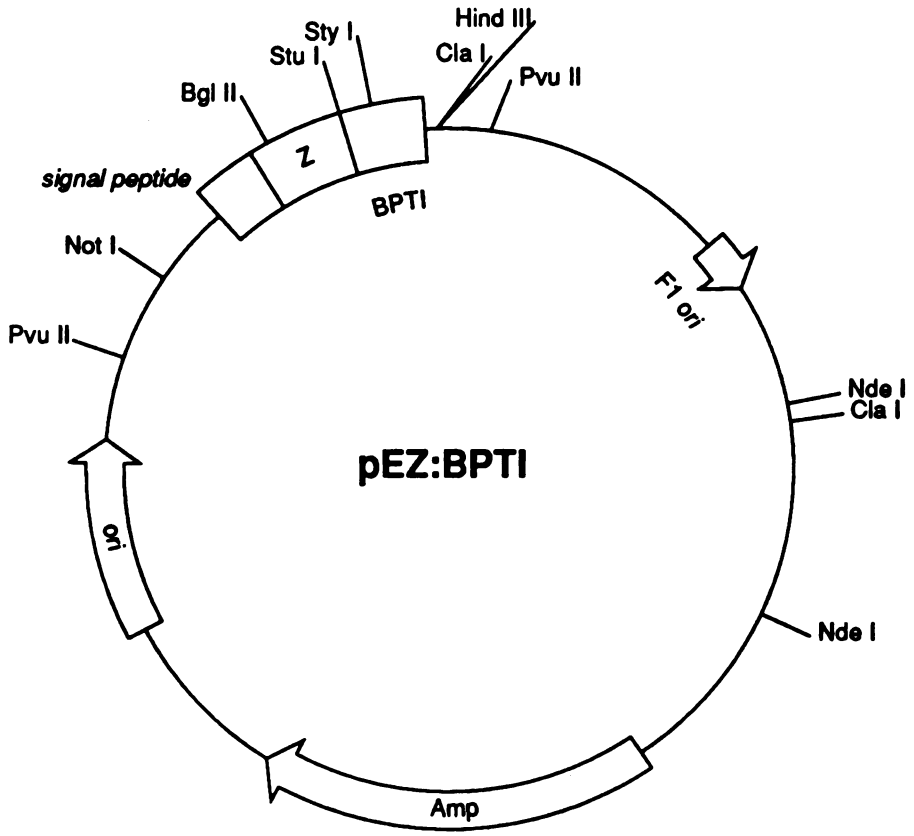


Figure 2B.2: The pEZ:BPTI expression plasmid

MAP restriction endonuclease sites:
Cut positions given are the last base of the fragment on the 5'-side (left side) of the cut.
The signal sequence and protein linker sequences are in normal font. *The Z domain is in italic type.* The BPTI domain is underlined.

The following enzymes were examined:

AATI	AATII	ACCIII	AFLII	ALWNI	AOSI
APAI	APALI	ASEI	ASP700I	ASP718I	ASUII
AVIII	AVRII	AXYI	BALI	BAMHI	BBEI
BBUI	BCLI	BGLI	BGLII	BSMI	BSPHI
BSPMI	BSPMII	BSSHII	BSTBI	BSTEII	BSTI
BSTXI	BSU36I	CLAI	CSP45I	CVNI	DRAI
DRAIII	DRDI	EAGI	EARI	ECO47III	ECO52I
ECO81I	ECONI	ECOO65I	ECORI	ECORV	ECOT22I
ESPI	FSPI	HINDIII	HPAI	KPNI	KSP632I
MLUI	MROI	NAEI	NARI	NCOI	NDEI
NHEI	NOTI	NRUI	NSII	NUNII	PAER7I
PFLMI	PSTI	PVUI	PVUII	RSPXI	RSRII
SACI	SACII	SALI	SAUI	SCAI	SFII
SMAI	SNABI	SPEI	SPHI	SPLI	SPOI
SSPI	SSTI	SSTII	STUI	STYI	TTH111I
XBAI	XHOI	XMAI	XMAIII	XMNI	XORII

C1 (1f): |>u 1>+++++ ss (4646 bases)+++++>u 4646>|

```
        10          20          30          40          50          60
AGCTTGGCAC TGGCCGTCGT TTTACAACGT ACTGACTGGG AAAACCCTGG CGTTACCCAA

        70          80          90         100         110         120
CTTAATCGCC TTGCAGCACA TCCCCCCTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC
                                    PVUII       KSP632I
                                                EARI

       130         140         150         160         170         180
ACCGATCGCC CTTCCCAACA GTTGCGTAGC CTGAATGGCG AATAATTCCA GACGATTGAG
     XORII                          BGLI
     PVUI

       190         200         210         220         230         240
CGTCAAAATG TAGGTATTTC CATGAGCGTT TTTCCTGTTG CAATGGCTGG CGGTAATATT
                                                                SSPI

       250         260         270         280         290         300
GTTCTGGATA TTACCAGCAA GGCCGATAGT TTGAGTTCTT CTACTCAGGC AAGTGATGTT

       310         320         330         340         350         360
ATTACTAATC AAAGAAGTAT TGCGACAACG GTTAATTTGC GTGATGGACA GACTCTTTTA

       370         380         390         400         410         420
CTCGGTGGCC TCACTGATTA TAAAAACACT TCTCAGGATT CTGGCGTACC GTTCCTGTCT

       430         440         450         460         470         480
AAAATCCCTT TAATCGGCCT CCTGTTTAGC TCCCGCTCTG ATTCTAACGA GGAAAGCACG

       490         500         510         520         530         540
TTATACGTGC TCGTCAAAGC AACCATAGTA CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC

       550         560         570         580         590         600
GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC AGCGCCCTAG CGCCCGCTCC

       610         620         630         640         650         660
TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC AAGCTCTAAA
                                    NAEI

       670         680         690         700         710         720
TCGGGGGCTC CCTTTAGGGT TCCGATTTAG TGCTTTACGG CACCTCGACC CCAAAAAACT

       730         740         750         760         770         780
TGATTAGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT
                      DRAIII

       790         800         810         820         830         840
GACGTTGGAG TCCACGTTCT TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA
     DRDI

       850         860         870         880         890         900
CCCTATCTCG GTCTATTCTT TTGATTTATA AGGGATTTTG CCGATTTCGG CCTATTGGTT

       910         920         930         940         950         960
AAAAAATGAG CTGATTTAAC AAAAATTTAA CGCGAATTTT AACAAAATAT TAACGTTTAC
                                                    SSPI

       970         980         990        1000        1010        1020
AATTTAAATA TTTGCTTATA CAATCTTCCT GTTTTTGGGG CTTTTCTGAT TATCAACCGG
    DRAI
        SSPI

      1030        1040        1050        1060        1070        1080
GGTACATATG ATTGACATGC TAGTTTTACG ATTACCGTTC ATCGATTCTC TTGTTTGCTC
       NDEI                                    CLAI

      1090        1100        1110        1120        1130        1140
CAGACTCTCA GGCAATGACC TGATAGCCTT TGTAGACCTC TCAAAAATAG CTACCCTCTC

      1150        1160        1170        1180        1190        1200
CGGCATGAAT TTATCAGCTA GAACGGTTGA ATATCATATT GATGGTGATT TGACTGTCTC
```

```
      1210        1220        1230        1240        1250        1260
CGGCCTTTCT CACCCGTTTG AATCTTTACC TACACATTAC TCAGGCATTG CATTTAAAAT
                                                          DRAI

      1270        1280        1290        1300        1310        1320
ATATGAGGGT TCTAAAAATT TTTATCCTTG CGTTGAAATA AAGGCTTCTC CCGCAAAAGT

      1330        1340        1350        1360        1370        1380
ATTACAGGGT CATAATGTTT TTGGTACAAC CGATTTAGCT TTATGCTCTG AGGCTTTATT

      1390        1400        1410        1420        1430        1440
GCTTAATTTT GCTAATTCTT TGCCTTGCCT GTATGATTTA TTGGATGTTG GAATTTGATG

      1450        1460        1470        1480        1490        1500
CGGTATTTTC TCCTTACGCA TCTGTGCGGT ATTTCACACC GCATATGGTG CACTCTCAGT
                                                NDEI    APALI

      1510        1520        1530        1540        1550        1560
ACAATCTGCT CTGATGCCGC ATAGTTAAGC CAGCCCCGAC ACCCGCCAAC ACCCGCTGAC

      1570        1580        1590        1600        1610        1620
GCGCCCTGAC GGGCTTGTCT GCTCCCGGCA TCCGCTTACA GACAAGCTGT GACCGTCTCC
             DRDI

      1630        1640        1650        1660        1670        1680
GGGAGCTGCA TGTGTCAGAG GTTTTCACCG TCATCACCGA AACGCGCGAG GCAGCTTGAA

      1690        1700        1710        1720        1730        1740
GACGAAAGGG CCTCGTGATA CGCCTATTTT TATAGGTTAA TGTCATGATA ATAATGGTTT
                                                RSPXI
                                                BSPHI

      1750        1760        1770        1780        1790        1800
CTTAGACGTC AGGTGGCACT TTTCGGGGAA ATGTGCGCGG AACCCCTATT TGTTTATTTT
       AATII

      1810        1820        1830        1840        1850        1860
TCTAAATACA TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA ATGCTTCAAT
                                  RSPXI
                                  BSPHI

      1870        1880        1890        1900        1910        1920
AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT
  SSPI       KSP632I
             EARI

      1930        1940        1950        1960        1970        1980
TTGCGGCATT TTGCCTTCCT GTTTTTGCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG

      1990        2000        2010        2020        2030        2040
CTGAAGATCA GTTGGGTGCA CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA
                    APALI

      2050        2060        2070        2080        2090        2100
TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC
                            ASP700I                       DRAI
                            XMNI

      2110        2120        2130        2140        2150        2160
TATGTGGCGC GGTATTATCC CGTGTTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC

      2170        2180        2190        2200        2210        2220
ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAAGCAT CTTACGGATG
                            SCAI

      2230        2240        2250        2260        2270        2280
GCATGACAGT AAGAGAATTA TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA
```

```
      2290        2300        2310        2320        2330        2340
ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG
                   XORII
                   PVUI

      2350        2360        2370        2380        2390        2400
GGGATCATGT AACTCGCCTT GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG

      2410        2420        2430        2440        2450        2460
ACGAGCGTGA CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGCGCAAA CTATTAACTG
                                                AVIII
                                                FSPI
                                                AOSI

      2470        2480        2490        2500        2510        2520
GCGAACTACT TACTCTAGCT TCCCGGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG
                                    ASEI

      2530        2540        2550        2560        2570        2580
TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTTATTGCT GATAAATCTG
                               BGLI

      2590        2600        2610        2620        2630        2640
GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT GGGGCCAGAT GGTAAGCCCT

      2650        2660        2670        2680        2690        2700
CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA CGAAATAGAC

      2710        2720        2730        2740        2750        2760
AGATCGCTGA GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT

      2770        2780        2790        2800        2810        2820
CATATATACT TTAGATTGAT TTAAAACTTC ATTTTTAATT TAAAAGGATC TAGGTGAAGA
                      DRAI                        DRAI

      2830        2840        2850        2860        2870        2880
TCCTTTTTGA TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT
                   RSPXI
                   BSPHI

      2890        2900        2910        2920        2930        2940
CAGACCCCGT AGAAAAGATC AAAGGATCTT CTTGAGATCC TTTTTTTCTG CGCGTAATCT

      2950        2960        2970        2980        2990        3000
GCTGCTTGCA AACAAAAAAA CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC

      3010        3020        3030        3040        3050        3060
TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC GCAGATACCA AATACTGTCC

      3070        3080        3090        3100        3110        3120
TTCTAGTGTA GCCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC

      3130        3140        3150        3160        3170        3180
TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCG TGTCTTACCG
                               ALWNI

      3190        3200        3210        3220        3230        3240
GGTTGGACTC AAGACGATAG TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT

      3250        3260        3270        3280        3290        3300
CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG
     APALI

      3310        3320        3330        3340        3350        3360
AGCATTGAGA AAGCGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG

      3370        3380        3390        3400        3410        3420
GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG GGGAAACGCC TGGTATCTTT

      3430        3440        3450        3460        3470        3480
ATAGTCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG
                                    DRDI
```

```
      3490        3500        3510        3520        3530        3540
GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT TTTACGGTTC CTGGCCTTTT

      3550        3560        3570        3580        3590        3600
GCTGGCCTTT TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA

      3610        3620        3630        3640        3650        3660
TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT

      3670        3680        3690        3700        3710        3720
CAGTGAGCGA GGAAGCGGAA GAGCGCCCAA TACGCAAACC GCCTCTCCCC GCGCGTTGGC
                KSP632I
                EARI

      3730        3740        3750        3760        3770        3780
CGATTCATTA ATCCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTGAGCGCA
         ASEI   PVUII

      3790        3800        3810        3820        3830        3840
ACGCAATTAA TGTGAGTTAC CTCACTCATT AGGCACCCCA GGCTTTACAC TTTATGCTTC
       ASEI

      3850        3860        3870        3880        3890        3900
CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT TCACACAGGA AACAGCTATG

      3910        3920        3930        3940        3950        3960
ACCATGATTA CGAATTAGCG GCCGCTCGAA ATAGCGTGAT TTTGCGGTTT TAAGCCTTTT
                      XMAIII
                      NOTI
                      ECO52I
                      EAGI

      3970        3980        3990        4000        4010        4020
ACTTCCTGAA TAAATCTTTC AGCAAAATAT TTATTTTATA AGTTGTAAAA CTTACCTTTA
                           SSPI                                  DRAI

      4030        4040        4050        4060        4070        4080
AATTTAATTA TAAATATAGA TTTTAGTATT GCAATACATA ATTCGTTATA TTATGATGAC

      4090        4100        4110        4120        4130        4140
                              Le uLysLysLys AsnIleTyrS erIleArgLys
TTTACAAATA CATACAGGGG GTATTAATTT GAAAAAGAAA AACATTTATT CAATTCGTAA
                           ASEI

      4150        4160        4170        4180        4190        4200
 LeuGlyVal GlyIleAlaS erValThrLe uGlyThrLeu LeuIleSerG lyGlyValThr
ACTAGGTGTA GGTATTGCAT CTGTAACTTT AGGTACATTA CTTATATCTG GTGGCGTAAC

      4210        4220        4230        4240        4250        4260
 ProAlaAla AsnAlaAlaG lnHisAspGl uAla*ValAsp* *AsnLysPheA* *snLysGluGln*
ACCTGCTGCA AATGCTGCGC AACACGATGA AGCCGTAGAC AACAAATTCA ACAAAGAACA
             BSPMI
                 AVIII
                 FSPI
                 AOSI

      4270        4280        4290        4300        4310        4320
 *GlnAsnAla* *PheTyrGluI* *leLeuHisLe* *uProAsnLeu* *AsnGluGluG* *lnArgAsnAla*
ACAAAACGCG TTCTATGAGA TCTTACATTT ACCTAACTTA AACGAAGAAC AACGAAACGC
         MLUI           BGLII

      4330        4340        4350        4360        4370        4380
 *PheIleGln* *SerLeuLysA* *spAspProSe* *rGlnSerAla* *AsnLeuLeuA* *laGluAlaLys*
CTTCATCCAA AGTTTAAAAG ATGACCCAAG CCAAAGCGCT AACCTTTTAG CAGAAGCTAA
             DRAI                      ECO47III
```

```
      4390        4400        4410        4420        4430        4440
  LysLeuAsn AspAlaGlnA laProLysVa lAspMetAsn GlyArgProA spPheCysLeu
AAAGCTAAAT GATGCTCAGG CGCCGAAAGT AGACATGAAT GGGAGGCCTG ACTTCTGCCT
                       NUNII                        STUI
                       NARI                         AATI
                        BBEI

      4450        4460        4470        4480        4490        4500
 GluProPro TyrThrGlyP roAlaLysAl aArgIleIle ArgTyrPheT yrAsnAlaLys
AGAGCCTCCA TATACGGGTC CCGCTAAGGC CAGAATTATC AGATACTTCT ACAACGCCAA
                                                              STYI

      4510        4520        4530        4540        4550        4560
 AlaGlyLeu CysGlnThrP heValTyrGl yGlyAlaArg AlaLysArgA snAsnPheLys
GGCTGGGCTC TGCCAGACCT TTGTATATGG TGGCGCTAGA GCTAAGAGAA ACAATTTCAA

      4570        4580        4590        4600        4610        4620
 SerAlaGlu AspCysMetA rgThrCysGl yGlyAla***
GAGCGCAGAG GACTGCATGA GGACCTGTGG TGGTGCTTAA GGGCCCCGGG AATTCTCATG
                                          AFLII APAI      ECORI
                                                SMAI
                                                  XMAI

      4630        4640
TTTGACAGCT TATCATCGAT AAGCTT
                  CLAI    HINDIII
```

AATII(GACGTC):	1745	FSPI(TGCGCA):	2443 4216
AFLII(CTTAAG):	4596	HINDIII(AAGCTT):	4641
ALWNI(CAGNNNCTG):	3142	MLUI(ACGCGT):	4266
APAI(GGGCCC):	4601	MSTI(TGCGCA):	2443 4216
APALI(GTGCAC):	1488 1996 3242	NAEI(GCCGGC):	635
ASEI(ATTAAT):	2492 3727 3786 4103	NARI(GGCGCC):	4399
		NDEI(CATATG):	1025 1482
BGLI(GCCNNNNNGGC):	149 2544	NOTI(GCGGCCGC):	3918
BGLII(AGATCT):	4278	PVUI(CGATCG):	123 2296
BSPHI(TCATGA):	1723 1828 2836	PVUII(CAGCTG):	93 3734
BSPMI(ACCTGC):	4201	SCAI(AGTACT):	2185
CLAI(ATCGAT):	1061 4635	SMAI(CCCGGG):	4605
DRAI(TTTAAA):	963 1253 2088 2780 2799 4017 4333	SSPI(AATATT):	235 946 967 1861 3986
		STUI(AGGCCT):	4424
DRAIII(CACNNNGTG):	738	STYI(CCWWGG):	4497
DRDI(GACNNNNNNGTC):	781 1568 3448	XMAI(CCCGGG):	4605
EAGI(CGGCCG):	3919	XMAIII(CGGCCG):	3919
EARI(CTCTTC):	109 1874 3678	XMNI(GAANNNNTTC):	2064
ECORI(GAATTC):	4610		
ESPI(GCTNAGC):	4643		

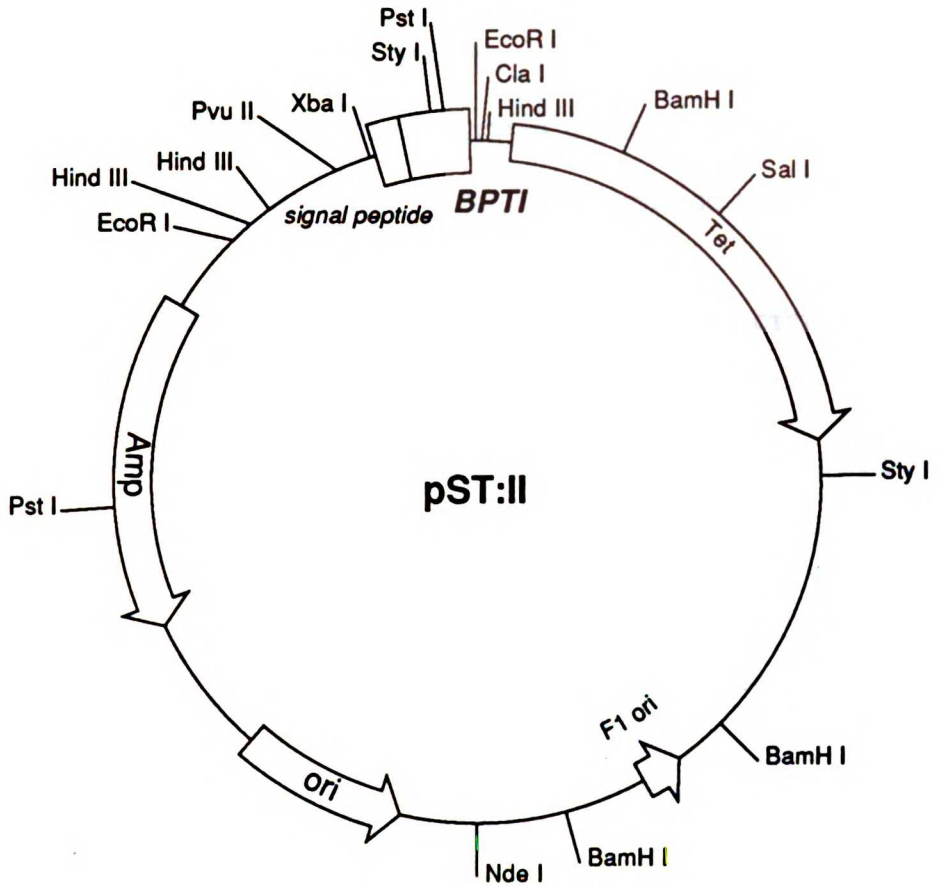


Figure 2B.3: The pST:II expression plasmid (the wild type plasmid is also called p657).

MAP restriction endonuclease sites:
Cut positions given are the last base of the fragment on the 5'-side (left side) of the cut.
The ST:II signal sequence is in italic type. The BPTI domain is underlined.

The following enzymes were examined:

AATI	AATII	ACCIII	AFLII	ALWNI	AOSI
APAI	APALI	ASEI	ASP700I	ASP718I	ASUII
AVIII	AVRII	AXYI	BALI	BAMHI	BBEI
BBUI	BCLI	BGLI	BGLII	BSMI	BSPHI
BSPMI	BSPMII	BSSHII	BSTBI	BSTEII	BSTI
BSTXI	BSU36I	CLAI	CSP45I	CVNI	DRAI
DRAIII	DRDI	EAGI	EARI	ECO47III	ECO52I
ECO81I	ECONI	ECOO65I	ECORI	ECORV	ECOT22I
ESPI	FSPI	HINDIII	HPAI	KPNI	KSP632I
MLUI	MROI	NAEI	NARI	NCOI	NDEI
NHEI	NOTI	NRUI	NSII	NUNII	PAER7I
PFLMI	PSTI	PVUI	PVUII	RSPXI	RSRII
SACI	SACII	SALI	SAUI	SCAI	SFII
SMAI	SNABI	SPEI	SPHI	SPLI	SPOI
SSPI	SSTI	SSTII	STUI	STYI	TTH111I
XBAI	XHOI	XMAI	XMAIII	XMNI	XORII

C1 (1f): |>u 1>+++++ ss (5525 bases)+++++>u 5525>|

```
        10          20          30          40          50          60
TTCTCATGTT TGACAGCTTA TCATCGATAA GCTTTAATGC GGTAGTTTAT CACAGTTAAA
 ECORI                     CLAI    HINDIII

        70          80          90         100         110         120
TTGCTAACGC AGTCAGGCAC CGTGTATGAA ATCTAACAAT GCGCTCATCG TCATCCTCGG

       130         140         150         160         170         180
CACCGTCACC CTGGATGCTG TAGGCATAGG CTTGGTTATG CCGGTACTGC CGGGCCTCTT

       190         200         210         220         230         240
GCGGGATATC GTCCATTCCG ACAGCATCGC CAGTCACTAT GGCGTGCTGC TAGCGCTATA
      ECORV                                                 NHEI
                                                             ECO47III

       250         260         270         280         290         300
TGCGTTGATG CAATTTCTAT GCGCACCCGT TCTCGGAGCA CTGTCCGACC GCTTTGGCCG
                       AVIII
                       FSPI
                       AOSI

       310         320         330         340         350         360
CCGCCCAGTC CTGCTCGCTT CGCTACTTGG AGCCACTATC GACTACGCGA TCATGGCGAC

       370         380         390         400         410         420
CACACCCGTC CTGTGGATCC TCTACGCCGG ACGCATCGTG GCCGGCATCA CCGGCGCCAC
                  BAMHI                       NAEI          NUNII
                  BSTI                                      NARI
                                                             BBEI

       430         440         450         460         470         480
AGGTGCGGTT GCTGGCGCCT ATATCGCCGA CATCACCGAT GGGGAAGATC GGGCTCGCCA
                NUNII
                NARI
                 BBEI

       490         500         510         520         530         540
CTTCGGGCTC ATGAGCGCTT GTTTCGGCGT GGGTATGGTG GCAGGCCCCG TGGCCGGGGG
            RSPXI
            BSPHI
               ECO47III

       550         560         570         580         590         600
ACTGTTGGGC GCCATCTCCT TGCATGCACC ATTCCTTGCG GCGGCGGTGC TCAACGGCCT
          NUNII           SPHI
          NARI            BBUI
           BBEI

       610         620         630         640         650         660
CAACCTACTA CTGGGCTGCT TCCTAATGCA GGAGTCGCAT AAGGGAGAGC GTCGACCGAT
                          ECONI                           SALI

       670         680         690         700         710         720
GCCCTTGAGA GCCTTCAACC CAGTCAGCTC CTTCCGGTGG GCGCGGGGCA TGACTATCGT

       730         740         750         760         770         780
CGCCGCACTT ATGACTGTCT TCTTTATCAT GCAACTCGTA GGACAGGTGC CGGCAGCGCT
                                                      NAEI  ECO47III

       790         800         810         820         830         840
CTGGGTCATT TTCGGCGAGG ACCGCTTTCG CTGGAGCGCG ACGATGATCG GCCTGTCGCT

       850         860         870         880         890         900
TGCGGTATTC GGAATCTTGC ACGCCCTCGC TCAAGCCTTC GTCACTGGTC CCGCCACCAA

       910         920         930         940         950         960
ACGTTTCGGC GAGAAGCAGG CCATTATCGC CGGCATGGCG GCCGACGCGC TGGGCTACGT
                                  NAEI          XMAIII
                                      BGLI      ECO52I
                                                EAGI
```

```
       970         980         990        1000        1010        1020
CTTGCTGGCG TTCGCGACGC GAGGCTGGAT GGCCTTCCCC ATTATGATTC TTCTCGCTTC
              SPOI
              NRUI

      1030        1040        1050        1060        1070        1080
CGGCGGCATC GGGATGCCCG CGTTGCAGGC CATGCTGTCC AGGCAGGTAG ATGACGACCA
                                      BSPMI

      1090        1100        1110        1120        1130        1140
TCAGGGACAG CTTCAAGGAT CGCTCGCGGC TCTTACCAGC CTAACTTCGA TCACTGGACC

      1150        1160        1170        1180        1190        1200
GCTGATCGTC ACGGCGATTT ATGCCGCCTC GGCGAGCACA TGGAACGGGT TGGCATGGAT
                           BGLI

      1210        1220        1230        1240        1250        1260
TGTAGGCGCC GCCCTATACC TTGTCTGCCT CCCCGCGTTG CGTCGCGGTG CATGGAGCCG
       NUNII
       NARI
        BBEI

      1270        1280        1290        1300        1310        1320
GGCCACCTCG ACCTGAATGG AAGCCGGCGG CACCTCGCTA ACGGATTCAC CACTCCAAGA
                           NAEI

      1330        1340        1350        1360        1370        1380
ATTGGAGCCA ATCAATTCTT GCGGAGAACT GTGAATGCGC AAACCAACCC TTGGCAGAAC
PFLMI                                    AVIII         PFLMI
                                         FSPI          STYI
                                         AOSI
                                          BSMI

      1390        1400        1410        1420        1430        1440
ATATCCATCG CGTCCGCCAT CTCCAGCAGC CGCACGCGGC GCATCTCGGG CAGCGTTGGG

      1450        1460        1470        1480        1490        1500
TCCTGGCCAC GGGTGCGCAT GATCGTGCTC CTGTCGTTGA GGACCCGGCT AGGCTGGCGG
     BALI        AVIII
                 FSPI
                 AOSI

      1510        1520        1530        1540        1550        1560
GGTTGCCTTA CTGGTTAGCA GAATGAATCA CCGATACGCG AGCGAACGTG AAGCGACTGC

      1570        1580        1590        1600        1610        1620
TGCTGCAAAA CGTCTGCGAC CTGAGCAACA ACATGAATGG TCTTCGGTTT CCGTGTTTCG

      1630        1640        1650        1660        1670        1680
TAAAGTCTGG AAACGCGGAA GTCAGCGCCC TGCACCATTA TGTTCCGGAT CTGCATCGCA
                                                 BSPMII
                                                 MROI
                                                 ACCIII

      1690        1700        1710        1720        1730        1740
GGATGCTGCT GGCTACCCTG TGGAACACCT ACATCTGTAT TAACGAAGCG CTGGCATTGA
                                                   ECO47III

      1750        1760        1770        1780        1790        1800
CCCTGAGTGA TTTTTCTCTG GTCCCGCCGC ATCCATACCG CCAGTTGTTT ACCCTCACAA

      1810        1820        1830        1840        1850        1860
CGTTCCAGTA ACCGGGCATG TTCATCATCA GTAACCCGTA TCGTGAGCAT CCTCTCTCGT

      1870        1880        1890        1900        1910        1920
TTCATCGGTA TCATTACCCC CATGAACAGA AATTCCCCCT TACACGGAGG CATCAAGTGA

      1930        1940        1950        1960        1970        1980
CCAAACAGGA AAAAACCGCC CTTAACATGG CCCGCTTTAT CAGAAGCCAG ACATTAACGC
```

```
      1990        2000        2010        2020        2030        2040
TTCTGGAGAA ACTCAACGAG CTGGACGCGG ATGAACAGGC AGACATCTGT GAATCGCTTC
                                                          ASP700I
                                                          XMNI

      2050        2060        2070        2080        2090        2100
ACGACCACGC TGATGAGCTT TACCGCAGGA TCCGGACGCG CCCTGTAGCG GCGCATTAAG
                                  BAMHI
                                  BSTI
                                     ACCIII
                                     MROI
                                     BSPMII

      2110        2120        2130        2140        2150        2160
CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG CCCTAGCGCC

      2170        2180        2190        2200        2210        2220
CGCTCCTTTC GCTTTCTTCC CTTCCTTTCT CGCCACGTTC GCCGGCTTTC CCCGTCAAGC
                                                 NAEI

      2230        2240        2250        2260        2270        2280
TCTAAATCGG GGGCTCCCTT TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA

      2290        2300        2310        2320        2330        2340
AAAACTTGAT TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTTCG
                               DRAIII

      2350        2360        2370        2380        2390        2400
CCCTTTGACG TTGGAGTCCA CGTTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC
             DRDI

      2410        2420        2430        2440        2450        2460
ACTCAACCCT ATCTCGGTCT ATTCTTTTGA TTTATAAGGG ATTTTGCCGA TTTCGGCCTA

      2470        2480        2490        2500        2510        2520
TTGGTTAAAA AATGAGCTGA TTTAACAAAA ATTTAACGCG AATTTTAACA AAATATTAAC
                                                          SSPI

      2530        2540        2550        2560        2570        2580
GTTTACAATT TCCGGATCCT GCCTCGCGCG TTTCGGTGAT GACGGTGAAA ACCTCTGACA
                BSPMII
                ACCIII
                MROI
                   BAMHI
                   BSTI

      2590        2600        2610        2620        2630        2640
CATGCAGCTC CCGGAGACGG TCACAGCTTG TCTGTAAGCG GATGCCGGGA GCAGACAAGC
                                                                DRDI

      2650        2660        2670        2680        2690        2700
CCGTCAGGGC GCGTCAGCGG GTGTTGGCGG GTGTCGGGGC GCAGCCATGA CCCAGTCACG
                                                       TTH111I

      2710        2720        2730        2740        2750        2760
TAGCGATAGC GGAGTGTATA CTGGCTTAAC TATGCGGCAT CAGAGCAGAT TGTACTGAGA

      2770        2780        2790        2800        2810        2820
GTGCACCATA TGCGGTGTGA AATACCGCAC AGATGCGTAA GGAGAAAATA CCGCATCAGG
    APALI
          NDEI

      2830        2840        2850        2860        2870        2880
CGCTCTTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT GCGGCGAGCG
             KSP632I
             EARI

      2890        2900        2910        2920        2930        2940
GTATCAGCTC ACTCAAAGGC GGTAATACGG TTATCCACAG AATCAGGGGA TAACGCAGGA

      2950        2960        2970        2980        2990        3000
AAGAACATGT GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC CGCGTTGCTG
```

```
      3010        3020        3030        3040        3050        3060
GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG CTCAAGTCAG
                                                        DRDI

      3070        3080        3090        3100        3110        3120
AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG AAGCTCCCTC

      3130        3140        3150        3160        3170        3180
GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCTT TCTCCCTTCG

      3190        3200        3210        3220        3230        3240
GGAAGCGTGG CGCTTTCTCA TAGCTCACGC TGTAGGTATC TCAGTTCGGT GTAGGTCGTT

      3250        3260        3270        3280        3290        3300
CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTTCAGC CCGACCGCTG CGCCTTATCC
                          APALI

      3310        3320        3330        3340        3350        3360
GGTAACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT GGCAGCAGCC

      3370        3380        3390        3400        3410        3420
ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG CTACAGAGTT CTTGAAGTGG
ALWNI

      3430        3440        3450        3460        3470        3480
TGGCCTAACT ACGGCTACAC TAGAAGGACA GTATTTGGTA TCTGCGCTCT GCTGAAGCCA

      3490        3500        3510        3520        3530        3540
GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC CGCTGGTAGC

      3550        3560        3570        3580        3590        3600
GGTGGTTTTT TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC TCAAGAAGAT

      3610        3620        3630        3640        3650        3660
CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG TTAAGGGATT

      3670        3680        3690        3700        3710        3720
TTGGTCATGA GATTATCAAA AAGGATCTTC ACCTAGATCC TTTTAAATTA AAAATGAAGT
        RSPXI                                       DRAI
        BSPHI

      3730        3740        3750        3760        3770        3780
TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTACCA ATGCTTAATC
  DRAI

      3790        3800        3810        3820        3830        3840
AGTGAGGCAC CTATCTCAGC GATCTGTCTA TTTCGTTCAT CCATAGTTGC CTGACTCCCC

      3850        3860        3870        3880        3890        3900
GTCGTGTAGA TAACTACGAT ACGGGAGGGC TTACCATCTG GCCCCAGTGC TGCAATGATA

      3910        3920        3930        3940        3950        3960
CCGCGAGACC CACGCTCACC GGCTCCAGAT TTATCAGCAA TAAACCAGCC AGCCGGAAGG
                                                            BGLI

      3970        3980        3990        4000        4010        4020
GCCGAGCGCA GAAGTGGTCC TGCAACTTTA TCCGCCTCCA TCCAGTCTAT TAATTGTTGC
                                                      ASEI

      4030        4040        4050        4060        4070        4080
CGGGAAGCTA GAGTAAGTAG TTCGCCAGTT AATAGTTTGC GCAACGTTGT TGCCATTGCT
                                          AVIII
                                          FSPI
                                          AOSI

      4090        4100        4110        4120        4130        4140
GCAGGCATCG TGGTGTCACG CTCGTCGTTT GGTATGGCTT CATTCAGCTC CGGTTCCCAA
  PSTI

      4150        4160        4170        4180        4190        4200
CGATCAAGGC GAGTTACATG ATCCCCCATG TTGTGCAAAA AAGCGGTTAG CTCCTTCGGT
```

```
      4210        4220        4230        4240        4250        4260
CCTCCGATCG TTGTCAGAAG TAAGTTGGCC GCAGTGTTAT CACTCATGGT TATGGCAGCA
       XORII
       PVUI

      4270        4280        4290        4300        4310        4320
CTGCATAATT CTCTTACTGT CATGCCATCC GTAAGATGCT TTTCTGTGAC TGGTGAGTAC
                                                              SCAI

      4330        4340        4350        4360        4370        4380
TCAACCAAGT CATTCTGAGA ATAGTGTATG CGGCGACCGA GTTGCTCTTG CCCGGCGTCA

      4390        4400        4410        4420        4430        4440
ACACGGGATA ATACCGCGCC ACATAGCAGA ACTTTAAAAG TGCTCATCAT TGGAAAACGT
                                    DRAI                      ASP700I
                                                              XMNI

      4450        4460        4470        4480        4490        4500
TCTTCGGGGC GAAAACTCTC AAGGATCTTA CCGCTGTTGA GATCCAGTTC GATGTAACCC

      4510        4520        4530        4540        4550        4560
ACTCGTGCAC CCAACTGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC TGGGTGAGCA
        APALI

      4570        4580        4590        4600        4610        4620
AAAACAGGAA GGCAAAATGC CGCAAAAAAG GGAATAAGGG CGACACGGAA ATGTTGAATA

      4630        4640        4650        4660        4670        4680
CTCATACTCT TCCTTTTTCA ATATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC
               KSP632I                                        RSPXI
               EARI   SSPI                                    BSPHI

      4690        4700        4710        4720        4730        4740
GGATACATAT TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC

      4750        4760        4770        4780        4790        4800
CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC CTATAAAAAT
                   AATII                                RSPXI
                                                        BSPHI

      4810        4820        4830        4840        4850        4860
AGGCGTATCA CGAGGCCCTT TCGTCTTCAA GAATTCAACT TCTCCATACT TTGGATAAGG
                                    ECORI           PFLMI

      4870        4880        4890        4900        4910        4920
AAATACAGAC ATGAAAAATC TCATTGCTGA GTTGTTATTT AAGCTTGCCC AAAAAGAAGA
                                                HINDIII       KSP632I
                                                              EARI

      4930        4940        4950        4960        4970        4980
AGAGTCGAAA GAACTGTGTG CGCAGGTAGA AGCTTTGGAG ATTATCGTCA CTGTAATGCT
                  BSPMI                HINDIII
                       AVIII
                       FSPI
                       AOSI

      4990        5000        5010        5020        5030        5040
TCGCAATATG GCGCAAAATG ACCAACAGCG GTTGATTGAT CAGGTAGAGG GGGCGCTGTA
                                            BCLI

      5050        5060        5070        5080        5090        5100
CGAGGTAAAG CCCGATGCCA GCATTCCTGA CGACGATACG GAGCTGCTGC GCGATTACGT
                  BSMI                                         SNABI

      5110        5120        5130        5140        5150        5160
AAAGAAGTTA TTGAAGCATC CTCGTCAGTA AAAAGTTAAT CTTTTCAACA GCTGTCATAA
                                                        PVUII
```

```
      5170        5180        5190        5200        5210        5220
AGTTGTCACG GCCGAGACTT ATAGTCGCTT TGTTTTTATT TTTTAATGTA TTTGTAACTA
             XMAIII
             ECO52I
             EAGI

      5230        5240        5250        5260        5270        5280
                                                        Met LysLysAsnIle
GTACGCAAGT TCACGTAAAA AGGGTATCTA GAGGTTGAGG TGATTTTATG AAAAAGAATA
SPEI                             XBAI

      5290        5300        5310        5320        5330        5340
  AlaPheLe uLeuAlaSer MetPheValP heSerIleAl aThrAsnAla TyrAlaArgPro
TCGCATTTCT TCTTGCATCT ATGTTCGTTT TTTCTATTGC TACAAATGCC TATGCACGGC

      5350        5360        5370        5380        5390        5400
  AspPheCy sLeuGluPro ProTyrThrG lyProCysLy sAlaArgIle IleArgTyrPhe
CTGACTTCTG CCTAGAGCCT CCATATACGG GTCCCTGCAA GGCCAGAATT ATCAGATACT

      5410        5420        5430        5440        5450        5460
   TyrAsnAl aLysAlaGly LeuCysGlnT hrPheValTy rGlyGlyCys ArgAlaLysArg
TCTACAACGC CAAGGCTGGG CTCTGCCAGA CCTTTGTATA TGGTGGCTGC AGAGCTAAGA
          STYI                                          PSTI

      5470        5480        5490        5500        5510        5520
   AsnAsnPh eLysSerAla GluAspCysM etArgThrCy sGlyGlyAla***
GAAACAATTT CAAGAGCGCA GAGGACTGCA TGAGGACCTG TGGTGGTGCT TAAGGGCCCC
                                                      AFLII  SMAI
                                                             APAI

GGGAA
 XMAI
```

aatII(GACGTC):	4756	hindIII(AAGCTT):	29 4901 4950
aflII(CTTAAG):	5509	naeI(GCCGGC):	401 769 929
ahaIII(TTTAAA):	3702 3721 4413		1283 2201
alwNI(CAGNNNCTG):	3356	narI(GGCGCC):	413 434 548
apaI(GGGCCC):	5514		1205
apaLI(GTGCAC):	2761 3259 4505	ndeI(CATATG):	2767
aseI(ATTAAT):	4009	nheI(GCTAGC):	229
balI(TGGCCA):	1444[dcm-]	nruI(TCGCGA):	972
bamHI(GGATCC):	375 2068 2534	pflMI(CCANNNNNTGG):	1315 1364 4844
bclI[dam-](TGATCA):	5017	pstI(CTGCAG):	4079 5447
bglI(GCCNNNNNGGC):	929 1163 3952	pvuI(CGATCG):	4205
bsmI(GAATGC):	1353 5061	pvuII(CAGCTG):	5149
bspHI(TCATGA):	489 3665 4673	salI(GTCGAC):	651
	4778	scaI(AGTACT):	4316
bspMI(ACCTGC):	1063 4942	smaI(CCCGGG):	5518
bspMII(TCCGGA):	1664 2071 2531	snaBI(TACGTA):	5096
claI(ATCGAT):	23	speI(ACTAGT):	5217
draI(TTTAAA):	3702 3721 4413	sphI(GCATGC):	562
draIII(CACNNNGTG):	2304	sspI(AATATT):	2512 4640
ecoNI(CCTNNNNNAGG):	622	styI(CCWWGG):	1369 5410
ecoRI(GAATTC):	4831 5523	tth111I(GACNNNGTC):	2689
ecoRV(GATATC):	185	xbaI(TCTAGA):	5247
fspI(TGCGCA):	260 1356 1454	xmaIII(CGGCCG):	939 5169
	4058 4939	xmnI(GAANNNNTTC):	2031 4433

Appendix 2C

2C. Future expression system designs

The pHAZY vectors represent the first attempt of the Anderson and Kuntz laboratories at intracellular expression of BPTI. There are a number of areas in which the plasmids could be redesigned or improved upon. These range from the rather trivial, such as cleaning up the 3' end of the wild type gene in pEZ and pHAZY that I mentioned in the previous appendix, to the use of a completely different promoter. In most cases, the modifications will not make a large difference in the chance of success of a certain project, such as the production of additional mutants for protein folding studies. Enough protein is made from the pHAZY vectors to proceed with most NMR, crystallography, and folding projects. However, it may be desirable to optimize the levels of protein production for projects which include expensive isotope labeling with deuterium, ^{15}N, or ^{13}C. In addition, the introduction of a single strand origin into the pHAZY plasmids would eliminate the subcloning steps from any mutagenesis project. This is a matter of small inconvenience in the production of single mutants, but is quite time consuming for large scale mutagenesis projects. I will discuss a number of the higher priority modifications in the sections to follow.

2C.1. Turning the pHAZY plasmid into a phagemid

Most modern *E. coli* expression plasmids are actually phage/plasmid chimeras termed phagemids (Vieira and Messing, 1987). These allow protein expression, oligo-directed mutagenesis, and single strand DNA sequencing starting from a copy of the same circular piece of DNA, eliminating the subcloning steps that were necessary in early mutagenesis experiments that required the ssDNA bacteriophage M13. The current version of pHAZY is based on pBR322 and does not contain the genetic sequence that directs packaging of single strand phagemids into M13KO7 helper phage heads.

Twice, I have attempted to subclone the phage F1 single strand origin from the pST:II series into the pHAZY plasmids. Both vectors are based upon pBR322, with the

F1 origin cloned into the *Pvu* II site of pST:II on *Bam*H I linkers. I attempted to move a larger *Sal* I-*Nde* I fragment containing the F1 origin from pST:II into the corresponding site in pHAZY, but I was not able to obtain transformants. The *Pvu* II site in pBR322 is within the coding region of the ROP (repressor of primer) protein that is involved in plasmid copy control (Castagnoli *et al.*, 1989). When copy control is removed, the limited number of trp repressors in the cell will not be able to bind to all of the plasmid borne *trp* operators and TrpLE-Z-BPTI expression will become constitutive and perhaps lethal (Warne *et al.*, 1986). This unproven hypothesis suggests that I may have obtained the desired construction if I had selected for transformants at 30 °C rather than 37 °C, which should decrease the plasmid copy number. I also recommend attempting the construction using different restriction fragments such as *Sal* I-*Pst* I. If both attempts fail, I would choose another tactic, such as placing the single strand origin in another location on the plasmid.

The simplest alternate construction is to isolate the *Bam*H I fragment containing the F1 origin from the pST:II phagemid and ligate it into the *Bam*H I site of pHAZY. The vector can be treated with a phosphatase to prevent recircularization of the plasmid, leaving the *tet* gene intact. Recombinants can be selected by transforming cells bearing F pili with the ligation mixture, isolating phage and phagemid DNA by superinfection with M13KO7 helper phage, and retransformation of F^- cells with single stranded DNA, selecting for ampicilin resistance. Insertion of the F1 origin at other sites will require additional linker oligonucleotides and other procecures.

Finally, I note that the lore from many labs is that the F1 origin cannot be inserted into their favorite expression plasmid, meaning that they have not been able to do so with a reasonable expenditure of effort (D. Shortle, personal communication). If after several attempts the F1 origin cannot be inserted into pHAZY, it may be worthwhile to create constructs in pEZ that facilitate subcloning into pHAZY. This might include removal or replacement of the protein A constitutive promoter in pEZ in order to turn off expression in

these plasmids, reducing the selection against them at 37 °C and therefore improving the yields of single stranded DNA. In this case, the chymotrypsin sensitive linker sequence should certainly be moved into the pEZ series.

2C.2. Alternative promoters

The *trp* promoter is at best leaky and is repressed in minimal media only by the addition of exogenous tryptophan (Brosius, 1988). Other promoters have been developed which are both more tightly regulated and more efficient. For isotope labeling experiments, any improvement in protein yield will be well worth the effort of developing a better expression system.

The hybrid *tac* promoter (de Boer *et al.*, 1983) utilizes the *lac* operator sequence which can be more tightly controlled than the *trp* promoter through the use of *lac*I^Q strains (Brosius, 1988). In addition, it is not necessary to add corepressors prior to induction by the addition of IPTG. The λ P_L promoter is even more efficient and can be controlled by the temperature sensitive λ repressor C_{I857}; induction in this system is accomplished by a shift in temperature to 42 °C, so no chemical inducers are needed. This system was successfully used by LeMaster in the expression of deuterated thioredoxin (LeMaster and Richards, 1988), and by Clore and Gronenborn to express IL-1β uniformly labeled with ^{13}C and ^{15}N (Clore *et al.*, 1990a).

The current state-of-the-art in *E. coli* expression is certainly the T7 system developed by Studier (Studier and Moffatt, 1986; Rosenberg *et al.*, 1987; Studier *et al.*, 1990). The T7 promoter is recognized by the RNA polymerase of bacteriophage T7 and not by the *E. coli* RNA polymerase. *E. coli* strains are available in which a single copy of the gene for T7 polymerase under *lac* control has been inserted into the chromosome; synthesis of the T7 polymerase is induced by the addition of IPTG, thereby inducing transcription from the plasmid borne T7 promoter. This system has been recently used to produce 50-75 mg/L of uniformly ^{13}C labeled carbonic anhydrase and the MetJ protein in a minimal/acetate

media (Venters *et al.*, 1991). Protein expression from the T7 promoter often results in the formation of inclusion bodies (Grumont *et al.*, 1988; Rentier-Delrue *et al.*, 1989; West *et al.*, 1989; Chang and Doi, 1990; Conner and Udey, 1990), which as we have seen is a desireable result for an intracellular BPTI expression system. BPTI might be expressed as the TrpLe-Z- fusion protein, or perhaps directly with an N-terminal methionine extension. The latter would be advantageous for mutants which do not refold to give active inhibitors.

In summary, I recommend an attempt at BPTI expression from the T7 promoter as the highest priority. This promoter is the most likely to increase the yield of BPTI in minimal media.

Appendix 2D

2D. Alternative fusion protein cleavage schemes

2D.1. Many mutants are not resistant to chymotrypsin

The single column chymotrypsin affinity purification and cleavage procedure was developed to process mutants which retain chymotrypsin binding activity when refolded. This has proven enourmously useful for a large number of mutants and isotopically labeled samples, but many interesting mutants will not bind to chymotrypsin, either because they do not refold to give stable, native-like structures, or because residues in the protease binding site have been altered. Specifically, we have been interested in mutants designed to mimic Creighton's folding intermediates—these include C14A/C55A and C38A/C55A (Creighton, 1975)—and three mutants which retain only a single disulfide bond. We are also interested in mutations which destabilize the core of BPTI, such as N43D. Most of these mutants are not expected to be active inhibitors when refolded.

Several alternative schemes are available to deal with the cleavage of inactive fusion proteins. The first is to express BPTI directly, without the fusion protein, such as with the T7 expression system described in the previous appendix. If this proves to be impractical, chemical and enzymatic cleavage schemes must then be considered. The hydroxylamine sensitive AsnGly sequence that is present in the pEZ vectors could be reinserted, or a methionine residue could be inserted for cleavage by cyanogen bromide, together with mutation of the Met-52 residue of wild type BPTI. This tactic has been used by Creighton (Darby *et al.*, 1991), but the cyanogen bromide procedure is rather harsh and may not yield materials of suitable purity for crystallization and NMR analysis.

2D.2. Attempts with Ala-64 Subtilisin

Nilsson and Carter have attempted to adapt the Z-BPTI fusion proteins for cleavage by Ala-64 subtilisin BPN', an active site mutant of the protease that utilizes a P2 histidine

residue in the substrate to "assist" in the cleavage of a peptide bond (Carter and Wells, 1987; Carter *et al.*, 1989). They failed to observe cleavage of the Z-BPTI fusion proteins and attributed this to unfavorable interactions of a proline at P2' (Pro-2 in the BPTI sequence) with the enzyme (Nilsson and Carter, personal communication).

I have attempted the Ala-64 subtilisin substrate-assisted cleavage of a P2A mutant of the same BPTI fusion protein used by Nilsson and Carter. The mutant was made in the C14A/C38A background in the pEZ:475 plasmid and then moved into the pHAZY expression vector on a *Bgl* II-*Eco*R I fragment; the final plasmid was named pHAZ:His3:475. While this represents a compromise of the "authentic" BPTI sequence, much as Creighton's M52R mutant or an N-terminal methionine extension, the N-terminus is not well ordered in the crystal or in solution, suggesting that the P2A mutation will have little or no structural consequence. Preliminary evidence from SDS-PAGE experiments suggest that the cleavage proceeds as designed with completely reduced and unfolded fusion proteins, but I have not yet isolated the desired BPTI fragment to confirm the cleavage. The results so far are very encouraging, and I recommend pursuing the Ala-64 subtilisin cleavage, trying both folded and unfolded fusion proteins.

2D.3. Other enzymatic cleavage possibilities

The cleavage of heterologously expressed fusion proteins is a general problem and a wide range of more or less specific enzymes have been used for this purpose (Carter, 1990). The most popular scheme uses the blood clotting Factor X_a, which has the tetrapeptide recognition sequence Ile-Glu-Gly-Arg (Nagai and Thøgersen, 1984). However, a recent report from a Keystone Symposium presents evidence that BPTI fusion proteins with a Factor X_a site next to Arg-1 of BPTI are cleaved at a site six residues N-terminal to the designed site (Lauritzen *et al.*, 1991). It seems other enzymatic cleavage schemes must be sought.

Chapter 3

A New Strategy for Spin Labeling BPTI

3. Introduction

Isotope labeling procedures such as those mentioned in chapter 1 are expensive. For this reason, high yield modification reactions are necessary in order to make the combination of isotope and spin labeling economically justifiable. In this chapter, I will describe the protein chemistry I have developed for the attachment of nitroxide spin labels to free thiols in mutant derivatives of BPTI. I will begin with a short survey of the previously available techniques for spin labeling of BPTI, primarily developed by Phyllis Kosen in the Kuntz laboratory. Following this, I will discuss a number of options for the introduction of free thiols into BPTI, which has 6 cysteines, each of which participates in a disulfide bond. I will then focus on two mutants chosen as targets for spin labeling, C55A and C38A. The labeling procedures for C38A proved to be extremely efficient, while attempts to label the Cys-5 thiol in C55A were eventually abandoned. However, the C55A mutant turns out to be particularly interesting from a protein folding point of view, and much of the discussion of this mutant will focus on the folding reaction of the completely reduced mutant with oxidized DTT. I will also discuss some of the folding/unfolding properties of the C38A and C14A mutants. In the next chapter, the results from NMR experiments with a spin labeled derivative of C38A will be presented.

3.1. Previous Non-Selective Labeling

The only pure preparations of mono- spin labeled derivatives of BPTI to date have come from the extensive purification of a mixture of species nonselectively labeled at the 5 amino groups found in the wild type protein. The amino group specific reagent succinimidyl 2,2,5,5-tetramethyl-3-pyrroline-1-oxyl-3-carboxylate was used for these experiments. Beginning with 650 mg of protein, only species monolabeled at the amino terminus (12%), lysine 15 (9%) and lysine 26 (12%) were produced in sufficient quantities for subsequent NMR experiments (Kosen *et al.*, 1986; Kosen, 1989). From these data, Kosen

concluded: "In general, the lack of specificity will limit the preparation of mono-derivatized proteins to those proteins which can be obtained in large quantities" (Kosen, 1989).

Not all of the lysine derivatives produced in this manner will be suitable for long range distance determination via NMR. The primary complication arises from the length of the tether connecting the paramagnetic center and the relatively rigid protein backbone. If the label is free to flop around and has a correlation time significantly different from the overall tumbling time of the protein, the quality of the distance information obtained from the experiment will be degraded. From this point of view, the lengthy lysine side chain represents possibly the worst case scenario, while the modification at the amino terminus is expected to produce a label that is much less mobile.

Kosen has also attempted to label tyrosine residues in BPTI using the reagent N-(2,2,5,5-tetramethyl-3-carbonylpyrrolidine-1-oxyl)imidazole (Kosen, 1989). These efforts were not successful, apparently due to extensive modification of lysine residues by the reagent. This problem may be circumvented by reversibly blocking the lysine residues with citraconic anhydride (Butler and Hartley,), with a possible degradation in yield with each individual reaction. These experiments have not been pursued.

3.2. Previous mutants for selective labeling

Wien, Morrisett, and McConnell (Wien *et al.*, 1972), and later Schmidt and Kuntz (Schmidt and Kuntz, 1984) have shown that the single His-15 of hen egg lysozyme can be labeled with the reagent 2,2,6,6-tetramethyl-4-(bromoacetamido)piperidine-1-oxyl and that useful distance information can be extracted from the NMR data obtained with this derivative. Wild type BPTI does not contain histidine residues, so the introduction of a single histidine by mutagenesis may provide a site for more specific labeling than was possible with the amino group specific reagents. Based upon the surface exposure and the occurrence of a number of charged residues at the corresponding sequence position in a number

of natural homologs (Creighton and Charles, 1987), Tyr-10 was chosen as the site for the mutation to histidine. Labeling of this mutant was impeded by two complications: (1) Met-52 was also reactive towards the reagent under the conditions used, and (2) the histidine labeled species was difficult to separate from the unlabeled material. The latter difficulty was attributed to the abnormally low pK_a value of His-10—the observed value was approximately 4— probably due to the number of positive charges surrounding the site of mutation. The double mutant Y10H/M52L was constructed to circumvent the first problem, but these experiments have not yet been pursued.

3.3. Odd Cysteine Mutants: Introduction

3.3.1. Rationale

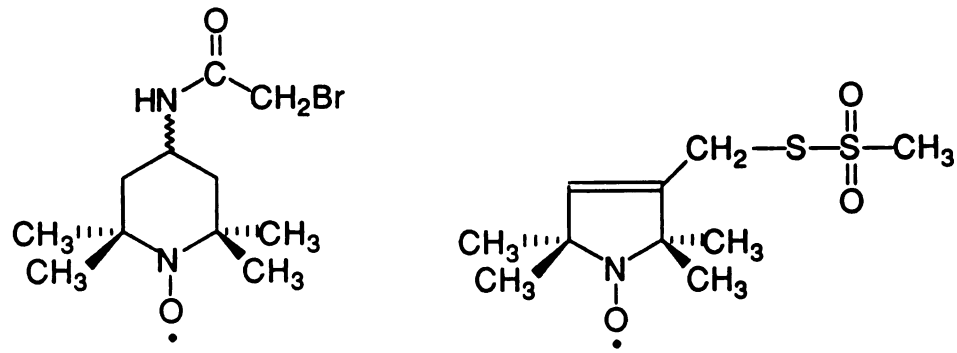
Recently, crystallographers have begun to introduce additional cysteine residues into proteins as specific attachment sites for heavy atoms—most often mercury derivatives—for the solution of the phase problem by multiple isomorphous replacement (Sun *et al.*, 1987; Hatfull *et al.*, 1989; Tucker *et al.*, 1989). Generally, a number of mutants must be screened to find suitable candidates based on the following criteria: (1) the added cysteine residue must react with the heavy atom reagent at a reasonably high occupancy, (2) the mutant must yield crystals which when modified with the heavy atom are isomorphous with the wild type protein, and (3) the modified crystals must diffract well. Many, perhaps most, of the cysteine mutants that have been constructed in the course of a structure determination project have failed on one or all of these criteria. Nevertheless, the method has provided a more rational approach to the discovery of suitable heavy atom derivatives, a process that has historically been more art and serendipity than science.

The criteria for suitable mutant derivatives for spin labeling of a protein as an aid in solution structure determination are more relaxed and less well defined than the criteria for cysteine-directed heavy atom derivatives for crystal structure determination. NMR spectro-

scopists do not have to worry about the vagaries of crystal packing interactions, yet they will always be less certain than their colleagues in crystallography that the derivatives they have prepared bear a detailed similarity to the unmodified and unmutated protein whose structure is desired. For the crystallographer, it is a difference between obtaining a structure of the unmodified protein or not obtaining the structure; once the phases of the reflections have been determined, the structural model can be refined with only data collected from the unmodified wild type protein. The NMR spectroscopist using data from spin labeled proteins can only check the data from the spin label experiment for consistency with all other available data. The best indication of a relative lack of structural perturbation induced by a spin label is found by a comparison of chemical shifts from the diamagnetic derivative to the unlabeled protein. Luckily, the wealth of data obtained from protein engineering experiments in the last ten years indicates that proteins can tolerate a wide variety of mutations with only small perturbations to the native structure; for example, mutants of BPTI that lack one of the native disulfides have a structure that is very similar to the wild type protein (Eigenbrot *et al.*, 1990). It is not yet known to what resolution structural perturbations can be defined with the NMR techniques (Hurle *et al.*, 1991).

Wild type BPTI has six cysteines which form three disulfide bonds. The sulfur atoms in these disulfide bonds are very much less active nucleophiles than are the thiolate anions in non-disulfide bonded cysteines at pH values above their pK_a's. If a free thiolate could be introduced into BPTI such that there was a minimal perturbation of the native structure, it ought to be possible to modify it in a very selective manner with an iodo- or bromoacetamide derivative of a nitroxide spin label. Compounds such as 2,2,6,6-tetramethyl-4-(bromoacetamido)piperidine-1-oxyl (BrAmTempo) are commercially available and have been used to modify cysteines, methionines, and histidines in proteins (Kosen, 1989). A compound which is even more specific for cysteine, (1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methane thiolsulfonate (proxyl MTS) has been described

(Berliner *et al.*, 1982; Kosen, 1989) but is not yet commercially available[1]. However, the specificity of the iodo- or bromoacetamide derivatives for a free cysteine introduced into BPTI is expected to be very good because BPTI does not contain histidines, and labeling at methionine occurs at a much slower rate than does the reaction at cysteine (Kosen, 1989). The specificity of the reaction of the cysteine residues of the BPTI folding intermediates with iodoacetamide or iodoacetate forms the basis for all of Creighton's studies.



2,2,6,6-tetramethyl-4(bromo-acetamido) piperidine-1-oxyl

BrAmTempo

(1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methane thiolsulfonate

Proxyl MTS

Figure 3.1: Two nitroxide compounds which specifically modify cysteine residues. The trivial names are not standard but are used in the text for clarity. Note the mixed stereochemistry in the BrAmTempo molecule.

Unpaired cysteines can be introduced into BPTI by either addition or subtraction—Cys-plus and Cys-minus mutants, respectively. Cys-plus mutants are constructed by oligo-directed mutagenesis (or semisynthesis), while Cys-minus mutants may be constructed by mutagenesis or chemical trapping. Cys-plus mutants provide a much greater range of potential sites than Cys-minus mutants, which are limited to positions in the sequence corresponding to the six cysteines in the wild type molecule. In either case, it is

[1] I strongly recommend the use of proxyl MTS in the future. For more comments on this reagent, see the appendix to this chapter (3.A)

wise to follow the dictum *caveat emptor*: the possibility always exists that an odd cysteine will destabilize the protein via disulfide rearrangements, though this is unlikely at suitable redox and pH conditions.

3.3.2. Cys-plus mutants

Among the several advantages to the addition of cysteines are the following: (1) a large range of choices, not limited to the position of the cysteines in the native sequence, (2) surface residues may be chosen for maximum reactivity with bulky spin labels and minimal structural perturbation, and (3) spin labeled Cys-plus mutants are more likely to retain the structural integrity and thermal stability of the wild type protein than are Cys-minus mutants. Of course, Cys-plus mutants have a greater potential for non-native disulfide bond formation in a combinatorial sense, but it is very likely that the stability of the native structure will greatly favor the formation of native disulfide bonds.

Experiments with Cys-plus mutants of BPTI have not been previously pursued. Difficulties with the expression of such mutants in the secretion systems were anticipated due to the reactivity of the free thiol; poorly understood reactions with molecular oxygen and intermolecular aggregation were of particular concern. With the pHAZY intracellular expression system, in which the disulfide bonds are formed *in vitro* using a controlled glutathione-based disulfide shuffling system, the anticipated unproductive side reactions can be minimized. However, mixed protein-glutathione disulfides can accumulate with these procedures, causing potential difficulties. I predict that it will not be possible to reduce the mixed disulfides with DTT^{SH}_{SH} while leaving the 14-38 protein disulfide intact; however, it may be possible to prepare the protein with 3 free thiols and selectively resynthesize the 14-38 disulfide with oxidized DTT. Alternatively, the mixed disulfide might be selectively reduced with a large excess of a linear thiol reagent such as glutathione. Because the effective concentration of the thiols in the native protein is so large—230 M for the 14-38 disulfide up to 4.6×10^5 for the 5-55 disulfide (Creighton and Goldenberg, 1984)—it is nearly

impossible to reduce them with linear disulfide reagents. In some cases, it may be worth investigating whether the 14-38 disulfide bond can be protected from reduction by DTT^{SH}_{SH} by complex formation with trypsin or chymotrypsin. The major expected complication to this procedure is the potential instability of the protease to reduction with DTT^{SH}_{SH}. The following protocol should serve in the best case: 1) refold the protein with glutathione under conditions which will favor the mixed disulfide, 2) dialyze out the remaining guanidine hydrochloride and glutathione reagents, 3) apply the refolded protein to the chymotrypsin column for fusion protein cleavage and affinity purification, including the gentle shaking overnight, 4) add DTT^{SH}_{SH} to the column wash buffer and incubate for approximately 30 minutes, 5) quickly wash the column with deoxygenated buffer to remove the DTT^{SH}_{SH}, and 6) elute the free thiol form of the protein in the normal 250 mM KCl, pH 1.7 solution. The labeling reaction should be carried out soon after elution to prevent side reactions involving the free thiol, including oxidation and intra- or intermolecular thiol disulfide exchange reactions.

Finally, I acknowledge the work of my colleague Elisha Haas at Bar Ilan University in Israel, who has constructed the BPTI Cys-plus mutant A58C and expressed it in the pHAZY expression vector. Haas is interested in attaching a fluorescent probe at the additional cysteine, but his on-going characterization of the thiol chemistry of this mutant should be valuable for future work with all Cys-plus mutants. This mutant will be available to the Kuntz lab.

3.3.3. Cys-minus mutants

3.3.3.1. Candidates

Solely based on the primary sequence of BPTI, each of the six cysteines 5, 14, 30, 38, 51, and 55 are potential targets for construction of Cys-minus derivatives for spin labeling; however, a number of features of the native structure effectively limit the possible sites even further. Based on the surface exposure of the 14-38 disulfide and the kinetic

stability of the N_{SH}^{SH} unfolding intermediate, cysteines 14 and 38 provide the candidates most likely to succeed for construction of the free thiol derivative and subsequent derivatization. The sulfur atoms of the remaining cysteine residues are completely buried in the native conformation (Deisenhofer and Steigmann, 1975) and therefore present less attractive sites for the introduction of a bulky TEMPO derivative. The 30 and 51 thiols in the off-pathway folding intermediate {5-55, 14-38} are completely unreactive to iodoacetamide and iodoacetate (States *et al.*, 1984), effectively precluding spin labeling at these sites. However, the two disulfide intermediate {30-51, 14-38} has been isolated in refolding reactions with the 5 and 55 thiols blocked with iodoacetate. At pH 3.8 this species has a melting temperature of 44°C, indicating that it is able to maintain some of its native structure (States *et al.*, 1987); the stability of the corresponding acetamide derivative is not known. The same intermediate with Cys-5 and Cys-55 both present as mixed disulfides with glutathione has been isolated from experiments where the completely reduced protein is refolded in 3.0 mM GSSG (Creighton, 1977b; Creighton, 1977a). Though it has not been shown that this species has a compact, native-like conformation, these data suggest that both Cys-5 and Cys-55 present possible sites for spin labeling. These spin labeled derivatives may significantly perturb the protein structure, but they may also yield better localized labels than have been observed when labeling lysine residues on the protein surface (Kosen *et al.*, 1986). As I will later show, experiments with labeling at Cys-14 were indeed successful, while attempts to attach a label at Cys-5 were thwarted, but for some very interesting reasons.

3.3.3.2. Chemical Strategy: Taking Advantage of the Unique 14-38 Disulfide Bond

Beginning with wild type BPTI, Creighton has prepared derivatives with either Cys-14 or Cys-38 blocked by partial alkylation of the N_{SH}^{SH} unfolding intermediate with iodoacetamide followed by trapping of the remaining thiol with oxidized glutathione

(Creighton, 1977a). The individual mono-alylkated species were subsequently purified from a mixture also containing di- and non-alkylated species. The mixed disulfides in the individual mono-alylkated species could then be selectively reduced with DTT^{SH}_{SH} and the liberated thiol subsequently alkylated with the appropriate spin label reagent. Due to the multiplicity of products, the yield of the mono-alylkated proteins may not be suitable for spin labeling of isotope labeled proteins; however, the method will be suitable for the commercially available wild type protein. Attempts to obtain the equivalent mono-alylkated derivatives by trapping the N^{SH}_{SH} species with a mixture of spin label reagent and iodoacetic acid were not successful for two reasons. The first was the apparent tendency of a fraction of the nitroxide groups to be reduced in an intramolecular reaction with the remaining un-trapped thiol. Furthermore, the $N^{SCOM}_{SAmTEMPO}$ and $N^{SAmTEMPO}_{SCOM}$ species could not be separated in preliminary experiments on a MonoS FPLC column (P. Kosen, personal communication).

3.3.3.3. Mutagenesis strategy

The cloning and expression of mutant BPTI genes provides a much more direct route to proteins which can be specifically labeled than does the chemical trapping of hemi-alkylated species described above. In this context, the C14A mutant is equivalent to the N^{SCOM}_{SH} species, as the C38A mutant is to the N^{SH}_{SCOM} species. Two approaches were available for the construction of the C14A and C38A mutants, oligo-directed mutagenesis from the wild type gene and splicing from existing mutants. The latter approach was chosen for its simplicity.

All of the genes coding for the removal of single disulfides in BPTI have been constructed in the pEZ and pHAZY vectors, replacing the cysteine pairs with alanines. In the pEZ vectors, there are single copy restriction sites spanning the codons for Ala25-Ala27 (*Sty* I), and Gly37-Arg39 (*Pst* I, absent in the C14A/C38A double mutant). (In the pHAZY vectors, theses sites exist at additional locations in the plasmid, such that using

them requires inconvenient partial digestions.) These sites allow convenient splicing of the double cysteine mutants with the wild type gene, creating single cysteine mutants which may then be spliced into the pHAZY vectors. In the constructions I will describe below, Cys-14 and Cys-38 were split at the *Sty* I site, while Cys-5 and Cys-55 were split at the *Pst* I site, though the *Sty* I site might also have been used. A schematic of the construction of the C38A mutant, which will be described in section 3..4.2.3, is shown in figure 3.2.

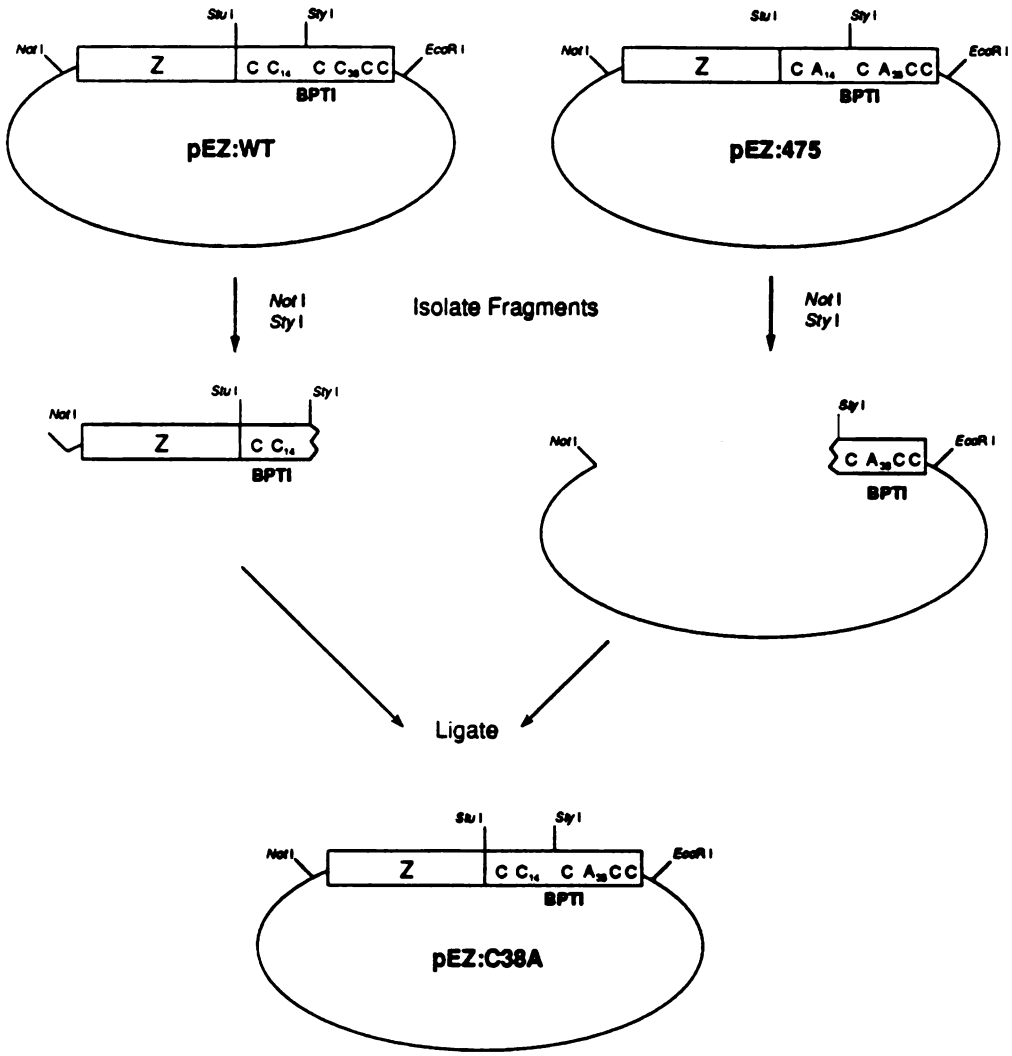


Figure 3.2: A schematic of the construction of the C38A mutant in the pEZ:BPTI expression vector by gene splicing. The synthetic protein A domain is designated "Z", and the cysteine residues (or their alanine replacements) in the BPTI domain are shown. The *Not* I and *Sty* I sites are used for this construction, while the *Stu* I and *Eco*R I sites are used to move the gene into the pHAZY vectors (not shown).

Unfortunately, the wild type gene in both the pEZ and pHAZY vectors lacks the *Eco*R I restriction site at the 3' end of the BPTI gene that is customarily used in transferring mutants from the pEZ to the pHAZY vectors; this relic from the initial construction of the pEZ plasmids (B. Nilsson, personal communication) was mentioned in an appendix to chapter 2. It represents a minor inconvenience that will touched upon below.

3.4. Two Cys-Minus Mutants

3.4.1. C55A BPTI

3.4.1.1. Construction of the C55A mutant

Apart from being a target for spin labeling, there were several reasons for constructing the C55A mutant. The C55A mutant is an intermediate in the reconstruction of the "S2", or C38A/C55A, mutant of BPTI. This mutant is designed to mimic one of the two predominant non-native two disulfide intermediates which immediately precede the N_{SH}^{SH} in the refolding pathway of BPTI. The other such intermediate, "S1" or C14A/C55A, was constructed by Björn Nilsson and is described in the previous chapter (the "S" stands for scrambled). Either of these molecules presents a unique opportunity to study the detailed structure of a folding intermediate. Finally, the C55A mutant is interesting in its own right. In the folding pathway, Cys-55 is involved in an intramolecular disulfide exchange reaction with both of the kinetically favored non-native two disulfide intermediates. Through prevention of those rearrangements, the C55A mutant will allow a more thorough investigation of the equilibrium properties of the collection of non-native two disulfide intermediates than has been previously possible. This will be discussed in more detail below.

The C55A mutant was constructed by splicing of the genes for wild type BPTI (pEZ:b3:wt) and for the C5A/C55A double mutant (pEZ:920) in the pEZ:BPTI vectors. The 616 bp *Not* I - *Pst* I fragment of pEZ:b3:wt was ligated to the 4030 bp *Not* I - *Pst* I

vector fragment of pEZ:920 to create the plasmid pEZ:C55A. The BPTI coding region was then moved into the pHAZY:BPTI vectors as a 186 bp *Stu* I - *Eco*R I fragment , using pHAZY:637 (C5V/C55A) as the vector, though any of the pHAZY plasmids would have been suitable. Both the C55A and C38A coding plasmids direct expression of fusion proteins of the proper length (by reducing SDS-PAGE) when compared with either the C5V/C55A or the C14A/C55A fusion proteins (figure 4.3; details in the text below).

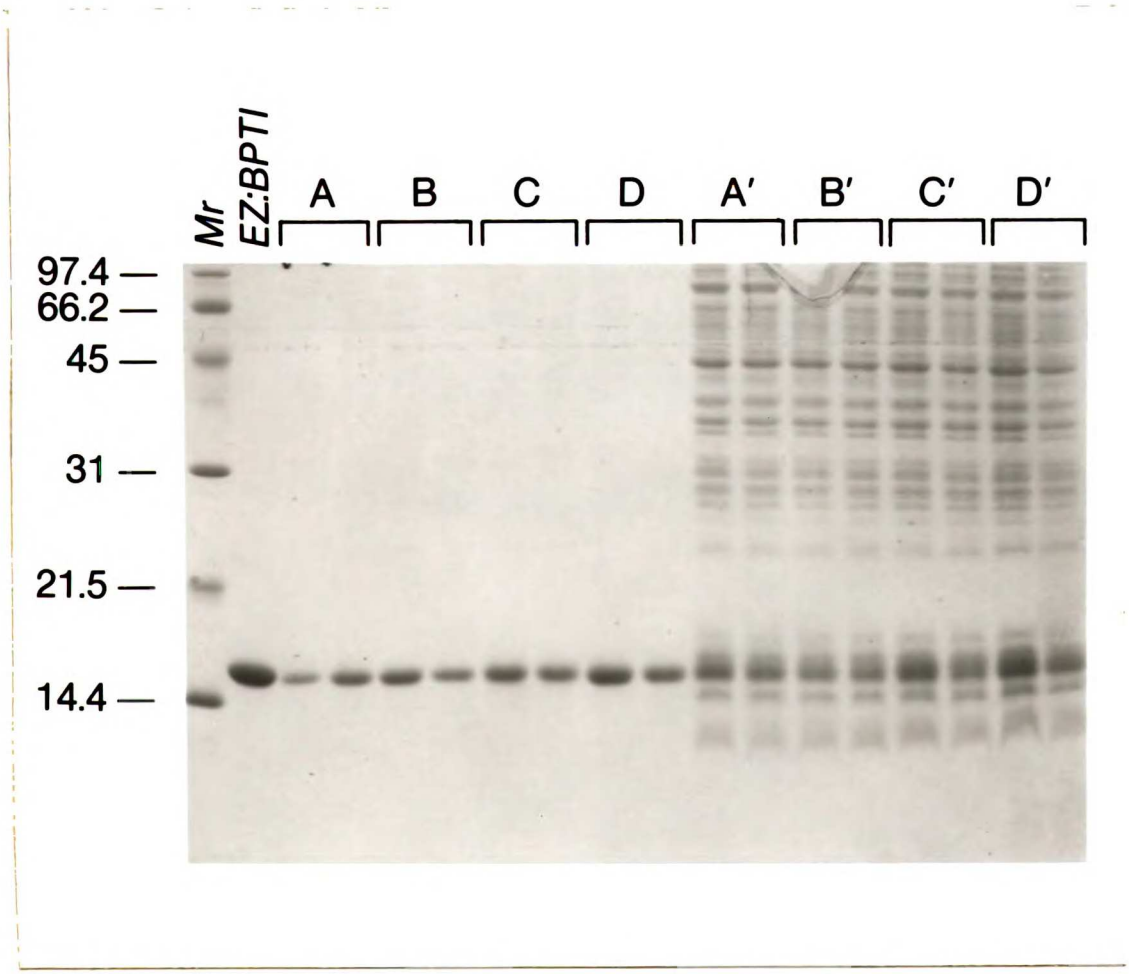


Figure 3.3: The pHAZY:C55A and pHAZY:C38A expression plasmids direct the expression of full length TrpLE:Z:BPTI fusion proteins. Shown is a 16% reducing SDS gel. The molecular weight markers are as in Figure 2.4. Expression from two independent isolates of each plasmid are shown. A set of inclusion body preps are shown, followed by a set showing total protein (designated with primes). (A-B) pHAZY:637 (C5V/C55A); (C-D) pHAZY:S1 (C14A/C55A); (E-F) pHAZY:C38A; (G-H) pHAZY:C55A

3.4.1.2. Expression

3.4.1.2.1. Refolding

Activity

The TrpLE-Z-C55A fusion protein was expressed in shaker flask expression experiments as described in chapter 2. Following folding of the protein from solubilized inclusion bodies, two samples were prepared for trypsin inhibition activity assays to check for DTT^{SH}_{SH} sensitivity of the refolded protein (Coplen *et al.*, 1990). In the control sample, the protein was simply desalted after the glutathione refolding step. The second sample was treated with 10 mM DTT^{SH}_{SH} prior to desalting. The wild type protein is expected to be fully active after this treatment, with only the 14/38 disulfide bond selectively reduced, but the C55A mutant was expected to be much less stable to inactivation by DTT^{SH}_{SH} (Creighton, 1977a) (Coplen *et al.*, 1990). The control sample gave stoichiometric trypsin inhibition at the level of 9.6 mg BPTI per liter, while the DTT^{SH}_{SH} treated sample was completely inactive (figure 3.4).

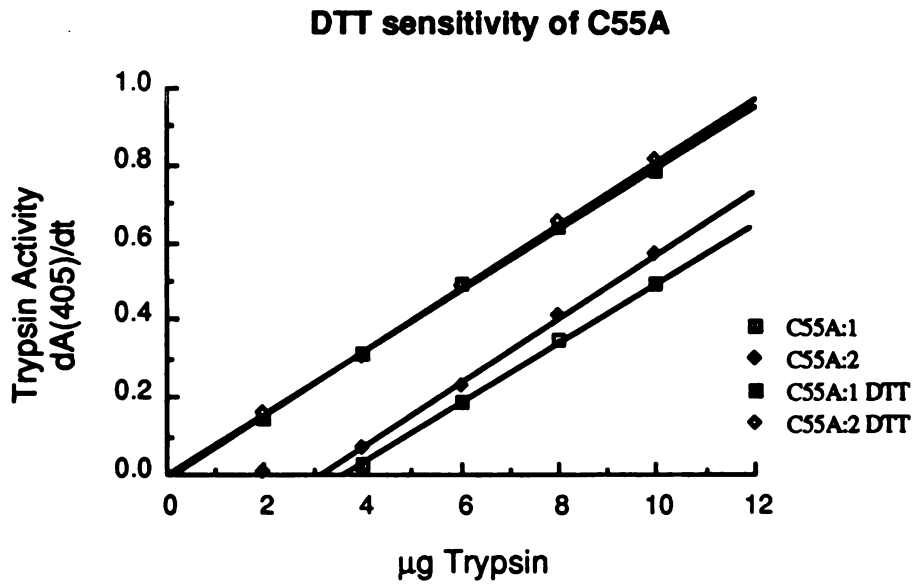


Figure 3.4: DTT sensitivity of the refolded C55A mutant. The DTT treated samples intercept the abscissa at 0 and have no trypsin inhibition activity. The samples not treated with DTT have an average x intercept at 3.41 µg trypsin in the standard trypsin inhibition assay. This corresponds to an expression level of 9.6 mg/l, based on the molecular weight of wild type BPTI.

3.4.1.2.2. Chymotrypsin cleavage

Gel evidence

The glutathione refolded TrpLE-Z-C55A fusion protein was cleaved and affinity purified on a chymotrypsin column as described in the previous chapter. The protein eluted off the chymotrypsin column gave single bands on Coomassie stained 14% nonreducing SDS and 15% native gels. In addition, a single peak was observed in the C18 HPLC chromatogram, which gave the correct N-terminal sequence.

3.4.1.2.3. Glutathione adduct

The cleaved C55A protein was less mobile on 15% native gels than the wild type protein (figure 3.5A), suggesting that the glutathione refolded protein forms a mixed disulfide adduct with glutathione. The negative charge of the glutathione reduces the mobility of the protein in the native gels, but not to the extent of the doubly IAA modified RCOM protein. In a 14% nonreducing SDS gel, the protein was slightly more mobile than wild type BPTI, perhaps because the added negative charge led to a lower level of SDS binding (figure 3.5B).

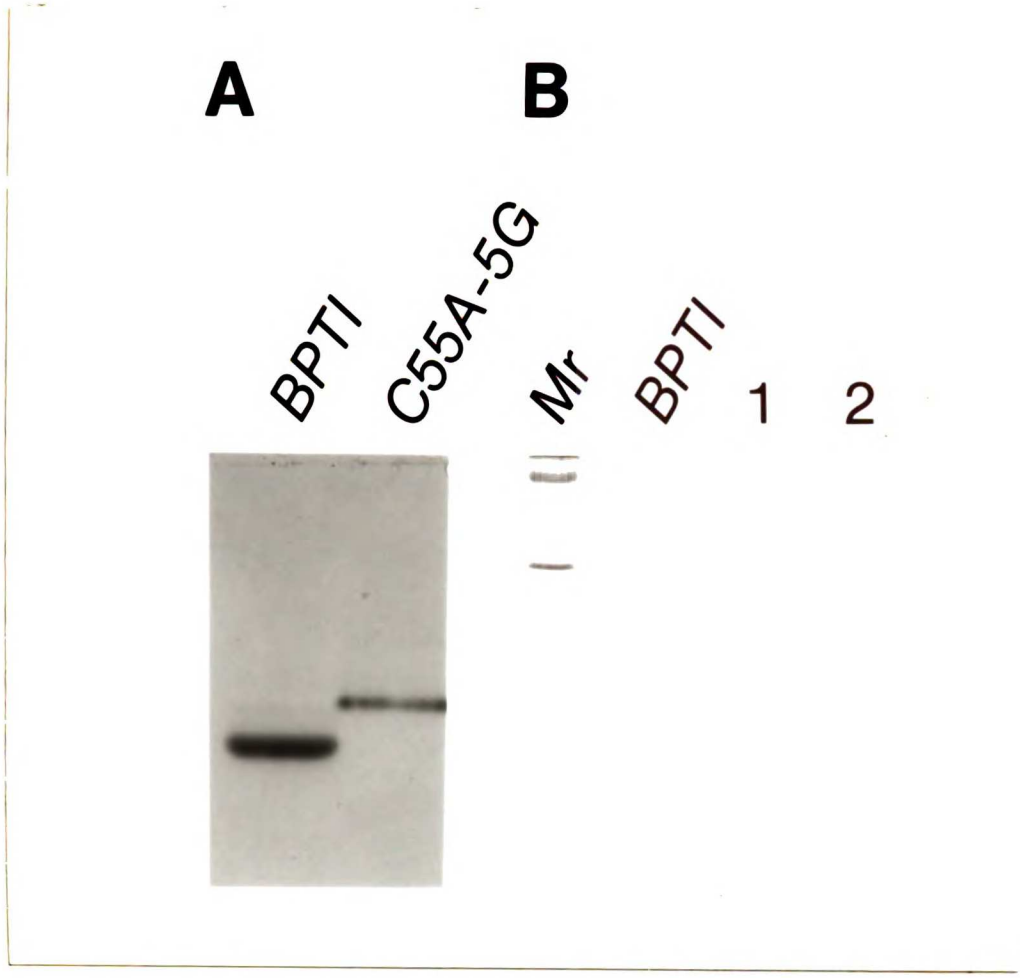


Figure 3.5: The C55A protein eluted from the chymotrypsin column contains a mixed disulfide with glutathione. Comparison of the mobility of the C55A-SG protein with commercial wild type BPTI (Traysylol) by electrophoresis through (A) a 15% native gel. Lane 1 - wild type BPTI. Lane 2 - the C55A-SG protein; (B) a 14% nonreducing SDS gel, with lanes as follows: Mr, molecular weight markers; BPTI; (1 & 2) two different quantities of protein from the chymotrypsin column eluate of the C55A protein.

Two other pieces of evidence strongly suggest that the C55A protein formed a mixed disulfide adduct with glutathione. First, the protein did not react with DTNB in 6M GdnHCl. Second, a low S/N and low resolution $[M+2H]^{+2}$ molecular ion peak in the mass spectrum was consistent with the glutathione adduct. The $[M+H]^{+1}$ peak was not observed. Additional confirmation could be obtained by refolding the protein with radioactive glutathione.

3.4.1.2.4. Refolding C55A with oxidized DTT

Rationale

The refolding of wild type BPTI with linear disulfide reagents such as glutathione is approximately ten times faster than with the cyclic disulfide reagent, oxidized DTT (Cleland, 1964) (Creighton, 1974). However, due to the stability of its intramolecular disulfide and a more favorable entropy, DTT does not readily form mixed protein disulfides (Creighton, 1974). This makes it the reagent of choice for the preparation of the free thiol form of C55A. The kinetic properties of the glutathione refolding reaction are still more suitable for the initial refolding step after solubilization of the inclusion bodies, and the utilization of the glutathione mixed disulfide as a protecting group during the chymotrypsin cleavage and affinity purification is probably essential. Once the properly cleaved and purified C55A-GSH protein is obtained, it may then be completely reduced by 10 mM DTT in the presence of 8 M urea, isolated in 10 mM HCl by buffer exchange on a G-25 column, and refolded utilizing oxidized DTT for disulfide bond formation.

Methods

A 10 ml solution of C55A-SG at a concentration of 0.6 mg/ml in the standard TKE buffer plus 8 M urea and 10 mM DTT^{SH}_{SH} was incubated at 37°C for 50 minutes. The buffer was exchanged for 10 mM HCl on an 80 ml G-25 column. The pooled protein fractions (~15 ml) were at a concentration of ~0.31 mg/ml (≈ 47 μM). A fraction of the completely reduced protein was refolded at a final concentration of 28 μM in TKE buffer with the addition of 20 mM oxidized DTT. The refolding reaction was incubated at room temperature for 48 hours. No attempt was made to exclude oxygen.

Results

Gel evidence for refolding

The observation of a band in a 15% native gel that had the same mobility as wild type BPTI (figure 3.6) suggested that an indeterminate fraction of the protein had regained

the native-like structure without the presence of the glutathione mixed disulfide. A series of controls utilizing a variety of iodoacetate (IAA) modified C55A and wild type samples were inconclusive, probably due to impurities in the IAA preparations. It is very important that the IAA stock solutions be prepared from recrystallized material just before use, to minimize the concentration of potentially reactive iodine species.

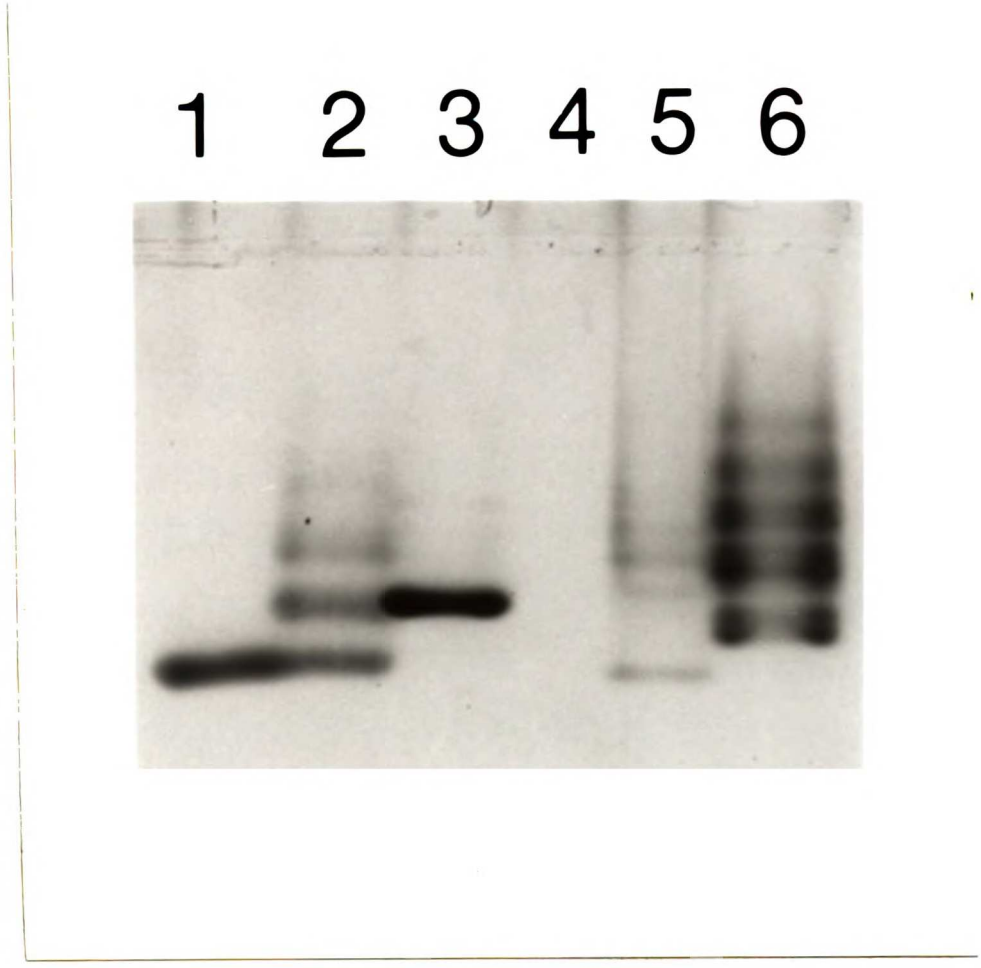


Figure 3.6: The completely reduced C55A protein can be refolded with oxidized DTT. A 15% native gel is shown documenting the refolding of the C55A mutant to a protein with a mobility identical to wild type BPTI. The lanes are (1) wild type BPTI, (2) BPTI + IAA, (3) C55A-SG, (4) reduced C55A + IAA, (5) C55A refolded with oxidized DTT, (6) C55A BPTI refolded with oxidized DTT + IAA.

Precipitation when concentrated

The 20 ml refolding reaction was concentrated on an Amicon YM-2 membrane. A visible precipitate formed when the total volume was approximately 8 ml. When the volume reached approximately 1.5 ml, the precipitate was separated by centrifugation for 10 minutes in a microcentrifuge. Not all of this precipitate was soluble in 1.5 ml TKE with 8 M urea. It is not clear how much of the protein was found in the precipitate, especially since some solid remained on the Amicon membrane.

Native gel analysis of the concentrated refolded proteins again suggested that as much as 70% of the soluble protein was refolded to a native like structure, but again, there were complications with the alkylation controls. Further characterization of the C55A protein was not done.

3.4.1.3. C55A and the refolding pathway

3.4.1.3.1. Optimizing the refolding conditions

The available evidence suggests that the oxidized DTT mediated refolding of C55A is worth pursuing. I expect that the C55A-SG protein is sufficiently pure for Creighton-type refolding kinetics experiments which do not require the isolation of purified C55A-SH. For structural characterization and spin labeling, the yields may already be sufficient for working with non-isotope labeled proteins, but the procedures need to be optimized. I would begin again with analytical scale refolding reactions which facilitate the exploration of a variety of refolding conditions such as temperature, solvent composition, and protein concentration. Exclusion of oxygen from the refolding reactions should serve to minimize irreversible side reactions. Careful attention should be paid to the source of the precipitated material, which might be assayed by SDS-PAGE of the urea solubilized protein with and without reducing agents. At this point, it is not clear whether the precipitated protein is from non-native monomers (which are expected to have a lower solubility) or from covalent aggregates formed largely from air oxidation.

3.4.1.3.2. Equilibrium between {30-51, 14-38} (II_n), and {30-51, 5-14} plus {30-51, 5-38} (II_i)

The length of the preparative refolding reaction should also be optimized. The Creighton-type refolding kinetics experiments will provide some guidelines but it should be possible to obtain this information without the full Creighton kinetic treatment. It is difficult to estimate the folding times based on a consideration of the refolding pathway and analogies with well characterized mutants. The most reasonable pathway for folding of C55A to a native-like structure is via formation of the 14-38 disulfide directly from the 30-51 one disulfide intermediate to give the two disulfide species {30-51, 14-38}, which I will call II_n. Two other two disulfide intermediates {30-51; 5-14} and {30-51; 5-38} (collectively termed II_i) are also expected to form. For the purposes of calculating rate constants, Creighton assumes that the species II_i and II_n in wild type BPTI are in rapid equilibrium; there is no proof of this, and some of Creighton's papers even contradict this notion. For example, after 6 minutes of refolding of the completely reduced inhibitor with 60 mM oxidized DTT, "1.0% of the protein was eluted (from a CM-cellulose column) in the position expected of intermediates {30-51, 5-14} plus {30-51, 5-38} and 1.6% as {30-51, 14-38}" (Creighton, 1977b). Creighton concludes "the greater accumulation of intermediate (II_n) indicates that it may be formed directly more rapidly than the other intermediates (II_i), as was also suggested by the kinetics of disulphide bond formation in modified forms of the inhibitor, and that the rate of its rearrangement to the latter two species is not rapid relative to their interconversion to N_{SH}^{SH}" (Creighton, 1977b). However, he also states that the three species {30-51, 14-38}, {30-51, 5-14}, and {30-51, 5-38} "appear to be kinetically homogeneous, probably in rapid equilibrium" (Creighton and Goldenberg, 1984), and "the individual rates of their interconversion are unknown" (Creighton, 1977b). In the C55A mutant, the rapid interconversion of II_i to N_{SH}^{SH} is precluded, changing the time scale for rapid equilibration. Of course, formation of a third disulfide bond is also pre-

cluded, perhaps allowing accumulation of the two disulfide bond intermediates in the oxidized DTT refolding reaction, in contrast to what is observed with wild type BPTI. Therefore, the C55A mutant should provide an excellent opportunity to measure the equilibrium constant for the reaction $II_n \leftrightarrow II_i$. Since this is an intramolecular rearrangement, the equilibrium constant does not depend on the redox conditions, suggesting it may not be possible to isolate pure preparations of II_n. This provides an explanation for the isolation of only C55A-SG and not C55A-SH on the chymotrypsin column. The native-like free thiol form is in equilibrium with non-native species, which are probably digested by chymotrypsin. It also suggests that spin labeled derivatives must be prepared by trapping the correct species from a mixture.

3.4.1.3.3. Stability of II_i plus II_n relative to R

Collectively, the species II_i and II_n are 1.2 kcal less stable ($K_{eq} = 0.13$) than completely reduced BPTI at a ratio of oxidized to reduced DTT of 10^3 (they are also 1.3 kcal less stable than the one disulfide intermediates, which are all in rapid equilibrium)(Creighton and Goldenberg, 1984). Therefore on the order of only 10% of the molecules will have two disulfide bonds when refolded with oxidized DTT. This suggests that the precipitate observed upon concentration of the refolding reaction is accounted for by the less soluble completely reduced protein, or by side products due to irreversible air oxidation. Creighton has shown differential stabilization of the II_n species, either at a lower temperature or through the use of stabilizing additives such as Hofmeister salts (Creighton, 1980). Other known stabilizers of the native conformation of proteins such as sucrose or glycerol may be helpful (Timasheff and Arakawa, 1989).

3.4.1.4. Possibilities with C5A

It may be easier to isolate the native-like C5A-SH protein than the C55A-SH protein. In wild type BPTI, Cys-55 does not form disulfides with any of the other cysteines in the early two disulfide intermediates. It's only participation in the kinetically significant

pathway is in the rate-limiting intramolecular transition from II_i to N_{SH}^{SH}. The only significant two disulfide species that the C5A mutant is predicted to form is II_n, in other words, the desired species. Once the proper restriction sites are replaced at the 3' end of the wild type BPTI gene in pEZ:b3:wt, creating the C5A mutant via gene splicing as described above will be easily done.

3.4.2. C14A and C38A BPTI

3.4.2.1. C14A, C38A, and the folding pathway

Either Cys-14 or Cys-38 participate in the rate limiting rearrangement of the species II_i to N_{SH}^{SH} in both the folding and unfolding reactions of BPTI. This rearrangement demonstrates the great flexibility of selected sections of the peptide chain in these intermediates as the participating cysteines are approximately 22 Å apart in the native inhibitor (Creighton, 1977a). Nevertheless, the rearrangement is 10^3 times faster in the folding direction, presumably reflecting the more numerous favorable interactions which must be broken in the unfolding direction. As mentioned above, Creighton has prepared chemically modified proteins in which either Cys-14 or Cys-38 are irreversibly blocked and has demonstrated that both modified proteins are kinetically competent for folding, at approximately the same rate. However, these experiments were done under conditions where the formation of the second disulfide bond was rate limiting (40 mM oxidized DTT) masking the potentially two orders of magnitude difference in the rate of the intramolecular rearrangement. It may be possible to get a more accurate picture of this rearrangement by measuring the rates of reduction by DTT_{SH}^{SH} of the C14A and C38A mutants. As a cautionary note, the differences in rate may not be great, as the pseudo first order rate constant for the reduction of C14S/C38S at 10 mM DTT_{SH}^{SH} is only 10 fold slower than the the rearrangement of N_{SH}^{SH} to II (see figure 6 (Goldenberg, 1988)).

3.4.2.2. C14A and C38A as sites for non-perturbing probes

Because the disulfide bond between Cys-14 and Cys-38 is on the surface and its selective reduction does not interfere with the native structure or activity of the molecule (Liu and Meienhofer, 1968; Vincent and Lazdunski, 1972), attachment of a spin label at either of these cysteines is much less likely to introduce structural perturbations than at either Cys-5 or Cys-55. In addition, Creighton's data on the refolding of the hemi-alkylated Cys-14 and Cys-38 species suggests that neither the C14A nor the C38A mutants should cause problems with refolding akin to those observed in the C55A mutant (Creighton, 1977a). Mixed disulfide formation should not present problems because the glutathione can be selectively removed from the C38A-SG or C14A-SG proteins with DTT^{SH}_{SH} and the resulting protein undergoes rearrangements only very slowly (Creighton, 1977b). On the other hand, the mixed disulfide with glutathione adds a bulky group and an extra negative charge near the Lys-15 combining site with trypsin and chymotrypsin, requiring alterations to the fusion protein cleavage and affinity purification procedure. Apart from this potential complication, this wealth of well characterized chemistry prompted me to pursue the C38A mutant before attempting to work with Cys-plus mutants.

3.4.2.3. Creation of the C14A and C38A mutants by splicing

The mutants were made by swapping of the 579 bp *Not* I -*Sty* I fragment from pEZ:b3:wt and pEZ:b3:475 (see figure 3.2). The *Sty* I site spans the codons for A25-A27, making it very useful for the separation of amino acid 14 from amino acid 38 (and for 5 from 55, though for this construction, I chose the *Pst* I site instead). This procedure could not be used in the pHAZY vectors because there is a *Sty* I site in the *tet* gene of the pHAZY vector which was derived from pBR322. The vector fragments were treated with calf alkaline phosphatase for 30 minutes prior to gel purification in 0.8% low melting point agarose gels using sterile TAE buffer. The smaller inserts were purified in 1.6% gels. The appropriate bands were cut from the gels, the vector fragments were diluted 8 fold with sterile TE

buffer, and the inserts were diluted 4 fold prior to melting of the agarose at 68 °C for 5 minutes. Ligations were performed directly with these samples, without phenol extractions. The background frequency for the ligations was at most 0.35%.

Once the C38A mutant was made in the pEZ:BPTI vector, it was moved into the pHAZY vector as a *Stu* I - *Eco*R I fragment. The pEZ:C14A mutant could not be moved in this manner because the parent wild type vector lacked the *Eco*R I site. There are several alternative strategies for constructing this plasmid. In the first strategy, it would be necessary to open the pEZ:C14A plasmid with *Hind* III, fill in with Klenow, cut the linear plasmid with *Stu* I and ligate the blunt-ended insert into the *Stu* I - *Eco*R I (pol) large vector fragment of any of the pHAZY:BPTI plasmids. The blunt insert could go in in either orientation, and the correct orientation could be obtained by restriction screening of the plasmid DNA or by examination of the size of the protein produced. Alternatively, the linearized *Hind* III (pol) plasmid could be cut with *Bgl* II and moved into the *Bgl* II - *Eco*R I (pol) vector fragment of pHAZY:BPTI, removing the chymotrypsin cleavage site. In a second step, a *Stu* I - *Bam*H I fragment from this plasmid could be cloned in the correct orientation into any of the pHAZY:BPTI plasmids. These laborious procedures were not done, in anticipation that the pHAZY:wt sequence with the *Eco*R I site at the 3' end would be constructed. This remains to be done.

3.4.2.4. Expression of C38A

3.4.2.4.1. Refolding and activity of C38A

The TrpLE-Z-C38A protein was expressed in small scale experiments and refolded in a three hour reaction as described in the previous chapter. The protein was tested for DTT sensitivity as described in the C55A section. The untreated sample possessed approximately 4.7 mg/l of BPTI activity, while the DTT treated sample contained 13.2 mg/l activity (figure 3.7). These data suggest that the glutathione adduct is not an active trypsin inhibitor, and that 64% of the TrpLE-Z-C38A from the glutathione refolding reactions is pre-

sent as a mixed disulfide with glutathione. I expect that the relative amount of mixed disulfide protein will be dependent on the length of the incubation and the oxygen content of the solutions, which were not controlled here. The inactivity of the glutathione adduct is not surprising in light of the inactivity of the selectively reduced and carboxymethylated RCOM derivative (Vincent and Lazdunski, 1972).

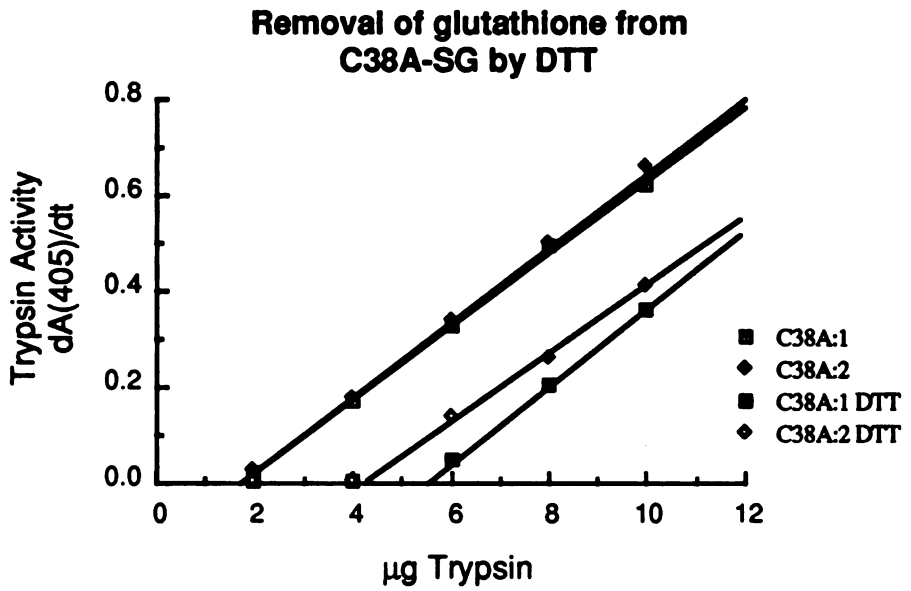


Figure 3.7: DTT treatment of the TrpLE:Z:C38A protein from the glutathione refolding reaction restores the inhibition activity of the protein.

Selective reduction of the C38A-SG mixed disulfide

In subsequent experiments, the selectively reduced C38A-SH molecule was blocked with iodoacetamide prior to assay. The selectively reduced and carboxyamidomethylated derivative of BPTI (RCAM) remains a stoichiometric inhibitor of trypsin under the conditions used here (Liu and Meienhofer, 1968).

3.4.2.4.2. Glutathione adduct prevents chymotrypsin binding of C38A-SG

In analogy with the lack of activity of the RCAM molecule with chymotrypsin (Vincent and Lazdunski, 1973) and the lack of activity of RCOM with trypsin and chy-

motrypsin (Vincent and Lazdunski, 1972), the glutathione adduct was not expected to inhibit chymotrypsin. This expectation was confirmed in chymotrypsin assays using N-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide as a substrate.

3.4.2.5. Approaches to cleavage of the TrpLE-Z-C38A-SG fusion protein

3.4.2.5.1. Rationale

Though the selectively reduced C38A-SH protein is expected to bind to chymotrypsin (Vincent and Lazdunski, 1973), I decided not to pursue this route because of anticipated problems with uncontrolable air oxidation. Instead, two approaches were chosen for the cleavage of the TrpLE-Z-C38A-SG fusion protein. In the first, the bulky and negatively charged glutathione group was replaced by the smallest, uncharged mixed disulfide possible, the -SMe group, which can be added with methyl methane thiosulfonate (MMTS), or Kenyon's reagent (Smith and Kenyon, 1974; Smith *et al.*, 1975; Kenyon and Bruice, 1977). It was initially hoped that this protecting group would not abolish the chymotrypsin inhibition activity of the protein, which could be cleaved and affinity purified on the chymotrypsin column as described. Though the observation that the RCAM protein is not a chymotrypsin inhibitor (Vincent and Lazdunski, 1973) suggested that the MMTS modification might not work in combination with chymotrypsin, the protecting groups are different enough in steric bulk to warrant the attempt. In the event that the C38A-SMe molecule was not a chymotrypsin inhibitor, a modification of this experiment was designed in with the fusion protein would be cleaved by chymotrypsin while immobilized on a trypsin column.

The second approach began with the hypothesis that while the C38A-SG protein is not a chymotrypsin inhibitor, the BPTI domain might be resistant to proteolysis, allowing cleavage of the fusion protein by enzymatic quantities of chymotrypsin in solution. The attractive feature of this approach is that many fewer steps are involved, affording a potentially greater yield. Subsequent purification is not expected to be a problem because the

BPTI domain does not measurably bind to the chymotrypsin and the protease, which is used in small quantities, can be removed by ion exchange chromatography.

3.4.2.5.2. Protection of C38A-SH with MMTS

C38A-SH reaction with MMTS: Problems of scale

The preparative scale refolding reactions of the TrpLE-Z-BPTI fusion proteins, beginning with 100 g of cell paste, are typically done in 2 liter volumes. These can be concentrated at least 10 fold by ultrafiltration on a YM-2 membrane, but this is a time consuming procedure which can take 2 days at room temperature, longer at 4 °C, and requires attention every several hours so that the reservoir does not go dry. The optimal design for the MMTS reaction is a multi-step procedure encompassing the following: 1) selective removal of the glutathione adduct from C38A-SG with DTT^{SH}_{SH}, 2) rapid removal of the DTT and protection of the free thiol by buffer exchange into 10 mM HCl on a G-25 column , 3) readjustment of the buffer pH and concomitant addition of the MMTS reagent, and 4) removal of the MMTS reagent by gel filtration. Obviously, the gel filtration steps are very inconvenient, even with volumes of 200 ml. Therefore, a compromise procedure was designed.

C38A-SH reaction with MMTS: Compromise procedure

The compromise procedure utilizes an excess of MMTS over all other thiols in the solution, including the reduced glutathione, the DTT^{SH}_{SH}, and the protein thiols. Reaction of MMTS with DTT^{SH}_{SH} will produce oxidized DTT and one equivalent of MeSH. A second equivalent of MMTS is needed to react with the MeSH. The large excess of MMTS is not optimal, due to the possibility of side reactions with other protein nucleophiles, especially amino groups. However, the presence of free thiols is even less desirable.

The refolding reactions were done with GSH and GSSG concentrations of 2.5 mM and 0.5 mM respectively, lower than is customary, in order to minimize the required concentration of MMTS. After refolding overnight at room temperature, all of the correctly re-

folded inhibitor molecules were present as mixed disulfides with glutathione. Solid DTT^{SH}_{SH} was added to a concentration of 2 mM to selectively reduce the mixed disulfide. After 30 minutes, neat liquid MMTS was added to a concentration of 8 mM. After an additional 30 minutes, the solutions were transferred to 3000 molecular weight cut off dialysis tubing, and were dialyzed at 4 °C against 4 changes of 50 liters of 50 mM Tris, 5 mM EDTA, pH 7.5 over the course of 3 days.

C38A-SMe is not an inhibitor of chymotrypsin

The -SMe protecting group creates a methionine isostere at Cys-14. Modification of Cys-14 and Cys-38 with iodoacetamide is known to inactivate the inhibitor towards chymotrypsin (Vincent and Lazdunski, 1973). The acetamide group contains a chain with one additional heavy atom and an extra branch compared to the -SMe group. Therefore, it was not possible to predict by analogy that the C38A-SMe protein is not a chymotrypsin inhibitor, but there was cause for suspicion. Assays showed that the molecule is not a chymotrypsin inhibitor, but that it does inhibit trypsin. SDS-PAGE analysis on 16% gels of the C38A-SMe protein incubated with chymotrypsin suggested that the protein is in fact efficiently cleaved by chymotrypsin, yielding peptides which are not fixed or stained in the gels.

Immobilization of the C38A-SMe on a trypsin column

The activity of C38A-SMe towards trypsin suggested a possible procedure for cleavage of the fusion protein by chymotrypsin at the designated site. Experiments were designed to test whether C38A-SMe could be bound to immobilized trypsin and then cut by chymotrypsin added to the column wash buffer. A conceptually similar experiment was done by (Dahlman *et al.*, 1989), who coincidentally cleaved a protein A fusion to the DNA binding domain of the glucocorticoid receptor with chymotrypsin while the fusion protein was bound to an immobilized IgG column. In their experiment, the desired protease resistant DNA binding domain was released by the chymotrypsin treatment, but in the TrpLE-Z-C38A-SMe experiment, the desired domain remains bound to the column.

The kinetics of the chymotrypsin cleavage of trypsin immobilized TrpLE-Z-C38A-SMe

In order to qualitatively test the kinetics of the chymotrypsin cleavage of the trypsin-Affigel 10 immobilized fusion proteins, 1 mg of TrpLE-Z-C38A-SMe activity (based on the wild type BPTI molecular weight) was bound to 0.4 ml of trypsin-Affigel 10 resin at 4 °C. Chymotrypsin at a concentration of 10 μg/ml in CWB was allowed to saturate the column, with a liquid head of approximately 1 ml. Incubation of the cleavage reactions for 48 hours on a gentle rocker, followed by elution of C38A-SMe at low pH, afforded approximately 2/3 full cleavage. SDS-PAGE and analytical Phenyl Superose hydrophobic interaction chromatography show the presence of anywhere from 4-6 additional cleavage sites in the Z domain or linker region.

Phenyl Superose clean up

The 58 amino acid long C38A-SMe was purified from the mixture described above on a Phenyl Superose 10/10 Hydrophobic Interaction Chromatography (HIC) column (Pharmacia). The mixture of properly and improperly cleaved inhibitors was dialyzed against 10 volumes 2 M $(NH_4)_2SO_4$, 50 mM TrisHCl, pH 8.0 (AST buffer). A small amount of precipitate was observed after the dialysis which was removed by centrifugation in a 50 ml Falcon tube at 7000 RPM for 10 minutes in a bench top Sorval centrifuge. The supernatant was then filtered through a 0.2 μm membrane. It was loaded onto the HIC column, previously equilibrated with AST buffer. Protein was eluted with a gradient of AST to 50 mM TrisHCl, pH 8.0, at a flow rate of 0.9 ml/min. The gradient ran from 0-30% Tris buffer in the first five minutes, followed by a 30-60% gradient over the next 40 minutes. Three major and several minor peaks were observed. The desired protein was found in the first peak.

3.4.2.5.3. Another procedure: cleavage of glutathione adduct (TrpLE-Z-C38A-SG) in solution

Rationale

The glutathione adduct, C38A-SG, is not a chymotrypsin inhibitor, but I hypothesized that it might still be resistant to the protease. If so, then it ought to be possible to cleave the fusion protein at the designed site by chymotrypsin in solution. I decided to use a molar ratio of 1:40, enzyme:fusion protein for all experiments. Experiments were then designed to test the stability of the BPTI domain towards the protease and the efficiency of cleavage at the designed site as a function of pH and temperature; these are the equivalent of Phase I and Phase II pharmaceutical trials.

Optimizing the conditions for the cleavage of TrpLE-Z-C38A-SG

Experiments were designed to test the solution phase cleavage as a function of temperature). An inhibitor concentration of 43 μM (or 0.28 mg/ml, based on the molecular weight of wild type BPTI) was used in combination with a chymotrypsin concentration of 1 μM (or 1/43 equivalent, based on the weight of lyophilized chymotrypsin) in a volume of 1 ml. The buffer was 100 mM TrisHCl and 10 mM $CaCl_2$. Samples with chymotrypsin were incubated at 4, 25, and 37 °C, and a control reaction without chymotrypsin was incubated at 25 °C. Aliquots of 100 μl were withdrawn at selected time intervals, added to 1 μl 3 mM TPCK (L-1-*p*-Tosylamino-2-phenyl-ethyl chloromethyl ketone) in 100% methanol (in an attempt to irreversibly inactivate the chymotrypsin—it didn't completely work, for an undetermined reason) and incubated at room temperature for 30 minutes. A 2 μl aliquot of 0.5 M DTT^{SH}_{SH} was then added to selectively reduce the mixed disulfide and the samples were incubated at room temperature for an additional 30 minutes. The reduction was quenched by the addition of 25 μl 0.5 M IAM (freshly prepared in 0.1 M TrisHCl, pH 8.0). After 2 minutes at room temperature, the samples were moved to dry ice and then stored in a -80°C freezer until assayed. The time intervals were 0, 1, 2, 5, 8, and 24 hours.

Trypsin inhibition assays were performed on 100 µl aliquots of samples. The concentration of C38A-RCAM in the samples was designed to inhibit 8.5 µg of trypsin. The control samples inhibited 10 µg trypsin by weight, indicating that the trypsin stock did not represent completely active enzyme. All of the samples from the time points at 4 and 25°C showed complete inhibition of up to 10 µg trypsin. The 37°C samples showed a time dependent degradation of inhibitor activity, as shown below.

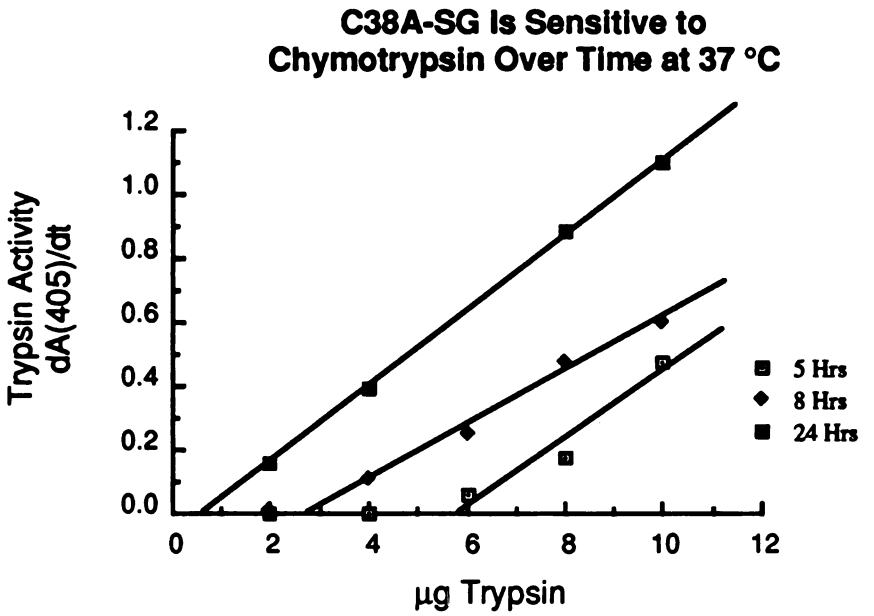


Figure 3.8: The TrpLE-Z-C38A-SG fusion protein glutathione adduct is digested over time by chymotrypsin. The x-intercept decreases for samples from the chymotrypsin digest as a function of time, suggesting proteolysis of the inhibitor domain.

The apparent site of chymotrypsin cleavage was assayed by electrophoresis of the samples through 12% native gels. The 4 °C samples showed only approximately 25% cleavage at the designed site, even after 24 hours. After 8 hours, the 25°C reactions showed approximately 1/2 of the molecules were cleaved at the designed site, while cleavage at the Try(-1)-Arg(1) peptide bond was nearly complete after 24 hours. The 37°C samples showed nearly complete cleavage at the designed site after 5 hours, but at this

time, measurable degradation of the inhibitor activity was observed. I concluded that the optimal cleavage condition at pH 8.0 is 25°C over a 24 hour period.

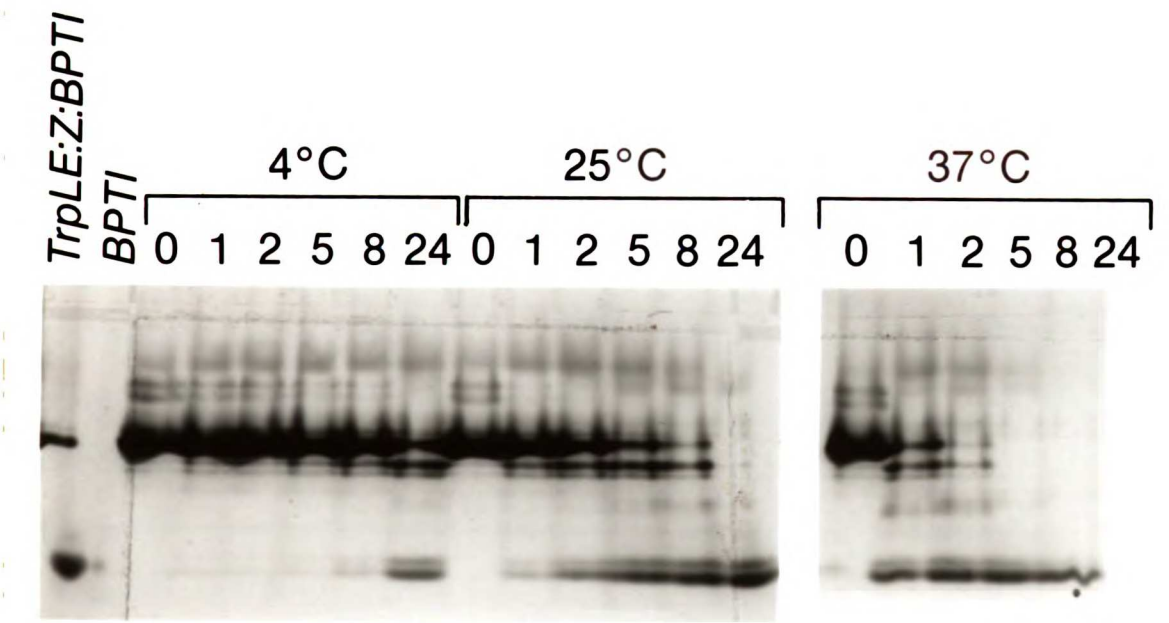


Figure 3.9: The TrpLE:Z:C38A-SG protein is cleaved at the Tyr(-1)—Arg(1) bond by chymotrypsin in solution. The gels are 12% native gels. Two standard lanes contain the full length fusion protein and wild type BPTI (Traysylol). The additional lanes are coded in terms of temperature and time of reaction, measured in hours.

Another set of experiments was designed to test the effect of pH on the safety and efficacy of the chymotrypsin cleavage reaction at 25 °C. The concentration of the fusion protein in this set of trial reactions was 0.714 mg/ml, based on activity assays and the molecular weight of wild type BPTI. Chymotrypsin was used at 68.7 µg/ml, or 1/40 molar equivalent. The residual inhibitor activity remaining in the reactions was followed over a period of 21 hours at three pH values: 7.48, 7.95, and 8.44. The buffer was 100 mM

TrisHCl and 10 mM $CaCl_2$ and the initial volume of the reactions was 0.5 ml. At several time points (0, 2, and 21 hours), 98 µl aliquots were withdrawn and placed on dry ice to stop the reaction. After the last time point, the aliquots were thawed and selectively reduced by the addition of 2 µl 0.5 M DTT^{SH}_{SH} (10 mM final concentration). After 30 minutes at room temperature, the reduction was quenched by the addition of 25 µl 0.5 M iodoacetamide (IAM), incubated for two minutes at room temperature and placed on ice till assay and electrophoresis.

No measurable degradation of the inhibitor activity remained after 21 hours at pH 7.48 and 7.95; approximately 50-60% of the inhibitor activity remained after 21 hours at pH 8.44 (data not shown). These results indicate that the stability of the C38A-SG domain is good, but not absolute, and is sensitive to pH. The result cannot be accounted for solely by the variation of chymotrypsin activity as a function of pH; the maximum of k_{cat}/K_M for chymotrypsin occurs at pH 7.8 (Fersht, 1985) (Fersht and Renard, 1974). The electrophoresis results show that complete cleavage occurred at the Tyr(-1)—Arg(1) site after 21 hours, but not after 2 hours, in all three samples.

Preparative scale cleavage of TrpLE:Z:C38A-SG

Preparative scale cleavages of the TrpLE:Z:C38A-SG fusion protein were done at a protein concentration of 114 µM, equivalent to 0.745 mg/ml wild type BPTI, in 20 or 100 ml; chymotrypsin was used at 2.8 µM, or 71 µg/ml. Further reaction conditions were 100 mM TrisHCl, pH 8.0, 10 mM $CaCl_2$, at a temperature of 25 °C. After 21 hours, trypsin activity assays of the selectively reduced and carboxyamidomethylated protein showed no loss of inhibitor activity. Further analysis by 12% native gel electrophoresis showed that approximately 90% of the fusion protein was cleaved at the Tyr(-1)-Arg(1) peptide bond.

The intact C38A-SG domain was purified by cation exchange chromatography on either a MonoS 10/10 column or a 2.5 cm diameter, 70 ml bed volume Fast Flow S-Sepharose column (packed in an OMNI column), attached to a Pharmacia FPLC. The cleavage reaction mixture was diluted with 1.5 volumes H_2O and brought to 50 mM MES

by the addition of the solid buffer. The measured pH of the solution was 5.78, fortuitously matching the pH of the column running buffers. The A buffer was 50 mM MES, pH 5.8, while the B buffer was buffer A plus 1 M NaCl. The column flow rate when using the MonoS column was 2.0 ml/minute, with a gradient of 0-12.5% B over the first 5 minutes, followed by 12.5-28.1% B over the next 25 minutes. The single major peak represented 88 percent of the total OD(280) eluting from the column. A similar gradient was calculated for the Fast Flow column: at a flow rate of 6 ml/min, the gradient was 0-12.5% B over 16.7 minutes, then 12.5-37.5% B over the next 133.3 minutes. Analysis of Coomassie stained native gel analysis suggested a contaminant level of approximately 5%, mostly from the full length fusion protein.

3.4.2.6. Modification of C38A-SH with a nitroxide spin label

The purified C38A-SG domain was prepared for modification with the 2,2,6,6-tetramethyl-4-(bromoacetamido)piperidine-1-oxyl (BrAmTempo) spin label by selective reduction of the mixed disulfide between Cys-14 and glutathione. C38A-SG at a concentration of 1.06 mM in 10 ml of 100 mM TrisHCl, 1 mM EDTA, pH 8.0 was reduced by the addition of 1 ml 400 mM DTT, followed by incubation at room temperature for 15 minutes. The selectively reduced inhibitor was isolated by buffer exchange into 10 mM HCl on an 80 ml G-25 column (2.5 cm diameter). The conditions for the reaction with BrAmTempo were: 0.158 mM protein (corresponding to 1.03 mg/ml), 0.2 mM BrAmTempo (1.26 equivalents) delivered from a 20 mM stock solution in isopropanol, 100 mM Tris HCl, 1 mM EDTA, pH 8.7 in a total volume of 50 ml. The reaction was carried out at room temperature in a foil wrapped flask. After 135 minutes, no free thiol was detectable with 5,5'-dithionitrobenzoic acid (DTNB) (Ellman, 1959; Creighton, 1989).

The possibility of side reactions with reactive bromine compounds or modification of the protein by the bromoacetamide derivative at sites other than Cys-14 makes it desirable to remove the BrAmTempo reagent and byproducts as quickly as possible, prohibiting

the use of dialysis. Due to the large volume of the reaction, gel filtration was not convenient (more concentrated reaction conditions might cause side reactions between the free thiol and the nitroxide). Instead, the residual BrAmTempo reagent and undesired reaction products were removed by cation exchange chromatography on a Fast Flow S-Sepharose column; the positively charged protein should stick to the column, but the uncharged BrAmTempo reagent and negatively charged byproducts should not bind. The reaction mixture was diluted 5 fold with H_2O and brought to 50 mM MES by the addition of the solid buffer. The unadjusted pH was 5.64 and the solution was still kept in the dark.

The solution was loaded onto the Fast Flow S-Sepharose column (2.5 cm diameter, 80 ml total volume) with the aid of a Rainin Rabbit Peristaltic Pump at a flow rate of approximately 8 ml/minute. The total loading time was approximately 30 minutes. Following loading, the column was connected to a FPLC and washed with approximately 100 ml of 50 mM MES, pH 5.8 buffer (buffer A). Elution was carried out with a simple linear gradient of 0-50% buffer B (buffer A+ 1M NaCl) over 60 minutes at a flow rate of 6 ml/minute. Fractions from the single major peak were pooled and concentrated to a total volume of 10 ml by ultrafiltration on a YM-1 membrane. The absorbance of the concentrated protein was 5.307, with an A_{278}/A_{251} ratio of 1.55.

The C38A-AmTempo molecule was further purified by HIC on a Phenyl Superose column using 50 mM Tris, 2M $(NH_4)_2SO_4$, pH 8.0 as HIC buffer A and 50 mM Tris, pH 8.0 as HIC buffer B. Concentrated stock solutions of Tris buffer sometimes contained significant amounts of UV absorbing material, so the chromatography buffers were always prepared fresh from UltraPure Tris base solid. The concentrated C38A-AmTempo sample from the previous ion exchange purification was diluted 5 fold into HIC buffer A and filtered through a 0.45 μm membrane and purified on the column in 3 separate runs. Following loading, the protein was eluted with a steep gradient of 0-40% B over 5 minutes, followed by a shallow gradient of 40-60% B over a period of 45 minutes, at a flow rate of 1.0 ml/minute. C38A-AmTempo eluted as the major peak at 31.7 minutes after the

beginning of the steep gradient. The early fractions of this peak were contaminated with protein from an overlapping peak and were therefore pooled and rechromatographed. The latter fractions were free of contaminants detectable by HPLC on a C18 column. A manually collected peak off the C18 column was dried down and submitted for mass spectral analysis on an electrospray mass spectrometer. The calculated molecular weight for the C38A-AmTempo protein is 6692.75. The observed weight in the mass spectrometer was 6691.99 ± 0.88 (s.d.) (figure 3.10).

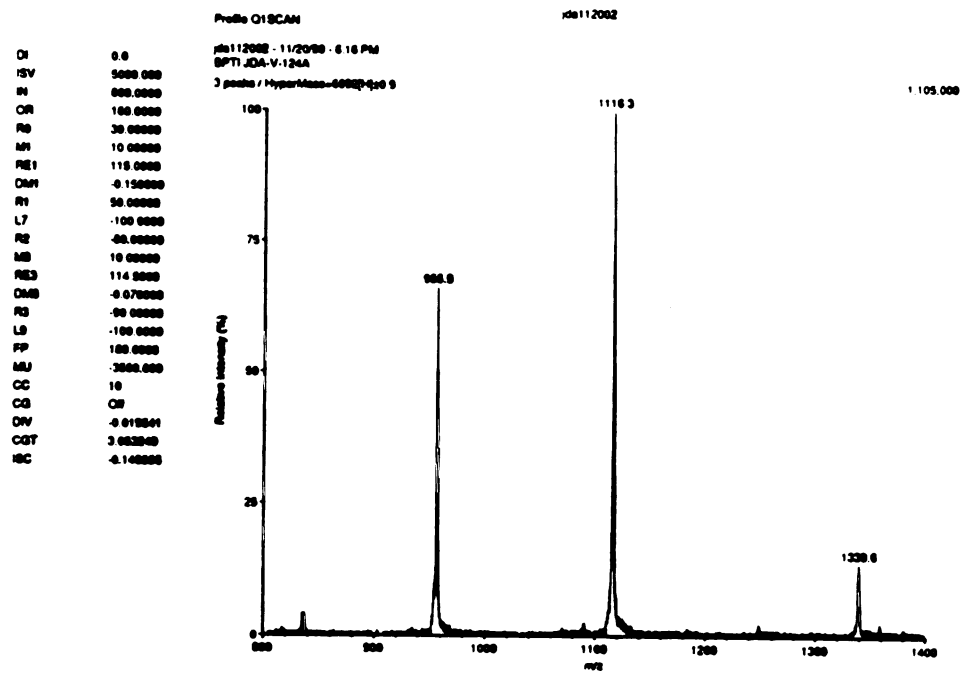


Figure 3.10: The mass spectrum of the C38A-Am-Tempo protein indicates a molecular weight of 6691.99 ± 0.88 (s.d.). The calculated molecular weight is 6692.75.
Methods: Electrospray mass spectra were obtained using a Sciex API-III triple quadrupole mass spectrometer equipped with the Ionspray interface operating at 5 kV and calibrated with a solution of polypropylene glycols. Each spectrum represents the summation of several scans of the first quadrupole acquired while the sample was infused at a rate of 1 - 2 ul/min from an acetonitrile / acetic acid solution (1 - 10 pmoles/ul).

The pooled fractions from the HIC purification of C38A-AmTempo were concentrated by ultrafiltration on a YM-1 membrane to approximately 6 ml. The buffer was exchanged for 0.5% ammonium bicarbonate on a G-25 desalting column and the protein containing fractions were pooled and lyophilized prior to storage.

An EPR spectrum of a 1.5 mM sample of C38A-AmTempo in a 25 µl capillary was taken at room temperature (figure 3.11). The spectrum is typical of nitroxide samples with correlation times in the range expected for the overall tumbling of a protein the size of BPTI.

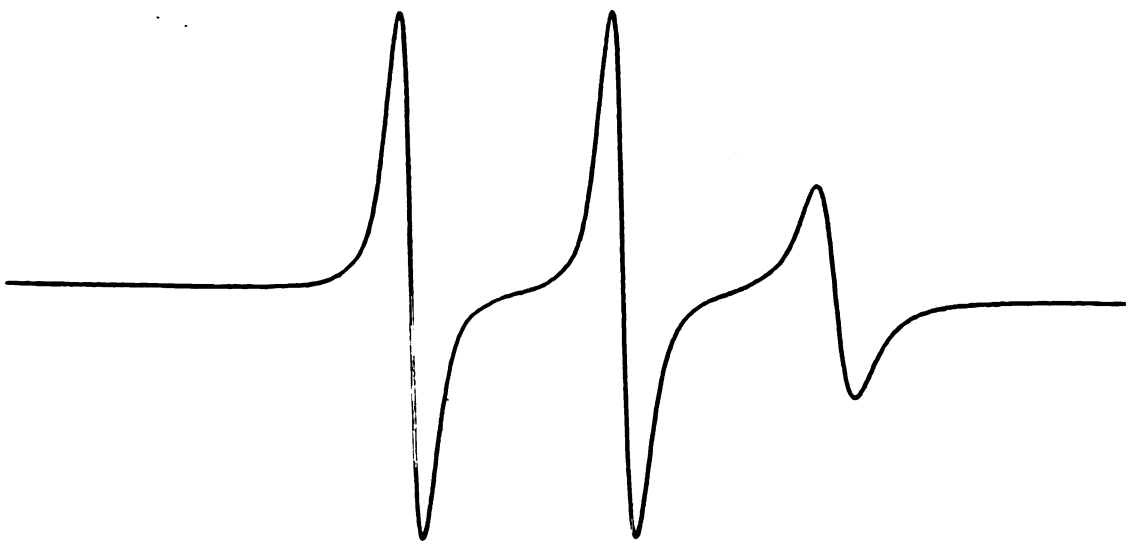


Figure 3.11: The EPR spectrum of the C38A-AmTempo protein at room temperature.

3.5. Conclusions

The C38A mutant of BPTI has been constructed and expressed in high yield in the pHAZY BPTI expression system. When refolded, the protein forms the native 30-51 and 5-55 disulfides, while the remaining Cys-14 thiol is commonly isolated as a mixed disulfide with glutathione. In contrast to most of the mutants discussed in chapter 2, the C38A-SG domain is not a chymotrypsin inhibitor and its fusion to the protein A domain cannot be cleaved on the chymotrypsin column. However, the TrpLE-Z-BPTI(C38A-SG) fusion protein is efficiently cleaved at the site designed in the construction of the pHAZY vector by enzymatic quantities of chymotrypsin in solution. The glutathione moiety can be removed by reaction with DTT, leaving the two remaining protein disulfides intact. The liberated Cys-14 thiol reacts selectively and stoichiometrically with a bromoacetamide TEMPO derivative. The high yields provided by these procedures satisfy the prerequisites for an experiment combining spin labeling with isotope labeling. These experiments have not yet been carried out, but in the next chapter, the spin labeled BPTI-C38A derivative (C38A-AmTempo) will be analyzed by homonuclear two-dimensional ^{1}H NMR.

Appendix 3A

3.1. Recommendation: In the future, use the (1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methanethiolsulfonate.

I have not pursued experiments with this compound, but I strongly recommend that future labeling of cysteines in BPTI be done with this compound. There are three important advantages to the use of the methane thiolsulfonate compound compared to the bromoacetamide derivatives: (1) the protein modification reaction is even more specific, (2) the modification is readily reversible, allowing recycling of the protein—even after reduction of the nitroxide moiety with ascorbate—by removal of the label with DTT$^{SH}_{SH}$, and (3) unlike the BrAmTempo molecule the methane thiolsulfonate derivative is not a mixture of enantiomers and therefore will not give a mixture of protein diastereomers. Finally, the single step synthesis of the compound from commercially available starting materials is exceedingly simple and provides a 70% yield of pure product (Berliner *et al.*, 1982). There may even be a commercial supplier of the compound in Hungary (Reanal) (Todd and Millhauser, 1991).

Chapter 4

NMR Results from C38A-Am-Tempo

In this chapter, I will focus on the results from NMR experiments with the C38A-Am-Tempo spin labeled protein, whose production I have described in the previous chapter. The goals of these experiments were twofold. First, I sought to determine whether this derivative is appropriate for the detailed analysis of quantitative distance information derived from paramagnetic broadening data. The primary test here is whether a single structural model can account for all of the spin label data. Not all spin label derivatives pass this test; indeed, two out of the three previously studied spin label derivatives of BPTI were rejected from further analysis for this reason (Kuntz, personal communication). Second, I sought to assess the contribution of paramagnetic broadening data to the precision of protein structure determination in a general and practical sense with real experimental data. Of course, the second goal depends on the success of the first. As I shall show, the C38A-Am-Tempo derivative does not yield data which are consistent with a single structural model, so the second question remains open. In the course of this work I have developed some opinions concerning the future application of the spin label technique, and these will be presented towards the end of the chapter.

The experimental work described in this chapter is divided into two major sections: (1) the assignment of a subset of the resonances from C38A-Am-Tempo and (2) the acquisition and analysis of the paramagnetic broadening data. All work on solution structure determinations by NMR spectroscopy is dependent on the assignment of a majority of the proton resonances of the molecule. For this reason, I will begin with a discussion of the assignment of C38A-Am-Tempo.

4.1. Assignment of the proton spectrum of C38A-Am-Tempo

4.2. Choice of conditions

A common set of considerations applies to the choice of sample conditions for most protein NMR experiments. These include the protein concentration, the effects of pH on hydrogen exchange (Molday *et al.*, 1972), salt concentration, and the effects of

temperature on correlation times and protein stability. Experiments with spin labeled proteins introduce a new set of constraints, particularly with respect to protein concentration. Whereas it is common to use protein concentrations as high as 20 mM for diamagnetic samples (the actual choice will depend to a large degree on the solubility of the protein), for spin labeled samples it is important to choose conditions that will minimize intermolecular interactions. This requires low protein concentrations—a practical limit, based on the low sensitivity of the NMR experiment itself, is 1-2 mM—and sometimes extreme pH and temperature values. The consideration of protein stability will of course limit the range of pH and temperature.

The previous studies of spin labeled BPTI derivatives were carried out at a concentration of 2 mM at pH 2.0 and 68°C. The low pH ensured that most carboxylate groups were protonated, increasing the positive charge on the molecule to maximize electrostatic repulsion. The high temperature was chosen to further disrupt intermolecular aggregates. All of these derivatives contained the three native disulfide bonds present in wild type BPTI. In contrast, the C38A-Am-Tempo derivative contains only two of the native disulfide bonds, and it is therefore expected to be less stable. The stabilities of a number of BPTI derivatives with the 14-38 disulfide bond selectively reduced and alkylated have been studied by calorimetry (Schwarz *et al.*, 1987) and by optical rotatory dispersion (Vincent *et al.*, 1971). Both studies indicate that BPTI with the 14-38 disulfide bond broken is largely unfolded at pH 2.0 and 68°C; this observation was independent of the choice of cysteine blocking reagent. It was therefore necessary to choose other, more suitable conditions for the experiments.

To facilitate the resonance assignment process, I decided to choose conditions where assignments are available for BPTI or any one of a number of derivatives. A list of conditions or derivatives, or both, where nearly full assignments are available is presented in Table 4.1.

Table 4.1

^1H chemical shift assignments of BPTI and derivatives—a survey[§].

Derivative	Conditions		Reference
	pH	Temp	
Wild Type	4.6	68	(Wagner and Wüthrich, 1982b)
Wild Type	4.6	36	(Wagner *et al.*, 1987), Eads[#]
Wild Type	2.0	68	(Kosen *et al.*, 1986), Kosen[#]
RCAM[†]	4.6	68	(Stassinopoulou *et al.*, 1984)
TRAM[‡]	4.6	68	(Stassinopoulou *et al.*, 1984)
Circular BPTI	4.6	36	(Chazin *et al.*, 1985)
Met-Ox (R & S)[*]	2.0	68	Kosen[#]
C14A/C38A	4.6	36	Naderi[#]
C30A/C51A	4.6	36	Eads[#]
C30V/C51A	4.6	36	Hurle[#]
C30T/C51A	4.6	36	Hurle[#]

[§] Additional assignments are presented for the Asn-44 side chain amide protons and the Gly-37 amide proton in (Tüchsen and Woodward, 1987) ; the aromatic ring protons for Phe-22 and Phe-33 were reassigned in (Wagner *et al.*, 1987).
[†] BPTI with the 14-38 disulfide bond reduced and the free cysteines carboxyamidomethylated
[‡] BPTI modified by transamination of the a-amino group of Arg-1
[*] BPTI with Met-52 oxidized to the sulfoxide level. Two diastereomers are produced, and can be separately assigned.
[#]Unpublished

From this list, I have chosen to use pH 4.6 and 36°C for the experiments with C38A-Am-Tempo. Since the 1987 paper from the Wüthrich lab (Wagner *et al.*, 1987), these have become "standard BPTI conditions" in the Kuntz lab. The availability of assignments of the structurally similar C14A/C38A mutant also made this an attractive choice. Finally, the previously cited work on the stability of derivatives lacking the 14-38 bond suggested that these conditions are well below the transition for the thermal unfolding of the protein.

4.3. NMR methodology

4.3.1. Sample preparation

Slightly acidic solutions of the C38A-Am-Tempo protein in H_2O were prepared from pure lyophilized stocks of the protein. The concentration of these stock solutions

was measured by UV spectroscopy of a 1:10 dilution, using an extinction coefficient at 280 nm of 5900 L cm^{-1} mol^{-1}, approximately the sum of the extinction coefficients for BPTI and the nitroxide (Kassell, 1970; Kosen *et al.*, 1986). Measured aliquots were removed from the stock solutions, such that the aliquot contained precisely the quantity of protein needed to prepare a 0.5 ml solution of protein at a specified concentration. When the desired volume was greater than 0.45 ml, the protein was lyophilized in small test tubes and redissolved in 0.45 ml 90% H_2O/10% D_2O. The pH of these solutions was adjusted by the addition of small measured aliquots of either 1 M HCl or 1 M NaOH; the measured pH was not corrected for the isotope effect arising from the 10% D_2O. The solutions were then made up to 0.5 ml with H_2O and spun in a microcentrifuge. Finally, 0.40 ml of these solutions was transferred to nitric acid washed NMR tubes. The sample concentrations were 3, 2, and 0.6 mM.

The paramagnetic nitroxide samples were reduced by the addition of small aliquots of a concentrated ascorbate solution. Ascorbate solutions were freshly prepared in H_2O at 100 times the concentration of the protein samples; the pH of these solutions was adjusted to 4.6 prior to bringing them up to the final volume. Two equivalents of ascorbate were added in a total volume of 8 μl; this limited the dilution of the protein sample to 2%.

NMR experiments with the paramagnetic and corresponding reduced diamagnetic samples were performed in succession to minimize differences in the settings of the magnet. When the paramagnetic sample was removed from the magnet, it was replaced with a D_2O sample to maintain the field lock. Further precautions were taken during the reduction procedure to minimize the change in volume of the sample in the probe and therefore necessary changes in shim settings. The paramagnetic sample was not removed from the sample spinner. Instead, the ascorbate was added directly to the tube while it was still in the spinner and the sample was mixed either by careful inversion of the resealed tube or by repeated pipetting through a capillary tube. Immediately after mixing,

the sample was again placed in the magnet, the field was locked, and the reduction reaction was allowed to proceed for 20-30 minutes. In most cases, it was necessary to touch up the lower order spinning shim settings.

Absolute value COSY spectra of oxidized and reduced samples were taken at 0.6, 2, and 3 mM protein concentrations. For assignment purposes, NOESY and HOHAHA data were acquired from the reduced 3 mM sample. The two dimensional experiments are described in detail below.

4.3.2. Setup of COSY, NOESY, and HOHAHA

COSY – Normally, DQF-COSY spectra (Rance *et al.*, 1983) would be used for assignment purposes instead of the absolute value COSY (Aue *et al.*, 1976; Kumar *et al.*, 1980b), but because the absolute value spectra were to be used for intensity comparisons of the oxidized and reduced sample (Weiss *et al.*, 1984; Kosen *et al.*, 1986), they were also used for assignments. Seven second relaxation delays were used to allow complete protein proton relaxation prior to the acquisition of each FID. The intense H_2O solvent signal was suppressed by presaturation with a low power coherent transmitter pulse during this delay (Zuiderweg *et al.*, 1986). Solvent suppression by preirradiation was not possible with t_1 increments greater than 42.6–56.8 ms—most likely due to paramagnetic relaxation of the solvent protons—preventing the acquisition of more than 300–400 t_1 slices with the paramagnetic sample. The same number of slices was also used for the reduced spin label sample.

HOHAHA – Two-dimensional Homonuclear Hartmann-Hahn (HOHAHA) spectra (Bax and Davis, 1985) were collected from the 3 mM sample of the diamagnetic C38A-Am-Tempo protein for assignment purposes. Two complete data sets with mixing times of 75 and 100 ms were collected sequentially, and the FIDs were added before processing. A 6-W amplifier (Astron Corp., CA) with a transmitter γB_1 of 8 kHz was used, providing constant power throughout the mixing time and the same power for the initial

90° pulse (Basus *et al.*, 1988). The trim pulses were 2.5 ms in duration. A 3 s, on resonance, DANTE pulse train (Morris and Freeman, 1978) was used to saturate the solvent water signal. The total relaxation delay, including the acquisition time, was 3.582 s. Phase sensitive data with quadrature detection in both dimensions was collected using the hypercomplex method of States *et al.* (States *et al.*, 1982). A total of 512 t_1 increments was collected, with 4 k complex points in each block of the data set. The data were processed with the Lorentz-Gauss transformation (Derome, 1987), using the parameters -7 and 0.2 in ω_2 and -7 and 0.1 in ω_1. The data were zero filled one time in each dimension, to give a final data size of 1 k × 4 k real data points. The spectral width in each dimension was 7042.25 Hz.

NOESY – Phase sensitive NOESY spectra (Jeener *et al.*, 1979; Macura and Ernst, 1980) with a mixing time of 160 ms were collected from the 3 mM diamagnetic C38A-Am-Tempo sample for assignment purposes. Quadrature detection in both dimensions was achieved by the method of States *et al.* (States *et al.*, 1982). The solvent signal was suppressed by a low power pulses with the transmitter followed by 5 ms electronic switching delays (Zuiderweg *et al.*, 1986) during the relaxation delay (3 s pulse) and the mixing time (150 ms pulse, sandwiched between two 5 ms delays). The total relaxation delay was 3.587 s, which includes the 0.582 s acquisition time. Because the spectrum was collected for qualitative assignment purposes only, the relaxation delay was kept relatively short, and no effort was made to remove the cross peaks due to zero quantum coherence (Otting *et al.*, 1990). The data were processed with with the Lorentz-Gauss transformation (Derome, 1987), using the parameters -7 and 0.2 in ω_2 and -7 and 0.1 in ω_1; the data were zero filled one time in each dimension.

4.4. Partial chemical shift assignments of C38A-AmTempo

The analysis of the paramagnetic relaxation data was restricted to the well resolved fingerprint region of the two-dimensional NMR spectrum (Kosen *et al.*, 1986).

While it is certainly possible to analyze other regions of the spectrum, the fingerprint should give a good first pass view of the quality of data available from the spin label. For this reason, assignment of the proton spectrum of C38A-Am-Tempo concentrated on the backbone α and amide protons, with β and other side chain protons included when needed to confirm the sequential assignment. All resonance assignments were made using spectra of the reduced protein.

4.4.1. Rigorous assignment by HOHAHA and NOESY

Many of the chemical shift assignments for C38A-Am-Tempo could be made by direct comparison of its spectra to spectra from C14A/C38A and wild type BPTI. Because these proteins are all similar in structure, their NMR "fingerprints" are also similar, allowing assignment by simple comparison. Ultimately, the sequential assignment procedure developed by Wüthrich and colleagues provides the most certain assignment of homonuclear two-dimensional protein NMR spectra (Billeter *et al.*, 1982; Wüthrich *et al.*, 1982). The now standard procedure has been well described by Wüthrich (Wüthrich, 1986) and others (Basus, 1989), so only an abbreviated outline will be presented here, primarily for the purposes of highlighting small deviations from the procedure that I followed. Briefly, the procedure starts with the identification in J-correlated spectra (COSY or the HOHAHA relay experiment) of amino acid spin systems; all such spin systems are bounded by the peptide bonds between residues. Sequential neighbors are then identified in NOESY spectra by searching for cross peaks between the amide proton of the i+1'th residue to the amide, α, or β protons of the i'th residue; the short hand notations for these connectivities are d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$, respectively (Stassinopoulou *et al.*, 1984). At least two of these connectivities are required for a 90-99% confident assignment (Billeter *et al.*, 1982). Finally, the sequentially connected main chain (NH, α, and β) protons are matched with their spin system assignments, and the protein primary sequence is searched for unique combinations.

The procedure is substantially simplified for the assignment of the spectra of closely related protein analogs (Stassinopoulou *et al.*, 1984). Most of the cross peaks for corresponding atoms occur in the same region of the spectrum, with chemical shift differences limited to less than ±0.05 ppm (Hurle *et al.*, 1991; Naderi *et al.*, 1991). The chemical shift difference can be significantly larger for protons near the site of mutation or modification, or both. I chose to start with confidently assigned "outlier" fingerprint cross peaks such those for Y23, K46, Y21, and C30 and to work my way both forwards and backwards in sequence to obtain the sequential assignment. Most of the NOESY cross peaks were found in the spectral regions expected based upon the assignments of the C14A/C38A mutant and the RCAM derivative (see Table 4.1). The spin systems of the sequentially connected residues were confirmed by inspection of COSY and especially HOHAHA spectra. All assignments were carried out manually based on visual inspection of the spectra with the aid of the computer program SPARKY, developed by Dr. Don Kneller of the Kuntz lab. The list of partial assignments is given in Table 4.2.

Table 4.2

Chemical shifts of the assigned ^1H NMR lines of C38A-Am-Tempo BPTI, pH 4.6, 36°C

Residue		NH	$C^{\alpha}H$	$C^{\beta}H$‡	$C^{\gamma}H$‡	$C^{\delta}H$‡
Arg	1					
Pro	2		4.33	2.05 0.93		
Asp	3	8.66	4.26	2.77		
Phe	4	7.83	4.61	3.36 2.96		
Cys	5	7.44	4.36	2.87 2.75		
Leu	6	7.58	4.50	1.86	1.71 0.97	0.87
Glu	7	7.52	4.61	2.27 2.19	2.59	
Pro	8					
Pro	9		3.78*	0.18 0.14*		
Tyr	10	7.95†#	4.86*	2.97		
Thr	11	8.76†	4.53	4.10	1.39	
Gly	12	7.10*	4.14† 3.36†			
Pro	13					
Xxx	14					
Lys	15					
Ala	16	8.08†	4.28	1.18		
Arg	17	8.16*	4.31	1.62	1.50 1.31	
Ile	18	8.31†#	4.29†	1.85 1.34		
				0.99		
Ile	19	8.65	4.38*#	1.96 1.47	1.39	
					0.73	
Arg	20	8.39		1.85* 0.88*		
Tyr	21	9.19	5.70	2.71		
Phe	22	9.77	5.28	2.91 2.82		
Tyr	23	10.54	4.31	3.46 2.75		
Asn	24	7.79	4.61	2.87 2.18		
Ala	25	8.75	3.77	1.57		
Lys	26	7.92	4.08	1.90	1.74 1.46	
Ala	27	6.82	4.30	1.19		
Gly	28	8.13	3.92 3.62†			
Leu	29	6.82	4.75	1.73 1.44		
Cys	30	8.42	5.60	3.67 2.67		
Gln	31	8.77	4.84	2.17# 1.73	2.24# 1.90	
Thr	32	8.05	5.27	4.04	0.59	
Phe	33	9.38	4.89	3.12* 2.95		
Val	34	8.34	3.93	1.95 0.83	0.70	
Tyr	35	9.30*	4.82*	2.65 2.50		
Gly	36	8.42†	3.67† §3.59†#			
Gly	37					
Ala	*38*	*7.83#*	*4.18†§*	*1.35*		
Arg	39	8.48†	4.30†#			
Ala	40	7.92†#	4.60†#			
Lys	*41*	*8.42†#*	*4.45†*	*2.21 1.65§*		
Arg	42	8.31#	3.67	1.04* 0.44	1.48 1.22	
Asn	43	7.22	5.04	3.30*		
Asn	44	6.79	4.91	2.75 2.55*		
Phe	45	9.92	5.14	3.42 2.79		
Lys	46	9.92	4.40	2.10 2.00	1.62 1.50*	
Ser	47	7.47	4.54	4.13 3.88		
Ala	48	8.15	3.16	1.05		
Glu	49	8.60#	3.87	2.02 1.85	2.34 2.18#	
Asp	50	7.86	4.29	2.88 2.73		
Cys	51	7.00	1.71	3.18 2.89*		
Met	52	8.59	4.17	2.07 1.97	2.70	
Arg	53	8.28	3.99	1.93 1.87		3.23
Thr	54	7.41	4.08	4.00	1.61	
Cys	55	8.23	4.63*	2.24 2.01		
Gly	56	7.95	3.86			
Gly	57	8.19	3.98 3.81			
Ala	58	7.94	4.01	1.31		

Table 4.2 (cont.): Legend

Chemical shifts are quoted relative to internal sodium 3-trimethylsilyl-(2,2,3,3-$^2H_4$)-propionate (TSP)
‡ Whenever only one number is given for a methylene group, it could not be established if the observed resonance corresponds to one or both protons.
* These chemical shifts differ by 0.05-0.10 ppm from those of the corresponding protons in wild type BPTI.
† These chemical shifts differ by greater than 0.10 ppm from those of the corresponding protons in wild type BPTI.
# These chemical shifts differ by 0.05-0.10 from those of C14A/C38A.
§ These chemical shifts differ by greater than 0.10 from those of C14A/C38A.

4.4.2. Results from the chemical shift assignment

Chemical shift assignments were easily obtained for residues at some distance from residues 14 and 38, the sites of mutation and chemical modification. The chemical shifts of all assigned protons from residues at i ± 5, where i equals 14 or 38, were all within 0.10 ppm of the published values for wild type BPTI. The resonance for the Gly-28 α2 proton in Table I in the assignment paper (pH 4.6, 36°C) from Wüthrich's lab (Wagner *et al.*, 1987) is listed as 3.91 ppm, while the value found here is 3.62 ppm. The published value is almost certainly a typographical error, as an independent assignment of the wild type protein under nearly identical conditions by Dr. Charles Eads of the Kuntz laboratory yielded a value of 3.62 ppm. An additional typographical error was found for the Arg-17 γ1 resonance; the published value of 1.28 should be corrected to 1.48, as found by Eads. A value of 1.50 was found for this resonance in the C38A-Am-Tempo derivative.

The chemical shifts for many protons near the site of modification remain unassigned. This is a particularly difficult region of the molecule to assign—even for wild type BPTI—for several reasons: (1) it is composed of two exposed loops, showing considerable mobility (Wagner and Wüthrich, 1982a; Wagner, 1983) (Wagner and Nirmala, 1989), and (2) the relatively low density of protons near these amino acids does not permit the observation of a large number of inter-residue NOEs that might be used for assignment purposes. Based on NOESY data with a mixing time of 100 ms from a 20 mM wild type BPTI sample at pH 4.6, 36°C, Wagner *et al.* report only one inter-residue connectivity for Cys-14 and Lys-15, three for Ala-16 and Arg-17, and none for Pro-13

(Wagner *et al.*, 1987). There are no connectivities reported from Ala-16 to Lys-15. Naderi *et al.* do report $d_{\alpha N}$ and $d_{\beta N}$ connectivities for Lys-15 to Ala-16 for the C14A/C38A double mutant at a concentration of 6 mM at pH 4.6, 36°C, using the comparatively long 200 ms mixing time. Using NOESY data acquired from a 3 mM solution of C38A-Am-Tempo at pH 4.6, 36°C with a 160 ms mixing time, I was unable to positively assign any of the resonances from Lys-15, the modified Cys-14, or Pro-13.

The density of protons is also low near the C38A mutation. In wild type BPTI, only four inter-residue connectivities are reported for Lys-41 and for Arg-39 (Wagner *et al.*, 1987). Two are reported for glycines 36 and 37, and one for Cys-38. In contrast, seven inter-residue connectivities are reported for Ala-40, including two sequential backbone connectivities to both Arg-39 and Lys-41. Despite this, I have so far been unable to assign any of the resonances from Gly-37 and the assignments are tentative for Ala-38 and Arg-39.

Gly-37 - The Gly-37 amide proton is strongly affected by the ring current field of Tyr-23; it is found at 4.23 ppm in wild type BPTI and at 4.46 in C14A/C38A, approximately 4 ppm upfield of the glycine random coil chemical shift (Wüthrich, 1986; Tüchsen and Woodward, 1987; Wagner *et al.*, 1987). In C38A-Am-Tempo, this resonance may be obscured by the strong t_1 noise band from the H_2O solvent peak.

Ala-40 - The fingerprint cross peak for Ala-40 is normally weak, probably due to the small $^3J_{HN\alpha}$ coupling constant (Pardi *et al.*, 1984). While searching for the Ala-40 fingerprint cross peak, I found it helpful to consider the chemical shifts of the amide and alpha protons of this residue in a number of previous studies. The most useful Ala-40 chemical shifts are given in Table 4.3.

Table 4.3
Chemical shifts of Ala-40 in a number of derivatives

Molecule	Conditions		Ala-40		Reference
	pH	°C	NH	HA	
Wild type	4.6	36	7.38	4.08	(Wagner *et al.*, 1987)
C14/C38	4.6	36	7.86	4.57	(Naderi *et al.*, 1991)
RCAM	4.6	68	7.92	4.70	(Stassinopoulou *et al.*, 1984)
TRAM	4.6	68	7.36	4.09	(Stassinopoulou *et al.*, 1984)
C38A-Am-Tempo	4.6	36	7.92	4.60	This work

Naderi's chemical shifts for Ala-40 in C14A/C38A are far from the wild type values, but are near the values reported for the more closely related RCAM molecule—which also lacks the 14-38 disulfide bond—at 68°C. Comparison of the RCAM chemical shifts to those from TRAM-BPTI (BPTI modified by transamination of the amino terminus; cross peak coordinate) suggests that the perturbation of the chemical shift is dominated by the disruption of the disulfide bond and not by the temperature coefficient (Stassinopoulou *et al.*, 1984). Therefore, I expected the chemical shifts for Ala-40 NH and HA to be in the neighborhood of 7.9 and 4.6. The key to the assignment of the Ala-40 spin system was the observation of an NOE between the amide proton of Lys-41 and a resonance at 1.29 ppm. A relay peak was observed at this position at an ω2 frequency of 7.92 in the HOHAHA spectrum. The α proton peak was then located at the position (4.60, 7.92) in both the HOHAHA and NOESY spectra. This fingerprint peak was absent from the COSY spectrum.

Ala-38 - The best analogy for the chemical shift of Ala-38 in the C38A-Am-Tempo molecule is the C14A/C38A mutant. Naderi's assignments for the Ala-38 spin system are 7.91, 4.70, and 1.31 ppm for the amide, alpha, and beta protons, respectively. The alpha proton is therefore within 0.02 ppm of the water peak at 36°C, a position where it is likely to be saturated during the solvent suppression period. I have tentatively identified a relay peak at (1.352, 7.822) in the HOHAHA that appears to be connected to a fingerprint peak at (4.170, 7.822). This peak is also found in the COSY, but not in the NOESY. The putative intraresidue HN–methyl connection is also missing from the

NOESY. However, there is a NOESY cross peak at the chemical shift position of the amide proton tentatively assigned to Arg-39 and the resonance at 1.35 ppm. This would appear to support the backbone assignments of both Ala-38 and Arg-39, but the evidence is not definitive.

Arg-39 - The assignments for Arg-39 are primarily based upon analogy with the assignment in C14A/C38A. They are supported by the assignment of an NOE between the amide proton and the methyl group of Ala-38 (see above), but the assignment must still be regarded as tentative.

Lys-41 - The assignment for Lys-41 is also based upon analogy with C14A/C38A, but it is substantiated by the observation of an NOE between its amide proton and the β methyl group of Ala-40. Relay peaks connected to the amide proton are observed in the HOHAHA spectrum and are consistent with the analogy.

Other unassigned residues - Assignments are still unavailable for Arg-1 and Pro-8 in C38A-Am-Tempo. Neither of these residues will yield fingerprint cross peaks in the COSY spectra.

Comparison of chemical shift differences - While it is not possible to directly interpret the chemical shifts in terms of a detailed structure, differences in corresponding chemical shifts between two similar structures are indicative of some structural change, even if that change cannot be described based solely on the chemical shift information. For this reason, chemical shift difference plots are frequently a useful description of the sites of structural change between two proteins. In Figure 4.1, the α and amide chemical shifts of C38A-Am-Tempo (S14 in the shorthand terminology of the figure) are compared to the C14A/C38A mutant and to wild type BPTI. The large differences observed in Figure 4.1A occur at one of the Gly-36 α protons and at the at Ala-38. These are both sites where the assignment is only tentative. The deviations in the amide chemical shifts (panel C) are centered around the sites of mutation, but they are relatively small. The

comparison of the C38A-Am-Tempo chemical shifts to those from the wild type protein reveal larger deviations, as expected, which are concentrated near the sites of mutation. Together, all of these data suggest that the perturbations in structure of the C38A-Am-Tempo derivative are limited to the site of mutation and modification. This bodes well for the inclusion of long range distance data derived from this species in the structure calculations of wild type BPTI.

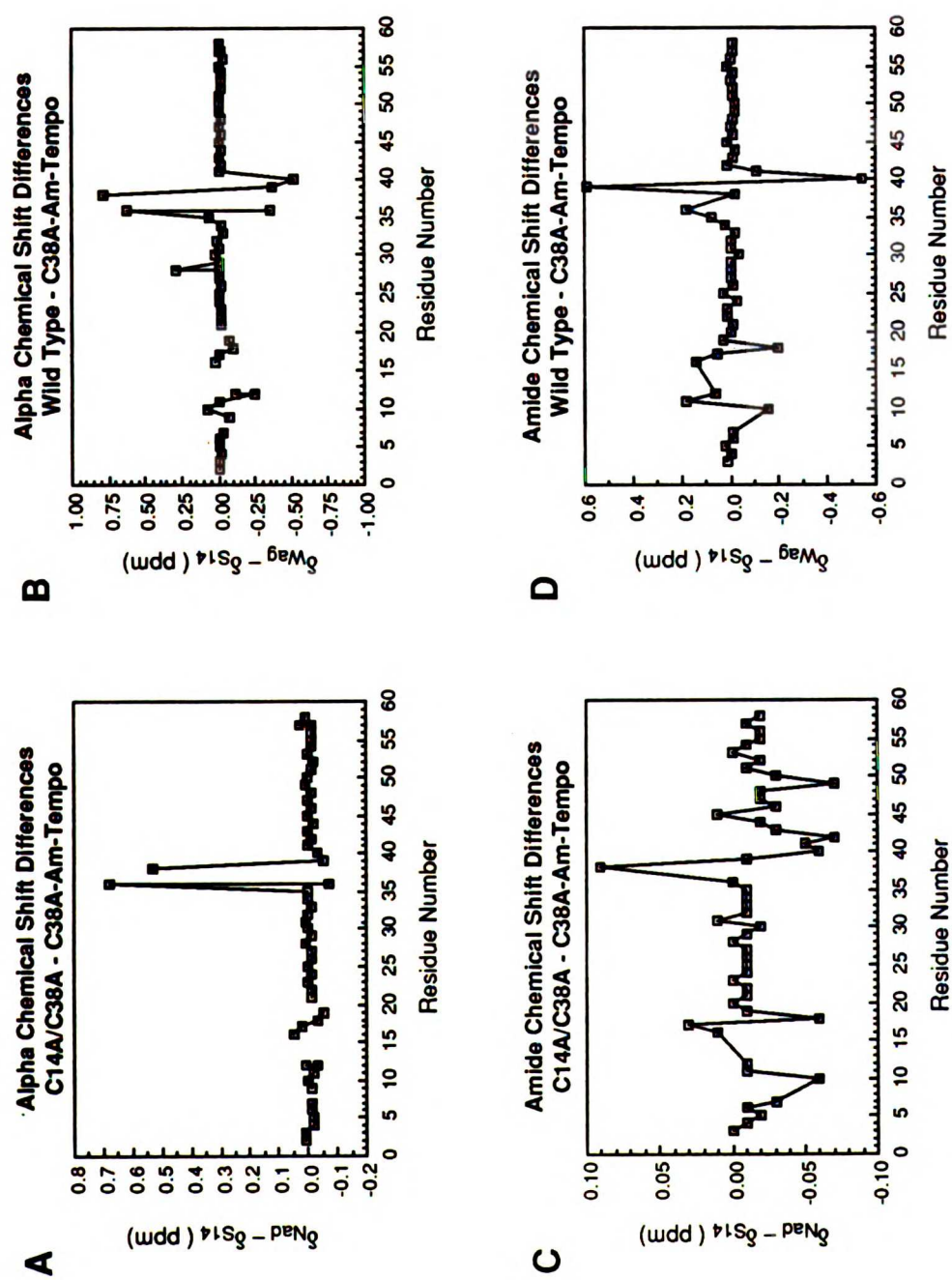


Figure 4.1: Chemical shift differences between the mutant C14A/C38A (Nad = Naderi), wild type (Wag = Wagner), and C38A-Am-Tempo (S14). All chemical shifts were determined at pH 4.6, 36 °C. (A) α proton shift differences between C14A/C38A and C38A-Am-Tempo; (B) α proton chemical shift differences between wild type and C38A-Am-Tempo; (C) amide proton chemical shift differences between C14A/C38A and C38A-Am-Tempo; (D) amide proton chemical shift differences between wild type and C38A-Am-Tempo.

4.4.3. Sequential connectivities

The data supporting the chemical shift assignments of proteins are customarily summarized in sequential connectivity diagrams. Figure 4.2 is a diagram of the observed sequential connectivities for C38A-Am-Tempo.

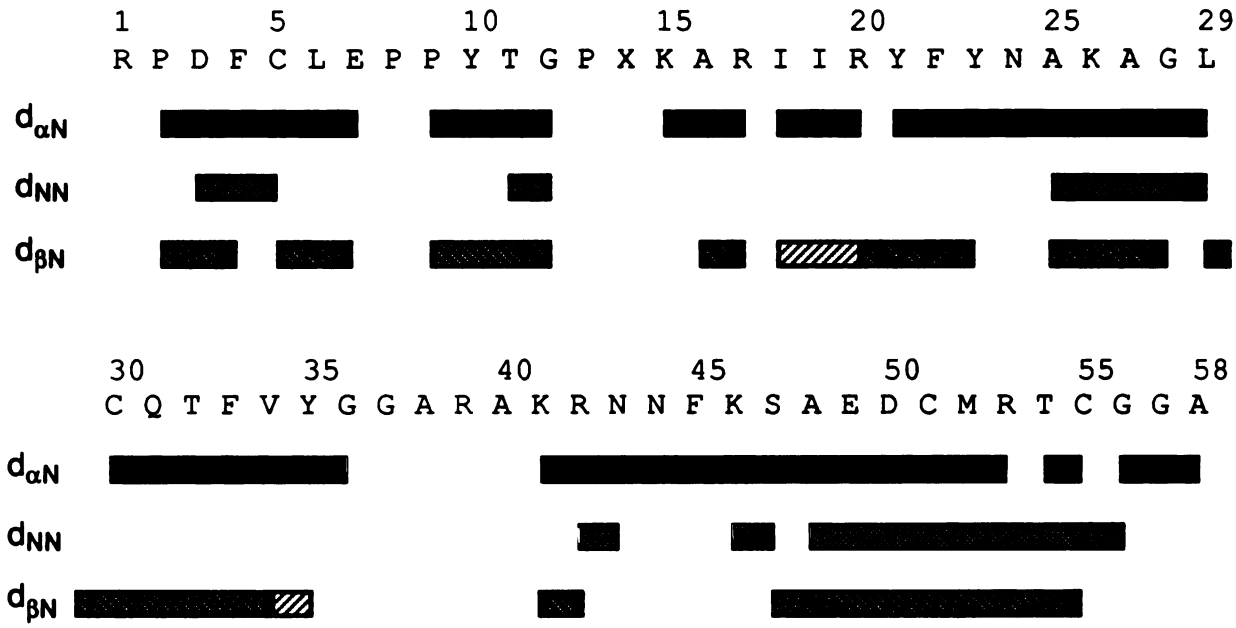


Fig 4.2. Observed sequential connectivities in C38A-Am-Tempo. Shading of the boxes is self explanatory, except for the appearance of ▨ in the d(βN) rows, which represents NOE's from the NH of the i + 1'th residue to a gamma methyl group of the i'th residue.

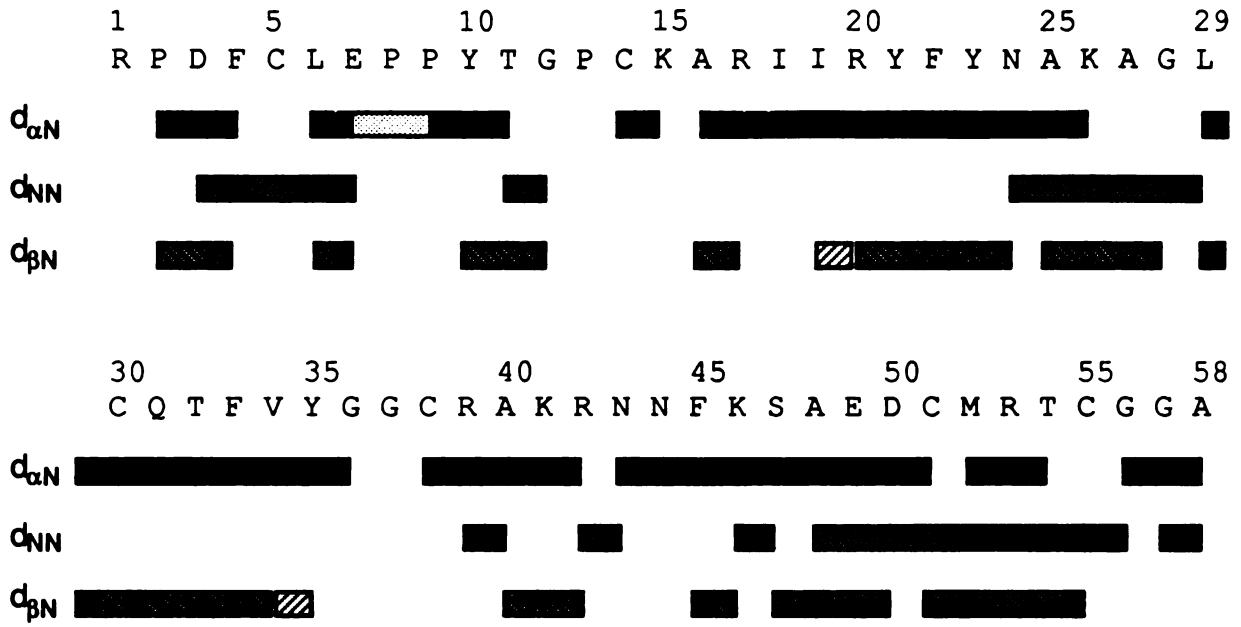


Fig 4.3. Observed sequential connectivities in wild type BPTI. Shading of the boxes is self explanatory, except for the appearance of ▨ in the d(βN) rows, which represents NOE's from the NH of the i + 1'th residue to a gamma methyl group of the i'th residue, and ▩ in the d(αN) rows, which represents a connection between an HD proton from a proline residue to the HA proton of the preceeding residue.

For comparison, a similar diagram, constructed from the data in (Wagner *et al.*, 1987), is shown in Figure 4.3.

The sequential connectivity diagrams are more useful for documenting the sequential assignment than they are for structure comparison; the picture of structures based on the diagrams suffers from an inherent reduction in dimensionality. Other problems arise in this specific instance for a number of reasons: (1) the assignments for C38A-Am-Tempo are incomplete, as noted above, (2) the NOESY spectra used to construct these diagrams were acquired with different mixing times, perhaps allowing the observation of NOEs between protons separated by longer distances in the C38A-Am-Tempo sample(τ_m = 160 ms, compared to 100 ms for the wild type data), by either a direct relaxation pathway or through spin diffusion, (3) the wild type data were acquired

with a 20 mM sample (compared to 3 mM for the C38A-Am-Tempo sample), increasing the signal-to-noise ratio and perhaps the overall effective correlation time, increasing the probability of the observation of weaker NOEs, counteracting the effect of number 2 above, and (4) t_1 noise in the individual spectra will obscure a subset of the real NOE peaks. These complications notwithstanding, the d_{NN} connectivity patterns are strikingly similar, exhibiting an absence of NOE cross peaks in the β sheet regions of the molecule, extending from residues 16-25 and 28-36 (using the 4pti secondary structure classification (Deisenhofer and Steigmann, 1975)), and showing NOEs in the turn region separating the two β strands and the helical regions near the N- and especially the C-terminus. These patterns are expected based on the secondary structure of BPTI (Wüthrich *et al.*, 1984).

4.4.4. Similarity of structures: conclusions

Only a small number of tertiary NOEs have been assigned for C38A-Am-Tempo (Table 4.4), and these do not permit the construction of even a moderate resolution structure for the molecule. The best evidence for the conservation of three dimensional structure between C38A-Am-Tempo and wild type BPTI is therefore not the NOEs but the lack of perturbation of most of the chemical shift values, as seen in Table 4.2 and Figure 4.1. Although it is sometimes possible to rationalize chemical shifts in terms of structure, particularly due to ring current shifts (Tüchsen and Woodward, 1987; Guiles *et al.*, 1990), it is not yet possible to assign structures based on chemical shifts. On the other hand, chemical shifts are a sensitive probe of changes of structure, and the large measure of agreement between the chemical shifts of C38A-Am-Tempo and wild type BPTI indicates that most of the structural details are conserved.

Table 4.4

Select tertiary NOEs assigned in C38A-Am-Tempo

Res-1	Atom-1	Res-2	Atom-2	Structural constraint
P2	HA	F4	HN	3_{10}-helix; $d_{\alpha N}(i,i+2)$
P2	HBa	F4	HN	3_{10}-helix; $d_{\beta N}(i,i+2)$
P2	HBb	F4	HN	3_{10}-helix; $d_{\beta N}(i,i+2)$
P2	HBb	C5	HN	3_{10}-helix; $d_{\beta N}(i,i+3)$
E7	HB2	N43	HN	Tertiary
Y21	HN	F45	HN	Tertiary
Y21	HB	F45	HN	Tertiary
I18	HN	Y35	HN	Antiparallel β sheet
I19	HA	Y35	HN	Antiparallel β sheet
R20	HN	F33	HN	Antiparallel β sheet
Y21	HA	F33	HN	Antiparallel β sheet
Y21	HA	T32	HA	Antiparallel β sheet
F22	HN	Q31	HN	Antiparallel β sheet
Y23	HA	C30	HA	Antiparallel β sheet
N24	HN	C30	HA	Antiparallel β sheet
A25	HA	G28	HN	Turn
S47	HA	E49	HN	C-terminal α helix; $d_{\alpha N}(i,i+2)$
S47	HB1	E49	HN	C-terminal α helix
E49	HA	M52	HN	C-terminal α helix; $d_{\alpha N}(i,i+3)$
E49	HA	R53	HN	C-terminal α helix; $d_{\alpha N}(i,i+4)$
D50	HA	R53	HN	C-terminal α helix; $d_{\alpha N}(i,i+3)$
D50	HA	T54	HN	C-terminal α helix; $d_{\alpha N}(i,i+4)$
C51	HA	T54	HN	C-terminal α helix; $d_{\alpha N}(i,i+3)$
M52	HA	G56	HN	C-terminal α helix; $d_{\alpha N}(i,i+4)$
F45	HA	S47	HN	
F45	HB2	S47	HN	

4.5. Analysis of spin label NMR data

4.5.1. Choice of experiment for spin label distance determination

As mentioned in chapter 1, Kosen (Kosen, 1989) concluded that the absolute value COSY experiment was the best two-dimensional experiment for the measurement of spin label broadening for the following reasons: (1) high signal-to-noise ratio, (2) spectral simulation is simple (Widmer and Wüthrich, 1986) (though we have not yet successfully applied spectral simulations for the extraction of line broadening factors),

and (3) the integrated peak intensity is strongly dependent on the linewidth when the linewidth is approximately equal to the coupling constant (Weiss *et al.*, 1984), a common situation for proteins the size of BPTI. The last reason is certainly the most compelling.

The primary disadvantage to the absolute value COSY experiment is that the contribution to the change in integrated peak intensity from the line broadening of the two coupled nuclei cannot be separated. As a consequence, it is necessary to use an averaged coordinate for the nuclei that give rise to each cross peak (Kuntz, personal communication). This probably represents a small contribution to the overall error of the measured distances, though this has not been analyzed rigorously. The effect of limited digital resolution on the quality of the spin label data has not been determined. A further problem with all two dimensional experiments is the need to acquire data from fully relaxed systems, requiring long relaxation delays; this effectively reduces the signal-to-noise ratio available for a two-dimensional experiment in any given amount of time by limiting signal averaging for each t_1 increment. Finally, enhanced relaxation of water protons by the paramagnetic sample during the frequency labeling period makes solvent suppression by presaturation difficult and can limit the number of t_1 increments which may be acquired. This effect ought to show a correlation with the concentration of the paramagnetic sample, but other effects such as field inhomogeneity will also yield a significant contribution. In practice, the limit for the C38A-Am-Tempo samples (within the concentration range 0.6–3 mM) was 300-400 t_1 increments with a vertical dwell time of 142 μs, corresponding to delays of 42.6–56.8 ms. Solvent suppression in NOESY experiments with spin labeled proteins will be less difficult because solvent irradiation may be employed during the mixing period or the $\pi/2$ read pulse may be replaced by a frequency selective pulse such as the 1-$\bar{3}$-3-$\bar{1}$ pulse (Hore, 1983; Hore, 1989).

4.5.2. Interpolation: comparison of oxidized spin label data to a series of artificially broadened reduced spin label data

The observed line broadening was quantitated by comparison of the volumes of the fingerprint peaks from the paramagnetic sample with those of the corresponding peaks in a series of artificially broadened spectra of the reduced protein. Custom digital filtering functions were constructed by multiplication of an unshifted sine bell—the window function universally used for absolute value COSY spectra—with a decaying exponential corresponding to a defined line broadening factor. This was done in both dimensions, with the identical line broadening factor used in both ω_1 and ω_2. The sine bell function was calculated to decay to zero at the last data point before zero-filling; in the t_1 dimension, this corresponds to data point 300 or 400. Line broadening factors of 0, 1, 2, 4, 8, and 16 Hz were used, yielding six complete data sets; cross peak volumes were determined separately for each data set.

We anticipated that antiphase cancellation would cause poor signal-to-noise ratios in the artificially broadened spectra. However, exponential multiplication causes a significant reduction in the noise level (the noise base plane was at 1e4 in the unbroadened spectra, but only at 5e2 in the 16 Hz broadened spectra), and no cross peaks were observed to vanish. A much more serious problem is caused by a degradation in resolution—already suboptimal in the absolute value COSY spectra—in the broadened spectra. Some peaks which are well resolved in the 0 Hz data set become seriously overlapped in the 8 and 16 Hz data sets, and are therefore difficult to integrate, even with routines that fit multiple two-dimensional peak shapes (D. Kneller, unpublished). This includes several peaks near the strip of noise from the residual water signal that appears in the ω2 dimension.

A possible alternative procedure which will create much less overlap, involves the following steps: individual cross peaks from the reduced spectrum, processed with no

line broadening, are (1) back transformed, (2) broadened by exponential multiplication, (3) retransformed, and (4) compared to the corresponding peak in the oxidized spectrum. This can be done in an iterative manner until the peak volumes match within a certain tolerance. A further advantage to this procedure is its reduced demand on disk space; instead of storing six full data sets, the back transformation procedure requires the equivalent of only two, the real and imaginary portions of the final spectrum. This procedure, although explored by Scheek and Manogaran (unpublished) was not chosen because it is not implemented in the current software packages available to the lab.

A major difficulty with either procedure is that the comparison of peak volumes involves quantities which are in a sense *extensive*, *i.e.* the peak volumes depend on the concentration of the sample. Even though reduction of the paramagnetic sample by ascorbate was accompanied by a dilution of only 1%, a quantity I have considered negligible, when a sample is removed from the magnet it must be regarded as a new sample. Solvent that had condensed on the side of the tube prior to sample removal is likely to be mixed with the bulk of the liquid, causing an irreproducible dilution of the sample. Similar effects can occur during mixing of the ascorbate with the protein solution, either by gentle inversion of the sample or via pipetting through a small capillary tube into a larger pipette tip. Both are likely to change the volume of the sample, affecting the optimal shim settings. In theory, it is desirable to use identical instrument settings, including the shims, for the acquisition of data from the oxidized and reduced samples. In practice, small adjustments of the lower order spinning shims are always necessary.

Difficulties attributable to changes in the concentration of the sample might be responsible for the observation that a number of peaks in the paramagnetic protein are more intense than the corresponding peaks in the reduced protein; this was seen for data taken at 0.6 and 3 mM, but not at 2 mM. In principle, it is possible to normalize the oxidized and reduced data sets, using cross peaks that are unaffected by the reduction. This technique is commonly used in studies of hydrogen exchange kinetics, where the NMR data

are normalized based on the intensity of a cross peak arising from two non-exchangeable protons, for instance a well resolved α–β cross peak (Wagner and Wüthrich, 1982a). Despite a number of attempts to find a suitable protein cross peak peak or set of peaks for normalization—based on the distance of the residue giving rise to the cross peak from the known site of spin label modification in the crystal structure—none could be found, and the data acquired at 0.6 and 3 mM were not used for the spin label calculations. A small molecule concentration standard yielding a well resolved cross peak might be useful, with the caveat that it is desirable to use protons with T_1 values similar to those of the protein. Alternatively, it ought to be possible to obtain normalization factors from one-dimensional data, perhaps using the standard TSP chemical shift resonance reference. The long T_1 values of the TSP peak do not present a problem because one-dimensional data can be obtained from fully relaxed conditions in quite reasonable time periods. These data may even provide better signal-to-noise ratios and more precise normalization factors.

4.5.3. Data from C38A-Am-Tempo

It was immediately apparent that a number of peaks were completely absent from the COSY spectra of the oxidized protein but not from the reduced protein, and that these residues cluster in space near the site of spin label attachment. These residues include Tyr-10, Thr-11, Arg-17, Ile-18, Ile-19, Phe-33, Val-34, and Tyr-35. In addition, there are a number of unassigned peaks in the fingerprint region which disappear in the oxidized spectrum. These probably arise from the unassigned residues Cys-14* , Lys-15, and Ala-38; these peaks were used as starting points for the extension of the sequential assignment mentioned in section 4.4, but further evidence indicating the identity of these

---

* Cys-14 is the site of modification by the spin label.

peaks could not be found in the NOESY spectra. Figure 4.4 shows a sample of peaks which are either obliterated or relatively unaffected by the state of the spin label.

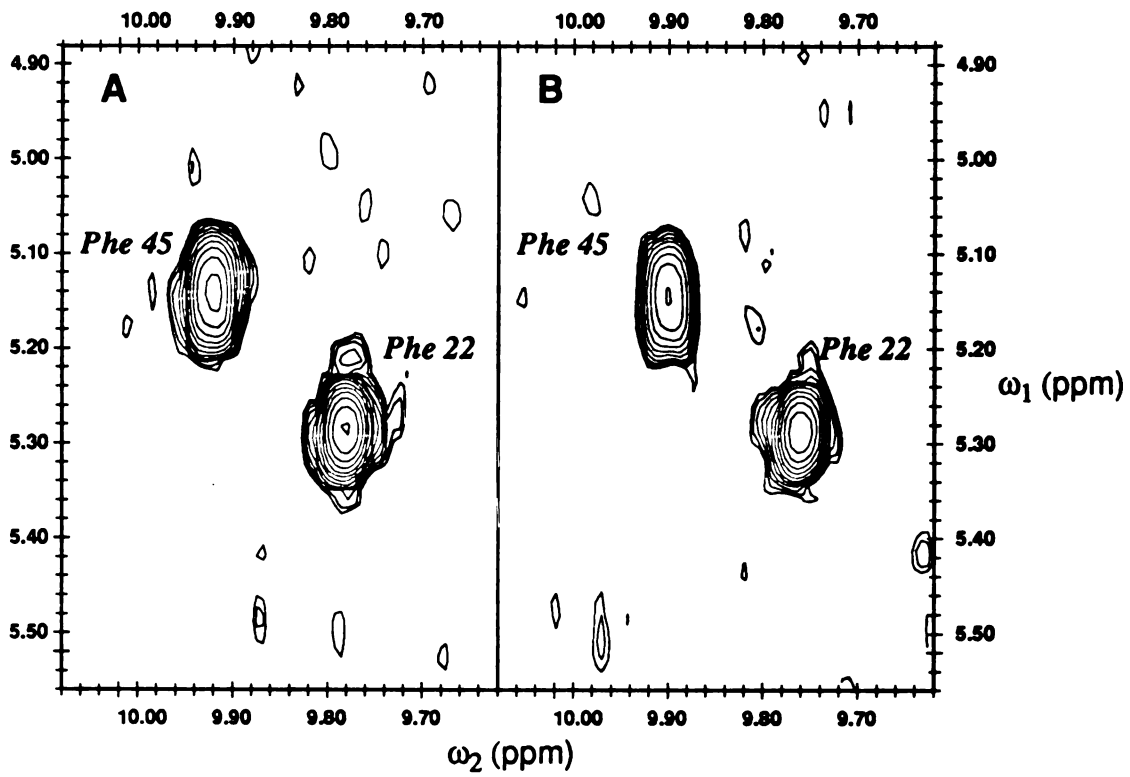


Figure 4.4 (A-F): Comparison of two-dimensional absolute value COSYspectra from diamagnetic and paramagnetic derivatives of C38A-Am-Tempo for several selected cross peaks. In each case, the diamagnetic spectra are on the left and the paramagnetic spectra are on the right. Peak labels in the figure are self-explanatory.

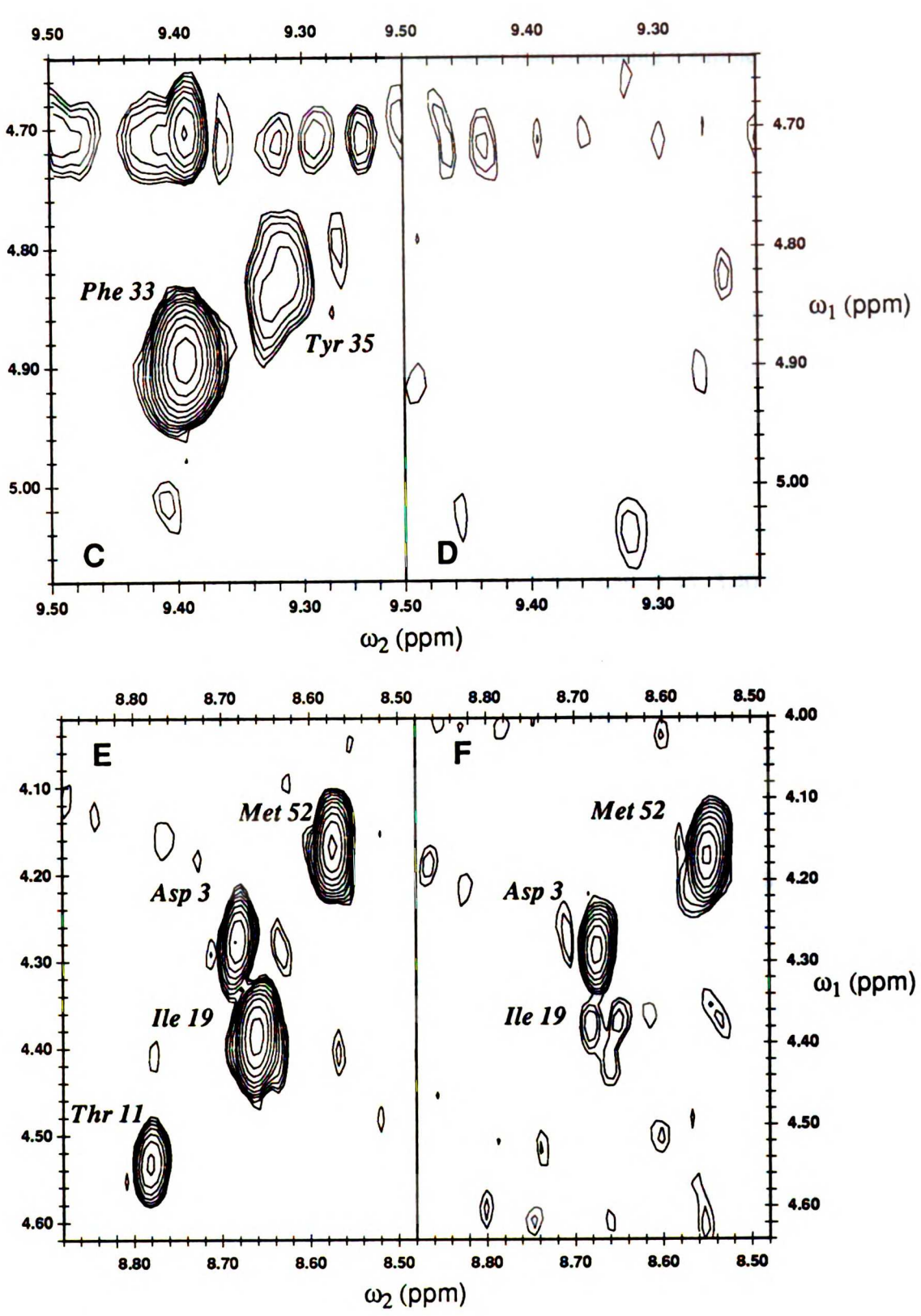


Figure 4.4, cont.

The volume of all peaks in the COSY spectra were approximated by fitting each resolved feature to a two-dimensional gaussian function using a simplex routine (D. Kneller, unpublished), followed by analytical integration using the parameters of the best fit lineshape. The peak shapes are certainly not gaussian, but it is necessary to assume some lineshape model in order to obtain quantitative data from overlapped peaks. The relative error introduced by this assumption is probably small. For cross peaks which contain more than one resolved feature, the volume of each component is determined first and the total volume of the cross peak is then the simple sum of the volume of each feature. For an example, see the treatment of the Gly-28 HAb peak in Figure 4.5.

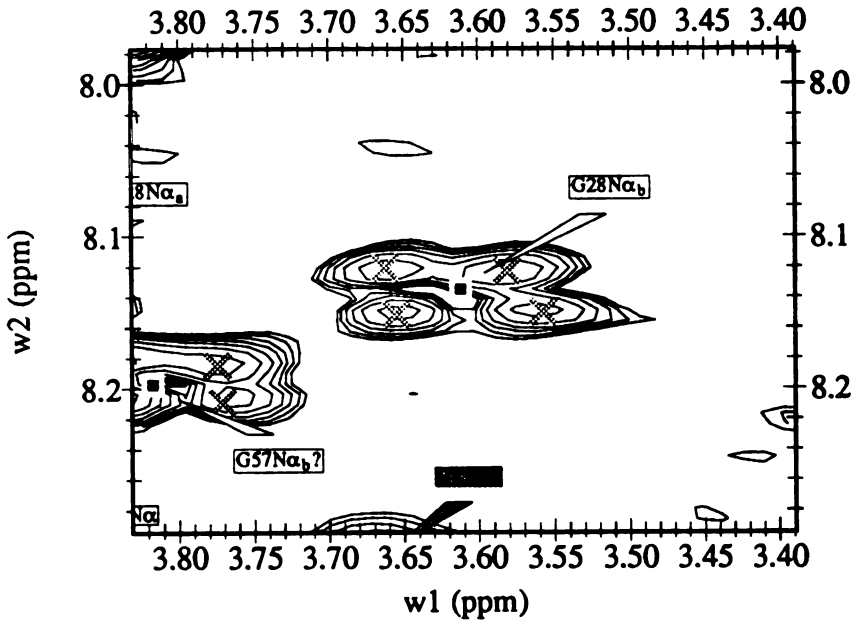


Figure 4.5: The intensity of cross peaks with resolved features is simply the sum of the intensities of the resolved features. The Gly-28 HAb cross peak is shown as an example.

The simplex routine is quite sensitive to the quality of the starting parameters, which represent reasonable guesses of the linewidth in each dimension. Starting guesses are usually uniform throughout the spectrum and can be chosen to reflect the characteristics of an "average" peak. In heavily overlapped regions, the following procedure was sometimes necessary: (1) the data was recontoured using a larger lowest contour level (hopefully removing some of the overlap), (2) better starting estimates of the lineshapes were obtained by line fitting at the higher contour level, and (3) the data was contoured at the original level (usually just above the noise level) and fit again. The fitting procedure frequently reports linewidth parameters which are obviously either too small or too large, or peak positions that are not consistent with those obtained from a visual inspection of the data. Such misinterpretations of the data could usually be corrected by repeating the fitting procedure, beginning with a new set of starting conditions. Unfortunately, the procedure is sensitive to which lowest contour level is chosen and the degree of overlap, so the reproducibility of the procedure is not great.

The intensity data for each individual cross peak was first plotted as a stacked plot, with the line broadening factor on the abscissa. Examples of such plots are shown in Figure 4.6 for Ala-48 and for Cys-30.

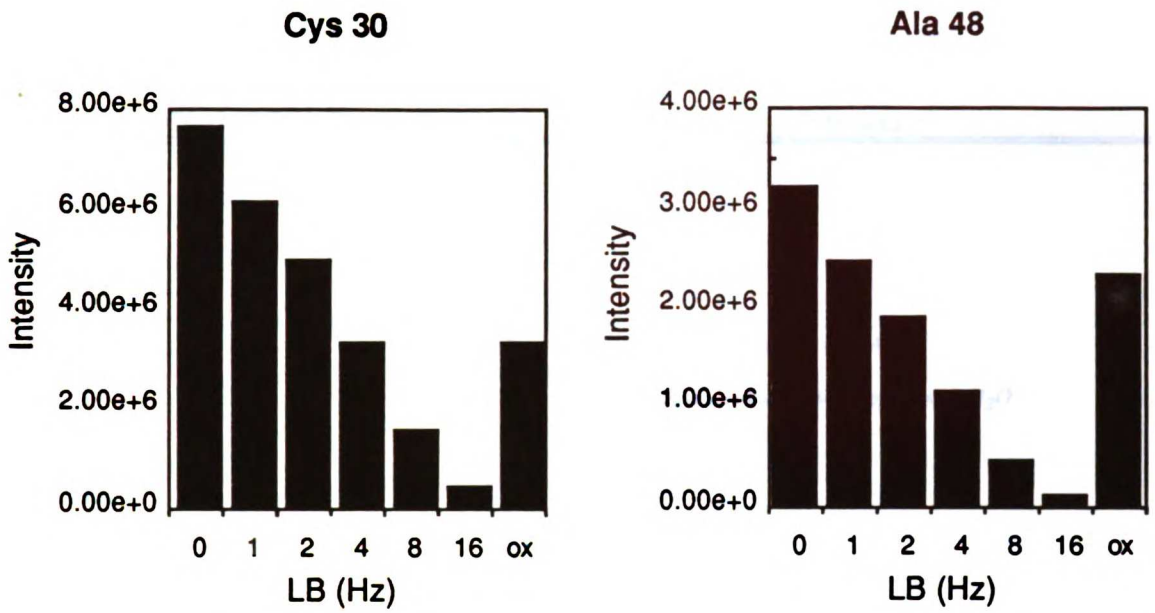


Figure 4.6: Two dimensional peak intensities (arbitrary units) in artificially broadened spectra. Peak intensities are plotted against the applied broadening factors for the fingerprint crosspeaks of Cys 30 and Ala 48. The intensity observed in the paramagnetic derivative is also included.

Plots for other resonances were generally similar in appearance. Based upon a visual inspection of these plots, the intensity data for the reduced protein was replotted as a function of the $\log_2$ of the artificial line broadening and then fit to a linear equation. I am unaware of any general analytical treatment of this situation. The observed paramagnetic line broadening for each peak was then calculated by substitution of the paramagnetic intensity into the corresponding linear equation, using (somewhat arbitrarily) the fit of the data using the $\log_2$ function. These data are reported in Table 4.5a. The average distance of the amide/alpha proton pair to the Cys-14 sulfur atom is also reported in this table. This is intended only as a gross approximation of the site of the paramagnetic center; the actual coordinate of the nitroxide will differ because of the size of the Tempo moiety and where it is positioned with respect to the rest of the protein.

Table 4.5a
Line broadening of non-vanishing fingerprint cross peaks

Residue		Distance from Cys-14 SG (Å)*	Paramagnetic line broadening (Hz)	Comment
Asp	3	25.4	0.7	
Cys	5	22.3	1.0	
Leu	6	23.2	1.7	
Tyr	21	17.4	2.3	
Phe	22	18.6	1.7	
Tyr	23	21.3	1.6	
Ala	25	25.0	2.0	
Lys†	*26*	*27.2*	*2.5*	*Peak is overlapped with A58*
Ala	27	28.0	2.5	
Leu	*29*	*26.9*	*3.0*	*Peak is weak and overlapped with H_2O.*
Cys	30	24.2	3.9	
Gln	31	21.9	4.5	
Thr	32	19.6	6.0	
Arg	42	15.5	1.5	
Asn	43	17.0	1.6	
Asn	44	15.6	1.6	
Phe	45	16.7	1.8	
Lys	46	17.2	1.8	
Ser	47	20.3	1.9	
Ala	48	22.6	1.2	
Glu	49	25.6	2.0	
Asp	50	23.8	1.3	
Cys	51	22.1	0.6	
Met	52	25.5	1.2	
Arg	53	27.4	2.3	
Thr	54	25.4	0.9	
Gly	56	27.6	0.9	
Ala	*58*	*33.6*	*1.1*	*Peak is overlapped with K26*

*The distance from the Cys-14 SG atom to the average position of the amide and alpha proton of the indicated residue, using the 5pti crystal structure (Wlodawer *et al.*, 1984).

†The uncertainty for the italicized residues is somewhat greater than for the others. Explanations are provided in each case.

Table 4.5b
Peaks not yielding quantitative data

Residue		Distance from Cys-14 SG (Å)*	Comment
Phe	4	21.9	overlapped with water; weak
Glu	7	20.5	Peak is weak and overlapped with S47 and H_2O
Tyr	10	12.6	vanished; weak peak in reduced protein
Thr	11	10.2	vanished
Gly	12	8.0	vanished; weak peak in reduced protein
Cys	14	2.7	not assigned; almost certainly vanished
Lys	15	5.3	not assigned; almost certainly vanished
Ala	16	6.8	vanished
Arg	17	10.0	vanished
Ile	18	11.0	vanished
Ile	19	14.0	vanished
Arg	20	15.1	α proton saturated by solvent suppression
Asn	24	22.9	not resolved
Gly	28	28.0	difficulty with integration due to peak fine structure
Phe	33	16.2	vanished
Val	34	13.4	vanished
Tyr	35	10.2	vanished
Gly	*36*	*6.9*	
Gly	37	6.2	not assigned
Ala	*38*	*4.3*	
Arg	*39*	*6.8*	
Ala	40	8.5	overlapped with water; weak
Lys	*41*	*11.9*	*vanished or overlapped*
Cys	55	25.5	reduced peak overlapped with water
Gly	57	31.6	

Ideally, quantitative data should be available for all non-proline residues in the protein sequence (excepting the amino terminal arginine residue). In practice, it was not possible to extract quantitative data for a number of cross peaks; a list of such peaks and the reasons for the difficulty associated with quantitation are presented above in Table 4.5b. The most significant set, as mentioned at the beginning of this section, are those cross peaks which are completely absent from the spectrum of the paramagnetic protein. Based upon peak heights in the artificially broadened spectra and the noise level in the paramagnetic spectrum, lower limits can be estimated for the line broadening in most of these cases. All but the strongest peaks (Arg-17, Tyr-21, Phe-22, Lys-26, Ala-27, Cys-30, Thr-32, Val-34, Phe-45, and Ala-58) will vanish if the paramagnetic broadening is at least 16 Hz. Weaker peaks,(*e.g.*, Glu-7, Tyr-10, Gly-12, Tyr-35, and Ala-40) will vanish if broadened by at least 4-8 Hz. However, the information garnered from all of the van-

ished peaks must be regarded as only qualitative. The location of the residues involved certainly provides further confirmation of the site of spin labeling.

The number of vanishing cross peaks in the data from C38A-Am-Tempo is significantly greater than was previously observed for a number of BPTI derivatives labeled at primary amino groups (Kosen *et al.*, 1986). The diagrams in Figure 7 of that paper, which plots the ratio of the intensity of the paramagnetic protein peaks to those of the diamagnetic derivatives as a function of amino acid position, show values of zero at positions where no intensity is observed for the paramagnetic protein. The plots are somewhat misleading because values of zero are also given in cases where resonances are not assigned (*e.g.*, due to chemical shift degeneracy with water) or where overlap of cross peaks did not permit the integration of peak volumes (determined by simple summation of intensity within a square defined around the cross peak). Taking this caveat into account, only a single case in three protein derivatives is noted where a cross peak has truly vanished from the paramagnetic spectrum—*i.e.*, Gly-56 in the spectrum of BPTI modified by a spin label at the amino terminus.

There are several probable causes for the qualitative difference in the number of vanishing peaks between C38A-Am-Tempo and the derivatives of wild type BPTI, most relating to differences in correlation times. Based on equations 1.2 and 1.4, in the spin diffusion regime ($\omega^2\tau_c^2 >> 1$), the paramagnetic contribution to linewidth is directly proportional to the effective correlation time (this also makes the signal intensity inversely proportional to the correlation time). There are two primary reasons why the effective correlation times are greater for the C38A-Am-Tempo system compared to the previous wild type data (Kosen *et al.*, 1986). First, the wild type data was collected at 68°C, while the C38A-Am-Tempo data was obtained at 36°C; based on Debye-Stokes theory, overall rotational correlation times of spherical objects are inversely proportional to absolute temperature (James, 1975). This corresponds to an approximate 10% increase in the paramagnetic contribution to the linewidth when decreasing the temperature from 68 °C

to 36 °C. The larger correlation time at 36 °C also increases the inherent line widths in the C38A-Am-Tempo spectra, decreasing the signal intensity due to antiphase cancellation, even in the absence of the spin label. The disappearance of peaks is caused by the sum of the contributions from the inherent line width and the paramagnetic interaction. Smaller paramagnetic contributions are required to cause the disappearance of peaks when the line widths are broader, as is the case for the data at 36 °C.

The second contribution to the increased effective correlation time in C38A-Am-Tempo is due to a reduction in internal mobility. Two of the three derivatives in the previous study were labeled at the ε-amino group of lysine residues, in contrast to the modification at a cysteine residue in C38A-Am-Tempo. The more numerous conformational degrees of freedom of the surface exposed lysine side chains will decrease the effective correlation time relative to spin labels attached to cysteine residues. The increase in temperature will also increase the internal mobility of the lysine-attached spin labels, further decreasing the correlation times. It is difficult to estimate the size of these conformational effects. Perhaps it is significant that the only completely vanished peak in the previous study was observed in the derivative modified at the amino terminus, where the attachment site is much more rigid. Finally, the shorter cysteine tether may actually decrease the distance between the spin label and the rest of the protein. The extreme case, that of a lysine side chain fully extended away from the protein probably does not apply because the hydrophobicity of the Tempo moiety will cause it to bury some of its surface area by nestling up to the protein.

4.5.4. First order analysis of observed broadening

The most striking observation in Table 4.5a is the large paramagnetic line broadening at Gln-31 and Thr-32. As shown by Figure 4.7, these peaks are both intense and well resolved in the diamagnetic spectra, so the reported values are not due to limitations in the signal-to-noise ratio or integration artifacts. Further evidence of the significance of

the observation is obtained from counting contour levels; there are 9 contour lines for Gln-31$_{dia}$ and 7 for Gln-31$_{para}$, while there are 13 for Thr-32$_{dia}$ and 9 for Thr-32$_{para}$. It is difficult to reconcile these data, together with data from the peaks which have disappeared in the paramagnetic spectra, with a model that relies completely on intramolecular interactions.

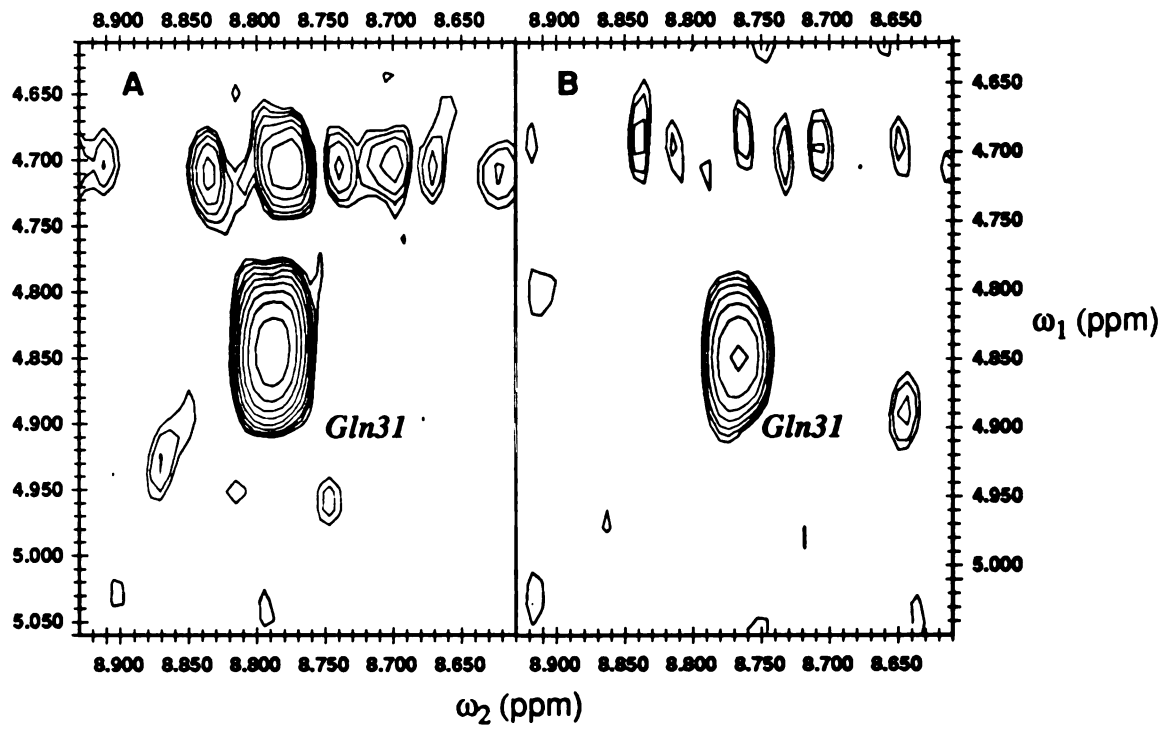


Figure 4.7: The fingerprint cross peaks from the Gln-31 and Thr-32 residues of C38A-Am-Tempo reveal the anomalous broadening at these sites. In each case, spectra from the diamagnetic sample is on the left and the paramagnetic sample is on the right. All spectra are contoured at the same level with a factor of 1.4 between levels.

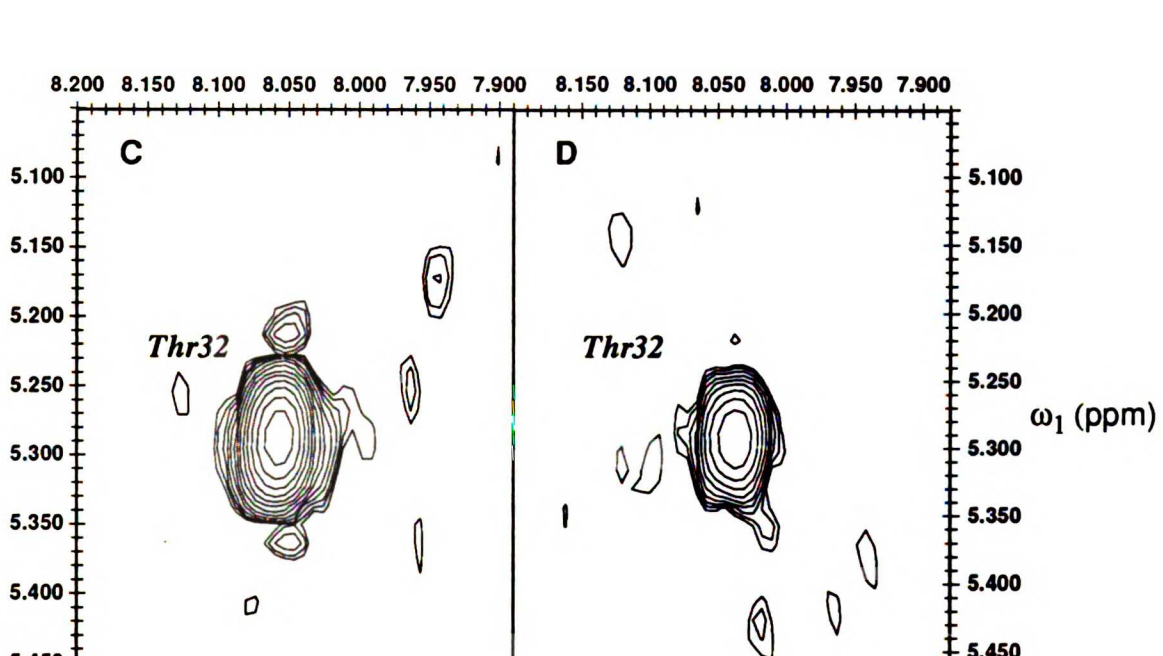


Figure 4.7, cont.

4.5.4.1. Intermolecular interactions

There is conflicting evidence on the formation of specific intermolecular interactions in BPTI. (Gallagher and Woodward, 1989) found no evidence for aggregates by dynamic light scattering. Wagner, on the other hand, notes the anomalous hydrogen exchange behavior of several solvent accessible amide protons and attributes this to intermolecular interactions: "For four protons of the latter group (Asp 3, Cys 30, Thr 32, and Val 34) the unexpected slow exchange seems to be due to an aggregation. The exchange experiments were normally carried out at 20 mM protein concentration. When the exchange was measured in 0.2 mM solution none of these four protons showed slow exchange." (Wagner, 1983).

Wagner's observation of intermolecular interactions at the Cys-30, Thr-32, and Val-34 patch may be caused by the same interaction which causes the anomalous broadening at Cys-30, Gln-31, and Thr-32, but this is only speculation. In neither set of data is the interacting surface on the complementary molecule(s) defined. In addition, the equilibrium concentration of aggregates will probably be lower in the 2 mM C38A-Am-Tempo sample than in the 20 mM wild type BPTI sample, unless the hydrophobic Tempo moiety becomes a major contributor to the interaction. The anomalous hydrogen exchange data will primarily be a function of the equilibrium concentration of the aggregates, while the spin label data will be a function of both the equilibrium constant and the exchange behavior of the interactions, much like the transferred NOE (Clore and Gronenborn, 1982; Clore and Gronenborn, 1983). Therefore, high concentrations of aggregates are not necessarily required for the observation of intermolecular interactions in the paramagnetic relaxation data.

Intermolecular interactions were previously observed in data from the Lys-26 spin labeled derivative of BPTI (Kosen, 1989). The broadening due to dimerization and intermolecular exchange was accounted for by the extrapolation of data obtained over the

protein concentration range of 1–5 mM to zero concentration. As previously noted, the 0.6 and 3 mM data sets from C38A-Am-Tempo could not be used for this type of analysis because of the inability to normalize the concentration differences between the oxidized and reduced samples. In addition, the signal-to-noise ratio for the 0.6 mM data set was particularly poor.

4.5.5. Modeling of the position of the spin label

The problem of positioning the spin label in Cartesian space based on the paramagnetic broadening data is formulated in a different manner from the way in which the spin label data is ultimately to be used. The purpose of acquiring spin label data is to provide long-range constraints for structure generation by distance geometry (or an alternative structure generation algorithm). Here, I have asked the question, can the spin label data be accounted for the by crystal structure of BPTI (Wlodawer *et al.*, 1984)? This formulation is entirely appropriate for our model system which has been designed to test the quality of data that may be obtained from the spin labeling approach.

Two basic parameters are allowed to vary in the construction of the structural models for the spin label: the Cartesian coordinate of the spin label and the number of sites needed to account for the data in a multiple site model. Multiple site models are considered because of the possibility that the spin label will reside in more than one distinct site on the protein. This type of model may also be necessary to account for the mixed population of stereoisomers introduced by the prochiral center in the 2,2,6,6-tetramethyl-4-(bromoacetamido)piperidine-1-oxyl reagent. I have included the rather restrictive assumption that the free energies of the individual models in the multiple site sets are equal—*i.e.*, the populations of all species are equal. I have also assumed that the structure of the spin labeled protein in solution is very close to that in the crystal and I have not allowed for protein flexibility. In light of the fact that I have broken a disulfide bond to introduce the site for selective spin labeling, this may be too restrictive an as-

sumption, but the recent solution of the crystal structure of the C14A/C38A mutant of BPTI provides further evidence that the perturbation caused by the removal of the disulfide is not large.

The best position for the spin label is modeled with the aid of the program SPINGRIDD (Kuntz, personal communication). The program moves the spin label around on a three-dimensional grid and attempts to minimize the difference between the sum of the calculated and observed paramagnetic line broadenings; the line broadening is calculated using the Solomon-Bloembergen equation presented in chapter 1 (equation 1.2). The difference between the calculated and observed broadening at each residue is weighted by the estimated error in the measurement. The non-gradient steepest descent algorithm is used to minimize the difference between the calculated and observed broadenings (Kuntz, personal communication). This algorithm was chosen because it will not get trapped in shallow local minima and because it is straightforward to code. However, the procedure will get caught in deep local minima, and the space around the protein is not searched exhaustively; therefore, the calculation is repeated at several initial spin label coordinate positions to find individual local minima.

The data for the calculation are entered in a file that is a modification of the standard PDB format for ATOM records (Bernstein *et al.*, 1977; Abola *et al.*, 1987). Columns 1-8 are as in the PDB. Column 9 contains the observed line broadening, column 10 contains the estimated error in the line broadening, and column 11 contains an integer necessary for the correct averaging of the atomic coordinates when data are obtained from cross peaks where the contribution from each nucleus cannot be separated, as is the case with the data from absolute value COSY peaks. When two atoms are involved, as is the case for non-glycine fingerprint cross peaks, the entry in column 11 for the first atom in the list is 2 and for the second occurrence, the entry is one. For glycine residues with resolved alpha protons, the countdown begins at 3. An excerpt from a typical data set is given in Table 4.6.

Table 4.6
Excerpt from a typical spingridd input data table

RECORD	NUM	NME	RES	RNUM	X	Y	Z	LB	ERR	I
ATOM	454	D	ALA	27	22.488	5.076	-9.180	2.5	2.0	2
ATOM	455	HA	ALA	27	20.090	5.015	-10.727	2.5	2.0	1
ATOM	463	D	GLY	28	22.531	7.298	-10.049	2.0	10.0	3
ATOM	464	HA1	GLY	28	22.992	9.579	-10.730	2.0	10.0	2
ATOM	465	HA2	GLY	28	21.483	9.541	-11.578	2.0	10.0	1
ATOM	474	D	LEU	29	21.510	8.788	-8.071	2.0	10.0	2
ATOM	475	HA	LEU	29	20.311	11.468	-7.252	2.0	10.0	1

Note that only the atoms for which paramagnetic broadening factors are observed are retained in the file; other atoms, including all heavy atoms are deleted.

A second parameter file contains a number of additional parameters necessary for the calculation: (1) the value of the constants in the Solomon-Bloembergen equation, including the factor of π needed to convert the $1/T_{2P}$ expression into a linewidth,

$$((1/15)S(S+1)\gamma^2 g^2\beta^2(4)(1/\pi) = 15{,}661{,}000\ \text{Å}^6\,s^{-1}ns^{-1}$$

(2) the overall rotational correlation time of the molecule, (3) the number of probe positions to be modeled, and (4) a number of different coordinate sets which describe the starting position of the spin label. In the current implementation, a number of UNIX™ shell and *awk* scripts are necessary to automatically run the calculation on more than one set of starting coordinates.

The overall rotational correlation time for this system was estimated to be roughly 2 ns. This is consistent with the shape of the EPR spectrum of the spin labeled protein (Poole and Farach, 1987) (see Figure 3.11) and the expected value based on the Debye-Stokes relation. Alternatively, the correlation time may be measured based on the frequency dependence of $1/T_{1P}$ or the ratio of T_{1P}/T_{2P} (James, 1975). The latter measurement is difficult to obtain for proteins because of the scarcity of singlet resonances. Correlation times can also be estimated from heteronuclear relaxation measurements (James *et al.*, 1978) (Nirmala and Wagner, 1988); such experiments suggested a value of 4 ns for BPTI at 36 °C (Nirmala and Wagner, 1988). For a given observed line broadening, the calculated distances are proportional to the sixth root of the correlation time.

Therefore, errors in the estimation of this parameter will have only a much smaller effect on the distance estimation.

The SPINGRIDD program does not use any chemical information that would provide additional (and quite stringent) constraints on the position of the spin label. For instance, the label is not constrained to be within covalent bonding distance—given the configuration and conformation of the acetamido-Tempo moiety—of the sulfur atom of Cys-14. A more sophisticated treatment of this situation might include a restrained molecular dynamics simulation where pseudo potentials are introduced to represent the line broadening and where the protein is held fixed, except for bond rotations of the side chain of the modified residue, including the acetamido-Tempo moiety itself. Subsequent analysis might allow for movement of portions of the remaining protein chain. SPINGRIDD is intended for a first—and computationally inexpensive—attempt to assess whether a single site model is appropriate for the data and where inconsistencies between the data and the model, such as those coming from intermolecular interactions, might arise.

Multiple site models can be distinguished from intermolecular effects in two ways. The first test is whether the individual sites are chemically reasonable, *i.e.*, can a spin label reach each site from its point of attachment to the protein. The second test comes from analysis of concentration dependent data. The failure of SPINGRIDD to produce a reasonable model may also derive from poor quality data; while non-quantitative but completely correct NOE data will often yield excellent NMR structures (Clore and Gronenborn, 1989a; Clore and Gronenborn, 1991a), it is of the utmost importance to obtain accurate, quantitative spin label data.

Spin label derivatives which yield multiple site models or evidence of strong intermolecular interactions which cannot be accounted for by extrapolation to zero concentration must be discarded as unsuitable for future use in structure generation algorithms. Of the three amino group labeled derivatives of BPTI previously analyzed,

only the Lys-26 derivative was not discarded on these grounds (Kuntz, personal communication). An additional complication arises because these cases may be difficult to distinguish from the bad data case. A judgement must therefore be made as to whether time is better spent acquiring new data or constructing new derivatives. With non-selective spin labeling schemes, the construction of new derivatives may not be possible. However, the construction of new mutant proteins for selective labeling, such as at free cysteines, may be well worth the effort.

4.5.6. Simulated tethering of the spin label and weighting of the data

For the purposes of the SPINGRIDD calculation, is possible to simulate the distance constraints that arise because of chemical connectivity by the inclusion of artificial line broadening parameters, *e.g.*, to atoms of the Cys-14 residue, even though the protons for this residue were not assigned. In fact, it is possible to choose atoms other than protons, such as the cysteine sulfur atom, for this treatment. The distance from the paramagnetic center to the Cys-14 sulfur atom, which I have used as an artificial tether, is not known. There is no precisely analogous compound in the Cambridge Crystallographic Database, but the coordinate set for a more distantly related molecule, di(2,2,6,6-tetramethyl-4-piperidinyl-1-oxyl) suberate, was used to provide a rough estimate of the required distance (Capiomont, 1972). A portion of this molecule may be considered isosteric with the cysteine acetamido Tempo and suggested a rough distance between the sulfur atom and the paramagnetic center of 7.5-8.8 Å. The sulfur atom of Cys-14 was then treated as magnetically equivalent to a proton, and the line broadening experienced by a proton at the mean of this distance range was entered in a line of the data input file as in Table 4.6. Note that it is not necessary to change the name of the sulfur atom; SPINGRIDD treats all atoms as protons, regardless of the atom name.

Distance ranges are not treated in an elegant fashion in SPINGRIDD. The concept of a range for the observed data is provided in the estimated error column of the input

data table and is expressed in line broadening, rather than distance space. Because of the r^{-6} dependence of the line broadening, it is not possible to express the error in the line broadening as a single number corresponding to the range of available distances. The current implementation is the most appropriate because the bulk of the data is expressed in line broadening space. Rather than work out an elegant solution to this problem, a number of input data sets were constructed in which the error limits (and the artificial line broadening) at the Cys-14 sulfur atom were varied.

The preceding paragraph only begins to address the assumptions made in the treatment of errors and weighting of the data by SPINGRIDD. Implicit in the expression of the error as a single value is the assumption that the errors are normally distributed; in this sense, the error may be regarded as a standard deviation of sorts. More importantly, the observed line broadening values in Table 4.5 are from a single measurement; multiple measurements were not possible due to the limited availability of the labeled protein sample at the time. This means that all of the "errors" in the input data sets are somewhat arbitrary estimates of the genuine errors. The relative size of the error, compared to the observed line broadening, was assigned based on criteria such as the signal to noise ratio of the observed peaks and the uncertainty introduced by spectral overlap. In practice, many different data sets were constructed in which the error estimates were varied, as was the case for the error for the Cys-14 sulfur pseudo broadening. It is of course not possible to carry out this procedure in an exhaustive manner—nor do the data warrant it. This remains a weak link in the treatment of the current NMR data from C38A-Am-Tempo.

The treatment of data from the vanished peaks constitutes perhaps an even weaker link. As pointed out earlier, it is possible to obtain a rough lower estimate of the broadening at these peaks; the true line broadening may be almost any value above the lower estimate, up to that experienced by a proton within van der Waals contact of the spin label. Therefore, with the current treatment of the data and weighting by errors in SPINGRIDD,

the lower bounds on line broadening have little utility. My approximate solution to this problem was to arbitrarily assign observed line broadening values between sixteen and thirty Hz and to assign relatively high errors so that the calculations are not heavily weighted by these values. In one data set, the line broadenings experienced by the vanished peaks were not included at all—only an artificial constraint to the Cys-14 sulfur atom was included.

Distance geometry provides a more natural solution to the problems posed by the vanished peaks. The lower bounds on the line broadenings may be reformulated as upper distance bounds, and the lower distance bounds may be taken, as is customary, as the sum of the van der Waals radii. Bound smoothing may then be used to reduce the upper bounds and raise the lower bounds were possible (Kuntz *et al.*, 1989). Like the restrained molecular dynamics approach, the distance geometry calculations are computationally very expensive compared to the much faster program SPINGRIDD. SPINGRIDD, despite its limitations, still provides the best first test of quality of the data.

4.5.7. Seven data sets designed to test the quality of the data

Six data sets were constructed, varying only the artificial line broadening applied to the Cys-14 sulfur atom. These data sets should account for differences between the approximate model discussed in the preceding paragraphs and more accurate descriptions of the "true" distance, including contributions from different conformations of the spin label molecule itself. A number of data sets are also necessary because of the nonlinear mapping of the distance to a line broadening factor. In each case, the error in the "line broadening" at this atom was arbitrarily set to one-fourth of the line broadening. A full set of distance constraints, with the amide proton records removed (because the information is redundant), from one of the files is given below in Table 4.7. A summary of the variation in the tethering constraint is presented in Table 4.8.

Table 4.7

Complete SPINGRIDD input constraints (α protons only)

Rec	NUM	NME	RES	RNUM	X	Y	Z	LB	LBER	CNT
ATOM	50	HA	ASP	3	35.144	9.745	-10.321	0.7	0.3	1
ATOM	66	HA	PHE	4	35.473	8.781	-5.789	0.7	30.0	1
ATOM	81	HA	CYS	5	30.964	10.266	-5.800	1.0	0.5	1
ATOM	94	HA	LEU	6	30.422	6.406	-8.558	1.7	0.8	1
ATOM	119	HA	GLU	7	32.837	4.055	-5.770	2.0	10.0	1
ATOM	169	HA	TYR	10	31.496	2.274	3.038	15.0	7.5	1
ATOM	185	HA	THR	11	29.423	3.601	6.562	15.0	7.5	1
ATOM	196	HA1	GLY	12	34.329	1.947	7.445	30.0	20.0	2
ATOM	217	SG	CYS	14	34.877	4.044	14.560	600.0	25.0	1
ATOM	251	HA	ALA	16	27.005	4.610	16.375	30.0	15.0	1
ATOM	267	HA	ARG	17	24.803	4.465	12.336	30.0	15.0	1
ATOM	288	HA	ILE	18	24.000	8.847	12.188	30.0	15.0	1
ATOM	307	HA	ILE	19	22.918	7.937	8.132	30.0	15.0	1
ATOM	354	HA	TYR	21	23.253	10.171	1.894	2.3	1.2	1
ATOM	374	HA	PHE	22	26.914	9.910	-0.832	1.7	0.8	1
ATOM	395	HA	TYR	23	24.818	10.226	-4.744	1.6	0.8	1
ATOM	412	HA	ASN	24	25.252	5.817	-5.749	3.0	99.0	1
ATOM	423	HA	ALA	25	25.558	7.786	-9.716	2.0	1.0	1
ATOM	437	HA	LYS	26	24.244	3.768	-11.951	2.0	4.0	1
ATOM	455	HA	ALA	27	20.090	5.015	-10.727	2.5	1.7	1
ATOM	464	HA1	GLY	28	22.992	9.579	-10.730	2.0	10.0	2
ATOM	475	HA	LEU	29	20.311	11.468	-7.252	2.0	10.0	1
ATOM	492	HA	CYS	30	22.835	10.632	-3.940	3.9	2.0	1
ATOM	506	HA	GLN	31	19.517	8.874	-1.128	4.5	2.2	1
ATOM	521	HA	THR	32	21.592	8.681	2.741	6.0	3.0	1
ATOM	539	HA	PHE	33	22.613	4.388	3.918	16.0	8.0	1
ATOM	555	HA	VAL	34	23.896	5.373	8.063	16.0	8.0	1
ATOM	576	HA	TYR	35	28.395	5.346	7.704	16.0	8.0	1
ATOM	639	HA	ALA	40	33.879	7.870	6.308	16.0	8.0	1
ATOM	653	HA	LYS	41	36.456	8.823	2.962	16.0	60.0	1
ATOM	677	HA	ARG	42	34.712	12.168	0.950	1.5	0.7	1
ATOM	698	HA	ASN	43	31.097	11.126	-1.506	1.6	0.8	1
ATOM	712	HA	ASN	44	28.831	10.413	2.219	1.6	0.8	1
ATOM	729	HA	PHE	45	28.527	14.889	2.741	1.8	0.9	1
ATOM	747	HA	LYS	46	25.458	14.851	5.822	1.8	0.9	1
ATOM	766	HA	SER	47	22.206	16.804	3.574	1.9	1.0	1
ATOM	776	HA	ALA	48	22.244	15.481	-0.310	1.2	0.6	1
ATOM	790	HA	GLU	49	21.885	19.789	-2.644	2.0	1.0	1
ATOM	805	HA	ASP	50	26.388	20.193	-1.435	1.3	0.6	1
ATOM	816	HA	CYS	51	26.840	15.447	-2.313	0.6	0.3	1
ATOM	832	HA	MET	52	24.027	15.986	-6.232	1.2	0.6	1
ATOM	857	HA	ARG	53	25.631	20.419	-6.896	2.3	1.2	1
ATOM	877	HA	THR	54	30.069	18.613	-6.258	0.9	0.5	1
ATOM	890	HA	CYS	55	29.253	14.775	-8.417	6.0	99.0	1
ATOM	899	HA1	GLY	56	25.526	14.160	-10.093	0.9	0.5	2
ATOM	906	HA1	GLY	57	23.651	17.180	-13.336	2.0	10.0	2
ATOM	915	HA	ALA	58	19.960	14.987	-13.050	0.5	1.0	1

The column names are as follows: REC - record; NUM - atom number (arbitrary); NME - atom name; RES - residue name; RNUM - residue number; X, Y, and Z - coordinates (Å); LB - observed line broadening; LBER - assigned error in line broadening determination; CNT - counter for coordinate averaging

Table 4.8
Tethering constraints at the Cys-14 SG atom

Data Set	LB (Hz)	LBERR (Hz)	Distance (Å)
A	600	150.0	6.11
B	100	25.0	8.24
C	50	12.5	9.25
D	30	7.5	10.07
E	20	5.0	10.78
F	10	2.5	12.10

The seventh data set was constructed to test the hypothesis that the most reliable data in the whole set come from the sequence between Cys-30 and Thr-32. Earlier, I made the argument that the broadening observed here was "anomalous", and probably due to intermolecular interactions. In data set G, only a small subset of the total is retained. Data from residues 21-23, 25, 30-32, 43-46, 51, and 52 are used. Significant broadening is observed at 30-32, while little broadening is observed at the other sites. The results from this data set were checked to see if any of the spin label positions generated from it could also explain the numerous vanishing peaks.

4.5.8. Calculated positions of the spin label

The coordinates of the paramagnetic center for each of the seven constraint sets described above were calculated starting from nine different initial coordinates. When the distance from the paramagnetic center to the Cys-14 sulfur atom was greater than 15 Å, the spin label coordinate was discarded from further consideration. In all cases, the two or three remaining minima were modeled from at least two different starting coordinates, indicating that the surface describing the optimal position of the spin label is rather smooth. The output from the SPINGRIDD program gives the calculated paramagnetic line broadening and the distance from the spin label of each of the atoms in the input coordi-

data set; proline residues were not included in this calculation, and the coordinate averaging was done over three atoms for the glycine residues. These distances were then compared to the distance of the average coordinate of the backbone atoms to the Cys-14 sulfur atom; these data are presented in Figure 4.8 A and B as "distance difference" plots.

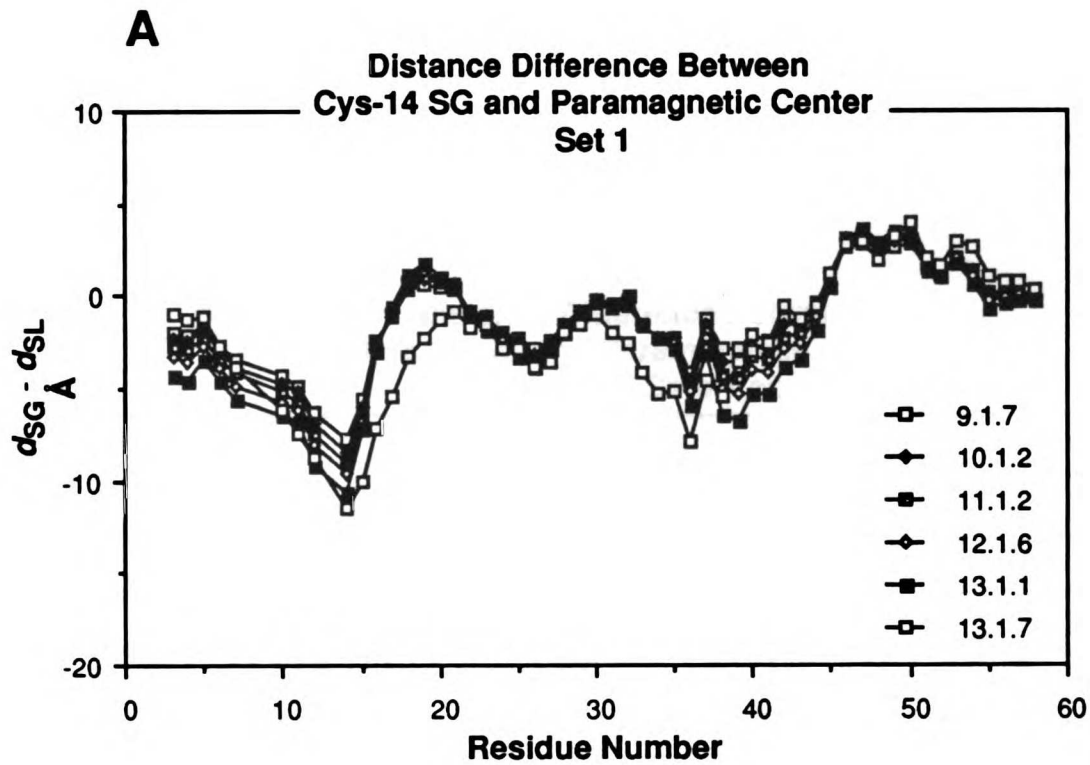
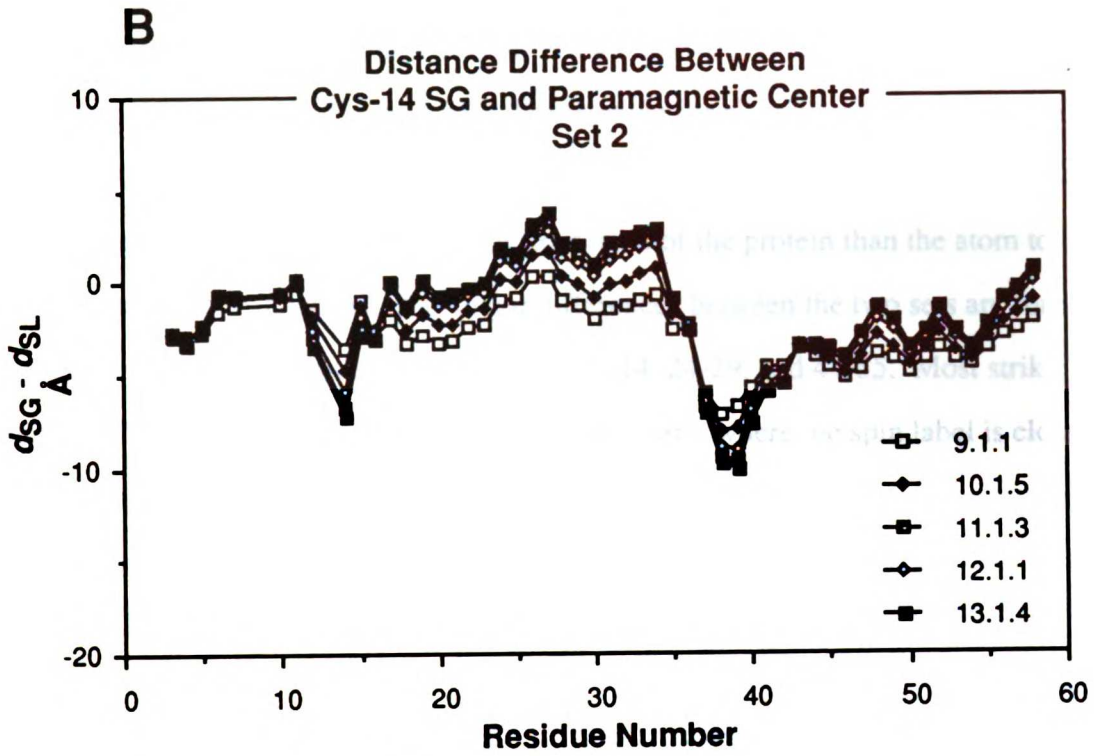


Figure 4.8: Difference distance plots. The distance from the average position of the amide and alpha proton atom from each residue in BPTI to the calculated spin label position and the sulfur atom of Cys-14 (4pti coordinates) are calculated. The difference between these distances is then plotted. Two clusters are observed; one cluster is plotted in (A), the second in (B), and the average of each of the two in (C).



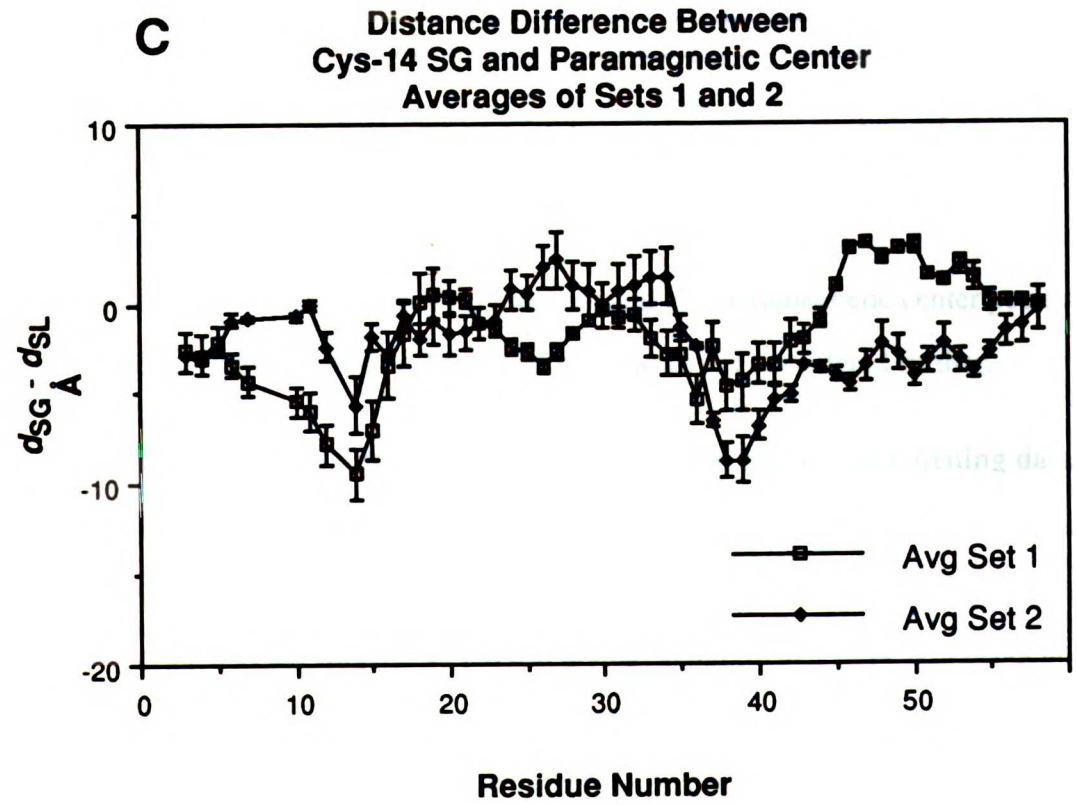


Figure 4.8, cont.

The designations of the coordinate sets in Figure 4.8 A and B are arbitrary. The outstanding feature of these plots is the clustering into two different sets. The average of each cluster is presented in Figure 4.8 C. The average for all residues is -1.6 ± 2.8 (SD) Å for set 1 and -2.1 ± 2.5 (SD) Å for set two; this indicates that the best fit position for the spin label is only slightly further away from most of the protein than the atom to which the spin label is attached. The major differences between the two sets are found in three regions encompassing the sequences from 6-14, 24-29, and 44-55. Most striking are the 44-55 region in set 1 and the 24-29 region in set 2 where the spin label is closer to the protein atoms than the Cys-14 sulfur atom. This indicates a position of the nitroxide portion of the Tempo moiety that is folded back on the protein. The average coordinate for each set in the 5pti coordinate system is given in Table 4.9 below.

Table 4.9
Coordinates of the spin label positions

Cluster		X	Y	Z
Set 1	Coord	32.004	13.205	17.601
	St. Dev. (6)	2.310	1.418	0.968
Set 2	Coord	28.823	-4.491	12.537
	St. Dev. (5)	2.693	0.196	0.757

The calculations from data set G did not yield any paramagnetic center positions closer to the Cys-14 SG atom than 15 Å, and these were not considered further.

4.5.8.1. The spin label coordinates do not fully describe the line broadening data

The SPINGRIDD program reports an error value that describes the quality of the fit of the spin label coordinate to the line broadening data; this error value is simply the sum of the errors at each of the fingerprint pseudo atoms. Because this error is dependent on the atom by atom line broadening error values and these vary among the input coordinate sets (see Table 4.7), the fitting error values can only be compared within the coordinate set generated from a single input coordinate set and not between sets. With the single ex-

ception of the spin label coordinate 13.1.7 (see Figure 4.8), all of the error values in set 1, Figure 4.8 are lower than the corresponding coordinate in set 2; coordinate 13.1.7 is clearly the outlier in set 1.

It is difficult to assess the quality of the fit of the paramagnetic center coordinate in a one dimensional error parameter. It is therefore necessary to back calculate the line broadening at each of the fingerprint pseudo atom coordinates and compare the calculated broadening with the observed broadening. This is how the individual error parameters are derived. These tables were then inspected for regions where the calculated broadening differed significantly from the observed broadening, indicating the failure of the model to explain the data.

As anticipated earlier, the line broadening observed between residues 30 thru 32 is difficult to explain with a single site model. This is clearly illustrated in Figure 4.9, where the calculated line broadening for each of the spin label positions previously analyzed in Figure 4.8 is plotted for residues 30 thru 34. It is clear that none of the spin label positions can explain the line broadening at these sites. This result, together with the calculations from data set G (where the primary constraints were to residues 30-34 and no "tethering" constraints to Cys-14 were included) indicate that a single position model is not consistent with the observed line broadening data.

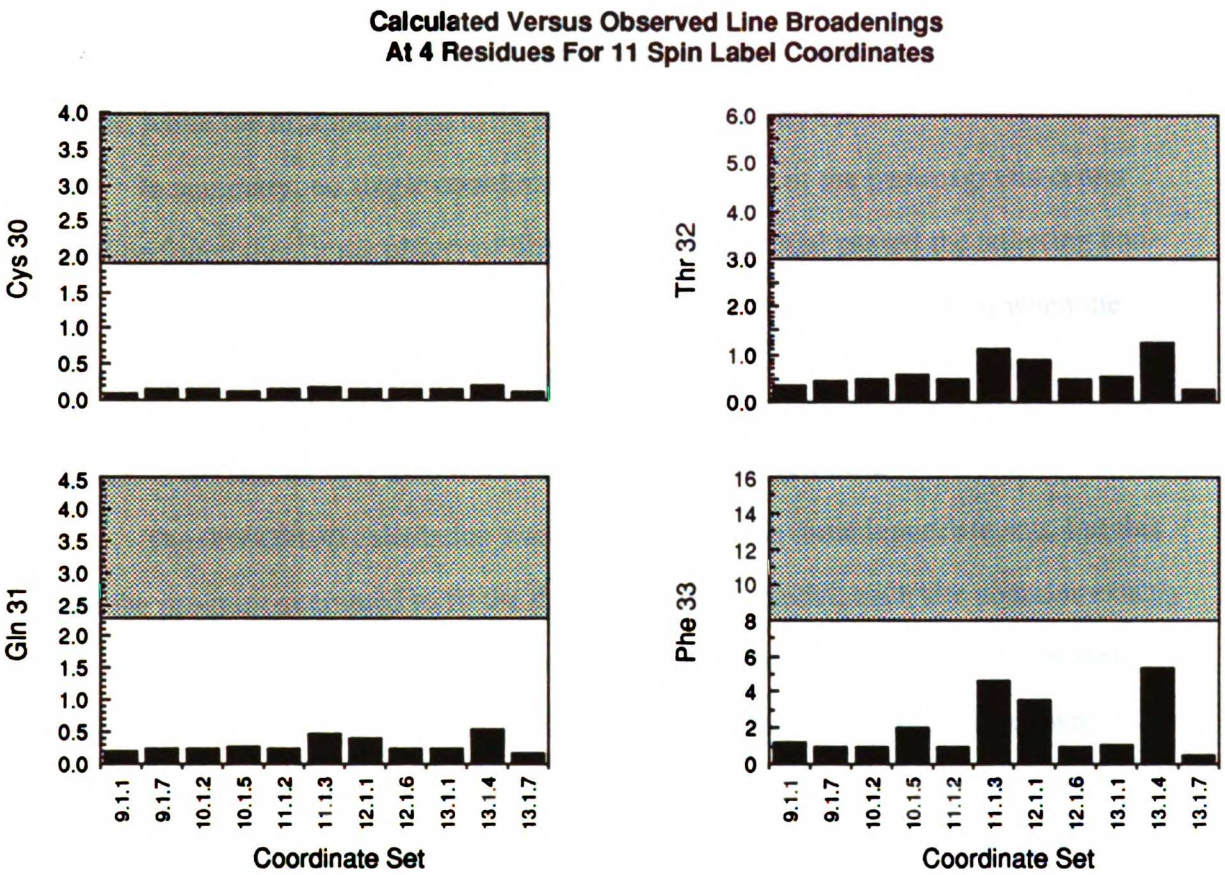


Figure 4.9: The calculated line broadenings in Hz due to a spin label at an arbitrarily designated coordinate at each of four amino acids is plotted in the bar graph. The shaded region represents the lower half of the error range assigned in the SPINGRIDD input data files. The uppermost value in each graph represents the observed line broadening.

The data from a number of the peaks which have vanished from the oxidized spectra are also not explained by any of the coordinate positions. Spin label coordinate set 2 explains the disappearance of the fingerprint peak of Thr-11 and Gly-12 (both weak in the reduced spectra) better than set 1, but the broadening at Ile-18 and Ile-19 is better explained by spin label set 1. The disappearance at Phe-33 is not explained by either cluster, while the disappearance at 34 is better explained by set 2.

In summary, no single coordinate for the position of the paramagnetic center could be found that fit the observed line broadening data and passed the tethering constraint for chemical sensibility. The inability to fit the data persisted even when the tethering constraint to the Cys-14 SG atom was removed.

4.5.8.2. Two site models also fail to explain the observed data

The SPINGRIDD calculation was repeated with the same input data sets, but this time the program attempted to fit the data to a two site model; each site contains exactly one half of the magnitude of the magnetic dipole compared to the single site model. Other fractional site models are certainly possible, including two site models with unequal spin populations, but are not warranted by the quality and quantity of the data. An initial starting coordinate was necessary for each of the two probe sites; these initial guesses were taken from the same file as was used in the one probe calculations.

Again, the error values between spin label coordinate data sets cannot be compared because of differences in the "observed" errors in the input data. In addition, a comparison of the "two spin" errors to the corresponding "one spin" errors is not warranted. Models with a greater number of parameters will always appear to fit the observed data better than those with fewer parameters, but that does not prove that the former model is warranted. Here, I have applied the same simple filter to test the results of the calculations for reasonableness that I used in the one spin case; *i.e.*, *both* spin probe sites must place the paramagnetic center within chemical bonding distance of the Cys-14

sulfur atom to the Tempo moiety, taken here as 15 Å. None of the calculated results passed this criterion.

4.5.9. Summary of the paramagnetic broadening analysis of C38A-Am-Tempo

At this juncture, the paramagnetic line broadening data from C38A-Am-Tempo cannot be used as a demonstration of the spin labeling approach for the extraction of a large number of quantitative long range distances in proteins. Qualitatively, the data is predictable, assuming a structure close to that observed in the crystal and the site of spin label attachment. However, a quantitative interpretation of the data was beset by difficulties arising from the quality of the data itself and the likely presence of intermolecular interactions whose effects must be factored out before reasonable distance information can be extracted from the data.

Apart from the theoretical problems described in Table 1.1, several additional problems were associated with the data itself. Of primary importance is the determination of the proper normalization factor required for a direct comparison of the peak intensities from the dia- and paramagnetic samples. Despite careful sample preparation, including a dilution of the protein by only 2% during the reduction of the spin label with ascorbate, it was not possible to derive a single normalization constant for data collected at two of three concentrations. Some of this affect may also arise from baseline distortions that remain in data acquired from H_2O solutions, even after baseline correction. This is only a remote possibility, since such baseline distortions were not readily apparent upon visual inspection of the data. An additional source of potential error is the peak integration routines in SPARKY which have not yet been fully validated; this includes the approximation of the peak shapes as gaussian, which they clearly are not, for the purpose of extracting data from overlapped peaks. However, this will still not account for the anomalously reduced intensities of the fingerprint cross peaks of Gln-31 and Thr-32 which are readily apparent upon inspection of Figure 4.7. Finally, there is a

problem, frequently encountered in multidimensional NMR data from proteins, of limited signal-to-noise ratios. This problem cannot easily be solved because of the necessary use of dilute solutions to avoid intermolecular artifacts and because of the long recycle times needed to ensure that the data are acquired from fully relaxed systems. Signal-to-noise problems varied from residue to residue due to the dependence of the peak intensities on conformation (Karplus, 1959), and were especially apparent at concentrations below 1 mM.

Peaks which have completely vanished from the spectra of the paramagnetic derivative complicate the interpretation of the data. The estimates I have made on the minimum broadening required for a peak to disappear provide a less than satisfactory quantitative treatment of data that is at best only qualitatively useful. These data suggest a limitation of the COSY experiment for the analysis of paramagnetic broadening in proteins with moderate correlation times and where the spin label is relatively immobile—the peak intensities (heights and volumes) appear to decrease too rapidly in the ≥8 Hz paramagnetic broadening range for this experiment to be used. This effect was not a large problem in the data from the previously described lysine derivatives (Kosen *et al.*, 1986) because there were far fewer missing peaks in that data, as discussed in section 4.5.3.

Estimates of the error of the paramagnetic line broadening measurement provide important data weighting parameters for the SPINGRIDD calculation but these data were only assumed and not determined from the data. The errors could certainly be determined by the collection of spectra from independent samples at otherwise nearly identical conditions, but at a high cost in instrument time and in protein, which is effectively destroyed by the ascorbate reduction. With the expression system described in chapter 2 and the labeling scheme described in chapter 3, the limitations in protein are less severe compared to the samples prepared by non-selective labeling (Kosen *et al.*, 1986), but substantial investments of time are still required.

Finally, a complete account of the possible intermolecular effects observed in the data C38A-Am-Tempo remains to be described. It is clear that one and two paramagnetic site models do not adequately explain the data. The major alternative explanation is that the data suggest a significant contribution from intermolecular interactions. Adequate models for this situation would include a description of the relative orientations of the interacting molecules, the equilibrium concentration of dimers and higher aggregates, and the exchange rates of the complexes. The paramagnetic line broadening NMR data are, and always will be, inadequate to describe these situations, so instead the intermolecular interactions must be factored out by extrapolation of data acquired at a number of different concentrations to zero concentration. My attempts at this have not been successful, so additional data must be acquired.

4.6. Future Directions

The use of the absolute value COSY experiment for the analysis of paramagnetic broadening has always been a compromise in which sensitivity of peak intensity to broadening was exchanged for an inability to separate the line broadening contribution from two or more J coupled nuclei. My efforts at protein deuteration were begun in an attempt to measure paramagnetic broadening in one dimensional spectra with improved resolution (compared to the natural abundance samples) (Schmidt and Kuntz, 1984), but these experiments have not been completed. In the interim, Gochin has developed a new experiment in which proton T_2's can be measured in a heteronuclear correlation experiment, provided that the heteronucleus does not have a large coupling to a like heteronucleus. This makes the experiment ideal for uniformly $^{15}N[\geq 98\%]$ labeled proteins; with the efficient expression system described in chapter 2 and the spin labeling chemistry described in chapter 3, this experiment has become very feasible and ought to be the first priority for further spin labeling experiments.

The ^{13}C version of Gochin's experiment has been applied to a 6 mM natural isotope abundance sample of BPTI spin labeled at the amino terminus. The results from these experiments are encouraging, though not all of the derived distances match predictions based upon the crystal structure (M. Gochin, personal communication). The major problems with the experiment arise again from intermolecular interactions, especially at a concentration of 6 mM, and from the relatively low signal-to-noise ratio of the spectra. The use of lower protein concentrations of course decreases the signal-to-noise, but the S/N can be increased by uniform ^{13}C enrichment. However, the experiment is limited to samples labeled at or below approximately 10% ^{13}C to minimize ^{13}C-^{13}C couplings. Under identical conditions, this will increase the signal-to-noise ratio tenfold over the natural abundance sample (1.1% ^{13}C). A clever method for uniform labeling of proteins expressed in *E. coli* with ^{13}C at or below the 50% level with suitable concentrations of [1-^{13}C, 99%] sodium acetate, [2-^{13}C, 99%] sodium acetate, and natural abundance sodium acetate as the sole carbon sources has recently been introduced (Venters *et al.*, 1991). This method is easily applicable to the production of 10% ^{13}C enriched and spin labeled BPTI. This should be the second priority for further spin labeling experiments with BPTI.

Finally, the third priority ought to be the design and production of a collection of new mutants of BPTI that will allow the attachment of spin labels at a number of sites. I would focus on mutants which retain the six cysteine residues of wild type BPTI—which will therefore form the three native disulfide bridges—and contain an additional cysteine residue which may be spin labeled. The additional free thiol may complicate the oxidative refolding of the mutant proteins, but in most cases the native structure will be strongly favored. These mutants will provide much more flexibility in terms of placement of the label and will allow the use of higher temperatures during the NMR experiments due to the greater stability of BPTI with all three native disulfides intact, compared to the two disulfide mutants. The use of elevated temperatures will contribute to the re-

duction, though not the elimination, of problems associated with intermolecular interactions.

4.6.1. A cautionary note on spin labeling in larger proteins

The work on spin labeled BPTI is still very much a "demonstration of the method" project in a model system. We hope to demonstrate that the spin labeling method can yield valuable long range distance information that will increase the precision of solution structure determinations and that the method is easily adaptable to other macromolecular systems. In this final short section, I will discuss some physical constraints that may limit the application of the technique.

The decrease in signal intensity due to antiphase cancellation that I have used in comparison of spectra from para- and diamagnetic spin label derivatives of BPTI also limits the application of homonuclear COSY-like experiments to small to moderate size proteins with linewidths less than about 12-15 Hz (T_2's of 21-26 msec). For a different reason—primarily due to pulse delays inherent in the experiment—Gochin's method for the measurement of proton T_2's is limited to values greater than or equal to 25 ms. This effectively limits the size of the protein to which the method can be applied. Even with extensively deuterated proteins, the T_2 for the residual protons, as measured by the heteronuclear correlation method, will be dominated by the attached heteronucleus; for the α proton, the ^{13}C alpha carbon magnetic dipole is equivalent to a proton 1.58 Å away (the sixth root of the square of the ratio of gyromagnetic ratios). Therefore, other methods will be required for the measurement of paramagnetic relaxation in larger proteins. Methods based on T_{1P} will probably not have enough sensitivity; the ratio $(1/T_{2P})/(1/T_1P)$ at large correlation times is directly proportional to $\tau_c^2\omega_I^2$, which is much greater than one. Direct measurement of linewidths in pure absorption phase two- and three-dimensional spectra, such as the NOESY experiment, will be beset by limited digital resolution and accuracy. Without unforeseen developments in experimental

methodology, it does not seem likely that distance determination via spin labeling will be practical for larger proteins which are currently the focus of much attention in the protein NMR field (Clore and Gronenborn, 1991a).

Chapter 5

Collaborations and conclusions

Introduction

In the course of any graduate study, there are inevitably odds and ends which don't fit into a full chapter of a Ph.D. thesis. In this last chapter, I present a brief description of some of those studies, with particular emphasis on the work of collaborators whose work relied on BPTI mutants I have produced in the new intracellular expression system. Following this discussion, I will present a summary of my work and a discussion of future directions.

5.1. Collaborations

Over the course of my studies, a group of closely associated collaborators at UCSF and Genentech formed the informal "BPTI group." At the nucleus of this group were the Anderson group at Genentech, with their interests in protein expression and folding, the Kuntz group at UCSF, whose primary interests were in NMR spectroscopy and protein folding, and the Kossiakoff group at Genentech, who pursued structural studies of BPTI by x-ray crystallography. Many of the studies of these groups were limited by the inability to produce a number of desirable mutants. Since my development of a new BPTI expression system, many of these studies have moved forward, and I take this chance to describe some of them.

5.1.1. Crystallography

The Kossiakoff group has been interested in the structural consequences of the removal of disulfide bonds from BPTI. With mutants produced from the old expression system, they were able to solve the structures of C30A/C51A (Eigenbrot *et al.*, 1990), C30V/C51A (Eigenbrot and Kossiakoff, unpublished), and more recently, the C14A/C38A mutant (Randal and Kossiakoff, unpublished). The primary conclusion from these studies was that changes in protein structure due to differences in crystal

packing interactions were often as large as those due to changes in primary structure (Kossiakoff, personal communication).

I have produced several 5/55 mutants that could not be produced in the previous expression systems (see chapter 2), and the structure of C5V/C55A was recently solved by Randal and Kossiakoff. Even though the 5/55 disulfide bond contributes more to the stability of BPTI than either the 30/51 or 14/38 disulfide bond (Creighton and Goldenberg, 1984), the structural changes due to this mutation are not large. Studies on the stability of the 5/55 mutants are currently under way in the Anderson lab at Rutgers.

We have also been interested in changes to residues in the core of BPTI that have large effects on the protein's stability. The wild type asparagine 43 residue of BPTI is completely buried and forms three hydrogen bonds to backbone atoms of Tyr-23 and Glu-7. Using the pHAZY expression system, I have produced the N43L mutant in order to study how the protein accommodates mutations which are not sterically compatible with the structure of the wild type protein. The structure of this mutant was recently solved to a resolution of 1.6 Å by Randal and Kossiakoff; the preliminary results from these studies are very interesting.

The thermal denaturation of the N43L protein at 25 μg/ml, pH 2.0 was studied by circular dichroism at 220 nm. The unfolding transition was reversible and gave a T_m of 35 °C. The T_m of the wild type protein could not accurately be measured under the same conditions, but the melting temperature was greater than 75 °C. A single mutation is therefore responsible for a decrease in thermal stability of over 40 °C.

The structures of the mutant and wild type protein near position 43 are shown in figure 5.1. Segments of the peptide chain from 4-8, 23-29, and 35-45 move slightly to accommodate the leucine side chain (data not shown). The most interesting feature of the structure is a close contact of 2.9 Å between a leucine methyl group and the carbonyl oxygen of Glu-7. A two dimensional grid search (in 10° increments) of possible χ_1 and

χ_2 angles for the leucine side chain failed to find conformations of the leucine side chain with less steric overlap.

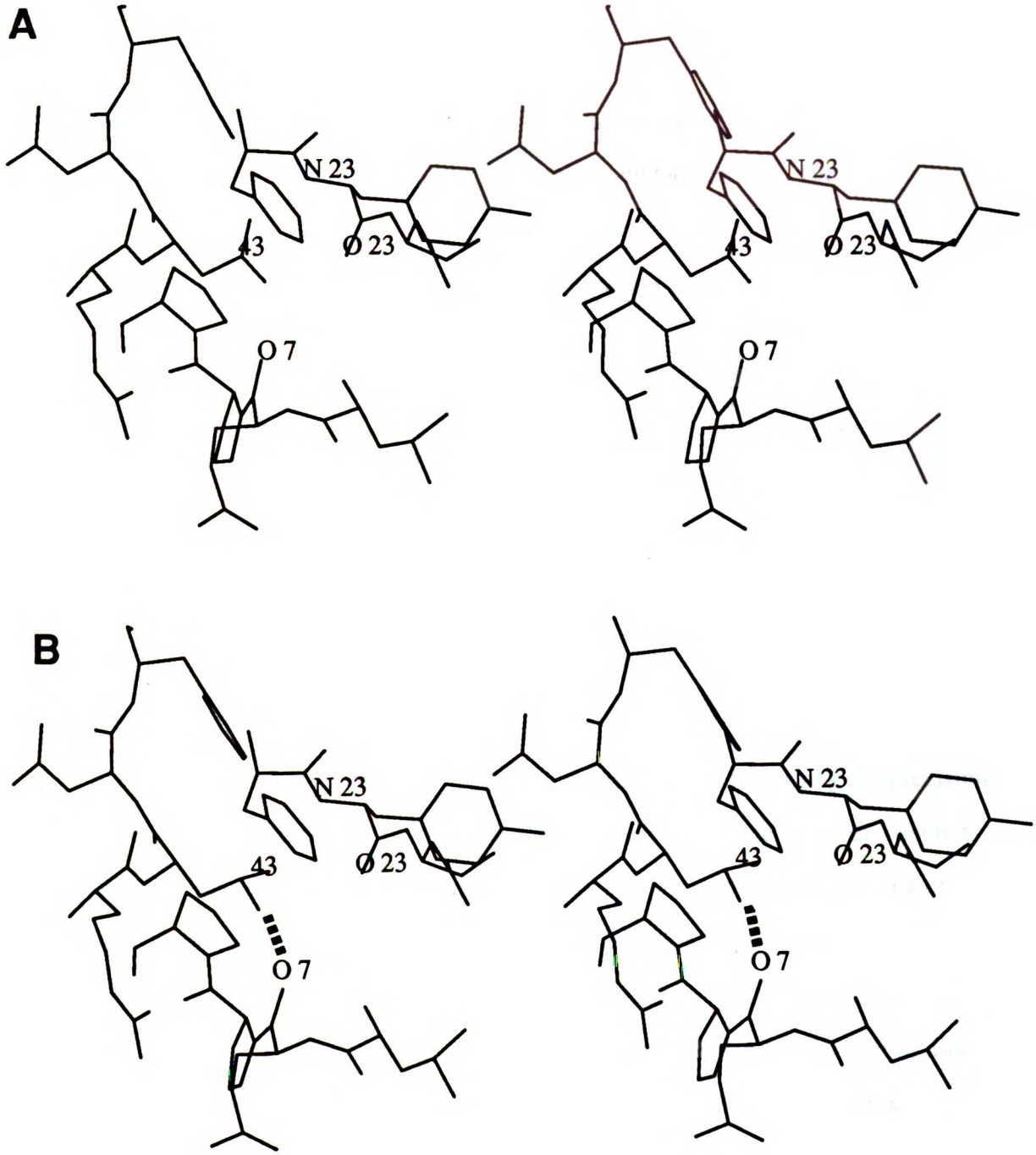


Figure 5.1: The structure of wild type BPTI and the N43L mutant near position 43. (A) The wild type protein. (B) The N43L protein. A close contact of 2.9 Å is observed between the CD2 methyl group of Leu-43 and the carbonyl oxygen of Glu-7.

5.1.2. Calorimetry

Stephen Anderson's lab has been studying the stability of a series of BPTI mutants, many of them produced in the pHAZY expression system, by differential scanning calorimetry. Their data are summarized in Table 5.1 below. The relative destabilization introduced by the removal of the 5/55 disulfide bond is consistent with data obtained from analysis of the BPTI folding pathway (Creighton and Goldenberg, 1984).

Table 5.1
Stabilities of BPTI mutants determined by Differential Scanning Calorimetry

	pH 3				pH 7			
	BPTI(wt)	A14/A38	A30/A51	V5/A55	BPTI(wt)	A14/A38	A30/A51	V5/A55
T_m (°C)	94	67	62	48	104	78	71	58
$\Delta H°_d$(kcal/mol)	69.2	47.8	58.6	47.2	~70	52.5	66.6	52.5
$\Delta S°_d$(cal/mol•K)	187	141	174	148	186	149	193	158
ΔH_{VH}(kcal/mol)	74.6	56.0	60.2	38.4	~70	60.4	70.9	52.2
$r(\Delta H°_d/\Delta H_{VH})$	1.07	1.17	1.02	1.24	~1	1.15	1.06	1.01

5.2. Conclusions

5.2.1. NMR and Biotechnology

In chapter 1, I reviewed a number of possible methods for increasing the precision of protein structure determination by NMR spectroscopy. Most of the methods—including the spin labeling experiments I proposed—rely on biosynthetic isotope labeling of heterologously expressed proteins. Protein structure determination by NMR spectroscopy has therefore become a multidisciplinary enterprise, requiring expertise in biotechnology and protein engineering, as well as advanced NMR methods. Until now, most NMR labs have gotten by through collaborations (*e.g.*, see (Montelione *et al.*, 1989)), but the experience of many x-ray crystallography labs shows the value of "in house" control and mastery of enough biotechnology to express mutants and isotope labeled samples of proteins under study in the lab.

My studies with BPTI were in fact slowed by an expression system which was not robust enough to stand up to the rigors of labeling in minimal media. My experience illustrates the difference between an expression system that is capable of producing adequate amounts of protein under optimized conditions—*e.g.*, in an industrial fermenter—and an expression system optimized for the production of protein, preferably in shake flasks, in minimal media. I have therefore concentrated on the development of a well controlled expression system that is capable of producing virtually any mutant of BPTI, even in minimal media. This system was described in detail in chapter 2. I further emphasize the success of my expression system by presenting in figure 5.2 a comparison of a section of the one-dimensional NMR spectrum of the C30V/C51A mutant at natural proton abundance with its 85% deuterated counterpart in 10% D_2O solution. I have not yet completed a detailed analysis of the deuterated molecule, including characterization of two dimensional spectra acquired from it, nor have I produced a spin labeled analog of the deuterated protein. The spectra are presented here merely to show that it is now possible to isotope label BPTI under the most demanding conditions—the deuterium isotope effect on the growth of *E. coli* is much greater than that for either ^{13}C or ^{15}N—and that it can be done with a reasonable amount of effort and expense. With BPTI, the question now becomes "which experiments should we do?" rather than "which experiments can we do?"

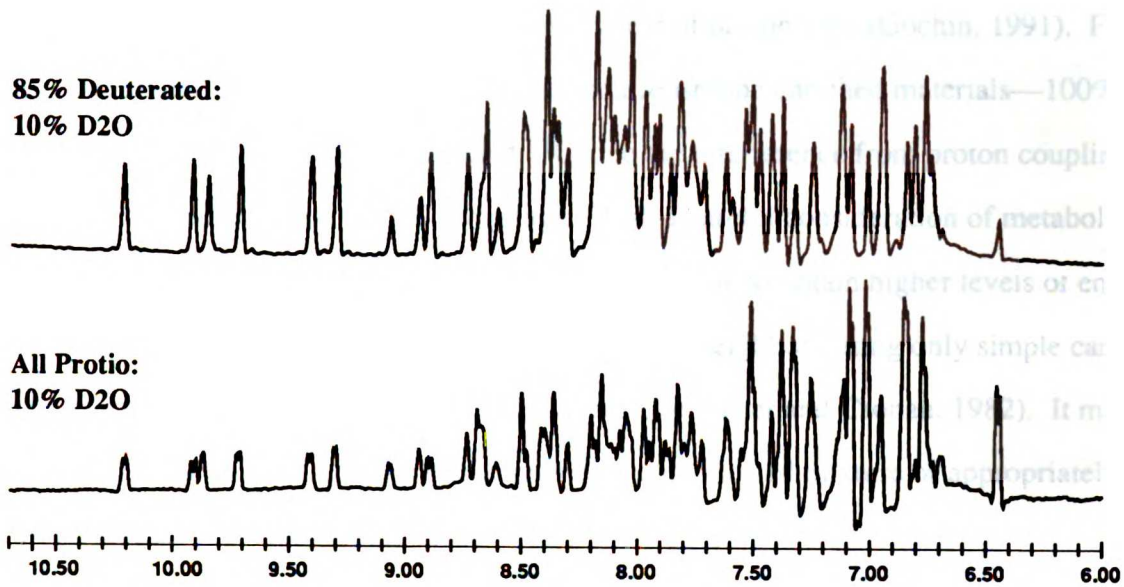


Figure 5.2: Comparison of the amide and aromatic region of a natural abundance and an 85% deuterated sample of BPTI(C30V/C51A). Spectra were acquired at 500 MHz, pH 4.6, 36 °C at approximately 2 mM protein concentration in 90% H_2O and 10% D_2O. All amide protons in the deuterated sample are fully protonated (to the level of the solvent). Comparison of the integrated intensity of the Tyr-23 H_ε resonance—the most downfield in the spectrum—to the Tyr-23 HN resonance—the furthest upfield resonance—gives the approximate level of deuteration. Note that all of the resolved doublets in the all protio control spectrum are singlets in the spectrum of the deuterated molecule.

5.2.2. Which experiments should be done?

At the beginning of my work, my answer to the question "which experiments should I do?" was to acquire and analyze one dimensional NMR data from a deuterated and spin labeled sample of BPTI. I didn't quite get to this end, but I have developed important methodologies which allow that end to be reached. The pHAZY expression system is obviously important, but the introduction and practical realization of the concept of nitroxide labeling at free thiols introduced by mutagenesis should not be overlooked as a contribution to the success of future spin labeling efforts.

In chapter 4, I concluded that attempts to acquire long range distance data from paramagnetic samples of deuterated proteins should be abandoned in favor of Gochin's new heteronuclear experiment for the measurement of proton T_2's (Gochin, 1991). For adequate signal-to-noise, these experiments require isotope enriched materials—100% for ^{15}N but only 10% for uniformly labeled ^{13}C, due the interference from proton coupling interactions mediated by ^{13}C-^{13}C couplings. With a careful consideration of metabolic pathways and appropriate strains of *E. coli* , it is possible to obtain higher levels of enrichment at specific positions with relatively few ^{13}C neighbors using only simple carbon precursors such as lactate, succinate, and acetate (LeMaster and Cronan, 1982). It may also be possible to label specific positions with a complicated mixture of appropriately carbon labeled precursor amino acids.

All of these experiments can now be done with BPTI and mutants such as C38A. However, I now believe that staphylococcal nuclease (SNase) has become a better system for further testing long range distance determination by spin labeling. There are three factors which make the SNase system as good or better than BPTI for these experiments: (1) a full set of proton and ^{15}N assignments are now available for SNase (Torchia *et al.*, 1989; Wang *et al.*, 1990), where few if any were available in 1985; (2) the expression of SNase is superior to BPTI, even with the pHAZY expression system (Hibler *et al.*, 1989); (3) wild type SNase does not have any cysteine residues—this allows introduction of cysteines by mutagenesis without disturbing disulfide folding pathways. The absence of disulfide bonds in the native protein has two additional advantages: (1) it allows inclusion of mercaptans such as β-mercaptoethanol or DTT^{SH}_{SH} throughout the purification to protect the new thiol introduced by mutagenesis, and (2) it facilitates possible recycling of proteins modified by the 1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methane thiolsulfonate reagent that I recommended in the appendix to chapter 3.

5.2.3. Spin labeling versus the heteronuclear revolution

The last 3-5 years have seen spectacular advances in heteronuclear two-, three-, and four-dimensional NMR spectroscopy that have allowed structures to be determined with increased precision (Clore and Gronenborn, 1991b). How does spin labeling compare and what will be its role in the future?

There is no question that the spin labeling method has yet to receive a clear and convincing demonstration that can compete with the success of the more general heteronuclear methods. All of the previous experiments have involved some sort of compromise—from analysis of poorly resolved one-dimensional paramagnetic difference spectra to unresolved effects in homonuclear two-dimensional COSY spectra. Gochin's experiment, combined with isotopically enriched samples, should finally provide the definitive test case. Early results from natural abundance samples suggest that conformational exchange and intermolecular interactions may prevent straightforward interpretation of the data, and that not all spin label derivatives will provide useful data, decreasing the general utility of the method. A final factor argues against the general application of spin labels for improving NMR structures: since it appears that isotope labeling is necessary to obtain good spin label data, as much data as is possible should be squeezed out of the isotope labeled sample before a spin labeling experiment is done. These higher priority experiments take a considerable amount of time to perform and interpret and may leave little time for spin labeling experiments.

Spin labeling may be worthwhile in special cases, such as the analysis of protein loop regions with low proton densities, using data from a spin label at a more distant site (T. James, personal communication). The analysis of nucleic acid structures, which reveal generally few tertiary interactions in their NMR spectra, may also be significantly aided by paramagnetic labeling.

Despite these generally pessimistic remarks, I urge my scientific successors to pursue the heteronuclear spin labeling experiments. These experiments will be the definitive and fairest test case of the spin label method and will in the worst case provide new and important data on proton relaxation and protein dynamics.

References

Abola, E. E., *et al.* (1987). in ***Crystallographic Databases -Information Content, Software Systems, Scientific Applications***. (Allen, F. H., Bergerhoff, G. and Seivers, R., eds.) Data Commission of the Int'l Union of Crystallography, Bonn/Cambridge/Chester, pp. 107-132.

Abrahmsén, L., Moks, T., Nilsson, B. & Uhlén, M. (1986). "Secretion of heterologous gene products to the culture medium of *Escherichia coli*." *Nucl. Acids Res.*, **14**, 7487-7500.

Altman, J. D., *et al.* (1991). "Intracellular expression of BPTI fusion proteins and single column cleavage/affinity purification by chymotrypsin." *Protein Engin.*, **4**, 593-600.

Anglister, J. (1990). "Use of deuterium labelling in NMR studies of antibody combining site structure." *Q. Rev. Biophys.*, **23**, 175-203.

Arseniev, A., *et al.* (1988). "Three-dimensional structure of rabbit liver [Cd7]metallothionein-2a in aqueous solution determined by nuclear magnetic resonance." *J. Mol. Biol.*, **201**, 637-657.

Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976). "Two-dimensional spectroscopy. Application to nuclear magnetic resonance." *J. Chem. Phys.*, **64**, 2229-2246.

Ausubel, F. M., *et al.* (1987). **Current Protocols in Molecular Biology**. Wiley and Sons, New York.

Basus, V. J. (1989). "Proton nuclear magnetic resonance assignments." *Methods Enzymol.*, **177**, 132-149.

Basus, V. J., *et al.* (1988). "Structural Studies of alpha-Bungarotoxin. 1. Sequence-Specific ^{1}H NMR Resonance Assignments." *Biochemistry*, **27**, 2763-2771.

Bax, A. & Davis, D. G. (1985). "MLEV-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy." *J. Magn. Reson.*, **65**, 355-360.

Berliner, L. J., Grunwald, J., Hankovszky, H. O. & Hideg, K. (1982). "A Novel Reversible Thiol-Specific Spin Label: Papain Active Site Labeling and Inhibition." *Anal. Biochem.*, **119**, 450-455.

Bernstein, F. C., *et al.* (1977). *J. Mol. Biol.*, **112**, 535-542.

Billeter, M., Braun, W. & Wüthrich, K. (1982). "Sequential Resonance Assignments in Protein ^{1}H Nuclear Magnetic Resonance Spectra: Computation of Sterically Allowed Proton-Proton Distances and Statistical Analysis of Proton-Proton Distances in Single Crystal Protein Conformations." *J. Mol. Biol.*, **155**, 321-346.

Bjorkman, P. J., *et al.* (1987). "Structure of the human class I histocompatibility antigen, HLA-A2." *Nature*, **329**, 506-512.

Boelens, R., *et al.* (1989). "Iterative procedure for structure determination from proton-proton NOEs using a full relaxation matrix approach. Application to a DNA octamer." *J. Magn. Reson.*, **82**, 290-308.

Borgias, B. A. & James, T. L. (1988). "COMATOSE: A Method for Constrained Refinement of Macromolecular Structure Based on Two-Dimensional Nuclear Overhauser Effect Spectra." *J. Magn. Reson.*, **79**, 493-512.

Borgias, B. A. & James, T. L. (1989). "Two-dimensional nuclear Overhauser effect: complete relaxation matrix analysis." *Methods Enzymol.*, **176**, 169-183.

Borgias, B. A. & James, T. L. (1990). "MARDIGRAS - a procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure." *J. Magn. Reson.*, **87**, 475-487.

Braun, W., Wider, G., Lee, K. H. & Wüthrich, K. (1983). "Conformation of glucagon in a lipid-water interphase by ^{1}H nuclear magnetic resonance." *J. Mol. Biol.*, **169**, 921-948.

Brosius, J. (1988). "Expression vectors employing lambda-, *trp*-, *lac*-, and *lpp*-derived promoters" in ***Vectors: A survey of molecular cloning vectors and their uses.*** (Rodriguez, R. L. and Denhardt, D. T., eds.) Butterworths, Boston.

Butler, P. J. G. & Hartley, B. S. "Maleylation of Amino Groups." *Methods Enzymol.*, XXV, 191-199.

Bystrov, V. F. (1976). "Spin-spin coupling and the conformational states of peptide systems." *Prog. NMR Spectros.*, **10**, 41-82.

Campbell, I. D., Dobson, C. M. & Williams, R. J. P. (1975). "Assignment of the ^{1}H NMR spectra of proteins." *Proc. R. Soc. Lond. A.*, **345**, 23-40.

Campbell, I. D., Dobson, C. M., Williams, R. J. P. & Xavier, A. B. (1973). "Resolution enhancement of protein PMR spectra using the difference between a broadened and a normal spectrum." *J. Magn. Reson.*, **11**, 172-181.

Capiomont, A. (1972). "Structure cristalline du radical nitroxide: suberate de ditétraméthyl-2,2,6,6 piperidinyl-4-oxyl-1." *Acta Cryst.*, **B28**, 2298-2301.

Carter, P. (1990). "Site-Specific Proteolysis of Fusion Proteins" in ***Protein Purification: From Molecular Mechanisms to Large-Scale Processes.*** (Ladisch, M. R., Willson, R. C., Painton, C.-d. C. and Builder, S. E., eds.) American Chemical Society Symposium Series, pp. 181-193.

Carter, P., *et al.* (1989). "Engineering subtilisin BPN' for site-specific proteolysis." *Proteins*, **6**, 240-248.

Carter, P. & Wells, J. A. (1987). "Engineering enzyme specificity by "substrate-assisted catalysis"." *Science*, **237**, 394-399.

Castagnoli, L., *et al.* (1989). "Genetic and structural analysis of the ColE1 Rop (Rom) protein." *EMBO J.*, **8**, 621-629.

Chang, B. Y. & Doi, R. H. (1990). "Overproduction, purification, and characterization of Bacillus subtilis RNA polymerase sigma A factor." *J. Bacteriol.*, **172**, 3257-3263.

Chang, C. N., *et al.* (1987). "High-level secretion of human growth hormone by *Escherichia coli.*" *Gene*, **55**, 189-196.

Chazin, W. J., Goldenberg, D. P., Creighton, T. E. & Wüthrich, K. (1985). "Comparative studies of conformaiton and internal mobility in native and circular basic pancreatic trypsin inhibitor by ^{1}H nuclear magnetic resonance in solution." *Eur. J. Biochem.*, **152**, 429-437.

Cleland, W. W. (1964). *Biochemsitry*, **3**, 480-482.

Clore, G. M., *et al.* (1990a). "Assignment of the side-chain ^{1}H and ^{13}C resonances of interleukin-1β using double- and triple-resonance heteronuclear three-dimensional NMR spectroscopy." *Biochemistry*, **29**, 8172-8184.

Clore, G. M., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990b). "Analysis of the backbone dynamics of interleukin-1β using two-dimensional inverse detected heteronuclear ^{15}N-^{1}H NMR spectroscopy." *Biochemistry*, **29**, 7387-7401.

Clore, G. M., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990c). "Low resolution structure of interleukin-1β in solution derived from ^{1}H-^{15}N heteronuclear three-dimensional nuclear magnetic resonance spectroscopy." *J. Mol. Biol.*, **214**, 811-817.

Clore, G. M. & Gronenborn, A. M. (1982). "Theory and applications of the transferred nuclear Overhauser effect to the study of the conformations of small ligands bound to proteins." *J. Magn. Reson.*, **48**, 402-417.

Clore, G. M. & Gronenborn, A. M. (1983). "Theory of the time-dependent transferred nuclear Overhauser effect: applications to structural analysis of ligand-protein complexes in solution." *J. Magn. Reson.*, **53**, 423-442.

Clore, G. M. & Gronenborn, A. M. (1989a). "Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy." *Crit. Rev. Biochem. Mol. Biol.*, **24**, 479-564.

Clore, G. M. & Gronenborn, A. M. (1989b). "How accurately can interproton distances in macromolecules really be determined by full relaxation matrix analysis of nuclear overhauser enhancement data." *J. Magn. Reson.*, **84**, 398-409.

Clore, G. M. & Gronenborn, A. M. (1991a). "Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy." *Science*, **252**, 1390-1399.

Clore, G. M. & Gronenborn, A. M. (1991b). "Two-, three- and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution." *Annu. Rev. Biophys. Biophys. Chem.*, **20**, 29-63.

Clore, G. M., Kay, L. E., Bax, A. & Gronenborn, A. M. (1991a). "Four-dimensional ^{13}C-^{13}C-edited nuclear Overhauser enhancement spectroscopy of a protein in solution - application to interleukin 1β." *Biochemistry*, **30**, 12-18.

Clore, G. M., Wingfield, P. T. & Gronenborn, A. M. (1991b). "High-Resolution Three-Dimensional Structure of Interleukin 1β in Solution by Three- and Four-Dimensional Nuclear Magnetic Resonance Spectroscopy." *Biochemistry*, **30**, 2315-2323.

Conner, G. E. & Udey, J. A. (1990). "Expression and refolding of recombinant human fibroblast procathepsin D." *DNA Cell Biol.*, **9**, 1-9.

Coplen, L. J., Frieden, R. W. & Goldenberg, D. P. (1990). "A Genetic Screen to Identify Variants of Bovine Pancreatic Trypsin Inhibitor With Altered Folding Energetics." *Proteins: Struct. Funct. Genet.*, **7**, 16-31.

Creighton, T. E. (1974). "Renaturation of Reduced Bovine Pancreatic Trypsin Inhibitor." *J. Mol. Biol.*, **87**, 563-577.

Creighton, T. E. (1975). "The Two-disulfphide Intermediates and the Foldign Pathway of Reduced Pancreatic Trypsin Inhibitor." *J. Mol. Biol.*, **95**, 167-199.

Creighton, T. E. (1977a). "Conformational Restrictions of the Pathway of Folding and Unfolding of the Pancreatic Trypsin Inhibitor." *J. Mol. Biol.*, **113**, 275-293.

Creighton, T. E. (1977b). "Energetics of Folding and Unfolding of Pancreatic Trypsin Inhibitor." *J. Mol. Biol.*, **113**, 295-312.

Creighton, T. E. (1980). "Role of the environment in the refolding of reduced pancreatic trypsin inhibitor." *J. Mol. Biol.*, **144**, 521-550.

Creighton, T. E. (1984a). "Disulfide Bond Formation in Proteins." *Methods Enzymol.*, **107**, 305-329.

Creighton, T. E. (1984b). ***Proteins. Structure and Molecular Principles.*** Freeman, New York, N. Y.

Creighton, T. E. (1989). "Disulphide Bonds between Cysteine Residues" in ***Protein Structure: A Practical Approach.*** (Creighton, T. E., es.) IRL Press, Oxford, pp. 155-167.

Creighton, T. E. & Charles, I. G. (1987). "Biosynthesis, processing, and evolution of bovine pancreatic trypsin inhibitor." *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 511-519.

Creighton, T. E. & Goldenberg, D. P. (1984). "Kinetic Role of a Meta-stable Native-like Two-disulfide Species in the Folding Transition of Bovine Pancreatic Trypsin Inhibitor." *J. Mol. Biol.*, **179**, 497-526.

Crespi, H. L., Daboll, H. F. & Katz, J. J. (1970). "The biosynthesis of isotope hybrid proteins for high resolution nuclear magnetic resonance studies by incorporation of [^1H] amino acids into fully deuterated algae." *Biochim. Biophys. Acta*, **200**, 26-33.

Crespi, H. L., Kostka, A. G. & Smith, U. H. (1974). "Proton magnetic resonance observations of hydrogen exchange rates and secondary structure in algal ferredoxin." *Biochem. Biophys. Res. Commun.*, **61**, 1407.

Crespi, H. L., Norris, J. R., Bays, J. P. & Katz, J. J. (1973). *Ann. N. Y. Acad. Sci.*, **222**, 800.

Crespi, H. L., R., N. J. & Katz, J. J. (1972). "Magnetic resonance of isotope hybrid flavoprotein ^{2}H-flavoprotein (^{1}H-flavin mononucleotide)." *Nature New Biol.*, **236**, 178-180.

Crespi, H. L., Rosenberg, R. M. & Katz, J. J. (1968). "Proton magnetic resonance of proteins fully deuterated except for ^{1}H-leucine side chains." *Science*, **161**, 795-796.

Dahlman, K., *et al.* (1989). "High level expression in *Escherichia coli* of the DNA-binding domain of the glucocorticoid receptor in a functional form utilizing domain-specific cleavage of a fusion protein." *J. Biol. Chem.*, **264**, 804-809.

Darby, N. J., van Mierlo, C. P. M. & Creighton, T. E. (1991). "The 5-55 single-disulphide intermediated in folding of bovine pancreatic trypsin inhibitor." *FEBS Lett.*, **279**, 61-64.

de Boer, H. A., Comstock, L. J. & Vasser, M. (1983). "The *tac* promoter: A functional hybrid derived from the *trp* and *lac* promoters." *Proc. Natl. Acad. Sci. USA*, **80**, 21-25.

Deisenhofer, J. & Steigmann, W. (1975). "Crystallographic Refinement Of The Structure Of Bovine Pancreatic Trypsin Inhibitor At 1.5 Angstroms Resolution." *Acta Crystallogr. sect. B*, **31**, 238-240.

Denk, W., Baumann, R. & Wagner, G. (1986). "Quantitative evaluation of cross-peak intensities by projection of two-dimensional NOE spectra on a linear space spanned by a set of reference resonance lines." *J. Magn. Reson.*, **67**, 386-390.

Denk, W. G., Wagner, G., Rance, M. & Wüthrich, K. (1985). "Combined suppression of diagonal peaks and t_1 ridges in two-dimensional nuclear Overhauser enhancement spectra." *J. Magn. Reson.*, **62**, 350-355.

Derome, A. (1987). ***Modern NMR Techniques for Chemistry Research***. Pergamon, Oxford.

DesJarlais, R. L., *et al.* (1990). "Structure-based design of nonpeptide inhibitors specific for the human immunodeficiency virus 1 protease." *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 6644-6648.

Driscoll, P. C., Clore, G. M., Beress, L. & Gronenborn, A. M. (1989a). "A proton nuclear magnetic resonance study of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*: sequential and stereospecific resonance assignment and secondary structure." *Biochemistry*, **28**, 2178-2187.

Driscoll, P. C., Gronenborn, A. M., Beress, L. & Clore, G. M. (1989b). "Determination of the three-dimensional solution structure of the antihypertensive and antiviral protein BDS-I from the sea anemone *Anemonia sulcata*: a study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing." *Biochemistry*, **28**, 2188-2198.

Driscoll, P. C., Gronenborn, A. M. & Clore, G. M. (1989c). "The influence of stereospecific assignments on the determination of three-dimensional structures of proteins by nuclear magnetic resonance spectroscopy. Application to the sea anemone protein BDS-I." *FEBS Lett.*, **243**, 223-233.

Eads, C. D. & Kuntz, I. D. (1989). "Programs for Computer-Assisted Sequential Assignment of Proteins." *J. Magn. Reson.*, **82**, 467-482.

Edison, A. S., Westler, W. M. & Markley, J. L. (1991). "Elucidation of Amino Acid Spin Systems in Proteins and Determination of Heteronuclear Coupling Constants by Carbon-Proton-Proton Three-Dimensional NMR." *J. Magn. Reson.*, **92**, 434-438.

Eigenbrot, C., Randal, M. & Kossiakoff, A. A. (1990). "Structural Effects Induced by Removal of a Disulfide Bridge; The X-ray Structure of the C30A/C51A Mutant of Basic Pancreatic Trypsin Inhibitor at 1.6 Å." *Protein Eng.*, **3**, 591-598.

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

Ernst, R. R., Bodenhausen, G. & Wokaun, A. (1987). ***Principles of Nuclear Magnetic Resonance in One and Two Dimensions***. Clarendon, Oxford.

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

Fersht, A. R. & Renard, M. (1974). *Biochemistry*, **13**, 1416.

Fesik, S. W. & Zuiderweg, E. R. P. (1988). "Heteronuclear three-dimensional NMR spectroscopy: a strategy for the simplification of homonuclear two-dimensional spectra." *J. Magn. Reson.*, **78**, 588.

Fesik, S. W. & Zuiderweg, E. R. P. (1989). "An approach for studying the active site of enzyme inhibitor complexes using deuterated ligands and 2D NOE difference spectroscopy." *J. Am. Chem. Soc.*, **111**, 5013-5015.

Fesik, S. W. & Zuiderweg, E. R. P. (1990). "Heteronuclear three-dimensional NMR spectroscopy of isotopically labelled biological macromolecules." *Q. Rev. Biophys.*, **23**, 97-131.

Folkers, P. J., *et al.* (1989). "Solution structure of recombinant hirudin and the Lys-47→Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study." *Biochemistry*, **28**, 2601-2617.

Forman-Kay, J. D., *et al.* (1989). "A proton nuclear magnetic resonance assignment and secondary structure determination of recombinant human thioredoxin." *Biochemistry*, **28**, 7088-7097.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T. & Gronenborn, A. M. (1991). "High-Resolution Three-Dimensional Structure of Reduced Recombinant Human Thioredoxin in Solution." *Biochemistry*, **30**, 2685-2698.

Forman-Kay, J. E., *et al.* (1990). "Studies on the solution conformation of human thioredoxin using heteronuclear ^{15}N-^{1}H nuclear magnetic resonance spectroscopy." *Biochemistry*, **29**, 1566-1572.

Gallagher, W. H. & Woodward, C. K. (1989). "The concentration dependence of the diffusion coefficient for bovine pancreatic trypsin inhibitor: a dynamic light scattering study of a small protein." *Biopolymers*, **28**, 2001-2024.

Gochin, M. (1991). "Determination of proton transverse relaxation times in homonuclear coupled spin systems." *J. Magn. Reson.*, in press.

Gochin, M., Zon, G. & James, T. L. (1990). "Two-Dimensional COSY and Two-Dimensional NOE Spectroscopy of $d(AC)_4 \cdot d(GT)_4$. Extraction of Structural Constraints." *Biochemistry*, **29**, 11161-11171.

Goldenberg, D. P. (1988). "Kinetic Analysis of the Folding and Unfolding of a Mutant Form of Bovine Pancreatic Trypsin Inhibitor Lacking the Cysteine-14 and -38 Thiols." *Biochemistry*, **27**, 2481-2489.

Gonzalez, C., *et al.* (1991). "Toward An NMR R Factor." *J. Magn. Reson.*, **91**, 659-664.

Gray, G. L., *et al.* (1985). *Gene*, **39**, 247.

Griesinger, C., Sørenson, O. W. & Ernst, R. R. (1985). "Two-Dimensional Correlation of Connected NMR Transitions." *J. Am. Chem. Soc.*, **107**, 6394-6396.

Gronenborn, A. & Clore, G. M. (1985). "Investigation of the solution structures of short nucleic acid fragments by means of NOE measurements." *Prog. Nucl. Magn. Reson. Spectrosc.*, **17**, 1-32.

Gronenborn, A. M., Bax, A., Wingfield, P. T. & Clore, G. M. (1989). "A powerful method of sequential proton resonance assignment in proteins using relayed ^{15}N-^{1}H multiple quantum coherence spectroscopy." *FEBS Lett.*, **243**, 93-98.

Grumont, R., Sirawaraporn, W. & Santi, D. V. (1988). "Heterologous expression of the bifunctional thymidylate synthase-dihydrofolate reductase from Leishmania major." *Biochemistry*, **27**, 3776-3784.

Guiles, R. D., Altman, J., Kuntz, I. D. & Waskell, L. (1990). "Structural studies of cytochrome b5: complete sequence-specific resonance assignments for the trypsin-solubilized microsomal ferrocytochrome b5 obtained from pig and calf." *Biochemistry*, **29**, 1276-1289.

Güntert, P., Braun, W., Billeter, M. & Wüthrich, K. (1989). "Automated Stereospecific ^{1}H NMR Assignments and Their Impact on the Precision of Protein Structure Determinations in Solution." *J. Am. Chem. Soc.*, **111**, 3997-4004.

Güntert, P., Braun, W. & Wüthrich, K. (1991a). "Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting CALIBA, HABAS, and GLOMSA." *J. Mol. Biol.*, **217**, 517-530.

Güntert, P., *et al.* (1991b). "Structure determination of the *Antp(C39→S)* homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS, and GLOMSA." *J. Mol. Biol.*, **217**, 531-540.

Habazettl, J., Cieslar, C., Oschkinat, H. & Holak, T. A. (1990). "^{1}H NMR assignments of sidechain conformations in proteins using a high-dimensional potential in the simulated annealing calculations." *FEBS Lett.*, **268**, 141-145.

Härd, T., *et al.* (1990). "^{1}H NMR studies of the glucocorticoid receptor DNA-binding domain: sequential assignments and identification of secondary structure elements." *Biochemistry*, **29**, 9015-9023.

Hatfull, G. F., *et al.* (1989). "Preparation of heavy-atom derivatives using site-directed mutagenesis. Introduction of cysteine residues into gamma delta resolvase." *J. Mol. Biol.*, **208**, 661-667.

Havel, T. F. & Wüthrich, K. (1985). "An Evaluation of the Combined Use of Nuclear Magnetic Resonance and Distance Geometry for the Determination of Protein Conformations in Solution." *J. Mol. Biol.*, **182**, 281-294.

Hibler, D. W., *et al.* (1989). "Isotopic labeling with hydrogen-2 and carbon-13 to compare conformations of proteins and mutants generated by site-directed mutagenesis, I." *Methods Enzymol.*, **177**, 74-86.

Hoch, J. C. (1989). "Modern spectrum analysis in nuclear magnetic resonance: alternatives to the Fourier transform." *Methods Enzymol.*, **176**, 216-241.

Hoffman, R. E., *et al.* (1989). "Application of the maximum likelihood method to a large 2D NMR spectrum using a parallel computer." *J. Magn. Reson.*, **83**, 586-594.

Hoffman, R. E. & Levy, G. C. (1989). "Spectral deconvolution by simulated annealing." *J. Magn. Reson.*, **83**, 411-417.

Holak, T. A., Gondol, D., Otlewski, J. & Wilusz, T. (1989a). "Determination of the complete three-dimensional structure of the trypsin inhibitor from squash seeds in aqueous solution by nuclear magnetic resonance and a combination of distance geometry and dynamical simulated annealing." *J. Mol. Biol.*, **210**, 635-648.

Holak, T. A., Nilges, M. & Oschkinat, H. (1989b). "Improved strategies for the determination of protein structures from NMR data: the solution structure of acyl carrier protein." *FEBS Lett.*, **242**, 218-224.

Hore, P. J. (1983). *J. Magn. Reson.*, **55**, 283.

Hore, P. J. (1989). "Nuclear magnetic resonance. Solvent suppression." *Methods Enzymol.*, **176**, 64-77.

Hultman, T., *et al.* (1990). "Solid phase in vitro mutagenesis using plasmid DNA template." *Nucleic Acids Res.*, **18**, 5107-5112.

Hurle, M. R., *et al.* (1991). "Comparison of solution structures of mutant bovine pancreatic trypsin inhibitor proteins using two-dimensional nuclear magnetic resonance." *Protein Sci.*, **submitted**,

Hyberts, S. G., Märki, W. & Wagner, G. (1987). "Stereospecific assignments of side-chain protons and characterization of torsion angles in Eglin c." *Eur. J. Biochem.*, **164**, 625-635.

Ikura, M., Krinks, M., Torchia, D. A. & Bax, A. (1990). "An efficient NMR approach for obtaining sequence-specific resonance assignments of larger proteins based on multiple isotopic labeling." *FEBS Lett.*, **266**, 155-158.

James, T. L. (1975). ***Nuclear Magnetic Resonance in Biochemistry***. Academic, New York.

James, T. L., Matson, G. B. & Kuntz, I. D. (1978). "Protein rotational correlation times determined in aqueous solution by carbon-13 rotating frame spin-lattice relaxation in the presence of an off-resonance radiofrequency field." *J. Am. Chem. Soc.*, **100**, 3590-3594.

Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). "Investigation of exchange processed by two-dimensional NMR spectroscopy." *J. Chem. Phys.*, **71**, 4546-4553.

Kalbitzer, H. R., Leberman, R. & Wittinghofer, A. (1985). "^{1}H-NMR spectroscopy on elongation factor Tu from *Escherichia coli*: Resolution enhancement by perdeuteration." *FEBS Lett.*, **180**, 40-42.

Kalk, A. & Berendsen, H. J. C. (1975). "Proton magnetic relaxation and spin diffusion in proteins." *J. Magn. Reson.*, **24**, 343-366.

Kaptein, R., *et al.* (1985). "A protein structure from nuclear magnetic resonance data: lac repressor headpiece." *J. Mol. Biol.*, **182**, 179-182.

Karplus, M. (1959). *J. Phys. Chem.*, **30**, 11-15.

Karplus, M. (1963). "Vicinal proton coupling in nuclear magnetic resonance." *J. Am. Chem. Soc.*, **85**, 2870.

Kassell, B. (1970). "Bovine trypsin-kallikrein inhibitor (Kunitz inhibitor, basic pancreatic trypsin inhibitor, polyvalent inhibitor from bovine organs)." *Methods Enzymol.*, **19**, 844-852.

Kay, L. E. & Bax, A. (1990). "New methods for the measurement of NH-C-alpha-H coupling constants in N-15-labeled proteins." *J. Magn. Reson.*, **86**, 110-126.

Kay, L. E., *et al.* (1989). "Measurement of NH-C_αH coupling constants in staphylococcal nuclease by two-dimensional NMR and comparison with x-ray crystallographic results." *J. Am. Chem. Soc.*, **111**, 5488-5490.

Kay, L. E., Clore, G. M., Bax, A. & Gronenborn, A. M. (1990a). "Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution ." *Science*, **249**, 411-414.

Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990b). "Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins." *J. Magn. Reson.*, **89**, 496-514.

Kay, L. E., Marion, D. & Bax, A. (1989). "Practical aspects of 3D heteronuclear NMR of proteins." *J. Magn. Reson.*, **84**, 72-84.

Kenyon, G. L. & Bruice, T. W. (1977). "Novel Sulfhydryl Reagents." *Methods Enzymol.*, **47**, 407-430.

Kim, Y. & Prestegard, J. H. (1989). "A dynamic model for the structure of acyl carrier protein in solution." *Biochemistry*, **28**, 8792-8797.

Kosen, P. A. (1989). "Spin labeling of proteins." *Methods Enzymol.*, **177**, 86-121.

Kosen, P. A., *et al.* (1986). "Two-dimensional ^{1}H NMR of three spin-labeled derivatives of bovine pancreatic trypsin inhibitor." *Biochemistry*, **25**, 2356-2364.

Kowalksy, A. (1965). *Biochemistry*, **4**, 2382.

Kowalsky, A. (1962). *J. Biol. Chem.*, **237**, 1807.

Kraulis, J., *et al.* (1989). "Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing." *Biochemistry*, **28**, 7241-7257.

Kumar, A., Ernst, R. R. & Wüthrich, K. (1980a). "A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules." *Biochem. Biophys. Res. Commun.*, **95**, 1-6.

Kumar, A., Wagner, G., Ernst, R. R. & Wüthrich, K. (1980b). "Studies of J-connectivities and selective ^{1}H-^{1}H Overhauser effects in H_2O solutions of biological macromolecules by two-dimensional NMR experiments." *Biochem. Biophys. Res. Commun.*, **96**, 1156-1163.

Kumar, A., Wagner, G., Ernst, R. R. & Wüthrich, K. (1981). "Buildup rates of the nuclear Overhauser effect measure by two-dimensional proton magnetic resonance spectroscopy: implications for studies of protein conformation." *J. Am. Chem. Soc.*, **103**, 3654-3658.

Kuntz, E. D., Crippen, G. M. & Kollman, P. A. (1979). "Application of distance geometry to protein tertiary structure calculations." *Biopolymers*, **18**, 939.

Kuntz, I. D., Thomason, J. F. & Oshiro, C. M. (1989). "Distance geometry." *Methods Enzymol.*, **177**, 159-204.

Lauritzen, C., Skovgaard, O., Tüchsen, E. & Hansen, P. E. (1991). "N-extension of BPTI. A model for folding and long-range interactions." *Abstracts of papers*, Frontiers of NMR in Molecular Biology-II, Keystone, Colorado. Wiley-Liss. p. 65.

LeMaster, D. M. (1987). "Chiral beta and random fractional deuteration for the determination of protein sidechain conformation by NMR." *FEBS Lett.*, **223**, 191-196.

LeMaster, D. M. (1988). "Protein NMR resonance assignment by isotropic mixing experiments on random fractionally deuterated samples." *FEBS Lett.*, **233**, 326-330.

LeMaster, D. M. (1989). "Deuteration in protein proton magnetic resonance." *Methods Enzymol.*, **177**, 23-43.

LeMaster, D. M. (1990a). "Deuterium labelling in NMR structural analysis of larger proteins." *Q. Rev. Biophys.*, **23**, 133-174.

LeMaster, D. M. (1990b). "Uniform and selective deuteration in two-dimensional NMR of proteins." *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 243-266.

LeMaster, D. M. & Cronan, J. E. (1982). "Biosynthetic production of carbon-13-labeled amino acids with site-specific enrichment." *J. Biol. Chem.*, **257**, 1224-1230.

LeMaster, D. M. & Richards, F. M. (1985). "^{1}H—^{15}N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type." *Biochemistry*, **24**, 7263-7268.

LeMaster, D. M. & Richards, F. M. (1988). "NMR Sequential Assignment of *Escherichia coli* thioredoxin utilizing random fractional deuteration." *Biochemistry*, **27**, 142-150.

Liu, W.-K. & Meienhofer, J. (1968). "Preparation of 14,38-Bis-[S-Carbamidomethyl]-(Basic Trypsin Inhibitor) Possessing Full Biological Activity." *Biochem. Biophys. Res. Commun.*, **31**, 467-473.

Macura, S. & Ernst, R. R. (1980). "Elucidation of cross-relaxation in liquids by two-dimensional NMR spectroscopy." *Mol. Phys.*, **41**, 95.

Marion, D., *et al.* (1989a). "Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H-^{15}N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1β." *Biochemistry*, **28**, 6150-6156.

Marion, D., *et al.* (1989b). "Three-Dimensional Heteronuclear NMR of ^{15}N-Labeled Proteins." *J. Am. Chem. Soc.*, **111**, 1515-1517.

Marion, D. & Wüthrich, K. (1983). "Application of Phase Sensitive Two-Dimensional Correlated Spectroscopy (COSY) for Measurements of ^{1}H-^{1}H Spin-Spin Coupling Constants in Proteins." *Biochem. Biophys. Res. Commun.*, **113**, 967-974.

Markley, J. L. (1989). "Two-dimensional nuclear magnetic resonance spectroscopy of proteins: an overview." *Methods Enzymol.*, **176**, 12-64.

Markley, J. L. & Jardetzky, O. (1970). "Nuclear magnetic resonance studies of the structure and binding sites of enzymes: XIV. Inhibitor binding to staphylococcal nuclease." *J. Mol. Biol.*, **50**, 223-233.

Markley, J. L., Putter, I. & Jardetzky, O. (1968). "High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease." *Science*, **161**, 1249-1251.

Marks, C. B., *et al.* (1987a). in ***Protein Structure, Folding and Design***. (Oxender, D. L., ed.) Alan R. Liss, New York, pp. 335-340.

Marks, C. B., *et al.* (1987b). "Mutants of Bovine Pancreatic Trypsin Inhibitor Lacking Cysteines 14 and 38 Can Fold Properly." *Science*, **235**, 1370-1373.

Marks, C. B., *et al.* (1986). "Production of Native, Correctly Folded Bovine Pancreatic Trypsin Inhibitor by *Escherichia coli*." *J. Biol. Chem.*, **261**, 7115-7118.

McDonald, C. C. & Philips, W. D. (1967). *J. Am. Chem. Soc.*, **89**, 6332-6341.

McIntosh, L. P. & Dahlquist, F. W. (1990). "Biosynthetic Incorporation of ^{15}N and ^{13}C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins." *Q. Rev. Biophys.*, **23**, 1-38.

McIntosh, L. P., *et al.* (1990). "Assignment of the backbone ^{1}H and ^{15}N NMR resonances of bacteriophage T4 lysozyme." *Biochemistry*, **29**, 6341-6362.

Meiboom, S. & Gill, D. (1958). *Rev. Sci. Instrum.*, **29**, 688.

Miller, J. H. (1972). ***Experiments in Molecular Genetics***. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Molday, R. S., Englander, S. W. & Kallen, R. G. (1972). "Primary structure effects on peptide group hydrogen exchange." *Biochemistry*, **11**, 150-159.

Montelione, G. T. & Wagner, G. (1989). "Accurate measurements of homonuclear H-N-H-alpha coupling constants in polypeptides using heteronuclear 2D NMR experiments." *J. Am. Chem. Soc.*, **111**, 5474-5475.

Montelione, G. T., Winkler, M. E., Rauenbuehler, P. & Wagner, G. (1989). "Accurate Measurements of long-range heteronuclear coupling constants." *J. Magn. Reson.*, **82**, 198-204.

Moonen, C. T. W., Scheek, R. M., Boelens, R. & Müller, F. (1984). "The use of two-dimensional nuclear-magnetic-resonance spectroscopy and two-dimensional difference spectra in the elucidation of the active center of *Megasphaera elsdenii* flavodoxin." *Eur. J. Biochem.*, **141**, 323-330.

Morris, G. A. & Freeman, R. (1978). "Selective excitation in Fourier transform nuclear magnetic resonance." *J. Magn. Reson.*, **29**, 433-462.

Muchmore, D. C., *et al.* (1989). "Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance." *Methods Enzymol.*, **177**, 44-73.

Mueller, L. (1987). "P.E.COSY, a simple alternative to E.COSY." *J. Magn. Reson.*, **72**, 191-196.

Naderi, H. M., *et al.* (1991). "^{1}H NMR assignments of Ala14/Ala38 bovine pancreatic trypsin inhibitor and three-dimensional structure based on two-dimensional NMR and distance geometry." in press.

Nagai, K. & Thøgersen, H. C. (1984). "Generation of β-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*." *Nature*, **309**, 810-812.

Neidhardt, F. C., Bloch, P. L. & Smith, D. F. (1974). "Culture Medium for Enterobacteria." *J. Bacteriol.*, **119**, 736-747.

Neri, D., Otting, G. & Wüthrich, K. (1990). "New Nuclear Magnetic Resonance Experiments for Measurements of the vicinal coupling constants ^{3}JHN-alpha in proteins." *J. Am. Chem. Soc.*, **112**, 3663-3665.

Neri, D., *et al.* (1989). "Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling." *Biochemistry*, **28**, 7510-7516.

Neuhaus, D., *et al.* (1985). "Systematic application of high-resolution, phase sensitive two-dimensional ^{1}H-NMR techniques for the identificatin of the amino-acid-proton spin systems in proteins: Rabbit metallothionein-2." *Eur. J. Biochem.*, **151**, 257-273.

Neuhaus, D. & Williamson, M. P. (1989). ***The Nuclear Overhauser Effect in Structural and Conformational Analysis***. VCH, New York.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988a). "Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. Circumventing problems associated with folding." *FEBS Lett.*, **239**, 129-136.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988b). "Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations." *FEBS Lett.*, **229**, 317-324.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1990). "^{1}H-NMR stereospecific assignments by conformational data-base searches." *Biopolymers*, **29**, 813-822.

Nilges, M., Gronenborn, A. M., Brunger, A. T. & Clore, G. M. (1988). "Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2." *Protein Eng.*, **2**, 27-38.

Nilsson, B. & Abrahmsen, L. (1990). "Fusions to staphylococcal protein A." *Methods Enzymol.*, **185**, 144-161.

Nilsson, B., Berman-Marks, C., Kuntz, I. D. & Anderson, S. (1991). "Secretion Incompetence of Bovine Pancreatic Trypsin Inhibitor Expressed in *Escherichia coli*." *J. Biol. Chem.*, **266**, 2970-2977.

Nilsson, B., *et al.* (1987). "A synthetic IgG-binding domain based on staphylococcal protein A." *Protein Eng.*, **1**, 107-113.

Nirmala, N. R. & Wagner, G. (1988). "Measurement of ^{13}C Relaxation Times in Proteins by Two-Dimensional Heteronuclear ^{1}H-^{13}C Correlation Spectroscopy." *J. Am. Chem. Soc.*, **110**, 7557-7558.

Noggle, J. H. & Schirmer, R. E. (1971). ***The Nuclear Overhauser Effect***. Academic, New York.

Oka, A., Sugisaki, H. & Takanami, M. (1981). "Nucleotide Sequence of the Kanamycin Resistance Transposon Tn*903*." *J. Mol. Biol.*, **147**, 217-226.

Olejniczak, E. T., Gampe, R. T. J. & Fesik, S. W. (1986). "Accounting for spin diffusion in the analysis of 2D NOE data." *J. Magn. Reson.*, **67**, 28-41.

Omichinski, J. G., *et al.* (1990). "High-resolution 3-dimensional structure of a single zinc finger from a human enhancer binding protein in solution." *Biochemistry*, **29**, 9324-9334.

Otting, G., Orbons, L. P. M. & Wüthrich, K. (1990). "Suppression of Zero-Quantum Coherence in NOESY and Soft NOESY." *J. Magn. Reson.*, **89**, 423-430.

Otting, G., Widmer, H., Wagner, G. & Wüthrich, K. (1986). "Origin of t_1 and t_2 ridges in 2D NMR spectra and procedures for suppression." *J. Magn. Reson.*, **66**, 187-193.

Pardi, A., Billeter, M. & Wüthrich, K. (1984). "Calibration of the Angular Dependence of the Amide Proton-C^{α} Proton Coupling Constants, $^{3}J_{HN\alpha}$, in a Globular Protein: Use of $^{3}J_{HN\alpha}$ for identification of Helical Secondary Structure." *J. Mol. Biol.*, **180**, 741-751.

Pauling, L. & Corey, R. B. (1951). "Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets." *Proc. Natl. Acad. Sci. U.S.A.*, **37**, 729-740.

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

Pearlman, D. A. & Kollman, P. A. (1991). "Are time-averaged restraints necessary for NMR refinement: a model study for DNA." *J. Mol. Biol.*, **in press**.

Pohl, T. (1990). "Concentration of proteins and removal of solutes." *Methods Enzymol.*, **182**, 68-83.

Poole, C. P., Jr. & Farach, H. A. (1987). ***Theory of Magnetic Resonance***. Wiley, New York.

Rance, M., *et al.* (1983). "Improved spectral resolution in COSY ^{1}H-NMR spectra of proteins via double quantum filtering." *Biochem. Biophys. Res. Comm.*, **117**, 479-485.

Rentier-Delrue, F., *et al.* (1989). "Tilapia prolactin: molecular cloning of two cDNAs and expression in *Escherichia coli*." *DNA*, **8**, 261-270.

Richardson, J. (1981). "The Anatomy and Taxonomy of Protein Structure." *Adv. Protein Chem.*, **34**, 167-339.

Rosenberg, A. H., *et al.* (1987). "Vectors for selective expression of cloned DNAs by T7 RNA polymerase." *Gene*, **56**, 125-135.

Saunders, M., Wishnia, A. & Kirkwood, J. G. (1957). *J. Am. Chem. Soc.*, **79**, 3289.

Scheek, R. M., van Gunsteren, W. F. & Kaptein, R. (1989). "Molecular dynamics simulation techniques for determination of molecular structures from nuclear magnetic resonance data." *Methods Enzymol.*, **177**, 204-218.

Schmidt, P. G. & Kuntz, I. D. (1984). "Distance Measurements in Spin-Labeled Lysozyme." *Biochemistry*, **23**, 4261-4266.

Schulz, G. E. & Schirmer, R. H. (1979). ***Principles of Protein Structure***. Springer-Verlag, New York.

Schwarz, H., *et al.* (1987). "Stability studies on derivatives of the bovine pancreatic trypsin inhibitor." *Biochemistry*, **26**, 3544-3551.

Seeholzer, S. H., *et al.* (1986). "NMR studies of a complex of deuterated calmodulin with melittin." *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 3634-3638.

Shine, J. & Dalgarno, L. (1975). "Determinant of cistron specificity in bacterial ribosomes." *Nature*, **254**, 34-38.

Smith, D. J. & Kenyon, G. L. (1974). *J. Biol. Chem.*, **249**, 3317.

Smith, D. J., Maggio, E. T. & Kenyon, G. L. (1975). "Simple Alkanethiol Groups for Temporary Blocking of Sulfhydryl Groups of Enzymes." *Biochemistry*, **14**, 766-771.

Smith, L. J., Sutcliffe, M. J., Redfield, C. & Dobson, C. M. (1991). "Analysis of ϕ and χ_1 torsion angles for hen lysozyme in solution from ^{1}H NMR spin-spin coupling constants." *Biochemistry*, **30**, 986-996.

Solomon, I. & Bloembergen, N. (1956). "Nuclear magnetic interactions in the HF molecule." *J. Chem. Phys.*, **25**, 261.

Sørensen, O. W. (1990). "Three-dimensional and four-dimensional NMR experiments for measurement of spin-spin coupling constants." *J. Magn. Reson.*, **90**, 433-438.

Stader, J. A. & Silhavy, T. J. (1990). "Engineering *Escherichia coli* to secrete heterologous gene products." *Methods Enzymol.*, **185**, 166-187.

Stassinopoulou, C. I., Wagner, G. & Wüthrich, K. (1984). "Two-dimensional ^{1}H NMR of two chemically modified analogs of the basic pancreatic trypsin inhibitor: sequence-specific resonance assignments and sequence location of conformation changes relative to the native protein." *Eur. J. Biochem.*, **145**, 423-430.

States, D. J., Creighton, T. E., Dobson, C. M. & Karplus, M. (1987). "Conformations of Intermediates in the Folding of the Pancreatic Trypsin Inhibitor." *J. Mol. Biol.*, **195**, 731-739.

States, D. J., Dobson, C. M., Karplus, M. & Creighton, T. E. (1984). "A New Two-disulphide Intermediate in the Refolding of Reduced Bovine Pancreatic Trypsin Inhibitor." *J. Mol. Biol.*, **174**, 411-418.

States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982). "A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants." *J. Magn. Reson.*, **48**, 286.

Studier, F. W. & Moffatt, B. A. (1986). "Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes." *J. Mol. Biol.*, **189**, 113-130.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." *Methods Enzymol.*, **185**, 60-89.

Sun, D. P., *et al.* (1987). "Use of site-directed mutagenesis to obtain isomorphous heavy-atom derivatives for protein crystallography: cysteine-containing mutants of phage T4 lysozyme." *Protein Eng.*, **1**, 115-123.

Thomas, P. D., Basus, V. J. & James, T. L. (1991). "Protein solution structure determination using distances from two-dimensional nuclear overhauser effect experiments - effect of approximations on the accuracy of derived structures." *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 1237-1241.

Timasheff, S. N. & Arakawa, T. (1989). "Stabilization of protein structure by solvents" in ***Protein Structure: A Practical Approach***. (Creighton, T. E., ed.) IRL Press, Oxford, pp.

Todd, A. P. & Millhauser, G. L. (1991). "ESR spectra reflect local and global mobility in a short spin-labeled peptide throughout the α-helix ⇒ coil transition." *Biochemistry*, **30**, 5515-5523.

Torchia, D. A., Sparks, S. W. & Bax, A. (1988). "Delineation of α–Helical Domains in Deuteriated Staphylococcal Nuclease by 2D NOE NMR Spectroscopy." *J. Am. Chem. Soc.*, **110**, 2320-2321.

Torchia, D. A., Sparks, S. W. & Bax, A. (1989). "Staphylococcal nuclease: sequential assignments and solution structure." *Biochemistry*, **28**, 5509-5524.

Torda, A. E., Scheek, R. M. & van Gunsteren, W. F. (1990). "Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat." *J. Mol. Biol.*, **214**, 223-235.

Tüchsen, E. & Woodward, C. (1987). "Assignment of asparagine-44 side-chain primary amide ^{1}H NMR resonances and the peptide amide N^{1}H resonance of glycine-37 in basic pancreatic trypsin inhibitor." *Biochemistry*, **26**, 1918-1925.

Tucker, A. D., *et al.* (1989). "Crystallographic phases through genetic engineering: experiences with colicin A." *Protein Eng.*, **2**, 399-405.

Venters, R. A., Calderone, T. L., Spicer, L. D. & Fierke, C. A. (1991). "Uniform ^{13}C isotope labeling of proteins with sodium acetate for NMR studies: application to human carbonic anhydrase II." *Biochemistry*, **30**, 4491-4494.

Vieira, J. & Messing, J. (1987). "Production of single-stranded plasmid DNA." *Methods Enzymol.*, **153**, 3-11.

Vincent, J.-P., Chicheportiche, R. & Lazdunski, M. (1971). "The Conformational Properties of the Basic Pancreatic Trypsin-Inhibitor." *Eur. J. Biochem.*, **23**, 401-411.

Vincent, J.-P. & Lazdunski, M. (1972). "Trypsin-Pancreatic Trypsin Inhibitor Association. Dynamics of the Interaction and Rold of Disulfide Bridges." *Biochemistry*, **11**, 2967-2977.

Vincent, J.-P. & Lazdunski, M. (1973). "The Interaction between α-Chymotrypsin and Pancreatic Trypsin Inhibitor (Kunitz Inhibitor): Kinetic and Thermodynamic Properties." *Eur. J. Biochem.*, **38**, 365-372.

von Wilcken-Bergmann, B., *et al.* (1986). "A synthetic operon containing 14 bovine pancreatic trypsin inhibitor genes is expressed in *E. coli*." *EMBO J.*, **5**, 3219-3225.

Wagner, G. (1983). "Characterization of the distribution of internal motions in the basic pancreatic trypsin inhibitor using a large number of internal NMR probes." *Q. Rev. Biophys.*, **16**, 1-57.

Wagner, G. (1990). "NMR Investigations of Protein Structure." *Prog. NMR Spectrosc.*, **22**, 101-139.

Wagner, G., *et al.* (1987). "Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN." *J. Mol. Biol.*, **196**, 611-639.

Wagner, G. & Bruhwiler, D. (1986). "Toward the complete assignment of the carbon nuclear magnetic resonance spectrum of the basic pancreatic trypsin inhibitor." *Biochemistry*, **25**, 5839-5843.

Wagner, G., Bruhwiler, D. & Wüthrich, K. (1987). "Reinvestigation of the aromatic side-chains in the basic pancreatic trypsin inhibitor by heteronuclear two-dimensional nuclear magnetic resonance." *J. Mol. Biol.*, **196**, 227-231.

Wagner, G. & Nirmala, N. R. (1989). "Studies Of Protein Dynamics By Heteronuclear NMR - Individual C-13 Relaxation Times And Evidence For Multiple Conformations In The Reactive Site Of BPTI." *Chemica Scripta*, **29A**, 27-30.

Wagner, G. & Wüthrich, K. (1982a). "Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance." *J. Mol. Biol.*, **160**, 343-361.

Wagner, G. & Wüthrich, K. (1982b). "Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra: basic pancreatic trypsin inhibitor." *J. Mol. Biol.*, **155**, 347-366.

Wang, J. F., LeMaster, D. M. & Markley, J. L. (1990). "Two-dimensional NMR studies of staphylococcal nuclease. 1. Sequence-specific assignments of hydrogen-1 signals and solution structure of the nuclease H124L-thymidine 3',5'-bisphosphate-Ca2+ ternary complex." *Biochemistry*, **29**, 88-101.

Warne, S. R., Thomas, C. M., Nugent, M. E. & Tacon, W. C. A. (1986). "Use of a modified *Escherichia coli trpR* gene to obtain tight regulation of high-copy-number expression vectors." *Gene*, **46**, 103-112.

Weber, P. L., Morrison, R. & Hare, D. (1988). "Determining stereo-specific ^{1}H nuclear magnetic resonance assignments from distance geometry calculations." *J. Mol. Biol.*, **204**, 483-487.

Weiss, M. A., Eliason, J. L. & States, D. J. (1984). "Dynamic filtering by two-dimensional ^{1}H NMR with application to phage λ repressor." *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 6019-6023.

West, D. K., Changchien, L. M., Maley, G. F. & Maley, F. (1989). "Evidence that the intron open reading frame of the phage T4 td gene encodes a specific endonuclease." *J. Biol. Chem.*, **264**, 10343-10346.

Wider, G. (1990). "Elimination of baseline artifacts in NMR spectra by oversampling." *J. Magn. Reson.*, **89**, 406-409.

Wider, G., Neri, D., Otting, G. & Wüthrich, K. (1989). "A heteronuclear three-dimensional NMR experiment for measurements of small heteronuclear coupling constants in biological macromolecules." *J. Magn. Reson.*, **85**, 426-431.

Widmer, H., Billeter, M. & Wüthrich, K. (1989). "Three-dimensional structure of the neurotoxin ATX Ia from Anemonia sulcata in aqueous solution determined by nuclear magnetic resonance spectroscopy." *Proteins*, **6**, 357-371.

Widmer, H. & Wüthrich, K. (1986). "Simulation of Two-Dimensional NMR Experiments Using Numerical Density Matrix Calculations." *J. Magn. Reson.*, **70**, 270-279.

Widmer, H. & Wüthrich, K. (1987). "Simulated Two-Dimensional NMR Cross-Peak Fine Structures for ^{1}H Spin Systems in Polypeptides and Polydeoxynucleotides." *J. Magn. Reson.*, **74**, 316.

Wien, R. W., Morrisett, J. D. & McConnell, H. M. (1972). "Spin-label-induced nuclear relaxation. Distance between bound saccharides, histidine-15, and tryptophan-123 on lysozyme in solution." *Biochemistry*, **11**, 3707-3716.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). "Solution conformation of proteinsase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry." *J. Mol. Biol.*, **182**, 295-315.

Wlodawer, A., Walter, J., Huber, R. & Sjolin, L. (1984). "Structure Of Bovine Pancreatic Trypsin Inhibitor. Results of Joint Neutron and X-ray Refinement of Crystal Form II." *J. Mol. Biol.*, **180**, 301.

Wüthrich, K. (1986). ***NMR of Proteins and Nucleic Acids***. Wiley, New York.

Wüthrich, K. (1989). "Protein structure determination in solution by nuclear magnetic resonance spectroscopy." *Science*, **243**, 45-50.

Wüthrich, K., Billeter, M. & Braun, W. (1983). "Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance." *J. Mol. Biol.*, **169**, 949-961.

Wüthrich, K., Billeter, M. & Braun, W. (1984). "Polypeptide secondary structure determination by nuclear magnetic resonance observation of short proton-proton distances." *J. Mol. Biol.*, **180**, 715-740.

Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982). "Sequential Resonance Assignments as a Basis for Determination of Spatial Protein Structures by High Resolution Proton Nuclear Magnetic Resonance." *J. Mol. Biol.*, **155**, 311-319.

Yip, P. & Case, D. A. (1989). "A New Method for Refinement of Macromolecular Structures Based on Nuclear Overhauser Effect Spectra." *J. Magn. Reson.*, **83**, 643-648.

Zilber, B., Scherf, T., Levitt, M. & Anglister, J. (1990). "NMR-derived model for a peptide-antibody complex." *Biochemistry*, **29**, 10032-10041.

Zuiderweg, E. R. & Fesik, S. W. (1989). "Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a." *Biochemistry*, **28**, 2387-2391.

Zuiderweg, E. R. P., Boelens, R. & Kaptein, R. (1985). "Stereospecific Assignments of ^{1}H-NMR Methyl Lines and Conforamtion of Valyl Residues in the *lac* Repressor Headpiece." *Biopolymers*, **24**, 601-611.

Zuiderweg, E. R. P., Hallenga, K. & Olejniczak, E. T. (1986). "Improvement of 2D NOE spectra of biomacromolecules in H_2O solution by coherent suppression of the solvent resonance." *J. Magn. Reson.*, **70**, 336-343.