Nuclear Magnetic Resonance Structural Studies of Allergen Proteins

by
Gregory L. Warren
B.S., Chemistry
Walla Walla College
(1986)

Submitted to the Department of Chemistry in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

September 1993

Signature of Author_____
Department of Chemistry
September 10, 1993

Certified by_____
Professor Robert G. Griffin
Thesis Supervisor

Accepted by_____
Glenn A. Berchtold
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee appointed by the Department of Chemistry as follows:

Professor Peter T. Lansbury, Jr. _____
Chairman

Professor Robert G. Griffin _____
Thesis Supervisor

Professor Gregory A. Petsko _____
Thesis Supervisor

Dr. Christopher J. Turner _____

Nuclear Magnetic Resonance Structural

Studies of Allergen Proteins

by

Gregory L. Warren

Submitted to the Department of Chemistry,
on September 10, 1993 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in Chemistry

Abstract:

This thesis presents a collection of studies using statistical analysis, solid-state NMR and solution-state NMR for the elucidation of structure-function relationships, dynamics, and macromolecular structure of proteins.

The first chapter contains a brief introduction to a number of biophysical techniques used to study structure-function relationships, dynamics and determine three dimensional structure.

Chapter 2 presents a statistical comparison of the amino acid composition in secondary structure of proteins stable at high temperatures with those which are not. This study has shown that the temperature dependent Zimm-Bragg helix propagation value **s** is not a good predictor for the amino acids found in thermostable α-helix. We have shown that residues tyrosine and glycine show a significant increase in α-helix residency in thermostable proteins over their non-thermostable counterparts.

In chapter 3 we have shown that chemical shift, in particular, phosphorus anisotropic chemical shift tensors are a good probe of structure and dynamics in the hydrolysis reaction of bovine pancreatic ribonuclease A. We have demonstrated that the inhibitor 2'-cytidine monophosphate changes ionization state upon binding at pH 5 and that the anisotropic chemical shift tensor σ_{11} is a sensitive probe of structural interactions at the binding site, while σ_{22} is an excellent indicator of the ionization state of phosphorus.

In chapter 4 an internuclear distance measurement is presented using solid-state NMR and R^2-SELTICS, a combination of the homonuclear dipolar coupling with sideband suppression. In this study we were able to measure an internuclear distance for two phosphorus atoms in glutamine synthetase of 2.7 ±0.2 Å. This distance was indicative of a P-O-P ester bond and clearly demonstrates, *in situ*, that the inhibitor phosphinothriacin (PPT) is phosphorylated during inactivation.

Finally, 3D ^1H solution-state NMR was used to determine the structure of a protein allergen isolated from *Ambrosia trifida* (*Amb. t.* V), a small (4,300 MW) disulfide rich protein. The four disulfide bond pairs have been unambiguously identified through Cys CβH NOEs. The structure consists of a triple stranded antiparallel β-sheet connected by

a type II β-turn and an unstructured loop. The β-sheet is held together by two disulfide bonds. A C-terminal nine residue α-helix lies over the top of the β-sheet and is held in place by two disulfide bonds. The backbone RMS deviation for this structure is 0.92 Å.

Thesis supervisor : Dr. Robert G. Griffin
Title : Professor of Chemistry

Thesis supervisor : Dr. Gregory A. Petsko
Title : Lucille P. Markey Professor of Biochemistry and Chemistry

Acknowledgments:

I would like to take this opportunity to acknowledge the people who have aided and abetted much of the work presented in this thesis. I would like to extend a general thanks to the current members of the Griffin lab; Leno, Dennis, John, Joanna, Jeff, Andrew, Janet, Phil, Doug, Bogin, Jingui, and Brendan. There are former members of this group which should also be included in particular Ling Zheng, Valerie Copié, Lynemarie Thompson, and Michele Auger. I would like to extend a special thanks for Ken Fishbein, because without his help most of the figures presented here and some of the experiments would have difficult, if not impossible, to do. Ajay Thakkar and Peter Allen deserve adulation for providing competent technical support and friendship in and outside of the Thirsty Ear. Thank you Annie for you help in times of need and for interesting discussions about the world outside this laboratory.

Ann McDermott played a pivotal role in my introduction to solid-state NMR and in the development of the glutamine synthetase project. Gary Gerfen deserves a special thanks for not only being a great scientist by helping me solve the glutamine synthetase problem, but for also being a good friend, confidant, and hostel manager over the last few months. There is not enough room in this section to list reasons why Lakshmi deserves a special acknowledgment. Other than a simple thank you I will not embarrass you further Lakshmi.

I would like to acknowledge and thank Kebede Beshah for his help and guidance during the early days of my graduate career. I would also like to thank Lawrence Goodfriend for generously providing samples of the ragweed allergens. Without Dave Ruben's hardware, software, and facile reminders how stupid I can sometimes be 90% of this work could not have been done. Thank you Dave and could we get this new 3-D artifact suppression technique I just....

In the acknowledgment section, when an individual states that the work presented could not have been done without the help of another individual, more often than not they are being polite. In this case I am not being polite when I say that the solution-state NMR work could not have been completed without the help and guidance of Dr. Christopher Turner. I owe most of the success of Chapter 5 to Chris. I hope you will be kind and gentle with me when the time comes to pay.

Last but not least, I would like to extend thanks to the people who provided financial support over the last six years. This includes Professor Gregory A. Petsko, who attempted to introduce me to X-ray crystallography unfortunately without success, and posthumously Dr. Leo Neuringer. This work could not have been completed without the support of Professor Robert G. Griffin.

TABLE OF CONTENTS Page

Page

LIST OF FIGURES

Page

CHAPTER 3

CHAPTER 4

CHAPTER 5

Page

Page

LIST OF TABLES

Page

CHAPTER 2

CHAPTER 3

CHAPTER 5

CHAPTER 1

Introduction to structural techniques for the study of allergen proteins

The study of the relationship between protein structure and function began almost as soon as the first proteins were isolated and purified. Advances in this field have been marked by three scientific events or epochs. The first was the discovery that proteins are biopolymers, whose component parts are twenty amino ethanoic acids, linked linearly via amide bonds. The elucidation of how cellular nuclear material determines the order and composition of this polymer chain was a part of this first event. The second event was and continues to be the use of physical means to study the structural and dynamic properties of proteins. There are a large number of physical techniques which have been adapted, over the last 50 years, to study or investigate protein structure and dynamics. The last event was the development of models and computer algorithms to simulate dynamics and calculate structure at atomic resolution.

The techniques used for the study of the structural properties of biopolymers can be divided into five categories which are: 1) Absorption spectroscopy, 2) other optical techniques, 3) magnetic resonance, 4) diffraction techniques, and 5) conformational analysis. We will briefly describe how these techniques can be used to study protein structure. A more detailed discussion of these techniques and their application to biological molecules can be found in the first two volumes of the three volume series by Cantor and Schimmel [1, 2].

Absorption spectroscopy is done by passing an energy wave through a substance and measuring the amount of energy absorbed by the substance. For biological molecules the region of the energy spectrum used goes from the infrared (IR) bandwidth to X-rays.

Information about the vibrational modes of bonds can be measured using IR or Raman spectroscopy [3, 4]. These techniques have been used, in conjunction with the Fourier transform (FTIR) counterpart, to measure molecular changes in enzymatic pathways. When combined with stop-flow techniques, it is possible to measure these spectra rapidly and thus gain an understanding of the molecular changes occurring with respect to time. Another bandwidth used for biological molecules is the UV bandwidth. From UV we can obtain information about the electronic states of the macromolecule. In proteins, the far UV region is dominated by amide bond absorption while in the near UV region contains information about aromatic sidechains. Again this technique has proved most useful when combined with stop-flow techniques for studying changes in the electronic state during enzymatic reactions. The last absorption technique used in biological macromolecules is a X-ray absorption technique called X-ray absorption fine structure (EXAFS). This technique uses a tunable high-flux X-ray source generally from a synchrotron and measures the X-ray absorption. This technique has proven useful for analyzing protein structure adjacent to metal atoms in metalloproteins [5].

Two optical techniques used in the study protein structure that do not measure absorption are circular dichroism (CD) and fluorescence spectroscopy. CD can be used to measure the amount of secondary structure (α-helix and β-sheet) present in a protein. This technique has been used extensively as a monitor of protein structure during protein folding experiments [6]. Fluorescence occurs when a chemical species absorbs energy at a particular wave length and then relaxes back to the ground state by emitting radiation. In proteins the residues phenylalanine, tyrosine, and tryptophan all absorb light between 250 and 280 nm and fluoresce between 280 and 350 nm. Fluorescence in more sensitive to the environment of the chromophore than absorption is. This property has been capitalized for the study of protein structure by using both fluorescent groups present in the molecule and external probes [2].

Nuclear and electron magnetic resonance can been used to study both the local and global aspects of protein structure and dynamics. With the development of multi-dimensional homo- and heteronuclear techniques and isotopically enriched proteins, solution state NMR has been used to determine protein structure at atomic resolution for proteins up to 30,000 MW [7-10]. Also important in structure determination by solution state NMR was the development of computer algorithms which calculate a three dimensional structure from the distance constraints generated by employing NMR dipolar coupling techniques [11].

Solution NMR is not the only magnetic resonance technique which can be used to obtain structure information about biological macromolecules. Solid state NMR, in conjunction with resolution enhancing magic angle spinning [12], can be used to study insoluble or membrane bound proteins. There are many techniques in the literature which use heteronuclear (DIPSHIFT, REDOR, TEDOR, FDR, etc.) and homonuclear (R^2, SEDRA, RFDR, etc.) dipolar couplings to obtain nuclear distance information. We will describe only two of the techniques developed over the last six years, Rotational Resonance (R^2) and Rotational-echo, double-resonance (REDOR) because, so far, these are the only methods which have been used in biological macromolecular systems to measure internuclear distances.

A description of R^2, a technique to selectively reintroduce homonuclear dipolar coupling, is found in the introduction to Chapter 4. REDOR is an experiment which measures heteronuclear spin pair distances by using rotor synchronized π pulses on one nucleus and measuring the dephasing effect the dipolar coupling as on the other nucleus (e.g. ^{13}C-^{15}N, etc.). For an isolated spin pair the dephasing is related to the distance between the spin pair by $1/r^3$. This heteronuclear technique is not chemical shift tensor dependent and does not require the heteronuclear coupling to be resolved within the sweep width of the spectrum [13-17]. This method has been used to measure nuclear spin pair

distances in an enzyme-inhibitor complex [18]. In practice, for ^{13}C-^{15}N spin pairs, distance measurements are limited to less than 6 Å.

There are a number of diffraction techniques which have been used to determine atomic resolution protein structures. In this case the protein structure is determined by measuring the diffraction pattern of waves passed through a crystalline lattice. The wave source can be X-rays or an electron or neutron beam. Electron diffraction has been used to determine the structure of proteins found in a 2-dimensional lattice such a lipid bilayer. The resolution of this technique is about 3 Å. X-ray diffraction is able to resolve heavy atoms which in the case of proteins is carbon, nitrogen and oxygen. Hydrogen can not be observed using X-ray diffraction. X-ray crystal structures of proteins have been obtained to 0.9 Å resolution. Last, in neutron diffraction hydrogen has a scattering length comparable to heavy atoms thus in this technique hydrogen position can be determined. The principle disadvantage of neutron diffraction is that available neutron fluxes are low requiring long exposure time and large protein crystals for the acquisition of scattering intensities sufficient for analysis. All the diffraction techniques, while providing detailed atomic resolution structural information, require an ordered protein crystal and require, in the case of new unrelated protein structures, heavy atom derivatives for the solution of the phase problem before the structure can be determined.

The last technique for studying protein structure and function is conformational analysis. This technique uses the results from structural techniques and attempts to organize this information through the use of statistics so as to understand how protein structure relates to and effects function [19]. A detailed description of this technique is found in Chapter 2.

The dynamics occurring in biological macromolecules can be studied using a number of different techniques. IR and resonance Raman spectroscopy are useful in studying dynamics as well as structure because these techniques measure vibrational modes in proteins which are the operational modes for protein dynamics. Magnetic resonance is

another technique which has been used to study dynamics in proteins. In solution NMR T_1 relaxation of ^{13}C and ^{15}N is a good probe for measuring the dynamics of the backbone of proteins. In solid state NMR, it is possible to study the dynamics of specific functional groups using quadrupolar nuclei such as deuterium (^2H). The size of the ^2H quadrupole echo interaction matches the time scale of dynamics in solids. Because the quadrupole coupling constant is $e^2qQ/\hbar \approx 2\times10^5$ Hz for C-D bonds, motions of CD_3 groups and two fold flips of aromatic rings, on the order of 10^{-3} to 10^{-8} sec [20-22], effect the line shape of ^2H spectra. Dynamic rates from 10^{-7} to 10^{-10} sec can be measured using ^2H spin lattice relaxation anisotropy [23, 24], while rates slower than 10^{-3} can be measured using 2-D chemical exchange experiments [25-33]. The T_1 inversion recovery and 2-D exchange experiments can be applied to spin 1/2 nuclei as well [34-36].

Protein dynamics can be modeled and simulated using computer algorithms [37, 38]. While it is not completely clear how this work will benefit the understanding of protein dynamics in the future, work on the dynamics of the light driven proton pump in bacteriorhodopsin and of the dynamics of the triose phosphate isomerase loop during the reaction cycle have provided interesting insights into how dynamics effect the function of these proteins and indicate these techniques hold a great potential for aiding in understanding how dynamics effects structure and function.

The work described in this thesis involves only two of the techniques described so far; conformational analysis and magnetic resonance to determine protein structure, study dynamics, and to gain some understanding of how structure relates to function. Conformational analysis was used in Chapter 2 in an effort to understand the mechanisms involved in increasing proteins stability at high temperatures. Solid state NMR was used in Chapter 3 to studying the dynamics and structural changes involved during hydrolysis of ribonucleic acid by ribonuclease A. Chapter 4 describes an experiment in which solid state NMR was used to determine the structure of an intermediate in the reaction pathway of

glutamine synthetase. Last, a three dimensional protein structure was determined using solution NMR of a protein allergen from ragweed pollen.

References:

(1) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; W.H. Freeman: New York, 1980; pp 341 .

(2) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry*; W.H. Freeman: New York, 1980; pp 846 .

(3) Yu, N. T. *Crit. Rev. Biochem.* **1977**, *4,* 229-280.

(4) Spiro, T. G.; Gaber, B. P. *Ann. Rev. Biochem.* **1977**, *46,* 553-572.

(5) Schulman, R. G.; Eisenberger, P.; Kincaid, B. M. *Ann. Rev. Biophys. Bioeng.* **1978**, *7,* 559.

(6) Creighton, T. E. *Proteins: Structure and Molecular Properties*; W.H. Freeman: New York, 1993; pp 507 .

(7) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986; pp 292 .

(8) Clore, G. M.; Gronenborn, A. M. *CRC Crit. Rev. Biochem. Mol. Biol.* **1989**, *24,* 479-564.

(9) Fesik, S. W.; Zuiderweg, E. R. P. *Quart. Rev. Biophys.* **1990**, *23,* 97-131.

(10) Clore, G. M.; Gronenborn, A. M. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20,* 29-63.

(11) Brünger, A. T.; Nilges, M. *Quart. Rev. Biophys.* **1993**, *26,* 49-125.

(12) Griffin, R. G.; Aue, W. P.; Haberkorn, R. A.; Harbison, G. S.; Herzfeld, J.; Menger, E. M.; Munowitz, M. G.; Olejniczak, E. T.; Raleigh, D. P.; Roberts, J. E.; Ruben, D. J.; Schmidt, A.; Smith, S. O.; Vega, S. in *Magic-Angle Sample Spinning*; C. Corso, Bologna, 1988; pp 203-266.

(13) Gullion, T.; Poliks, M. D.; Schaefer, J. *J. Magn. Reson.* **1988**, *80,* 553-558.

(14) Gullion, T.; Schaefer, J. *J. Magn. Reson.* **1989**, *81,* 196-200.

(15) Bork, V.; Gullion, T.; Hing, A.; Schaefer, J. *J. Magn. Reson.* **1990**, *88,* 523-532.

(16) Holl, S. M.; McKay, R. A.; Gullion, T.; Schaefer, J. *J. Magn. Reson.* **1990**, *89,* 620-626.

(17) Pan, Y.; Gullion, T.; Schaefer, J. *J. Magn. Reson.* **1990**, *90,* 330-340.

(18) Christensen, A. M.; Schaefer, J. *Biochemistry* **1993**, *32,* 2868-2873.

(19) Chou, P. Y.; Fasman, G. D. *Ann. Rev. Biochem.* **1978**, *47,* 251-276.

(20) Beshah, K.; Olejniczak, E. T.; Griffin, R. G. *J. Chem. Phys.* **1987**, *86,* 4730-4736.

(21) Rice, D. M.; Meinwald, Y. C.; Scheraga, H. A.; Griffin, R. G. *J. Am. Chem. Soc.* **1987**, *109,* 1636-1640.

(22) Beshah, K.; Griffin, R. G. *J. Magn. Reson.* **1989**, *84,* 268-274.

(23) Siminovitch, D. J.; Ruocco, M. J.; Olejniczak, E. T.; Das Gupta, S. K.; Griffin, R. G. *Biophys. J.* **1988**, *54,* 373-381.

(24) Griffin, R. G.; Beshah, K.; Ebelhahser; Huang, T. H.; Olejniczak, E. T.; Rice, D. M.; Siminovitch, D. J.; Wittebort, R. J. in *Deuterium NMR studies of dynamics in solids*; Long, G. J. Grandjean, F.; Kluwer Academic Publishers, 1988; pp 81-105.

(25) Szeverenyi, N. M.; Sullivan, M. J.; Mciel, G. E. *J. Magn. Reson.* **1982**, *47,* 462-475.

(26) Edzes, H. T.; Bernards, J. P. C. *J. Am. Chem. Soc.* **1984**, *106,* 1515-1517.

(27) Clayden, N. J. *J. Magn. Reson.* **1986**, *68,* 360-362.

(28) Schmidt, A.; Smith, S. O.; Raleigh, D. P.; Roberts, J. E.; Griffin, R. G.; Vega, S. *J. Chem. Phys.* **1986**, *85,* 4248-4253.

(29) Schmidt, C.; Blümich, B.; Spiess, H. W. *J. Magn. Reson.* **1988**, *79,* 269-290.

(30) Voelkel, R. *Angew. Chem. Int. Ed. Engl.* **1988**, *27,* 1468-1483.

(31) Wefing, S.; Spiess, H. W. *J. Chem. Phys.* **1988**, *89,* 1219-1233.

(32) Wefing, S.; Kaufmann, S.; Spiess, H. W. *J. Chem. Phys.* **1988**, *89,* 1234-1244.

(33) Blümich, B.; Spiess, H. W. *Angew. Chem. Int. Ed. Engl.* **1988**, *27,* 1655-1672.

(34) Cullis, P. R.; De Kruijff, B. *Biochim. Biophys. Acta* **1978**, *513,* 31-42.

(35) Blume, A.; Rice, D. M.; Wittebort, R. J.; Griffin, R. G. *Biochemistry* **1982**, *21,* 6220-6230.

(36) Blume, A.; Griffin, R. G. *Biochemistry* **1982**, *21,* 6230-6242.

(37) van Gunsteren, W. F.; Karplus, M. *Nature* **1981**, *293,* 677-678.

(38) Karplus, M.; McCammon, J. A. *Crit. Rev. Biochem.* **1981**, *9,* 293-349.

CHAPTER 2

Conformation analysis of α-helix in thermophilic organisms

Introduction

With the solution of the three dimensional structure of myoglobin [1] by X-ray crystallography came the realization that the amino acid sequence in some way determines protein structure. This realization spawned many attempts at prediction of protein structures based on the amino acid sequences. Structure prediction, however, turned out to be an extremely complex problem. Solutions were sought by looking for a less complex substructure with the hope that prediction of this substructure would be a more amenable problem.

It was observed that the myoglobin structure contained regions of what appeared to be an intermediate or secondary structure along regions of irregular structure. Early studies noted that certain amino acids show a tendency towards this secondary structure or α-helix [2]. Attention focused on finding methods by which α-helix segments could be predicted based on the amino acid sequence. Chou and Fasman [3,4] performed statistical analysis on the tendency of amino acids for a specific secondary structure based on X-ray structures of proteins, and developed a method by which secondary structure was predicted from the primary sequence. Levitt [5] extended the data base from 20 to 66 proteins. However, the accuracy of the prediction methods tended to peak at about 60-70%, with the boundaries of the secondary structure elements most often wrong.

It was recognized that other factors, such as long range effects which were not easily accounted for, had a large effect on secondary structure formation. However, for a time

after the publication of the Chou and Fasman [3,4] secondary structural analysis method, it was thought that secondary structure could more or less be predicted, and from that tertiary structure could be predicted by packing the secondary structure so as to optimize various energy parameters.

Presently, the field of secondary structure prediction has resigned itself to optimizing its predictive capabilities to about 80% through the addition of long range electrostatic and energy parameters, and through algorithm optimizations [6,7]. A few amino acid conformational and composition studies have been done recently but they have not dealt with protein secondary structure [8,9].

Scheraga [10] published a list of experimentally measured helix-coil stability constants for the amino acids in randomly polymerized host-guest copolymers. This study generated Zimm-Bragg [11] **s** value, a measure of helix propagation propensity, and σ value, a measure of helix nucleation tendency, values over a temperature range of 0-70° C. The information of interest to this study was the change in helix stability for various amino acids at high temperatures. Because the Zimm-Bragg **s** value is temperature dependent [12], this leads to the prediction that a thermally stable protein should show a shift in the amino acid composition of its α-helices.

A literature search was done to determine if the amino acid composition of α-helix in thermostable proteins had been performed. No literature was found on the amino acid composition of proteins which are catalytically stable at over 60° C.

The present study of ten thermostable proteins was undertaken to determine the conformational preference of amino acids in thermophilic organisms (thermophilic defined as having an optimal growth temperature above 60° C). The data generated were compared to the conformation frequency parameters for mesophilic organisms (mesophilic defined as having an optimal growth temperature below 45° C) and used to determine if Scheraga's [10] data accurately predicts the shifts observed in the helix residency. Analysis

of the change in the frequency parameters, along with a few postulates for the frequency changes, are shown later in this chapter.

Experimental Procedures

A survey of the α-helix amino acid composition of proteins from thermophiles was made based on ten proteins whose amino acid sequence and, for eight of the proteins, X-ray crystal structures are known: thermolysin [13], tyrosyl-tRNA synthetase [14,15], holo-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* [16] and *Thermus acquaticus* [17], manganese superoxide dismutase [18], thermitase [19], rubredoxin [20,21], phosphoglyceral kinase from *Bacillus stearothermophilus* [22] and *Thermus thermophilus* [23,24], and L30 ribosomal protein [25]. The regions of α-helix were determined on the basis of their main-chain dihedral angles and were visually verified.

Counting Statistics. Whether or not a particular amino acid, *i*, shows a statistical preference for α-helix, *j*, depends on the natural abundance of the amino acid in the protein, N_i, and the number of times the amino acid is found in α-helix, n_{ij}. This is the frequency for *i* in *j*, or f_{ij}. Levitt [5] found that if *i* amino acids randomly occur in *j* α-helix, then the probability of finding *i* amino acid in *j* structure could be approximated by the binomial distribution function

$$P(n_{ij}) = \left[\frac{N_i!}{n_{ij}!(N_i - n_{ij})!}\right] f_{ij}^{n} (1 - f_{ij})^{N_i - n_{ij}} \quad \textbf{(2-1)}$$

Because n_{ij} and N_i are determined by counting, it is the value of f_{ij} which needs to be determined. By assuming that all values of fij were equally likely, Levitt [5] normalized equation 2-1 so that $\int_0^1 P(f_{ij})df = 1$ which yields:

$$P(f_{ij}) = \left[\frac{(N_i + 1)!}{n_{ij}(N_i - n_{ij})!}\right] f_{ij}^{n} (1 - f_{ij})^{N_i - n_{ij}} \quad \textbf{(2-2)}$$

Lindley [26] showed that the above distribution (equation 2-2) closely approximates a Gaussian distribution for a new variable $x = \ln[f_{ij}/(1 - f_{ij})]$. Since the most likely value for

$P(f_{ij})$ will be $f_{ij} = n_{ij}/N_i$, the mean for the above variable is $\bar{x} = \ln[n_{ij}/(N_i - n_{ij})]$ and the variance $\sigma_x^2 = [1/(n_{ij}+1)] + [1/(N_i - n_{ij}+1)]$. Maxfield and Scheraga [27] have used this approximation, with a high degree of accuracy, to determine frequency values.

The f_{ij} calculated is the particular preference an amino acid *i* has for α-helical secondary structure. The probability of an amino acid which has no preference for α-helix occurring randomly in that structure is $f_j = N_j/N_T$, where N_j is the number of residues found in α-helix and N_T is the total number of residues in the statistical sample. If f_{ij} is 10% greater than f_j, or if f_{ij} is 10% less than f_j then the particular amino acid is considered helix favoring or helix breaking, respectively. If f_{ij} is greater than 0.9 f_j but less than 1.1 f_j then the *i* amino acid is considered helix indifferent.

Integrating the probability function, $P(f_{ij})$, is a method for determining the certainty of the previous preference determination. The integrals are defined as follows: $I_{0.0}^{0.9}$ is f_{ij} integrated from 0 to 0.9 f_j, $I_{0.9}^{1.1}$ is f_{ij} integrated from 0.9 f_j to 1.1 f_j, and $I_{1.1}^{\infty}$ is f_{ij} integrated from 1.1 f_j to 1.0. These probability integrals demonstrate the probability that the amino acid breaks, is indifferent to, or forms α-helix.

Because it is not possible to compare the f_{ij}'s of different data sets due to the effects of the natural abundance of *j* in a structure on f_{ij}, Chou and Fasman [3] normalized the function f_{ij} to yield $P_{ij} = f_{ij}/f_j$. The standard deviation, for the normalized function P_{ij}, would be $\sigma_{P_{ij}} = \sqrt{f_{ij}(1-f_{ij})/N_i}(N_T/N_j)$. The normalized values correspond to the previous classification scheme: $P_{ij} < 0.9$ is α-helix breaking, for $0.9 < P_{ij} < 1.1$ the amino acid is indifferent, and if $P_{ij} > 1.1$ the amino acid favors α-helix.

Population Difference Test. The Fisher-Irwin population proportion test can be used to determine whether there was a statistical difference in the P_{ij} value between proteins from thermophilic organisms and proteins from mesophilic organisms. The P_{ij} values published by Levitt [5] were used for the mesophile population proportion values. Because the n value (the sum of the number of *i* amino acids from both populations) was large, i.e.

greater than 90, the exact Fisher-Irwin test values were not calculated. Instead a large sample number approximation was used [28].

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_i} + \frac{\hat{p}_2\hat{q}_2}{N'_i}}} \quad \textbf{(2-3)}$$

Where $\hat{p}$ is the frequency f_{ij} and $\hat{q} = 1 - \hat{p}$. There is a problem with this test, in that the frequency values f_{ij} between different data sets and between different amino acids are not comparable. Consequently a normalized test statistic shown below must be used.

$$Z = \frac{\hat{p}_a - \hat{p}_b}{\left(\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_i} + \frac{\hat{p}_2\hat{q}_2}{N'_i}}\right) \Big/ \left[(N_j + N'_j)/(N_T + N'_T)\right]} \quad \textbf{(2-4)}$$

The variables $\hat{p}_a$ and $\hat{p}_b$ correspond to the P_{ij} for each population. The $\hat{p}_1, \hat{p}_2$ values are f_{ij} from the populations being compared and $\hat{q} = 1 - \hat{p}$. The P_{ij} value difference, between the populations, was considered significant $\left(\mathrm{H_0}: \hat{p}_1 - \hat{p}_2 = 0, \mathrm{H_a}: |\hat{p}_1 - \hat{p}_2| > 0\right)$ if the test variable Z was greater than 1.25, which corresponds to a 90% confidence level that the populations are different.

Results

Thermophile Amino Acid Shift Predictions. The data from host-guest copolymer stability constants [10] can be used to predict a shift in the α-helix forming amino acid population for proteins from mesophilic to thermophilic organisms. These data are shown in Table 2-1. Using the **s** values we would predict that the residues favoring α-helix would be leucine, glutamic acid, valine and isoleucine. From the Δs values the residues, valine, threonine, asparagine, glycine, and proline, are expected to show a slight increase in α-helix residency over the mesophilic counterparts. All other residues would be predicted to show a decrease in α-helix residency in thermophilic organisms with glutamic acid, methionine, glutamine, tyrosine, and histidine predicted to be much less stable in α-helices at high temperatures.

TABLE 2-1: α-Helix Forming Tendency From Zimm-Bragg **s** Values

Amino Acids[a] In Order of α-Helix Forming Tendency at 60° C	s[b] at 60° C	$\sigma \times 10^4$	Δs 20 to 60° C
L	1.09	33	-0.05
E	1.07	100	-0.28
V	1.06	1.0	0.11
I	1.01	55	-0.13
A	1.00	8.0	-0.06
F	1.00	18	-0.09
M	1.00	54	-0.20
R^+	0.94	0.1	-0.09
W	0.92	77	-0.19
E^-	0.93	6.0	-0.04
C	0.91	1.0	-0.08
K^+	0.91	1.0	-0.03
T	0.86	0.1	0.04
N	0.85	210	0.07
Q	0.78	33	-0.20
S	0.73	0.1	-0.03
Y	0.65	66	-0.37
G	0.63	0.1	0.04
D	0.63	210	-0.15
H^+	0.59	0.1	-0.10
H	0.55	210	-0.30
D^-	0.53	70	-0.15
P	0.25	0.1	0.06

[a] Amino acids are listed in the standard single letter code.

[b] **s** > 1.0 is helix forming, **s** < 1.0 is helix breaking [10,29-48].

A comparison of mesophile α-helix probability and stability values is shown in Table 2-2. From these data the residues glutamic acid, alanine, leucine, methionine, glutamine, lysine, arginine show a statistical tendency for α-helix. The residues phenylalanine, isoleucine, and tryptophan are borderline residues, meaning that by some scales they show helix tendencies, but in others are helix indifferent.

α-Helix Probability Values. Normalized probability values were calculated using the formula

$$P_{ij} = \frac{(n_{ij}/N_i)}{(N_j/N_T)} \quad \textbf{(2-5)}$$

P_{ij} and $\sigma_{P_{ij}}$ values are found in Table 2-3. There are obvious differences between the mesophilic data and the thermophilic data. First, lysine and methionine shift from being strongly helix favoring (H) to helix indifferent (I) or H/I in the case of lysine. Isoleucine moves from H/I to the helix indifferent helix breaking (I/B) border. Aspartic acid, asparagine, and valine more from the I to B category while tyrosine moves from strongly B to I.

The change observed in cysteine from B to I/B is probably not significant, as a quick comparison between Table 2-2 and Table 2-3 shows that the mesophile probability value lies within one standard deviation of thermophile value. The significance of the change observed for cysteine can be tested by integrating the probability function.

The probability values, $I_{0.0}^{0.9}$, $I_{0.9}^{1.1}$, and $I_{1.1}^{\infty}$ were calculated and are shown in Table 2-4. The predictions from the probability function integration are slightly different than the predictions from Table 2-3. Phenylalanine, which is helix forming based on P_{ij} is actually a borderline H/I residue based on the integration values. The same change in structural code classification occurs for histidine and cysteine. As is seen in Table 2-4 these residues are now defined as I/B and B/I instead of I as determined based on P_{ij} alone.

TABLE 2-2: Mesophile α-Helix stability and probability values

Amino Acids	Williams[49]	Levitt[5]	Richardson[50]	O'Neil[51]	Scheraga[10]
E	1.59	1.44	0.80	-0.27	0.97
A	1.41	1.29	1.80	-0.77	1.07
L	1.34	1.30	1.20	-0.62	1.14
M	1.30	1.47	1.50	-0.50	1.20
Q	1.27	1.27	1.30	-0.33	0.98
K	1.23	1.23	1.10	-0.65	0.94
R	1.21	0.96	1.30	-0.68	1.03
F	1.16	1.07	1.30	-0.41	1.09
I	1.09	0.97	1.20	-0.23	1.14
H	1.05	1.22	1.00	-0.06	0.69
W	1.02	0.99	1.50	-0.45	1.11
D	0.99	1.04	1.00	-0.15	0.68
V	0.98	0.91	1.20	-0.14	0.95
N	0.76	0.90	0.90	-0.07	0.78
T	0.76	0.82	1.00	-0.11	0.82
Y	0.74	0.72	0.80	-0.17	1.02
C	0.66	1.11	0.70	-0.23	0.99
S	0.57	0.82	0.60	-0.35	0.76
G	0.43	0.56	0.50	0.0	0.59
P	0.34	0.52	0.30	≈3	0.19

TABLE 2-3: Statistical Preference for α-Helix in Thermophiles

Amino acid	n_{ij}	N_i	P_{ij} [a]	$\sigma_{P_{ij}}$	$P_{ij} - \sigma_{P_{ij}}$	$P_{ij} + \sigma_{P_{ij}}$
E	106	177	1.50	0.09	1.41	1.59
A	164	299	1.38	0.07	1.31	1.45
Q	31	61	1.28	0.16	1.12	1.44
L	120	239	1.26	0.08	1.18	1.34
R	62	132	1.18	0.11	1.07	1.29
F	38	82	1.16	0.14	1.02	1.30
K	70	159	1.10	0.10	1.00	1.20
W	13	30	1.09	0.23	0.86	1.32
M	17	43	0.99	0.19	0.80	1.18
Y	37	98	0.95	0.12	0.83	1.07
H	24	64	0.94	0.15	0.79	1.09
C	4	11	0.91	0.36	0.55	1.28
I	49	142	0.87	0.10	0.77	0.97
N	36	113	0.80	0.11	0.69	0.91
T	48	151	0.80	0.10	0.70	0.90
V	71	226	0.79	0.08	0.71	0.87
D	51	165	0.78	0.09	0.69	0.87
G	69	246	0.70	0.07	0.63	0.77
S	36	132	0.68	0.10	0.58	0.78
P	23	113	0.51	0.10	0.41	0.61

[a] N_j =1,069, N_T =2,683

TABLE 2-4: Integration Values for α-Helices of Thermophilic Organisms

Amino acids	P_{ij}	$I_{0.0}^{0.9 f_{ij}}$	$I_{0.9 f_{ij}}^{1.1 f_{ij}}$	$I_{1.1 f_{ij}}^{\infty}$	structural code
E	1.50	4.16E-11	9.16E-06	1.000	H
A	1.38	1.12E-11	6.57E-05	1.000	H
Q	1.28	7.96E-03	0.126	0.866	H
L	1.26	2.77E-06	0.024	0.977	H
R	1.18	4.18E-03	0.226	0.770	H
F	1.16	0.024	0.294	0.682	H,I
K	1.10	0.016	0.459	0.525	H,I
W	1.09	0.185	0.324	0.491	I,H
M	0.99	0.291	0.414	0.296	I
Y	0.95	0.335	0.549	0.116	I
H	0.94	0.375	0.466	0.160	I,B
C	0.91	0.442	0.224	0.334	B,I
I	0.87	0.619	0.369	0.013	B,I
N	0.80	0.803	0.192	5.03E-03	B
T	0.80	0.845	0.153	1.34E-03	B
V	0.79	0.916	0.084	7.25E-05	B
D	0.78	0.904	0.096	3.68E-04	B
G	0.70	0.995	5.07E-03	1.90E-07	B
S	0.68	0.980	0.020	4.92E-05	B
P	0.51	1.000	2.05E-04	9.85E-08	B

TABLE 2-5: Fisher-Irwin Population Proportion Hypothesis Test

Amino acids	n_{ij}	N_i	n'_{ij} [a]	N'_i [a]	Z	H_0 at [b] α=0.05	P H_a true
E	106	177	126	282	0.432	r	0.666
A	164	299	186	464	0.834	r	0.797
Q	31	61	67	171	0.046	r	0.516
L	120	239	152	378	-0.332	r	0.629
R	62	132	50	168	1.336	a	0.910
F	38	82	65	193	0.473	r	0.681
K	70	159	147	385	-0.950	r	0.829
W	13	30	26	84	0.328	r	0.629
M	17	43	39	84	-1.768	a	0.962
Y	37	98	48	98	1.111	r	0.844
H	24	64	50	131	-1.288	a	0.902
C	4	11	42	121	-0.449	r	0.674
I	49	142	90	296	-0.708	r	0.761
N	36	113	68	241	-0.647	r	0.742
T	48	151	90	351	-0.153	r	0.560
V	71	226	124	440	-1.085	r	0.860
D	51	165	107	330	-1.997	a	0.977
G	69	246	90	519	1.437	a	0.925
S	36	132	112	439	-1.082	r	0.860
P	23	113	36	219	-0.075	r	0.532

[a] Values obtained from Levitt [5]

[b] r = H_0 is rejected, a = H_0 is accepted

Population Difference Test. The results from the Fisher-Irwin and hypothesis tests are found in Table 2-5. Because n_{ij} is small, less than 100 for some of the amino acids in the thermophile sample, the applicability of the large sample number approximation had to be tested. If both $\hat{p}_1 n_1$ and $\hat{q}_0 n_1$, where $\hat{q}_0 = \left[1 - \left((n_{ij} + n'_{ij})/(N_i + N'_i)\right)\right]$, are greater than 5 then the large sample approximation is appropriate [52]. All the residues passed the large sample size approximation test (data not shown).

From Table 2-5 arginine and glycine show a statistically significant increase in α-helix residency in thermophilic organisms, while methionine, histidine, and aspartic acid show a significant decrease. Aspartic acid had the largest statistical significance, Z value of 2.0, corresponding to 98% probability that the mesophile and thermophile conformation probability are different. Residues that had at least an 80% probability but less than the 90% statistical significance cutoff are alanine and tyrosine, showing an increase in α-helix residency, and lysine, valine, and serine which were found to have decreased residency in α-helix. Residues having a trend towards difference from the mesophile data are phenylalanine with increasing helix tendency, and cysteine, isoleucine, and asparagine with a smaller residency.

Discussion

Though ten thermostable proteins were used for the statistics of this study, only nine have known structures. For glyceraldehyde-3-phosphate dehydrogenase from *Thermus acquaticus* only the primary sequence is known. Chothia and Lesk [53] have shown that if a protein has >50% sequence homology with another protein, the two proteins will have nearly identical structures. This is especially true if much of the core of the protein is made up of α-helix. There is 62.1% sequence identity and 70% sequence similarity between the glyceraldehyde-3-phosphate dehydrogenases of *Bacillus stearothermophilus* and *Thermus*

acquaticus. The major secondary structure surrounding the active site of glyceraldehyde-3-phosphate dehydrogenase is α-helix. Using the relationship formula $\Delta = 0.40e^{1.87H}$ [53], where Δ is the RMS deviation in Å and H is fraction of mutated residues, a 0.79 Å RMS deviation in the backbone atomic positions between these two proteins would be expected. A sequence alignment of the two proteins thus allowed us to assign the regions we expect to be α-helical.

Comparison of Thermostable Probability and Zimm-Bragg s Values. Since the purpose of this study was to test whether or not the predictions taken from Scheraga's [10,29-48] α-helix stability constants could accurately predict mutations required to stabilize α-helices at high temperatures, it is important to examine the data in more detail, in particular, the **s** values from Table 2-1 and the P_{ij} values found in Table 2-4. Though the Zimm-Bragg **s** values obtained by Scheraga [10,29-48] accurately predicted glutamic acid (E), alanine (A), leucine (L), and phenylalanine (F) as helix forming, the **s** values wrongly predicted valine (V) and isoleucine (I) as helix forming while failing to predict arginine (R) as helix forming. In fact, Scheraga's [44] data predicted that tyrosine would shift from helix forming at low temperatures to helix breaking at high temperatures, instead of the helix indifferent value observed. Though there are minor differences between the **s** values found in Table 2-1 and the probability coefficients in Tables 2-2 and 2-3, e.g. I and R, it is the large differences for V and Y which are of the greatest concern.

The inaccuracy for V is a result of a positive Δ**s** value for this residue. A comparison of the mesophilic probability values with the **s** value (Table 2-2) shows a similarity between the measurements. The probability value for the thermostable proteins studied is significantly less than the mesophilic value. This is supported by the integration values (Table 2-4) showing V having 92% of its probability in the helix breaking category. Thus it is that **s** values predict that V will be more stable in an α-helix at high temperature, while in thermostable proteins the opposite is true. A similar but opposite problem is present between the **s** and probability values for Y. In this case, Y is predicted to be much less

stable (Δs = -0.37) in α-helix at high temperatures. Based on a **s** value of 0.65 we would predict Y to be helix breaking. Instead we find it to be helix indifferent. Explanations for these differences will be discussed latter.

Comparison of Mesophile and Thermophile Probability Values. The mesophile α-helix probability study which contains the largest number of proteins and residues, and is expected to be the most accurate, is the study by Williams et. al. [49] summarized in Table 2-2. Comparison of the thermophilic data with this data set was not possible as the paper does not contain the counting statistics necessary for a Fisher Irwin population proportion test. As a result the comparison found in Table 2-5 uses the data found in Levitt [5] as the mesophile data set. Levitt's [5] probability values were obtained from a set of 36 proteins, while Williams et. al. [49] values were calculated from a data set containing, effectively, 75 proteins. Although the Williams et. al. values could not be used in the Fisher-Irwin test they can be used to test the comparisons involving Levitt's data. Table 2-5 shows the residues glycine > arginine > tyrosine > alanine > phenylalanine increasing in α-helix residency as temperature increases, while aspartic acid > methionine > histidine > valine = serine > lysine > isoleucine > asparagine show a decrease in residency.

It is now necessary to compare the P_{ij} values of Williams with the thermophilic value to test if the differences apply for both the Williams and Levitt data. For glycine the Williams P_{ij} is less than (0.43 versus 0.56) Levitt's value so the result showing an increase in glycine holds. For arginine, Williams value is within one standard deviation (SD) of the thermophilic value (1.21 and 1.18), so the increase seen versus Levitt's data is probably not correct. There is no significant difference between the P_{ij} values of Levitt and Williams (0.72 and 0.74) so the increase observed tyrosine stands. For alanine, the Williams data moves to within one SD of the thermophilic P_{ij} (1.41 versus 1.38). The P_{ij} values of Williams and the thermophilic data are identical for phenylalanine. To summarize, there is an increase in α-helix residency for glycine and tyrosine, but when compared with Williams data no increase in residency is observed for arginine, alanine, and phenylalanine.

Continuing the comparison for the residues which showed a decrease in residency for thermophilic proteins, there is little difference between Williams and the Levitt data for aspartic acid (0.99 versus 1.04). However, a large difference for methionine (1.30 versus 1.47) was observed. The thermophilic P_{ij} value is 0.99 ± 0.19. Williams value is still greater than one SD from the thermophilic data which suggests that methionine shows a decrease in residency. For histidine this is not the case, as the Williams data lies within one SD of the thermophilic data (1.05 and 0.94±0.15). Valine is not a concern, as the Williams data is greater than Levitt's (0.98 versus 0.91) so the statistical decrease is still valid. For serine there is an interesting problem. When compared with Levitt's data the thermophilic residency shows a decrease (0.82 versus 0.68±0.10). However, when compared with Williams data the thermophilic data shows a increase of greater than one SD (0.57 versus 0.68±0.10). Since the thermophilic value lies between the two mesophilic values, for the case of serine it appears that there is no statistical difference between the residency of mesophiles or thermophiles. The P_{ij} mesophile values for lysine are identical so the statistical decrease observed for Levitt's data holds. For isoleucine, the Williams value is greater than the Levitt value so the decrease observed is even more statistically significant. The Williams value, for asparagine, is closer to the thermophilic value so the tendency towards a decrease (probability of difference 0.742) is not significant. In summary, the decreases observed for aspartic acid, methionine, valine, lysine, and isoleucine are valid while the decreases for histidine, serine, and asparagine, when compared with Williams data, are not significant.

There is one residue which showed no significant difference when compared with Levitt's data, but when the proline thermophilic P_{ij} value is compared with the Williams value, proline shows a significant increase in residency (1.5 SD difference) for thermophilic proteins. What is interesting about this increase is that Scheraga [30] predicted that proline would show more stability in α-helix (Δs=0.06) at high temperatures. The

comparison between Williams P_{ij} value and the thermophilic P_{ij} seems to support this prediction.

Literature Predictions of Changes Required to Increase Thermostability. The residues for which a statistical increase in α-helix residency have been observed are glycine > tyrosine, and possibly proline. The residues aspartic acid > methionine > valine > lysine ≈ isoleucine all show decreased residency over mesophilic proteins. There have been two papers by Argos et. al. and Menéndez-Arias et al. [54,55] which used sequence comparison to look at the sequential changes required for increased thermostability. It is interesting to compare our results with the results of these papers.

Argos et. al. [54] found that the preferred changes to increase helix stability were changes from glycine to alanine, serine to alanine, aspartic acid to glutamic acid, and valine to alanine. All of these changes, with the exception of the change from valine to alanine, were found to increase hydrophobicity and internal packing or bulkiness of the side chain. There was a valine to isoleucine change which was found to have a small ΔP_α (+0.06) for α-helix. The Argos et. al. data would lead to the expectation of a decrease in the α-helix residency for glycine, serine, aspartic acid, and valine while seeing a concomitant increase in alanine and possibly glutamic acid.

Menédez-Arias and Argos [55] found shifts of isoleucine to valine, glutamic acid to arginine, serine to alanine, lysine to arginine, aspartic acid to arginine increased thermostability of the protein and had a preferred conformation of α-helix. Shifts of lysine to alanine and glycine to alanine were found to improve thermostability, but did not have a preferred location; being found in both α-helix and β-sheet. All of these changes, with the exception of the isoleucine to valine change, were found to increase hydrophobicity. For the isoleucine to valine change stabilization was thought to come about through a decrease in side chain flexibility. From these data we would expect to see decreases in α-helix residency for isoleucine, glutamic acid, serine, lysine, aspartic acid and possibly for glycine. Increases are expected for valine, alanine and arginine residues.

The results from these two studies and the results from this study are summarized in Table 2-6.

TABLE 2-6: A Comparison of Helix Residency and Thermostability Studies

Current	Study	Argos		Menédez	
increase	decrease	increase	decrease	increase	decrease
G	D	A	G	A	D
Y	M	E	S	R	E
P(?)	V		D	V	G(?)
	K		V		I
	I				K
					S

With the exception of aspartic acid none of the studies show much agreement. In the Argos et. al. and the Menédez-Arias study alanine is shown to increase thermostability while serine and aspartic acid decrease thermostability. Other than these three residues there is little agreement. If we assume that the Menédez-Arias study is the most accurate, since it used a larger data base, then comparisons can be made with the current study. The residues which decrease thermostability and are found less often in thermostable helices are very similar, with the exception of methionine and valine. Valine had the largest ratio of observed to expected for helix. However, its proposed mechanism of stabilizing, a reduction of stability, was different from all other observed helix thermostable residue shifts. Also Argos et. al. has valine as a residue to shift away from to increase thermostability. None of these studies predicted the increase in glycine and tyrosine we observed in this study.

One explanation for the increased use of the aromatic amino acid tyrosine is that thermal stability is increased through aromatic-aromatic interactions which provide

approximately 1 kcal/mol of interactions [56]. The favorable enthalpy is produced when an aromatic hydrogen, a nitrogen, or a sulfur atom interacts with the π electron cloud of another other ring perpendicular to the plane of the ring. The structure of thermolysin, which has the largest percentage of α-helix and the highest percentage of aromatic residues in helix of the protein studied, was checked visually for the presence of such aromatic interactions. Only two near perpendicular ring-ring interactions within 3.7 Å were observed for all the 41 aromatic residues in thermolysin. Neither interaction had a perfect 90° angle, *para* hydrogen perpendicular to and pointing into the plane of the π electron cloud, nor where the residues which showed a possible interaction found in α-helix. Though aromatic interactions may be an explanation for the increased use of tyrosine in thermostable proteins, these interactions were not observed to exist extensively, at least in thermolysin.

The use of tyrosine in an α-helix might be favored by thermophilic organisms because of the ability of this residue to hydrogen bond. Although tyrosine is unable to form the energetically more favorable salt bridge because it lacks a charge, it is capable of forming hydrogen bonds. The structure of thermolysin was again checked to see if many of the α-helical tyrosine residues formed hydrogen bonds. Two of the α-helical tyrosine were observed to form hydrogen bonds and one (Y 83) formed an intra-helix hydrogen bond with histidine. Thus, the formation of hydrogen bonds is one possible explanation for why tyrosine over phenylalanine is favored in thermostable proteins. Presumably, the amphipathic nature of tyrosine is also important since one would expect, if hydrogen bonding alone was the major stabilizing factor, that residues like glutamine, asparagine, histidine, and tryptophan would have shown statistically significant α-helix residency increases also, as these residues are not ionic and readily form hydrogen bonds.

One argument for the increased use of hydrophobic residues by thermophilic organisms would be to stabilize α-helices and thus their proteins through the hydrophobic

effect [57]. If the core of the protein becomes more hydrophobic then the stability of the core can be increased and the temperature at which the core is stable increases as well.

While all of the arguments listed above work well to explain the increase in residency for tyrosine there are no good explanations why glycine should show increased residency in thermophilic proteins. Menédez-Arias and Argos [55] list, in addition to the hydrophobic effect, a decrease in flexibility of the backbone atoms as a thermal stabilizing factor. Glycine is the most flexible of the amino acids, so if anything we would expect glycine to decrease in residency for thermostable proteins. The only study which would predict an increase in stability for glycine at high temperatures is the Zimm-Bragg **s** value measured by Scheraga [31]. For glycine the Δ**s** value is positive and from it we would predict a increase in the glycine residency for thermostable proteins, as observed

The prediction from the Zimm-Bragg negative Δ**s** value [36] that methionine would move from a helix forming to a helix indifferent residue was accurate (Tables 2-1 and 2-5). While the most obvious and reasonable explanation for its decrease in helix residency at high temperatures is that the sulfur present in the side chain oxidizes to a sulfoxide or sulfone, this is not a complete explanation. The host-guest polymers were carefully tested so as to exclude any oxidization products or the effects those products might have on the **s** value measured. The fact that, in this case, this technique predicted the observed effect and the magnitude of that effect forces us to place in question oxidization as the complete answer for the methionine temperature dependent decrease in helix residency. In general, most of the amino acids with large aliphatic side chain showed a decrease in helix residency. In the case of methionine, oxidation may work to increase the magnitude of the effect.

In summary, we have shown that Scheraga's Zimm-Bragg **s** value or Δ**s** value cannot be used to predict whether a residue will be thermally stable in α-helices. While the prediction was accurate for glycine, the direction of Δ**s** was wrong for tyrosine and valine and thus predicted exactly the opposite of that observed statistically. The lack of agreement

with other studies of thermal stability is puzzling. However, the fact that the two studies cited do not agree with each other reveals an inherent flaw in all the work on thermal stability done so far. In all these studies, the sample size is small. There is very little structural data about thermostable proteins, be that primary sequence or three-dimensional structure. Another problem may be in the method by which we have attempted to solve the problem. In addition the use of a large statistical study, may be hiding what are small cumulative changes. A better method of study might be to compare a single protein from organisms in cold environments to extreme thermophiles searching for the changes made in the individual protein as it increases in thermostability. Another method, which has already been proposed, is the mutagensis of a thermostable protein back to a mesostability. This reductive method would provide information about a pathway otherwise impossible to obtain. The current study provides useful information to those designing thermostable proteins and peptides. It has been shown for the first time that glycine and tyrosine show a net increase in thermostable α-helix residency, while valine, a residue previously described as increasing α-helix temperature stability, appears not to do so.

References

(1) Kendrew, J. C.; Dickerson, R. E.; Strandberg, B. E.; Hart, R. G.; Davies, D. R.; Phillips, D. C.; Shore, V. C. *Nature* **1960**, *185*, 422-427.

(2) Ptitsyn, O. B. *J. Mol. Biol.* **1769**, *42*, 501-510.

(3) Chou, P. Y.; Fasman, G. D. *Biochemistry* **1974**, *23*, 211-222.

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

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

(6) Vasquez, M.; Pincus, M. R.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 351-371.

(7) Yada, R. Y.; Jackman, R. L.; Nakai, S. *Int. J. Protein Res.* **1988**, *31*, 98-108.

(8) Vasquez, M.; Pincus, M. R.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 373-386.

(9) Chuknyiski, D. *Dokl. Bolg. Akad. Nauk* **1987**, *40*, 115-118.

(10) Wójcik, J.; Altmann, K. H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121-134.

(11) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526-535.

(12) Kidera, A.; Mochizuki, M.; Hasegawa, R.; Hayashi, T.; Sato, H.; Nakajima, A.; Fredrickson, R. A.; Powers, S. P.; Lee, S.; Scheraga, H. A. *Macromolecules* **1983**, *16*, 162-172.

(13) Holmes, M. A.; Matthews, B. W. *J. Mol. Biol.* **1982**, *160*, 623-639.

(14) Irwin, M. J.; Nyborg, J.; Reid, B. R.; Blow, D. M. *J. Mol. Biol.* **1976**, *105*, 577-586.

(15) Bhat, T. N.; Blow, D. M.; Brick, P. *J. Mol. Biol.* **1982**, *158*, 699-709.

(16) Skarzynski, T.; Moody, P. C. E.; Wonacott, A. J. *J. Mol. Biol.* **1987**, *193*, 171-187.

(17) Hecht, R. M.; Garza, A.; Lee, Y. H.; Miller, M. D.; Pisegna, M. A. *Nucl. Acids Res.* **1989**, *17*, 10123.

(18) Ludwig, M. L.; Metzger, A. L.; Pattridge, K. A.; Stallings, W. C. *J. Mol. Biol.* **1991**, *219*, 335-358.

(19) Teplyakov, A. V.; Kuranova, I. P.; Harutyunyan, E. H.; Frömmel, C.; Höhne, W. E. *FEBS Lett.* **1989**, *244*, 208-212.

(20) Blake, P. R.; Day, M. W.; Hsu, B. T.; Joshuator, L.; Park, J. B.; Hare, D. R.; Adams, M. W. W.; Rees, D. C.; Summers, M. F. *Prot. Sci.* **1992**, *1*, 1522-1525.

(21) Day, M. W.; Hsu, B. T.; Joshuator, L.; Park, J. B.; Zhou, Z. H.; Adams, M. W. W.; Rees, D. C. *Prot. Sci.* **1992**, *1*, 1494-1507.

(22) Davies, G. J.; Gamblin, S. J.; Littlechild, J. A.; Watson, H. C. *Proteins* **1993**, *15*, 283-289.

(23) Bowen, D.; Littlechild, J. A.; Fothergill, J. E.; Watson, H. C.; Hall, L. *Biochem. J.* **1988**, *254*, 509-517.

(24) Watson, H. C.,

(25) Wilson, K. S.; Appelt, K.; Badger, J.; Tanaka, I.; White, S. W. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 7251-7255.

(26) Lindley, D. V. *Ann. Math. Stat.* **1964**, *35*, 1622-1643.

(27) Maxfield, F. R.; Scheraga, H. A. *Biochemistry* **1976**, *15*, 5138-5153.

(28) Devore, J. L. In *Probability and Statistics for Engineering and the Sciences,* Brooks/Cole Publishing Co.: Monterey, CA, 1982; pp 301-309.

(29) Alter, J. E.; Andreatta, R. H.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 564-570.

(30) Altmann, K. H.; Wójcik, J.; Vásquez, M.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 107-120.

(31) Anathanarayanan, V. S.; Andreatta, R. H.; Poland, D.; Scheraga, H. A. *Macromolecules* **1971**, *4*, 417-424.

(32) Denton, J. B.; Powers, S. P.; Zweifel, B. O.; Scheraga, H. A. *Biopolymers* **1982**, *21*, 51-77.

(33) Dygert, M. K.; Taylor, G. T.; Cardinaux, F.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 794-801.

(34) Fredrickson, R. A.; Chang, M. C.; Powers, S. P.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 625-632.

(35) Hecht, M. H.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1978**, *11*, 545-551.

(36) Hill, D. J. T.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 2447-2467.

(37) Hughes, L. J.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 187-197.

(38) Kobayashi, Y.; Cardinaux, F.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1271-1283.

(39) Konishi, Y.; van Nispen, J. W.; Davenport, G.; Scheraga, H. A. *Macromolecules* **1977**, *10*, 1264-1271.

(40) Matheson Jr., R. R.; Nemenoff, R. A.; Cardinaux, F.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 1567-1585.

(41) Maxfield, F. R.; Alter, J. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1975**, *8*, 479-493.

(42) Nagy, J. A.; Powers, S. P.; Zweifel, B. O.; Scheraga, H. A. *Macromolecules* **1980**, *13*, 1428-1440.

(43) Platzer, K. E. B.; Anathanarayanan, V. S.; Andreatta, R. H.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 177-187.

(44) Scheule, R. K.; Cardinaux, F.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 23-33.

(45) Scheraga, H. A. *Pure Appl. Chem.* **1978**, *50*, 315-324.

(46) Sueki, M.; Lee, S.; Powers, S. P.; Denton, J. B.; Konishi, Y.; Scheraga, H. A. *Macromolecules* **1984**, *17*, 148-155.

(47) Van Nispen, J. W.; Hill, D. J.; Scheraga, H. A. *Biopolymers* **1977**, *16*, 1587-1592.

(48) Van Wart, H. E.; Taylor, G. T.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 266-273.

(49) Williams, R. W.; Chang, A.; Juretic, D.; Loughran, S. *Biochem. Biophys. Acta* **1987**, *916*, 200-204.

(50) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648-1652.

(51) O'Neil, K. T.; DeGrado, W. F. *Science* **1990**, *250*, 646-651.

(52) Marascuilo, L. A.; McSweeny, M. In *Nonparametric and Distribution-Free Methods for the Social Sciences,* Brooks/Cole Publishing Co.: Monterey, CA, 1977; pp 96-126.

(53) Chothia, C.; Lesk, A. M. *EMBO J.* **1986**, *5*, 823-826.

(54) Argos, P.; Rossmann, M. G.; Grau, U. M.; Zuber, H.; Frank, G.; Tratschin, J. D. *Biochemistry* **1979**, *18*, 5698-5703.

(55) Menéndez-Arias, L.; Argos, P. *J. Mol. Biol.* **1989**, *206*, 397-406.

(56) Burley, S. K.; Petsko, G. A. *Science* **1985**, *229*, 23-28.

(57) Tanford, C. In *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*; 2 ed. Wiley: New York, 1980; pp 21-28.

CHAPTER 3

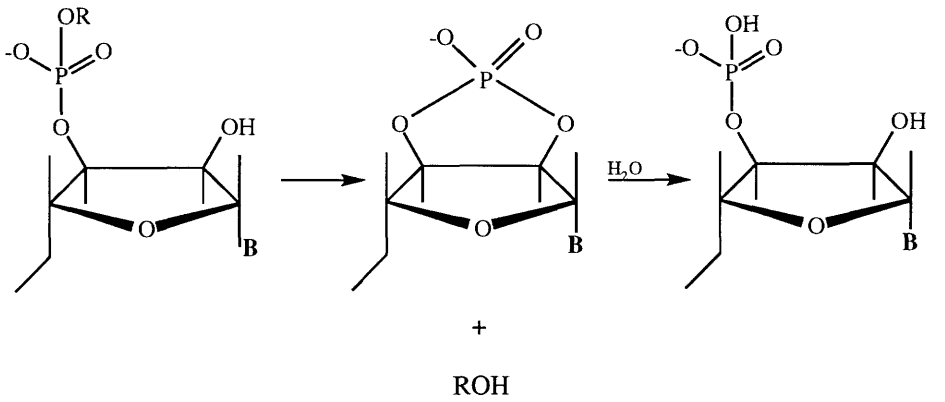
Protein Structure and Dynamics Studies By Solid State NMR

Introduction

In solid state NMR, as in solution, two methods exist for obtaining structural information; chemical shift, and/or magnetization exchange processes. While the chemical shift is the most sensitive method, registering structural changes on the order of hundredths to thousandths of an angstrom, there is no method by which the chemical shift can be turned into quantifiable structure information. On the other hand, magnetization exchange processes, in particular dipolar coupling exchange, provide quantifiable structural information but are insensitive, with one exception [1], to changes in structure of less than 0.1 Å [2-5]. The next two chapters are concerned with both processes in an effort to understand something about the dynamics and structure of two proteins: ribonuclease A from bovine pancreas (RNase), and glutamine synthetase from *Escherichia coli.*

Bovine pancreatic ribonuclease A (RNase A) is a 13 kD enzyme that consists of 124 amino acid residues and catalyses the hydrolysis of RNA via a two step process. Its small size, availability and ruggedness make RNase a good model for physical and chemical studies of enzyme catalysis. Extensive enzymatic and structural studies [6] reveal that it does not catalyze the hydrolysis of DNA which lacks the 2' hydroxyl group that is essential for this reaction. It has also been observed that it has a strong specificity for the pyrimidine, i.e. U or C, ribonucleotide bases. The first step in the hydrolysis of RNA is the formation of a cyclic phosphate intermediate during an in-line attack by the 2'-hydroxyl group as seen in Equation 3-1. The cyclization step is generally accepted as

being much faster than the hydrolysis step, with the histidine residues H-12 and H-119 acting as general-base and general-acid catalysts, respectively, for cyclization. H-12 is the general-acid and H-119 the general-base catalyst for the hydrolysis step. This reversal of roles between cyclization and hydrolysis is expected based on the principle of microscopic reversibility [6].



(3-1)

Of interest in this study are the binding properties of the substrate. Cryogenic crystallographic studies have shown that crystallographic B-factor changes as a function of temperature in a biphasic fashion with a relatively flat region from 210 to 230° K [7]. Previous solid state ^{31}P NMR investigations of 2'-cytidine monophosphate (2'-CMP) bound to RNase in the crystalline state indicate a dianionic state for the bound inhibitor [8]. The study of chemical shift anisotropy would be of value in understanding the structural and dynamic aspects of such enzyme substrate complexes. ^{31}P is an ideal nucleus for such a study owing to its high sensitivity and high natural abundance.

In nuclear magnetic resonance, the most sensitive but least quantifiable probe of structural change is the chemical shift both isotropic and anisotropic. Structural change measured by chemical shift is not quantifiable because, to date, there is no theory as to how to calculate, on first principles, chemical shift from structure or vice versa. In solid state NMR the chemical shift has two components, a measure of the local directionally independent magnetic environment (isotropic chemical shift) and a directional component

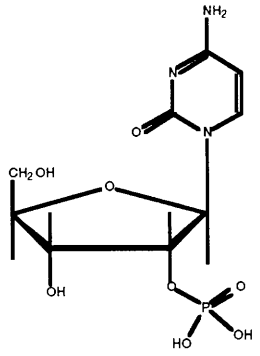
(anisotropic chemical shift tensors). Andrew et. al. [9,10] showed that the intensity of heteronuclear dipolar $\|H_{IS}\|$ and chemical shift anisotropy $\|H_{ZS}\|$ components of the line to first order go as $(3\cos^2\theta - 1)/2$, where θ is angle of the sample and NMR coil with respect to the static B_0 magnet field. If the sample is spun at an angle of $\cos^2\theta = 1/3$ or $\theta = 54.735°$ and at a speed of $\omega_r > \|H_{ZS}\|, \|H_{IS}\|$ these components are averaged to zero. θ is called the "magic" angle and spinning at this angle is referred to as magic angle spinning (MAS). If $\|H_{ZS}^{aniso}\| > \omega_r > \|\overline{H}_{ZS}^{iso}\|$ then analysis of the rotational side bands yields a measure of the local magnetic environment of the nuclei (isotropic chemical shift) and orientation information (anisotropic chemical shift) [11]. Thus, a qualitative measure of binding dependent structural changes in RNase A can be obtained by studying changes in chemical shift tensors in the bound and unbound states.

Experimental Procedures

Protein Crystallization. All solid state NMR studies performed used crystallized bovine pancreatic ribonuclease A type X11-A from Sigma (St. Louis, MO). RNase A was crystallized as follows. RNase was dissolved in distilled water to a concentration of 24 mg/mL. Methanol was added slowly to this solution until the concentration was 50% (v/v). The pH was adjusted to 5.5 and the solution filtered using a 0.22 μm Millipore filter. Aliquots (1.0 mL) were placed in siliconized glass vials. Methanol was added drop wise to the vials until the solution remained opalescent for approximately 10 seconds after addition of methanol. This occurred at approximately 52% (v/v) methanol. The solutions were seeded, sealed and left to crystallize for two weeks. Final protein concentration was 12 mg/mL.

Ribonuclease A Inhibition Standard. RNase (100 mg) was dissolved in 2.0 mL distilled water. 2'-cytidine monophosphate (2'-CMP, **1**), purchased from Sigma (St. Louis, MO), was added to the solution (2.6 mg, 1.1 eq). The solution was allowed to

stand at room temperature for 30 minutes and frozen in a liquid nitrogen bath. The sample was lyophilized and stored at -20° C.



1 2'-CMP

^{31}P NMR. ^{31}P cross polarization [12] magic angle spinning (CPMAS) solid state NMR (SSNMR) experiments with continuous wave (CW) proton decoupling [13] were performed on a home built spectrometer operating at 317 MHz for protons and 128.6 MHz for ^{31}P. A home built variable temperature double resonance probe ($^{1}H/^{31}P$) was used with a spinning assembly from Doty Scientific Inc. (Columbia, SC). The probe and sample were cooled using a low temperature device described previously [14]. Typical ^{1}H and ^{31}P pulse lengths were 3.5 and 5 μs respectively. A 3 s recycle delay was used which was adequate for full relaxation (>5 ^{1}H T_1) with an optimal cross polarization time of 2 ms for the enzyme-inhibitor complex. The RNase sample was prepared as follows for the NMR experiment. A pooled sample from a 720 mg crystallization containing protein crystals and supernatant was centrifuged at 5,000 G for 30 minutes. The supernatant was decanted and approximately 180 mg of the resulting crystal slurry packed into a 7 mm single crystal sapphire rotor. The probe and rotor were cooled to 195° K and then 5 mg of 2'-CMP in 200 μL 60% (v/v) methanol (1.2 eq) was added to the rotor. The cold sample and probe were placed in the magnet and a CPMAS spectrum was obtained. Spectra at higher temperatures were obtained by warming the probe while in the magnet and then acquiring the spectrum.

Chemical Shift Tensor Analysis. The chemical shift tensors were determined by analysis of side band intensities based on the method of Herzfeld and Berger[15]. Chemical shift anisotropy, δ, and the asymmetry parameter, η, follow the Spiess convention. The asymmetric tensor elements represented by R_{xx}, R_{yy}, and R_{zz} were assigned such that $|R_{zz} - R_{iso}| \geq |R_{xx} - R_{iso}| \geq |R_{yy} - R_{iso}|$, $\delta = R_{zz}$, $\eta = (R_{yy} - R_{xx})/(R_{zz})$ [16].

Results

Since small changes in ^{31}P isotropic chemical shifts occur during binding and inhibition of RNase to 2'-CMP [8], our experiments were concerned with interpreting the large changes in chemical shift tensors to elucidate changes in structure and/or in the ionization state during the enzymatic reaction . The isotropic chemical shifts for the bound and free inhibitor are identical to those obtained by Dobson and Lian [8,17,18], where they performed a direct comparison between the bound and unbound inhibitor in solid and solution states. For 2'-CMP the chemical shift of the bound phosphate group was observed to be very close to that of the dianionic state of free 2'-CMP in solution, a result rationalized on the basis of interactions with Lys41 and His119 in the active site [17]. In the present study, the aim is to utilize the chemical shift tensors in the free and bound states to provide a detailed description of the structural changes that occur upon binding. The chemical shift parameters, including both the isotropic chemical shift and the anisotropic chemical shift tensors, were analyzed from three experiments and are found in Table 3-1.

The first experiment was a comparison of the unbound inhibitor 2'-CMP, measured in a frozen solvent similar to the enzyme crystallization conditions, with the inhibition standard described previously. Comparisons of these parameters to that of 2'-CMP bound to RNase are found in Table 3-1. There is no observable change in isotropic chemical shift of the 2'-CMP phosphate group. This disagrees with the results from Dobson and Lian [8] in solution, which state that at a pH of 5.5, free 2'-CMP in solution

exists in a combination of the monoanionic and dianionic form while 2'-CMP bound to RNase in the crystalline state exists almost completely in the dianionic form.

In the second experiment, the spectrum of the solubilized free inhibitor was compared with that of a crystallized RNase-inhibitor mixture at 295 K. Once again, it can be seen that there is no major change in the isotropic chemical shift difference of the 2'-CMP phosphate group. In the third experiment , the inhibitor was allowed to soak into the RNase A crystals over night and spectra of the soaked enzyme-inhibitor complex were obtained. It should be noted, from Table 3-1, that once again there was no change in the isotropic chemical shift, σ_{iso}. This would be expected if 2'-CMP was bound to the enzyme in the dianionic state since the single pH value used for all the samples was well above the titration point for bound 2'-CMP [8] in a methanol solution.

From the second experiment, it was observed that the anisotropic chemical shift tensors (in Table 3-1), σ_{11} and σ_{33} for the 295° K data for the crystallized enzyme and inhibitor mixture, are similar to the solubilized inhibitor values (experiment 1). The downfield and upfield extremes of tensors, σ_{11} and σ_{33}, have previously been shown to vary with bond length and bond angle [19,20]. The absence of a change in these values indicates that the bond lengths and bond angles of the phosphate group in the inhibitor remain unchanged in the crystallized enzyme-inhibitor mixture. It has also been demonstrated that a change in the σ_{22} element does not correlate well with changes in the bond lengths and angles , however, σ_{22} is sensitive to the ionization state of the phosphate group. As can be seen in Table 3-1, σ_{22} remains practically unchanged on going from the solubilized inhibitor to the crystallized enzyme-inhibitor mixture which indicates that there is no change in the ionization state of the inhibitor.

However, it is interesting to note that the tensor elements σ_{11} and σ_{22} are markedly different for the soaked crystal and free solubilized inhibitor data (experiment 3). On comparing the values of σ_{22} for urea phosphoric acid complexes, BDEP, calcium hydrogen phosphate dihydrate and calcium dihydrogen phosphate monohydrate [21,22], it

can be seen that σ_{22} shifts upfield upon deprotonation. Therefore it can be concluded that a change from 11.7 ppm (in the free inhibitor) to -25.9 ppm (in the soaked crystal mixture) indicates a change in the ionization state of the phosphate group of 2'-CMP upon binding. A 50 ppm downfield shift in σ_{11} on going from the free inhibitor to the soaked crystal mixture is indicative of a change in the bond lengths and angles. Variations of bond lengths in phosphates and their esters result from reduction in or elimination of the capacity of one or more of the oxygen to contribute to d-p π-bonding. This occurs when one or more of the of the oxygen p orbitals, which contribute to the d-p π-bonding molecular orbitals, must also accommodate a bond resulting from protonation [19]. In this case, we believe the variation in σ_{33} to be a measure of the ability of the oxygen O1's (Figure 3-1), which is involved in an ester linkage with the sugar, contribution to d-p π-bonding. The downfield shift of σ_{11} is the result of a decrease in the ability of O1 to contribute to d-p π bonding. This in turn would tend to increase the d-p π-bond order on the other 3 oxygen atoms and thus a larger change would be observed in the σ_{11} - σ_{22} plane.

A change from 11.7 ppm (in the free inhibitor) to -25.9 ppm (in the soaked RNase crystals-inhibitor mixture) for σ_{22} indicates a change in the protonation state of the phosphate group of 2'-CMP. The upfield direction of this shift is indicative of deprotonation in the soaked crystals mixture. This is in agreement with the suggestion by Dobson and Lian [8] that 2'-CMP exists in the dianionic state when bound to RNase. This result demonstrates that the isotropic chemical shift is a poor indicator of the ionization state of 2'-CMP since no difference was observed between free and bound inhibitor. There are significant differences for the chemical shift anisotropy δ and the asymmetry parameter η on going from the free inhibitor to the soaked crystals. The δ and η values, both show that the free inhibitor has a larger anisotropy and is more asymmetric than the bound form. This is to be expected if bound 2'-CMP is in the dianionic state. A dianionic phosphate group is more symmetric and can undergo a tensor narrowing hopping motion.

This was evidenced by the fact that the width of the tensor for the inhibitor was considerably narrower in the bound form than in the free form.

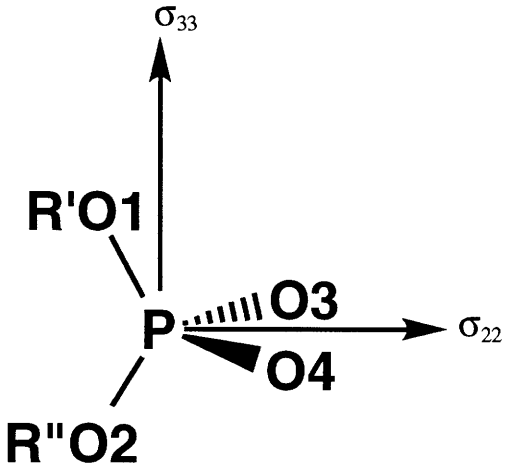


Figure 3-1: The chemical shift tensor orientation with respect to the molecular frame. The tensor σ_{11} is perpendicular to the page pointing out towards the viewer [20]. R'=P,C,H and R"=P,C,H. For 2'-CMP R' is the sugar moiety and R" is H for the monoanionic state.

TABLE 3-1: ^{31}P Chemical Shift Parameters[a]

Species	Temp. (°K)	σ_{iso}	σ_{11}	σ_{22}	σ_{33}	δ	η
2'CMP 52% MeOH	295	0.0	-71.1	11.7	59.4	71.1	0.67
2'CMP RNase	295	0.0	-76.4	14.4	61.9	76.4	0.62
2'CMP RNase lyph	295	0.0	-51.4	-5.5	56.9		
2'CMP RNase	195	0.0	-66.8	10.6	56.1	66.8	
2'CMP RNase	215	0.0	-59.6	11.0	48.5	59.6	
2'CMP RNase	230	0.0	-62.2	12.9	49.2	62.2	
2'CMP RNase	245	0.0	-58.9	10.6	48.2	58.9	
2'CMP RNase	255	0.0	-60.2	11.7	48.5	60.2	
2'CMP RNase soak	295	0.0	-33.3	-25.9	59.2	59.2	0.12
Urea-H_3PO_4 [21]		-5.2	-44.6	2.5	26.6		
Ba-diethylPO_4 [21]		-5.5	-109.8	17.5	75.9		
$CaHPO_4 \cdot 2H_2O$ [22]		-1.0	-53.0	-12.0	70.0		
$Ca(H_2PO_4)_2 \cdot H_2O$ [22]		0.7	-48.8	1.0	49.0		
$Ca(H_2PO_4)_2 \cdot H_2O$ [22]		0.0	-66.0	7.0	59.0		

[a] ^{31}P NMR chemical shift parameters of the phosphorus containing inhibitor and inhibitor-protein complexes: 2'CMP 52% MeOH is the inhibitor solvated in 52% methanol solution, 2'CMP RNase lyph is the inhibition standard described previously, 2'CMP RNase is a mixture of crystallized RNase and a 2'-CMP solution described previously, 2'CMP RNase soak is a mixture of RNase crystals and 1.2 eq of 2'-CMP soaked over night in mother liquor.

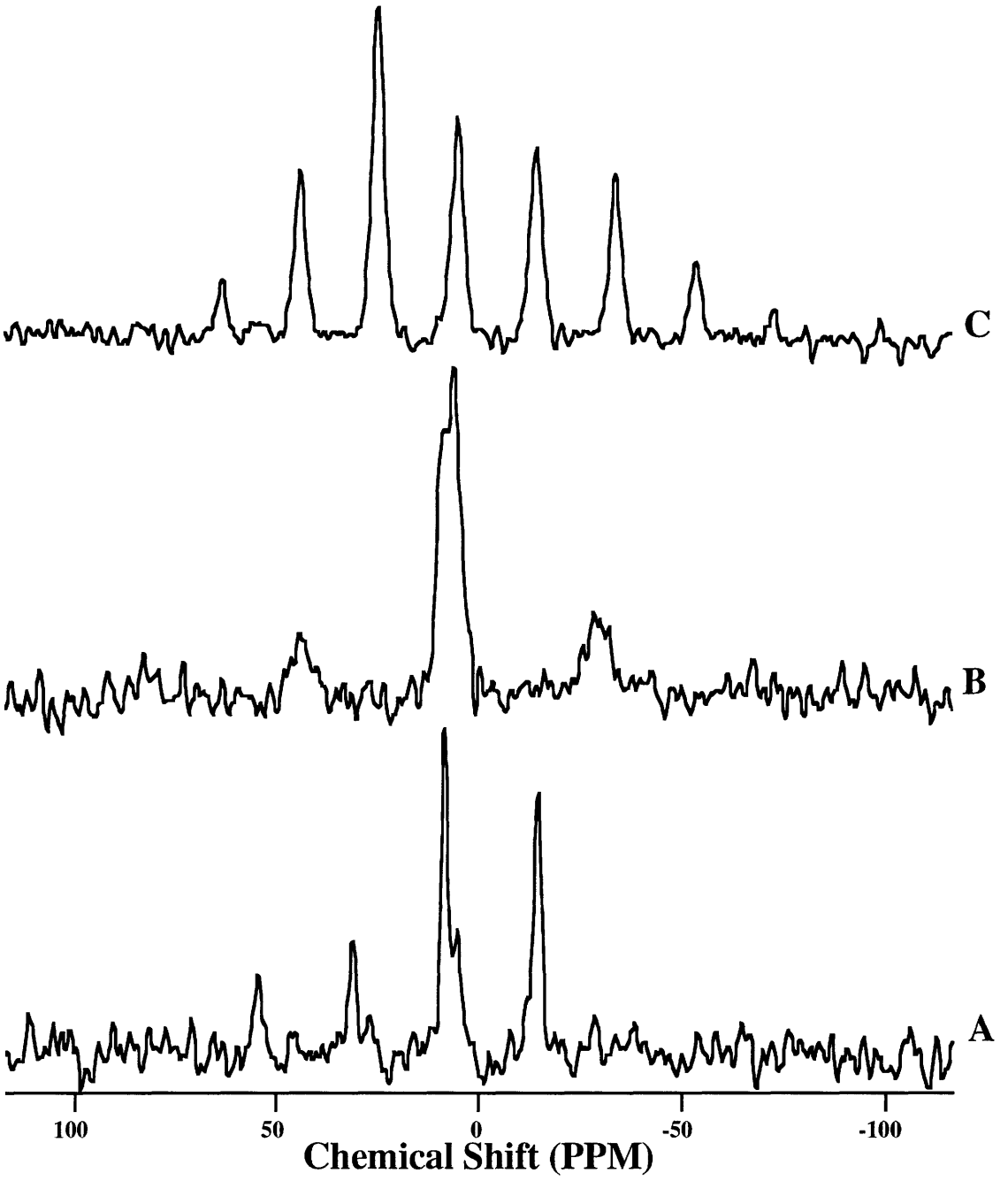


Figure 3-2: ^{31}P CPMAS spectra of inhibitor and enzyme-inhibitor complexes. Spectrum A is of RNase crystals soaked with 1.2 eq. 2'-CMP dissolved in mother liquor over night. Spinning speed for this spectrum was 2.979 KHz. Spectrum B is of the standard lyophilized RNase 2'-CMP complex. The spinning speed was 4.931 KHz. Spectrum C is of 100 mg 2'-CMP solvated with 200 μL 52% (v/v) MeOH. Spinning speed was 2.490 KHz.

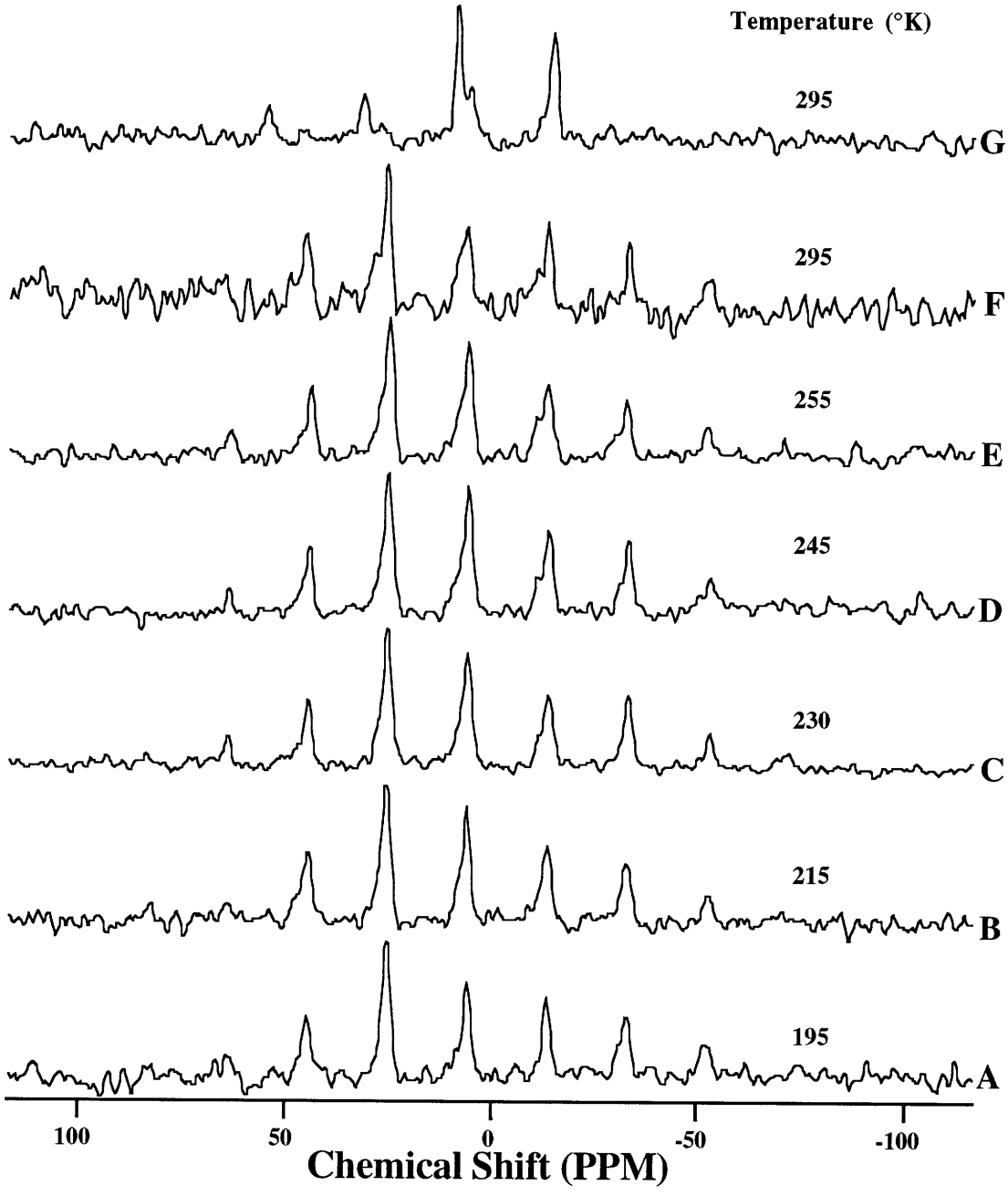


Figure 3-3: ^{31}P CPMAS temperature dependent spectra of crystallized RNase and a 2'-CMP mixture is 52% (v/v) MeOH. Spectra A to F are from a single sample of crystallized RNase cooled to 180 °K. An inhibitor solution was added and the sample warmed to 195 °K. The spinning speeds (SS) for spectra A to F are as follows: A SS=2.49 KHz, B SS=2.44 KHz, C SS=2.49 KHz, D SS=2.49 KHz, E SS=2.49 KHz, and F SS=2.54 KHz. Spectrum G is the soaked RNase crystal sample described previously. The spinning speed for this sample was 2.98 KHz.

Spectra of solubilized 2'-CMP, the standard lyophilized enzyme-inhibitor complex, and the overnight soaked crystal data are found in Figure 3-2. The standard inhibition sample and soaked crystal sample have similar spectra, while the solubilized inhibitor 2'-CMP differs greatly. The difference in the spectra is in the width of the chemical shift tensor (it is narrower for the standard lyophilized enzyme-inhibitor complex and the overnight soaked crystals and broader for the free solubilized inhibitor). There is an asymmetry or splitting of the center band in both spectrum A and B. This can be explained by the fact that the overnight soaked crystals might be contain predominantly the dianionic form of the inhibitor in the bound state but the presence of some unbound monoanionic inhibitor in the sample cannot be ruled out.

Temperature dependent spectra are shown in Figure 3-3. There is no noticeable differences in spectra A through E, which cover a temperature range from 195° K to 255° K. There is a shift in all the resonance lines for spectrum F. Though the spinning speed increased by 1.6%, the differences seen in this spectrum are not due to an increase in the spinning speed but to a change in the shift tensors as is seen in Table 3-1. All of these spectra (A-F) are similar to solvated 2'-CMP spectrum (see spectrum B in Figure 3-2) and are very different from the soaked crystal spectrum G. It should be noted that an asymmetry in the line shape starts in spectrum E and continues to develop in spectrum F. The asymmetry is similar to the one seen in Figure 3-2 spectrum B, but different from the splitting of the center band observed in spectrum G of Figure 3-2. The asymmetry is probably due to an unresolved isotropic chemical shift difference between the monoanionic and dianionic states of 2'-CMP.

Discussion

The questions we wished to address in this study were: Does 2'-CMP undergo structural and dynamic changes upon binding to RNase and would there be noticeable

changes in the chemical shift tensors in the bound and unbound states. If so, can a qualitative measure of binding dependent changes in RNase be obtained by studying changes in the chemical shift tensors of the phosphate group of the inhibitor.

We observed that the isotropic chemical shift in the solid state is not very sensitive to the changes that occur upon binding although our chemical shift for the free solubilized inhibitor and the bound lyophilized inhibition standard are in agreement with those obtained by Dobson and Lian [8].

We have demonstrated that, for the 2'-CMP RNase system, anisotropic chemical shift tensor elements are a sensitive register of changes in structure. Chemical shift tensor elements, for this system, seem to be a better probe for structural change because the difference for σ_{11} and σ_{22} between bound and unbound for 2'-CMP is large. σ_{22} for 2'-CMP in the free state and the unbound state is at 11.7 ppm and 14.4 ppm respectively whereas it shifts upfield to -25.9 ppm in the bound state. This change in σ_{22} is indicative of deprotonation of the 2'-CMP phosphate group. This result is in agreement with the suggestion by Dobson and Lian [8] that the inhibitor exists in the dianionic form in the bound state. A downfield shift of 50 ppm in σ_{11} on going from the free state to the bound state of the inhibitor indicates that there is a change in the d-p π bond order in the σ_{11} - σ_{22} plane and that there is considerable deshielding owing to the interactions with Lys41 and His119 in the active site of RNase [17].

Furthermore, the CSA and the asymmetry parameter are smaller in the bound state as compared to the unbound state. This is explained by the fact that the dianionic form of the inhibitor is more symmetric than the monoanionic form. The width of the chemical shift tensor is narrower in the bound state than in the unbound state and this suggests that there might be motional averaging (e.g. three site hopping motion in the phosphate group) in the bound dianionic state.

It is of interest to understand how temperature and thus the dynamic state of the enzyme effects the binding of substrate. We have demonstrated that it is possible to study

small structural changes using anisotropic chemical shift. Future work in this field will include the use of anisotropic chemical shift to study temperature dependent structural changes. One suggestion is the use of 3'-CMP as the inhibitor for measuring the anisotropic tensor values over a temperature range for inhibited crystallized RNase. Since 3'-CMP binds less tightly it may have smaller induced structural changes due to binding. We would it expect it to be a sensitive probe of temperature related structural change in that any restriction of motion in the active site would be readily observed. Because inhibited enzyme would be used freezing of the sample, which would occur at low temperatures, would in no way affect the results.

In summary, we have demonstrated that it is possible to use anisotropic chemical shift tensors as a sensitive measure of changes in protein structure . We have shown it is possible to obtain useful information about protein structure and temperature related dynamics using ribonuclease A.

References

(1) Copié, V. C.; Kolbert, A. C.; Drewry, D. H.; Bartlett, P. A.; Oas, T. G.; Griffin, R. G. *Biochemistry* **1990**, *29*, 9176-9184.

(2) Raleigh, D. P.; Harbison, G. S.; Neiss, T. G.; Roberts, J. E.; Griffin, R. G. *Chem. Phys. Letters* **1987**, *138*, 285-290.

(3) Raleigh, D. P.; Levitt, M. H.; Griffin, R. G. *Chem. Phys. Letters* **1988**, *146*, 71-76.

(4) Colombo, M. G.; Meier, B. H.; Ernst, R. R. *Chem. Phys. Letters* **1988**, *146*, 189-196.

(5) Gullion, T.; Poliks, M. D.; Schaefer, J. *J. Magn. Reson.* **1988**, *80*, 553-558.

(6) Fersht, A. *Enzyme Structure and Mechanism;* Second ed.; W.H. Freeman and Co.: New York, 1985, pp 426-431.

(7) Alber, T.; Dao-pin, S.; Nye, J. A.; Muchmore, D. C.; Matthews, B. W. *Biochemistry* **1987**, *26*, 3754-3758.

(8) Dobson, C. M.; Lian, L. Y. *FEBS Lett.* **1987**, *225*, 183-187.

(9) Andrew, E. R.; Bradbury, A.; Eades, R. G. *Nature* **1958**, *182*, 1659.

(10) Lowe, I. J. *Phys. Rev. Lett.* **1959**, *2*, 285-287.

(11) Ernst, R. R.; Bodenhausen, G.; Wokaun, A. In *Principles of Nuclear Magnetic Resonance in One and Two Dimensions,* Clarendon Press: Oxford, 1987; pp 387-399.

(12) Pines, A.; Gibby, M. G.; Waugh, J. S. *J. Chem. Phys.* **1973**, *59*, 569-590.

(13) Schaefer, J.; Stejskal, E. O. *J. Am. Chem. Soc.* **1976**, *98*, 1031-1033.

(14) Allen, P. J.; Creuzet, F.; DeGroot, H. J. M.; Griffin, R. G. *J. Magn. Reson.* **1991**, *92*, 614-617.

(15) Herzfeld, J.; Berger, A. E. *J. Chem. Phys.* **1980**, *73*, 6021-6030.

(16) McDermott, A. E.; Creuzet, F.; Griffin, R. G.; Zawadzke, L. E.; Ye, Q.; Walsh, C. T. *Biochemistry* **1990**, *29*, 5767-5775.

(17) Haar, N.; Thompson, J. C.; Mauver, N.; Ruterjans, H. *Eur. J. Biochem.* **1973**, *40*, 259-266.

(18) Gorenstein, D. G.; Wyrwicz, A. M.; Bode, J. *J. Am. Chem. Soc.* **1976**, *98*, 2308-2314.

(19) Cruickshank, D. W. J. *J. Chem. Soc.* **1961**, 5486.

(20) Un, S. Thesis, University of California, Berkeley, 1987.

(21) Herzfeld, J.; Griffin, R. G.; Haberkom, R. A. *Biochemistry* **1978**, *17*, 2711.

(22) Rothwell, W. P.; Waugh, J. S.; Yesinowski, J. P. *J. Am. Chem. Soc.* **1980**, *102*, 2637.

CHAPTER 4

Structure Determination In *Escherichia coli* Glutamine Synthetase By Rotational Resonance ^{31}P Solid State NMR

Introduction

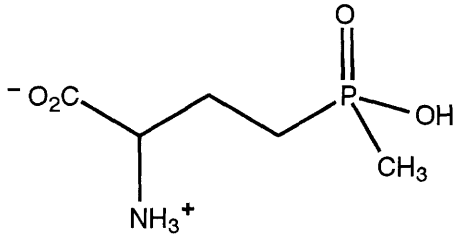
As discussed in Chapter 3, there are two NMR methods of obtaining structural information for biological macromolecules: chemical shift (isotropic or anisotropic) and magnetization exchange processes, namely exchange through dipolar coupling. This chapter describes work which uses dipolar coupling, employing one of the many methods currently available, to obtain structural information about a large protein complex.

Glutamine synthetase (GS) from *Escherichia coli* is a soluble high molecular weight (MW~600 kD) protein with 12 subunits. The enzyme is thought to form a tetrahedral adduct during the two step ATP-dependent process of converting glutamate to glutamine [1]. Inhibitors which, when phosphorylated, mimic this postulated tetrahedral adduct form a tightly bound complex halting the enzymatic reaction in the step just prior to release of product (Equation 4-1).

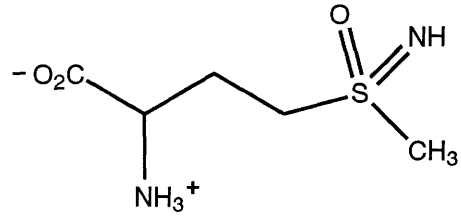
$$\mathbf{E + I \rightleftharpoons E{\bullet}I \longrightarrow E{\bullet}I\text{-}P \longrightarrow E^{*}{\bullet}I\text{-}P \longrightarrow E{+}I{+}P} \qquad \mathbf{(4\text{-}1)}$$

The hydrolysis of ATP is necessary for inhibition. This was demonstrated spectroscopically using a stable phosphorylated inhibitor, D,L-2-amino-4-phosphonobutyric acid (APB), which was isolated from inactivated enzyme and its structure determined by NMR [2]. It has been shown that another class of inhibitors, L-

phosphinothricin (L-2-amino-4-(hydroxymethylphosphinyl) butanoic acid, PPT, **1**) [2,3] and L-methionine-S-sulfoximine (MSOX, **2**) [4], inactivate GS in the presence of ATP. Both PPT and MSOX have been isolated from inactivated enzyme and characterized by NMR [1]. However, direct attempts to observe phosphorylated PPT (PPT-P) bound to the enzyme in solution using ^{31}P NMR have failed due to extreme line broadening of the ^{31}P resonance as a result of the large molecular weight of the complex. Solid state MAS NMR is uniquely suited for the solution of this problem as the line width for the ^{31}P resonance is unrelated to the molecular weight of the solid being studied.



1 PPT



2 MSOX

Magic angle spinning (MAS) removes second rank tensor interactions, including isolated spin pair dipolar coupling, from the spectrum. Consequently dipolar coupling, if it is to be used, must be reintroduced. One of the first methods found for reintroducing homonuclear dipolar coupling was discovered by Andrew et. al. [5]. They found that if the sample was spun at an integer multiple of the chemical shift difference, i.e. $\omega_{\Delta}^{iso} = n\omega_r$, of two coupled homonuclear spins their resonance lines broaden. Recently it has been shown that resonance line broadening or the magnetization transfer rate can be modeled and the distance between the homonuclear spin pair determined [6-10]. This method, Rotational Resonance (R^2), has been used to measure spin pair distances in peptides, membrane proteins, and in a photointermediate of bacteriorhodopsin [11-13]. Magnetization transfer for the spin pair goes as $1/r^6$ and, for values of $n > 1$ of $n\omega_r$, shows a chemical-shift tensor

orientation dependence. Distances of up to 6 Å have been successfully measured for ^{13}C-^{13}C and ^{31}P-^{31}P spin pairs.

In the present study we employ a homonuclear magnetization transfer method to determine if there is coupling between the phosphinate and phosphorus species present in the PPT ATP inactivated enzyme. The homonuclear magnetization transfer technique used is a modification of the R^2 technique described previously [6-13]. This method allowed us to demonstrate that PPT is phosphorylated in the inactivated enzyme. We have measured a P-P distance of 2.7 ± 0.2 Å between the phosphinate of PPT and phosphate, which is indicative of a P-O-P ester linkage between the phosphinate and phosphate.

Experimental Procedures

Enzyme Inactivation. Glutamine synthetase from *Escherichia coli* was purified and inactivated as described previously, [1,2] with one exception. The samples used for solid state MAS NMR had a lower salt concentration and were prepared as follows: 47 mg of GS in 1.5 mL of 1.2 mM $MgCl_2$ and 10 mM Hepes buffer at pH 7.5. The sample was inactivated by incubating with 1.2 mM DL-PPT and 1.08 mM ATP for 5 minutes at 37° C and was then frozen and lyophilized at room temperature. The GS had a starting specific activity of 114.4 with an adenylation state of approximately 1. After inactivation the specific activity was 35.4.

R^2 SELTICS Development. The experiments described were performed on a home-built spectrometer operating at 317 MHz for ^1H and 79.9 MHz for ^{13}C. All sample spinning was performed using rotors and stators from Doty Scientific Co. (Columbia, SC) and the spinning speed was stabilized to within ±0.05% [14]. R^2 SELTICS is R^2 [7-9] with the sideband suppression technique SELTICS [15] added just before acquisition. The pulse sequence and phase cycle modes are shown in Figure 4-1. There are two forms of SELTICS side band suppression, the basic SELTICS and SELTICSπ which contains a

refocusing π pulse in the middle of the sequence to correct for linear dephasing errors. Both methods were extensively tested for differences. Side band suppression was performed by measuring the spinning speed and then calculating the pulse lengths and delays based on the pulse sequence in Figure 4-1. The parameters were then optimized to obtain the best sideband suppression possible. Testing of these sequences was done using two model compounds, (1,2-^{13}C)-Zinc acetate (ZnAc) [9] and 15% (4'-$^{13}C_1$)-tyrosine (1'-$^{13}C_1$)ethyl ester (TEE) [9]. Magnetization exchange rates were determined by integrating the peak area for the rotationally resonant spin pair with respect to the mixing time. The peak areas were summed, normalized to the initial value, and plotted (Figures 4-2 and 4-3) along with numerical simulations [10]. A number of rotational resonance conditions, i.e. $n\omega_r$, were tested for both model compounds (TEE n=2,3,4; ZnAc n=3,4,5,6).

NMR of Glutamine Synthetase. All ^{31}P NMR spectra were acquired on a home-built spectrometer operating at 212 MHz for ^{1}H and 85.7 MHz for ^{31}P. The spinning speed was stable to better than ±0.5%. Typical pulse lengths were 3 μs for ^{1}H and 3.8 μs for ^{31}P with a soft inversion pulse length of 650 μs. The optimum cross polarization time was measured to be 0.5 ms for this sample. ^{31}P spectra acquired with and without SELTICS are shown in Figure 4-4. The resonance lines, for all the GS samples, were broadened due to chemical shift inhomogeneity, resulting in line widths on the order of 600 to 1000 Hz. All R^2 SELTICS data were acquired at the n=1, $n\omega_r$, R^2 condition and a spinning speed of 5.41 KHz. At mixing times of 2 ms all of the magnetization had exchanged (Figure 4-6).

Typically 9600 transients were acquired per mixing time. The magnetization exchange rate was determined by integrating the area under the phosphinate resonance with respect to time. The integrals were normalized with respect to the initial value and plotted versus time.

Results

R^2 SELTICS Method Development. Since there was considerable concern that any side band suppression technique would influence the R^2 magnetization rate extensive testing was done to prove that this was not the case. The model compounds studied were chosen, because they had been well characterized and because they had dipolar coupling constants at either extreme of the scale (2,300 and 58.9 Hz for ZnAc and TEE, respectively). The differences observed for ZnAc at short mixing times i.e. from 0.0 to 10.0 ms, between the exchange curves with and with out SELTICS were all within the noise inherent in each data set (data not shown). For low values of **n** in $n\omega_r$, the exchange profiles are similar except at long mix times where SELTICS data tended to continue exchanging for a period of time after R^2 without SELTICS had ceased to exchange. This trend continued and became more pronounced at higher **n** values in ZnAC where the rotational sidebands are increasingly more intense (data not shown). The differences are more apparent in the low coupling constant TEE sample where there is a much larger side band overlap.

A typical example of this lowering of the magnetization exchange asymptote and lengthening of the mixing time required to reach the asymptote for R^2-SELTICS is found in Figures 4-2 and 4-3 for TEE at the **n**=2 condition. In the R^2 with no sideband suppression (Figure 4-2) the magnetization exchange profile exchanges more rapidly than the simulation crossing the simulation line and reaching an exchange asymptote at 30 ms mixing time. The R^2 data acquired using SELTICS does not deviate from the numerical analysis value even at a mixing time as long as 40 ms. The result is an increase in the precision of the distance measurement. These data, which were acquired using R^2 SELTICS, are the most precise distance measurement for TEE obtained to date.

Distance Measurement in Glutamine Synthetase. Because the chemical shift anisotropy for ^{31}P is large (generally on the order of 15 KHz) overlapping sidebands are a severe problem especially when the R^2 condition is matched. For the typical GS sample,

the signal intensity for the unbound inhibitor, PPT, and the phosphate present from ADP and ATP was 10 to 20 times that of the bound PPT. The result was spinning sidebands for the phosphate peak equivalent in intensity to the bound inhibitor peak (Figure 4-4). However, with the use of SELTICS before acquisition the spinning sideband intensity was less than one percent of the center band (Figure 4-4). Eight mixing times from 0.0 to 2.0 ms at the **n**=1 R^2 condition were acquired, four of which are shown in Figure 4-6. Comparing these spectra (Figure 4-6) with the off R^2 spectra found in Figure 4-5 shows that the bound PPT signal decreases over time, but only when the R^2 condition was matched. Spectra were acquired at longer mix times (5, 10 and 15 ms) to check to make sure there were no magnetization exchange oscillations and for complete exchange. No oscillations were observed (data not shown) so it was assumed that complete exchange occurs between 1.75 and 2.0 ms.

Since the zero quantum T_2 (T_{2ZQ}) can be estimated using line width, we needed to determine if the line broadening observed was due to relaxation, T_{2ZQ}, or to chemical shift inhomogeneity. This was checked by applying a highly selective pulse to the unbound PPT peak. The result was a hole burned in the resonance line which is indicative of line broadening due to inhomogeneity (data not shown). This inhomogeneity could be generated either by the magnetic field or by an overlap of many different chemical shifts. Because a phosphorus standard, barium diethyl phosphate (BDEP), measured under similar conditions had a narrow line width, 93 Hz, it is safe to assume that inhomogeneity in the magnetic field was not the cause of the broad lines observed for GS. Thus line broadening in GS was the result of chemical shift inhomogeneity. Since T_{2ZQ} is one of the experimental parameters required for the numerical simulation analysis some estimate of this value had to be made. We chose twice the value of BDEP, 186 Hz, as starting value. The final value used for the simulations was 860 μs which corresponds to a line width of 195 Hz. This value is significantly different from the previous R^2 study of D-Ala-D-Ala ligase [11] where T_{2ZQ} corresponded to a line width of 37 Hz. Since the line width

measured for the standard BDEP sample was much larger than 37 Hz a comparison of the T_{2ZQ} between these two experiments is not appropriate. The T_{2ZQ} value used in the numerical simulations was checked by performing a covariance analysis of dipolar coupling versus T_{2ZQ}. This value, 195 Hz, lies within the region of minimum RMS deviation between the experimental and simulated exchange (data not shown).

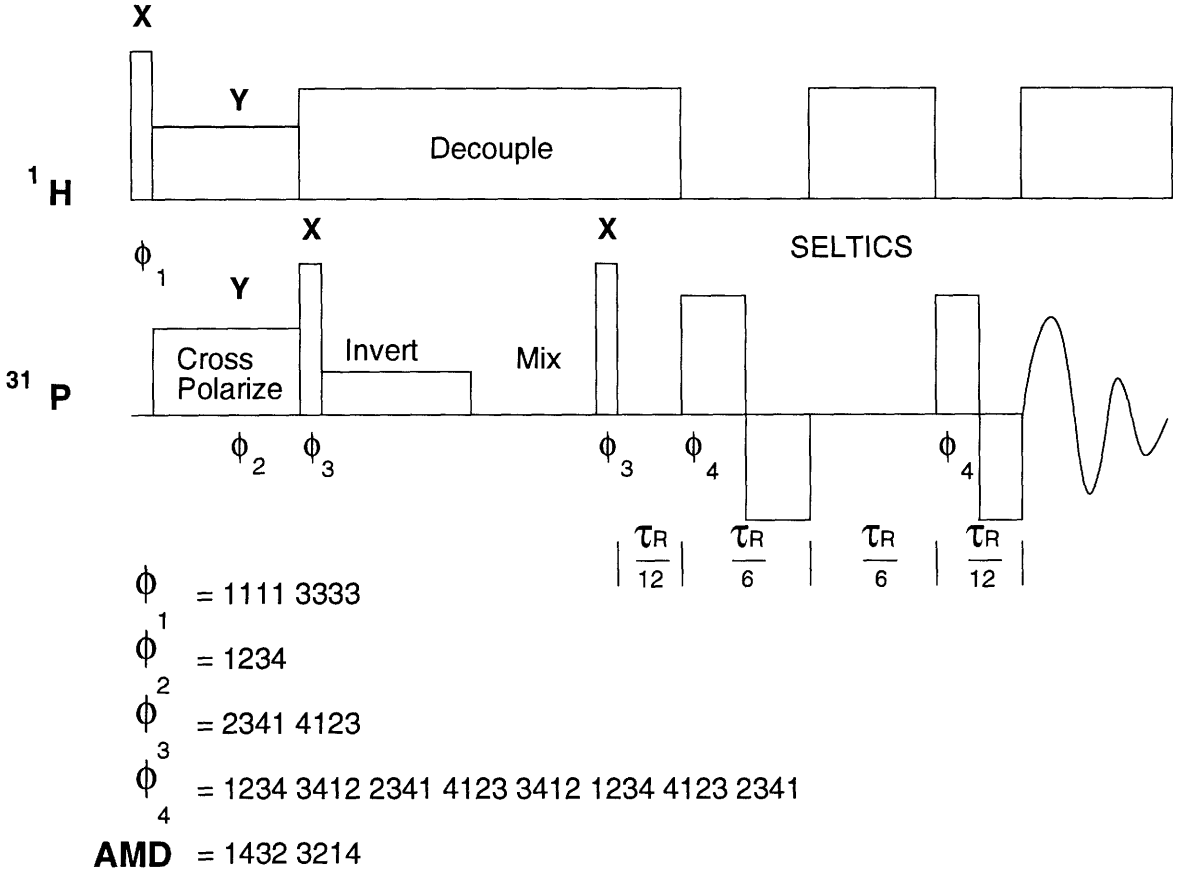


Figure 4-1: Pulse sequence and transmitter-receiver (ϕ, AMD) phase modes for the R^2 SELTICS experiment. The transmitter modes 1,2,3,4 represent the phases 0,90,180,270° or $x, y, \bar{x}, \bar{y}$. The receiver modes 1,4,3,2 are defined as 0,90,180,270° or $x, y, \bar{x}, \bar{y}$. The variable τ_R is the time required for one rotor cycle in seconds.

The anisotropic chemical shift tensor and principle crystallographic orientation values are also required information for numerical analysis. We were unable to measure the tensors because the lines in the phosphate region of the spectrum were not resolved and because the bound PPT peak was weak. Consequently, we used the published value for a similar inhibitor-enzyme system [11]. The anisotropic chemical shift values were scaled for field and the principle crystallographic orientation values were used directly. Because we were matching the R^2 condition at n=1 there was a smaller orientation dependence than when if the n=2 or 3 conditions were used. Consequently, the accuracy of these parameters need not be high. One of the anisotropic chemical shift parameters, η, was tested versus dipolar coupling using covariant analysis. The value used for η lies in the region of minimum RMS deviation between experimental and simulated. Covariant analysis reveals that the data is dependent on dipolar coupling but independent of η. A variation of as much as 20% for η is still within the minimum RMS region (data not shown).

The experimental magnetization exchange values and numeral analysis curves are found in Figure 4-7. The dipolar coupling constant between the phosphinate of PPT and a β-phosphate species is 1000 Hz. With the exception of the 0.25 ms mixing time the data agrees well with a distance of 2.7 ± 0.2 Å. Since the coupling is between the phosphinate and chemical shift typical for β-phosphate and the distance between the species is 2.7 Å, the phosphinate of PPT is linked via a P-O-P bond to β-phosphate.

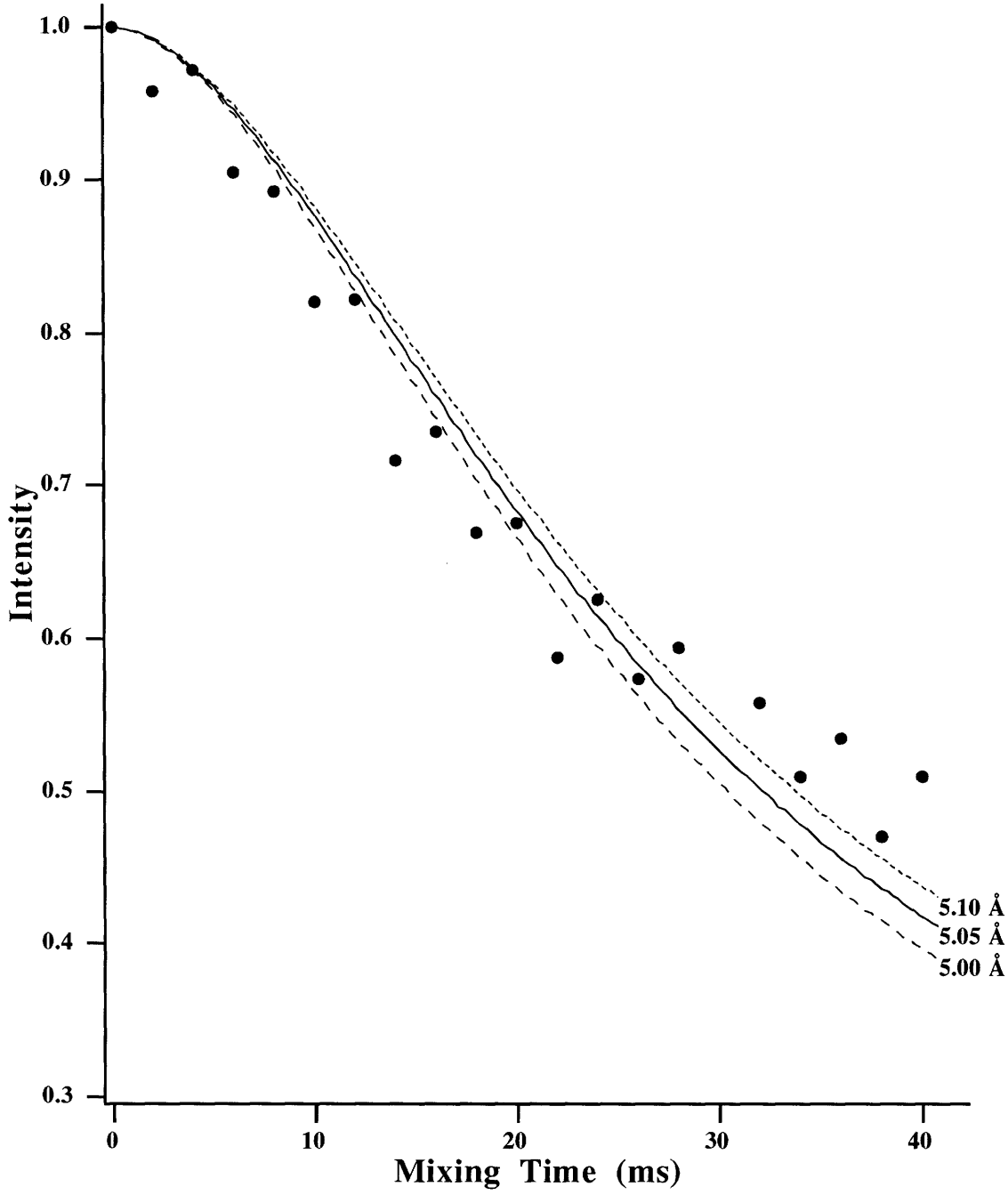


Figure 4-2: Calculated and experimental magnetization exchange profiles consisting of normalized peak area differences of the center bands for ^{13}C in TEE at the R^2 condition of **n**=2, $n\omega_r$ and no side band suppression. The upper and lower bounds correspond to 5.10 and 5.00 Å.

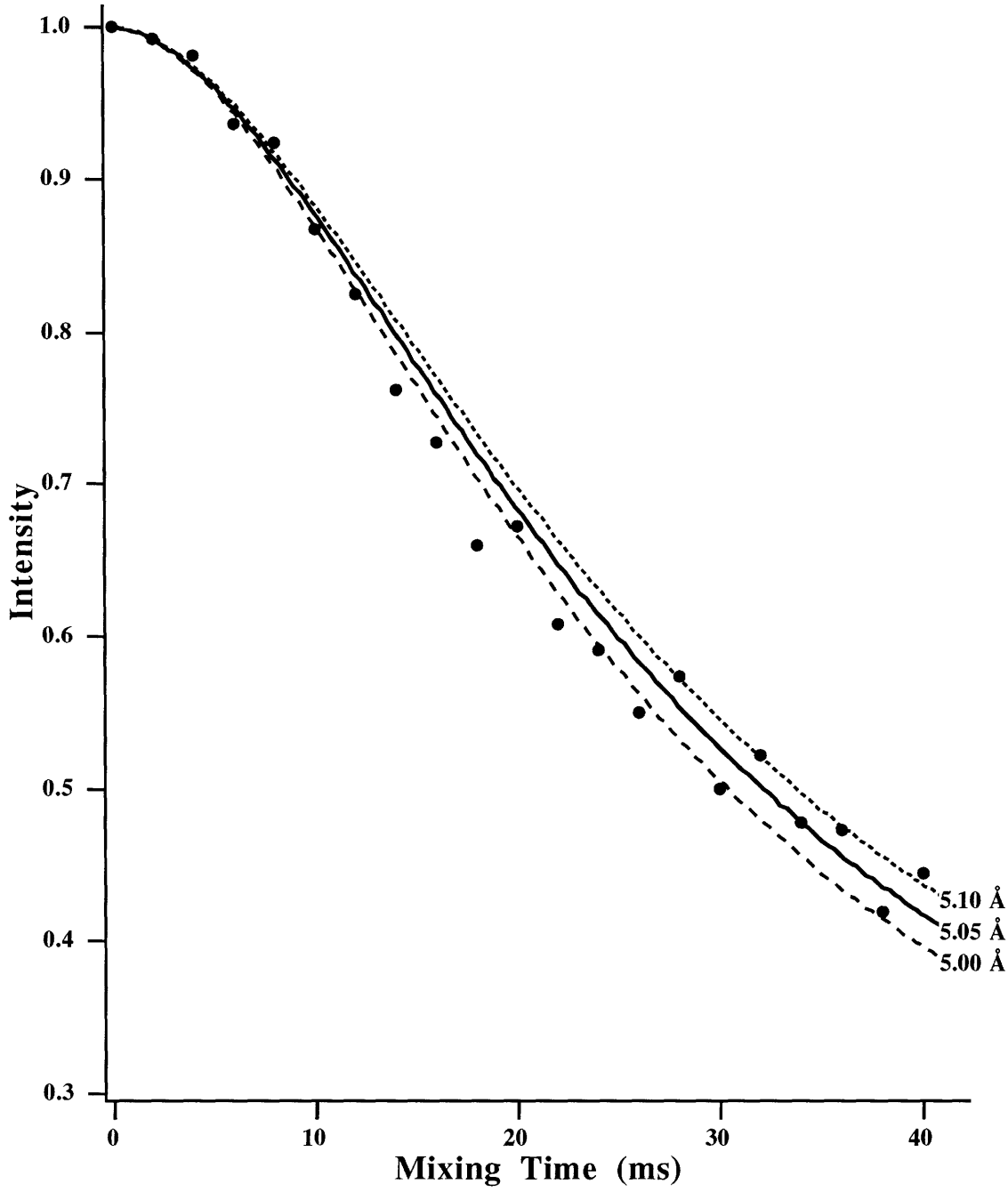


Figure 4-3: Magnetization exchange profiles for the normalized center band integral differences of TEE at the R^2 condition of **n**=2 using SELTICS to suppress the sidebands. The upper and lower bounds correspond to 5.05 ± 0.05 Å. The data show excellent agreement through out the mixing time displayed.

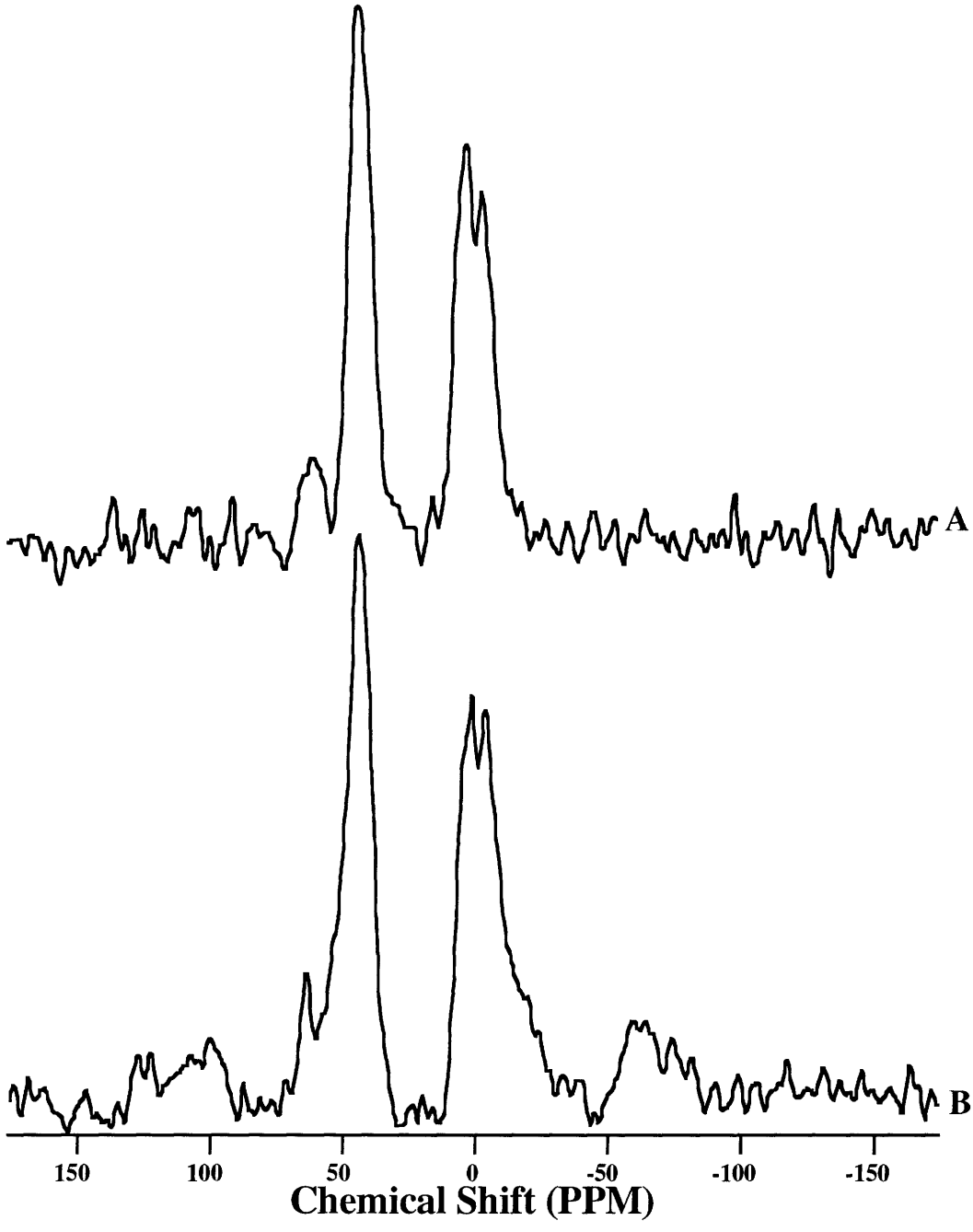


Figure 4-4: ^{31}P MAS NMR spectra of PPT inactivated GS. Spectrum A is a CPMAS spectrum with SELTICS sideband suppression and a spinning speed of 5.41 KHz. Spectrum B is a CPMAS spectrum with no sideband suppression and a spinning speed of 5.1 KHz. Chemical shifts for spectrum A are PPT-P 60.1, PPT 42.0, β-PO_4 1.48, and ADP α-PO_4 -4.44 ppm. For spectrum B the chemical shifts are PPT-P 63.1, PPT 42.3, β-PO_4 0.30, and α-PO_4 -5.03.

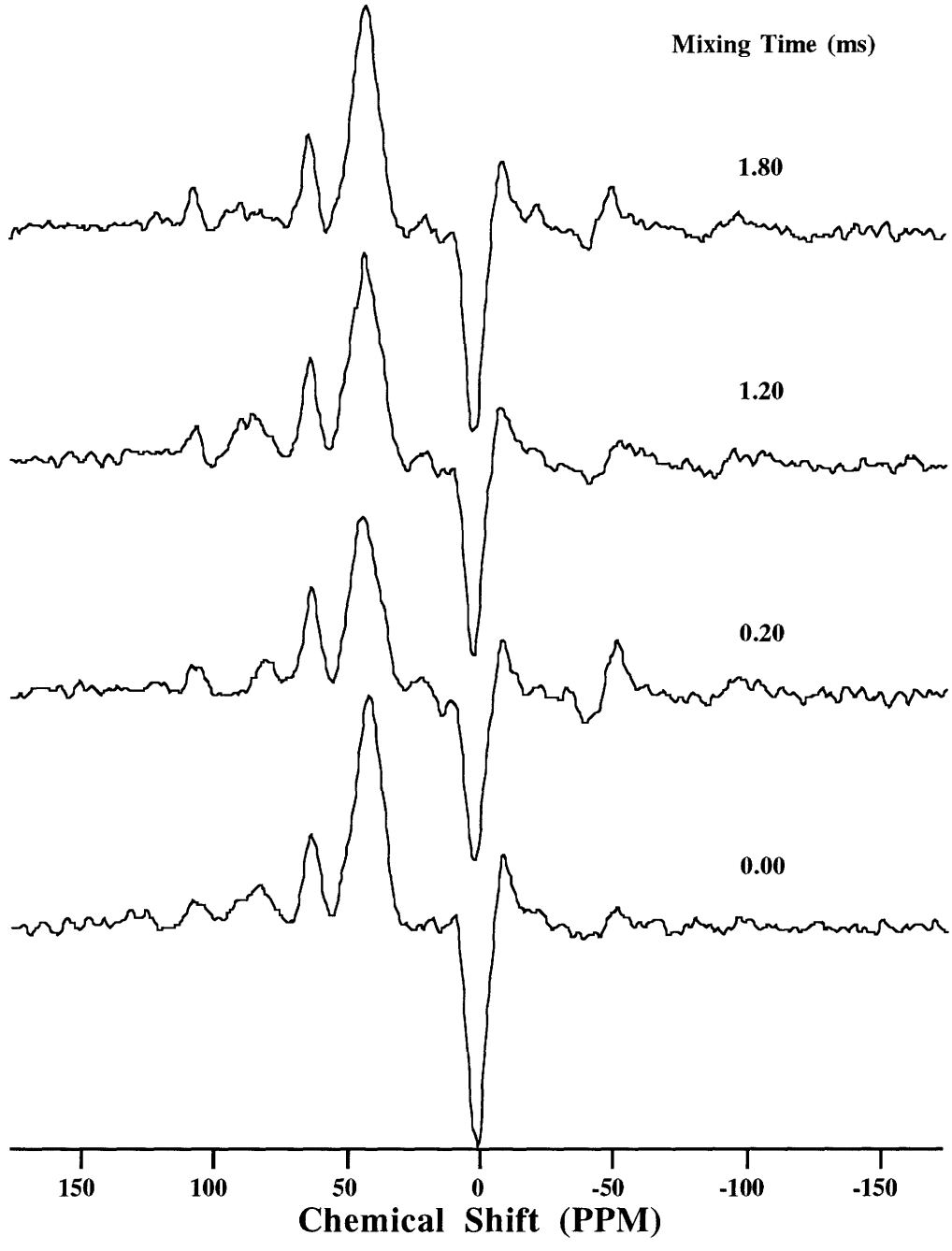


Figure 4-5: MAS ^{31}P NMR off R^2 spectra from 0.0 to 1.8 ms mixing time. The spinning speed was 5.1 KHz and no sideband suppression was performed.

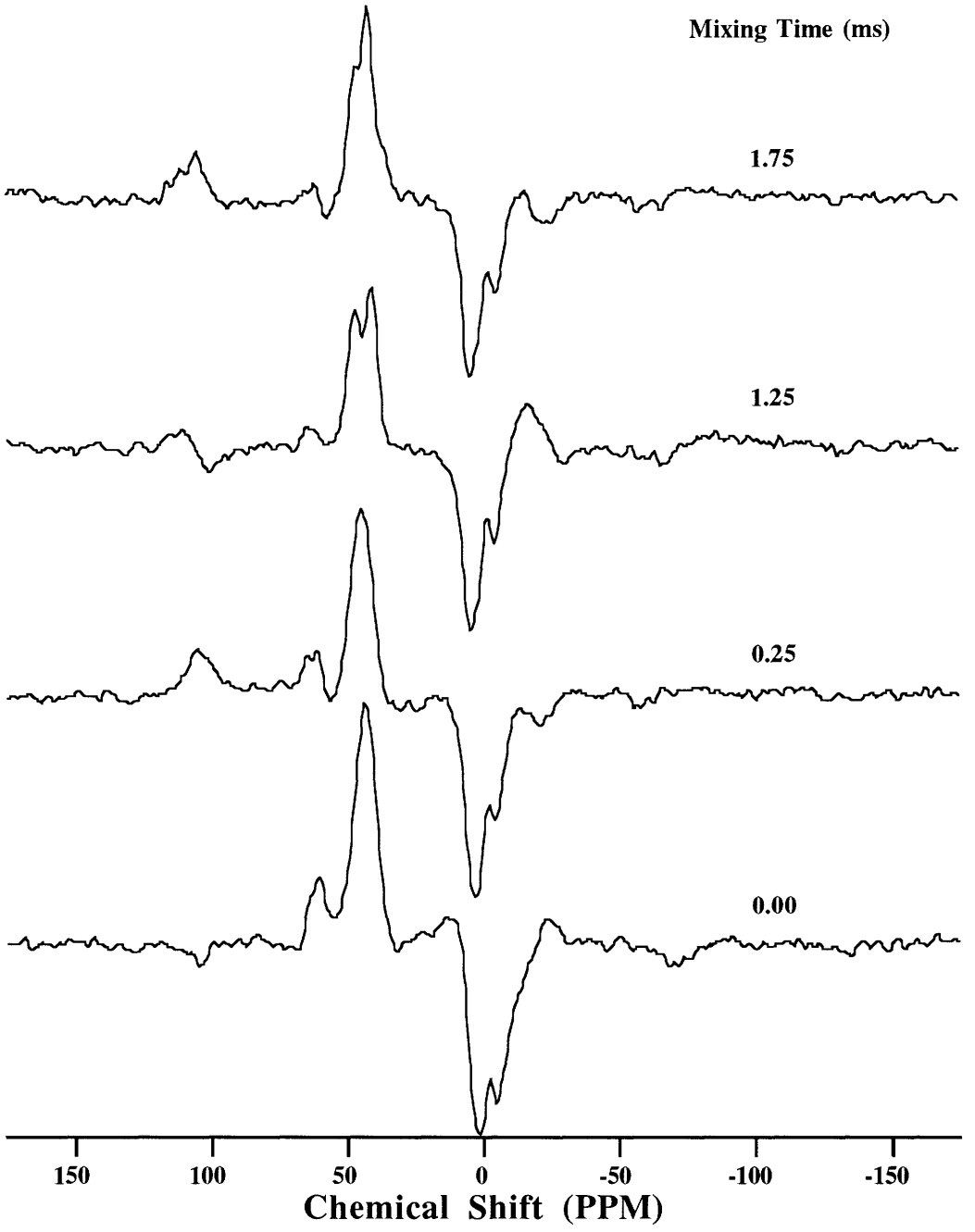


Figure 4-6: ^{31}P MAS R^2 SELTICS spectra of PPT and GS. The **n**=1 R^2 condition was matched requiring a spinning speed of 5.41 KHz. The bound PPT signal exchanges with the negative β-PO_4 peak over time until the peak intensity is zero at 1.75 ms mixing time.

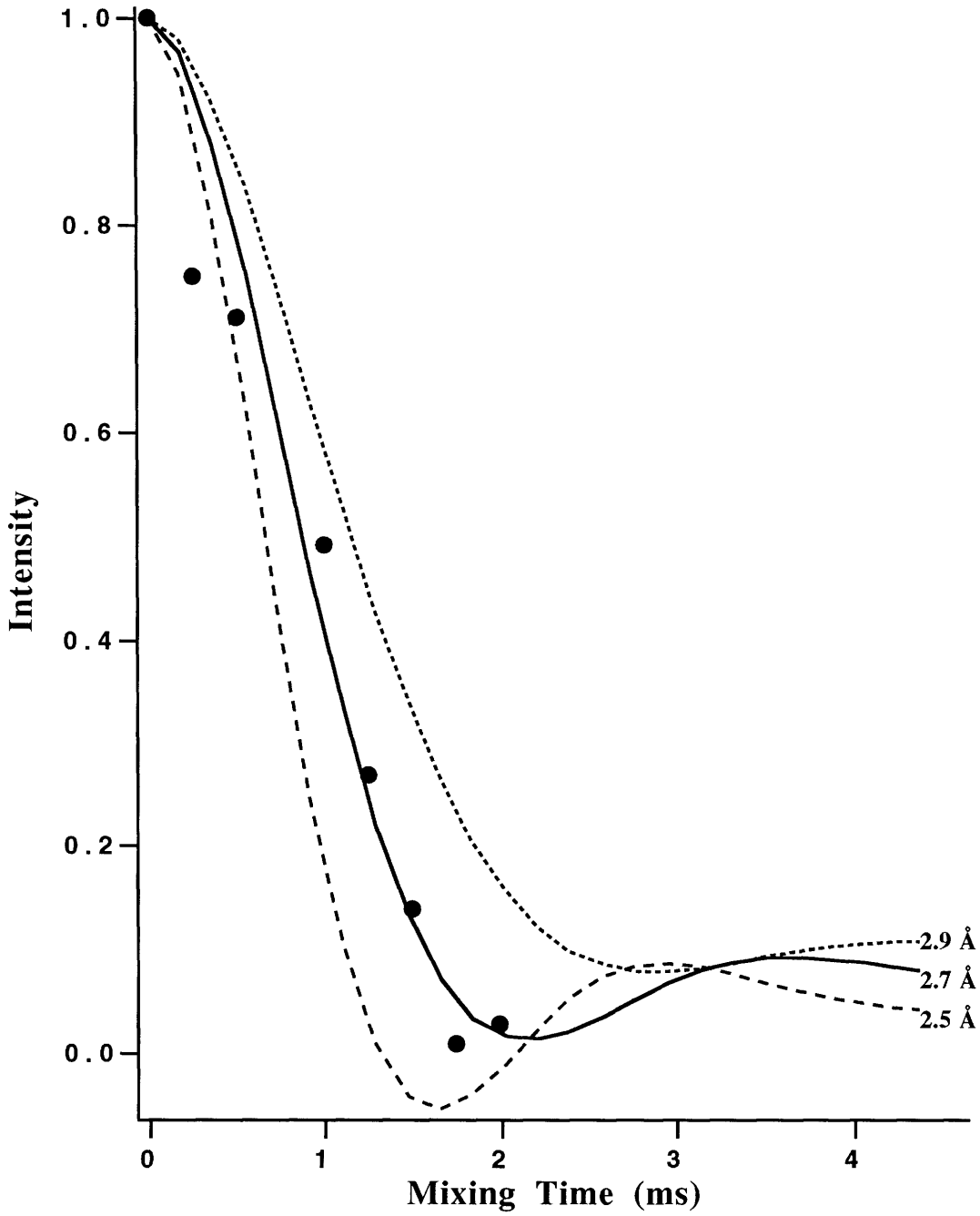


Figure 4-7: Experimental and calculated magnetization evolution profiles for the bound PPT-GS ^{31}P-^{31}P spin pair. The upper and lower bounds correspond to 2.7 ±0.2 Å. The spin pair dipolar coupling was 1000 Hz and T_{2ZQ} 860 μs.

Discussion

We have introduced two modifications to the R^2 technique in this chapter. The first and most important is the use the side band suppression technique SELTICS to remove sidebands from species with large anisotropic chemical shift which would bury low intensity coupled signals. In the case of GS where the β-PO_4 sidebands are equal in intensity to the bound PPT peak it was essential that the side bands be suppressed. There are other methods of side band suppression, i.e. TOSS [16,17], however SELTICS is the only technique which mixes the sidebands back into the centerband instead of using dispersive averaging as does TOSS to remove the sidebands. In addition SELTICS is a much shorter sequence, requiring only half a rotor cycle to complete, the result of which is a smaller loss in signal. Biological macromolecular systems suffer from weak signals, thus the use of a suppression technique which involves a large loss in signal is unacceptable. In our hands R^2 SELTICS suffered from a 46% signal loss over the traditional R^2 experiment. This results in data acquisition times which are 2 to 3 times longer than if R^2 without sideband suppression was performed. Any sideband suppression technique involving a larger signal loss would make distance measurements using R^2 as the dipolar coupling technique unfeasible. Using R^2 SELTICS we have been able to improve the precision of the R^2 technique by an order of magnitude from ± 0.5 to ± 0.05 Å for TEE [9]. In cases where a large anisotropy overlap exists, we have shown that it is vital to address the problem of damping of the magnetization exchange caused by overlapping sidebands.

The second modification to the currently published R^2 technique is the use of one site analysis. For this study the magnetization exchange profile was obtained by examining the change in the area under the bound or PPT-P peak with respect to time. The use of one site analysis is not a novel idea as attempts to use it have been studied in the past. However, most of these attempts determined that the change in the bulk magnetization of the sample,

as measured by difference in the integrals of the two coupled resonance lines (intensity difference), remained the same for both R^2 and the off R^2 data before attempting to use it. Though the intensity difference remained within ±5% of the initial value for the GS-PPT sample, the contribution of the bound PPT peak on the intensity difference parameter was less than the variability seen for this parameter in the off R^2 data. This raised questions as to whether for this system the conventional method of determining the bulk magnetization present was appropriate. Since the magnetization of the β-phosphate region remains within 5% of the initial value for even long mixing times (up to 25 ms, data not shown) for both on and off R^2 it is reasonable to assume that the bulk magnetization is constant through out the experiment.

We have observed PPT-P bound to GS and have been able to answer questions about this complex previously unanswerable in solution. We have observed a dipolar coupling of 1000 Hz in PPT ATP inactivated enzyme between the phosphinate and phosphate peaks. This coupling has been simulated and corresponds to a distance of 2.7 ±0.2 Å which is indicative of a P-O-P ester bond. This distance measurement demonstrates, *in situ*, that PPT is phosphorylated by *E. coli* GS to produce a PPT-P inactivated enzyme.

We also demonstrate the feasibility of SSNMR, in particular the novel technique of R^2-SELTICS, for studying structure in high MW (~600kD) proteins. This is the first such study performed to date. R^2-SELTICS opens for study, by SSNMR, a class of proteins never before studied by solution or solid state NMR. It is now possible to ask specific structural questions about proteins of molecular weights up to 600 kD. This expansion of the molecular weight feasibility limits makes accessible for study all except the very largest protein complexes. It also opens the way for work on a class of proteins relatively unstudied structurally, namely insoluble or membrane bound proteins and protein complexes.

References

(1) Schineller, J. B.; Villafranca, J. J., Personal Communication.

(2) Logusch, E. W.; Walker, D. M.; McDonald, J. F.; Franz, J. E.; Villafranca, J. J.; DiIanni, C. L.; Colanduoni, J. A.; Li, B.; Schineller, J. B. *Biochemistry* **1990**, *29*, 366-372.

(3) Colanduoni, J. C.; Villafranca, J. J. *Bioorg. Chem.* **1986**, *14*, 163-169.

(4) Ronzio, R. A.; Rowe, W. B.; Meister, A. *Biochemistry* **1969**, *8*, 1066-1075.

(5) Andrew, E. R.; Clough, S.; Farnell, L. F.; Gledhill, T. D.; Roberts, I. *Phys. Letters* **1966**, *21*, 505-506.

(6) Raleigh, D. P.; Harbison, G. S.; Neiss, T. G.; Roberts, J. E.; Griffin, R. G. *Chem. Phys. Letters* **1987**, *138*, 285-290.

(7) Raleigh, D. P.; Levitt, M. H.; Griffin, R. G. *Chem. Phys. Letters* **1988**, *146*, 71-76.

(8) Colombo, M. G.; Meier, B. H.; Ernst, R. R. *Chem. Phys. Letters* **1988**, *146*, 189-196.

(9) Raleigh, D. P.; Creuzet, F.; Das Gupta, S. K.; Levitt, M. H.; Griffin, R. G. *J. Am. Chem. Soc.* **1989**, *111*, 4502-4503.

(10) Levitt, M. H.; Raleigh, D. P.; Creuzet, F.; Griffin, R. G. *J. Chem. Phys.* **1990**, *92*, 6347-6364.

(11) McDermott, A. E.; Creuzet, F.; Griffin, R. G.; Zawadzke, L. E.; Ye, Q.; Walsh, C. T. *Biochemistry* **1990**, *29*, 5767-5775.

(12) Creuzet, F.; McDermott, A.; Gebhard, R.; van der Hoef, K.; Spijker-Assink, M. B.; Herzfeld, J.; Lugtenburg, J.; Levitt, M. H.; Griffin, R. G. *Science* **1991**, *251*, 783-786.

(13) Lakshmi, K. V.; Auger, M.; Raap, J.; Lugtenburg, J.; Griffin, R. G.; Herzfeld, J. *Biophys. J.* **1993**, *64*, A212-A212.

(14) de Groot, H. J. M.; Copié, V.; Smith, S. O.; Winkel, C.; Lugtenburg, J.; Herfeld, J.; Allen, P. J.; Griffin, R. G. *J. Magn. Reson.* **1988**, *77*, 251.

(15) Hong, J.; Harbison, G. S. *J. Magn. Reson.* **1993**, In press.

(16) Raleigh, D. P.; Olejniczak, E. T.; Griffin, R. G. *J. Chem. Phys.* **1988**, *89*, 1333-1350.

(17) Raleigh, D. P.; Olejniczak, E. T.; Griffin, R. G. *J. Magn. Reson.* **1991**, *93*, 472-484.

CHAPTER 5

Protein Structure Determination By Multidimensional ^{1}H NMR Spectroscopy of Ragweed Allergen *Amb. t. V.*

Introduction

Of the plethora of foreign substances which assault the immune system there are a few (allergens) which elicit a deleterious hypersensitive response. Though exposure is under extremely low-dose conditions the response can be as mild as onset of sneezing and watering eyes or as severe as anaphylactic shock and death. This is called the allergic response. Though a response is elicited, the part allergen structure plays in triggering this detrimental mechanism is unknown. It was suggested many years ago that because the allergen is presented under extremely low-dose conditions, recognition is not an acquired response but occurs only for epitopes already present, i.e. immunodominant epitopes, at the B and T cell level [1]. B-cell epitopes, known more commonly as antibodies or immunoglobulins (Ig), are primarily surface structures making them structurally or chemically species-specific [2]. They bind with high avidity only to native structure [3]. This means that peptides or protein fragments containing the chemical species found in the binding site, but without the native 3-dimensional structure, bind with less avidity. B cell epitopes are generated by clonal selection, e.g. immunoglobulin is generated when an antigen binds to a B cell surface receptor, have a repertoire estimated to be a large as 10^7, and a high specificity being able to recognize single atom differences [3,4]. There are several types of B cell epitopes, IgA subclass IgA1 and IgA2, IgD, IgE, IgG with subclasses IgG1, IgG2, IgG3, including the rare IgG4, and last IgM.

These classes are determined by the immunoglobulin heavy chain type. It is of interest to note that allergic reactions are triggered exclusively by the binding of IgE. No other Ig class triggers an allergic response. B cell epitope idiotype is determined by the lymphokines produced by the T helper cell [5,6].

In contrast, "Ia/T cell" epitopes are generally peptide fragments processed by the antigen-presenting cell (APC) from the native protein and presented to the T cell within the groove of the major histocompatibility complex (MHC, transplantation antigen, or Ia) surface protein [7,8]. Genetically encoded in three classes, MHC is referred to as class I, class II, or class III [9] and is involved in the presentation of antigens to the immune system for both the B and T cell systems. MHC is the molecule detected by the immune system for the determination of self and non-self and was first identified as being involved in graft rejection. Allergens are presented in MHC class II molecules. Though the Ia/T cell epitope has a specificity equivalent to the B cell epitope its repertoire is smaller, on the order of 10^3 to 10^4 [3]. For binding avidity, structural conformation appears to be unimportant. Binding has been observed for peptides of the antigen region, protein fragments and native protein [9,10]. A T cell response is elicited when a sufficient number of antigens are bound to and presented by the MHC for recognition by a helper T cell receptor [11,12]. It is not known how the APC processes the native protein so as to generate the Ia/T cell epitope. However, processing must take place as binding of the antigen to the MHC occurs only a high antigen concentration once the MHC is transported to the cell surface. Based primarily on mouse studies which used ovalbumin as the antigen, it has been proposed that amphipathic peptides, primarily amphipathic α-helix, are the preferred Ia/T cell epitopes [13]. In summary, the allergic response is triggered by binding of the B cell epitope (IgE) and is a surface interaction. However before this surface interaction can become an allergy, the selection of the antibody idiotype for IgE by the T/Ia cell epitope and the T helper cell must be made. This selection is based on chemistry; native allergen structure playing a poorly understood role in this process.

The low molecular weight protein allergens *Amb. t.* V and *Amb. a.* V (formerly Ra5G and Ra5S) isolated from *Ambrosia* (ragweed) *trifida* (giant) and *artemisiifolia* (short), respectively, are good models for studying the structure function relationship between B and Ia/T cell epitopes [14-18]. *Amb. t.* V and *Amb. a.* V are small molecular weight (40 residues MW~4300, 45 residues MW~5000 respectively) highly homologous (50% sequence identity) polypeptides [19-21]. They are extremely thermostable [22] (1 hour at 100° C in 50% glycerin) and contain four disulfide bonds. These proteins are not cross-reactive at the B cell level [23] but are recognized by the same HLA-DR2/Dw2-associated MHC class II molecule, HLA-DR(α,β_1*1501), which suggests they have the same Ia-binding epitope [19,24-28]. This implies that the two proteins have different surface structures, but share a piece of internal structure or chemistry which is similar.

Understanding the structure-function relationships between these two proteins and the immune system requires detailed 3D structure information. These proteins do not readily crystallize so the X-ray crystal structure can not be determined. They are, however, highly soluble. With the development of multidimensional NMR techniques, evolution of data analysis protocols, and design of computational algorithms, it has become possible to use ^{1}H NMR to elucidate 3-dimensional structure in solution [29]. We have undertaken the structure determination of *Amb. t.* V using 2 and 3D ^{1}H NMR in an effort to clarify the structural basis and elements involved in the interaction of these proteins with the immune system.

Experimental Procedures

Sample Preparation. A 40 mg sample of highly purified *Amb. t.* V from *A. trifida* isolated and purified, as described previously, was generously provided by L. Goodfriend [19,23]. For the NMR experiments 20 mg each of *Amb. t.* V was dissolved in 600 µL of an NMR buffer (50 mM acetate, 1 mM NaN_3, 11 mM EDTA, pH 4.0 at 20° C) to make two

samples of a final concentration of 7.8 mM. One the samples contained 10 % ^2H_2O included for an internal lock signal. The other was lyophilized, dissolved in 99.8% ^2H_2O and quickly placed in the spectrometer so as to measure amide exchange rates. The fully exchanged sample was prepared as follows. The exchange sample was lyophilized, dissolved in 99.9% ^2H_2O, warmed to 50° C and lyophilized twice, and finally dissolved in 99.996% ^2H_2O. 5-mm NMR tubes model 535-PP purchased from Wilmad Glass (Buena, NJ) were used.

^1H NMR Spectroscopy. ^1H NMR experiments were performed on home built spectrometers operating at 501.9 and 591.1 MHz for ^1H. 2D NMR spectra were acquired with quadrature detection in both dimensions by the hypercomplex method [30] in the phase-sensitive mode. 3D NMR spectra were acquired using the TPPI-States method [31]. For the 2D spectra, water suppression was performed using a low power presaturation pulse for 2.0-s during the recycle delay. The α protons were allowed to recover with the addition of the scuba pulse train just after the presaturation pulse [32]. Most 2D data were acquired as 2048 complex in t_2, and 512 complex FIDs in t_1 over a 6024 Hz sweep width. At least 64 transients were acquired for each t_1 increment. The same sweep width was used for the 3D experiments and 8 transients were acquired per time increment. The 3D acquisition parameters were 1024 complex points in t_3, with 128 complex FIDs in t_2 and 64 complex FIDs in t_1. All data were acquired at 20° C.

For the 2D data "no-diagonal" ND-COSY [33], "primitive" PCOSY [34], "primitive exclusive" PECOSY [35,36], "double quantum filtered" DQFCOSY [37,38], TOCSY [39,40], NOESY [41-43], and "1,1 jump-return" JRNOESY [44-46] experiments were performed following the standard methods. A composite 180° pulse was applied during the middle of the mixing, in the NOESY experiment, to prevent recovery of the solvent peak. Four NOESY mixing times, 50, 100, 150, and 300 ms were acquired. A single mixing time, 150 ms, was used for the JRNOESY experiment. Most resonance assignments and structural constraints were obtained using these data. The spin-locking period during the

TOCSY experiments involved either MLEV-17 [40] or z-filtered DIPSI-2 sequences [47-50]. Two TOCSY mixing times were used, 50 and 80 ms, with a 8.8 KHz average spin lock field.

The 3D homonuclear data were acquired using NOESY-TOCSY [51-54] and TOCSY-NOESY [54-57] experiments. Water suppression was achieved with a low power presaturation pulse for 1.5-s during the recycle delay and the α protons were allowed to recover with the addition of the scuba pulse train just after the presaturation pulse [32]. Both experiments used a 1,1 jump return selective excitation pulse immediately following the NOESY portion of the experiment for an additional improvement in water suppression. These experiments will be referred to as NJRTOC and TOCNJR for NOESY-TOCSY and TOCSY-NOESY, respectively. The pulse sequence for the TOCNJR experiment is found in Figure 5-1. The selective excitation pulse sequence allowed for the use of a weaker presaturation pulse resulting in less bleaching of the α proton resonances near the solvent peak. In the 3D experiments the mixing times were 150 ms for NOESY and 50 ms for TOCSY.

Data processing was done using in-house software written by David J. Ruben on VAX station 4000/90 and VAX station 3100 M38. A low frequency convolution filter was applied to the acquisition dimension (t_2 for 2D and t_3 for 3D) of all data acquired with H_2O as the solvent so as to remove residual water peaks [58]. Ridge artifacts were removed by applying baseline fixing after Fourier transformation in both t_2 and t_1 for 2D experiments. Baseline fixing was applied only to the t_3 transform of the 3D data sets. All 2D data, except the ND-COSY, PCOSY, and PECOSY were linearly predicted to 1024 points in t_1 and then zero filled to 2048 resulting in a 2048×2048 Fourier transformed matrix. An exponential/Gaussian apodization function was applied in both time domains before Fourier transformation. The ND-COSY and PCOSY were linearly predicted to 4096 in t_2 and 1024 in t_1 then Fourier transformed into a 4096×4096 matrix by zero filling. A sine bell apodization function was applied to both time domains

before Fourier transformation. All 3D data was linearly predicted to 256 points in t_2 and 128 point in t_1 and Fourier transformed into a 1024 × 512 × 256 matrix by zero filling. A Gaussian apodization function equal to 1/(acquisition time) in Hz was applied to all three time domains before Fourier transformation.

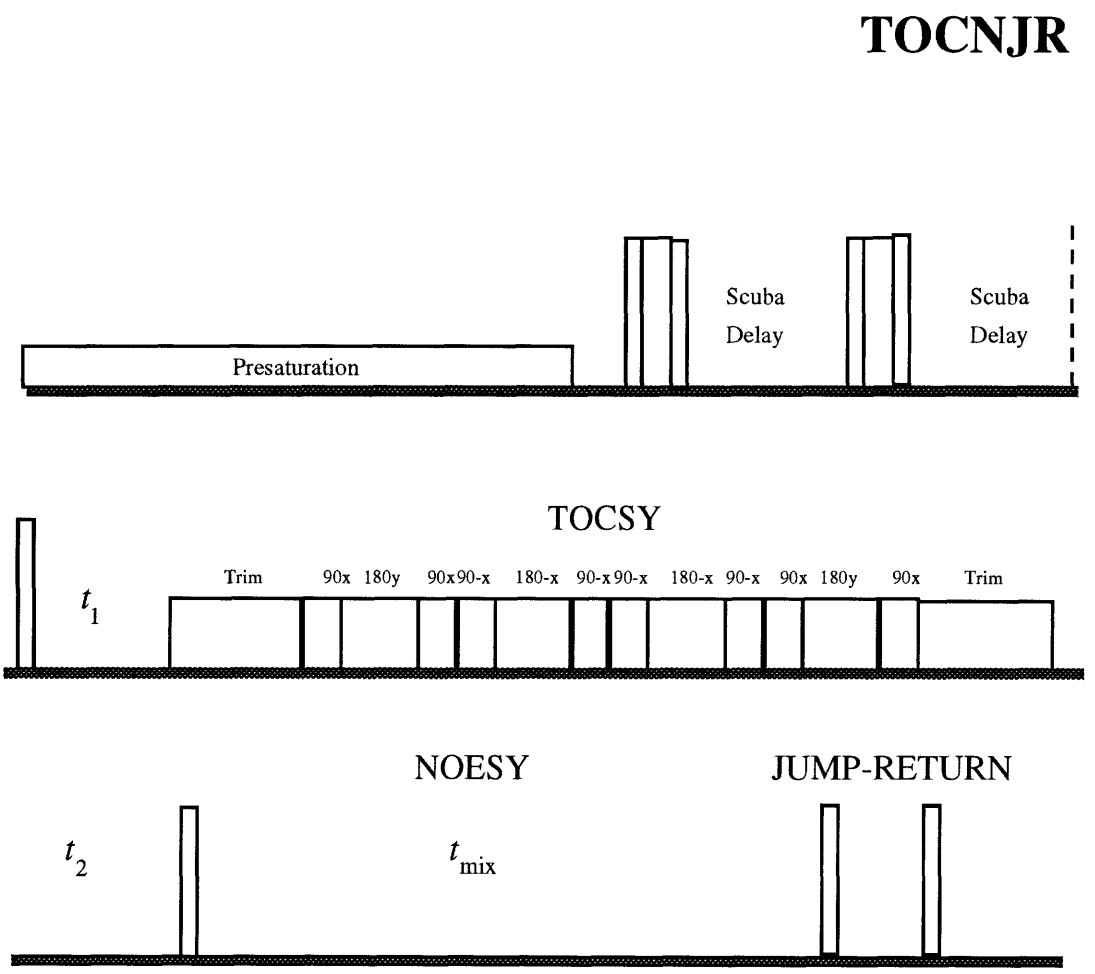


Figure 5-1: Pulse sequence for the 3D ^{1}H NMR TOCSY-NOESY experiment. This sequence contains a 1,1 jump-return selective excitation pulse instead of a 90° read pulse at the end of the NOESY experiment for improved solvent suppression.

Resonance Assignments. Resonance assignments were made in the manner described by Wüthrich [29]. This involved the classification of spin systems using 2D ND-COSY,

DQFCOSY, and PCOSY experiments. These were verified using 2D TOCSY and the TOCSY portion of the 3D TOCNJR, NJRTOC experiments. Sequence specificity was determined using medium mixing time (150 ms) NOESY data from both 2 and 3D. NOESY data were also used to identify intraresidue assignments for the δNH protons of asparagine and the aromatic protons of tyrosine. This resulted in assignment of 236 intraresidue NOE cross peaks.

For the 3D experiments, TOCSY assignments were obtained from $\omega_1 \times \omega_3$ cross-sections along ω_2 axis of both the NJRTOC and TOCNJR experiments, while sequential and nonsequential NOESY assignments were obtained from observing $\omega_2 \times \omega_3$ planes along the ω_1 axis of the TOCNJR experiment (see Figures 5-2, 5-3, 5-4, and 5-5). The $\omega_1 \times \omega_2$ cross sections were not extensively used in either of the 3D experiments because signal from the solvent peak tended to reside in these cross-sections. They were checked to confirm what was observed in the $\omega_1 \times \omega_3$ and $\omega_2 \times \omega_3$ cross-sections (Figure 5-5). The convention used has ω_1 as the acquisition dimension corresponding to t_3, with ω_2 as the slow dimension corresponding to t_2, and ω_3 as the frequency of very slow dimension with the related time domain being t_1.

A procedure similar to the one described by Kline et al. [59] was used to assign the nonsequential NOE cross-peaks. Approximately 60% of the cross peaks were assigned unambiguously based on unique chemical shifts present in uncrowded regions of the spectrum. The distance constraints were calculated using cross peak intensities and preliminary structures calculated. The remaining crosspeaks were assigned using the structures generated in an interactive process as described below. This procedure led to the identification of 80 sequential and 100 nonsequential NOE cross peaks.

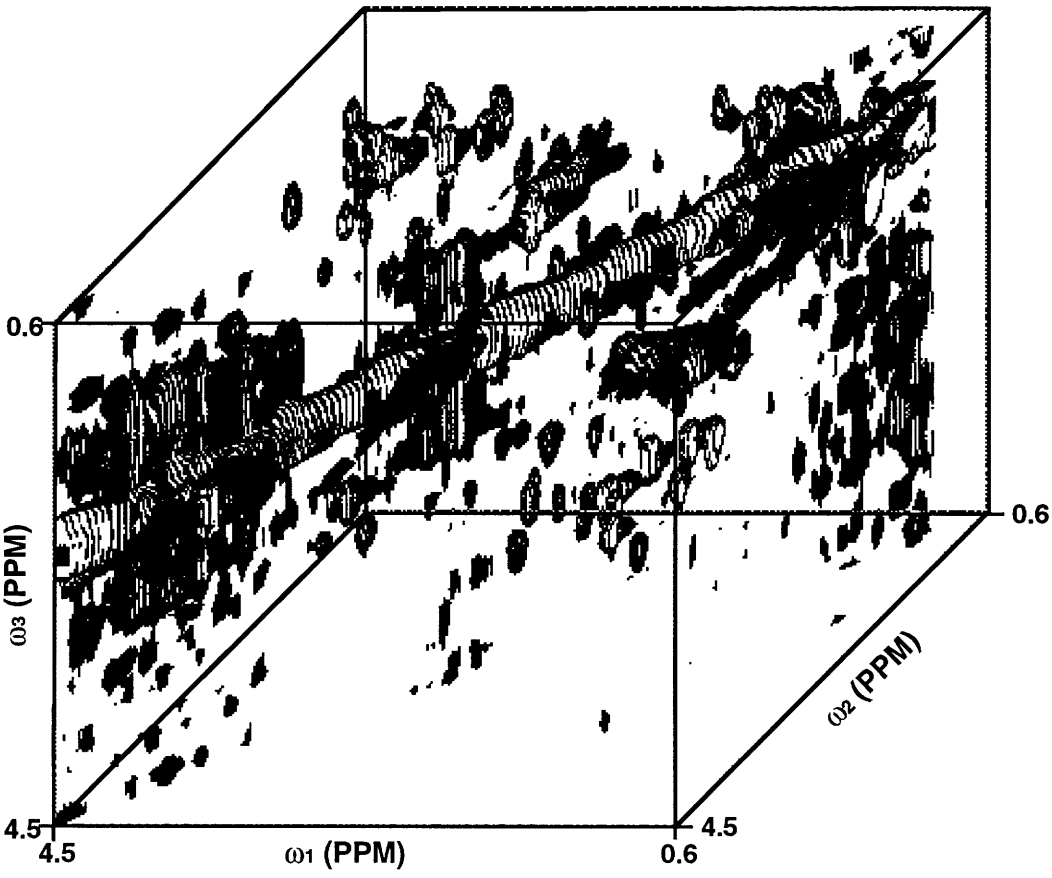


Figure 5-2: The aliphatic region of a 3D ^{1}H NMR NOESY-TOCSY experiment in D_2O at 20° C. The NOESY mixing time was 150 ms and the TOCSY mixing time 80 ms. The TOCSY spin lock mixing sequence used was DIPSI-2 [49]. The processed data matrix size was 512 X 256 X 32.

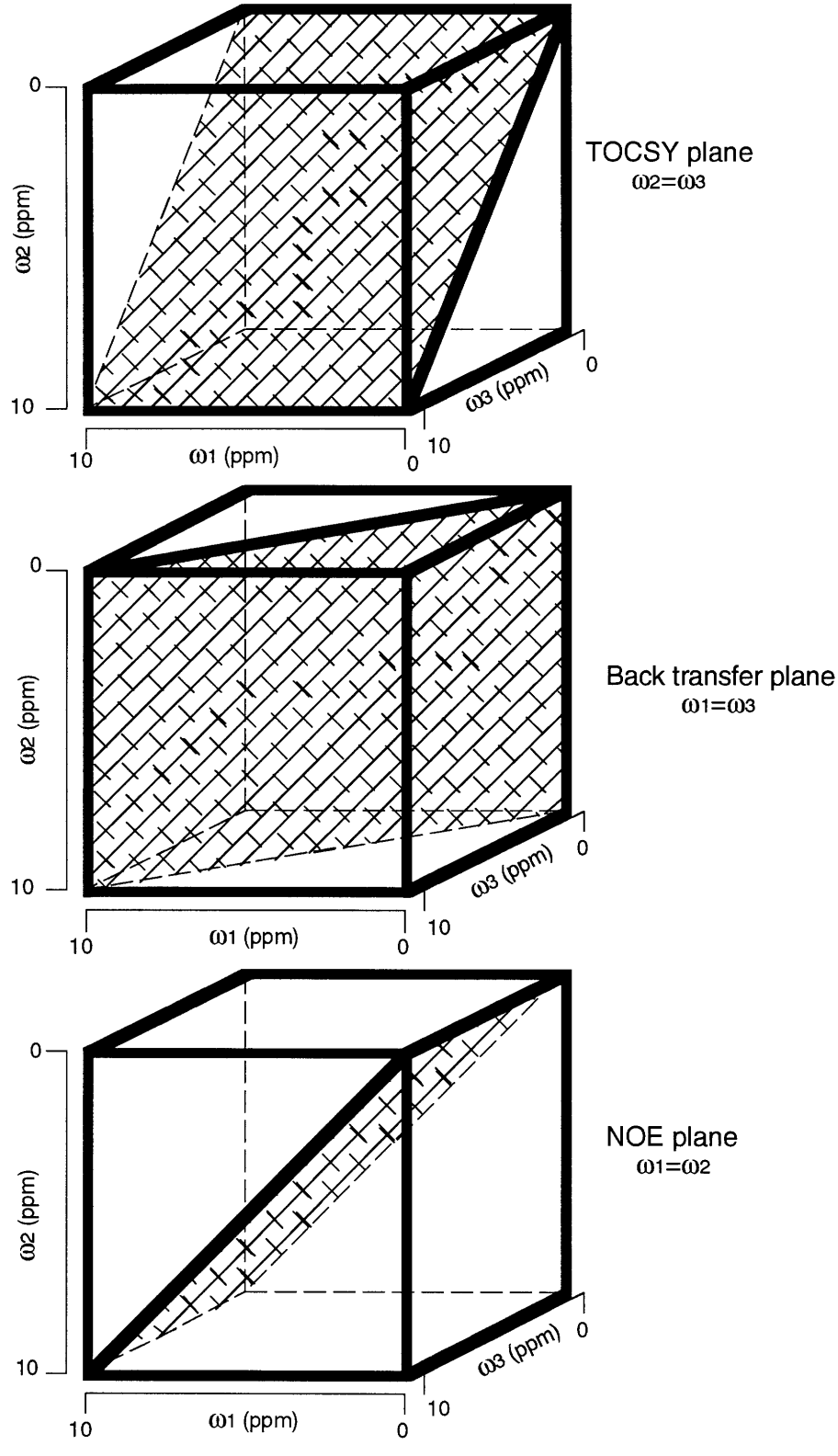


Figure 5-3: A diagram illustrating the magnetization transfer planes for the 3D TOCSY-NOESY experiment.

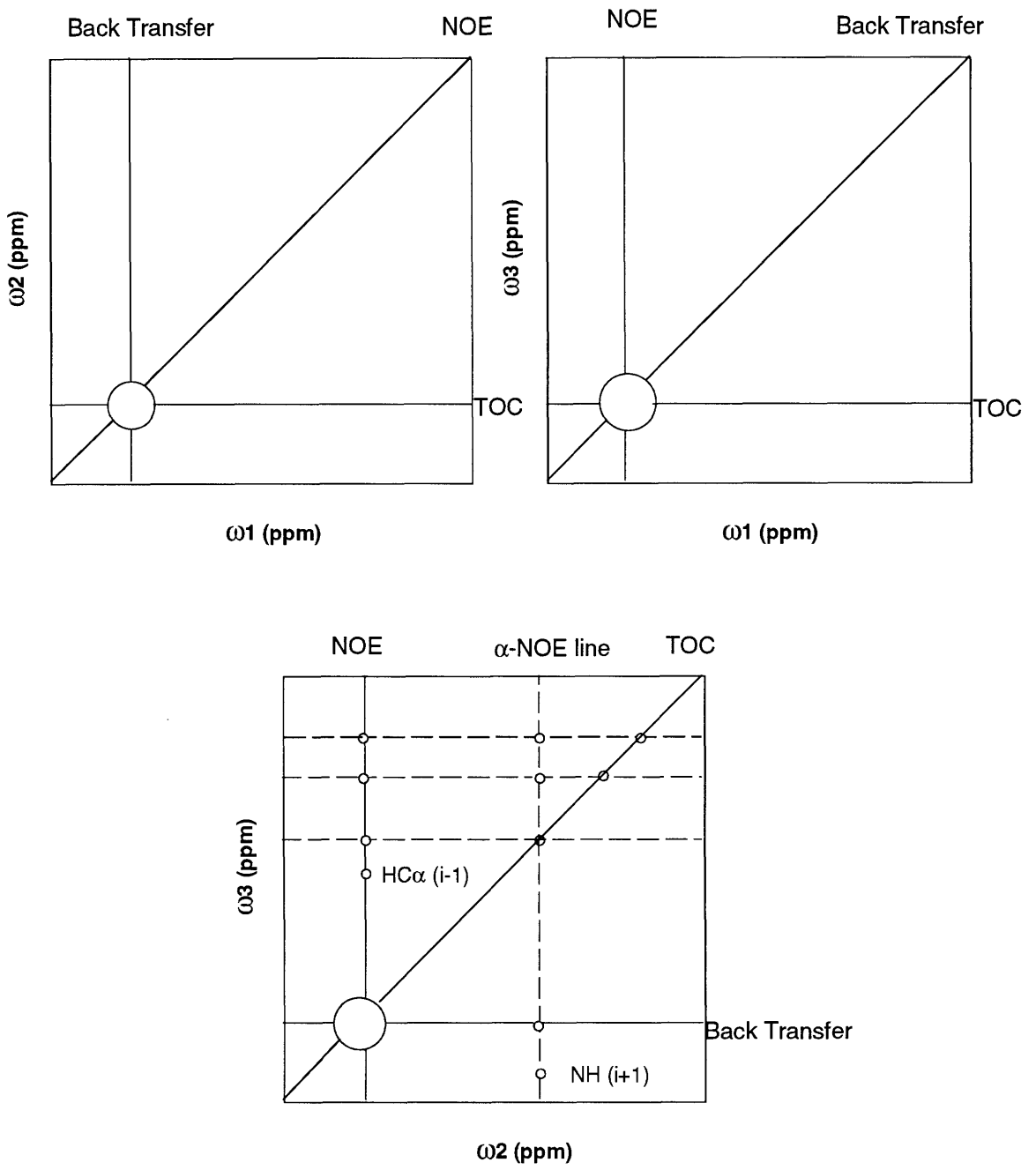


Figure 5-4: A schematic diagram illustrating the magnetization transfer pathways as observed in 2D planes from a 3D TOCSY-NOESY experiment. The $\omega_2 x \omega_3$ plane illustrates a one possible sequential assignment strategy.

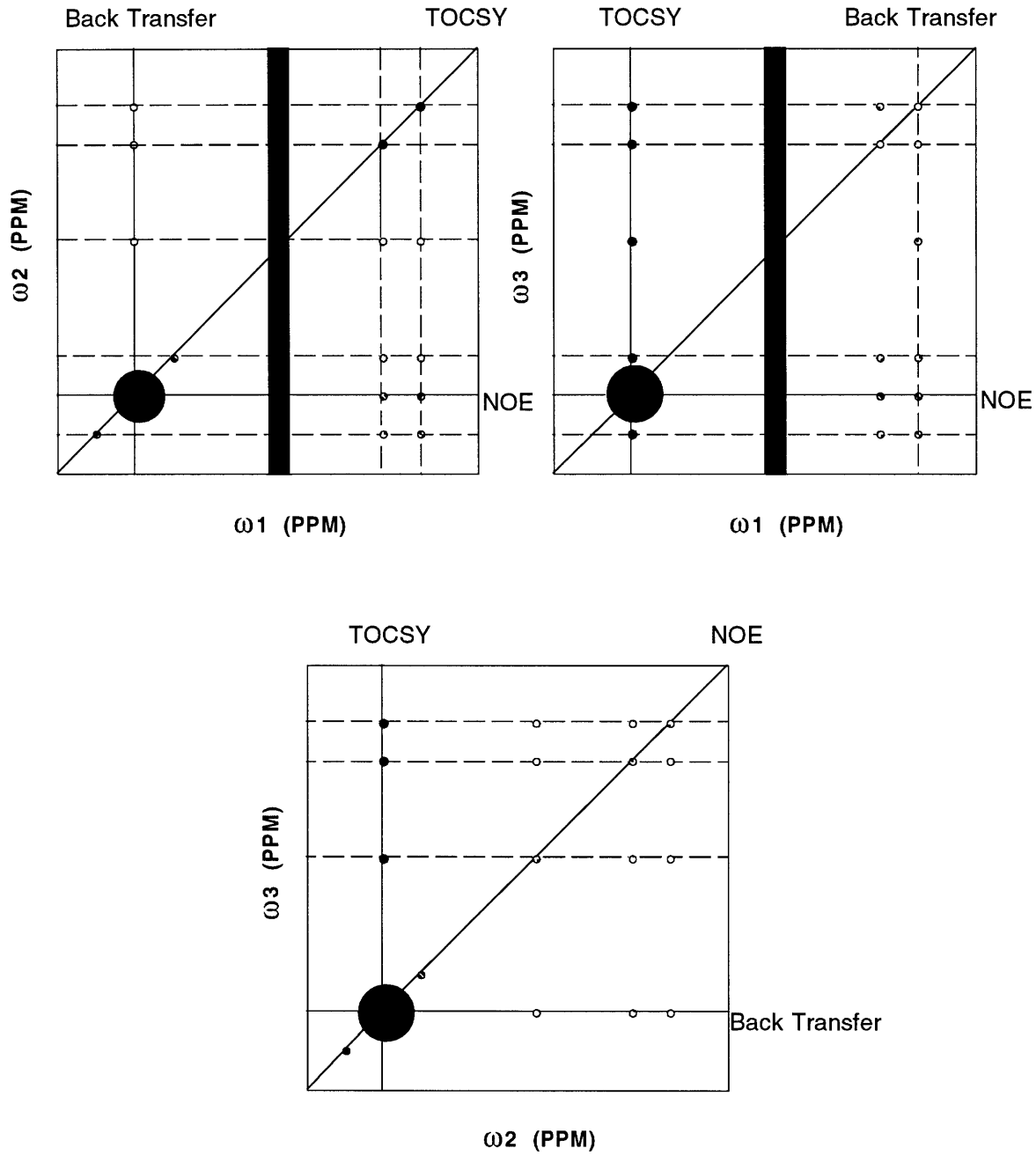


Figure 5-5: A diagram illustrating the magnetization transfer pathways for the ^{1}H 3D NOESY-TOCSY experiment. The cross peak patterns shown are the patterns expected for α-helical secondary structure. The dark line in the upper two spectra represents the residual solvent peak and the filled circles indicate a TOCSY transfer pathway. The peaks above and below the NH resonance in the $\omega_1 \times \omega_3$ slice represent the NHNH (*i*, *i* +1) and (*i*, *i* - 1) cross peaks. It should be noted that there is no residual solvent peak observed in the $\omega_2 \times \omega_3$ plane.

Structural Constraint Generation. There are three types of structural constraints that can be obtained from NMR data: interproton distance constraints from NOE data, torsion angle constraints from J-coupled spectra, and hydrogen bond constraints from exchange data.

Initial distance constraints were calculated using resolved β-methylene protons based on the isolated two spin model at a NOESY mixing time of 150 ms. These constraints were further refined using a number of intensities from protons present in the antiparallel β-sheet. The NOE cross peak volumes were not calculated for the 3-D data. Instead it was assumed that the peak amplitude and volume were linearly related. Since the digital resolution in the NOE plane was low it was assumed that the multiplicities did not introduce large errors into this relationship. The experimental intensities were comparable to calculated intensities. The resulting distances were divided into three classes, strong between 1.8 and 2.7 Å, medium from 1.8 to 3.4 Å, and weak from 1.8 to 5.0 Å [60-62]. NOEs observed at 300 ms mixing time, but not at 150 ms, were initially included in the weak class to prevent distance misassignment due to spin lattice diffusion. Distances were refined in latter structures. For the tyrosine aromatic and methyl group protons 0.6 and 1.0 Å were added to the upper distance constraints, respectively [61-63]. The increase in the distance constraint is due to the multiple unresolved protons (three in the case of the methyl and two for the tyrosine aromatic) involved in exchange which produces measurable exchange at longer distances than for a single spin pair.

The backbone torsion constraint Φ was obtained by measuring the value of $^3J_{HN\alpha}$ using high resolution DQF-COSY and PCOSY experiments for 19 residues. The Φ constraint ranges were -120° ± 40° for $^3J_{HN\alpha} > 8.0$ Hz, and -60° ± 40° for $^3J_{HN\alpha} < 4.5$ Hz. These constraints were not used in the structure calculation, but instead were used to test the final structures. Stereochemical assignments of the β-methylene protons were made using NOE intensities from HN-β and α-β at 150 ms mixing time [61] and confirmed by measuring $^3J_{\alpha\beta_1}$ with PECOSY. This resulted in the stereochemical assignment of 11

nondegenerate β-methylenes and the determination of χ_1 angle constraints (either 60°, -60,° or 180° ± 40°). These constraints were included in the final structure calculations.

Hydrogen bond constraints were not included in the structure calculations because these bonds can not be determined directly using ^{1}H NMR. They are inferred based on long amide exchange times and NOE patterns indicative of types of secondary structure, i.e. β-turns, β-sheet, and α-helix. While long exchange times, i.e. greater than 24 hours, were observed for a small number of amide protons hydrogen bonds, for these protons, were not included in the structure calculations.

Structure Calculations. All structure calculations were performed on a Silicon Graphics Crimson R-4000 computer using the program XPLOR 3.1 [64]. Ten initial structures included only unambiguous β-protons between cysteine disulfide pairs and putative secondary structure NOE constraints. Dihedral Φ and χ_1 constraints were not included in these calculations. The structures were checked for global fold and for disulfide bonding pattern. Additional constraints were gradually added to the structure calculations while checking the generated structures to make sure there were few distance constraint violations and low energy values. Structures with more than 2 NOE violations greater than 0.5 Å or with E_{NOE} greater than 50 and/or E_{TOTAL} greater than 200 kcal/mole were excluded. After the first three cycles of this process, pseudo-NOE constraints for the cysteine sulfur atoms of the observed disulfide bonds were included. This process was repeated until 50% of the structures generated fulfilled the above criteria and had a RMS deviation for all atoms of less than 3.5 Å. A set of 30 structures were generated using this final constraint set. The structures were then refined using high temperature simulated annealing with a slow-cooling portion at the end. The refinement included dihedral Φ, χ_1 constraints and disulfide bonds. A set of ten structures were generated using an *ab initio* simulated annealing protocol against a straight chain template [65] for comparison with the distance geometry/simulated annealing regularized structures.

The refinement procedure used has been described previously [66] and was used with a few modifications. A straight chain polypeptide template was generated in which the cysteine sulfur atoms were not protonated. A file containing structure information, i.e. atom types, bonding and angle information, was generated in which the cysteine sulfur atoms were modified so that they were unprotonated, had no charge, and had a nonbonded van der Waals radii of 1.1 Å. Ten structures were generated using an embedded distance geometry (substructure) protocol (dg_sub_embed.inp) and regularized using a simulated annealing protocol (dgsa.inp) [64,66,67]. The steps are as follows: 1) distance geometry (DG) substructure embedding, 2) minimization against a DG energy term, 3) high temperature simulated annealing (SA) against the template structure described previously, 4) a cooling stage followed by Powell minimization, and 5) testing for the correct enantiomer. The last three steps are the SA regularization for the DG substructures mentioned previously. Covalent disulfide bonds were not included during any of the procedure described above. The structures were further refined using a 5 fs time step for Langevin dynamics starting at 1000° K and slowly cooling to 100° K. A final minimization included two hundred steps of the Powell conjugate gradient method.

Relaxation Matrix Back-Calculation. The average coordinates were further refined against a NOE intensity relaxation matrix. This was accomplished by calculating cross peak intensity between spins i and j, I_{ij}^{c}, from refined atomic coordinates by means of a relaxation matrix $\Re$ [68,69]:

$$I_{ij}^{c} \propto \left[\exp\left(-\Re\tau_m\right)\right]_{ij}, \qquad \textbf{(5-1)}$$

where τ_m is the NOE mixing time. The transition rates Ω^{ij}, of which the matrix $\Re$ is a function,

$$\Re_{ij} = \begin{cases} \Omega_2^{ij} - \Omega_0^{ij} & \text{if } i \neq j \\ \sum_{k \neq j} \Omega_0^{kj} + 2\Omega_1^{kj} + \Omega_2^{kj} + R_{leak} & \text{if } i = j \end{cases} \qquad \textbf{(5-2)}$$

are determined by dipolar coupling strengths and spectral densities [64,70]:

$$d_{ij} = \frac{\gamma^4 \hbar^2}{10 r_{ij}^6} \qquad \textbf{(5-3)}$$

and

$$\begin{aligned} \Omega_0^{ij} &= d_{ij} J(0) \\ \Omega_1^{ij} &= \tfrac{3}{2} d_{ij} J(\omega) \\ \Omega_2^{ij} &= 6 d_{ij} J(2\omega) \end{aligned} \qquad \textbf{(5-4)}$$

where γ is the gyromagnetic ratio for the nucleus of interest (in this case protons) and r_{ij} is the distance between nuclei *i* and *j*. For this calculation, the Lipari and Szabo [64,71] "model free" approach was used to describe the non-NOE magnetization losses in the lattice, R_{leak}, where internal motion is modeled using two parameters, an effective correlation time τ_e and an order parameter S^2:

$$J(\omega) = S^2 \frac{\tau_c}{1 + \omega^2 \tau_c^2} + \left(1 - S^2\right) \frac{\tau_e}{1 + \omega^2 \tau_e^2} \qquad \textbf{(5-5)}.$$

The variable τ_c assumes that the macromolecule being studied has a single isotropic correlation time. For this calculation it was assumed that internal motion was much faster than the overall molecular tumbling rate $\left(\tau_e \ll \tau_c\right)$ so that the value of the second term of equation 5-5 becomes negligible [64]. Because this calculation is computationally intensive, the calculation was performed for a single mixing time, 150 ms. The correlation times (τ_c) used for backbone to backbone, backbone to sidechain, and sidechain to sidechain, were 0.85, 0.80, and 0.65 seconds, respectively.

Results

Resonance Assignments. ^{1}H resonance assignments for *Amb. t.* V were made following the standard procedures described in detail by Wüthrich [29]. This first requires associating amino acid residue spin systems into classes using ND-COSY, PCOSY, and TOCSY spectra. The spin systems are initially identified from the *J*-coupling and cross

peak pattern observed using one of the variants of the COSY experiment. Figure 5-6 shows a portion of a PECOSY spectrum with some of the AMX spin systems identified. The spin system assignments were verified using a long mixing times (80 ms) TOCSY spectrum. Sequence specific assignments were made from a NOESY experiment with a medium length mixing time (150 ms). Two of the spin systems provided unique starting points for the sequential connectivities, in particular alanine 33 (A33) and glutamic acid 7 (E7), because they occurred only once in the sequence. Other spins systems such as the 4 glycine with nondegenerate Cα protons (G8, G12, G15, and G23), the two isoleucine residues (I22 and I34), and the two threonine residue (T9 and T40) provided other starting positions. In addition, identification of the aromatic $d_{\delta\beta}$ NOEs for the four tyrosine and the $d_{\delta\beta}$ for two of the three asparagine helped to reduce the degeneracy in the 20 AMX spin systems present in this peptide and provided additional starting positions. NOESY spectra with a 150 ms mixing time were used to identify the sequential $d_{\alpha N}$, d_{NN}, and $d_{\beta N}$ connectivities. We were able to assign 70% of the residues sequence specifically. 2D contour plots of some of the connectivities, $d_{\alpha N}$ and d_{NN} are shown in Figures 5-7 and 5-8, respectively. Additional sequential assignments were made using 3D data (see Figures 5-9, 5-10, 5-11, 5-12, 5-13, 5-14, and 5-15) leading to the identification of sequential connectivities for 39 of the 40 residues (sequential connectivities to D1 were not observed). The result was the identification of 80 sequential (i, i+1 and i, i-1) NOE connectivities. The determination of the sequence specific resonance assignments allowed for the identification of 236 intraresidue NOE connectivities. The resonance assignments for *Amb. t.* V are summarized in Table 5-1.

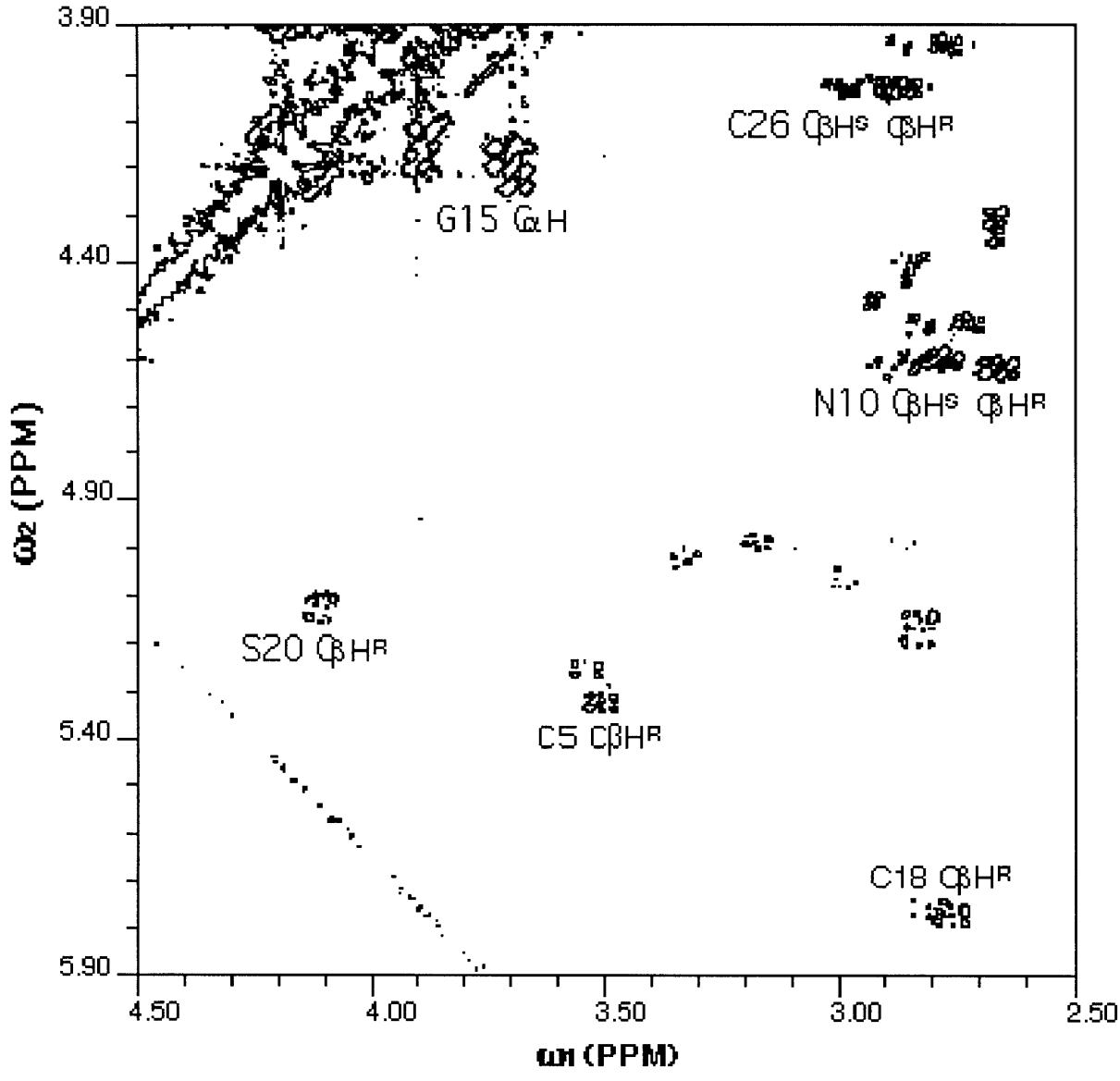


Figure 5-6: The HC_α to HC_β region of a 2D PECOSY spectrum in D_2O at 20 °C. The cross peaks highlighted are characteristic for an AMX spin system and their sequential assignments are as shown.

Table 5-1: Proton Chemical Shifts (PPM) for *Amb. t.* V[a]

residue	NH	$C_\alpha H$	$C_\beta H$	$C_\gamma H$	$C_\delta H$	other
D1		4.24	2.88, 2.80			
D2	8.03	4.13	1.47, 0.92			
G3	8.27	3.98				
L4	8.28	4.44	1.67	1.66	0.92, 0.89	
C5	8.45	5.32	3.51^R, 2.73^S			$C_\varepsilon H$ 6.79
Y6	8.93	5.08	3.10^R, 2.99^S		7.09	
E7	9.02	3.55	1.77^R, 2.18^S	2.54, 2.06		
G8	8.39	3.76, 3.64				
T9	8.88	3.67	4.17	1.19		
N10	9.00	4.59	2.77^R, 2.84^S			$N_\delta H$ 7.35, 6.95
C11	7.89	4.85	2.87^R, 3.28^S			
G12	8.11	4.15, 3.89				
K13	7.63	4.58	1.37^R, 1.55^S	1.07	1.28	$C_\varepsilon H$ 2.86
V14	8.03	3.55	1.89	1.07, 0.90		
G15	9.27	4.22, 3.69				
K16	7.48	4.69	1.70, 2.00	1.26, 1.15	1.36	$C_\varepsilon H$ 2.78
Y17	8.40	4.77	2.38, 2.55		6.70	$C_\varepsilon H$ 6.85
C18	9.44	5.74	2.76^R, 2.32^S			
C19	9.31	5.04	3.09^R, 3.35^S			
S20	9.27	5.10	4.12^R, 4.24^S			
P21		4.61	2.36, 1.99	2.10, 1.85	3.55, 3.49	
I22	8.16	3.90	1.81	1.60	1.26	$C_\gamma H_3$ 0.93
G23	8.69	4.18, 3.72				
K24	8.48	4.64	1.44^R, 1.34^S	1.77	1.66	$C_\varepsilon H$ 2.97;$N_\varepsilon H$ 7.59
Y25	7.90	4.86	2.93		7.10	$C_\varepsilon H$ 6.79
C26	8.24	3.89	2.71^R, 3.17^S			
V27	8.77	3.99	2.21	0.84, 0.82		
C28	8.29	5.15	2.83^R, 2.52^S			
Y29	9.20	4.77	2.57^R, 3.21^S		7.13	$C_\varepsilon H$ 6.69
D30	9.06	4.61	2.61^R, 2.81^S			
S31	7.45	4.94	4.20^R, 3.92^S			
K32	8.89	3.19	0.28^R, 1.27^S	1.08, 0.66	1.37	$C_\varepsilon H$ 2.80
A33	8.34	4.03	1.32			
I34	7.58	3.50	1.62	1.54, 1.06	0.92	$C_\gamma H_3$ 1.06
C35	7.84	3.15	3.15, 3.00			
N36	8.75	4.30	2.68			$N_\delta H$ 7.31, 7.03
K37	7.31	4.07	1.77	1.59	1.43	$C_\varepsilon H$ 2.91 $N_\varepsilon H$7.53
N38	7.33	4.70	1.96^R, 2.83^S			$N_\delta H$ 7.22, 6.78
C39	7.99	4.98	2.94^R, 3.18^S			
T40	7.60	4.25	4.26	1.13		

[a] Assignments for *Amb. t.* V were made in 50 mM acetate buffer (pH=4.0) at 20 °C in 90% H_2O. *R* and *S* denote protons assigned stereospecifically.

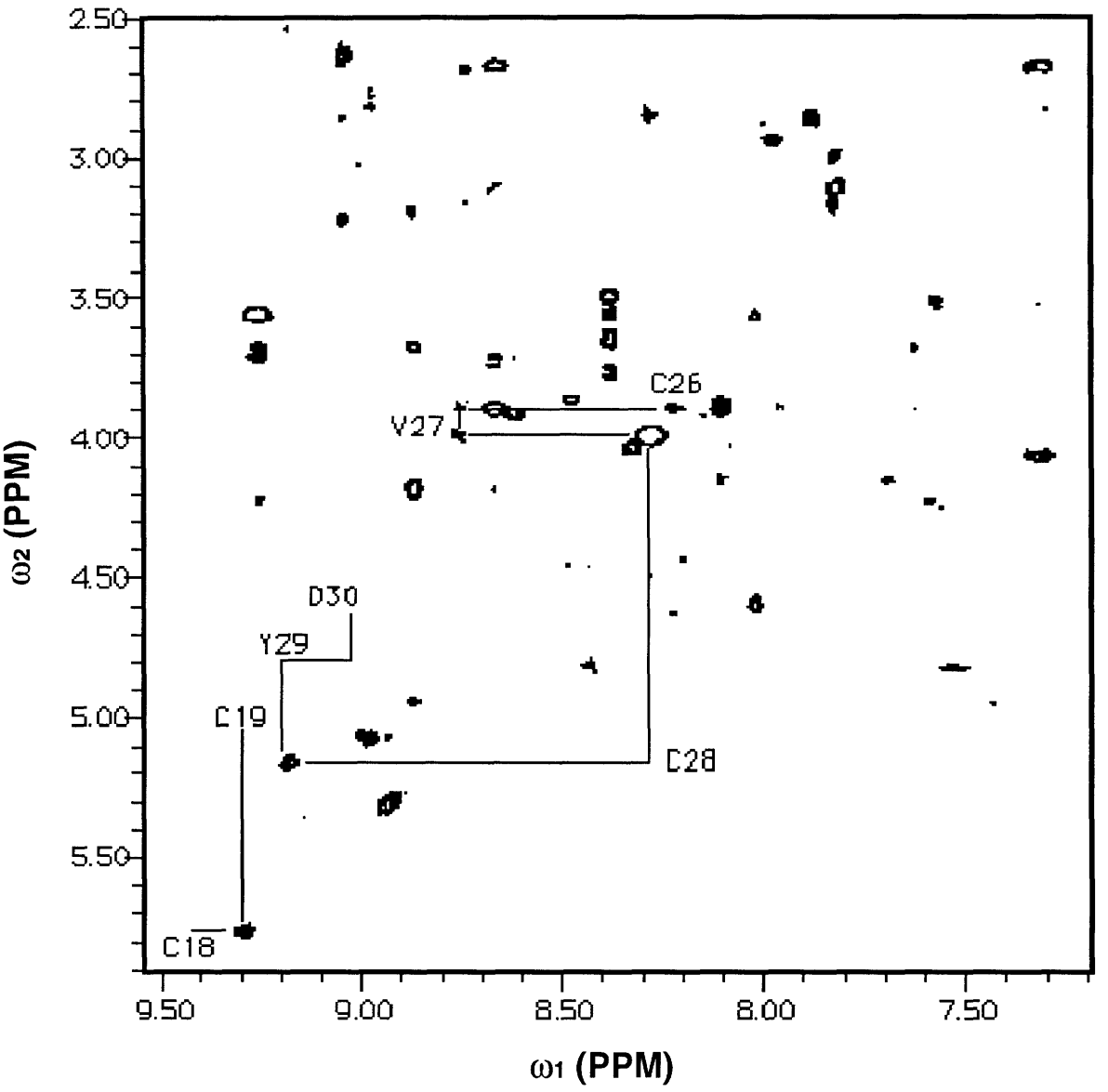


Figure 5-7: An expansion of the fingerprint region of a 2D NOESY spectrum with a 150 ms mixing time. Sequential $d_{\alpha N}$ connectivities from C18 to C19 and from C26 to D30 are shown. No cross peaks were observed for two of the sequential connectivities, Y29 and S30, at this mixing time.

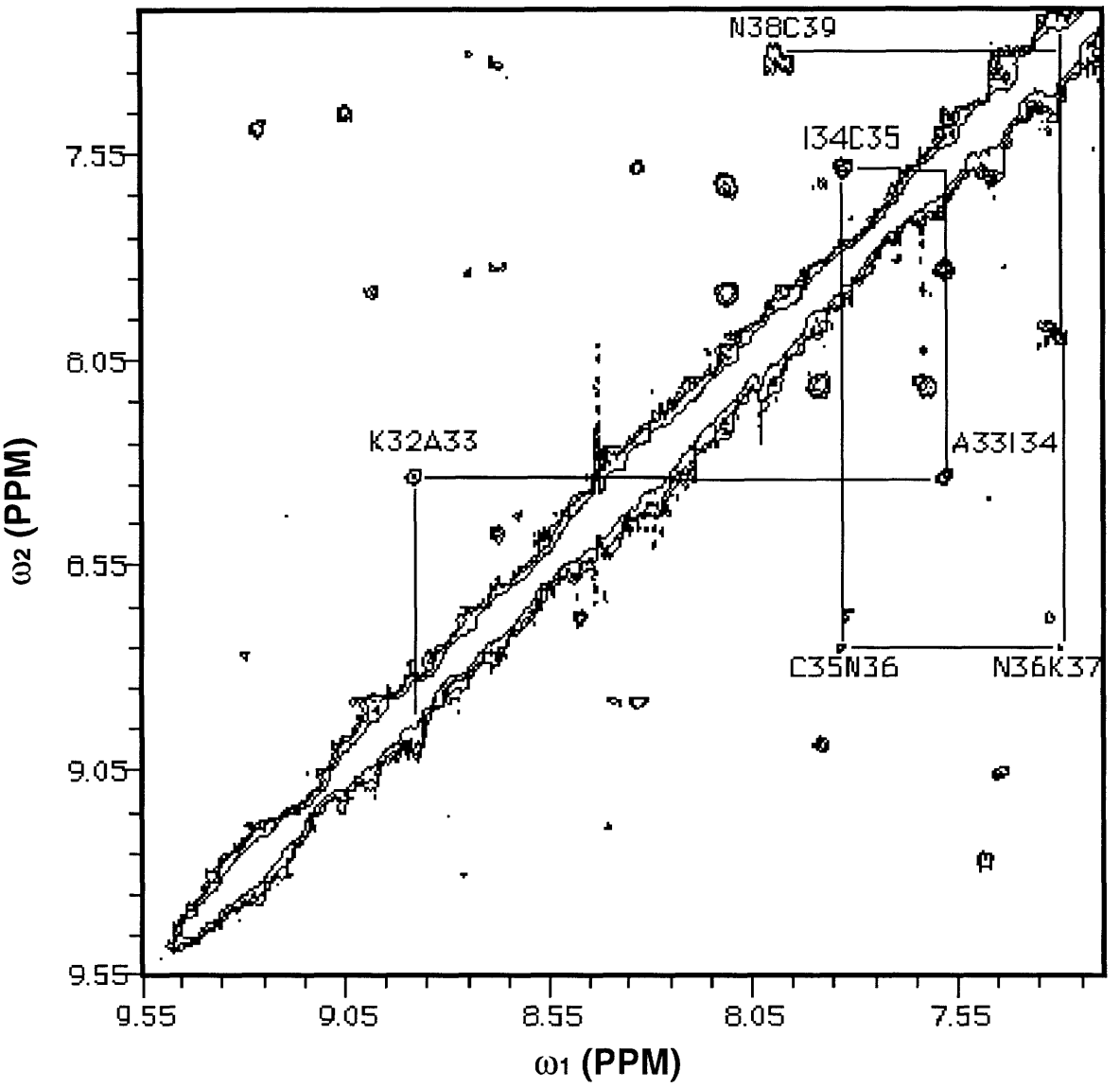


Figure 5-8: An expansion of the amide region from a NOESY spectrum with a 150 ms mixing time. The sequential medium strength NOEs highlighted are indicative of α-helix.

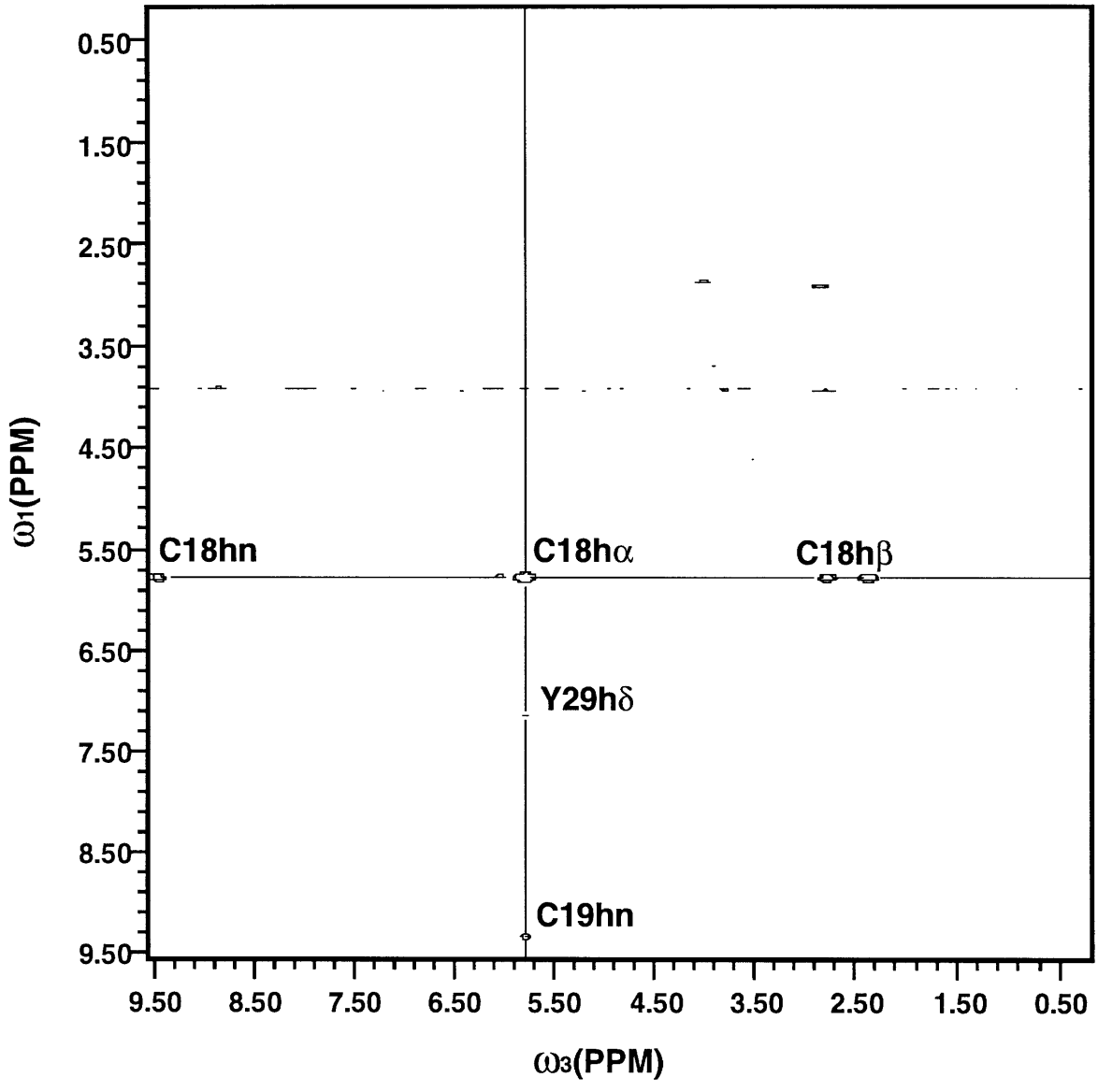


Figure 5-9: The Hα resonance plane along ω_2 from a 3D NJRTOC experiment. The vertical line indicates a NOE transfer pathway while the horizontal line delineates the TOCSY pathway. Note that in addition to the sequential connectivities a long-range NOE from C18 to Y29 is observed.

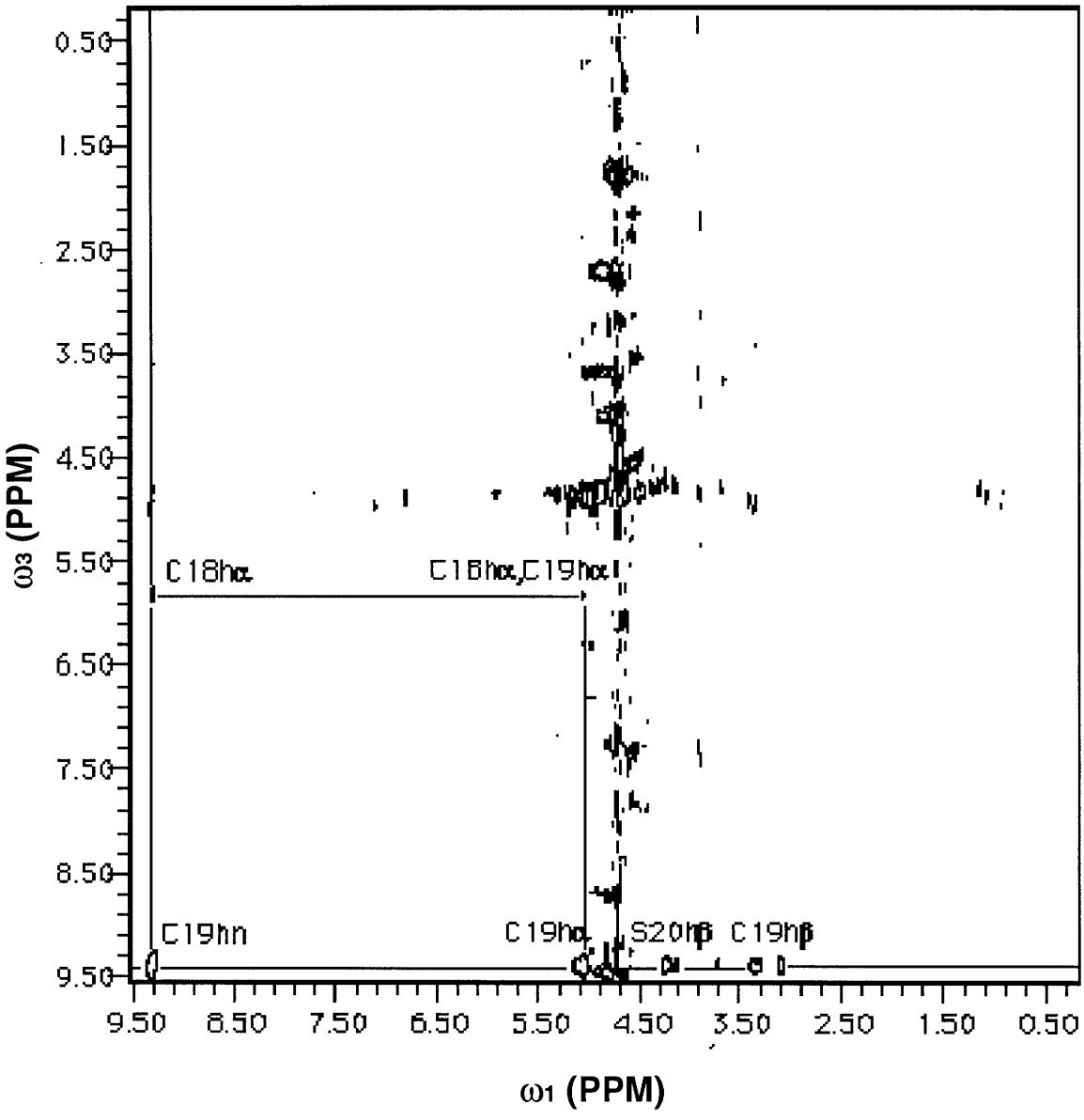


Figure 5-10: The sequential identification of residues 18,19, and 20 are shown in this ω_2 plane from the TOCNJR experiment at the C19 amide resonance. Note that the S20 Hα resonance could not be identified due to interference from residual solvent and the small difference between the amide resonance for C19 and S20. This interference would not be present in the $\omega_2 \times \omega_3$ plane.

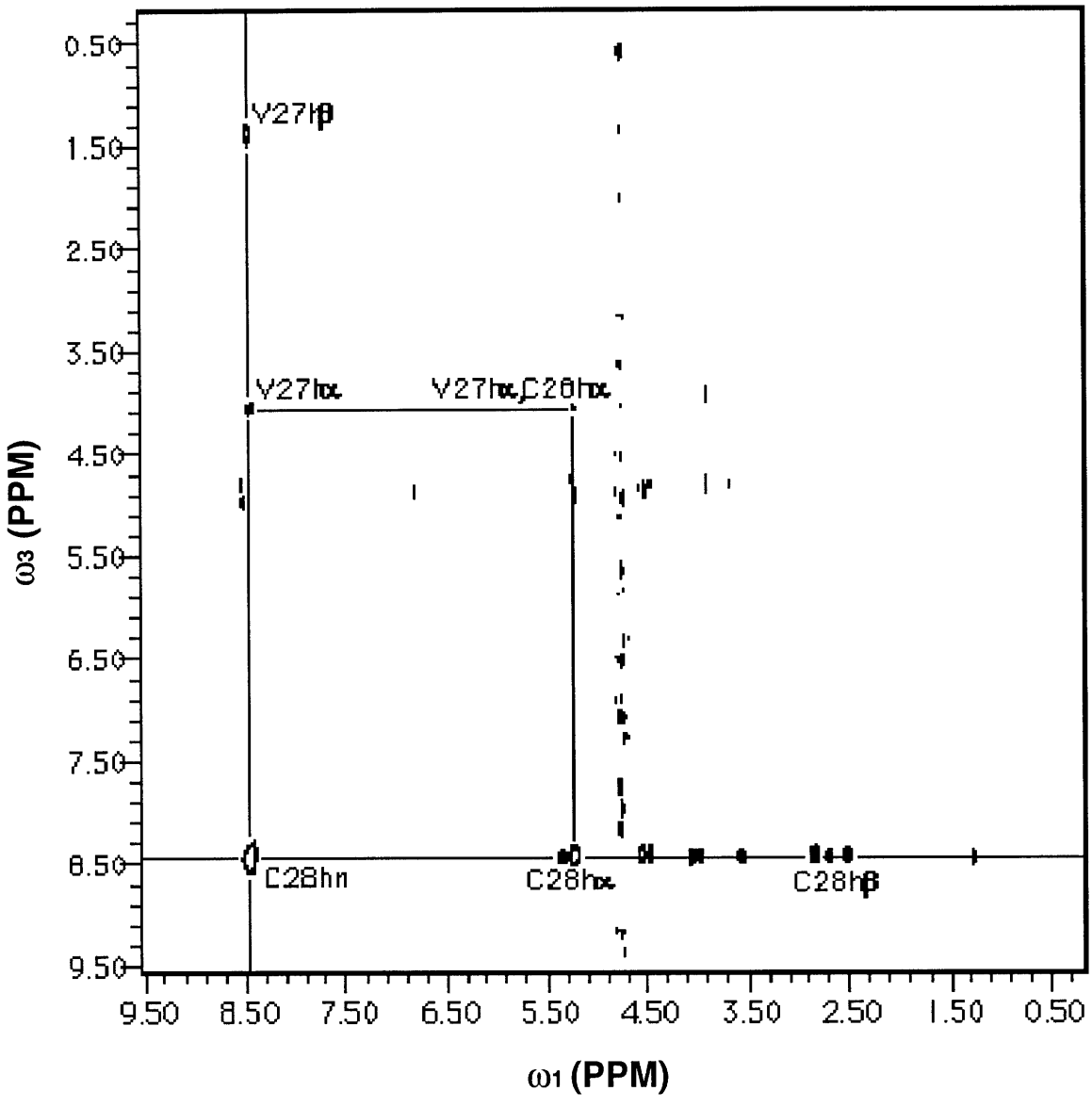


Figure 5-11: An $\omega_1 \times \omega_3$ plane along ω_2 from the TOCNJR experiment at the amide resonance for C28 showing the sequential resonance assignment of V27 and C28. The horizontal line signifies a TOCSY transfer pathway while the vertical line indicates a NOE transfer pathway.

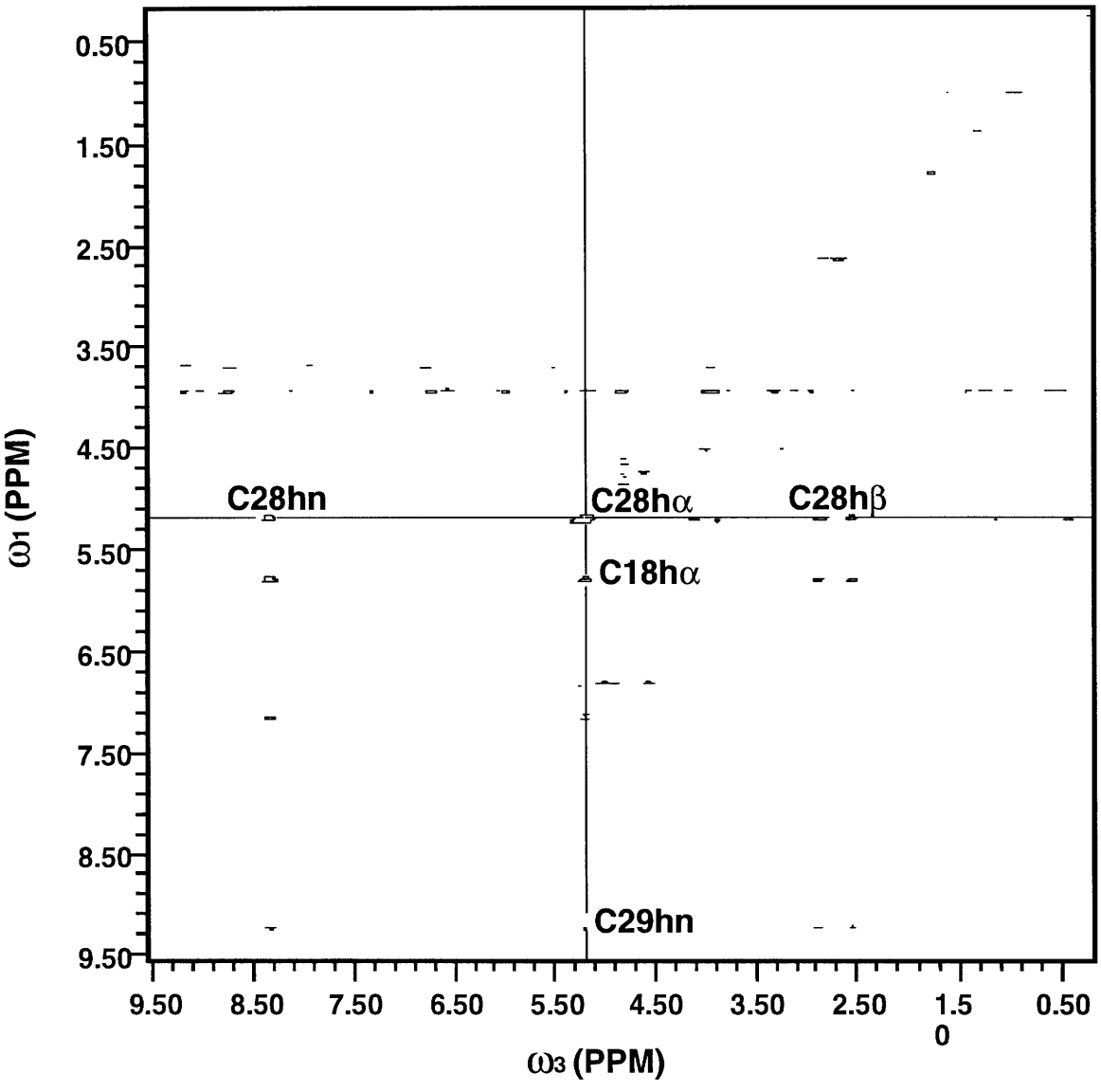


Figure 5-12: A plane along ω_2 from the NJRTOC experiment at the Hα resonance for C28. Note that in addition to the sequential cross peak to the amide of Y29 there is an intense long range NOE to the Hα of C18. This NOE is indicative of anti-parallel β-sheet.

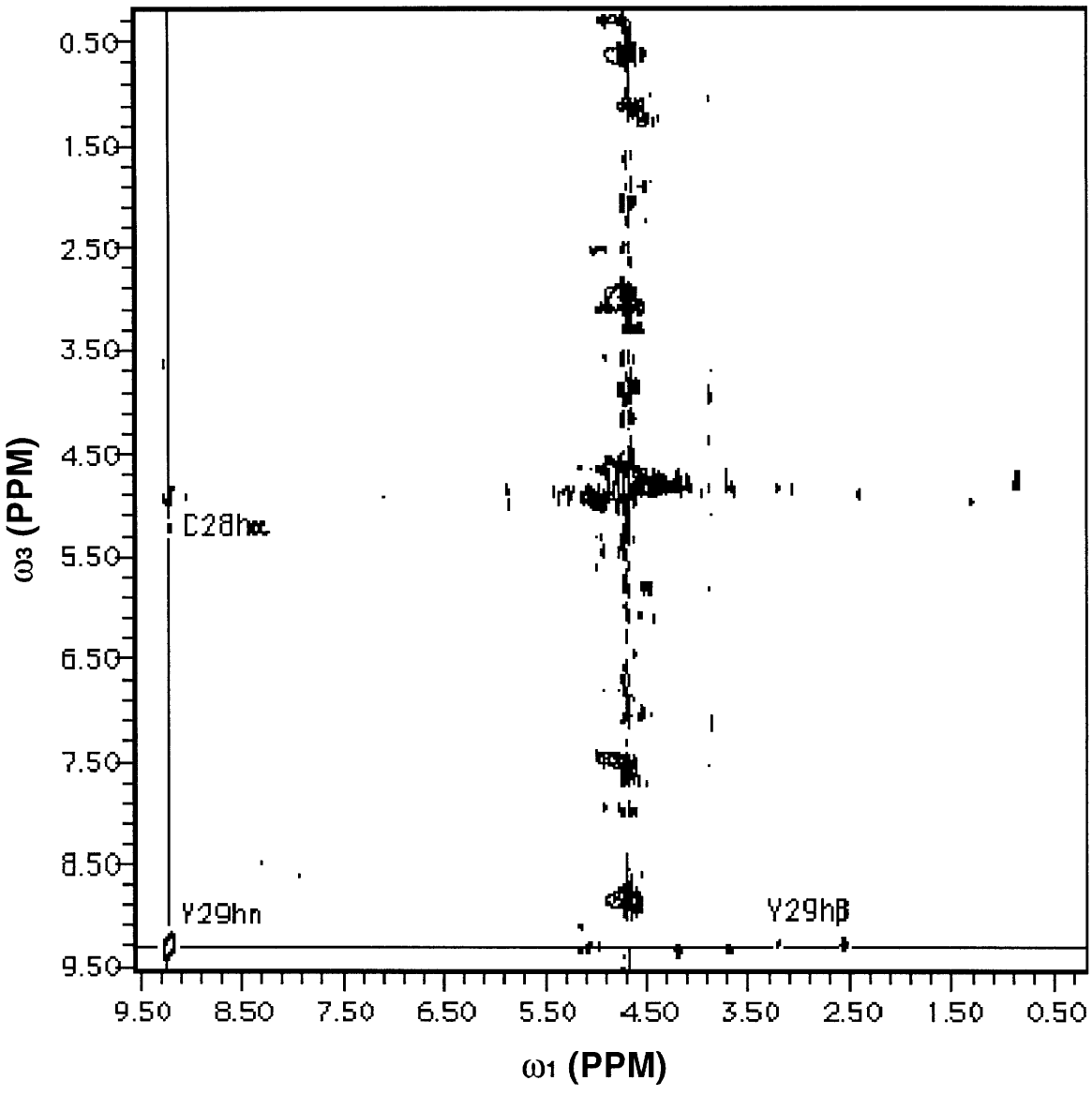


Figure 5-13: This ω_2 plane from the TOCNJR experiment at the Y29 amide resonance confirms the sequential crosspeak observed at the Cα proton resonance. The missing Hα to HN crosspeak between Y29 and D30 may be lost in the solvent signal since the Y29 Hα resonance lies under the solvent peak.

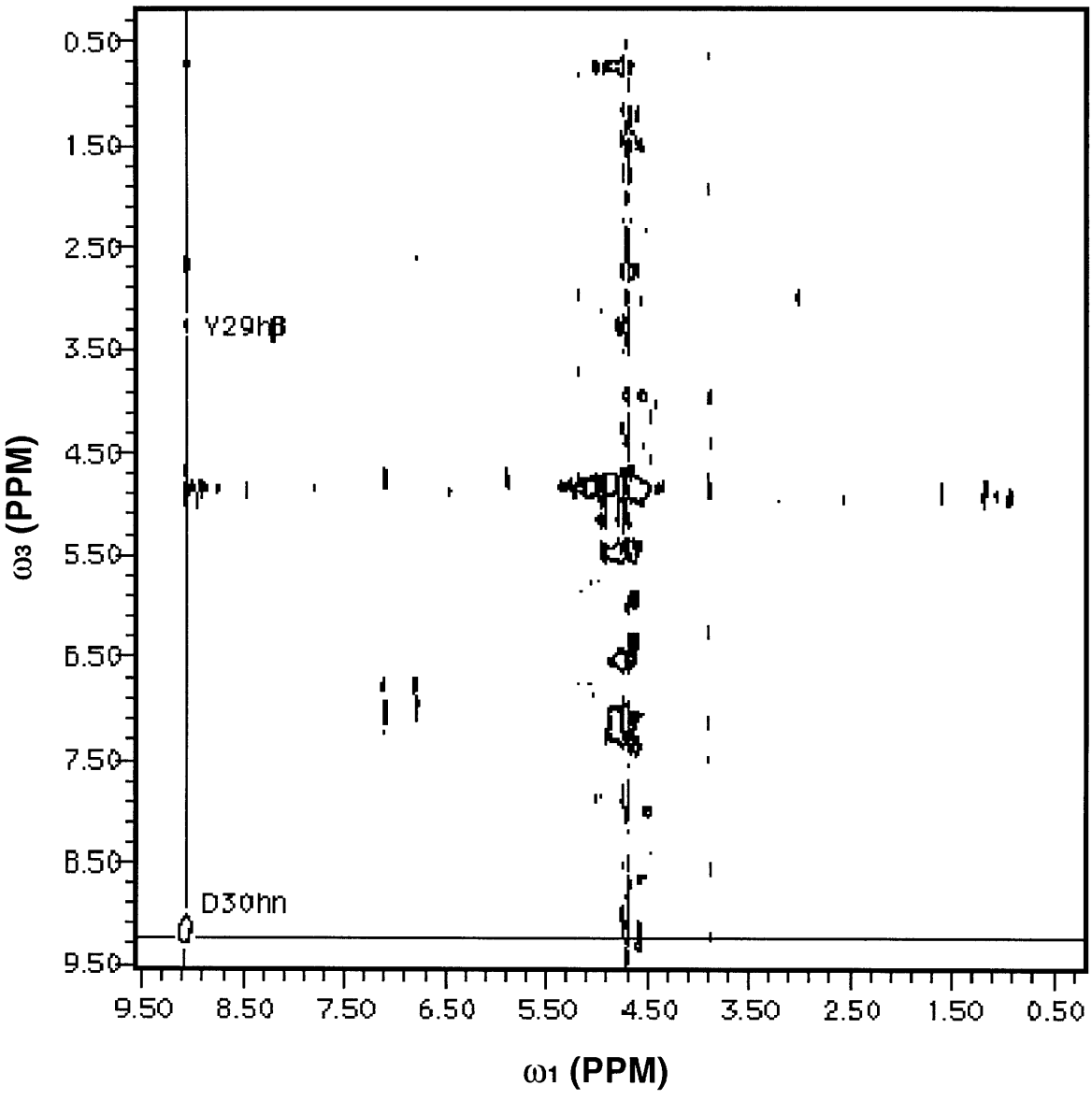


Figure 5-14: The sequential resonance assignment for residues Y29 and D30 from the ω_2 plane of a NJRTOC experiment at the chemical shift corresponding to the amide resonance for Y29. This Hα to HN cross peak was not observed in the fingerprint region of 2D NOESY spectrum seen in Figure 5-7.

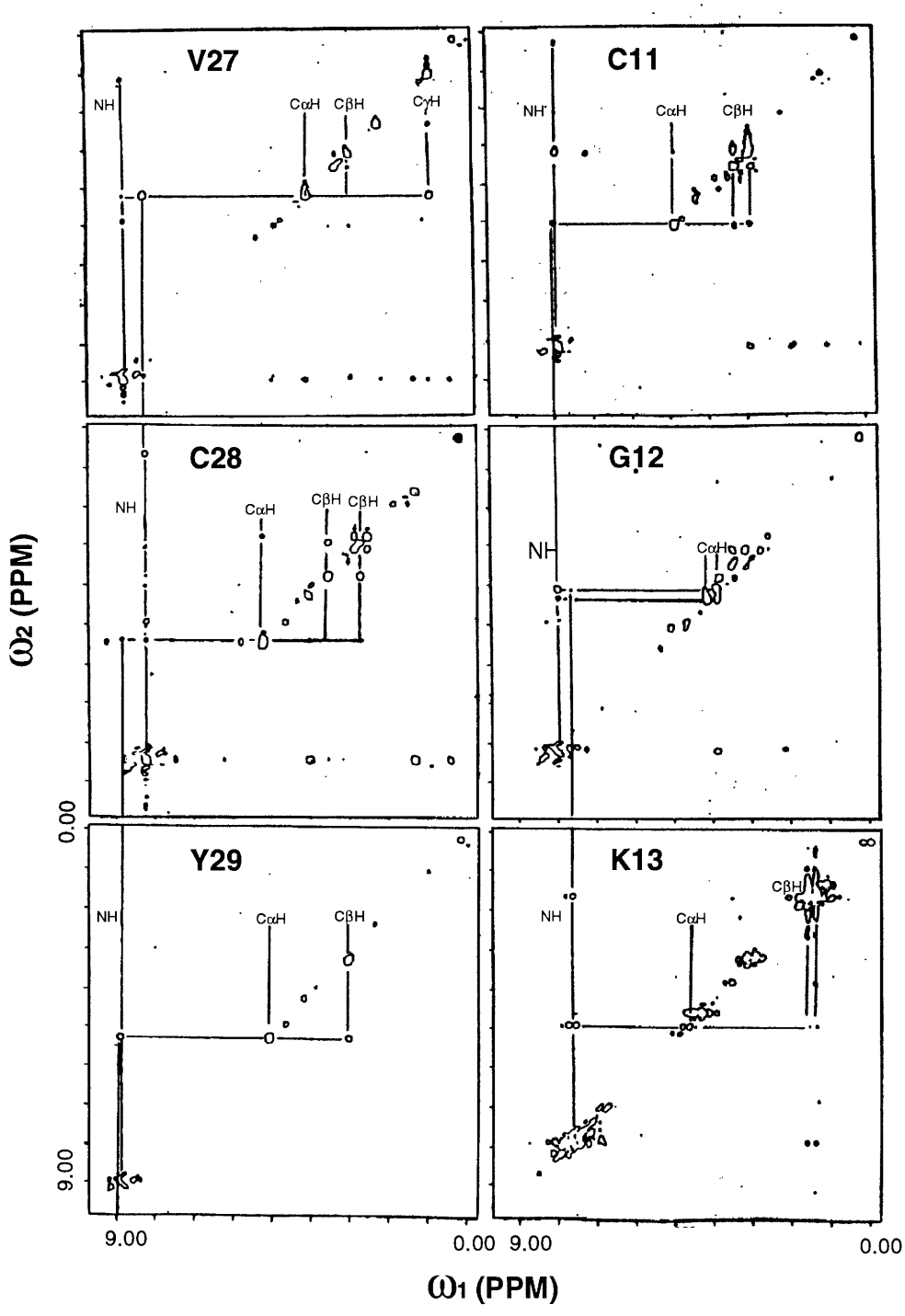


Figure 5-15: Six ω_3 planes from a NJRTOC experiment illustrating the sequential assignment of six residues, V27 to Y29 and C11 to K13. The planes intersect at a ω_3 chemical shift corresponding to the amide resonance of the residue listed in the left hand corner of the spectrum.

Secondary Structure Conformation Analysis. The presence of slowly exchanging amide proton is indicative of secondary structure or a large tightly packed hydrophobic core. Since this is a small peptide and the presence of a large hydrophobic core is unlikely, it was assumed that the slowly exchanging amide proton observed in this case were probably due hydrogen bonding present in secondary structure. However, there are examples of small peptides with slowly exchanging amides due to a tightly packed hydrophobic core and not hydrogen bonding [72].

The presence of α-helix was indicated by the observation of medium strength d_{NN} NOEs observed from residues K32 to C39. This was substantiated by the identification of a stretch of medium range NOEs, $d_{\alpha\beta}$ (i, i+3) and $d_{\alpha N}$ (i, i+3), from residues D30 to C39. The small values observed for $^3J_{HN\alpha}$ $d_{\alpha\beta}$ in some of the residues in this region confirms the presence of α-helix for this region.

The β-sheet was identified by the presence of strong sequential $d_{\alpha N}$ NOEs, weak intraresidue $d_{\alpha N}$ NOEs, and large values for $^3J_{HN\alpha}$. These conditions were observed over three regions of the sequence; residues C5 to G8, Y17 to S20, and C26 to Y29. The presence of strong $d_{\alpha\alpha}$ NOEs far apart in the sequence is indicative of the presence of anti-parallel β-sheet. Two strong $d_{\alpha\alpha}$ NOEs were observed between C18 and C28 and between E7 and Y17. This would indicate the presence of a triple-stranded anti-parallel β-sheet. This was confirmed by the presence of strong $d_{\alpha N}(i,j)$, $d_{NN}(i,j)$ between residues adjacent to the $d_{\alpha\alpha}$ NOEs.

The last secondary structure feature identified was a Type II β-turn. This was identified by the presence of a small $^3J_{HN\alpha}$ value for V14 with a strong $d_{\alpha N}$ NOE between V14 and G15, and a strong d_{NN} NOE between G15 and K16. The last identifying NOE was a medium strength $d_{\alpha N}$ (i, i+2) NOE between V14 and K16.

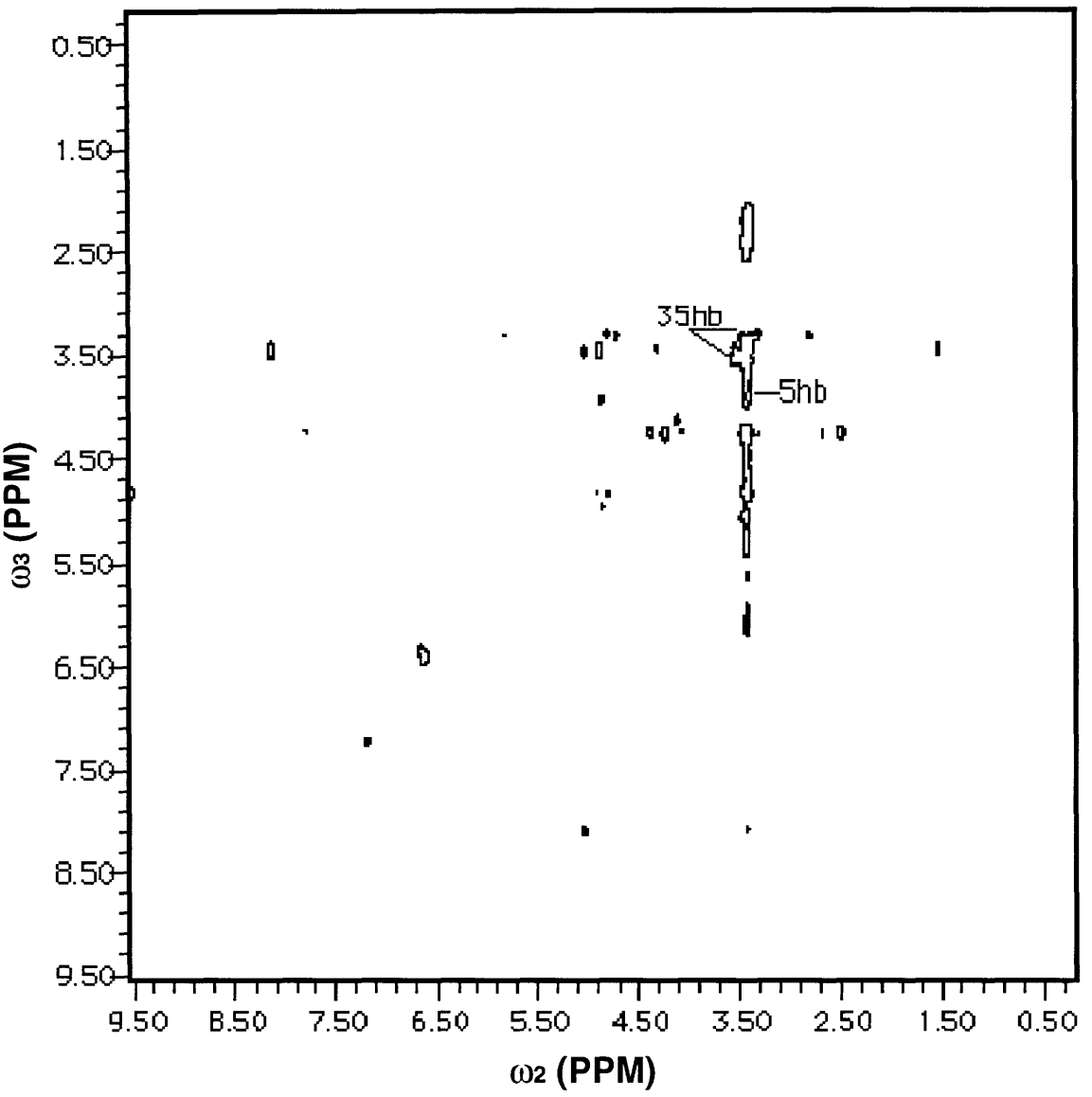


Figure 5-16: An ω_1 plane corresponding to the chemical shift of the Hβ of C35 (3.15 PPM) in the NJRTOC experiment. The cross peak observed is a NOE backtransfer peak from the Hβ resonance at 3.00 PPM to the $H\beta^R$ of C5 at 3.51 PPM. A backtransfer NOE and NOE cross peaks were observed to the $H\beta^S$ C5 resonance as well.

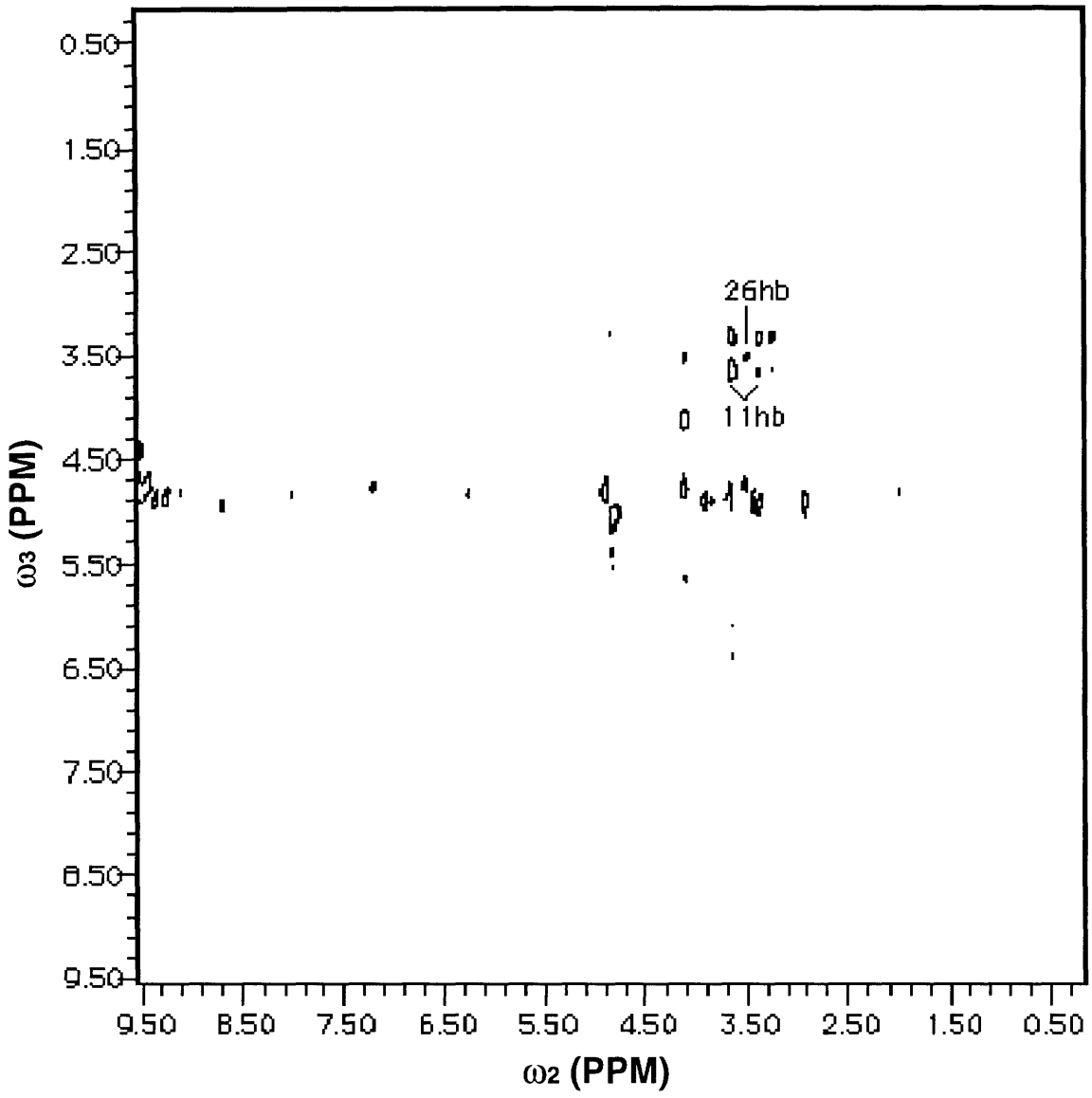


Figure 5-17: The $H\beta^S$ to $H\beta^S$ C11-C26 NOE cross peak is high lighted in this 2D plane along an ω_1 chemical shift of 3.28 PPM. A NOE cross peak from C11 $H\beta^S$ to the $H\alpha$ of C26, not shown, was also identified.

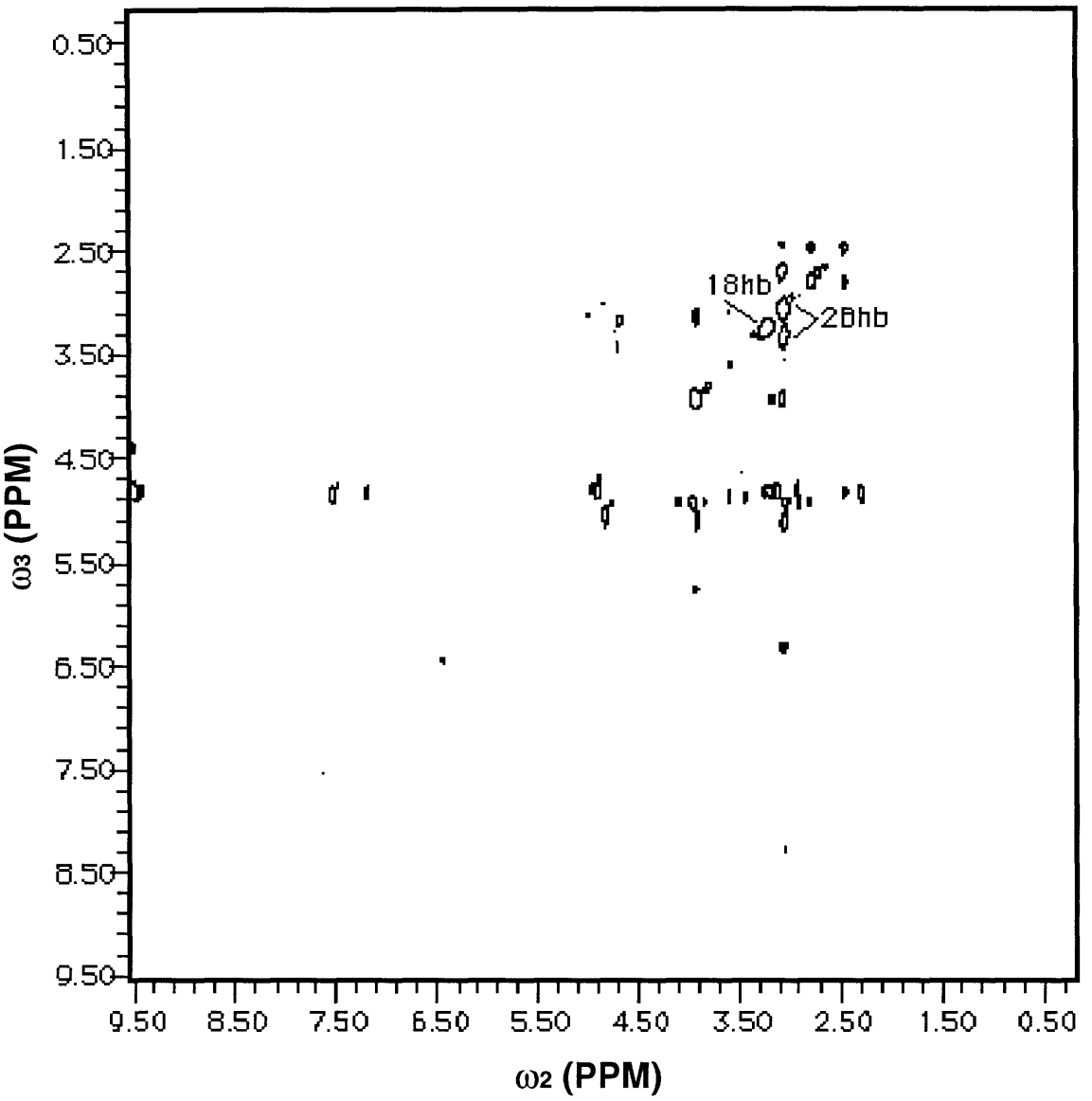


Figure 5-18: An ω_1 plane, from the NJRTOC experiment, corresponding to the chemical shift of the $H\beta^S$ of C28 (2.52 PPM) reveals a NOE cross peak between the C28 $H\beta^S$ and the C18 $H\beta^R$ protons. A NOE cross peak was also observed between the C18 $H\beta^S$ and the C28 $H\beta^S$ protons.

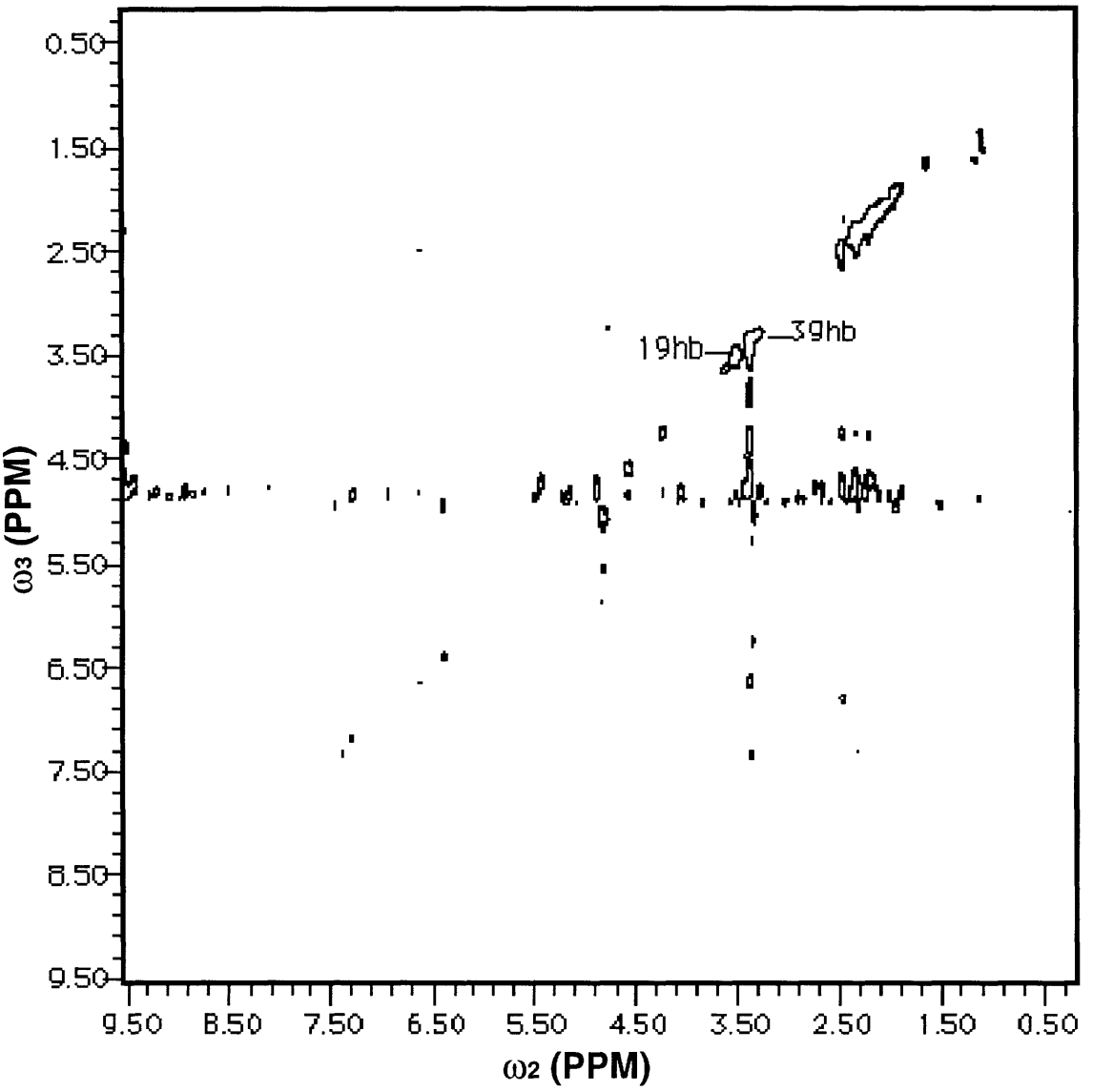


Figure 5-19: The NOE cross peak between C39 $H\beta^R$ and C19 $H\beta^R$ found in the ω_1 plane corresponding to 2.94 PPM is more ambiguous than previous Hβ-Hβ disulfide identifications. However, it serves as a verification that the three previous disulfide assignments are correct.

The disulfide bonding pattern was determined by identifying $d_{\beta\beta}$ NOEs between cysteine residues. Seven such constraints were observed, four of which are shown in Figures 5-16, 5-17, 5-18, and 5-19. In addition two $d_{\alpha\beta}$ NOEs between cysteine residues were identified from 26 Hα to 11 Hβ and from 18 Hα 28 Hβ. The disulfide bonding pattern, identified based on these 9 NOE constraints, was: $C_5 - C_{35}$, $C_{11} - C_{26}$, $C_{18} - C_{28}$, and $C_{19} - C_{39}$.

Structure Determination. A total of 416 NOE structural constraints were identified using NOESY 2- and 3-D data. The NOE data set used for the structure calculations included 80 sequential, 53 medium-range $(|i-j| \leq 5)$ $(|i-j| > 5)$, and 47 long-range $(|i-j| > 5)$ constraints for a total of 180 interresidue constraints. A set of 30 DG-SA starting structures were generated and refined by high temperature dynamic simulated annealing using the program XPLOR as described previously. Before refinement there were 14 structures defined as accepted i.e. no NOE constraint violations greater than 0.5 Å, E_{NOE} less than 50 kcal/mole, and E_{TOTAL} less than 200 kcal/mole. After refinement there were 21 structures which met the above criteria. Of these 21 structures 5 were chosen and an average refined structure calculated. It is of interest to note, from the structure calculation statistics found in Table 5-2, that the RMSD for all atoms dropped by 1.1 Å during the refinement as well as E_{TOTAL}, but E_{NOE} increased slightly and the RMSD for all NOE constraints remained constant. The average RMSD for all backbone atoms was 0.92 Å. This included residues 1 and 2 which are disordered and for which no NOE constraints have been identified. The RMSD for all nonhydrogen atoms was 1.50 Å and for all atoms 2.76 Å. A superposition of the Cα atoms and the backbone atoms for the 5 final refined structures is shown in Figure 5-20 and 5-21, respectively. Figure 5-22 contains a Richardson plot displaying the secondary structure of a representative "best" structure.

Figure 5-20: A view of the C_α atoms from 5 refined structures of *Amb. t.* V. These structures were superimposed so as to minimize the RMSD of all atoms. Note the disorder in the first two C_α atom positions due to a lack of NOE constraints for these residues.

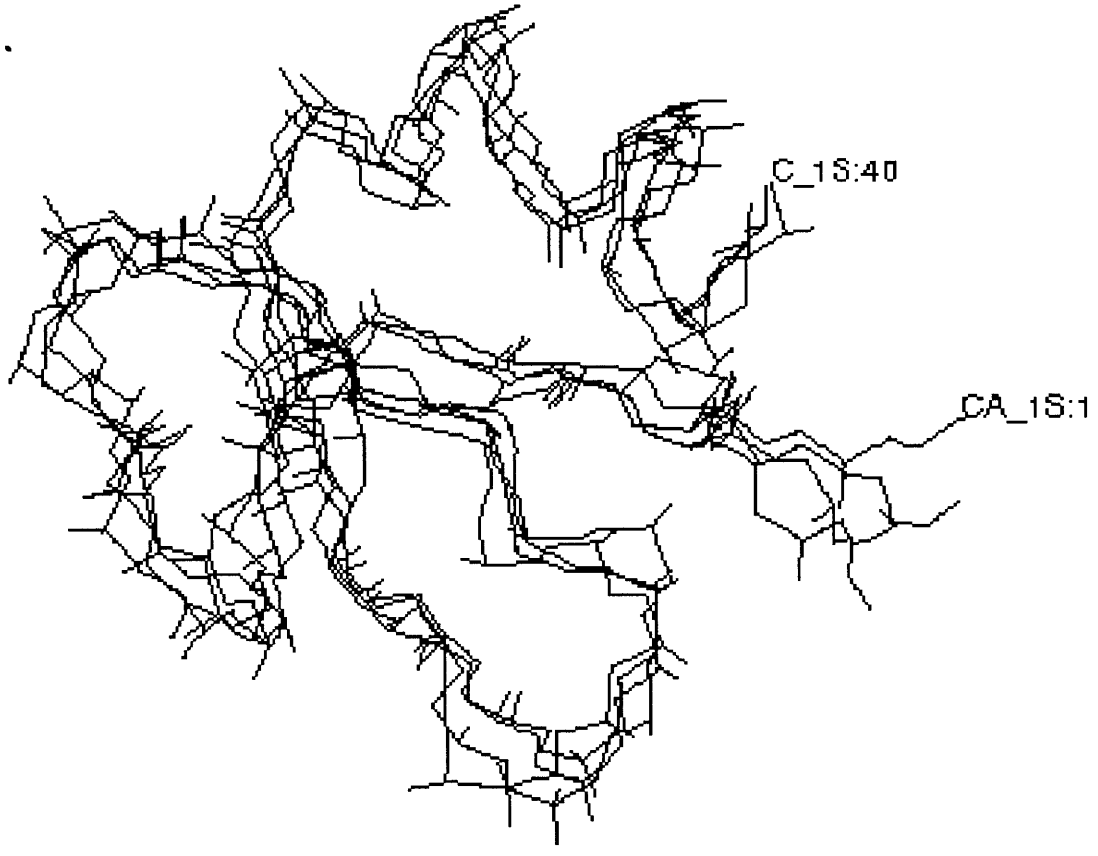


Figure 5-21: A view of the backbone atoms (HN, N, HC_α, C_α, and C) of 5 of the final refined structures of *Amb. t.* V. The structures were superimposed in a manner similar to that described in Figure 5-20. Note that in the β-strands (5-7 and 27-29 in particular) the HN and HC_α protons are highly resolved.

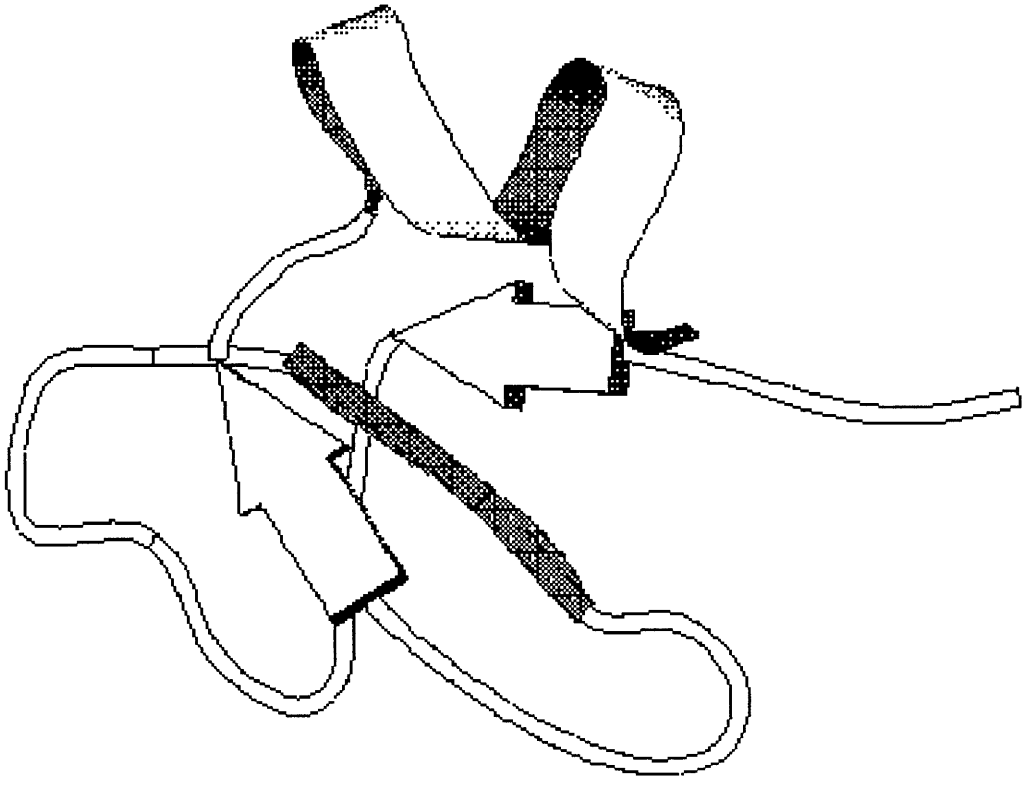


Figure 5-22: A Richardson diagram of a representative structure. The triple stranded antiparallel β-sheet (residues 5-7, 17-20, and 27-29) is represented by arrows and the α-helix (residues 31-39) by a spiral tape. The α-helix lies over strand 5-7 and 17-20 of the β-sheet.

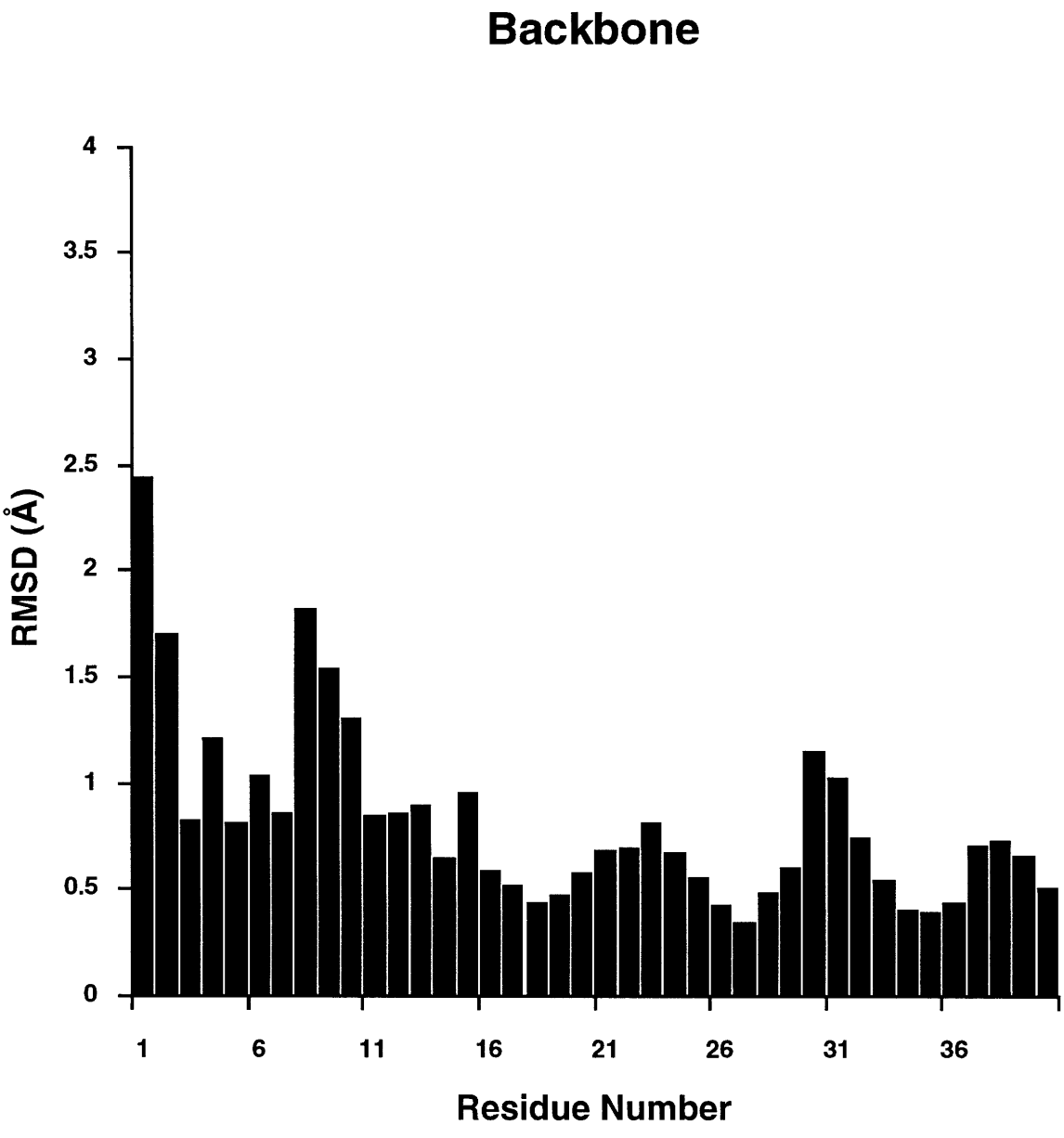


Figure 5-23: A histogram of the RMSD of the atomic positions for the backbone atoms (N, C_α, C, O). The structure is well defined for part of the β-sheet (residues 17-20 and 27-29) and in the α-helical region.

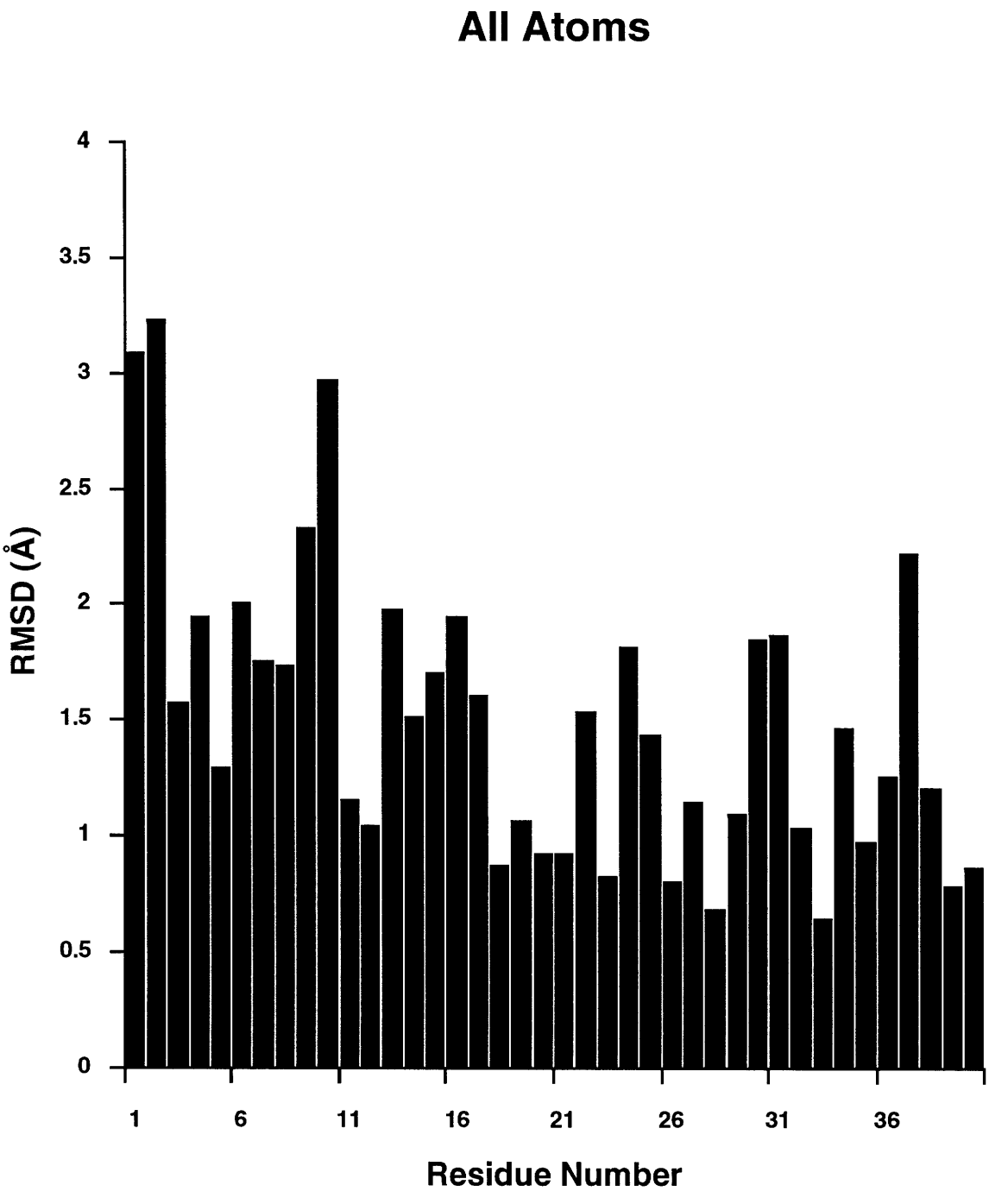


Figure 5-24: A histogram of the RMSD of the atomic positions for all atoms. The structure is well defined in the β-strand from residues 17-20.

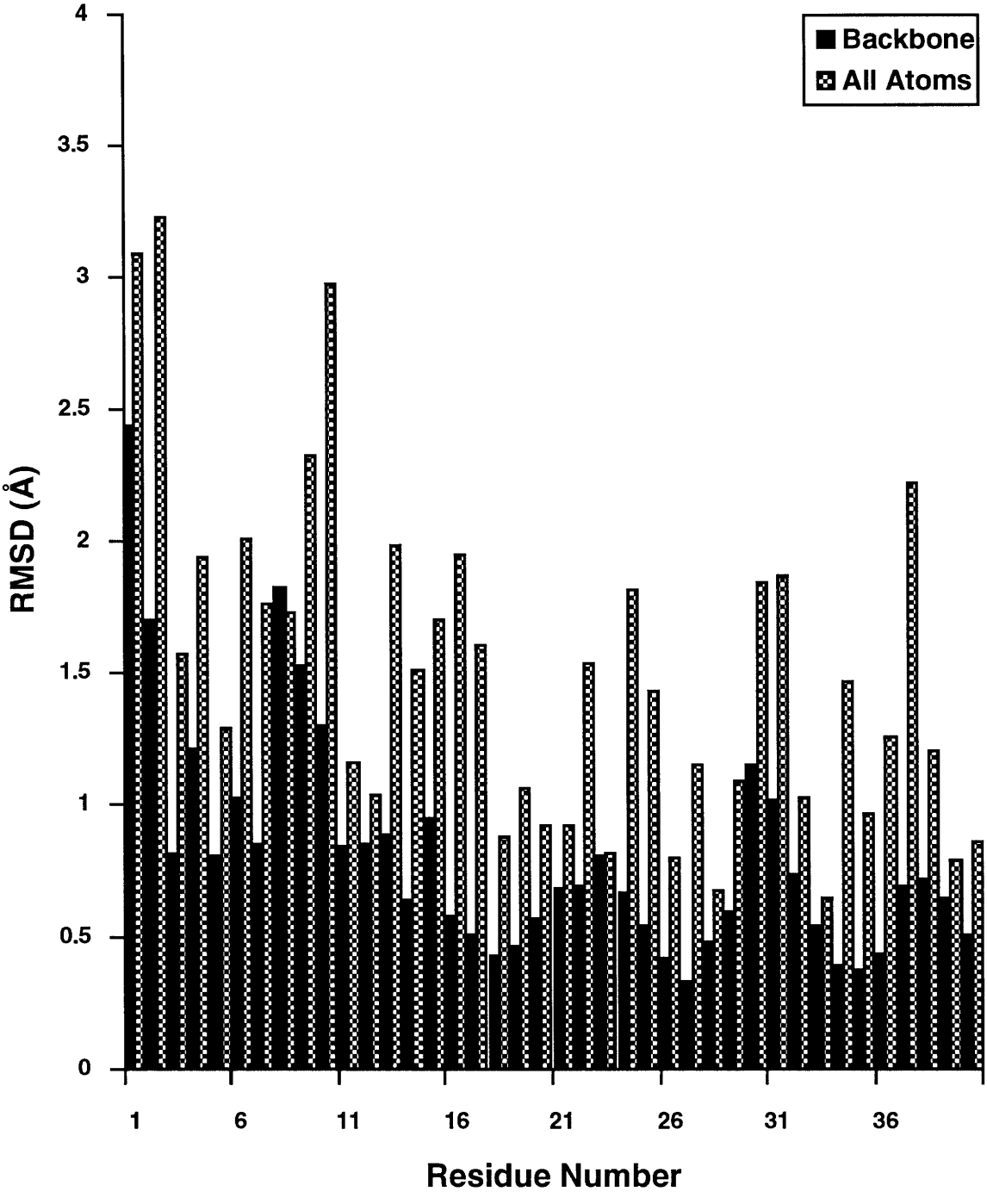


Figure 5-25: A combined histogram for the RMSD of the atomic positions of the backbone, in black, and all atoms, cross hatched.

Table 5-2: Structural Statistics

	$\langle SA \rangle$ [a]	$\langle SA \rangle_r$ [b]	Metzler [73]
RMSD			
violations for all NOE constraints	0.060	0.060	0.047
bond violations [64]		0.004	0.004
angle violations [64]		0.98	2.78
improper violations [64]		0.64	1.06
Energy			
E_{TOTAL}	127.2	112.9	
E_{NOE}	28.6	30.1	48.0
RMSD for backbone atoms		0.92	0.64[c]
RMSD for all nonhydrogen atoms		1.50	
RMSD for all atoms	3.56	2.76	

[a] DG/SA regularized structures.
[b] High temperature SA refined structures.
[c] This value includes only residues 5 to 38.

The structure was chosen on the basis that it had no NOE violations > 0.5 Å, no χ_1 constraint violations > 40°, and the lowest E_{NOE} and E_{TOTAL} force field values out of a family of structures. The structure is best defined in two of the strands of the triple stranded β-sheet, residues Y17 to S20 and V27 to Y29 as well as in the α-helix from residues K32 to C39. Plots of the RMS distance deviations are found in Figures 5-23, 5-24 and 5-25. For the backbone atoms, the structure is well defined except for the first two residues, residue in a loop region from 8 to 10 and a short loop consisting of residues 29 and 30.

The sidechain atoms are disordered in this structure except in the central strand of the β-sheet, residues 17 to 20, and in the last two residues of the α-helix, residues 39 and 40.

This is due to the difficulty, especially in the 3D data, of unambiguously identifying structural NOEs between sidechain protons. Another factor contributing to the disorder seen in the sidechains is the absence of the 236 intraresidue NOE constraints identified from the structure calculations. Inclusion of these constraints should improve sidechain resolution. These constraints were omitted from the current distance constraint set in a effort to decrease structure computation time.

Structural Details. At a gross structural level *Amb. t.* V can be viewed as two structural elements held together by four disulfide bonds. The elements are a short triple stranded antiparallel β-sheet connected by two irregular loops regions from residue 1 to 30 and an α-helix from residues 31 to 40 lying over the top of the sheet. The β-sheet is held together by two disulfide bonds C11-C26 and C18-C28. The two structural elements are held in place by the other two disulfide bonds C5-C35 and C19-C39.

The α-helical region from 31 to 39 is not a 3-10 helix. This was determined by the detection of one $d_{\alpha N}(i, i+4)$ NOE and by the detection of several weak $d_{\alpha N}(i, i+2)$ NOEs. The $d_{\alpha N}(i, i+2)$ NOE is expected to be medium in strength if the helix is 3-10. The α-helix extends from residue 31 to 39 with residue 40 being slightly disordered.

The NMR secondary structure analysis predicted the presence of a short triple stranded anti-parallel β-sheet. The three strands of the sheet are observed to run from residues 5 to 7, 17 to 20, and 27 to 29. As is seen in the structure (Figures 5-20 and 5-22) two strands 5 to 7 and 17 to 20 parallel with the third strand 27 to 29 angled about 10° out of the plane. Two of the strands, 17 to 20 and 27 to 29, are in well-defined regions of the structure (Figure 5-23).

The three β-sheet strands are connected by two irregular loops. One of the loops, residues 8 to 16, contains a type II β-turn [74] between residues 13 and 16 which was predicted previously based on NOE cross peak patterns. There is a puckering of the loop just prior to the β-turn in the sequence. However, this puckering does not fall within the standard classification for a well-defined turn [74]. The second loop, residues 21 to 26,

contains no standard secondary structure elements. Though this loop is not as well-defined as the β-sheet and α-helix regions it is not as disordered as the N-terminus of this protein. There are a number to medium- and long-range NOEs to this loop. Much of the conformational variability is a result of the ϕ torsional flexibility of G23.

The N-terminal residues 1-3 are quite disordered, though there are two long-range NOEs observed for residues 2 and 3. There are no medium- or long-range NOEs to residue 1 which explains the greater than 2 Å deviation observed for the backbone atoms of this residue (Figure 5-23). The flexibility observed in further confirmed by the degenerate shifts for the α protons of G3. The degeneracy suggests that the torsion angles are conformationally averaged on the NMR timescale in solution.

The disulfide bonds for this protein have not, to date, been chemically determined. Thus it was of extreme importance that care be taken in determining the disulfide bonding pattern using NMR since there was no independent check of the assignments. Initial structures included only the unambiguous cysteine $d_{\beta\beta}$ NOEs observed in the 3D NJRTOC experiment (Figures 5-16, 5-17, 5-18, and 5-19). When it was obvious from the preliminary structures that the $d_{\beta\beta}$ NOEs observed placed the cysteine sulfur atoms within bonding distance to form pairs than pseudo-NOEs connecting the sulfur atoms were added to the constraint set. These sulfur-sulfur distance constraints were used for all subsequent structure calculations.

Quality of Structural Calculations. The quality of a structure can be determined by analyzing the number of distance and torsion angle constraint violations and by determining the RMSD for the distance and torsion constraints for a set of structures. However, these criteria do not necessarily define a "good" structure, especially if misassignment of ambiguous constraints has occurred. In an effect to provide additional checks of the quality of the structure we have looked at other structure determination criteria. The first is an examination of the structures for deviation from ideal covalent geometry. In this case, this was done automatically by accepting and selecting only those

structures which deviate from ideal geometry by less than 0.05 Å for all covalent bonds and by less than 5° from dihedral and improper (this refers to planarity and chirality constraints) geometry [64]. Second we checked that the total force field energy values where approximately equal to or less than the values expected for "good" structure. Third, a Ramachandran plot (Figure 5-26) has been used to demonstrate that the Φ and Ψ torsion angles for the average structure lie, with a few exceptions close to sterically allowed regions. The exceptions are residues 2, 29, and 31. Residue 31 is the first residue in the C-terminus α-helix. There are no obvious reason why this residue is found outside the sterically allowed region for α-helix except that the current structure has the backbone reversing direction in two residues from β-strand in residue 29 to α-helix lying over β-strands 5-7 and 17-20. This reversal of direction in two residues in similar to a γ-turn which would allow for the steric strain observed for residue 31. Residues 2 and 29 also lie outside any sterically allowed region. This error is not as bad as it might appear for Y29. There is a sterically allowed region in the lower left-hand corner of the plot which Y29 lies just outside of. Again for similar reasons as found for residue 31, the steric strain observed for this residue was expected. Since D2 is in the disordered N-terminus and has only one long-range constraint finding it in a sterically forbidden region is not unexpected and can be explained as the result of positional disorder due to a lack of distance constraints.

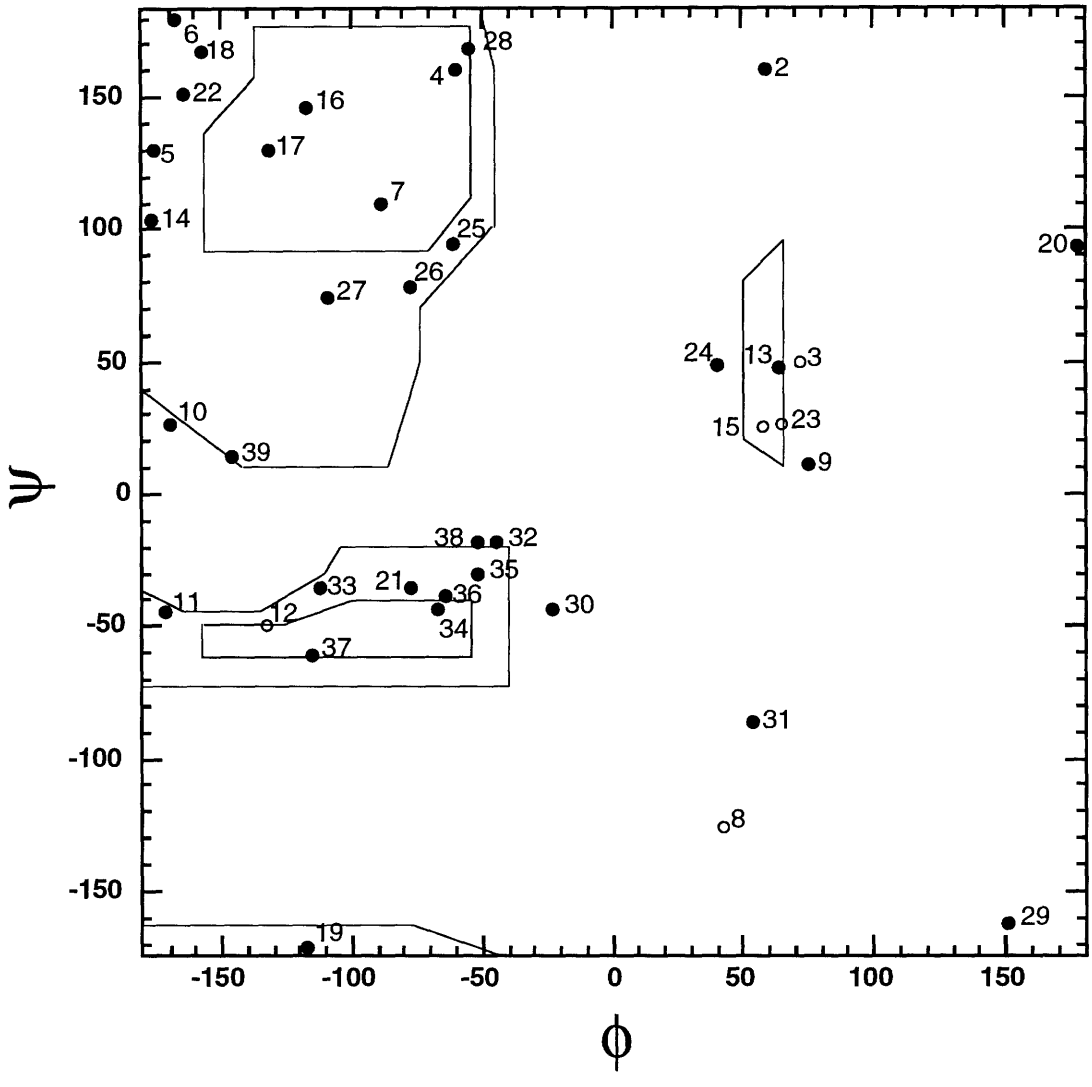


Figure 5-26: A Ramachandran plot of a geometrically averaged, high temperature SA refined structure. Each amino acid is labeled with its residue number. The open circles represent glycine residues and the closed circles represent all other amino acids.

A derived structure should not only fulfill all structural constraints it must also explain any observed chemical shift anomalies. The present structure fulfills that requirement. For example, the β protons for K32 and D2 are shifted upfield substantially. In the present structure the β protons of K32 lie above the π-cloud of Y17. Thus the upfield shift is due to ring current effects. In the current structure the only long-range NOE constraint to D2 is from T40. This places D2 directly at the end of α-helix. Thus one explanation for the anomalous chemical shift observed for D2 is that the β protons are effected by the α-helix dipole moment.

Discussion

The low molecular weight *Amb.* V protein allergens from the pollen of *Ambrosia* (ragweed) are good models for studying the structure function relationships involved in the human immune response. Because these proteins are small and contain a large number of disulfide bonds they are of particular interest for studying the effect of disulfide bonds on immune epitopes. These interactions are not amenable for study by any other systems, in particular peptides.

To date the disulfide bonding pattern has not been determined biochemically. We have demonstrated for the first time that the disulfide bonding pattern can be determined by NMR alone. Our disulfide bonding pattern has two disulfide bonds, C11-C26 and C18-C28, linking and stabilizing the triple stranded β-sheet. The α-helix is held in place and stabilized by the other two disulfide bonds, C5-C35 and C19-C39. This disulfide bonding pattern is in agreement with the 2D NMR solution structure published previously [73]. It does not agree with sequence analysis which suggested a pairing of C5-C19 and C35-C39 [75] or with the 1D NMR work which proposed a pairing of C5-C39 and C19-C35 [76]. The other two disulfide pairings, C11-C26 and C18-C28, correspond to those proposed previously.

The structure presented here is almost identical to the structure published using 2D NMR techniques. The published structure will be referred to as the 2D structure since 2D NMR was used as compared to the 3D NMR techniques used in this study. A comparison of the structural calculation statistics is found in Table 5-2. The differences are minor. The E_{NOE} term obtained for our structure was 60% of the value obtained for the 2D structure. However, the NOE RMSD for the 2D structure was 20% smaller than the value for this structure. The differences in the structure are found in the β-sheet and in the C-terminus. We were able to identify NOE constraints for both residues 39 and 40. Consequently, these residues are well defined structurally. The 2D structure is poorly defined at both the N- and C-terminus e.g. > 4 Å RMSD for backbone at the N-terminus and 2.5 Å RMSD for the C-terminus. The other structural difference is that the 2D structure has tilt in the residues 5 to 7 portion of the β-sheet, with residues 17 to 20 and 27 to 29 being flat and well defined. The current structure has the 5 to 7 and 17 to 20 portion of the β-sheet as flat and the 27 to 29 portion as tilted. This difference is a result of the use of hydrogen bonds in the structure calculation constraint file for the 2D structure. No hydrogen bonds were used in the constraint set for the current structure in an effort to not bias the structure using constraints which can only be inferred and not measured. We have been able to generate a well defined structure with out the use of hydrogen bonds, though some of the tilt observed in the β-strands (Figure 5-22) may result from the lack of these constraints.

The C-terminal region of *Amb. t.* V has been implicated as being the Ia/T cell epitope [26]. It is of interest to note that in the current structure the C-terminal helix is an amphipathic helix. This agrees with the model proposed by Margalit et. al. [13] for Ia/T cell epitopes. This would imply that it is the C-terminal helical region which binds to the MHC class II molecule. It must be emphasized that recent crystallographic data on the related MHC class I molecule indicates that the peptides maybe bound in an extended conformation [77]. This would imply that native structure may not be important for binding

in the MHC molecule. Native structure may, however, be important in the processing required before binding of the epitope to the MHC.

In the current study we have shown that it is possible to identify NOEs to and determine the disulfide bonding pattern in proteins using 3D ^{1}H NMR. The structure of *Amb. t.* V contains a rigid triple-stranded antiparallel β-sheet stabilized by two disulfide bonds from C11-C26 and C18-C28. An eight residue two turn α-helix is tethered to the β-sheet by two disulfide bonds, C5-C35 and C19-C39, such that the charged residues lie on the surface of the molecule and are exposed to the solvent.

References

(1) Marsh, D. G. In *The Antigens*; M. Sela, Ed.; Academic Press: New York, 1975; Vol. III; pp 271-359.

(2) Benjamin, D. C.; Berzofsky, J. A.; East, I. J.; Gurd, F. R.; Hannum, C.; Leach, S. J.; Margoliash, E.; Michael, J. G.; Miller, A.; Prager, E. M.; Reichlin, M.; Sercarz, E. E.; Smith-Gill, S. J.; Todd, P. E.; Wilson, A. C. *Annu. Rev. Immunol.* **1984**, *2*, 67-101.

(3) Kimball, J. W. *Introduction to immunology;* 2nd ed.; Macmillan: New York, 1986, pp 564.

(4) Blackwell, T. K.; Alt, F. W. In *Molecular Immunology*; B. D. Hames and D. M. Glover, Ed.; IRL Press: Oxford, 1988; pp 1-60.

(5) Cebra, J.; Komisar, J.; Schweitzer, P. *Annu. Rev. Immunol.* **1984**, *2*, 493.

(6) Teale, J. M.; Abraham, K. M. *Immunol. Today* **1987**, *8*, 122.

(7) Bjorkman, P. J.; Saper, M. A.; Samraoui, B.; Bennett, W. S.; Strominger, J. L.; Wiley, D. C. *Nature* **1987**, *329*, 506-512.

(8) Brown, J. H.; Jardetsky, T.; Saper, M. A. *Nature* **1988**, *332*, 845-850.

(9) Guillemot, F.; Auffray, C.; Orr, H. T.; Strominger, J. L. In *Molecular Immunology*; B. D. Hames and D. M. Glover, Ed.; IRL Press: Oxford, 1988; pp 81-144.

(10) Davis, M. M. In *Molecular Immunology*; B. D. Hames and D. M. Glover, Ed.; IRL Press: Oxford, 1988; pp 61-80.

(11) Unanue, E. R.; Allen, P. M. *Science* **1987**, *236*, 551-557.

(12) Sette, A.; Buus, S.; Colon, S.; Smith, J. A.; Miles, C.; Grey, H. M. *Nature* **1987**, *328*, 395-399.

(13) Margalit, H.; Spouge, J. L.; Cornette, J. L.; De Lisi, C.; Berzofsky, J. A. *J. Immunol.* **1987**, *138*, 2213-2229.

(14) Marsh, D. G.; Bias, W. B.; Hsu, S. H.; Goodfriend, L. *Science* **1973**, *179*, 691-693.

(15) Marsh, D. G.; Bias, W. B.; Santilli, J.; Schacter Jr, B.; Goodfriend, L. *Immunochemistry* **1975**, *12*, 539-543.

(16) Goodfriend, L. *Dev. Biol. Standard* **1975**, *29*, 41-53.

(17) Goodfriend, L. *Proc. IX Int. Congr. Allergy* **1976**, 151-155.

(18) Adolphson, C.; Goodfriend, L.; Gleich, G. J. *J. Allergy Clin. Immunol.* **1978**, *62*, 197-210.

(19) Goodfriend, L.; Choudhury, A. M.; Klapper, D. G.; Coulter, K. M.; Dorval, G.; Del Carpio, J.; Osterland, C. K. *Molec. Immun.* **1985**, *22*, 899-906.

(20) Lapkoff, C. B.; Goodfriend, L. *Int. Arch. Allergy* **1974**, *46*, 215-229.

(21) Mole, L. E.; Goodfriend, L.; Lapkoff, C. B.; Kehoe, J. M.; Capra, J. D. *Biochemistry* **1975**, *14*, 1216-1220.

(22) Baer, H.; Anderson, M. C.; Hale, R.; Gleich, G. J. *J. Allergy Clin. Immunol.* **1980**, *66*, 281-285.

(23) Roebber, M.; Klapper, D. G.; Goodfriend, L.; Bias, W. B.; Hsu, S. H.; Marsh, D. G. *J. Immunol.* **1985**, *134*, 3062-3069.

(24) Marsh, D. G.; Hsu, S. H.; Roebber, M.; Kautzky, E. E.; Freidhoff, L. R.; Meyers, D. A.; Pollard, M. K.; Bias, W. B. *J. Exp. Med.* **1982**, *155*, 1439-1451.

(25) Coulter, K. M.; Yang, W. H.; Dorval, G.; Drouin, M. A.; Osterland, C. K.; Goodfriend, L. *Molec. Immunol.* **1987**, *24*, 1207-1210.

(26) Huang, S. K.; Marsh, D. G. *Immunology* **1991**, *73*, 363-365.

(27) Huang, S. K.; Zwollo, P.; Marsh, D. G. *Eur. J. Immunol.* **1991**, *21*, 1469-1473.

(28) Zwollo, P.; Ehrlich-Kautzky, E.; Ansari, A. A.; Scharf, S. J.; Erlich, H. A.; Marsh, D. G. *Immunogenetics* **1991**, *33*, 141-151.

(29) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986, pp 292.

(30) States, D. J.; Haberkorn, R. A.; Ruben, D. J. *J. Magn. Reson.* **1982**, *48*, 286-292.

(31) Marion, D.; Ikura, M.; Tschudin, R.; Bax, A. *J. Magn. Reson.* **1989**, *85*, 393.

(32) Brown, S. C.; Weber, P. L.; Mueller, L. *J. Magn. Reson.* **1988**, *77*, 166-169.

(33) Friedrichs, M.; Metzler, W. J.; Mueller, L. *J. Magn. Reson.* **1991**, *95*, 178-183.

(34) Marion, D.; Bax, A. *J. Magn. Reson.* **1988**, *80*, 528-533.

(35) Mueller, L. *J. Magn. Reson.* **1987**, *72*, 191-196.

(36) Bax, A.; Lerner, L. *J. Magn. Reson.* **1988**, *79*, 429-438.

(37) Piantini, U.; Sorensen, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1982**, *104*, 6800-6801.

(38) Rance, M.; Sorenson, O. W.; Bodenhausen, G.; Wanger, G.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *69*, 979-987.

(39) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521-528.

(40) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355-360.

(41) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546-4553.

(42) Kumar, A.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1-6.

(43) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95-117.

(44) Plateau, P.; Guéron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310.

(45) Guéron, M.; Plateau, P.; Decorps, M. *Prog. NMR Spect.* **1991**, *23*, 135-209.

(46) Guéron, M.; Plateau, P.; Kettani, A.; Decorps, M. *J. Magn. Reson.* **1992**, *96*, 541-550.

(47) Sørensen, O. W.; Rance, M.; Ernst, R. R. *J. Magn. Reson.* **1984**, *56*, 527-534.

(48) Rance, M. *J. Magn. Reson.* **1987**, *74*, 557-564.

(49) Shaka, A. J.; Lee, C. J.; Pines, A. *J. Magn. Reson.* **1988**, *77*, 274-293.

(50) Rucker, S. P.; Shaka, A. J. *Molec. Phys.* **1989**, *68*, 509-517.

(51) Vuister, G. W.; Boelens, R.; Kaptein, R. *J. Magn. Reson.* **1988**, *80*, 176-185.

(52) Oschkinat, H.; Cieslar, C.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1989**, *81*, 212-216.

(53) Vuister, G. W.; Boelens, R.; Padilla, A.; Kleywegt, G. J.; Kaptein, R. *Biochemistry* **1990**, *29*, 1829-1839.

(54) Simorre, J. P.; Marion, D. *J. Magn. Reson.* **1991**, *94*, 426-432.

(55) Oschkinat, H.; Cieslar, C.; Griesinger, C. *J. Magn. Reson.* **1990**, *86*, 453-469.

(56) Mooren, M. M. W.; Hilbers, C. W.; van der Marel, G. A.; van Boom, J. H.; Wijmenga, S. S. *J. Magn. Reson.* **1991**, *94*, 101-111.

(57) Wijmenga, S. S.; van Mierlo, C. P. M. *Eur. J. Biochem.* **1991**, *195*, 807-822.

(58) Marion, D.; Ikura, M.; Bax, A. *J. Magn. Reson.* **1989**, *84*, 425-428.

(59) Kline, T. P.; Brown, F. K.; Brown, S. C.; Jeffs, P. W.; Kopple, K. D.; Mueller, L. *Biochemistry* **1990**, *29*, 7805-7813.

(60) Clore, G. M.; Nilges, M.; Sukumaran, D. K.; Brünger, A. T.; Karplus, M.; Gronenborn, A. M. *EMBO J.* **1986**, *5*, 2729-2735.

(61) Clore, G. M.; Gronenborn, A. M. *CRC Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 479-564.

(62) Clore, G. M.; Gronenborn, A. M. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 29-63.

(63) Brünger, A. T.; Nilges, M. *Quart. Rev. Biophys.* **1993**, *26*, 49-125.

(64) Brünger, A. T. *X-PLOR Version 3.1 A System for X-ray Crystallography and NMR*; Yale University: New Haven, 1992, pp 382.

(65) Nilges, M.; Kuszewski, J.; Brünger, A. T. In *Computational Aspects of the Study of Biological Macromolecules by NMR*; J. C. Hoch, Ed.; Plenum Press: New York, 1991.

(66) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *229*, 317-324.

(67) Crippen, G.; Havel, T. *Distance Geometry and Molecular Conformation*; Research Studies Press: Tauton, Somerset, England, 1988, pp 541.

(68) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95-117.

(69) Nilges, M.; Habazettl, J.; Brünger, A. T.; Holak, T. A. *J. Mol. Biol.* **1991**, *219*, 499-510.

(70) Solomon, I. *Phys. Rev.* **1955**, *99*, 559-565.

(71) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546-4559.

(72) Creighton, T. E. *Proteins Structures and Molecular Principles*; W.H. Freeman and Company: New York, 1984, pp 515.

(73) Metzler, W. J.; Valentine, K.; Roebber, M.; Friedrichs, M. S.; Marsh, D. G.; Mueller, L. *Biochemistry* **1992**, *31*, 5117-5127.

(74) Richardson, J. S. *Adv. Protein Chem.* **1981**, *34*, 167-339.

(75) Drenth, J.; Low, B. W.; Richardson, J. S.; Wright, C. S. *J. Biol. Chem.* **1980**, *255*, 2652-2655.

(76) Vidusek, D. A.; Roberts, M. F.; Goodfriend, L. *Biochemistry* **1985**, *24*, 2747.

(77) Madden, D. R.; Gorga, J. C.; Strominger, J. L.; Wiley, D. C. *Nature* **1991**, *353*, 321-325.