1-1-1994

Nanoscale scaffolding by folding of monodisperse and sequentially precise poly[(AG)3EG(GA)3EG] :: biosynthesis and characterization by x-ray diffraction, FTIR and NMR/

Ajay Dattatraya Parkhe
University of Massachusetts Amherst

NANOSCALE SCAFFOLDING BY FOLDING OF MONODISPERSE AND SEQUENTIALLY PRECISE POLY{$(AG)_3EG(GA)_3EG$}: BIOSYNTHESIS AND CHARACTERIZATION BY X-RAY DIFFRACTION, FTIR AND NMR

A Dissertation Presented by

AJAY DATTATRAYA PARKHE

Submitted to the Graduate School of the University of Massachusetts
Amherst in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

September 1994

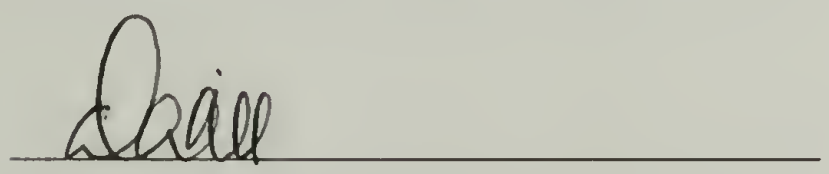
Polymer Science and Engineering

NANOSCALE SCAFFOLDING BY FOLDING OF MONODISPERSE AND SEQUENTIALLY PRECISE POLY{(AG)$_3$EG(GA)$_3$EG}: BIOSYNTHESIS AND CHARACTERIZATION BY X-RAY DIFFRACTION, FTIR AND NMR

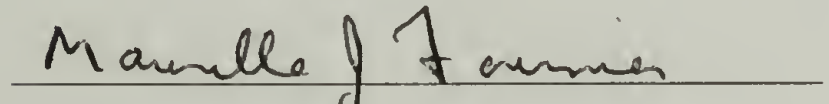
A Dissertation Presented by

AJAY DATTATRAYA PARKHE

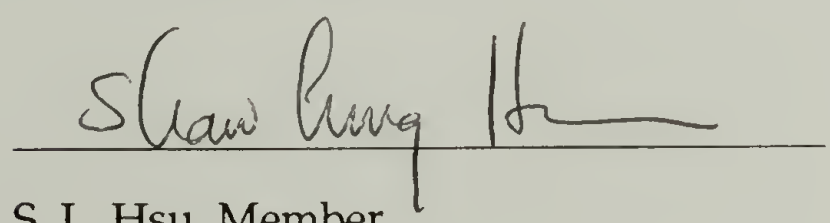
Approved as to the style and content by:



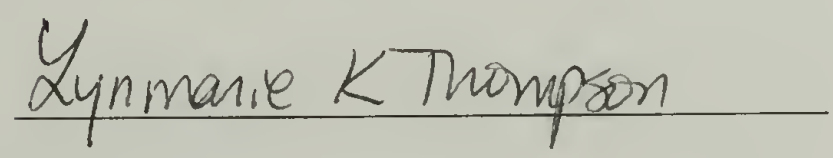
D. A. Tirrell, Committee Chairperson



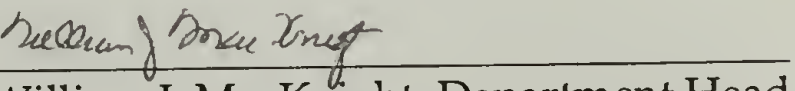
M. J. Fournier, Member



S. L. Hsu, Member



L. K. Thompson, Member



William J. MacKnight, Department Head
Polymer Science and Engineering

ACKNOWLEDGMENTS

I would like to thank Professor Tirrell for giving me the opportunity to work on a very exciting problem and for all the valuable advice. However, an even more valuable experience, for which I will always be indebted, was the professional association with him; it inculcated an appreciation of perseverance, discipline, precision, ethics and modesty- virtues which transcend the realm of scientific education.

I would like to thank Professor Fournier for all the valuable advice on biosynthesis and for the opportunity to work in his laboratory; this gave me the opportunity to interact with biochemists and widen my horizons.

I would like to thank Yasumasa Fukushima and Mike Dougherty for the initial help in biosynthesis. I would like to thank Professor Atkins and Alan Waddon for valuable discussions in x-ray diffraction. I would like to thank Professor Thompson and Jianxin Wang for a fruitful collaboration in NMR, Professor Hsu for the discussions in FTIR and Professor Mason for the discussions in biochemistry.

I would like to thank my colleagues Jim Thomas, Vince Conticello, Susan Dawson, Geni Dessipri, Mark Krejchi, Laurie Gower, Wendy Petka, Seenu Kothakota and Andrey Balakin for innumerable discussions, both scientific and otherwise.

ABSTRACT

NANOSCALE SCAFFOLDING BY FOLDING OF MONODISPERSE AND SEQUENTIALLY PRECISE POLY{(AG)3EG(GA)3EG}: BIOSYNTHESIS AND CHARACTERIZATION BY X-RAY DIFFRACTION, FTIR AND NMR

SEPTEMBER 1994

AJAY DATTATRAYA PARKHE, B. TECH. (CHEMICAL ENGINEERING), INDIAN INSTITUTE OF TECHNOLOGY, KANPUR, INDIA

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor David A. Tirrell

Synthesis of three dimensional structures which are precise on the length scale of nanometers could be of great scientific and technological importance; no existing polymeric materials can self-assemble into precise nanoscale scaffolds. We have designed repetitive polypeptides which would self assemble into a unique three dimensional structure. The *de novo* design of these polypeptides is based on existing information on protein chain folding. It is anticipated that this will lead to a novel class of ordered polymeric materials with order on the length scale of nanometers.

As a first step to designing more complicated scaffolds, we are interested in designing repetitive polypeptides which would self assemble into lamellae of uniform, predetermined thickness. The design of the polypeptide repeating unit has been based in part on work on the structure of silk (and its analogues) and in part on the literature on reverse turns in globular proteins. The polypeptides in this work comprise alternating beta

sheet forming segments and turn forming segments with exact periodicity. Repetitive alanyl-glycyl diads are known to form a beta sheet structure. It is anticipated that these would reside in the lamella and define its thickness. The stronger the thermodynamic driving force for the chains to form alternating beta sheets and turns, the more is the probability of obtaining a unique scaffold.

Broadly, there are two aspects to the problem; the first is to design and synthesize polymers which can form nanoscale scaffolds; the second is to prove that when these polymers crystallize, they do indeed form the desired scaffolds.

In this thesis, DNA sequences encoding the synthesis of repetitive units of **A**, **B**, **C** and **D** have been synthesized. Polypeptides with ten repeats of **A**, four repeats of **D** and three repeats of **C** were successfully synthesized in *E. coli*.

-(AlaGly)$_3$GluGly(GlyAla)$_3$GluGly-	**A**
-(AlaGly)$_3$AspGly(GlyAla)$_3$AspGly-	**B**
-(AlaGly)$_3$GluGly(AlaGly)$_3$ValGly-	**C**
-(AlaGly)$_3$GluGly(AlaGly)$_3$MetGly-	**D**

The polypeptide containing ten repeats of **A** (referred to as **A-10** in the Abstract) has been characterized in the solid state by x-ray diffraction, FTIR and NMR.

The polypeptide **A-10** forms an antiparallel beta sheet structure when stirred in 70% formic acid; electron microscopy shows the morphology of the crystallites to be needle like. The unit cell deduced from the diffraction patterns is similar to those from previous diffraction patterns on silks and silk-like polypeptides. The unit cell which best explains the diffraction data for **A-10** is monoclinic with a=9.56 Å, b=10.68 Å and c=7.0 Å with the angle between b and c equal to 80°. Qualitatively, the intensities of the reflections suggest the presence of apolar sheets.

FTIR spectroscopy on crystalline **A-10** shows the polypeptide predominantly adopting the antiparallel beta sheet structure. The repetitive polypeptide **A-10** and a ^{13}C=O enriched analogue have been synthesized biologically in *Escherichia coli*. The two analogues have been blended in solution and co-crystallized. FTIR spectra (in the amide I region) have been studied as a function of blend composition. The results indicate the folding of the chains to be predominantly in an adjacent reentry fashion.

Solid state CPMAS NMR experiments were carried out on the amorphous and crystalline forms of **A-10**. The data suggests that glutamic acids are in identical environments in the crystalline and amorphous samples. The fast relaxation rates of their alpha, beta and gamma carbons suggest that glutamic acid residues are excluded from the crystalline regions in both the samples. Hence, the data supports the presence of glutamic acid residues at the turn positions.

Static solid state NMR experiments have been carried out on oriented mats of crystalline **A-10**. The orientation of the crystallites has been studied by the analysis of lineshapes. The lineshape analysis is consistent with the crystallites being an imperfect form of an orientation in which the crystallographic *b* axis (Figure 1.6) is perpendicular to the mat and there is free rotation around *b*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The phenomenon of polymer chain folding can lead to the formation of nanoscale scaffolds. Precisely engineered nanoscale scaffolds cannot be synthesized using the current polymeric materials and processing techniques due to their inherent heterogeneity. An attempt to synthesize precise scaffolds by polymeric chain folding is therefore a research topic which could have far reaching consequences in basic and applied research.

In the first chapter, after defining nanoscale scaffolds, the reasons for pursuing this dissertation research have been presented in general terms. The importance of the research has been illustrated with the help of a few examples in which nanoscale scaffolding is used. This general discussion is then followed by specifics of polymer design and structural expectations in order to generate the scaffolds.

The second chapter describes polymer synthesis and purification.

The polymers described in this dissertation have been designed to fold into a particular scaffold. Chapters 3,4 and 5 probe for the structure of the crystallized polymers. Chapter 3 deals with determination of unit cells from x-ray diffraction, Chapter 4 with the problem of chain folding and Chapter 5 with the determination of the location of the turns.

1.1 Nanoscale Scaffolds

The most important characteristic property of any polymer chain in comparison to other materials like metals or ceramics is the connectivity of the monomer subunits. In polymers, adjacent monomer subunits are held together by strong chemical bonds (with bonding energies on the order of 100 kcal/mol[1]) while the secondary structure is held together by weak secondary bonds like hydrogen and Van der Waals bonds (with bonding energies on the order of 5 kcal/mol[1]). This asymmetry of bonding which causes bonding along the chain direction to be stronger than along other directions endows polymers with attributes which are not present in metals or ceramics. This unique feature of polymers allows one to string together a predetermined series of monomers along a copolymer chain; the more the diversity in the monomers, more is the variation along a chain. The importance of this feature will be apparent in future discussions.

A nanoscale scaffold is defined as a skeleton in three dimensional space on the length scale of nanometers; obviously, the requirement for its synthesis is to create structural features on the required length scale.

1.1.1 Generation of Nanoscale Scaffolds

The goal of this study is to generate unique nanoscale structures which would serve as precise molecular scaffolds on to which chromophores, catalytic groups or recognition moieties could be grafted. Since the scaffolds would be unique, the positions of the reactive handles, if

any, and hence the grafted moieties would be placed precisely in three dimensional space.

Features on nanometer length scales can be formed by phase separation or by controlled chain folding of polymers; the latter leads to more precise scaffolds which could potentially be unique if the design of the polymer is optimal.

1.1.1.1 Nanoscale Phase Separation

Nanophase separation of ceramics or metals, or their synthesis from atomic clusters can give rise to nanophase materials[2]. Polymer blends also show similar phase separation during spinodal decomposition[3]; although binodal phase separation can form phase separated systems, the length scales are larger due to the nucleation and growth phenomenon[3]. Polyurethanes with hard and soft segments are also known to form small phase separated domains[3]. Interpenetrating networks which can be formed by simultaneous cross-linking of two or more miscible monomers by using two or more different initiation pathways can also lead to nanometer size phases[4]. Block copolymers of styrene and butadiene also show microphase separation[5]; the phase separated morphology is due to the immisciblity of polystyrene with polybutadiene and the small domain sizes are due to the fact that the two blocks are covalently connected.

1.1.1.2 Scaffolding by Folding of Polypeptides

Generation of scaffolds by folding of polymers into unique structures is prevalant in most polypeptides in living organisms. As will be discussed later, the process of crystallization of a precise polymer into a scaffold is not as dependent on the statistics of diffusion or kinetics of reaction as the structures generated by phase separation or crosslinking are. As a consequence of this, the structures generated by controlled crystallization of well designed polymers could be unique; this would require the polymer to fold into a unique conformation with all other conformations being energetically less favorable.

1.1.2 Importance of Nanoscale Scaffolds

1.1.2.1 Photosynthesis

Photosynthesis is responsible for the assimilation of ca. 200 billion tons of carbon every year; this amount is equivalent to ca. 0.1% of the radiative energy transfer by the sun to the earth[6]. The photosynthetic system comprises a light reaction and a dark reaction. The light reaction is responsible for converting light energy to ATP and NADPH while the dark reaction is responsible for the utilization of these energetic molecules and production of reduced carbon compounds from carbon dioxide and water. The light reaction is carried out in photosynthetic reaction centers, two types of which are present in green plants, called PS-I and PS-II. PS-I is activated by light between 680-700 nm wavelength and PS-II by light of a

wavelength of 650 nm. Some photosynthetic bacteria contain only one type of reaction center, the PS-II. A very simplified schematic of the photosynthetic system responsible for the light reactions is shown in Figure 1.1. Antenna chlorophyll molecules harness the energy from sunlight and cause the active chlorophyll molecule (generally referred to as P_{650}), shown in Figure 1.1, to be excited. The active P_{650} gets oxidized to P^{+}_{650} and an electron is transferred to a quinone through the pathway shown in Figure 1.2 (refer to the next section). The oxidant P^{+}_{650} oxidizes water and returns to the uncharged state. The active chlorophyll molecule in PS-I is referred to as P_{700}. When this molecule is excited, it releases an electron which eventally is transferred, through a few intermediates, to a thylakoid membrane bound ferredoxin. The oxidized P^{+}_{700} accepts an electron through an electron transport chain linking PS-I and PS-II which is shown in Figure 1.1 and returns to the uncharged state. The pH gradient generated due to the oxidation of water is used to generate ATP from ADP and inorganic phosphate. A more detailed explanation of the electron transfer pathways and the energetics during photosynthesis can be found elsewhere[6].

1.1.2.1.1 Nanoscale Ordering in PS II

The importance of nanoscale order in three dimensions is evident in photosynthetic reaction centers of purple membrane of *Halobacterium halobium*[7], photosynthetic membranes of *Rhodopsedomonas viridis*[8] and thylakoid membranes of higher plants. The best studied part of the photosynthetic machinery is the reaction center PS-II.

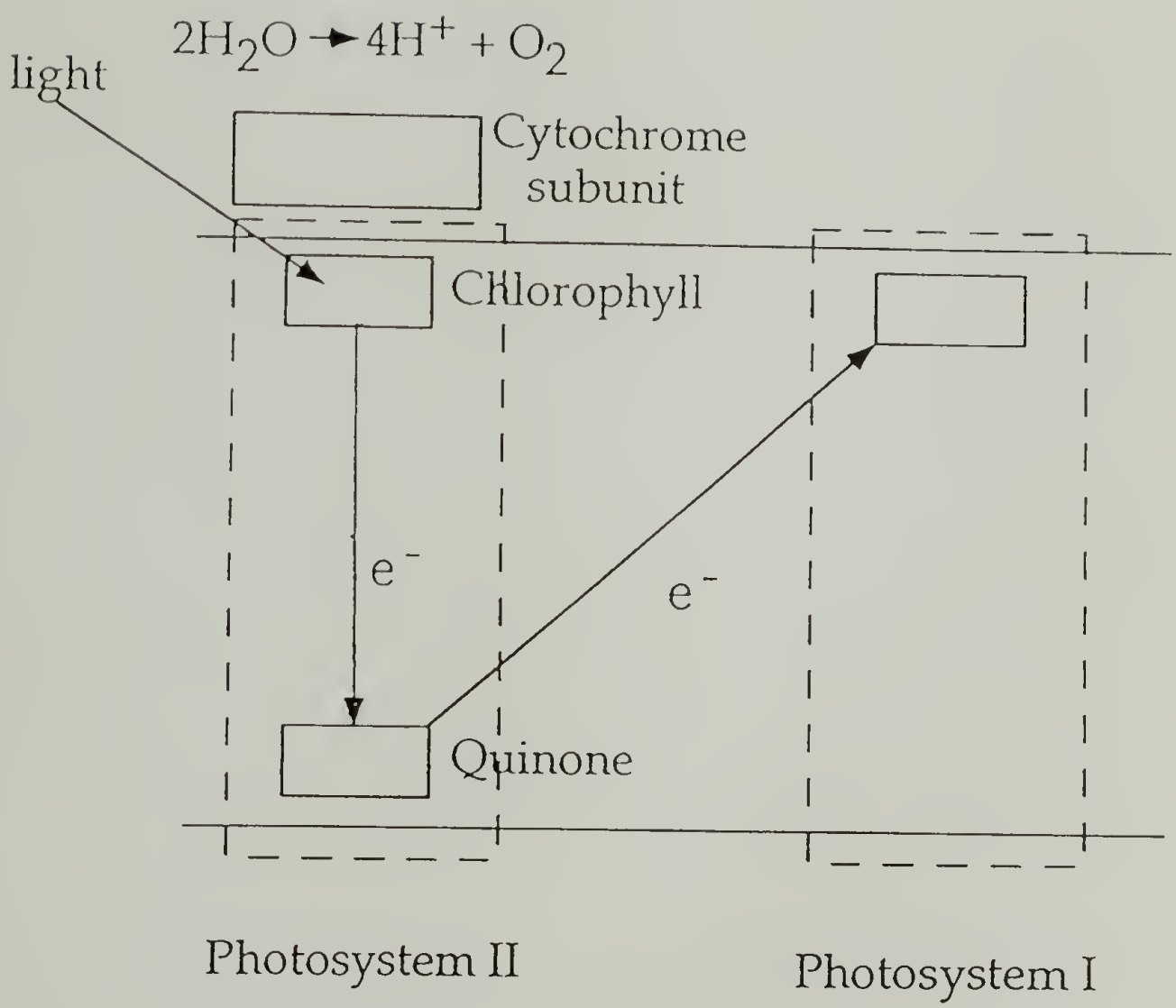


Figure 1.1 Schematic of the light driven photosynthetic center.

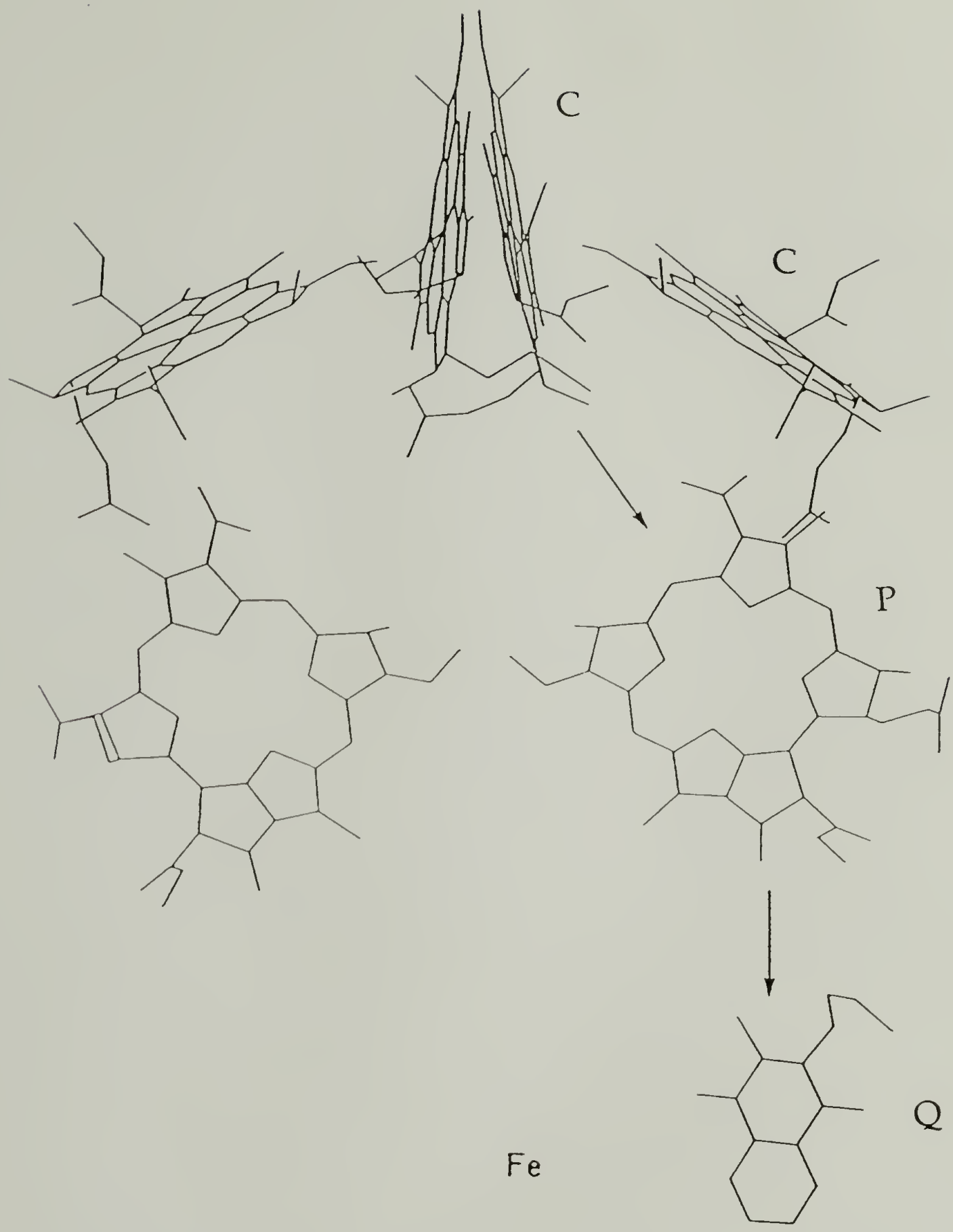


Figure 1.2 Relative placement of groups involved in the electron transfer during photosynthesis[8]. C=Chlorophyll, P=Pheophytin, Q=Quinone.

The chromophores (bacteriochlorophylls and bacteriopheophytins) and other cofactors involved in electron transfer (quinones, carotenoid and a non-heme iron) are precisely organized with respect to distance and angle. This precise placement, as evidenced by the crystal structure (determined by x-ray diffraction) of the reaction center of *Rhodopseudomonas viridis*[9], is due to three dimensional protein scaffolding. The relative placement of the chromophores in space is shown in Figure 1.2[9]. Figure 1.2 shows the two arms of the PS-II reaction center. The four topmost molecules in the figure are chlorophyll molecules and the two pheophytins are placed below them. The path the electron is believed to take when excited P_{650} is oxidized, is shown in Figure 1.2[10]; the quinone serves as the final acceptor of the electron. A more elaborate description of the PS-II machinery can be found elsewhere[10]. The chromophores, which are derivatives of tetrapyrrole (Figure 1.3) are located in a trans-membrane, elliptical complex with the two axes being 70 Å and 30 Å; the total length of the subunit is ca. 130 Å. Within the complex, the four bacteriochlorophylls, two bacteriopheophytins and the other cofactors are placed in exact positions by four protein subunits. The photosynthetic system serves as a paradigm for devices to effect photon-induced charge separation[11]. Bacteriorhodopsin has been used to make photo detectors, spatial light modulators and optical digital memories[7].

The current state of materials science does not allow us to place chromophores at locations as precisely as in the photosynthetic reaction center. Controlling the placement of chromophores in space with relative distances on length scales of nanometers could be significant due to a few

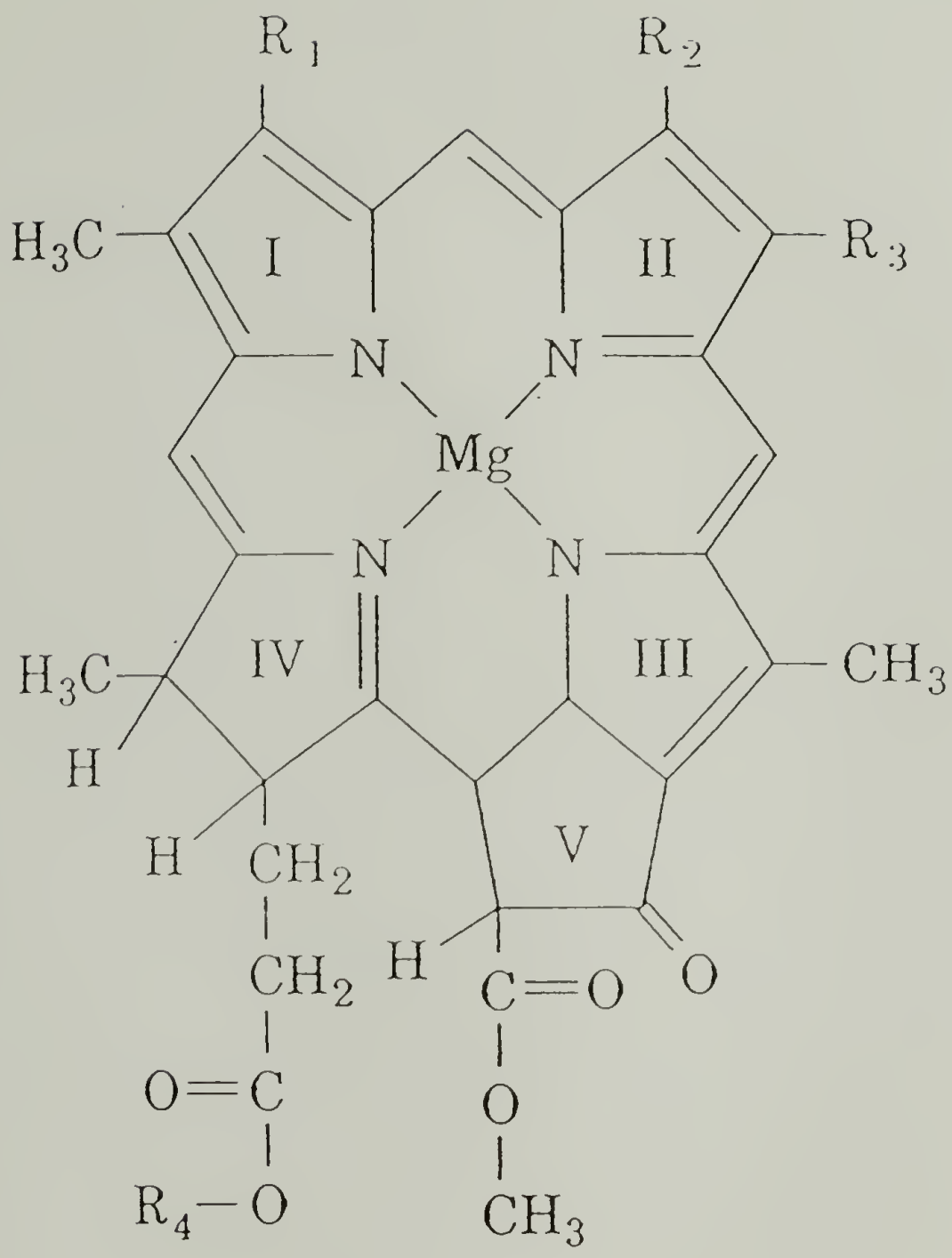


Figure 1.3 Structure of chlorophyll; substitution of the R groups on the tetrapyrrol ring yield different types of chlorophyll.

reasons. From a practical standpoint, this could help in the synthesis of artificial centers to harvest light energy into electrical energy. From a theoretical standpoint, control over the placement of chromophores could help understand the photosynthetic reaction center better[10]; this would involve systematically altering the locations of the cofactors, their dielectric environment etc. and monitoring the corresponding changes in properties of the reaction center.

1.1.2.1.2 Mimicking the Photosynthetic Reaction Center

Mimicking this very efficient system of conversion of light energy into charge separation and chemical energy could be useful in the fabrication of solar cells and photodetectors. One strategy to mimic the reaction center and make well designed modifications to the cofactors involved in the reaction center is by using nanoscale scaffolds as shown in Figure 1.4. The copolymer comprises of a domain which can attach to a surface and a domain which can fold precisely into the desired structrue; the copolymer features reactive handles at precise locations along the chain which are present at precise locations in space on the scaffold. The derivatized forms of the chromophores are then reacted on to these reactive handles on the scaffolds. Current polymers do not afford such options because of their inability to form precise scaffolds; this inability arises because of the inability to synthesize precise copolymers with sufficient diversities in the monomers.

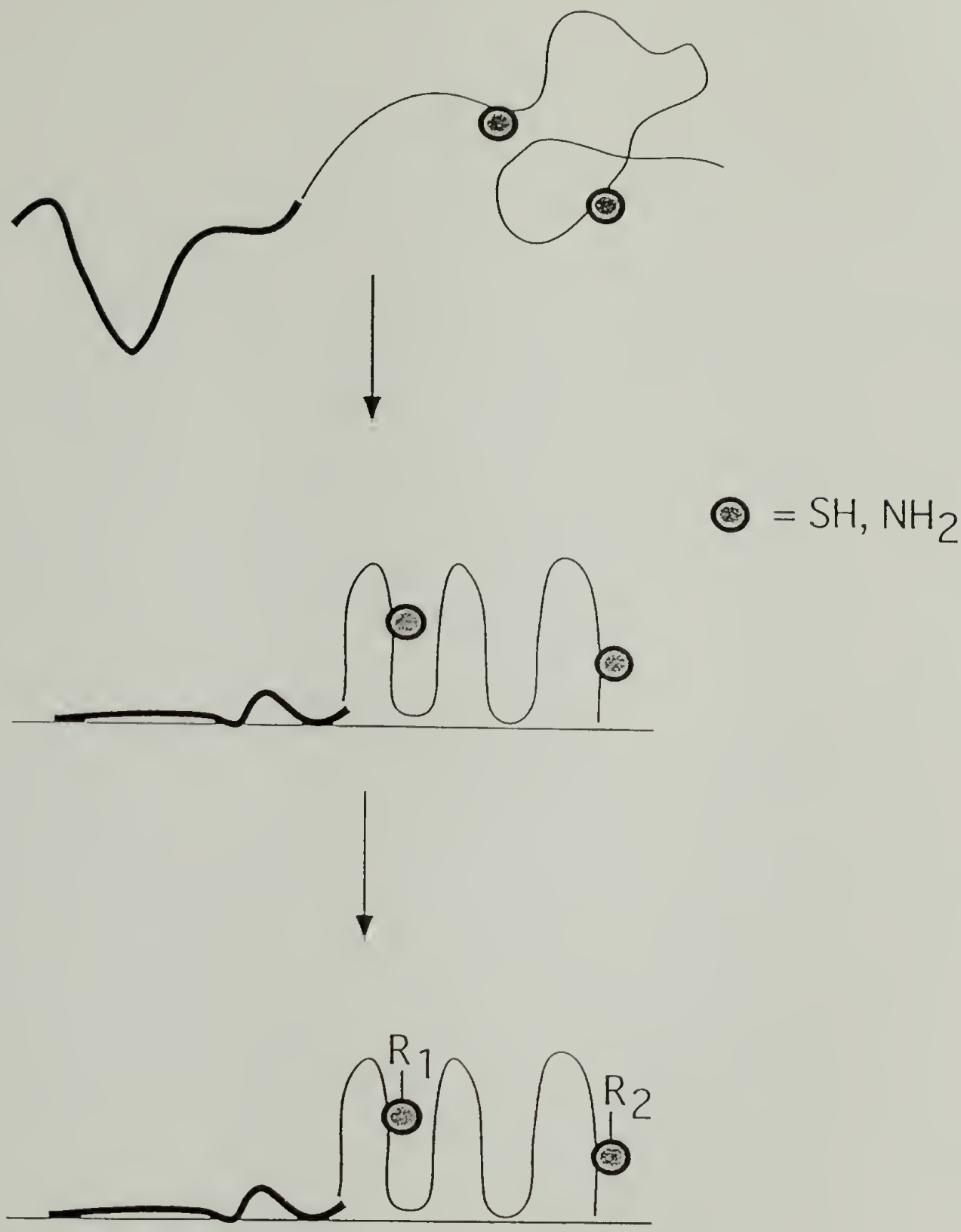


Figure 1.4 Use of precise nanoscale scaffolding to precisely place reactive handles in space; appropriately derivatized chromophores can then be grafted on to these scaffolds.

1.1.2.2 Catalysis on Solid Supports

The synthesis of precise nanoscale scaffolds also holds promise in the synthesis of precise catalytic cavities. Polymer supported catalysts have been used in the form of micro- and macroporous beads. These heterogeneous catalytic systems afford advantages like ease of product isolation, control of the catalytic microenvironment and reusability of catalyst[12]. Typically, the required functional group is introduced on to a support in two ways. 1) Incorporation of catalyst during polymerization of the support: One can functionalize the polymer during the synthesis of the support. When a vinyl polymer is used as a polymer support, imidazole, pyridine or morpholine can be incorporated by copolymerization of the vinyl monomers with the vinyl-derivatized catalytic moieties; many groups have been grafted on styrene by electrophilic substitution for this purpose. Polystyrene-supported catalysts have been synthesized in suspension to yield catalytic porous beads containing pyridine and phosphines[13]. 2) Chemical grafting of catalyst on support after its synthesis: Chemical modification of the support after its synthesis is the preferred method of immobilizing a catalyst. Either linear polymers or cross-linked polymers can be modified to incorporate the catalysts in this way.

1.1.2.2.1 Synthetic Polymer Supported Catalysts

Although the kinetics of the reactions which are facilitated by the matrix supported heterogenous catalysts are slower in many cases than their homogeneous counterparts, they do have some advantages over the homogeneous catalytic systems in terms of specificity. Since the polymer support provides a three dimensional scaffold, the microenvironment within the catalytic cavity can be well tailored; this could lead to a reaction specificity similar to those observed in naturally occurring enzymes. Polymer supported catalysts which have relatively high specificity have been synthesized to carry out geometrically controlled organic reactions[14-17]. Most of these approaches involve the synthesis of scaffolds by chemical cross-linking of catalytic groups within the cavities of cyclodextrins, crown ethers, cyclophanes or similar ring systems.

It is apparent from the literature that geometrically selective organic catalysis is in its infancy. The ability to recognize and react specific molecular geometries is hampered by the inability of polymeric materials to form precise three dimensional scaffolds.

1.1.2.2.2 Catalysis in Natural Enzymes

In contrast, natural enzymes are very precise in their catalytic function.

Chymotrypsin, a serine protease, is a 241 amino acid long polypeptide which is globular in the native state; the hydrophilic residues are on the surface and the hydrophobic residues form an interior core of the ellipsoidal molecule. The tertiary structure is held in place by several disulphide bridges. The catalysis takes place in an active site in which three amino acids (serine, histidine and aspartic acid) are directly involved in the catalysis[18]. The placement of the three residues as shown in Figure 1.5 is controlled by the scaffolding provided by the protein.

In general, the amino acids involved in catalytic function are serine, cystein, glutamic acid, aspartic acid, lysine, histidine and tyrosine. In an attempt to imitate enzyme-like esterolysis, vinyl polymers containing imidazoles were first synthesized by Overberger[19]; subsequently, a large number of variants have been synthesized and the effects of different microenvironments in the catalytic cavity have been studied[12]. Although monofunctional catalysis has been carried out on polymer substrates with some success, multifunctional catalysis like that in enzymes which involves cooperative interactions of multiple catalytic residues has met with limited success because the current polymer technology cannot position the catalytic residues precisely with respect to each other.

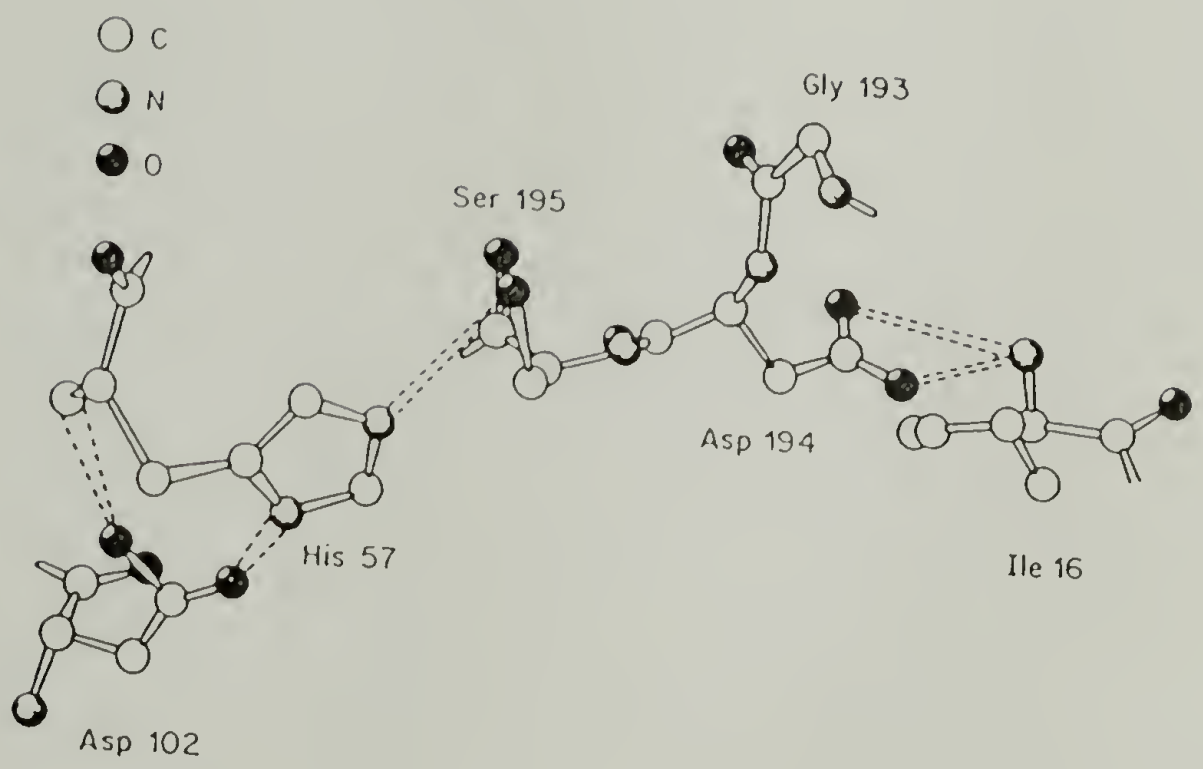


Figure 1.5 Multifunctional catalysis in chymotrypsin which requires the placement of three amino acid residues in precise locations with respect to each other[18].

1.1.3 Technological Impact of Nanoscale Scaffolding

The placement of chromophores or catalytic groups at precise relative locations in space on the length scale of nanometers has great technological and scientific importance. In order to control the placement of these prosthetic groups it is essential to form precise three dimensional scaffolds; an elegant strategy to synthesize such scaffolds in the solid state would be by formation of stable and unique tertiary structures due to polymer chain folding. It is anticipated that this will lead to a novel class of ordered polymeric materials with order on the length scale of nanometers.

1.2 Goals of the Work

1.2.1 Three Dimensional Structural Control

The synthesis of three dimensional scaffolds as shown in Figure 1.6 requires polymers which can form unique structures when they crystallize; in particular, one needs to synthesize copolymers comprising a stretch of monomers with a tendency to form an extended structure alternating with a stretch of monomers with a tendency to form a turn.

1.2.2 Requirements on Synthetic Strategy

The requirement for the formation of a unique three dimensional structure, places strict restrictions on the purity of the polymers which can be used; the polymers are required to possess precise sequence and

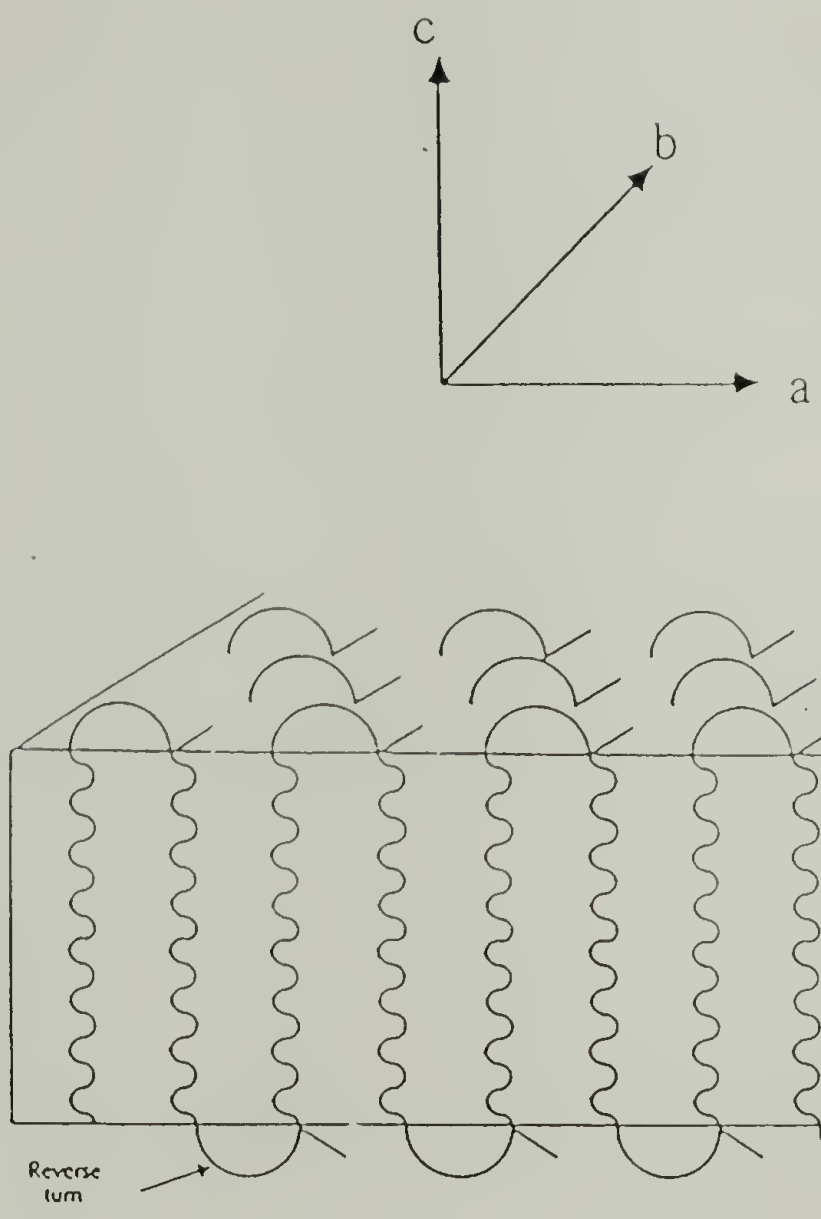


Figure 1.6 Desired three dimensional scaffold; the reference directions shown in this figure will be used in the rest of the thesis; *c* is the direction along the chain, *b* the direction perpendicular to the sheets and *a*is the third orthogonal direction.

chirality in order to fold in accordance to the design. We chose polypeptides due to three reasons: 1) A large amount of information is available on the structure-property relationships of polypeptides due to studies on globular and fibrous proteins. 2) The recent advances in recombinant DNA technology can permit the synthesis of precise polypeptides with excellent sequence, chiral and molecular weight fidelity. 3) The repertoire of monomers that are available in nature is diverse with respect to polarity, charge and sterics. Figure 1.7 shows the 20 natural amino acids that are available as monomers in the design of copolypeptides; furthermore the repertoire can be expanded by incorporating non-natural amino acids[20].

1.3 Structure-Property Relationships in Polypeptides

1.3.1 Secondary Structures

The conformation of a polypeptide is defined by the three dihedral angles ϕ, ψ and ω for every amino acid in that polypeptide. These are the dihedral angles corresponding to the N-C α bond, C α-CO bond and CO-N bond respectively. Although the polypeptide can sample all of the ϕ, ψ and ω space, a few conformations are seen to occur with a high frequency; a few of these secondary structures are discussed below.

1.3.1.1 Alpha Helix

The alpha-helix with 3.6 residues per turn with a rise per residue of 1.5 Å was first proposed by Pauling[21,22] as a stable structure in

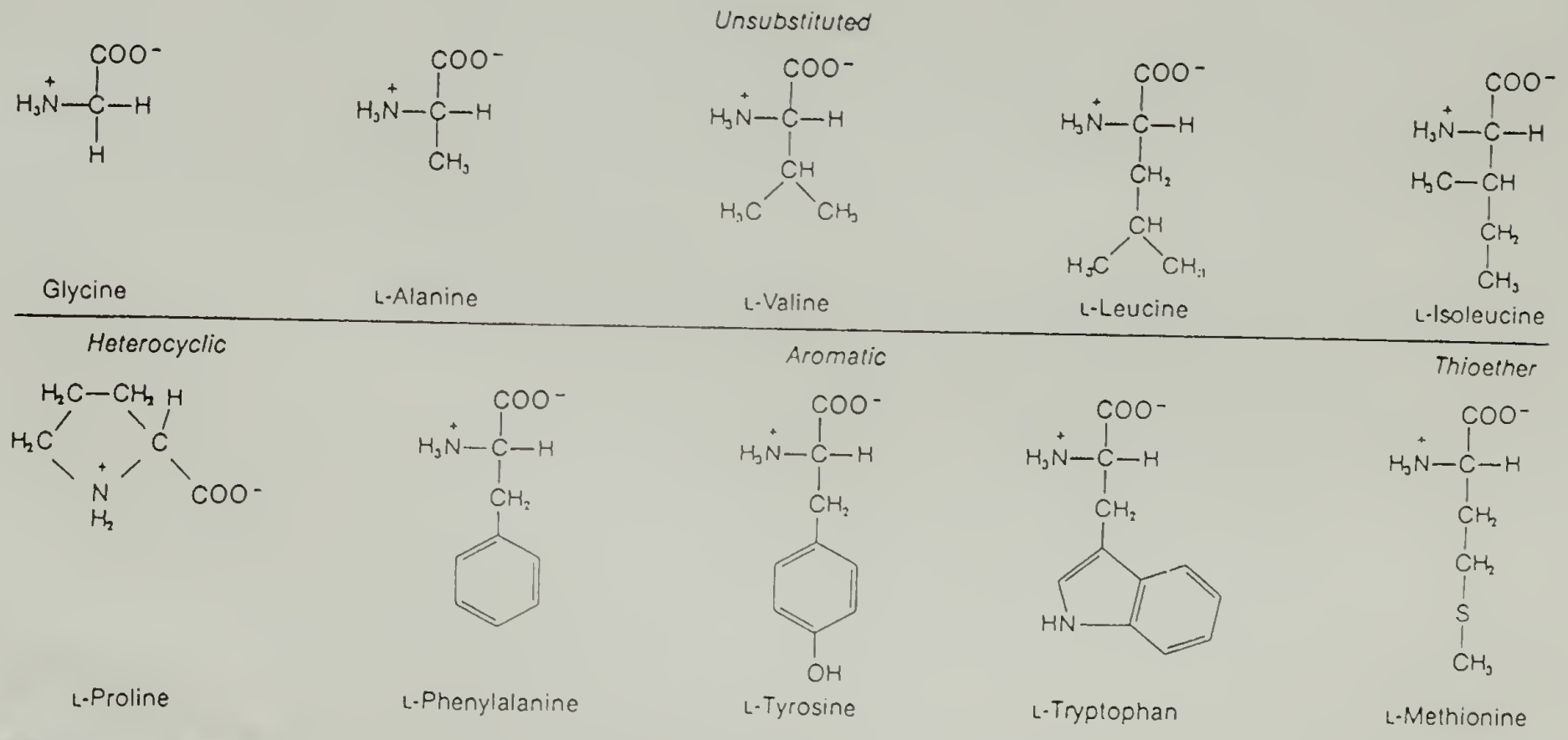
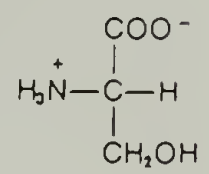


Figure 1.7 Natural amino acids (at pH=7.0); these form the repertoire of monomers for the synthesis of the copolymers; adapted from work by Voet[6].

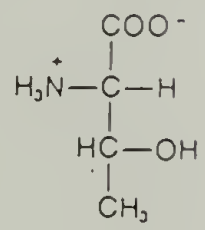
Hydroxy

Mercapto

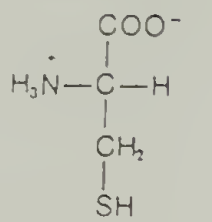
Carboxamide



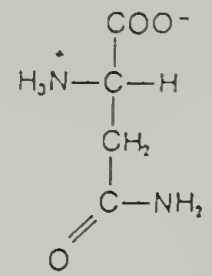
L-Serine



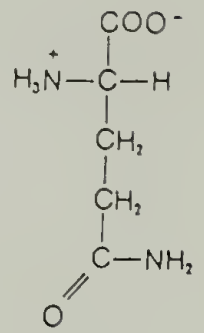
L-Threonine



L-Cysteine



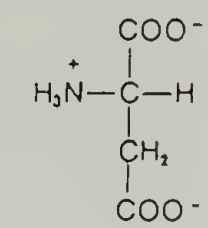
L-Asparagine



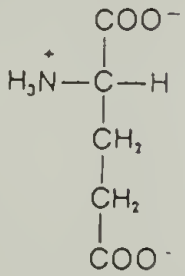
L-Glutamine

Monoamino, dicarboxylic

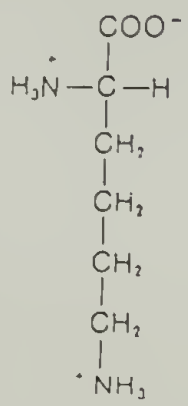
Diamino, monocarboxylic



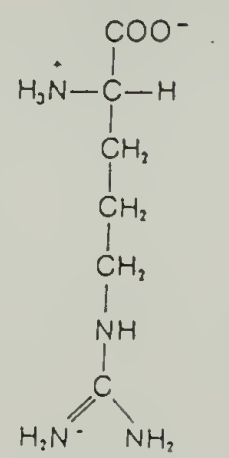
L-Aspartate



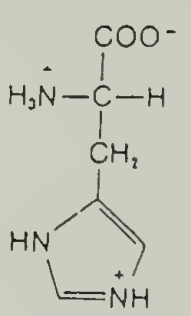
L-Glutamate



L-Lysine



L-Arginine



L-Histidine

Figure 1.7 (contd.)

polypeptides. The proposed structure which is repetitive with ϕ and ψ values of -60° and -40° respectively, is shown in Figure 1.8. The structure facilitates hydrogen bonding between an N-H and the C=O three residues apart along the chain. Since the helix is a rigid secondary structure which is formed by intramolecular hydrogen bonding, a large amount of work has been subsequently carried out on its formation and properties.

The alpha-helix has a macrodipole associated with it. It is often observed that alpha-helices have residues with negative charges at their N-termini and positive residues at their C-termini; this is attributed to the favorable interaction of these charges with the macrodipole of the helix[24]. The correlation, however, is far from perfect because stable helices have been observed without stabilization charges at either end of the helix; the stability of the helices in these cases has been attributed to the hydrogen bonding and coulombic interactions between the side chains of amino acids three residues apart along a chain[25]. It has been proposed that the residues flanking the N-terminus and C-terminus should possess side chains capable of hydrogen bonding to the unbonded N-H and C=O groups at the ends of the helix[26]. This effect is more pronounced in shorter helices because the eight unbonded residues at the ends of the helix significantly reduce the stabilization energy[26]; for example, a helix 12 residues long, contains eight internal N-H···O=C hydrogen bonds; however, the first four N-H and the last four C=O do not have any internal hydrogen bonding partners.

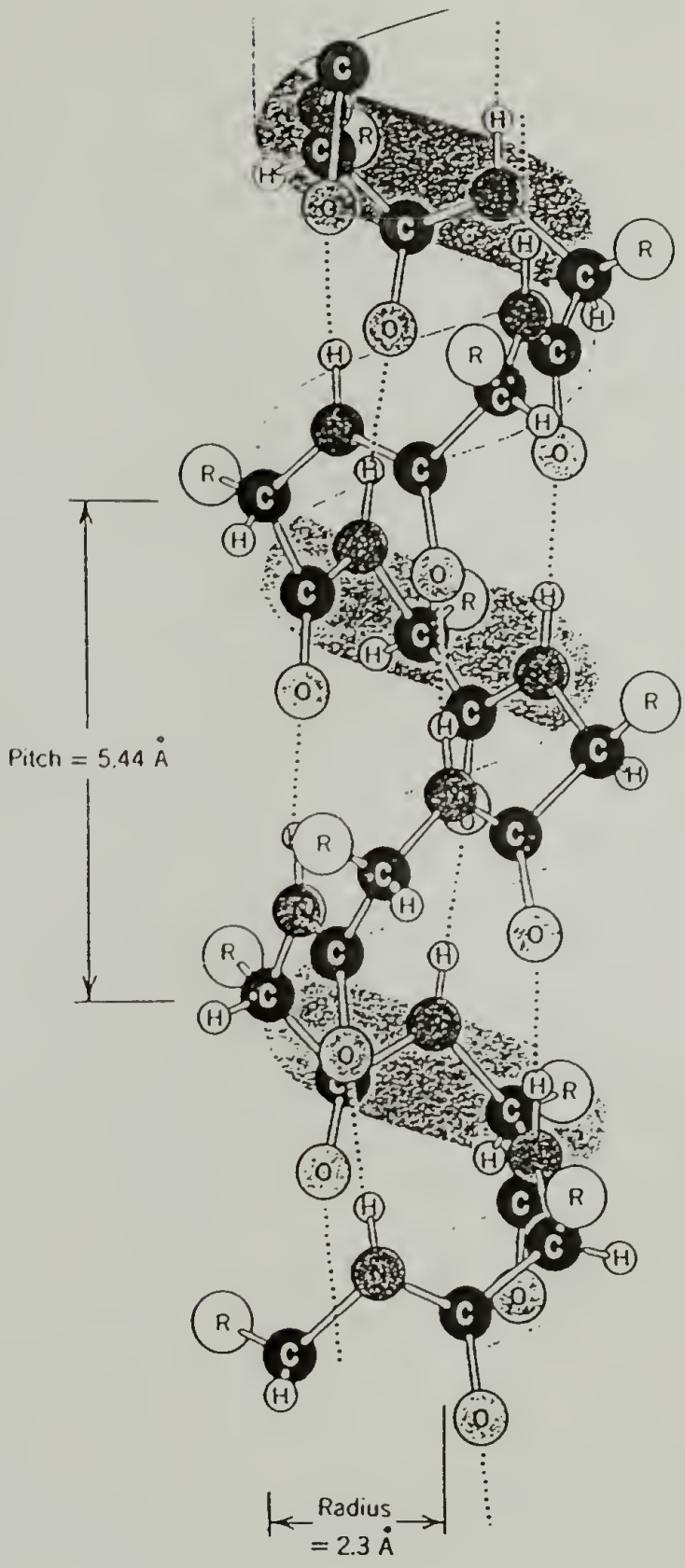


Figure 1.8 Structure of the alpha helix; adapted from work by Lewin[23].

It is generally assumed that the formation of helices is determined by the primary sequence of the chain and is not dependent on the formation of any other tertiary structure. The shortest number of residues required to form a helix is dependent on the sequence and varies between 10 and 20[26].

1.3.1.1.1 Use of Helices to Control Tertiary Structure

Monomeric helices and multimeric helices joined together by flexible spacers have been used to obtain predetermined tertiary structures. DeGrado and coworkers have designed oligopeptides, 12 or 16 residues long, to form amphiphilic helices with a tendency to form four helix bundles[27,28]. The sequences are designed such that the helices which form are rich in leucine on one surface, and in glutamic acid and lysine on the other. This favors the formation of four helix bundles in aqueous environments; the bundles have a hydrophobic core and a hydrophilic exterior. Their work shows that helices can be designed *de novo*.

1.3.1.1.2 Use of Synthetic Helices to Study Proton Channeling

The ability to design helices *de novo* has been used to study proton channelling in membranes by Chung[29]. The ability to design helices and introduce a fluorescent tryptophan in its sequence without distorting its secondary structure has helped assess the role of short helices in proton channelling. Their studies involved the interaction of helices with lipid membranes. It was observed that the helices lie parallel to the planar lipid bilayer when no voltage is applied across the membrane. With the

application of a voltage, a fraction of the helices align themselves to a transmembrane orientation; a few of the helices then cluster to form an ion conducting channel across the membrane.

Multimeric helices, however, responded differently to the same experiments. When four helix monomers were attached to a rigid tetraphenylporphyrin ring[30], the ion conduction was nearly voltage independent. This suggests that the rigid system of four clustered helices exists in a transmembrane orientation even without the application of an external voltage.

1.3.1.1.3 Other Uses of Synthetic Helices

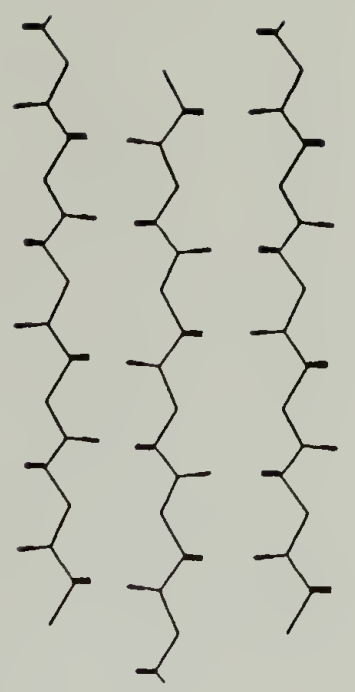
The ability to design helices has also been used to study the binding of calmodulin-binding peptides[31]. In an attempt to mimic the hydrophobic cores of enzymes, four helices have been interconnected with various bridges to reduce the configurational entropy of monomeric helices, and thereby drive the four helices to form bundles with more stability. There has been success in self-assembly of four monomeric helices[28]and it has been suggested that future generation of similar structures could be used to provide a semirigid matrix for positioning funtional groups in order to facilitate catalysis and binding[27].

1.3.1.2 Collagen

Another form of secondary structure with a large persistence length is seen in collagen which is present in tendon, bone, cornea and other locations in animals which require tensile strength and reversible extensibility. The characteristic meridional reflection in the x-ray diffraction patterns of oriented collagen is always between 2.82 Å and 2.90 Å. The principal equatorial reflection however, varies between 10.4 Å and 15.5 Å depending on the sample preparation; when the sample is wet, the larger distance is observed. The propensity for the unit cell to expand laterally led Pauling to propose that the hydrogen bonds were satisfied intramolecularly[32]. Analysis of density leading to the determination of mass per unit rise of a chain lead to the proposal that there are about three residues per unit in the structure of collagen. The conclusion then was that collagen consisted of three chains which were intertwined together with a fiber axis repeat of 2.86 Å. It has been observed that the collagen type proteins diplay the presence of only two hydrogen bonds per 2.86 Å whereas alpha-helices show three hydrogen bonds per element; a compositional evaluation of collagens shows that one-third of the amino acids are either proline and hydroxyproline; since proline is devoid of an N-H hydrogen it cannot hydrogen bond. Subsequent work has shown that polypeptides containing proline at every third residue do indeed form these intertwined structures[33].

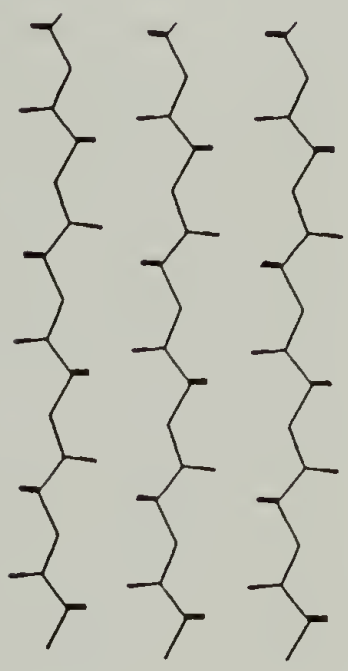
1.3.1.3 Beta Sheet

Beta strands are characterized by dihedral angles of ϕ and ψ values of -139° and 135° respectively and all trans amide bonds. The assembly of beta strands can lead to antiparallel pleated beta sheets, parallel pleated beta sheets (Figure 1.9)[34], parallel rippled beta sheets or antiparallel rippled beta sheets. In all these structures the polypeptide chain has a two fold screw axis and the hydrogen bonds are intermolecular. Figure 1.9 shows the antiparallel and parallel pleated beta sheet conformations for polyglycine. If one defines the directionality of a chain as the direction from the N-terminus to the C-terminus, the anti parallel beta sheet features neighboring chains (in the hydrogen bonding direction) with opposite orientation; all the chains in a sheet are in the same orientation in the parallel beta sheet structure. In the antiparallel structure, all the N-H····O=C are colinear; there is a slight deviation from this perfect angle situation when the sheet contains parallel chains and should cause the parallel beta sheet structure to be less stable than the antiparallel structure. Although the figure shows the most simple case, amino acids other than glycine which also favor the beta sheet conformation can be incorporated in the beta sheet structure;



(i)

(ii)



(iii)

(iv)

Figure 1.9 Antiparallel pleated beta sheet structure for polyglycine; (i) shows the projection along a direction which is perpendicular to the plane containing the hydrogen bonds of a sheet comprising three chains (conventionally referred to as the 'b' direction); (ii) shows the same sheet viewed along the direction which is parallel to the hydrogen bonding direction (conventionally referred to as the 'a' direction); (iii) and (iv) show similar projections for parallel beta sheets.

the side groups of these amino acids would then decorate the two sides of the sheet with varying density depending on the sequence of the amino acids in the polypeptide; in some cases the sheets are decorated only on one side with the side chains to give rise to polar sheets and in some cases the sheets are decorated on both the sides to yield apolar sheets. These aspects for copolypeptides of alanine and glycine are discussed in Chapter 3. The repeat distance along a chain in a beta sheet conformation is 6.9 Å and the repeat distance along the hydrogen bonded axis is 9.4 Å for antiparallel and 4.7 Å for parallel orientations; the dimension orthogonal to these two (the intersheet dimension) is dependent on the specific side chains of the amino acids comprising the copolypeptides.

It was found that polypeptides rich in alanine and glycine form beta sheets; Pauling was the first to observe that silk from *Bombyx mori* favored the antiparallel pleated beta sheet morphology in the crystalline state[35]. Subsequent studies on *Bombyx mori* and synthetic copolypeptides of alanine and glycine have confirmed this observation[36-38]. A more detailed discussion of the crystal structure of these polypeptides is presented in Chapter 3.

1.3.1.4 Beta Turns

Most of the work on the characterization of turns in polypeptides has been based on a systematic analysis of the crystal structure of globular proteins. Wilmot and Thornton[39] have classified beta turns into seven conventional categories and one unconventional category depending on the dihedral angles. A sequence of residues is defined to be in a turn by these workers if the distance between an alpha carbon (of the i^{th} residue) and the alpha carbon three residues ($i+3^{rd}$ residue) removed from it along the chain is less that 7 Å, and if the sequence is not a part of an alpha-helix. The seven conventional categories that they define have been tabulated in Table 1.1 (refer to section 1.3.1 for definition of ϕ, ψ and ω). They observe that types I' and II' turns are not very common in general but are very common in beta-hairpins where the chain reverses its direction leading to the formation of antiparallel beta sheet structures.

An analysis by Chou and Fasman[40] which is similarly based on structure-sequence relationships from known protein structures, classifies beta turns into eleven categories. Sibanda and Thornton[41] have specifically analyzed beta hairpin turns in great detail and assigned preferences for amino acids to reside in various beta hairpin turn types.

All the analyses discussed above categorize the natural amino acids on the basis of their tendency to reside in (or to be excluded from) turns, their preferred turn types and locations within the various turn types.

Table 1.1

Definition of beta turn types in polypeptides (refer to 1.3.1.4 for details)

(Adapted from work by Wilmot and Thornton[39])

Turn Type	ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
I	-60	-30	-90	0
I'	60	30	90	0
II	-60	120	80	0
II'	60	-120	-80	0
VIa	-60	120	-90	0
VIb	-120	120	-60	0
VIII	-60	-30	-120	120

A more detailed analysis and review of the literature on turns is presented by Dougherty[42].

1.3.2 Prediction of Secondary Structures

Various approaches have been used to derive general rules for the relation between secondary structure of a protein and its primary sequence.

1.3.2.1 Chou and Fasman

Chou and Fasman have assigned probabilities to various amino acids depending on their tendency to reside in or interrupt alpha helices ($P\alpha$) and beta sheets($P\beta$)[43]. They then predict that if four 'helix formers'

are clustered together, a helix propagates in both the directions until it is terminated by the presence of a tetrapeptide 'helix breaker'; tetrapeptides are designated 'breakers' or 'formers' based on the probabilities of the corresponding amino acids. A similar analysis is carried out for the tendency for formation of beta sheets. The values of their probabilities are listed in Table 1.2; the higher the probability, the larger the tendency to be a helix or beta sheet 'former"; the lower the probability, the higher the tendancy to be a helix or beta sheet 'breaker'.

1.3.2.2 Eisenberg

Eisenberg and coworkers[44] observe that proteins that fold into similar structures can have large differences in their primary sequences. They analyze known protein structures and assign each residue a score based on its hydrophobic and polar environment, and the local secondary structure; this computation converts the three dimensional information on the protein into a one dimensional string. The one dimensional string is used to predict the similarity of structure with any other polypeptide; the larger the 'one dimensional similarity' between two proteins, the higher the probability of them folding into similar tertiary structures. Using this method they detect structural similarities between actins and heat shock proteins even though they have no sequence similarity.

Table 1.2

$P\alpha$ and $P\beta$ values for amino acids

(Adapted from Chou and Fasman[43])

Amino Acid	$P\alpha$
Glu	1.53
Ala	1.45
Leu	1.34
His	1.24
Met	1.20
Gln	1.17
Trp	1.14
Val	1.14
Phe	1.12
Lys	1.07
Ile	1.00
Asp	0.98
Thr	0.82
Ser	0.79
Arg	0.79
Cys	0.77
Asn	0.73
Tyr	0.61
Pro	0.59
Gly	0.53

Amino Acid	$P\beta$
Met	1.67
Val	1.65
Ile	1.60
Cys	1.30
Tyr	1.29
Phe	1.28
Gln	1.23
Leu	1.22
Thr	1.20
Trp	1.19
Ala	0.97
Arg	0.90
Gly	0.81
Asp	0.80
Lys	0.74
Ser	0.72
His	0.71
Asn	0.65
Pro	0.62
Glu	0.26

1.3.2.3 Levitt

Levitt[45] states that 19 out of the 20 naturally occurring amino acids have preferences to reside in alpha-helices, beta sheets or reverse turns. His analysis shows that 1) valine, isoleucine, tyrosine, phenylalanine, tryptophan and threonine prefer beta sheet structures, 2) serine, aspartic acid, aspargine, glycine and proline prefer reverse turns, 3) arginine has no preference and 4) all other residues prefer alpha helices. The conformational preferences as predicted by his work are shown in Table 1.3

1.3.3 Formation of Unique Structures *In Vitro*

Most polypeptides synthesized *in vivo* adopt unique secondary and tertiary structures. The kinetics of folding and the role of intracellular machinery to achieve the structure are ill-understood; hence, it is not certain whether the unique structure obtained *in vivo* is the state with the lowest free energy or a state kinetically frozen-in due to interactions with other components of the cell. The complexity of the folding process is reduced *in vitro;* in this case the secondary and tertiary structures are determined to a very large extent only by the primary sequence of the polypeptide; under the appropriate conditions of solvent, temperature and concentration the folding process can then be controlled by the thermodynamics of folding which are controlled soley by the primary sequence. The studies on protein folding which have been discussed above suggest that even in cases where the

Table 1.3

Conformational preferences of amino acids

(Adapted from work by Levitt[45])

Type of Structure	Residues Favoring	Residues Indifferent	Residues Breaking
α-helix	Ala, Leu, Met, His, Gln, Lys, Cys	Val, Ile, Phe, Trp, Asp, Asn, Arg	Tyr, Thr, Gly, Ser, Pro
β-sheet	Val, Ile, Phe, Trp, Tyr, Thr	Ala, Leu, Met, His, Gly, Ser, Arg	Glu, Gln, Lys, Asp, Asn, Pro, Cys
Reverse turn	Gly, Ser, Asp, Asn, Pro	Gly, Gln, Lys, Tyr, Thr, Arg	Ala, Leu, Met, His, Val, Ile, Phe, Trp, Cys, Arg

cellular machinery does not play a role in the folding process, the structures could be unique. The prediction of these structures is a very difficult task and as previous studies indicate, at the current time is in its infancy.

1.4 Design Strategy to Achieve Formation of Lamellar Scaffolds

As a first step to designing more complicated scaffolds we are interested in designing repetitive polypeptides which would self assemble into lamellae of uniform, predetermined thickness (Figure 1.6). The design comprises of a 'stem' region and a 'turn' region. The stem region would comprise a sequence of amino acids which would form an extended secondary structure with a high persistence length, eg. an alpha helix or a beta strand. Work at this laboratory focused on beta strands being the 'stem' regions. Repetitive alanyl-glycine diads were known to favor beta sheet formation and were therefore selected to form the stem. This work, in particular, focuses on the synthesis and characterization of a polypeptide **1** comprising ten repeats of **1-m**. (The polypeptides are denoted by a number, and the repeating monomer by the number followed by **-m**.)

Previously, McGrath synthesized repetitive peptides of **3-m**[46], Krejchi synthesized a polypeptide comprising 36 repeats of **2-m**[47]. Creel synthesized repetitive polypeptides of **4-m**[48] which were variants of **3** with different numbers of alanyl-glycyl diads in the stem region.

Dougherty synthesized variants of the repetitive octapeptides **2** with different turn compositions based on predictions by Sibanda and Thornton[41]; these polypeptides (5) contained various residues in positions denoted by X and Y; the propensity for turn formation was explored as a function of X and Y.

-(AlaGly)$_3$GluGly(GlyAla)$_3$GluGly-	**1-m**
-(AlaGly)$_3$GluGly-	**2-m**
-(AlaGly)$_3$ProGluGly-	**3-m**
-(AlaGly)$_n$ProGluGly-	**4-m**
-(AlaGly)$_3$XY	**5-m**

In addition, this work also involved the synthesis of repetitive polypeptides of **6-m**, **7-m** and **8-m**, although not much structural characterization was carried out on these.

-(AlaGly)$_3$AspGly(GlyAla)$_3$AspGly-	**6-m**
-(AlaGly)$_3$GluGly(AlaGly)$_3$ValGly-	**7-m**
-(AlaGly)$_3$GluGly(AlaGly)$_3$MetGly-	**8-m**

Various aspects of the structural characterization and the comparison to similar polypeptides will be discussed in different chapters in this thesis. The two polypeptides **1** and **2** are compositionally identical but vary in the local sequence of beta-sheet-forming-segments in parts of the repeating unit sequences (alternate beta-sheet-forming-segments in **1** have been designed with local sequence inversions). Formation of turns at

the glutamic acid residue could lead to interesting systematic differences in the crystal structures of **1** and **2**; hence, they form an important contrasting set of polypeptides.

All the polypeptides above have been designed to form lamellae as shown in Figure 1.6. If the chains did fold as predicted, we anticipated **7** and **8** to yield lamellae with different functionalities on the two surfaces.

The structural characterization in this work is limited to **1**. Broadly, there are two aspects to the problem; the first is to design and synthesize polymers which can form nanoscale scaffolds; the second is to prove that when these polymers crystallize, they do form the desired scaffolds. In the second chapter, the synthesis of the copolymers which have a potential to scaffold as depicted in Figure 1.6 has been discussed. In the rest of the chapters, an attempt is made to structurally assess the formation of this scaffold. In the third chapter, the unit cell of **1** has been determined. A comparison of the unit cell with that of **2** suggests that, if the chains are folding in the crystalline forms of the two polymers, the types of turns are limited to ones with odd residues. The third section deals with the proof of adjacent reentry of the chains by using FTIR. The fourth section, which uses solid state NMR to determine dynamics of side chains, suggests that the glutamic acid residues are at the fold surface.

1.5 References

(1) March, J. *Advanced Organic Chemistry*; John Wiley & Sons: New York, 1986.

(2) Siegel, R. W. *MRS Bulletin* **1990**, *October*, 60-67.

(3) Paul, D. R.; Barlow, J. W. *J. Macromol. Sci.-Rev. Macromol. Chem.* **1980**, *C18(1)*, 109-168.

(4) Mason, J. A.; Sperling, L. H. *Polymer Blends and Composites*; Plenum: New York, 1976.

(5) Billmeyer, F. W. *Textbook of Polymer Science*; John Wiley & Sons: New York, 1984.

(6) Voet, D.; Voet, J. G. *Biochemistry*; John Wiley & Sons: New York, 1990.

(7) Birge, R. *Annu. Rev. Phys. Chem.* **1990**, *41*, 683-733.

(8) Diesenhofer, J.; Michel, H. *Science* **1989**, *245*, 1463-1473.

(9) Diesenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *J. Mol. Biol.* **1984**, *180*, 385-398.

(10) *Electron Transfer in Inorganic, Organic and Biological Systems*; Bolton, J. R.; Mataga, N.; McLendon, G., Ed.; ACS: Washington, 1991; Vol. 228.

(11) Hopfield, J. J.; Onuchic, J. N.; Beratan, D. N. *Science* **1988**, *241*, 817-819.

(12) *Polymer Supported Reactions in Organic Synthesis*; Hodge, P.; Sherrington, D. C., Ed.; John Wiley & Sons: New York, 1980.

(13) Greig, J. A.; Sherrington, D. C. *Polymer* **1978**, *19*, 163-175.

(14) Frechet, J. M. J.; Darling, G. D.; Itsuno, S.; Lu, P.; Meftahi, M. V.; Rolls, W. A. *Pure Appl. Chem.* **1988**, *60(3)*, 353-364.

(15) Deratani, A.; Darling, G. D.; Horak, D.; Frechet, J. M. J. *Macromolecules* **1987**, *20*, 767-772.

(16) *ACS Symposium Series*; Ford, W. T., Ed.; American Chemical Society: Washington, 1986; Vol. 308, pp 187-230.

(17) Breslow, R. *Science* **1982**, *218*, 532-537.

(18) *The Stereochemistry of Macromolecules*; Ketley, A. D., Ed.; Marcel Dekker: New York, 1967.

(19) Overberger, C. G.; Pierre, T. S.; Vorshheimer, N.; Yaroslavsky, S. *J. Am. Chem. Soc.* **1963**, *85*, 3513.

(20) Dougherty, M. J.; Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *Macromolecules* **1993**, *26*, 1779-1781.

(21) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci.* **1951**, *37*, 261-271.

(22) Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci.* **1951**, *37*, 205-211.

(23) Lewin, B. *Genes III*; John Wiley and Sons: New York, 1987.

(24) Shoemaker, K. R.; Peter, S. K.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563-567.

(25) Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. *Science* **1990**, *250*, 669-673.

(26) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632-1641.

(27) Ho, S. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 6751-6758.

(28) Regan, L.; DeGrado, W. F. *Science* **1988**, *241*, 976-978.

(29) Chung, L. A.; Lear, J. D.; DeGrado, W. F. *Biochemistry* **1992**, *31*, 6608-6616.

(30) Akerfeldt, K. S.; Kim, R. M.; Camac, D.; Groves, J. T.; Lear, J. D.; DeGrado, W. F. *JACS* **1992**, *114*, 9656-9657.

(31) O'Neil, K. T.; Wolfe, H. R.; Ericson-Viitanen, S.; DeGrado, W. F. *Science* **1987**, *236*, 172-174.

(32) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci.* **1951**, *37*, 273-281.

(33) Devlin, T. M. *Textbook of Biochemistry with Clinical Correlations;* 3rd ed.; Wiley-Liss: New York, 1992.

(34) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci.* **1951**, *37*, 729-740.

(35) Marsh, R. E.; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1-34.

(36) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706-712.

(37) Lotz, B.; Cesari, F. C. *Biochimie* **1979**, *61*, 205-214.

(38) Lotz, B.; Brack, A.; Spach, G. *J. Mol. Biol.* **1974**, *87*, 193-203.

(39) Wilmot, C. W.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221-232.

(40) Chou, P. Y.; Fasman, G. D. *J. Mol. Biol.* **1977**, *115*, 135-175.

(41) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759-777.

(42) Dougherty, M. J. *Univ. of Mass.,Amherst* **1993**, *Ph. D. Thesis*

(43) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13 (2)*, 222-245.

(44) Bowie, J. U.; Luthy, R.; Eisenberg, D. *Science* **1991**, *253*, 164-170.

(45) Levitt, M. *Biochemistry* **1978**, *17(20)*, 4277-4285.

(46) McGrath, K. P. *Univ. of Mass.,Amherst* **1990**, *Ph. D. Thesis*

(47) Krejchi, M. T. *Univ. of Mass., Amherst* **1993**, *Ph. D. Thesis*

(48) Creel, H. S. *Univ. of Mass., Amherst* **1993**, *Ph. D. Thesis*

CHAPTER 2

SYNTHESIS

2.1 Materials

All reagents, chemicals and enzymes have been categorized according to source.

2.1.1 Sigma Chemical Company

Acrylamide

Adenosine triphosphate

Agarose

Ammonium chloride

Ammonium persulphate

Ammonium phosphate

Ampicillin; sodium salt

Antifoam A

Bisacrylamide

Boric acid

Bovine serum albumin (BSA)

5-Bromo-4-chloro-3-indolyl-β-galactopyranoside (X-gal)

Calcium chloride

Chloramphenicol

Coomassie Brilliant Blue R-250

Cyanogen bromide

Dithiothreitol (DTT)

D-(+)-Glucose

Ethylenediaminetetraacetic acid (EDTA); sodium salt

Glycine

β-Isopropylthiogalactoside (IPTG)

L-Alanine

L-Arginine

L-Aspartic acid

L-Cystein

L-Glutamic acid

L-Glutamine

L-Histidine

L-Isoleucine

L-Leucine

L-Lysine

L-Methionine

L-Phenyl Alanine

L-Proline

L-Serine

L-Threonine

L-Tryptophan

L-Tyrosine

L-Valine

2-Mercaptoethanol

3-(N-morpholino)propane sulfonic acid (MOPS)

2-(N-morpholino)ethane sulfonic acid (MES)

Phenol

Potasium phosphate

Sodium dodecylsulfate (SDS)

Sodium hydroxide

Sodium phosphate

Sodium chloride

N,N,N',N' Tetramethylethylenediamine (TEMED)

Thiamine (Vitamin B1)

Tris (hydroxymethyl) aminomethane (Tris)

Urea

2.1.2 United States Biochemical

Calf intestinal alkaline phosphatase

Sequenase Version 2.0 (DNA sequencing kit)

2.1.3 Qiagen, Inc.

'Midi' and 'Maxi' Preparation Plasmid kit

2.1.4 Fisher Scientific

Acetic acid

Ethanol (100 %)

Formic acid (98%)

Glycerol

HCl

n-Butanol

2-Propanol

2.1.5 Difco Laboratories

Bacto-tryptone

Yeast extract

2.1.6 Du Pont NEN

{^{3}H}-Glycine

Rapid Autoradiography Enhancer

2.1.7 New England Biolabs

2.1.7.1 Enzymes

*Ava*I

*Bam*HI

*Ban*I

*Eco*RI

*Hin*dIII

T4 DNA Ligase

T4 Polynucleotide Kinase

2.1.7.2 Primers

M13 Primers

2.1.7.3 Molecular Weight Markers

Lambda DNA-*Hin*dIII digest

Lambda DNA-*Bst*EII digest

pBR322 DNA-*Msp*I digest

2.1.8 Cambridge Isotope Laboratories

L-Alanine (2-^{13}C, 99%)

L-Alanine (^{15}N, 95-99%)

Formic acid (formyl-D, 98%)

Glutamic acid (^{15}N, 95-99%)

Glycine (1-^{13}C, 99%)

Glycine (^{15}N, 99%)

2.2 Solutions

All solutions and buffers were constituted in deionized, distilled water (dd water) unless otherwise stated. In some instances, the typically used concentrations have been specified as a dilution ratio; i.e. a stock buffer or solution was diluted 'n' times if the dilution ratio is specified as n X for routine use. In some instances, the weights or volumes of the

constituent components have been specified; these were the typical scales of volumes that were used routine storage of the solutions.

2.2.1 Autoclaved Solutions (120 °C for 30 min)

Calcium chloride	1M
Glycerol	30% volume
Magnesium chloride	1M
Magnesium sulfate	1M
Sodium acetate	3M; pH=4.8
Sodium chloride	3M
Sodium dodecyl sulfate (SDS)	20%
Sodium hydroxide	2 M
Zinc chloride	1M

2.2.2 Filter Sterilized Solutions

Adenosine triphosphate	10 mM
Ampicillin, Sodium salt	200 mg/ml
Chloramphenicol	25 mg/ml in 100% ethanol
Dithiothreitol (DTT)	0.4 M
Glucose	1M
IPTG	0.2 M (stored at -20 °C)
Thiamine (vitamin B1)	5 mg/ml
X-gal	40 mg/ml in DMF

2.2.3 Other Solutions

Amino acid solution 20 X

500 mg/L of each of the 20 amino acids was added to dd water; the pH was adjusted to 7.0 by addition of 2 M NaOH

Acrylamide solution 40 wt%

The stock solution was 40 wt%; the solution was stored in the dark at 4 °C.

EDTA (pH=8.0) 0.5 M

After dissolution of EDTA the pH was adjusted with NaOH.

Ethidium Bromide 10 mg/ml

Phenol solution

Phenol was melted and was extracted repetitively with Tris-HCl (pH=7.5) until the pH was 7.5; this was stored at 4 °C in the dark.

M9 salts solution 5 X

15 g Na_2HPO_4 (anhydrous), 7.5 g KH_2PO4 (anhydrous), 1.25 g NaCl, 2.5 g NH_4Cl were dissolved in 500 ml dd water and autoclaved; after cooling, 1 ml of 0.25 M $CaCl_2$ and 2.5 ml of 1 M $MgSO_4$ was added.

Tris 1M

Tris-HCl was prepared at pH=7.4, 7.6 and 8.0 by the addition of HCl to Tris solution; most electrodes do not accurately measure the pH of Tris solutions; hence the pH was ascertained by use of pH paper[1].

Staining solution (for protein gels)

12.5 g Coomassie Brilliant Blue R-250, 400 ml methanol and 70 ml acetic acid were added to dd water and the final volume was brought to 1 L.

Destaining solution (for protein gels)

200 ml methanol and 100 ml acetic acid were added to dd water and the final volume was brought to 1 L.

2.2.4 Enzymes and Enzyme Buffers

2.2.4.1 Enzymes

The restriction enzymes were used as purchased and were diluted as mentioned in the Section 2.4.

RNAse 10 mg/ml

RNAse was dissolved in Tris-HCl (pH=7.5) and heated to 100 °C for 10 min, allowed to cool and was stored at -20 °C.

DNAse 1 mg/ml

DNAse was dissolved in 0.15 M NaCl solution and was stored at -20 °C.

2.2.4.2 Buffers

Dephosphorylation buffer 10 X

The buffer contained 10 mM $ZnCl_2$, 10 mM $MgCl_2$ and 100 mM Tris-HCl (pH=8.3).

T4 polynucleotide kinase buffer 10 X

The buffer contained 200 mM Tris-HCl (pH=7.6), 50 mM, $MgCl_2$, 50 mM DTT and bovine serum albumin (500 μg/ml); ATP was added at the time of use at a concentration of 1 mM.

*Bam*HI 1 X

The buffer contained 150 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1mM DTT and was supplemented with 100 μg/ml of BSA.

*Ban*I 1 X

The buffer contained 50 mM potassium acetate, 20 mM Tri-acetate, 10 mM magnesium acetate and 1 mM DTT.

*Ava*I 1 X

The buffer contained 50 mM potassium acetate, 20 mM Tri-acetate, 10 mM magnesium acetate and 1 mM DTT.

*Eco*RI 1 X

The buffer contained 150 mM NaCl, 100 mM Tris-HCl, 5 mM $MgCl_2$ and 0.025% triton X-100.

T4 DNA Ligase 1 X

The buffer contained 50 mM Tris-HCl (pH=7.8), 10 mM $MgCl_2$, 20 mM DTT, 1 mM ATP and 50 μg/ml of BSA.

2.2.5 Other Common Buffers

TBE buffer 5 X

The buffer contained 54 g Tris base, 27.5 g boric acid and 20 ml of 0.5 M EDTA (pH=8.0) in 1 L.

TAE buffer 50 X

The buffer contained 242 g Tris base, 57.1 ml glacial acetic acid, 100 ml of 0.5 M EDTA (pH=8.0) in 1 L.

Gel elution buffer 1 X

The buffer contained 0.5 M NH_4OAc, 1 mM EDTA (pH=8.0).

GTE buffer 1 X

The buffer contained 4.5 g glucose, 7.5 ml 1M Tris HCl (pH=8.0), 6 ml of 0.5 M EDTA in a total volume of 300 ml; the buffer was stored after autoclaving.

Stacking gel buffer (protein gels) 4 X

The buffer contained 0.4 g SDS, 1.6 ml 0.5 M EDTA and 6 g Tris base; the final volume was made to 80 ml and the pH to 6.8 (with HCl).

Separating gel buffer (protein gels) 4 X

The buffer contained 36.3 g Tris base, 8 ml 0.5M EDTA and 3.2 ml 25% SDS; the pH was made to 8.8 by addition of HCl; the final volume was 200 ml.

Running buffer (protein gels) 5 X

The buffer contained 57.6 g glycine, 2 g SDS and 2.97 g EDTA; the pH was made to 8.6 by addition of Tris and the final volume was 2 L.

Sample buffer (protein gels) 2 X

The buffer contained 2.5 ml of stacking gel buffer, 2 ml 50% glycerol, 0.4 g SDS, 0.3 g DTT and 0.1 mg bromophenol blue.

TBF1 buffer (for preparation of competent cells)

The buffer contained 100 mM MES (pH=6.2), 100 mM $RuCl_2$, 50 mM MnCl, 10 mM $CaCl_2$; the pH was made to 5.8 with glacial acetic acid.

TBF2 buffer (for preparation of competent cells)

The buffer contained 0.418 g MOPS, 0.242 g $RbCl_2$, 1.665 g $CaCl_2$, 30 ml glycerol; the pH was made to 6.5; the final volume was 200 ml.

Loading buffer (for DNA) 6 X

The buffer contained 0.25 wt% bromophenol blue, 0.25 wt% xylene cyanol and 30 vol% glycerol.

2.2.6 Media

YT media

The liquid media contained 8 g Bacto tryptone, 5 g yeast extract and 5 g NaCl; the final volume was 1 L.

M9AA media

The liquid media contained 8 g glucose, 200 ml 20 X amino acid solution, 800 ml of M9 salt solution and 0.8 ml of 5 mg/ml vitamin B1; the total volume was 4 L; this was filter sterilized prior to use.

2.3 Methods

2.3.1 General Methods

Unless otherwise stated, all preparative and analytical gels were prepared as described by Sambrook, Fritsch and Maniatis[1]. Decisions regarding the compositions of gels and their operating variables were made according to their recommendation for the expected molecular weights of interest of the DNA or protein. Ampicillin was used for selection of pUC18 and pET-3a at a concentration of 200 μg/ml in liquid and solid media. Chloramphenicol was used for selection of pLysS at a concentration of 25

μg/ml in liquid and solid media. X-gal, the chromogenic substrate, was used at a concentration of 40 ng/ml. The scheme used for DNA manipulation for the synthesis of **1**, **6**, **7** and **8** is shown in Figure 2.1. All DNA manipulations were performed as described by Sambrook, Fritsch and Maniatis[1] unless otherwise stated. All centrifugations in the microcentrifuge were carried out at RT (approximately 25 °C) at speeds of 13,500 x g unless otherwise stated. Lambda DNA-*Hin*dIII digest and pBR322 DNA-*Msp*I digest were used as molecular weight markers for the analysis of DNA unless otherwise stated. All the gels used for the analysis of double stranded DNA were stained with ethidium bromide and visualized by UV fluorescence.

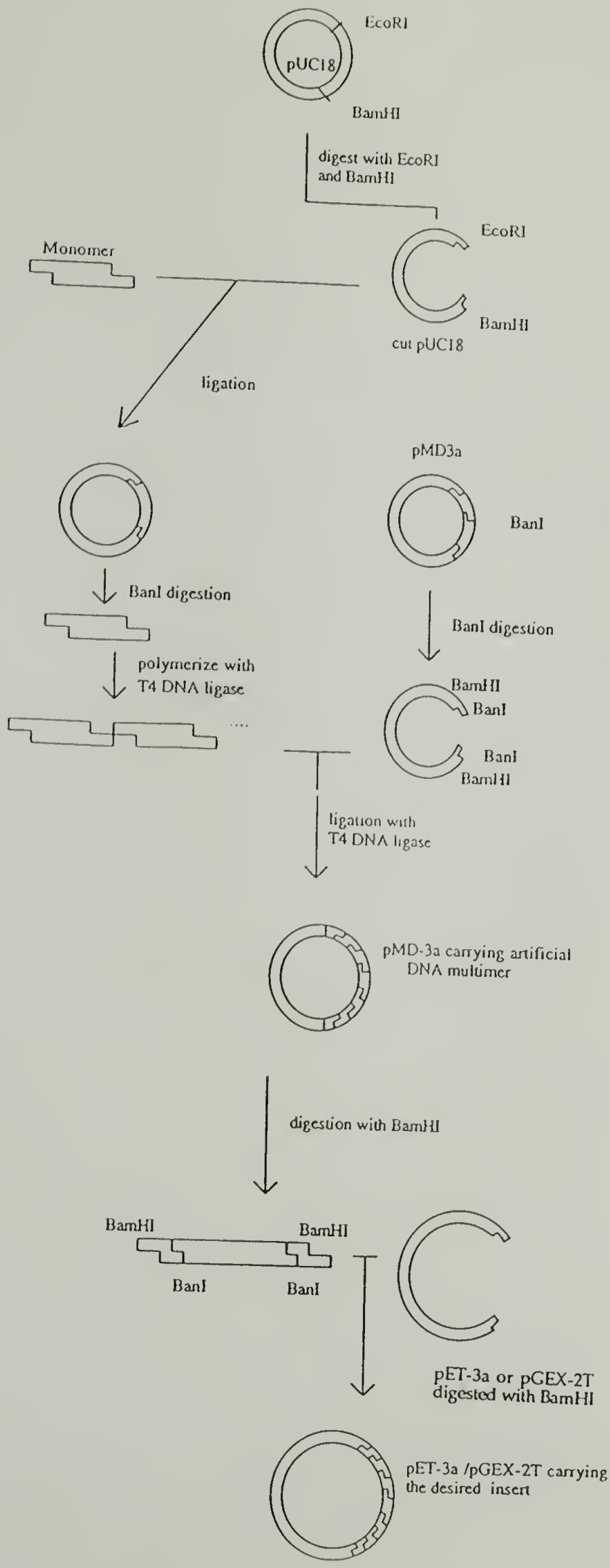


Figure 2.1 DNA manipulation scheme used for biosynthesis of 1, 6, 7 and 8.

2.3.1.1 Measurement of Cell Density

The cell densities were estimated by measuring the optical densities on a Hitachi U2000 spectrophotometer at 600 nm. The path length was 1 cm.

2.3.1.2 Precipitation of DNA

DNA in solution was mixed with two volumes of 100% ethanol and was incubated at -20 °C for at least 10 h. The suspension was spun down for 30 min at 13500 x g on a microcentrifuge in 1.5 ml polypropylene microfuge tubes or at 9000 x g in 250 ml polycarbonate bottles in a GSA rotor. The pellet was washed twice with 70% ethanol and was dried either in a centrifuge speed-vac or a vacuum oven at RT. The dried DNA was then dissolved in TE buffer or dd water.

2.3.1.3 Purification of DNA by Dialysis

The DNA extracted by most procedures was often contaminated with high concentrations of salt. The DNA was purified before use in subsequent steps by the following procedure: The DNA was solubilized in dd water. A petri dish was filled with dd water and a 0.025 μm, 1 cm diameter teflon filter paper (manufactured by Millipore Corp.) was placed at the air-water interface. The impure DNA solution was spotted on the filter paper and the solution was dialysed for 2 h. This reduced the quantity of salt in the DNA.

In some preparations, the DNA solution was contaminated with enzymes and other proteins. Any residual enzyme in the reaction solution was removed by phenol/chloroform extraction followed by chloroform extraction.

2.3.1.4 Preparation of Competent Cells

Cell were grown to saturation at 37 °C in YT media. An aliquot of these cells was used to inoculate 20 ml of fresh YT media. The cells were grown in a shaker flask at 37 °C until the optical density was 0.4 units. The cell suspension was spun at 4000 x g at 4 °C for 5 min and the cells were resuspended in cold (4 °C) TBF1 solution. The cells were held at 4 °C for 20 min and were subsequently spun down at 4000 x g at 4 °C for 5 min. The pellet was resuspended in cold TBF2 solution and aliquoted in 100 μl quantities and stored at - 80 °C. In this entire procedure care was taken to handle the cells gently and keep them at 4 °C.

2.3.1.5 Transformation

Approximately 1 ng of plasmid DNA was added to a 100 μl aliquot of competent cells. The mixture was gently agitated and placed on ice for 30 min and subsequently heat shocked for 90 s at 42 °C. To the heat shocked cells, 1 ml of YT media was added and the suspension was incubated at 37 °C in order for the cells to develop antibiotic resistance.

The transformation mix was then plated on to YT plates containing the appropriate antibiotic selection.

2.3.1.6 Isolation of Plasmid DNA

Cell were grown to saturation in culture tubes at 37 °C. A volume of 1.5 ml of the suspension was spun down in a microcentrifuge at 13500 x g for 5 min. The cell pellet was resuspended in 100 μl GTE buffer. In order to lyse the cells, NaOH (20 μl of 2 M solution) and SDS (20 μl of 10% solution) were added and the mixture was incubated at RT for 2 min, followed by incubation on ice for 10 min. The genomic DNA was precipitated by the addition of sodium acetate (150 μl of 3 M (pH=4.8) solution); after incubation on ice for 20 min, this suspension was spun down at 13500 x g. The supernatant was transferred to a fresh microfuge tube and DNAase free RNAase was added (1 μl of 10 mg/ml solution). The solution was incubated at 37 °C for 1 h. Subsequently, the solution was extracted with 450 μl of a 1:1 phenol/chloroform mixture and then with 450 μl of chloroform. The supernatant was mixed with two volumes of 100% ethanol and icubated at -20 °C for 12 h. The suspension was spun down at RT for 30 min at 13500 x g and the suspended DNA was washed several times with 70 % ethanol. The DNA was finally dried in a centrifuge speed-vac.

2.3.1.7 Quantification of DNA

Estimates of DNA were made by comparing with known amounts of molecular weight standards; the DNA was visualized by staining of electrophoresis gels with ethidium bromide. Larger quantities of DNA were quantified by measuring absorbance at 260 nm. A solution containing 50 μg/ml of double stranded DNA has an absorbance of 1.0 at 260 nm; a solution containing 40 μg/ml of single stranded DNA has an absorbance of 1.0 at 260 nm[1]. These experiments were conducted on a Hitachi U2000 spectrophotometer in quartz cuvettes with a path length of 1 cm.

2.3.1.8 Large Scale Isolation of Plasmids

Maxi and midi preparation kits supplied by Qiagen, Inc. were used for isolation of plasmids from large volumes of cultures. The maxi preparation kit was used for 150-500 ml of culture. The cell suspension was spun down by centifugation at 9000 x g for 30 min. The cell pellet was resuspended in 10 ml of buffer P1 (50 mM Tris-HCl, 10 mM EDTA pH=8.0) and 10 ml of buffer P2 (0.2 M NaOH, 1 % SDS) was added and the mixture was incubated for 10 min at RT. To this suspension, 10 ml of buffer P3 (3.0 M KAc, pH=5.5) was added and the mixture was centrifuged for 10 min immediately. The supernatant was centrifuged once again in order to obtain a particle free lysate, and then passed through a modified silicagel anion-exchange resin column provided by Qiagen and was subsequently washed thrice with a wash buffer (1.0 M NaCl, 50 mM MOPS 15% ethanol, pH=8.5). The DNA was finally eluted by 15 ml of an elution buffer (1.25 M

NaCl, 50mM Tris/HCl, 1 mM EDTA, pH=8.0) and 0.7 volume of isopropanol was added to precipitate the DNA. The DNA suspension was spun at 9000 x g, the DNA was washed with 70 % ethanol and subsequently dried. The yield of plasmid was approximately 500 μg per 500 ml of saturated culture containing pUC18.

2.3.1.9 Purification of DNA Fragments

2.3.1.9.1 Polyacrylamide Gels

The band of interest was excised from the gel by a razor blade and was transferred to a microfuge tube. The gel was crushed and gel elution buffer was added to soak the gel, which was incubated at 37 °C for 4 h. The gel particles were removed by centrifugation and the supernatant was transferred to a fresh microfuge tube and mixed with oyster glycogen and two volumes of 100% ethanol; the solution was subsequently incubated at -20 °C overnight. The DNA suspension was centrifuged for 30 min and the pellet was washed with 70% ethanol. This process yielded DNA with low salt contamination (in comparison to DNA from electroelution) and the efficiency of recovery from the gel was the best at low molecular weights (up to 1000 base pairs).

2.3.1.9.2 Agarose Gels

Electroelution was used for larger DNA fragments which were separated by agarose gel electrophoresis. The band of interest was excised from the gel with a razor and was placed in the sample well of an IBI Model UEA electroeluter. The DNA was eluted in the wells containing elution buffer. The voltage used for electroelution was typically 100-150 V; TBE or TAE buffers were used. After 1 h of electroelution, the elution buffer was removed with a syringe. Two volumes of 100% ethanol were added to this and the DNA was allowed to precipitate at -20 °C overnight. The DNA was recovered by centrifugation for 30 min, washed with 70% ethanol and dried. The DNA extracted by this procedure often was contaminated with high concentrations of salt. In order to reduce the salt contamination, the DNA was solubilized in dd water and was dialyzed on a 0.025 μm teflon filter paper (manufactured by Millipore Corp.).

2.3.1.10 Preservation of Strains

2.3.1.10.1 Stab Cultures

Bacto-agar (1.5 wt%) was added to YT solution. The suspension was autoclaved and the molten solution was aliquoted in 1 ml volumes into sterile vials. The vials were closed under sterile conditions and were allowed to cool gradually. Cells were grown to saturation and the stab vial was inoculated with a cauterized platinum loop. The cells were allowed to

grow in the vial overnight and the vial was sealed air-tight and stored at 4 °C. Strains can be stored for up to two years using this procedure[1].

2.3.1.10.2 Frozen Stock

Cells were grown to an optical density of 0.8 units. Autoclaved glycerol (300 μl) was mixed with 700 μl of the culture in a sterile cryovial; the mixture was gently agitated to ensure even mixing. This was then quenched in liquid nitrogen to rapidly freeze the solution. The vials were stored at -70 °C. Cells can be stored practically indefinitely in this state[1]. In order to recover the bacteria, the frozen surface was scraped with a sterile platinum loop.

2.3.1.11 Double Stranded DNA Sequencing

Sequencing of plasmid DNA was carried out in pUC18 using the protocol and chemicals provided in the 'Sequenase Version 2.0' kit manufactured by United States Biochemical. The kit employs an improved version of the Sanger dideoxy chain termination method[2]. A modified version of the wild type T7 DNA polymerase which is devoid of any 3'-5' exonuclease activity is used for chain extension.

2.3.1.11.1 Denaturation of Plasmid DNA

Approximately 500-1000 ng of plasmid DNA was used for sequencing. The plasmid was denatured in 0.2 M NaOH for 30 min at 37

°C. The mixture was subsequently neutralized by addition of 0.1 volume of 3 M sodium acetate (pH=4.8) and the DNA precipitated by the addition of 3 volumes of 100% ethanol (-70 °C for 15 min). The DNA was washed with 70% ethanol, dried and reconstituted in 10 μl dd water.

2.3.1.11.2 Annealing the Primer to Template

The primer/template molar ratio that yielded the best sequencing results was 20:1. The -40 universal primer sold by New England Biolabs was used. To the denatured template, 2 μl of annealing buffer and 2 μl primer at a concentration of 30 μg/ml was added. The solution was annealed at 37 °C for 20 min.

2.3.1.11.3 Sequencing Reactions

The reactions conditions and the chemicals were as specified in the 'Sequenase Version 2.0' kit manufactured by United States Biochemical. {^{35}S}ATP was used to incorporate radioactivity in the DNA. The DNA sequences used in this work are rich in guanine and cytosine; this leads to extensive secondary structure which cannot be denatured even under the denaturing electrophoresis conditions. To circumvent this difficulty, parts of the sequence were analysed by carrying out the sequencing reactions with the 7-deaza analogues of the deoxynucleotides.

2.3.1.11.4 Gel Electrophoresis

It should be noted that the quality of the gel electrophoresis is often the limiting factor which determines the readability of a sequence. All the reagents used were electrophoresis grade. Since acrylamide solutions increase in conductivity over time, fresh monomer solutions were used. Samples were analysed on 7% denaturing polyacrylamide gels containing 7 M urea. The gel was preheated at a constant power of 60 W for 1 h before loading the samples. After electrophoresis, the gel was transferred from the glass plates on to 3MM Whatman filter paper and dried at 85 °C for 1.5 h. The gel was then placed in close contact with an x-ray film and exposed overnight at -70 °C.

2.3.1.12 Colony Purification

Colony purification was used in order to ensure the selection of pure clones. After transformation of the host cells, the transformation mix was plated on YT medium (Section 2.3.1.5); since the density of viable cells on the tranformation mix was high, a colony could comprise two or more clones.

A colony from a primary transformation plate was selected for colony purification. A cauterized platinum loop was used to scrape a small number of cells from the plate and streaked on a fresh YT plate containing the same antibiotics as the primary plate.

The plate was incubated under the same conditions as the primary plate. The colonies obtained on this plate were considered pure, i.e. each colony was initiated by a single viable cell.

2.3.2 Cloning of **1**, **6**, **7** and **8**

The nomenclature used for the polypeptide sequences has been described in Section 1.4.

2.3.2.1 Synthesis of Oligonucleotides

The eight complementary strands corresponding to **1-m**, **6-m**, **7-m** and **8-m** were synthesized on a Biosearch 8700 DNA synthesizer which utilizes β-cyanoethylphosphoramidite chemistry[3]. The synthesis was carried out on glass beads (1 μM scale) funtionalized with the nucleotide at the 3' end of the DNA strand. After the automated step-wise synthesis of the oligonucleotides, they were cleaved from the glass beads by reaction with concentrated ammonium hydroxide at 70 °C for 24 h. The glass beads were separated from the supernatant by centrifuging on a microcentrifuge for 5 min. The supernatant was placed in fresh microfuge tubes and the ammonium hydroxide was removed by drying in a centrifuge speed-vac at 60 °C for 12 h. The DNA thus obtained was dissolved in dd water and quantified on the Hitachi U2000 spectrophotometer.

2.3.2.2 Purification of Oligonucleotides

The oligonucleotides synthesized on the synthesizer are polydisperse due to some premature termination steps and need to be purified in order to obtain the desired oligopeptide in the pure form. This was accomplished by gel electrophoresis of the crude oligonucleotides in a 15% polyacrylamide denaturing gel containing 7 M urea. Oligonucleotides (6 μl of 40 mg/ml oligonucleotide solution) were mixed with 6 μl of 2 X sample loading buffer (50 vol% formamide, 0.2% bromophenol blue, 0.2% xylene cyanol and 10 vol% glycerol). The samples were heated at 95 °C prior to loading. The conditions of electrophoresis were constant current (15 mA) with TBE as the electrolyte; the current was stopped when the xylene cyanol was at the bottom of the 14 cm gel. The gel was peeled from the glass plates and visualized by UV shadowing. The bands of interest were excised from the gel and placed in fresh microfuge tubes. In each case, the gel was crushed and then soaked in two volumes of gel elution buffer for 12 h at 37 °C. The DNA was isolated by the procedure described in section 2.3.1.9.1.

2.3.2.3 Phosphorylation of Oligonucleotides

The oligonucleotides were phosphorylated at the 5' end by T4 polynucleotide kinase. The polynucleotide buffer used for this reaction contained 80 mM Tris-HCl (pH=7.5), 20 mM DTT, 10 mM $MgCl_2$ and 1 mM ATP. To a 50 μl mixture of purified oligonucleotides (15 μg) and polynucleotide buffer, 20 units of the enzyme were added and the solution

was incubated at 37 °C for 1 h. The enzyme was inactivated by heating at 70 °C for 1 h. The DNA was precipitated as described in section 2.3.1.2.

2.3.2.4 Annealing of Oligonucleotides

Complementary strands at a concentration of 0.2 μg/ μl each were mixed in a solution comprising 50 mM NaCl and 20 mM $MgCl_2$. The total volume was 40 μl. The mixture was heated to 95 °C for 5 min in a DNA Thermal Cycler manufactured by Perkin Elmer Cetus and the temperature was uniformly ramped to RT in 4 h. The DNA was precipitated by mixing with two volumes of ethanol.

2.3.2.5 Preparation of pUC18 for Cloning

The cloning vector pUC18 was purchased from New England Biolabs. The plasmid (20 μg) was cleaved at the *Bam*HI site in the polylinker region by incubation with 20 units of *Bam*HI overnight in the buffer recommended by NEB. The DNA was precipitated and freed of salt by dialysis on a teflon filter. The linearized plasmid was further cleaved overnight by 20 units of *Eco*RI. The linearized plasmid was separated from the uncut and supercoiled plasmids by electrophoresis of the mixture and the band of interest was excised and electroeluted to yield the required linearized plasmid. The DNA thus obtained was dephosphorylated by calf intestinal alkaline phosphatase (CIP). The conditions used for this procedure were as provided by United States Biochemicals. The linearized dephosphorylated plasmid was precipitated and purified, and stored at -20

°C for further use. The yield of the linearized dephosphorylated plasmid was estimated at 9 μg.

2.3.2.6 Insertion of DNA Duplex in pUC18

In the case of each polypeptide, the duplex DNA monomer was ligated into the plasmid by incubation of 100 ng plasmid, 100 ng monomer and 0.5 μl (400,000 U/ml) of T4 DNA ligase in a total volume of 40 μl at 15 °C for 12 h; the buffer recommeded by New England Biolabs (which included 1 mM ATP) was used for ligation.

2.3.2.7 Transformation of *E. coli* DH5 αF' with pUC18 Containing DNA Monomer

The DNA from the ligation mix was precipitated and spot dialyzed on a teflon filter in order to remove salts and glycerol. Approximately 10 ng of DNA was used to transform competent cells as described before. The cells were spread on YT plates containing ampicillin (200 μg/ml), X-gal (40 μg/ml) and IPTG (0.1 mM). The plates were incubated at 37 °C overnight and were subsequently cooled to 4 °C in order to enhance the blue color developed by the reaction with X-gal. For every monomer, 10 white colonies were selected and were colony purified on fresh plates in order to ensure purity of transformants.

2.3.2.8 Confirmation of Monomer Inserts

Four colony purified clones were selected in each case for further analysis. Saturated overnight 5 ml YT cultures were grown for each clone and the plasmid DNA was isolated using the previously described protocol. The DNA obtained from 0.75 ml of saturated culture was digested with 20 units of *Ban*I. The digested DNA was electrophoresed along with Lambda DNA-*Hin*dIII digest and pBR322 DNA-*Msp*I digest as molecular weight markers. All the colonies showed the presence of an approximately 48 base pair fragment corresponding to the monomers cloned in the vectors.

The 16 clones were sequenced using the double stranded sequencing protocol described before. At least one of each showed the correct sequence. The plasmids which contained the monomer with the correct sequence were labelled pAJ-PUC-1-m, pAJ-PUC-8-m, pAJ-PUC-6-m and pAJ-PUC-7-m in direct correspondance to **1-m**, **8-m**, **6-m** and **7-m**; the strains containing these plasmids were labeled DH5αF'-pAJ-PUC-1-m, DH5αF'-pAJ-PUC-8-m, DH5αF'-pAJ-PUC-6-m and DH5αF'-pAJ-PUC-7-m respectively; these strains were resistant to ampicillin. No indications of hot spots for deletion or insertion at specific locations were observed; however, 1 clone of **1-m**, 1 clone of **6-m**, 2 clones of **7-m** and 2 clones of **8-m** did not show the correct sequence. It could not be stated with certainty if this was due to genetic instability or due to artifacts in the sequencing protocols. One clone of each was chosen for further analysis.

2.3.2.9 Large Scale Synthesis of DNA Monomers

The clones that were chosen were grown in 500 ml of 2 X YT medium with ampicillin selection. The plasmid DNA was isolated from the cells by using a Qiagen Maxi Preparation kit. The recombinant plasmids were cleaved with *Ban*I; typically 500 μg of plasmid were dissolved in 1 ml of buffer (provided by New England Biolabs) and 1000 units of enzyme were used for cleavage. The reaction was allowed to proceed overnight and the fragments were electrophoresed on a 12% polyacrylamide gel at 200 V with TBE as the electrolyte, until the xylene cyanol band was a few centimeters away from the bottom of the gel. The bands of interest were excised from the gel and the DNA was eluted by soaking in elution buffer. In each case, isolated DNA monomer was saved for further use in ligation.

2.3.2.10 Preparation of pMD-3a for Cloning

The vector pMD-3a was synthesized by Dougherty[4]. The plasmid (20 μg) was cleaved overnight at the *Ban*I site in the polylinker region with 20 units of *Ban*I in the buffer recommended by NEB. The DNA was precipitated and was freed of salt by dialysis on a teflon filter. The linearized plasmid was separated from the uncut and supercoiled plasmids by electrophoresis of the mixture and the band of interest was excised and electroeluted to yield the required plasmid. The DNA thus obtained was dephosphorylated by calf intestinal alkaline phosphatase (CIP). The linearized dephosphorylated plasmid was precipitated and purified by spot dialysis on a teflon filter, and stored at -20 °C for further use.

2.3.2.11 Preparation of Multimers

An aliquot of DNA monomers (50 μg) was dissolved in 50 μl of ligation buffer. To this solution 1000 units of T4 DNA ligase was added and the solution was incubated at 14 °C overnight. The ligated fragments were analysed on an 8% acrylamide gel. A wide distribution of multimer sizes was observed as evidenced by the presence of a ladder with the degree of polymerization varying between 2 and 30. The highest degree of polymerization could not be ascertained accurately on this gel because of its acrylamide concentration; a lower concentration acrylamide gel or an agarose gel would have been better suited for this purpose. The bands containing the multimers were excised and the multimers were divided into two groups, one with degree of polymerization lower than 15 and one higher. The DNA multimers were then eluted by soaking in elution buffer.

2.3.2.12 Insertion of Multimers in pMD-3a in *E. coli* HB101

The linearized dephosphorylated vector pMD-3a (100 ng) was mixed with the two sets of multimers in two fresh microfuge tubes. The total volume of each ligation reaction was 40 μl. The buffer recommended by NEB was used. To this solution, 200 units of T4 DNA ligase were added and the ligation was allowed to occur at 14 °C overnight. After the ligation, the DNA was purified by spot dialysis on a teflon filter; it was observed that the efficiency of transformation depended strongly on the purity of the DNA.

Max Efficiency competent HB101 cells purchased from Gibco BRL were thawed on ice. Approximately 1 ng of plasmid DNA was added to a 100 μl aliquot of competent cells. The mixture was gently agitated and placed on ice for 30 min and subsequently heat shocked for 90 s at 42 °C. To the heat shocked cells, 1 ml of YT medium was added and the suspension was incubated at 37 °C in order for the cells to develop chloramphenicol resistance. The transformation mix was then spread on YT plates containing chloramphenicol.

2.3.2.13 Selection of the Multimerized Inserts

The plates were incubated at 37 °C for 20 h. In the case of each monomer, approximately 40 colonies were selected and the transformants were grown overnight in 5 ml YT medium with chloramphenicol selection. The plasmid DNA was isolated as explained before. To the plasmid DNA obtained from 1.5 ml YT culture, 20 units of *Bam*HI were added and the reaction was carried out in 20 μl of total volume at 37 °C overnight in the buffer provided by NEB. The DNA fragments were analysed on a 7% polyacrylamide gel. The largest insert was selected for each monomer for further analysis. This corresponded to the decamer for **1**, tetramer for **8** and **6**, and trimer for **7**. The plasmids which contained the multimers with the correct sequence were labelled pAJ-PMD-1, pAJ-PMD-8, pAJ-PMD-6 and pAJ-PMD-7 in direct correspondence to the decamer for **1**, tetramer for **8** and **6**, and trimer for **7**; strains containing these plasmids were labeled HB101-pAJ-PMD-1, HB101-pAJ-PMD-8, HB101-pAJ-PMD-6 and HB101-pAJ-PMD-7 respectively; these were resistant to chloramphenicol.

2.3.2.14 Insertion of Multimers in Expression Vector pET-3a

2.3.2.14.1 Preparation of Vector

The expression vector pET-3a (20 μg) was cleaved overnight at the *Bam*HI site in the polylinker region with 40 units of *Bam*HI in 1 ml total volume in the recommended buffer by NEB. The DNA was precipitated and was freed of salt by dialysis on a teflon filter. The linearized plasmid was separated from the uncut and supercoiled plasmids by electrophoresis (on a 1.5% agarose gel) and the band of interest was excised and electroeluted to yield the required plasmid. The DNA thus obtained was dephosphorylated by calf intestine phosphatase. The conditions used for this procedure were as provided by United States Biochemicals. The linearized dephosphorylated plasmid was precipitated and purified, and stored at -20 °C for further use.

2.3.2.14.2 Isolation of Multimerized Insert

The clones that were chosen in section 2.3.2.13 were grown in 100 ml of YT medium with chloramphenicol selection. The plasmid DNA was isolated from the cells by using Qiagen Midi Preparation kit. The recombinant plasmids were cleaved with *Bam*HI; typically 100 μg of plasmid was dissolved in 200 μl of buffer (provided by New England Biolabs) and 200 units of enzyme were used for cleavage. The cleavage was allowed to proceed overnight and the fragments were electrophoresed on a

12% polyacrylamide gel at 200 V with TBE as the electrolyte, until the xylene cyanol band was a few centimeters away from the bottom of the gel. The bands of interest were excised from the gel and the DNA was eluted by soaking in elution buffer.

2.3.2.14.3 Insertion of Multimerized Inserts into Expression Vector

A measured quantity of the linearized dephosphorylated vector pET-3a (100 ng) was mixed with the multimerized insert from section 2.3.2.14.2 in a fresh microfuge tube. The total volume of each ligation reaction was 40 μl. The buffer recommended by NEB was used. To this solution, 200 units of T4 DNA ligase were added and the mixture was incubated at 14 °C overnight. Subsequently, the DNA was precipitated and purified.

2.3.2.15 Determination of Orientation of Multimerized Inserts in pET-3a

Competent DH5αF' cells prepared in house were thawed on ice. Approximately 1 ng of plasmid DNA (from section 2.3.2.14.3) was added to a 100 μl aliquot of competent cells. The mixture was gently agitated and placed on ice for 30 min and subsequently heat shocked for 90 s at 42 °C. To the heat shocked cells, 1 ml of YT medium was added and the suspension was incubated at 37 °C in order for the cells to develop ampicillin resistance. The transformation mix was then spread on YT plates containing ampicillin. The plates were incubated overnight at 37 °C.

Five transformants each, were selected for the decamer of **1**, tetramers of **8** and **6**, and trimer of **7** and were colony purified.

The plasmid DNA was isolated by the DNA isolation protocol (section 2.3.1.6) and the DNA from 1.5 ml of YT culture was cleaved with 20 units of *Ava*I in a total volume of 20 μl. Once again the buffer employed was as recommended by NEB. The linker region in pMD-3a contains an *Ava*I site which is asymmetrically placed within the linker (Figure 2.3). Another *Ava*I site is present in pET-3a, 1425 base pairs upstream of the *Bam*HI site. Hence, if the insert were present in the wrong orientation, an *Ava*I digest would show the presence of a band of 1170 base pairs; in the event of the correct orientation, the band would be longer than 1170 by the length of the insert. The clones with the correct orientation were selected for further experiments. These plasmids were labeled pAJP-3a-1, pAJP-3a-8, pAJP-3a-6 and pAJP-3a-7 in direct correspondence to the decamer of **1**, tetramers of 8 and 6, and trimer of 7.

The strains containing these plasmids were grown in 5 ml YT cultures and the plasmid DNA was isolated in μg quantities for each plasmid. Competent BL21(DE3)pLysS cells prepared in house were thawed on ice. Approximately 1 ng of plasmid DNA was added to a 100 μl aliquot of competent cells. The mixture was gently agitated and placed on ice for 30 min and subsequently heat shocked for 90 s at 42 °C. To the heat shocked cells, 1 ml of YT medium was added and the suspension was incubated at 37 °C in order for the cells to develop ampicillin resistance. The transformation mix was then spread on YT plates containing ampicillin

and chloramphenicol. The plates were incubated overnight at 37 °C. One transformant each was selected for the decamer of **1**, tetramers of **8** and **6**, and trimer of **7** and were colony purified. The strains were labeled BL21(DE3)pLysS-pAJP-3a-1, BL21(DE3)pLysS-pAJP-3a-8, BL21(DE3)pLysS-pAJP-3a-6 and BL21(DE3)pLysS-pAJP-3a-7 in correspondence to the decamer of **1**, tetramers of 8 and 6, and trimer of 7; these strains were resistant to ampicillin and chloramphenicol. These strain were stored for short term in a stab culture at 4 °C and for long term in 50% glycerol at -70 °C.

2.3.2.16 Insertion into Expression Vector pGEX-2T

The plasmid pGEX-2T was prepared for ligation and the inserts were ligated using the same protocol as for pET-3a. The orientation of the inserts was checked by cleavage with *Ava*I. When the insert is in the correct orientation, the cleavage with *Ava*I produces a fragment which is 11 base pairs long; in the wrong orientation, the cleavage produces a fragment almost the same size as the insert. The 11 base pair fragment is too small to be seen on a gel. Hence, the orientation was determined by two simultaneous digestions of the plasmids (for each multimer), in two microfuge tubes, by *Bam*HI and *Ava*I along with a control which proved the efficacy of *Ava*I; the correct orientation led to no observed band on a 7% polyacrylamide gel with *Ava*I digestion and a band of approximately the size of the insert when digested with *Bam*HI; the incorrect orientation led to a band of approximately the size of the insert when digested with *Bam*HI and also with *Ava*I. The plasmids in the correct orientation were labelled

pAJX-3a-1, pAJX-3a-8, pAJX-3a-6 and pAJX-3a-7 in direct correspondence to the decamer of **1**, tetramers of 8 and 6, and trimer of 7. The plasmids were transferred into the expression strain using the same protocol used for the pET-3a system with the exception that the strain used was DH5αF'. The strains were labeled DH5αF'-pAJX-3a-1, DH5αF'-pAJX-3a-8, DH5αF-pAJX-3a-6 and DH5αF'-pAJX-3a-7 in direct correspondence to the decamer of **1**, tetramers of 8 and 6, and trimer of 7; these were resistant to ampicillin. These strains were stored for short term use in a stab culture at 4 °C and for long term in 50% glycerol at -70 °C.

2.3.2.17 Expression of Polypeptides

2.3.2.17.1 Expression in BL21(DE3)pLysS

Transformants (BL21(DE3)pLysS-pAJP-3a-1, BL21(DE3)pLysS-pAJP-3a-8, BL21(DE3)pLysS-pAJP-3a-6 and BL21(DE3)pLysS-pAJP-3a-7) containing the expression vectors carrying the multimerized inserts in the correct orientation were grown under ampicillin and chloramphenicol resistance in 20 ml of YT medium. Cultures were induced with IPTG (to a final concentration of 0.4 mM) in their mid-log phase (ca. 10^8 cell/ml, OD_{600}=0.8). {^{3}H}glycine (to a final concentration of 5 μCi/ml) was added at the time of induction in order to monitor the amount of protein synthesized as a function of time. Cells equivalent to 0.8 OD units were aliquoted at various time points between 0 min and 150 min. These were lysed in 40 μl of sample buffer comprising 1% SDS, 1% mercaptoethanol, 10mM Tris pH=7.5 and 10% glycerol and analyzed by a 10% discontinuous

SDS denaturing polyacrylamide gels as described by Sambrook, Fritsch and Maniatis[1]. The proteins in the gels were visualized by staining with Coomassie Brilliant Blue R-250 in the staining solution (Section 2.3.5) followed by destaining in the protein gel destaining solution. Subsequently, the gels were also visualized by fluorography; the gels were soaked in Rapid Autoradiography Enhancer (Du Pont/NEN), dried at 65 °C on Whatman filter paper for 2 h and were exposed to x-ray film for 3 days.

2.3.2.17.2 Expression in DH5 αF'

Transformants containing the expression vectors carrying the multimerized inserts in the correct orientation (DH5 αF'-pAJX-3a-1, DH5 αF'-pAJX-3a-8, DH5 αF'-pAJX-3a-6 and DH5 αF'-pAJX-3a-7) were grown under ampicillin resistance in 20 ml of YT media. Cultures were induced with IPTG (to a final concentration of 0.4 mM) in their mid-log phase (ca. 10^8 cell/ml, OD_{600}=0.8). Cells equivalent to 0.8 OD units were aliquoted at various time points between 0 min and 150 min. These were lysed in 40 μl of sample buffer comprising of 1% SDS, 1% mercaptoethanol, 10mM Tris pH=7.5 and 10% glycerol and analyzed by a 10% discontinuous SDS denaturing polyacrylamide gels as described by Sambrook, Fritsch and Maniatis[1]. The proteins in the gels were visualized by staining with Coomassie Brilliant Blue R-250 in the staining solution followed by destaining in the protein gel destaining solution. These gels were not visualized by autoradiography because the glutathione S-

transferase tag at the N-terminus of the protein of interest could be stained with Coomassie Brilliant Blue R-250.

2.3.2.18 Large Scale Synthesis and Purification of **1**

The pET-3a expression system was chosen for large scale expression and only polypeptide **1** was synthesized on a large scale.

Large scale fermentations were carried out in a 12 L 'Microferm fermentor' produced by New Brunswick Scientific Co. (New Brunswick, NJ). All fermentations were carried out in M9AA minimal medium under ampicillin and chloramphenicol selection. The sterile medium in the fermentor was inoculated with 500 ml of a saturated culture of cells containing the gene for expression of **1** (BL21(DE3)pLysS-pAJP-3a-1), grown overnight in M9AA minimal medium under ampicillin and chloramphenicol selection. The cells were grown in the fermentor at 37 °C into mid-log phase (ca. 10^8 cells/ml) before being induced by 0.4 mM IPTG (1200 mg of IPTG were added to the 12 L total volume). The cells were grown for an additional 2.5 h before being harvested; the final optical density of the culture was approximately 1.8. The culture was centrifuged at 5000 rpm on a Sorvall RC-3 centrifuge with a HG-4L rotor at 4 °C in 1 L centrifuge bottles for 20 min to sediment the cells. Cells obtained from 12 L of culture were resuspended in 150 ml of a buffer comprising 100 mM NaCl, 50 mM Tris pH=8.3 and 1mM EDTA. The suspension of cells was frozen and the cells were lysed by thawing them. $MgCl_2$ was added to a final concentration of 2.5 mM. In order to reduce the viscosity of the suspension, the DNA and RNA were cleaved by the addition of DNAse

and RNAse. The protein was purified by a four step reduction in pH followed by centrifugation at every step; the precipitate was discarded at each step. Ethanol was added to the supernatant at pH=4.0, to a final concentration of 40% and the mixture was incubated at 4 °C for 1 h. Ethanol was added to the supernatant obtained after centrifugation of the above to a final concentration of 80% and the mixture was incubated overnight at -20 °C. The purified protein was collected by centrifugation of the 80% ethanol suspension at 8000 rpm in 250 ml bottles and dried overnight in a vacuum oven at room temperature; this yielded a precursor polypeptide with a 25 amino acid fusion, FUS, at the N-terminus, due to the DNA sequence, DNA-FUS, between the start codon and the N-terminal methionine codon of the target polypeptide. The polypeptide of interest was obtained from the precursor by cleavage with cyanogen bromide. Cyanogen bromide was added to a ca. 2.5 mg/ml solution of the polypeptide in 70% formic acid to a concentration of ca. 2 mg/ml. The system was purged with nitrogen and the flask was allowed to stand at room temperature in the dark for 24 h. Solvent was removed on a rotary evaporator and distilled water was added to the dried protein fragments to make the concentration of the protein ca. 6 mg/ml. The mixture was stirred for 24 h to extract the soluble proteins. Centrifugation at 8000 rpm gave a clear supernatant which contained **1**, which was precipitated by addition of ethanol to a final concentration of 80%. The precipitate was dried *in vacuo* at room temperature and then over P_2O_5 in a drying pistol with refluxing ethanol. The purity of the sample was confirmed by amino acid compositional analysis, combustion analysis and ^1H NMR spectrometry.

Start Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Gly Ser Tyr Val Cys Gly Arg Lys Tyr Ser Arg Asp Pro Met (FUS)

ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGC GGA TCC TAT GTT TGC GGC CGC AAA TAT TCT CGC GAT CCG ATG

(DNA-FUS)

2.3.2.19 Large Scale Synthesis and Purification of **1** with Isotopically Enriched Amino Acids

The polypeptide was synthesized and purified as described in the previous section; the medium was modified according to the required isotope enrichment. The presence of amino acids of natural abundance in the medium yielded a precursor of polypeptide containing the natural abundance isotopes; large scale fermentation with ^{13}C-enriched or ^{15}N-enriched amino acids in the growth medium yielded a precursor of the polypeptide containing the heavier isotope at the required position in the corresponding amino acid.

2.4 Results and Discussion

2.4.1 Design and Synthesis of DNA Sequences

There exists a multiplicity in the repertoire of codons corresponding to each amino acid, thereby, leaving the designer the option of selecting the codon for the amino acid in the synthetic polypeptide[5-7]. Some codons are found in more abundance than others in a given microorganism. Sharp and Li[6] found the abundance of codons to correlate strongly with the amount of t-RNA corresponding to the particular codon. They propose that the cell uses these codons as regulatory devices to control the rate of synthesis of proteins which they encode. It has been suggested that the codons which are 'rare' are the rate determining elements in the translation of mRNA. Since the goal in this work was to overproduce the polypeptide of interest, care was taken to use only those codons that were abundant in *E.coli*.

Krejchi[8] and Dougherty[4] have succesfully synthesized polymers **2** and **5**. To ensure the successful biosynthesis of the polypeptides **1**, **6**, **7** and **8**, considerable homology to their sequences was maintained. The homology could not be 100% because of constraints imposed by the cloning strategy and the placement of restriction sites therein. The DNA strands corresponding to the four polypeptides and the choice of codons are shown in Figure 2.2. The ends of the monomer were designed to be *Ban*I restriction sites. The non-palindromic nature of this site favors head to tail

(i) $\{(AG)_3EG(GA)_3EG\}_n$

```
            Stop Gly  Ala Gly  Ala  Gly  Ala  Gly Glu  Gly Gly  Ala  Gly
AAT TCG TAA GGT GCC GGC GCT GGT GCG GGC GAA GGT GGC GCT GGT
      GC ATT CCA CGG CCG CGA CCA CGC CCG CTT CCA CCG CGA CCA
EcoR I          Ban I

Ala Gly  Ala  Glu Gly  Ala
GCG GGC GCT GAA GGT GCC G
CGC CCG CGA CTT  CCA CGG CCTAG
                 Ban I      BamH I
```

(ii) $\{(AG)_3DG(GA)_3DG\}_n$

```
            Stop Gly  Ala Gly  Ala  Gly  Ala  Gly Asp  Gly Gly  Ala  Gly
AAT TCG TAA GGT GCC GGC GCT GGT GCG GGC GAT GGT GGC GCT GGT
      GC ATT CCA CGG CCG CGA CCA CGC CCG CTA CCA CCG CGA CCA
EcoR I          Ban I

Ala Gly  Ala  Asp Gly  Ala
GCG GGC GCT GAT GGT GCC G
CGC CCG CGA CTA CCA CGG CCTAG
                 Ban I      BamH I
```

Figure 2.2 Sequence of DNA for (i) **1-m**, (ii) **6-m**, (iii) **7-m** and (iv) **8-m**.

Continued, next page

(iii) $\{(AG)_3EG(AG)_3VG\}_n$

```
            Stop Gly  Ala Gly   Ala   Gly Ala   Gly  Glu  Gly  Ala  Gly   Ala Gly
AATTCG TAA GGT GCC GGC GCT GGT GCG GGC GAA GGT GCA GGC GCT GGT
       GC ATT CCA CGG CCG CGA CCA CGC CCG CTT CCA CGT CCG CGACCA
EcoR I          Ban I
```

```
Ala  Gly  Val  Gly  Ala  Stop
GCG GGC GTT GGT GCC TAA CG
CGC CCG CAA CCA CGG ATT GCCTAG
          Ban I              BamH I
```

(iv) $\{(AG)_3EG(AG)_3MG\}_n$

```
            Stop Gly  Ala Gly   Ala   Gly Ala   Gly  Glu  Gly  Ala  Gly   Ala Gly
AATTCG TAA GGT GCC GGC GCT GGT GCG GGC GAA GGT GCA GGC GCT GGT
       GC ATT CCA CGG CCG CGA CCA CGC CCG CTT CCA CGT CCG CGACCA
EcoR I          Ban I
```

```
Ala  Gly  Met  Gly  Ala  Stop
GCG GGC ATG GGT GCC TAA CG
CGC CCG TAC CCA CGG ATT GCCTAG
          Ban I              BamH I
```

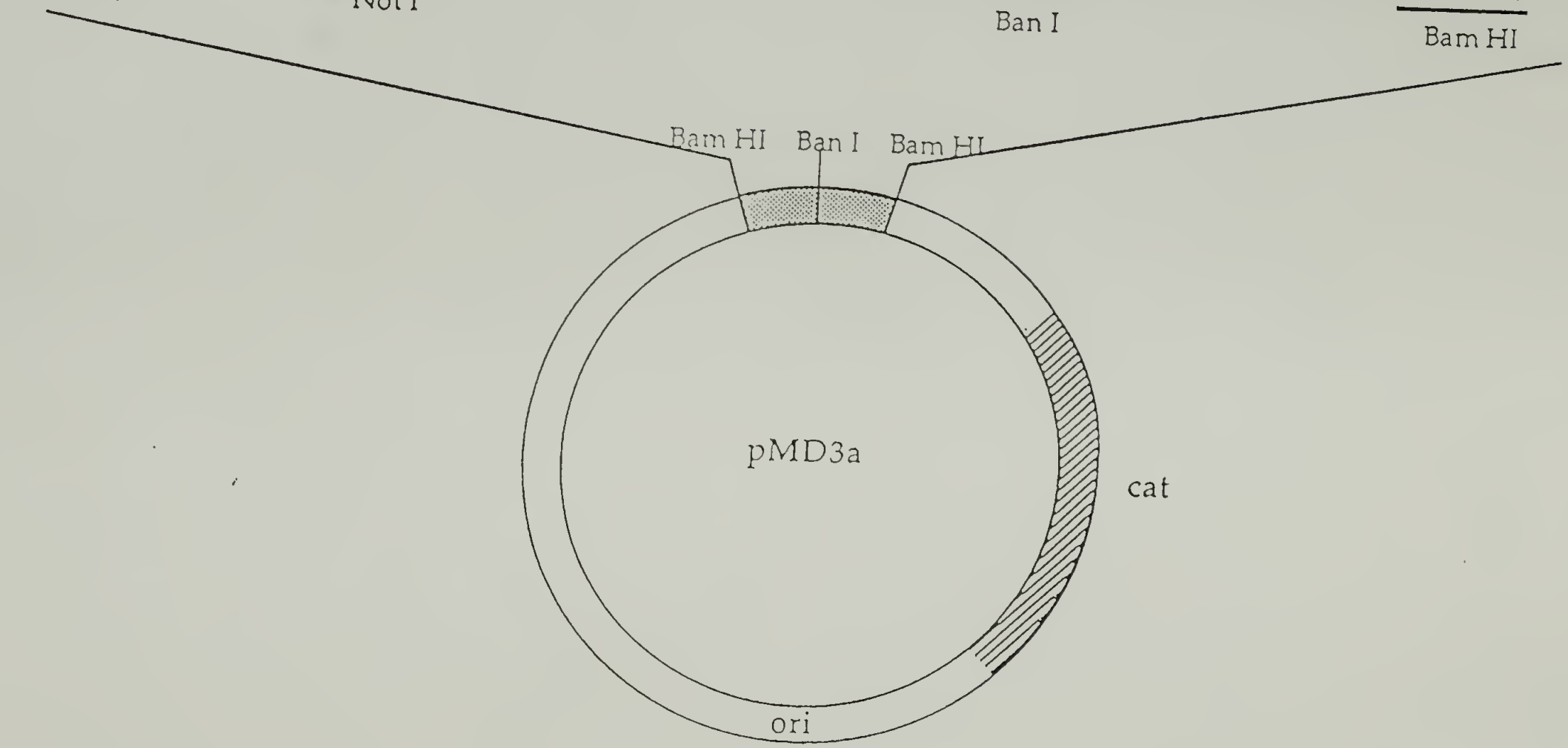
Figure 2.2 Contd.

Figure 2.3 Polylinker region of pMD-3a. The *Ban* I site was used to insert the DNA multimer and the *Bam* H I sites were used to insert the sequence in the pET-3a expression vector. The downstream STOP codon causes the gene product to be a polypeptide with no C-terminal fusion. The introduction of methionine at the N-terminus facilitates the cleavage of the N-terminal fusion with CNBr. The chloramphenicol resistance gene (responsible for the production of chloramphenicol acetyl transferase) aids the selection of cells containing the plasmid. (Adapted from work by Dougherty[4])

Arg Gly Ser Tyr Val Cys Gly Arg Lys Tyr Ser Arg Asp Pro Met Gly Ala Stop Stop

CGT GGA TCC TAT GTT TGC GGC CGC AAA TAT TCT CGC GAT CCG ATG GGT GCC TAA TAA CCC GGG GGA TCC

Bam HI — Not I — Ban I — Bam HI



ligation of these units in the self ligation step and maintains the orientation after insertion into pMD-3a.

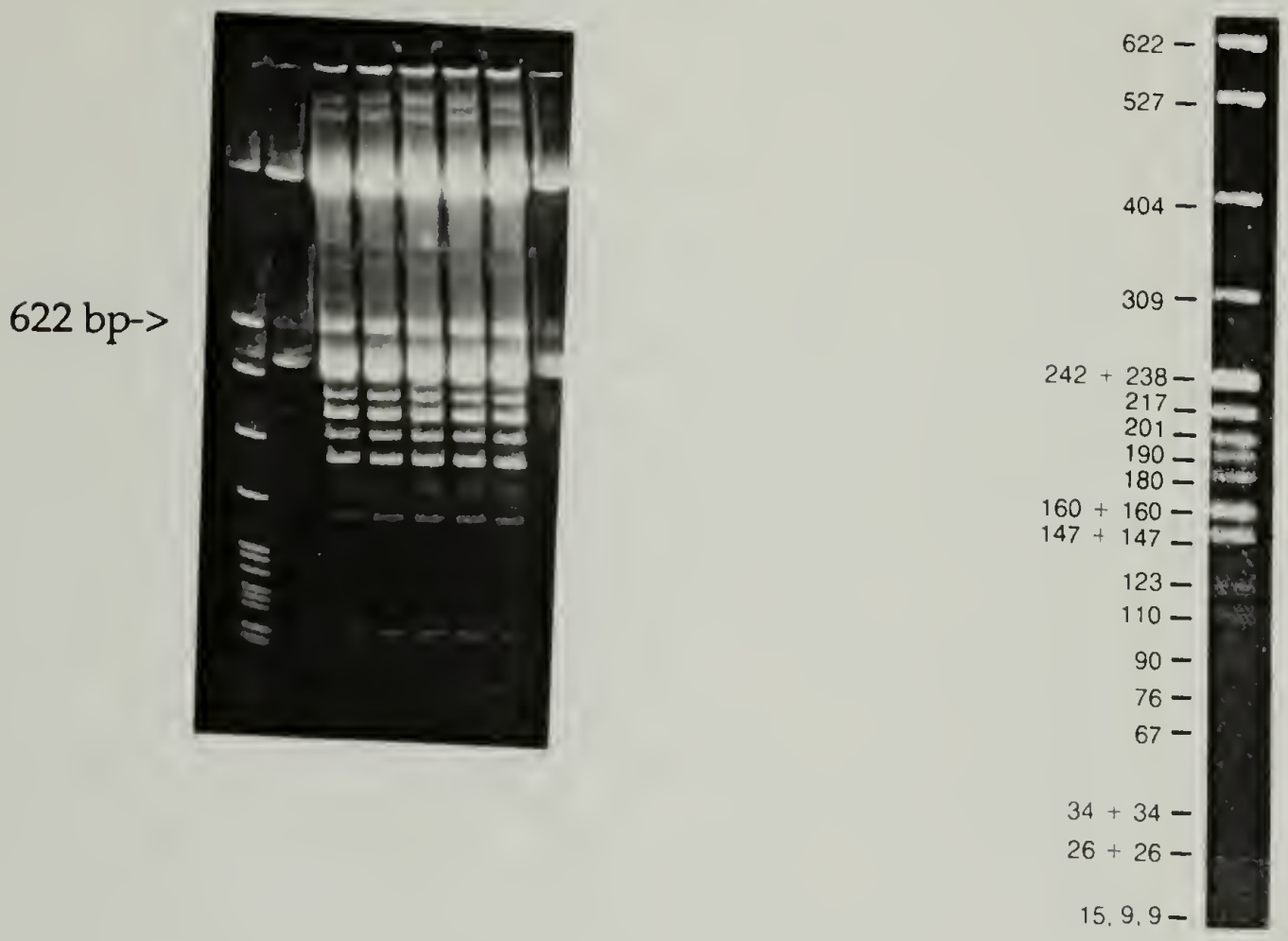
2.4.2 Cloning Strategy

The cloning strategy is shown in Figure 2.1.

The monomers were successfully cloned in pUC18 and at least one stable clone was observed from the sequencing experiments in each case. As described in the Methods Section, these were then ligated and inserted in pMD-3a.

The vector pMD-3a was synthesized by Dougherty[4] by modifying the vector p937.51 from Protein Polymer Technologies. The vector has a unique *Ban*I site which is flanked by two *Bam*HI sites (Figure 2.3). The *Bam*HI sites were used to excise the multimerized insert from the cloning vector and to transfer it to the expression vector (Figure 2.1). An *Ava*I site is present at the downstream end of the linker in pMD-3a, which was used as a diagnostic for orientation.

The multimerized inserts displayed some interesting behavior in regard to their stability in pMD-3a. Larger inserts were notably unstable in this vector and we had no success in cloning multimers in excess of 550 base pairs. Figure 2.4(i) shows a complete *Bam*HI digestion of plasmid DNA from six clones of an 11-mer of **1-m** after subcloning in pMD-3a. The lane adjacent to the molecular weight standard (pBR322 DNA *Msp*I digest)



(i) (ii)

Figure 2.4 (i) *Bam*HI digestion of clones of *E. coli* strain HB101 containing pMD-3a with an 11-mer of **1-m**. This is an analogue of HB101-pAJ-PMD-1 (refer to section 2.3.2.14.1 for nomenclature) with a different size of multimer insert. The first lane on the left is the molecular weight standard (pBR322 DNA *Msp*I digest) which is shown in (ii). The topmost band in the molecular weight standard lane is due to contamination from the second lane; hence, the second band from the top in this lane corresponds to 622 base pairs.

shows a clone with lesser DNA loading than the next five lanes; the strong band at 540 bp corresponds to the 10-mer of **1-m**. In all the sample lanes, a ladder is seen with nine discreet bands lower in length than the 10-mer of **1-m**. No further attempts were made to probe the stability of the insert in pMD-3a; however, it is obvious from Figure 2.4 that the multimer is cleaved to generate discreet molecular weight fragments. Since the clones were colony purified, we can rule out the possibility of simultaneous culturing of multiple clones containing plasmids with varying molecular weight inserts. The instability ceased after the inserts were transferred to pET-3a and pGEX-2T. (The reader might refer to Figure 2.8; in all the sample lanes, only one band is observed in the region below 622 pb.) After screening for inserts we selected a decamer with $\{(AG)_3$ EG $(GA)_3$ EG$\}$ as the repeat, a tetramer with $\{(AG)_3$ EG $(AG)_3$ MG$\}$ as the repeat and a trimer with $\{(AG)_3$ EG $(AG)_3$ VG$\}$ as the repeat. We had poor success in cloning any inserts larger than dimers in the case of $\{(AG)_3$ DG $(GA)_3$ DG$\}$ as the repeat.

As stated in section 2.3.2.11, the ligation ladder in each of the four cases, was divided into two aliquots, one comprising inserts smaller than 15-mer and the other larger than 15-mer; the two fractions were separately ligated into pMD-3a and *E. coli* strain HB101 transformed with the ligation mix as stated in section 2.3.2.12. In all four instances, no viable colonies were obtained when the fraction corresponding to inserts larger than 15-mer was used. In each of the four instances, approximately 100 clones were screened for insert size from the transformation experiments with ligation mixes corresponding to inserts smaller than 15-mer. The sizes of the largest

inserts obtained have been reported above. No further attempts to increase the size of the inserts were made.

2.4.3 Expression of Polypeptides

Two expression systems were used in this work.

2.4.3.1 Expression in pGEX-2T

This system uses the vector pGEX-2T designed by Smith and Johnson[9]. The vector is designed to produce the target polypeptide as the C terminal fusion to glutathione S-transferase, which is a 26 kD protein encoded by the parasitic helminth, *Schistosoma japonicum*. The target protein can be purified by affinity chromatography on immobilized glutathione and can be cleaved from the support by reaction with thrombin or blood coagulation factor X_a (Figure 2.5). The vector has a *tac* promoter (which can be induced with IPTG), and the beta lactamase gene Ap^R which confers ampicillin resistance on the cell. The expression was carried out in *E. coli* DH5αF′. The results of this expression experiment are shown in Figure 2.6. The expression experiment was carried out with four strains. The first four lanes correspond to the control which is induced and produces the glutathione S-transferase fragment without any fusion. The next three sets, of four lanes each, display the induction of GST fusion proteins which correspond to $\{(AG)_3\ EG\ (GA)_3\ EG\}_{10}$ (GST-1; molecular weight =26.0+11.4 kD), $\{(AG)_3\ EG\ (AG)_3\ MG\}_4$ (GST-8; molecular weight =26.0+4.4kD) and $\{(AG)_3\ EG\ (AG)_3\ VG\}_3$ (GST-7; molecular weight =26.0+3.4kD). Though the migration of these in the gel is dominated by

Thrombin

pro lys ser asp leu val pro arg gly ser pro gly ile his
CCA AAA TCG GAT CTG GTT CCG CGT GGA TCC CCG GGA ATT CAT ....

BamHI EcoRI

AvaI

→

Direction of translation

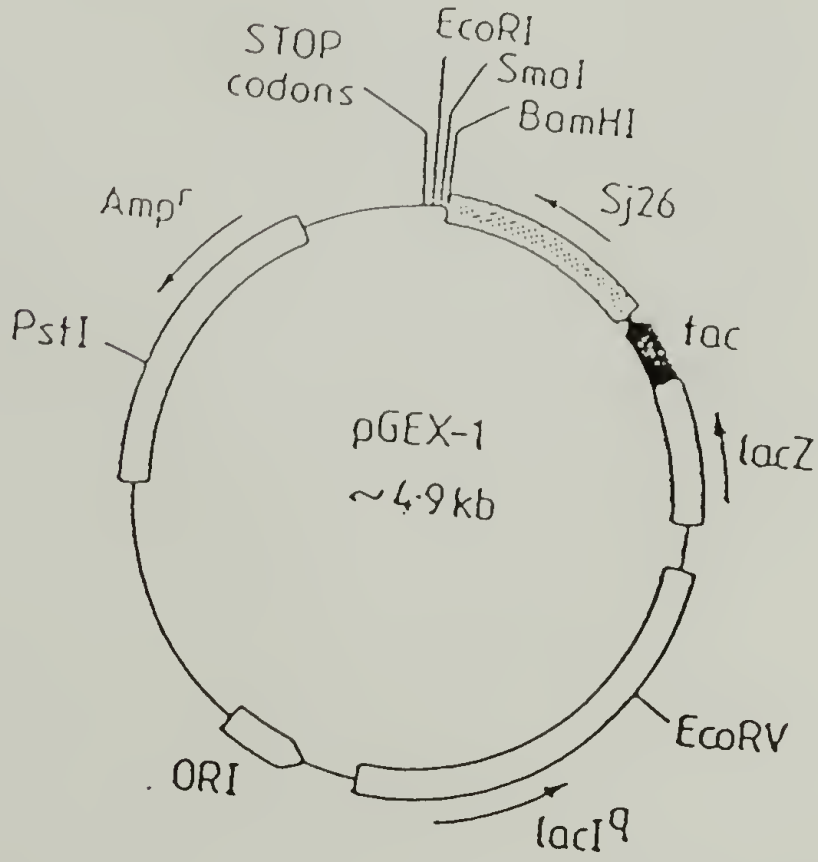
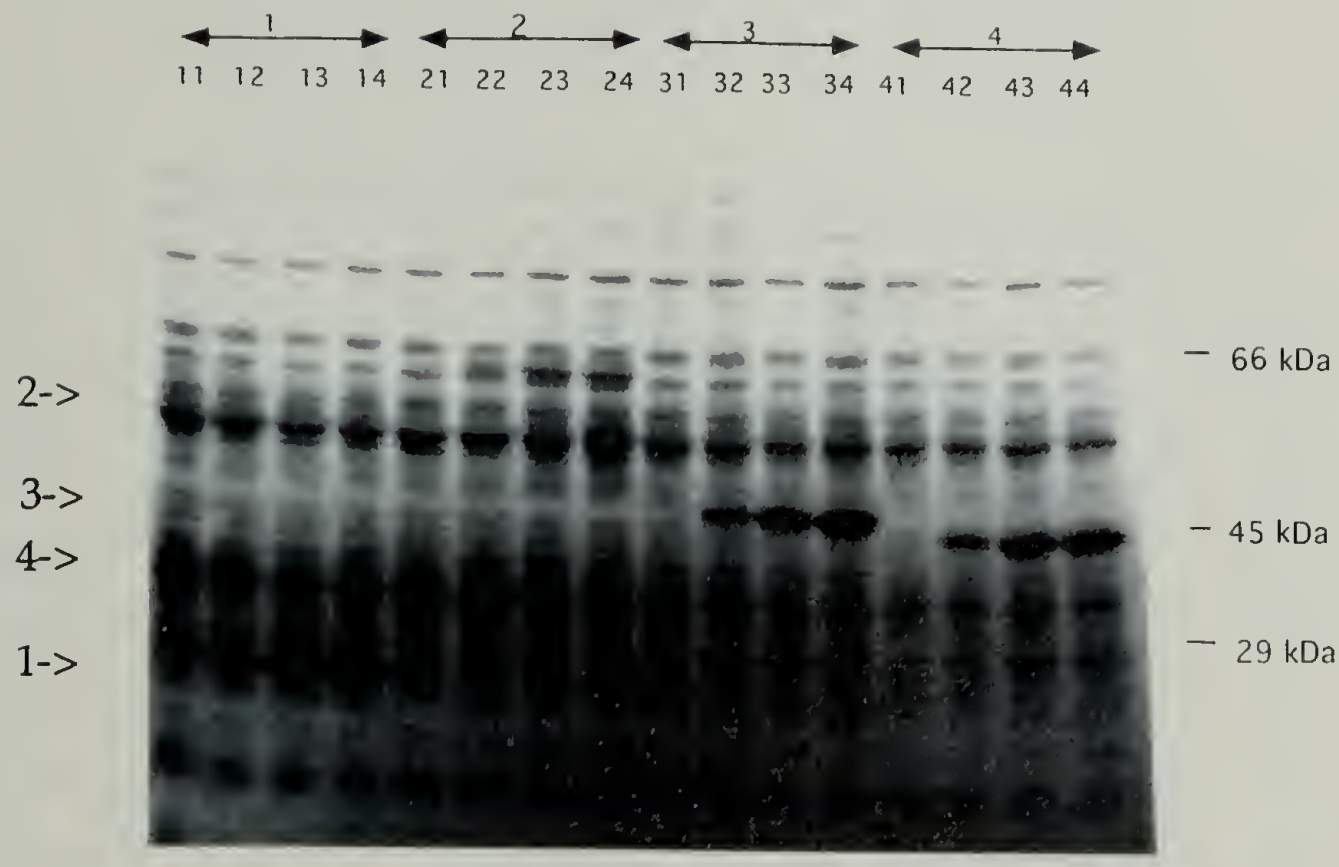


Figure 2.5 Cloning site in pGEX-2T. Adapted from work by Smith and Johnson[9].

Figure 2.6 The results of the expression experiment in the pGEX-2T expression system. The cell extracts were analyzed by SDS-PAGE and stained with Coomassie Brilliant Blue. In all four sets of four lanes each, the first lane represents data at 5 min, the second at 20 min, the third at 90 min and the fourth at 110 min after induction. Lanes 11 through 14 correspond to the positive control. The positive control is the glutathione S-transferase (GST) fragment without any fusion. Note the temporal accumulation of a protein at 26 kDa. Lanes 21 through 24 which correspond to the expression of GST fused with $\{(AG)_3EG(GA)_3EG\}_{10}$ display the accumulation of a protein at approx. 66 kDa. Lanes 31 through 34 which correspond to the expression of GST fused with $\{(AG)_3EG(AG)_3MG\}_4$ display the accumulation of a protein at approx. 46 kDa. Lanes 41 through 44 which correspond to the expression of GST fused with $\{(AG)_3EG(AG)_3VG\}_3$ display the accumulation of a protein at approx. 43 kDa.

In order to aid the reader in locating the induced bands, the figure has been labeled with 1, 2, 3 and 4 in direct correspondence with the first, second, third and fourth set of lanes.


1
2
3
4
11 12 13 14 21 22 23 24 31 32 33 34 41 42 43 44
66 kDa
45 kDa
29 kDa
2->
3->
4->
1->

the characteristics of migration of the GST fragment, one can readily discern the various molecular weights.

2.4.3.2 Expression in pET-3a

We also carried out expression experiments in the pET system developed by Studier[10]. The vector in frame with the inserts was pET-3a. The pET vectors have an Ap^R gene which confers ampicillin resistance as a marker and a T7 promoter upstream of the *Bam*HI site. It also has a terminator downstream of the cloning site. The promoter, which is very strong, promotes the transcription of a coat protein in the T7 phage. The synthesis of the T7 RNA polymerase is under the control of a lacUV promoter which is inducible by IPTG. The expression is carried out in *E. coli* BL21(DE3)pLysS. The lacUV promoter is leaky and produces a basal amount of T7 RNA polymerase in the absence of IPTG. This is counteracted by the production of lysozyme, coded on the pLysS plasmid, which acts as an inhibitor for the RNA polymerase; this reduces leaky expression of the polypeptide. After induction, however, the concentration of the polymerase increases and renders the inhibition ineffective.

We tried the expression of all the three constructs, as in the other expression system, in addition to the control. The auto radiogram for this experiment is shown in Figure 2.7. As has been stated before, the target protein has a 25 amino acid N terminal fusion (FUS) and no fusion at the C terminus because of the stop codons engineered in the linker region in pMD-3a (Figure 2.3). The migration of the polypeptides is slower than

expected; similar migration behavior has been observed for similar proteins at this laboratory. The first four lanes correspond to the positive control which is a protein of approximately 17 kD. The next two sets, of four lanes each, show the expression results for the strains encoding $\{(AG)_3EG\ (GA)_3\ EG\}_{10}$ (molecular weight =2.5+10.8 kD) and $\{(AG)_3E\ (AG)_3\ MG\}_4$ (molecular weight =2.5+4.4kD). The strain encoding $\{(AG)_3\ EG\ (AG)_3\ VG\}_3$ (molecular weight =2.5+3.4kD) did not show any evidence of expression.

In order to check the size of the DNA multimers and ascertain their stability, *Bam*HI digestion products of pAJP-3a-1, pAJP-3a-7 and pAJP-3a-8 were analyzed on an 8% acrylamide gel. Figure 2.8(i) shows the results of the *Bam*HI digestion; the first lane corresponds to the molecular weight standard (pBR322 DNA *Msp*I digest); the next ten lanes (2 through 11) correspond to clones corresponding to pAJP-3a-8, the next eleven (12 through 22) to pAJP-3a-7 and the last six (23 through 28) to pAJP-3a-1. Lanes 2 and 3 have been contaminated with the molecular weight standard, but 5 and 10 show the presence of an insert which corresponds to a tetramer of 8. Lanes 13 through 21 show the presence of an insert corresponding to a trimer of 7. Lanes 23, 27 and 28 show the insert corresponding to the decamer of 1. In all these instances, the orientation was confirmed by *Ava*I digestion. The polypeptide which accumulates after induduction in expression experiments with BL21(DE3)pLysS-pAJP-3a-8 is seen to migrate at approximately 29 kD; the reasons for this anomalously slower migration are unclear.

The expression results with BL21(DE3)pLysS-pAJP-3a-1 (refer to Figure 2.7) show accumulation of a two other proteins in addition to the prominant accumulation at approximately 29 kD; these bands can be seen in Figure 2.7 at approximately 10 kD; no further experiments were carried out to determine the composition of these proteins. The extraneous bands are not observed in experiments with BL21(DE3)pLysS-pAJP-3a-8; hence, they appear to be specific to the expression of BL21(DE3)pLysS-pAJP-3a-1. The combustion analysis, amino acid analysis and ^{1}H NMR show **1** to be very pure (*vide infra*). Hence, these extraneous bands could be a result of sequence specific *in vivo* degradation of **1**.

2.4.4 Purity Analysis of **1**

For further analysis, polypeptide **1** was selected because of its size and the sequence correlation to **2**. The pET expression system was selected in the large scale synthesis due to the following reasons: 1) The gene constructs used in this work were such that the polypeptides could be synthesized in the pET expression system with very short leader and no tail amino acid sequence; cleavage of the precursor polypeptides therefore yielded the desired polypeptides with only slight contamination by other amino acids; the pGEX-2T expression system did not provide this advantage. 2) The purification of the polypeptide from the pET expression system could be carried out by gradual titration of the glutamic acid residues (*vide infra*); on the contrary, the pGEX-2T expression system required use of expensive glutathione colomns for purification; the pGEX-2T system indeed offers advantages for purification of polypeptides which

cannot be denatured during purification but this was not a constraint in the work reported in this thesis. 3) Finally, the yields in the pET expression system are more than those in the pGEX-2T expression system because the T7 promoter is stronger than the *tac* promoter[9,10].

The fusion protein was expressed on a large scale in BL21(DE3)pLysS and was purified and cleaved as discussed in section 2.3.2.18. The yield of the protein was estimated at 40 mg/L. The purity of the polypeptide was confirmed by combustion analysis, ^{1}H NMR and amino acid analysis. Table 2.1 shows the amino acid analysis of the purified crystallized polypeptide; the protocol described in Chapter 3 was used for crystallization. Table 2.2 shows the combustion analysis of the crystallized polypeptide. Tables 2.3 and 2.4 show the analysis of the amorphous purified polypeptide. The crystalline sample shows an additional enrichment of the desired amino acids. Figure 2.9 shows the TGA for crystallized **1**. The TGA for the sample which was dried at room temperature for 20 h shows the presence of 3-4% water in the sample; this sample was used in the combustion and amino acid analyses. The combustion analysis is consistent with the presence of 3.5% water. The ^{1}H NMR of the amorphous sample is shown in Figure 2.10. All these analyses show that the level of impurities in the sample is very low.

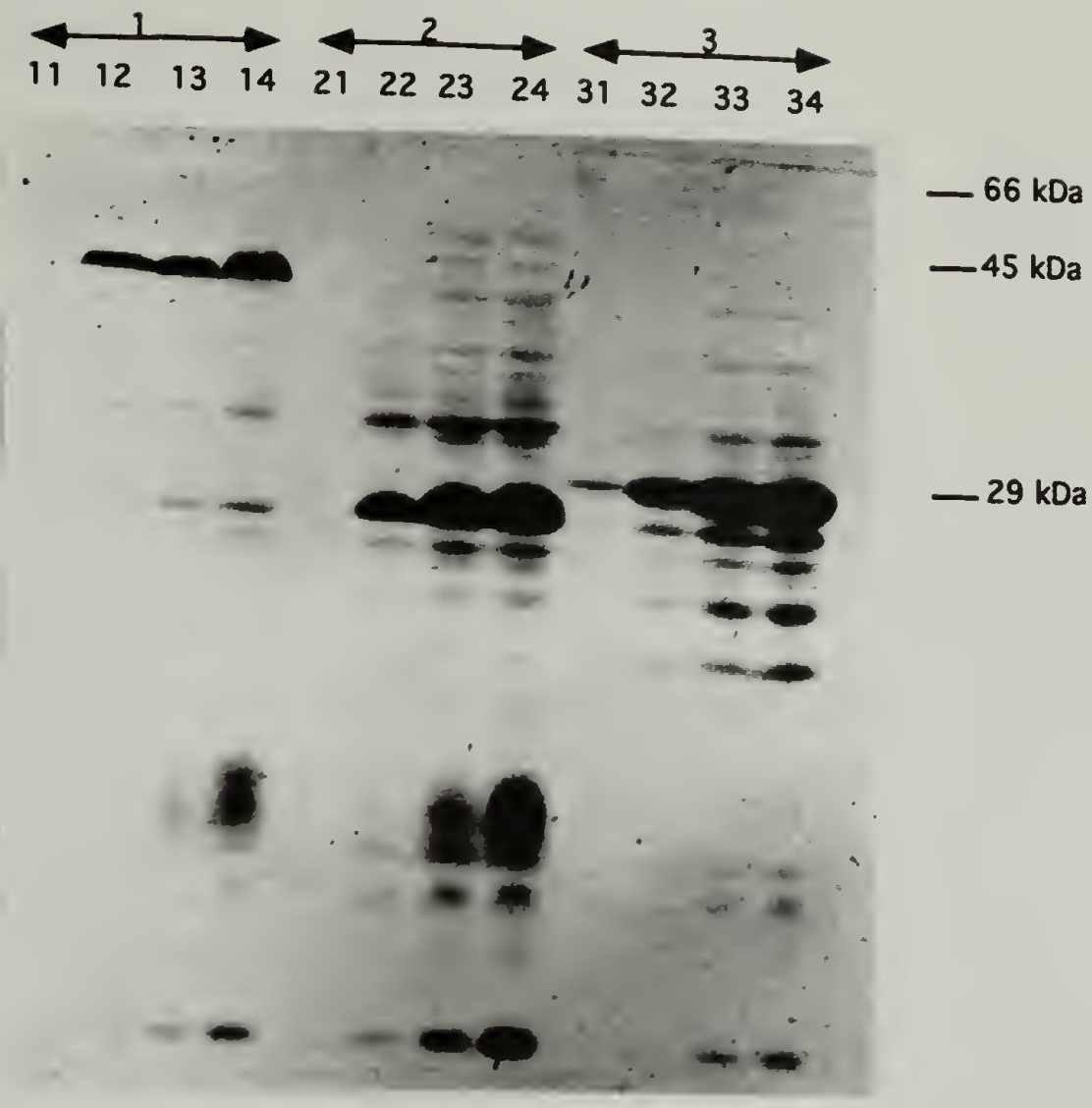
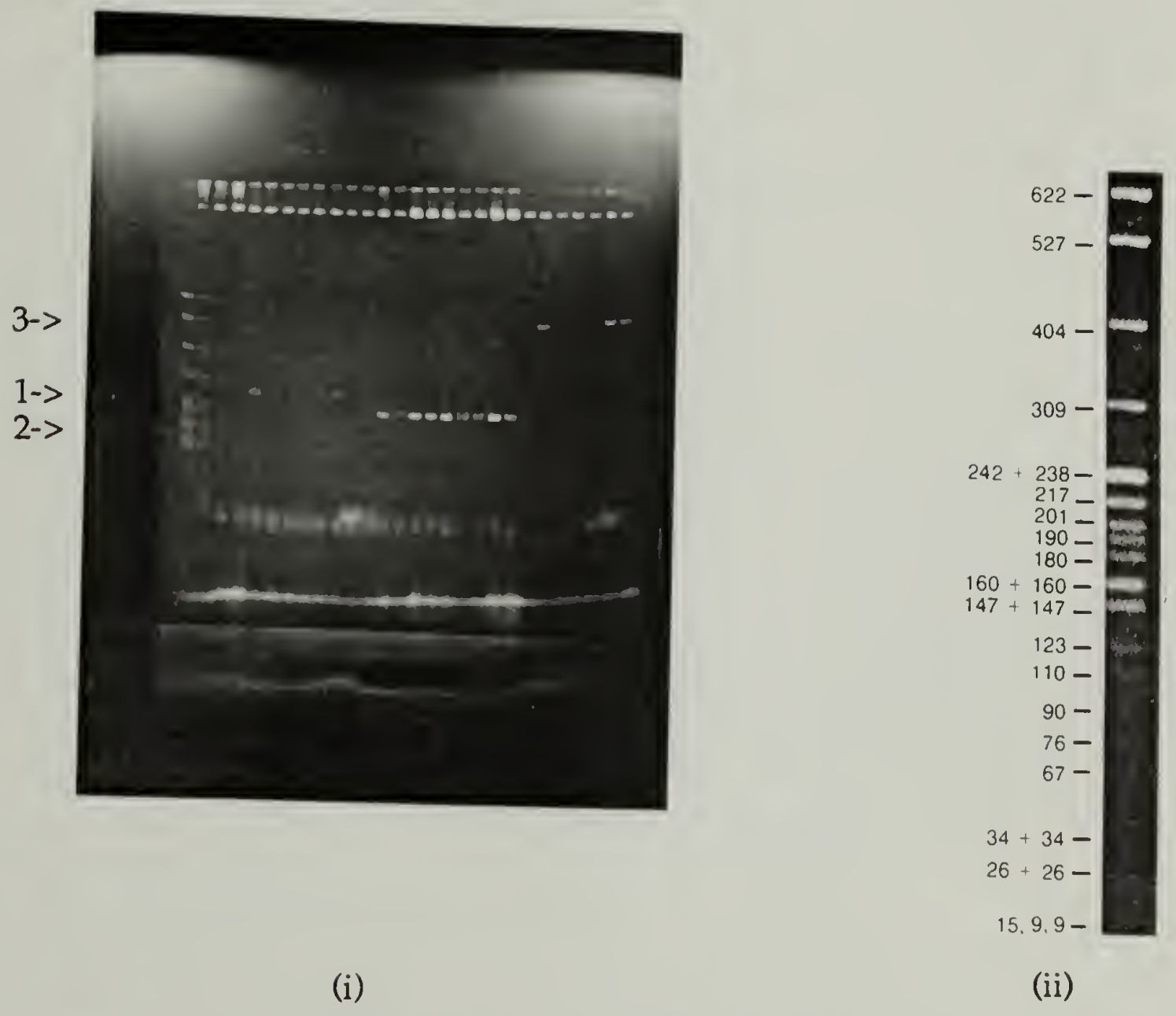


Figure 2.7 The results of the expression experiment in the pET system. The quantity of the protein was visualized by *in vivo* labeling with {^{3}H}glycine followed by autoradiography. The cell extracts were analyzed on a denaturing SDS 15% acrylamide gel. In all three sets of four lanes each, the first lane represents data at 5 min, the second at 20 min, the third at 90 min and the fourth at 110 min after induction. Lanes 11 through 14 correspond to the positive control; note the temporal accumulation of a protein at 44 kDa. Lanes 21 through 24 which correspond to the expression of $\{(AG)_3EG(GA)_3EG\}_{10}$ display the accumulation of a protein at approx. 29 kDa. Lanes 31 through 34 which correspond to the expression of $\{(AG)_3EG(AG)_3MG\}_4$ display the accumulation of a protein at approx. 31 kDa.



(i) (ii)

Figure 2.8 *Bam*HI digestion of inserts in pET-3a; the first lane correspond to the molecular weight standard (pBR322 DNA *Msp*I digest); the next ten lanes (2 through 11) correspond to clones corresponding to pAJP-3a-8, the next eleven (12 through 22) to pAJP-3a-7 and the last six (23 through 28) to pAJP-3a-1. (ii) Molecular weight standard (pBR322 DNA *Msp*I digest).

Multimers in the plasmids pAJP-3a-8, pAJP-3a-7 and pAJP-3a-1 have been marked in the figure as 1, 2 and 3 respectively.

Table 2.1

Amino acid compositional analysis of crystallized **1**

Amino acid[a]	Expected mole %	Observed mole %
Glutamic acid	12.5	12.08
Glycine	50.0	50.62
Alanine	37.5	37.30
NORL[b]	Internal Std.	Internal Std.

[a] All other amino acids were below the threshold. The amino acid analysis was carried out at the MCB core facility at the University of Massachusetts, Amherst on an Applied Biosystems 420A/130A PTC separation unit. The absorption threshold was set at 2000 μAU. The highest peak (glycine) observed with this setting was 52284 μAU. Hence, no amino acid contamination less than 3.8% of glycine could be detected.

[b] NORL is the internal calibration standard

Table 2.2

Elemental analysis of crystallized **1**

Element	Theoretical %	Theoretical % (4% Water)	Theoretical % (3.5% Water)	Observed %
Carbon	46.23	44.38	44.6	44.70
Hydrogen	6.01	6.20	6.18	5.99
Nitrogen	19.60	18.80	18.90	18.41

Table 2.3

Amino acid compositional analysis of amorphous **1**

Amino acid[a]	Expected mole %	Observed mole %
Glutamic acid	12.5	12.26
Glycine	50	47.35
Arginine	0	3.31
Alanine	37.5	31.33
Proline	0	4.38
Lysine	0	1.37
NORL[b]	Internal Std.	Internal Std.

[a] All other amino acids were below the threshold. (refer to Table 2.1)

[b] NORL is the internal calibration standard used

Table 2.4

Elemental analysis of amorphous **1**

Element	Theoretical %	Theoretical % (4% Water)	Theoretical % (3.0% Water)	Observed %
Carbon	46.23	44.38	44.7	45.11
Hydrogen	6.01	6.2	6.1	6.12
Nitrogen	19.6	18.8	19.1	17.76

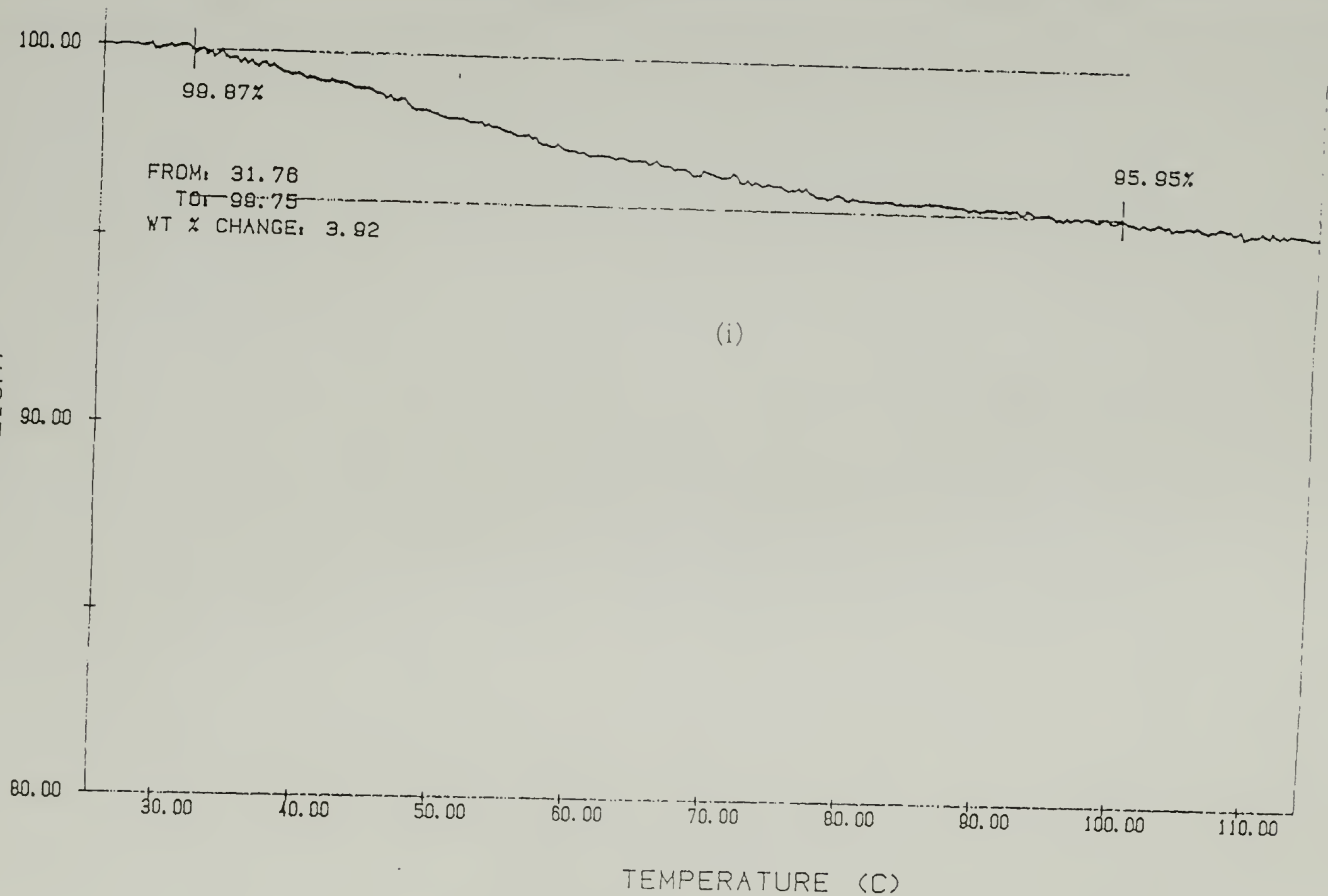


Figure 2.9 Thermogravimetric analysis of **1**. (i) and (ii) Amorphous sample dried at RT for 20 h. (iii) Amorphous sample dried at 80 °C for 12 h.

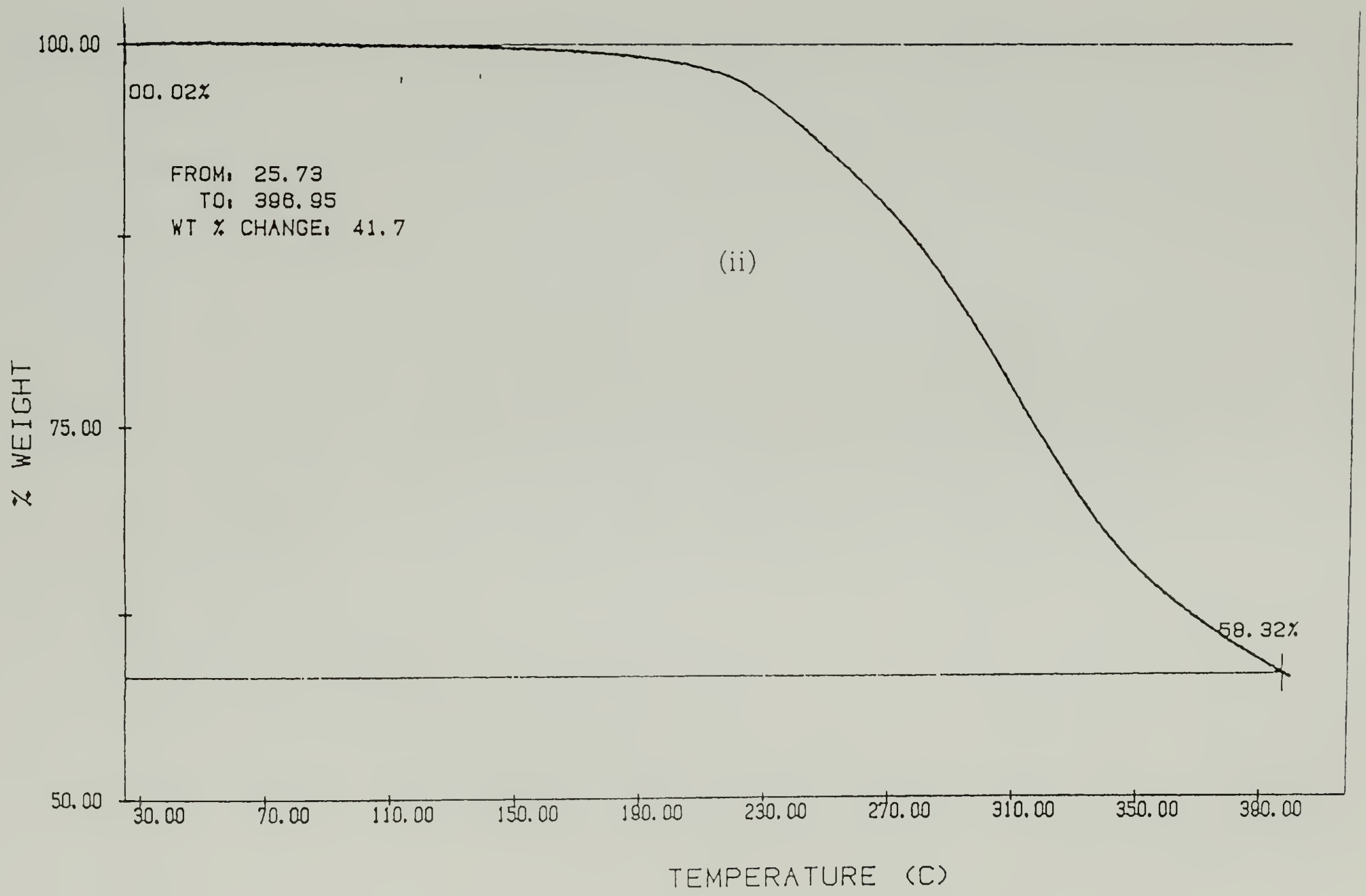


Figure 2.9 (contd)

Continued, next page

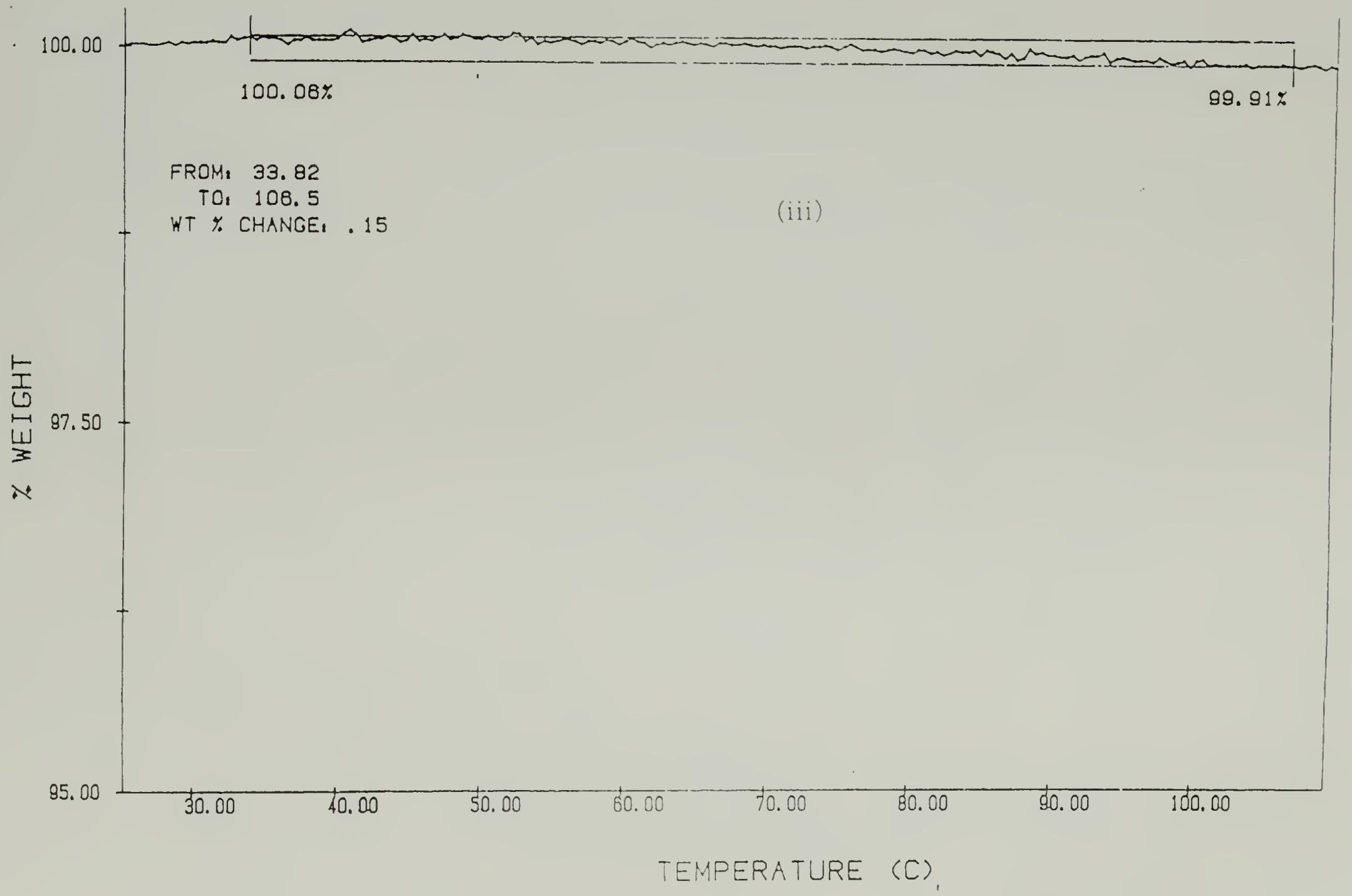
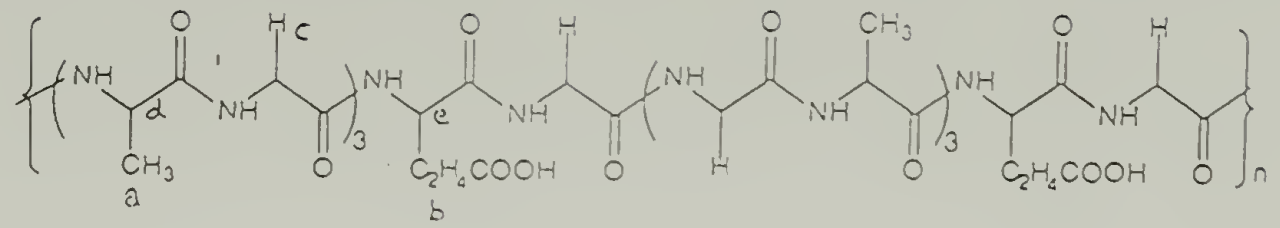


Figure 2.9 (contd)



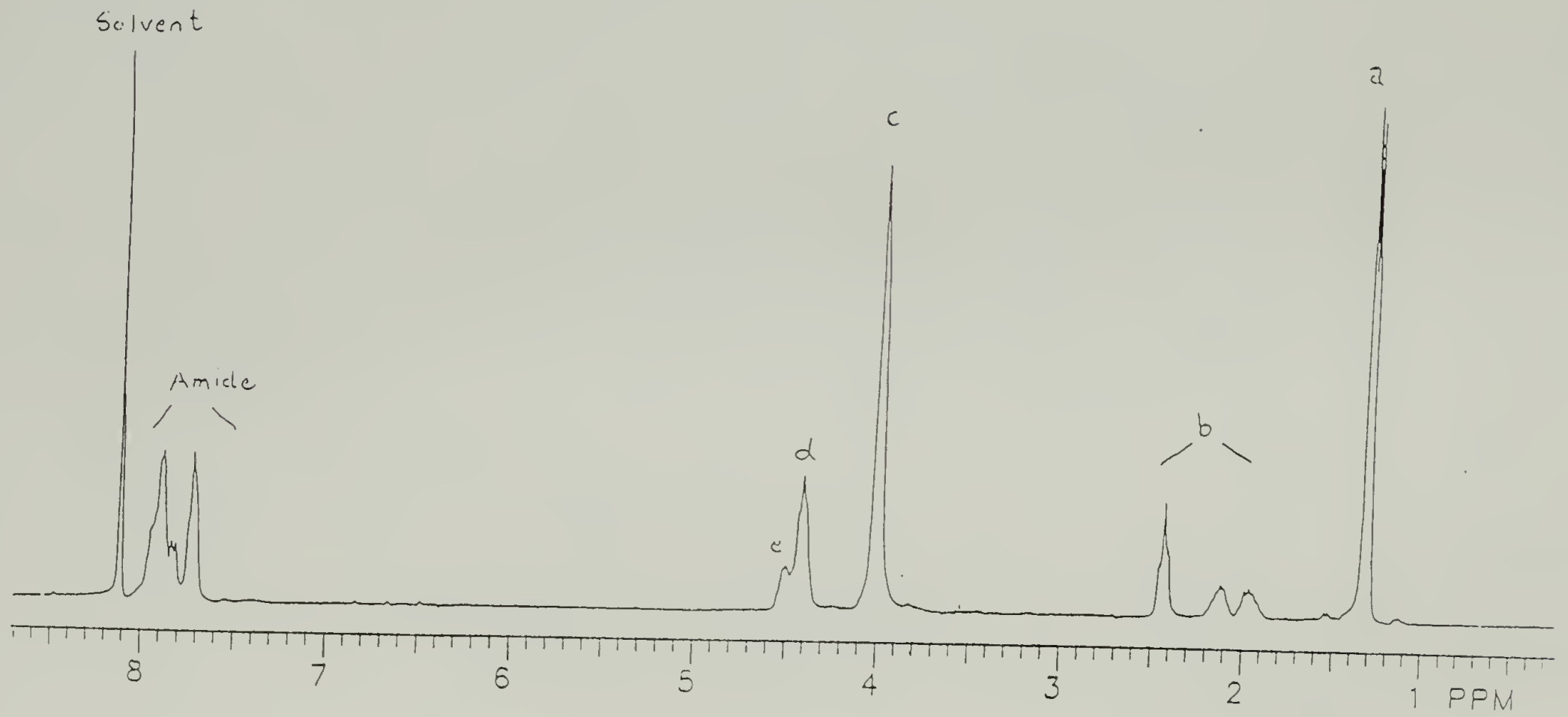


Figure 2.10 The 300 MHz. ^1H NMR spectrum of **1** in deuterated formic acid.

2.4.5 Thermal Analysis of 1

Figure 2.9 (ii) shows the polypeptide to begin degrading at approximately 180 °C. Before carrying out differential scanning calorimetry on the amorphous sample, it was dried at 80 °C for 12 h over P_2O_5 in order to remove residual moisture. Figure 2.11 shows the DSC thermogram for the first heating cycle from 25 °C to 140 °C at 20 °C/min. A glass transition is seen at 60 °C and a broad first order transition is observed at 111 °C; the transition appears to be due to the vaporization of molecules of water bound to the polypeptide. The amount of bound water estimatedfrom the endotherm at 111.5 °C is 0.05 wt% from the DSC; this is in fair agreement with the amount observed in the TGA measurements (Figure 2.9 (iii)). Both the glass transition and the melting transition are not observed in the cooling cycle and the subsequent heating cycles; however, the only heating scans reported here are from 40 °C to 140 °C. The glass transition could be explained by the plasticizing effect of the residual water which is lost at 111 °C; since this water is lost in the first heating cycle, the sample is no longer plasticized in the subsequent cycles and therefore does not show any thermal transitions.

We cannot rule out the possibility of a glass transition temperature at temperatures above 140 °C; true second order thermal transitions have been observed at approximately 175 °C for 4[11] (refer to section 1.4). DSC experiments were carried out on 1 with heating scans from 40 °C to 190 °C; no second order thermal transitions were observed; however, inspection of the sample after the experiment showed a change in color from

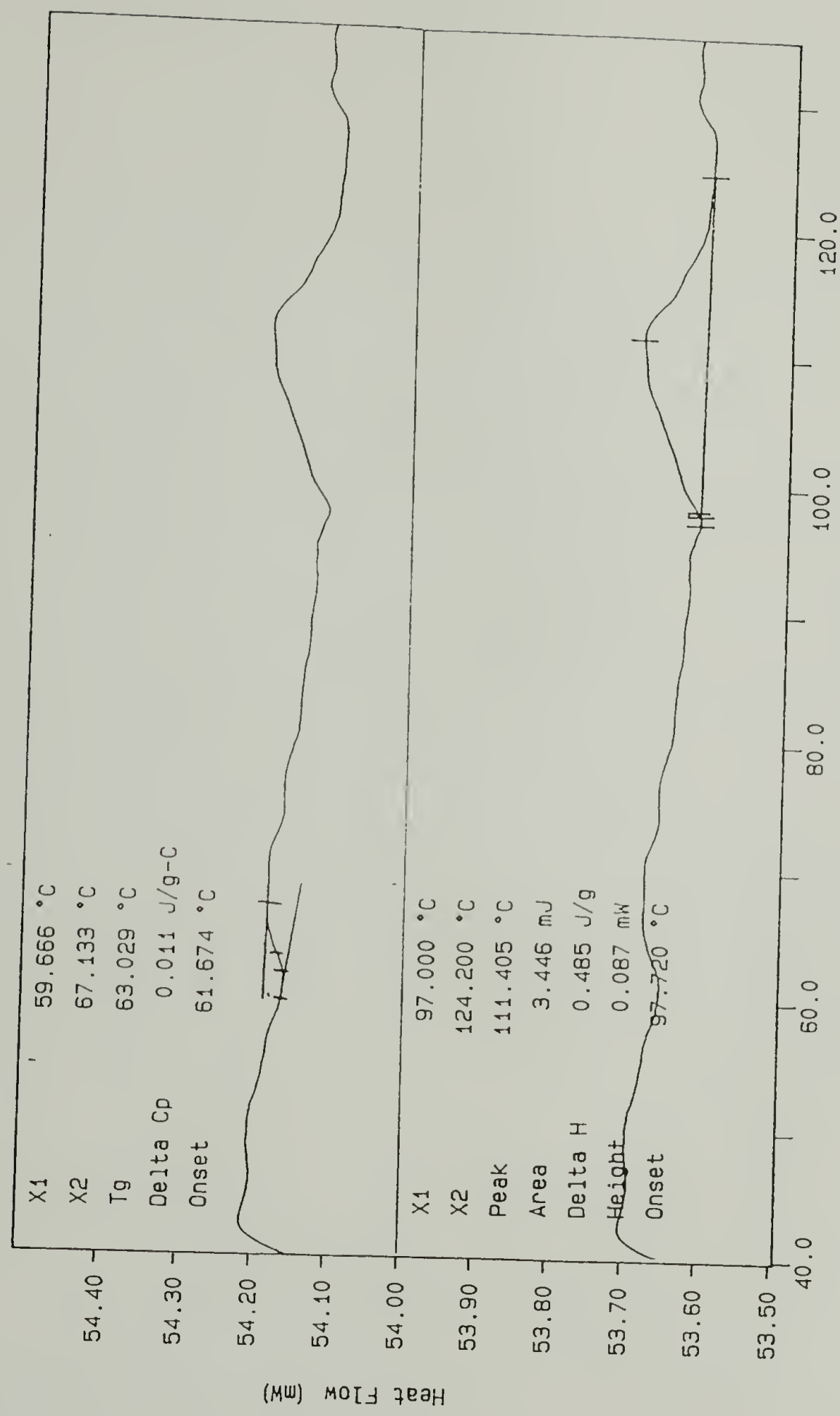


Figure 2.11 DSC thermogram for 1. First heating cycle at 20 °C/min from 25 °C to 140 °C.

white to light brown; the DSC also showed a gradually increasing endotherm indicative of degradation Hence, this data cannot be considered reliable in determination of a glass transition temperature.

2.5 Conclusion

The DNA corresponding to **1**, **6**, **7** and **8** was successfully cloned in pUC18. The DNA was multimerized in each case and **1**, **7** and **8** were successfully expressed in *E. coli*. The polypeptide **1** was selected for large scale expression and was further analyzed for purity and composition. The amino acid analysis and the elemental analysis are consistent with a protein rich in alanine, glycine and glutamic acid. The yield of the protein was estimated at 40 mg/L. The results of the expression experiment show the induction and accumulation of a protein with a uniform molecular weight.

2.6 References

(1) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning -A Laboratory Manual;* 2nd ed.; Cold Spring Harbor Laboratory Press: New York, 1989.

(2) Sanger, F.; Niklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci.* **1977**, *74*, 5463-5467.

(3) McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.

(4) Dougherty, M. J. *Univ. of Mass.,Amherst* **1993**, *Ph. D. Thesis*

(5) Ikemura, T. *Mol. Biol. Evol.* **1985**, *2(1)*, 13-34.

(6) Sharp, P. M.; Li, W. H. *Nucleic Acids Res.* **1986**, *14*, 7737-7749.

(7) Andersson, S. G. E.; Kurkland, C. G. *Microbiol. Rev.* **1990**, *54 (2)*, 198-210.

(8) Krejchi, M. T. *Univ. of Mass., Amherst* **1993**, *Ph. D. Thesis*

(9) Smith, D. B.; Johnson, K. S. *Gene* **1988**, *67*, 31-40.

(10) Rosenburg, A. H.; Lade, B. N.; Chui, D.; Lin, S.; Dunn, J. J.; Studier, F. W. *Gene* **1987**, *56*, 125-135.

(11) Creel, H. S. *Univ. of Mass., Amherst* **1993**, *Ph. D. Thesis*

CHAPTER 3

DETERMINATION OF CRYSTAL STRUCTURE BY X-RAY DIFFRACTION

The rationales for the synthesis of the polypeptides have been outlined briefly in previous chapters. In this chapter the intention is to present an introduction to the crystal structures observed for related polypeptides, to analyze the solid state structure of **1** by x-ray diffraction and place it in the context of previous work. In this regard, **2** is the most important.

3.1 Introduction

3.1.1 Crystal Structure of Silk and its Analogues

The propensity for silk to exist in the antiparallel beta sheet structure was first shown by Pauling and coworkers[1]. Silk derived from *Bombyx mori* is known to be rich in the amino acid residues alanine, glycine and serine[2]. The repetitive part of the polypeptide can be well represented by $(Ala\text{-}Gly\text{-}Ala\text{-}Gly\text{-}Ser\text{-}Gly)_n$, although imperfections in the sequence are known to exist. Lotz and coworkers[3] have characterized its crystalline morphology and concluded that it can exist either in the antiparallel beta pleated sheet structure or the silk I structure. The former is a well characterized structure with respect to dihedral angles and crystallographic spacings. They have also characterized periodic copolypeptides of alanine and glycine with different repeat sequences of alanine and glycine, viz. Ala_2Gly, AlaGly, Ala_2Gly_2, Ala_3Gly_3, Ala_2Gly_3, $AlaGly_2$, and $AlaGly_3$; the most favored

crystalline structure is antiparallel beta sheet in all these instances with varying intersheet spacings; the chain direction in oriented samples of these silk analogues is predominantly in the direction of the fiber. The diffraction patterns of poly(alanyl-glycine) have been indexed to an orthorhombic unit cell containing polypeptide chains in an antiparallel beta sheet conformation[4]. Geddes et al[5] have studied the morphology of a silk rich in serine (41%), alanine (20 %) and glycine (23 %) obtained from *Chrysopa flava*, a green lace-wing fly; the studies indicate that the silk exists in a cross beta strucuture with the chain direction orthogonal to the fiber direction; ribbons ca. 25 Å thick are formed due to periodic folding of the polypeptides.

3.1.2 Crystal Structure of Repetitive Polypeptides Containing Alternate 'Stem' and 'Turn' Sequences

Three sets of polypeptides have been synthesized by previous workers at this laboratory. All these polypeptides were designed such that the chains would fold periodically and the structure formed in between the folds would be antiparallel beta sheet; this required repetitive sequences of amino acids comprising a 'stem' alternating with a 'turn' sequence. In all three sets, the selection of residues in the 'stem' region was based on studies of silks and copolymers of alanine and glycine. A more detailed account of the design rationale has been presented in Chapter 1.

3.1.2.1 Poly{$(AG)_n$PEG}[6,7]

The selection of the tripeptide- PEG- in the turn region was due to studies by Chou and Fasman[8]. Based on assignments of probability for proline to reside in turns and serve as a beta sheet breaker[8,9], it was anticipated that proline would initiate a turn in the polypeptide. This series was synthesized with n=3,4,5,6. It was observed that the polypeptide with n=3 did not crystallize under the processing conditions employed. The propensity to crystallize increased as the value of n increased from 4 to 6[6]. The spacings (from x-ray diffraction experiments) observed in the crystalline domains correspond to beta sheet morphology.

3.1.2.2 Poly {$(AG)_n$EG} [10,11]

In this series of polypeptides, crystalline domains were observed when n=3. In addition to the short range order which was observed by wide angle X-ray diffraction, order was also observed in the range of 35 Å. This would be the characteristic thickness of the lamellae if the chains folded at every turn position. Furthermore, the wide angle diffraction patterns could be indexed on an orthorhombic unit cell and showed the presence of a very strong 010 and a strong 020 reflection (refer to section 3.3.1 and Figure 3.4 for a detailed discussion of the unit cells). This would correspond to the formation of beta sheets decorated with the methyl groups (of the alanine residue) only on one side of the sheet; packing of these polar sheets would lead to an assembly corresponding to a

periodicity of approximately 10 Å along the *b* direction. Similar behavior was observed in the wide angle patterns of polypeptides with n=4, 5, 6.

3.1.2.3 Poly {$(AG)_3DG$}, Poly {$(AG)_3VG$}, Poly {$(AG)_3MG$} and Poly {$(AG)_3NG$}[12]

These polypeptides along with other type I′ and type II′[13] turn forming residues at the turn position were synthesized to observe the effect of turn composition on the turn efficiency. The structural analysis of poly {$(AG)_3DG$} and poly {$(AG)_3MG$} also showed the formation of beta sheet structures.

3.1.3 Importance of Structural Analysis of **1**

The studies discussed above show that the unit cells of crystalline polypeptides containing large fractions of alanine and glycine are very similar and the structures are predominantly antiparallel beta sheets when appropriately processed. The dimensions of the unit cells along the *a* and *c* axes are identical but there is some difference in the *b* direction (Figure 1.6) depending on the groups decorating the sheet surface. Furthermore, the ability to crystallize is itself dependent on the sequence of the polypeptide.

In the context of previous studies at this laboratory regarding periodic chain folding, the structural analysis of **1** was particularly useful due to the inversion of alternate alanyl-glycyl diads in comparison with **2**. None of the polypeptides synthesized prior to this study contained

alternating inverted diads sequences. We were interested in ascertaining the formation of beta sheets in crystalline **1** and in determining the unit cell dimensions. Formation of a turn similar to that proposed for **2** at the glutamic acid residue would cause formation of sheets which were decorated with methyl groups (of alanine) evenly on both sides; these apolar sheets would then be anticipated to pack in the solid state with a periodicity of approximately 5 Å along the *b* direction; the corresponding wide angle diffraction patterns would show an absence of 0k0 reflections when the value of k was odd. The presence of all 0k0 reflections in the solid structure of **2** and the systematic absence of 0k0 reflections when k was odd in the structure of **1** would support the formation of periodic turns leading to the formation of polar and apolar sheets,respectively, for the two polypeptides which differed from each other due to diad inversions in **1**.

The number of repeats of the alanyl-glycyl diad was three for **1** and **2**. In the event of the formation of a lamella as shown in Figure 1.6 , the thicknesses would therefore be indentical for both the polypeptides. It was also our intention to address this issue.

3.2 Methods

3.2.1 Sample Preparation

Two kinds of sample preparations were employed to obtain two different types of morphologies.

3.2.1.1 Amorphous Preparation

The purification of **1** has been described in 2.4.2.18. The polypeptide **1** was amorphous after precipitation by ethanol.

3.2.1.2 Crystalline Preparation

In the second type of sample preparation the sample was dissolved in 70% formic acid at a protein concentration of 40 mg/ml and stirred for ca. 24 h at ca. 700 rpm on a Thermix stirrer model 220T (Fisher Scientific). The sample formed a gel which was precipitated by adding excess water. This was sedimented by centrifugation and washed several times with water and then with methanol. Finally the gel was resuspended in methanol. This was spun down at 13500 x g for 10 min, the methanol decanted and the gel dried in the presence of P_2O_5 at 78 °C overnight for use in powder diffraction experiments.

Oriented diffraction patterns were obtained by carrying out diffraction experiments on oriented 'mats' of semicrystalline polypeptides. In order to obtain oriented wafers, the wet gel obtained after centrifugation of the suspension in methanol was placed between glass slides covered with teflon filters; the plates subsequently were clamped and the oriented gel was allowed to dry in air at room temperature. The oriented thin wafers were peeled off from the teflon filters and a few of these were stacked on top of each other between two pieces of teflon filters; the thin wafers were fused into a thicker 'mat' by placing a drop of 50% formic acid

on the filter, clamping the assembly between glass plates and drying the mat at room temperature.

3.2.2 X-ray Diffraction Analysis

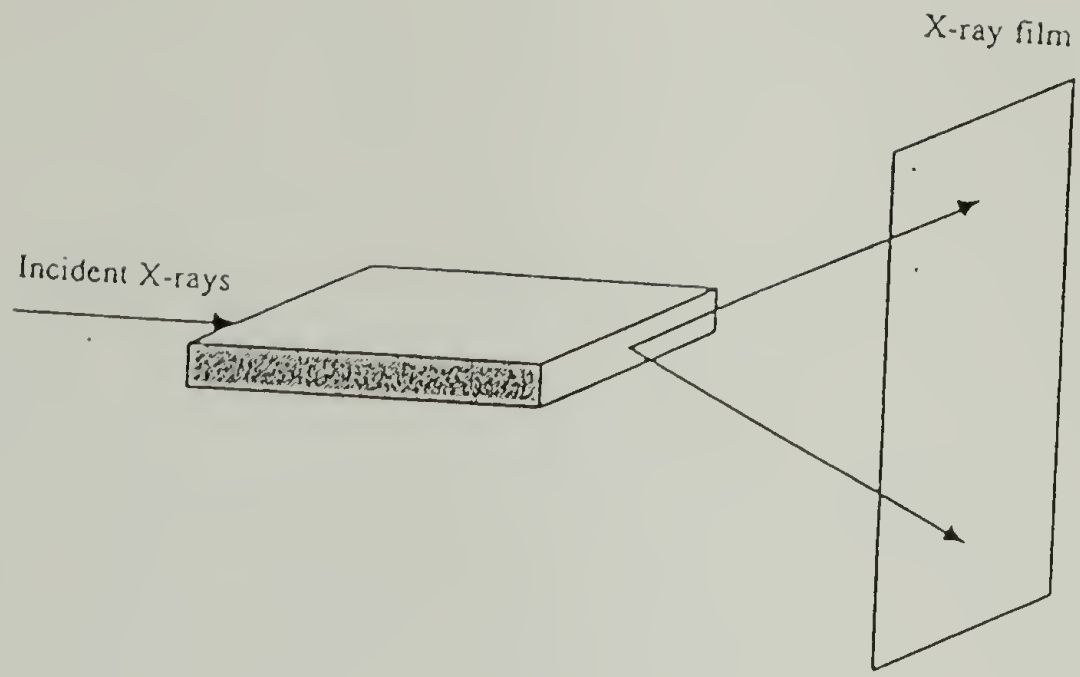
Glass capillaries of 1.5 mm internal diameter were used to hold powder samples in the X-ray experiments. In the diffraction experiments with oriented samples, the mat was mounted directly on the pin hole collimater. The geometry of the diffraction experiment on the oriented sample is shown in Figure 3.1. X-ray diffraction patterns were recorded on film in a Statton camera, with nickel filtered Cu K α radiation.

3.2.3 Electron Microscopy

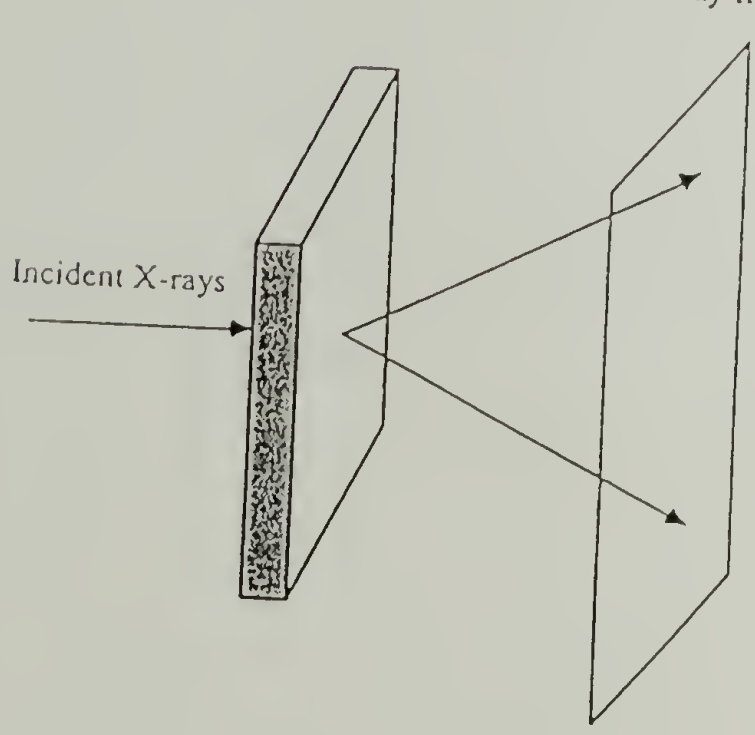
The size of the crystallites was determined by transmission electron microscopy. Crystallites from the suspension in methanol were sedimented on carbon coated TEM grids and were obliquely shadowed with Pt-Pd alloy and were examined in a JOEL FX 2000 transmission electron microscope operating at an accelerating voltage of 200 kV. The vertical dimensions on the micrographs were determined by shadow length measurements of polystyrene latex spheres 10690 Å in diameter.

3.3 Results and Discussion

An electron micrograph of crystallites of **1** is shown in Figure 3.2. The crystallites are asymmetric, needle-like objects on the order



(i)



(ii)

Figure 3.1 Geometry of the x-ray diffraction experiment; (i) Incident x-rays parallel to the mat; (ii) Incident x-rays perpendicular to the mat

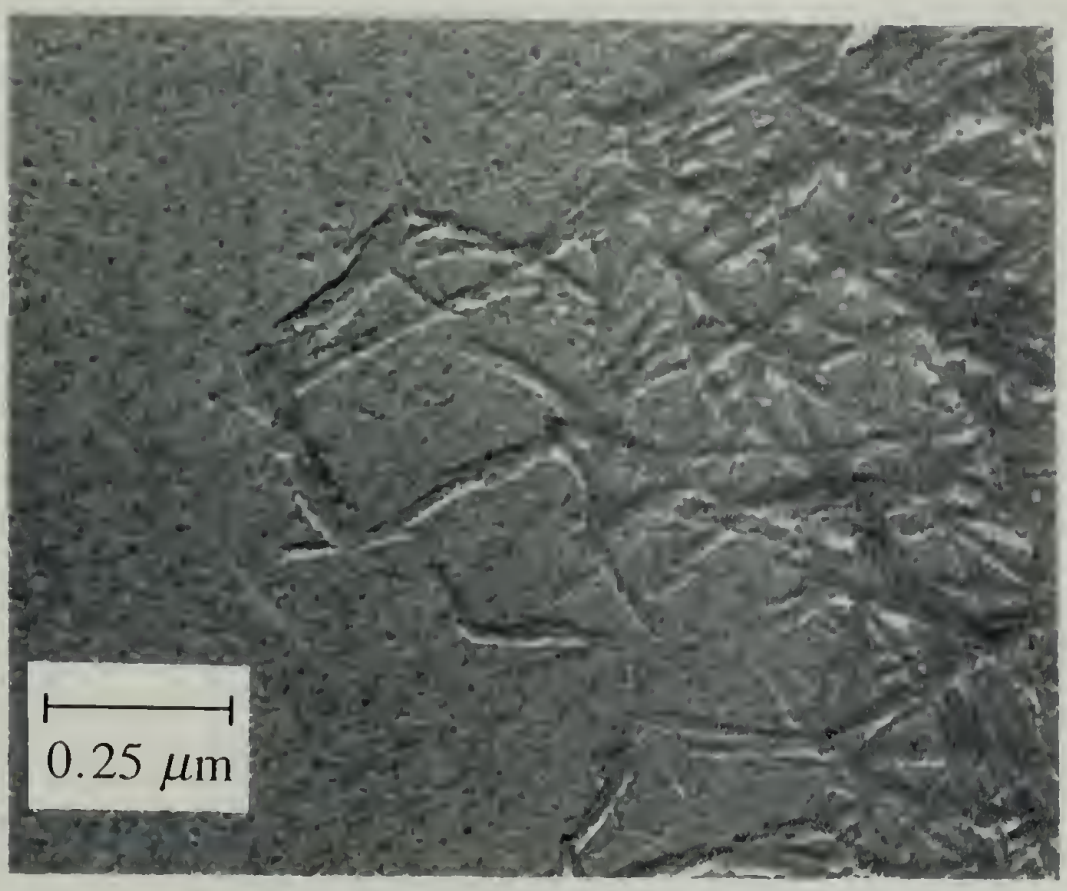


Figure 3.2 Transmission electron micrograph showing needle-like morphology of the crystalline aggregates. The morphology is identical in **1** and **2**.

of 2500 Å long. The other two dimensions vary between 50 and 250 Å in the isolated aggregates. The grain size of the Pt-Pd alloy which is used for shadowing leads to an error in the measurement of the shadow lengths which is on the order of 15-50 Å[14]. Isolated needle-like structures are more prevalent at the edges of the sample; in the inner areas, aggregates of the needle-like crystallites with larger dimensions are observed. Discrete thicknesses which are an integral multiple of the smallest thickness of the needle-like crystallites have been observed in 2[14]; this leads to the obvious speculation that the aggregates with larger thickness are formed due to layering of unit cystallites. The observations of layering effects of crystallites of 2 are preliminary[14] and have not been rigourously proved; hence, the electron micrograph reported in Figure 3.2 has not been analyzed for layering effects. This thesis does not address the crystal morphology and the layering effects of crystallites (leading to thicknesses of crystallites which are an integral multiples of the smallest thickness) in any detail.

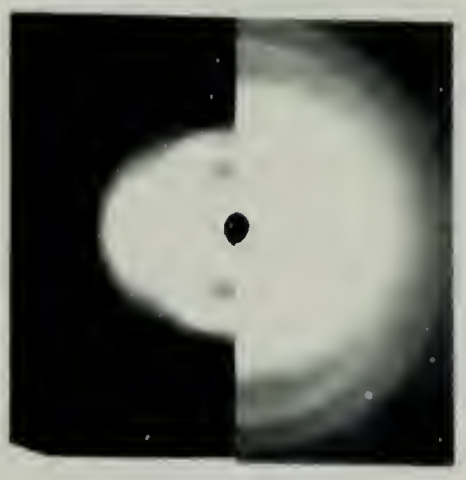
The wide angle x-ray diffraction patterns obtained for oriented samples of **1** and **2** when the incident x-ray beam is along the mat (Figure 3.1) are shown in Figure 3.3.

When the beam is perpendicular to the mat, the wide angle diffraction pattern is symmetric (Figure 3.3(iii)); the pattern resembles the powder pattern of the respective polypeptide with the exception that some of the reflections are missing. (After indexing of the patterns it is believed that the diffraction pattern obtained is a two dimensional powder pattern with the 0k0 reflections missing). The wide angle patterns (Figure 3.3(i)

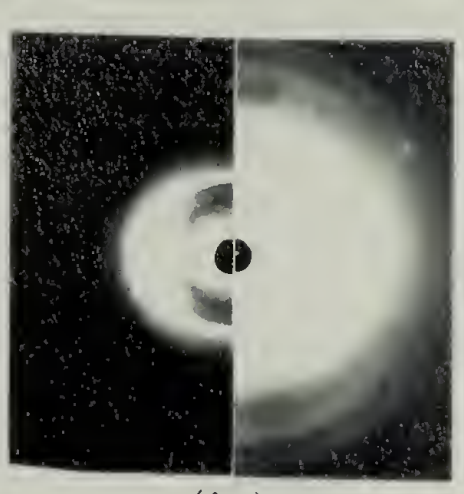
and 3.3(iv)) are similar for the two polypeptides. However, the inner reflections in the wide angle region (Figure 3.3(ii) and 3.3(v)) are different. Along the equator, the pattern for **1** (Figure 3.3(ii)) shows strong intensity at very low angles (greater than 11 Å), followed by a weak but sharp reflection (at 10.52 Å) at a larger angle, followed by a strong, sharp reflection (at 5.26 Å) followed by a strong, sharp reflection (at 4.58Å). Along the equator, the pattern for **2** (Figure 3.3(v)) shows practically no intensity in the low angle region (greater than 11Å), shows a very strong and sharp reflection (at 10.6 Å), followed by a weaker but sharp reflection (at 5.3 Å), followed by a weak, sharp reflection (at 4.21 Å). At approximately 45° to the equator, **1** shows a sharp, very strong reflection at 4.39 Å and a sharp, strong reflection at 3.76 Å, while **2** shows a sharp but weaker reflection at 4.33 Å but only a broad,weak reflection beyond it; these features are more apparent in Figure 3.3(i) and 3.3(iv). The observed reflections for **1** have been listed in Tables 3.1 and 3.2.

Previous work on related polypeptides like **2**[15], silks and silk analogues[3,16] indexed the diffraction patterns based on an orthorhombic lattice. Hence, our first attempt was to index the diffraction pattern of **1** to an orthorhombic unit cell. It was later determined that the orthorhombic unit cell could not explain the diffraction pattern but a monoclinic unit cell could. In this chapter, the orthorhombic system is discussed first. The monoclinic unit cell is discussed subsequently.

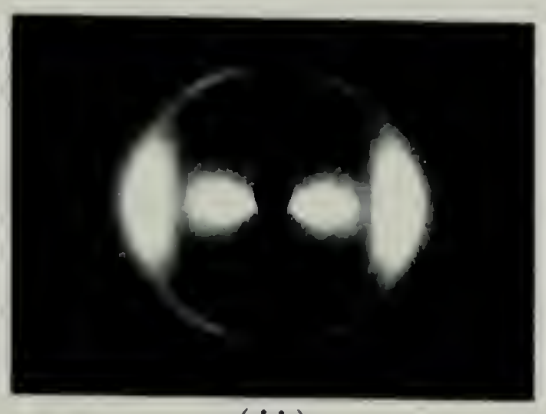
Figure 3.3 X-ray diffraction patterns of oriented crystalline aggregates of **1** and **2** recorded on film in a Statton camera, with a Ni filter and Cu Kα radiation. (i)-(iii) are for **1**, (iv) and (v) are for **2**. (iii) is obtained with the diffraction geometry as shown in Figure 3.1(ii); the rest are using the diffraction geometry as shown in Figure 3.1(i). (ii) and (v) show the inner reflections in (i) and (iv) respectively.



(i)



(iv)



(ii)

(v)

(iii)

3.3.1 Analysis Based on Orthorhombic Unit Cell for **1**

Most of the reflections in the diffraction patterns of **1** could be indexed on an orthorhombic unit cell with a=9.56 Å, b=10.68 Å and c=7.0 Å; a is the hydrogen bonding direction, b is the intersheet direction and c is the chain direction. Indexing of the patterns is not complete until the issue of orientation of crystallites is addressed; this is discussed in section 3.3.9; however, the reader might benefit by reading that section and Figure 3.11 briefly, in order to become familiar with the locations of the reflections. Figure 3.3(iv) and 3.3(v) show that the 010 reflection in **2** is the strongest 0k0 reflection on the equator while Figure 3.3(i) and 3.3(ii) show that it is weak in **1** (which features a strong 020 reflection). Furthermore, the intensities of the 021, 211 and 210 reflections in **2** are reduced in comparison to those in **1**. Orthorhombic unit cells for the two polypeptides are shown in Figure 3.4. Table 3.1 shows the calculated and observed d-spacings for polypeptide **1**. The corresponding spacings for **2** have been reported elsewhere[15]; this polypeptide has been shown to be indexed on an orthorhombic unit cell with the unit cell dimensions a=9.48 Å, b=10.6 Å, c=6.95 Å.

3.3.2 Disagreement of X-ray Data with Orthorhombic Unit Cell

Figure 3.5(i) shows the crystal packing corresponding to an orthorhombic unit cell with apolar sheets; the calculated diffraction pattern is shown in Figure 3.5(ii). Figure 3.6(i) shows the crystal packing corresponding to an orthorhombic unit cell with polar sheets; the calculated diffraction pattern is shown in Figure 3.6(ii). The difference in

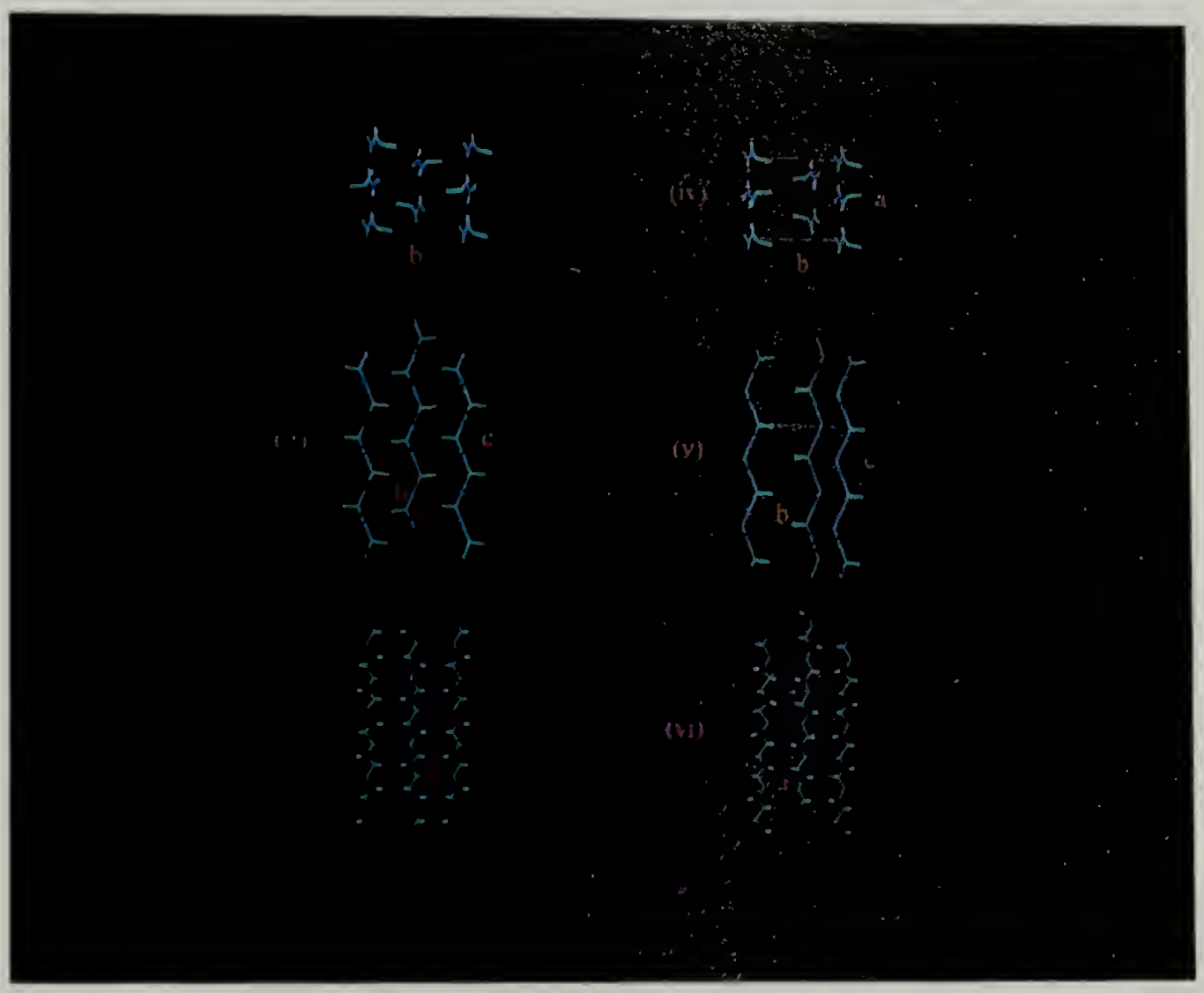
Table 3.1

Calculated and observed d-spacings

Orthorhombic unit cell with a=9.56 Å, b=10.68 Å, c=7.0 Å

	kl	d-spacing (Calculated) (Å)	d-spacing (Observed) (Å)	Comment
h=0	10	10.68		
	20	5.34	5.26	Sharp (s)
	21	4.24	4.58	Sharp (s)
	30	3.56		
	02	3.5		
	22	2.92	3.12	Broad (w)
h=2	00	4.78	4.77	Very sharp (w)
	10	4.36	4.39	Sharp (vs)
	11	3.70	3.76	Sharp (s)
h=3	01	2.9	2.78	Broad (m)
	10	3.05	2.78	Broad (m)
	11	2.8	2.78	Broad (m)
	12	2.3	2.30	Sharp (w)
	22	2.16		
h=4	00	2.39	2.39	Very sharp (w)
	10	2.33	2.30	Sharp (w)
	20	2.18	2.11	Sharp (w)

Figure 3.4 Proposed orthorhombic unit cells for **1** and **2**. Figures (i), (ii) and (iii) show the projections of the unit cell for **1** on the three principal orthogonal planes. Figures (iv), (v) and (vi) show similar projections for **2**. For the purpose of clarity, in Figures (iii) and (vi), only one hydrogen bonded sheet is shown.


b
(iv)
a
b
c
(v)
c
b
(vi)

intensities of the 010 reflections between the polar and apolar packing is apparent in the calculated diffraction patterns; this is due to the difference in the crystallographic repeat along the b axis. The intensities observed in the experimental diffraction patterns do agree with the formation of apolar sheets. However, there is a discrepancy in the locations of the reflections. Table 3.1 shows that the 021 reflection for **1** should occur at 4.24 Å while the reflection is actually observed at 4.58 Å. One way to account for this is to employ a larger unit cell; however, this causes the 020, 211 and 210 reflections to be at larger spacings also. Hence, no orthorhombic unit cell can fit the observed x-ray data satisfactorily.

3.3.3 Monoclinic Unit Cell for **1**

The x-ray diffraction pattern for **1** can be best explained by a monoclinic unit cell with a=9.56 Å, b=10.68 Å, c=7.0 Å and the angle between c and b equal to 80°. This corresponds to a translation of 0.91 Å in the c direction of every sheet with respect to the adjacent sheet. In the orthorhombic system, however, there is no displacement between adjacent sheets. Figure 3.7(i) shows the crystal packing corresponding to a monoclinic unit cell with apolar sheets; the calculated diffraction pattern is shown in Figure 3.7(ii). Figure 3.8(i) shows the crystal packing corresponding to a monoclinic unit cell with apolar sheets; the calculated diffraction pattern is shown in Figure 3.8(ii). Table 3.2 shows the calculated and observed d-spacings for the monoclinic unit cell.

Table 3.2

Calculated and observed d-spacings

Monoclinic unit cell with a=9.56 Å, b=10.68 Å, c=7.0 Å; Angle between c and b=80°

	kl	d-spacing (Calculated) (Å)	d-spacing (Observed)	Comment
h=0	10	10.52		
	20	5.26	5.26	Sharp (s)
	21	4.58	4.58	Sharp (s)
	30	3.51		
	02	3.45		
	22	3.15	3.12	Broad (w)
h=2	00	4.78	4.77	Very sharp (w)
	10	4.35	4.39	Sharp (vs)
	11	3.81	3.76	Sharp (s)
h=3	01	2.89	2.78	Broad (m)
	10	3.05	2.78	Broad (m)
	11	2.84	2.78	Broad (m)
	12	2.34	2.30	Sharp (w)
	22	2.24		
h=4	00	2.39	2.39	Very sharp (w)
	10	2.33	2.30	Sharp (w)
	20	2.18	2.11	Sharp (w)

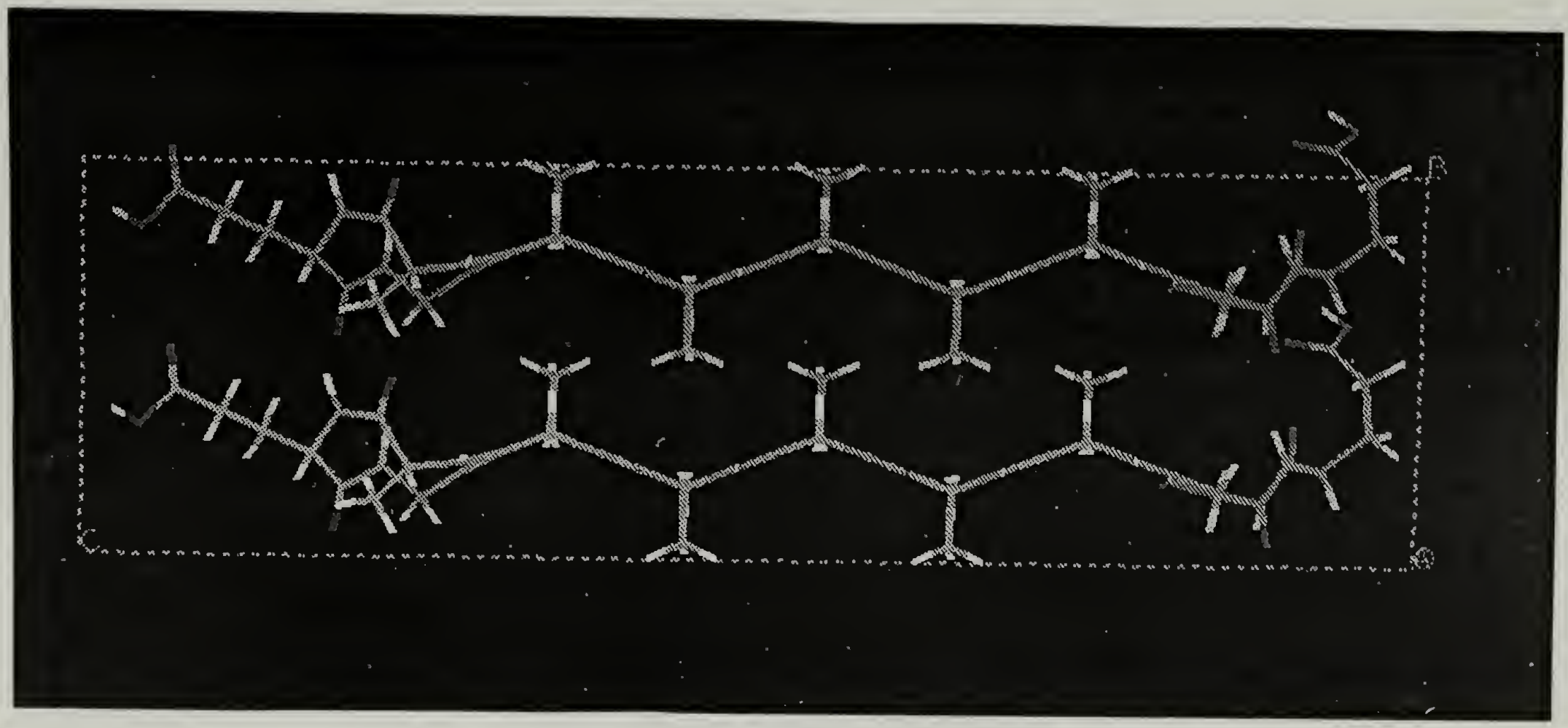


Figure 3.5 Orthorhombic apolar sheets

Figure 3.5(i) Crystal packing for an orthorhombic unit cell of **1** with apolar beta sheets (Reproduced with permission from E. D. T. Atkins, Bristol, England)

Continued, next page

Figure 3.5(ii) Calculated diffraction pattern for 3.5(i) (Reproduced with permission from E. D. T. Atkins, Bristol, England)

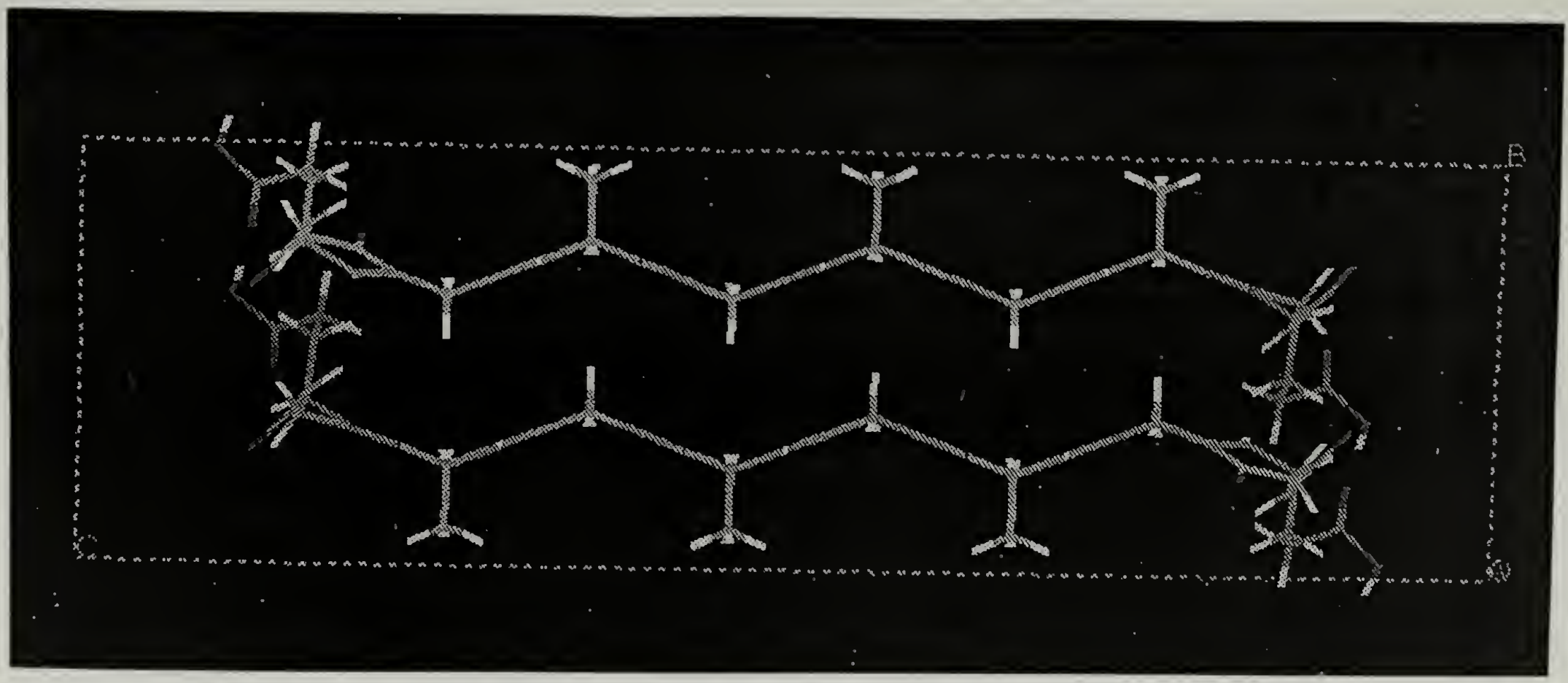


Figure 3.6 Orthorhombic polar sheets

Figure 3.6(i) Crystal packing for an orthorhombic unit cell of **1** with polar beta sheets. (Reproduced with permission from E. D. T. Atkins, Bristol, England)

Continued, next page

Figure 3.6(ii) Calculated diffraction pattern for 3.6(i). (Reproduced with permission from E. D. T. Atkins, Bristol, England)

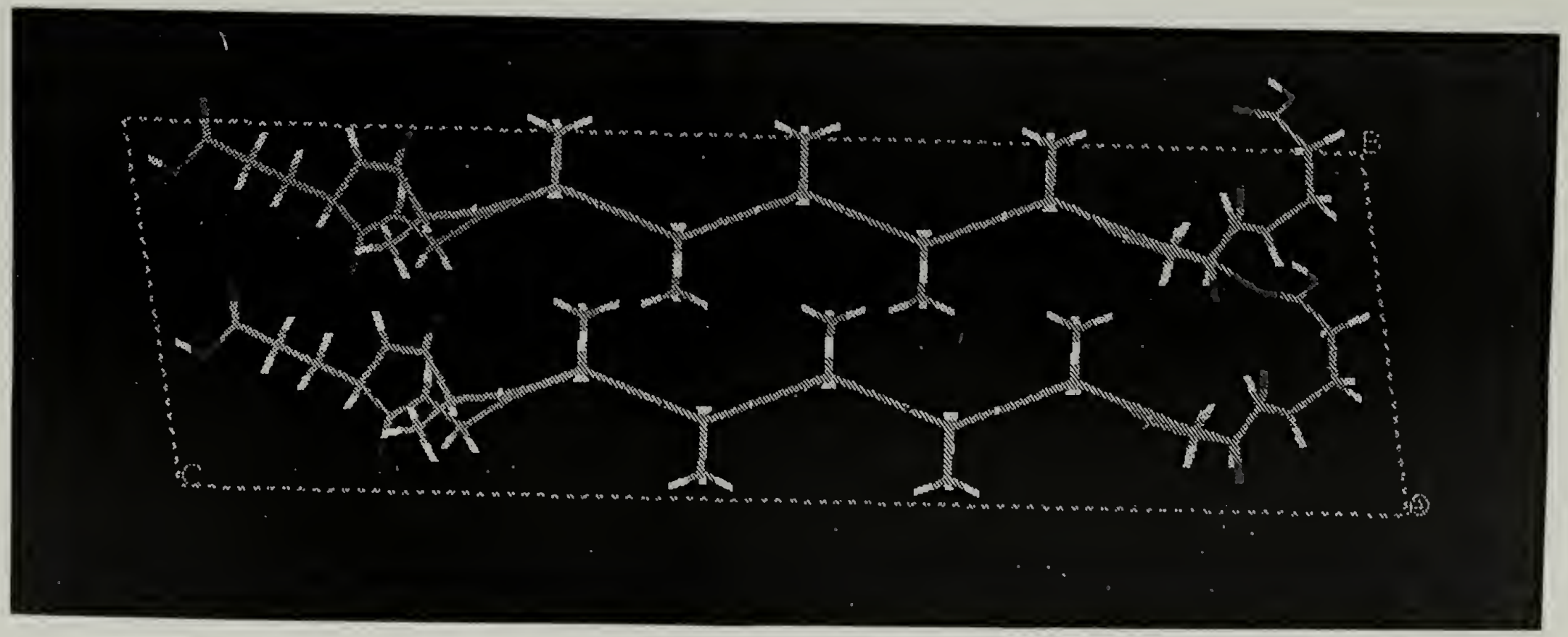


Figure 3.7 Monoclinic apolar sheets

Figure 3.7(i) Crystal packing for a monoclinic unit cell (a=9.56 å, b=10.68 å, c=7.0 å; angle between c and b=80°) of **1** with apolar beta sheets.

(Reproduced with permission from E. D. T. Atkins, Bristol, England)

Continued, next page

Figure 3.7(ii) Calculated diffraction pattern for 3.7(i). (Reproduced with permission from E. D. T. Atkins, Bristol, England)

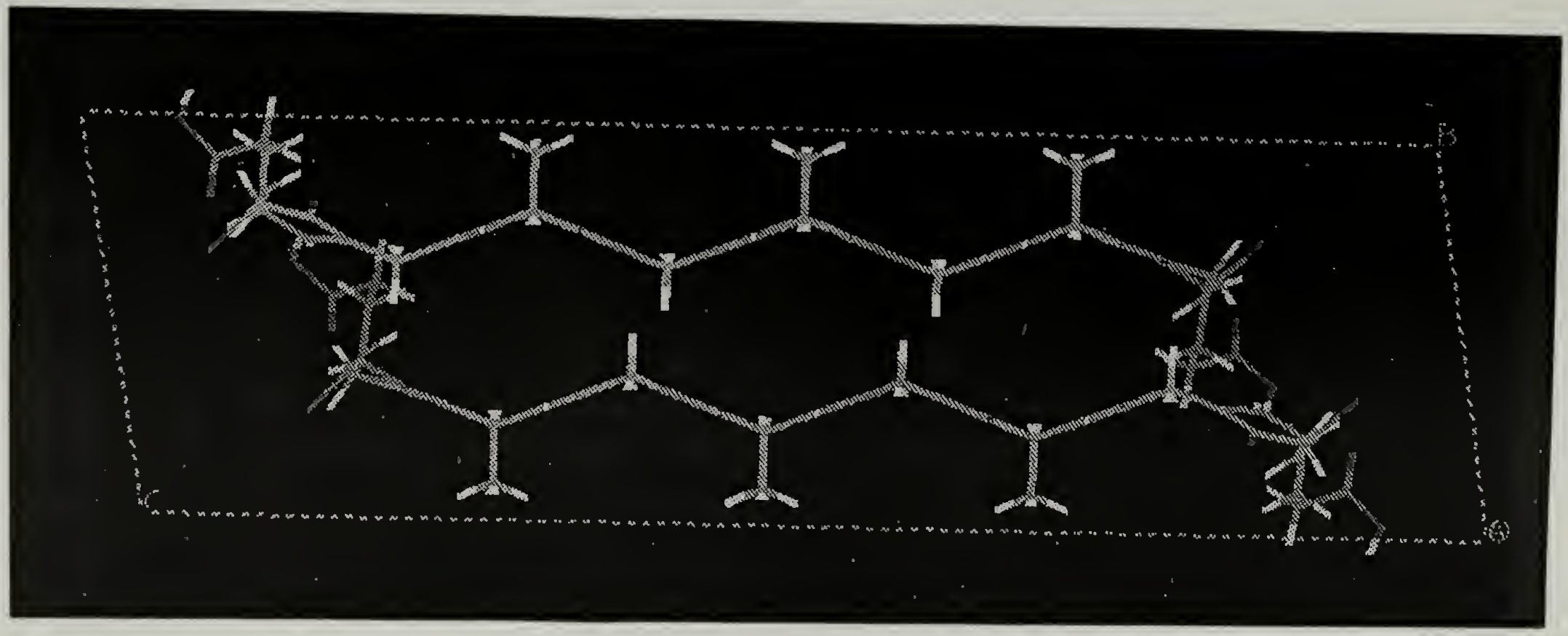


Figure 3.8 Monoclinic polar sheets

Figure 3.8(i) Crystal packing for a monoclinic unit cell (a=9.56 å, b=10.68 å, c=7.0 å; angle between c and b=80°) of **1** with polar beta sheets.

(Reproduced with permission from E. D. T. Atkins, Bristol, England) Continued, next page

Figure 3.8(ii) Calculated diffraction pattern for 3.8(i). (Reproduced with permission from E. D. T. Atkins, Bristol, England)

Table 3.2 shows that all the positions of the reflections can be explained by employing the monoclinic unit cell.

3.3.4 Scope and Limitations of X-ray Analysis

The diffraction patterns are very similar to those observed in silk II forms of *Bombyx mori* and poly (alanineglycine)[1,3,4,16]. Table 3.3 shows the observed and calculated diffraction spacings for poly(alanyl-glycine)[4]. A comparison of the reflections between those listed in Table 3.2 and those in Table 3.3 indeed shows some similarity. The main reflections, 020, 021, 200, 210 and 211 are observed in both instances; the unit cell in the *b* direction, is expanded in the structure of **1**. The expansion of the unit cell could be due to the effects of turns or due to the incorporation of glutamic acids as explained below. We are unable to rationalize the monoclinic packing in **1** in comparison to orthorhombic for poly(alanylglycine).

In order for the polypeptide chains to form unique and predictable scaffolds, the chains had to fold and form the expected crystal structure. Of course, the formation of the expected crystal structure was a necessary but not a sufficient condition; other conformations of chains which did not involve periodic folding could also lead to the formation of silk-like crystal structures (Figure 3.9). Figure 3.9(i) shows the projection of a single sheet along the *a* axis of crystalline **2**; the figure shows a sheet comprising chains of **2** which do not fold at the glutamic acid residue (the glutamic acid residue is the one with the larger side chain). These chains can then be incorporated into a crystal structure as shown in Figure 3.9(ii). Figure

Table 3.3

Calculated and observed d-spacings for poly(alanyl-glycine)

(Adapted from work by Fraser and coworkers)

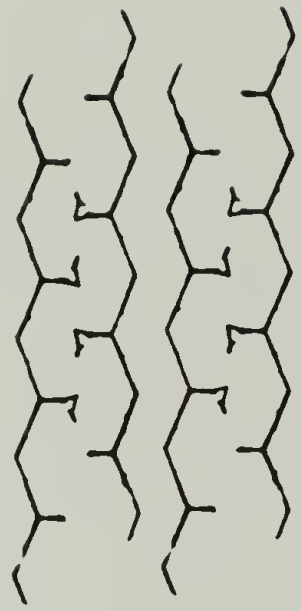
Orthorhombic unit cell with a=9.42 Å, b=8.87 Å, c=6.95 Å

vw=very weak, w=weak, m=medium, s=strong

	kl	d-spacing (Calculated) (Å)	d-spacing (Observed)	Comment
h=0	10	8.87	8.87	m
	20	4.44	4.45	vvs
	21	3.74	3.76	m
	30	2.96	2.97	m
	40	2.22	2.25	vw
	06	1.16	1.16	vw
h=2	00	4.71	4.71	m
	10	4.23	4.22	vs
	11	3.57	3.61	m
	15	1.32	1.31	vw
h=3	10	2.96	2.95	w
	11	2.72	2.74	w
	12	2.25	2.25	m
	22	2.06	2.07	w
h=4	00	2.22	2.25	vw
h=5	22	1.55	1.55	w

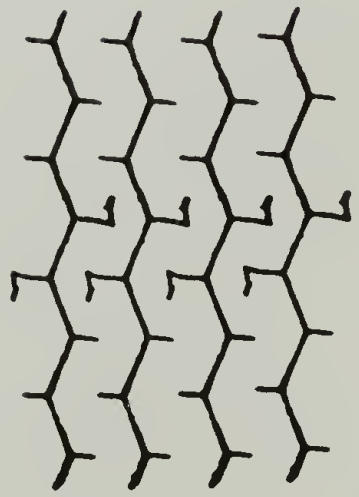
Figure 3.9 Alternate schemes of chain folding leading to beta sheet structure; (i) Projection of a sheet comprising a few chains of **2** along the *a* direction; the shorter side-chain corresponds to the methyl group of alanine and the larger side chain is due to glutamic acid; the hydrogen due to glycine is not shown; (ii) A combination of chains shown in (i) which shows the formation of polar sheets and the ensuing packing; (iii) Projection of a sheet comprising a few chains of **1** along the *a* direction; (iv) A combination of chains shown in (iii) which shows the formation of apolar sheets and the ensuing packing.

(i)



(ii)

(iii)



(iv)

3.9(iii) shows similar unfolded chains of **1** and the subsequent crystalline packing (Figure 3.9(iv)); in this case also, the glutamic acid side chains are incorporated within the crystalline domains. The structures depicted by Figure 3.9(ii) and 3.9(iv) would lead to a beta sheet structure with a unit cell slightly expanded in the *b* direction. A large number of such combinations of the chains in different orientation can lead to the formation of polar or apolar sheets in both the polypeptides; combinations of these sheets can explain the observed diffraction patterns for both the polypeptides. The combinations shown in Figure 3.9 are not exhaustive but serve to illustrate the multiplicity of solutions to the diffraction patterns. It is important to underscore at this juncture, the purpose and the limitations of the x-ray diffraction experiments. The presence of silk-like structures was directly determined by these experiments but the folding pattern was not. Although the folding habits have been previously determined in polymers by x-ray diffraction[17], this work did not address this issue directly.

3.3.5 Comparison of Crystal Structures of **1** and **2**

It has been previously pointed out that **1** and **2** are different because of the inverted diad sequence in alternate beta-sheet-forming-segments along the polypeptide chain (alternation of $(AG)_3$ and $(GA)_3$ along the chain). Previous work[15] has suggested that **2** forms polar sheets which are decorated with methyl groups only on one surface. The x-ray data suggests that in the crystal structure of **2**, the turns are constrained to place the methyl groups (of alanine) of adjacent beta sheet strands on one side of the sheet and the hydrogen atoms (of glycine) on the other side of the sheet

which leads to the formation of polar sheets; these sheets are therefore decorated with methyl groups on one side and hydrogen atoms on the other. These polar sheets then pack such that the methyl-group-decorated side of the sheet faces the methyl-group-decorated side of an adjacent sheet and the hydrogen-decorated side of the sheet faces the hydrogen-decorated side of an adjacent sheet. This periodicity leads to the strong 010 reflection observed. The crystal structure generated with **1** is different because of the inverted diad sequence in alternate beta-sheet-forming-segments along the polypeptide chain (alternation of $(AG)_3$ and $(GA)_3$ along the chain). Periodic chain folding of **1** leads to the formation of apolar sheets; these sheets have equal numbers of methyl groups decorating both sides of the sheet. The packing of these sheets in the crystal structure leads to a weakening of the 010 reflection and the presence of a strong 020 reflection.

This phenomenon of polar and apolar sheet packing of beta sheets has been observed previously for silk and its analogues[1,3,4,18]. However, this is the first time that beta sheet forming chains have been interrupted at exact periodic intervals with turn forming residues causing them to fold and generate predicted three dimensional structures with precise scaffolding on the length scale of nanometers.

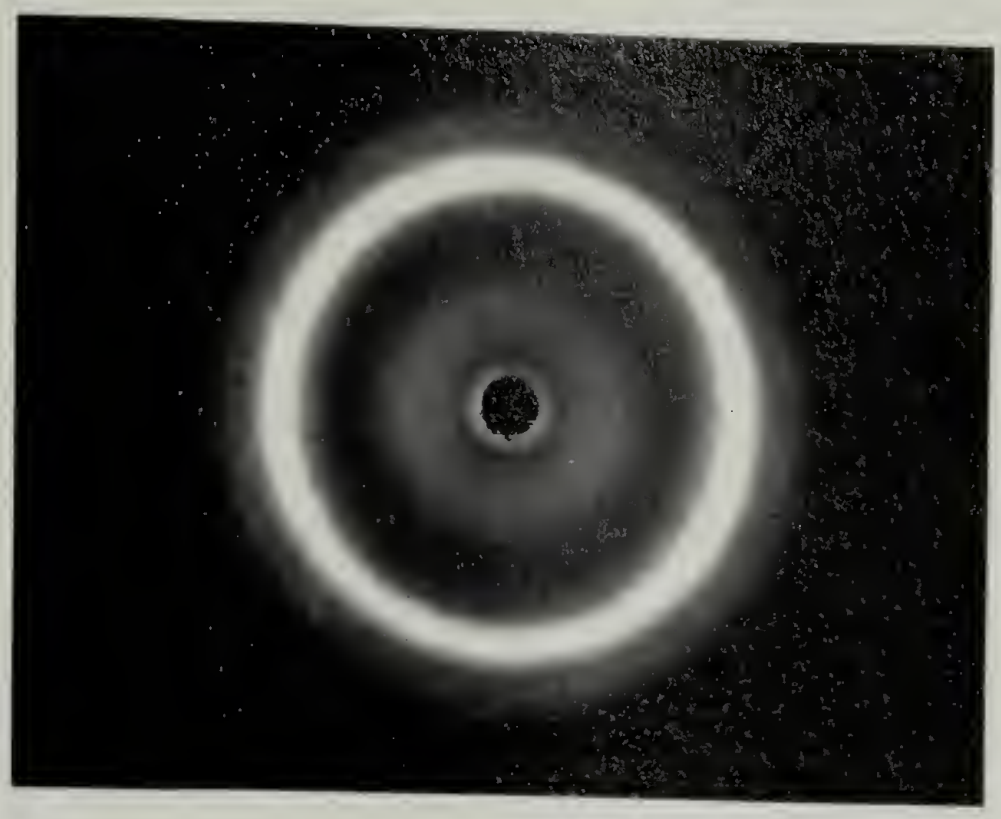
These patterns can be explained by a few other forms of chain folding also. These have been depicted in Figure 3.9. The incorporation of the bulky glutamic acid side-chain in the lattice could lead to an expanded unit cell in the inter-sheet dimension. At this juncture this possibility cannot be ruled out on a structural basis. However, the glutamic acid side-

chain is polar in comparison to the other amino acids. In aqueous environments, it is anticipated that the glutamic acid resides at the turn, if the turn does form.

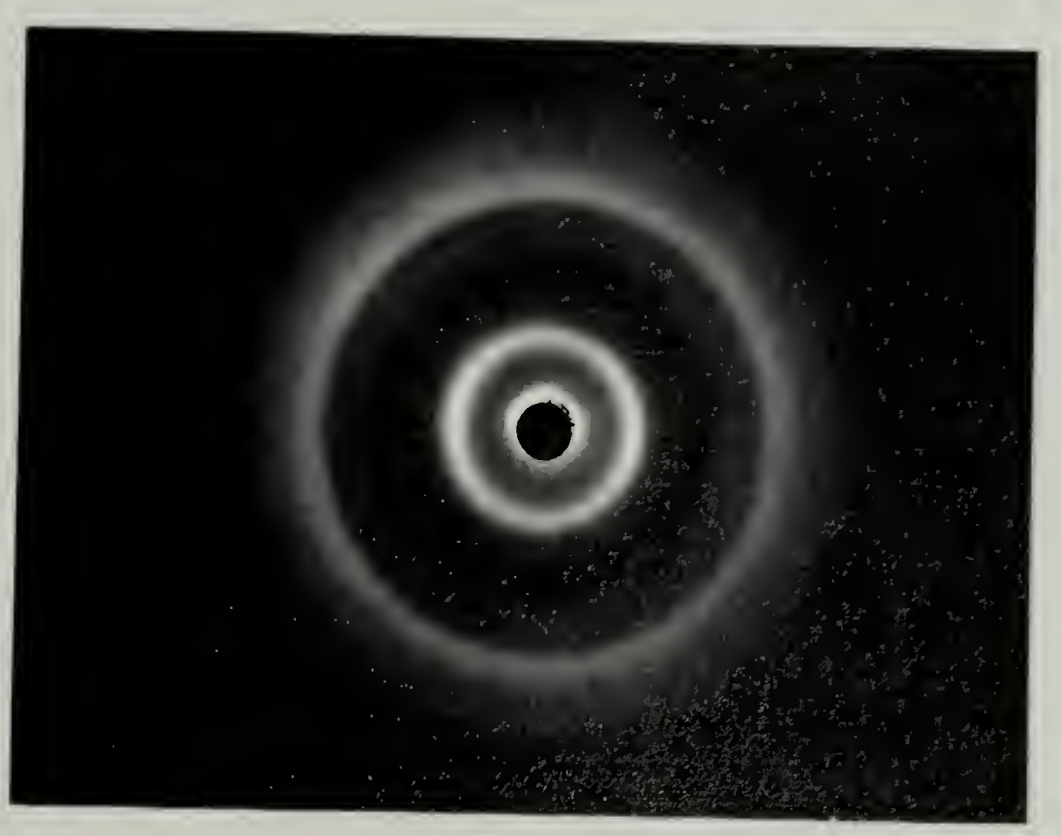
It is unexpected that **1** forms a monoclinic unit cell; this corresponds to a translation of 0.91 Å in the c direction of every sheet with respect to the adjacent sheet. In the orthorhombic system, however, there is no displacement between adjacent sheets. At this juncture we have no explanation for this change of the unit cell from orthorhombic for **2** to monoclinic for **1**.

3.3.6 Comparison of Powder Diffraction Patterns of **1** and **2**

The powder diffraction patterns for **1** and **2** are shown in Figure 3.10(i) and 3.10(ii) respectively. The powder diffraction patterns show the presence of rings due to the occurrence of all possible orientations of the crystallites with equal probability. Figure 3.10(ii) shows a prominent 010 diffraction pattern for **2** (the innermost ring is this reflection); the diffraction pattern for **1** does not show any 010 reflection. The 020, 021, 200 and 210 reflections coalesce into one intense ring for **1** (Figure 3.10(i)); the intensity of this ring is lower for **2** (Figure 3.10(ii)) due to the lower intensity of the 021, 200 and 210 reflections. In the diffraction pattern corresponding to **1** (Figure 3.10(i)) a third sharp ring corresponding to the 211 reflection is seen outside the strong broad ring; this is absent in the diffraction pattern corresponding to **2** probably due to the weak 211 reflection as seen in Figure 3.3.



(i)



(ii)

Figure 3.10 (i) Powder diffraction patterns for crystalline **1**; (ii) Powder diffraction patterns for crystalline **2**; X-ray diffraction patterns were recorded on film in a Statton camera, with a Ni filter and Cu K α radiation; Sample to film distance=76 mm;

3.3.7 Small Angle Reflections in **1**

In addition to the wide angle reflection, a set of small angle reflections is also observed in the oriented diffraction pattern of **1** (Figure 3.3(ii)). The reflections are listed in Table 3.4.

Table 3.4

Small angle reflections for **1**

Observed d-spacing (Å)	Comment
21	Broad (vs)
10.5	Very broad (s)
7	Very broad (w)

The reflections are very broad and decrease in intensity with decrease in d-spacing.

Figure 3.10(i) does not show any reflections corresponding to those listed in Table 3.3. The absence of small angle reflections from the powder pattern leads us to conclude that these reflections are not due to any crystallographic features but arise as a consequence of the orientation process. The waning intensities as a function of increasing scattering angle suggest that the 21 Å, 10.5 Å and the 7 Å reflections are the first, second and third orders of order on the length scale of 21 Å. The intensities and character of these reflections do not change on annealing for 12 hours at 80

°C. These small angle reflections are not observed for similar sample preparations of **2** (Figure 3.3(v)). At this juncture, we cannot speculate on the cause of the reflections.

3.3.8 Crystal Packing and Morphology for **1** and **2**

As is apparent from the diffraction patterns, the crystal packing in **1** and **2** is different. However, the morphology is identical in both the cases (Figure 3.2). Moreover, the dimensions of the needle-like crystallites are similar for both the polypeptides. This observation is important because it proves that the differences between the two patterns are not due to different numbers of coherent scattering centers along a direction; the numbers would be different if the sizes of the crystallites were different and one would have to scale the intensities and breadth of the reflections in order to compare the **1** and **2**.

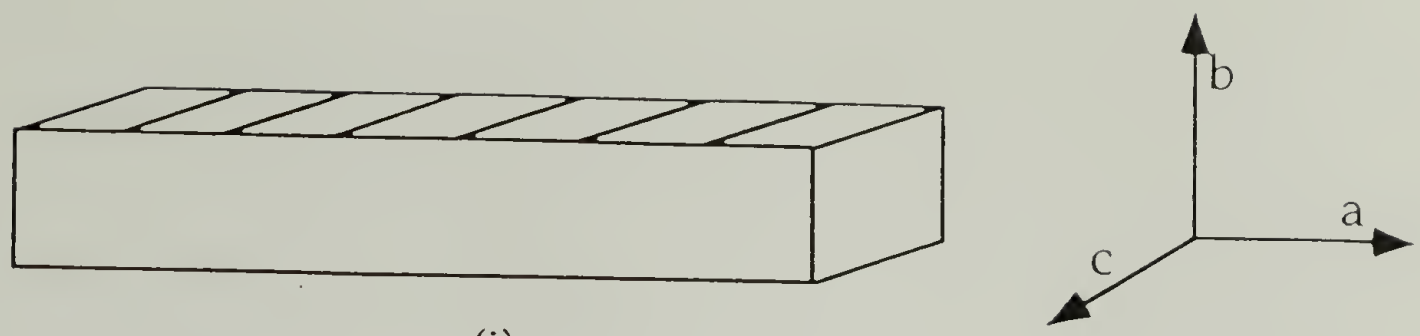
3.3.9 Orientation of the Asymmetric Crystallites within the Oriented Mat

The crystallites depicted in Figure 3.2 were oriented by clamping the gel between two glass plates (re. Methods section). The flow fields leading to this orientation have not been ascertained. The diffraction patterns can be explained by two modes of crystallite orientation, as shown in Figure 3.11, within the mat. The schematics of the diffraction patterns are shown in Figure 3.12. The presence of crystallites as shown in Figure 3.11(iii) in conjunction with the appropriate scaling of the reflection intensity by the Lorentz factor can predict the oriented diffraction patterns correctly. This

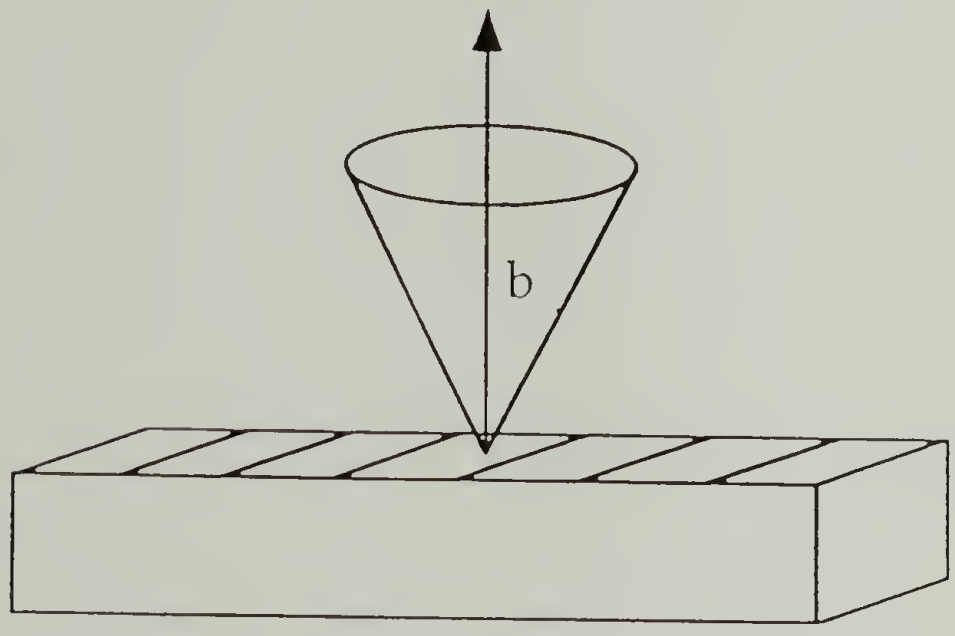
requires the crystallites to be oriented such that the *a* axis in every crystallite is along the plane of the mat; the crystallites are dispersed randomly when viewed orthogonal to the mat; rotation around the *a* axis gives rise to different crystallite orientations, all of which are present with equal probability.

Figure 3.11(ii) shows another model which can adequately explain the intensities in the diffraction patterns. The crystallites are imperfectly oriented such that the *b* axis is predominantly perpendicular to the plane of the mat; the probability of the orientation drops as the angle subtended by the *b* axis with the perpendicular direction increases.

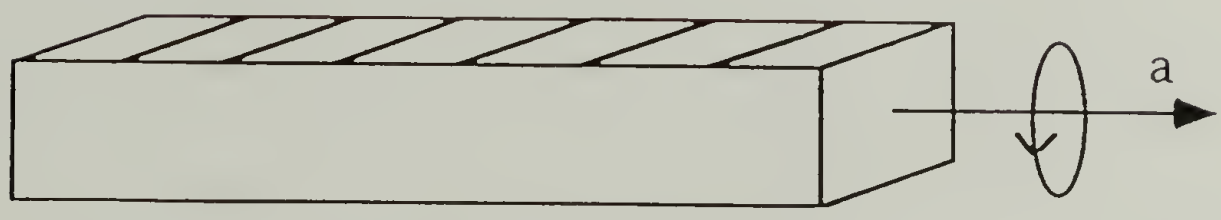
It should be noted, however, that the intensities as a function of the azimuthal angle have not been accurately quantified and any attempt to deduce the exact orientation of crystallites can be qualitative at best.



(i)

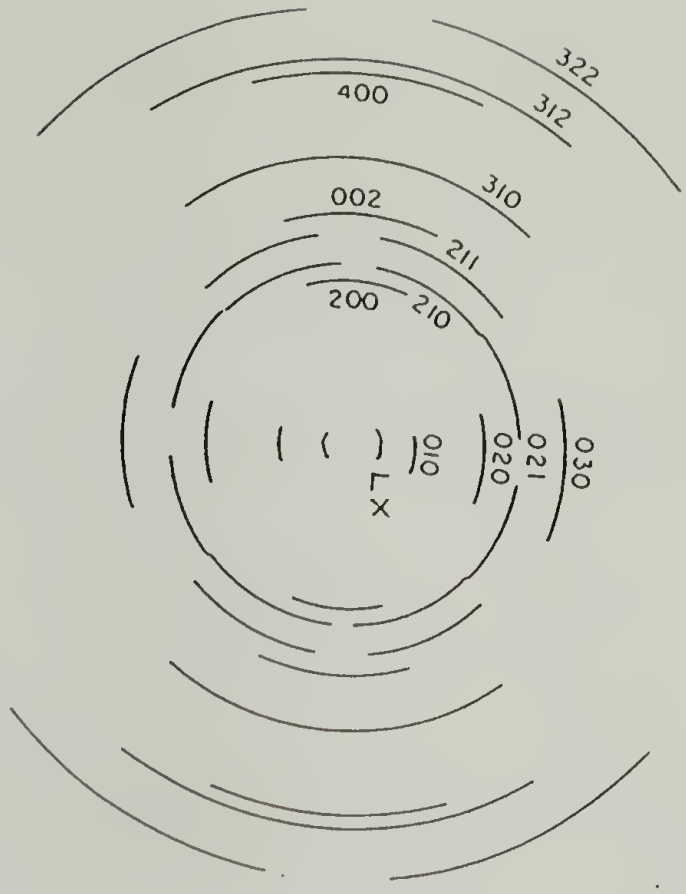


(ii)



(iii)

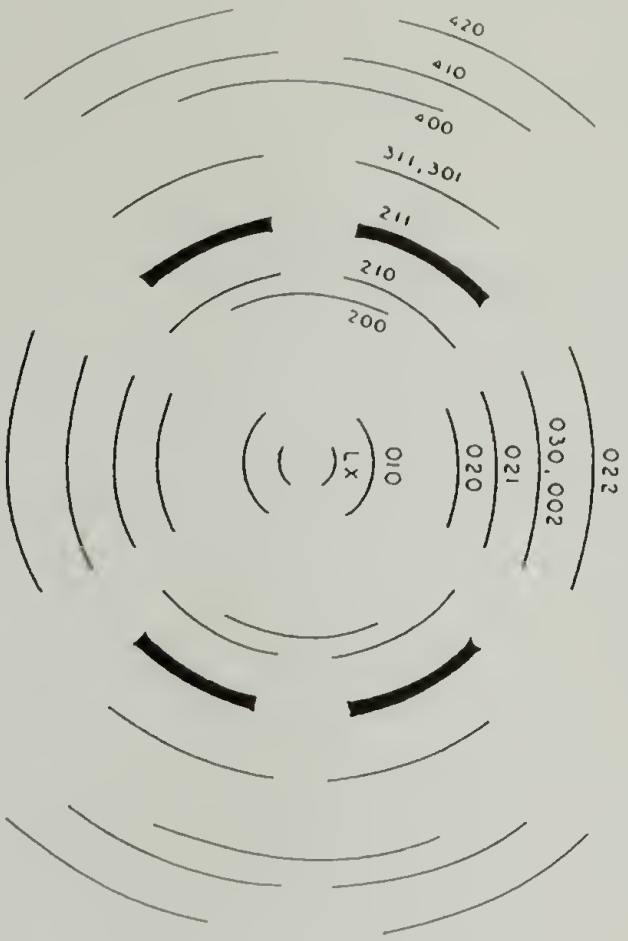
Figure 3.11 Possible orientation of crystallites within the mat; Refer to text for details. The mat is a horizontal plane perpendicular to the plane of the figure.



(i)

Figure 3.12 Schematic of the observed reflections and indexing of the patterns based on the two orientations of the crystallites depicted in Figure 3.11. (i) corresponds to the orientation as in Figure 3.11(ii); (ii) corresponds to the orientation as in Figure 3.11(iii). In these schematics, all the reflections are shown with almost equal intensity; however, the intensities do vary between **1** and **2** as has been discussed in previous sections. Furthermore, since there is a difference between the dimensions of the unit cells and their lattices (orthorhombic or monoclinic), the actual positions may be different from those seen in Figure 3.3, 3.5(ii), 3.6(ii), 3.7(ii) or 3.8(ii). The Figures 3.5(ii), 3.6(ii), 3.7(ii) and 3.8(ii) are generated by using the orientation as shown in Figure 3.11(iii).

Continued, next page



(ii)

Figure 3.12 (Contd.)

3.4 Conclusions

The polypeptide **1** forms a silk II structure when processed in 70% formic acid. The unit cell deduced from the diffraction patterns is similar to those from previous diffraction patterns on silks and silk-like polypeptides. There are systematic differences between the diffraction patterns of **1** and **2**. These differences can be explained on the basis of a model with a three residue turn of the polypeptide chain leading to the formation of a beta sheet structure. As in most diffraction analyses, this one does not rule out a large number of other possibilities, some of them not even requiring the formation of periodic turns. The most definite inference that arises from this analysis is the presence of crystalline beta sheet regions and the presence of apolar sheets for **1**. The issues of the degree of crystallinity and the mode of connectivity in between or within these crystalline regions have not been addressed. The unit cell which best explains the diffraction data for **1** is monoclinic with a=9.56 Å, b=10.68 Å and c=7.0 Å with the angle between b and c=80°.

Previous work[10] on **2** suggests lamellar thicknesses on the order of 30 Å. This would be the expected thickness, if the polypeptide were to fold as predicted. The small angle x-ray diffraction results in this work, however, were not conclusive and no inference regarding the lamellar thickness of **1** can be drawn.

3.5 References

(1) Marsh, R. E.; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1-34.

(2) Lotz, B.; Cesari, F. C. *Biochimie* **1979**, *61*, 205-214.

(3) Lotz, B.; Brack, A.; Spach, G. *J. Mol. Biol.* **1974**, *87*, 193-203.

(4) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706-712.

(5) Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343-358.

(6) Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Manuscript in preparation*

(7) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727-733.

(8) Chou, P. Y.; Fasman, G. D. *J. Mol. Biol.* **1977**, *115*, 135-175.

(9) Levitt, M. *Biochemistry* **1978**, *17(20)*, 4277-4285.

(10) Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* In press.

(11) Krejchi, M. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polymer Preprints* **1991**, *32(1)*, 411-412.

(12) Dougherty, M. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Manuscript in preparation*

(13) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759-777.

(14) Waddon, A. J. *Personal discussions*

(15) Krejchi, M. T. *Univ. of Mass., Amherst* **1993**, *Ph. D. Thesis*.

(16) Lotz, B.; Keith, H. D. *J. Mol. Biol.* **1971**, *61*, 195-200.

(17) Atkins, E. D. T.; Hill, M.; Hong, S. K.; Keller, A.; Organ, S. *Macromolecules* **1992**, *25*, 917-924.

(18) Warwicker, J. O. *Acta. Cryst.* **1954**, *7*, 565-573.

CHAPTER 4

DETERMINATION OF CHAIN FOLDING HABIT BY FTIR

4.1 Introduction

4.1.1 Scope of Infrared Analysis

X-ray diffraction analysis of the polypeptides showed that **1** has a propensity to crystallize into antiparallel beta sheets when stirred in 70% formic acid. However, it did not address the issue of chain folding in any detail; precise chain folding is essential for the chains to form the predicted lamellar scaffold. The infrared experiments which will be discussed in this chapter were designed to address this issue. In addition to addressing the issue of chain folding, the infrared spectra could be used to corroborate the presence of the antiparallel beta sheet structures in the crystalline samples.

In addition to the repetitive polypeptide **1**, an analogue in which the alanine and glycine were enriched with $^{13}C{=}O$ (**1-AG**) was synthesized using the same biological synthetic machinery. The two analogues were blended in solution and co-crystallized. FTIR spectra (in the amide I region) were studied as a function of blend composition. Crystallization and the infrared analysis of **1** enriched with $^{13}C{=}O$ at the glycine position in varying fractions (**1-Gx**; **x**=100, 60, 30, 0) has also been carried out; these results prove that the location of absorption of $^{13}C{=}O$ in an antiparallel beta sheet is dependent on the fraction of $^{13}C{=}O$ in the unit cell. The results on the blends indicate the presence of beta sheet structure in the

crystallized polypeptide and the folding of the chains to be predominantly in an adjacent re-entry fashion.

The results presented in this chapter are specific to **1** and the generality of the chain folding habit to other related polypeptides has yet to be established.

The series of experiments involving **1-Gx** could serve as a model system for studying the effect of well defined isotopic substitution in antiparallel beta sheets. We are unaware of any work which predicts the location and intensities of the absorptions of $^{13}C{=}O$ and $^{12}C{=}O$ as a function of their ratios in a unit cell. In the experiments involving the FTIR of **1-Gx** the intensity of the $^{13}C{=}O$ absorption is higher than that predicted by the addition of the spectra of **1-AG** and **1** in the appropriate ratios; due to the lack of any available predictions, we cannot address the issue of anomalous intensities in this work.

4.1.2 Use of Model Compounds for Prediction of Secondary Structures

The correlation of infrared spectra of polypeptides with the secondary structure has been studied since the 1960's[1-5]. The approach in most of these studies has been a comparison of predicted infrared spectra from normal mode analyses with the experimentally observed ones for model compounds; the model compounds which have used with the most success in prediction of beta sheet FTIR spectra are N-methylacetamide, polyglycine and polyalanylglycine. A comparison between the calculated

and predicted frequency values for N-methylacetamide, the shortest peptide, is shown in Table 4.1[1]. The amide A mode, which is entirely a stretching vibration, is a completely localized mode whereas the amide B which occurs in the vicinity of the amide A is thought to be a Fermi resonance of combinations of amide II[1]. The amide I is a vibration comprising predominantly the CO stretch and some contributions from CN stretch and CCN deformation. The amide II vibration has major contributions due to NH in-plane bend and CN stretch and minor contributions due to CC stretch, CO in-plane bend and NC stretch. Amide III comprises largely of NH in-plane bend and the amide IV the CO in-plane bend. The amide V predominantly comprises the CN torsion and the amide VI, the CO out-of-plane bend.

4.1.3 Use of FTIR as a Diagnostic for Secondary Structure

Changes in structure, both primary and secondary, of the polypeptide can cause changes in the vibration frequencies. However, the complexity of the spectrum and its dependence on the sequence of the polypeptides (especially for low frequency vibrations) prevents the tool from being an absolute diagnostic for the presence of alpha helices, antiparallel rippled sheets, parallel rippled sheets, antiparallel pleated sheets or parallel pleated sheets, although some fingerprints can be observed. Fortunately, however, infrared spectroscopy can be very useful for determination of beta sheet structures; its limited use in analyzing other secondary structures is reviewed by Bandekar and Krimm[1].

Table 4.1

Observed and calculated frequencies for N-methylacetamide[1]

(S=Strong, M=Medium, W=Weak)

Mode	Observed frequency (cm^{-1})	Observed frequency (cm^{-1})
Amide A	3236 S	3254
Amide I	1653 S	1646
Amide II	1567 S	1515
Amide III	1299 M	1269
NC stretch	1096 M	1070
NH stretch, CC stretch	881 W	908
Amide V	725 S	721
Amide IV	627 W	637
Amide VI	600 M	655
CCN deformation	436 W	498
CNC deformation	289 W	274
Amide VII	206 W	226

4.1.4 The Amide I and Amide II Vibrations

4.1.4.1 Transition Dipole Coupling

The amide I vibration appears to be the most useful for detection of antiparallel beta sheet conformation. First order perturbation analyses have been used to predict the vibration frequencies in the amide I and amide II regions of polypeptides for various secondary structures[6-9]. Normal mode analysis employing ordinary valence force fields cannot account for the presence of the large splitting of the amide I vibration in antiparallel beta sheets. In order to account for the large splitting, transition dipole coupling (TDC) has been invoked. The observed frequencies of vibration can be predicted by equation (1)[6,9].

$$\begin{aligned} \nu(\delta,\delta') &= \nu_0 + \sum_{s,t} D_{st}\cos(s\delta)\cos(t\delta') \\ &= \nu_0 + D_{10}\cos\delta + D_{01}\cos\delta' + D_{11}\cos\delta\cos\delta' \end{aligned} \qquad (1)$$

where, ν_0 is the unperturbed frequency, δ is the phase angle between group motions and D_{st} are the interchain and intrachain interaction constants. The unit cell corresponding to the antiparallel beta sheet conformation and the corresponding interaction constants are shown in Figure 4.1(i).

When only the first neighbors are considered in the expansion, (1) gives rise to four possible absorption frequencies, viz. $\nu(0,0)$, $\nu(0,\pi)$, $\nu(\pi,0)$, $\nu(\pi,\pi)$. The four allowed modes of vibration are shown in Figure 4.1 (ii).

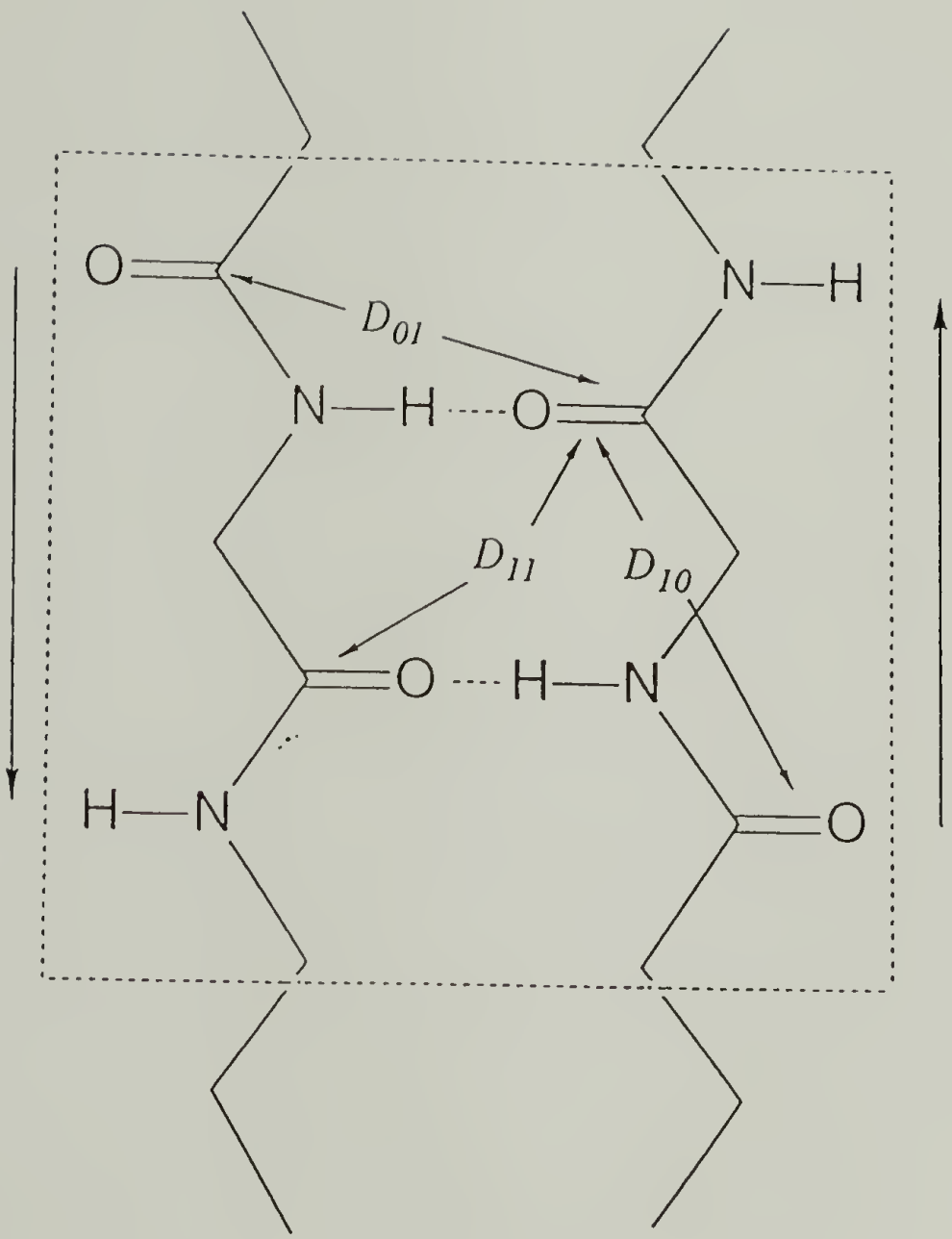


Figure 4.1(i) Unit cell of the antiparallel beta sheet structure. The interactions constants between the nearest neighbors, D_{st}, have been shown.

Figure 4.1 Antiparallel beta sheets Continued, next page

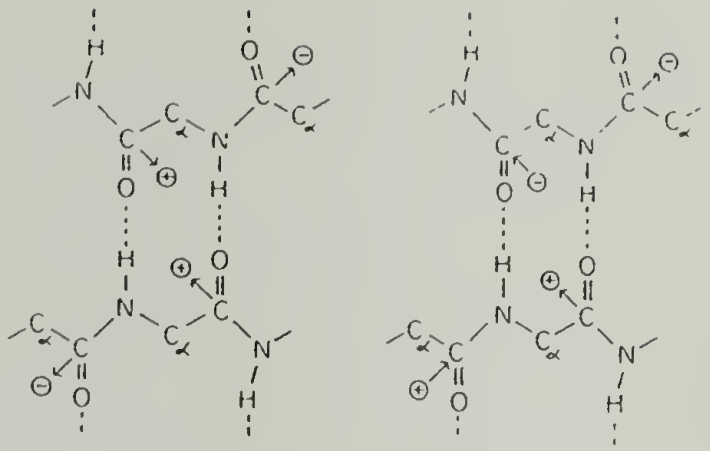


Figure 4.1(ii) The vibrational modes for antiparallel beta sheets. The arrows denote the components of the transition dipole in the plane of the figure; the plus and minus signs denote the component of the transition dipole outward and inward respectively. The phase angles of the motions are also shown; 0 is in phase and π is out of phase. (Adapted from work by Miyazawa and Blout[9])

4.1.4.2 Transition Dipole Coupling in Beta Sheet Polypeptides

The TDC analysis predicts the observed amide I frequencies for the antiparallel pleated beta sheet structure of *Bombyx mori* silk (1634 cm^{-1} and 1695 cm^{-1}) and the parallel pleated beta sheet in steam-stretched horsehair (1630 cm^{-1} and 1645 cm^{-1})[9]. Hence, the presence of both the 1634 cm^{-1} and 1695 cm^{-1} vibrations is characteristic of antiparallel pleated beta sheet structure. In order for the splitting between the two absorbances to be of ca. 60 wavenumbers, two factors are important. First, the amide groups should exist in relative locations such that the characteristic transition coupling occurs, i.e. in the antiparallel pleated beta sheet form. Second, each of the amides involved in the amide I vibration should resonate at identical frequencies, i.e. the carbon, oxygen, nitrogen and hydrogen isotopes involved in the vibration should be identical[1,10]. (Little is known regarding the strength of TDC as a function of the frequency mismatch of the amides that are involved; the references that have been cited are not complete proof of the requirement for the amides to be identical.) Hence, the degree of the transition dipole coupling, i.e. the amount of frequency splitting, can be reduced in either of two ways: First, by altering the placement of the amide groups in space from the antiparallel pleated beta sheet form, i.e. changing the secondary structure and second, by changing the unperturbed frequency of the amide groups involved in the amide I vibration, i.e. changing the isotopes of the atoms involved. Isotopic substitution of the carbon in some of the carbonyls surrounding an amide should result in a reduction of the frequency shift due to TDC of that

amide. This phenomenon was first utilized by Lansbury and coworkers to determine locations of beta sheet regions in a model amyloid protein[10].

4.1.5 Folding Habit

In the design of the repetititive polypeptides, it was anticipated that the repetitive alanyl-glycyl diads would adopt beta sheet structures, and that stacks of such sheets would constitute the lamella and define its thickness; the glutamyl-glycyl diad or the prolyl-glutamyl-glycyl triad would force the chain to fold and re-enter the lamella. The success of the assembly process (to yield structures such as that shown in Figure 1.6) depends on the ability of the chains to fold and crystallize in an adjacent re-entry fashion.

4.1.5.1 Folding in Polyethylene Crystals

Polyethylene single crystals were the first to be studied with respect to the folding habit during crystallization. Characterization of polyethylene single crystals precipitated from dilute solutions showed the formation of folded-chain lamellae with thicknesses on the order of 100 Å[11]. The preference for adjacent re-entry of the chains was demonstrated by Krimm and coworkers[12] who carried out infrared studies on blends of polyethylene with deuterated polyethylene. A comparison of their spectroscopic results with those of a normal mode analysis of unit cells containing varying fractions of deuterated polyethylene chains, led to the conclusion that the chains prefer to fold with adjacent re-entry along

certain crystallographic planes[12,13]. The elegance of these deductions is underscored by the observation that the preferred fracture planes of the crystals are coincident with the fold planes inferred from the infrared data.

4.1.5.2 Folding in Nylon Crystals

Single crystals have also been obtained and characterized for various nylons and polypeptides, by gradual precipitation of these polymers from dilute solutions[14-18]. The folding patterns for several nylons have been deduced by Atkins and coworkers from x-ray diffraction results[18,19]. We are, however, unaware of any work on nylons or polypeptides which probes for the folding habit directly.

4.1.5.3 Folding in Crystals of 1

Since the morphology desired in this work requires the chains to fold in adjacent re-entry fashion, we are especially interested in the folding habit of the polypeptide chains. Furthermore, most work on single crystals has been based on crystallization from dilute solutions. At higher concentrations the folding analysis by diffraction methods is complicated by the formation of sheaf-like or spherulitic aggregates of the single crystals[11].

4.1.5.4 Use of FTIR to Determine Folding Habit

The biosynthetic strategy allows us to place isotopically enriched amino acid residues in the sequence by supplementing the growth medium with the corresponding enriched amino acids. In this chapter are presented the infrared data for blends of the natural abundance (**1**) and ^{13}C enriched (**1-AG**) forms of the polypeptide **1** and an analysis of the folding habit during crystallization from solutions at concentrations of ca. 4 wt%. The results indicate that the chains fold predominantly in an adjacent re-entry fashion even at the high polymer concentrations studied.

4.2 Methods

4.2.1 Large Scale Synthesis of 1 and its Enriched Analogues

Fermentation on a 12L scale was carried out as described previously in M9 minimal growth medium[20] supplemented with L-amino acids (20 μg/mL), glucose (2 mg/mL) and vitamin B1 (1 μg/mL). The biosynthetic strategy yielded a precursor of **1** bearing a short (25-residue) N-terminal residue sequence (**FUS**). The presence of amino acids of natural abundance in the medium yielded a precursor of **1**; large scale fermentation with ^{13}C-enriched glycine and alanine in the growth medium (99% ^{13}C, Cambridge Isotope Laboratories, Cambridge, MA) yielded a precursor of **1-AG**, containing the heavier isotope at the carbonyl positions of alanine and glycine. In order to synthesize precursors of **1-G100**, **1-G60** and **1-G30**, the media were supplemented with 100% 1-^{13}C glycine, 60% 1-^{13}C glycine and

30% 1-^{13}C glycine. The protein in each case was purified by a four step reduction in the pH of the cell extract leading to precipitation of impurities at every step; at every step 100 mL of glacial acetic acid was added to the solution but the pH was not measured at any of the intermediate steps. Ethanol was added to the supernatant at pH=4.0 to a final concentration of 40% and the mixture was incubated at 4 °C for 1 h. After centrifugation, ethanol was added to the supernatant to a final concentration of 80% and the mixture was incubated overnight at -20 °C. The purified protein was collected by centrifugation of the 80% ethanol suspension at 8000 rpm and dried overnight in a vacuum oven at room temperature. The polypeptide of interest was obtained from the precursor by cleavage with cyanogen bromide. Cyanogen bromide was added to a ca. 2.5 mg/mL solution of the polypeptide in 70% formic acid to a concentration of ca. 2 mg/mL. The system was purged with nitrogen and the flask was allowed to stand at room temperature in the dark for 24 h. Solvent was removed on a rotary evaporator and distilled water was added to the dried protein fragments to make the concentration of the protein ca. 6 mg/mL. The mixture was stirred for 24 h to extract the soluble proteins. Centrifugation at 8000 rpm gave a clear supernatant which contained 1, which was precipitated by addition of ethanol to a final concentration of 80%. The precipitate was dried *in vacuo* at room temperature and then over P_2O_5 in a drying pistol with refluxing ethanol.

4.2.2 Crystallization of **1**, Blends of **1** with **1-AG**, and with **1-Gx**; **x**=100, 60, 30

Polypeptide blends of the natural abundance (**1**) and ^{13}C enriched (**1-AG**) forms of the polypeptide **1** were crystallized from formic acid as follows: The mixture was dissolved in 90% formic acid and the concentration of the formic acid was reduced to 70% by the addition of distilled water to obtain a 40 mg/mL protein solution. The solution was sheared for ca. 24 h by continuous stirring at ca. 700 rpm on a Thermix stirrer model 220T (Fisher Scientific). The sample formed a gel which was washed with water several times and then with methanol. The gel was sedimented by centrifugation and the precipitate was washed several times with water and methanol and dried over P_2O_5 in a drying pistol with refluxing ethanol. The same crystallization procedure was followed for the copolymers.

4.2.3 FTIR Spectroscopy

FTIR spectra were obtained on a Nicolet IR44 bench driven by PC/IR Version 3.00 software at a resolution of 1 cm^{-1} with 50 scans per spectrum. Sample pellets were made with IR grade KBr on a 'Macro-micro KBr die' (Aldrich Chemical Co.) on a 12 ton two column laboratory press.

4.3 Results and Discussion

4.3.1 Determination of Labeling Efficiency

The incorporation of ^{13}C at the carbonyl positions of alanine and glycine in **1-AG** was confirmed by ^{13}C NMR. The labelling efficiency of the carbonyl carbon in the enriched polypeptide was estimated from the ^{13}C NMR spectrum to be 90 ±10%. The incorporation efficiency of the expression strain was 90±10% for glycine. Figure 4.2(i) shows a solution ^{13}C NMR spectrum of **1-G100**; Figure 4.2(ii) shows a solution ^{13}C spectrum of **1-AG**; the resonances between 170-172 ppm are due to the glycine carbonyl and those between 174-176 ppm are due to the alanine carbonyl; Since the incorporation efficiency of the expression strain was close to 100% for glycine, it implied that the ratio of the enriched glycine to the unenriched glycine in **1-Gx** was equal to the concentration ratio of the amino acids in the medium. This was confirmed by NMR; Figure 4.2(iii)-(v) show the solid state ^{13}C NMR CPMAS spectra of **1-G100**, **1-G60** and **1-G30** respectively. In all these samples, the carbonyl region was dominated by the carbonyl of glycine; the ratios of the integrated carbonyl intensity to the integrated beta carbon (of alanine) intensity for the three samples scaled as 22.24: 13.32: 6.42 (=100: 59.9: 28.9). This result in conjunction with almost 100% incorporation as seen from (i), leads us to conclude that the samples were enriched with 1-^{13}C glycine at levels of 100% (**1-G100**), 60% (**1-G60**) and 30% (**1-G30**). It was assumed that the placement of the labeled glycine along the chain was random.

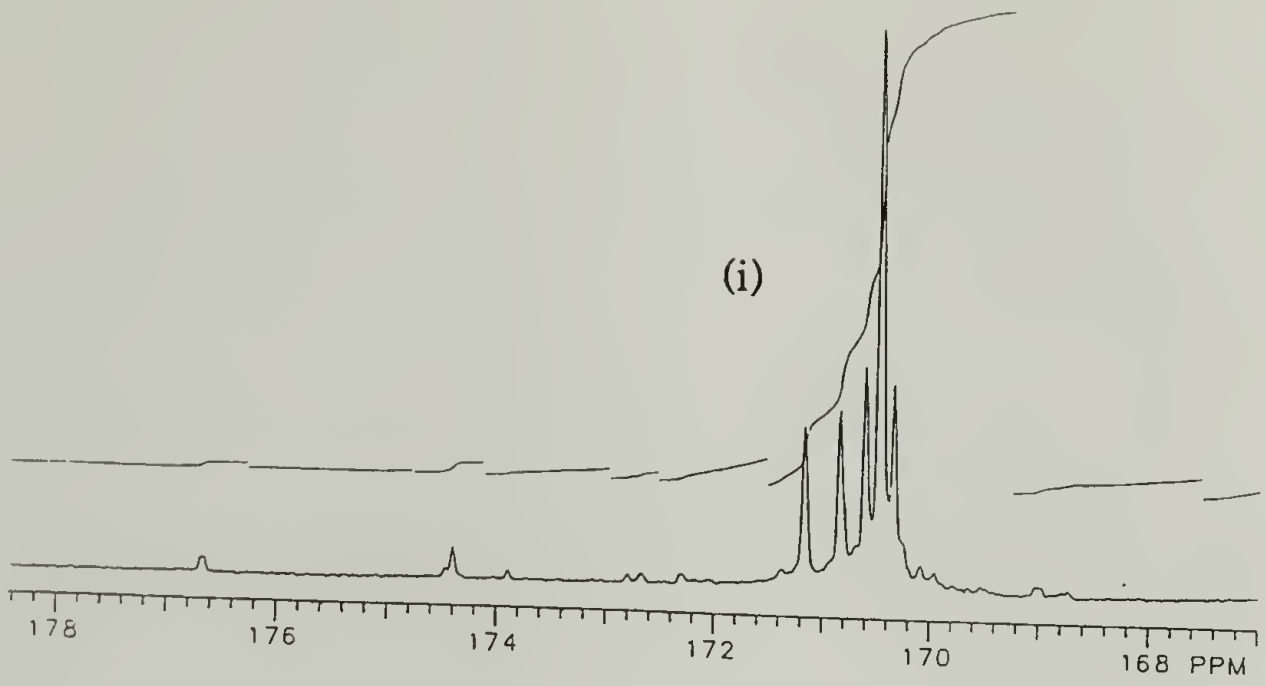
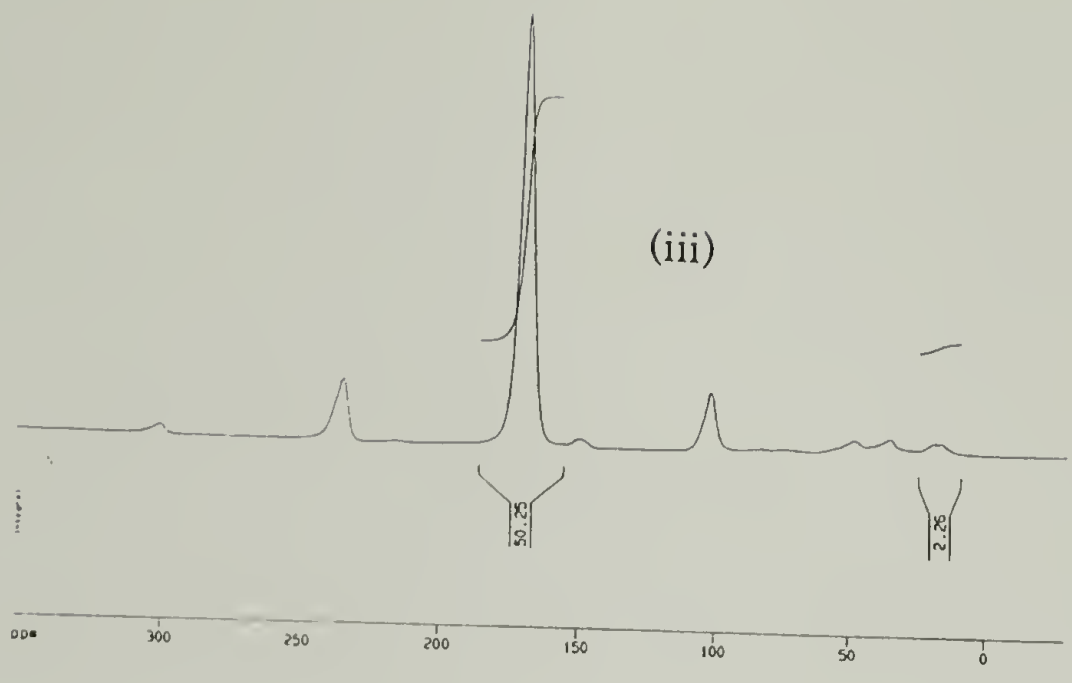
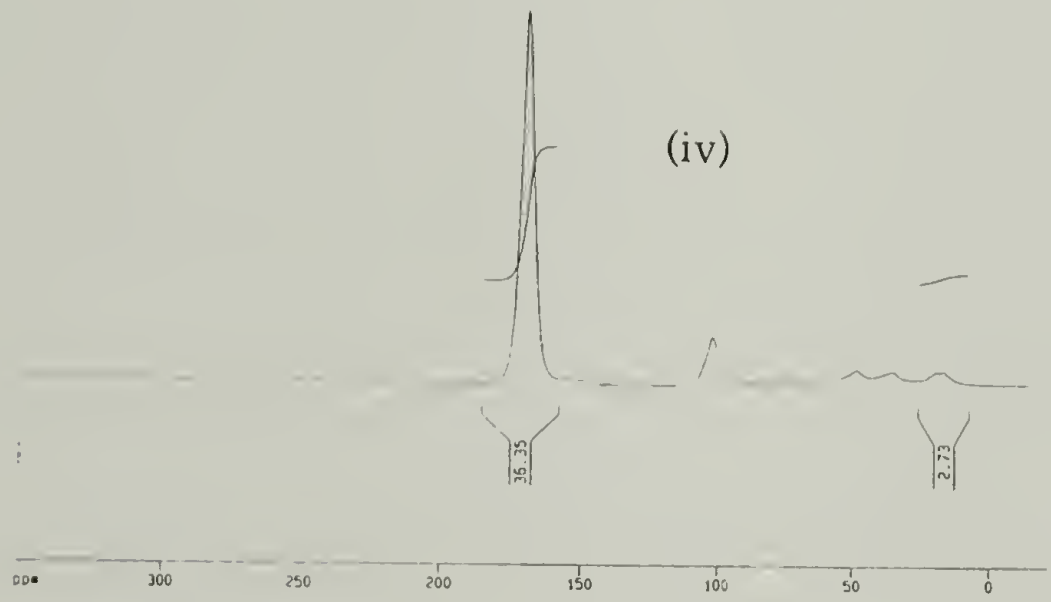


Figure 4.2 NMR spectra of labelled polypeptdies. (i) ^{13}C solution NMR of **1-G100**; (ii) ^{13}C solution NMR of **1-AG**; (iii) ^{13}C CPMAS spectrum of **1-G100**; (iv) ^{13}C CPMAS spectrum of **1-G60**; (v) ^{13}C CPMAS spectrum of **1-G30**.

Continued, next page





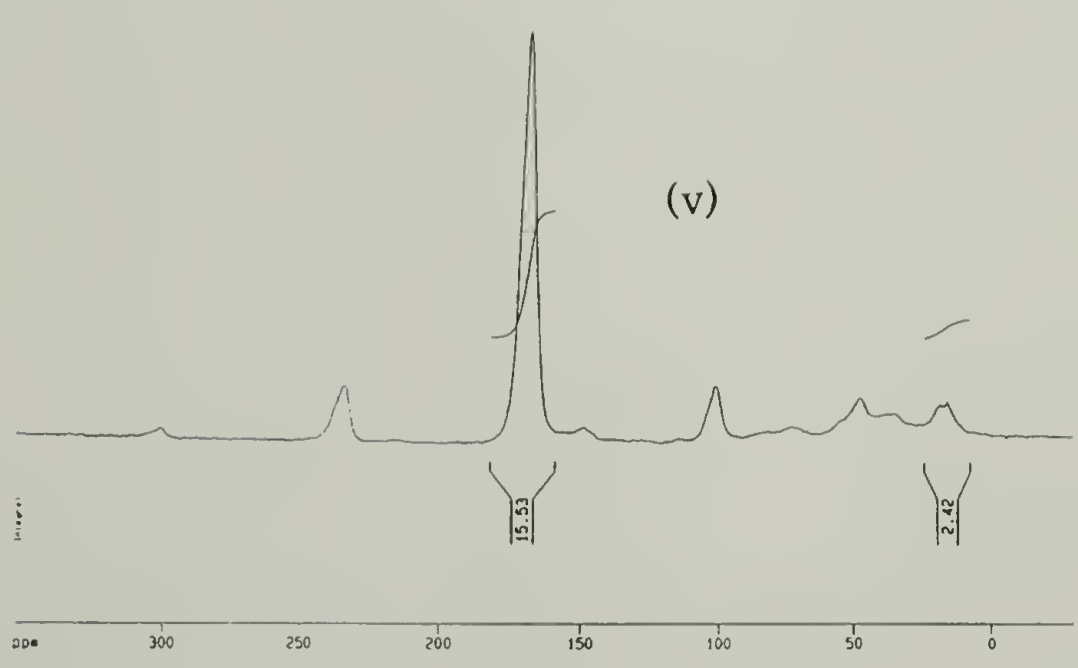


Figure 4.2 (Contd.)

4.3.2 Determination of Secondary Structure in Crystallized 1

The infrared spectrum of 1, as crystallized from 70% formic acid is shown in Figure 4.3. The amide I vibration, shows the characteristic splitting corresponding to antiparallel beta sheet structure (the vibrations occur at 1626 cm^{-1} and 1702 cm^{-1}). The amorphous form of 1 shows a broad absorption at 1655 cm^{-1}. The detection of antiparallel beta sheets by FTIR corroborates the evidence from the x-ray diffraction experiments.

4.3.3 Chain Folding of 1

4.3.3.1 Effect of Isotope Substitution on Transition Dipole Coupling

Figure 4.1 shows a unit cell of an antiparallel beta sheet structure. The accompanying transition dipole coupling has been discussed in the introduction. All the experiments in the literature to study the phenomenon of transition dipole coupling have used polypeptides with $^{13}C{=}O$ in natural abundance[1,8,9]; since the natural abundance of $^{13}C{=}O$ is only 1%, this means that all four carbonyls in Figure 4.1(i) will be of the same frequency. There have been no systematic efforts made to determine the change in the amide I vibration as a function of the $^{13}C{=}O$ composition in the unit cell. Since there are four carbonyls in the unit cell, this leads to 2^3 possible combinations of unit cells containing varying amonts of $^{13}C{=}O$ in varying positions in the unit cell. *Ab initio* calculations on model compounds have been performed to predict the amide I region of natural

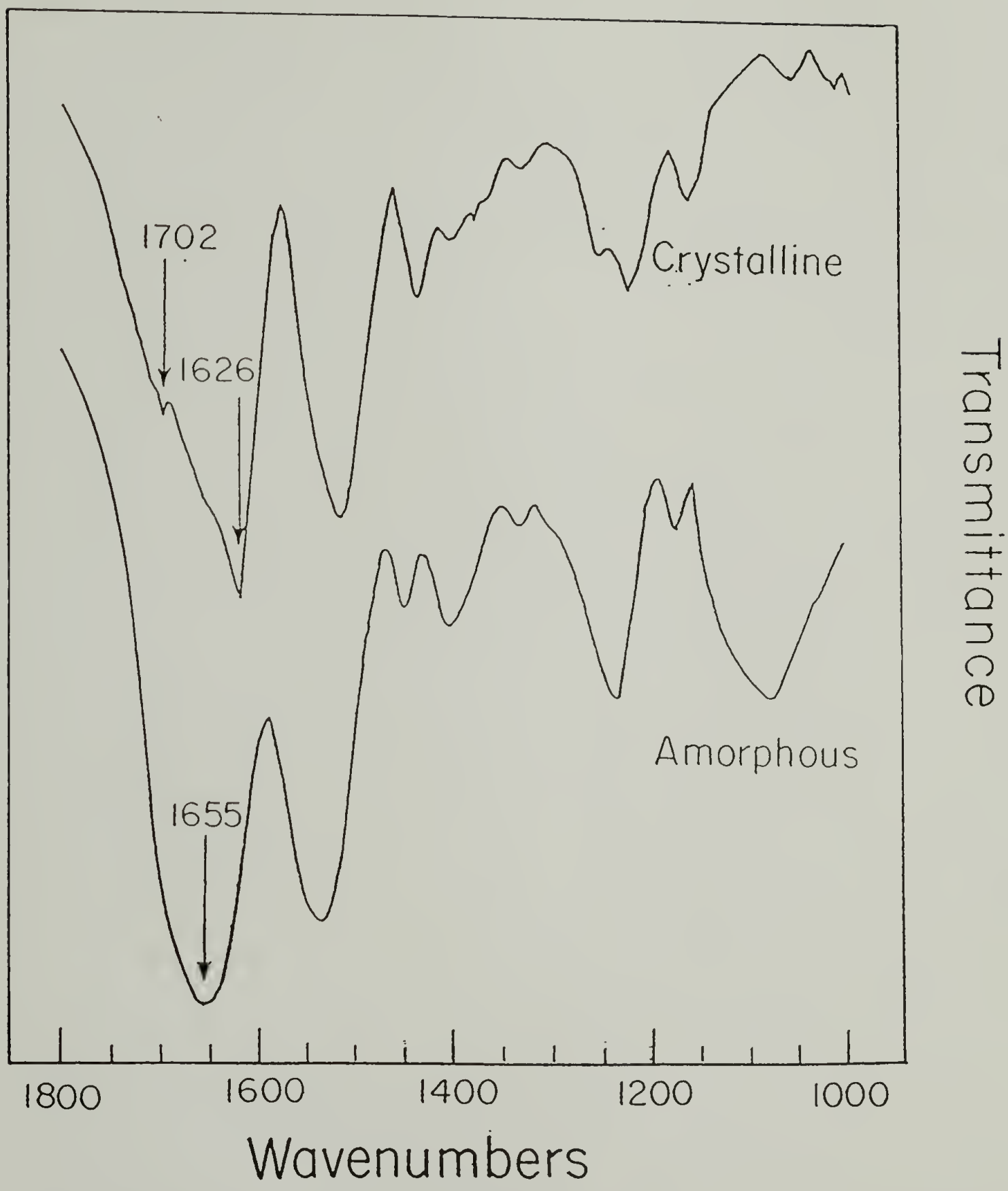


Figure 4.3 FTIR spectrum of **1** as crystallized from formic acid (crystalline) and as quenched from an aqueous solution by addition of ethanol (amorphous).

abundance polypeptides[7] but the variation of the spectrum as a function of the ^{13}C composition has not been predicted.

Since the amide I mode of vibration has a large contribution from the CO stretch, its frequency is dependent on the mass of the carbonyl carbon atom. The stretching frequency is given by $\vartheta_{AB} = 1/2\pi\sqrt{\frac{k}{\mu}}$; where $\mu = \frac{M_A M_B}{M_A + M_B}$ and k is the force constant for the vibration. The unperturbed frequency for the ^{13}C=O amide I vibration is expected to occur at ca. 1620 cm^{-1} because of the higher mass of the carbon atom[1,10]; the actual frequency may vary slightly because the amide I vibration is not entirely a stretching vibration. When all four carbonyls in a unit cell are ^{13}C=O, the antiparallel beta sheet conformation should give rise to two characteristic vibrations, with the more prominent at ca. 1591 cm^{-1}[10]. In this work the prominent vibrations for ^{12}C=O and ^{13}C=O will be used for analysis, viz. the 1626 cm^{-1} and 1596 cm^{-1} respectively.

4.3.3.2 Deconvolution of Spectra of Crystalline **1** and **1-AG**

The actual and deconvoluted spectra of the amide I region for natural abundance **1** in the antiparallel beta sheet are shown in Figure 4.4(i); the corresponding spectra for **1-AG** are shown in Figure 4.4(ii). The spectrum for **1-AG** is not a simple wavenumber shifted spectrum of **1**; although the prominent absorbance at approximately 1626 cm^{-1} is shifted to approximately 1596 cm^{-1}, the sharp absorbance shoulder at 1702 cm^{-1}

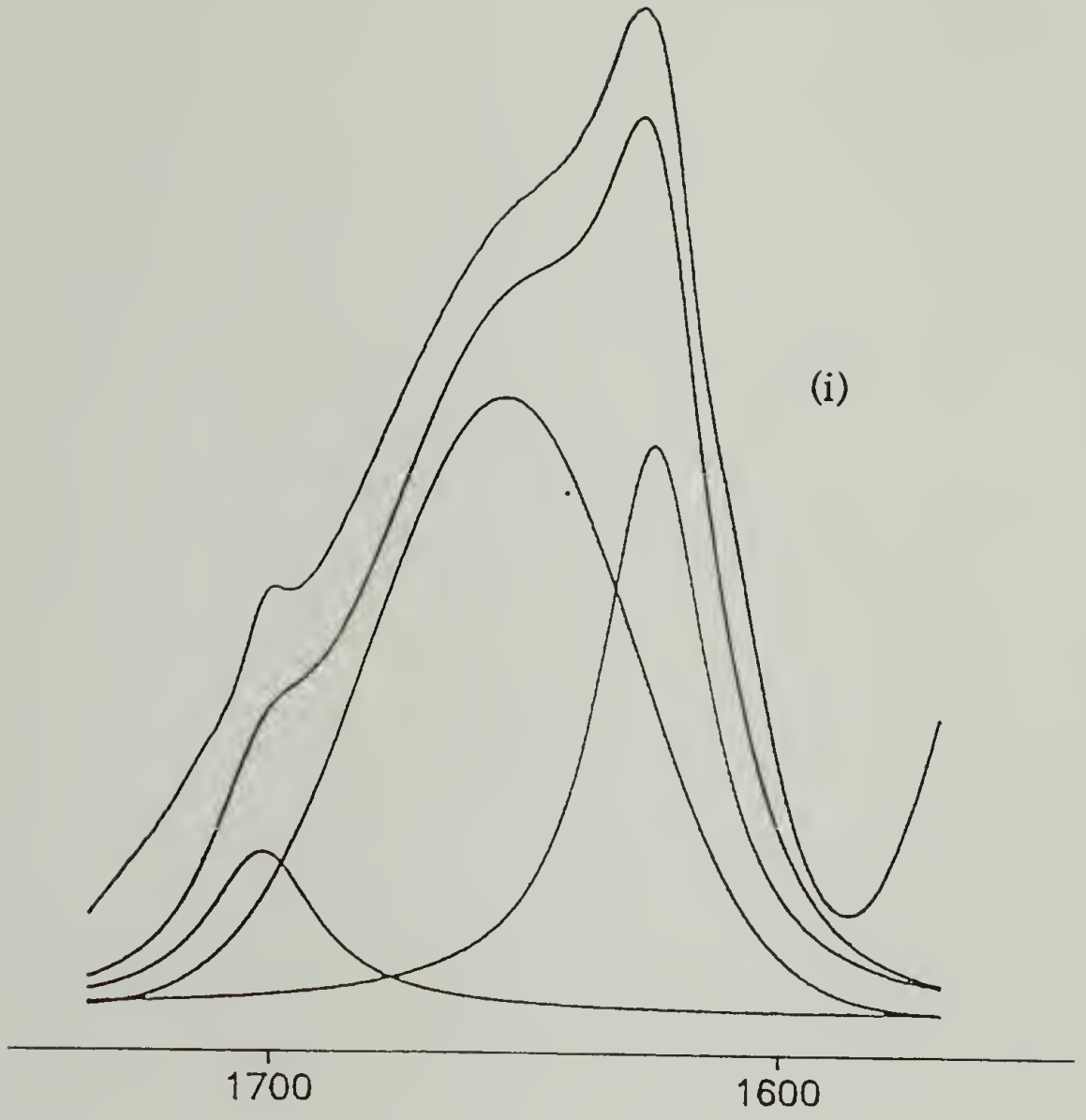


Figure 4.4 Deconvolution: (i) The upper spectrum is the observed spectrum for crystalline **1** and the lower spectrum is the addition spectrum generated from three constituent absorbances. (ii) shows a similar analysis for **1-AG**.

Continued, next page

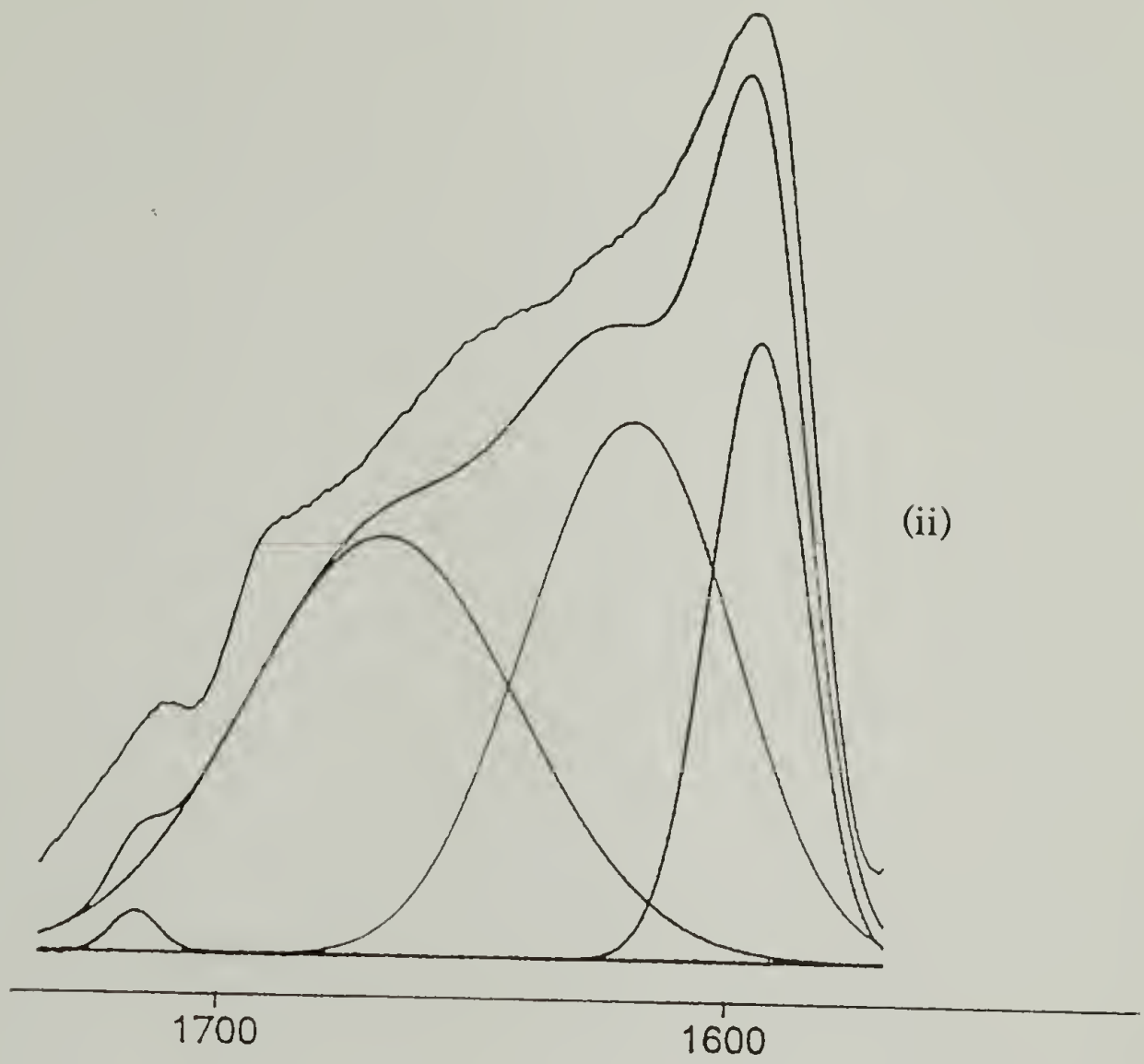


Figure 4.4 Contd.

in Figure 4.4(i) does not have a counterpart at 1670 cm^{-1} in Figure 4.4(ii). The reasons for the absence of this shoulder are unclear; this is the first time a spectrum of a polypeptide in the antiparallel beta sheet structure, with all the $^{12}C=O$ replaced with $^{13}C=O$ is being reported. As has already been stated before, there are some aspects of the data which cannot be explained due to the lack of a theoretical framework in the literature.

4.3.3.3 Analysis of Spectra of Crystalline 1-Gx

Figure 4.5 shows the spectra in the amide I region of the series of polypeptides **1-Gx** in the beta sheet form. Figure 4.5(a) shows the spectrum for **1**; the strong absorbance is observed at approximately 1626 cm^{-1} and the weak shoulder is seen at 1702 cm^{-1}. Figure 4.5(e) shows the spectrum for **1-AG**; the strong absorption is observed at 1596 cm^{-1} but no sharp shoulder is observed; a weak broad shoulder is seen at 1662 cm^{-1} and another one at 1712 cm^{-1}. In the spectrum corresponding to **1-G100** (Figure 4.5(d))two major absorbances are observed, one at 1607 cm^{-1} and the other at 1636 cm^{-1}. The absorbance at 1607 cm^{-1} remains at the same location, irrespective of the concentration of the copolymer composition, while the one at higher wavenumber changes from 1636 cm^{-1} to 1630 cm^{-1} as the amount of $^{13}C=O$ glycine in the copolymer is reduced.

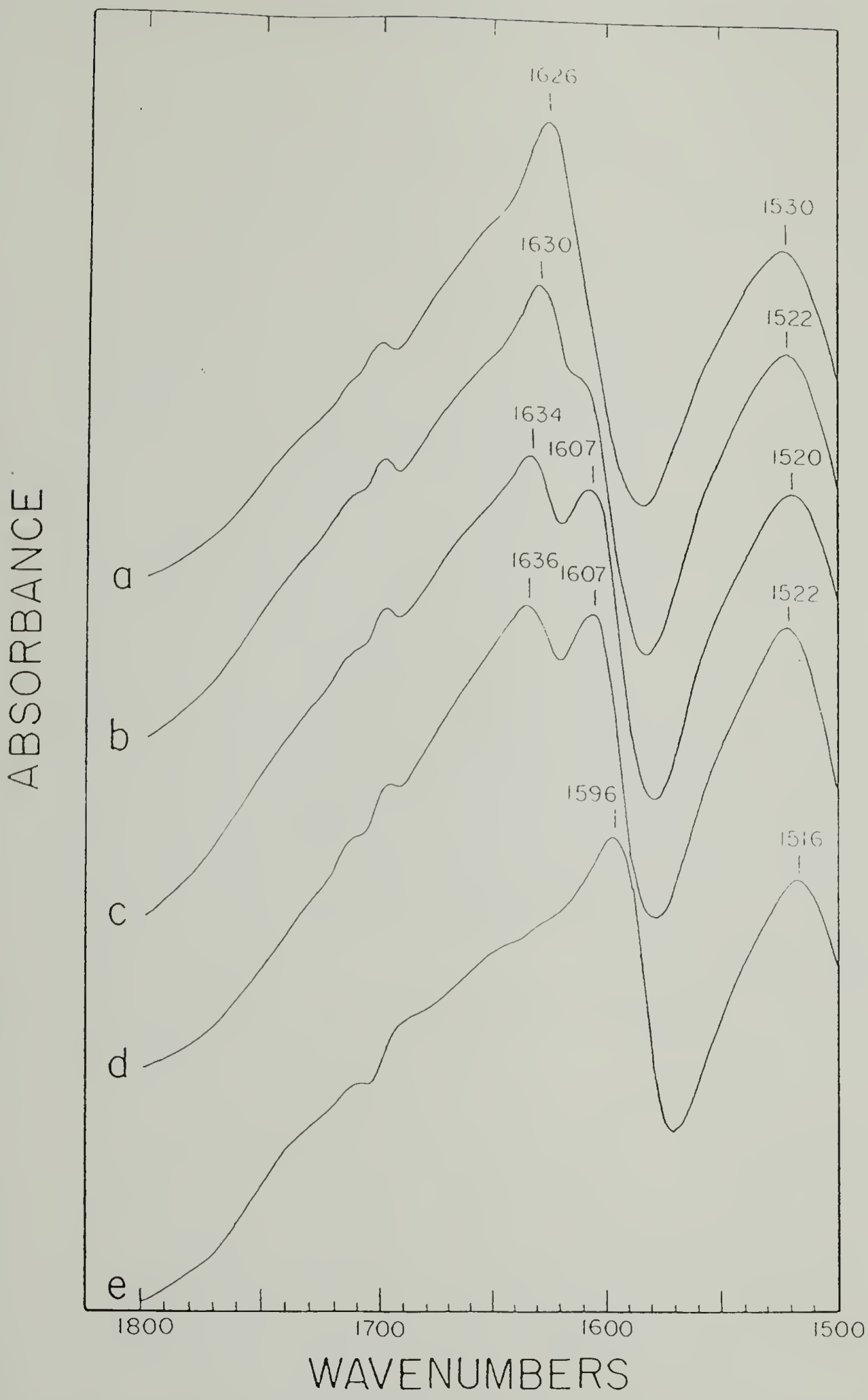
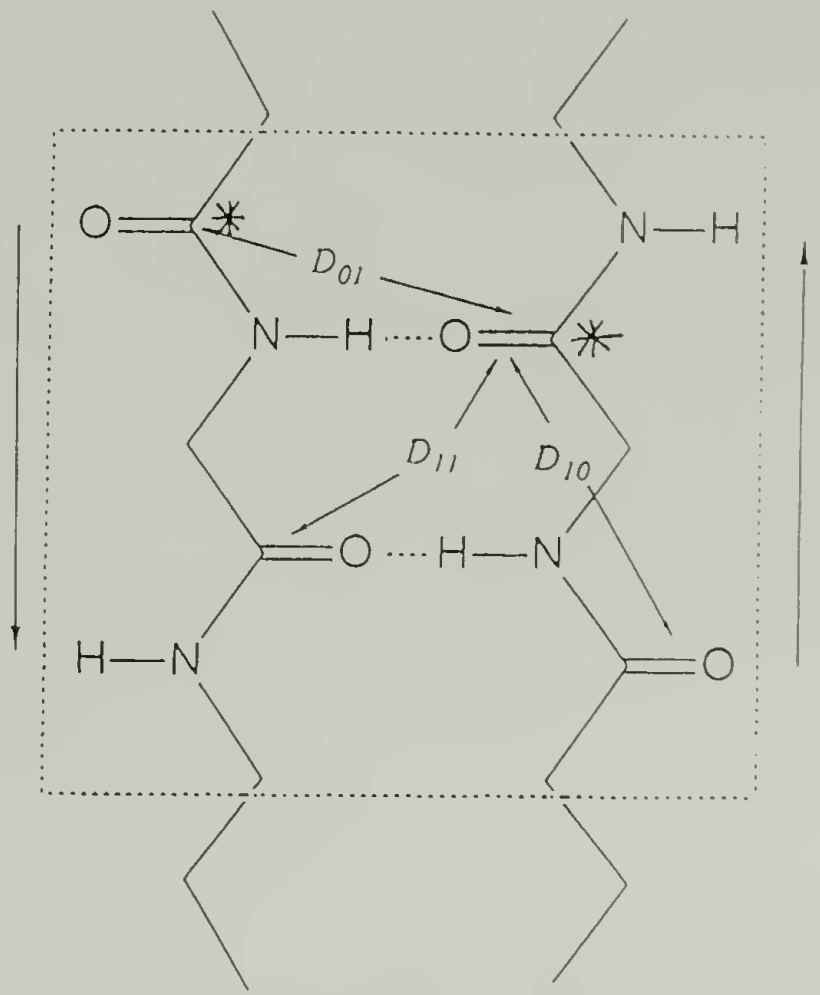


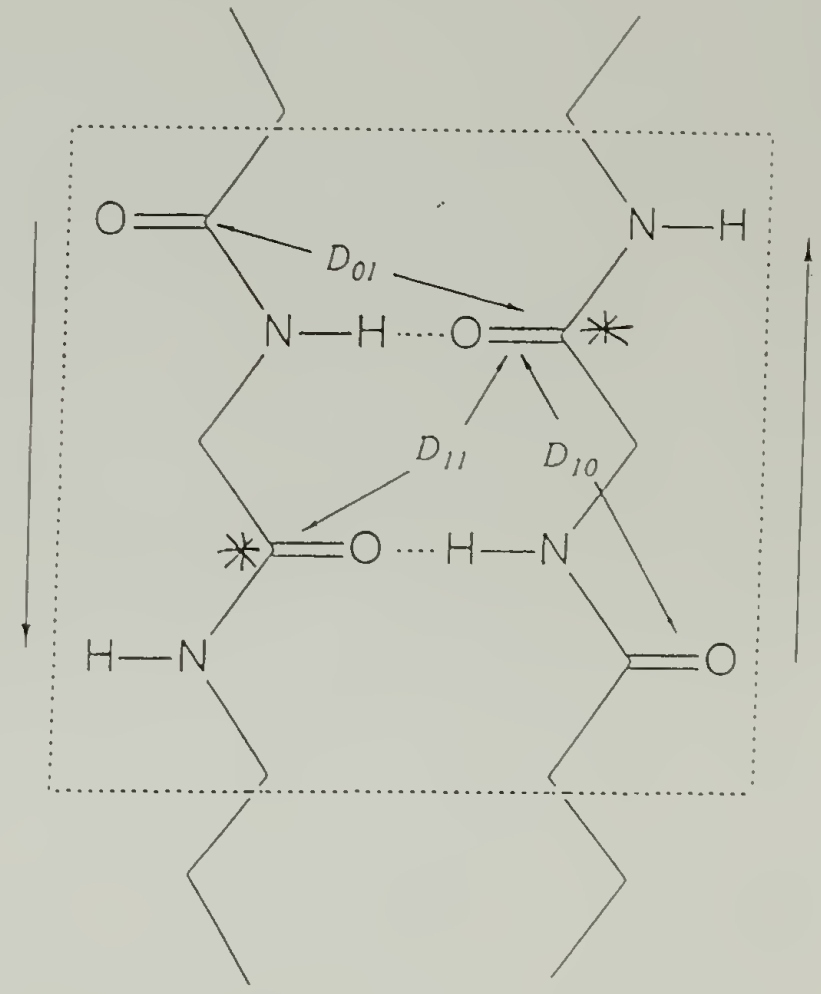
Figure 4.5 FTIR spectra of (a) crystalline 1; (b) crystalline 1-G30; (c) crystalline 1-G60; (d) crystalline 1-G100; (e) crystalline 1-AG;

4.3.3.4 Consequence of Labeling with $^{13}C{=}O$ Glycine in **1-G100**

The unit cell in the antiparallel beta sheet structure of **1-G100** contains two $^{13}C{=}Os$ and two $^{12}C{=}Os$. Since the alanines and glycines alternate along a chain, the $^{13}C{=}Os$ and the $^{12}C{=}Os$ alternate along a chain also. Hence, the unit cell has to correspond to either the unit cell depicted in Figure 4.6(i) or Figure 4.6(ii). If the crystalline structure of **1** is as shown in Figure 4.6(ii), the methyls of alanine would decorate only one face of the antiparallel beta sheet; this would lead to a polar sheet. However, if the registry of adjacent chains is as shown in Figure 4.6(i), apolar sheets would be formed. The evidence from x-ray diffraction suggests formation of apolar sheets (refer to Chapter 3). Hence, the glycines are placed in adjacent chains within a sheet as shown in Figure 4.6(i). This would then support the hypothesis that all the antiparallel beta sheet unit cells in the crystalline sample of **1-G100** are as shown in Figure 4.6(i). Such a unit cell would reduce the transition dipole coupling for both $^{13}C{=}Os$ and $^{12}C{=}Os$ in comparison to the case where the unit cells are comprised entirely of $^{13}C{=}Os$ or $^{12}C{=}Os$; this should be manifested in movement of the prominant absorption to a higher wavenumber. This is indeed observed, since the $^{13}C{=}Os$ absorb at 1607 cm^{-1} instead of 1596 cm^{-1} and the $^{12}C{=}Os$ absorb at 1636 cm^{-1} instead of 1626 cm^{-1}.



(i)



(ii)

Figure 4.6 Predicted unit cells for **1-G100** in antiparallel beta sheet structure. ^{13}C is denoted by an asterisk (*).

4.3.3.5 Correlation of Amide I Intensity and Location with $^{13}C{=}O$ Content of Unit Cell

The deconvolution of the spectra corresponding to crystalline **1** and **1-AG** have been discussed in Section 4.3.3.2 and Figure 4.4. The real spectra are fit by addition of three peaks (in case of **1**) and four peaks (in case of **1-AG**); because this corresponds to six and eight fitting parameters respectively, the uncertainty developed due to the large number of parameters limits the use of deconvolution and prevents more quantitative analysis. At this juncture, very little is known about the lineshapes of the component peaks arising from transition dipole coupling; the presence of $^{13}C{=}Os$ and $^{12}C{=}Os$ in statistical variation within the unit cell further confounds the prediction of lineshapes since their interaction is ill-characterized. This statistical variation becomes important in the spectra of partially labeled polypeptides, **1-G60** and **1-G30** and in the spectra of blends of **1** with **1-AG**. In the region between 1626 cm^{-1} and 1700 cm^{-1} absorption due to a large uncharacterized population of different unit cells could exist. However, the region between 1596 cm^{-1} and 1626 cm^{-1} is predominantly due to the $^{13}C{=}O$ transition dipole coupling; hence, we believe that this region is the most diagnostic and does not suffer from the shortcomings of the region between 1626 cm^{-1} and 1700 cm^{-1}.

We believe that the transition dipole coupling is reduced in **1-G100** due to the reduction in the content of $^{13}C{=}O$ in the unit cell as evidenced by the absorption at 1607 cm^{-1} instead of 1596 cm^{-1}; although the $^{12}C{=}O$

absorbs at 1636 cm^{-1} instead of 1626 cm^{-1} as anticipated, we do not think that it is a reliable piece of evidence (due to reasons outlined above). In Figure 4.5, while the position of the 1636 cm^{-1} absorption varies slightly, the position of the 1607 cm^{-1} absorption is at the same location, although the intensity is seen to reduce in **1-G100** through **1-G30**; the reduction in intensity is due to reduction in the fraction of unit cells containing two ^{13}C=Os. Since the deconvolution and the subsequent quantitative estimation is not robust, we will base our analysis on the location of the absorption rather than the intensity of the absorption; we believe that the analysis based on the location of the maximum between 1596 cm^{-1} and 1630 cm^{-1} is very robust.

There is no previous work in the literature which correlates the position of the ^{13}C=O absorption to the content of ^{13}C=O in the antiparallel beta sheet unit cell. If the absorption of the antiparallel beta sheet ^{13}C=O were not coupled with its neighbors, we would anticipate it not to change its frequency of absorption; this would imply that the ^{13}C=O would absorb at 1596 cm^{-1} irrespective of its local neighborhood. Furthermore, this would require the absorption spectrum of **1-Gx** to be a linear combination of the spectra of **1** and **1-AG** with the appropriate ratios.

Figure 4.7 shows the spectra of **1-G100**, **1-G60** and **1-G30** respectively along with the spectra generated by the addition of the spectra of **1** and **1-AG**; in the addition spectra corresponding to **1-Gx**, the ratio of **1-AG**:**1** was x/(200-x). Table 4.2 lists the observed and simulated maxima in absorbance. In all instances, the observed low frequency absorption is at

1607 cm^{-1}, while the absorption in the simulated spectra varies slightly around 1598 cm^{-1}. The slight deviation from 1596 cm^{-1} is due to the mathematical reason mentioned in section 4.3.3.6.2. We cannot explain the additional intensity of the 1607 cm^{-1} absorption. This data suggests that the position of the low frequency absorption is dependent on the $^{13}C{=}O$ environment and increases in frequency as the amount of $^{13}C{=}O$ in the unit cell is reduced; a similar trend is observed for the $^{12}C{=}O$ absorption. The position of the $^{12}C{=}O$ absorption is however not fixed at 1636 cm^{-1} for all copolymer compositions. The reason for this is suggested in section 4.3.3.5.

Table 4.2

Location of absorption maxima for the copolymer series **1-Gx**

	$^{12}C{=}O$ Absorption (cm^{-1})		$^{13}C{=}O$ Absorption (cm^{-1})	
	Sample	Simulation	Sample	Simulation
1-AG			1596	
1-G100	1636	1627	1607	1598
1-G60	1634	1626	1607	Shoulder (~1598)
1-G30	1630	1626	1607	1598
1	1626		-	

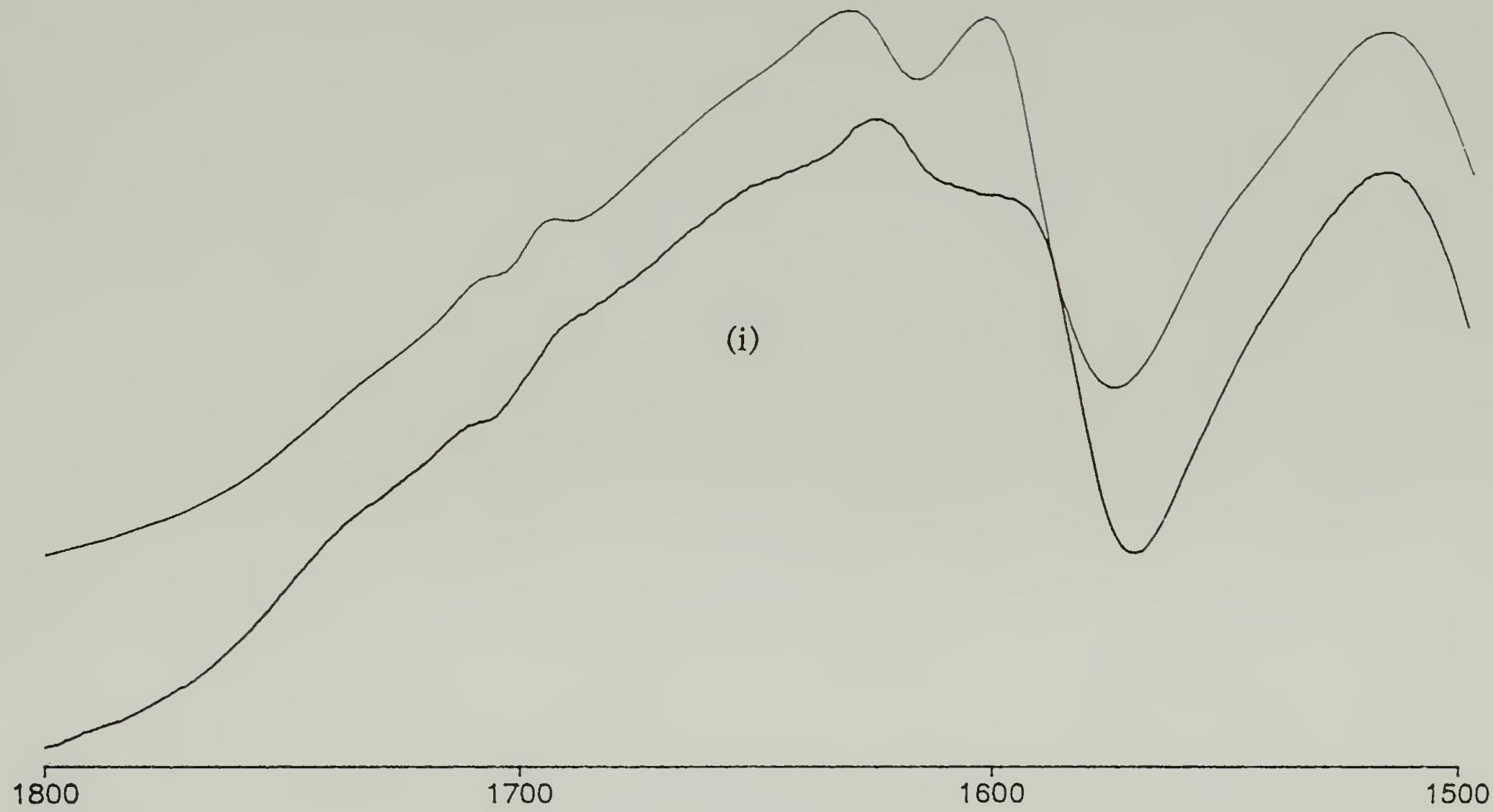


Figure 4.7 Computer generated addition absorption spectra of 1 and 1-AG in the ratios of the copolymer compositions; the actual spectrum (above) is also shown in each case. Refer to the text for details. The spectra correspond to (i) 1-G100; (ii) 1-G60; (iii) 1-G30. Continued, next page

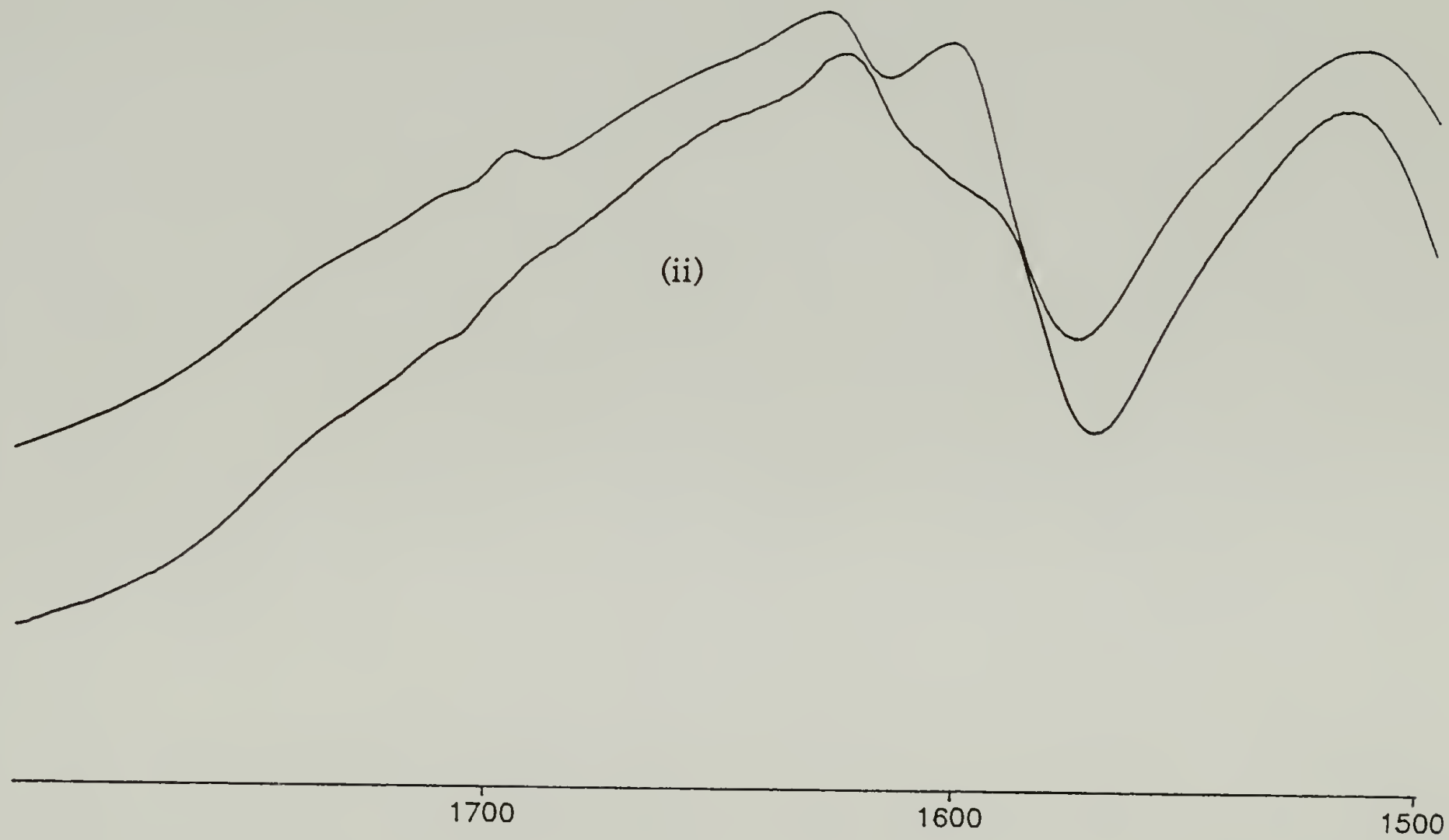


Figure 4.7 (Contd.)

Continued, next page

Figure 4.7 (Contd.)

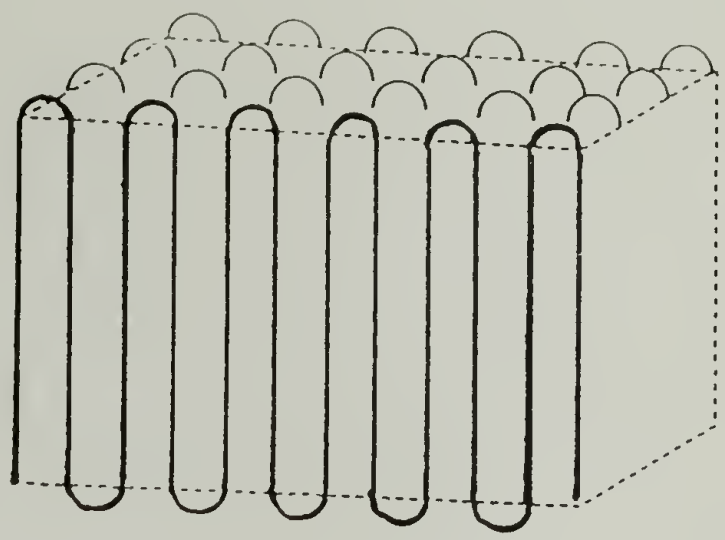
4.3.3.6 Cocrystalliztion of **1** and **1-AG**

In the experiment to determine the folding habit of **1**, varying fractions of **1** were solution blended with **1-AG**. A schematic of the self assembled lamellar morphology is shown in Figure 4.8(i).

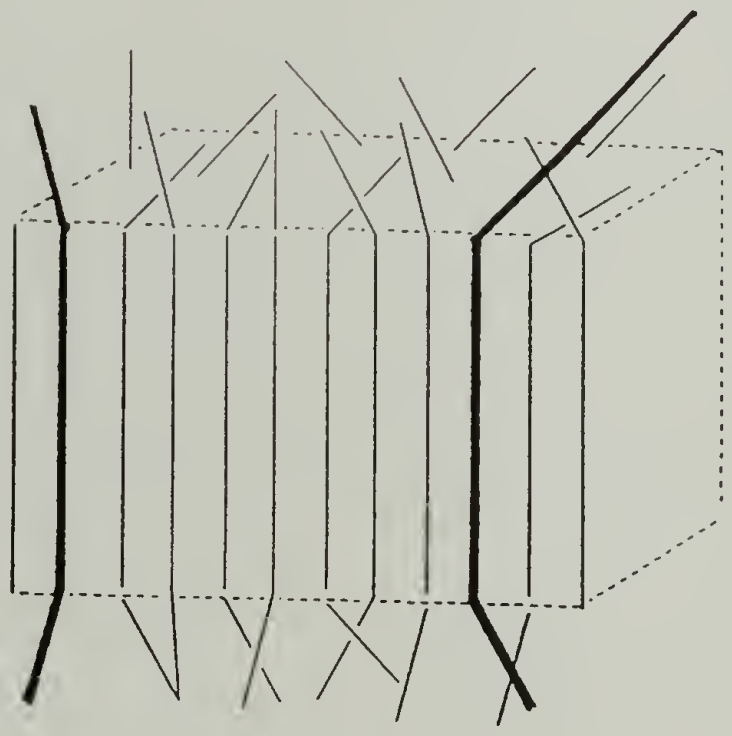
4.3.3.6.1 Consequence of Folding Habit on Relative Placement of $^{13}C{=}O$ and $^{12}C{=}O$

Figure 4.8(i) shows the crystallization of the repetitive polypeptide with adjacent re-entry. In this model, the 'stem' of the polypeptide consists of three alanyl-glycyl or glycyl-alanyl diads and the 'turn' region consists of a glycyl-glutamyl-glycyl triad. The thicker line represents the ^{13}C-enriched stem; the thinner line represents the natural abundance polypeptide. The corresponding unit cells are shown in Figure 4.7(i) and (ii); ^{13}C labelled chain segments are always present adjacent to other ^{13}C labelled segments from the same chain in this mode of chain folding; similarly, ^{12}C chain segments are always present adjacent to other ^{12}C segments from the same chain. Any other form of chain folded morphology would lead to some ^{12}C chain segments being adjacent to ^{13}C labelled chain segments.

Figure 4.8(ii) shows a fringed micellar morphology and Figure 4.9(iii) and (iv) show the corresponding unit cells. In this case, the segments shown with a thicker line are not always adjacent to other segments with thicker lines.



(i)

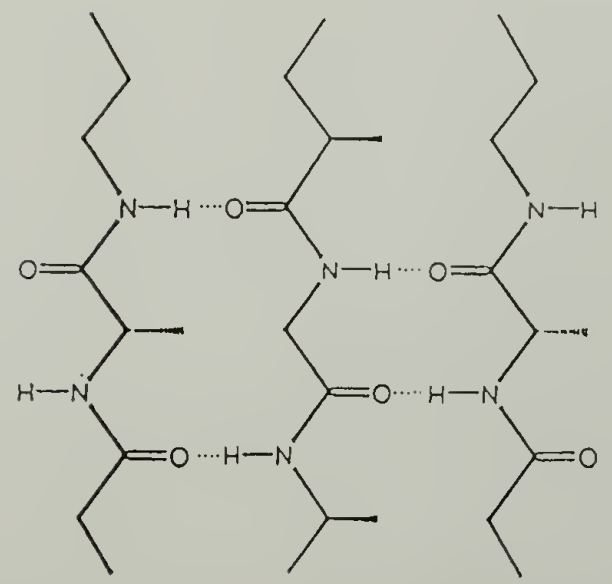


(ii)

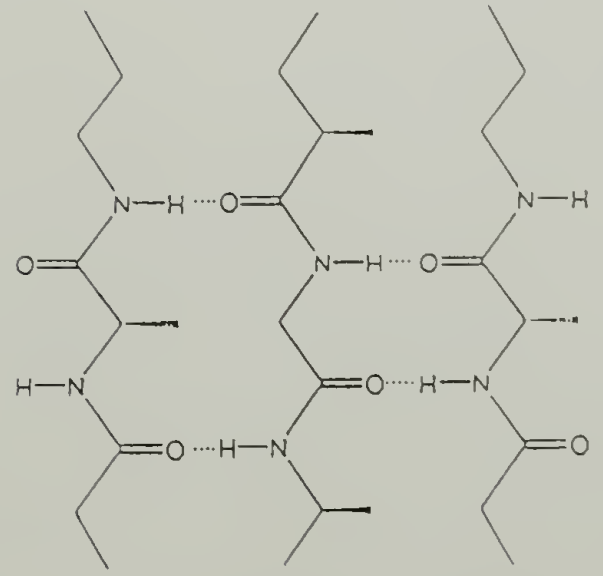
Figure 4.8 Formation of lamellae with: (i) adjacent reentry and (ii) fringed micellar morphology. The bold line represents **1-AG** and the thinner lines represent **1**.

Figure 4.9 Unit cells corresponding to formation of lamellae by: (i) and (ii) adjacent reentry and (iii) and (iv) fringed micellar morphology.

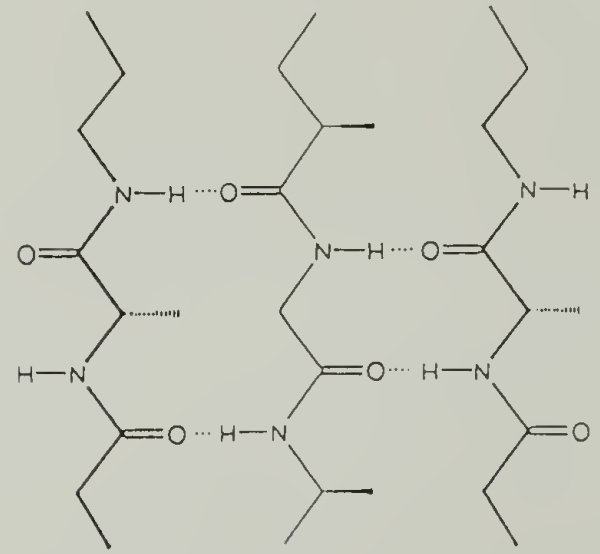
(I)



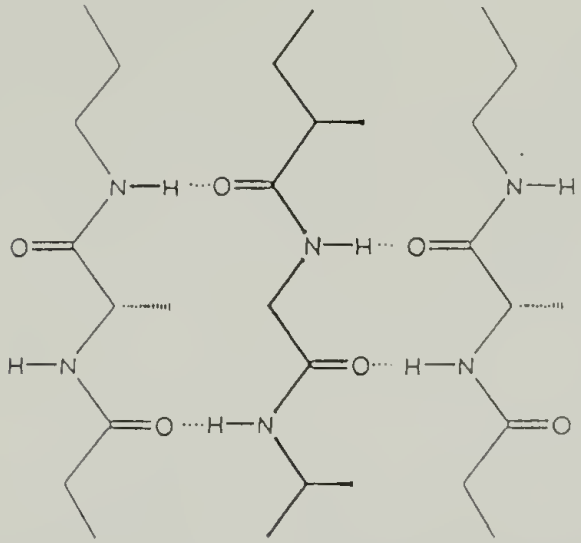
(II)



(III)



(IV)



4.3.3.6.2 FTIR Spectra of Blends of **1** and **1-AG**

We have studied the amide I region of the infrared spectra of blends containing **1** and **1-AG** in varying ratios. The spectrum corresponding to 100% of **1** shows the characteristic absorption at 1629 cm^{-1} due to the TDC of ^{12}C carbonyls in the antiparallel beta sheet structure (Figure 4.10). The TDC of ^{13}C carbonyls in the antiparallel beta sheet structure leads to an absorption at 1596 cm^{-1} in **1-AG** as shown in the sample containing 100% of **1-AG**. All intermediate blend compositions show the presence of both bands in the amide I region and the relative absorption intensities scale qualitatively with the composition of the blend. The positions of the two main absorptions are not fixed at 1629 cm^{-1} and 1596 cm^{-1} in all of the blend compositions. The slight changes in the frequencies of absorption as a function of blend composition may be due to two reasons: 1) When two distributions are added, it is not uncommon to see slight movements of the position of the apparent maximum towards the direction of the distribution with higher integrated intensity; the effect is more pronounced when the distributions are asymmetric[21]. 2) Contributions from intersheet interactions between carbonyls; these have never been accounted for in any of the transition dipole coupling studies that have discussed before.

Figure 4.11 shows computer generated addition spectra of **1** and **1-AG** with varying ratios and the absorption spectra obtained for the blends. Table 4.3 tabulates the observed the calculated absorption locations.

Figure 4.10 Infrared spectra of blends of **1** and **1-AG**. The weight percentage of ^{13}C-enriched polypeptide **1-AG** in the blend is reported. Note the presence of two absorption maxima in spectra of all blend compositions.

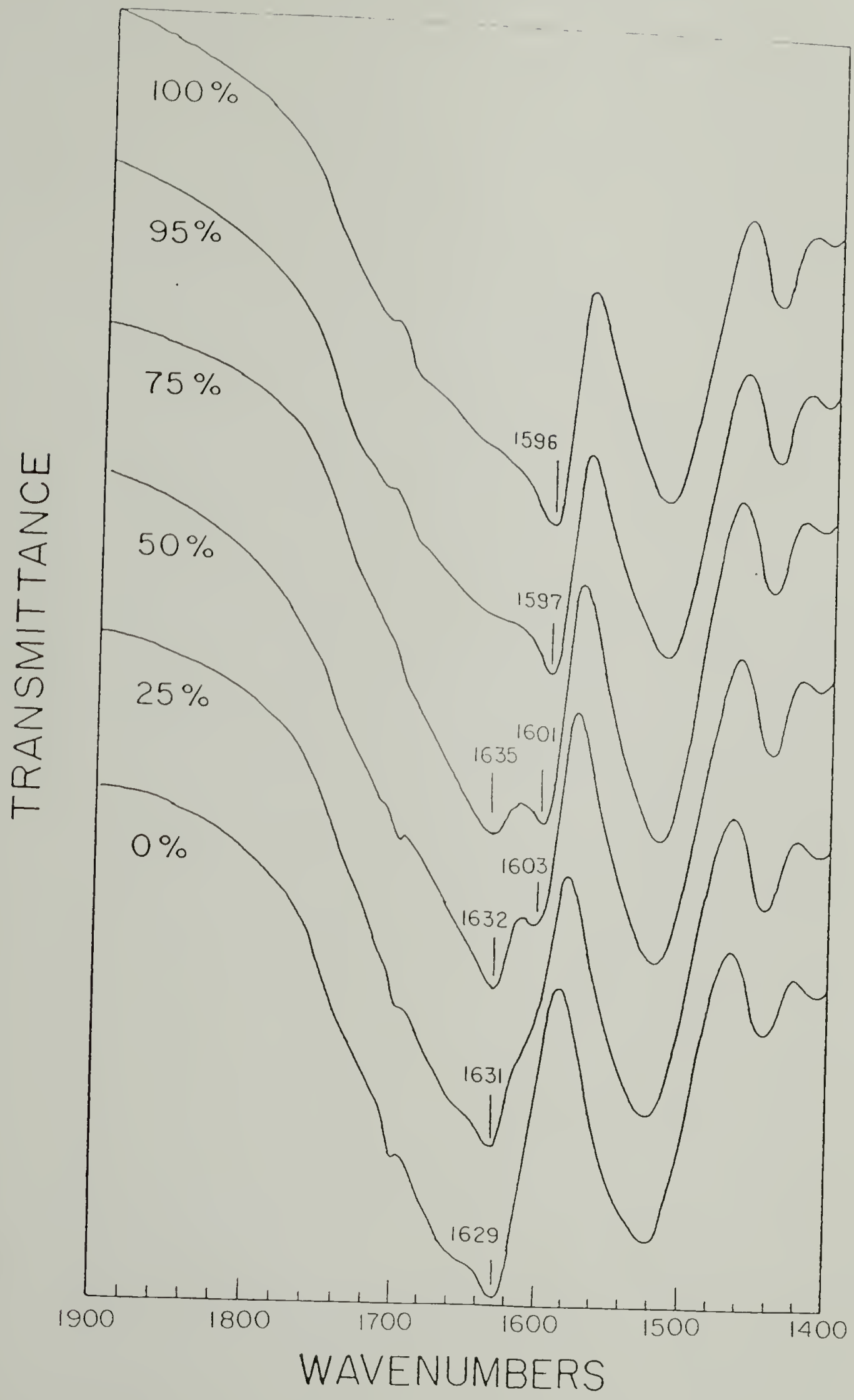

TRANSMITTANCE
100%
95%
75%
50%
25%
0%
1596
1597
1601
1635
1603
1632
1631
1629
1900
1800
1700
1600
1500
1400
WAVENUMBERS

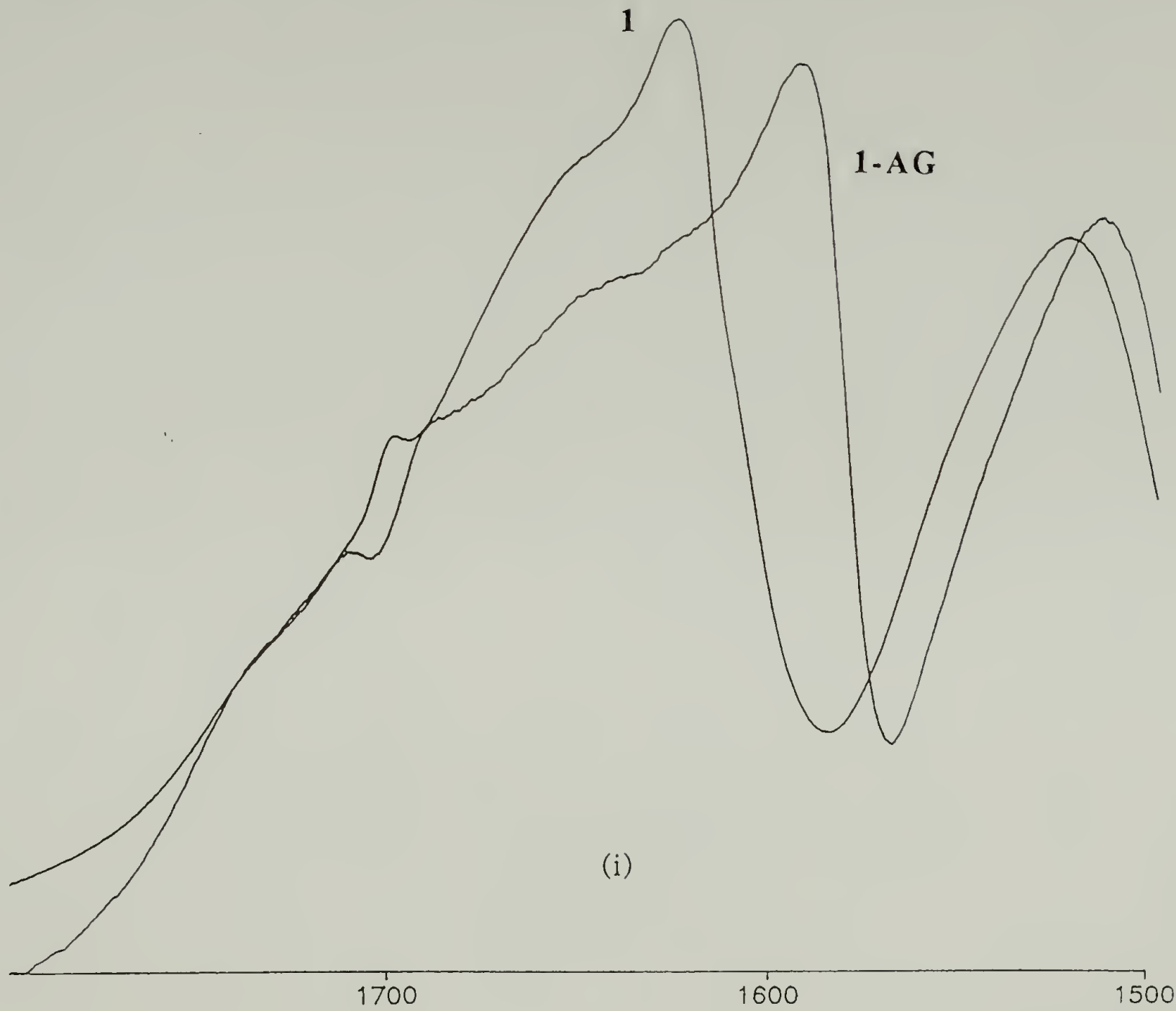
Table 4.3

Location of absorption maxima for blends of **1** with **1-AG**

Percent **1-AG**	^{12}C=O Absorption (cm^{-1})		^{13}C=O Absorption (cm^{-1})	
	Sample	Simulation	Sample	Simulation
100 wt%	-	-	1596	
95 wt%	Shoulder[a]	Shoulder[a]	1597	1596
75 wt%	1635	1629	1601	1597
50 wt%	1632	1629	1603	1598
25 wt%	1631	1629		Shoulder (~1598)
None	1629		-	

[a] The location of the shoulder cannot be determined exactly.

Figure 4.11 Computer generated absorption addition spectra of **1** and **1-AG** in the ratios of the blend compositions; the actual spectrum (below) is also shown in each case. (i) Spectra of **1** and **1-AG** used for addition.



(i)

Continued, next page

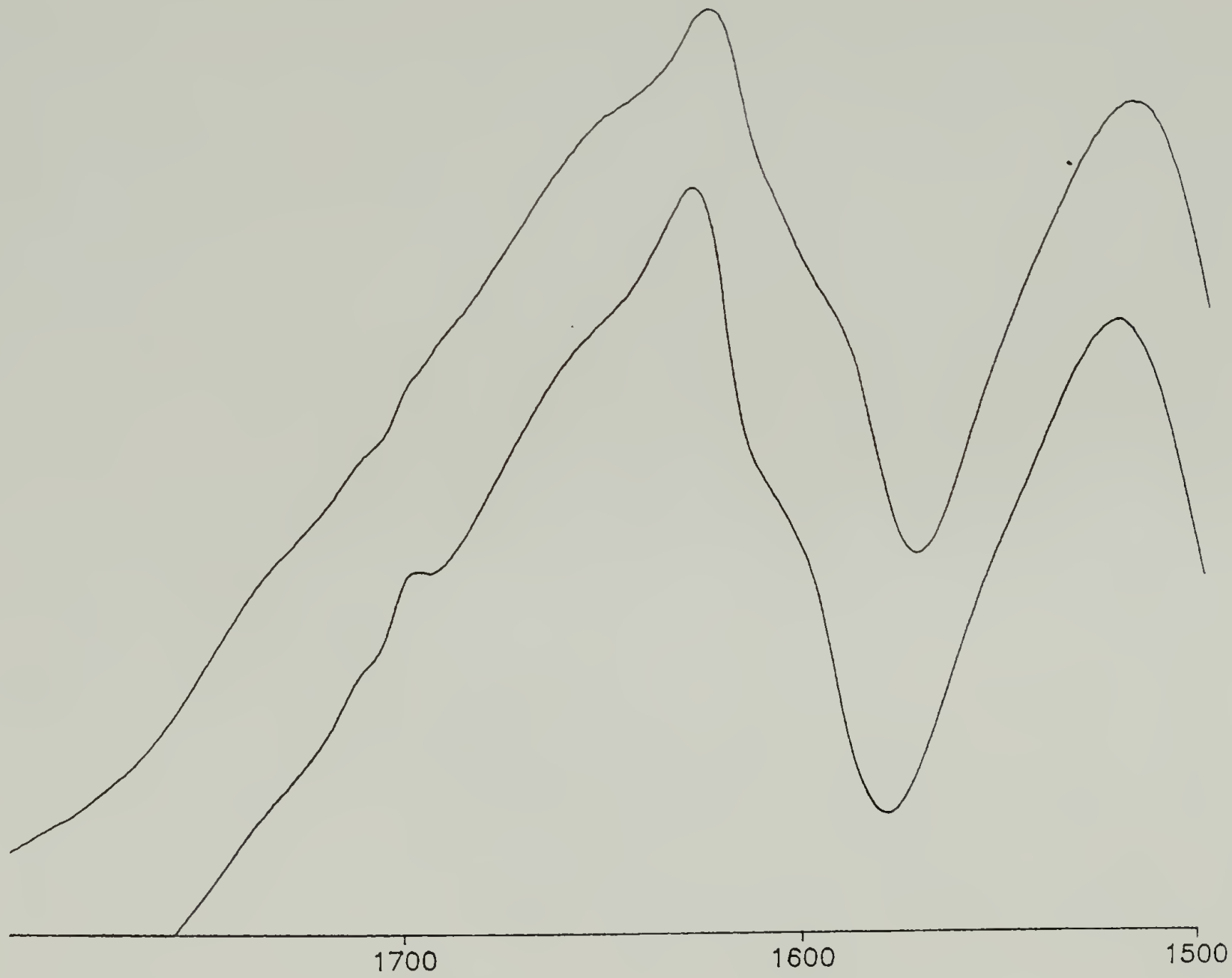


Figure 4.11 (ii) 25% 1-AG

Continued, next page

Figure 4.11 (iii) 50% 1-AG

Continued, next page

Figure 4.11 (iv) 75% 1-AG

Continued, next page

Figure 4.11 (v) 95% **1-AG**

The agreement of the blend spectrum with the computer generated spectrum is fair in all the cases; the agreement of the absorbance at approximately 1600 cm^{-1} is indeed very good. This leads us to conclude that the absorption intensities of the transition dipole coupled $^{13}C{=}O$ scale qualitatively with the composition of the blend.

4.4 Conclusion

The crystalline regions of 1 are predominantly antiparallel beta sheet.

The data from the blending experiments lead us to propose that majority of the chains are folding in adjacent reentry fashion. Any other form of chain folding- as in a fringed micellar morphology or in a 'switch board' re-entry model[11]- would reduce the magnitude of the frequency shift due to TDC and shift the amide I absorption toward the unperturbed frequency. Adjacent re-entry of chains places chain segments containing ^{12}C carbonyls adjacent to one another and chain segments containing ^{13}C carbonyls adjacent to one another. The frequency shift due to intrasheet interactions, is therefore, not diminished and the characteristic splitting is observed for amides containing either of the isotopes.

4.5 References

(1) Bandekar, J.; Krimm, S. In *Advances in Protein Chemistry*; C. B. Anfinsen, J. T. Edsall and F. M. Richards, Ed.; Academic Press: Orlando, 1986; Vol. 38; pp 183-364.

(2) Jakes, J.; Krimm, S. *Spectrochim. Acta* **1971**, *27A*, 35-63.

(3) Jakes, J.; Krimm, S. *Spectrochim. Acta* **1971**, *27A*, 19-34.

(4) Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15*, 2465-2483.

(5) Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15*, 2439-2464.

(6) Abe, Y.; Krimm, S. *Biopolymers* **1972**, *11*, 1817-1839.

(7) Cheam, T. C.; Krimm, S. *Chem. Phys. Lett.* **1984**, *107 (6)*, 613-616.

(8) Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci. USA* **1972**, *69 (10)*, 2788-2792.

(9) Miyazawa, T.; Blout, E. R. *J. Am. Chem. Soc.* **1961**, *83*, 712-719.

(10) Halverson, K. J.; Sucholeiki, I.; Ashburn, T. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1991**, *113*, 6701-6703.

(11) Keller, A. *Repts. Progr. Phys. Part 2* **1968**, *31*, 623-704.

(12) Bank, M. I.; Krimm, S. *J. Polym. Sci. A-2* **1969**, *7*, 1785-1809.

(13) Tasumi, M.; Krimm, S. *J. Polym. Sci. A-2* **1968**, *6*, 995-1010.

(14) Dreyfuss, P. *J. Polym. Sci. (Polym. Phys. Ed.)* **1973**, *11*, 201-216.

(15) Geil, P. H. *J. Polym. Sci.* **1960**, *44*, 449-458.

(16) Padden, F. J.; Keith, H. D.; Giannoni, G. *Biopolymers* **1969**, *7*, 793-804.

(17) Keith, H. D.; Giannoni, G.; Padden, F. J. *Biopolymers* **1969**, *7*, 775-792.

(18) Atkins, E. D. T.; Hill, M.; Hong, S. K.; Keller, A.; Organ, S. *Macromolecules* **1992**, *25*, 917-924.

(19) Atkins, E. D. T.; Keller, A.; Sadler, D. M. *J. Polym. Sci. A-2* **1972**, *10(5)*, 863-875.

(20) Miller, J. H. *Experiments in Molecular Genetics*; Cold Spring Harbor Laboratory: New York, 1972.

(21) Kreyszig, E. *Advances Engineering Mathematics*; John Wiley & Sons, Inc.: New York, 1975.

CHAPTER 5

CHAIN DYNAMICS AND ORIENTATION BY NMR

It is anticipated that polypeptide **1** will self-assemble into lamellae of precise thickness, with the 'stem' consisting of three alanyl-glycyl or glycyl-alanyl diads and the 'turn' consisting of a glycyl-glutamyl-glycyl triad. In the previous chapters, the presence of crystalline regions comprising antiparallel beta sheets has been ascertained. Furthermore, the infrared experiments show that the chains fold predominantly with adjacent reentry. However, the design requires, in addition to the presence of antiparallel beta sheets and adjacent reentry, that the glutamic acid residues reside at the turn position.

In the first part of this chapter, the solid state structures of various preparations of **1** have been analyzed by magic angle spinning ^{13}C NMR spectroscopy. The dynamics of the main chains and side-chains have been monitored by determining the rates of relaxation of ^{13}C nuclei in the rotating frame. The mobilities of the side-chains suggest that the glutamic acids are excluded from the crystalline regions during crystallization. This explanation is in agreement with the formation of nanoscale lamellae with the glutamic acids decorating the surfaces.

In Chapter 3 we showed **1** to crystallize in the form of needle-like entities. The orientation of these asymmetric needles was not entirely determined by the x-ray diffraction analyses. Static ^{13}C lineshape

experiments were carried out on oriented samples in order to determine the orientation; the second part of this chapter discusses the static NMR experiments.

The reader should refer to an NMR text for the theory and practice of solid state NMR[1,2].

5.1 Solid State Magic Angle Spinning NMR of **1**

5.1.1 Introduction

5.1.1.1 Scope of CPMAS NMR Experiments

The analysis by x-ray diffraction and FTIR suggests the formation of unique scaffolds generated by adjacent reentry of the chains during the crystallization of **1**. However, there are some important questions that are left unanswered in spite of the analyses. We are not certain about the glutamic acid residues being preferentially excluded from the interior of the crystallites and being placed at their surfaces. We are also not certain about the glycine and alanine residues being preferentially included in the interior of the crystallites. In this chapter we present solid state NMR relaxation data on **1** which suggests formation of the predicted scaffolds.

5.1.1.2 NMR of Silks and Their Analogues

The conformations of *Bombyx mori* silk and its analogues have been analyzed by NMR in solution[3-5] and in solids[6-9]. It has been believed that the ^{13}C chemical shifts of polypeptides are influenced primarily not by the primary sequence but by the conformation of the chain[10]. Saito and coworkers[8] have discussed the conformational dependence of ^{13}C chemical shifts of oligo(L-alanine) and its copolymers with D-alanine. Solid state NMR investigation leads them to conclude that the ^{13}C chemical shifts of the alpha, beta and carbonyl carbons are a function of the conformation of the oligomer of the alanine: beta sheet, alpha helix or random coil. Their assignments as a function of the conformation are shown in Table 5.1.

Subsequent work on solid state NMR spectroscopy of *Bombyx mori* silk and poly (alanylglycine)[9], and the comparison with results from FTIR and x-ray diffraction confirmed the dependence of ^{13}C chemical shifts on the conformation of the polypeptide. Table 5.2 summarizes the findings by these workers on conformation dependent chemical shifts; in their nomenclature, silk II is the antiparallel beta sheet form.

A more recent survey of the conformation dependence of chemical shifts is presented by Ishida and coworkers[7]. These workers have used the conformation dependence of chemical shifts in silk fibroin from *Bombyx mori* and *Philosomia cynthia ricine* to study the mechanical treatment induced transitions in these silks.

Table 5.1

^{13}C Chemical shifts in ppm as a function of conformation for oligo (L-alanine) (Adapted from work by Saito[8])

	Alpha Helix	Beta Sheet	Random Coil
C α	52.4	48.2	51.1
C β	14.9	19.9	15.7
C=O	176.4	171.8	176.1

Table 5.2

^{13}C Chemical shifts in ppm as a function of conformation for *Bombyx mori* silk and poly (AlaGly) (Adapted from work by Saito; [a] $(AlaGly)_nI$ is defined to be silk II like and $(AlaGly)_nII$ is defined to be silk I like[9])

	Silk II	$(AlaGly)_n$ I[a]	Silk I	$(AlaGly)_n$ II[a]
Alanine				
C α	48.9	48.5	50.1	50.5
C β	20.0	20.0	15.7	16.6
C=O	172.5	171.9	171.5	177.1
Glycine				
C α	43.5	43.3	43.0	43.7
C=O	169.8	169.2		171.9
Serine				
C α	53.1		54.4	
C β	64.1		59.2	

5.1.1.3 Relaxation Studies on Silks

Relaxation studies on silks and their analogues have been carried out in order to determine the solid state dynamics of these semi-crystalline systems[11,12]. However, all these studies on the dynamics involve the measurement of only the spin-lattice (T_1) and spin-spin relaxation (T_2) times in the laboratory frame. These experiments probe for motions in solids which cause magnetic fluctuations in the megahertz frequency range. Most of the motions in polymeric systems are in the mid-kilohertz range; hence, these experiments are not sensitive to most polymeric motions. This is indeed observed to be the case in the work by Asakura and coworkers[12]. These workers have carried out relaxation studies in the laboratory frame; they conclude that the only motions observed are the rapid rotations of the methyl groups; since most of the motions have low spectral density in this range, the chains appear immobile in these experiments.

5.1.1.4 Relaxation Studies of Polymeric Systems

5.1.1.4.1 Relaxations in the Laboratory Frame of Reference

A precessing ^{13}C nucleus can transfer its magnetization to the lattice if there is a magnetic fluctuation in the lattice which matches the ^{13}C precessing magnetization. The frequency matching occurs if the motions in the lattice are of the same frequency as those of the precessing nucleus. The determination of the spectral density of the lattice, the determination of the exact quantum mechanical matching conditions required for the transfer of

magnetization and the mechanism of transfer of magnetization are beyond the scope of this work. The precessing frequency (ω) of a ^{13}C nucleus is given by the product of the gyromagnetic ratio and the magnetic field ($\omega = \gamma B$); the larger the magnetic field, the higher the precession frequency. Since the spin lattice relaxations in the laboratory frame involve precession in the large external Zeeman magnetic field, the frequency in question is high and the magnetic fluctuations required in the lattice which match the fluctation caused due to the precession of the ^{13}C magnetization are of a high frequency.

5.1.1.4.2 Relaxations in the Rotating Frame of Reference

On the contrary, Schaefer[13,14] has shown that relaxations in the rotating frame are due to fluctuations in the lattice which are of a lower frequency. In these experiments the ^{13}C magnetization is spin locked by a magnetic field which is three orders of magnitude lower than the strong Zeeman magnetic field; any spin-lattice relaxations that can occur are due to magnetic fluctuations in the lattice which are on the order of the precession frequency, ω, where $\omega = \gamma B$. However, the magnetic field in question is the spin lock field which is three orders of magnitude lower than the external magnetic field; hence, the motions which cause these fluctuations are also three orders of magnitude lower in frequency than the ones required for spin-lattice relaxations in the laboratory frame. It has been shown that the spin-lattice relaxation in the rotating frame ($T_{1,\rho}$) probes for motions in solids in the low to mid-kilohertz range[13,14]. Since glassy polymers have been known to show high spectral density in this

frequency range, $T1,\rho$ relaxation studies have already been used to study polymer dynamics and polymer miscibility[13,15-18].

5.1.1.4.3 $T1,\rho$ Relaxation Studies in Polymers

Schaefer and coworkers[13] have carried out room temperature $T1,\rho$ relaxation studies on glassy polymers. Their studies involve poly(methyl methacrylate), poly(phenylene oxide), polycarbonate (of bisphenol-A), poly(vinyl chloride), polystyrene, polysulfone and poly(ether sulfone). They observe that in most cases the $T1,\rho$ relaxations are composed of more than one component. These nonlinearities have been attributed to multiplicity of relaxation times; however, the workers rule out static heterogeneity in the sample as a reason for the nonlinearity. They claim that the nonlinearity is due to the molecular dynamics and in particular, to the inherent dynamic heterogeneity of a polymer glass. A complete discussion of this reasoning is presented by Schaefer[13]. In spite of the nonlinearity in the $T1,\rho$ relaxation behavior, the experiment does yield useful results. In the $T1,\rho$ relaxation studies with poly(methyl methacrylate), the quaternary carbon relaxes faster than the rest in spite of lack of adjoining protons. This is attributed to a crankshaft motion involving four carbons (or three bonds) along the main chain. The analysis of poly(vinyl chloride) and polystyrene shows that the main chain $T1,\rho$s are comparable to those observed for poly(aryl ethers); however, the T1s and T_{CH}s are shorter for for poly(vinyl chloride) and polystyrene. This suggests that the spectral density of poly(vinyl chloride) and polystyrene is considerably greater at high frequency than that of polyaryl ethers. It is

speculated that this is the case because the operational motion unit in poly(vinyl chloride) and polystyrene is shorter and involves fewer carbon units than that for the poly(aryl ethers).

5.1.1.4.4 Use of $T_{1,\rho}$ Relaxations for Prediction of Mechanical Properties

The low frequency motions studied by $T_{1,\rho}$ relaxation have been important in the understanding of mechanical properties of polymeric materials. Conventionally, these motions have been studied by mechanical spectroscopy. The impact resistance of a polymer is dependent on its ability to dissipate energy; this is manifested in the display of mechanical loss peaks indicative of main chain motions. Schaefer and coworkers[13] use the NMR relaxation data to successfully predict the trends in mechanical properties of the glassy polymers; this method has obvious advantages over mechanical methods which cannot assign the losses to the specific units within the chain.

5.1.1.4.5 Use of $T_{1,\rho}$ Relaxations to Study Polymer Miscibility

Polymer blend miscibility has been extensively studied by using the $T_{1,\rho}$ relaxation phenomenon. Stejskal and coworkers[18] have studied miscibility in poly(phenylene oxide)/polystyrene blends by monitoring the proton $T_{1,\rho}$s. This analysis permits determination of heterogeneities on the order of 30 Å; if a blend is homogeneous on the length scale of 30 Å, only one proton $T_{1,\rho}$ is observed; if there is phase separation on the length scale of 30 Å, the proton $T_{1,\rho}$ is bicomponent. When the system is

phase separated, the two phases display different average mobilities which are manifested in two values of $T_{1,\rho}$. Dickinson and coworkers have used the same phenomenon to determine the phase behavior of poly(4-methylstyrene)/poly(2,6-dimethylphenylene oxide) blends[16] and poly(vinyl methyl ether)/polystyrene blends[15].

5.1.1.4.6 $T_{1,\rho}$ Relaxation Studies on Polypeptides

Polypeptides in their solid state possess a higher density of hydrogen bonds and are therefore different from the previously studied glassy polymers. Since we are unaware of any $T_{1,\rho}$ data on such highly hydrogen bonded solids, it is also our intention to present $T_{1,\rho}$ data for silk-like polypeptides. Our primary interest, however, is to determine the morphology and the folding habit of **1** in semi-crystalline samples by observing the $T_{1,\rho}$ relaxation of the alanine, glycine and glutamic acid residues.

5.1.2 Methods

5.1.2.1 Synthesis, Purification and Crystallization of **1**

The details of the polymer synthesis, purification and crystallization have been outlined in previous chapters.

5.1.2.2 CPMAS NMR

Solid state NMR spectra were obtained on a Bruker ASX 300 spectrometer with a ^{13}C frequency of 75.426 MHz. The samples were typically spun at the magic angle at 5000 Hz in a solid state probe in a either a 4 mm or a 7 mm zirconia rotor fabricated by Bruker. The pulse sequences employed for the $T_{1,\rho}$ measurements are shown in Figure 1[13]. The spin lock field, $B_{1,C}$ was equivalent to a ^{13}C precession frequency of 50 KHz; the 90° pulse was 5 μs. All spectra except the Bloch decay were obtained with cross polarization; broad band proton decoupling was employed in all instances. Typically, recycle times of 3 s were employed for the cross polarized spectra; the recycle time was 15 s for the Bloch decay experiments. Spin-lattice relaxation times (T_1) were measured by the method of Torchia[19]. All spectra were calibrated using (para)ditertiary butyl benzene as a standard and the shift values were assigned with respect to the TMS carbon (at 0 ppm). All deconvolutions were performed by employing Bruker software.

Experiments were carried out on amorphous samples of **1** and on crystallized samples of **1** in order to compare the structures in the two phases.

Carbon $T_{1,rho}$ (Relaxation in the Rotating Frame)

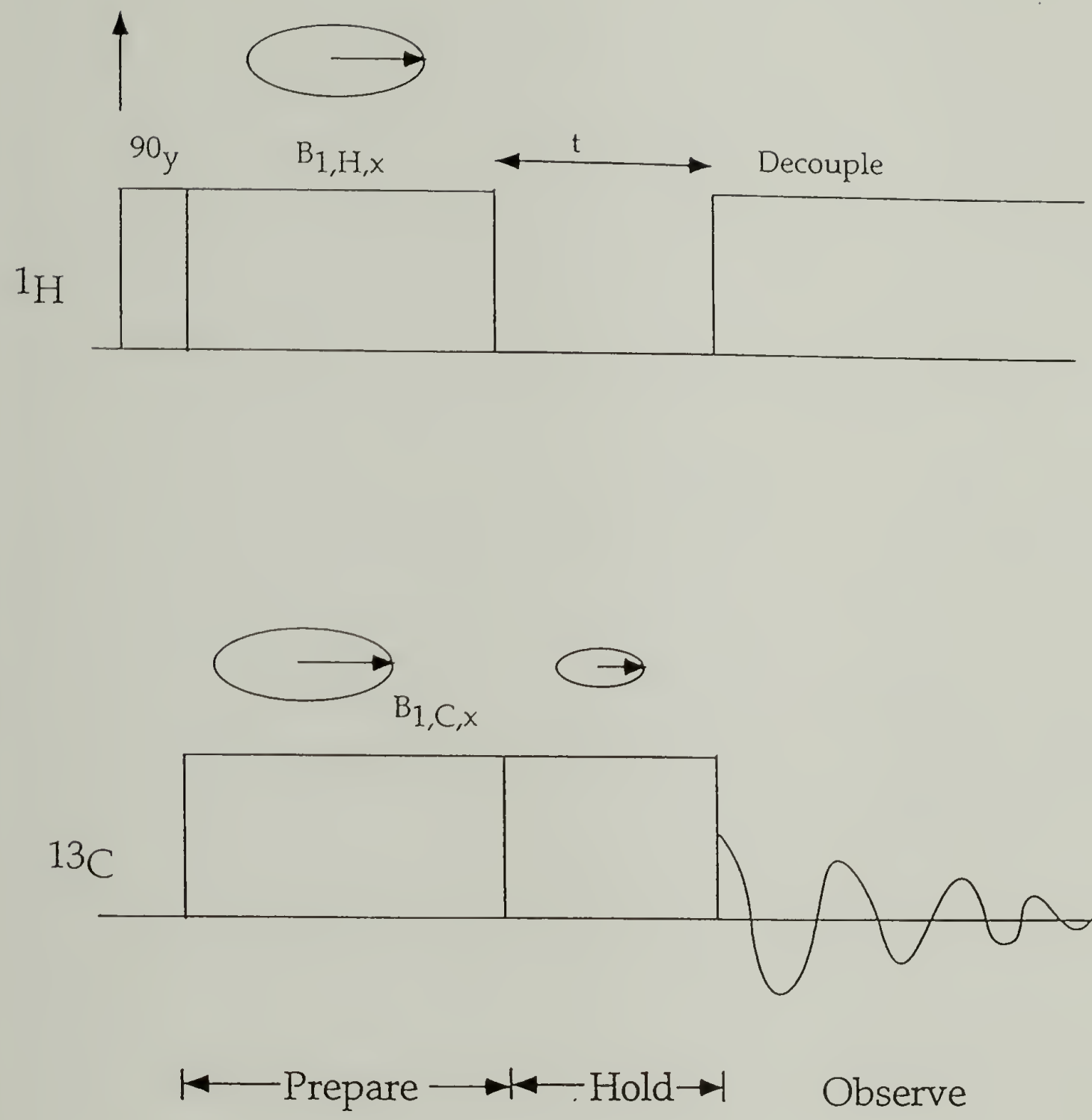
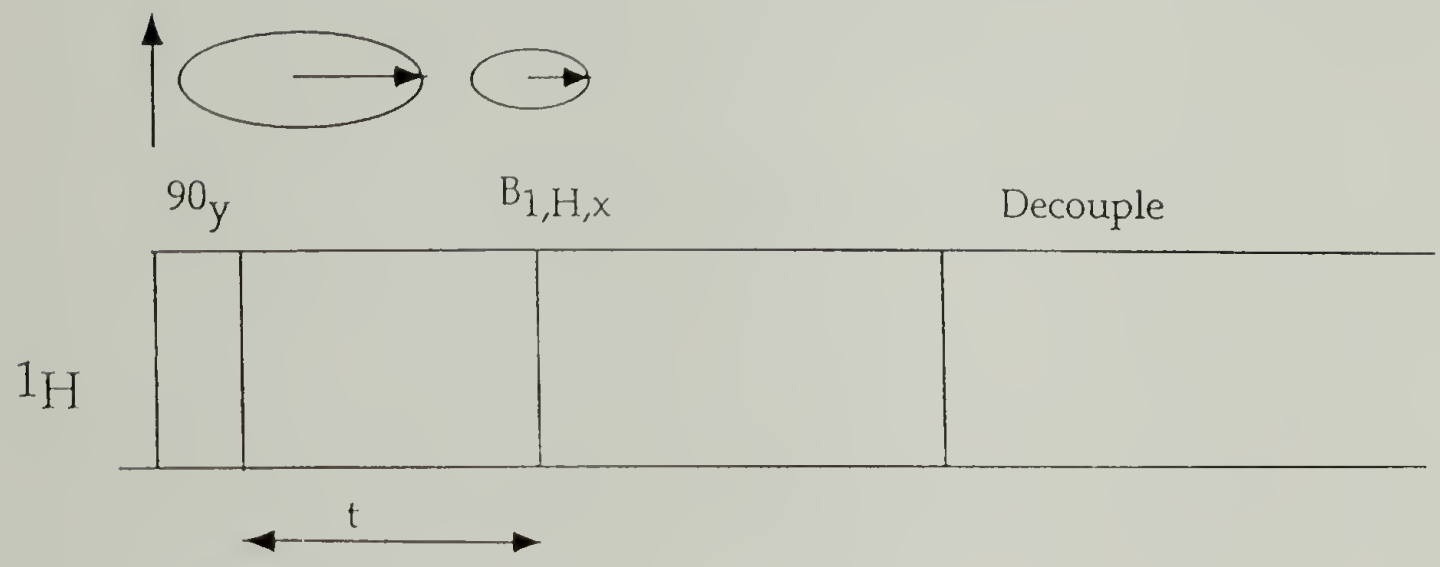


Figure 5.1(i) Pulse sequence used for the measurement of carbon-$T_{1,\rho}$

Figure 5.1 Pulse sequences

Continued, next page

Proton $T_{1,rho}$ (Relaxation in the Rotating Frame)



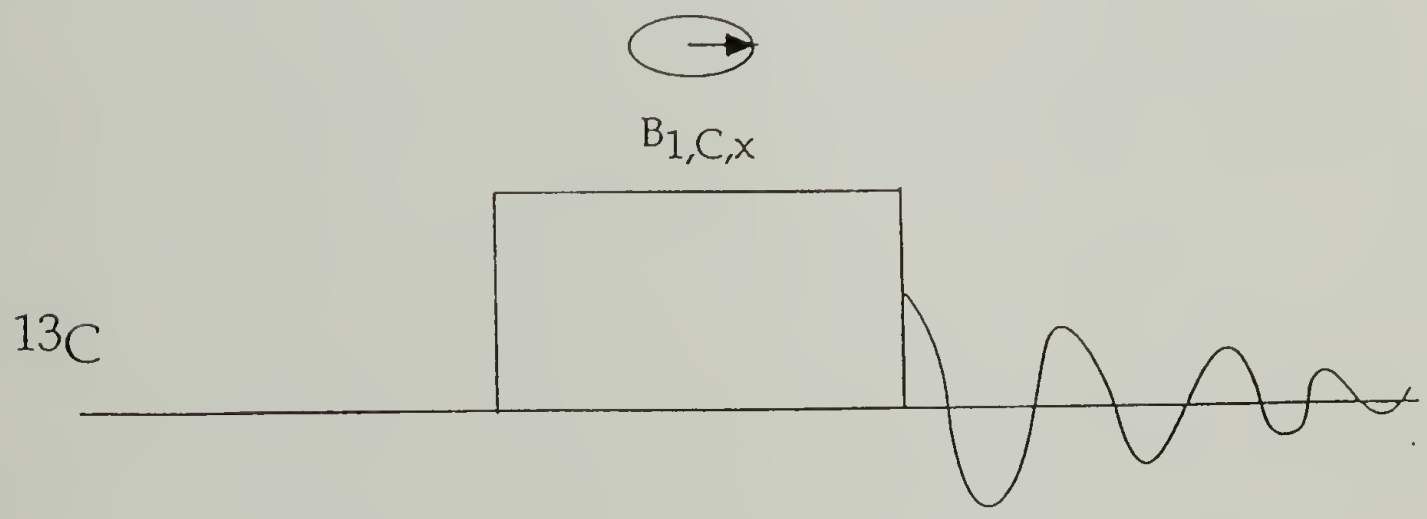


Figure 5.1(ii) Pulse sequence used for the measurement of proton-$T_{1,\rho}$

5.1.3 Results and Discussion

5.1.3.1 Summary of Previous Results on 1

The crystallization of **1** from formic acid has been shown to from lath-like structures with large aspect ratios[20]. The long direction has been observed to be approximately 0.25 μm and the thickness and width have been estimated to be between 50 Å and 200 Å (Chapter 3). The x-ray diffraction pattern indexes on a monoclinic unit cell consistent with the formation of antiparallel beta sheet structure[20]. Formation of antiparallel beta sheets causes the carbonyl residues to be placed at positions within the unit cell such that they lead to a characteristic transition dipole coupling in the FTIR spectra. The absorptions in the amide I region at 1629 cm^{-1} and 1702 cm^{-1} corroborate the formation of antiparallel beta sheets in crystalline **1**.

5.1.3.2 CPMAS of 1

Figures 5.2(i) and 5.2(iv) show the CPMAS spectra for amorphous and crystalline **1** respectively. Figures 5.2(ii) and 5.2(iii) show the spectra without cross polarization (Bloch decay) for amorphous and crystalline **1** respectively. In the Bloch decay experiments, the ^{13}C nuclei are polarized by a direct pulse instead of cross polarization from the ^{1}H nuclei. Since the efficiency of cross polarization is a function of the immobility of the nuclei between which polarization is occurring, the intensity ratios are different in the two experiments. In Bloch decay experiments, ratios of intensities of

carbons at different chemical shifts are directly proportional to the ratios of the fractions of those carbons in the sample; however, in CPMAS experiments they are not proportional because the intensity is a function of the efficiency of magnetiztion transfer in addition to the fractions of those carbons in the sample. A spectrum of in-house synthesized poly(glutamic acid)[21] was used to assign the 57 ppm and 29 ppm resonances in Figure 5.2; the rest of the resonances were assigned using the literature cited in the Section 5.1.1. Figure 5.3 shows the CPMAS spectrum of amorphous poly(glutamic acid); the resonance at 56.5 ppm is due to the alpha carbon, the one at 29 ppm is due to the beta and gamma carbons, and the 175.7 ppm resonance is due to the carbonyl carbon. These values of carbon resonances are in fair agreement with those reported in the literature[10] for glutamic acid residues in proteins. The assignments for the resonances of **1** are listed in Table 5.3.

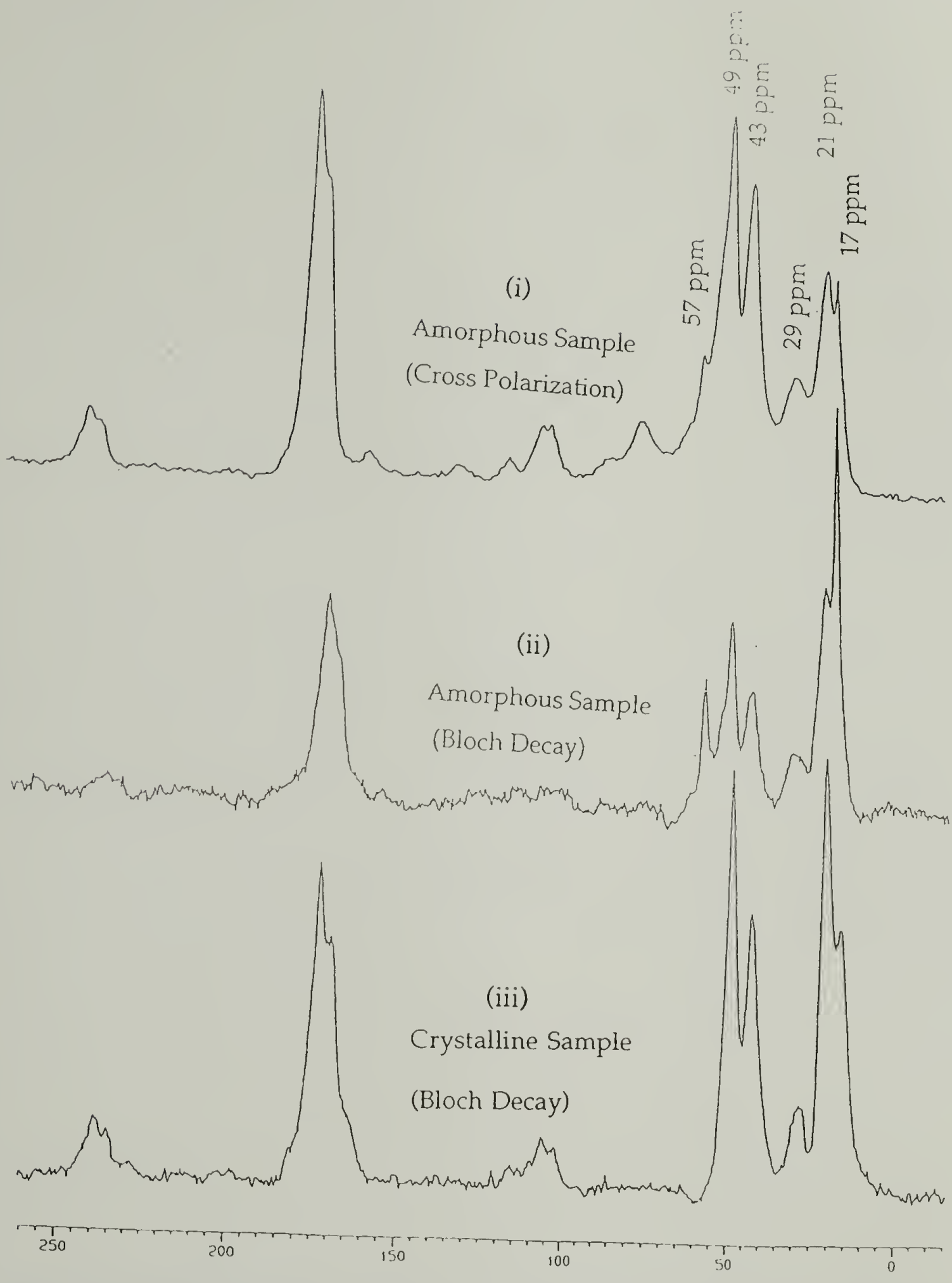


Figure 5.2 Solid state magic angle spinning (MAS) spectra for amorphous ((i) and (ii)) and crystalline ((iii) and (iv)) samples of **1**. (i) and (iv) are acquired with cross polarization from the ^{1}H nuclei and (ii) and (iii) are acquired after direct polarization of the ^{13}C nuclei.

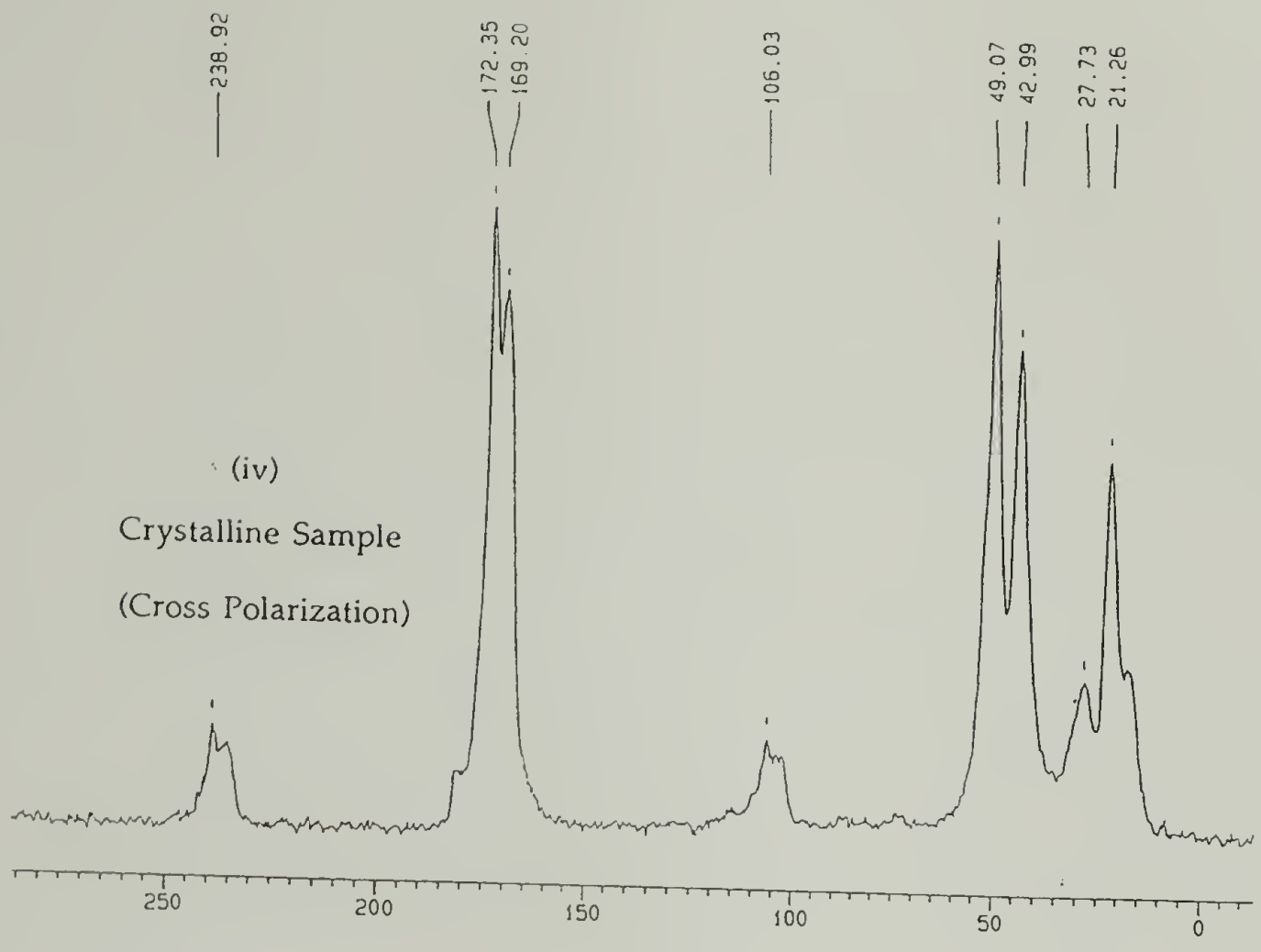


Figure 5.2 (Contd.)

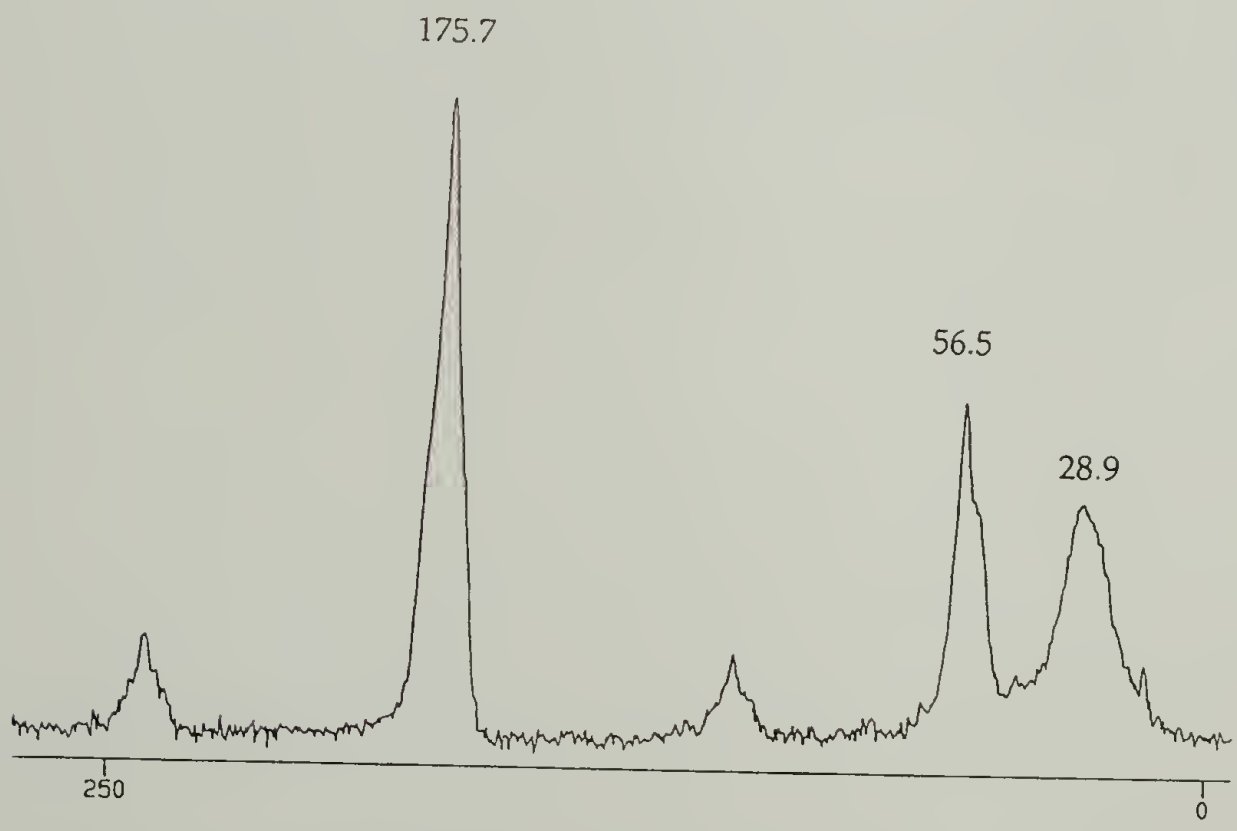


Figure 5.3 CPMAS spectrum of poly(glutamic acid).

5.1.3.2.1 Determination of Crystalline Fraction

Figure 5.4(i) shows the deconvolution of the amorphous Bloch decay spectrum and Figure 5.4(ii) shows the deconvolution of the crystalline Bloch decay spectrum. As has been stated previously, since the Bloch decay experiments polarize the ^{13}C nuclei directly (instead of cross polarization by adjoing protons), the intensities observed in the Bloch decay experiments are a direct function of the fraction of the corresponding carbons in the sample, provided the carbon magnetization is allowed to relax completely after every acquisition. The spin lattice relaxation times for all the carbons other than the beta carbons of alanine were on the order of 50 s. Under the condition of the experiments, none of these except the beta carbons of the alanine relaxes entirely after every acquisition; hence, reliable quantitative estimates can be obtained only in the case of the beta carbons of alanine. Table 5.4 lists the relative fractions of the deconvoluted components for the amorphous Bloch decay spectrum and Table 5.5 lists the relative fractions of the deconvoluted components for the crystalline Bloch decay spectrum.

Table 5.3

Assignment of resonances in the solid state CPMAS spectrum of **1**

Peak Position	Assignment	Peak Position	Assignment
172 ppm	Alanine C=O	43 ppm	Glycine C α
170 ppm	Glycine C=O	29 ppm	Glutamic Acid C γ, C β
57 ppm	Glutamic Acid C α (disordered)	21 ppm	Alanine C β (β-sheet)
49 ppm	Alanine C α	17 ppm	Alanine C β (disordered)

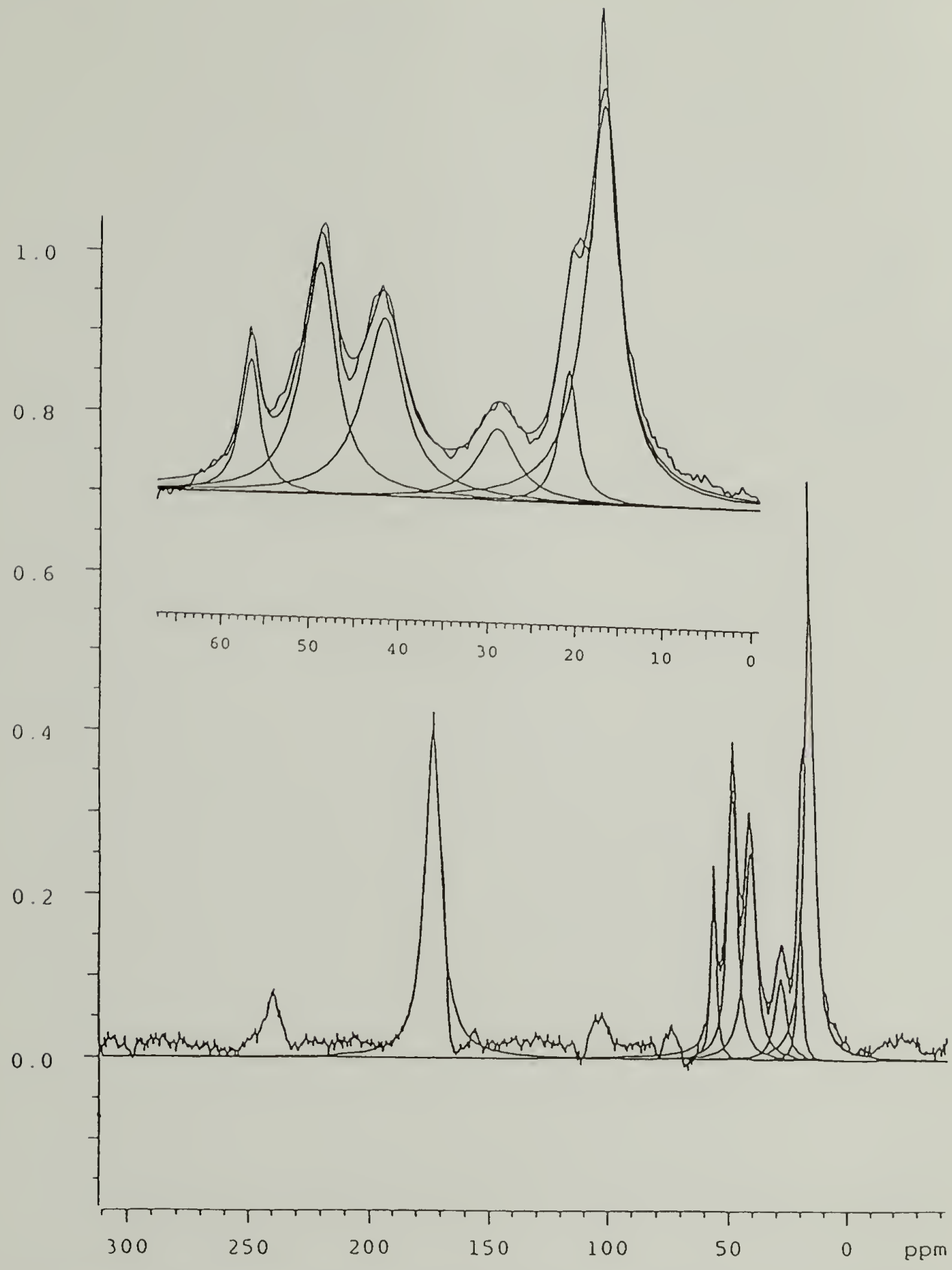
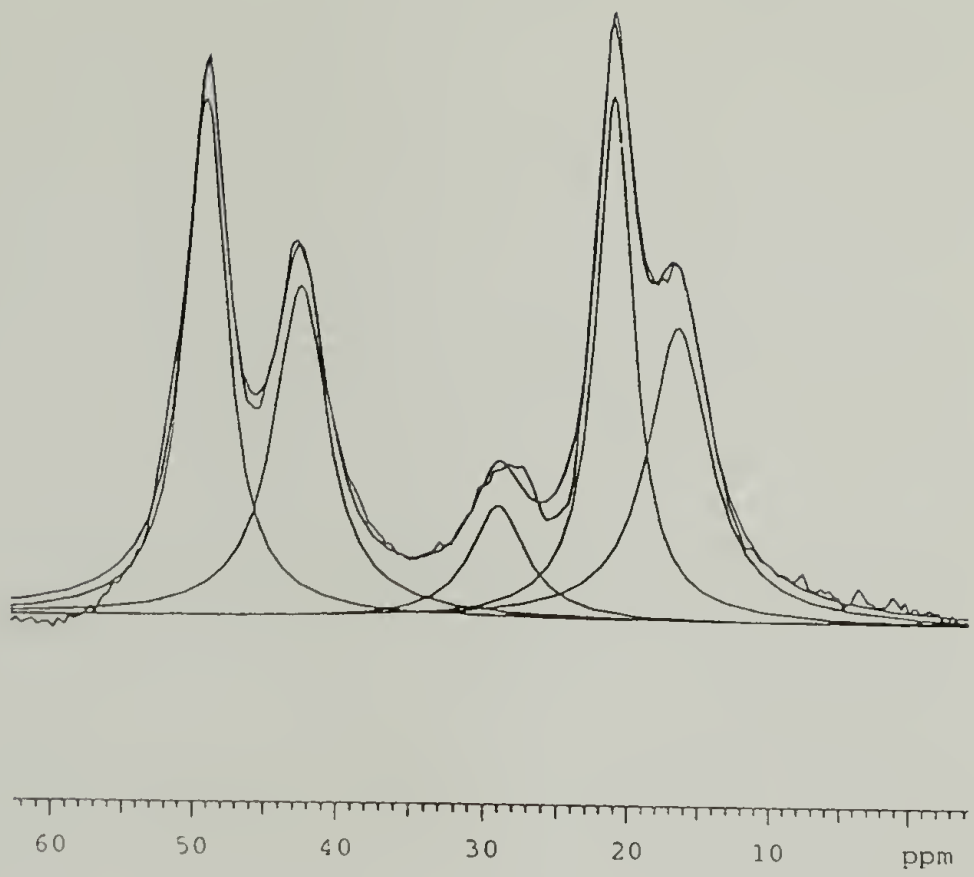


Figure 5.4 (i) Deconvolution of Bloch decay MAS spectra of amorphous 1.

Figure 5.4 Deconvolutions Continued, next page



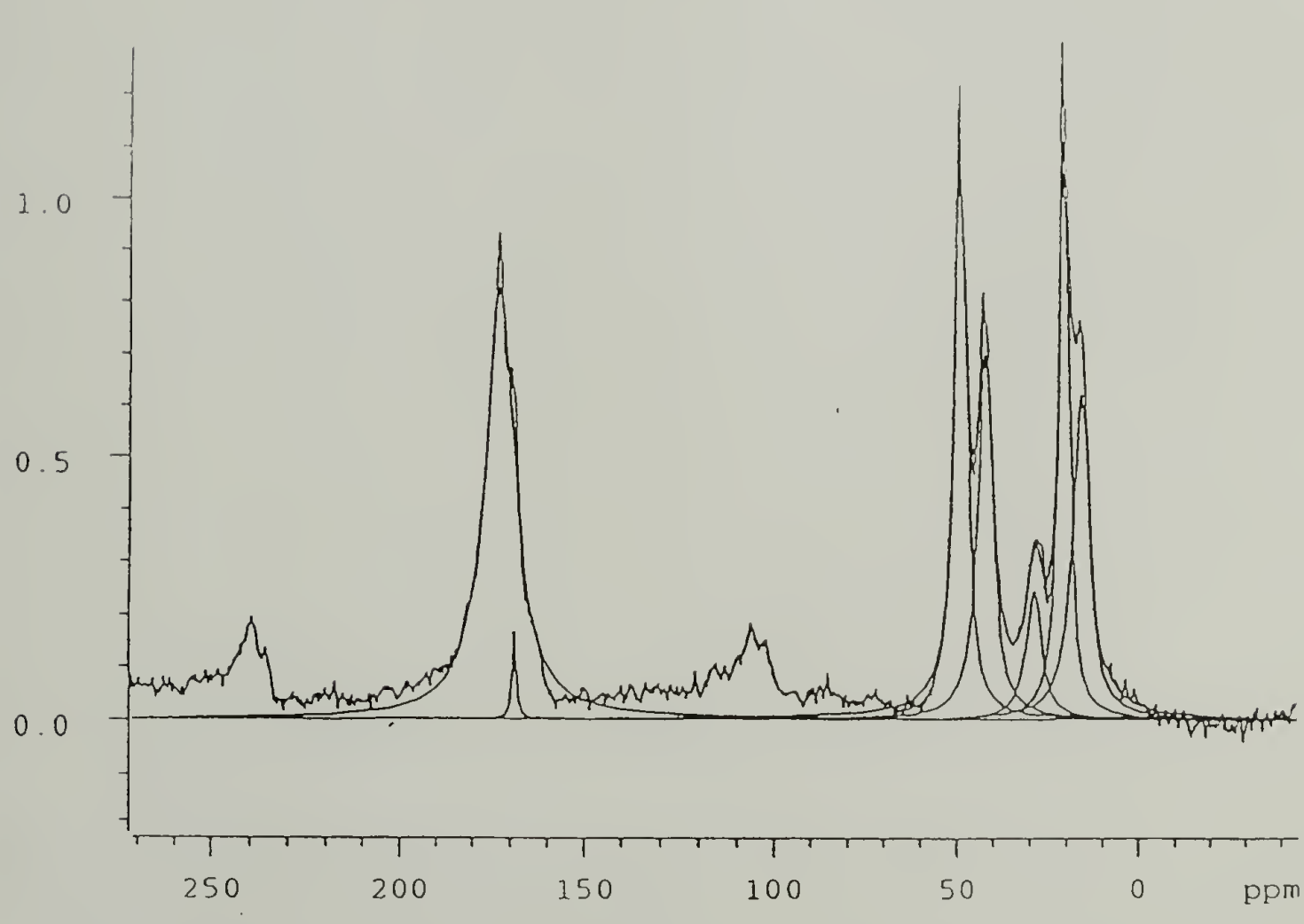


Figure 5.4 (ii) Deconvolution of Bloch decay MAS spectra of crystalline 1.

5.1.3.2.2 Differences between Crystalline and Amorphous Samples of **1**

The amorphous sample shows the glutamic acid alpha carbon to resonate at 57 ppm; however, in the crystalline sample, the resonance appears to have shifted to approximately 49 ppm (*vide infra*). A comparison of the relative areas of the alpha carbons between Table 5.4 and Table 5.5 does indeed agree qualitatively with this hypothesis; however, since the alpha carbons have not relaxed fully in these experiments, the numbers cannot be very rigourous. From Tables 5.4 and 5.5 it can be seen that 51.5% of alanine beta carbons are in beta sheet conformation in the crystalline sample and 13.9% are in the beta sheet conformation in the amorphous sample. If these polypeptides, during crystallization, were to fold as predicted, a fraction of the alanine residues would be constrained to reside in disordered regions (or turns) of the sample; hence, it is plausible that the sample would never display an entirely crystalline spectrum. Furthermore, there are bound to be some regions of the sample which are non crystalline due to the kinetics of crystallization; these regions would also contribute to the intensity at 17 ppm in the crystalline sample. The amorphous sample displays some intensity at 21 ppm which is indicative of crystal-like regions; however, these samples do not display any reflections in x-ray diffraction experiments. The x-diffraction experiments and NMR experiments probe order on different length scales; short range beta-strand-like structure might lead to some intensity in the solid state NMR spectrum at 21 ppm, while the x-ray diffraction experiment will not show any reflections until a larger aggregate of beta strands is formed to yield periodic planes of electron density.

Table 5.4

Deconvolution areas of amorphous Bloch decay spectrum of **1**

	Peak Maximum (ppm)	Normalized Area (%)
Peak #1	173.19	30.38
Peak #2	56.97	4.83
Peak #3	49.32	14.7
Peak #4	42.10	13.93
Peak #5	29.24	5.38
Peak #6	21.07	4.28
Peak #7	17.23	26.49

Table 5.5

Deconvolution areas of crystalline Bloch decay spectrum of **1**

	Peak Maximum (ppm)	Normalized Area (%)
Peak #1	172.64	32.88
Peak #2	168.75	0.87
Peak #3	49.14	17.51
Peak #4	42.47	14.18
Peak #5	29.04	5.23
Peak #6	20.92	15.11
Peak #7	16.8	14.21

5.1.3.3 $T_{1,\rho}$ Relaxation Studies of 1

5.1.3.3.1 Spectral Density and the Efficiency of Relaxation

As has already been discussed in the Section 5.1.1, polymeric solids display a large spectral density in the low to mid-kilohertz range[13] as opposed to small molecules which have more spectral density in the megahertz range. Hence, T_1 relaxation pathways that exist due to fluctuations on the order of Larmor frequencies are not very efficient in most polymeric systems. However, some exceptions do exist for nuclei positioned on short side chains and at chain ends. The spin lock field $B_{1,C}$ is three orders of magnitude smaller than the strong external magnetic field causing the Zeeman splitting; this causes relatively slower precession around $B_{1,C}$. Hence, the spin-lattice relaxation in the rotating frame occurs due to fluctuations on the order of kilohertz. Polymeric motions which lead to these relaxation pathways can therefore be studied by monitoring these relaxation rates.

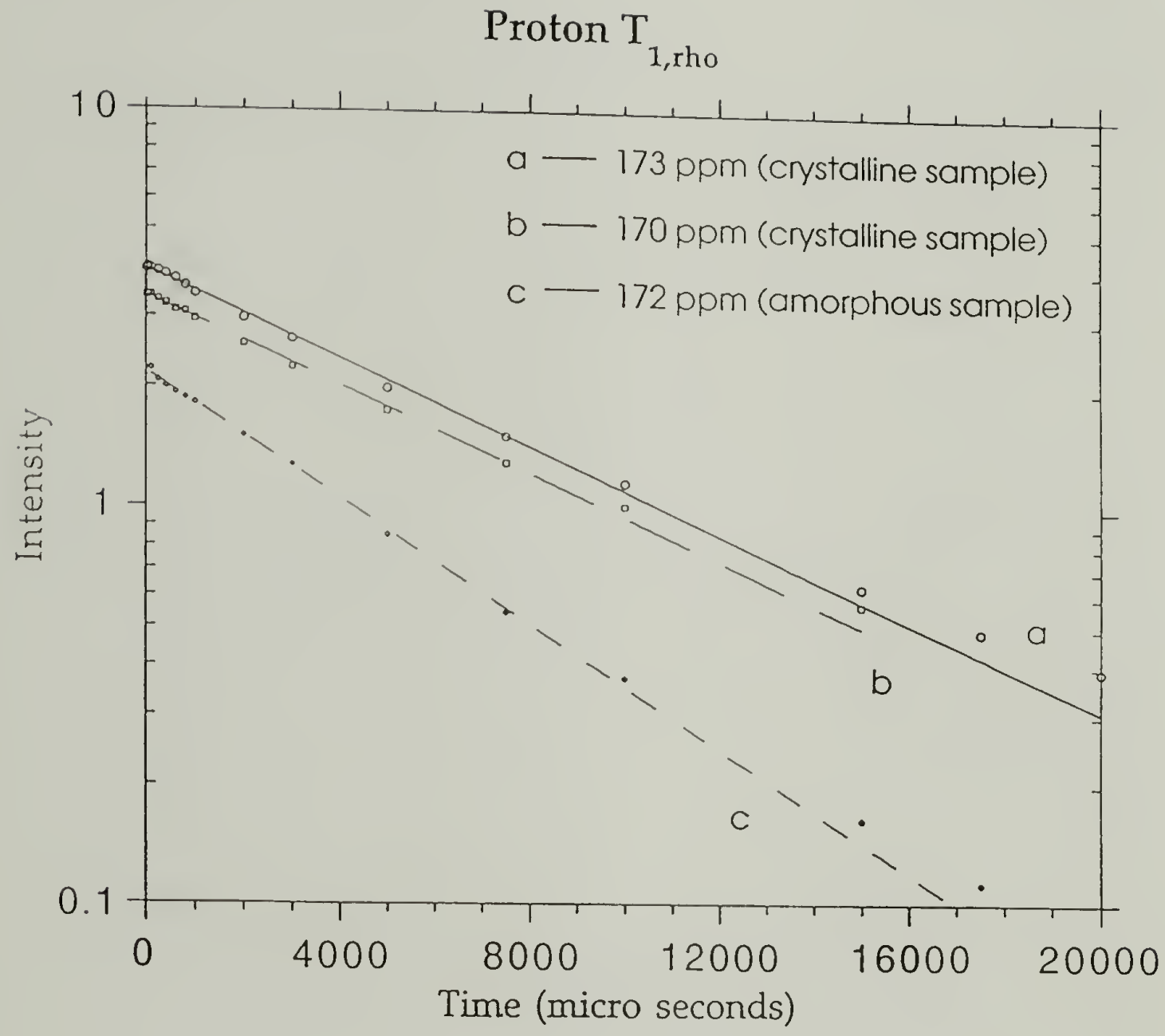
5.1.3.3.2 Relaxation Pathways in the Rotating Frame of Reference

Typically, rotating frame relaxations are caused due to transfer of magnetization to adjacent ^1H spins- this process of transfer being independent of motions-, or to the lattice- this process being enhanced due to kilohertz motions. Hence, the contribution of the spin lattice relaxation processes to the total $T_{1,\rho}$, determines the sensitivity of the experiment to chain motions; i.e. the experiment is the most effective when all the

magnetization transfer is to the lattice and none to the adjacent protons; conversely, the experiment is not at all sensitive to motions if all the magnetization transfer occurs to the adjacent protons. In order to determine the efficacy of the experiment, adiabatic relaxation times in the rotating frame (T_{ADRF}) have been measured for glassy polymeric solids; from these measurements and the subsequent analysis, it has been concluded that the spin-spin contributions (transfer to adjacent protons) to $T_{1,\rho}$, although present, do not hamper their utility in determining and analyzing motions in these solids[14].

5.1.3.3.3 Proton $T_{1,\rho}$ Relaxation of 1

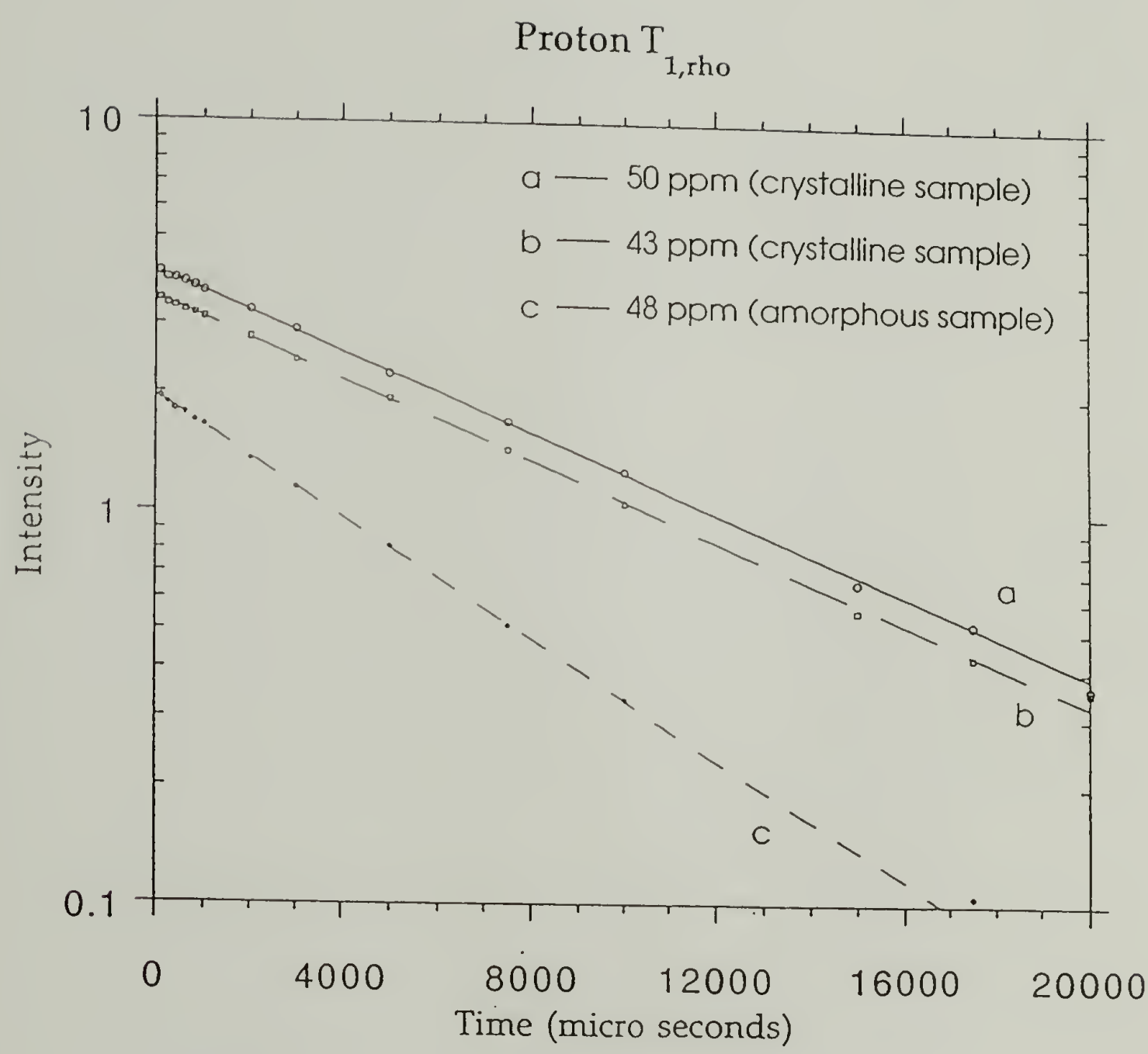
The proton $T_{1,\rho}$ relaxation behavior of 1 is linear in all instances and can therefore be precisely defined by one time constant only; hence, instead of displaying all carbon resonances, we have arbitrarily chosen three carbonyl carbons, three alpha carbons and three side chain carbons. Figure 5.5 shows these ^1H $T_{1,\rho}$ relaxation trends and Table 5.6 lists all the ^1H $T_{1,\rho}$ relaxation results on 1. The proton relaxation times are on the order of 3 ms for the amorphous sample and 8 ms for the crystalline sample. Diffusion of ^1H spins occurs due to the close proximity of ^1H nuclei in solids; therefore, all the carbons relax at similar rates in a given sample. The average mobility of the protons appears lower in the crystalline sample than in the amorphous sample. The difference in average mobilities between them is manifested in the shorter relaxation times for the amorphous sample.



(i)

Figure 5.5 ^{1}H $T_{1,\rho}$ for crystalline and amorphous **1**. (i) Carbonyl carbons; (ii) Alpha carbons; (iii) Side chain carbons;

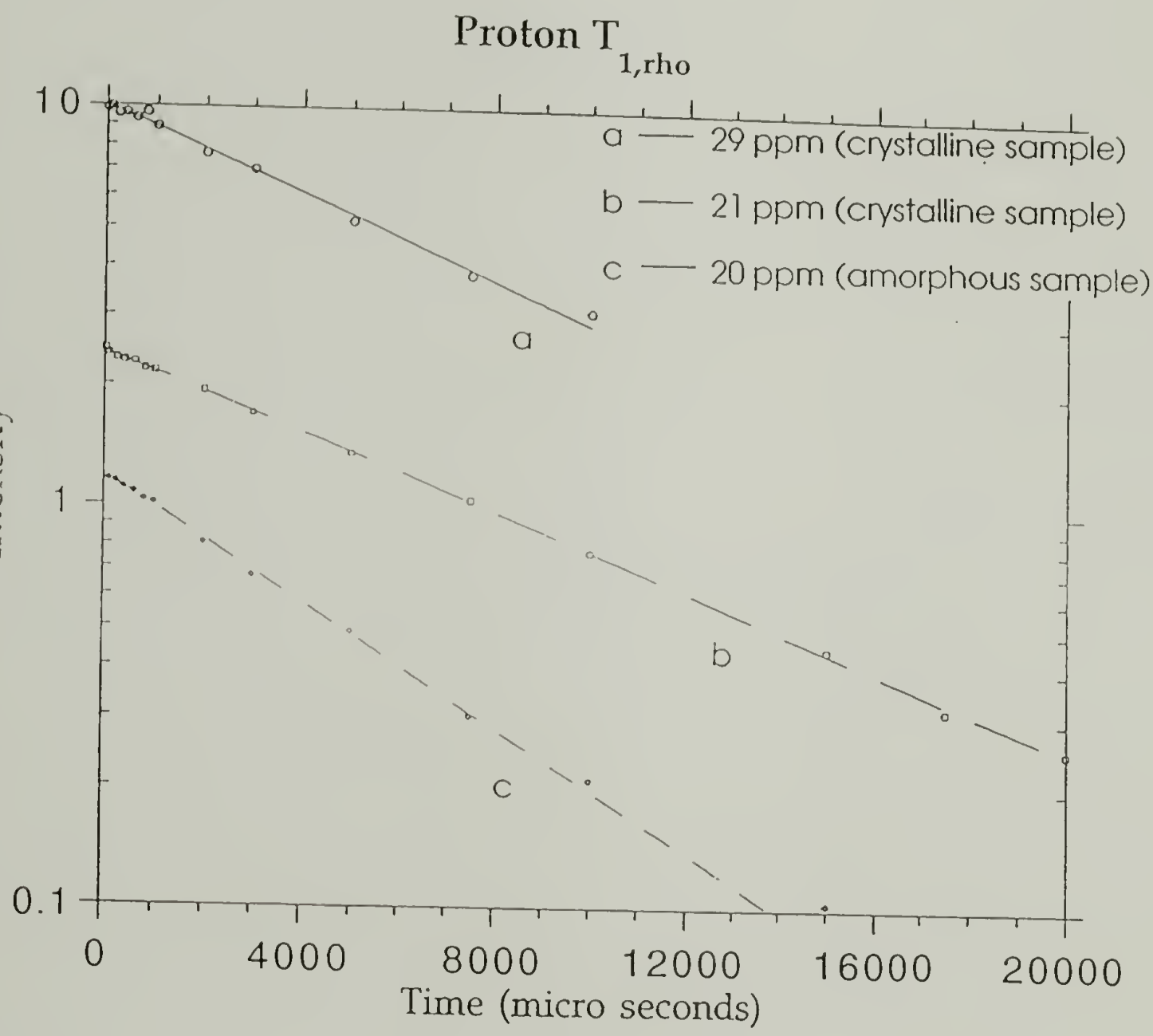
Continued, next page



(ii)

Figure 5.5 (contd.)

Continued, next page



(iii)

Figure 5.5 (contd.)

Table 5.6

Proton $T_{1,\rho}$ for amorphous and crystalline samples of **1**

Peak Position	[a] Proton $T_{1,\rho}$ (ms)
Amorphous Sample:	
172 ppm	2.67
57 ppm	2.23
49 ppm	2.75
43 ppm	2.81
29 ppm	2.65
21 ppm	2.8
17 ppm	2.24
Crystalline Sample:	
172 ppm	7.83
170 ppm	7.86
49 ppm	8.46
43 ppm	8.40
29 ppm	8.02
21 ppm	8.84

[a] The intensity data is fit to the form $I\exp(-t/\tau)$. Refer to Figure 5.1 for the definition of t. The uncertainty in $T_{1,\rho}$ is on the order of 10%.

5.1.3.3.4 Carbon $T_{1,\rho}$ Relaxation of **1**

Figure 5.6, Figure 5.7 and Table 5.7 show the ^{13}C $T_{1,\rho}$ results on **1**. Vis-a-vis the ^{1}H $T_{1,\rho}$, the ^{13}C $T_{1,\rho}$ values are different for different carbons due to the low abundance of ^{13}C nuclei and the consequently lower rate of spin diffusion.

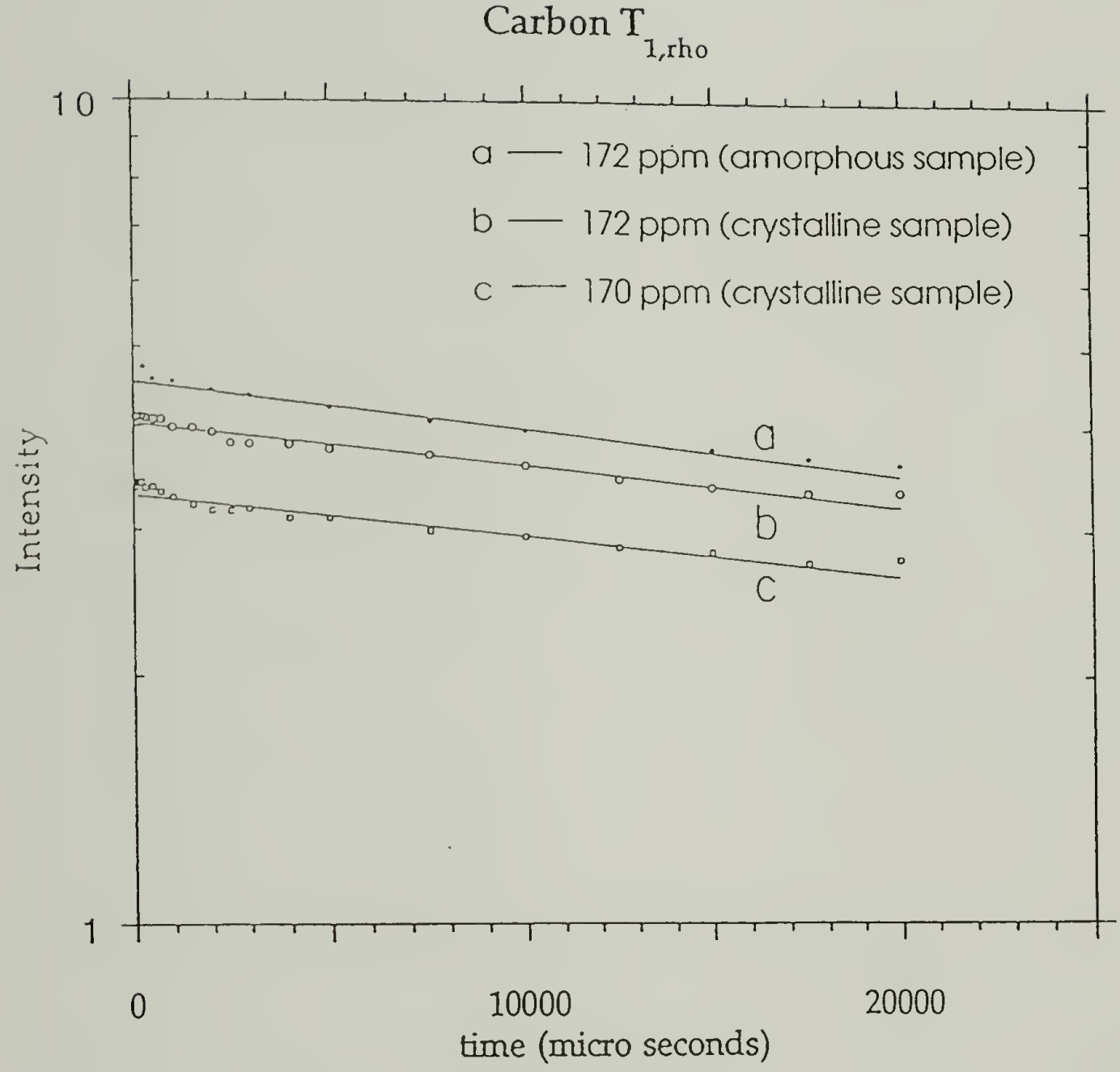
5.1.3.3.4.1 Carbonyl Carbons

The ^{13}C $T_{1,\rho}$ values for the carbonyl carbons are the highest due to the lower mobility of the backbone due to high density of hydrogen bonds, and the absence of attached ^{1}H nuclei. A comparison of the values for the amorphous and crystalline samples shows the carbonyls to exibit similar mobilities in both cases; in both the cases, the relaxation is single component; we believe that most of the carbonyls are hydrogen bonded in the amorphous state also; however, the long range order of the crystal is absent.

5.1.3.3.4.2 Alpha Carbons

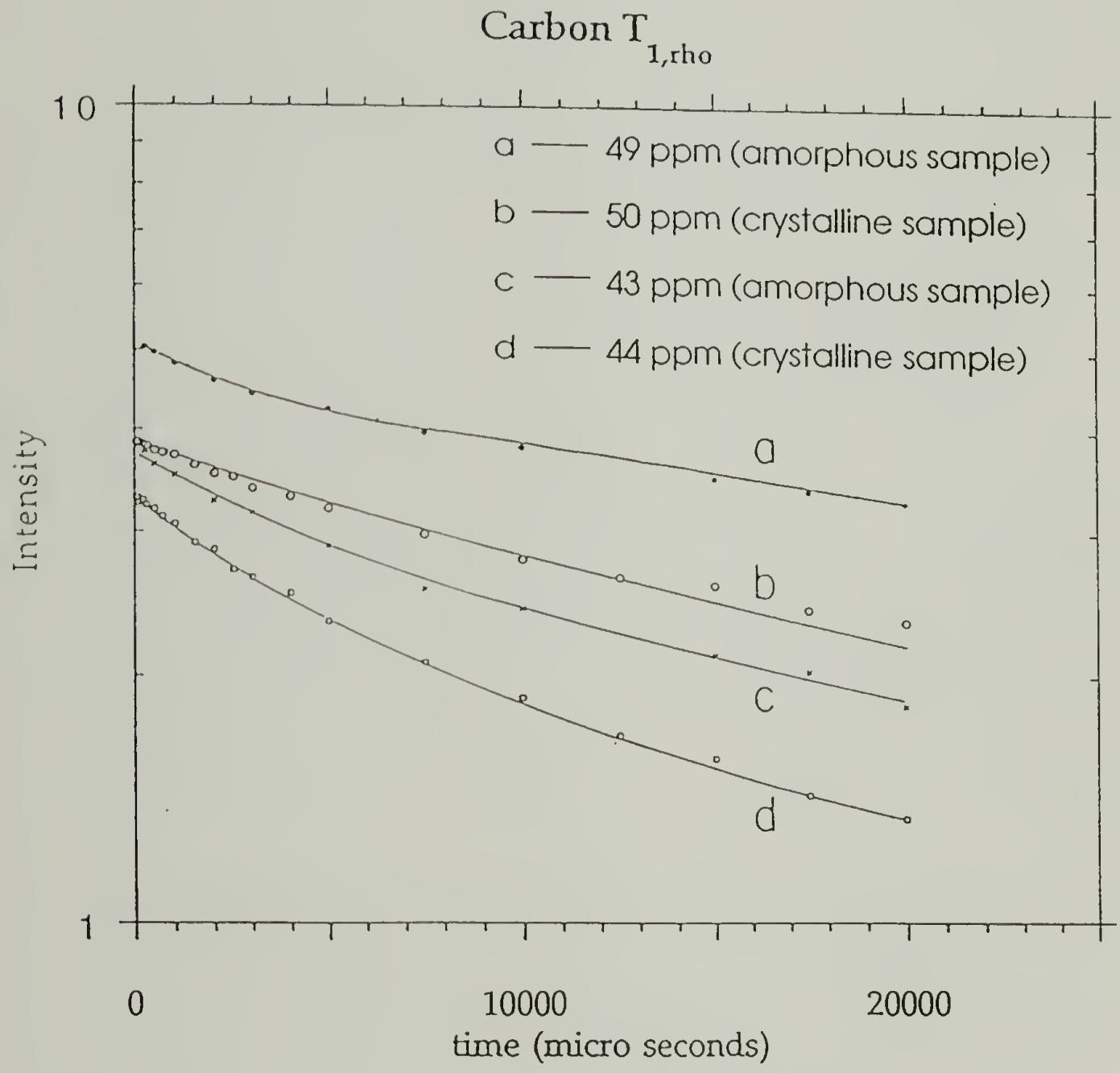
Most of the alpha carbons in both amorphous and crystalline samples display a bicomponent exponential decay of intensity. It was not the intention of this study to investigate the dynamic heterogeneity which leads to such multiexponential behavior. The amorphous and crystalline alpha carbons display very similar behavior; the only exception to this is

Figure 5.6 ^{13}C $T1,\rho$ for main chain carbons of crystalline and amorphous 1. (i) Carbonyl carbons; (ii) Alpha carbons; (iii) Alpha carbon of glutamic acid; (iv) Alpha carbon of glutamic acid (early points)- note the linearity in comparison to the rest.



(i)

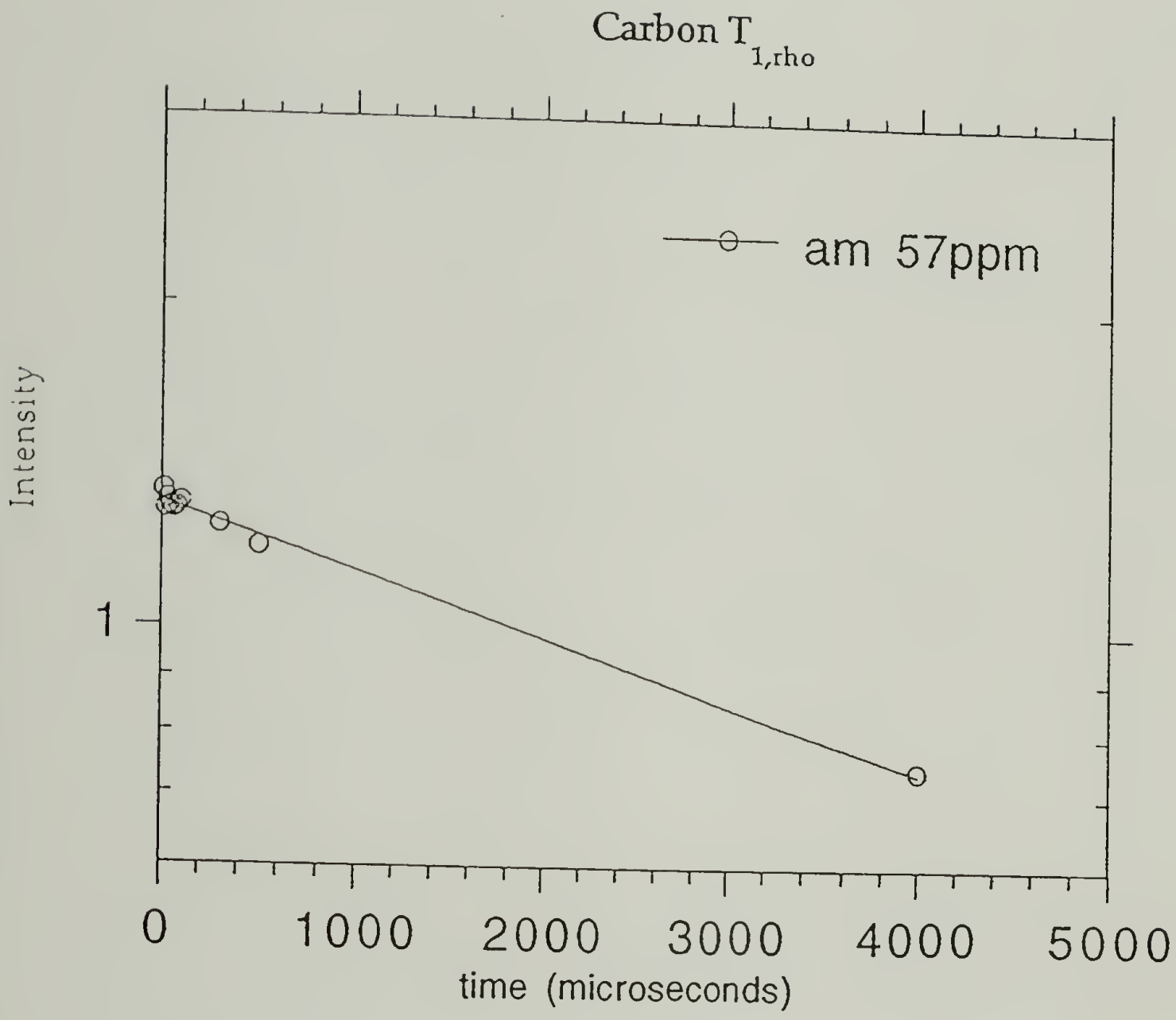
Continued, next page



(ii)

Figure 5.6 (contd.)

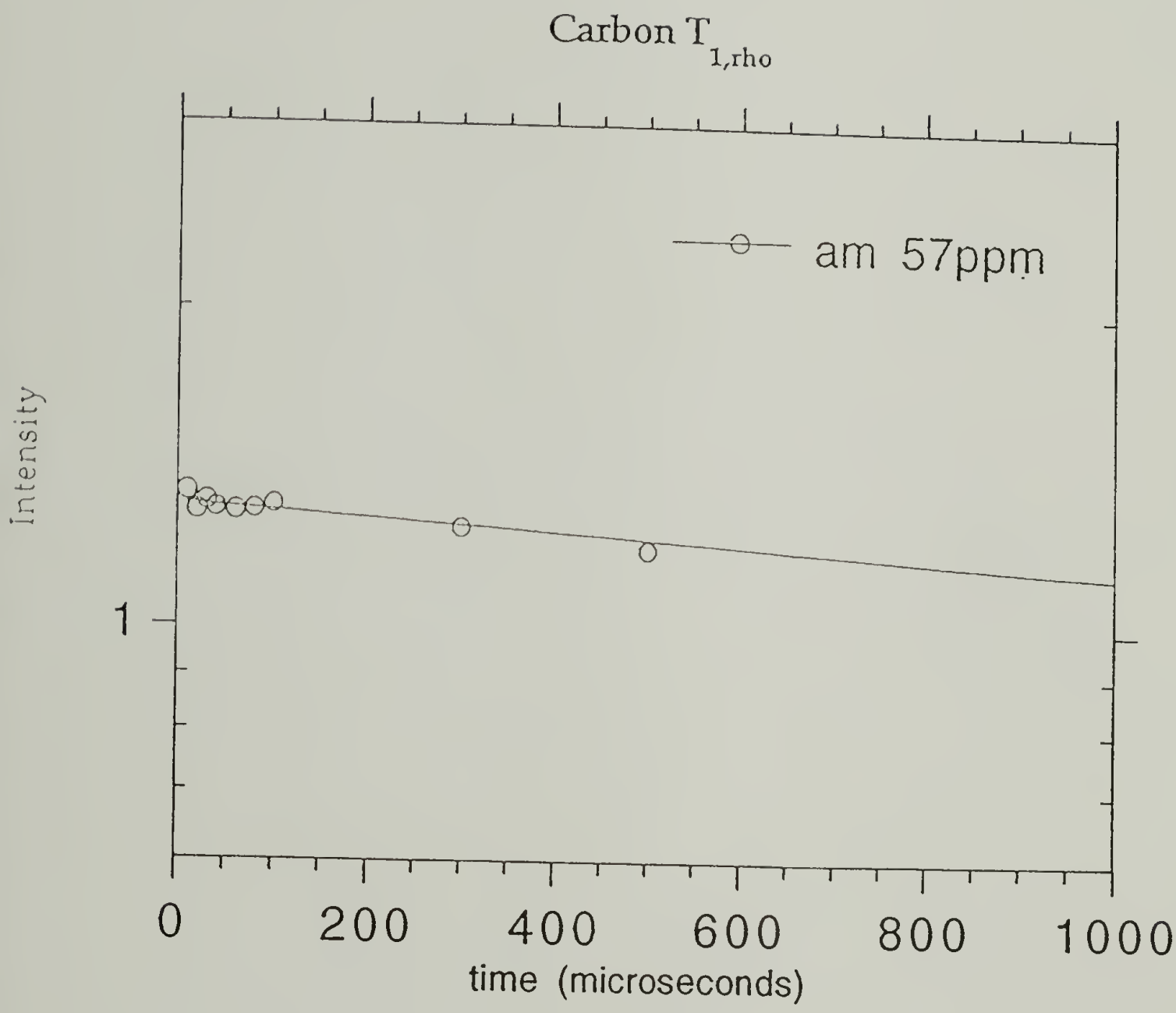
Continued, next page



(iii)

Figure 5.6 (contd.)

Continued, next page



(iv)

Figure 5.6 (contd.)

Table 5.7

Carbon $T1,\rho$ for Amorphous and Crystalline Samples of **1**

Peak Position	[a]Carbon $T1,\rho$ (ms)
Amorphous Sample:	
172 ppm	87 (100%)
57 ppm	7 (100%)
49 ppm	5 (6%); 50 (94%)
43 ppm	5 (26%); 50 (74%)
29 ppm	3 (69%); 50 (31%)
21 ppm	50 (100%)
17 ppm	6 (38%); 61 (62%)
Crystalline Sample:	
172 ppm	87 (100%)
170 ppm	88 (100%)
49 ppm	38 (100%)
43 ppm	7 (43%); 50 (57%)
29 ppm	3 (62%); 51 (38%)
21 ppm	50 (100%)
17 ppm	5 (28%); 61 (72%)

[a] When the decay is bicomponent, the intensity data is fit to the form $I_1 \exp(-t / \tau_1) + I_2 \exp(-t / \tau_2)$. The two relative intensities and relaxation times are reported. Refer to Figure 5.1 for the definition of t. The uncertainity in $T1,\rho$ is on the order of 10%.

the glutamic acid alpha carbon in the amorphous sample which relaxes significantly faster and shows a single component relaxation. The glutamic acid alpha carbon cannot be monitored in the crystalline sample because of the shift of its resonance to 49 ppm where it overlaps with the more intense alanine alpha carbon resonance. One method to monitor the glutamic acid alpha carbon would be to incorporate 100% 2-^{13}C glutamic acid into the polymer; if this were successful, the intensity at 49 ppm would be dominated by the glutamic acid alpha carbon; although labeling experiments were not carried out with 100% 2-^{13}C glutamic acid, all labeling experiments with glutamic acid analogues lead us to conclude that the method of synthesis does not allow us to incorporate labeled glutamic acid into the strain of *E. coli* used for the expression of the polypeptide.

The lack of difference in mobilities between crystalline and amorphous samples is different from similar studies on other less hydrogen bonded polymers; in aliphatic and aromatic semi-crystalline polymers like isotactic polystyrene and poly(phenylene sulfide)[13], the $T_{1,\rho}$s are longer and the T_2s are shorter than their glassy analogues; this is expected due to the reduction in low frequency motion in the crystalline polymer. In our experiments, the alanine and glycine alpha carbons and all the carbonyl carbons relax at the same rate in the crystalline and amorphous samples, probably because of high density of hydrogen bonds; the hydrogen bonds could still be present in the amorphous phase which would reduce mobility. The glutamic acid by virtue of the bulky side chain could be present in regions of relatively larger local void volumes and lower local

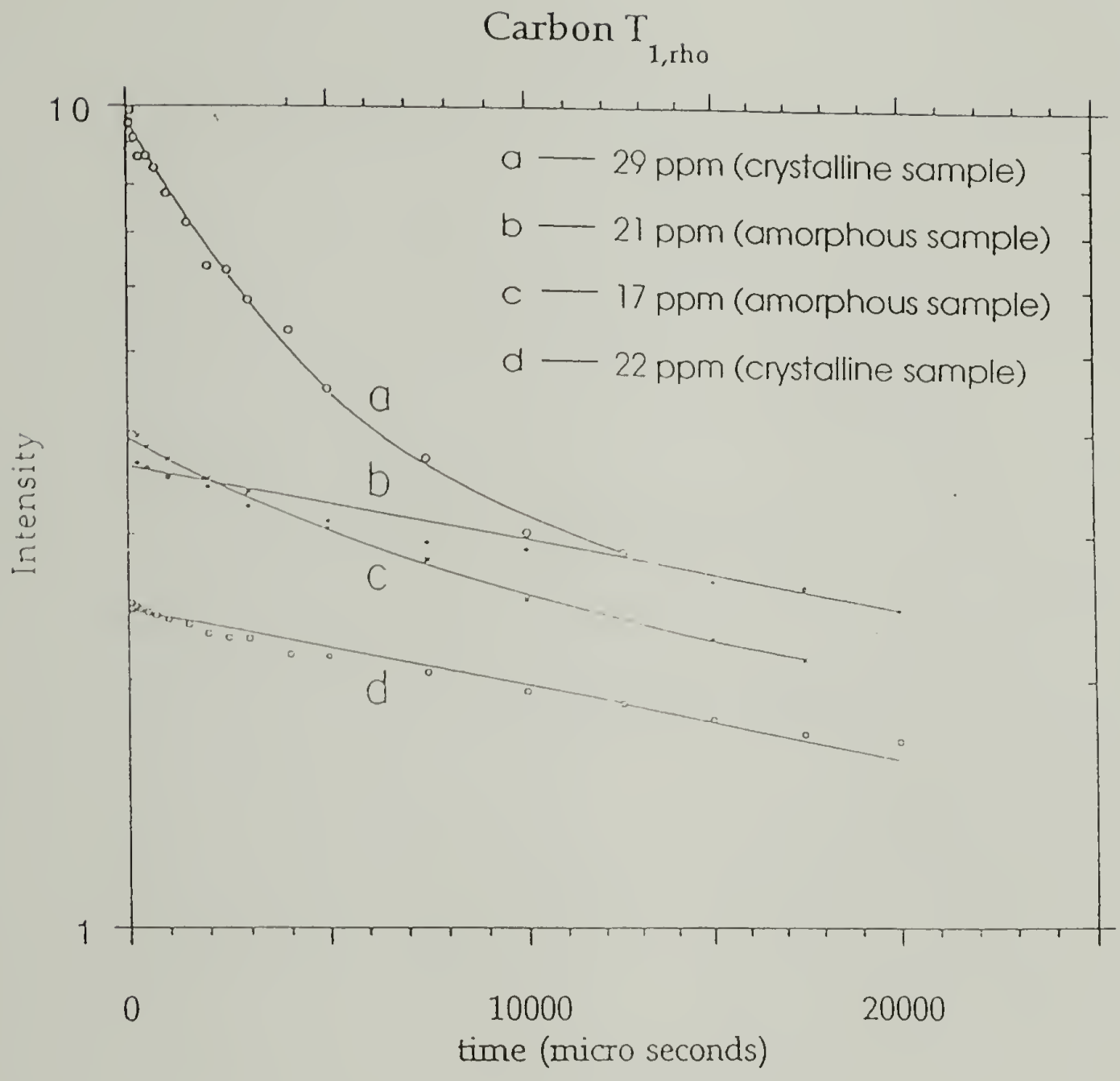
hydrogen bond density; this would explain the faster relaxation in the amorphous sample.

5.1.3.3.4.3 Side Chains of Alanine and Glutamic Acid

The alanyl and glutamyl side chains display the most interesting features with respect to mobility differences between the two phases. These have been depicted in Figure 5.7. The alanine beta carbon corresponding to the amorphous phase (17 ppm) shows a two component exponential decay in both the samples. The shorter ^{13}C $T_{1,\rho}$ could be due to alanine residues in pockets of higher free volume; this could lead to higher mobility and lower relaxation time. On the contrary, the alanine beta carbon resonance corresponding to the beta sheet region (21 ppm) shows a single component relaxation. The resonance at 29 ppm comprises the unresolved beta and gamma glutamyl carbons. The decay of intensity for this resonance is bicomponent and is the fastest; we believe that the bicomponent decay is due to the two unresolved yet dynamically different carbons. Most interestingly for the glutamyl side chains, however, the decay is identical for the crystalline and amorphous samples.

5.1.3.3.4.4 Implications of the Side Chain Mobilities

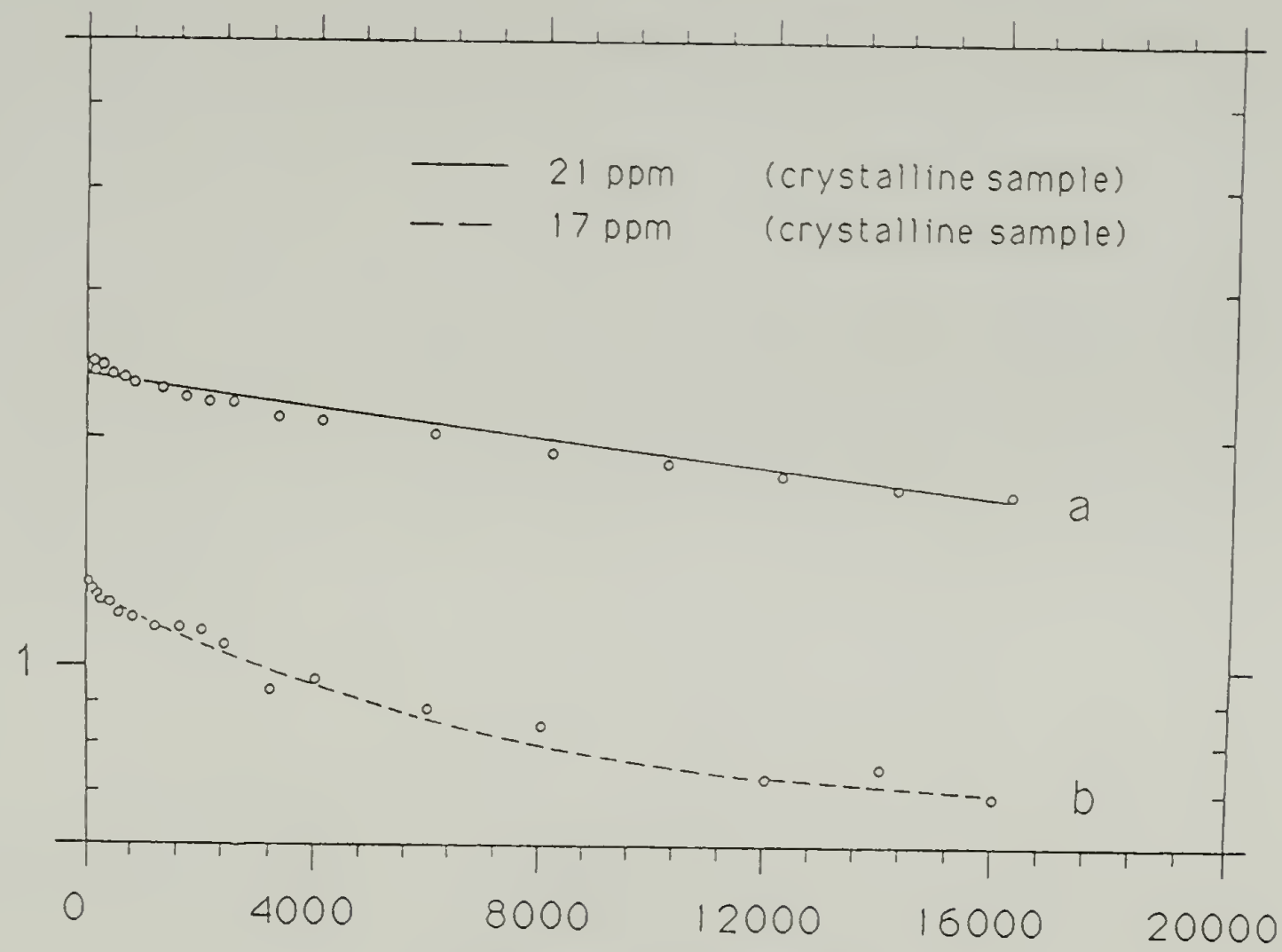
The presence of the glutamic acid residues in crystalline regions would have caused the side chains to pack with less free volume than in the amorphous regions; this should have reduced the mobility and increased the ^{13}C $T_{1,\rho}$ of the side chain carbons in the



(i)

Figure 5.7 ^{13}C $T_{1,\rho}$ for side chain carbons of crystalline and amorphous **1**. (i) shows the comparison between the mobilities of crystalline beta carbon (b,d), amorphous beta carbon (c), and the glutamyl beta and gamma carbons (a). (ii) shows that in the crystalline sample, the crystalline beta carbon (a) relaxes slower than the amorphous beta carbon (b). (iii) shows the glutamyl beta and gamma carbons.

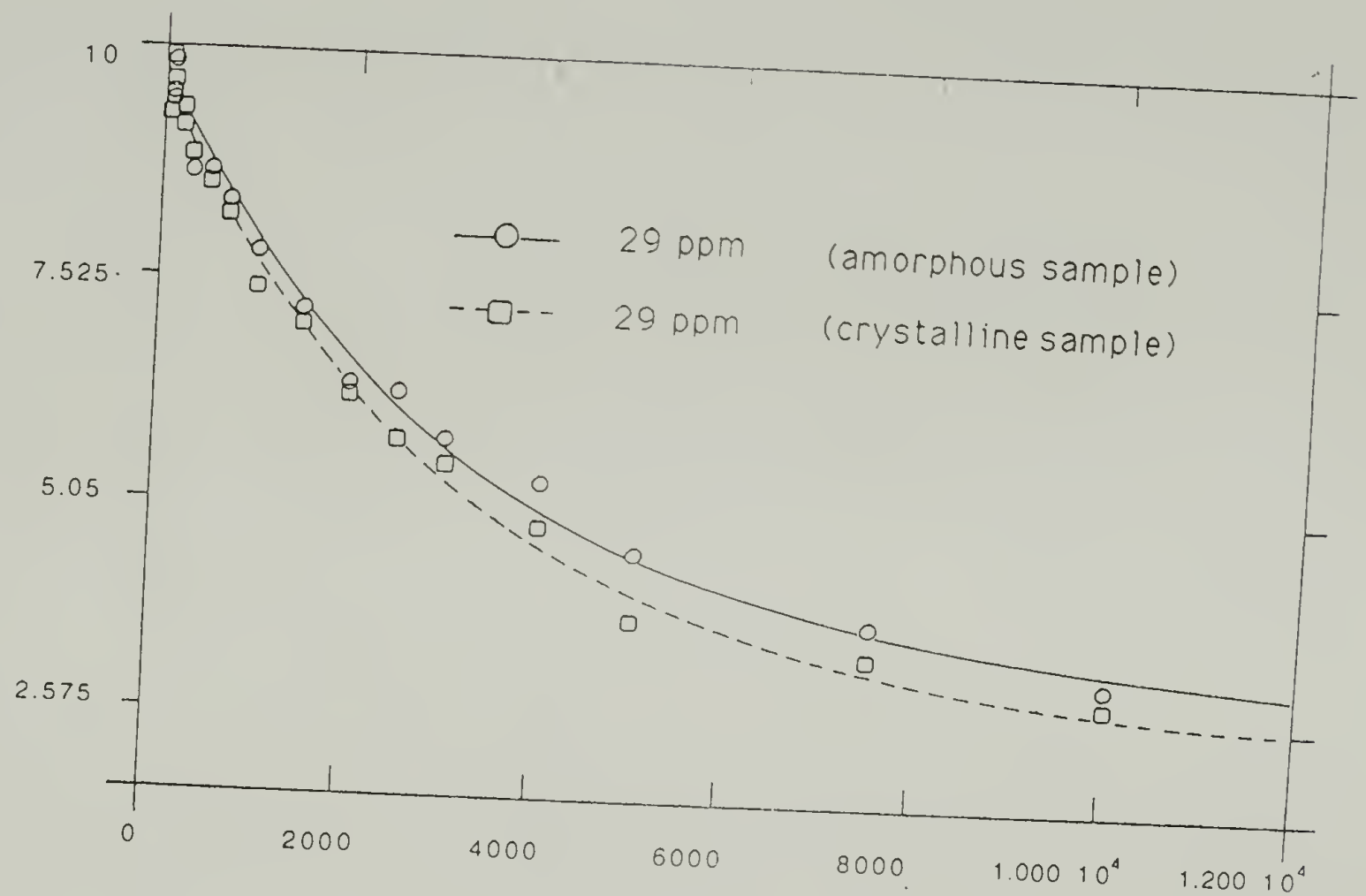
Carbon $T_{1, rho}$



(ii)

Figure 5.7 (contd.)

Continued, next page



(iii)

Figure 5.7 (contd.)

crystalline sample. We have already shown that the alanine side chains suffer a change in their dynamic behavior due to their being included in crystalline regions. The glutamyl side chain should therefore display a similar effect on mobility if it were included in crystalline regions. We would have anticipated a larger effect on the ^{13}C $T_{1,\rho}$ of the glutamyl side chain than that of alanine for two reasons: 1) the side chain is longer than that of alanine and would therefore be more constrained in a crystalline environment. 2) The ^{13}C $T_{1,\rho}$ of the glutamyl side chains is more sensitive to motions as evidenced by the short relaxation time (the spectral density of the glutamyl side chains appears to be better matched than any other carbons at the spin lock $B_{1,C}$ used in the experiment). However, identical relaxation behavior is observed for amorphous and crystalline samples. Hence, we conclude that the glutamyl residues are in amorphous regions in the crystalline sample of **1**.

5.1.3.4 Carbon T_1 Relaxation of **1** in the Laboratory Frame of Reference

The ^{13}C T_1 results are shown in Table 5.8(i). The values are in agreement with those observed in previous work on silks (Table 5.8(ii))[11]. As observed by other workers, the exceptionally short alanine beta-carbon spin-lattice relaxation times are due to rapid C_3 rotation of the methyl groups[11].

Table 5.8

Values of carbon T_1

Table 5.8 (i) Values of carbon T_1 for **1**

Amorphous Sample		Crystalline Sample	
Peak Position	T_1 (s)	Peak Position	T_1 (s)
172 ppm	35.56	172 ppm	46.25
49 ppm	16.23	170 ppm	55.46
43 ppm	19.56	49 ppm	17.59
29 ppm[a]	56.1 (49%)	43 ppm	29.94
	6.0 (51%)	29 ppm[a]	56.1 (30%)
21 ppm	0.611		6.0 (70%)
17 ppm	0.762	21 ppm	0.402

[a] These resonances display a bicomponent decay of intensity. The data is fit to the form $I_1 \exp(-t / \tau_1) + I_2 \exp(-t / \tau_2)$. The two characteristic times are reported.

Continued, next page

Table 5.8 (ii) Values of carbon T_1 for silk II[11]

Residue	Carbon	T_1 (s)
Alanine	Alpha	14
	Beta	0.53
	Carbonyl	20
Serine	Alpha	11
	Beta	0.56, 31
	Carbonyl	20
Glycine	Alpha	20
	Carbonyl	46

5.1.4 Conclusion

The differences between the ^{13}C NMR spectra of amorphous and crystalline forms of **1** have been studied. The crystalline sample shows as high as 52% of alanines in the beta sheet form. The glutamic acid alpha carbon changes its position of resonance from 57 ppm (in the amorphous form) to approximately 49 ppm (in the crystalline form); the beta and gamma glutamyl carbons resonate at 29 ppm in both the forms.

^{13}C $T_{1,\rho}$ measurements show that all the carbonyl carbons relax at the same rate irrespective of the form of the sample. The alanine and glycine alpha carbons also relax relax at similar rates in both the forms; the glutamic acid alpha carbon in the amorphous sample relaxes an order of magnitude faster than the other alpha carbons. The side chains display the most interesting ^{13}C $T_{1,\rho}$ behavior; the 21 ppm resonance in both the crystalline and amorphous samples, which is due to the crystalline alanine beta carbon, relaxes at the same rate; the relaxation of the 17 ppm resonance, which is due to the disordered phase, is identical in both the samples and is faster than the 21 ppm resonance; the glutamic acid side chain carbons relax at identical rates in both the samples and are faster than any other carbon.

The data lead us to propose that the glutamic acid is in identical environments in the crystalline and amorphous samples. The fast relaxation rates of its alpha, beta and gamma carbons lead us to propose that glutamic acid residues are excluded from the crystalline regions in the both the samples. Hence, the data supports the presence of glutamic acid residues at the turn positions.

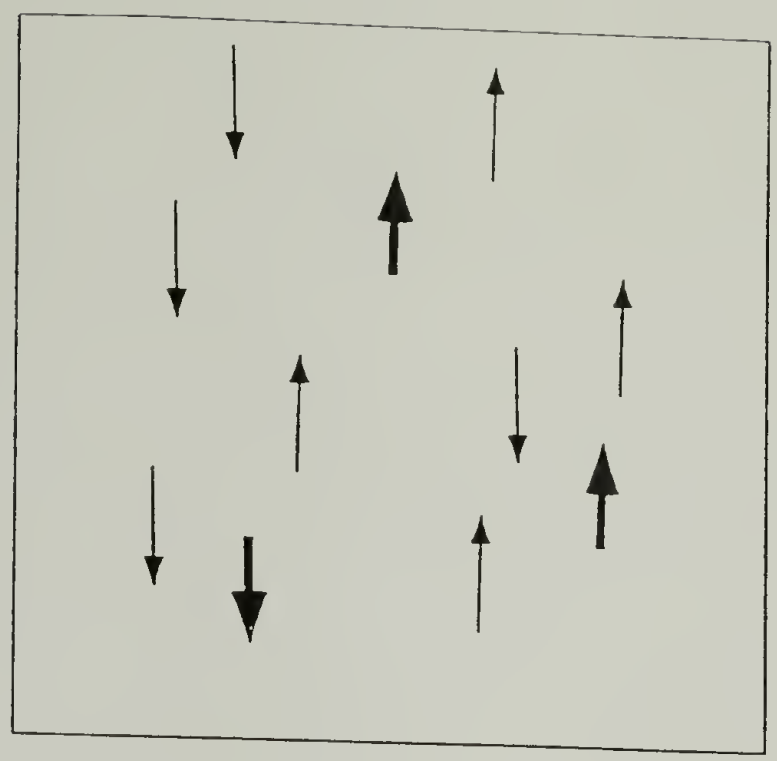
5.2 Cross Polarization Static NMR Spectroscopy of Oriented Mats of **1**

5.2.1 Introduction

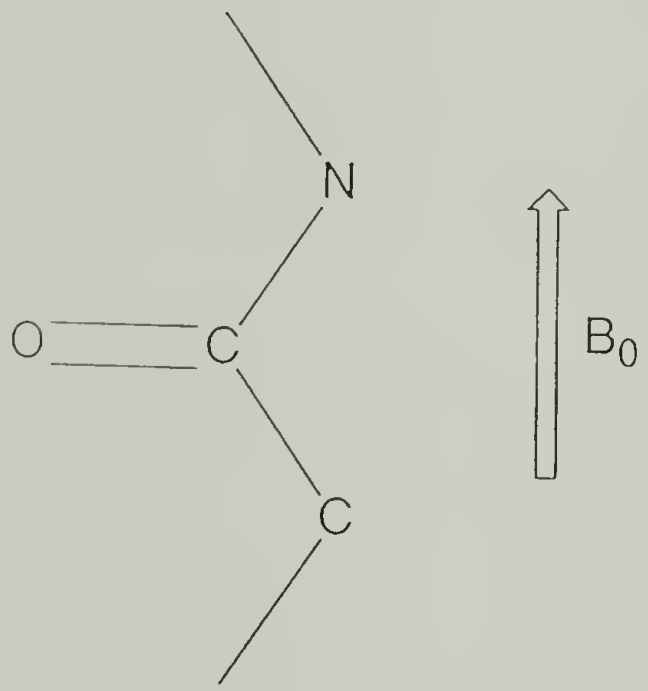
5.2.1.1 The Nuclear Spin Hamiltonian

Figure 5.8 shows the anisotropic contributions to the nuclear spin Hamiltonian[1]. All discussion in this section is for ^{13}C nuclei with the external magnetic field causing the Zeeman splitting to be equivalent to 75.426 MHz. In solid state NMR, resonances which would have been narrow are broadened due to dipolar coupling and chemical shift anisotropy.

The dipolar coupling is due to the effect of neighboring spins at the nucleus in question; in ^{13}C NMR of organic materials, this is predominantly due to neighboring protons and causes line broadening on the order of 50 KHz. In Figure 5.8(i) the carbon nucleus is shown in bold arrows; each nucleus experiences a magnetic field due to adjacent protons which are depicted with lighter arrows; the broadening occurs because the carbon sites are not identical. The chemical shift anisotropy causes the shielding Hamiltonian to vary as a function of the relative orientation of the molecule with respect to the magnetic field; this leads to line broadening on the order of 5 KHz; of course, the exact value of the chemical shift anisotropy (CSA) tensor depends on the specifics of the carbon functionalities. In Figure 5.8(ii), the chemical shift of the carbonyl carbon varies as a function of the angle the peptide bond makes with the magnetic field because the electron density around the ^{13}C is not spherical.



(i)



(ii)

Figure 5.8 Anisotropic contributions to the nuclear spin Hamiltonian. (i) Dipolar coupling due to neighboring protons causes ^{13}C line broadening. The carbon nucleus is shown in bold and the proton in light. (ii) Anisotropy at the carbonyl carbon in a peptide bond. The chemical shift of the carbonyl carbon varies as a function of the angle the peptide bond makes with the magnetic field because the electron density around the ^{13}C is not spherical.

The three principal values of the chemical shift tensor for an amide carbon[2] are given by $\sigma_{11} = 243; \sigma_{22} = 179; \sigma_{33} = 94$; these are the three principal chemical shifts for a carbon nucleus in an amide bond environment. In most solid state experiments, the dipolar coupling is removed by high power broadband dipolar decoupling and the chemical shift anisotropy is reduced by magic angle spinning.

5.2.1.2 Molecular Orientation and its Effect on Lineshapes

Lines obtained from dipolar decoupled spectra without magic angle spinning are broad but contain information regarding the distribution of orientation of the molecules within the sample with respect to the magnetic field. The function for the lineshape when all orientations of the molecule are possible with equal probability (powder lineshape) is given below[1] (σ is the chemical shift):

For $\sigma_{22} < \sigma < \sigma_{33}$,

$$I(\sigma) = \pi^{-1}[(\sigma_{33} - \sigma_{22})(\sigma - \sigma_{11})]^{-1/2} K\left\{\arcsin\left[\frac{(\sigma_{33} - \sigma)(\sigma_{22} - \sigma_{11})}{(\sigma_{33} - \sigma_{22})(\sigma - \sigma_{11})}\right]^{1/2}\right\}$$

and for $\sigma_{11} < \sigma < \sigma_{22}$,

$$I(\sigma) = \pi^{-1}[(\sigma_{33} - \sigma)(\sigma_{22} - \sigma_{11})]^{-1/2} K\left\{\arcsin\left[\frac{(\sigma_{33} - \sigma_{22})(\sigma - \sigma_{11})}{(\sigma_{33} - \sigma)(\sigma_{22} - \sigma_{11})}\right]^{1/2}\right\}$$

where $K(k)$ is a complete elliptic integral of the first kind[22].

A typical powder lineshape is shown in Figure 5.9.

In oriented samples all the molecular orientations are not present with equal probability and the lineshape is a deviation from the powder lineshape; indeed, in the event of only one orientation present in the sample, the line is very narrow even without magic angle spinning. The value of the observed chemical shift is given by

$$\sigma_{obsv} = \overline{B}.\overline{\overline{\sigma}}.\overline{B}^{T}$$

where $\overline{B}$ is the normalized external magnetic field and σ_{obsv} is the observed chemical shift.

In the principal frame of reference, this is equivalent to

$$\sigma_{obsv} = \begin{bmatrix} B_x & B_y & B_z \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix} \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} \quad \text{where } \left|\overline{B}\right| = 1$$

An intermediate situation between a powder and one unique orientation would cause some orientations to be preferred and some orientations to be excluded in the sample. Integration over all these existing orientations would yield $I(\sigma)$ which would be the corresponding lineshape. A more elaborate mathematical treatment can be obtained from other sources[1].

5.2.1.3 Use of Static CP NMR to Determine Orientation

Gramicidin A (gA) is a 15 amino acid peptide[23] that forms a channel in a lipid environment. Solid state NMR experiments have been instrumental in the determination of the structure of this peptide within a dimyristoyl phosphatidylcholine (DMPC) bilayer[24]. Labeled polypeptides containing ^{13}C or ^{15}N in different residues along the peptide were synthesized by solid phase peptide synthesis and static NMR experiments were carried out on them[23-25]. Samples were prepared by dissolving DMPC and gA (in molar ratios of 8:1) in an organic solvent; the samples were oriented by spreading on glass coverslips. Since the orientation was very good, static experiments yielded sharp lines which were then correlated to the local orientation of the molecule at the location of the magnetic nucleus. Analysis of lineshapes from a large number of labeling experiments helped assign a secondary structure to the peptide. The results show that the channel in the lipid bilayer comprises dimers, each comprising a single stranded helix with six/seven residues per turn. Opella and coworkers have used a similar strategy to determine the structure and orientation of the antibiotic peptide, magnin, in membranes[26]. Magnin 2 is an antibiotic peptide comprising 23 amino acids and is known to disrupt electrochemical gradients across cell membranes of bacteria. Solid state static NMR experiments[26] show that the peptide is alpha-helical and is parallel to the lipid bilayer.

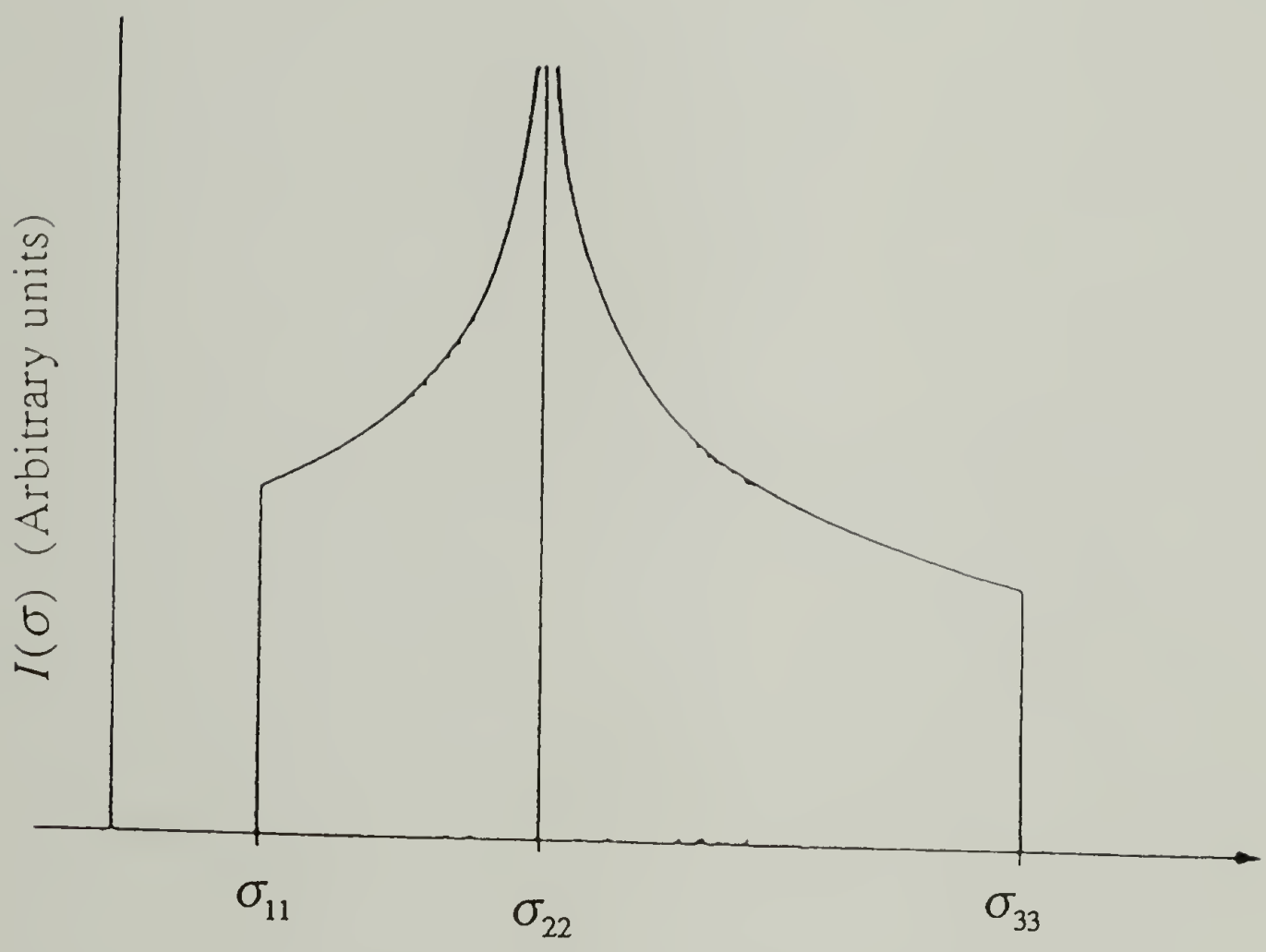


Figure 5.9 Typical powder lineshape[1].

5.2.1.4 Static CP NMR and Determination of Orientation and Folding Habit of **1**

We were interested in using this technique for two purposes:

1) We have already shown in Chapter 3 that **1** forms needle-like crystallites when crystallized from formic acid. It was also stated in Chapter 3 that we were uncertain of the orientation of the needles within the plane of the mat. We wanted to determine the orientation of the needle-like crystallites within the mats.

2) If the chains were to fold at the anticipated locations, the glutamic acid residues would be sampling orientation space that was different from that of alanine or glycine. A comparison between the lineshapes of the carbonyl carbon of glutamic acid and those of alanine or glycine would help in determining the presence or absence of the glutamic acid in the turn.

5.2.2 Methods

5.2.2.1 Synthesis of Labeled **1**

Polypeptides labeled with $^{13}C{=}O$ at the alanine position (**1-A**) and with $^{13}C{=}O$ at the glycine position (**1-G**) were synthesized as described in Chapter 4.

5.2.2.2 Crystallization of 1

Polypeptide **1** was dissolved in 70% formic acid at 40 mg/ml and stirred for ca. 24 h at ca. 700 rpm on a Thermix stirrer model 220T (Fisher Scientific). The sample formed a gel which was precipitated by adding excess water. This was sedimented by centrifugation and washed several times with water and then with methanol. Finally the gel was resuspended in methanol. This was spun down at 13500 x g for 10 min.

5.2.2.3 Sample Preparation for Static CP NMR Experiments

In order to obtain oriented samples, the wet gel obtained after centrifugation of the suspension in methanol was placed between glass slides covered with a teflon filter on one side and a glass coverslip on the other; the plates were subsequently clamped and the oriented gel was allowed to dry in air at room temperature. This process was repeated several times to yield multiple oriented layers of the crystalline polypeptide, stacked on top of each other on a glass coverslip. The sample coated glass coverslip was then cut into several rectangular pieces of 10 mm × 5 mm; a few of these glass pieces were stacked in a home-made sample holder in a static NMR probe . Experiments were carried out with the giass coverslips perpendicular to the magnetic field and also with the glass coverslips parallel to the magnetic field.

5.2.2.4 Solid State Static CP NMR

Solid state NMR spectra were obtained on a Bruker ASX 300 spectrometer with a ^{13}C frequency of 75.426 MHz. A solid state static probe fabricated by Bruker was used and a home-made glass sample holder held the oriented polypeptide at the required angle to the magnetic field. Spectra were obtained with broad band proton decoupling and cross polarization.

5.2.3 Results and Discussion

5.2.3.1 Synthesis of **1-A** and **1-G**

Analogues of polypeptide **1** were synthesized with the carbonyl of the alanine position enriched with ^{13}C (**1-A**) and with the carbonyl of the glycine position enriched with ^{13}C (**1-G**). The levels of enrichment in these polypeptides have been discussed in Chapter 4. All attempts at incorporating encriched glutamic acid were unsuccessful.

5.2.3.2 Models for Orientation of Crystallites of **1**

Figure 5.10 shows the principal directions for a ^{13}C in a carbonyl carbon in a peptide bond[27]. Figure 5.11 shows two possible orientations for the needle-shaped crystallites (Chapter 3) within the plane of the mat which can explain the results from x-ray diffraction experiments.

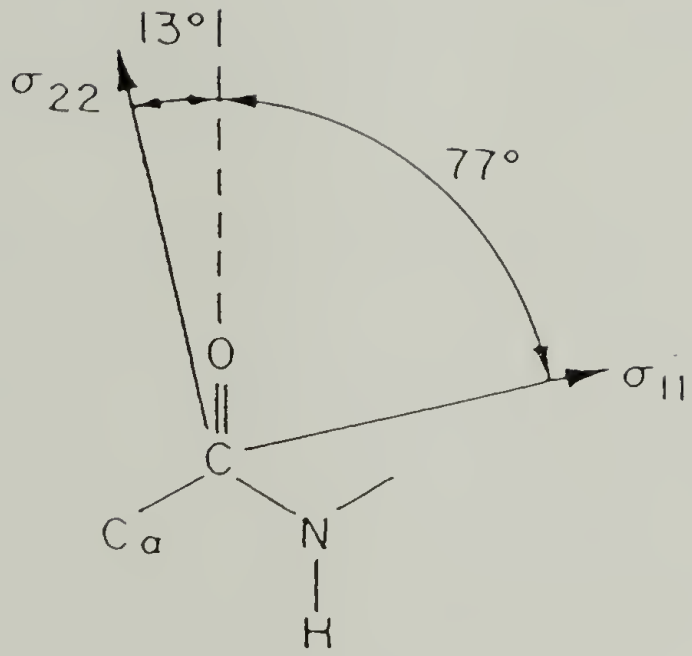


Figure 5.10 Principal directions in the chemical anisotropy shift tensor of the peptide carbonyl ^{13}C. σ_{33} is perpendicular to the plane of the figure (Adapted from work by Stark[27])

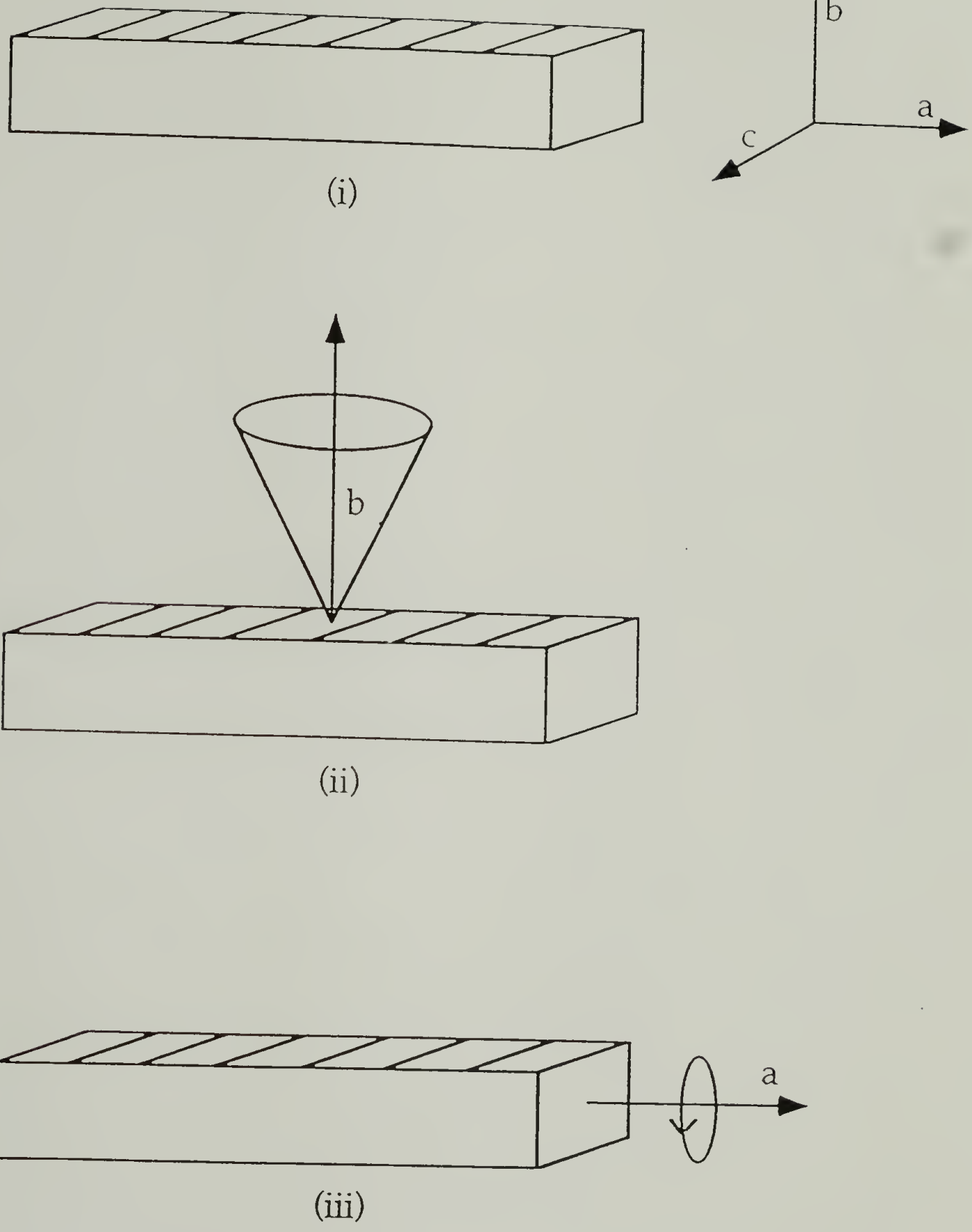


Figure 5.11 Possible orientation of crystallites within the plane of the mat. The mat is a horizontal plane perpendicular to the plane of the figure. Refer to the text for details.

Figure 5.11(i) shows a needle-like crystallite of **1**. Figure 5.11(ii) shows a model which can qualitatively explain the intesities of the diffraction patterns. The crystallites are imperfectly oriented such that the *b* axis is predominantly perpendicular to the plane of the mat; the probability of the orientation drops as the angle subtended by the *b* axis with the mat normal increases.

The presence of crystallites as shown in Figure 5.11(iii) in conjunction with the appropriate scaling of the reflection intensity by the Lorentz factor can also predict the oriented diffraction patterns qualitatively; since no intensity measurements were carried out on the diffraction patterns, the prediction is only qualitative. This requires the crystallites to be oriented such that the *a* axis in every crystallite is along the plane of the mat; the crystallites are dispersed randomly when viewed orthogonal to the mat; rotation around the 'a' axis gives rise to different crystallite orientations, all of which are present with equal probability.

5.2.3.3 Correlation between Orientation of Crystallites and the CSA Tensor

5.2.3.3.1 Imperfect *b* Axis Orientation

If all the crystallites were oriented as shown in Figure 5.11(ii), the peptide bond would be such that σ_{33} points along the *b* axis. The needles are oriented along the mat such that the *a* axis is always along the plane of the mat. Such an orientation results in the most perfect form of Figure 5.11(ii); Figure 5.12 shows this situation with the magnetic field

perpendicular to the plane of the mat. The magnetic field, $\overline{B}$, is always parallel to *b* for all crystallites; free rotation of the crystallites around the *b* axis leads to the isotropy on the a-c plane. Hence, the molecular orientation is such that all orientations of the CSA tensor involving rotations around *b* are possible with equal probability; the fate of the chemical shift tensor is shown in Figure 5.12(ii). Hence, the magnetic field for all the crystallites is orthogonal to σ_{11} and σ_{22}, thereby causing no contribution of these two components to the observed chemical shift. Figure 5.13(i) shows the lineshape when the magnetic field is perpendicular to the mat. The result is a sharp line at σ_{33}. Figure 5.13(ii) shows the lineshape when the glass coverslip is parallel to the magnetic field; in this case, σ_{33} is orthogonal to the magnetic field and thereby does not contribute to the chemical shift; all chemical shifts between σ_{11} and σ_{22} are possible with equal probability as shown in Figure 5.12.

A more realistic case is depicted in Figure 5.11(ii); in this instance, the orientation is not perfect; there exists a distribution of the needle-like crystallites as a function of the angle between the direction *b* and the direction orthogonal to the mat; although most of the crystals are oriented with *b* perpendicular to the mat, there are a finite number of crystals which are not. Hence, the lineshape is not as perfect as depicted in Figure 5.13; the exact lineshape depends on the exact distribution of the crystals. However, the features of the lineshapes will be similar to those depicted in Figure 5.13; one would anticipate the line shape to be skewed towards σ_{33}

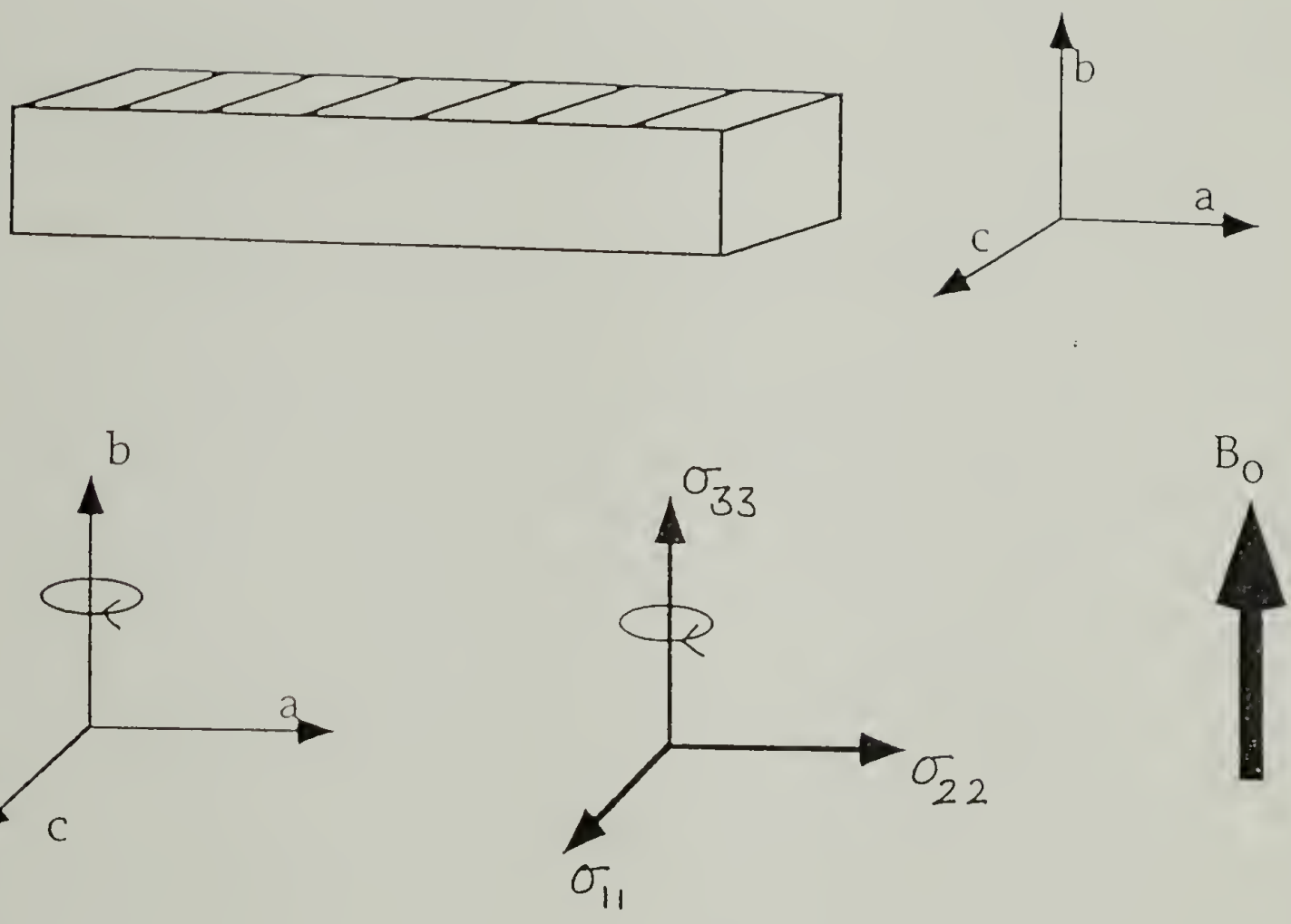


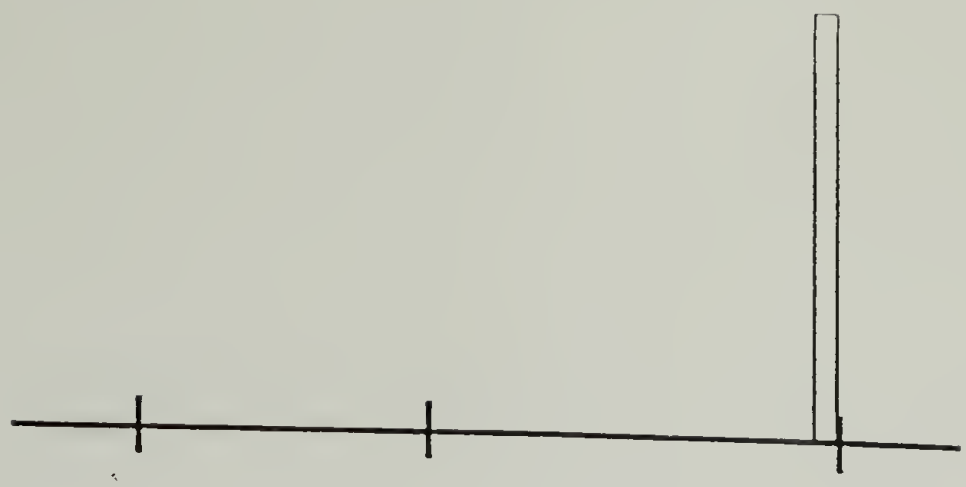
Figure 5.12 Geometric and chemical shift constraints for a situation with orientation as shown in Figure 5.11(ii).

when the magnetic field is perpendicular to mat; one would anticipate the line shape to be such that most of the intensity is between σ_{11} and σ_{22} when the magnetic field is parallel to mat.

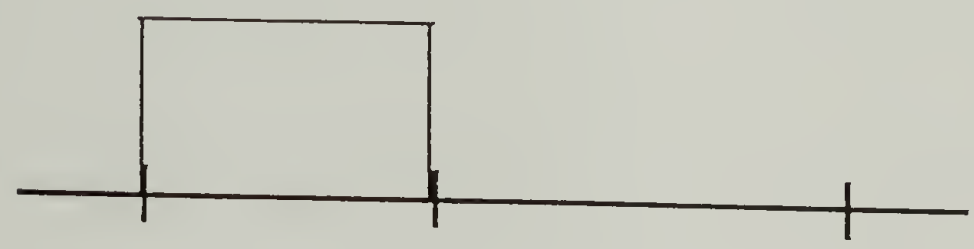
All the analysis above has been carried out with a model in which the peptide bond is perpendicular to *b*. Figure 5.14(v) shows the projection of a pleated beta sheet on the b-c plane. The angle between *b* and the normal to the peptide bond is calculated to be 24.3°. Figure 5.14(ii) and (iii) show the orientation of the chemical shift anisotropy tensor as a result of the pleats in the beta sheet structure. When the magnetic field is perpendicular to the mat, the component of σ_{33} along the magnetic field would be such that the sharp line is observed not at σ_{33} but at a value at lower field than σ_{33}.

5.2.3.3.2 Free Rotation Around the Crystallographic *a* Axis

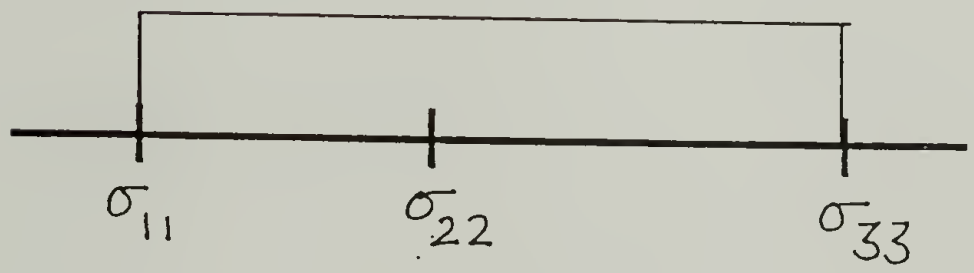
The other possible mode of orientation for the crystallites is shown in Figure 5.11(iii). The crystallographic directions and the CSA tensor are shown in Figure 5.15. A qualitative analysis on the lineshape of this kind of orientation leads us to propose the lineshape as shown in Figure 5.13(iii) when the magnetic field is perpendicular to the mat; one would anticipate a wide range of chemical shifts between σ_{33} and σ_{11}; since σ_{11} is 13° removed from the chain direction (Figure 5.10), the exact lineshape will differ in detail from this qualitative estimate but the trend should be conserved.



(i)

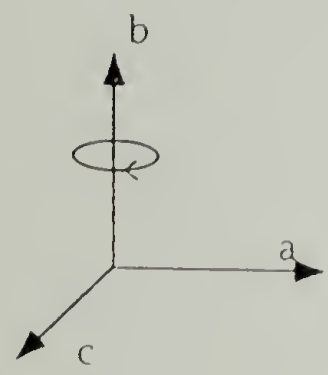


(ii)

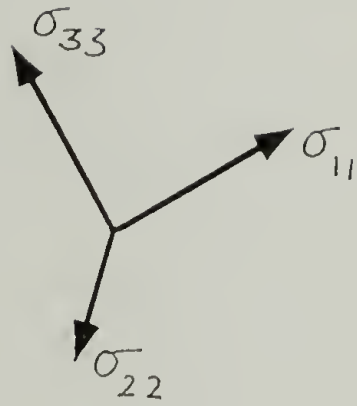


(iii)

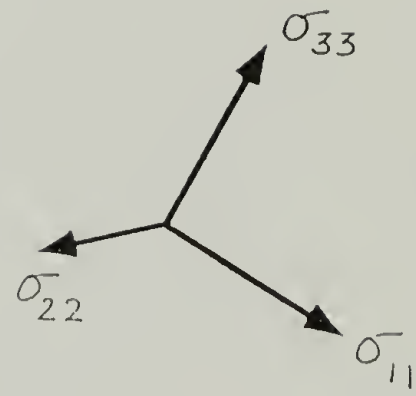
Figure 5.13 Predicted lineshapes for (i) orientation as in Figure 5.12 and the magnetic field perpendicular to the mat, (ii) orientation as in Figure 5.12 and the magnetic field parallel to the mat, (iii) orientation as shown in Figure 5.15.



(i)

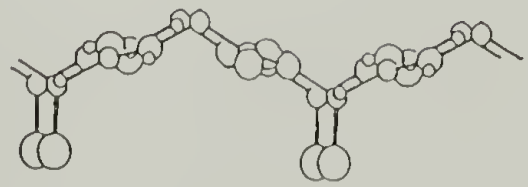


(ii)



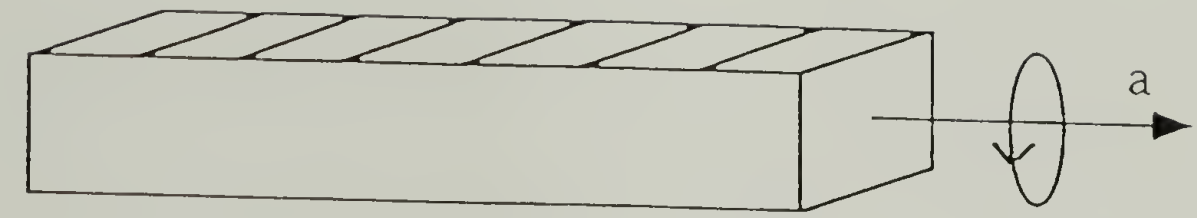
(iii)

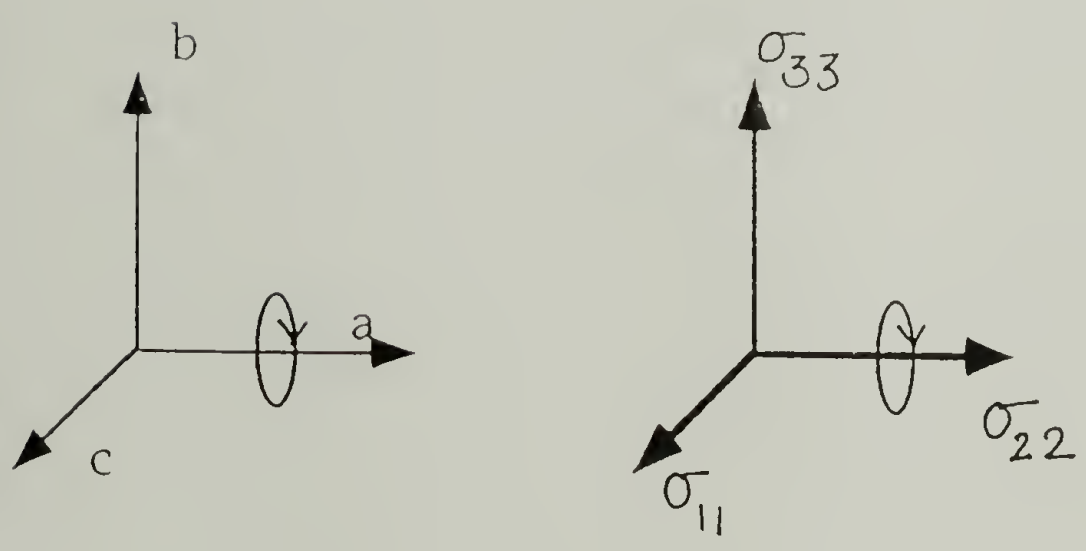
(iv)



(v)

Figure 5.14 Crystallographic and CSA constraints due to formation of pleats in the antiparallel beta sheet structure. (i) shows the free rotation around 'b' as depicted in Figure 5.11(ii). (ii) and (iii) show the constraints on the CSA tensor due to the formation of pleats as shown in (v). (iv) shows the direction of the magnetic field which is perpendicular to the mat.





(i) (ii) (iii)

Figure 5.15 Geometric and chemical shift constraints for a situation with orientation as shown in Figure 5.11(iii). (i) The crystallite can rotate around the 'a' axis freely; in any one of these orientations, the crystallite can rotate freely around a direction perpendicular to the mat. (ii) The fate of the CSA tensor with this kind of rotation. (iii) If the magnetic field is perpendicular to the mat, a line shape as shown in Figure 5.13(iii) would be observed.

When the mat is parallel to the magnetic field, the situation is more complicated and the qualitative prediction of the lineshape cannot be carried out with out an elaborate simulation; however, the mode of orientation requires σ_{33} to be along the magnetic field in many instances; this implies that σ_{33} is not always orthogonal to the magnetic field; hence, we would anticipate a fair amount of intensity in the region between σ_{33} and σ_{22}.

5.2.3.4 Observed Lineshapes for Oriented Samples of 1

Figure 5.16 shows the lineshapes observed for a crystalline, oriented sample of **1-G**; the lineshapes are identical for a similar experiment with **1-A**. Figure 5.16(i) shows the powder lineshape; the lineshape shown in Figure 5.16(ii) is observed when the magnetic field is parallel to the plane of the mat; when the mat is perpendicular to the magnetic field, the lineshape shown in Figure 5.16(iii) is observed. From the data presented in Figure 5.4 and Table 5.5, we estimate the amorphous content in the crystalline sample at 49%; this amount would contribute a powder lineshape in the oriented spectra. Figure 5.17 shows the lineshapes obtained after subtraction of the corresponding powder fraction from the lineshapes.

The lineshape corresponding to the mat being oriented parallel to the magnetic field (Figure 5.17(ii)) displays practically no intensity in the region between σ_{33} and σ_{22}; the line shape is similar to the one depicted in Figure 5.13(ii) with the exception that there is higher intensity around σ_{22}.

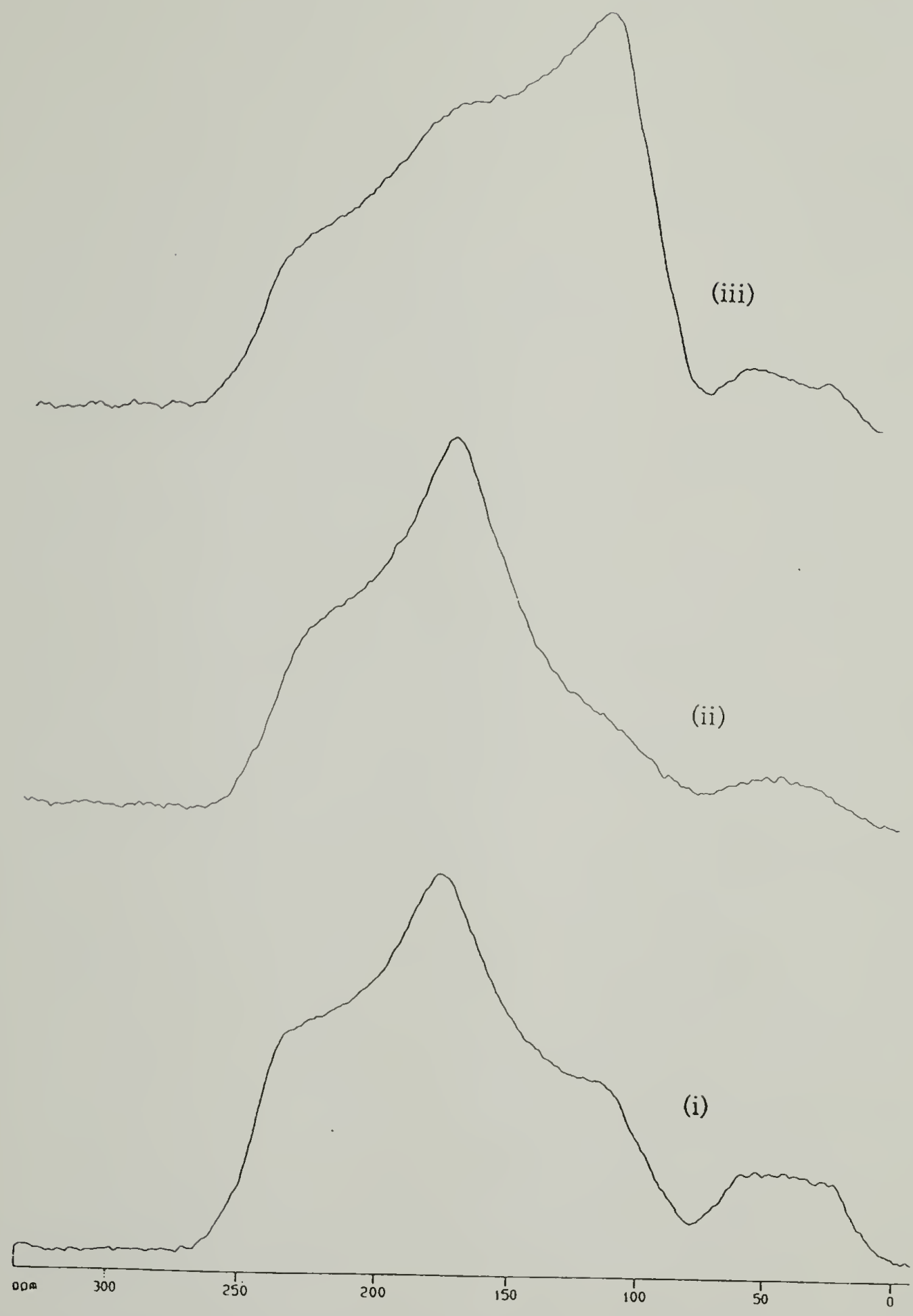


Figure 5.16 Lineshapes obtained for CP spectra of **1-G** (i) Powder lineshape. (ii) Magnetic field parallel to the mats. (iii) Magnetic field perpendicular to the mats.

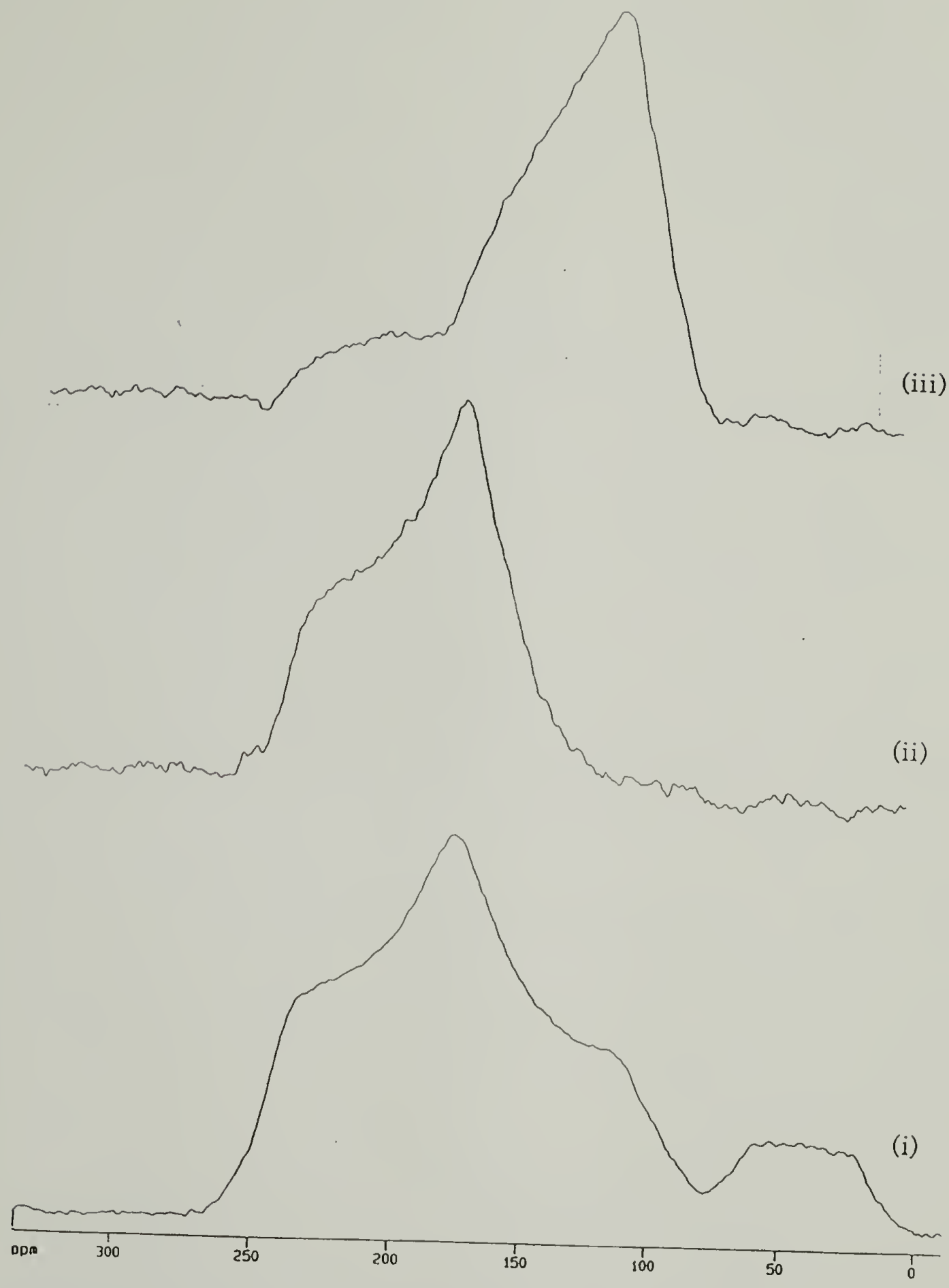


Figure 5.17 Lineshapes obtained after subtraction of 49% powder contribution due to amorphous content for CP spectra of **1-G** (i) Powder lineshape. (ii) Magnetic field parallel to the mats. (iii) Magnetic field perpendicular to the mats. Note: $\sigma_{11} = 243; \sigma_{22} = 179; \sigma_{33} = 94$ for the amide carbon.

Lack of perfect orientation as shown in Figure 5.11(ii) leads to a contribution of σ_{33} to the observed chemical shift in some of the crystallites; we believe that this causes the deviation of the observed lineshape form the predicted one.

The lineshape observed when the mat is perpendicular to the magnetic field (Figure 5.17(iii)) is practically devoid of intensity in the region between σ_{11} and σ_{22}. The line shape shows a sharp maximum at 110 ppm. The breadth of the line is due the presence of multiple orientations. The value of σ_{33} for a carbonyl in peptide bond has been estimated at 94 ppm^2; however, the maximum in Figure 5.17(iii) is at 110 ppm; this is due to the mismatch between the crystallographic *b* axis (the magnetic field axis) and the σ_{33} direction as outlined in Figure 5.14.

5.2.4 Conclusion

Crystalline 1 has been shown to form needle-like crystallites in Chapter 3. The orientation of these crystallites has been studied in this section of this chapter. Mats of oriented samples can be made for CP static NMR experiments. The lineshape analysis is consistent with the crystallites being an imperfect form of the following orientation: The crystallographic *b* axis is perpendicular to the mat and there is free rotation around *b*.

5.3 References

(1) Haeberlen, U. *High Resolution NMR in Solids*; Academic Press: New York, 1976.

(2) Bovey, F. A. *Nuclear Magnetic Resonance Spectroscopy*; Academic Press: San Diego, 1988.

(3) Asakura, T. *Makromol. Chem., Rapid Commun.* **1986**, *7*, 755-759.

(4) Asakura, T.; Watanabe, Y.; Itoh, T. *Macromolecules* **1984**, *17*, 2421-2426.

(5) Asakura, T.; Suzuki, H.; Watanabe, Y. *Macromolecules* **1983**, *16*, 1024-1026.

(6) Saito, H.; Iwanaga, Y.; Tabeta, R.; Narita, M.; Asakura, T. *Chem. Lett. (Chem. Soc. Jap.)* **1983**, 427-430.

(7) Ishida, M.; Asakura, T.; Yokoi, M.; Saito, I. *Macromolecules* **1990**, *23*, 88-94.

(8) Saito, H.; Tabeta, R.; Shoji, A.; Ozaki, T.; Ando, H. *Macromolecules* **1983**, *16*, 1050-1057.

(9) Saito, H.; Tabeta, R.; Asakura, T.; Iwanaga, Y.; Shoji, A.; Ozaki, T.; Ando, I. *Macromolecules* **1984**, *17*, 1405-1412.

(10) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311-333.

(11) Saito, H.; Ishida, M.; Yokoi, M.; Asakura, T. *Macromolecules* **1990**, *23*, 83-88.

(12) Asakura, T.; Demura, M.; Watanabe, Y.; Sato, K. *J. Polym. Sci. Part B* **1992**, *30*, 693-699.

(13) Schaefer, J.; Stejskal, E. O.; Buchdahl, R. *Macromolecules* **1977**, *10 (2)*, 384-405.

(14) Schaefer, J.; Stejskal, E. O.; Steger, T. R.; Sefcik, M. D.; McKay, R. A. *Macromolecules* **1980**, *13*, 1121-1126.

(15) Chu, C. W.; Dickinson, L. C.; Chien, J. C. W. *J. Appl. Polym. Sci.* **1990**, *41*, 2311-2325.

(16) Dickinson, L. C.; Yang, H.; Chu, C. W.; Stein, R. S.; Chien, J. C. W. *Macromolecules* **1987**, *20*, 1757-1760.

(17) Poliks, M. D.; Schaefer, J. *Macromolecules* **1990**, *23*, 3426-3431.

(18) Stejskal, E. O.; Schaefer, J.; Sefcik, M. D.; McKay, R. A. *Macromolecules* **1981**, *14*, 275-279.

(19) Torchia, D. A. *J. Mag. Reson.* **1978**, *30*, 613-616.

(20) Parkhe, A. D.; Krejchi, M. T. K.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Manuscript in preparation*

(21) Zhang, G. *Personal discussion*

(22) Gradshteyn, I. S.; Ryzhik, I. M. *Table of Intgrals, Series and Products*; Academic Press: New York, 1980, pp 626-627.

(23) Mai, W.; W., H.; C., W.; Cross, T. A. *Protein Science* **1993**, *2*, 532-542.

(24) Ketchem, R. R.; Hu, W.; Cross, T. A. *Science* **1993**, *261*, 1457-1460.

(25) Nicholson, L. K.; Moll, F.; Mixon, T. E.; LoGrasso, P. V.; Lay, J. C.; Cross, T. A. *Biochemistry* **1987**, *26*, 6621-6626.

(26) Bechinger, B.; Zasloff, M.; Opella, S. J. *Protein Science* **1993**, *2*, 2077-2084.

(27) Stark, R. E.; Jelinski, L. W.; Ruben, D. J.; Torchia, D. A.; Griffin, R. G. *J. Magn. Res.* **1983**, *55*, 266-273.

CHAPTER 6

CONCLUSIONS

Polypeptides have been designed to adopt unique tertiary structures in the solid state. The design was based on existing information on protein chain folding. Polypeptides which would self assemble into lamellae with predetermined thickness (Figure 1.6) were designed. DNA sequences encoding the synthesis of polymers of **1-m**, **6-m**, **7-m** and **8-m** have been synthesized. In the first part of this dissertation, polypeptides with ten repeats of **1-m**, four repeats of **8-m** and three repeats of **7-m** were successfully synthesized in *E. coli*.

-(AlaGly)3GluGly(GlyAla)3GluGly-	**1-m**
-(AlaGly)$_3$ GluGly-	**2-m**
-(AlaGly)$_3$AspGly(GlyAla)$_3$AspGly-	**6-m**
-(AlaGly)$_3$GluGly(AlaGly)$_3$ValGly-	**7-m**
-(AlaGly)$_3$GluGly(AlaGly)$_3$MetGly-	**8-m**

In the rest of this dissertation, an attempt was made to structurally assess the formation of lamellae during crystallization. Polypeptide **1** (containing ten repeats of **1-m**) was characterized in the solid state by x-ray diffraction, FTIR and NMR to study the formation of lamellae of uniform, predetermined thickness.

In the third chapter, the unit cell of **1** was determined and was compared with that of **2** (containing thirty six repeats of **2-m**). A

crystalline forms of the two polymers, the types of turns are limited to ones with odd residues. The polypeptide **1** forms an antiparallel beta sheet structure when stirred in 70% formic acid; electron microscopy shows the morphology of the crystallites to be needle-like. The unit cell deduced from the diffraction patterns is similar to those of silks and silk-like polypeptides. The unit cell which best explains the diffraction data for **1** is monoclinic with a=9.56 Å, b=10.68 Å and c=7.0 Å with the angle between b and c equal to 80°. Qualitatively, the intensities of the reflections in **1** suggest the presence of apolar sheets.

The fourth chapter dealt with the proof of adjacent reentry of the chains by using FTIR. The repetitive polypeptide **1** and a ^{13}C=O enriched analogue also synthesized biologically in *Escherichia coli*, were blended in solution and co-crystallized. FTIR spectra (in the amide I region) were studied as a function of blend composition. The results indicate the folding of the chains to be predominantly in an adjacent reentry fashion. In addition to probing the folding habit, FTIR spectroscopy on crystalline **1** corroborates the results from x-ray diffraction experiments and shows the polypeptide to predominantly adopt the antiparallel beta sheet structure.

The fifth chapter, in which solid state NMR was used to determine dynamics of side chains, suggests that the glutamic acid residues are at the fold surface. Solid state CPMAS NMR experiments were carried out on the amorphous and crystalline forms of **1**. The data suggests that glutamic acids are in identical environments in the crystalline and amorphous samples. The fast relaxation rates of their alpha, beta and gamma carbons

suggest that glutamic acid residues are excluded from the crystalline regions in both the samples. Hence, the data supports the presence of glutamic acid residues at the turn positions.

Static solid state NMR experiments were carried out on oriented mats of crystalline **A-10**. The orientation of the crystallites was studied by the analysis of lineshapes. The lineshape analysis is consistent with the crystallites being an imperfect form of an orientation with the crystallographic *b* axis (Figure 1.6) being perpendicular to the mat and with free rotation around *b*.

BIBLIOGRAPHY

Abe, Y.; Krimm, S. *Biopolymers* **1972**, *11*, 1817-1839.

Akerfeldt, K. S.; Kim, R. M.; Camac, D.; Groves, J. T.; Lear, J. D.; DeGrado, W. F. *JACS* **1992**, *114*, 9656-9657.

Alexander, L. E. *X-Ray Diffraction Methods in Polymer Science*; Robert E. Krieger Publishing Co.: Malabar, Florida, 1969.

Anderson, J. M.; Chen, H. H.; Rippon, W. B.; Walton, A. G. *J. Mol. Biol.* **1972**, *67*, 459-468.

Andersson, S. G. E.; Kurkland, C. G. *Microbiol. Rev.* **1990**, *54 (2)*, 198-210.

Asakura, T.; Suzuki, H.; Watanabe, Y. *Macromolecules* **1983**, *16*, 1024-1026.

Asakura, T.; Watanabe, Y.; Itoh, T. *Macromolecules* **1984**, *17*, 2421-2426.

Asakura, T. *Makromol. Chem., Rapid Commun.* **1986**, *7*, 755-759.

Asakura, T.; Demura, M.; Watanabe, Y.; Sato, K. *J. Polym. Sci. Part B* **1992**, *30*, 693-699.

Atkins, E. D. T.; Keller, A.; Sadler, D. M. *J. Polym. Sci. A-2* **1972**, *10(5)*, 863-875.

Atkins, E. D. T.; Hill, M.; Hong, S. K.; Keller, A.; Organ, S. *Macromolecules* **1992**, *25*, 917-924.

Baer, E.; Hiltner, A.; Keith, H. D. *Science* **1987**, *235*, 1015-1022.

Bandekar, J.; Krimm, S. In *Advances in Protein Chemistry*; C. B. Anfinsen, J. T. Edsall and F. M. Richards, Ed.; Academic Press: Orlando, 1986; Vol. 38; pp 183-364.

Bank, M. I.; Krimm, S. *J. Polym. Sci. A-2* **1969**, *7*, 1785-1809.

Bechinger, B.; Zasloff, M.; Opella, S. J. *Protein Science* **1993**, *2*, 2077-2084.

Billmeyer, F. W. *Textbook of Polymer Science*; John Wiley & Sons: New York, 1984.

Birge, R. *Annu. Rev. Phys. Chem.* **1990**, *41*, 683-733.

Bolton, J. R.; Mataga, N.; McLendon, G., Ed.; in *Electron Transfer in Inorganic, Organic and Biological Systems*; ACS: Washington, 1991; Vol. 228.

Boor, J.; Youngman, E. A. *J. Polym. Sci.* **1964**, *B2*, 262-268.

Bovey, F. A. *Nuclear Magnetic Resonance Spectroscopy*; Academic Press: San Diego, 1988.

Bowie, J. U.; Luthy, R.; Eisenberg, D. *Science* **1991**, *253*, 164-170.

Breslow, R. *Science* **1982**, *218*, 532-537.

Brown, L.; Trotter, I. F. *Trans. Faraday Soc.* **1956**, *52*, 537-548.

Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469-487.

Caravatti, P.; Neunschwander, P.; Ernst, R. R. *Macromolecules* **1985**, *18 (1)*, 119-122.

Cheam, T. C.; Krimm, S. *Chem. Phys. Lett.* **1984**, *107 (6)*, 613-616.

Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *13 (2)*, 222-245.

Chou, P. Y.; Fasman, G. D. *J. Mol. Biol.* **1977**, *115*, 135-175.

Chu, C. W.; Dickinson, L. C.; Chien, J. C. W. *J. Appl. Polym. Sci.* **1990**, *41*, 2311-2325.

Chung, L. A.; Lear, J. D.; DeGrado, W. F. *Biochemistry* **1992**, *31*, 6608-6616.

Closs, G. L.; Miller, J. R. *Science* **1988**, *240*, 440-447.

Colombo, M. G.; Meier, B. H.; Ernst, R. R. *Chem. Phys. Lett.* **1988**, *146*, 189-196.

Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Manuscript in preparation*

Creel, H. S.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1991**, *24*, 1213-1214.

Creel, H. S. *Univ. of Mass., Amherst* **1993**, *Ph. D. Thesis.*

Dawson, J. *Science* **1988**, *240*, 433-439.

Decius, J. C. *J. Chem. Phys.* **1955**, *23 (7)*, 1290-1294.

DeGrado, W. E.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622-628.

Deratani, A.; Darling, G. D.; Horak, D.; Frechet, J. M. J. *Macromolecules* **1987**, *20*, 767-772.

Devlin, T. M. *Textbook of Biochemistry with Clinical Correlations;* 3rd ed.; Wiley-Liss: New York, 1992.

Dickinson, L. C.; Yang, H.; Chu, C. W.; Stein, R. S.; Chien, J. C. W. *Macromolecules* **1987**, *20*, 1757-1760.

Diesenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *J. Mol. Biol.* **1984**, *180*, 385-398.

Diesenhofer, J.; Michel, H. *Science* **1989**, *245*, 1463-1473.

Dougherty, M. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Manuscript in preparation*

Dougherty, M. J.; Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *Macromolecules* **1993**, *26*, 1779-1781.

Dougherty, M. J. *Univ. of Mass.,Amherst* **1993**, *Ph. D. Thesis*,

Dows, D. A. *J. Chem. Phys.* **1958**, *29 (3)*, 484-489.

Dreyfuss, P. *J. Polym. Sci. (Polym. Phys. Ed.)* **1973**, *11*, 201-216.

Fossey, S. A.; Nemethy, G.; Gibson, K. D.; Scheraga, H. A. *Biopolymers* **1991**, *31*, 1529-1541.

Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C.; Suzuki, E. *J. Mol. Biol.* **1965**, *11*, 706-712.

Frechet, J. M. J.; Darling, G. D.; Itsuno, S.; Lu, P.; Meftahi, M. V.; Rolls, W. A. *Pure Appl. Chem.* **1988**, *60(3)*, 353-364.

Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. *J. Mol. Biol.* **1968**, *32*, 343-358.

Geil, P. H. *J. Polym. Sci.* **1960**, *44*, 449-458.

Goldman, M.; Shen, L. *Phy. Rev.* **1966**, *144*, 321-331.

Gradshteyn, I. S.; Ryzhik, I. M. *Table of Intgrals, Series and Products;* Academic Press: New York, 1980, pp 626-627.

Greig, J. A.; Sherrington, D. C. *Polymer* **1978**, *19*, 163-175.

Gupta, S. C.; Weigh, H. L.; Somerville, R. L. *Biotechnology* **1983**, *9*, 602-608.

Gust, D.; Moore, T. A. *Science* **1989**, *244*, 35-41.

Haeberlen, U. *High Resolution NMR in Solids*; Academic Press: New York, 1976.

Halverson, K. J.; Sucholeiki, I.; Ashburn, T. T.; Lansbury, P. T. *J. Am. Chem. Soc.* **1991**, *113*, 6701-6703.

Harbison, G. S.; Jelinski, L. W.; Stark, R. E.; Torchia, D. A.; Herzfeld, J.; Griffin, R. G. *J. Magn. Reson.* **1984**, *60*, 79-82.

Hexter, R. M. *J. Chem. Phys.* **1960**, *33 (6)*, 1833-1841.

Ho, S. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1987**, *109*, 6751-6758.

Hodge, P.; Sherrington, D. C., Ed.; in *Polymer Supported Reactions in Organic Synthesis*; John Wiley & Sons: New York, 1980.

Hopfield, J. J.; Onuchic, J. N.; Beratan, D. N. *Science* **1988**, *241*, 817-819.

Hornig, D. F. *J. Chem. Phys.* **1948**, *16 (11)*, 1063-1076.

Hrostowski, H. J.; Pimentel, G. C. *J. Chem. Phys.* **1951**, *19*, 661-662.

Ikemura, T. *Mol. Biol. Evol.* **1985**, *2(1)*, 13-34.

Ishida, M.; Asakura, T.; Yokoi, M.; Saito, I. *Macromolecules* **1990**, *23*, 88-94.

Jakes, J.; Krimm, S. *Spectrochim. Acta* **1971**, *27A*, 35-63.

Jakes, J.; Krimm, S. *Spectrochim. Acta* **1971**, *27A*, 19-34.

Joran, A. D.; Leland, B. A.; Felker, P. M.; Zewail, A. H.; Hopfield, J. J.; Dervan, P. B. *Nature* **1987**, *327*, 508-511.

Keith, H. D.; Giannoni, G.; Padden, F. J. *Biopolymers* **1969**, *7*, 775-792.

Keller, A. *Repts. Progr. Phys. Part 2* **1968**, *31*, 623-704.

Kennedy, J. P.; Smith, R. A. *J. Polym. Sci. Polym. Chem. Ed.* **1980**, *18*, 1523-1537.

Kennedy, J. P.; Smith, R. A. *J. Polym. Sci. Polym. Chem. Ed.* **1980**, *18*, 1539-1546.

Ketchem, R. R.; Hu, W.; Cross, T. A. *Science* **1993**, *261*, 1457-1460.

Ketley, A. D., Ed.; in *The Stereochemistry of Macromolecules*; Marcel Dekker: New York, 1967.

Krejchi, M. T.; Atkins, E. D. T.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Science* in press.

Krejchi, M. T.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polymer Preprints* **1991**, *32(1)*, 411-412.

Krejchi, M. T. *Univ. of Mass., Amherst* **1993**, *Ph. D. Thesis*.

Kreyszig, E. *Advances Engineering Mathematics*; John Wiley & Sons, Inc.: New York, 1975.

Krimm, S.; Abe, Y. *Proc. Natl. Acad. Sci. USA* **1972**, *69 (10)*, 2788-2792.

Krimm, S.; Ching, J. H. C. *Macromolecules* **1972**, *5(2)*, 209-211.

Levitt, M. *Biochemistry* **1978**, *17(20)*, 4277-4285.

Lewin, B. *Genes III*; John Wiley and Sons: New York, 1987.

Linder, M.; Henrichs, P. M.; Hewitt, J. M.; Massa, D. J. *J. Chem. Phys.* **1985**, *82 (3)*, 1585-1598.

Lotz, B.; Keith, H. D. *J. Mol. Biol.* **1971**, *61*, 195-200.

Lotz, B.; Keith, H. D. *J. Mol. Biol.* **1971**, *61*, 201-215.

Lotz, B.; Brack, A.; Spach, G. *J. Mol. Biol.* **1974**, *87*, 193-203.

Lotz, B.; Cesari, F. C. *Biochimie* **1979**, *61*, 205-214.

Lyu, P. C.; Liff, M. I.; Marky, L. A.; Kallenbach, N. *Science* **1990**, *250*, 669-673.

Mai, W.; W., H.; C., W.; Cross, T. A. *Protein Science* **1993**, *2*, 532-542.

March, J. *Advanced Organic Chemistry*; John Wiley & Sons: New York, 1986.

Marsh, R. E.; Corey, R. B.; Pauling, L. *Biochim. Biophys. Acta* **1955**, *16*, 1-34.

Mason, J. A.; Sperling, L. H. *Polymer Blends and Composites*; Plenum: New York, 1976.

McBride, L. J.; Caruthers, M. H. *Tetrahedron Lett.* **1983**, *24*, 245.

McGrath, K. P. *Univ. of Mass.,Amherst* **1990**, *Ph. D. Thesis*,

McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727-733.

Merrifield, R. B. *Pure and Appl. Chem.* **1978**, *50*, 643-653.

Miller, J. H. *Experiments in Molecular Genetics*; Cold Spring Harbor Laboratory: New York, 1972.

Miyazawa, T.; Blout, E. R. *J. Am. Chem. Soc.* **1961**, *83*, 712-719.

Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15*, 2465-2483.

Moore, W. H.; Krimm, S. *Biopolymers* **1976**, *15*, 2439-2464.

Moore, J. M.; Case, D. A.; Chazin, W. J.; Gippert, G. P.; Havel, T. F.; Powls, R.; Wright, P. E. *Science* **1988**, *240*, 314-317.

Natta, G.; Pino, P.; Corradini, P.; Danusso, F.; Mantica, E.; Mazzanti, G.; Moraglio, G. *J. Am. Chem. Soc.* **1955**, *77*, 1708-1720.

Natta, G.; Mazzanti, G.; Valvassori, A.; Sartori, G.; Fiumani, D. *J. Poly. Sci.* **1960**, *51*, 411-427.

Natta, G.; Dall'Asta, G.; Mazzanti, G.; Pasquon, I.; Valvassori, A.; Zambelli, A. *J. Am. Chem. Soc.* **1964**, *83*, 3343-3345.

Nicholson, L. K.; Moll, F.; Mixon, T. E.; LoGrasso, P. V.; Lay, J. C.; Cross, T. A. *Biochemistry* **1987**, *26*, 6621-6626.

O'Neil, K. T.; Wolfe, H. R.; Ericson-Viitanen, S.; DeGrado, W. F. *Science* **1987**, *236*, 172-174.

Odian, G. *Principles of Polymerization*; John Wiley & Sons: New York, 1981.

Overberger, C. G.; Pierre, T. S.; Vorshheimer, N.; Yaroslavsky, S. *J. Am. Chem. Soc.* **1963**, *85*, 3513.

Padden, F. J.; Keith, H. D.; Giannoni, G. *Biopolymers* **1969**, *7*, 793-804.

Parkhe, A. D.; Krejchi, M. T. K.; Waddon, A. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Manuscript in preparation*

Parkhe, A. D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Polymer Preprints* **1993**, *34 (1)*, 150-151.

Parkhe, A. D.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *Macromolecules* **1993**,

Paul, D. R.; Barlow, J. W. *J. Macromol. Sci.-Rev. Macromol. Chem.* **1980**, *C18(1)*, 109-168.

Pauling, L.; Corey, R. B.; Branson, H. R. *Proc. Natl. Acad. Sci.* **1951**, *37*, 205-211.

Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci.* **1951**, *37*, 251-256.

Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci.* **1951**, *37*, 261-271.

Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci.* **1951**, *37*, 273-281.

Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci.* **1951**, *37*, 729-740.

Poliks, M. D.; Schaefer, J. *Macromolecules* **1990**, *23*, 3426-3431.

Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632-1641.

Raliegh, D. P.; Levitt, M. H.; Griffin, R. G. *Chem. Phys. Lett.* **1988**, *146 (1,2)*, 71-76.

Regan, L.; DeGrado, W. F. *Science* **1988**, *241*, 976-978.

Rosenburg, A. H.; Lade, B. N.; Chui, D.; Lin, S.; Dunn, J. J.; Studier, F. W. *Gene* **1987**, *56*, 125-135.

Saito, H.; Iwanaga, Y.; Tabeta, R.; Narita, M.; Asakura, T. *Chem. Lett. (Chem. Soc. Jap.)* **1983**, 427-430.

Saito, H.; Tabeta, R.; Shoji, A.; Ozaki, T.; Ando, H. *Macromolecules* **1983**, *16*, 1050-1057.

Saito, H.; Tabeta, R.; Asakura, T.; Iwanaga, Y.; Shoji, A.; Ozaki, T.; Ando, I. *Macromolecules* **1984**, *17*, 1405-1412.

Saito, H.; Ishida, M.; Yokoi, M.; Asakura, T. *Macromolecules* **1990**, *23*, 83-88.

Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning -A laboratory Manual;* 2nd ed.; Cold Spring Harbor Laboratory Press: New York, 1989.

Sandeman, I.; Keller, A. *J. Polym. Sci.* **1956**, *19*, 401-435.

Sanger, F.; Niklen, S.; Coulson, A. R. *Proc. Natl. Acad. Sci.* **1977**, *74*, 5463-5467.

Schaefer, J.; Stejskal, E. O.; Buchdahl, R. *Macromolecules* **1977**, *10 (2)*, 384-405.

Schaefer, J.; Stejskal, E. O.; Steger, T. R.; Sefcik, M. D.; McKay, R. A. *Macromolecules* **1980**, *13*, 1121-1126.

Schildnecht, C. E.; Skeist, I., Ed.; in *Polymerization Processes*; John Wiley & Sons: New York, 1977.

Sharp, P. M.; Li, W. H. *Nucleic Acids Res.* **1986**, *14*, 7737-7749.

Shoemaker, K. R.; Peter, S. K.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987**, *326*, 563-567.

Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759-777.

Siegel, R. W. *MRS Bulletin* **1990**, *October*, 60-67.

Smith, D. B.; Johnson, K. S. *Gene* **1988**, *67*, 31-40.

Stark, R. E.; Jelinski, L. W.; Ruben, D. J.; Torchia, D. A.; Griffin, R. G. *J. Magn. Res.* **1983**, *55*, 266-273.

Stejskal, E. O.; Schaefer, J.; Sefcik, M. D.; McKay, R. A. *Macromolecules* **1981**, *14*, 275-279.

Suwelack, D.; Rothwell, W. P.; Waugh, J. S. *J. Chem. Phys.* **1980**, *73(6)*, 2559-2569.

Suzuki, S.; Iwashita, Y.; Shimanouchi, T.; Tsuboi, M. *Biopolymers* **1966**, *4*, 337-350.

Szwarc, M. *Nature* **1956**, *178*, 1168-1167.

Tasumi, M.; Krimm, S. *J. Polym. Sci. A-2* **1968**, *6*, 995-1010.

Thompson, L. K.; McDermott, A. E.; Raap, J.; van der Wielen, C. M.; Lugtenburg, J.; Herzfeld, J.; Griffin, R. G. *Biochemistry* **1992**, *31*, 7931-7938.

Torchia, D. A. *J. Mag. Reson.* **1978**, *30*, 613-616.

Voet, D.; Voet, J. G. *Biochemistry*; John Wiley & Sons: New York, 1990.

Waddon, A. J. *Personal discussions*

Warwicker, J. O. *Acta. Cryst.* **1954**, *7*, 565-573.

Warwicker, J. O. *Trans. Faraday Soc.* **1956**, *52*, 554-557.

Warwicker, J. O. *J. Mol. Biol.* **1960**, *2*, 350-362.

Wilmot, C. W.; Thornton, J. M. *J. Mol. Biol.* **1988**, *203*, 221-232.

Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311-333.

Zhang, G. *Personal discussion*