UNIVERSITY OF CALIFORNIA
RIVERSIDE

Structure and Dynamics of Biological Macromolecules Using NMR Spectroscopy

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Ye Tian

August 2010

Dissertation Committee:

Dr. Leonard J. Mueller, Chairperson
Dr. Christopher J. Bardeen
Dr. Eric Chronister

The Dissertation of Ye Tian is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

The Text of this thesis, in part or in full, is a reprint of the material as it appears in:

1. Journal of Physical Chemistry B 113, 2596 (2009).

2. Physical Chemistry Chemical Physics 11, 7078 (2009).

The co-author, Dr. L.J. Mueller and Dr. M.E. Hatcher directed and supervised the research which forms the foundation for this dissertation.

Dr. L.L. Chen designed pulse sequences for backbone correlations.

Dr. J.F. Lai performed most of the calculations for theoretical scalar couplings.

Dr. J.M. Kaiser built the triple resonance probe for doing NMR experiments.

Dr. C.M. Rienstra and his students J.M. Boettcher generously provided the sample of uniform labeled GB1 and their valuable suggestions on experiments.

There are many people who contribute to this work. First and foremost, I would like to thank my advisor Professor Leonard Mueller, who is a great teacher, and warm-hearted friend. Without his help it wouldn't be possible to accomplish this work. His devotion to science always encourages me. And he himself shows a good pattern how to do research. I hope, with the knowledge I learned from him and the virtues I witnessed, the completion of my Ph.D. is just the beginning of my journey in science.

I would also like to thank our group, Dr. Lingling Chen, Dr. Michael Kaiser, Arun Agarwal, our very talented new members Yachong Wang, Lingchao Zhu, Olivia Alley. A special portion gives to my good friend, Dr. Jinfeng Lai, who shares his knowledge and wonderful time with me in all of these years. In addition, I also benefited a lot from the help of Dr. Dan Borchardt, who supported me to do NMR experiments and solve problems whenever I need his help; also Dr. Chad Rienstra in UIUC, who provides the

protein sample and put in his professional advice and wise ideas. Dr. Mary Hatcher initializes and guides the DNA work, and generously provides us the sample and her undergraduate helpers.

The assistance of my family is always important, my parents who are been coming from China to help us take care of children, supporting us behind the curtain and bringing blessings to me. My daughters, Christina and Teresa, give me a lot of joy in the busy days. The noise they make turns my entire life into music. The last but never the least, lots of affection and appreciation go to my wife Bing, who is willing to build together with me in love. The time we read books in the evening really makes us into oneness. I believe our life journey will be full of pleasure and enjoyments into the uttermost.

ABSTRACT OF THE DISSERTATION

Structure and Dynamics of Biological Macromolecules Using NMR Spectroscopy

by

Ye Tian

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, August, 2010
Dr. Leonard J. Mueller, Chairperson

Nuclear magnetic resonance is a robust tool in a broad range of scientific area from biochemistry, material science to MRI in medical application. In this thesis, two distinct works are presented in both solid sate and liquid state NMR. Firstly, benefiting from the fast MAS and cleverly designed devices, both resolution and sensitivity of solid state NMR spectra have been improved many folds, so that it is available to study complex biological solids, such as proteins. However there is always a need to increase resolution in order to work on more complicate systems. Here, we introduce a series of highly resolved scalar-based three-dimensional homonuclear correlation experiments for ^{13}C sidechain correlation in solid-state proteins. These experiments are based on a sensitive constant-time format, in which homonuclear scalar couplings are utilized for polarization transfer, but decoupled during chemical shift evolution, to yield high resolution in indirect dimensions and band selectivity as desired. Together with the

experiments designed to obtain backbone correlations, we could fully assign the backbone and aliphatic sidechain chemical shifts with the 3D spectra that are collected on 9.4T magnet (^{1}H frequency 400MHz) for model protein GB1. We also discuss the method of chemical shift based structural refinement.

Another practice of NMR in my work is using ^{31}P Dynamic NMR to characterize the backbone conformation and dynamics in DNA Dickerson Dodecamer. The results confirm solid-state ^{2}H-NMR experiments showing that the C3pG4 and C9pG10 steps experience unique dynamics. And cytosine methylation has significant impact on the local dynamics. The results also show that ^{31}P Dynamic NMR is an efficient way to extract DNA backbone dynamic information, and provide detail knowledge to study DNA-protein interactions.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS…… iv

ABSTRACT …… vii

TABLE OF CONTENTS …… ix

LIST OF FIGURES …… xii

LIST OF TABLES AND SCHEMES……. xiv

CHAPTER I: INTRODUCTION

1.1 OVERVIEW…… 1

1.2 BASIC TECHNIQUES AND METHODS…….. 3

1.2.1 MAGIC ANGLE SPINNING……. 3

1.2.2 CROSS POLARIZATION……..…. 4

1.2.3 DECOUPLING……..…. 5

1.2.4 SCALAR COUPLINGS IN SOLID STATE NMR……. 6

1.2.5 CHEMICAL EXCHANGE……. 7

1.2.6 NUCLEAR OVERHAUSE EFFECT AND HSQC…… 8

REFERENCE…… 10

CHAPTER II: THROUGH-BOND CORRELATION SPECTROSCOPY IN SOLID-STATE PROTEINS AND CHEMICAL SHIFT BASED STRUCTURE REFINEMENT

ABSTRACT……. 13

2.1 INTRODUCTION……... 14

2.2 J-BASED CORRELATION EXPERIMENTS…… 18

2.2.1 PULSE SEQUENCES AND PRODUCT OPERATORS…… 18

2.2.2 EXPERIMENTS…….. 25

2.2.3 SAMPLE PREPARATION…….…. 26

2.2.4 T2 MEASUREMENTS…… 28

2.3 CHEMICAL SHIFT ASSIGNMENTS OF SOLID GB1……. 30

2.3.1 BACKBONE CHEMICAL SHIFT ASSIGNMENTS……..... 30

2.3.2 SIDECHAIN CHEMICAL SHIFT ASSIGNMENTS……..... 33

2.3.3 COMPARISON TO DIPOLAR-DRIVEN METHOD…… 36

2.3.4 DISCUSSION……. 37

2.4 STRUCTRUE REFINEMENT WITH CHEMICAL SHIFT…….. 45

2.4.1 INTRODUCTION ……. 46

2.4.2 STRUCTURE REFINEMENT …….. 47

2.4.3 DISCUSSION …… 51

2.5 CONCLUSION…….. 61

REFERENCE…….. 62

CHAPTER III: ^{31}P NMR INVESTIGATION OF BACKBONE DYNAMICS IN DNA BINDING SITES

ABSTRACT…….. 70

3.1 INTRODUCTION……. 71

3.2 THEORY OF TWO-SITE FAST EXCHANGE MODEL IN NMR…….. 76

3.3 APPLICATION IN DNA BACKBONE DYNAMICS……. 81

3.3.1 SAMPLE PREPARATION…… 81

3.3.2 EXPERIMENTAL SETUP……. 81

3.3.3 TEMPERATURE AND SPECTRA CALIBRATION…… 82

3.3.4 NOESY AND HSQC ASSIGNMENTS…… 84

3.3.5 LINEWIDTH ANALYSIS……. 87

3.3.6 DISCUSSION……..... 100

3.4 CONCLUSION……. 104

REFERENCE…….. 105

APPENDIX

A.1 PULSE CODES…… 109

A.2 CSROSETTA INPUT FILES…… 120

A.3 MATHEMATICA CODES FOR DYNAMICS ANALYSIS…… 132

A.4 GB1 SIDECHAIN WALK FIGURES…… 136

LIST OF FIGURES

CHAPTER 2:

Figure 2.1 Pulse sequences of constant-time *J*-MAS CBCACO and CACBCG 3D homo-nuclear correlation experiments…… 19

Figure 2.2 CBCACO and CACBCG 3D correlation spectra for GB1 at 9.4 T…… 24

Figure 2.3 2D NCA correlation spectrum of N-terminal region of uniform labeled GB1…… 32

Figure 2.4 Backbone sequential walk for GB1…… 33

Figure 2.5 Sidechain walk for GB1…….. 34

Figure 2.6 CBCACO correlation spectrum of TS at 9.4T…… 41

Figure 2.7 Constant-time J-MAS CACO correlation spectrum of TS at 21.15 T… 43

Figure 2.8 Constant-time J-MAS CACO correlation spectrum of TS at 14.1 T….. 44

Figure 2.9 J-gb1 of 10 lowest structure ensemble from CS-Rosetta…… 49

Figure 2.10 Superimposed GB1 structures…….. 50

Figure 2.11 Superimposed CS-2JSV and J-gb1…… 51

Figure 2.12 Plot of 3-residue fragments bbRMSD of CS-2JSV and 2JSV, relative to 2QMT…….. 53

Figure 2.13 Plots of Rosetta all-atom energy versus C_α RMSD…….. 55

Figure 2.14 Accuracy of structures generated in CSrosetta……. 56

Figure 2.15 Plots of 3- and 9-residue fragments bbRMSD of J-gb1 with incomplete chemical shift input……... 60

CHAPTER 3:

Figure 3.1 Potential energy diagram of two-state conformational exchange…… 76

Figure 3.2 Temperature dependent ^{31}P NMR spectra…… 81

Figure 3.3 Temperature dependent ^{31}P 1D spectra of H_3PO_4 and DNA sample…… 83

Figure 3.4 Distances of H2'(T8)-H3'(T8), H2'(T7)-H6(T8) and H2'(T8)-H6(T8).. 84

Figure 3.5 A representative NOESY spectrum of the native Dickerson sequence… 85

Figure 3.6 A representative HSQC spectrum of the native Dickerson sequence….. 86

Figure 3.7 The C9pG10 ^{31}P resonances Lorentzian fits and residuals…….. 87

Figure 3.8 ^{31}P chemical shifts versus temperature for seven sequences……... 88

Figure 3.9 The calculated BII percentage in the native Dickerson…… 90

Figure 3.10 The calculated BII percentage in C3 methylated Dickerson……. 91

Figure 3.11 The calculated BII percentage in C3 methylated Dickerson……. 91

Figure 3.12 The percent BII at 25 °C as a function of position…… 96

Figure 3.13 Line widths as a function of temperature for the C9pG10 step……. 97

Figure 3.14The energy difference between the transition state and most stable conformation for each nucleotide step…… 98

LIST OF TABLES AND SCHEMES

CHAPTER 2:

Table 2.1 Phase Cycling for CACBCO and CACBCG…… 20

Table 2.2 Echo T2(T2') for GB1…….. 29

Table 2.3 Types of correlations on 3D spectra……. 30

Table 2.4 Comparison of bbRMSD…….. 52

Scheme 2.1 Sidechain labeling and typical J-coupling constants…… 16

Scheme 2.2 Backbone sequential walk strategy…….. 31

Scheme 2.3 Magnetization transfers pathways…… 36

CHAPTER 3:

Table 3.1 %BII for each nucleotide step in native sample…… 94

Table 3.2 Calculated thermodynamic parameters for each nucleotide step…….. 95

CHAPTER I

INTRODUCTION

1.1 OVERVIEW

Since the first description and measurement of Nuclear Magnetic Resonance in molecular beams in 1938, NMR has gained great success and rapid development in the past seven decades, three Nobel Prizes have issued to the work in this field. Together with its “twin brother”, Magnetic Resonance Imaging (MRI), the practice of NMR covers a broad range from academic research to medical applications. NMR also powers the development of technology in physics, material science, biochemistry etc. Profiting from the high field magnet, which is made possible by superconductors, Fourier transform and multi-dimensional NMR spectroscopy, both resolution and sensitivity of NMR have improved hundreds of folds comparing to its initial prototype. This revolutionary breakthrough highlights its ability in studying macromolecules, not only in structure determination but also dynamics. Due to the unique information that is provided by high resolution liquid state NMR, it plays an important role to discover structure and function of biological molecules. Solid state NMR inherits the success from the solution state application, and stands on the firm theoretical foundation; since the beginning of this century, it exhibits the great capacity of structural refinement of

solid proteins, both in the pure state and in the biological environment. Less requirements on the sample quality makes solid state NMR more accessible to variety of biological solids. This technique is becoming an essential probe to structural biology, and function study of bio-molecules.

In this thesis, I discuss the applications of both solid sate and liquid state NMR in protein structure refinement and DNA dynamics study respectively. In chapter II, we describe a novel way to achieve backbone chemical shift assignments, also pulse sequences to get aliphatic sidechain correlations, and the probability of obtaining high resolution structures with help of computation chemistry. Unlike most of the methods utilized in solid state NMR, we design our approach based upon scalar coupling interactions instead of dipolar coupling interactions. Brief introduction can be found in this chapter, and details are discussed in the content of chapter II. Chapter III shows the method to extract dynamic information of DNA double strand complexes, in the sense of fast chemical exchange. Some basic techniques and methods that are employed in the content are first introduced in the following sessions.

1.2 BASIC TECHNIQUES AND METHODS

1.2.1 MAGIC ANGLE SPINNING

As the name shows, solid state NMR study samples in solid phase. Unlike the condition in solution state, in solid state molecules undergoes restricted motions, therefore each molecule is experiencing slightly different electronic environment, and gives different NMR resonances. Thus, broad lines are observed on solid state NMR spectra, and make the analysis impossible. However, this isn't a problem in solution state NMR, for the fast molecular motions average out this difference. To solve this problem, brilliant idea of Magic Angle Spinning comes into play, first reported by Raymond Andrew[1] in Nottingham University and Lowe[5] in Washington University. The NMR Hamiltonian[6,7,8] is described as following,

$$\mathcal{H} = \mathcal{H}_Z + \mathcal{H}_{rf} + \mathcal{H}_{CS} + \mathcal{H}_Q + \mathcal{H}_{SR} + \mathcal{H}_D + \mathcal{H}_J$$

In NMR rotating frame[2,6], we assume there are only five terms for the nuclear spin-1/2 Hamiltonian of solid, they are,

$$\mathcal{H} = \mathcal{H}_{CSA} + \mathcal{H}_{ICS} + \mathcal{H}_{rf} + \mathcal{H}_D + \mathcal{H}_J$$

$\mathcal{H}_{CSA}$ and $\mathcal{H}_{ICS}$ are chemical shift anisotropy[3,4] and isotropic chemical shift respectively. $\mathcal{H}_D$ and $\mathcal{H}_J$ are dipolar (direct) and scalar (indirect) coupling Hamiltonians. $\mathcal{H}_{rf}$ is the radio frequency Hamiltonian. In solids, instead of simplifying this express, another operator is added, and it spins the sample rapidly in the magnetic field. The sample tube

(rotor) lean an angle θ_m (which is ~54.74^o , here $\cos^2 \theta_m = 1/3$) to the static field (B_0), and spin at a frequency from 1 to 70 kHz. The results of this are, first chemical shift anisotropy bands are collected to its isotropic chemical shift values; second dipolar couplings are cancelled. Thus the solid state spectra gain shape lines similar to what is in the liquid state NMR. MAS is also applied in quadrupolar cases which associated with two angles (30.6^o and 70.1^o), yet it beyond the discussion of this thesis.

1.2.2 CROSS POLARIZATION

Another routine technique in most of solid state NMR experiments is cross polarization (CP)[10], invented by Michael Gibby and Alexander Pines[9] during their graduate study. By doing this, the polarization of dilute spins (such as ^{13}C) is boosted by the abundant spin (such as ^{1}H). Hence, signal of the rare spins is increased significantly. This technique has become an essential module in all the solid state pulse sequences. In practice, to achieve this effect spin locks (which are series of π pulses to lock the magnetization on transverse plane) are carried out for both the abundant and dilute spins. If the Hartmann-Hahn condition is satisfied according to

$$\gamma_I H_{1I} = \gamma_S H_{1S}$$

the rapid process that are proceeded through dipolar interactions will reach equilibrium quickly. The result is the increased magnetization of rare spin (S-spin), and slightly

decreased magnetization of rich spin (I-spin). Besides enhancing polarization of rare spins, this technique is also used in multi-dimensional correlation spectroscopy to transfer magnetization between nuclei. In this work, we apply cp at the beginning of each pulse sequence with the incremental power for two milliseconds. The cp conditions are optimized for each experiment by seeking the highest intensity of 1D spectrum.

1.2.3 DECOUPLING

The cross polarization happens because of the efficient X-H dipolar interaction. However, in solid state NMR spectroscopy, this effect could also cause significant line broadening, and kill the spectral resolution. Hence many efforts have been put on design broadband heteronuclear decoupling sequences to minimize this effect, such as CW decoupling, TPPM[11], XiX[12], SPINAL64[13], SDROOPY-1[14]. All of these decoupling sequences are sharing similar idea of rapidly flipping proton magnetization, so that the other nuclei experience an average dipolar interaction. In all of our J-based experiments, we use SPINAL64 at high decoupling power on proton. The scheme of SPINAL64 is described as below, according to the literature[13],

$$\text{SPINAL64} = Q\overline{Q}\overline{Q}Q \quad \overline{Q}QQ\overline{Q}$$

where Q is the basic element that is composed of eight pulses with opposite phases $\overline{Q}$ is the counterpart of Q. The scheme of Q and $\overline{Q}$ blocks are shown as:

$$Q = P(10^o)\ P(-10^o)\ P(15^o)\ P(-15^o)\ P(20^o)\ P(-20^o)\ P(15^o)\ P(-15^o)$$

$$\overline{Q} = P(-10^o)\ P(10^o)\ P(-15^o)\ P(15^o)\ P(-20^o)\ P(20^o)\ P(-15^o)\ P(15^o)$$

Values in the brackets are the phases of each pulse, and the pulse (P) is optimized for each individual experiment by setting pulse length as 3 microseconds and searching for best decoupling power.

1.2.4 SCALAR COUPLINGS IN SOLID STATE NMR

Scalar coupling[15,16], also known as J-coupling and through-bond coupling, is in much small magnitude (ten to hundred Hertz) comparing to the dipolar coupling (a few hundred Hertz to thousand) in solid state. And it is often ignored when do solid state NMR experiments. However, in the solid sate NMR work discussed in this thesis, we provide good results from J-based pulse sequences showing good reasons we shouldn't ignore it. And they are briefly mentioned as following. Firstly, under fast MAS condition (70kHz is commercially available up to the date of completion of this thesis), the lines become narrow enough to make the scalar couplings even measurable in 9.4T (^{1}H sequence: 400MHz) magnet. Secondly, scalar coupling provide unique covalent bonding information of molecules. At the meantime, dipolar coupling based experiments build correlations between nuclei that are close in space, this nature makes the dipolar spectra unacceptable complicate when we apply it on big bio-molecules, sometimes impossible

to do assignment. This isn't a problem in the scalar coupling based experiments, J-based spectrum only shows the connectivity that is possible through covalent bonds, therefore it is easier to analyze. This feature is especially useful in assigning spectra of big proteins. By introducing series of 3D J-based experiments, this advantage above highlights the contribution of this work to solid state NMR in solving protein structures. No matter what, taking the advantage of fast MAS and novel proton decoupling methods, scalar coupling based experiments become equally important as its dipolar coupling cousins. To thoroughly understand a biological process, it really requires both.

1.2.5 CHEMICAL EXCHANGE

In chapter III, I focus on a method of extracting dynamic information of DNA binding sites. Although NMR is the method that applies to the equilibrium systems, strong evidence is also shown on some NMR spectra that molecules undergoes chemical exchange process[17-24]. That is certain nuclear spin behave as a label through it we could monitor the local or global molecular movement from one chemical site or conformation to another. According to different exchange rate, this process could be divided as slow and fast exchange. In the between they are connected by the coalescence point. When the exchange rate between two site I and II is slow enough so that its magnitude is smaller than the chemical shift difference $|\delta_I-\delta_{II}|$, then the exchange process is in the slow

exchange region. And two distinct lines will be observed on the spectra. In general, the exchange rate is higher at higher temperature, thus we could see the change if we run the same experiment at different temperatures. On the spectra, while the temperature increases, the two distinct lines will start broadening and merge into each other. And the result of this is one broad line, this is a sign that the system is at the coalescence point. Pass coalescence point, the system enter into fast exchange region, where the exchange rate is fast so that $(\tau_{ex})^{-1} >> |\delta_I - \delta_{II}|$, in this region we could observe the line narrows up by increasing temperature. The system we are interested in is in the fast exchange region. In chapter III, I introduce a model that could be applied to the fast exchange and extract barrier heights of DNA binding sites to help understanding it biological process.

1.2.6 NUCLEAR OVERHAUSER EFFECT AND HSQC

In the solution state NMR study, two important methods are routinely used, they are Nuclear Overhauser Effect Spectroscopy (NOESY)[25-32] and Heteronuclear Single Quantum Correlation Spectroscopy (HSQC)[33-36]. They are also the major tools in the DNA dynamics study of this thesis. These methods are developed based on two different mechanisms. NOESY is based on based on dipolar interaction, and the cross peaks on 2D NOESY are generated from two nuclei that are close in space, one spin can experience the local magnetic field of the other. Therefore the cross peaks are the

reflection of the distance of two spins, the closer they are the stronger intensity is. Unlike NOESY, HSQC is correlating the through bond coupling spins and identifying directly connected nuclei. The advantage of HSQC for this project is both ^{1}H and ^{31}P have >99% nature abundance, so there is no need to make isotropic labeled samples. The coupling constant of ^{1}H-^{31}P is matched to gain maximum sensitivity on the 1D version of HSQC. The details are discussed in the content.

REFERENCES

1. Andrew, E. R.; Bradbury, A.; Eades, R. G. *Nature* **1958**, *182*, 1659-1659.
2. Becher, E.D. *High Resolution NMR, Theory and Chemical Applications.* Second Edition, Academic Press, New York, **1980.**
3. Herzfeld, J.; Berger, A. E. *Journal of Chemical Physics* **1980**, *73*, 6021-6030.
4. Maricq, M. M.; Waugh, J. S. *Journal of Chemical Physics* **1979**, *70*, 3300-3316.
5. Lowe, I. J. *Physical Review Letters* **1959**, *2*, 285-287.
6. Demco, D. E.; Tegenfeldt, J.; Waugh, J. S. *Physical Review B* **1975**, *11*, 4133-4151.
7. Mehring, M. *Principle of High Resolution NMR in Solids*, Second Edition, **1983**.
8. Haeberlen, U. *High Resolution NMR in Solids: Selective Averaging*, **1976**
9. Pines, A.; Gibby, M. G.; Waugh, J. S. *Journal of Chemical Physics* **1973**, *59*, 569-590.
10. Pines, A.; Waugh, J. S.; Gibby, M. G. *Journal of Chemical Physics* **1972**, *56*, 1776-&.
11. Bennett, A. E.; Rienstra, C. M.; Auger, M.; Lakshmi, K. V.; Griffin, R. G. *Journal of Chemical Physics* **1995**, *103*, 6951-6958.
12. Detken, A.; Hardy, E. H.; Ernst, M.; Meier, B. H. *Chemical Physics Letters* **2002**, *356*, 298-304.

13. Fung, B. M.; Khitrin, A. K.; Ermolaev, K. *Journal of Magnetic Resonance* **2000**, *142*, 97-101.

14. De Paepe, G.; Lesage, A.; Emsley, L. *Journal of Chemical Physics* **2003**, *119*, 4833-4841.

15. Hahn, E. L.; Maxwell, D. E. *Physical Review* **1952**, *88*, 1070-1084.

16. Alexander, S. *Journal of Chemical Physics* **1962**, *37*, 967-&.

17. Gutowsky, H. S.; Belford, G. G.; McMahon, P. E. *Journal of Chemical Physics* **1962**, *36*, 3353-&.

18. Newmark, R. A.; Sederhol.Ch *Journal of Chemical Physics* **1965**, *43*, 602-&.

19. Alger, T. D.; Gutowsky, H. S.; Vold, R. L. *Journal of Chemical Physics* **1967**, *47*, 3130-&.

20. Govil, G.; Bernstei.Hj *Journal of Chemical Physics* **1967**, *47*, 2818-&.

21. Binsch, G. *Journal of the American Chemical Society* **1969**, *91*, 1304-&.

22. Vega, A. J.; Fiat, D. *Journal of Magnetic Resonance* **1974**, *13*, 260-267.

23. Mueller, L. J.; Weitekamp, D. P. *Science* **1999**, *283*, 61-65.

24. Bain, A. D. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2003**, *43*, 63-103.

25. Overhauser, A. W. *Physical Review* **1953**, *92*, 411-415.

26. Solomon, I. *Physical Review* **1955**, *99*, 559-565.

27. Freeman, R.; Pachler, K. G. R.; Lamar, G. N. *Journal of Chemical Physics* **1971**, *55*, 4586-&.

28. Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *Journal of Chemical Physics* **1979**, *71*, 4546-4553.

29. States, D. J.; Haberkorn, R. A.; Ruben, D. J. *Journal of Magnetic Resonance* **1982**, *48*, 286-292.

30. Bodenhausen, G.; Kogler, H.; Ernst, R. R. *Journal of Magnetic Resonance* **1984**, *58*, 370-388.

31. Bax, A.; Davis, D. G. *Journal of Magnetic Resonance* **1985**, *63*, 207-213.

32. K. Wuthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**

33. Bodenhausen, G.; Ruben, D. J. *Chemical Physics Letters* **1980**, *69*, 185-189.

34. Kay, L. E.; Keifer, P.; Saarinen, T. *Journal of the American Chemical Society* **1992**, *114*, 10663-10665.

35. Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. *Journal of Magnetic Resonance* **1991**, *93*, 151-170.

36. Kontaxis, G.; Stonehouse, J.; Laue, E. D.; Keeler, J. *Journal of Magnetic Resonance Series A* **1994**, *111*, 70-76.

CHAPTER II

THROUGH-BOND CORRELATION SPECTROSCOPY IN SOLID-STATE PROTEINS AND CHEMICAL SHIFT BASED STRUCTURE REFINEMENT

ABSTRACT

Scalar-based three-dimensional homonuclear correlation experiments are introduced for ^{13}C sidechain correlation in solid-state proteins. These experiments are based on a sensitive constant-time format, in which homonuclear scalar couplings are utilized for polarization transfer, but decoupled during chemical shift evolution, to yield highly resolved indirect dimensions and band selectivity as desired. The methods therefore yield spectra of high quality that give unique sets of sidechain correlations for small proteins even at 9.4 Tesla (400 MHz ^{1}H frequency). We demonstrate versions of the pulse sequence that enable correlation from the sidechain to the backbone carbonyl as well as purely sidechain correlation sets; together these two data sets provide the majority of ^{13}C–^{13}C correlations for assignment. The polarization transfer efficiency is approximately 30% over two bonds. In the protein GB1 (56 residues), we find essentially all cross peaks uniquely resolved. We find similar efficiency of transfer (~30%) in the 140 kDa tryptophan synthase (TS), since the relaxation rates of immobilized solid proteins are not sensitive to global molecular tumbling, as long as the correlation time is

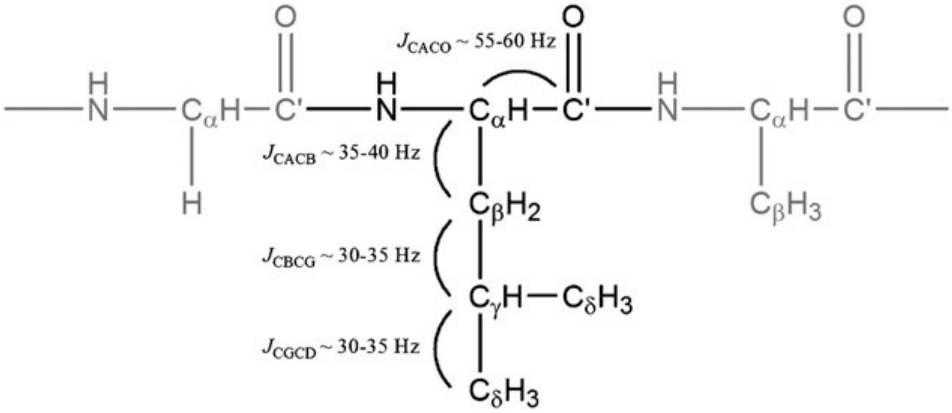
much longer than the magic-angle spinning rotor period. In 3D data sets of TS at 400 MHz, some peaks are resolved and, in combination with higher field data sets, we anticipate that assignments will be possible; in this vein, we demonstrate 2D ^{13}C–^{13}C spectra of TS at 900 MHz that are well resolved. These results together provide optimism about the prospects for assigning the spectra of such large enzymes in the solid state.

2.1 INTRODUCTION

In structural studies involving nuclear magnetic resonance (NMR), chemical shift assignment is an essential step, usually performed first in the analysis of experimental data. Multi-dimensional correlation spectroscopy plays a central role in this task, and a variety of experiments have been developed to transfer magnetization between pairs of nuclear spins and establish their connectivity via cross-peaks in multidimensional NMR spectra. Both the indirect spin–spin (J-) coupling and the dipolar coupling between nuclear spin pairs provide mechanisms to effect this transfer. The J-coupling denotes through-bond covalent contact, while the dipolar coupling reports on spatial proximity. Correlation spectroscopy in solids has traditionally favored the use of the dipolar coupling,[1–7] which, at 42 kHz for directly bonded carbons, is much larger than the typical 30–75 Hz J-coupling for the same spin pair. Over the last ten years, however, there has been growing interest in scalar-based methods,[8-28] which provide high sensitivity and

information complementary to dipolar experiments. While dipolar methods report on spatial proximity and require relatively short mixing times, they can be averaged by molecular motion, a phenomenon common not only in proteins, but also organic small molecules and inorganic materials.[17,19,23,29] The J-coupling, however, has a non-zero isotropic average, enabling applications in situations where dipolar methods fail. Another distinction of the scalar coupling is that it is restricted to the covalent network, allowing for correlation in the absence of intermolecular transfers. This can provide for the identification (and assignment) of inequivalent conformations in the same crystalline unit cell.[13,30] Of course, the distance dependence of the dipolar coupling makes it an important parameter for developing 3D structural models, and most full structure determinations in NMR rely on a combination of dipolar and scalar techniques. While the majority of this themed issue focuses on materials chemistry applications of solid-state NMR, this article is meant to present a comparative example of related methods used in biological solids. As in materials chemistry applications, solid-state NMR is capable of providing atomic-level 3D structural information even for macroscopically disordered samples. To accomplish this, a number of robust two-dimensional experiments for ^{13}C–^{13}C correlation in biological solids have been developed, based on both the strong and weak scalar coupling Hamiltonians.[8,9,11–14,16,18,20] Scalar based ^{15}N–^{13}C experiments have also been demonstrated as 2D experiments and,

more recently, as 3D experiments for assigning backbone resonances in solid-state proteins.[17,19,21–23] 3D correlation spectroscopy has proven to be an effective tool in the assignment of chemical shifts, benefiting from the increase in resolution as resonances are dispersed into three frequency dimensions. Backbone assignment is critical from the standpoint of overall conformational fold and 3D structure. Equally important to protein function are the sidechains, which mediate protein recognition and enzymatic activity. The ^{13}C chemical shifts of the sidechains provide information on direct conformational restraints (such as insights into sidechain rotameric states), and the majority of valuable long-range distances rely upon spectral resolution of the sidechain ^{13}C chemical shifts, especially for methyls and aromatics, which are principally located in the hydrophobic core of the protein. Therefore complete assignments of sidechains are highly valuable for structural studies.



Scheme 2.1 Sidechain labeling and typical homo-nuclear ^{13}C J-coupling constants, highlighted here for leucine in a three amino acid segment, –glycine–leucine–alanine–.

In this work, we use scalar-based sidechain to backbone carbonyl (Scheme 2.1) and purely sidechain selective 3D correlation experiments for assignment of chemical shifts in solid-state proteins under magic-angle-spinning (MAS) conditions. In the nomenclature of biological NMR, the former experiment is referred to as CBCACO correlation as it correlates the sidechain β-carbon (C_β) to the α-carbon (C_α) and then the backbone carbonyl (C'). The latter experiment is referred to as CACBCG correlation (for C_α– C_β– C_γ), although it actually traces out stepwise connectivity along the entire aliphatic sidechain. In both correlation experiments, we make use of a constant-time format in which the indirect evolution and transfer periods are combined into a single constant-time interval, resulting in high sensitivity and increased resolution. In the β1 immunoglobulin binding domain of protein G (GB1), this allows us to trace out the connectivity in every aliphatic sidechain of this 56 residue protein, including peaks lying close to the diagonal, such as the methyl resonances on leucine sidechains. We highlight an important distinction between these scalar methods and the popular dipolar methods based on spin diffusion. Specifically, multi-bond transfers are significantly attenuated in scalar transfers relative to those using spin diffusion. Even in smaller proteins, such as GB1, multiband transfers in spin diffusion experiments can lead to assignment ambiguities that are cleared up by scalar methods. In the substantially larger tryptophan synthase, an α2β2 heterodimer which has 665 distinct residues and a dimer

mass of 143 kDa, good correlation can still be observed in the 3D experiment, even at 9.4T.

2.2 J-BASED HOMO- AND HETERO-NUCLEAR CORRELATION EXPERIMENTS

2.2.1 PULSE SEQUENCES AND PRODUCT OPERATORS

Hard π/2 and π pulses at 80 kHz for ^{13}C were used throughout, along with 150 kHz SPINAL64 ^{1}H decoupling[35] during the constant time intervals and the final τ_4–π–τ_4 refocusing period, and 100 kHz SPINAL64 decoupling during the 16 ms z-filter and acquisition (t_3). Selective 260 and 420 μs r-SNOB pulses were used for the aliphatic and carbonyl π pulses, respectively, and rotor-synchronized as described in ref. 36. In these two experiments, the carrier frequency for carbon was placed at the center of the aliphatic carbon resonances (46 ppm). Pure phase spectra were obtained using the method of States *et al*.[37] by incrementing the phase of the spin lock pulse on ^{13}C (f_1) and the π/2 pulse immediately preceding t_2 (f_2). Minimal phase cycles, indicated in Fig. 1 and discussed below, include the first selection of double and zero quantum and a combined phase cycling of the three π pulses. Additional acquisition and processing parameters for each spectrum are included in the respective Figure captions. Assignments were performed with Sparky version 3.[38]

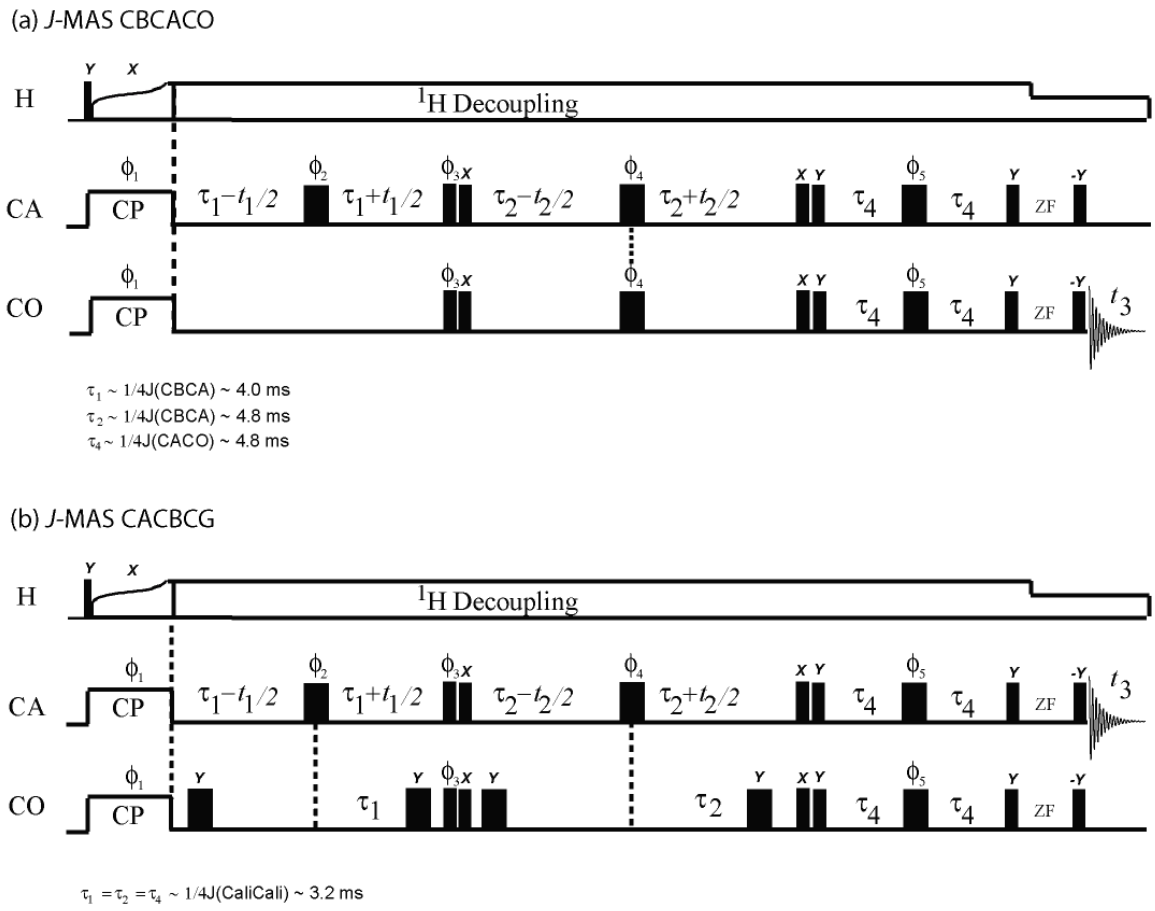


Figure 2.1 Pulse sequences of onstant-time *J*-MAS CBCACO and CACBCG 3D homo-nuclear correlation experiments.

In these pulse sequences, thin vertical lines indicate π/2 pulses and wide vertical lines indicate π pulses that are either selective (single carbon channel) or broadband (shown as simultaneous CA and CO pulses). The indirect evolution increments, τ refocusing, and z-filter periods (ZF) are all rotor synchronized. For CBCACO, τ_1= 4.0ms,

τ_2= 4.8ms, and τ_4= 4.8ms, and for CACBCG, $\tau_1=\tau_2=\tau_4$= 3.2ms. The fixed phases are shown on Figure 2.1. Then, phases ϕ_1, ϕ_2 and ϕ_3 are cycled to create a mixture of zero and double quantum, and the phase of three π pulses are cycled together to choose coherences that change the overall coherence order by ± 2.

Table 2.1 Phase Cycling for CBCACO and CACBCG

Phase	Cycle
ϕ_1	x, -x, x, -x, x, -x, x, -x
ϕ_2	x, -x, y, -y, -x, x, -y, y
ϕ_3	y, -y, y, -y, y, -y, y, -y
ϕ_4	x, x, y, y, -x, -x, -y, -y
ϕ_5	x, x, y, y, -x, -x, -y, -y
ϕ_{rec}	x, x, -x, -x, x, x, -x, -x

As with 2D and 3D J-based MAS experiments, the *J*-MAS 3D CBCACO and CACBCG experiments are inspired by, but not identical to, their solution-state counterparts.[39,40] The spin dynamics are similar for both the CBCACO and CACBCG experiments, so we begin with the former, tracing the desired coherence pathway, beginning on the sidechain C_β and ending on the backbone carbonyl. After cross-polarization from ^1H to ^{13}C to create transverse magnetization on C_β, I_x^{CB} , this initial state is allowed to evolve during the first constant-time echo period under the weak scalar coupling Hamiltonian to the α-carbon, $H_J = 2\pi J_{CACO} I_z^{CA} I_z^{CB}$. This evolution

creates magnetization that is anti-phase with respect to this active coupling, while the chemical shift on C_β is encoded by time-incrementing the central π pulse location:

$$I_x^{CB} \xrightarrow{(\tau_1 - t_1/2) - \pi - (\tau_1 + t_1/2)} \cos(\omega_{CB} t_1) 2 I_y^{CB} I_z^{CA}$$

The transverse component of the anti-phase term is then transferred from C_β to C_α by the two hard $\pi/2$ pulses:

$$\cos(\omega_{CB} t_1) 2 I_y^{CB} I_z^{CA} \xrightarrow{\pi/2 I_y} \cos(\omega_{CB} t_1) 2 I_y^{CB} I_x^{CA}$$
$$\xrightarrow{\pi/2 I_x} \cos(\omega_{CB} t_1) 2 I_z^{CB} I_x^{CA}$$

This transfer is implemented using two pulses, rather than a single pulse, so that the phases ϕ_1, ϕ_2, and ϕ_3 can be cycled to selectively preserve the signal, after the first $\pi/2$ pulse, from states that are combinations of zero and double quantum coherences, such as $I_y^{CA} I_x^{CB}$, while rejecting single-quantum coherences, such as I_x.

During the second constant-time period, evolution proceeds under an additional scalar coupling Hamiltonian between the α and carbonyl carbons, $H_J = 2\pi J_{CACO} I_z^{CA} I_z^{CO}$. The combined evolution under the two scalar coupling interactions during the second constant-time interval achieves two purposes. First, it rephases the anti-phase magnetization on C_α, with respect to the C_β coupling, into in-phase transverse magnetization on C_α. Second, it creates a coherence that is anti-phase with respect to the

coupling between C_α and C'. Once again, the chemical shift of the transverse magnetization (on C_α here) is encoded by time-incrementing the central p pulse location:

$$\cos(\omega_{CB}t_1)2I_z^{CB}I_x^{CA} \xrightarrow{(\tau_2-t_2/2)-\pi-(\tau_2+t_2/2)} \cos(\omega_{CB}t_1)\cos(\omega_{CA}t_2)2I_x^{CA}I_z^{CO}$$

The transverse component of the anti-phase magnetization is transferred from C_α to C' once more by the two hard $\pi/2$ pulses:

$$\cos(\omega_{CB}t_1)\cos(\omega_{CA}t_2)2I_x^{CA}I_z^{CO} \xrightarrow{\pi/2I_x} -\cos(\omega_{CB}t_1)\cos(\omega_{CA}t_2)2I_x^{CA}I_y^{CO}$$
$$\xrightarrow{\pi/2I_y} \cos(\omega_{CB}t_1)\cos(\omega_{CA}t_2)2I_z^{CA}I_y^{CO}$$

and again allowed to refocus into in-phase transverse magnetization, now on the carbonyl carbon, for final detection during t_3:

$$\cos(\omega_{CB}t_1)\cos(\omega_{CA}t_2)2I_z^{CA}I_y^{CO} \xrightarrow{\tau-\pi-\tau} \cos(\omega_{CB}t_1)\cos(\omega_{CA}t_2)I_x^{CO}$$
$$\xrightarrow{t_3} \text{detection}$$

The net result is the detection of magnetization that is frequency labeled by the C_β, C_α, and C' chemical shifts in the t_1, t_2, and t_3 time-domains, respectively, and can be Fourier transformed into a three-dimensional frequency-domain spectrum, giving a cross-peak at coordinate (f_1, f_2, f_3) = (ω_{CB}, ω_{CA}, ω_{CO}). This cross-peak indicates that these three carbons share a covalent bonding network and can be assigned to the same amino acid.

Compared to the CBCACO experiment, the CACBCG experiment has two additional π pulses on the CO channel during the first and second constant-time periods. These

pulses serve to refocus evolution, under the scalar coupling Hamiltonian, to the carbonyl spins as the central π pulse on C_α moves from $t_1 = 0$ or $t_2 = 0$. The first of these π pulses is stationary in time and its sole purpose is to compensate for off-resonance effects from the second pulse, which is maintained at a fixed delay, τ, from the C_α π pulse.

One of the distinguishing features of these experiments is that both indirect dimension evolution times are embedded in constant time periods, providing homo-nuclear decoupling in both indirect dimensions; in previous dipolar versions, homo-nuclear decoupling was achieved but with standard time evolution.[41,42] In both cases, the scalar decoupling improves resolution significantly, but since in the *J*-based version the evolution and coherence transfer are combined into a single constant-time period for each dimension, this increase in resolution does not compromise sensitivity.

Figure 2.2 shows the application of the *J*-MAS CBCACO and CACBCG to uniformly ^{13}C, ^{15}N labeled GB1. Negative intensity in the CACBCG cube corresponds to magnetization that is transferred back to its initial spin during t_2, rather than to a third spin and hence shows up at $\omega_1 = \omega_3$.

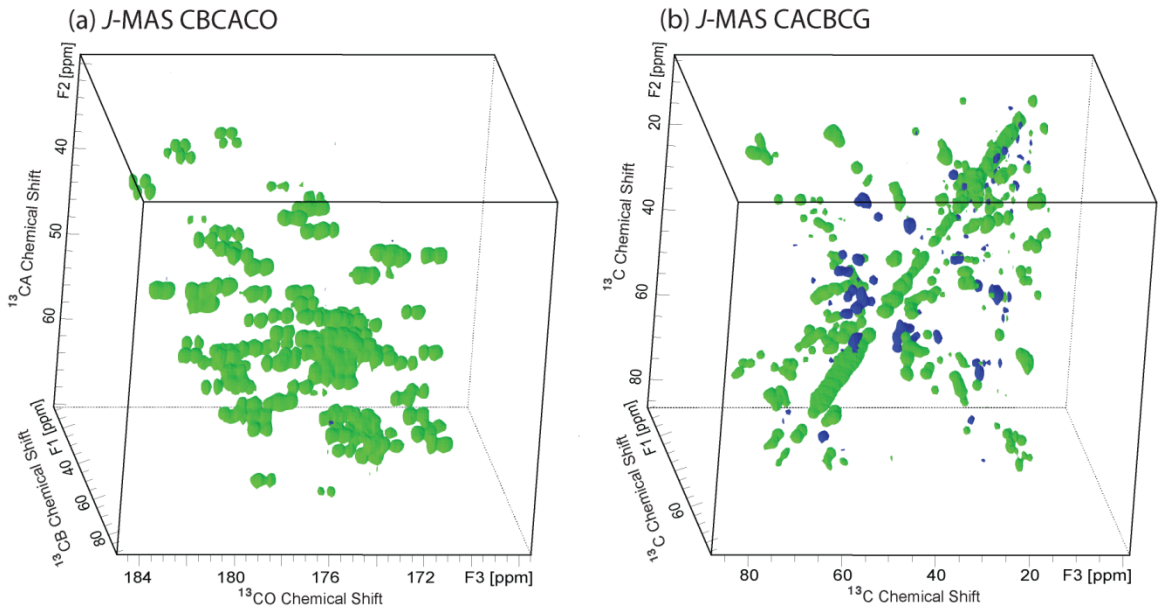


Figure 2.2 Constant-time J-MAS (a) CBCACO and (b) CACBCG 3D homo-nuclear correlation spectra for GB1 at 9.4 T.

Experiment specific parameters for CBCACO: τ_1=4.0ms, τ_2=τ_4= 4.8ms; 8 scans per transient; 2048 complex points with a dwell of 20 μs in t_3 (spectral width:50 kHz, total acquisition time: 41.0 ms), 38 complex points with a dwell of 240 μs in t_2 (spectral width: 4166.667 Hz, total acquisition time: 9.1 ms), and 64 complex points with a dwell of 120μs in t_1 (spectral width: 8333.33 Hz; total acquisition time: 7.7 ms). Total experiment time: 69.5 h. Acquisition parameters for CACBCG: τ_1=τ_2=τ_4=3.2 ms; 8 scans per transient; 1024 complex points with a dwell of 20 μs in t_3 (spectral width: 50 kHz, total acquisition time: 20.5 ms), 46 complex points with a dwell of 120 μs in t_2 (spectral width: 8333.33 Hz, total acquisition time: 5.5 ms), and 46 complex points with a

dwell of 120 μs in t_1 (spectral width: 8333.33 Hz; total acquisition time: 5.5 ms). Total experiment time is 50.5 h.

2.2.2 EXPERIMENTS

3D NMR experiments for GB1 and Tryptophan Synthase were performed using the pulse sequences in Figure 2.1 (described in detail in the discussion section) at 9.4 T on a Bruker DSX spectrometer (^{1}H frequency of 400.13 MHz), with an ^{1}H–^{13}C–^{15}N triple resonance 2.5 mm MAS probe. All experiments were performed with an MAS rate of 25 kHz and with the temperature of the bearing gas cooled to 269 K, resulting in an effective sample temperature of 287±4 K, due to frictional heating as calibrated using a methanol chemical shift thermometer.[34] Chemical shifts were referenced on the DSS (2,2-dimethylsilapentane-5-sulfonic acid) scale using adamantane as an external secondary chemical shift standard at 40.48 ppm (downfield peak).

2D CACO correlation experiments for Tryptophan Synthase were performed using the pulse sequence reported in ref. 18 at 21.15 T on a Bruker Avance II spectrometer (^{1}H frequency of 900.13 MHz) with an ^{1}H–^{13}C–^{15}N triple resonance 2.5 mm MAS probe. All experiments were performed with an MAS rate of 20 kHz and with the temperature of the bearing gas again cooled to 269 K. Hard π/2 and π pulses at 80 kHz for ^{13}C were used

along with 100 kHz SPINAL64 ^{1}H decoupling throughout. Additional acquisition and processing parameters are included in previous session.

2.2.3 SAMPLE PREPARATION

Two ^{13}C and ^{15}N uniform labeled proteins were used in this work, the procedure of preparing these samples are briefly introduced as following.

GB1 sample preparation: The 56 residue β1 immunoglobulin binding domain of protein G was prepared uniformly—^{13}C, ^{15}N—enriched through over-expression and purification as reported.[4] In particular, protein expression will utilize the BL21(DE3) cell line, while DNA amplification (Miniprep) will use NovaBlue cell line (both from Novagen). The cells are first grow in 1L rich media (LB), then transferred into 1L minimal media (2 g/L 15NH4Cl, 2 g/L ^{13}C glucose) with 10 ml 10× ^{13}C and ^{15}N enriched BioExpress (Cambridge Isotope Laboratories). Expression induced by 1 ml of 0.5 M IPTG (isopropyl-beta-D-thiogalactopyranoside). After 4 hrs induction at 37 ℃, the cells are centrifuged down for purification. Since GB1 is a thermal stable protein, the aliquot was put in ~90 ℃ for 10 minutes to lyse cells. After heat, in the supernatant, it is about 90% GB1 already. Fast Protein Liquid Chromatography (FPLC), is used for further purification. We use a Sephacryl S-100 gel filtration column with a 120 mL column volume by GE Healthcare. GB1 fragment is concentrated and dialysis for 3

times with 4 L of 50 mM sodium phosphate buffer (pH=5.5). Before crystallization, GB1 solution is concentrated > 25 mg/mL, determined by UV at 280nm. Here we use 3500 MWCO Centricon YM-3 Centrifugal Concentrator for concentration. A few methods are available to convert protein from solution into solid, microcrystalline precipitation is found to give the best NMR spectra. For our purpose, we used 50% MPD (methyl-2,4-pentanediol) and 25% IPA (isopropyl alcohol) with GB1 at a concentration of 25 mg/mL to obtain microcrystallines. Finally, about 10-15mg of microcrystal is packed into 2.5mm rotor to collect spectra.

TS sample preparation: Escherichia coli CB149 containing plasmid pEBA-10[31] were grown on minimal media (2 g/L 15NH4Cl, 2 g/L ^{13}C glucose) supplemented with 200 mL of 10× BioExpress rich media (Cambridge Isotope Laboratories). Protein expression was induced in the late log-phase with the addition of 0.2 mM isopropyl b-D-1-thiogalactoside for 8–12 h. Tryptophan synthase was purified as previously described32 with the following modifications. In step II, after the initial crystals were formed, the polyethylene glycol (PEG) concentration was increased to 7%. The subsequent wash step was performed with the same PEG concentration. The crystals were dissolved in B buffer, pH 9.0, and stirred at room temperature (RT) for 30 min prior to heating. The heating step was performed at 40–45℃ for up to 20 min to insure that all the crystals had dissolved completely (effectiveness of the heating step was monitored

by centrifuging small aliquots until yellow crystals were no longer visible in the pellet). In step III, recrystallization was carried out for at least 48 h. Pyridoxal-50-phosphate (PLP) was omitted from all buffers in order to retain the labeled cofactor. 1 L of culture yielded typically 170 mg of protein with 27% PLP content. The protein was judged to be >90% pure by SDS-PAGE. Microcrystals were prepared by a modification of the procedure of Mozzarelli et al., 1989.[33] To reconstitute the holoenzyme, concentrated enzyme (300–500 mM) was incubated 1 : 1 with PLP for 30 min at RT. This solution was subsequently mixed 1 : 1 with 50 mM Na–Bicine, pH 7.8, containing 14% PEG, 1.5 mM spermine, 0.3 mM PLP, and 20 mM dithiothreitol (DTT), and stored for at least 3 h at 4 1C. The microcrystals were collected by centrifugation and washed with 50 mM Na–Bicine, pH 7.8, containing 10–20% PEG and 10 mM DTT.

2.2.4 T2 MEASUREMENTS

In each case of these 3D experiments, including backbone experiments (NCACO, NCOCA, CANCO) and sidechain experiments (CBCACO and CACBCG), the pulse sequence takes up to 20 ms (depends on the delay times and length of ZF) before acquisition. This nature requires long T2 relaxation time to obtain good resolution and sensitivity. Therefore, it is necessary to determine T2 for C', C_α and N before running multi dimensional experiments. Uniform labeled GB1 was tested in 9.4T spectrometer,

at different spin rates and decoupling powers. T2 measurements were carried out using a simple τ–π–τ dephasing period following cross polarization from proton. Here π pulse is selective pulse optimized for carbonyl and alpha carbon, centered at C' and C_α, respectively. Hard π pulse is used for nitrogen T2 measurement. A previous member in our group uses multiple point measurement to extract T2 from an exponential fitting to different delay times (τ), typically 6-10 τ values. Her work shows the data points are well fitted to a single exponential function. This indicates each individual site in the whole protein shares similar decaying rate. So we didn't attempt to measure site-specific T2. In Table 2.2, shows T2 values of each nuclei, data collected using single point measurement instead of exponential fitting. Here we see the advantage of using high decoupling power and high MAS rate. And T2 values also reflects the quality of the sample. Generally dramatically short T2 means the microcrystalline may be destroyed.

Table 2.2 Echo T2(T2') for GB1

	MAS 9 kHz			**MAS 18 kHz**			**MAS 25 kHz**		
^{1}H Dec. (kHz)	**N**	**CO**	**CA**	**N**	**CO**	**CA**	**N**	**CO**	**CA**
75	29	22	13	28	30	9	24	28	8
100	55	33	20	68	62	25	43	44	16
125	51	36	21	96	69	33	90	83	32
150	76	40	22	101	84	30	104	96	36

2.3 CHEMICAL SHIFT ASSIGNMENTS OF GB1

2.3.1 BACKBONE CHEMICAL SHIFT ASSIGNMENTS

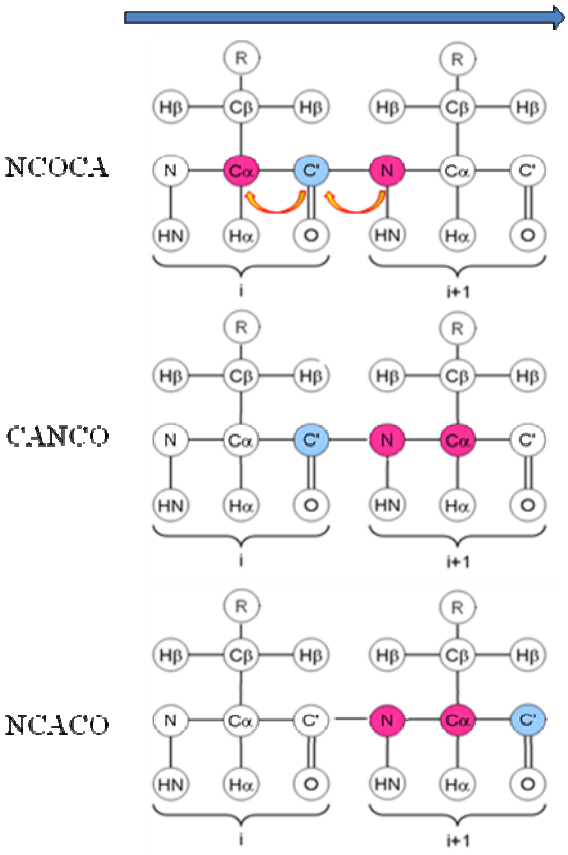
As shows in the previous session, these *J*-MAS experiments provide high resolution and sensitivity. Together with all the backbone correlation spectra, we are able to assign the entire protein backbone, in the sense of sequential walk. This also gives the ability to make chemical shift assignment of any unknown proteins. On the 3D spectra each resonance provides three chemical shift values from each dimension. Table 2.3 shows correlations can be observed in each 3D experiment.

Table 2.3 Types of correlations on 3D spectra

	Depth	**Search**	**Measure**
NCOCA	CA_i	CO_i	N_{i+1}
CANCO	CO_i	N_{i+1}	CA_{i+1}
NCACO	N_{i+1}	Ca_{i+1}	CO_{i+1}
NCOCA	CA_{i+1}	CO_{i+1}	N_{i+2}

The three backbone correlation spectra are used together to carry out the sequential walk process and make backbone chemical shift assignments. The way to visualize and analyze 3D spectra is to set one dimension as the depth and study the 2D slices. For example, on NCOCA spectrum, we set $C_{\alpha(i)}$ (C_α carbon of *i* th amino acid) chemical shift

as depth. Assuming chemical shift of $C'_{(i)}$ is known, then we can measure $N_{(i+1)}$ chemical shift on the 2D slice from NCOCA by search $C'_{(i)}$ chemical shift along C' dimension. Similarly, with two chemical shift values of $C'_{(i)}$ and $N_{(i+1)}$ we can measure $C_{a(i+1)}$ on CANCO by doing the same procedure. Then move to NCACO, measure $C'_{(i+1)}$ chemical shift (Scheme 2.2). Keep doing this until the end of the sequence.



Scheme 2.2 Backbone sequential walk strategy using NCOCA, CANCO and NCACO 3D correlation spectra.

Backbone sequential walk is initialized from 2D NCA spectrum, as shows in Figure 2.3. The N-terminal on protein gives a characteristic ^{15}N chemical shift at about 40 ppm, while the other backbone ^{15}N chemical shifts are all above 100ppm. Although some sidechain ^{15}N-^{13}C correlations from lysines will show in the regime, they are still easy to distinguish on the spectrum, as Figure 2.3 shows the resonances from sidechains of K13 and K4.

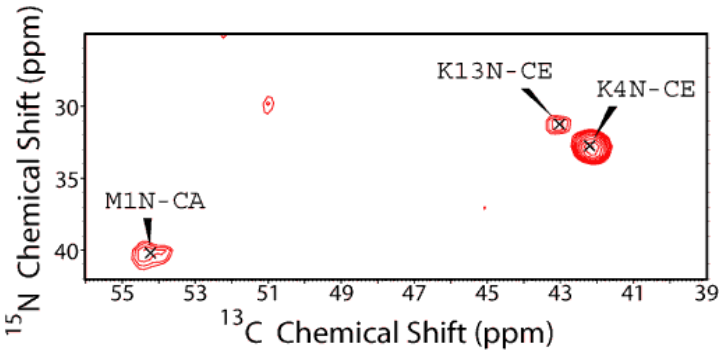


Figure 2.3 2D NCA correlation spectrum of N-terminal region of uniform labeled GB1.

With the N and C_α chemical shifts of N-terminal, we could measure C' chemical shift of the first amino acid on NCACO spectrum and enter into the sequential walk cycle. An example of first five amino acids backbone walk is showing in Figure 2.4.

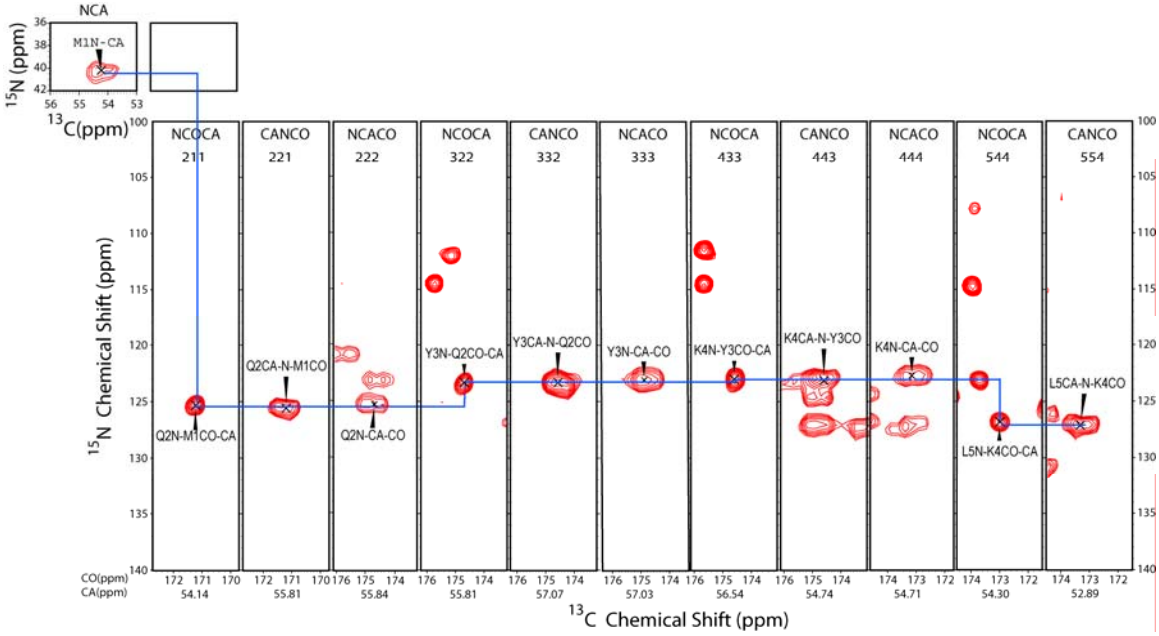


Figure 2.4 Backbone sequential walk for GB1.

2.3.2 SIDE-CHAIN CHEMICAL SHIFT ASSIGNMENTS

Under 25 kHz spinning rate and 150 kHz proton decoupling (conditions where the relaxation rate for the transverse magnetization is minimized[43]), strong sidechain–sidechain and sidechain–backbone correlations are found consistently throughout the spectra. Indeed, beginning with the previously reported J-MAS NCACO, NCOCA, and CANCO experiments to establish backbone C_0 and C_α assignments,[21] sidechain assignments for each of the 46 aliphatic residues in this 56 amino acid protein can be traced out.

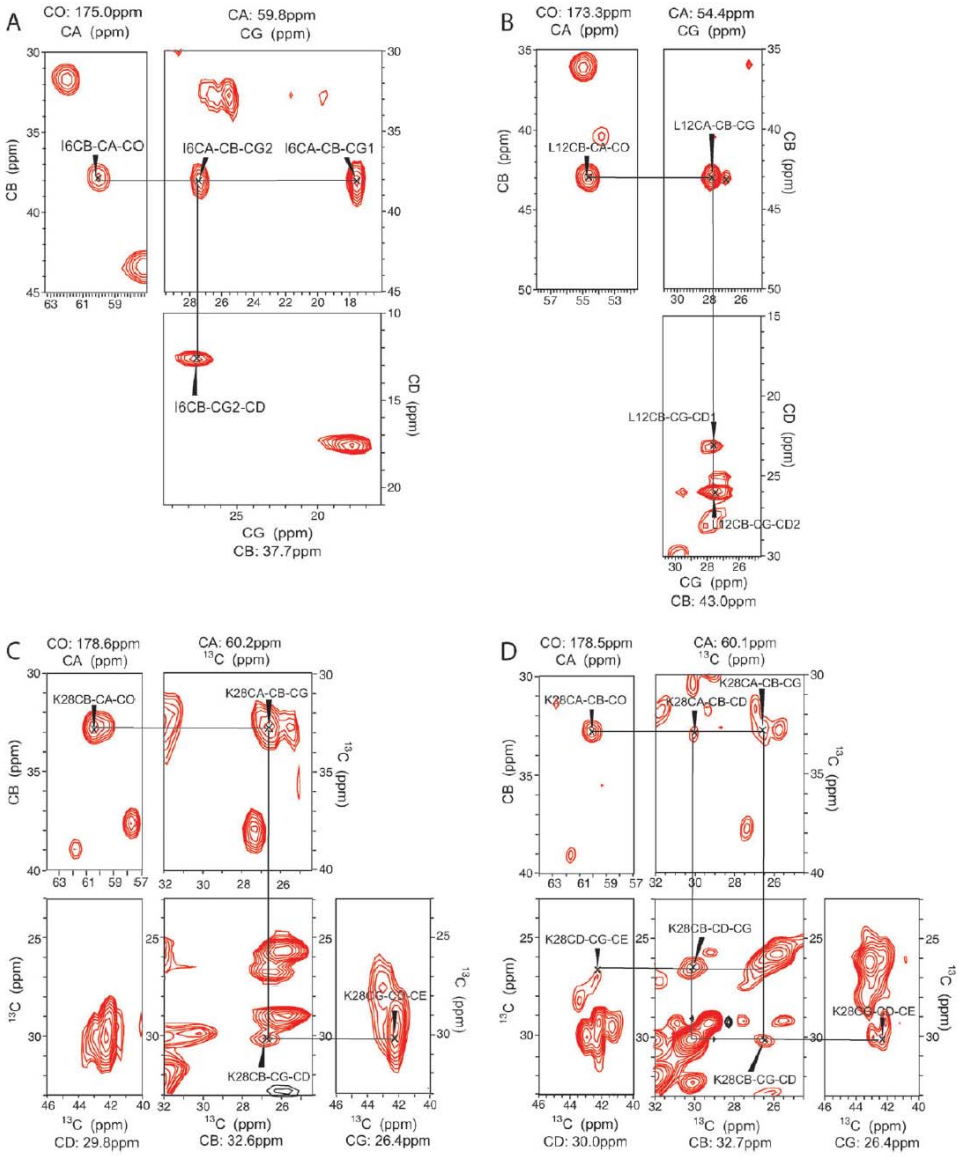


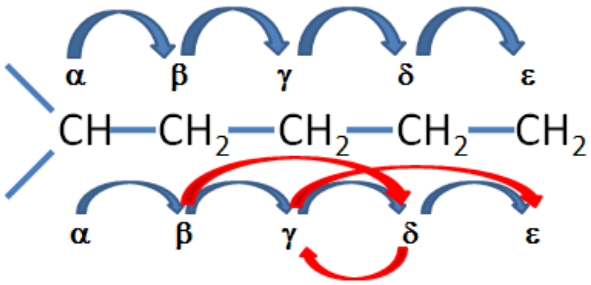
Figure 2.5 Sidechain assignment walk for (A) isoleucine-6, (B) leucine-12, and (C) lysine-28 in GB1, with 2D planes extracted from the 3D J-MAS CBCACO and CACBCG correlation spectra. (D) The corresponding assignment planes for dipolar-driven correlation in lysine-28,[42] showing additional multi-bond transfers not present in the scalar-coupling-driven experiments.

Three of these residues, isoleucine-6 (I6), leucine-12 (L12), and lysine-28 (K28), are shown in Figure 2.5, while all of the chemical shifts are tabulated in the ESI, along with complete sidechain assignment walks for all 46 residues. Nearly all of these assignments agree (within the digital resolution of the spectra) with the previously assigned GB1 resonance list deposited on the Biological Magnetic Resonance Data Bank,[44] which contains several values updated from the original report,[4] based on higher quality spectra collected from sparsely labeled samples.[45]

As noted, the assignment of resonance frequencies in the sidechains begins with the previously determined C' and C_α resonances along the protein backbone. From the backbone C' and C_α, C_β is uniquely identified in the CBCACO spectrum. In practice, this is accomplished by searching the two-dimensional CO–CB (f_1–f_3) plane, extracted from the 3D cube at f_2 equal to the α-carbon frequency, for a C_β resonance at the known C_0 frequency. To obtain C_γ, the CB–CG (f_2–f_3) plane is extracted from the 3D CACBCG cube at f_1 equal to the a-carbon frequency. Once more, a cross-peak to the assigned C_β resonance is sought to identify the γ-carbon. This process continues on until the sidechain is fully assigned. To verify assignments, the corresponding cross-peak in the CA–CG (f_1–f_3) plane, chosen at f_2 equal to the C_β frequency, is also confirmed. This assignment strategy is particularly straightforward to implement for short sidechain amino acids and those containing hetero-atoms (e.g., asparagine, glutamate, aspartate).

Long sidechains are more challenging to assign due to resonances close to the diagonal and with near degeneracies. δ-carbon methyls in leucine sidechains are another similarly congested and difficult to resolve example, yet because of the improved resolution of the constant-time J-MAS experiments, these can all be uniquely observed even on a 9.4 T spectrometer. See appendix for more sidechain walk figures.

2.3.3 COMPARISON TO DIPOLAR-DRIVEN CORRELATIONS



Scheme 2.3 Magnetization transfers pathways of scalar- and dipolar-coupling driven correlations of aliphatic sidechains. One and two bonds transfer are both allowed in dipolar-coupling driven experiments (bottom). Scalar-coupling driven experiments can be designed to allow only one bond transfer (top).

A salient feature of the scalar-coupling-driven correlation (See example in Figure 2.5C) is that multi-bond transfers are significantly attenuated, which underscores their

importance in the assignment of resonances for biological solids. For example, in Figure 2.5D, the assignment planes of lysine-28 taken from the band-selective 3D dipolar-driven spin diffusion correlation experiment (at ^{1}H frequency of 750 MHz) are shown for comparison.[42] In this experiment, there is significant intensity for two bond transfers and it is not unusual for preliminary assignments based on dipolar-driven correlation to be made in which several resonance assignments are transposed. In the scalar-coupling-driven experiments, these additional two-bond transfers are not seen, leading to far fewer assignment ambiguities.

2.3.4 DISCUSSION

There are several key experimental details that merit discussion. One of the most important is the choice of delays, τ, for the J transfers. Even in an ideal two-spin system, these transfer periods act as filters, optimally passing coherences through a scalar coupling, $J = 1/4\tau$, while completely attenuating those with $J = 1/2\tau$. This problem is compounded by the array of multispin dynamics in the sidechain spin systems, where both the active and passive couplings modulate the transfer. Typical values of scalar couplings along the protein sidechains are illustrated in Scheme 2.1. The efficiency of transfer during a $\tau-\pi-\tau$ period exhibits a sinusoidal dependence for the desired transfer

through the active coupling and a cosinusoidal dependence for each additional passive coupling,

$$T_{IS} = \exp\left(-2\tau/T_2'\right)\sin\left(2\pi J_{IS}\tau\right)\prod_{n_i}\cos\left(2\pi J_{In_i}\tau\right)$$

Here, T_{IS} is the efficiency of transverse magnetization beginning on spin I (e.g., I_x) to become anti-phase magnetization through the active coupling with spin S (e.g., I_yS_z), while also being passively coupled to as many as ni additional spins. This expression is also valid for the reverse process, in which the anti-phase magnetization on spin I rephases, but in general $T_{IS} \neq T_{SI}$ due to distinct passive spins on I and S. For an entire 3D sequence, the overall transfer efficiency is the product of successive mixing steps, and the combined transfer during τ_2 has an efficiency that is the product of both individual transfers. While selective pulses on the carbonyls and aliphatics can help mitigate the loss by effectively removing them from the passive spin network, multispin dynamics remain an issue within the aliphatics. Taking representative values of J_{CACO} = 55 Hz and J_{CACB} = 35 Hz, theoretical maxima for CBCACO transfers are τ_1 =7.14 ms, 3.57 ms and 2.80 ms for primary, secondary and tertiary CB, respectively, τ_2=5.31 ms, and τ_4=4.55 ms. As the choice of τ also puts an upper limit of 2τ on the length of acquisition in the indirect dimensions and hence the achievable resolution, we tend to choose slightly longer values of t, accepting a 5–10% decrease in transfer efficiency in some cases in order to improve transform limited resolution. For CBCACO correlation,

we set τ_1 = 4.0 ms, τ_2 = 4.8 ms, and τ_4 = 4.8 ms, finding an overall efficiency of >30%. For CACBCG correlation, optimum values for all τ again depend on the number of aliphatic couplings and here too τ = 7.14 ms, 3.57 ms and 2.80 ms are best for primary, secondary and tertiary carbons, respectively. For correlation in the aliphatic regions, we choose τ = 3.2 ms as a compromise. Finally, we note that the choice of optimum delay is affected by relaxation. If the value of τ is too large, relaxation will attenuate the efficiency. For this reason, optimized decoupling is critical.

As with most 3D experiments, it is desirable to use the shortest possible phase cycle,[46] as this, rather than signal-to-noise considerations, often determines the number of transients required at each t_1, t_2 point, and hence the total experimental time. At a minimum, we find the eight-step phase cycle given in Table 2.1 to be necessary to eliminate most artifacts. The first element of the phase cycle selects bilinear I_xS_y type terms between the first two back-to-back $\pi/2$ pulses. The selection of these mixed zero quantum–double quantum terms is accomplished by phase cycling the first three phases together in a two step cycle, $\{\phi, \phi + \pi\}$, while the receiver phase is held constant. Omitting this cycle results in significant intensity at ω_2=0 in the 3D cube that is due to unwanted coherences that pass through longitudinal magnetization in t_2. The second indirect dimension has, in principal, a similar requirement. Provided the z-filter element is robust for selection of longitudinal terms, however, this second filter can be omitted.

Imperfections in the π pulses also lead to significant artifacts in the 3D spectrum, producing skewed-diagonal peaks with a slope of 1/2 and an offset-dependent phase within the spectral regions of interest. These are due to spins that experience a π/2 pulse, rather than a full p pulse during the (τ-t/2)–π–(τ+t/2) intervals, and therefore only evolve for half the time in the indirect dimensions. Full independent phase cycling of all three π pulses is prohibitively long. Rather, we cycle them simultaneously according to {ϕ, ϕ + π/2, ϕ + π, ϕ + 3π/2}, while the receiver alternates {ϕ_{rec}, ϕ_{rec}+π, ϕ_{rec}, ϕ_{rec}+π}. This does not completely remove all artifacts, but reduces them to second order in the pulse error, which we find acceptable in these experiments. Equally important in maintaining the pure in-phase absorption mode line shapes in these experiments is the final z-filter.[47,48] Generally, this period must be long enough to dephase the unwanted transverse coherences, but short enough to avoid spin diffusion. We favor lowering the decoupling power to moderate levels (70–100 kHz), to prevent spin diffusion, and using relatively long (16 ms) delays.

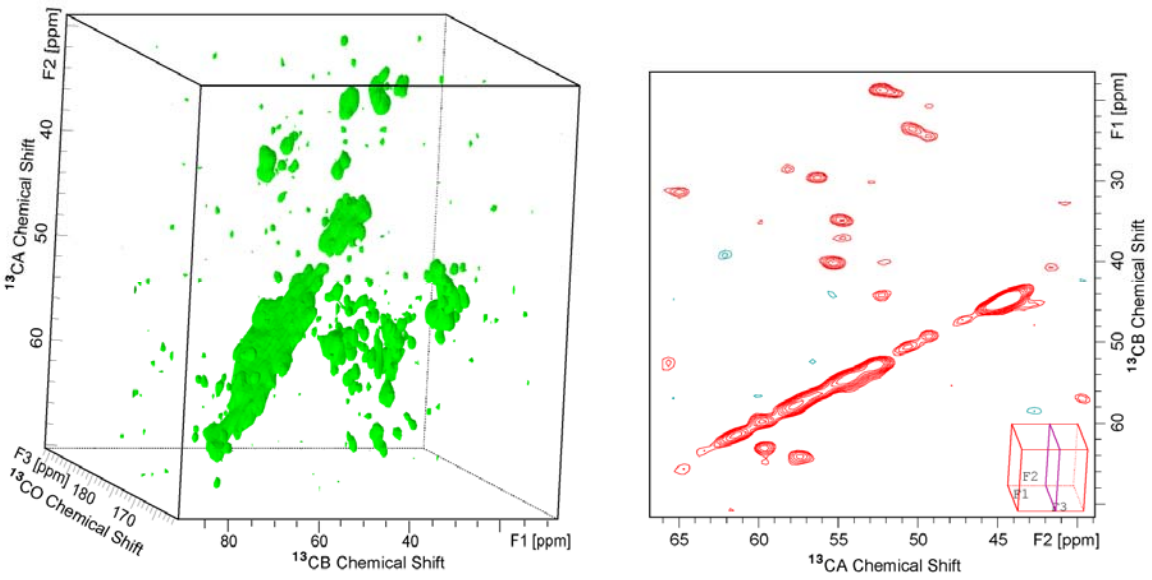


Figure 2.6 Constant-time J-MAS CBCACO homo-nuclear correlation spectrum of TS at 9.4 T showing the full 3D cube and a 2D plane extracted at ϕ_3=173.15 ppm. Experiment specific parameters for CBCACO: τ_1=4.0 ms, τ_2=τ_4=4.8 ms; 32 scans per transient; 2048 complex points with a dwell of 20 μs in t_3 (spectral width: 50 kHz, total acquisition time: 41.0 ms), 38 complex points with a dwell of 240 μs in t_2 (spectral width: 4166.667 Hz, total acquisition time: 9.1 ms), and 64 complex points with a dwell of 120 μs in t_1 (spectral width: 8333.33 Hz; total acquisition time: 7.7 ms). Total experiment time: 278 h.

One advantage of solid-state NMR compared to solution state NMR is that the transverse relaxation time, T2, does not depend on the size of the protein (presuming that the global molecular tumbling is slow compared to the MAS rate). Global molecular tumbling in solution-state NMR ultimately limits the size of biomolecules that can be

interrogated before resolution and sensitivity are lost. Although transverse relaxation optimized spectroscopy (TROSY) methods[49] have been devised in order to create situations where tumbling has less influence on the linewidths, there is always some significant contribution from the global correlation time in solution; in contrast, solid-state NMR does not share this limitation. For example, in tryptophan synthase, a 144 kDa $\alpha_2\beta_2$ heterodimer that catalyzes the last two steps in the biosynthesis of l-tryptophan,[50–54] echo T2 periods of 95 ms for backbone ^{15}N, 80 ms for ^{13}CO, and 35 ms for ^{13}CA are observed, comparable to those for the much smaller GB1. So, although the residue-specific spin density (and hence the signal-to-noise ratio at a specific site) decreases with size, there is no inherent limitation to implementing J-MAS experiments and Figure 2.6 shows preliminary 3D CBCACO correlation spectra in TS at 9.4 T. While not completely resolved at this field, it is clear that the experiments are effective even on a protein system this large. Indeed, the overall efficiency of polarization transfer over two bonds is still >30%.

Furthermore, initial 2D CACO correlation spectra at 21.15 T (900 MHz ^1H frequency, Figure 2.7) show that individual sites in TS can be resolved at high field even in 2D experiments. Taken together, there is considerable reason to expect that the combination of high field and 3D J-MAS correlation experiments will enable the assignment of proteins in this size regime.

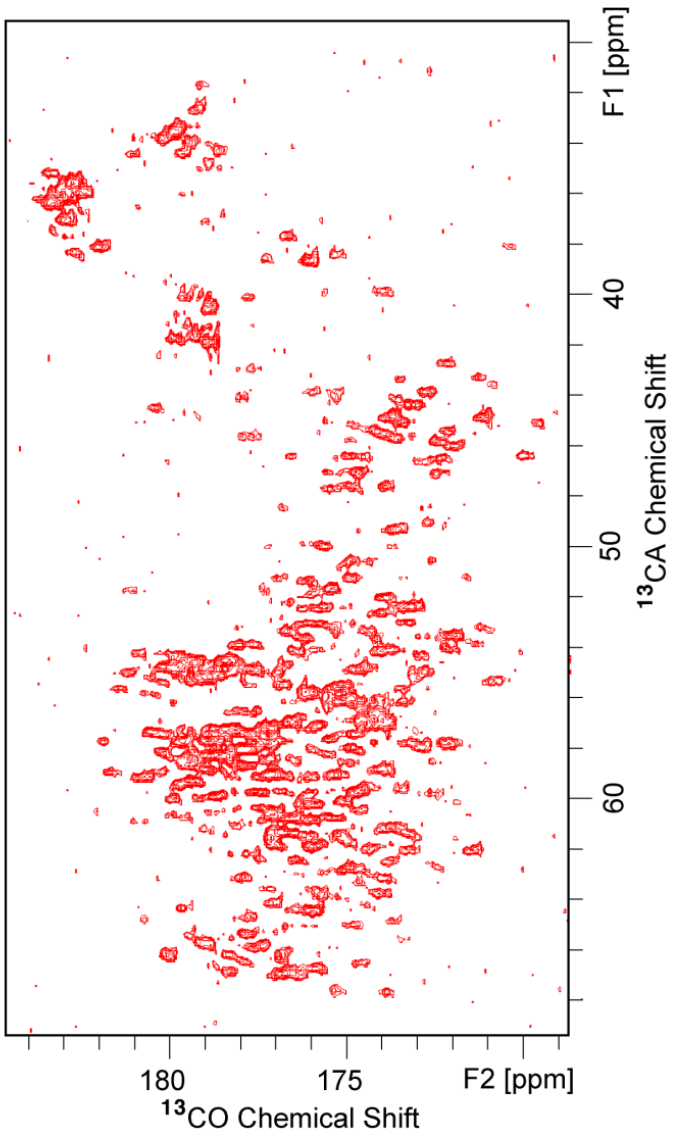


Figure 2.7 Constant-time J-MAS CACO homo-nuclear correlation spectrum of TS at 21.15 T. Experimental parameters $\tau_1 = \tau_2 = 4.8$ ms; 64 scans per transient; 4096 complex points with a dwell of 7.2 μs in t_2 (spectral width: 69.44 kHz, total acquisition time: 29.6 ms), 440 complex points with a dwell of 20 μs in τ_1 (spectral width: 50 kHz, total acquisition time: 8.8 ms). Total experiment time is 47 h.

Same experiment is also done on the departmental 14.1 T (600 MHz ^{1}H frequency) magnet. Sum of 4 individual 2D CACO correlation spectra is shown in Figure 2.8.

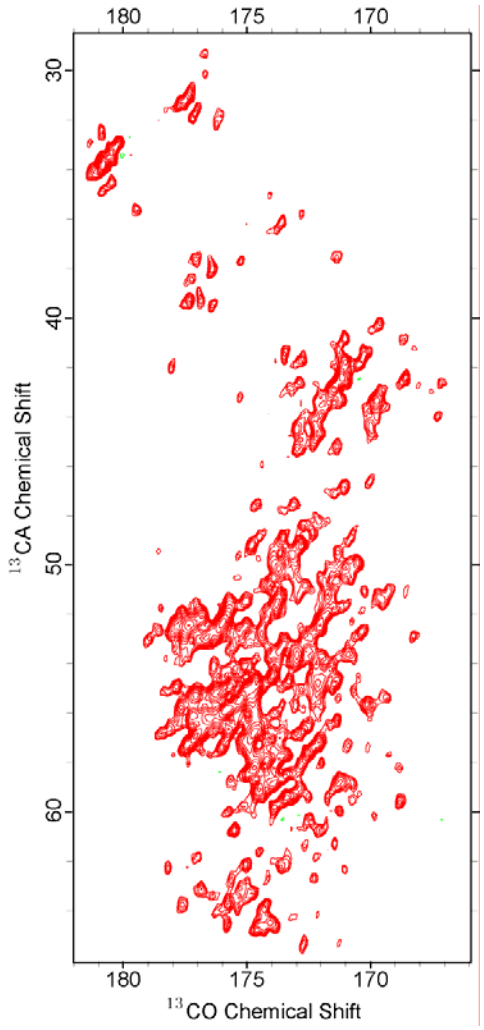


Figure 2.8 Constant-time J-MAS CACO homo-nuclear correlation spectrum of TS at 14.1 T. Sum of 4 individual 2D spectra Total experiment time: 24 h.

Although the spectrum is not assignable, we could see some peaks start to resolve even on the 2D spectrum. Moreover, comparable resolution to the spectrum from high field (900 MHz) is obtained, due to higher decoupling power applied for experiments on 14.1T. This gives a hope for us to assign proteins in this size using 14.1 T magnet, by applying high decoupling power and faster MAS spin rate to the 3D experiment.

2.4 STRUCTRUE REFINEMENT WITH CHEMICAL SHIFT

Various technologies are used in biophysics aiming at high resolution structures. Each one of them has its particular advantage and also limitations. For example, X-ray crystallography is able to solve high resolution protein structures, yet it requires high quality of 3D crystals, and limited by sample heterogeneity and flexibility. Liquid state NMR doesn't require crystals, however it requires isotropic labeled samples and sample solubility, and limited by the size of molecules (for protein sample, generally smaller than 30 kDa). While applying liquid state NMR to membrane proteins, it fails due to the low solubility of this family. In these cases solid state NMR, is playing more important role in structural biology. In the past ten years, there small proteins are solved by solid state NMR (such as SH3[55], Ubiquitin[56] and GB1[57]), it opens the window for solid state NMR.

Solving structures of biomolecules in NMR requires three types of major geometric constraints: Chemical shift, Distance constraint and Torsion angles. When solving

protein structures, Chemical shift constraints are not used directly, rather they are transferred into secondary chemical shift[58] which is then used for predicting secondary structures. Traditionally, these predicted secondary structures are combined with distance and torsion angle constraints for further structure refinement. With all the efforts that focus on extracting structural information from proteins, there are more than eight thousand structures on Protein Data Bank are solved by NMR. This provides powerful database for the technique of NMR to discover more in structural biology.

2.4.1 INTRODUCTION

With the rapid development of computational chemistry, nowadays it is possible to predict low resolution structures by de novo calculation using only the amino acid sequence. Recently, three programs are developed to combine Chemical shift constraints with computation chemistry for solving high resolution protein structures.

In our work, we are able to obtain high resolution spectra with the J-based pulse sequences. It makes the assignment work much easier than using the dipolar based method, due to the lower spectral complexity. Here we only use chemical shift constraints for structural refinement. The program we used is CSrosetta[59,60] which is developed in Bax's group. A web based program CS23D[61] developed by University of Alberta is also used for evaluation. Both of them use Rosetta[62-67] as the major

computation environment. As the literature[61] says, CSrosetta performs better in recognizing blind structures, which meet our needs for assigning and solving unknown proteins. In the following sections we test CSrosetta by considering GB1 as the model protein, and all the chemical shift assignments are obtained from the backbone walk as described in the previous sessions.

Besides Rosetta, another major program SPARTA[68] is used in CSrosetta, it predicts chemical shifts based on 3D structures. The workflow of CSrosetta is briefly introduced below: firstly, the chemical shift input file is used to generate 3-residue and 9-residue fragments from searching in structure database to prepare the initial package for the subsequent Rosetta structure generation; secondly, perform the Rosetta structure prediction and generate all-atom structures; finally, the all-atom structures will be evaluated in Rosetta by calculating all-atom energy and backbone C_α RMSD for convergence check and extraction of the best result. Convergence has to be confirmed for each prediction.

2.4.2 STRUCTURE REFINEMENT

The CSrosetta input files utilize six different types of chemical shifts as input, they are backbone ^{15}N, ^{1}HN, ^{1}HA, ^{13}CA, ^{13}CB and $^{13}C'$. In the J-based, homo- and hetro-nuclear correlation experiments we only look at the correlations between ^{15}N and

^{13}C, so in our input files we only include backbone ^{15}N, ^{13}CA, ^{13}CB and ^{13}C' chemical shifts. And we consider this kind of input file as complete input in this chapter, which compares to the input files that only have partially assigned chemical shifts of backbone ^{15}N, ^{13}CA, ^{13}CB and ^{13}C'. We call the input files that use partially assigned chemical shifts as incomplete chemical shift inputs. These incomplete inputs are used to test the program as well.

The first structure calculation we did was utilizing the entire backbone ^{15}N and ^{13}C chemical shifts of GB1 as input. One thousand structures are generated from Rosetta, for comparison, we extract the 10 lowest energy structures to calculate ensemble RMSDs, and superimpose the best structure with references for evaluation. In Figure 2.9, the J-gb1 ensemble (predicted structures using chemical shift assignments from J-based spectra) has overall 0.94±0.24Å backbone RMSD (bbRMSD), and 1.67±0.33Å all heavy atom RMSD.

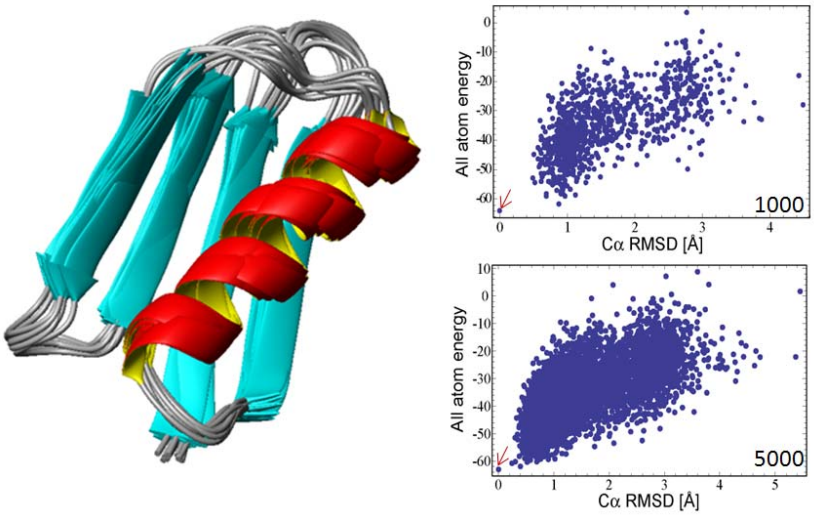


Figure 2.9 J-gb1 ensemble of 10 lowest energy structures, with all-atom energy plot of 1000 and 5000 structures.

The convergence check is processed with 1000 and 5000 structure calculation, and use the predicted structure as reference which has the lowest all-atom energy. (The evaluation can be done by using a reference from database as well.) In both cases, the results are clustered around the lowest energy structure and give overall low C_α RMSD (< 5Å). That says the calculations converge, and the 10 lowest energy structures are accepted. For more comparison and evaluation we extract the lowest energy structure and superimpose it with the structures on Protein Data Bank that are solved by solid state NMR (2JSV)[57] and X-ray diffraction (2QMT)[69], as shows in Figure. 2.10

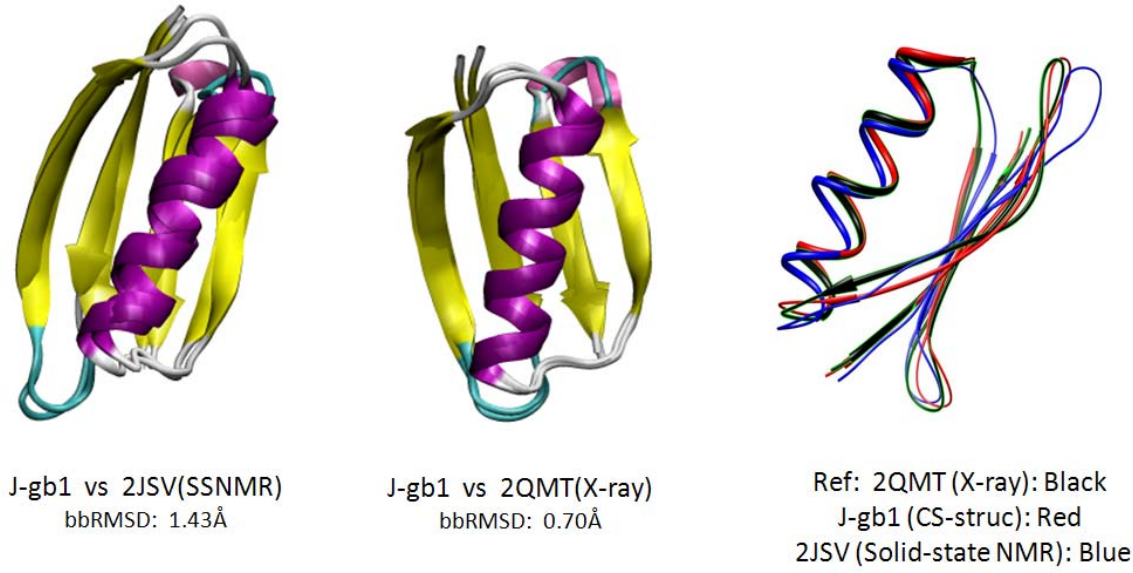


Figure 2.10 Superimposed GB1 structures.

The predicted GB1 structure shows similar agreement to the result from X-ray crystallography (2QMT) (bbRMSD : 0.70Å) than solid state NMR (2JSV) (bbRMSD : 1.43Å). The difference between 2JSV and 2QMT is about 1.44 Å of bbRMSD. This could be caused by the different level of sample's crystallinity. For solid state NMR, microcrystalline is allowed to obtain high quality spectra, while X-ray crystallography requires single crystal. Furthermore, the NMR sample is hydrated, and packed into the rotator with the mother liquor present. All these factors could make slightly change of the protein structure. However it isn't complete clear at this stage, more study is needed.

2.4.3 DISCUSSION

Considering the 1.43Å bbRMSD between J-gb1 and 2JSV, this is higher than both J-gb1 and 2JSV ensembles bbRMSD, which are 0.94 Å and ~0.5Å respectively. Then we might be looking at a real structure difference here. Because of different hydration level and errors during assignments, the chemical shift values from our J-based method are different from the reference on Biological Magnetic Resonance Data Bank (BMRB ID: 15156). This reference relates to the 2JSV on the Protein Data Bank. The backbone ^{15}N has overall 0.6 ppm difference, backbone ^{13}C shows average of 0.1 ppm difference. We use the chemical shift assignments on BMRB as input and generate a new structure, and call it CS-2JSV. As it shows in Figure 2.11, superimposed CS-2JSV and J-gb1, we find overall 0.48 Å bbRMSD between the two structures, both of this two structure report around 0.7 bbRMSD to the X-ray structure (2QMT).

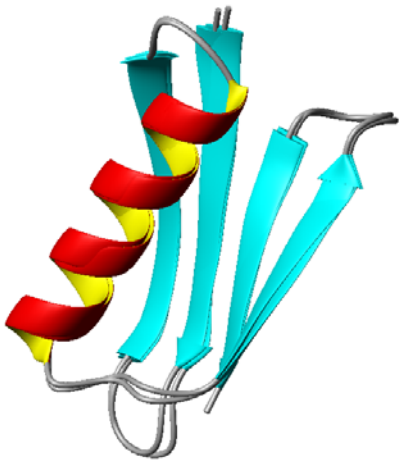


Figure 2.11 Superimpose lowest energy structures of CS-2JSV and J-gb1 from CSrosetta.

In these results, J-gb1 is closer to 2QMT, for it has 0.68 Å bbRMSD, which is slightly better than CS-2JSV (0.75 Å bbRMSD to 2QMT). However it is not clear that J-gb1 is a better structure, due to the high bbRMSD of the ensemble (0.94 Å). No matter what, these chemical shift differences, up to the level we observed in our experiments, doesn't cause a dramatic impact on the all-atom structure.

The interesting argument is on CS-2JSV, it uses the chemical shifts that are used to solve the high resolution structure of GB1 together with other geometric constraints from solid state NMR. And the structure that is solve by solid state NMR using all geometric constraints, PDB ID: 2JSV, has 1.4 Å bbRMSD to the X-ray structure (2QMT). However, we use its chemical shifts to predict a structure in CSrosetta, the result (CS-2JSV) only has 0.75 bbRMSD to the X-ray structure. All the bbRMSDs are listed in Table 2.4

Table 2.4 Comparison of bbRMSD

	J-gb1	2JSV	CS-2JSV
Ensemble	0.94	0.50	0.76
Vs 2QMT	0.70	1.44	0.75
Vs 2JSV	1.43	--	1.61

The calculations from CSrosetta are closer to X-ray structure, not solid state structure. If we consider the X-ray structure as the reference, then these results are saying by putting more geometric constraints we are pushing GB1 structure away from the reference. Results using only chemical shift data give better agreement to the X-ray reference, no matter with chemical shift data sets are used. It seems some information used to in obtaining 2JSV is redundant, and it drives the calculation to another direction that is away from X-ray reference. In Figure 2.12, the 3-residue fragments bbRMSD plot of CS-2JSV and 2JSV (relative to 2QMT), shows 2JSV has higher bbRMSD for the ending residues where are more flexible than the middle residues (α-helix). Particularly, these flexible residues are either the beginning/ending of β sheets or random coils.

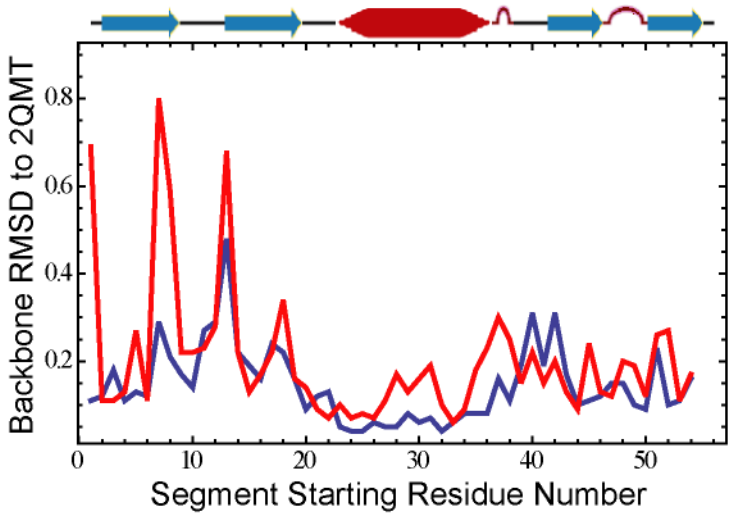


Figure 2.12 Plot of 3-residue fragments RMSD of CS-2JSV (blue) and 2JSV(red), relative to 2QMT

We need to be careful to draw conclusions here. Because 2JSV is calculated in Xplor-NIH[70,71]. And all the previous results from CS-Rosetta are mainly calculated in Rosetta, since Rosetta is the kernel computation program in both CS-Rosetta and CS23D. The structure difference between 2JSV and CS-2JSV (or J-gb1) may be just caused by these distinct programs. We wouldn't be able to tell until we combine chemical shifts and dipolar constraints in Rosetta to generate more structures for comparison. This part of work is still on-going. However, we do find out, as the literature promised, using the backbone ^{15}N and ^{13}C chemical shifts and CS-Rosetta we could generate high resolution structures (~1.5 Å). Together with our J-based method, it becomes possible for fast protein structure determination in solid state NMR.

All the results we discussed above are using the complete chemical shift input files, which have full chemical shift assignment of backbone ^{15}N and ^{13}C. In fact, for proteins that are in bigger sizes, the local molecular motions could affect spin-spin relaxation rate, to a certain extent that for some specific amino acids their backbone nuclei have shorter T2 times. Since the spectral sensitivity strongly depends on T2, in the extreme cases, the short T2 could lead to signal vanish on spectra. In all the J-based experiments, the constant time periods are optimized for the average J-coupling constant through the entire protein. While there could be a few sites that their signal vanishes before acquisition. Although we still could process the sequential walk to continue backbone assignments

starting from some characteristic sites (such as Glycine), or modify the constant time period for particular sites, eventually we may get an incomplete/imperfect chemical shift[59] table. Therefore, it is necessary to understand that up to what level CSrosetta is still able to function properly.

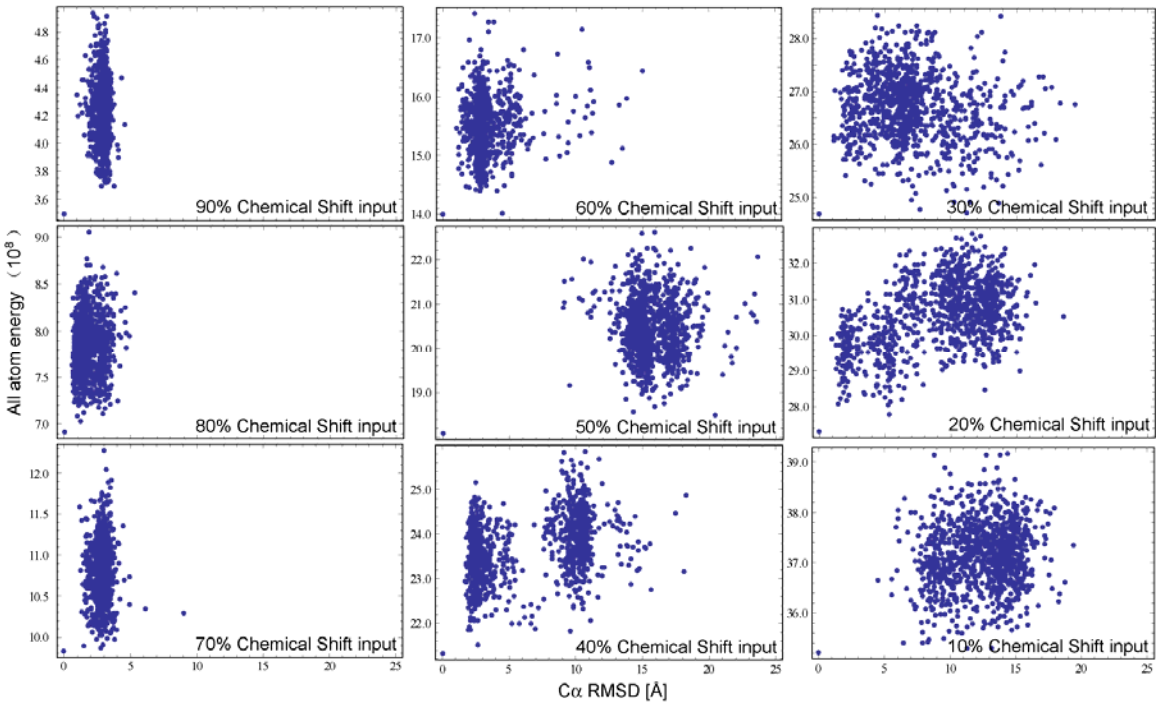


Figure 2.13 Plots of Rosetta all-atom energy versus C_α RMSD relative to the lowest energy structure from calculation, using the incomplete chemical shift input for J-gb1.

So we randomly pick some amino acids (the random residue numbers are generated in Mathematica.), and remove the chemical shifts of these chosen sites from the input file to make a group of imperfect input files that are 90-10% completed. Still use the GB1

domain and the chemical shift assignments from J-based 3D spectra, 90% completion means 6 sites' chemical shifts are deleted from the input file. We ran these nine calculations to generate 1000 structures in each case, the all-atom energy to backbone C_α plots are shown in Figure 2.13. In these plots, we found out the calculation starts with the 60% chemical shift input file is still showing the convergent pattern, although its bbRMSD to 2QMT (1.21Å) is higher than the complete chemical shift calculation (0.7Å).

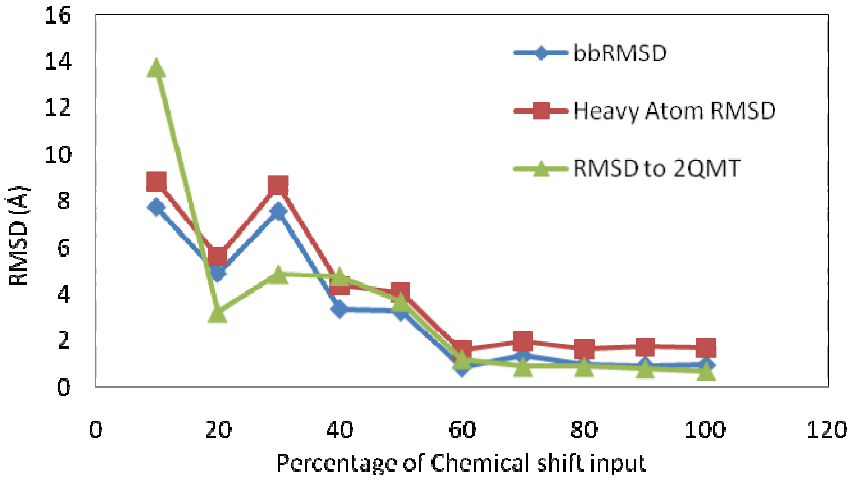


Figure 2.14 Plot of Accuracy of structures generated in CSrosetta with chemical shift assignments for J-gb1

Once we use 50% completed chemical shift input, the calculation becomes non-convergent suddenly. In calculations that are using less chemical shift inputs, the

program is trying to find the best structure, yet due to the lack of structural information, it fails to report acceptable result with convergent calculation. Finally, calculation becomes completely non-convergent in the case of 10% chemical shift input. And Figure 2.14, the accuracy of structures generated from imperfect chemical shift input files (90-10%), is also showing that with less than 60% of chemical shift input the structure is different from 2QMT in the sense of high backbone RMSD.

The way of CS-Rosetta works is combining the NMR constraints (Chemical shift, or Residue Dipolar Couplings) and de novo computation strategies. While there are enough NMR constraints, CS-Rosetta is just doing structural refinement that is driven by these constraints. However, if the constraints are not enough, the computation part of CS-Rosetta, which is Rosetta itself, will take over and perform ab initio calculation. Then the calculation depends on Rosetta, more specifically the fragment candidates generated at the beginning of each calculation. Rosetta is able to do de novo calculations with just amino acid sequences, to perform a good de novo calculation it requires much more fragment candidates than what we have in the regular CS-Rosetta calculation. The Hybrid mode of CS-Rosetta is designed to perform calculations with imperfect chemical shift inputs, and more like standard Rosetta calculations. The Hybrid calculations are initiated by generating 2000 Rosetta fragment candidates for each 3-residue and 9-residue target fragment, while the regular mode generates only 200

candidates with the chemical shift constraint filtering. Then these candidates will be assembled in Rosetta and generate full structures. This work of exploring the hybrid mode calculation is not yet completed when I am writing this dissertation. However, according to the literature, hybrid calculation provides better results.

To find out the fragment accuracy of each calculation with imperfect chemical shift input, we plot out the 3 and 9-residue fragments bbRMSD of J-gb1 with different chemical shift input files (100%-60%). In Figure 2.15 (A-F), the highlighted residues are those sites without chemical shift information in the input files. From these plots, we couldn't see a strong dependence of fragment accuracy on the local chemical shift information. That is saying the residues which have no chemical shift information in the input files don't have higher bbRMSD (or lower fragment accuracy), shows in Figure 2.15 A-E. The program can manage the fragment candidates and generate full structures according to the given chemical shift information which compensate the residues have no chemical shift information in the input. And overall, the middle residues in each case have better agreement to the reference structure (2QMT), as it show in Figure 2.15 A-E. To prove our point here, we remove the chemical shift values of residue 20 through 39 from input file, to make the 64% chemical shift input which is comparable to the 60% case. And residue 20 through 39 cover the middle part which has better fragment bbRMSD in each calculation. The result (F in Figure 2.15) is very similar to we see in

E of Figure 2.15. (Result from 64% chemical shift input calculation has overall 0.67 Å ensemble bbRMSD, 1.34 Å ensemble heavy atom RMSD and 0.79 Å bbRMSD to 2QMT.) The middle residues are still showing better agreement to reference (2QMT). In practice, this is very helpful, even if we couldn't obtain the full chemical shift assignment, with 60% completed chemical shift input file CS-Rosetta still can report acceptable structures. And the assigned residues could be a big block or scattered sites through the whole protein sequence. The former may be seen more frequently in the real work. For the spectral density increased linearly by the size of protein in the J-based experiments, undistinguishable overlapped resonances are still possible to be found even on the 3D spectra, if we apply them on much bigger proteins. In such conditions, it may have no choice but skip a few spots, and restart the sequential walk assignment from some characteristic residue, such as Glycine, it has negative peak on the NCACO and CANCO spectra.

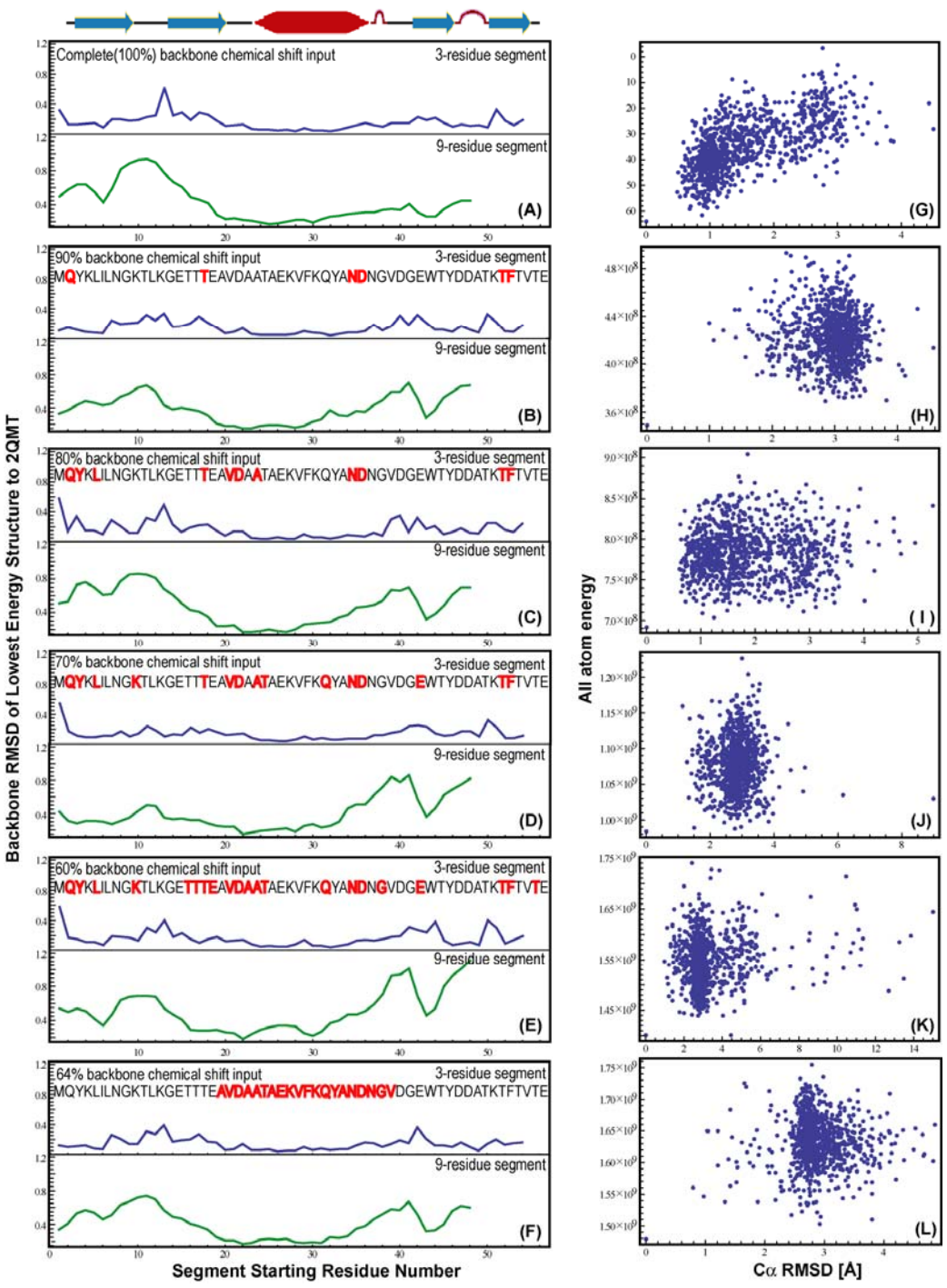


Figure 2.15 (A-F) Plots of 3- and 9-residue fragments bbRMSD of J-gb1 with incomplete chemical shift input. (G-L) Plots of Rosetta all-atom energy versus C_α RMSD

2.5 CONCLUSION

Despite being orders of magnitude smaller than the dipolar coupling, the through-bond scalar coupling provides a robust and efficient mechanism for correlation spectroscopy in proteins. Combined with our previously introduced J-MAS NCACO, NCOCA, and CANCO experiments, the J-MAS CBCACO and CACBCG experiments allow us to trace out the entire backbone and aliphatic sidechains in GB1. Because the evolution and transfer periods are combined into a single, efficient constant-time period in these experiments, high sensitivity and resolution are obtained, and even larger proteins such as tryptophan synthase are amenable to these methods. Indeed, preliminary 3D correlation at 9.4 T and 2D CACO correlation at 21.15 T show promise for obtaining the sensitivity and resolution necessary to completely assign solid state proteins greater than 650 amino acids. Furthermore, if combines the unique advantage of CSrosetta with our J-based high resolution 3D spectra, it is possible to solve full structures of proteins in much bigger size in solid state NMR. Hopefully, in the near future, solid state NMR will be highlighted as the major ‘explorer’ in structural biology.

REFERENCES

1. J. Pauli, M. Baldus, B. van Rossum, H. de Groot and H. Oschkinat, *Chembiochem*, **2001**, 2, 272-281.

2. M. Baldus, *Prog Nucl Mag Res Sp*, **2002**, 41, 1-47.

3. T. I. Igumenova, A. E. McDermott, K. W. Zilm, R. W. Martin, E. K. Paulson and A. J. Wand, *J Am Chem Soc*, **2004**, 126, 6720-6727.

4. W. T. Franks, D. H. Zhou, B. J. Wylie, B. G. Money, D. T. Graesser, H. L. Frericks, G. Sahota and C. M. Rienstra, *J Am Chem Soc*, **2005**, 127, 12291-12305.

5 M. Ernst, A. Detken, A. Bockmann and B. H. Meier, *J Am Chem Soc*, **2003**, 125, 15807-15810.

6 A. E. McDermott, *Curr Opin Struc Biol*, **2004**, 14, 554-561.

7 C. E. Hughes and M. Baldus, *Ann R Nmr S*, **2005**, 55, 121-158.

8 M. Baldus and B. H. Meier, *J Magn Reson Ser A*, **1996**, 121, 65-69.

9 E. H. Hardy, R. Verel and B. H. Meier, *J Magn Reson*, **2001**, 148, 459-464.

10 A. Lesage and L. Emsley, *J Magn Reson*, **2001**, 148, 449-454.

11 A. Lesage, M. Bardet and L. Emsley, *J Am Chem Soc*, **1999**, 121, 10987-10993.

12 A. Detken, E. H. Hardy, M. Ernst, M. Kainosho, T. Kawakami, S. Aimoto and B. H. Meier, *J Biomol Nmr*, **2001**, 20, 203-221.

13 R. A. Olsen, J. Struppe, D. W. Elliott, R. J. Thomas and L. J. Mueller, *J Am Chem Soc*, **2003**, 125, 11784-11785.

14 L. J. Mueller, D. W. Elliott, G. M. Leskowitz, J. Struppe, R. A. Olsen, K. C. Kim and C. A. Reed, *J Magn Reson*, **2004**, 168, 327-335.

15 B. Elena, A. Lesage, S. Steuernagel, A. Bockmann and L. Emsley, *J Am Chem Soc*, **2005**, 127, 17296-17302.

16 A. Bockmann, M. Juy, E. Bettler, L. Emsley, A. Galinier, F. Penin and A. Lesage, *J Biomol Nmr*, **2005**, 32, 195-207.

17 O. C. Andronesi, S. Becker, K. Seidel, H. Heise, H. S. Young and M. Baldus, *J Am Chem Soc*, **2005**, 127, 12965-12974.

18 L. L. Chen, R. A. Olsen, D. W. Elliott, J. M. Boettcher, D. H. H. Zhou, C. M. Rienstra and L. J. Mueller, *J Am Chem Soc*, **2006**, 128, 9992-9993.

19 M. Etzkorn, S. Martell, O. C. Andronesi, K. Seidel, M. Engelhard and M. Baldus, *Angew Chem Int Edit*, **2007**, 46, 459-462.

20 G. Pintacuda, N. Giraud, R. Pierattelli, A. Bockmann, I. Bertini and L. Emsley, *Angew Chem Int Edit*, **2007**, 46, 1079-1082.

21 L. Chen, J. M. Kaiser, T. Polenova, J. Yang, C. M. Rienstra and L. J. Mueller, *J Am Chem Soc*, **2007**, 129, 10650-10651.

22 L. L. Chen, J. M. Kaiser, J. F. Lai, T. Polenova, J. Yang, C. M. Rienstra and L. J. Mueller, *Magnetic Resonance in Chemistry*, **2007**, 45, S84-S92.

23 L. Zhong, V. V. Bamm, M. A. M. Ahmed, G. Harauz and V. Ladizhansky, *Biochimica Et Biophysica Acta-Biomembranes*, **2007**, 1768, 3193-3205.

24 F. Fayon, I. J. King, R. K. Harris, J. S. O. Evans and D. Massiot, *Comptes Rendus Chimie*, **2004**, 7, 351-361.

25 F. Fayon, G. Le Saout, L. Emsley and D. Massiot, *Chemical Communications*, **2002**, 1702-1703.

26 P. Florian, F. Fayon and D. Massiot, *Journal of Physical Chemistry C*, **2009**, 113, 2562-2572.

27 V. Agarwal and B. Reif, *J Magn Reson*, **2008**, 194, 16-24.

28 R. Linser, U. Fink and B. Reif, *J Magn Reson*, **2008**, 193, 89-93.

29 L. J. Mueller, D. W. Elliott, K. C. Kim, C. A. Reed and P. D. W. Boyd, *J Am Chem Soc*, **2002**, 124, 9360-9361.

30 R. K. Harris, S. A. Joyce, C. J. Pickard, S. Cadars and L. Emsley, *Phys Chem Chem Phys*, **2006**, 8, 137-143.

31 L. H. Yang, S. A. Ahmed and E. W. Miles, *Protein Expression and Purification*, **1996**, 8, 126-136.

32 E. W. Miles, H. Kawasaki, S. A. Ahmed, H. Morita, H. Morita and S. Nagata, *Journal of Biological Chemistry*, **1989**, 264, 6280-6287.

33 A. Mozzarelli, A. Peracchi, G. L. Rossi, S. A. Ahmed and E. W. Miles, *Journal of Biological Chemistry*, **1989**, 264, 15774-15780.

34 A. L. Vangeet, *Analytical Chemistry*, **1968**, 40, 2227-&.

35 B. M. Fung, A. K. Khitrin and K. Ermolaev, *J Magn Reson*, **2000**, 142, 97-101.

36 Y. Li, B. J. Wylie and C. M. Rienstra, *J. Magn. Reson.*, **2006**, 179, 206-216.

37 D. J. States, R. A. Haberkorn and D. J. Ruben, *J Magn Reson*, **1982**, 48, 286-292.

38 D. G. Kneller and I. D. Kuntz, *Journal of Cellular Biochemistry*, **1993**, 254-254.

39 I. Bertini, L. Duma, I. C. Felli, M. Fey, C. Luchinat, R. Pierattelli and P. R. Vasos, *Angew Chem Int Edit*, **2004**, 43, 2257-2259.

40 W. Bermel, I. Bertini, I. C. Felli, R. Kummerle and R. Pierattelli, *J Magn Reson*, **2006**, 178, 56-64.

41 S. K. Straus, T. Bremi and R. R. Ernst, *Chem Phys Lett*, **1996**, 262, 709-715.

42 D. H. H. Zhou, K. D. Kloepper, K. A. Winter and C. M. Rienstra, *J Biomol Nmr*, **2006**, 34, 245-257.

43 G. De Paepe, A. Lesage and L. Emsley, *J Chem Phys*, **2003**, 119, 4833-4841.

44 E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Y. Yao and J. L. Markley, *Nucleic Acids Research*, **2008**, 36, D402-D408.

45 W. T. Franks, B. J. Wylie, H. L. F. Schmidt, A. J. Nieuwkoop, R. M. Mayrhofer, G. J. Shah, D. T. Graesser and C. M. Rienstra, *P Natl Acad Sci USA*, **2008**, 105, 4621-4626.

46 G. Bodenhausen, H. Kogler and R. R. Ernst, *J Magn Reson*, **1984**, 58, 370-388.

47 O. W. Sorensen, M. Rance and R. R. Ernst, *J Magn Reson*, **1984**, 56, 527-534.

48 S. Cadars, J. Sein, L. Duma, A. Lesage, T. N. Pham, J. H. Baltisberger, S. P. Brown and L. Emsley, *J Magn Reson*, **2007**, 188, 24-34.

49 K. Pervushin, R. Riek, G. Wider and K. Wuthrich, *P Natl Acad Sci USA*, **1997**, 94, 12366-12371.

50 T. R. M. Barends, M. F. Dunn and I. Schlichting, *Current Opinion in Chemical Biology*, **2008**, 12, 593-600.

51 M. F. Dunn, D. Niks, H. Ngo, T. R. M. Barends and I. Schlichting, *Trends in Biochemical Sciences*, **2008**, 33, 254-264.

52 E. W. Miles, *Advances in Enzymology and Related Areas of Molecular Biology*, **1991**, 64, 93-&.

53 E. W. Miles, S. Rhee and D. R. Davies, *Journal of Biological Chemistry*, **1999**, 274, 12193-12196.

54 P. Pan, E. Woehl and M. F. Dunn, *Trends in Biochemical Sciences*, **1997**, 22, 22-27.

55 Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98-102.

56 Manolikas, T.; Herrmann, T.; Meier, B. H. *Journal of the American Chemical Society* **2008**, *130*, 3959-3966.

57 Franks, W. T.; Wylie, B. J.; Schmidt, H. L. F.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, C. M. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 4621-4626.

58 Cornilescu, G.; Delaglio, F.; Bax, A. *Journal of Biomolecular Nmr* **1999**, *13*, 289-302.

59 Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 4685-4690.

60 Shen, Y.; Vernon, R.; Baker, D.; Bax, A. *Journal of Biomolecular Nmr* **2009**, *43*, 63-78.

61 Wishart, D. S.; Arndt, D.; Berjanskii, M.; Tang, P.; Zhou, J.; Lin, G. *Nucleic Acids Research* **2008**, *36*, W496-W502.

62 Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387-1391.

63 Bradley, P.; Misura, K. M. S.; Baker, D. *Science* **2005**, *309*, 1868-1871.

64 Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364-1368.

65 Barth, P.; Wallner, B.; Baker, D. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 1409-1414.

66 Barth, P.; Schonbrun, J.; Baker, D. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, *104*, 15682-15687.

67 Yarov-Yarovoy, V.; Schonbrun, J.; Baker, D. *Proteins-Structure Function and Bioinformatics* **2006**, *62*, 1010-1025.

68 Shen, Y.; Bax, A. *Journal of Biomolecular Nmr* **2007**, *38*, 289-302.

69 Schmidt, H. L. F.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. A. *Journal of Physical Chemistry B* **2007**, *111*, 14362-14369.

70 Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *Journal of Magnetic Resonance* **2003**, *160*, 65-73.

71 Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2006**, *48*, 47-62.

CHAPTER III

^{31}P NMR INVESTIGATION OF BACKBONE DYNAMICS IN DNA BINDING SITES

ABSTRACT

The backbone conformation of DNA plays an important role in the indirect readout mechanisms for protein-DNA recognition events. Thus, investigating the backbone dynamics of each step in DNA binding sequences provides useful information necessary for the characterization these interactions. In this work we use ^{31}P Dynamic NMR to characterize the backbone conformation and dynamics in the Dickerson Dodecamer, and confirm solid-state ^{2}H-NMR results showing that the C3pG4 and C9pG10 steps experience unique dynamics and that these dynamics are quenched upon cytosine methylation. In addition, cytosine methylation affects the conformation and dynamics of neighboring nucleotide steps but this effect is localized to only near neighbors and base pairing partners. The results demonstrate that ^{31}P Dynamic NMR provides a robust and efficient method for characterizing the backbone dynamics in DNA. This allows simple, rapid determination of sequence-dependent dynamical information, providing a useful method for studying trends in protein-DNA recognition events.

3.1 INTRODUCTION

DNA-protein interactions are central to biology and play a crucial role in cellular tasks such as DNA replication and transcription. Understanding how proteins recognize particular DNA sequences holds tremendous promise for understanding basic cell cycle events. X-ray diffraction and high resolution NMR experiments have provided detailed structures of numerous DNA sequences. These studies have demonstrated that DNA is not simply a rigid rod of nucleotides for direct readout. Instead, local structural variations exist within DNA strands that most likely play a significant role in recognition of these sequences by proteins.[1,2] What has also become clear in recent years is that this recognition process is also dependent upon the dynamics of the DNA and proteins involved[3]. One part of this indirect readout is the conformational equilibrium between the BI and BII sub-states of the DNA backbone[4,5] and this sequence-dependent flexibility contributes to an indirect readout by cognate proteins.[6-10]

The local and global dynamics of polynucleotides has been investigated by high-resolution NMR,[11-17] solid-state NMR,[10,18-20] Fluorescence Polarization Anisotropy and Dynamic Light Scattering,[21,22] Electron Paramagnetic Resonance,[23] Molecular Dynamics,[24,25] and X-ray Diffraction.[26] However, obtaining accurate information about local dynamics is difficult because the interpretation of X-ray crystallography, molecular dynamics, and high-resolution NMR data requires numerous assumptions about the

structure and relaxation mechanisms of the molecule.[18] These assumptions become increasingly complicated in the realm of biological macromolecules.[21,22]

Solid-state deuterium NMR has been successful in providing accurate dynamical information because it exploits the deuterium quadrupole moment, a single nucleus property. This allows 2H-SSNMR to provide local dynamic information that is free of structural assumptions.[18] Using systematic isotopic labeling schemes, the local dynamics of the base, sugar, and backbone moieties of individual nucleotides within a sequence can be investigated via 2H-SSNMR. Such studies have uncovered local regions of large amplitude dynamics of the sugar and backbone moieties within the EcoRI and HhaI DNA binding sites.[19,20] In addition, deuterium NMR has been used to demonstrate the quenching of these dynamics due to cytosine methylation.[10,20]

Solution NMR studies have also been conducted on a number of DNA sequences in an effort to understand conformational changes.[11,27,28] Much of this work has been focused on the BI→BII transitions of the phosphate backbone. These conformations are defined by the ε and ζ torsion angles, C4'-C3'-O3'-P and C3'-O3'-P-O5', respectively. BI, the most common conformation in B-form DNA, is trans in ε and gauche- in ζ (t,g-) while BII is gauche- in ζ and trans in ζ (g-,t). A more general definition of these conformations is determined by the difference in these torsion angles, (ε–ζ), the difference being ~ -90° for the BI conformation and >0 for the BII conformation.[4,12]

Phosphorus chemical shifts and proton-phosphorus scalar couplings have been widely used to characterize these backbone conformations. A review of numerous ^{31}P NMR studies demonstrates that higher frequency ^{31}P resonances are indicative of nucleotide steps residing in the BII conformation or in conformational exchange where BII conformations are prevalent.[12] In addition, scalar couplings between the phosphorus and H3' of the sugar ring (^{3}J(P-H3')) are larger for BII conformations. A reparameterized Karplus equation for these coupling constants has been used to determine the percentage of BI conformer present in solution.[13] The need to extract accurate ^{3}J(P-H3') couplings led to development and use of an array of pulse sequences to carefully measure this scalar coupling.[29-31] However, carefully determining scalar couplings is still a difficult task with often unreliable accuracy[32] so using chemical shifts is desirable. Recently, a relationship between the ^{31}P chemical shift and the %BII has been developed allowing the translation of δP in terms of BI/BII ratios.[32] The ability to assign a %BII is extremely useful, but percent population is an equilibrium property that does not provide the detailed dynamical information needed to address the role of DNA dynamics in protein-DNA binding events. We demonstrate the use of ^{31}P dynamic NMR to extract this detailed dynamical information. We have used the relationship derived by the Hartmann group as a starting point to constrain a two-site model of the conformational

exchange between BI and BII in the backbones of nucleotide steps using dynamic NMR.[32]

We examine the backbone dynamics in the Dickerson Dodecamer, $[d(CGCGAATTCGCG)]_2$, the first synthetic oligonucleotide crystallized in the biologically relevant B-form helix.[1,2,26] This structure revealed the presence of sequence-dependent structural variations including; a large twist angle between T8 and C9, a splay in the G4-C9 base pair and an overall bend of 19°.[2] An important feature of the Dickerson Sequence is that it contains the binding site for the EcoRI restriction endonuclease, -G$^{\downarrow}$AATTC-, where the arrow indicates the cutting site. Restriction endonucleases bind and cleave DNA with remarkable selectivity and accuracy making this class of proteins good candidates for preliminary studies of DNA-protein interactions.

Solid-State ^{2}H-NMR has probed the local dynamics of each nucleotide in the Dickerson Dodecamer,[10,18,19,33,34] providing quality data with which to compare the results obtained from our ^{31}P dynamic NMR technique. However, Solid-State ^{2}H-NMR experiments require extensive sample preparation, including the synthesis of isotopically labeled nucleotides and phosphoramidites, and allow the study of only one nucleotide step at a time. Solution ^{31}P NMR allows the study of each backbone step simultaneously, using natural abundance DNA samples. This proves to be exceedingly

beneficial as it permits the analysis of a large number of DNA sequences rapidly and efficiently and enables the investigation of changes in flanking sequences, methylation states, pH, counter-ion and numerous other variables that may be important in protein-DNA recognition events.

A phosphorus NMR study of the Dickerson Dodecamer has been reported previously.[35] It was seen that the ^{31}P resonances for the CpG steps (phosphates 3 and 9) were at higher field than expected, indicating a change in backbone conformation at these steps. The crystal structure of this dodecamer also showed differences at these steps. The crystal demonstrated a regular B-type structure except for a break at steps 3 and 9 which had unusually small twist angles.[1] Increased temperature resulted in an overall narrowing and a shift to higher field of most resonances suggesting an increase in the BII conformer in the conformational equilibrium of the backbone. Phosphates 3 and 9 show the smallest chemical shift change, presumably due to their having a high population of BII even at lower temperatures. This shift and line narrowing with increased temperatures were also observed in the CpG step of the octamer $d(CTTCGAAG)_2$ and was described as a conformational exchange between the two backbone conformations.[36] It is this conformational exchange which we have characterized via dynamic NMR in order to extract rate and thermodynamic parameters.

The wealth of information already known about the Dickerson Sequence makes it the ideal sequence to test the validity of our dynamic NMR approach. Comparison of this sequence and its methylated counterpart with previous deuterium NMR results demonstrates the robustness and efficiency of the method described here. Additional studies demonstrating changes in backbone dynamics with changes in flanking sequences are already underway in our lab.

3.2 THEORY OF TWO-SITE FAST EXCHANGE MODEL IN NMR

The BI/BII conformational exchange in DNA oligonucleotides falls into the fast exchange regime, where the exchange rate is larger than the difference in chemical shifts between the two sites.[37-39] In this limit, a single resonance line is observed experimentally for each XpY step of the flanking sequence. We employ a two-state model to describe this exchange using the potential energy well shown in Figure 3.1.

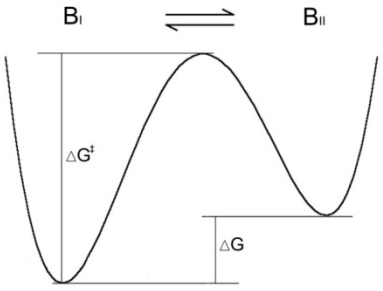


Figure 3.1. The potential energy diagram describing the two-state conformational exchange between BI and BII.

In this model, BI is taken as the zero-energy state, ΔG is the free energy difference between the BI and BII conformers, and $\Delta G^{\neq}$ is the free energy of activation, again defined relative to the BI state. The system is in dynamic equilibrium, with

$$B_I \underset{k_r}{\overset{k_f}{\rightleftarrows}} B_{II} \tag{3.1}$$

and the observed magnetization in the NMR experiment can be written as the sum of unitary magnetization precession, relaxation, and exchange.[40-44]

$$\frac{d}{dt}\begin{pmatrix} M_{BI} \\ M_{BII} \end{pmatrix} = -(iL + R + K)\begin{pmatrix} M_{BI} \\ M_{BII} \end{pmatrix} \tag{3.2}$$

given respectively by the Liouvillian matrix,

$$L = -\begin{pmatrix} \omega_{BI} & 0 \\ 0 & \omega_{BII} \end{pmatrix} \tag{3.3}$$

relaxation matrix and the exchange matrix,

$$R = \begin{pmatrix} \frac{1}{T_2} & 0 \\ 0 & \frac{1}{T_2} \end{pmatrix} \tag{3.4}$$

$$K = \begin{pmatrix} k_f & -k_r \\ -k_f & k_r \end{pmatrix} \tag{3.5}$$

T2 is the spin-spin relaxation time, which we assume to be equal for both conformers (this assumption can be easily relaxed), ω is the chemical shift for each of the exchanging conformer sites (in rad/s), and the k's are the exchange rates defined by equation 3.1. This set of linear coupled differential equations can be solved by the eigenvector method, in which the eigenvalues and eigenvectors of the combined matrix,

$$-(i\mathbf{L}+\mathbf{R}+\mathbf{K}) = \begin{pmatrix} i\omega_{BI} - k_f - \frac{1}{T_2} & k_r \\ k_f & i\omega_{BII} - k_r - \frac{1}{T_2} \end{pmatrix} \tag{3.6}$$

define the resulting lineshapes in the time and frequency domains.[40,42,43] The eigenvalues of this matrix are

$$\Lambda_{+/-} = -\frac{1}{T_2} - \frac{1}{2}\left(k_f + \frac{k_f}{K} - i(\omega_{BI} + \omega_{BII}) \pm \sqrt{\left(k_f + \frac{k_f}{K}\right)^2 + 2i\left(k_f - \frac{k_f}{K}\right)(\omega_{BII} - \omega_{BI}) - (\omega_{BII} - \omega_{BI})^2} \right) \tag{3.7}$$

where the equilibrium constant K = kf / kr. As is well-known in chemical exchange, the positive root leads to a quickly decaying time-domain signal with vanishing intensity in fast-exchange, while the negative root leads to a line that continues to narrow.[37] Taking

this fast exchange limit with $k_f, k_r >> (\omega_{BII} - \omega_{BI})$, the real part of equation 3.7 gives an effective line width,

$$r = \frac{1}{\pi T_2} + \frac{K^2 (\omega_{BII} - \omega_{BI})^2}{\pi (1+K)^3 k_f} \tag{3.8}$$

while the imaginary part gives the resonance offset leading to the observed averaged chemical shift

$$\langle \omega \rangle = \frac{\omega_{BI}}{(1+K)} + \frac{K\ \omega_{BII}}{(1+K)} \tag{3.9}$$

The connection with the activation energy is provided by Eyring's equation,[37-39,42,4345-47] with

$$k_f = \frac{K_b T}{h} \exp(-\Delta G^{\neq} / RT) \tag{3.10}$$

Here Kb is Boltzmann's constant, h is Planck's constant, and R is the gas constant, which allows the effective line width to be written as

$$r = \frac{1}{\pi T_2} + \frac{K^2 (\omega_{BII} - \omega_{BI})^2}{(1+K)^3} \frac{h}{\pi K_b T} e^{\Delta G^{\neq} / RT} \tag{3.11}$$

While equation 3.11 is valid in the fast-exchange limit, its application can be troublesome if independent measures of K and the slow-exchange chemical shifts are not

available.[12,32] Assuming that the pure BI and BII conformations each have a single known ^{31}P chemical shift, ω_{BI} and ω_{BII}, respectively, that is independent of sequence, one can use these to write a linear relationship between the observed fast-exchange chemical shift and the %BII conformation:

$$\%BII = \frac{1}{\omega_{BII} - \omega_{BI}}\langle\omega\rangle - \frac{\omega_{BI}}{\omega_{BII} - \omega_{BI}} \tag{3.12}$$

as well as the equilibrium constant

$$K = \frac{\langle\omega\rangle - \omega_{BI}}{\omega_{BII} - \langle\omega\rangle} \tag{3.13}$$

Such a linear relationship for %BII was recently introduced by Heddi *et al*.[32] as an empirical equation, but as shown above it is also the result of a two-site exchange model. Estimates of these values will be obtained experimentally from the extrema of a large sampling of observed chemical shifts. Along with equation 3.12, they provide the needed independent measures of K and $(\omega_{BII} - \omega_{BI})$ that allow us to fit the ^{31}P line widths in the fast exchange limit and extract T_2 and $\Delta G^{\neq}$ for each step in the sequence.

3.3 APPLICATION IN DNA BACKBONE DYNAMICS

3.3.1 SAMPLE PREPARATION

DNA samples are obtained from Alpha DNA (Montreal, Canada). A 1-2 mM solution of DNA was prepared in 500 μM of a 100 mM phosphate buffer (pH 7.4). The sample was lyophilized overnight and re-dissolved in 0.6 ml of 99.9% D_2O immediately before use.

3.3.2 EXPERIMENTAL SETUP

Backbone dynamics of DNA fall into the region of fast exchange, where the line from each resonance will narrow up by increasing temperature. The phenomenon is observable even on the 1D experiments, Figure 3.2.

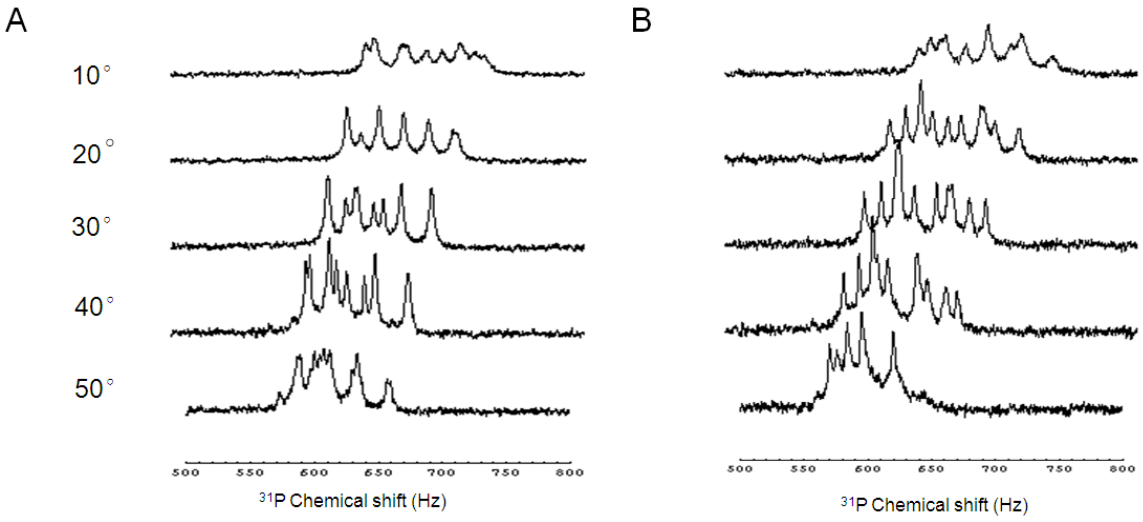


Figure 3.2 Temperature dependent ^{31}P NMR spectra. (A) $d(CGCGAATTCGCG)_2$ native Dickerson sequence. (B) $d(CGCGAATTC*GCG)_2$,C9 methylated Dickson sequence

NMR experiments were carried out on a Bruker Avance spectrometer (^{1}H frequency 600 MHz) with a 5mm gradient probe. 1D ^{1}H and ^{31}P experiments were acquired to set the spectral window at each temperature. 2D ^{1}H nuclear Overhauser effect (NOESY) experiments were performed with mixing times of 500 ms and a gradient pulse applied during the mixing time. A total of 4096 data points were collected in the t_2 dimension, and 512 increments, in t_1. The matrix was zero-filled to 4096 × 1024 points before double Fourier transformation. Phase sensitive ^{31}P-^{1}H HSQC were recorded using Echo/Antiecho- TPPI gradient selection[16,48,49] and decoupling during acquisition. All HSQC were performed with 2048 points in the t_2 dimension and 256 t_1 increments using a relaxation delay of 4s. The chemical shift of H_2O and the center of the DNA phosphorus resonances were used to set the center of HSQC and NOESY spectra. A spectral width of 9 ppm was used for all 1D and both dimensions of 2D experiments.

3.3.3 TEMPERATURE AND ^{31}P CHEMICAL SHIFT CALIBRATIONS

Sample Temperatures are controlled by an FTS Airjet temperature preconditioner (FTS,Stone Ridge, NY). Electronic thermometer (Omega HH509) was used to measure temperature in the spectrometer, from 10℃ to 50℃.

As IUPAC recommended, H_3PO_4 is used as ^{31}P chemical shift reference, which is defined as 0 ppm. In these experiments, ^{31}P chemical shifts change by temperature. In

order to calibrate ^{31}P chemical shifts, we use co-axis tube, with 85% H_3PO_4 in the inner tube and DNA sample with phosphate buffer (pH=7.4) in the outer tube. Single pulse ^{31}P experiments were ran through the whole temperature range, and ^{31}P chemical shifts of H_3PO_4 and phosphate buffer were measured. Figure 3.3 shows the shifts of 85% phosphoric acid and the buffer peak. Due to the concentration of DNA and the scale in this figure, DNA peaks are not seen at low temperature, where they are located at -1 ppm. However they become visible at higher temperatures.

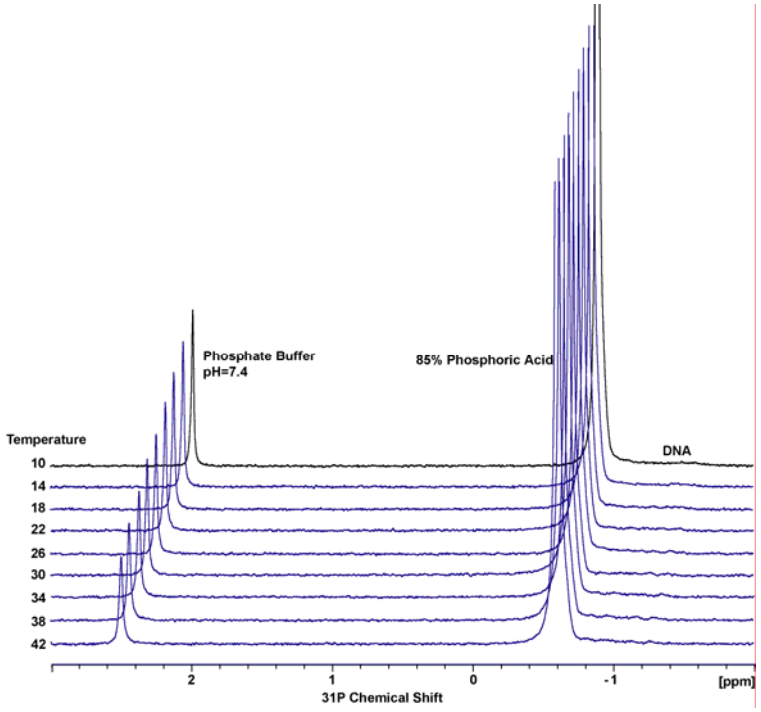


Figure 3.3 Temperature dependent ^{31}P 1D spectra of 85% H_3PO_4 (before calibration ~-0.6ppm as shows) and DNA sample in pH=7.4 phosphate buffer, from 10 to 42℃. Spectra were taken using co-axis tube. DNA peaks are at -0.8 ~ -1.4 ppm, which have much lower intensity than the buffer peak.

3.3.4 NOESY AND HSQC ASSIGNMENTS

Resonance assignments are achieved by combining NOESY and HSQC. Since H2' and H2'' are strongly coupled with H1', H3', and H5' (sugar protons) and to the base protons, we are able to assign protons on the NOESY using a sequential walk.[50] Because of the nature of NOESY experiment, cross peaks on spectrum are from protons that are close to each other in space. So the observable correlations are H2'-H3', H2'(i)-H6/H8(i), and H2'(i)-H6/H8(i+1) etc. In an structure of Dickerson sequence in protein data band (PDB id: 1NAJ)[52] which is solved by solution NMR, the three distances above are 2.31Å, 2.09Å and 3.00Å respectively, using T7-T8 step as a sample, shows as Figure 3.4 below.

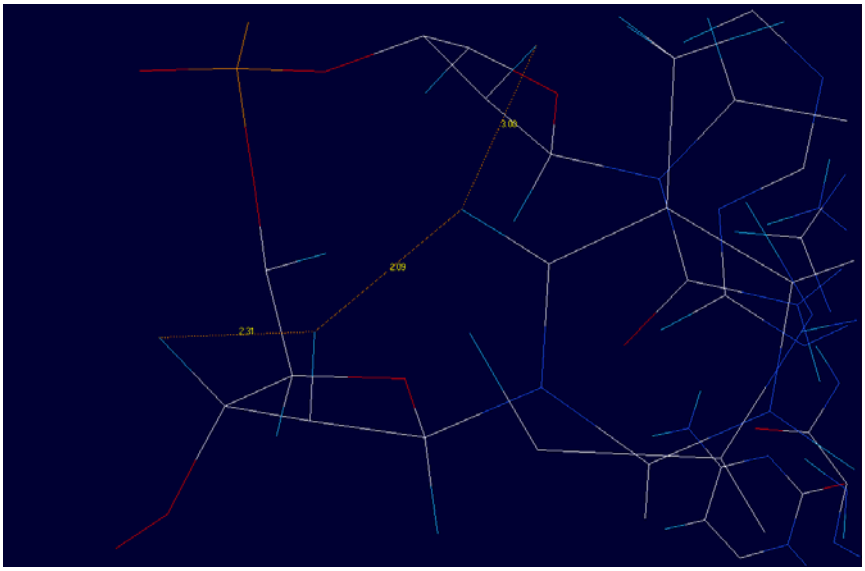


Figure 3.4 Distances of H2'(T8)-H3'(T8), H2'(T7)-H6(T8) and H2'(T8)-H6(T8). Model from Protein Data Bank, id=1NAJ.

NOESY spectrum was taken at 18 and 34℃ for ^{1}H assignments, and H2'-H3', H2'(i)-H6/H8(i), and H2'(i)-H6/H8(i+1) correlations were used in sequential walk. The sequential walk is initialized at two characteristic resonances which are from H2'-H6 correlation on the first cytosine. For the first step on DNA sequence only has the H2'(1)-H6/H8(1) correlation peaks along the indirect dimension, while the other steps has both H2'(i)-H6/H8(i), and H2'(i-1)-H6/H8(i) along the indirect dimension, this is a recognition for the first step on NOESY spectrum.

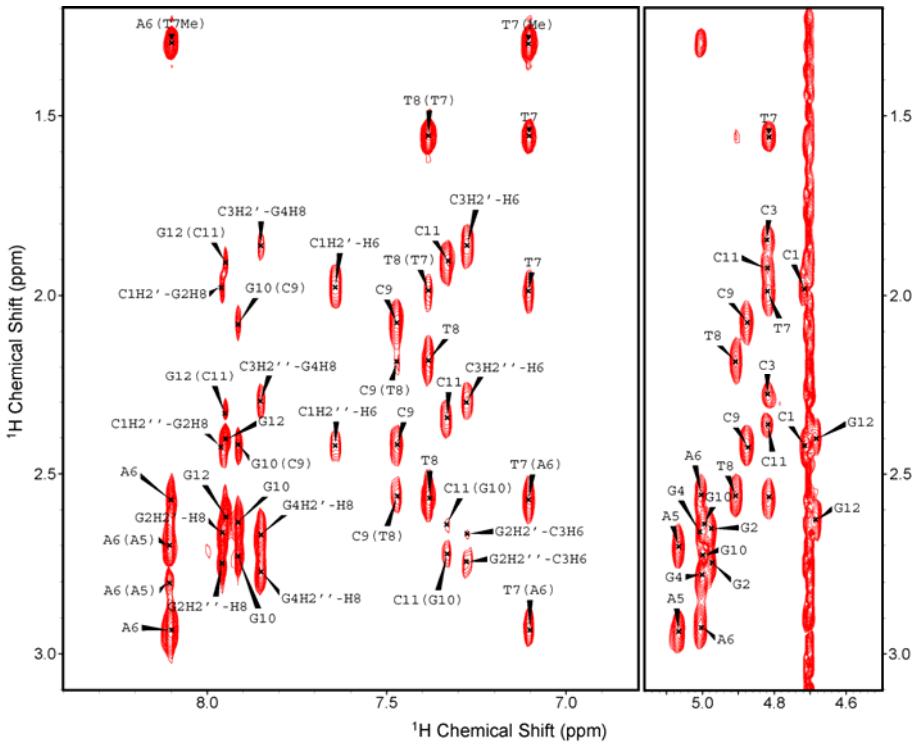


Figure 3.5 A representative NOESY spectrum of the native Dickerson sequence at 34°C. The two regions are showing H2'(1.5-3.0ppm)-H6/H8(7.0-8.5ppm) , and H2'(1.5-3.0ppm)-H3'(4.6-5.2ppm) correlations respectively.

The strong correlations between ^{31}P and H3' and H5'/H4' then enabled us to assign the HSQC using the proton assignments determined from the NOESY (Figure 3.5). The difficulty in linewidth analysis caused by overlapping resonances in the HSQC was mitigated somewhat by the fact that each phosphorus is coupled to both H3' and H4', which provides two HSQC cross peaks for phosphorus. Resonances in the region from 4.6 to 5.2 ppm on ^{1}H dimension are from H3'-P $^{3}J(^{1}H-^{31}P)$ correlations, and in the region from 4.0 to 4.6 ppm on ^{1}H dimension are from H5'-P $^{3}J(^{1}H-^{31}P)$ correlations. Both of the two regions are used alternatively to make assignments (Figure 3.6).

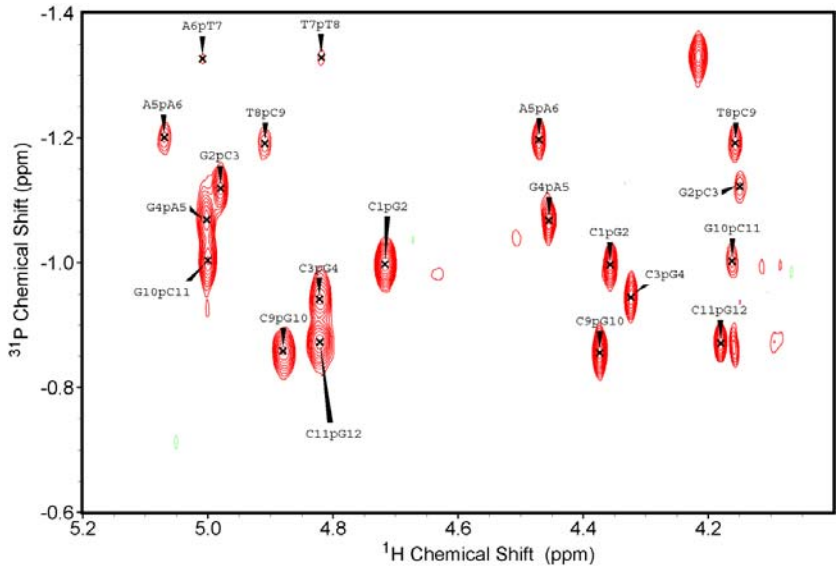


Figure 3.6 A representative HSQC spectrum of the native Dickerson sequence at 30 °C showing the DNA backbone ^{31}P resonances vs the H3'/H5' region.

3.3.5 LINEWIDTH ANALYSIS

1D ^{31}P traces were extracted from the indirect dimension of the HSQC spectra for each XpY step of the DNA sequences and fit to a single Lorentzian plus a simple baseline offset in Mathematica (Wolfram Research, Champaign, IL). Lorentzian half-widths from this fitting procedure (Figure 3.7) were then fit as a function of temperature to equation 3.11 to extract T2 and the free energy of activation. Due to the accuracy of our temperature and chemical shift data, the percent BII is reported with three significant figures, and the error bars for the transition state parameters were determined through a numerical calculation of the covariance matrix and are reported as 1 standard error.

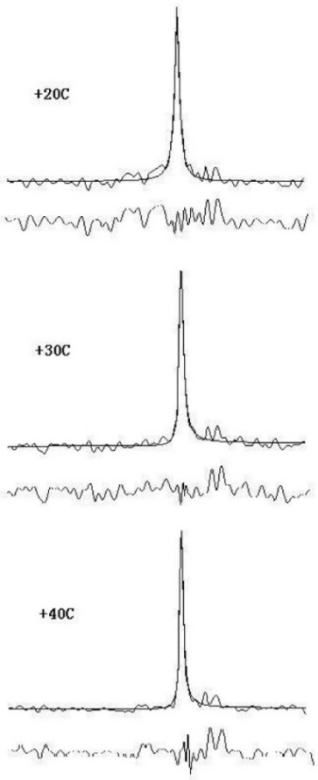


Figure 3.7 The C9pG10 ^{31}P resonances at 20, 30, and 40 °C Lorentzian fits and residuals.

Unlike the slow exchange experiments, we assume the two confirmations are unequally populated, and we observe this from %BII calculation using empirical equations. In order to fit for energy of activations, this model we describe requires both linewidth and equilibrium constants as input. So following steps were taken to calculate equilibrium constants.

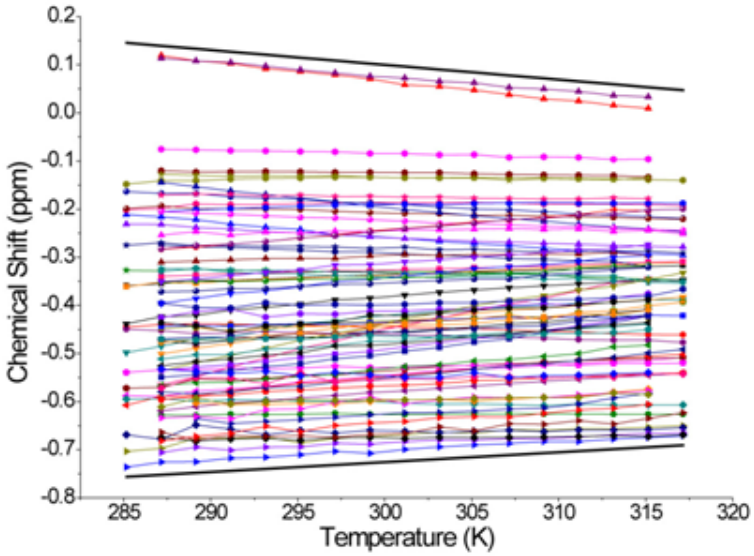


Figure 3.8 ^{31}P chemical shifts vs temperature for seven sequences: $[d(CGCGAATTCGCG)]_2$ native, C3 and C9 methylated; $[d(TTCGAATTCGAA)]_2$ native and C9 methylated; $[d(AAAGAATTCTTT)]_2$ native and C9 methylated.

In this procedure, the equilibrium constants were calculated at each temperature using equation 3.13. ^{31}P chemical shifts were referenced to 85% phosphoric acid using

a coaxial tube insert at each temperature. No attempt was made to separate out entropy and enthalpy terms in the free energy, which was assumed to be negligibly temperature dependent over this rather restricted temperature range.

Figure 3.8 shows a plot of ^{31}P chemical shifts vs. temperature for the following 7 sequences: $[d(CGCGAATTCGCG)]_2$ native, C3 methylated and C9 methylated; $[d(TTCGAATTCGAA)]_2$ native and C9 methylated; $[d(AAAGAATTCTTT)]_2$ native and C9 methylated. These shifts all lie between the two flanking lines:

$$\delta P_{BI} = 0.002031\,T(K) - 1.345 \tag{3.14}$$

and

$$\delta P_{BII} = -0.003079\,T(K) + 1.023 \tag{3.15}$$

which we take as the temperature-dependent limiting shifts for the pure BI and BII conformations. This gives for equation 3.12

$$\%B_{II} = \frac{1}{2.368 - 0.005110T}\delta P(ppm) + \frac{1.345 - 0.002031T}{2.368 - 0.005110T} \tag{3.16}$$

which compares fairly well with the original equation of Heddi et al near room temperature,

$$\%B_{II} = 143\delta P(ppm) + 85 \tag{3.17}$$

although we find that Heddi's equation leads to %BII values outside the range of 0 to 100%.

A representative HSQC spectrum of the native Dickerson Sequence at 30℃ is shown in Figure 3.6. From these, the phosphorus 1D spectra were extracted for line shape analysis. For example, Figure 3.7 shows the C9pG10 phosphorus resonance at 20, 30 and 40℃ and the Lorentzian fit for each, providing the chemical shift and line width for the subsequent analysis.

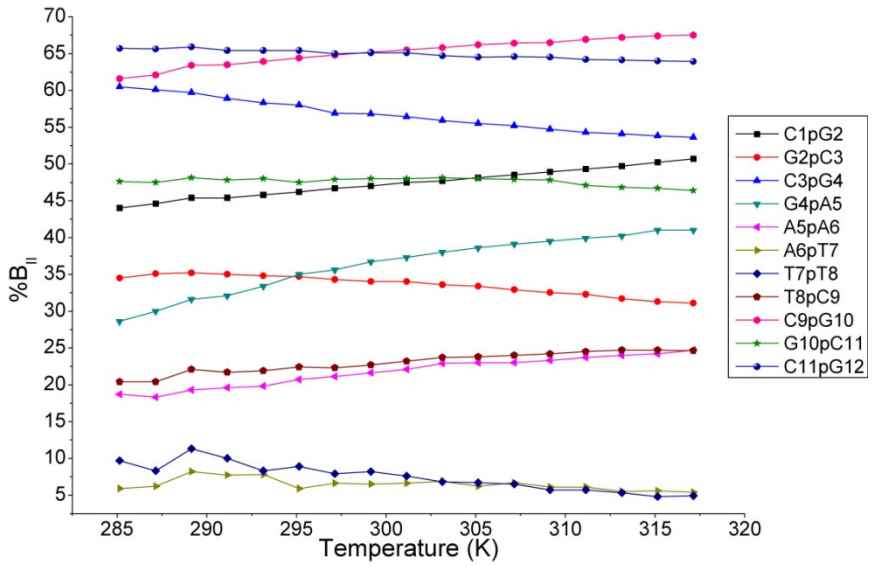


Figure 3.9 The calculated percent BII for each nucleotide step in the native Dickerson sequenceas as a function of temperature.

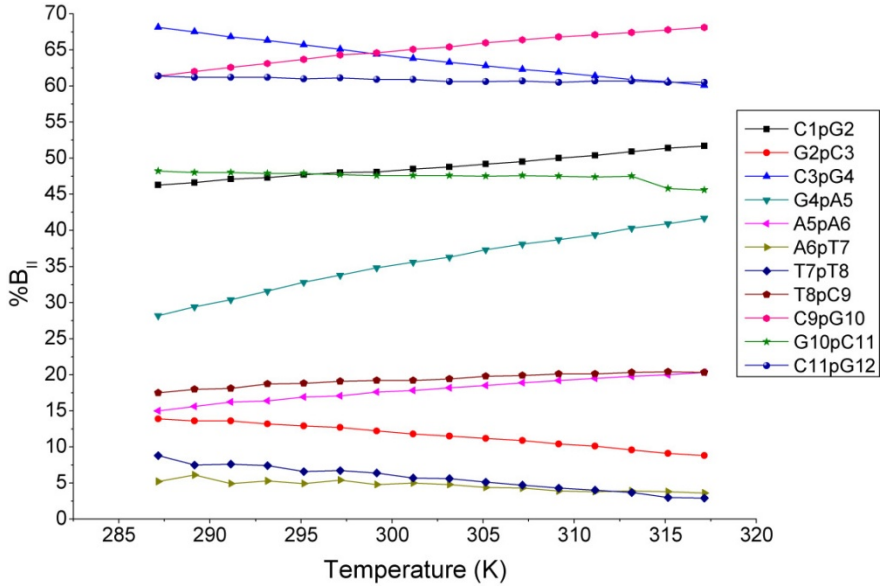


Figure 3.10 The calculated percent BII for each nucleotide step in the C3 methylated Dickerson Sequences as a function of temperature.

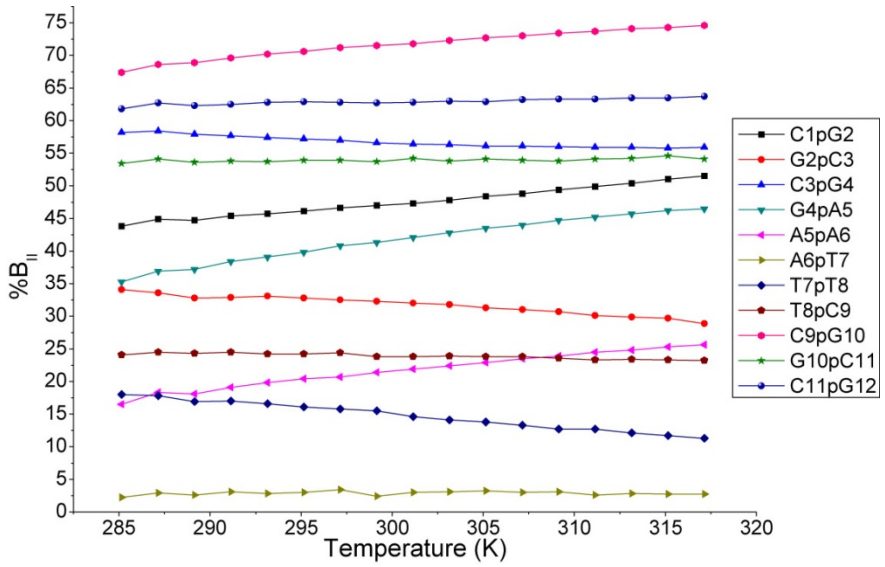


Figure 3.11 The calculated percent BII for each nucleotide step in the C9 methylated Dickerson Sequences as a function of temperature.

The percent BII (calculated from equation 3.16) as a function of temperature for each resonance in the native Dickerson sequence is shown in Figure 3.9. Similar plots for the C3 and C9 methylated sequences can be found in Figures 3.10 and 3.11, respectively. From Figure 3.9, it can be seen that as the temperature is increased, the majority of the backbone steps are shifting to higher frequency, toward a greater percentage of BII. However C3pG4 and C11pG12, the resonances already experiencing a great extent of percent BII at 12 °C (60.5 and 65.7%, respectively), actually shift to lower frequency, demonstrating an increased sampling of the BI conformation. The phosphate step at the C9pG10 position also shows a large percent BII at low temperatures, but its chemical shift remains relatively constant across the temperature range studied (-0.137 ppm). The phosphate steps at G2pC3 and G10pC11, the base pair partners of C11pG12 and C3pG4, also shift toward lower frequency with increasing temperature, although, like C11pG12, they experience a much smaller change in chemical shift than C3pG4. As expected, the percent BII is lowest in the center of the sequence and increases as we move out toward the ends of the sequence and toward decreased base stacking. For example, note the lines for A6pT7 and T7pT8 and for A5pA6 and T8pC9. These center resonances, presumed to be the most stable by virtue of their position and the increased base stacking conferred upon them, have chemical shift profiles that have the lowest percent BII and lie nearly on top of each other. G4pA5 has the next highest percent BII, but C9pG10 is

significantly higher. The variation is also large for the C3/G10 base pair. C3 methylation (Figure 3.10) causes an increase in percent BII for the C3pG4 and C1pG2 steps and adecrease in the G2pC3 step. C9 Methylation (Figure 3.11) causes an increase in percent BII for the C9pG10, G10pC11, T7pT8, and G4pA5 steps. Table 3.1 provides the percent BII as a function of temperature for all 11 resonances. Similar tables for the C3 and C9 methylated sequences can be found in the appendix. These trends in percent BII can be better observed in Figure 3.10 and 3.11.

Table 1. The %BII for each nucleotide step in the native sample.

TempK	C1pG2	G2pC3	C3pG4	G4pA5	A5pA6	A6pT7	T7pT8	T8pC9	C9pG10	G10pC11	C11pG12
285.2	44.0	34.5	60.5	28.6	18.7	5.9	9.7	20.4	61.6	47.6	65.7
287.2	44.6	35.1	60.1	30.0	18.3	6.2	8.3	20.4	62.1	47.5	65.6
289.2	45.4	35.2	59.7	31.6	19.3	8.2	11.3	22.1	63.4	48.1	65.9
291.2	45.4	35.0	58.9	32.1	19.6	7.7	10.0	21.7	63.5	47.8	65.4
293.2	45.8	34.8	58.3	33.4	19.8	7.8	8.3	21.9	63.9	48.0	65.4
295.2	46.2	34.7	58.0	35.0	20.7	5.9	8.9	22.4	64.4	47.5	65.4
297.2	46.7	34.3	56.9	35.6	21.1	6.6	7.9	22.3	64.8	47.9	65.0
299.2	47.0	34.0	56.8	36.7	21.6	6.5	8.2	22.7	65.2	48.0	65.1
301.2	47.5	34.0	56.4	37.3	22.1	6.6	7.6	23.2	65.5	48.0	65.1
303.2	47.7	33.6	55.9	38.0	22.9	6.9	6.8	23.7	65.8	48.1	64.7
305.2	48.1	33.4	55.5	38.6	23.0	6.2	6.7	23.8	66.2	48.0	64.5
307.2	48.5	32.9	55.2	39.1	23.0	6.7	6.5	24.0	66.4	47.9	64.6
309.2	48.9	32.5	54.7	39.5	23.3	6.1	5.7	24.2	66.5	47.8	64.5
311.2	49.3	32.3	54.3	39.9	23.7	6.1	5.7	24.5	66.9	47.1	64.2
313.2	49.7	31.7	54.1	40.2	24.0	5.5	5.3	24.7	67.2	46.8	64.1
315.2	50.2	31.3	53.8	41.0	24.2	5.6	4.8	24.7	67.4	46.7	64.0
317.2	50.7	31.1	53.6	41.0	24.7	5.4	4.9	24.6	67.5	46.4	63.9

Table 2. Calculated thermodynamic parameters for each nucleotide step.

	Native Dickerson			C9Me Dickerson			C3Me Dickerson		
	$G^{\ddagger}$ (kcal/mol)	G (kcal/mol)	T_2 (sec)	$G^{\ddagger}$ (kcal/mol)	G (kcal/mol)	T_2 (sec)	$G^{\ddagger}$ (kcal/mol)	G (kcal/mol)	T_2 (sec)
C1pG2	12.17±0.05	0.0783	0.0371	12.30±0.05	0.0801	0.0365	12.61±0.04	0.0478	0.0262
G2pC3	12.57±0.10	0.383	0.0319	12.69±0.07	0.431	0.0324	13.67±0.06	1.14	0.0310
C3pG4	12.08±0.10	-0.165	0.0312	12.23±0.03	-0.166	0.0318	12.19±0.06	-0.368	0.0321
G4pA5	12.93±0.08	0.349	0.0376	12.75±0.10	0.219	0.0341	13.01±0.06	0.396	0.0390
A5pA6	13.28±0.08	0.779	0.0365	13.44±0.07	0.792	0.0339	13.51±0.08	0.930	0.0324
A6pT7	14.32±0.16	1.57	0.0306	15.31±0.11	1.98	0.0290	14.89±0.17	1.69	0.0265
T7pT8	13.98±0.14	1.45	0.0262	13.77±0.06	0.989	0.0304	14.44±0.21	1.56	0.0260
T8pC9	13.54±0.06	0.737	0.0362	13.69±0.05	0.667	0.0308	13.56±0.04	0.854	0.0331
C9pG10	12.21±0.08	-0.360	0.0330	12.44±0.04	-0.534	0.0365	12.22±0.07	-0.347	0.0319
G10pC11	12.06±0.09	0.0502	0.0355	12.26±0.08	-0.0925	0.0313	12.37±0.06	0.0534	0.0322
C11pG12	11.93±0.07	-0.366	0.0406	12.18±0.08	-0.309	0.0366	11.99±0.10	-0.266	0.0285

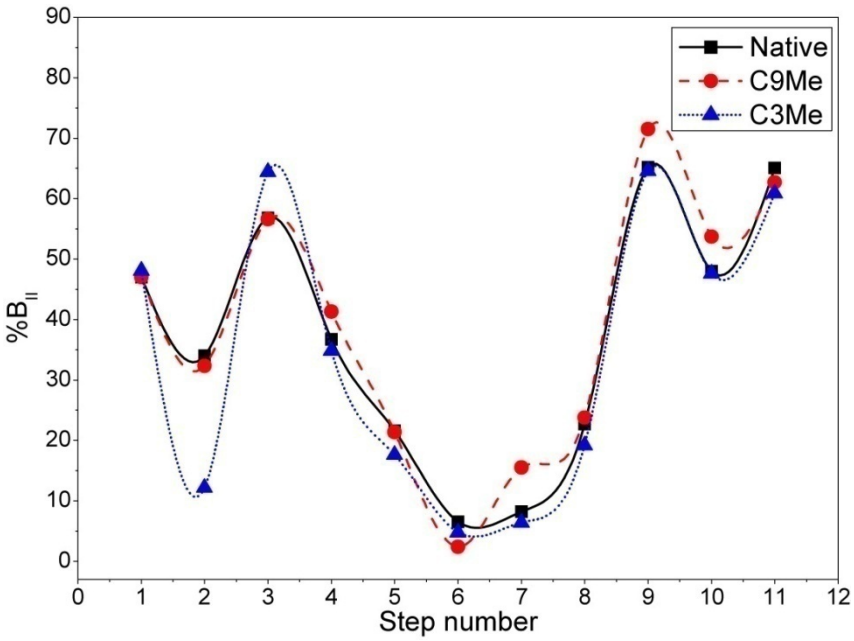


Figure 3.12 The percent BII at 25 °C as a function of position within the sequence.

Figure 3.12 shows the effect on the percentage of BII at 25°C as a function of sequence position. The plot is shown for the native and two methylated sequences. The plot of the native sequence shows the dramatic increase in percent BII at positions C3pG4 and C9pG10. Methylation of C3 results in an increase in percent BII at C3pG4, but also a significant decrease in percent BII of the C1pG2 and G2pC3 steps. However, methylation of C3 does not affect the backbone conformation of steps on the other end of the sequence. Methylation of C9 causes an increase in the percent BII of the C9pG10 and G10pC11 steps. Again, this effect is localized to the nucleotides near the site of methylation.

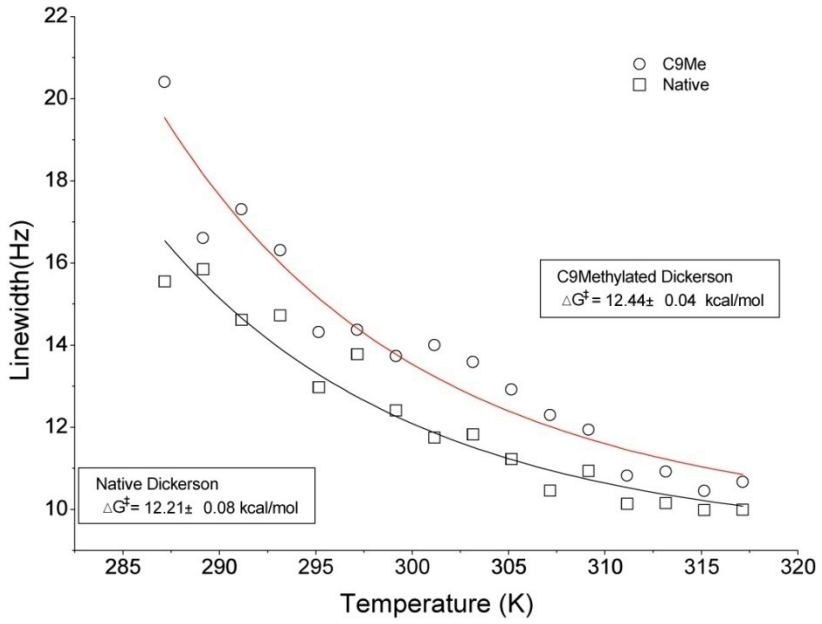


Figure 3.13 Line widths as a function of temperature for the C9pG10 step in the native (bottom) and C9 methylated (top) Dickerson sequences.

The dynamic behavior of the phosphate steps can be quantified using the two-site exchange model described in the Experimental Methods section. Representative plots of the line width as a function of temperature for the native and C9 methylated Dickerson sequence are shown in Figure 3.13. These fits allow us to determine both $\Delta G^{\neq}$ and T2 for this phosphate step and Table 3.2 summarizes the same fits for each step in the native and C3 and C9 methylated samples. The center nucleotides have the highest $\Delta G^{\neq}$ and are the least dynamic, as expected for phosphate steps experiencing the greatest base stacking. The $\Delta G^{\neq}$ values of C3pG4 and C9pG10 steps in the native Dickerson

sequence are similar to those of the ends and in the case of the C3pG4, lower, explaining the increased dynamics observed for these nucleotides in 2H-SSNMR studies.[19]

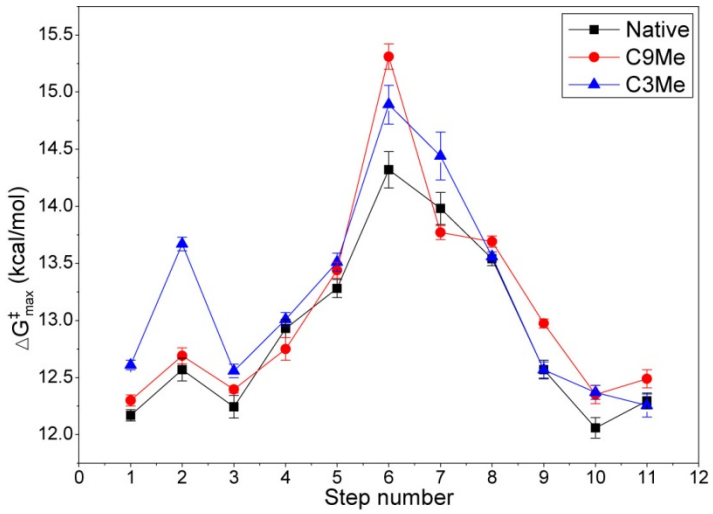


Figure 3.14 The energy difference between the transition state and most stable conformation for each nucleotide step. The squares represent the native Dickerson Sequence, the circles the C9 methylated sequence, triangles are C3 methylated sequence.

Methylation causes an increase in the barrier for the C3pG4 and C9pG10 steps. However, these changes in $\Delta G^{\neq}$ shown in Table 2 are deceptively low. This is because the activation barrier was defined in relation to the BI conformation. The C3pG4 and C9pG10 phosphate steps have a more stable BII state (as evidenced by their negative ΔG values). Thus, a better measure of the conformational locking that occurs upon methylation is an activation barrier defined as the difference between the lowest energy

state and the transition state. This definition clearly shows significant changes in the activation barriers for the C3pG4 and C9pG10 steps upon methylation. This is evident in Figure 3.14, which illustrates the change in the barrier between the most stable state and the transition state for each nucleotide step.

The increase in the barrier for conformational exchange upon methylation at C9pG10 is in agreement with the methylation induced dynamics quenching of the C9 step observed via ^{2}H-SSNMR.[10] However, since 2H-SSNMR can observe only one nucleus at a time, information about the increased conformational barrier of other steps local to the site of methylation has not been previously observed.

Another insight unobtainable from ^{2}H-SSNMR is the differences and changes in the ΔG values. Our technique provides information not only about the dynamics of the backbone but also about the conformation at each phosphate step. Most significantly, the negative ΔG values for the native C3pG4 and C9pG10 steps indicate that at these steps, BII is the most stable conformation, even before methylation. The C11pG12 step also has negative ΔG values; however, it is unlikely that this is sequence-specific but is, instead, a result of their position in the sequence; the ends of DNA sequences fray, resulting in a loss of base stacking and, thus, an increase in conformational freedom. Methylation of C9 results in significant decreases in the ΔG of C9pG10 and G10pC11, its 5' neighbor, and G4pA5, its base pairing partner, indicating an increase in the stability of

the BII conformation for these steps. However, methylation of C9 does not cause any significant changes in the ΔG of phosphate steps on the other (3') end of the sequence, suggesting that cytosine methylation causes only local structural perturbation of the sequence. Methylation of C3 also results in a drop in the ΔG of the C3pG4 step and increases the ΔG of its 3' neighbor G2pC3. However, there is no significant change in the ΔG of the phosphate steps at the other (5') end of the sequence.

3.3.6 DISCUSSION

Our results clearly indicate that the C3pG4 and C9pG10 steps of the Dickerson dodecamer are conformationally unique with respect to the other nucleotides in the sequence. They favor the BII conformation and have activation energies that are lower than other internal phosphate steps. The increased dynamics at these sites is expected in light of ^{2}H-SSNMR studies of this same sequence, which found that C3 and C9 experience unique, large-amplitude (38° half angle of libration) furanose dynamics and that C9 had large amplitude backbone dynamics as compared to other backbone steps previously studied (the backbone dynamics of C3 were not studied).[10] In addition to obtaining results consistent with ^{2}H-SSNMR studies of the native dodecamer, our methylation results are also consistent with a ^{2}H-SSNMR study of C9 methylation effects. This study showed increased T1z relaxation times for the C9pG10 backbone step upon

C9 methylation (0.03 and 0.055 ms, native and methylated, respectively) and a significant increase in the quadrupole coupling constant, QCC, indicative of a loss of largeamplitude dynamics.[10] This ^{2}H-SSNMR study investigated the effects of only C9 methylation (not C3 methylation) and, due to the quadrupole's being a single nucleus property, could not address the effects of this methylation on other steps in the sequence.

In addition to confirming that ^{31}P dynamic NMR detects increased dynamics of the phosphates at these nucleotide steps, our data uncover increased stability of the BII conformation for these steps. Our results also confirm that although these two nucleotide steps are uniquely dynamic, they are not the same. This difference between these steps is further confirmed by analysis of the temperature-induced conformational changes. The C3pG4 step shows a dramatic decrease (60.5-53.6%) in the percent BII as a function of temperature, whereas the C9pG10 step increases (61.6-67.5%). Interestingly, this response to temperature change is reversed in the base pairing partners of these steps. The G4pA5 step, the base pairing partner of C9, shows a dramatic increase in percent BII (28.6-41.0%), whereas the G10pC11 step, the base pairing partner of C3, shows little change in percent BII (47.6-46.4%). The low percent of BII of the G4pA5 step coincides with the stiffness of this nucleotide step observed via ^{2}H-SSNMR. It must be noted, however, that the G10 step was not studied via ^{2}H-SSNMR due to its position so near to the end of the sequence. Its location at the end of the sequence may

explain its relatively high percent of BII, even at low temperatures and, thus, its small changes in percent of BII with increasing temperature.

It may be very significant that at physiological temperature (37 °C), C3 is strongly BII but G4 is strongly BI (54.7 and 39.5%, respectively). This creates a step with a sharp switch in backbone conformation at this location, the step that is recognized by the protein, which cuts between G4 and A5. The only other step with such a significant switch in backbone conformation is the T8pC9 step (24.2 and 66.5%, respectively), across from the G4 cutting site. Enzymologists, recognizing that restriction endonucleases transiently bind to nonselective sites but do not cleave, have proposed that enzymes roll along DNA strands, dropping in and out of nonselective sites on the DNA. However, the enzyme binds more strongly when it comes to a high-affinity site, postulated to be a region of increased dynamics where the protein can distort the DNA to enhance necessary contacts for binding and cleavage.[51] This site is necessarily the result of local structure and dynamics in the DNA. Since the EcoRI binds to its palindromic DNA sequence as a dimer, it is not surprising to have discovered regions of enhanced dynamics at either end on both strands of this sequence.

It is important at this point to add two caveats. The first is that the accuracy of the kinetic and thermodynamic parameters we extract depends on the validity of equation 3.16. We have tested the sensitivity of our analysis to the exact form of this empirical

relationship by examining alternate paramerizations of percent BII versus chemical shift. In particular, the observed values of extrema for the chemical shifts do serve as limits of these values, but it is quite probable that the actual range is larger. By increasing the width of this range by 0.10 ppm on each edge, we find that the general trends remain intact, while the exact values vary by about 0.4 kcal/mol.

The second caveat is to emphasize that we used a two-site jump model, exchange between BI and BII, to model the dynamics of the phosphate backbones in this sequence and extract an activation energy. This is no doubt a very simplistic model for such a complex system and clearly has several shortcomings. That said, a two-site jump model, including the estimate of percent BII, provides a robust and convenient model for studying dynamics in this complicated biological system and provides a quantitative comparison of activation energies and dynamics. For example, the G2pC3 step in the C3 methylated sequence shows an increase of ~1.2 kcal/mol for the activation energy, corresponding to rates of exchange that are only 15% of those in the native sequence. Although we caution against the over interpretation of these numbers, our experiment and analysis do provide a consistent and efficient way to compare the dynamics of each backbone step in an oligonucleotide, affording important sequence-dependent information. This screening method is complementary to previously reported methods,

and once backbone steps with increased dynamics are identified, ^{2}H-SSNMR, for example, can be used if more complete modeling of the dynamics is desired.

3.4 CONCLUSION

We have presented a robust and efficient method of studying the local dynamics in DNA backbones. We have confirmed the presence of unique dynamics at the C3pG4 and C9pG10 steps of the Dickerson sequence. In addition, we have shown that these steps exist in the less common BII conformation, even at low temperatures, and that cytosine methylation at C3 and C9 increases the percent of BII at these and neighboring steps. Our goal is to provide sequence-dependent dynamical information about DNA sequences in an effort to uncover any trends in protein-DNA recognition. The method described herein is an efficient means of collecting this data, allowing us to look at a myriad of sequences under a variety of conditions.

REFERENCES

1 Dickerson, R. E.; Drew, H. R. *J. Mol. Biol.* **1981**, *149*, 761.

2 Drew, H. R.; Samson, S.; Dickerson, R. E. *Proc. Natl. Acad. Sci* **1982**, *79*, 4044.

3 von Hippel, P. H. *Science* **2004**, *305*, 350.

4 Hartmann, B.; Piazzola, D.; Lavery, R. *Nucleic Acids Research* **1993**, *21*, 561.

5 Djuranovic, D.; Hartmann, B. *Biopolymers* **2003**, *73*, 356.

6 Schroeder, S. A.; Roongta, V.; Fu, J. M.; Jones, C. R.; Gorenstein, D. G. *Biochemistry* **1992**, *31*, 8292.

7 Karslake, C.; Botuyan, M. V.; Gorenstein, D. G. *Biochemistry* **1993**, *32*, 6863.

8 Tisne', C.; Delepierre, M.; Hartmann, B. *Journal of Molecular Biology* **1999**, *293*, 139.

9 Banyay, M.; Graslund, A. *Journal of Molecular Biology* **2002**, *324*, 667.

10 Geahigan, K. M.; Meints, G. M.; Hatcher, M. E.; Drobny, G. P. *Biochem.* **2000**, *39*, 4939.

11 Bax, A.; Lerner, L. *J. Magn. Reson.* **1988**, *79*, 429.

12 Gorenstein, D. G. *Chemical Reviews* **1994**, *94*, 1315.

13 Isaacs, R. J.; Spielmann, H. P. *Journal of Molecular Biology* **2001**, *311*, 149.

14 Wu, Z.; Tjandra, N.; Bax, A. *Journal of the Ameican Chemical Society* **2001**, *123*, 3617.

15 Schwieters, C. D.; Clore, G. M. *Biochem.* **2007**, *46*, 1152.

16 Palmer, A. G.; Cavanaugh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *93*, 151.

17 Nevzorov, A. A.; Moltke, S.; Brown, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 4798.

18 Alam, T. M.; Drobny, G. P. *Chem. Rev* **1991**, *91*, 1545.

19 Hatcher, M. E.; Mattiello, D. L.; Meints, G. M.; Orban, J. O.; Drobny, G. P. *J. Am. Chem. Soc.* **1998**, *120*, 9850.

20 Meints, G. A.; Drobny, G. P. *Biochemistry* **2001**, *40*, 12436.

21 Allison, S. A.; Shibata, J.; Wilcoxon, J.; Schurr, J. M. *Biopolymers* **1982**, *21*, 721.

22 Allison, S. A.; Schurr, J. M. *Chem. Phys.* **1979**, *41*, 35.

23 Okonogi, T. M.; Alley, S. C.; Reese, A. W.; Hopkins, P. B.; Robinson, B. H. *Biophysical Journal* **2002**, *83*, 3446.

24 Withka, J. M.; Swaminathan, S.; Beveridge, D. L.; Bolton, P. H. *J. Amer. Chem. Soc.* **1991**, *113*, 5041.

25 Sen, S.; Nilsson, L. *Biophysical Journal* **1999**, *77*, 1782.

26 Holbrook, S. R.; Dickerson, R. E.; Kim, S.-H. *Acta Crystallogr.* **1985**, *B41*, 255.

27 Mauffret, O.; Hartmann, B.; Convert, O.; Lavery, R.; Fermandjian, S. *J. Mol. Biol.* **1992**, *227*, 852.

28 Dingley, A. J.; Peterson, R. D.; Grzesiak, S.; Feigon, J. *J. Am. Chem. Soc.* **2005**, *127*, 14466.

29 Gotfredsen, C. H.; Meissner, A.; Duus, J. O.; Sorensen, O. W. *Magnetic Resonance in Chemistry* **2000**, *38*, 692.

30 Sklenar, V.; Bax, A. *J. Amer. Chem. Soc.* **1987**, *109*, 2221.

31 Live, D.; Greene, K. *J. Magn. Reson.* **1989**, *85*, 604.

32 Heddi, B.; Foloppe, N.; Bouchemal, N.; Hantz, E.; Hartmann, B. *J. Amer. Chem. Soc.* **2006**, *128*, 9170.

33 Kintanar, A.; Huang, W.-C.; Schindele, D. C.; Wemmer, D. E.; Drobny, G. P. *Biochemistry* **1989**, *29*, 282.

34 Huang, W.-C.; Orban, J. O.; Drobny, G. P. *J. Am. Chem. Soc.* **1990**, *112*, 9059.

35 Ott, J.; Eckstein, F. *Biochemistry* **1985**, *24*, 2530.

36 Antri, S. E.; Mauffret, O.; Monnot, M.; Lescot, E.; Convert, O.; Fermandjian, S. *Journal of Molecular Biology* **1993**, *230*, 373.

37 Bain, A. D. *J. Phys. Chem.* **1994**, *98*, 7458.

38 Leskowitz, G. M.; Ghaderi, N.; Olsen, R. A.; Pederson, K.; Hatcher, M. E. M., L.J.,. *J. Phys. Chem.* **2004**, *109*, 1152.

39 Olsen, R.; Liu, L.; Ghaderi, N.; Johns, A.; Hatcher, M. E.; Mueller, L. *J. Amer. Chem. Soc.* **2003**, *125*, 10125.

40 Gutowsky, H. S.; Vold, R. L.; Wells, E. J. *J. Chem. Phys.* **1965**, *43*, 4107.

41 Inglefield, P. T. *Mol. Phys.* **1968**, *15*, 65.

42 Binsch, G. A. *J. Amer. Chem. Soc.* **1969**, *1969*, 1304.

43 Bain, A. D. *Prog. Nucl. Magn. Reson.* **2003**, *43*, 63.

44 Allerhan, A.; Gutowsky, H. S. *J. Chem. Phys.* **1965**, *42*, 1587.

45 Pechukas, P. *Annual Reviews of Physical Chemistry* **1981**, *32*, 159.

46 Wiberg, K. B. *J. Amer. Chem. Soc.* **1995**, *117*, 4261.

47 Wiberg, K. B.; Laidig, K. E. *J. Am. Chem. Soc.* **1987**, *109*, 5835.

48 Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663.

49 Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletsky, O.; Glaser, S. J.; Sorensen, O. W.; Griesinger, C. *Journal of Biomolecular NNR* **1994**, *4*, 301.

50 Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

51 Stryer, L. *Biochemistry*; Wh Freeman: New York, 1975.

52 Wu, Z. G.; Delaglio, F.; Tjandra, N.; Zhurkin, V. B.; Bax, A. *Journal of Biomolecular Nmr* **2003**, *26*, 297-315.

APPENDIX

A.1 PULSE CODES

CBCACO:

```
;rcp2.ctCBCACO.cpd

#include <Avance.incl>

"d21=30m"
"d0=3u"
"d10=3u"
"p27=p28-0.2"
"p25=p26-0.2"
"in11=in0"
"in12=in10"
"l3=(td1/2)"
"l4=(td2/2)"

define delay dtau1
"dtau1=((1s/cnst31)*l5 - 6u)"
define delay dtau3
"dtau3=((1s/cnst31)*l6 -3u)"
define delay dtau7
"dtau7=((1s/cnst31)*l8 )"
define delay dtau15
"dtau15=(dtau1-l3*in11)"
define delay dtau16
"dtau16=(dtau3-l4*in12)"
define delay dtauz
"dtauz=((1s/cnst31)*l9-p5-3u)"
"d11=dtau1"
"d12=dtau3"
define delay dtau11
"dtau11=((1s/cnst31)*(l11+0.5))"
"p11 = dtau11"
define delay dtau13
"dtau13=((1s/cnst31)*(l13+0.5))"
"p13 = dtau13"
```

```
1 ze
  dtau1
  dtau3
  dtau11
  dtau13
  dtau7
2 d1 do:f2
3 d21
4 d21
5 d21 do:f2
6 d21 do:f2 pl21:f2
  dtau15
  dtau16
  d0
  d11
  d10
  d12

; cp
  (p1 ph1):f2
  3u
  (p3:sp1 ph2):f1 (p3:sp3 ph3):f2
  3u cpds3:f2 pl26:f2 pl15:f1
; select CB
; (p13:sp13 ph22):f1
; 3u

; t1
  d0
  dtau1
  (p11:sp11 ph5):f1
  d11 pl15:f1
  3u
  3u
  (p5 ph6):f1
  3u
  (p5 ph7):f1
; t1 rephase t2 evolve
```

```
  d10
  dtau3
  (p5 ph8):f1
  (p5 ph8):f1
  d12
  3u
  (p5 ph17):f1
  3u
  (p5 ph18):f1

; t2 repahse
  dtau7
  (p5 ph20):f1
  (p5 ph20):f1
  dtau7 pl15:f1

;    zf
  (p5 ph25):f1
  3u cpds2:f2 pl28:f2
  dtauz ;
  (p5 ph26):f1

;    t3
  go=2 ph31 ;cpds2:f2 pl28:f2
  3m do:f2
  d1 wr #0 if #0 ip2 zd
  d21
  lo to 5 times 2
  d21 id0
  d21 dd11
  lo to 6 times l3
  d21 rd0
  d21 rd11
  d21 rp2
  d21 ip17
  d21 ip18
  d21 ip20
  d21 ip25
  lo to 3 times 2
```

```
  d21 id10
  d21 dd12
  lo to 4 times l4
exit

;dq/zq fileter
;ph1=1 3
;ph3=0 2
;ph2=0 0
;ph4=1 3
;ph5=0 0 1 1 2 2 3 3
;ph6=1 1
;ph7=0 2 ;0 0 2 2 2 2
;ph8=0 2 ;0 0 2 2 2 2
;ph11=0 2
;ph10=1 3
;ph15=0 2
;ph16=1 3
;ph17=0 2 ;0 0 2 2 2 2
;ph18=1 3
;ph20=0 2
;ph25=1 3
;ph26=3 1
;ph31=0 2 2 0

ph1=1
ph3=0 0
ph2=0 2
ph4=1 1
ph5=0 2 1 3 2 0 3 1
ph6=1 3
ph7=0 0 0 0
ph8=0 0 1 1 2 2 3 3
ph11=0 0
ph10=1 1
ph15=0 0
ph16=1 1
ph17=0 0 0 0
ph18=1 1
```

```
ph20=0 0 1 1 2 2 3 3
ph25=1 1
ph26=3 3
ph31=0 0 2 2 0 0 2 2

;pl21 : (1H) f2 chanrsnel 90 pulse
;pl15 : (X) f1 Pi/2 mixing pulses
;pl26: (1H) indirect decoupling
;pl28: (1H) observe decoupling
;p1 : f2 channel -    90 degree high power pulse
;p3: f1(f2) channel - contact pulse
;p5: (X) f1 channel Pi/2 mixing pulses
;p26:    1H TPPM pulse about a 160 degree pulse at pl26
;p28:    1H TPPM pulse about a 160 degree pulse at pl28
;cpdprg2:    1H direct cpd program
;cpdprg3:    1H indirectcpd program
;sp1: 13C CP peak power level
;spoffs1:    0
;spnam1:    name of the shape file
;sp3: 1H CP peak power level
;spoffs1:    0
;spnam1:    name of the shape file
;sp11:    selective active pi
;spoffs11:    CAli
;spnam11:    name of the shape file
;sp13:    selective CO select pi/2
;spoffs13:    CO
;spnam13:    name of the shape file
;d1 : relaxation delay; 1-5 * T1
;d21 : delay for disk I/O                                    [30 msec]
;cnst31 : MAS spinning speed
;NS : 1 * 4
;DS : none
;d0 : incremented delay (2D)                                 [3 usec]
;l3 : loop for phase sensitive 2D using States-TPPI method: l3 = td1/2
;l5 : rotor periods tau1
;l6 : rotor periods tau3
;l8 : rotor periods tau7
;l9 : number of rotor periods for z-filter
```

;l11 : selective pi pulse active
;l13 : selective pi/2 pulse
;in0 : 1/(1 * SW) = 2 * DW
;nd0 : 1
;td1 : number of experiments
;MC2 : States-TPPI

CACBCG:

;rcp2.ctNCC.cpd.1d

#include <Avance.incl>

"d21=30m"
"d0=3u"
"d10=3u"
"p27=p28-0.2"
"p25=p26-0.2"
"l3=(td1/2)"
"l4=(td2/2)"
define delay dtau1
"dtau1=((1s/cnst31)*l5)"
define delay dtau2
"dtau2=(dtau1-p13 - 6u)"
define delay dtau3
"dtau3=((1s/cnst31)*l6)"
define delay dtau4
"dtau4=(dtau3-p13-6u)"
define delay dtau7
"dtau7=((1s/cnst31)*l8)"
define delay dtauz
"dtauz=((1s/cnst31)*l9-p5-3u)"
"d11=dtau2"
"d12=dtau4"
define delay dtau15
"dtau15=(dtau2-l3*in11)"

```
define delay dtau16
"dtau16=(dtau4-l4*in12)"
define delay dtau11
"dtau11=((1s/cnst31)*(l11+0.5))"
"p11 = dtau11"
define delay dtau13
"dtau13=((1s/cnst31)*(l13+0.5))"
"p13 = dtau13"

1 ze
  dtau1
  dtau3
  dtau11
  dtau13
  dtau7
2 d1 do:f2
3 d21
4 d21
5 d21 do:f2
6 d21 do:f2 pl21:f2
  dtau15
  dtau16
  d0
  d11
  d10
  d12

; cp
  (p1 ph1):f2
  3u
  (p3:sp1 ph2):f1 (p3:sp3 ph3):f2
  3u cpds3:f2 pl26:f2 pl15:f1

; t1
  d0
  (p13:sp13 ph22):f1
  dtau2 pl15:f1
  (p11:sp11 ph5):f1
  d11
```

```
  (p13:sp13 ph22):f1
  6u pl15:f1
  (p5 ph6):f1
  3u
  (p5 ph7):f1

; t1 rephase t2 evolve
  3u
  d10
  (p13:sp13 ph22):f1
  dtau4
  (p11:sp11 ph8):f1
  d12
  (p13:sp13 ph22):f1
  6u pl15:f1
  (p5 ph17):f1
  3u
  (p5 ph18):f1

; t2 repahse
  dtau7
  (p11:sp11 ph20):f1
  dtau7 pl15:f1

;    zf
  (p5 ph25):f1
  3u cpds2:f2 pl28:f2
    dtauz ;
  (p5 ph26):f1

;    t3
  go=2 ph31 ;cpds2:f2 pl28:f2
  3m do:f2
  d1 wr #0 if #0 ip2 zd
  d21
  lo to 5 times 2
  d21 id0
  dtau4
  d21 dd11
```

```
  lo to 6 times l3
  d21 rd0
  d21 rd11
  d21 rp2
  d21 ip17
  d21 ip18
  d21 ip20
  d21 ip25
  lo to 3 times 2
  d21 id10
  d21 dd12
  lo to 4 times l4
exit

;dq/zq fileter - Lingling's
;ph1=1 1 3 3
;ph3=0 0 2 2
;ph2=0 2 2 0
;ph22 = 1 1 3 3
;ph4=1 1 3 3
;ph5=0 2 2 0
;ph6=1 3 3 1
;ph7=0 0 0 0
;ph8=0 0 0 0
;ph11=0 0 2 2
;ph10=1 1 3 3
;ph15=0 0 2 2
;ph16=1 1 3 3
;ph17=0 0 0 0
;ph18=1 1 3 3
;ph20=0 0 2 2
;ph25=1 1 3 3
;ph26=3 3 1 1
;ph31=0 0 2 2

ph1=1 1
ph3=0 0
ph2=0 2
ph22=1
```

```
ph4=1 1
ph5=0 2 1 3 2 0 3 1
ph6=1 3
ph7=0 0 0 0 ;2 2 2 2
ph8=0 0 1 1 2 2 3 3
ph11=0 0
ph10=1 1
ph15=0 0
ph16=1 1
ph17=0 0 0 0 ;2 2 2 2
ph18=1 1
ph20=0 0 1 1 2 2 3 3
ph25=1 1
ph26=3 3
ph31=0 0 2 2

;pl21 : (1H) f2 chanrsnel 90 pulse
;pl15 : (X) f1 Pi/2 mixing pulses
;pl26: (1H) indirect decoupling
;pl28: (1H) observe decoupling
;p1 : f2 channel -    90 degree high power pulse
;p3: f1(f2) channel - contact pulse
;p5: (X) f1 channel Pi/2 mixing pulses
;p26:    1H TPPM pulse about a 160 degree pulse at pl26
;p28:    1H TPPM pulse about a 160 degree pulse at pl28
;cpdprg2:    1H direct cpd program
;cpdprg3:    1H indirectcpd program
;sp1: 13C CP peak power level
;spoffs1:    0
;spnam1:    name of the shape file
;sp3: 1H CP peak power level
;spoffs1:    0
;spnam1:    name of the shape file
;sp11:    selective active pi
;spoffs11:    CAli
;spnam11:    name of the shape file
;sp13:    selective CO select pi/2
;spoffs13:    CO
;spnam13:    name of the shape file
```

;d1 : relaxation delay; 1-5 * T1
;d21 : delay for disk I/O [30 msec]
;cnst31 : MAS spinning speed
;NS : 1 * 4
;DS : none
;d0 : incremented delay (2D) [3 usec]
;l3 : loop for phase sensitive 2D using States-TPPI method: l3 = td1/2
;l5 : rotor periods tau1
;l6 : rotor periods tau3
;l8 : rotor periods tau7
;l9 : number of rotor periods for z-filter
;l11 : selective pi pulse active
;l13 : selective pi/2 pulse
;in0 : 1/(1 * SW) = 2 * DW
;nd0 : 1
;td1 : number of experiments
;MC2 : States-TPPI

A.2 CSROSETTA INPUT FILES

gb1.fasta:

>
MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFT
VTE

gb1.tab (for J-gb1):

DATA SEQUENCE MQYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD
GEWTYDDATK
DATA SEQUENCE TFTVTE

VARS RESID RESNAME ATOMNAME SHIFT
FORMAT %4d %1s %4s %8.2f

1 M CA 54.26
1 M CB 32.22
1 M C 171.21
2 Q N 125.11
2 Q CA 55.88
2 Q CB 30.16
2 Q C 174.86
3 Y N 122.88
3 Y CA 57.15
3 Y C 174.89
4 K N 122.53
4 K CA 54.83
4 K CB 36.17
4 K C 173.16
5 L N 126.25
5 L CA 53.01
5 L CB 42.47
5 L C 174.77
6 I N 125.95
6 I CA 59.98
6 I CB 37.86
6 I C 175.08

7	L	N	126.52
7	L	CA	54.84
7	L	CB	43.20
7	L	C	174.94
8	N	N	123.98
8	N	CA	50.77
8	N	CB	38.47
8	N	C	175.99
9	G	N	109.07
9	G	CA	44.75
9	G	C	172.88
10	K	N	120.60
10	K	CA	59.29
10	K	CB	32.77
10	K	C	178.73
11	T	N	105.75
11	T	CA	62.09
11	T	CB	69.50
11	T	C	172.99
12	L	N	127.15
12	L	CA	54.58
12	L	CB	43.08
12	L	C	173.70
13	K	N	122.89
13	K	CA	53.47
13	K	CB	39.12
13	K	C	175.60
14	G	N	105.12
14	G	CA	45.09
14	G	C	170.91
15	E	N	120.56
15	E	CA	53.97
15	E	CB	34.35
15	E	C	173.85
16	T	N	114.54
16	T	CA	60.28
16	T	CB	70.84
16	T	C	172.05
17	T	N	115.98

17	T	CA	60.36
17	T	CB	72.90
17	T	C	174.12
18	T	N	115.76
18	T	CA	61.47
18	T	CB	70.96
18	T	C	171.03
19	E	N	124.87
19	E	CA	54.50
19	E	CB	30.95
19	E	C	175.86
20	A	N	124.89
20	A	CA	50.88
20	A	CB	23.56
20	A	C	177.40
21	V	N	115.07
21	V	CA	63.73
21	V	CB	31.92
21	V	C	174.84
22	D	N	114.74
22	D	CA	52.58
22	D	CB	42.37
22	D	C	175.14
23	A	N	122.66
23	A	CA	54.58
23	A	CB	18.46
23	A	C	179.79
24	A	N	120.07
24	A	CA	54.48
24	A	CB	18.17
24	A	C	181.18
25	T	N	116.94
25	T	CA	67.66
25	T	CB	67.75
25	T	C	175.51
26	A	N	123.34
26	A	CA	55.18
26	A	CB	17.49
26	A	C	177.18

27	E	N	115.89
27	E	CA	59.27
27	E	CB	29.01
27	E	C	177.68
28	K	N	116.79
28	K	CA	60.40
28	K	CB	32.75
28	K	C	178.50
29	V	N	117.88
29	V	CA	66.47
29	V	CB	31.80
29	V	C	178.10
30	F	N	117.70
30	F	CA	57.49
30	F	C	178.73
31	K	N	120.06
31	K	CA	60.33
31	K	CB	31.80
31	K	C	179.56
32	Q	N	120.65
32	Q	CA	58.96
32	Q	CB	29.13
32	Q	C	177.36
33	Y	N	120.30
33	Y	CA	61.71
33	Y	C	178.30
34	A	N	121.89
34	A	CA	56.15
34	A	CB	17.96
34	A	C	179.64
35	N	N	117.77
35	N	CA	57.15
35	N	CB	39.56
35	N	C	179.34
36	D	N	120.71
36	D	CA	56.06
36	D	CB	38.47
36	D	C	175.69
37	N	N	114.32

37	N	CA	53.71
37	N	CB	40.65
37	N	C	174.02
38	G	N	107.54
38	G	CA	46.88
38	G	C	173.80
39	V	N	121.26
39	V	CA	61.91
39	V	CB	31.80
39	V	C	174.89
40	D	N	130.47
40	D	CA	52.85
40	D	CB	41.45
40	D	C	175.21
41	G	N	107.83
41	G	CA	45.26
41	G	C	172.69
42	E	N	119.03
42	E	CA	55.37
42	E	CB	31.56
42	E	C	177.66
43	W	N	124.73
43	W	CA	57.62
43	W	C	176.76
44	T	N	108.13
44	T	CA	61.00
44	T	CB	72.78
44	T	C	173.69
45	Y	N	118.11
45	Y	CA	57.88
45	Y	C	171.85
46	D	N	125.87
46	D	CA	51.11
46	D	CB	42.95
46	D	C	175.91
47	D	N	123.08
47	D	CA	54.45
47	D	CB	43.13
47	D	C	177.11

48	A	N	118.16
48	A	CA	53.96
48	A	CB	19.14
48	A	C	179.00
49	T	N	103.28
49	T	CA	60.61
49	T	CB	69.87
49	T	C	175.33
50	K	N	118.32
50	K	CA	55.50
50	K	CB	27.71
50	K	C	175.18
51	T	N	111.81
51	T	CA	62.56
51	T	CB	71.93
51	T	C	174.06
52	F	N	129.42
52	F	CA	56.61
52	F	C	175.59
53	T	N	111.32
53	T	CA	60.38
53	T	CB	72.02
53	T	C	171.92
54	V	N	117.36
54	V	CA	58.73
54	V	CB	32.59
54	V	C	172.53
55	T	N	123.47
55	T	CA	61.35
55	T	CB	72.29
55	T	C	174.20
56	E	N	130.31
56	E	CA	57.64
56	E	CB	33.51
56	E	C	180.16

gb1.tab (for CS-2JSV, Chemical shift data downloaded from Biological Magnetic Resonance Data Bank, BMRB ID: 15156):

DATA SEQUENCE MQYKLILNGK TLKGETTTEA VDAATAEKVF KQYANDNGVD GEWTYDDATK
DATA SEQUENCE TFTVTE

VARS RESID RESNAME ATOMNAME SHIFT
FORMAT %4d %1s %4s %8.2f

1 M C 171.36
1 M CA 54.33
1 M CB 32.50
2 Q C 175.08
2 Q CA 55.88
2 Q CB 30.45
2 Q N 125.24
3 Y C 174.98
3 Y CA 57.01
3 Y N 123.34
4 K C 173.31
4 K CA 54.88
4 K CB 36.25
4 K N 122.67
5 L C 174.80
5 L CA 52.97
5 L CB 42.54
5 L N 126.98
6 I C 175.31
6 I CA 59.97
6 I CB 37.85
6 I N 126.26
7 L C 175.12
7 L CA 54.66
7 L CB 42.90
7 L N 127.08
8 N C 176.43

8	N	CA	50.72
8	N	CB	38.37
8	N	N	125.11
9	G	C	173.22
9	G	CA	44.58
9	G	N	109.62
10	K	C	179.17
10	K	CA	59.27
10	K	CB	32.85
10	K	N	121.07
11	T	C	173.42
11	T	CA	61.91
11	T	CB	69.49
11	T	N	106.40
12	L	C	173.85
12	L	CA	54.43
12	L	CB	43.07
12	L	N	127.76
13	K	C	175.83
13	K	CA	53.34
13	K	CB	38.83
13	K	N	123.25
14	G	C	171.37
14	G	CA	44.89
14	G	N	105.61
15	E	C	174.11
15	E	CA	53.89
15	E	CB	34.16
15	E	N	121.07
16	T	C	172.04
16	T	CA	60.12
16	T	CB	70.51
16	T	N	115.18
17	T	C	174.25
17	T	CA	60.32
17	T	CB	72.64
17	T	N	116.05
18	T	C	171.27
18	T	CA	61.31

18	T	CB	70.84
18	T	N	116.26
19	E	C	175.90
19	E	CA	54.33
19	E	CB	30.56
19	E	N	125.35
20	A	C	177.79
20	A	CA	50.69
20	A	CB	23.67
20	A	N	125.88
21	V	C	175.10
21	V	CA	63.47
21	V	CB	31.96
21	V	N	116.28
22	D	C	175.08
22	D	CA	52.53
22	D	CB	42.30
22	D	N	115.52
23	A	C	179.73
23	A	CA	54.55
23	A	CB	18.20
23	A	N	122.82
24	A	C	181.53
24	A	CA	54.54
24	A	CB	18.20
24	A	N	120.75
25	T	C	175.92
25	T	CA	67.23
25	T	CB	67.82
25	T	N	117.37
26	A	C	177.28
26	A	CA	55.02
26	A	CB	17.55
26	A	N	123.99
27	E	C	177.82
27	E	CA	59.09
27	E	CB	29.06
27	E	N	116.35
28	K	C	179.24

28	K	CA	60.19
28	K	CB	32.79
28	K	N	117.38
29	V	C	178.98
29	V	CA	66.31
29	V	CB	31.92
29	V	N	119.34
30	F	C	179.09
30	F	CA	57.54
30	F	N	118.68
31	K	C	179.57
31	K	CA	60.11
31	K	CB	31.61
31	K	N	120.76
32	Q	C	177.58
32	Q	CA	58.92
32	Q	CB	28.97
32	Q	N	121.29
33	Y	C	178.82
33	Y	CA	61.59
33	Y	N	120.99
34	A	C	179.62
34	A	CA	56.07
34	A	CB	18.10
34	A	N	122.68
35	N	C	179.56
35	N	CA	57.04
35	N	CB	39.25
35	N	N	118.20
36	D	C	176.18
36	D	CA	55.91
36	D	CB	38.31
36	D	N	121.08
37	N	C	174.28
37	N	CA	53.51
37	N	CB	40.35
37	N	N	115.04
38	G	C	174.03
38	G	CA	46.80

38	G	N	108.35
39	V	C	175.20
39	V	CA	61.72
39	V	CB	31.94
39	V	N	121.79
40	D	C	174.89
40	D	CA	52.77
40	D	CB	41.71
40	D	N	131.05
41	G	C	172.81
41	G	CA	45.10
41	G	N	108.10
42	E	C	177.94
42	E	CA	55.10
42	E	CB	31.48
42	E	N	119.02
43	W	C	177.42
43	W	CA	57.46
43	W	N	124.95
44	T	C	174.03
44	T	CA	60.94
44	T	CB	73.09
44	T	N	109.21
45	Y	C	171.91
45	Y	CA	57.84
45	Y	N	118.61
46	D	C	176.30
46	D	CA	50.89
46	D	CB	42.60
46	D	N	126.34
47	D	C	177.51
47	D	CA	54.64
47	D	CB	43.00
47	D	N	123.39
48	A	C	179.70
48	A	CA	53.99
48	A	CB	19.00
48	A	N	118.98
49	T	C	175.84

49 T	CA	60.27
49 T	CB	69.90
49 T	N	104.23
50 K	C	175.47
50 K	CA	55.59
50 K	CB	28.03
50 K	N	119.71
51 T	C	174.44
51 T	CA	62.47
51 T	CB	71.71
51 T	N	112.00
52 F	C	175.81
52 F	CA	56.58
52 F	N	130.28
53 T	C	172.25
53 T	CA	60.39
53 T	CB	71.89
53 T	N	112.21
54 V	C	172.62
54 V	CA	58.48
54 V	CB	32.72
54 V	N	118.43
55 T	C	174.17
55 T	CA	61.32
55 T	CB	72.09
55 T	N	124.10
56 E	C	180.50
56 E	CA	57.56
56 E	CB	33.31
56 E	N	131.08

A.3 MATHEMATICA CODES FOR DYNAMICS ANALYSIS

DNA Dynamics
--Native Dickerson

----------Constants----------
h=6.626*10^{-34};
R=1.9872;
Kb=1.38*10^{-23};
γP=17.235 ;
γH=42.576;

----------ReadinData----------
LineWidth1={12.9706,11.9637,11.1956,11.2911,10.9477,10.7186,10.4366,9.80456,9.85
43,9.48347,9.36983,9.57841,9.22227,8.83011,8.89549};
CS1={-0.35927,-0.35447,-0.34839,-0.34836,-0.34537,-0.34267,-0.33943,-0.33726,-0.33
402,-0.33321,-0.33063,-0.32842,-0.32621,-0.32401,-0.3218,-0.31934,-0.31639};

LineWidth2={17.6008,14.275,15.9148,14.8064,12.86,13.0057,12.8481,12.2301,11.5997,
11.3435,11.5583,11.5371,11.5723,10.034,9.66134,9.5484};
CS2={-0.44544,-0.43932,-0.43835,-0.43902,-0.44056,-0.44065,-0.44347,-0.44525,-0.44
506,-0.44767,-0.44843,-0.45144,-0.45403,-0.45505,-0.45885,-0.46047,-0.46144};

LineWidth3={15.9691,14.5951,12.9322,13.827,13.862,12.2344,13.0203,11.7761,12.162
6,11.6461,11.2141,10.9629,10.7704,10.4017,10.1842,10.3498};
CS3={-0.21122,-0.21621,-0.22208,-0.23131,-0.23775,-0.24262,-0.25342,-0.25624,-0.26
103,-0.26672,-0.27164,-0.27544,-0.28121,-0.286,-0.28882,-0.29209,-0.29539};

LineWidth4={16.7026,16.4248,13.7616,14.537,13.1138,13.5511,11.7658,11.5567,11.37
1,10.1653,10.6683,9.85077,9.7781,10.1089};
CS4={-0.49826,-0.48498,-0.46984,-0.46479,-0.45225,-0.43798,-0.43247,-0.42327,-0.41
805,-0.41168,-0.40663,-0.40245,-0.39925,-0.39601,-0.39385,-0.38793,-0.38844};

LineWidth5={16.9359,16.6742,15.3565,14.5397,14.877,12.3818,11.8697,12.4611,12.08
09,10.2491,9.83151,11.1307,11.0298,10.8665,9.29527,9.25894};
CS5={-0.58773,-0.58947,-0.57816,-0.57373,-0.56913,-0.55968,-0.5545,-0.5483,-0.5421,
-0.53409,-0.53167,-0.52946,-0.52528,-0.52105,-0.51686,-0.51379,-0.50845};

LineWidth6={15.954,14.8187,13.3462,13.5536,11.9505,11.049,11.0148,11.8778,10.878
2,11.7001,10.7883,10.2671,11.3053};
CS6={-0.70321,-0.69676,-0.67578,-0.67666,-0.67293,-0.6857,-0.67649,-0.6733,-0.6691
1,-0.6632,-0.66568,-0.65849,-0.65929,-0.65609,-0.65689,-0.65307,-0.65049};

LineWidth7={19.8177,19.9675,18.4746,18.6885,17.417,15.5683,15.5266,12.5739,13.30
54,12.0333,10.6768,12.3574,12.9404,12.1248,11.2231,12.0371};
CS7={-0.669,-0.67824,-0.64852,-0.65714,-0.66873,-0.65972,-0.6655,-0.6593,-0.66108,-
0.66402,-0.66169,-0.65948,-0.66229,-0.6591,-0.65886,-0.65867,-0.65449};

LineWidth8={24.9549,23.7031,23.0319,20.3722,18.7898,15.7217,13.9798,12.2355,12.9
924,11.9673,12.0857,11.8068,12.889,10.8501,11.4799,10.3936};
CS8={-0.57241,-0.56996,-0.55297,-0.55504,-0.55172,-0.54568,-0.5445,-0.53929,-0.533
08,-0.52717,-0.52467,-0.52147,-0.51828,-0.51504,-0.51184,-0.50963,-0.50944};

LineWidth9={15.5523,15.8495,14.6156,14.7226,12.9729,13.7782,12.4053,11.7441,11.8
23,11.2245,10.4501,10.9346,10.131,10.1476,9.9828,9.99233};
CS9={-0.20126,-0.19843,-0.18939,-0.19113,-0.19007,-0.18799,-0.18742,-0.1862,-0.187,
-0.18714,-0.18662,-0.18741,-0.1892,-0.18901,-0.18939,-0.19043,-0.19238};

LineWidth10={13.8903,12.3703,12.0986,11.2362,10.8744,10.9583,10.8761,9.78233,10.
3122,9.67689,9.51883,9.75628,9.52859,9.82958,8.911,10.0865};
CS10={-0.32716,-0.32841,-0.32426,-0.32745,-0.32676,-0.33147,-0.32946,-0.32923,-0.3
3003,-0.33017,-0.33162,-0.33345,-0.33523,-0.34101,-0.34382,-0.34553,-0.34842};

LineWidth11={12.3273,10.6388,11.2707,10.6534,10.1545,9.75489,9.7213,9.51703,9.09
216,8.83138,8.41143,8.66324,8.21657,8.03709,8.26345,8.13084};
CS11={-0.16358,-0.16722,-0.16761,-0.17437,-0.17731,-0.17942,-0.18541,-0.18723,-0.1
9001,-0.19554,-0.19959,-0.20141,-0.20521,-0.21,-0.21278,-0.21621,-0.21939};

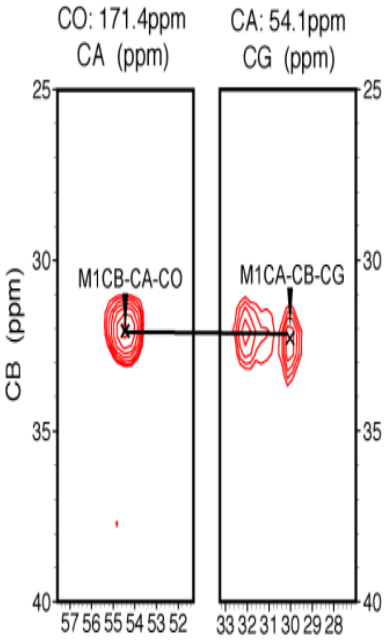
LineWidth=Join[{LineWidth1},{LineWidth2},{LineWidth3},{LineWidth4},{LineWidth
5},{LineWidth6},{LineWidth7},{LineWidth8},{LineWidth9},{LineWidth10},{LineWid
th11}];
CS=Join[{CS1},{CS2},{CS3},{CS4},{CS5},{CS6},{CS7},{CS8},{CS9},{CS10},{CS1
1}];
CS25={-0.33933,-0.44632,-0.24916,-0.43682,-0.55371,-0.67497,-0.66173,-0.54103,-0.1
8929,-0.33197,-0.18586};
StartTemp={16,14,14,18,14,20,14,14,14,14,14};

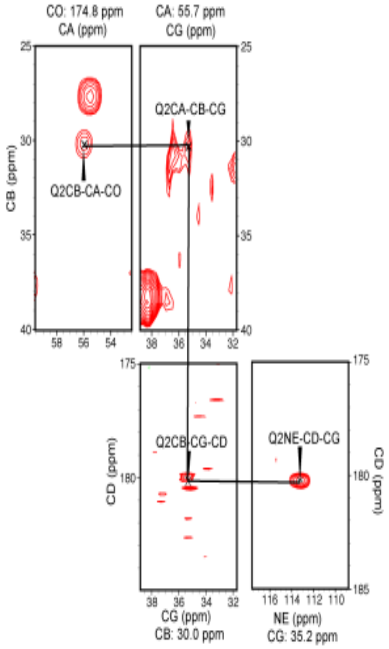
----------Calculate K---- --------
K1={0.826096119,0.826213172,0.837955209,0.848688882,0.861735946,0.87057772,0.883936689,0.887306299,0.898119931,0.907481792,0.916936459,0.926441898,0.936085931,0.946935115,0.960106732};
K2={0.529041701,0.531699674,0.529862768,0.525657284,0.525412224,0.517773354,0.512990918,0.513499966,0.50653709,0.504521614,0.496591961,0.489835375,0.487191194,0.4774224,0.473296718,0.470837412};
K3={1.544751361,1.501040433,1.435266358,1.391386227,1.359239513,1.290943397,1.273756664,1.245147171,1.212083995,1.184270283,1.163262269,1.132124788,1.106948717,1.092403029,1.07578569,1.059281356};
K4={0.462406715,0.494472321,0.532715976,0.548011819,0.574243116,0.589525662,0.608581849,0.624017003,0.63701889,0.64711405,0.657463051,0.664434937,0.683847241,0.682157091};
K5={0.20528411,0.224818947,0.232644215,0.240876302,0.258137634,0.267804733,0.27957252,0.291560812,0.307385565,0.312243137,0.316710825,0.325244746,0.333994105,0.342775372,0.349283119,0.360754348};
K6={0.078366977,0.061268067,0.073545027,0.077863803,0.083589503,0.091769803,0.088322129,0.098378102,0.097250039,0.101776251,0.100641201,0.106082247,0.109787628};
K7={0.071190461,0.112633689,0.100286977,0.084111789,0.096644662,0.088571631,0.097235953,0.094734371,0.090627431,0.09387971,0.096982465,0.093040361,0.097517744,0.097856085,0.098124086,0.104053388};
K8={0.239382827,0.270688557,0.266790035,0.273054378,0.284611282,0.286893632,0.297068508,0.309408519,0.321372446,0.326499385,0.333120212,0.339786459,0.346625787,0.353449564,0.35820275,0.358612954};
K9={1.686993955,1.765591506,1.750107873,1.759519722,1.778176755,1.783333621,1.79443572,1.787145684,1.785873837,1.790603702,1.783424263,1.767292813,1.768996214,1.765591506,1.756316029,1.739091159};
K10={0.907524363,0.92535699,0.911620069,0.914574738,0.894585634,0.903064734,0.90403982,0.900652513,0.900060973,0.893955894,0.88630664,0.878925428,0.855350562,0.844101843,0.83732305,0.825979081};
K11={1.979316986,1.975272215,1.906867737,1.878089593,1.857784634,1.80167215,1.785056833,1.760054395,1.711630563,1.677230662,1.662054568,1.630916372,1.592688618,1.571007367,1.544751361,1.52088363};
K={};
K=Join[{K1},{K2},{K3},{K4},{K5},{K6},{K7},{K8},{K9},{K10},{K11}];

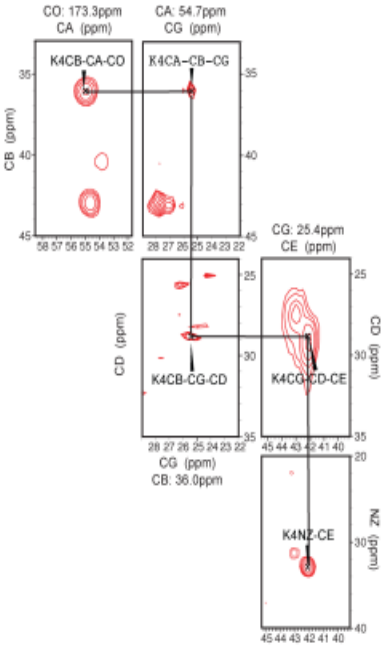
----------Fitting Data---- --------

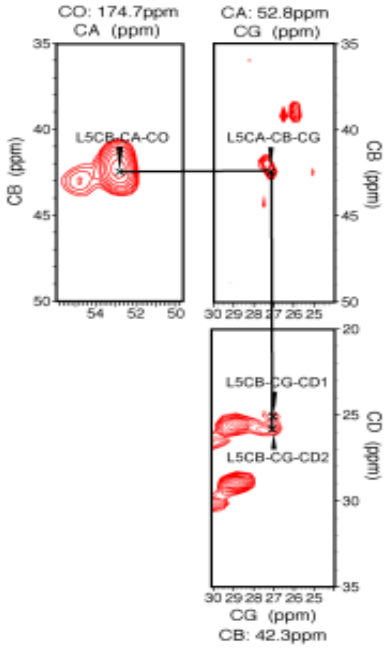
```
dBII=0.11922;
dBI=-0.73634;
δ=π (dBII-dBI) (600 γP/γH);
r=1/(3.14 T2)+(4y δ2)/(1+y)^3 h/(π Kb x) e^ΔΔG/(R x);
<<NonlinearRegression`
Results={};
Do[{
   datalist={};
   length=Length[LineWidth[[j]]];
   STemp=StartTemp[[j]];
   Temp=Table[273.15+STemp+2*i,{i,0,length-1}];
   datalist=Transpose[Join[{Temp},{K[[j]]},{LineWidth[[j]]}]];
   Fitting = FindFit[datalist,r,{{T2,0.03},{ΔΔG,12500} }, {x,y}];
   Results=Join[Results,{Fitting}];
   },{j,1,11}]
Results
ddG={};
Do[{
     temp=Join[{i},{ΔΔG/1000/.Results[[i,2]]}];
     ddG=Join[ddG,{temp}];
     },{i,1,11}];
a1=ListPlot[ddG,AxesLabel→{Steps,"ΔG^135(kcal/mol)"},PlotStyle→Directive[Red,PointS
ize[Large]],PlotRange→All];
a2=ListPlot[ddG,Joined→True,PlotRange→All];
Show[a1,a2]
MaxDistance=Table[ddG[[i,2]]-(ΔG[[i]]-Abs[ΔG[[i]]])/2,{i,1,11}]
MaxD={};
Do[{
   t=Join[{i},{MaxDistance[[i]]}];
   MaxD=Join[MaxD,{t}]
   },{i,1,11}]
d1=ListPlot[MaxD,AxesLabel→{Steps,"Max
Distance(kcal/mol)"},PlotStyle→Directive[Red,PointSize[Large]],PlotRange→All];
d2=ListPlot[MaxD,Joined→True,PlotRange→All];
Show[d1,d2]
Table[ddG[[i,2]],{i,1,11}]
```

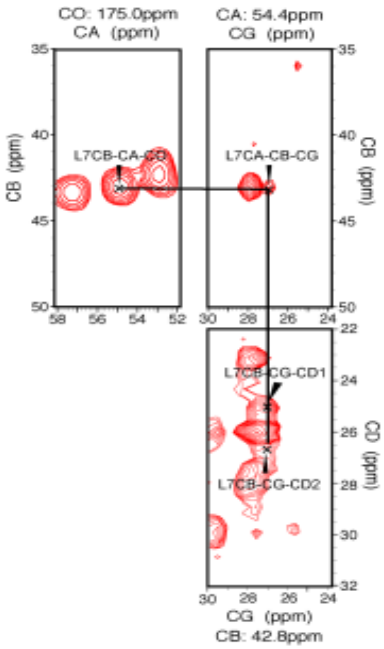
A.4 GB1 SIDECHAIN WALK FIGURES

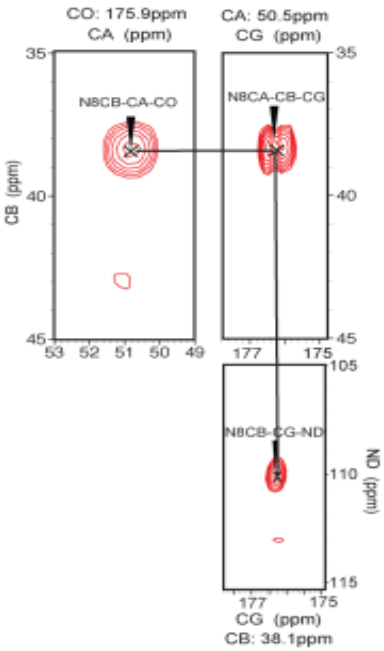


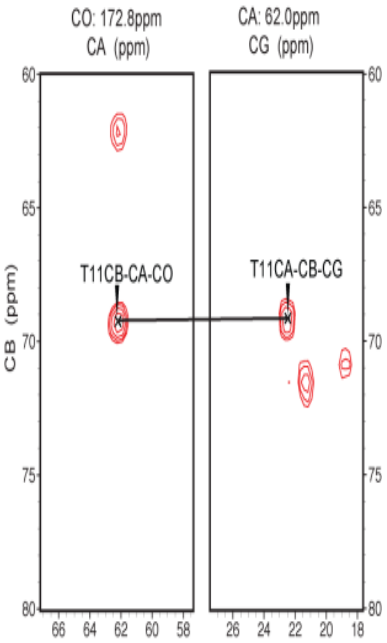


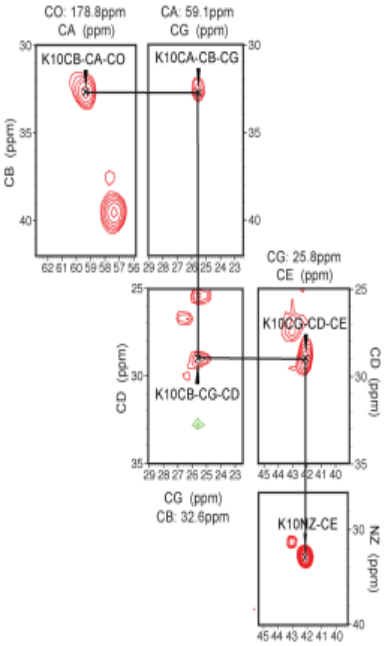


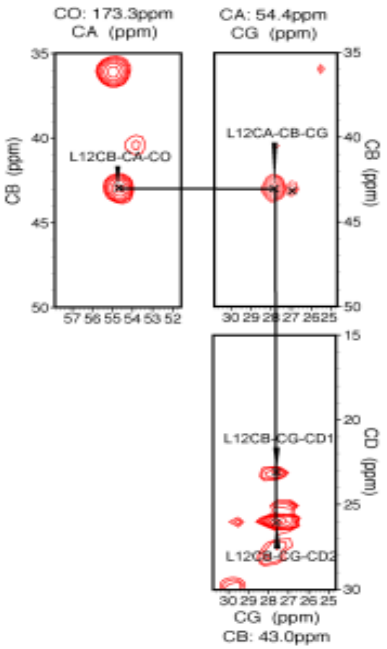


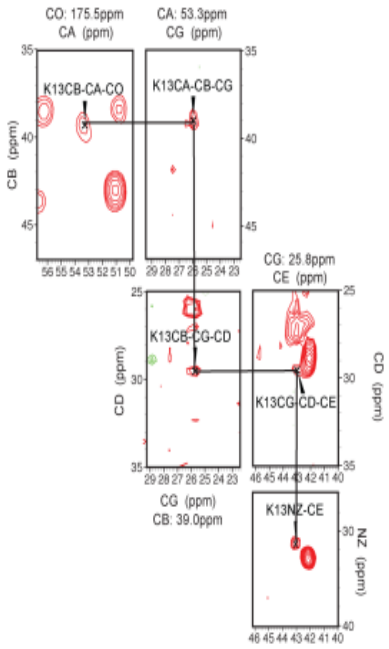

CO: 175.0ppm
CA (ppm)
CA: 54.4ppm
CG (ppm)
L7CB-CA-CO
L7CA-CB-CG
L7CB-CG-CD1
L7CB-CG-CD2
CB (ppm)
CD (ppm)
CG (ppm)
CB: 42.8ppm

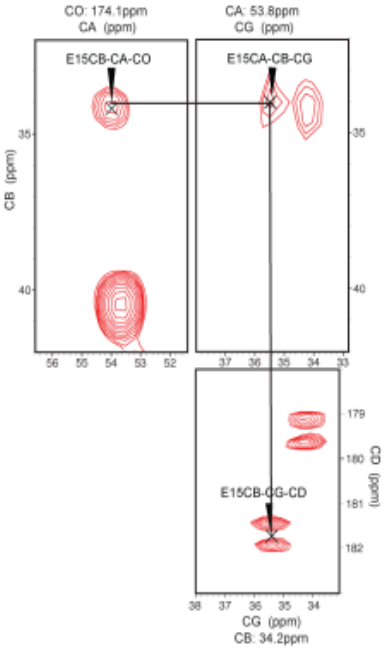

CO: 175.9ppm
CA (ppm)
CA: 50.5ppm
CG (ppm)
N8CB-CA-CO
N8CA-CB-CG
N8CB-CG-ND
CB (ppm)
ND (ppm)
CG (ppm)
CB: 38.1ppm

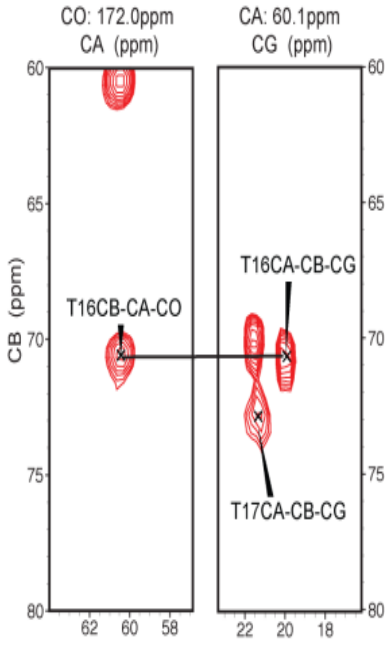

CO: 172.8ppm
CA (ppm)
CA: 62.0ppm
CG (ppm)
T11CB-CA-CO
T11CA-CB-CG
CB (ppm)

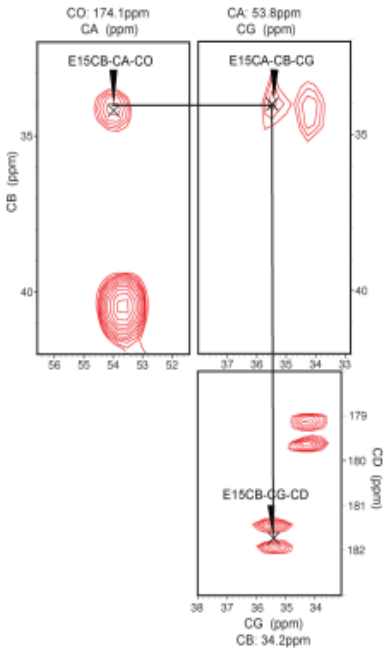

CO: 178.8ppm
CA (ppm)
CA: 59.1ppm
CG (ppm)
K10CB-CA-CO
K10CA-CB-CG
CG: 25.8ppm
CE (ppm)
K10CG-CD-CE
K10CB-CG-CD
K10NZ-CE
CB (ppm)
CD (ppm)
NZ (ppm)
CG (ppm)
CB: 32.6ppm

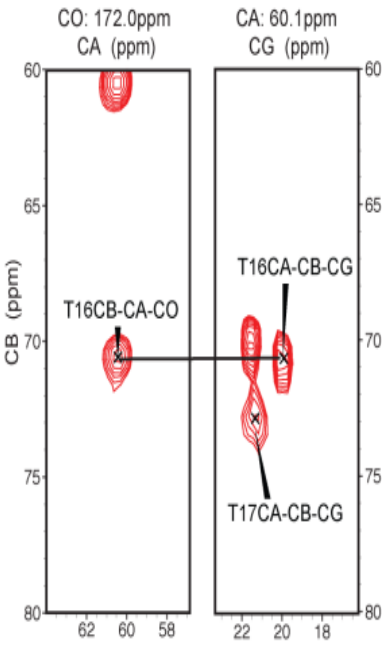

CO: 173.3ppm
CA (ppm)
CA: 54.4ppm
CG (ppm)
L12CB-CA-CO
L12CA-CB-CG
L12CB-CG-CD1
L12CB-CG-CD2
CB (ppm)
CD (ppm)
CG (ppm)
CB: 43.0ppm

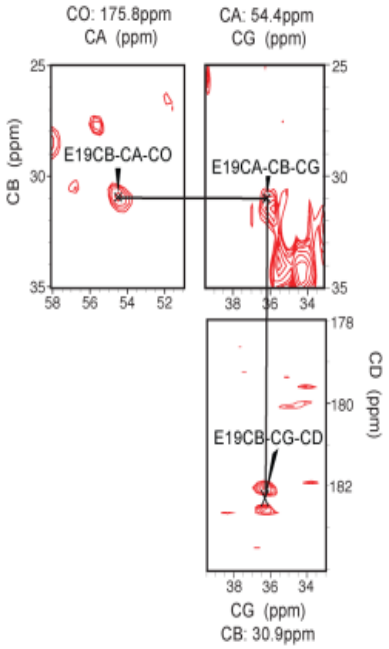

CO: 175.5ppm
CA (ppm)
CA: 53.3ppm
CG (ppm)
K13CB-CA-CO
K13CA-CB-CG
CG: 25.8ppm
CE (ppm)
K13CB-CG-CD
K13CG-CD-CE
K13NZ-CE
CB (ppm)
CD (ppm)
NZ (ppm)
CG (ppm)
CB: 39.0ppm

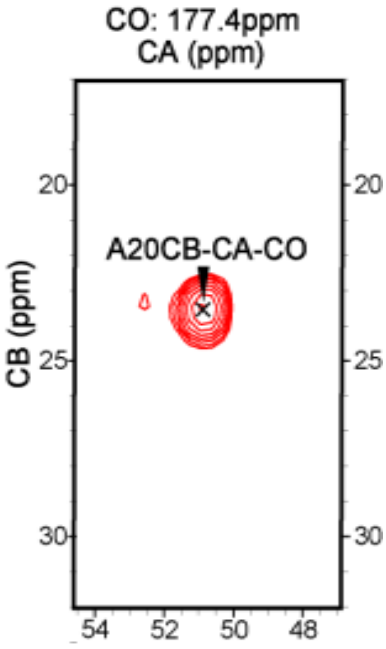

CO: 174.1ppm
CA (ppm)
CA: 53.8ppm
CG (ppm)
E15CB-CA-CO
E15CA-CB-CG
E15CB-CG-CD
CB (ppm)
CD (ppm)
CG (ppm)
CB: 34.2ppm

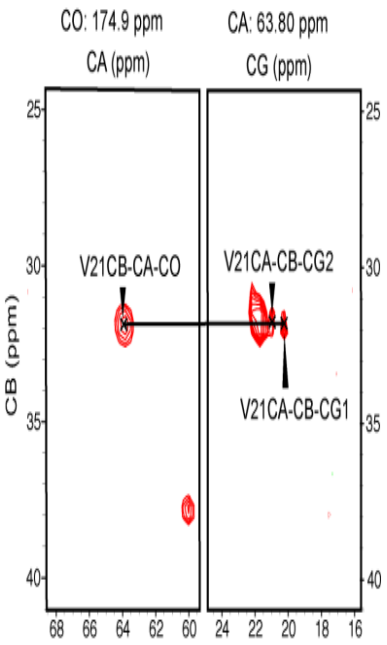

CO: 172.0ppm
CA (ppm)
CA: 60.1ppm
CG (ppm)
T16CB-CA-CO
T16CA-CB-CG
T17CA-CB-CG
CB (ppm)

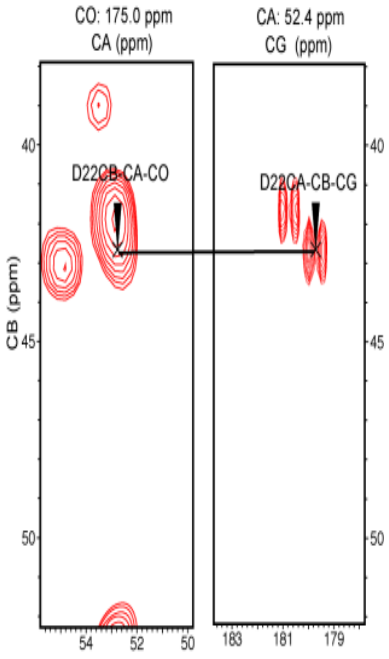

CO: 174.1ppm
CA (ppm)
CA: 53.8ppm
CG (ppm)
E15CB-CA-CO
E15CA-CB-CG
CB (ppm)
E15CB-CG-CD
CD (ppm)
CG (ppm)
CB: 34.2ppm

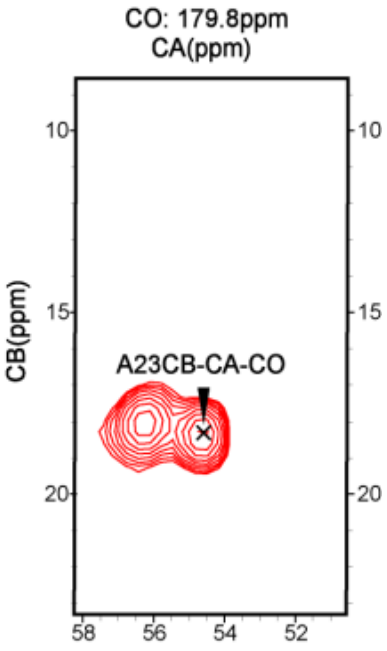

CO: 172.0ppm
CA (ppm)
CA: 60.1ppm
CG (ppm)
CB (ppm)
T16CB-CA-CO
T16CA-CB-CG
T17CA-CB-CG

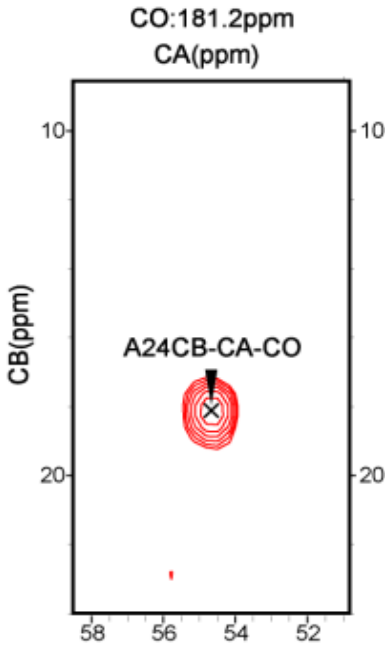

CO: 175.8ppm
CA (ppm)
CA: 54.4ppm
CG (ppm)
CB (ppm)
E19CB-CA-CO
E19CA-CB-CG
E19CB-CG-CD
CD (ppm)
CG (ppm)
CB: 30.9ppm

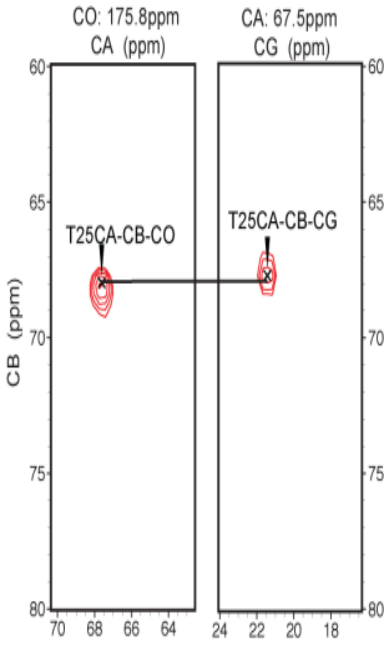

CO: 177.4ppm
CA (ppm)
A20CB-CA-CO
CB (ppm)

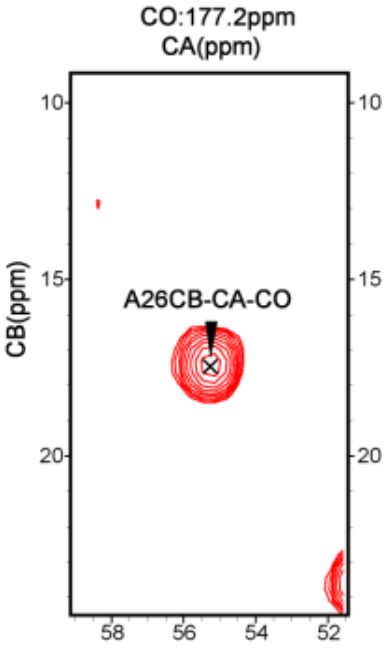

CO: 174.9 ppm
CA (ppm)
CA: 63.80 ppm
CG (ppm)
V21CB-CA-CO
V21CA-CB-CG2
V21CA-CB-CG1
CB (ppm)

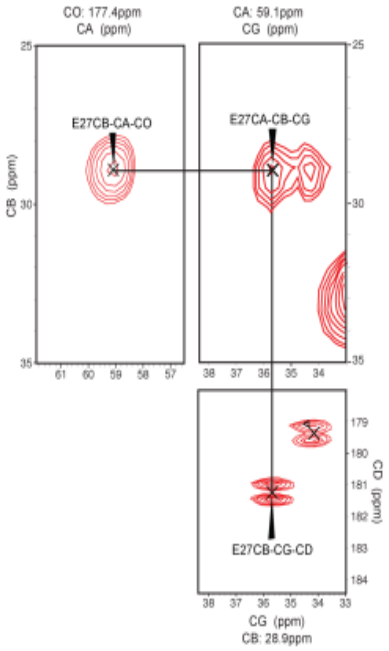

CO: 175.0 ppm
CA (ppm)
CA: 52.4 ppm
CG (ppm)
D22CB-CA-CO
D22CA-CB-CG
CB (ppm)

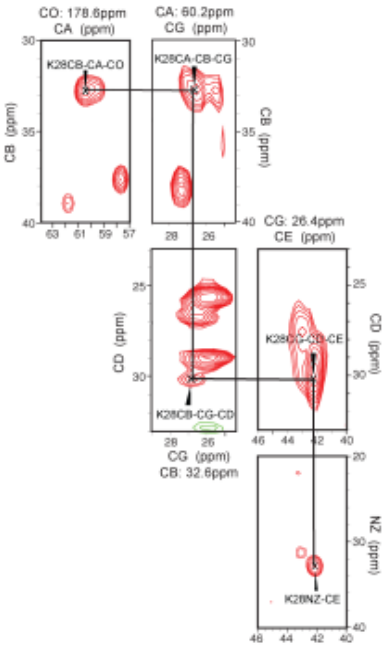

CO: 179.8ppm
CA(ppm)
A23CB-CA-CO
CB(ppm)

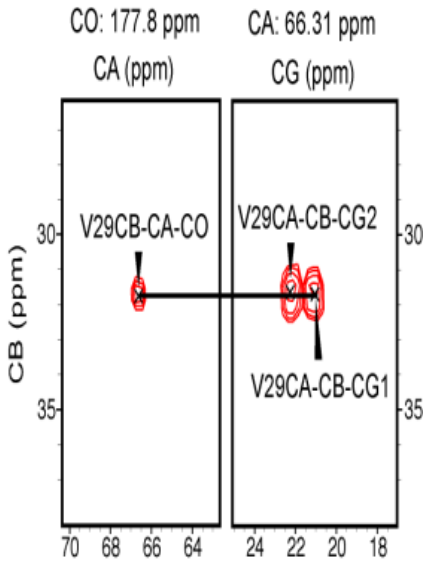

CO:181.2ppm
CA(ppm)
A24CB-CA-CO
CB(ppm)

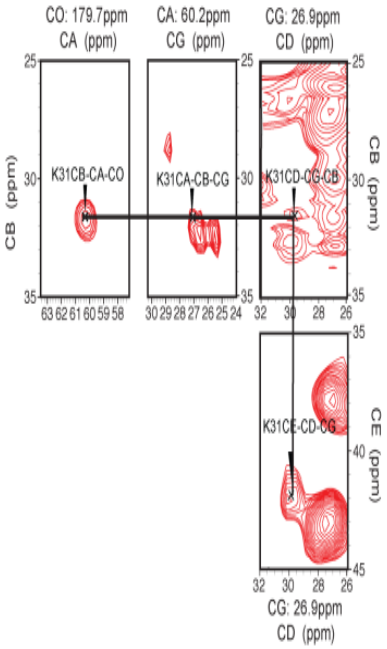

CO: 175.8ppm
CA (ppm)
CA: 67.5ppm
CG (ppm)
T25CA-CB-CO
T25CA-CB-CG
CB (ppm)

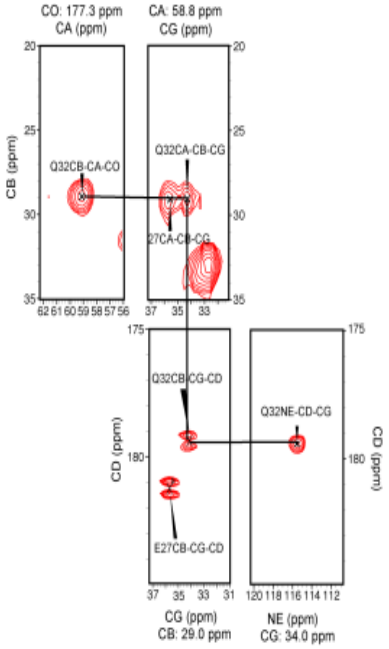

CO:177.2ppm
CA(ppm)
A26CB-CA-CO
CB(ppm)

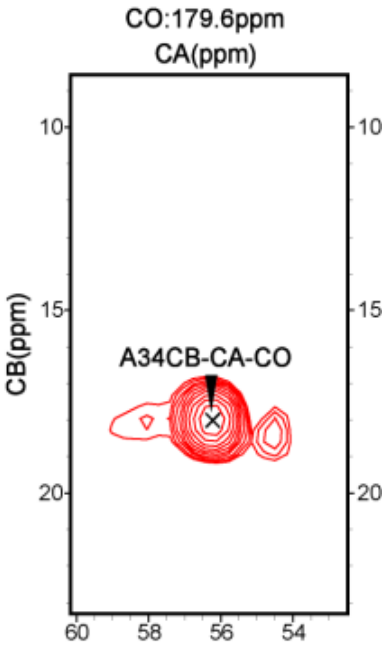

CO: 177.4ppm
CA (ppm)
CA: 59.1ppm
CG (ppm)
E27CB-CA-CO
E27CA-CB-CG
E27CB-CG-CD
CB (ppm)
CD (ppm)
CG (ppm)
CB: 28.9ppm

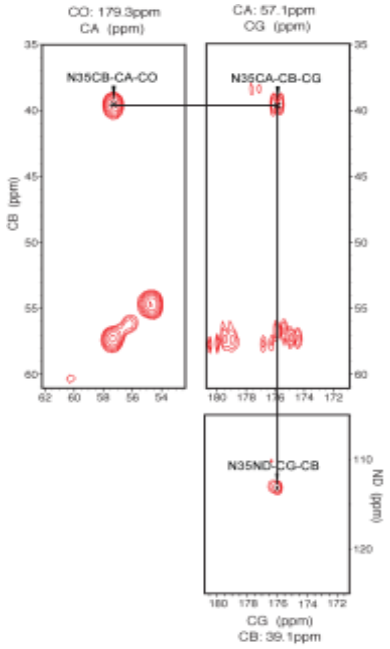

CO: 178.6ppm
CA (ppm)
CA: 60.2ppm
CG (ppm)
K28CB-CA-CO
K28CA-CB-CG
CG: 26.4ppm
CE (ppm)
K28CB-CG-CD
K28CG-CD-CE
K28NZ-CE
CG (ppm)
CB: 32.6ppm
NZ (ppm)

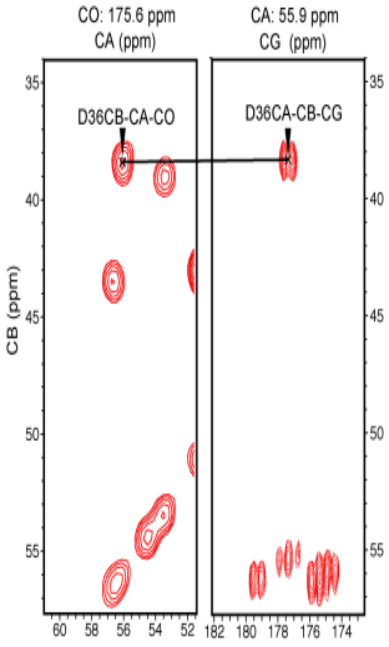

CO: 177.8 ppm
CA (ppm)
CA: 66.31 ppm
CG (ppm)
V29CB-CA-CO
V29CA-CB-CG2
V29CA-CB-CG1
CB (ppm)

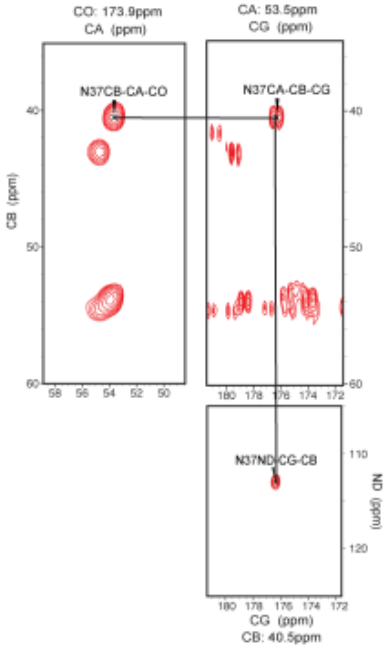

CO: 179.7ppm
CA (ppm)
CA: 60.2ppm
CG (ppm)
CG: 26.9ppm
CD (ppm)
K31CB-CA-CO
K31CA-CB-CG
K31CD-CG-CB
K31CE-CD-CG
CB (ppm)
CE (ppm)

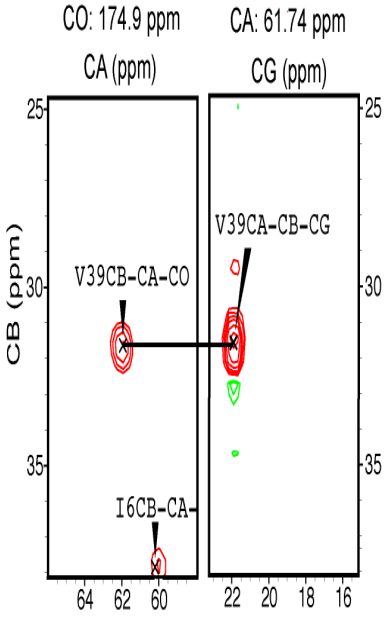

CO: 177.3 ppm
CA (ppm)
CA: 58.8 ppm
CG (ppm)
Q32CB-CA-CO
Q32CA-CB-CG
27CA-CB-CG
Q32CB-CG-CD
Q32NE-CD-CG
E27CB-CG-CD
CB (ppm)
CD (ppm)
CG (ppm)
CB: 29.0 ppm
NE (ppm)
CG: 34.0 ppm

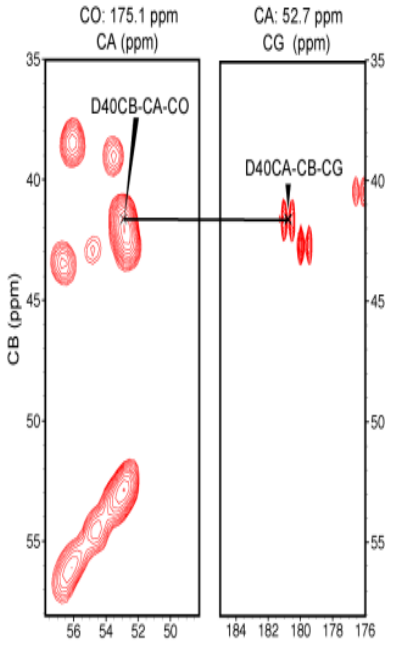

CO:179.6ppm
CA(ppm)
A34CB-CA-CO
CB(ppm)

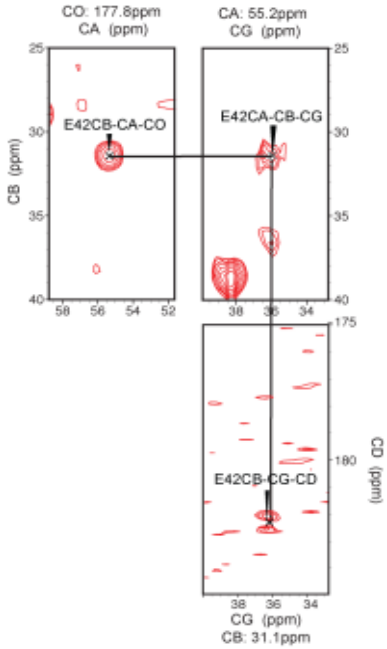

CO: 179.3ppm
CA (ppm)
CA: 57.1ppm
CG (ppm)
N35CB-CA-CO
N35CA-CB-CG
N35ND-CG-CB
CB (ppm)
ND (ppm)
CG (ppm)
CB: 39.1ppm

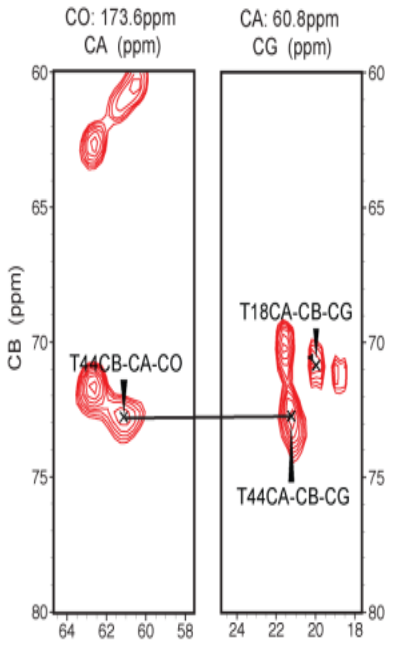

CO: 175.6 ppm
CA (ppm)
CA: 55.9 ppm
CG (ppm)
D36CB-CA-CO
D36CA-CB-CG
CB (ppm)

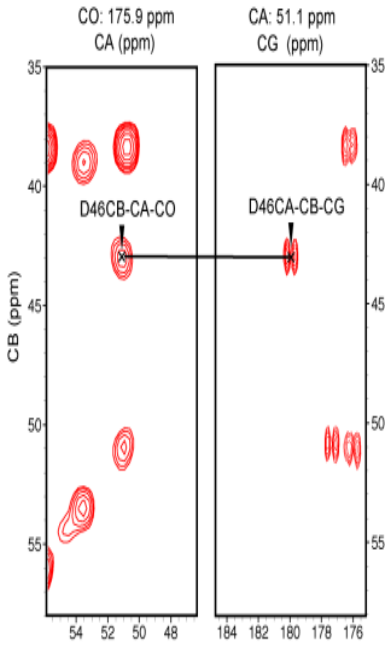

CO: 173.9ppm
CA (ppm)
CA: 53.5ppm
CG (ppm)
N37CB-CA-CO
N37CA-CB-CG
N37ND-CG-CB
CB (ppm)
ND (ppm)
CG (ppm)
CB: 40.5ppm

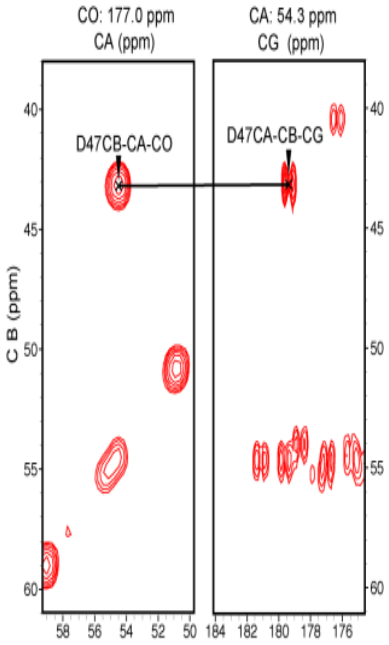

CO: 174.9 ppm
CA (ppm)
CA: 61.74 ppm
CG (ppm)
V39CB-CA-CO
V39CA-CB-CG
I6CB-CA-
CB (ppm)

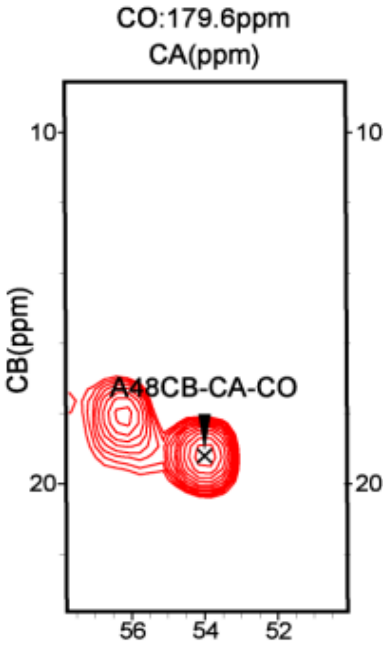

CO: 175.1 ppm
CA (ppm)
CA: 52.7 ppm
CG (ppm)
D40CB-CA-CO
D40CA-CB-CG
CB (ppm)

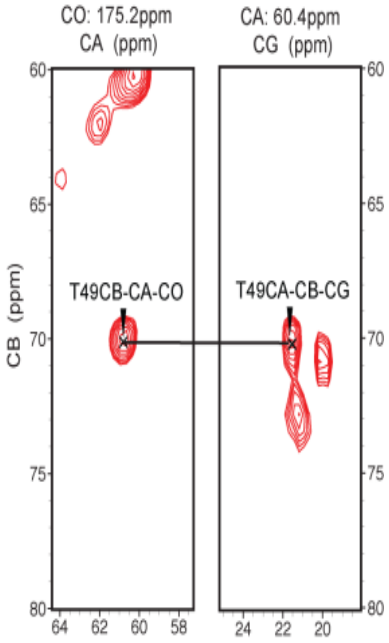

CO: 177.8ppm
CA (ppm)
CA: 55.2ppm
CG (ppm)
E42CB-CA-CO
E42CA-CB-CG
CB (ppm)
E42CB-CG-CD
CD (ppm)
CG (ppm)
CB: 31.1ppm

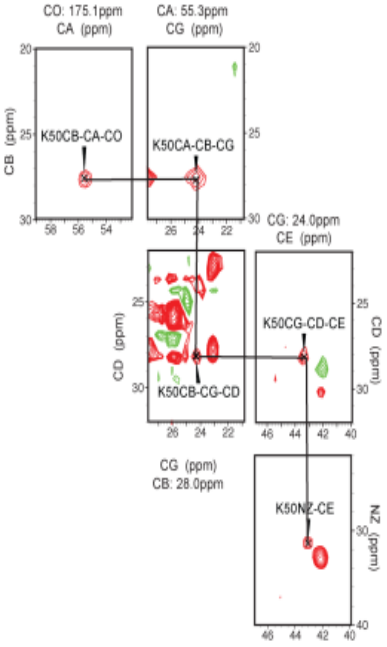

CO: 173.6ppm
CA (ppm)
CA: 60.8ppm
CG (ppm)
T18CA-CB-CG
T44CB-CA-CO
T44CA-CB-CG
CB (ppm)

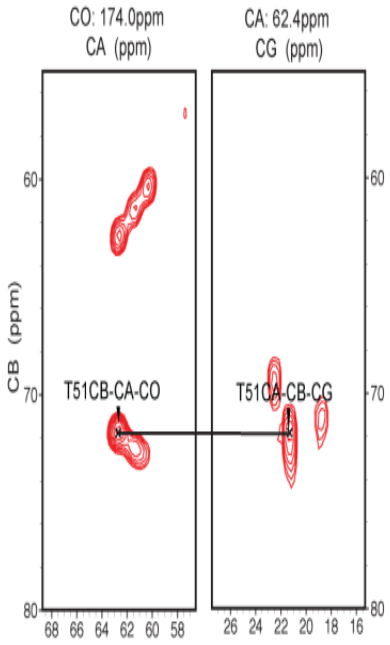

CO: 175.9 ppm
CA (ppm)
CA: 51.1 ppm
CG (ppm)
D46CB-CA-CO
D46CA-CB-CG
CB (ppm)

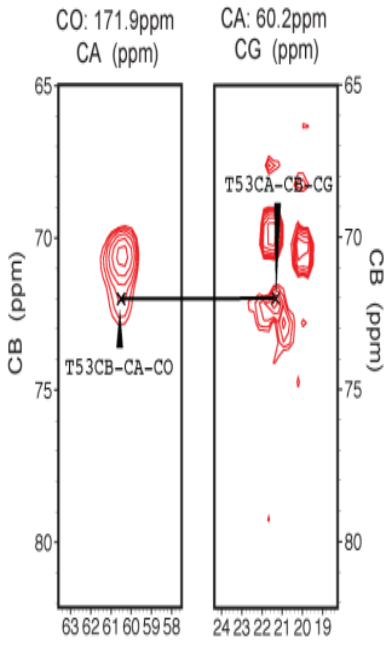

CO: 177.0 ppm
CA (ppm)
CA: 54.3 ppm
CG (ppm)
D47CB-CA-CO
D47CA-CB-CG
C B (ppm)

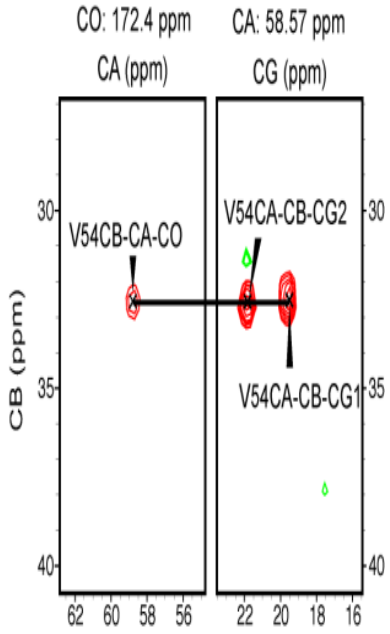

CO:179.6ppm
CA(ppm)
CB(ppm)
A48CB-CA-CO

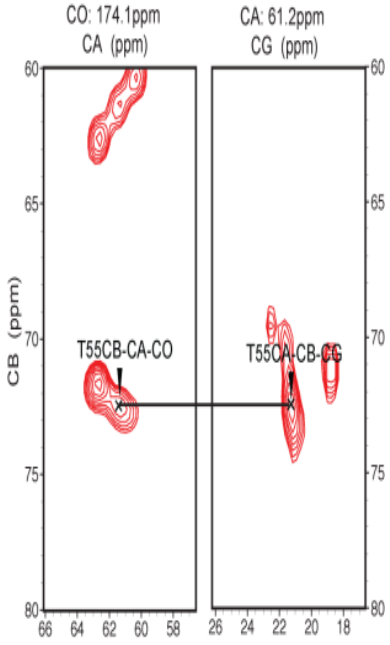

CO: 175.2ppm
CA (ppm)
CA: 60.4ppm
CG (ppm)
T49CB-CA-CO
T49CA-CB-CG
CB (ppm)

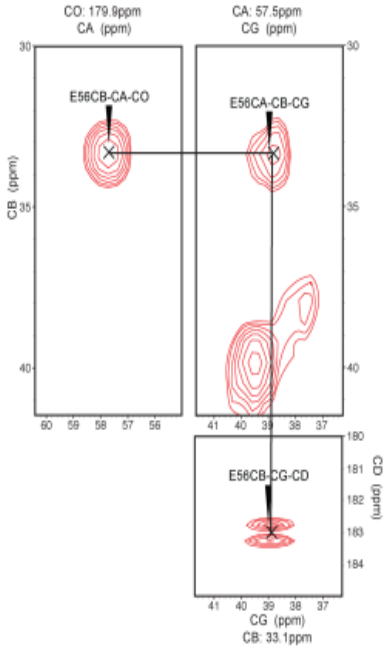

CO: 175.1ppm
CA (ppm)
CA: 55.3ppm
CG (ppm)
K50CB-CA-CO
K50CA-CB-CG
CB (ppm)
CG: 24.0ppm
CE (ppm)
K50CG-CD-CE
K50CB-CG-CD
CD (ppm)
CG (ppm)
CB: 28.0ppm
K50NZ-CE
NZ (ppm)


CO: 174.0ppm
CA (ppm)
CA: 62.4ppm
CG (ppm)
CB (ppm)
T51CB-CA-CO
T51CA-CB-CG


CO: 171.9ppm
CA (ppm)
CA: 60.2ppm
CG (ppm)
CB (ppm)
T53CA-CB-CG
T53CB-CA-CO


CO: 172.4 ppm
CA (ppm)
CA: 58.57 ppm
CG (ppm)
CB (ppm)
V54CB-CA-CO
V54CA-CB-CG2
V54CA-CB-CG1


CO: 174.1ppm
CA (ppm)
CA: 61.2ppm
CG (ppm)
CB (ppm)
T55CB-CA-CO
T55CA-CB-CG


CO: 179.9ppm
CA (ppm)
CA: 57.5ppm
CG (ppm)
E56CB-CA-CO
E56CA-CB-CG
E56CB-CG-CD
CB (ppm)
CD (ppm)
CG (ppm)
CB: 33.1ppm