

MSU
1
2001

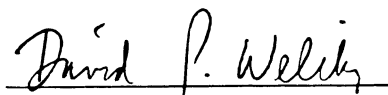
This is to certify that the
thesis entitled
SYNTHETIC AND STRUCTURAL SOLID STATE
NMR STUDIES OF THE STREPTOCOCCAL
PROTEIN G B1 DOMAIN

presented by

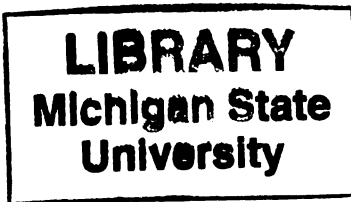
Bhagyashree A. Khunte

has been accepted towards fulfillment
of the requirements for

M.S. degree in Chemistry


Major professor

Date April 19, 2001



PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

SYNTHETIC AND STRUCTURAL SOLID STATE NMR STUDIES OF THE
STREPTOCOCCAL PROTEIN G B1 DOMAIN

By

Bhagyashree A. Khunte

A THESIS

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

MASTER OF SCIENCE

Department Of Chemistry

2001

ABSTRACT

SYNTHETIC AND STRUCTURAL SOLID STATE NMR STUDIES OF THE STREPTOCOCCAL PROTEIN G B1 DOMAIN

By

Bhagyashree A. Khunte

Solid state nuclear magnetic resonance (NMR) spectroscopy is a novel approach for determination of atomic level structure and dynamics in biological systems. We pursue ways of using solid-state magic angle spinning nuclear magnetic resonance (MAS-NMR) spectroscopy to help to determine the structure of molecules that cannot otherwise be studied by X-ray crystallography or solution NMR spectroscopy because they cannot be crystallized or because they are prohibitively large. We are investigating the dependence of measured chemical shift anisotropy (CSA) principal values on local secondary structure and hydrogen bonding environments in proteins. For these studies, we are using the fifty-five-residue B1 domain of Streptococcal Protein G as a model because it has a known 1.1 Å crystal structure and because it can be readily synthesized with specific labels on an automated peptide synthesizer.

We have developed a straightforward chemical synthesis of Protein G B1 Domain using solid phase methods with 9-fluorenylmethoxycarbonyl (Fmoc) strategy. Using ^{13}C labels at specific lysine carbonyl positions, we have observed some correlation between the CSA principal values and the local secondary structure of the Protein G B1 domain.

Copyright by
BHAGYASHREE A. KHUNTE
2001

To Aai

ACKNOWLEDGMENTS

I want to thank all the members of the Weliky group for their friendship, support and advice during the course of this work.

I would like to thank Dr. David Weliky for his constant guidance and support during this work. I have enjoyed the last three years and appreciate all of the responsibility, freedom, and confidence he gave me.

I am especially thankful for my parents and sister for the encouragement and my mother and father in law for their invaluable help and support.

Finally, I wish to thank my husband Ajay whose love and encouragement were vital to the completion of this degree.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
Chapters	
1. Introduction	1
2. Synthesis of Protein G B1 Domain by Linear Solid Phase Peptide Synthesis	
2.1 Introduction	10
2.2 Materials and Methods	17
2.3 Results and Discussion	26
2.4 References	27
3. Synthesis of Protein G B1 Domain by Solid-Phase Fragment Condensation	
3.1 Introduction	28
3.2 Materials and Methods	36
3.3 Results and Discussion	46
3.4 References	49
4. Chemical Shift Anisotropy Principal Values in Protein Secondary Structure Determination	
4.1 Introduction	50
4.2 Materials and Methods	64
4.3 Results and Discussion	66
4.4 References	70

LIST OF TABLES

Tables	Pages
3.1 Peptides and Proteins Synthesized on 2-Chlorotrityl Chloride Resin by the SPFC Strategy	35
4.1 Concentration and Line widths of the Unlabeled and Labeled Protein G samples	66
4.2 Measured Chemical Shift Anisotropy Principal Values for the ^{13}C Labeled Protein G Samples.....	66

LIST OF FIGURES

Figure	Pages
1.1 Protein G B1 Domain (A) Structure and (B) hydrogen bonding pattern	4
1.2 Ribbon drawing of the three-dimensional structure of the B1-domain	5
2.1 (a) Structure of the HMP Resin (b) Structure of Preloaded HMP Resin	12
2.2 9-fluorenylmethoxycarbonyl (Fmoc) Protecting Group	13
2.3 Linear Solid-Phase Peptide Synthesis	15
2.4 Reversed-Phase HPLC Spectra of Protein G (PG) fragments	20
2.5 Mass Spectra of Protein G fragments	22
2.6 (a) RP-HPLC and (b) Mass spectrum of Protein G B1 Domain (55 amino acids)...	25
3.1 Convergent peptide synthesis with C- to N-terminal chain extension	30
3.2 Structure of 2-Chlorotrityl Chloride Resin	32
3.3 Synthesis of 5+5-mer peptide	38
3.4 Synthesis of 11+11-mer peptide	41
3.5 Mass Spectra a) 12-mer On High Substituted Resin with no capping b) 12-mer On Low Substituted Resin with capping	45
4.1 Electronic Shielding	54
4.2: Chemical Shift Anisotropy	56
4.3 Dipolar Coupling	58
4.4: Magic Angle Spinning	61
4.5 Protein G B1 Domain Hydrogen Bonding Pattern	63
4.6: NMR Spectra of the Five-Labeled Protein G (Frozen Solution) Samples	67

LIST OF ABBREVIATIONS

Ala	Alanine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Glu	Glutamic acid
Gln	Glutamine
Gly	Glycine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine
AcOH	Acetic acid
DCC	Dicyclohexylcarbodiimide
DIC	Diisopropylcarbodiimide
DIEA	Diisopropylethylamine
DMAC	N, N-dimethylacetamide
DMF	N, N-dimethylformide
DMSO	Dimethylsulfoxide
FMOC	9-Fluorenylmethoxycarbonyl
HOBt	Hydroxybenzotriazole
HBTU	2-(1 H-benzotriazol-1-yl)-1, 1, 3, 3-tetramethyluroniumhexafluorophosphate
NMP	N-methyl pyrrolidone
t-Bu	Tertiary butyl
tBoc	Tertiary butoxycarbonyl
TBTU	O-benzotriazolyl-N, N, N', N', tetramethyluronium-tetrafluoroborate
CSA	Chemical Shift Anisotropy
MAS	Magic angle spinning
REDOR	Rotational-echo double resonance

Chapter 1

INTRODUCTION

Over the past twenty years, there has been an explosive increase in the number of reports of structure determination of biological macromolecules, and these structures have made important contributions to our understanding of biochemical processes and function. Our research focuses on the secondary structure determination of biological macromolecules by solid state nuclear magnetic resonance spectroscopy. The goal of my research was to develop methods to synthesize the Protein G B1 domain using 9-fluorenylmethoxycarbonyl (Fmoc) chemistry on an automated peptide synthesizer and to study the secondary structure by measuring chemical shift anisotropy (CSA) principal values for specific carbonyl ($^{13}\text{C}=\text{O}$) nuclei in the protein. This thesis represents the steps taken to achieve that goal.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for structure determination of biological macromolecules. Structure determination plays a key role in understanding the biochemical processes and function of the biomolecules. Solid state nuclear magnetic resonance spectroscopy is a novel approach for determination of atomic-level structure and dynamics in biological systems. The major advantages of solid state NMR over the more established techniques of X-ray crystallography and solution NMR spectroscopy are:

1. Crystals are not required.
2. Large (>30,000 molecular weight) systems can be studied.^{1,2}

3. Application to systems that are difficult to characterize by X-ray crystallography or solution NMR including, membrane, aggregated and partially ordered proteins.

In recent years, solid state NMR has been used to study systems such as membrane-bound channels gramicidin³ and colicin,⁴ β -amyloid fibrils,^{5,6} the E. coli serine receptor,⁷ the enzymes triosephosphate isomerase⁸ and EPSP synthase,^{9,10} and a peptide/neutralizing antibody complex.¹¹

In liquid state NMR, small, highly soluble proteins like Ubiquitin (76-residues) have served as model systems for methods development. There is also a need for such a model system in solid state NMR. A membrane protein might be a good choice except that there are fewer than twenty integral membrane proteins with known crystal structures. Bacteriophage coat protein (55-residue) has been used by Opella and coworkers.¹² However, the protein does not have a β -sheet secondary structure and also lacks an extensive tertiary structure. Another possible model system is the protein bacteriorhodopsin (248-residues), which has a high-resolution crystal structure. Griffin and coworkers¹ have extensively studied this protein with solid state NMR. Unfortunately, it is too large to synthesize chemically with isotopic labeling. Uniform labeling complicates the clear spectral interpretation needed in methods development. In addition, the uniformly labeled protein will have poor spectral resolution because of its large size.

We have selected the 55-residue streptococcal protein G B1 domain as the model system for solid state NMR studies. The reasons for selecting protein G are as follows:

1. It has a 1.1 Å resolution crystal structure.¹³

2. There are two other X-ray structures¹⁴ and a solution NMR structure¹⁵ with consistency between all structures.
3. The protein is extremely stable with a melting temperature of 87 °C.¹⁵
4. The protein has a novel structural motif of four beta strands with a helix laid across them.
5. The protein lacks cysteine and proline, so the folding occurs rapidly without formation of any disulfide bonds and without cis-trans proline isomerization.
6. It can be chemically synthesized in order to obtain specifically isotopically labeled samples.

Figure 1.1: Protein G B1 Domain (A) Structure and (B) hydrogen bonding pattern¹⁵

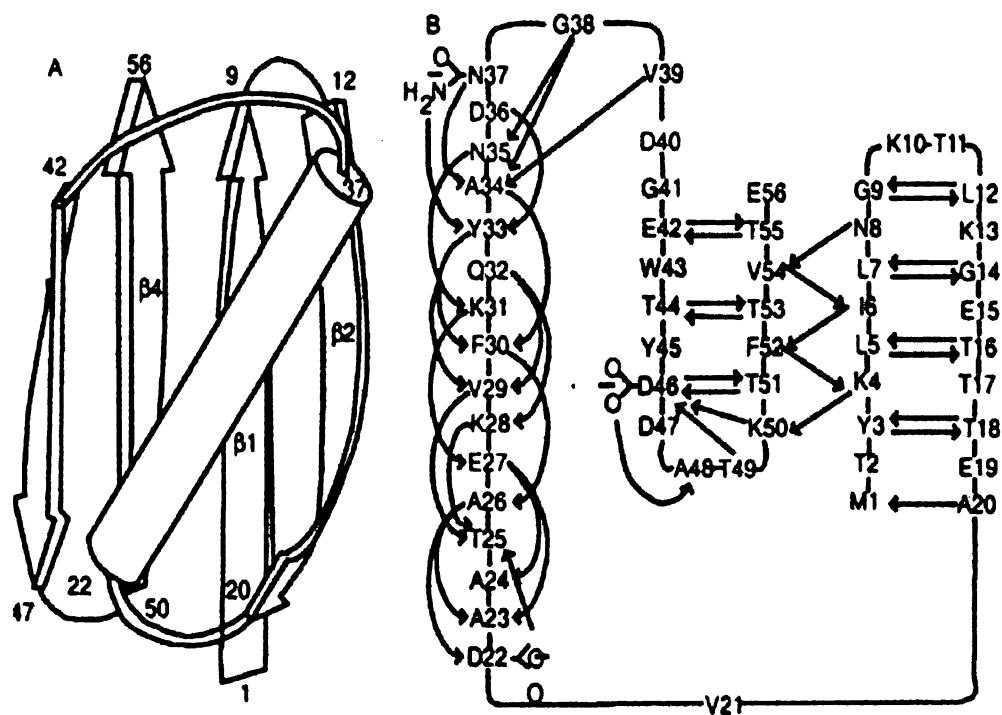


Figure 1.2: Ribbon drawing of the three-dimensional structure of the B1-domain¹⁶



The protein sequence is **TYKLI LNGKT LKGET TTEAV DAATA EKVFK QYAND NGVDG EWTYD DATKT FTVTE**. The N-terminal Met is not included as it is absent in 30% of the biologically expressed protein and the absence does not affect the protein structure.¹⁵

Three groups have reported the chemical syntheses of Protein G B1 Domain.^{16,17,18} Only one group has reported a purified yield of 2% using the tertiary butoxycarbonyl (tBoc) chemistry.¹⁶ We have developed a straightforward chemical synthesis of Protein G B1 Domain that routinely gives a 10% purified yield. The protein is synthesized by solid phase methods using the 9-fluorenylmethoxycarbonyl (Fmoc) methodology. The tBoc chemistry requires highly acidic conditions such as Hydrofluoric acid for cleavage. With Fmoc such highly acidic conditions are not required and this largely simplifies the experimental setup.

We are working on devising methods to synthesize Protein G B1 Domain using solid phase fragment condensation methods. Condensation of fully protected peptide fragments on a solid support is an efficient alternative to stepwise solid phase peptide synthesis (SPPS). The main advantage of this approach is that fully protected fragments are coupled instead of individual amino acids, thereby reducing the number of coupling steps. This reduction will potentially increase the yield and purity of the target peptide with respect to the stepwise approach.¹⁹

As discussed earlier, protein G B1 Domain is used as a model system for structural methods development by solid state NMR. The protein G model system is used to correlate the nuclear chemical shift anisotropy (CSA) principal values of specifically labeled ¹³C nuclei to secondary structure and hydrogen bonding

environments. These 1D CSA principal value measurements have great utility because they require an order of magnitude less material than 2D solid state NMR structural techniques and are generally applicable to any specifically labeled protein system. Protein G is well suited for this methods development because it contains eleven threonines and six lysines, and each residue type is found in a variety of secondary structure and hydrogen bonding environments.

The goals for the study of chemical shift anisotropy principal (CSA) values are:

1. To develop a straightforward CSA method for determining structural constraints in selectively labeled peptides and proteins.
2. To compare the measurements and known protein structure with recent theoretical calculations of CSA principal values vs. structure.

In summary, we have developed methods to synthesize specifically labeled Protein G B1 Domain and calculated the CSA principal values for five protein G samples, each labeled at a different lysine residue. We have observed some correlation of the principal values with the secondary structure. The methods employed and results obtained are described in the following chapters.

REFERENCES

1. Griffin, R. G. *Nat. Struct. Biol.* **1998**, 5 Suppl, 508-512.
2. Marassi, F. M. and Opella, S. J. *Curr. Opin. Struct. Biol.* **1998**, 8, 640-648.
3. Ketchum, R. R., Hu, W. and Cross, T.A. *Science* **1993**, 261, 1457-1460.
4. Kim, Y., Valentine, K., Opella, S. J., Schendel S. L., and Cramer, W. A. *Protein Science* **1998**, 7, 342-348.
5. Lansbury, P. T. Jr., Costa, P. R., Griffiths, J. M., Simon, E. J., Auger, M., Halverson, K.J., Kocisko, D. A., Hendsch, Z. S., Ashburn, T. T., Spencer, R. G., and et al. *Nat. Struct. Biol.* **1995**, 2, 990-998.
6. Benzinger, T. L., Gregory, D. M., Burkoth, T.S., Miller-Auer, H., Lynn, D. G., Botto, R. E., and Meredith, S. C. *Proc. Natl. Acad. Sci. USA* **1998**, 95, 13407-13412.
7. Wang, J., Balazs, Y. S., and Thompson, L. K. *Biochemistry* **1997**, 36, 1699-1703.
8. Tomita, Y., Oconnor, E. J., and McDermott, A. *J. Am. Chem. Soc.* **1994**, 116, 8766-8771.
9. Studelska, D. R., McDowell, L. M., Espe, M. P., Klug, C. A., and Schaefer, J. *Biochemistry* **1997**, 36, 15555-15560.
10. Jakeman, D. L., Mitchell, D. J., Shuttleworth, W. A., and Evans, J. N. *Biochemistry* **1998**, 37, 12012-12019.
11. Weliky, D. P., Benett, A. E., Zvi, A., Anglister, J., Steinbach, P. J., and Tycko, R. *Nat. Struct. Biol.* **1999**, 6, 141-145.
12. Marassi, F. M., Ramamoorthy, A., and Opella, S. J. *Proc. Natl. Acad. Sci. USA* **1997**, 94, 8551-8556.
13. Derrick, J. P., Wigley, D. B. *J. Mol. Biol.* **1994**, 243, 906-918.
14. Gallagher, T., Alexander, P., Bryan, P., and Gilligand, G. L. *Biochemistry* **1994**, 33, 4721-4729.
15. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield P. T., and Clore, G. M. *Science* **1991**, 253, 657-661.

16. Boutillon, C., Wintjens, R., Lippens, G., Drobecq, H., and Tartar A. *Eur. J. Biochem.* **1995**, 231, 166-180.
17. Dahiyat, B. I., and Mayo, S. L. *Proc. Natl. Acad. Sci. USA* **1997**, 94, 10172-10177
18. Kobayashi, N., Honda, S., Yoshii, H., Uedaira, H., and Munekata, E. *FEBS Letters* **1995**, 366, 99-103.
19. Barlos, K., Gatos, D. *Biopolymers* **1999**, 51, 266-278.

Chapter 2

LINEAR SOLID-PHASE PEPTIDE SYNTHESIS OF PROTEIN G B1 DOMAIN

2.1 Introduction

Peptide chemistry plays a major role in pharmaceutical research. The applications are in the areas of immunology, for the preparation of vaccines, and the study of biologically active molecules like enzymes and hormones. The traditional method of manufacturing proteins is expressing the protein using recombinant technologies. Due to the continuing progress in peptide chemistry, chemical synthesis is a widely recognized alternative method for synthesizing small proteins and protein domains.

Three groups have reported the chemical synthesis of Protein G B1 Domain.^{1,2,3} Only one group has reported a purified yield of 2% by using the tertiary butoxycarbonyl (tBoc) strategy.¹ We have developed a straightforward chemical synthesis of Protein G B1 Domain that routinely gives a 10% purified yield. The protein is synthesized by using solid-phase peptide synthesis methods with 9-fluorenylmethoxycarbonyl (Fmoc) strategy. The protein is synthesized from the C-terminus to the N-terminus on an automated Applied Biosystems 431A peptide synthesizer in our laboratory.

Solid phase peptide synthesis (SPPS)

R. B. Merrifield developed solid phase peptide synthesis in 1963 and has taken a major place in peptide synthesis within the last thirty-seven years.⁴ This method simplified and accelerated the synthesis and also made the synthesis of long peptides practical due to automation. The fundamental principle of the SPPS is that amino acids

can be assembled into a peptide of any sequence. SPPS generally involves the following four steps:

1. Chain assembly
2. Cleavage from resin and removal of side chain protecting groups
3. Purification
4. Characterization

The peptide is assembled from the C-terminus towards the N-terminus. The C-terminus is anchored to an insoluble support. After all the amino acids in the sequence are attached to the support, a reagent is used to cleave the peptide from the support.

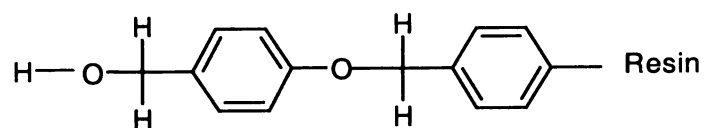
The advantages of using SPPS are that all the reactions involved in the synthesis can be brought to 100% completion. There is no mechanical loss of material since all the synthesis can be carried out in the same vessel and all the laborious purification in intermediate steps in the synthesis are eliminated. The major disadvantage is that in order to ensure complete introduction of the amino acid, excessive amount of amino acids and reagents must be used.⁵

Chemistry

The solid support is a synthetic polymer that has reactive groups. The α -carboxyl group of the amino acid is attached to the support. Many different types of resins are commercially available. We used preloaded (first amino acid already attached) HMP (p-alkoxybenzyl alcohol) resin developed by Wang. It is composed of polystyrene beads with 1% divinyl-benzene, a cross-linking agent.

Figure 2.1:

(a) Structure of the HMP Resin



(b) Structure of Preloaded HMP Resin

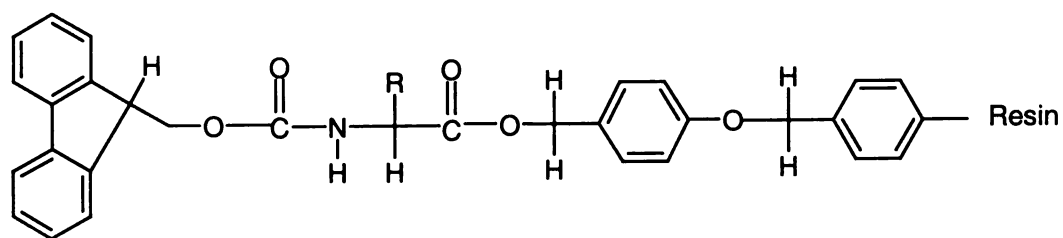
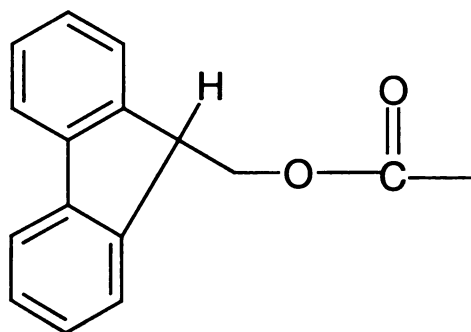


Figure 2.2: 9-fluorenylmethoxycarbonyl (Fmoc) Protecting Group

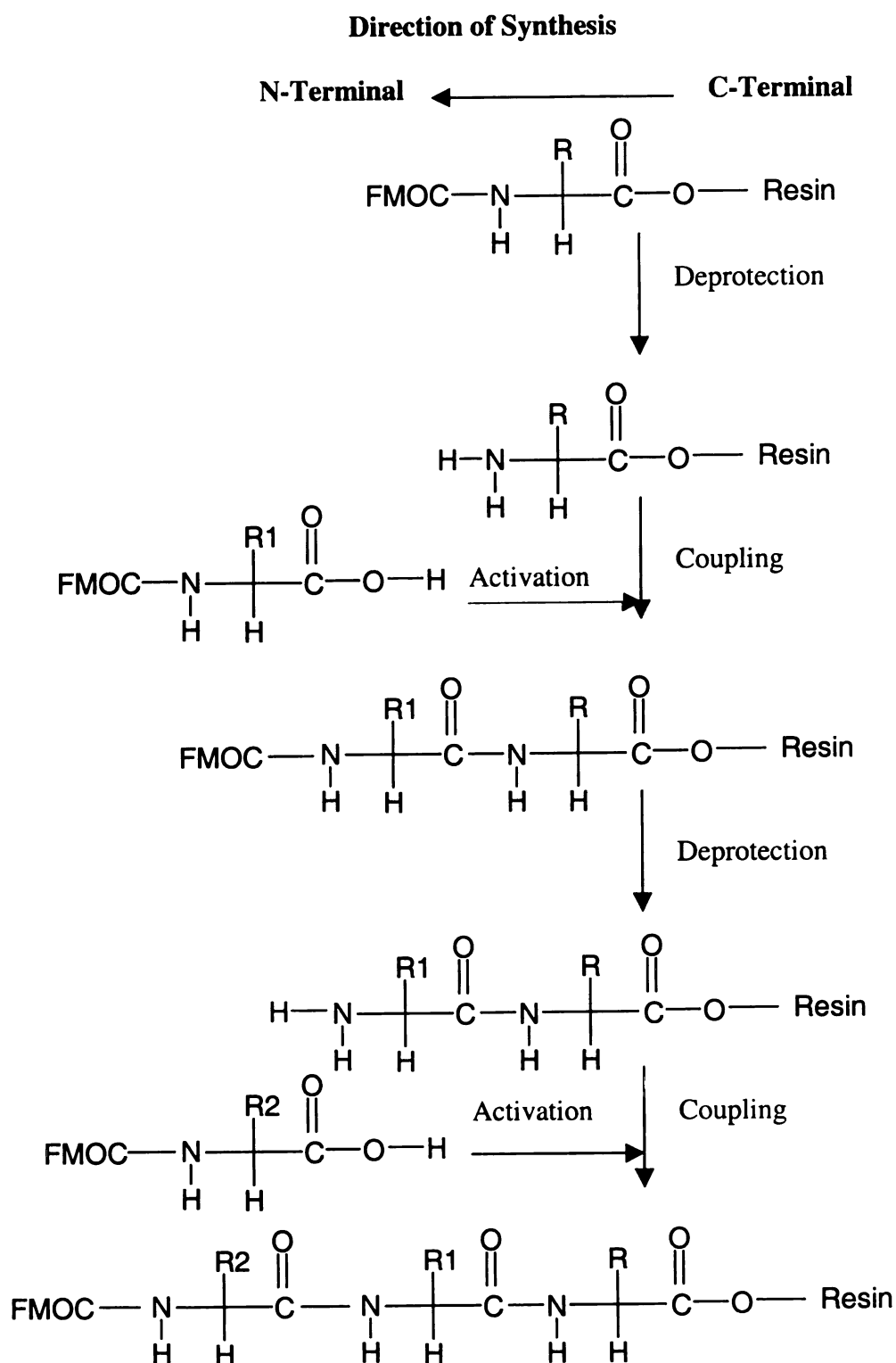


Chain Assembly

9-fluorenylmethoxycarbonyl (FMOC) strategy was used for the chain assembly. FMOC is a protecting group attached to the N-terminus of the amino acid. The synthesis is started with a FMOC protected amino acid attached to the resin. The steps involved in the chain assembly are:

1. Deprotection: Removal of the FMOC protecting group by a base such as piperidine.
2. Activation: The amino acid is activated using HBTU [2-(1 H-benzotriazol-1-yl)-1, 1, 3, 3-tetramethyluroniumhexafluorophosphate] FastMoc activation. The activating agent converts the carboxyl group of the amino acid to an ester.
3. Coupling: The next amino acid is then coupled to the deprotected amino end of the growing peptide chain and forms a peptide bond. The coupling time can be varied depending on the reactions and sequence of the peptides.

Figure 2.3: Linear Solid-Phase Peptide Synthesis



After introducing the desired sequence of amino acids, the final step is deprotection (removal of the Fmoc protecting group on the last amino acid). The protein is washed several times with dichloromethane and dried in a vacuum dessicator. Finally, the protein is cleaved from the resin (solid) support using acidic conditions like trifluoroacetic acid (TFA). Some scavengers like 1, 2-ethanedithiol or thioanisole are also used depending upon the amino acids present in the sequence. The scavengers are used to destroy the free radicals formed during the cleavage of the side chain protecting groups in certain amino acids. After cleavage, the protein is precipitated using tertiary butyl methyl ether and centrifuged three times. It is then purified by reversed-phase high performance liquid chromatography (RP-HPLC). The fractions obtained are lyophilized and the mass is detected by matrix assisted laser desorption ionization (MALDI) mass spectrometry.

2.2 Materials and Methods

The protected amino acids, amino acid resins and the activating agents HOBt and HBTU were obtained from Peptides International. The labeled amino acids were obtained from Cambridge Isotope Labs. The synthesis grade solvents piperidine, N-methyl pyrrolidone (NMP), Dichloromethane (DCM), 2M DIEA/NMP, and acetic anhydride were obtained from Applied Biosystems, Incorporated. TFA and 1, 2-Ethanedithiol (EDT) were purchased from Aldrich.

Our initial approach was to progressively synthesize fragments of the protein and by examination of the fragment chromatogram, discover the size at which synthetic purity and presumably coupling yields decrease.

Following is the Protein G 55 amino acid sequence from the C- to the N- terminus:

ETVTF TKTAD DYTWE GDVGN DNAYQ KFVKE ATAAD VAETT TEGKL
TKGNL ILKYT

To begin, we took 0.1 mmol Wang HMA resin and coupled the last 9 C-terminus residues (ETVTFKTAD) of the protein. 0.5 mmol (five-time excess) of amino acids were used. The coupling time of two hours was used for each amino acid. After the synthesis was complete, the resin was washed with dichloromethane for several times and then dried in a vacuum dessicator. The weight gain was measured and then only ~ 10 mg of the 10-mer fragment was cleaved from the resin using cleavage solution (0.25 mL water + 4.75 mL TFA). The cleavage mixture was rotatorqued (rotated) for 2 hours at room temperature. After filtration, the resin was washed three times with TFA (2 mL each), and the filtrate and washings were concentrated to a small volume by purging with nitrogen. The peptide was precipitated and washed with tertiary butyl methyl ether (20 mL) and centrifuged three times. Reversed-phase HPLC was performed on a Beckman

421 HPLC with a Vydac C18 column. The elution profile was monitored at 280nm. Solvent systems: Buffer A, 10% acetonitrile, 0.1 % TFA and Buffer B, 90% acetonitrile, 0.1 % TFA were used. The peptide was eluted with a linear gradient of 10% B to 40% B in 25 min at a flow rate of 10 mL/min. The main peak was observed at 20% acetonitrile. The fractions were combined and lyophilized using a Labconco Freeze Dryer 18. The molar mass of the peptide (1112.2 g/mol) was confirmed by using a PerSeptive Biosystems Voyager Elite MALDI mass spectrometer as seen in Figure 2.5A.

The next 9 residues (DYTWEGDVG) were then coupled to the 10-amino acid fragment attached to the resin. Coupling time and other procedures after the synthesis were similar. Only due to the presence of tryptophan, cleavage solution (0.125 mL EDT + 0.125 mL water + 4.75 mL TFA) was used. In RP-HPLC, the main peak was observed at 28 % acetonitrile. The anticipated molar mass of the 19-mer fragment is 2135.2 g/mol and was confirmed by MALDI as observed in Figure 2.5B.

The next 10 residues (NDNAYQKFVK) were coupled to the 19-amino acid fragment on the resin. To ensure higher coupling yields and purity, the coupling time was extended from two to four hours for the two Asn residues (Asn35 and Asn37) and one Lys31 residue in the synthesis. These three amino acids, Asn35, Asn37 and Lys31 were reported to couple unreliably in an earlier tBoc synthesis¹ and so were double-coupled in addition to the extended coupling time. (Double couple is to apply the same amino acid twice to ensure total coupling). The next 10 amino acids (EATAADVAET) were coupled directly during two hours of coupling time for each residue. Among the next ten residues (TTEGKLTKGN) coupled to the 39-amino acid fragment, residues Thr11, Lys13, and Thr17 were double coupled and the coupling time was extended to

four hours. The final six amino acids were attached to the 49mer fragment with four hours coupling time and double couple for residues Ile6 and Thr2. The main peak for the 55-mer was observed at 44 % acetonitrile. The actual mass of the 55-mer is 6065 g/mol and was confirmed by MALDI mass spectrometry as seen in Figure 2.5D.

Figure 2.4: Reversed-Phase HPLC Spectra of Protein G (PG) fragments

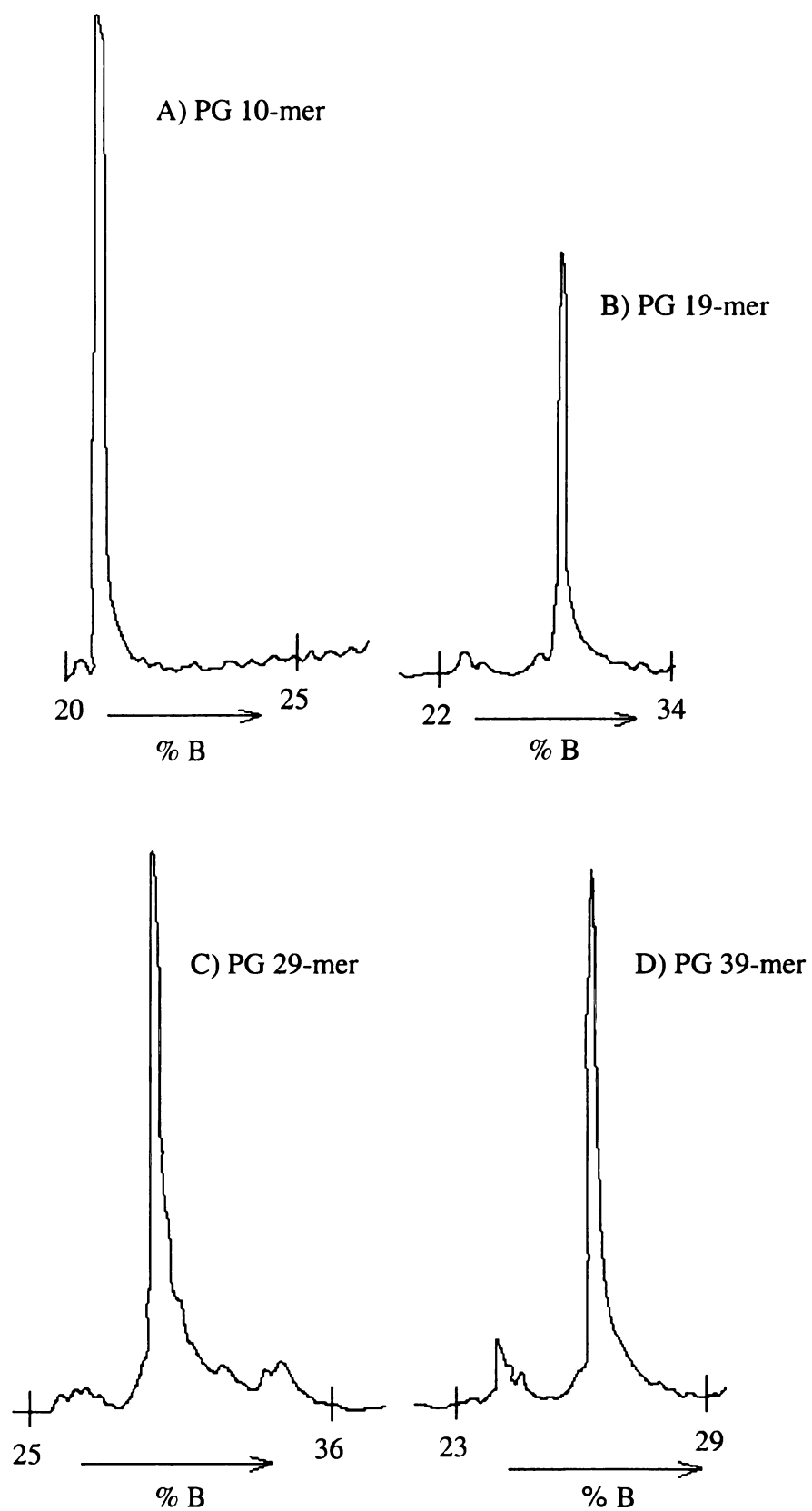


Figure 2.4: (continued)

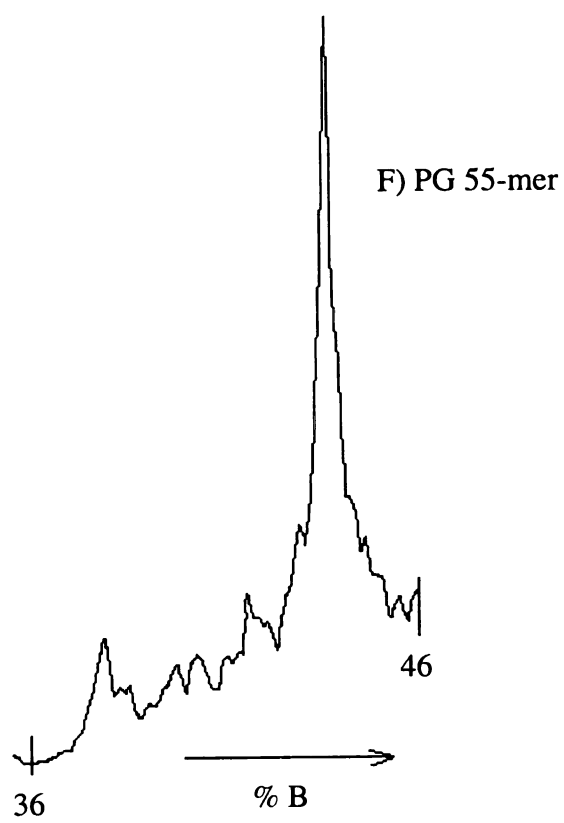
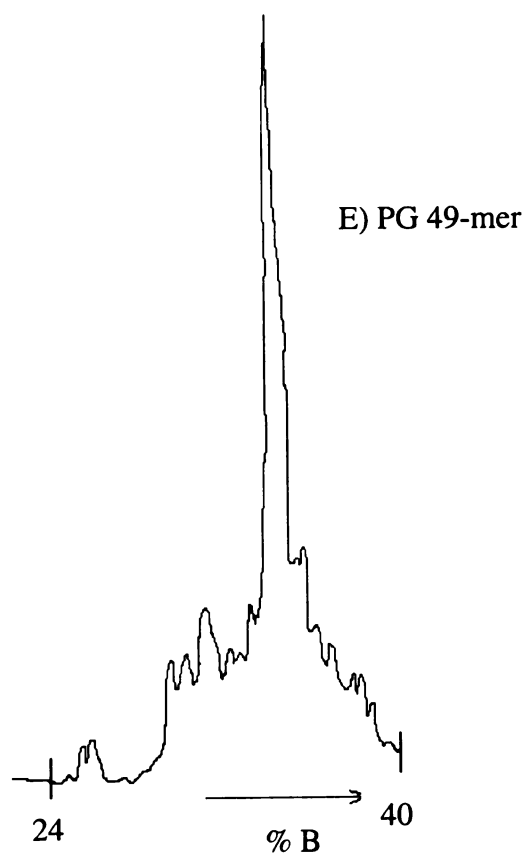
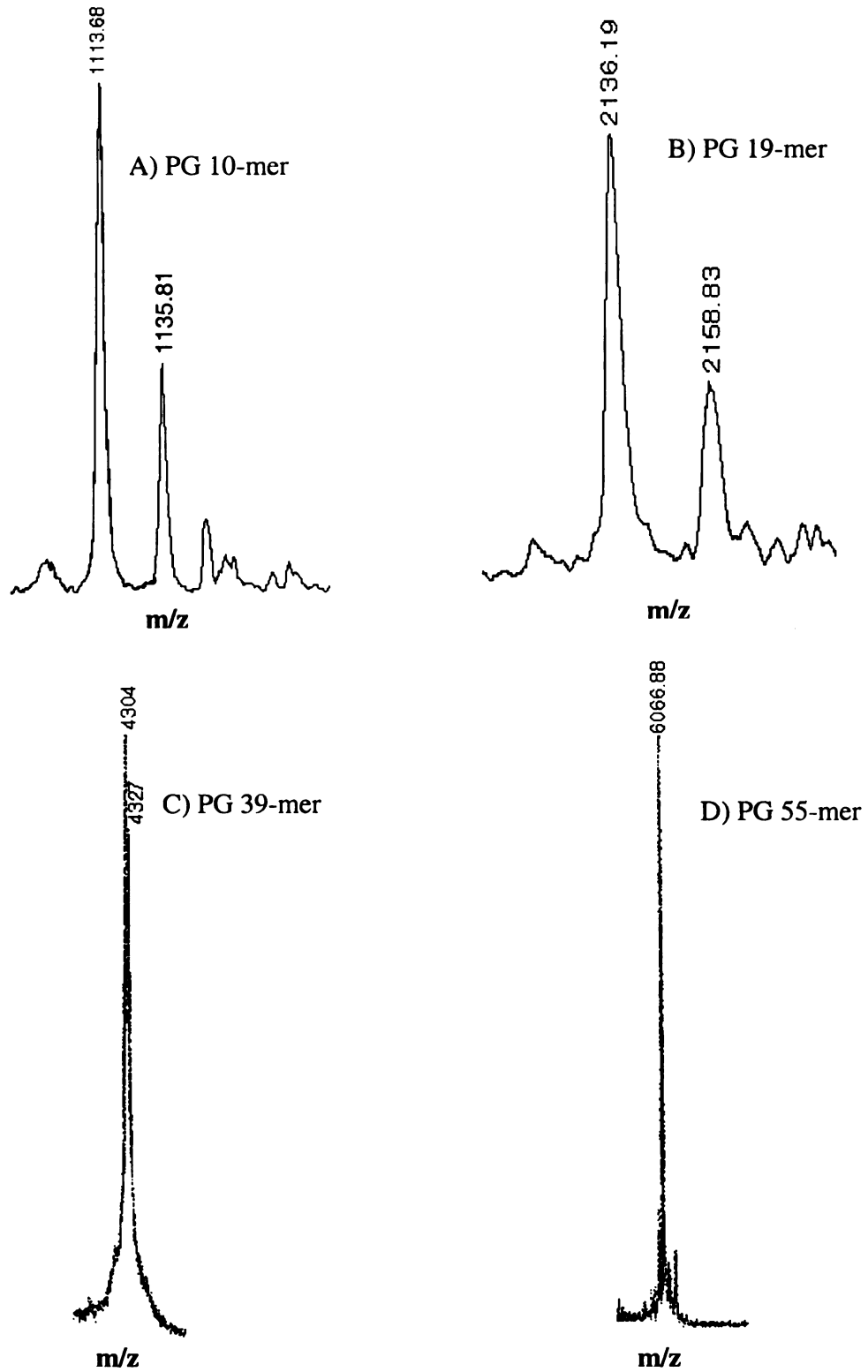


Figure 2.5: Mass Spectra of Protein G fragments

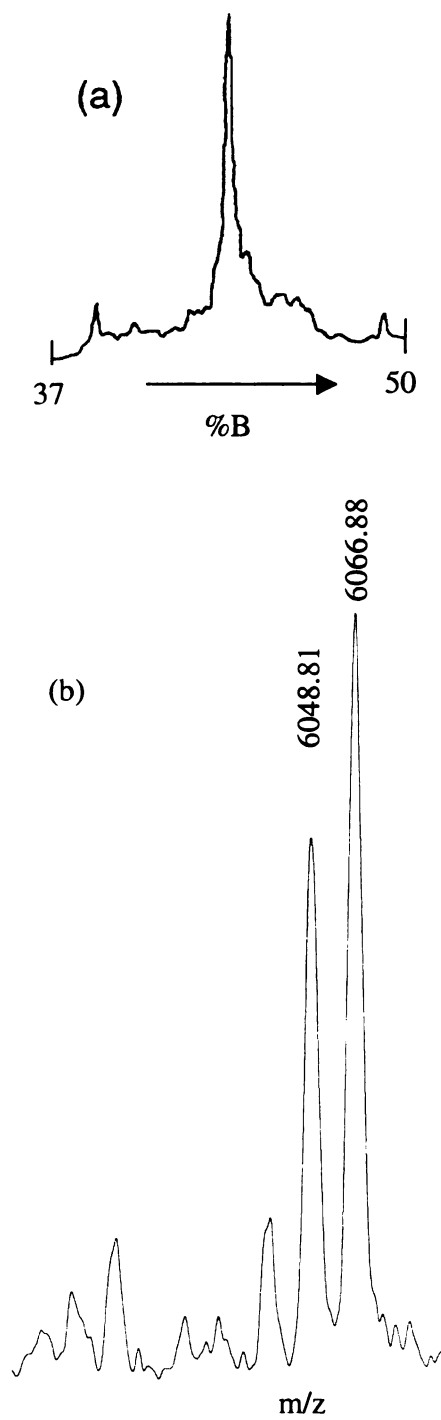


The synthetic approach was then extended to include specific isotopic labels using only 0.2mmol of amino acid/label. For these syntheses, we wished to make five different specifically labeled proteins, each ^{13}C -carbonyl labeled at a particular lysine. Peptide synthesis was started by coupling the last 25 (C-terminal) amino acids to Fmoc-glutamate Wang resin (0.2 mmol/g, 0.500g, 0.1 mmol). The resin-25 mer was then split into five equal portions. The final 30 amino acids were then individually coupled to each portion. Each of the five syntheses only differed in the position of a 1- ^{13}C labeled lysine (Lys-4, Lys-10, Lys-13, Lys-28 or Lys-31). Lys-28 and Lys-31 are in a α -helix, Lys-10 (H-bonded to the solvent and on a turn) and Lys-13 (H-bonded to the solvent) are in a β -sheet and Lys-4 is in a β -sheet (H-bonded to the peptide). Only 0.2mmol of amino acid/label was used. In these syntheses the amino acids were in 25-fold excess while the labeled Lysine (94mg) was in ten-fold excess. To ensure higher coupling yields and purity, the coupling time was extended from two to four hours for the last sixteen residues of the synthesis. The seven amino acids Ile6, Thr11, Lys13, Thr17, Lys31, Asn35, and Asn37 that were reported to couple unreliably in an earlier tertiary butoxy carbonyl (tBoc) synthesis¹ were double coupled.

The peptide was then cleaved from the resin by treating it with a mixture of 40:1:1 TFA:EDT:H₂O with shaking on a rotator at room temperature for 2 hours. After filtration, the resin was washed three times with TFA (2 ml each), and the filtrate and washings were concentrated to a small volume by purging with nitrogen. The peptide was precipitated with tertiary butyl methyl ether (20 ml) and centrifuged three times. Reversed-phase HPLC was performed and the elution profile was monitored at 280nm. The actual molar mass of the labeled Protein G is 6066g/mol and was confirmed

by MALDI mass spectrometry as seen in Figure 2.6B. Using the 280 nm absorbance assay, we calculated a > 10 % yield of purified product for each of these syntheses.

Figure 2.6: (a) RP-HPLC and (b) Mass spectrum of Protein G B1 Domain (55 amino acids).



2.3 Results and Discussion

We found that we could easily make a fragment containing the last 39 amino acids of the protein, but not one that contained the last 49 amino acids. As can be seen in figure 1, the 49-mer fragment of protein G B1 domain would contain the helix, all the β strands $\beta 2$, $\beta 3$ and $\beta 4$, as well as half of $\beta 1$. This may be a large enough fragment to partially fold in the organic solvent of the synthesis (N-methyl pyrrolidone) which may lead to lower coupling yields and purity. Our successful solution to this low coupling problem was to simply extend the coupling time from two to four hours for the last sixteen residues of the synthesis and also double couple the amino acids, which were reported to be difficult in an earlier synthesis.¹ Using the 280nm absorbance assay, we calculated a 10% yield of purified product for the complete synthesis. This corresponds to about 2 μ mole of purified protein. A full synthesis takes about ten days.

We are devising methods to improve the yield and decrease the synthesis time of Protein G B1 Domain. Our approach is the solid-phase condensation of the C-terminal 33-mer, middle 11-mer, and N-terminal 11-mer peptide fragments. Each fragment can be individually labeled and then coupled with the other fragment. The details and work done using this novel method is described in chapter 3.

REFERENCES

1. Boutillon, C., Wintjens, R., Lippens, G., Drobecq, H., and Tartar A. *Eur. J. Biochem.* **1995**, 231, 166-180.
2. Dahiyat, B. I., and Mayo, S. L. *Proc. Natl. Acad. Sci. USA* **1997**, 94, 10172-10177
3. Kobayashi, N., Honda, S., Yoshii, H., Uedaira, H., and Munekata, E. *FEBS Letters* **1995**, 366, 99-103.
4. Barlos, K., Gatos, D. *Biopolymers* **1999**, 51, 266-278.
5. Atherton, E., and Sheppard, R. C. Solid phase peptide synthesis : A practical approach *Publisher Oxford, England ; New York : IRL Press at Oxford University Press*, **1989**.

Chapter 3

SYNTHESIS OF PROTEIN G B1 DOMAIN BY SOLID-PHASE FRAGMENT CONDENSATION

3.1 Introduction

Condensation of fully protected peptide fragments on a solid support is an efficient alternative to stepwise solid phase peptide synthesis (SPPS). Stepwise SPPS is often efficient for peptides that are 20 to 30 amino acids in length. However, for longer peptides and small proteins (50 amino acid residues and more) the SPPS is not very efficient.¹ The main disadvantage of SPPS is the problem of non-quantitative yield of the amino acid coupling and N^α deprotection steps. Also, it is not possible to retain the optical integrity of the amino acids during the coupling reaction.² These problems can be overcome by solid-phase fragment condensation strategy. The main advantages of this approach are that instead of amino acids fully protected fragments are coupled instead of individual amino acids. Reduction of the coupling steps should potentially increase the yield and purity of the target peptide.

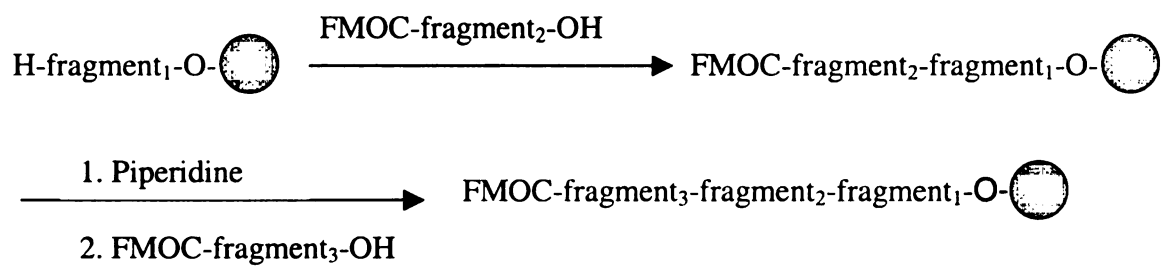
The synthesis depends on several parameters crucial for its success. The parameters include length of peptide, resin or solid support used, purity of peptide fragments, how soluble they are in the solvent, and activating agents used. The assembly of the protected fragments to the target protein can be started in three ways:

- (a) From the C-terminal fragment by extension of the peptide chain toward the N-terminus
- (b) By chain extension from the N- to the C-terminus

(c) From a middle region toward both directions

Among them only method (a) has been successfully applied so far for the production of proteins.³

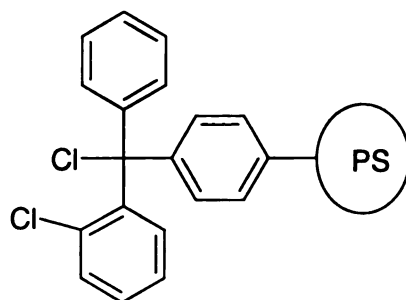
Figure 3.1: Convergent peptide synthesis with C- to N-terminal chain extension³



In solid-phase fragment condensation it is very important to divide the peptide into suitable fragments. Peptide fragments that contain preferentially glycine and proline as the C-terminal amino acid are selected in order to minimize the risk of epimerization.³ (Epimerization is a process in which the products formed are stereochemical isomers). If this cannot be achieved, then amino acids with small and nonfunctionalized side-chains like leucine or alanine are considered as the next best choice at C-terminal positions.² According to the studies conducted regarding the different tendencies of amino acids to racemize, it is concluded that the amino acids His, Cys, Phe, Tyr, Trp, Asp, Asn and Gln are especially prone to side-reactions and should not be considered for C-terminal positions in protected peptide fragments. Also, the amino acids Gln, Pro, Val, and Ile are avoided in the N-terminal positions.²

A wide variety of resins are commercially available. Due to the base lability of the Fmoc group and the acid sensitivity of amino-acid side chain protection of the tBu, trityl type, resins are chosen that can be cleaved under extremely mild conditions. The resin 2-chlorotrityl chloride is most commonly used. The esterification of the resin can be performed without racemization by treating it with Fmoc-amino acids and diisopropylethylamine (DIPEA) in dichloromethane (DCM). Other resins like Sasrin, benzyl-alcohol, oxime, amide, Sieber resins are also used depending on the chemistry required for the condensation reactions.²

Figure 3.2: Structure of 2-Chlorotrityl Chloride Resin



The selected fragments are then synthesized by linear solid phase peptide synthesis (SPPS) using 9-fluorenylmethoxycarbonyl (FMOC) or tertiary butoxycarbonyl (BOC) chemistry. The peptide can be cleaved from the resin by treatment with mixtures of acetic acid (AcOH)/ trifluoroethanol (TFE)/DCM or with TFE/DCM. The peptide cleaved by treatment with TFE/DCM (2:8) is obtained in a 70-80% yield. Fast, selective and quantitative cleavage occurs using AcOH/TFE/DCM (1:2:7), however AcOH contained in the cleavage mixture is very difficult to remove. Reprecipitation of the fragment from TFE/Water or DMF/water is necessary for its complete removal.⁴

The purity and nature of the solvents and activating agents used in the condensation reactions are very critical. Different solvents like N, N-dimethylformide (DMF), N, N-dimethylacetamide (DMAC) and N-methylpyrrolidone (NMP) are used. However, studies show that dimethylsulfoxide (DMSO) and DCM are the best solvents to reduce the formation of several unwanted by-products.³ Activating agents such as O-benzotriazolyl-N, N, N', N',-tetramethyluronium-tetrafluoroborate (TBTU), hydroxybenzotriazole (HOBt), dicyclohexylcarbodiimide (DCC), diisopropylcarbodiimide (DIC), diisopropylethylamine (DIEA) are most commonly used.²

In most cases, the protected fragments can be cleaved from the resin with >97% purity.³ Therefore, these segments can be used in the condensation reactions without further purification. For peptide fragments that need purification, preparative reverse phase HPLC using acetonitrile/water system, as the eluent is best suited for purifying FMOC/tBu-protected segments.

Many proteins have been successfully synthesized by this solid-phase convergent approach. The syntheses were carried out using different combinations of resins, solvents and activating agents. Table 3.1 summarizes the list of peptides and proteins synthesized according to the convergent strategy on 2-chlorotrityl chloride resin.

Table 3.1

Peptides and Proteins Synthesized on 2-Chlorotrityl Chloride Resin by the SPFC Strategy³

Year	Peptide	Residues
1990	Human Prothymosin α	109
1993	Gly ³³ human calcitonin	33
1993	T-cell receptor Hv β 2	109
1994	Antifreeze protein type III	64
1994	TNF- α	157
1998	Tyr ⁰ -Atriopeptin (rat)	24
1998	Tetanus toxin MUC-1 oligomers	115

3.2 Materials and Method

2-Chlorotrityl chloride resin with 1.49mmol/g substitution was obtained from Peptides International. The solvents Dimethylsulfoxide (DMSO), diisopropylethylamine (DIEA) and diisopropylcarbodiimide (DIC) were from Aldrich. Synthesis grade DCM and N-methyl pyrrolidone (NMP) were obtained from PE Biosystems. Activating agents Hydroxybenzotriazole (HOBt), 2-(1 H-benzotriazol-1-yl)-1, 1, 3, 3-tetramethyluroniumhexafluorophosphate (HBTU) and O-benzotriazolyl-N, N, N', N', tetramethyluronium-tetrafluoroborate (TBTU) were obtained from Peptides International.

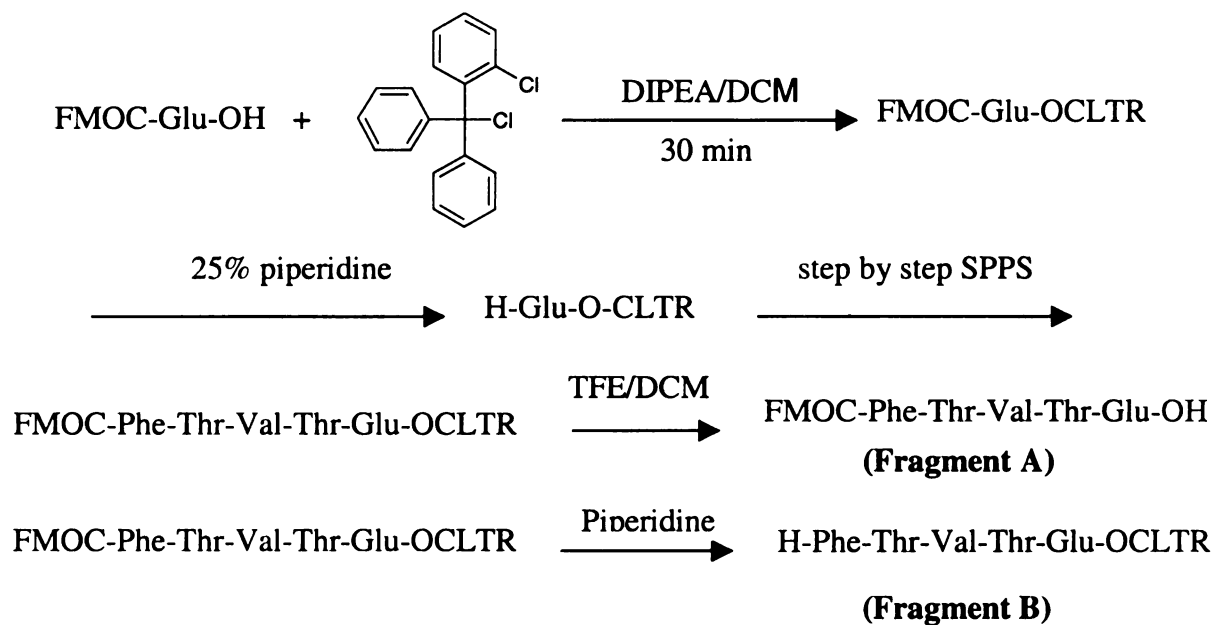
The peptide fragments were synthesized on an automated Applied Biosystems 431A peptide synthesizer. The synthesis program FastMoc was used. The program was modified such that the capping cycle (acetylation of the free amino/N-terminus by acetic anhydride) was eliminated as acetic anhydride can cause cleavage of the peptide chain from the 2-chlorotrityl chloride resin. The final deprotection step (removal of Fmoc from the last amino acid) was removed for certain peptide fragments. RP-HPLC was performed on a Beckman 421 HPLC with a Vydac C18 column. The elution profile was monitored at 280nm. Solvent systems: Buffer A, 10% acetonitrile, 0.1 % TFA and Buffer B, 90% acetonitrile, 0.1 % TFA were used. The molar mass of the peptides was confirmed by using PerSeptive Biosystems Voyager Elite MALDI mass spectrometer.

In order to devise a method for synthesizing Protein G with SPFC strategy, we started from condensing a 5-amino acid (5-AA) fragment to another 5-AA fragment. The 5-amino acid fragments were the last five C-terminal residues (ETVTF) of Protein G. Both the fragments were synthesized on the 2-chlorotrityl chloride resin. For the first fragment (fragment A), the deprotection step was not performed at the end to obtain a

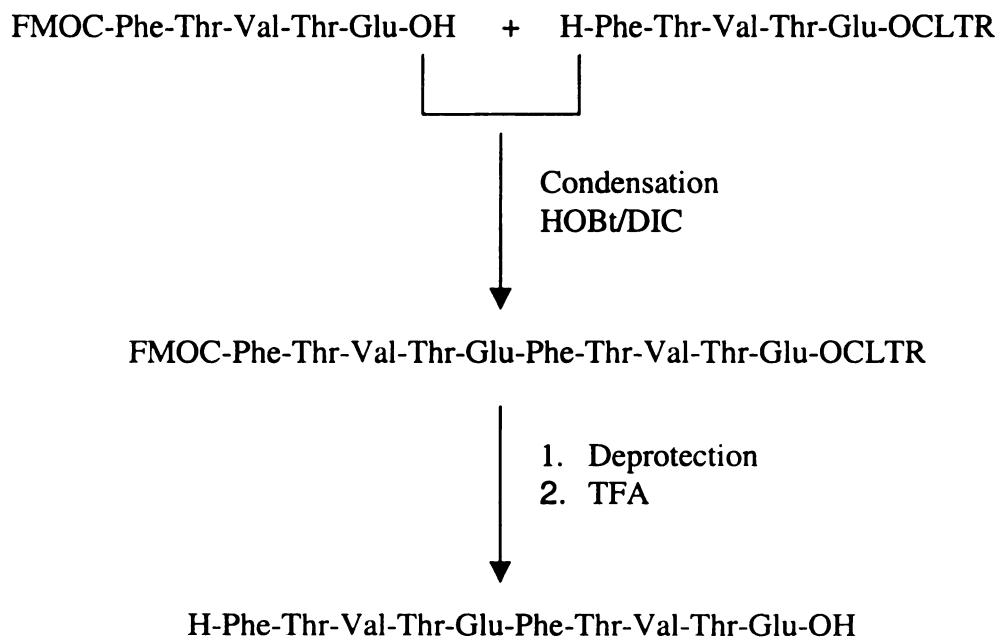
Resin-5AA-FMOC peptide. The other 5-AA fragment (fragment B) was completely deprotected. Resin-5AA-FMOC was then cleaved with 10 mL of a TFE:DCM (2:8) mixture with shaking at room temperature for three hours. After filtration, the resin was washed twice with TFE/DCM mixture and then three times with DCM. The filtrate and the washings were combined and concentrated to a small volume by purging with N₂. The peptide was precipitated with 25mL tertiary butyl methyl ether and centrifuged three times. The N-terminus FMOC protected 5-AA fragment was obtained with about 70% yield. Due to the presence of FMOC and other protecting groups, fragment A was insoluble in water and also in water/acetonitrile mixture. As a result, RP-HPLC could not be performed to purify the fragment prior to using it in the condensation reaction. A small-scale (pilot) synthesis was performed by condensing a 3-fold molar excess of protected fragment A over the resin bound fragment B. Fragment A was dissolved in DMSO and with condensing agents HOBt and DIC (1:2:2) and applied to fragment B. After 24 hours, the completion of condensation reaction was confirmed by performing the Kaiser test on the resin beads. The amino-terminal FMOC group was removed by treatment with 25% piperidine in NMP. The resin-bound peptide was then cleaved with a TFA/water solution and analyzed by RP-HPLC. The molar mass was detected by MALDI mass spectrometry. The anticipated molar mass of the 10-mer was 1173 g/mol and was confirmed by MALDI as observed in Figure 3.3c.

Figure 3.3: Synthesis of 5+5-mer Peptide

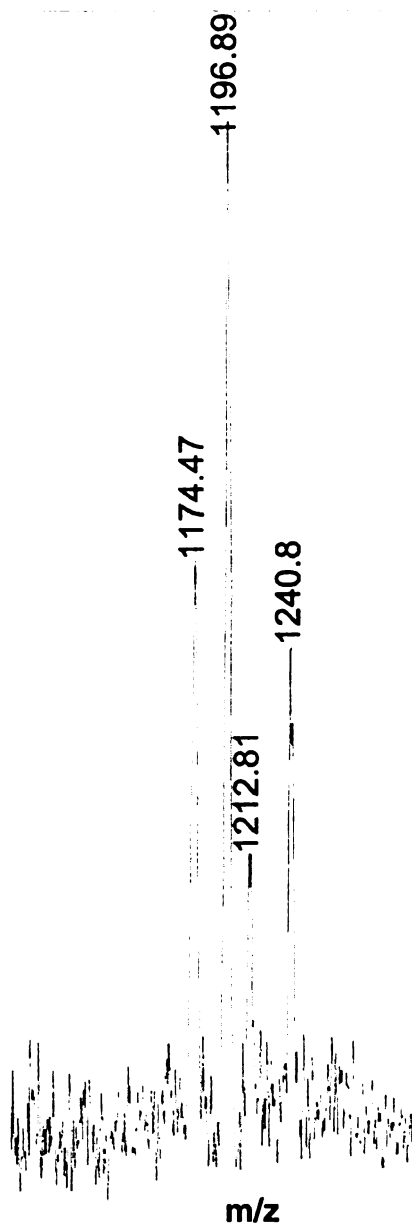
a) Strategy for Fmoc synthesis of basic peptide fragments (A and B)



b) Fragment condensation strategy for synthesis of 10-mer peptide



c) Mass Spectrum of the 10-mer peptide

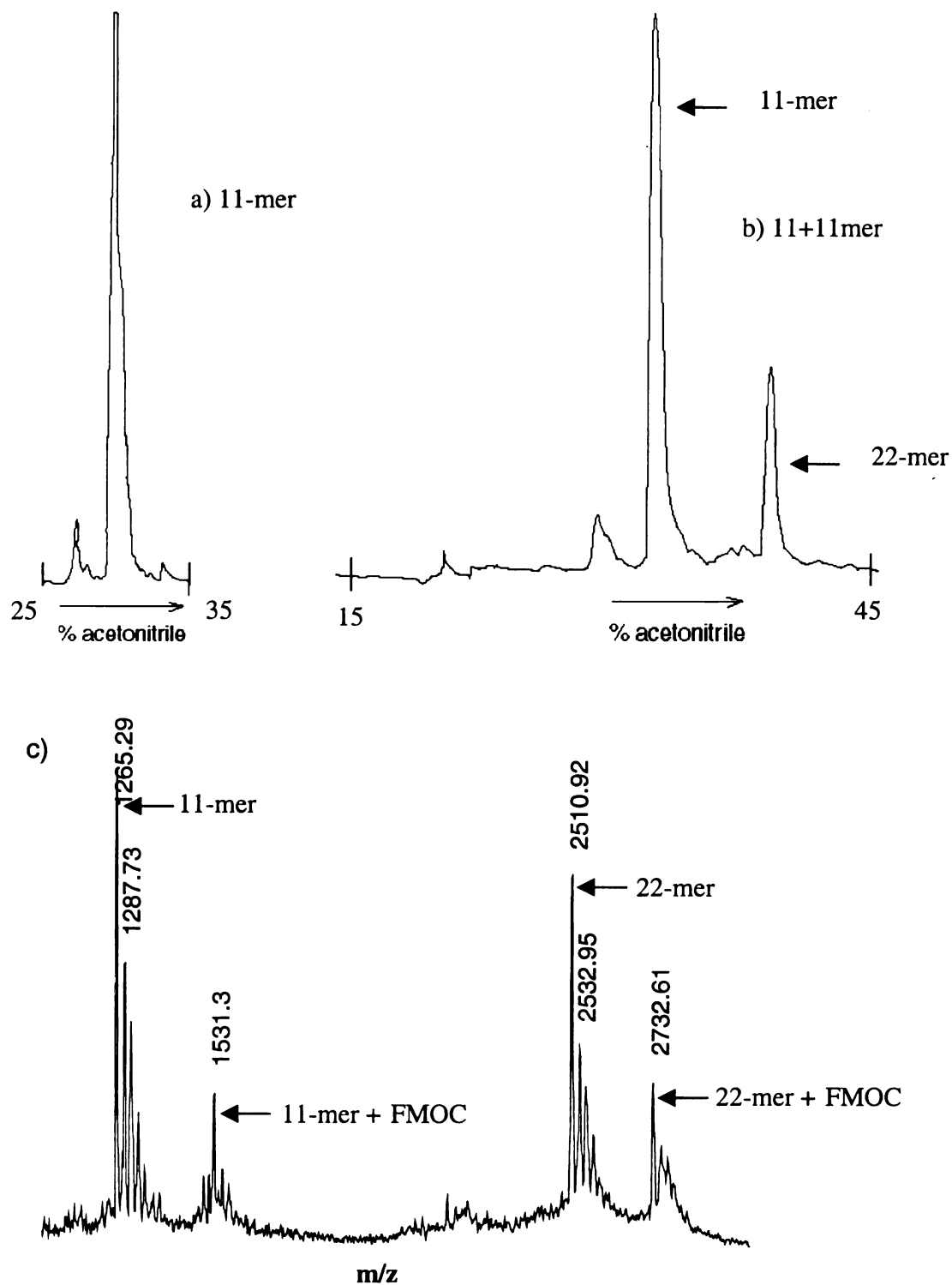


The SPFC strategy was then extended to attach an 11-AA fragment (**HGRVGIYFGMK**) to another 11-AA (**HGRVGIYFGMK**) peptide. Both the 11-AA peptides were synthesized on 2-chlorotrityl chloride resin by an automated solid-phase Fmoc strategy (431A Peptide Synthesizer). As described earlier, one 11AA fragment (fragment A) was cleaved from the resin using TFE/DCM mixture to obtain a fully protected peptide fragment with N-terminal Fmoc group. The other fragment (fragment B) was kept resin-bound. In order to enhance the solubility of the peptide in DMSO, fragment A was washed with acetonitrile/water (90:10) mixture three times and lyophilized. The condensation strategy used was same as described above. Both the RP-HPLC and MALDI confirmed the formation of a 22-mer peptide (molar mass 2511 g/mol) as seen in Figure 3.4 b and c.

Figure 3.4: Synthesis of 11+11-mer Peptide

a) RP-HPLC of 11-mer peptide b) RP-HPLC spectrum of 22-mer and

c) Mass Spectrum of 22-mer.



In order to synthesize Protein G with SPFC methods, the peptide was divided into three fragments. Following is the sequence of 55-residue Protein G B1 Domain from the C- to the N- terminus.

**ETVTF TKTAD DYTWE GDVGN DNAYQ KFVKE ATAAD VAETT TEGKL
TKGNL ILKYT**

The three fragments were as follows:

1. Last 18 C-terminal residues ETVTF TKTAD DYTWE GDV
2. Middle 24-mer GNDNAYQ KFVKE ATAAD VAETTTE
3. First 13 N-terminal amino acids GKL TKGNL ILKYT

Fragments 2 and 3 were synthesized on 2-chlorotrityl chloride resin using the Fmoc strategy. Fragment 2 was deprotected to remove the N-terminal Fmoc group (free N-terminus) and was kept resin bound. Fragment 3 with Fmoc on N-terminus was cleaved from the resin with TFE/DCM. The protected fragment 3 was washed several times with acetonitrile/water mixture and lyophilized. The solubility of fragment 3 in DMSO significantly increased after the washings. The protected fragment was dissolved in a minimum amount of DMSO with coupling agents HOBt/DIC (1:2:2) and applied the resin bound 24-mer. The coupling reaction was allowed to run for 24 hours. However, the RP-HPLC and mass spectrum revealed that the 37-mer was not formed. In order to carry out the reactions under varied conditions, three reactions were set up simultaneously.

1. Fragment 3 was dissolved in DMSO with coupling agents HOBt/DIC (1:2:2) and applied the resin bound 24-mer.

2. Fragment 3 was dissolved in DMSO with coupling reagents TBTU/HOBt/DIEA (1:1:1:1.8) and allowed to couple with the 24-mer.
3. Reaction 1 was repeated with the difference that it was carried out at a constant temperature of about 35 degrees Celsius.

All the coupling reactions were conducted for 24 hours and the protected fragment was applied in 4-fold molar excess over the resin-bound 24-mer. However, none of the three reactions was successful.

To explore suitable fragments, Protein G was then divided into three different fragments.

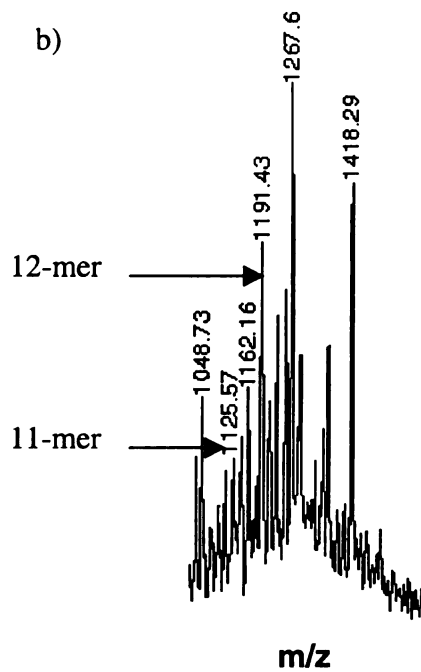
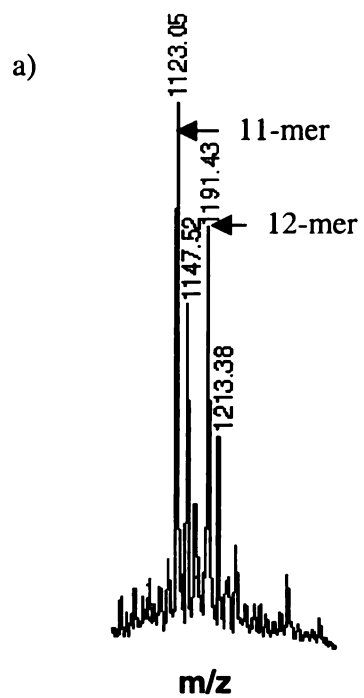
1. Last 33 C-terminal residues ETVTF TKTAD DYTWE GDVGN DNAYQ KFVKE
ATA
2. Middle 11mer ADVAE TTTEGK
3. First 11 N-terminal amino acids LTKGNL ILKYT

Fragment 1 was kept resin-bound on 2-chlorotriyl chloride resin and has a free N-terminus. The 2nd fragment was fully protected and cleaved from the resin. Two reactions were set up. In reaction I, fragment 2 was dissolved in minimum amount of DMSO with coupling agents HOBt/DIC and was applied to the resin-bound fragment in five fold molar excess. In reaction II, fragment 2 was dissolved in 1:1 DCM:NMP solvent mixture with the same coupling agents and was applied to fragment 1 in five-fold molar excess. However, RP-HPLC and MALDI revealed that both the reactions were unsuccessful. The MALDI spectrum showed the attachment of the 11-mer to the resin directly.

In order to study the conformational stability of the 11-mer fragment, we performed a pilot synthesis of condensing the 11-mer fragment with one amino acid (Ala) attached to resin. The solvent used was DMSO and 2 equivalents of HOBt and DIC with respect to the protected peptide were added. Formation of the 12-mer peptide was confirmed by MALDI mass spectrometry as seen in Figure 3.5. The actual mass of the 11-mer is 1121 g/mol and that of the 12-mer is 1192g/mol.

Minor changes were made to the automated synthesis procedure while synthesizing the 11-mer and 33-mer fragments. The capping cycle was performed during the attachment of the first amino acid to the resin and also during the attachment of the last amino acid. The condensation reactions using HOBt/DIC coupling in DMSO were repeated for Resin-Ala + 11-mer and resin-33mer + 11-mer peptides with the modified fragments.

Figure 3.5: Mass Spectra a) 12-mer On High Substituted Resin with no capping
b) 12-mer On Low Substituted Resin with capping



3.3 Results and Discussion

Solid-Phase fragment condensation is based on the principle that protected peptide fragments corresponding to the entire protein sequence can be condensed sequentially on a suitable solid support.⁵ We have attempted to apply this method to synthesize the 55-residue Protein G B1 domain using 2-chlorotrityl resin as the solid support by the Fmoc strategy.

The first choice of the three fragments (18-mer, 24-mer, and 13-mer) for Protein G was based on having glycine at the C-terminus of the 24-mer and 13-mer to minimize epimerization. The optical stability of glycine at the C-terminal position of the electrophilically activated fragment would ensure the optical purity of the resulting oligomers. It was observed that the solubility of the N-terminal 13-mer protected fragment in DMSO was not very efficient. Also this indicated that the solubility of the 24-mer + 13-mer after condensation would be even more difficult. Therefore, we changed the strategy and the fragments were divided as resin-bound 33-mer, middle 11-mer and N-terminal 11-mer. The rationale for this choice was that (1) the solubility of both the protected 11-mer fragments in DMSO greatly improved and (2) amino acids alanine and leucine were at the C-terminus of fragment 2 (11-mer) and fragment 3 (11-mer), respectively. Both alanine and leucine have small and nonfunctionalized side-chains.

Since both the SPFC reactions using solvents DMSO and DCM/NMP mixture were unsuccessful, we tested the conformational stability of the fragment 1 (11-mer) by condensing the 11-mer fragment to Ala attached to the resin. The formation of the 12-mer peptide proved that there were no problems with the activation, however, the mass

spectrum showed that the 11-mer fragment also attached to the resin directly as seen in Figure 3.5a. This shows that there is a competition between the resin and the N-terminus of the 33-mer to which the 11-mer fragment can get attached. In order to avoid this competition, we made a low substituted 2-chlorotrityl chloride-Ala-FMOC resin and also performed capping. Capping converts all the free amino groups on the resin to acetates. The condensation reaction of the low substituted resin-Ala + 11-mer fragment was repeated. The mass spectrum confirms that the attachment of the 11-mer to the resin has significantly reduced as seen in Figure 3.5b. The 33-mer fragment was synthesized again using the low substituted 2-chlorotrityl chloride-Glu-FMOC resin with capping. However, the 11-mer fragment did not get attached to the low substituted resin 33-mer.

It is clear from the synthesis of 11 + 1-mer fragment that the 11-mer is activated by the combination of activating agents used. However, the attachment of the 11-mer to the resin shows that the resin is not capped properly and some unwanted products, which can lead to side reactions, might be formed during the coupling reaction. Also it is suggested that even though only the solubilisation of protected peptide is necessary to achieve successful reaction, problems with the low solubility of intermediates persist in many cases.²

The exact reason for the failure of the SPFC reaction of 33-mer + 11-mer is still unknown. We need to optimize the techniques including distillation of the solvent (DMSO) used so that there are no side reactions, perform the reaction under Argon so that no air or moisture is introduced during the condensation process, use different solvent or solvent mixtures, and activating agents. One approach to consider is dividing the chain of 55 amino acids into five or six fragments. This seems to be reasonable in

terms of balance, between synthesizing fragments effectively in a stepwise manner, solubility of the protected peptide fragments, and effective coupling of the fragments. Also it will be clear from the stepwise approach, where exactly the SPFC reaction begins to fail for Protein G B1 domain.

REFERENCES

1. Atherton, E., Sheppard, R. C. *Solid Phase Peptide Synthesis: A Practical Approach* IRL Press at Oxford University Press: Oxford, **1989**
2. Benz, H. *Synthesis* **1994**, 337-358.
3. Barlos, K., Gatos, D. *Biopolymers* **1999**, 51, 266-278.
4. Chan, W. C., and White, P. D. Fmoc Solid Phase Peptide Synthesis: A Practical Approach, *Publisher Oxford University Press* **2000**
5. Krambovitis, E., Hatzidakis, G., and Kleomenis, B. *The Journal of Biological Chemistry* **1998**, 273, 10874-10879.

Chapter 4

CHEMICAL SHIFT ANISOTROPY PRINCIPAL VALUES IN PROTEIN SECONDARY STRUCTURE DETERMINATION

4.1 Introduction

Chemical shift anisotropy (CSA) values can be straightforwardly measured by solid state NMR and may provide useful constraints for structure determination of proteins in the solid state. We are investigating the dependence of measured CSA principal values on local secondary structure and hydrogen bonding environments in proteins. For these studies, we are using the fifty-five-residue B1 domain of streptococcal protein G as a model because it has a known 1.1 Å crystal structure and because it can be readily synthesized with specific labels on an automated peptide synthesizer. Using protein labels at specific lysine carbonyl positions, we have observed some correlation between the CSA principal values and local secondary structure and hydrogen bonding.

The use of chemical shifts as constraints in secondary structure determination of proteins is well characterized. Beginning with the work of Ando and coworkers,¹ and followed with studies by Bax² and Wishart and Skyes,³ the shifts of H^N , H^α , C^α , C^β , CO, and ^{15}N amide have all been correlated with secondary structure.⁴ Concurrently, there have been significant improvements in *ab initio* calculations to theoretically understand these correlations.⁵⁻⁷

There is also strong evidence for correlation between the nuclear CSA principal values δ_{11} , δ_{22} , and δ_{33} and local secondary structure and/or hydrogen bonding. This is reasonable because the chemical shift $\delta = (\delta_{11} + \delta_{22} + \delta_{33})/3$. Ando and coworkers

studied a number of di- and tripeptides as well as some synthetic biopolymers (e.g., polyalanine) and deduced correlations between the carbonyl principal values and the distance between the carbonyl O and the amide N of its hydrogen bond partner.^{8,9} For carboxylic acids, McDermott and coworkers found a correlation between carboxyl CSA principal values and hydrogen bond length.¹⁰ More recently, there has been solution NMR work on the correlation between the H^N CSA and its hydrogen bond length¹¹ and the C^α CSA and its local secondary structure.¹² In the solution NMR studies, the H^N and C^α CSA is the difference between the chemical shift parallel and perpendicular to the H-N and C-H bonds, respectively. A recent experimental/theoretical investigation of the C^α CSA principal values for the central Ala of the Ala-Ala-Ala and Gly, Ala, Val crystalline tripeptides correlated with dihedral angles which were within 12° of those observed crystallographically.^{13,14}

All of the previous studies point to the possibility that solid state NMR measurements of CSA principal values of nuclei in full proteins should provide useful constraints on local secondary structure and hydrogen bonding. We are making these measurements on protein G B1 domain to observe whether the correlation exists and can provide practical structural information in a real non-crystalline protein.

Measurements of CSA principal values have several potential advantages:

1. They can be done on any sample with a few specific labels. More sophisticated techniques such as 2D MAS (magic angle spinning) exchange,¹⁵ or REDOR (rotational-echo double resonance),¹⁶ require labeling of particular nuclei at particular positions. CSA measurements can be done on any of the samples specially prepared for these more sophisticated techniques and do not require preparation of additional

samples. This is significant because sample preparation is often the rate-limiting step in biological solid state NMR. The data from more complicated techniques are often consistent with a few different distinct structures¹⁷ and the CSA principal value measurements should help to distinguish between them.

2. The principal values are obtained from a straightforward 1D CP (cross polarization)-MAS (magic angle spinning)¹⁸ sequence, which is available on all commercial spectrometers and is easy to set up, even for non-solid state NMR spectroscopists. Principal values are obtained from analysis of sideband intensities¹⁹ using widely available computer programs.
3. The measurements have high sensitivity and are suited to the limited sample quantities of biological samples. The principal values can be obtained from the peak intensities of a few 1D MAS spectra at different spinning speeds.
4. In addition to their use in solid state NMR protein structure determination, the measurements should provide insight into liquid state NMR CSA measurements and their correlation with protein structure.^{11, 12}

Chemical Shift

The phenomenon of chemical shift arises because of shielding of the nuclei from the external magnetic field by the electrons. Electrons are induced to circulate around the nucleus about the longitudinal axis of the applied field B_0 . The angular velocity is given by

$$\omega_i = (e/2m_e) B_0$$

Moving electrons produce a moving current and the moving current induces a magnetic field. Therefore, the effective magnetic field experienced by the nucleus is given by the equation:

$$B = B_0(1-\sigma)$$

Where σ , a dimensionless number (usually listed in parts per million), is known as the shielding constant. Since the shielding effect is caused by electronic environment, values of σ vary with the position of the nucleus in the molecule. Thus, the shielding constant for the proton of an aldehyde group is lower than that for the protons of a methyl group. Variations in σ cause variations in resonance frequencies and this leads to the occurrence of chemical shifts.

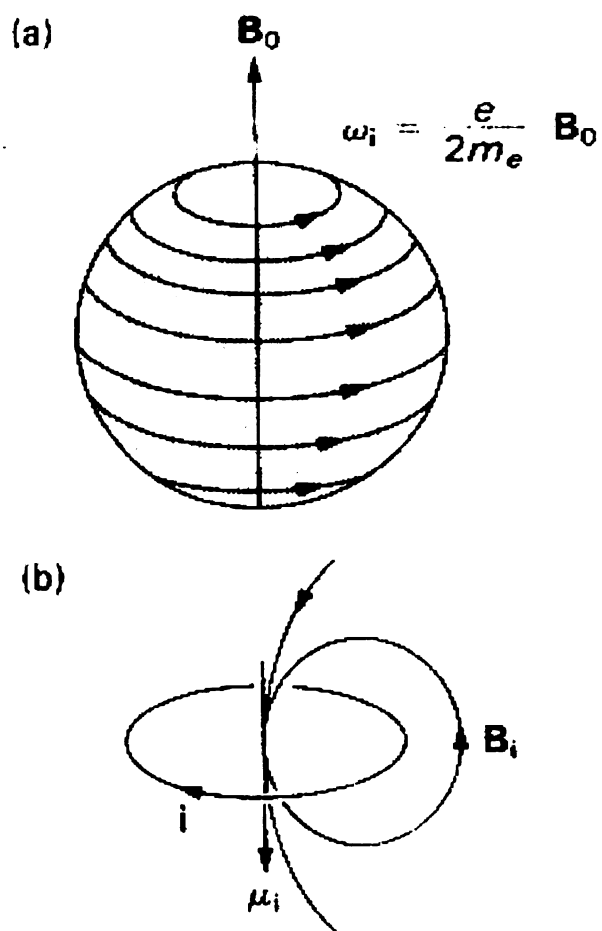
$$\nu_j = |\gamma/2\pi| B_0(1-\sigma_j)$$

ν_j = resonance frequency (Larmor frequency)

σ_j = shielding constant

Figure 4.1: Electronic Shielding²⁰

- (a) Circulation of the electronic charge cloud under the influence of a magnetic field
- (b) The secondary magnetic field produced by the precession



Chemical Shift Anisotropy

Chemical shift anisotropy is the dependence of the chemical shift on the orientation of the functional group, relative to the direction of the external magnetic field. Thus, it is the dependence of shielding on the orientation of the functional group. Nuclear shielding can be expressed as a second-rank tensor. In solid state, non-spinning cross polarization (CP) measurements can yield values for the individual components of the shielding tensor, σ , as expressed by:

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

σ_{ii} are the principal tensor components such that $\sigma_{11} < \sigma_{22} < \sigma_{33}$.

The rapid isotropic molecular motion in liquid averages the isotropic nuclear shielding and is expressed as $\sigma_{ISO} = (\sigma_{11} + \sigma_{22} + \sigma_{33})/3$

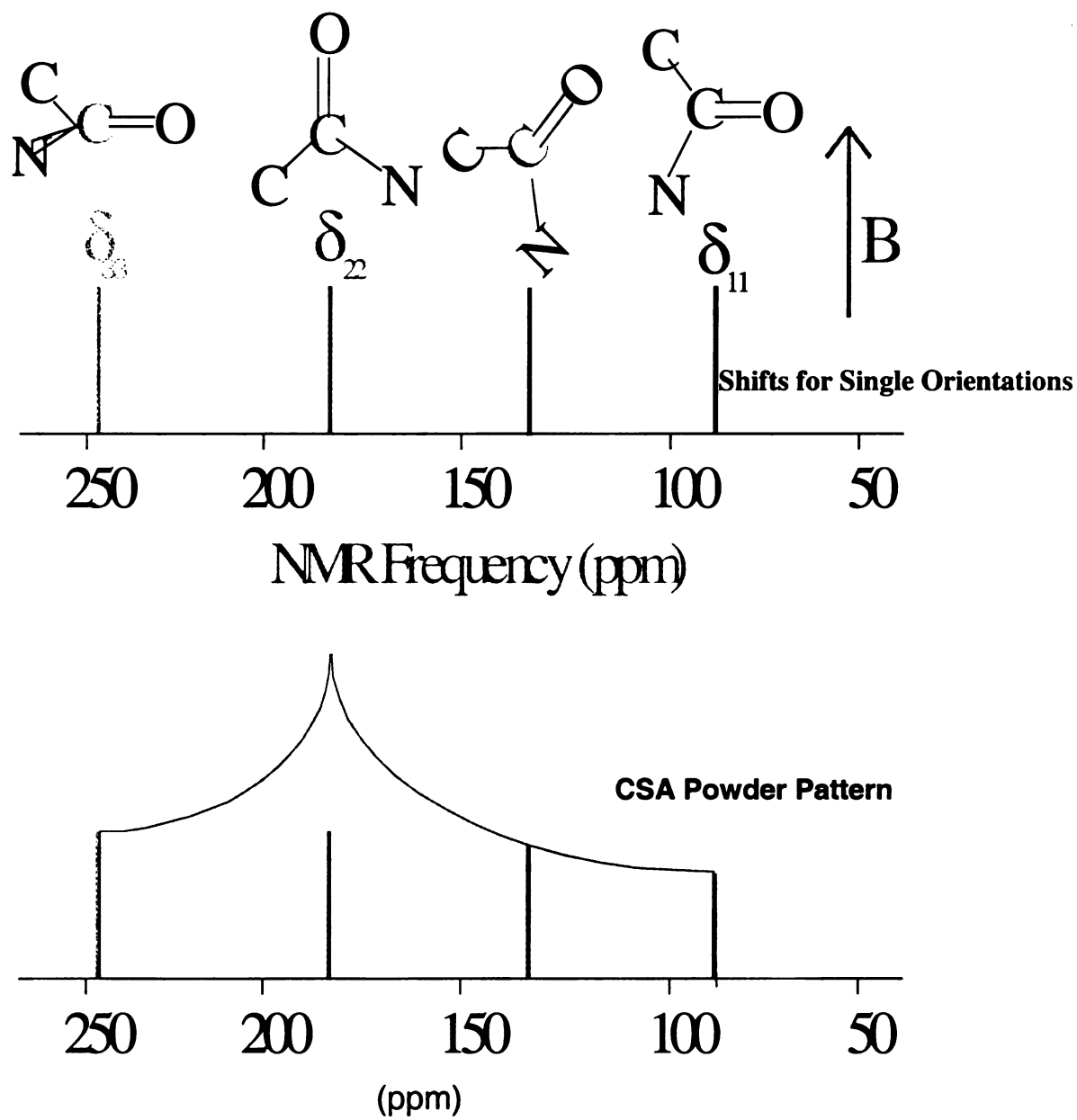
The chemical shift is given by the equation

$$\delta_A = \sigma_0 - \sigma_A$$

Where σ_0 is the shielding constant of a reference compound and δ_A and σ_A are the chemical shift and shielding constant of sample A.

In principle, the calculated chemical shifts are obtained as the nine elements of the chemical shielding tensor. The three principal values of the shielding tensor can be obtained by finding its eigen values.

Figure 4.2: Chemical Shift Anisotropy

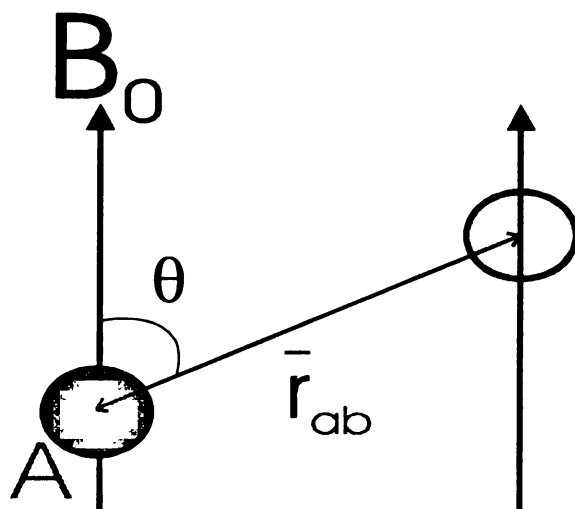


Magic Angle Spinning

For static protein solids, NMR signals are usually greatly broadened because of the absence of molecular tumbling. Typically, this broadening is greatly reduced by magic angle spinning (MAS). With MAS, signals at the isotropic chemical shift are observed and for low spinning speeds, additional signals are observed which are separated from the isotropic shifts by integral multiples of the spinning speed and are known as sidebands.

Dipolar coupling and chemical shift anisotropy are the two major factors that contribute to line broadening in solid state NMR. MAS averages out both CSA and dipolar coupling. As seen in figure 4.3, dipolar coupling is dependent on the term $(3\cos^2\theta - 1)$ where θ is the angle between the internuclear vector and external magnetic field, and the internuclear distance r . In liquids, due to the rapid isotropic molecular tumbling $\langle 3\cos^2\theta - 1 \rangle$ averages out to zero and this results in very narrow NMR lines. In solids, if the sample is rotated about an axis at an angle of 54.7° with respect to the external magnetic field then $\cos^2\theta = 1/3$ and the term $\langle 3\cos^2\theta - 1 \rangle$ averages out to zero and so dipolar broadening is eliminated, giving much higher resolution. This situation is known as magic angle spinning (MAS) and 54.7° is the magic angle.

Figure 4.3: Dipolar Coupling



Equations for Homonuclear and Heteronuclear Dipolar Coupling

Homonuclear Dipolar Coupling

$$\hat{H}_D^{ab} = -\frac{\mu_0 \hbar \gamma^2}{4\pi r_{ab}^3} \frac{1}{2} (3\cos^2 \theta - 1) (3\hat{I}_z^a \hat{I}_z^b - \hat{I}^a \cdot \hat{I}^b)$$

Heteronuclear Dipolar Coupling

$$\hat{H}_D^{ab} = -\frac{\mu_0 \hbar \gamma^a \gamma^b}{4\pi r_{ab}^3} \frac{1}{2} (3\cos^2 \theta - 1) 2\hat{I}_z^a \hat{I}_z^b$$

In Magic Angle Spinning:

$$\omega_{\text{NMR}(t)} = \omega_{\text{iso}} + \omega_{\text{aniso}(t)}$$

$\omega_{\text{NMR}(t)}$ = Angular frequency variant with time

ω_{iso} = Angular frequency with no time dependence

$\omega_{\text{aniso}(t)}$ = Angular frequency periodic in time with ω_r

(As the rotor is spinning, the angular frequency depends on the position of the rotor with respect to time.) As a result, the FID (fourier induction decay) will be periodic in time with angular frequency ω_r .

After a Fourier transform the spectrum will have peaks with spacing ν_r , where ν_r is the Larmor frequency such that:

$$\nu_{\text{iso}} + n\nu_r \quad n = \text{an integer dependent on the spinning speed}$$

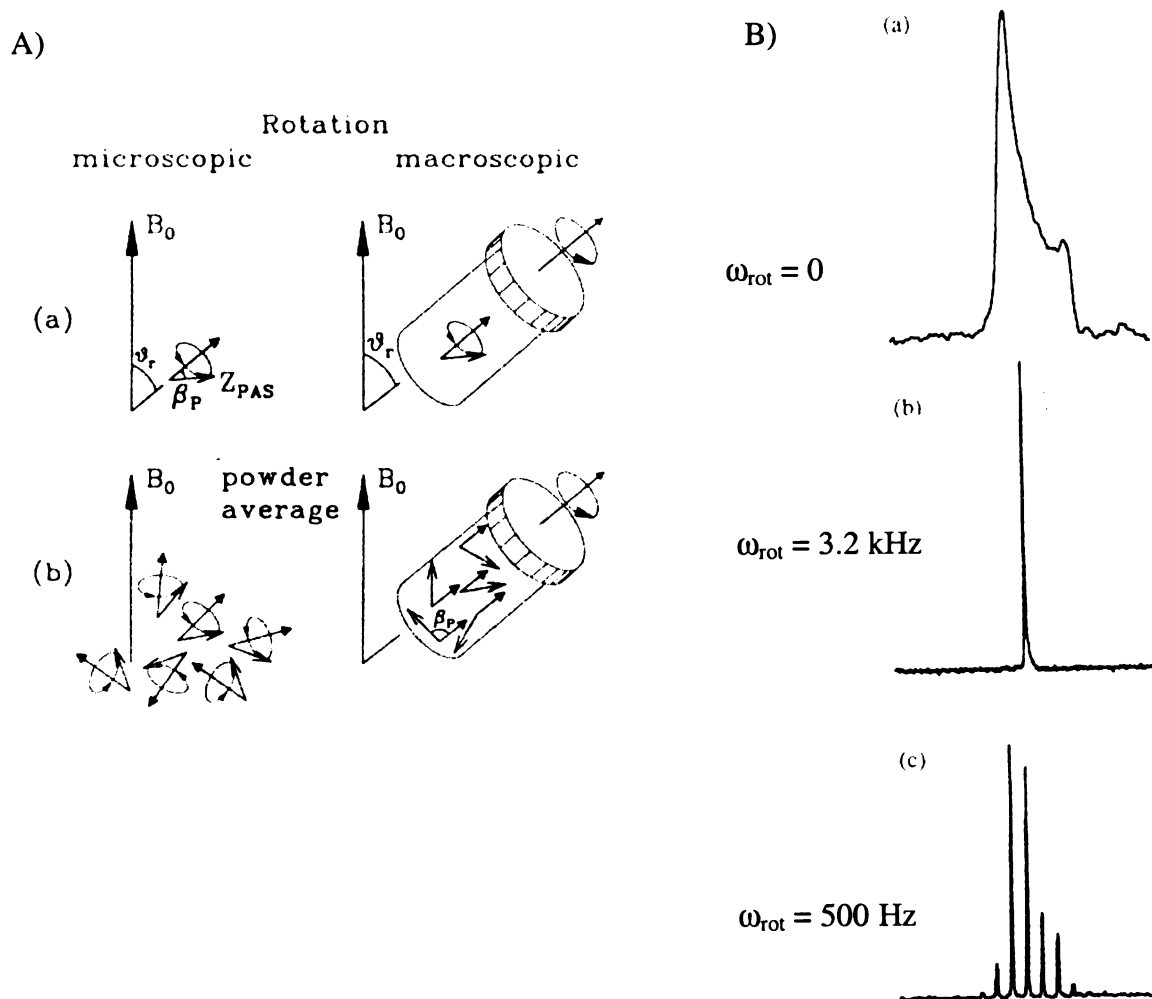
$\omega_{\text{aniso}(t)}$ depends on:

1. Principal values
2. Relative orientation of the principal axes to the spinning axes and external magnetic field direction.

When the spinning speed is zero, a powder spectrum (as shown in figure 4.4Ba) is obtained. The powder spectrum is sum of all the orientations. When the sample is rotated at a very high speed an isotropic (single) shift is obtained (as shown in figure 4.4Bb) and with intermediate spinning speed a chemical shift with sidebands is obtained as illustrated in figure 4.4Bc. The amplitudes of the spinning sidebands depend on orientation of the functional group relative to the spinning axis. The spinning axis is fixed relative to the direction of the external magnetic field. In magic angle spinning spectra, the exact principal values are determined from analysis of the relative intensities

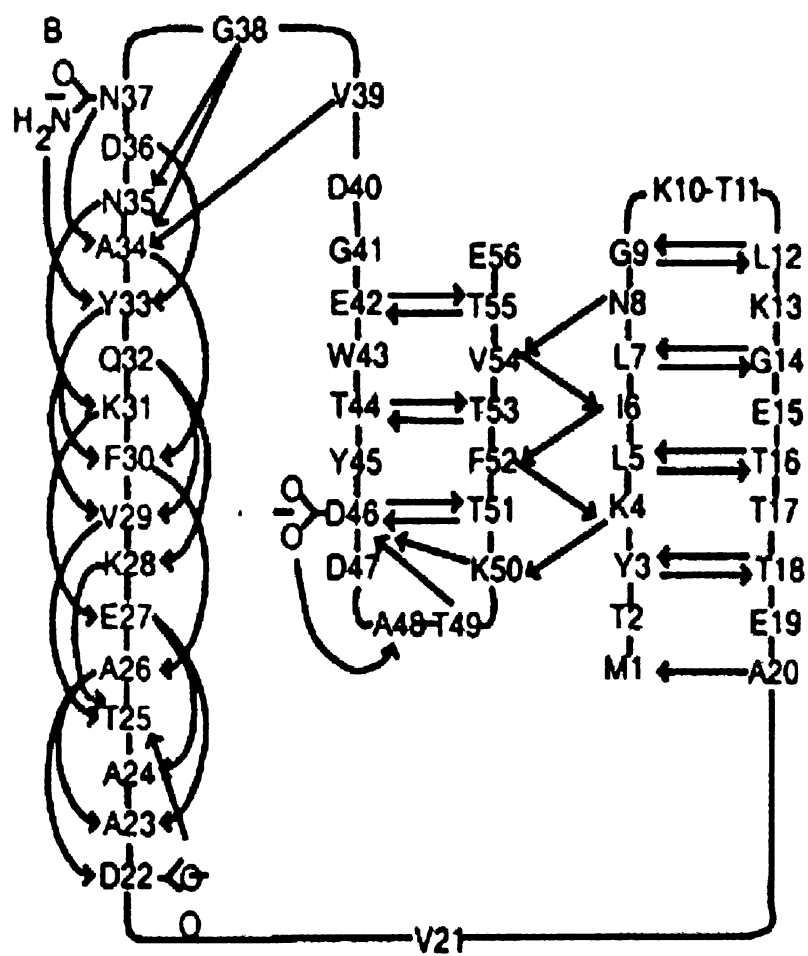
of the spinning sidebands as a function of spinning frequency. (Spinning sideband Herzfeld-Berger analysis). By spinning the sample at intermediate speed, chemical shift anisotropy is not completely eliminated and the intensities of the spinning sidebands can be used in calculating the CSA principal values.

Figure 4.4: Magic Angle Spinning^{21, 22}



We have targeted the five lysine residues in Protein G B1 Domain for measuring the carbonyl CSA principal values and correlating them to the secondary structure. Of the five lysine residues, two are in β sheet, two are α helical, and one is in a turn conformation.

Figure 4.5: Protein G B1 Domain Hydrogen Bonding Pattern²³



4.2 Materials and Methods

The protected amino acids and amino acid resins were obtained from Peptides International. The labeled amino acids were obtained from Cambridge Isotope Labs. Samples for study (Protein G B1 Domain) were synthesized on an automated Applied Biosystems 431A synthesizer. A standard synthesis program FastMoc was used. Solid phase methods were employed with 9-fluorenylmethoxycarbonyl (Fmoc) protection coupled with [2-(1H benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate] (HBTU) activation. We were successful in producing five different specifically labeled proteins, each ^{13}C -carbonyl labeled at a particular lysine residue. Each of the five syntheses only differed in the position of a $1\text{-}^{13}\text{C}$ labeled lysine (Lys-4, Lys-10, Lys-13, Lys-28 or Lys-31). Lys-28 and Lys-31 are in a α -helix, Lys-10 (H-bonded to the solvent and on a turn) and Lys-13 (H-bonded to the solvent) are in a β -sheet and Lys-4 is in a β -sheet (H-bonded to the peptide). All the five specifically labeled proteins were synthesized by the method described in chapter 2.

NMR Samples:

Frozen solution samples were prepared by dissolving the lyophilized peptide in a minimum volume of 20mM phosphate buffer at pH 7 with 0.03% azide and centrifuging to ensure total dissolution. The approximate sample concentration was then determined by measurement of the A_{280} . The solutions were quick-frozen by immersion in liquid nitrogen prior to NMR measurements. ^{13}C spectra were acquired on a 400-MHz Varian NMR spectrometer using 200 μL of 4mM protein solution at $-50\text{ }^{\circ}\text{C}$ and a spinning speed

of 2.5 kHz. A 90° pulse of 5.2 μs and decoupling power of 75 kHz was applied. A one second delay was used and the spectra represent ~12 hours of signal averaging.

4.3 Results and Discussion

Table 4.1: Concentration and Linewidth of the Unlabeled and Labeled Protein G Samples

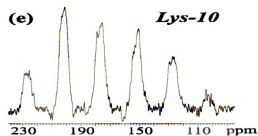
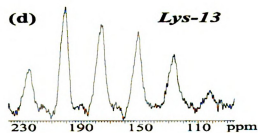
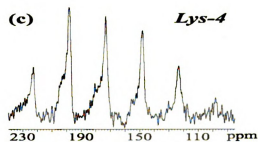
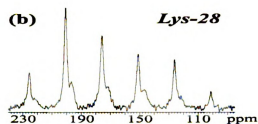
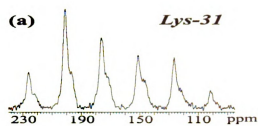
Sample	Concentration (mM)	Linewidth (Hz)
Unlabeled Prot G	6.22	824
Prot G Lys-31	3.79	372
Prot G Lys-28	4.17	232
Prot G Lys-13	4.65	584
Prot G Lys-10	3.21	653
Prot G Lys-4	2.13	349

Table 4.2: Measured Chemical Shift Anisotropy Principal Values for the ^{13}C Labeled Protein G Samples

Sample	CSA Principal Values (PPM)			Local Secondary Structure
	δ_{11}	δ_{22}	δ_{33}	
PG Lys-31	240	196	93	Alpha-Helix
PG Lys-28	239	195	93	Alpha-Helix
PG Lys-4	246	181	93	Beta-Sheet

Note: The uncertainty in the measured CSA principal values is $\sim \pm 2$ ppm

Figure 4.6: NMR Spectra of the Five-Labeled Protein G (Frozen Solution) Samples



The spectra were processed with 25-Hz line broadening. The carbonyl region contains sharp downfield signals from the label as well as a broad natural abundance background, which is about 60% of the labeled signal. The relative intensities of the sharp peaks were used to calculate the CSA principal values. The sharp linewidths are typical for what has been previously observed in frozen solutions of well-structured proteins^{17, 24} and significantly narrower than the ~ 6 ppm linewidths of a denatured protein.²⁴

As observed from Table 4.2, the CSA principal values obtained for Lys-31 and Lys-28 are very similar. Also a clear difference in the principal values is observed between Lys-31 and Lys-28 (residues in a helix) and Lys-4, which is in a β -sheet conformation. The similarities between the CSA principal values for the two helical sites and the difference in the CSA principal values for the residues in a helix and β -sheet are very encouraging. These results provide evidence that a correlation exists between the backbone carbonyl CSA principal values and local protein secondary structure.

The CSA principal values for Lys-10 and Lys-13 are not included in Table 4.2. As seen in Figure 4.6 and Table 4.1, the spectra for Lys-10 and Lys-13 have very broad lines. The broad linewidths may be due to the structural inhomogeneity, which leads to a wide distribution of chemical shifts. Therefore, in such cases, calculating CSA principal values may not be useful as one predominant structure does not exist.

The occurrence of broad linewidths may be due to partial protein unfolding during the freezing process. Better-folded protein would likely be formed in crystalline rather than frozen solution samples. However, samples containing large single crystals are difficult to prepare. An efficient alternative for this may be to study the protein as an

ammonium sulfate precipitate. Ammonium sulfate precipitate sample is in a “microcrystalline” form. The main advantage of this approach is that sample preparation of small crystals is much simpler and straightforward as compared to large single crystals. One of the problems we may encounter is that the samples may be conductive which can interfere with tuning the samples to the right resonance frequency. In order to overcome this problem, we may need to modify the probe and also optimize sample preparation.

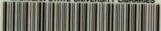
In addition to the CSA principal values calculated for the five labeled lysine samples, we need to obtain more data to confirm whether a correlation really exists between the CSA principal values and secondary structure. Therefore, future studies will include labeling the eleven threonine residues in Protein G to obtain their carbonyl CSA principal values and correlate them to the secondary structure. The $^{13}\text{C}=\text{O}$ labeled, Fmoc protected threonine is available and also the eleven threonines provide a variety of local secondary structure. For the eleven threonines, eight are in the β -sheet conformation, one is α -helical, and two are in turns.

REFERENCES

1. Ando, I., Kameda, T., Asakawa, N., Kuroki, S., and Kurosu, H. *Journal of Molecular Structure* **1998**, 441, 213-230.
2. Spera, S., and Bax, A. *Journal of the American Chemical Society* **1991**, 113, 5490-5492.
3. Wishart, D. S., Skyes, B. D., and Richards, F. M. *Journal of Molecular Biology* **1991**, 222, 311-333.
4. Beger, R. D., and Bolton, P. H. *Journal of Biomolecular NMR* **1997**, 10, 129-142.
5. Oldfield, E. *Journal of Biomolecular NMR* **1995**, 217-225.
6. DeDios, A.C. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1996**, 29, 229-278.
7. Case, D. A. *Curr. Opin. Struct. Biol.* **1998**, 8, 624-630.
8. Asakawa, N., Kuroki, S., and Kurosu, H., Ando, I., Shoji, A., and Ozaki, T. *Journal of the American Chemical Society* **1992**, 114, 3261-3265.
9. Kameda, T., Takeda, N., Kuroki, S., and Kurosu, H., Ando, I., Shoji, A., and Ozaki, T. *Journal of Molecular Structure* **1996**, 384, 17-23.
10. Gu, Z. T., Zambrano, R., and McDermott, A. *Journal of the American Chemical Society* **1994**, 116, 6368-6372.
11. Tjandra, N., and Bax, A. *Journal of the American Chemical Society* **1997**, 119, 8076-8082.
12. Tjandra, N., and Bax, A. *Journal of the American Chemical Society* **1997**, 119, 9576-9577.
13. Heller, J., Laws, D. D., Tomaselli, M., King, D. S., Wemmer, D. E., Pines, A., Havlin, R. H., and Odfield, E. Tjandra, N., and Bax, A. *Journal of the American Chemical Society* **1997**, 119, 7827-7831.
14. Havlin, R.S., Le, H. B., Laws, D. D., DeDios, A.C., Odfield, E. *Journal of the American Chemical Society* **1997**, 119, 11951-11958.

15. Tycko, R., Weliky, D.P., and Berger, A.E. *Journal of Chemical Physics* **1996**, 105, 7915-7930
16. Gullion, T., and Schaefer, J. *Journal of Magnetic Resonance* **1989**, 81, 196-200.
17. Weliky, D. P., Benett, A. E., Zvi, A., Anglister, J., Steinbach, P. J., and Tycko, R. *Nat. Struct. Biol.* **1999**, 6, 141-145.
18. Pines, A., G. M. C., and Waugh, J. S. *Journal of Chemical Physics* **1973**, 59, 569-590.
19. Herzfeld, J., and Beger, A. E. *Journal of Chemical Physics* **1980**, **73**, 6021-6030.
20. Harris, R. Nuclear Magnetic Resonance Spectroscopy, *Publisher John Wiley & Sons, Inc.* **1986**
21. Schmidt-Rohr & Spiess, Multidimensional solid-state NMR and polymers. *Publisher Academic Press.* **1994**
22. Sanders and Hunter, Modern NMR Spectroscopy: A guide for Chemists. 2nd edition, *Publisher Oxford University Press.* **1993**
23. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield P. T., and Clore, G. M. *Science* **1991**, 253, 657-661.
24. Long, H. W., and Tycko, R. *Journal of the American Chemical Society* **1998**, 120, 7039-7048.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02112 3215