LBNL-39117
UC-408

Structural and Dynamic Characterization of Eukaryotic Gene Regulatory Protein Domains in Solution

Andrew Loyd Lee
Ph.D. Thesis

Department of Chemistry
University of California, Berkeley

and

Structural Biology Division
Ernest Orlando Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94720

May 1996

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Structural and Dynamic Characterization of Eukaryotic Gene Regulatory Protein Domains in Solution

by

Andrew Loyd Lee

B.A. (Pomona College) 1989

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:

Professor David E. Wemmer, Chair
Professor Robert T. Tjian
Professor Raymond C. Stevens

1996

Structural and Dynamic Characterization of Eukaryotic Gene Regulatory Protein Domains in Solution

by

Andrew Loyd Lee

Reprinted in part with permission from *Biochemistry* 1994, *33*, 13775-13786.

Abstract

Structural and Dynamic Characterization of Eukaryotic Gene Regulatory Protein Domains in Solution

by

Andrew Loyd Lee

Doctor in Philosophy in Chemistry

University of California, Berkeley

Professor David E. Wemmer, Chair

Solution NMR was primarily used to characterize structure and dynamics in two different eukaryotic protein systems: the δ-a1-ε activation domain from c-jun and the *Drosophila* RNA-binding protein Sex-lethal. Circular dichroism and 2D NMR studies revealed that the c-jun acitivation domain has a significant helical propensity. The NMR results also suggest that there is a negligible amount of tertiary structure. In addition, fluorescence data suggests that this activation domain cannot be considered a molten globule. The second system is the *Drosophila* Sex-lethal (Sxl) protein, an RNA-binding protein which is the "master switch" in sex determination. Sxl contains two adjacent RNA-binding domains (RBDs) of the RNP consensus-type. The NMR spectrum of the second RBD (Sxl-RBD2) was assigned using multidimensional heteronuclear NMR, and an intermediate-resolution family of structures was calculated from primarily NOE distance restraints. The overall fold was determined to be similar to other RBDs: a βαβ-βαβ pattern of secondary structure, with the two helices packed against a 4-stranded anti-parallel β-sheet. In addition, ^{15}N T_1, T_2, and $^{15}N/^{1}H$ NOE relaxation measurements were carried

out to characterize the backbone dynamics of Sxl-RBD2 in solution. RNA corresponding to the polypyrimidine tract of *transformer* pre-mRNA was generated and titrated into 3 different Sxl-RBD protein constructs. Combining Sxl-RBD1+2 (both RBDs) with this RNA formed a specific, high affinity protein/RNA complex that is amenable to further NMR characterization. The backbone ^{1}H, ^{13}C, and ^{15}N resonances of Sxl-RBD1+2 were assigned using a triple-resonance approach, and ^{15}N relaxation experiments were carried out to characterize the backbone dynamics of this complex. The changes in chemical shift in Sxl-RBD1+2 upon binding RNA are observed using Sxl-RBD2 as a substitute for unbound Sxl-RBD1+2. This allowed the binding interface to be qualitatively mapped for the second domain.

For My Father

Edward K.C. Lee

Table of Contents

Acknowledgments

As an incoming chemistry graduate student at Berkeley, I thought I had a pretty good idea of the direction in which I was headed; with an interest in physical chemistry, and having discovered biochemistry only months before, I was going to do laser spectroscopy applied to some kind of interesting biological system. I would not have guessed that I would end up doing protein NMR. (My previous encounters with NMR typically left me in a state of confusion.) Graduate school has therefore been filled with countless unexpected turns, incidents, dead ends, disappointments, and the ever-elusive successes. For me, this is the nature of graduate school, and science.

During this time, I have also discovered that although science itself holds the revered laws and theories which explain so many natural phenomena, it is only *fun* because of the people who work alongside of you. It has been these people with whom I have shared moments of inspiration and discovery, the bleak reality of failed experiments, and everything in-between. Above all I would like to thank my professor, David Wemmer. He took me into his research group when he didn't need or particularly want any more students, and I will always appreciate this. He is a natural teacher and an extraordinary intellect. I believe that from him I have gained insight into his way of thinking—a manner of approaching problems logically, fearlessly. Also, he has helped me to form specific ideas about what constitutes good or "useful" research; this will be quite helpful to me over the next few years. I look forward to interacting with Dave in future meetings.

The most rewarding experience for me in graduate school was learning the basics of heteronuclear NMR with two senior "peptide/protein people" in the group, Brian Volkman and Fred Damberger. Brian and Fred were my mentors. We all had similar projects moving more or less in parallel, and therefore we shared the excitement of learning how to do 3D and 4D heteronuclear NMR, protein assignment, and structure calculations. In

retrospect, this was an unusually stimulating and collegial environment for this to occur, in both time and place. I will always look back on this year or two as a "golden period." I would like to send special thanks to Brian for saving the Sex-lethal project during the last 8 months. Our collaboration has been surprisingly successful, in addition to being the most intense few months I have experienced in science. Most importantly, Brian has become a great friend over the last several years. I would also like to send special thanks to Jeff Pelton. I know that Fred, Brian, and I would have made much slower progress had it not been for Jeff, who returned to Berkeley with his skills in triple-resonance NMR just before we were about to undertake these formidable experiments. I owe much of what I have learned to him.

On the biological side of things, Roland Kanaar, professor Don Rio, and David Rudner all were extremely generous in helping me make protein and RNA samples for the Sex-lethal project. Roland was instrumental in getting the project off the ground by helping me find a purification for RBD2, and he also worked out the "RNase-free" protocol for RBD1+2. I appreciate the fact that Roland and David always relinquished an area on their benches for me. The Rio lab was always a pleasant and well-equipped place to work. In the same building, David King has "mass spec'd" so many samples for me over the years, that I lost count long ago. He has also helped to solve purification problems for me and countless other students. He is an institution within the biochemical and biophysical community at Berkeley, and I would like to thank him for everything.

The Wemmer group has provided a very pleasant atmosphere for doing science. I believe that, aside from Dave's easygoing nature, this largely came from the enthusiasm of Marty Pagel and Mark Kubinec, who seemed committed to teaching and distributing information to anyone in the group who was willing to learn. Ho Cho is another enthusiastic fellow student, and I was fortunate to be paired with him to work on transcription factors when I joined the group. Ho taught me basic molecular biology techniques and "showed me the ropes" in general. Two other students—Mike Nohaile and

Patty Fagan—proved to be excellent companions for thesis-writing. I'm sure that writing this thesis would have seemed more painful if Mike and Patty had not been writing at the same time; it was always a relief to be able to console each other in our "thesis doldrums." Mike and I in particular have spent a lot of time over the last 5 or 6 years discussing and disputing many things involving proteins and NMR, since we share these passions. Our conversations were informative and fun, and I will miss them. I wish the best of luck to Mike as a postdoc. I would also like to thank Elena Rodriguez, previously an undergraduate in the lab, for her efforts in generating *in vitro* transcribed RNA. Corey, Bernie, Nat, Werner, Tim, Qing, Rafa, Christian, Jonathan, John, Cheryl, Dorothy, Pete, and Ping, it has been a pleasure to work alongside of you. I would especially like to thank Stephanie Robertson for showing an interest in the Sex-lethal project. I believe that the project is in excellent hands, and I wish Steph the best of luck.

Finally, I would like to thank my family for the loving support that they have provided for me over the years. Nothing could ever replace this, and I am forever grateful. I would also like to thank my housemates at Thomas street for their friendship: Ron, Pia, The Mave, Dave, Lilith, Ted, Janice, Liz, Craig, and Louisa. These have been good times!

May 1996

Chapter 1

Introduction to Protein NMR

In my brief scientific experience, nothing has captured my imagination to the extent that proteins have. To me, the idea that so many different proteins can be constructed from the same 20 amino acids, and carry out the diverse tasks constituting life, is remarkable. The beauty of this idea lies somewhere in between the simplicity of the amino acids themselves and the complexity of the permutations they find themselves in. Many scientists have recognized the intellectual attractiveness as well as the biological implications of understanding protein function, and therefore proteins have been under intense scrutiny by biologists, biochemists, chemists, and biophysicists for much of this century. In the last 50 years, proteins have been primarily understood within the context of their three-dimensional structures at atomic resolution as determined by X-ray crystallographic methods. More recently, nuclear magnetic resonance (NMR) spectroscopy as applied to proteins in solution has emerged as a technique which holds the promise of obtaining the combination of structural and dynamic information.

In this introductory chapter, I hope to lay down the basic principles and working ideas of NMR as applied to proteins in solution. As I mainly wish to communicate what kind of information the various NMR experiments yield so that the information can be appreciated and *used*, I will largely refrain from delving into the quantum mechanical evolutions of the nuclear spin ensembles. This is not the most obvious forum for a detailed discussion of "spin physics", so I will reference books and papers which I have found useful for understanding NMR phenomena in detail. To start, I will present the concept of the density matrix, upon which all NMR experiments are based, in order to emphasize that the methods used are indeed grounded in quantum mechanics, but from then on I will focus

on the qualitative aspects of how the experiments work and why they yield the information that they do. Hopefully this will give most readers a basic foundation for understanding the essence of the experiments to be discussed later. In the chapters that follow I will inevitably use NMR "language" to describe important details for any future NMR spectroscopist readers. In fact, some sections of later chapters could be considered to be an extension of this introductory chapter. The topic of relaxation/dynamics, for example, will be introduced and explored in Chapter 6. These additional topics will be dealt with when they become relevant. Finally, a collection of NMR pulse sequences can be found in Appendix D, and in certain instances I will refer to them instead of giving complete descriptions in the main text. With regard to this technical "stuff", I apologize in advance to those who do not spend minutes, hours, or days worrying about the workings of NMR.

The nuclear spin density matrix

An ensemble of nuclear spins is most rigorously described by the density matrix, ρ (Blum, 1981; Goldman, 1988). For an ensemble of isolated spins with spin quantum number of 1/2, the density matrix takes on the form

$$\begin{array}{c} \\ |\alpha\rangle \\ |\beta\rangle \end{array} \begin{array}{c} \begin{array}{cc} |\alpha\rangle & |\beta\rangle \end{array} \\ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{array}.$$

The two columns and two rows of this matrix correspond to the two different spin states, $|\alpha\rangle$ and $|\beta\rangle$, available to a spin-1/2 particle. The elements of this matrix are related to the probabilities of finding a spin in the ensemble in a pure state (diagonal elements a and d) or in a "mixed" or coherent state (off-diagonal elements b and c). It is the coherent states that give rise to observable signals, or coherences, in NMR. In Fourier transform (FT) NMR, otherwise known as pulsed NMR, these coherences relax back to an equilibrium population of pure states as the coherences are being detected.

As structural biologists, however, we are not terribly interested in isolated spins. Polypeptides have many protons, and if two or more spins are connected by less than 3 or 4 bonds, they are "aware" of each other's spin states. This is known as spin-spin or J-coupling (Harris, 1986), and is manifested as a "splitting" of the observed peak, resonance, coherence (these are synonymous terms) into two components corresponding to the two spin states of the other spin. For structural biologists who usually prefer to work at high magnetic field strengths, splittings from J-couplings are often small in comparison to the difference in chemical shift between the two J-coupled spins. In this case, it is useful to consider the density matrix for two "weakly" coupled spins,

$$\rho = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{pmatrix}.$$

Now there are 4 states for each of the 4 columns and rows: $|\alpha\alpha\rangle$, $|\alpha\beta\rangle$, $|\beta\alpha\rangle$, and $|\beta\beta\rangle$. $|\alpha\alpha\rangle$ corresponds to the two coupled spins both being in the α state, $|\alpha\beta\rangle$ corresponds to the first spin being in the α state and the second in the β state, and so on. It is interesting to note at this point that with two spins instead of one, the number of possible states and coherences has increased from four to sixteen! Furthermore, there are now many more different types of coherences, a feature that can be exploited for various purposes with clever experimental design. With three J-coupled spins there will be 9x9=81 possible "states". Fortunately, for most applications in NMR, and certainly for protein NMR, pretending that there are only 2-spins at a time and that they are weakly J-coupled is a good enough approximation for many of the types of experiments we do.

In practice it is often more useful to use the density matrix in a different incarnation. By doing an expansion of the density matrix in a Cartesian-based basis set, the density matrix can be tracked with the visual aid of x,y,z vectors, and a shorthand notation

developed in the mid-eighties for tracking the matrix elements in this basis set has proven to be extremely convenient. This shorthand is known as "product operators" (Sorenson et al., 1983; Ernst et al., 1987). In 2D and ND NMR, the experiments have always been developed with product operators or the density matrix as a "map" of where spin magnetization or coherences exist at all times.

Modern pulsed 2D and ND NMR

The simplest pulsed NMR experiment is the "1-pulse" experiment, in which the pulse is a short (~10 us) radiofrequency (RF) "90°" pulse which creates magnetization perpendicular to the static magnetic field (i.e. transverse magnetization). As this net magnetization precesses and decays (the free induction decay), it induces a current in the surrounding coil, which in turn can be digitized by the spectrometer computer. This is a one-dimensional NMR experiment, i.e. it yields a simple one-dimensional NMR spectrum that all chemists are familiar with. Figure 1.1 shows a 1-pulse experiment on an 11 kD protein in solution, as well as the data processsing Fourier transform (FT) required to convert the protein signals from the time domain (free induction decay) into the frequency domain (spectrum).

Figure 1.2 illustrates a general two-dimensional (2D) pulse sequence and how the 1-pulse element is often at the end of experiments of higher dimensionality for the same purpose as in the "1-pulse experiment"—to create signals to observe. The four fundamental elements which comprise a 2D experiment are labeled (Ernst et al., 1987): preparation, evolution, mixing, and detection. Briefly, the role of the preparation period is to create non-equilibrium magnetization. The role of the evolution period is to record chemical shift evolution, similar to t_1 in the 1-pulse experiment (Figure 1.1), but done in an indirect manner. By incrementing the delay between the first and second pulses (black bars) and recording the 1D spectrum (t_2) for each increment of t_1, chemical shift information is encoded in the series of 1D spectra. "Decoding" this information is

1D NMR

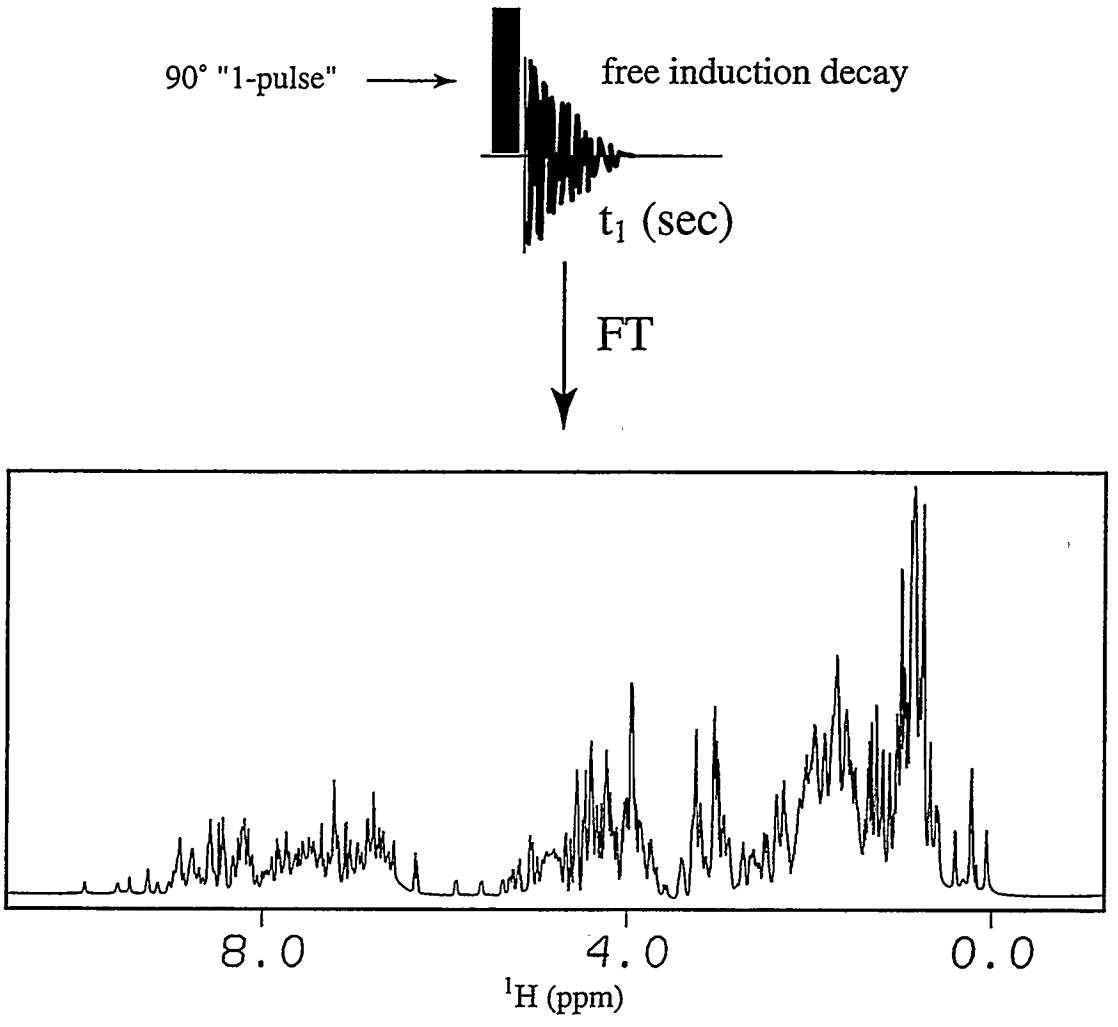


Figure 1.1. One-dimensional NMR pulsing methodology. A 90° radiofrequency (RF) pulse generates an observable free induction decay which can be Fourier transformed to yield a frequency domain spectrum. A 1D frequency domain spectrum of an 11 kD protein (Sxl-RBD2) is shown.

2D NMR

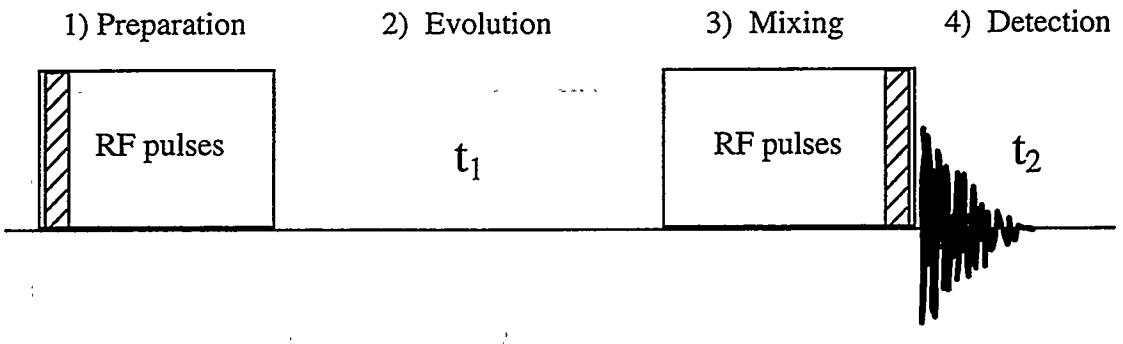


Figure 1.2. Organization of pulses required for a two-dimensional NMR experiment. An RF pulse or a group of pulses is represented by a box. The exact sequence of pulses inside of the boxes will be different for different experiments. The hatched bars represent 90° pulses which are commonly found in 2D experiments.

accomplished by doing an FT along the each dimension, and the resultant 2D data takes on the form of a contour map. It should be mentioned that chemical shifts are not necessarily recorded during the evolution period. Sometimes J-coupling or spin relaxation can "evolve" instead, but for our purposes chemical shift always evolves during t_1. This t_1 evolution essentially "frequency labels" a spin's magnetization before it is transferred to another spin, which puts a "second label" on the magnetization in t_2. The nature of the mixing period is what makes 2D experiments different from each other. NOESY, COSY, and TOCSY are all 2D experiments which will be discussed, and they all have different arrangements of RF pulses in this period. Basically, the different arrangements of pulses in the mixing period create magnetization transfer between spins through different types of mechanisms. In this way, different types of correlations (e.g. through-space in NOESY, through-bond in COSY) between different spins can be made. *Correlation* spectroscopy is at the heart of biomolecular NMR. Finally, the detection period is obvious: signal detection is required for observing anything at all.

Having made the conceptual "leap" from 1D to 2D NMR, going from 2D to 3D or 4D NMR is comparatively quite easy (Clore & Gronenborn, 1991). Because the key to the "2" in a 2D experiment is the ability to record chemical shift evolution in an indirect manner, it follows logically that a 3D experiment must have a *second* indirectly detected chemical shift evolution period. This second evolution period must be preceded by the first evolution and mixing periods (these could be considered to collectively form the second preparation period) as well as a second preparation period, and it must be followed by some sort of additional mixing period to ensure that the extra dimension has non-trivial information in it. Clearly, there is never more than one detection period. Therefore, a 4D NMR experiment will have 3 preparation, evolution, and mixing periods each, and it will end with the detection period. As dimensions are added, the length of time required to complete the experiment increases because each evolution period must be incremented

independently, and therein lies the primary drawback to high dimensionality NMR experiments.

The original motivation (Wüthrich, 1986) for moving from 1D to 2D for biomolecules was that these large molecules could not have their resonances assigned very easily since there is overlap of peaks in the 1D spectrum (Figure 1.1). This resolution problem was the same motivation for moving from 2D to 3D and 4D, albeit at a higher MW threshold. Adding dimensions spreads out the peaks so that they can be identified individually. The manner in which the peaks are distributed in the "extra" dimensions is determined by what happens during the mixing time(s). Specific examples will be discussed in more detail below. An equally important (and related) reason for moving to the second dimension was that the development of the 2D NOESY experiment (Kumar et al., 1980) made possible structure determination of biomolecules by NMR through use of the distance dependent nuclear Overhauser effect (Neuhaus & Williamson, 1989). Finally, it should be mentioned that pulsed NMR experiments are not restricted to a single type of nucleus, namely ^{1}H. For proteins in particular, moving to 3D and 4D experiments with evolution on ^{13}C and/or ^{15}N nuclei has turned out to be a powerful combination, as will be described below. Experiments of any dimensionality and with different types of mixing periods can be devised using many different groups of nuclei, such as ^{13}C, ^{15}N, ^{31}P, ^{19}F, and even ^{2}H and ^{3}H as well as ^{1}H. The possibilities are seemingly endless.

The assignment problem

The first goal in most peptide or protein NMR research projects is to obtain sequence specific resonance assignments (Wüthrich, 1986). Without knowing the relationship between specific nuclei in the molecule and specific chemical shifts in the NMR spectrum, nothing can be learned about specific sites in the molecule. Assignments can be obtained by using two types of information: "through-bond" and "through-space". Both approaches rely on making correlations between spin pairs, whether they are linked

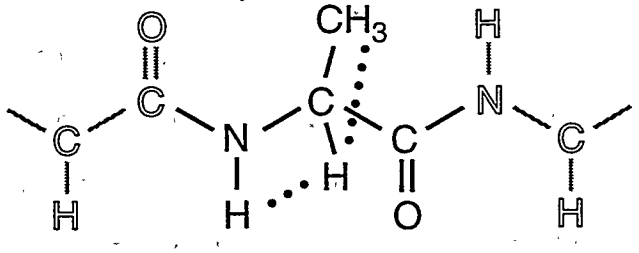
covalently or are close in space, and they both rely on two-dimensional spectroscopy. Naturally, higher dimensionality can be used as well since any 2D experiment can formally be incorporated into a 3D experiment (Clore & Gronenborn, 1991). Already knowing what kind of information is available, I would first like to describe the nature of the general assignment strategy, that is how the information is to be used to obtain sequence-specific assignments. Then in the following two sections I will describe how the different types of information are experimentally obtained using 2D NMR.

Consider the organization of ^1H spins in an alanine residue within the context of a polypeptide as shown in Figure 1.3A. Protons separated by 3 or fewer bonds generally have large enough J-couplings (coupling constants) such that magnetization from one proton can be transferred to the other proton. The mechanism for the magnetization transfer (Ernst et al., 1987) is difficult to describe without making explicit use of the density matrix, so I will simply state that: *magnetization can be transferred between protons which share a J-coupling which is on the order of or larger than the linewidths (in Hz) of the two protons.* As an example, ^1HN-$^1H_\alpha$ J-coupling constants or splittings are typically in the range of 5-9 Hz. Using J-couplings for magnetization transfer is the essence of through-bond correlation. Since many amino acids have networks of J-couplings throughout their sidechains, entire amino acids can be identified via through-bond ^1H-^1H correlations (Wüthrich, 1986). Moreover, different amino acids have different J-coupling network patterns, allowing for the various networks to be classified into different amino acid types, or "spin-systems". The combination of these patterns with characteristic chemical shifts for various amino acid substituents makes identifying specific amino acid spin-system types relatively easy, at least for shorter polypeptides. Finally, it is important to note that protons in neighboring amino acids are always separated by at least 4 bonds. This means that interresidue correlations cannot be made with simple ^1H-^1H J-couplings. This would be a moot point if the polypeptide did not have multiple occurances of particular amino acid types, since in this case all residues could be assigned

^{1}H Assignment Strategy

A)

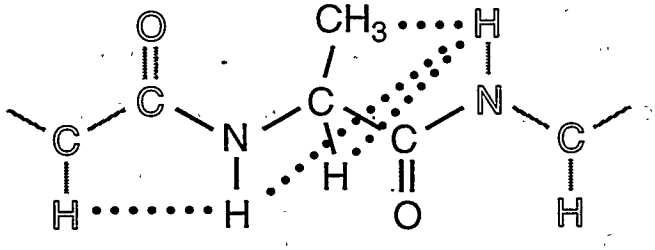
Ala



Through-Bond
(J-coupling)

B)

Ala



Through-Space
(NOE)

Figure 1.3. "Through-bond" (A) and "through-space" (B) correlations for an alanine residue. Dotted lines represent magnetization transfer pathways which allow spin pairs to be correlated via a crosspeak between the two protons in a 2D spectrum. In B, only interresidue correlations are shown.

unambiguously based on spin-system information only. In reality this is rarely the case. Therefore we need to make connections between residues in some other manner so that spin-systems can be placed "sequence-specifically" into the polypeptide. This is where the "through-space" correlations become useful.

The nuclear Overhauser effect, or NOE, is observed between two spins which are close in space (Neuhaus & Williamson, 1989). In short, perturbing the relative populations of the $|\alpha\rangle$ and $|\beta\rangle$ states in one spin ensemble will affect the relative populations of states in the second spin ensemble. The magnitude of this effect is governed by a system of interconnected relaxation rates between the four spin states of the two proximal spins. These rates are consicely described in Solomon's set of equations (Solomon, 1955). Combining the population effects of the NOE with 2D spectroscopy gave birth to the 2D NOESY (NOE SpectroscopY) experiment (Kumar et al., 1980) which revolutionized biomolecular NMR (Wüthrich, 1986) in the 1980's by enabling macromolecular solution structures to be determined from large sets of interproton distance restraints. The NOESY will be discussed in greater detail shortly. Returning to the assignment problem, the NOE (via NOESY) can be used to make through-space correlations between proximal spins (Figure 1.3B). This criterion of being close in space (< 5 Å) is not limited to intraresidue correlations as in the ^{1}H J-coupling networks, and therefore the NOE can be used to connect neighboring or "sequential" residues (Wüthrich, 1986) by making through-space correlations or connectivities. It should be added that having prior knowledge of the amino acid sequence makes this process considerably easier and the resulting assignments more reliable. Now that it is apparent how through-bond and through-space correlations are to be used in the resonance assignment process, it is time to take a closer look at the basic 2D experiments as well as the data, i.e. correlations, that they produce.

"Through-bond" experiments

COSY. The 2D COSY experiment (Aue et al., 1976; Wemmer, 1989) has already been mentioned. It is the simplest way to achieve through-bond magnetization transfer between two J-coupled spins. The COSY pulse sequence is given in Figure 1.4A, labeled with the four fundamental elements of a 2D experiment. This is the simplest of all 2D experiments. The mixing period is the second pulse, which transfers magnetization from one ^1H spin to the other. To summarize, chemical shift on the starting spin is recorded during t_1, magnetization is transferred to the second spin with the second pulse, and then the chemical shift on the second spin is recorded during t_2. This occurs for all J-coupled spin pairs simultaneously as the experiment runs (typically 10-24 hours). An example of the expected connectivities in the doubly Fourier transformed COSY spectrum ($t_1 \rightarrow f1$, $t_2 \rightarrow f2$) for an alanine spin system is illustrated in Figure 1.4B. The three peaks along the diagonal line are equivalent to the one-dimensional spectrum of alanine. The other peaks, which are off the diagonal, are called "crosspeaks". These crosspeaks *correlate* the spin pairs which have sufficiently large J-couplings, namely ^1HN-$^1H_\alpha$ and $^1H_\alpha$-$^1H_\beta$ in alanine. In other words, they contain all of the interesting information in the COSY experiment. In addition, there are two crosspeaks per correlation. This arises from the fact that magnetization originally started on each spin (t_1) and was subsequently transferred to the other. This symmetry is inherent to homonuclear 2D spectroscopy.

TOCSY. Another and perhaps more useful through-bond experiment is the 2D TOCSY, short for TOtal Correlation SpectroscopY (Braunschweiler & Ernst, 1983; Griesinger et al., 1988). The TOCSY uses J-couplings just as COSY does, but correlates *all* spins in a spin-system instead of just spin pairs. This turns out to be very helpful when spectra become very crowded. The mixing time in this experiment consists of a rapid series of lower-power ^1H pulses that are repeated for 30-100 milliseconds (for comparison, the 2nd pulse in the COSY lasts for ~10 microseconds). In this "pulse train", the spins in the J-coupled network tend to lose their identity and so magnetization is transferred freely

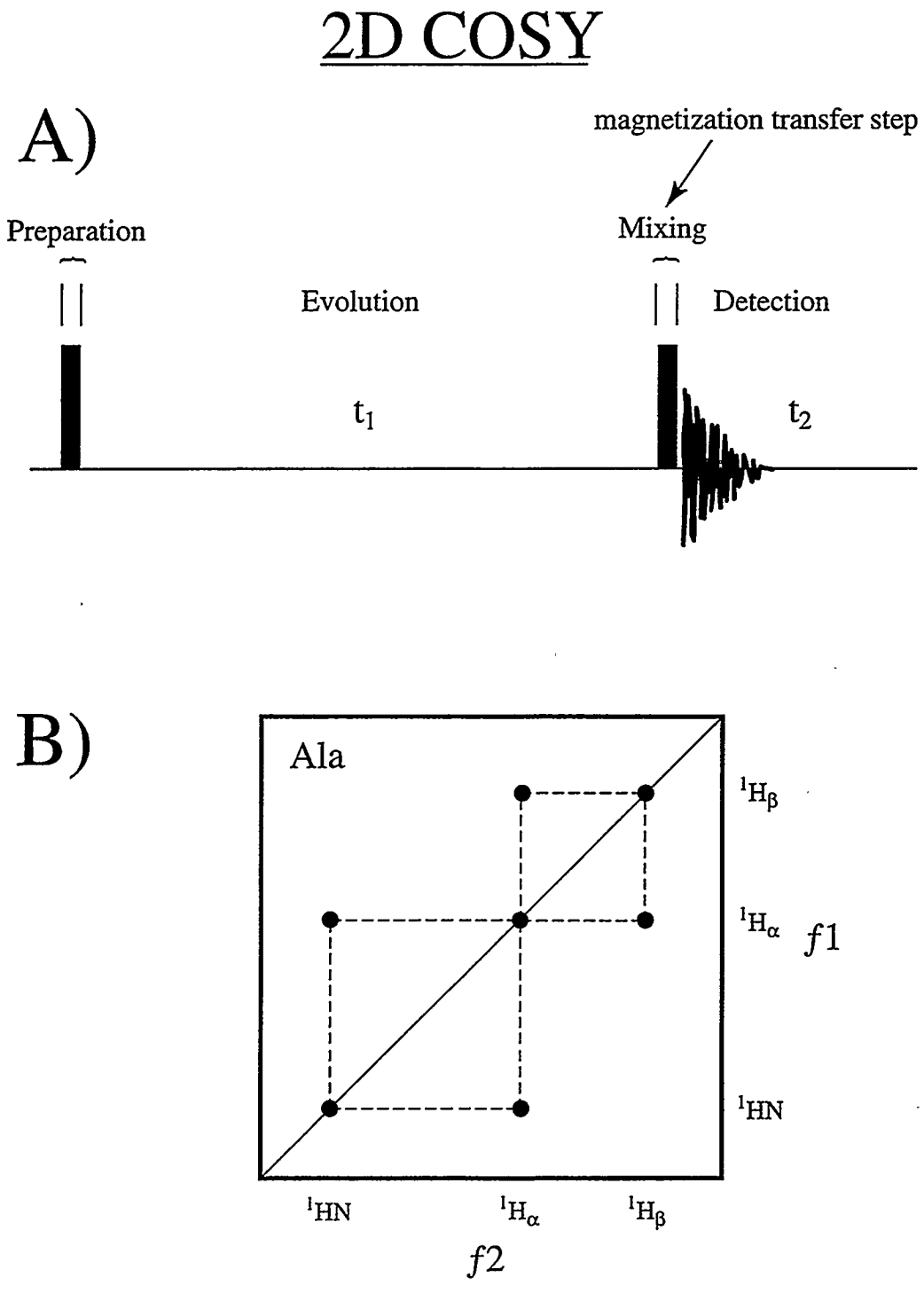


Figure 1.4. Two-dimensional COSY experiment. A) The pulse sequence is presented, and the four fundamental elements of a 2D experiment are shown. Both pulses are 90° pulses. B) The resulting Fourier transformed COSY spectrum of an alanine residue. Off diagonal "peaks" correlate (dashed lines) two individual spins.

among them until the detection period is begun. In the TOCSY, all correlations can be found as long as one of the spins in the spin-system is resolved. Crosspeaks in a spin-system are found in a straight line of crosspeaks instead of in the step-wise manner in the COSY experiment. Therefore, in Figure 1.4B, a TOCSY spectrum would have additional crosspeaks between ^{1}HN and $^1H_\beta$ spins. Nevertheless, the type of information obtained in the TOCSY is the same as in the COSY. The difference is that the information is more redundant in the TOCSY and is therefore often easier to interpret.

"Through-space" NOESY

The pulse sequence for the 2D NOESY experiment is presented in Figure 1.5 along with a sample spectrum of a small protein (~11 kD). The first pulse generates transverse magnetization and t_1 evolution is carried out just as in the COSY and TOCSY experiments. The second pulse initiates the mixing or NOE buildup period by putting magnetization along the z direction. As always, this is the stage in which magnetization is transferred from one spin to another. The difference in the NOESY is that the magnetization transfer is a through-space transfer mediated by a dipolar "cross-relaxation" mechanism (one of the rates in the Soloman equations). The third and final 90° pulse generates transverse magnetization that is detected and then digitized by the spectrometer computer.

The NOESY is the most important experiment for NMR structure determination (Wüthrich, 1986) since a large number of interproton distances (< 5 Å) can be estimated from the intensities of NOESY crosspeaks. Wüthrich and coworkers (Williamson et al., 1985) were first to use NOESY to determine the solution structure of a small protein, the 57 residue Bull Seminal Inhibitor (BUSI). NOESY spectra generally have many crosspeaks (Figure 1.5) since a given spin can be close to a relatively large number of other spins. Therefore, it is common to have crowded regions in the spectrum, such as the region around 1-2 ppm in Figure 1.5 or in Figure 1.7. It seems then that despite this elegant method for spreading out NOEs in two dimensions, getting assignments and

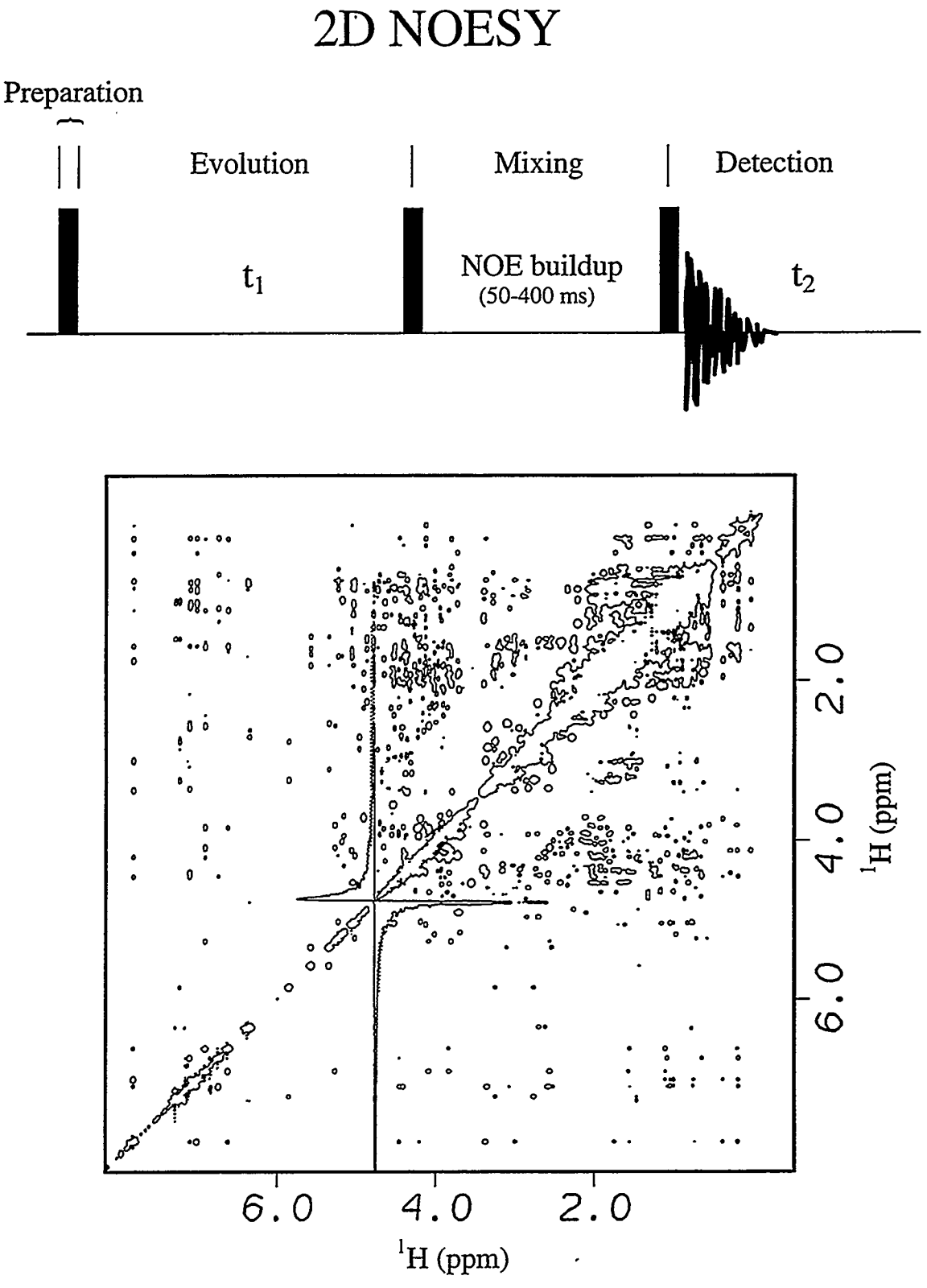


Figure 1.5. Two-dimensional ^{1}H-^{1}H NOESY experiment. The NOESY spectrum shown is of an 11 kD protein (Sxl-RBD2) in D_2O buffer.

determining structures for proteins greater than 10 kD from 2D COSY, TOCSY, and NOESY can turn into a Herculean task. This was the major driving force behind the development of higher dimensional NMR experiments, especially those which incorporated other spin-1/2 nuclei such as ^{13}C and ^{15}N. The development of methods for specific and uniform incorporation of these nuclei into proteins (McIntosh & Dahlquist, 1990; Hibler et al., 1989) cleared the path for the onset of multidimensional heteronuclear NMR in the late 1980s and early 1990s.

2D heteronuclear NMR

2D HSQC. Given a protein sample with uniform incorporation of ^{15}N (overexpressed in *E. coli* grown on media with $^{15}NH_4Cl$ as the sole nitrogen source), an experiment very similar to the COSY experiment can be done in approximately 30 minutes to 2 hours (~15-fold less time than COSY). This experiment shown in Figure 1.6A is called the HSQC experiment, short for Heteronuclear Single Quantum Coherence (Bodenhausen & Ruben, 1980). The "single quantum" in HSQC has to do with the nature of the ^{15}N magnetization during the evolution period. In fact, this ^{15}N magnetization is completely analogous to ^{1}H magnetization during the COSY, TOCSY, or NOESY t_1 evolution periods. Because the HSQC involves ^{1}H and ^{15}N spins, RF pulses must be applied to each nucleus type separately. This is because ^{1}H and ^{15}N spins precess at very different radio frequencies as a result of their different gyromagnetic ratios. Generally speaking, different nuclei have different gyromagnetic ratios, and therefore different nuclei can usually be manipulated independently. Even though this requires more sophisticated equipment, this turns out to be an incredible advantage over homonuclear NMR since it gives the spectroscopist greater control over heteronuclear spin pairs or triples which are J-coupled (Bax & Grzesiek, 1993). For these reasons, heteronuclear experiments or pulse sequences involving two different nuclei have two sets of pulses, just as in the HSQC in Figure 1.6A.

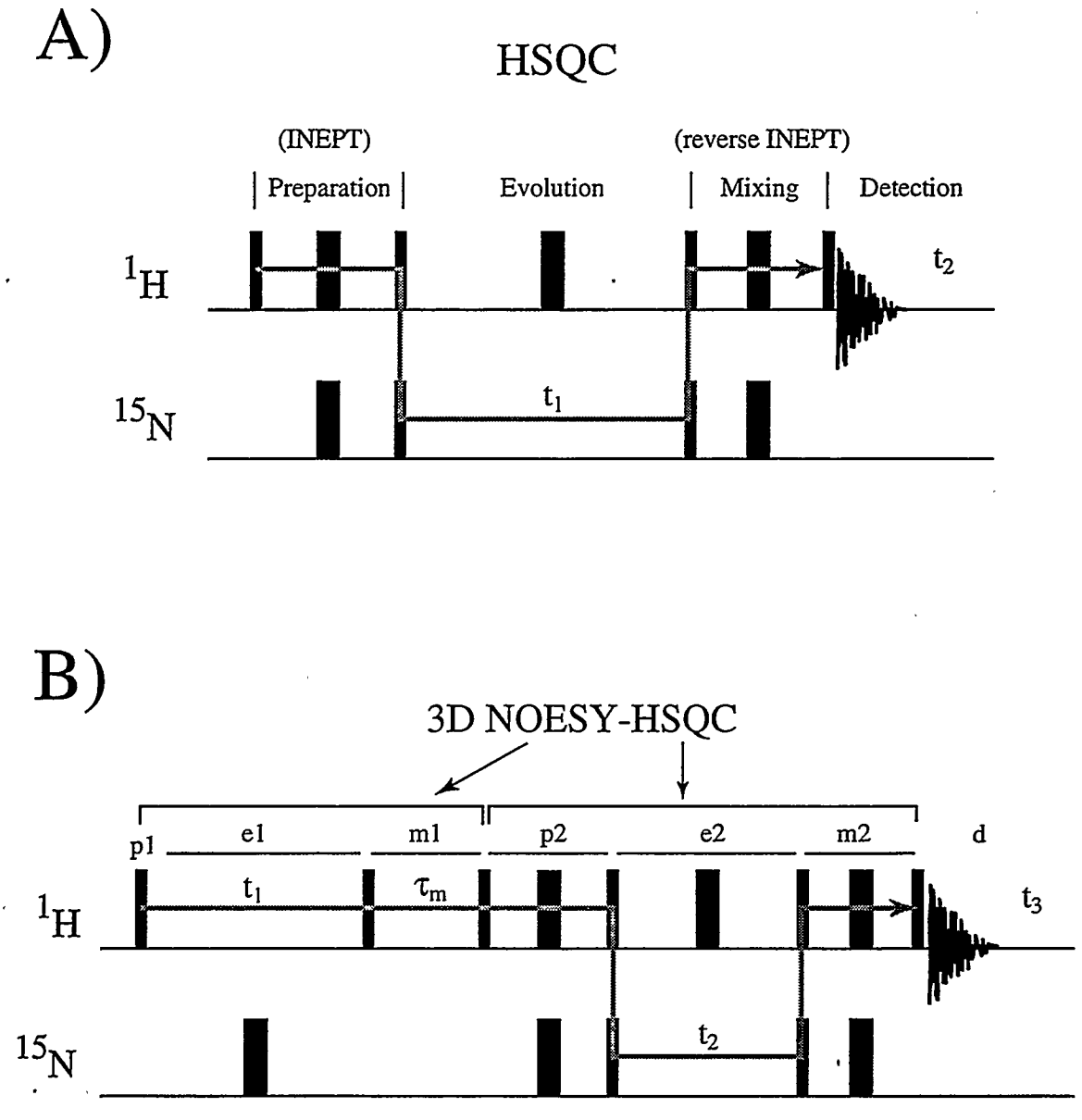


Figure 1.6. A) Pulse sequence for 2D HSQC (Heteronuclear Single Quantum Coherence). The four fundamental elements for a 2D sequence are labeled. An HSQC spectrum for an 11 kD protein (Sxl-RBD2) is shown in Figure 4.1. B) Pulse sequence for 3D NOESY-HSQC. "p1" stands for the first preparation period, "e1" corresponds to the first evolution period, etc.. See Figure 1.6 for a typical "^{15}N plane" from this 3D experiment. In both sequences, thin and thick bars correspond to 90° and 180° pulses, respectively. Gray lines represent the path of magnetization during the course of the experiments.

The gray line in Figure 1.6A represents the magnetization transfer pathway. Initially, coherent magnetization begins on ^1H with the 90° pulse. It is worth observing that in this case, unlike in the homonuclear sequences, the preparation period lasts for more than a single ^1H pulse. During the first few ^1H and ^{15}N pulses, ^1H magnetization is being *prepared* for transfer to ^{15}N. This group of pulses comprising the preparation period is called INEPT (Insensitive Nuclei Enhanced by Polarization Transfer). The INEPT building block is used in nearly all heteronuclear experiments because it has the effect of *increasing* the net amount of magnetization on the ^{15}N nucleus (Sanders & Hunter, 1987). This is why magnetization is generated on ^1H before ^{15}N in the HSQC. Because ^1H is a nucleus with greater sensitivity (greater gyromagnetic ratio) than ^{15}N, and because we wish to record ^1H chemical shifts (as t_2), after recording ^{15}N chemical shift information during t_1, magnetization is transferred back to ^1H during the mixing period with a reverse INEPT set of pulses. Of course, these heteronuclear transfers are possible for amide protons and nitrogens because the ^1H-^{15}N one-bond J-coupling constant is approximately 92 Hz. This is a much stronger coupling than between protons separated by three bonds, and this is the reason why the HSQC is such an efficient experiment. HSQC spectra can be seen for an 11 kD protein in Figure 4.1 and for a 21 kD protein in Figures 8.2A and 8.2B.

The HSQC is a correlation experiment in the same sense as the COSY is; it is a $^1H/^{15}N$ through-bond correlation experiment. In fact, the HSQC is essentially a heteronuclear COSY experiment. The usefulness in the HSQC lies in the dispersion of peaks. There is no diagonal since the two dimensions correspond to different nuclei, and so only true correlation peaks are observed. In proteins, a single peak for each backbone amide proton is generally observed, and therefore the HSQC of a protein is considered to be a spectral "fingerprint" for that particular protein. Because it is so efficient, it can be used for monitoring titrations, screening conditions, or any other purpose that used to be reserved for the 1D experiment based on experiment time considerations. In some well behaved proteins, such as Sxl-RBD2, every peak in the HSQC is resolved. This

"resolving of peaks" is the inherent useful feature of the HSQC, as there is truthfully not very much interesting information in a simple $^{1}H/^{15}N$ correlation spectrum. Therefore, the HSQC element will be used for dispersing peaks in a NOESY experiment, namely a 3D NOESY-HSQC experiment.

ct-HSQC. Before moving on to 3D NMR spectroscopy, I would like to briefly mention another type of HSQC experiment. There is no reason why the basic HSQC sequence cannot be used to generate $^{13}C/^{1}H$ correlation spectra. This can be done by simply pulsing at ^{13}C frequencies instead of ^{15}N frequencies. However, in proteins uniformly labeled in ^{13}C, there are additional 35-40 Hz couplings between ^{13}C nuclei along the sidechains, such as in lysine. These couplings are large enough to show splittings along the ^{13}C dimension ($f1$) in a high quality HSQC, creating unnecessary spectral overlap by making multiplets of all peaks. A method has been developed to collect $^{13}C/^{1}H$ HSQC spectra with this splitting removed. It is called "constant-time" evolution. In a constant-time experiment (Vuister & Bax, 1992), the actual time of the evolution period remains constant (~50 ms). The effect of pure chemical shift evolution emerges at the end of the evolution period, without any effect of ^{13}C-^{13}C J-coupling evolution (splittings). $^{13}C/^{1}H$ correlation spectra of very high resolution can be collected in this manner. An example of a constant-time evolution period can be seen in the lower pulse sequence of Figure 1.8B.

HMQC. Finally, the HMQC (Heteronuclear Multiple Quantum Coherence) (Mueller, 1979) is an alternate method to HSQC to obtain 2D heteronuclear correlation spectra. The details of this experiment are slightly different, but the end result is essentially the same. Therefore, in the following pages, HMQC might be substituted for HSQC and vice versa. This is really of no consequence.

3D heteronuclear NMR

Probably the most important feature of the HSQC is how it resolves protein resonances that would normally be overlapped in a 1D spectrum (Figure 4.1). This "degeneracy"-breaking feature can be combined with any normal ^{1}H homonuclear experiment, such as NOESY, to form a 3D experiment. The pulse sequence for a "^{15}N-separated" 3D NOESY-HSQC (Kay et al., 1989; Marion et al., 1989b) is given in Figure 1.6B. The gray line follows the path of the magnetization. In the 3D sequence, there are now *two* preparation, evolution, and mixing periods, and the three dimensions correspond to t_1 (^{1}H), t_2 (^{15}N), and t_3 (^{1}H). The resultant 3D spectrum contains only NOEs that end on ^{15}NH moieties, regardless of what kind of ^{1}H spin the magnetization may have started on (the spin encoded during t_1). The reason for this isotope "filtration" is that the HSQC part of the sequence selects only for ^{15}N-bound protons. Therefore, not only does a typical heteronuclear 3D spectrum disperse peaks over three dimensions, there are fewer peaks to disperse due to the "filtration". One is then left with a substantially sparser data matrix to analyze, relative to a 2D NOESY. Naturally, one can just as easily do a ^{13}C-separated 3D NOESY-HSQC (or NOESY-HMQC) if one wishes.

Figure 1.7 gives a visual representation of the change in the appearance of NOESY data when moving from 2D NOESY to 3D NOESY-HSQC. Separation of NOESY peaks into "^{15}N planes" clearly allows individual peaks to be fully resolved. If a ^{15}N-separated 3D TOCSY-HSQC is also collected to provide through-bond correlations to backbone amide protons (and ^{15}N), sequence-specific assignments can be obtained using the original strategy (Figure 1.3) proposed using 2D methods. This is essentially the way in which Sxl-RBD2 (~11 kD) was assigned (Figure 4.2).

Triple resonance NMR

The 2D and 3D strategy just described works reasonably well for obtaining protein assignments. One caveat to this approach is that it relies upon through-space NOEs to

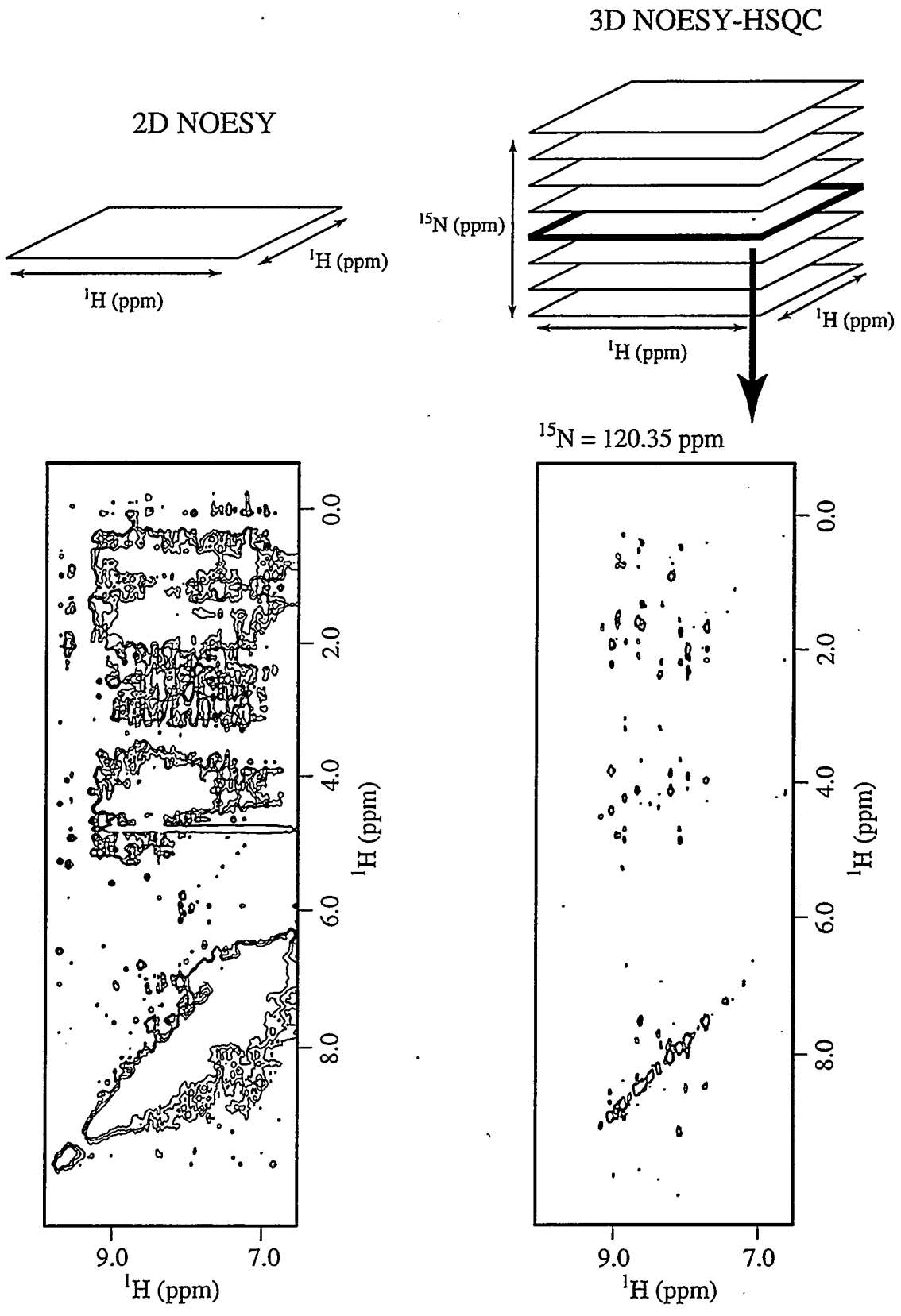


Figure 1.7. Demonstration of the resolution gains from 3D NMR. Both spectra are of a 21 kD protein (Sxl-RBD1+2) bound to RNA (5'-GUUUUUUUUC-3').

sequentially link residues. Unfortunately, the types of sequential NOEs that are utilized (H_α-HN, H_β-HN, HN-HN) are conformationally dependent (Wüthrich, 1986) (Chapter 4). Therefore, in some instances some NOEs may be ambiguous. In the worst case scenario, entire sections of polypeptide could be skipped or "jumped over".

Experiments pioneered by Ad Bax's research group circumvented this problem entirely by implementing a completely different assignment strategy for the polypeptide backbone (Kay et al., 1990; Ikura et al., 1990; Bax & Grzesiek, 1993). The premise for this new strategy was making correlations using *one-bond* J-couplings between ^1H, ^{13}C, and ^{15}N backbone spins so as to bridge sequential residues. The one-bond couplings throughout the backbone and sidechains of polypeptides are generally larger than the three-bond ^1H-^1H J-couplings used in the COSY and TOCSY experiments, and therefore magnetization can be transferred in several sequential steps (and recorded in the intervening evolution periods) before all of the magnetization is lost to relaxation. One can choose to start magnetization just about anywhere on the backbone or sidechain and send the magnetization to nearly any nearby spin, gathering chemical shift information and correlating spins in the process. With the right combinations of these 3D triple resonance experiments (Ikura et al., 1990), the entire backbone and much of the sidechains can be sequentially linked and assigned. One advantage of this approach is that through-space or conformationally dependent correlations are not used at all. A second advantage is that the 3D spectra are extremely sparse in terms of signal density because there are usually only one or two peaks from each residue. For these reasons, the data analysis is very straightforward, and assigning proteins of much larger size (> 30 kD) is tractable. In fact, getting assignments from these data sets is in principle so straightforward that the process could be (and is being) fully automated.

The nomenclature for the names of these experiments reflect the route that the magnetization takes during the course of the experiment. More specifically, the capital letters correspond to the nuclei where chemical shift information is recorded (i.e. evolved

on). If letters are inside parentheses, then the corresponding spin is used only for a transfer step, not for recording chemical shift information. Therefore, in an HNCO experiment, magnetization spends time on amide ^{1}H and ^{15}N spins, and then goes to the carbonyl of the preceding residue (the carbonyl of the same residue is two bonds away from the amide nitrogen). In an HN(CO)CA experiment, magnetization continues beyond the carbonyl to the previous residue's $^{13}C_{\alpha}$, and instead of evolving chemical shift on the carbonyl, $^{13}C_{\alpha}$ chemical shifts are recorded. In this nomenclature, an HSQC experiment would be called an "HN" experiment.

As an example of a set of 3D triple resonance experiments that can be used to assign most of the backbone of a protein, two experiments, HNCACB (Wittekind & Mueller, 1993; Muhandiram & Kay, 1994) and CBCA(CO)NH (Grzesiek & Bax, 1992; Muhandiram & Kay, 1994), are presented in Figure 1.8. These two experiments are complementary. The boxed regions in part A of Figure 1.8 show the path that the magnetization maps out, as do the gray lines in part B. (In HNCACB, the interresidue path found in CBCA(CO)NH also occurs, but with lower efficiency than the intraresidue path. For clarity, this path will be ignored here). Furthermore, the spins for which chemical shifts are recorded during an evolution period (t_1, t_2, t_3) are filled in as black, whereas the spins used only for transfer steps appear as outlined letters. Because the groupings of spins in these two experiments overlap, chemical shifts of the overlapping spins can be matched, and as a result the correlations are extended. For example, given a 3D HNCACB peak (intraresidue correlations) for residue i, a CBCA(CO)NH 3D peak with matching $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts can extend the correlation to the NH of residue i+1. At that point, a HNCACB peak can be found to match the ^{1}H and ^{15}N chemical shifts for residue i+1, and this extends assignments out to $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ of residue i+1. This process is repeated to i+2 and can be extended continuously until either a proline or a C-terminus is encountered. The respective pulse sequences for HNCACB and CBCA(CO)NH are given in part B of Figure 1.8.

Triple Resonance Assignment Strategy

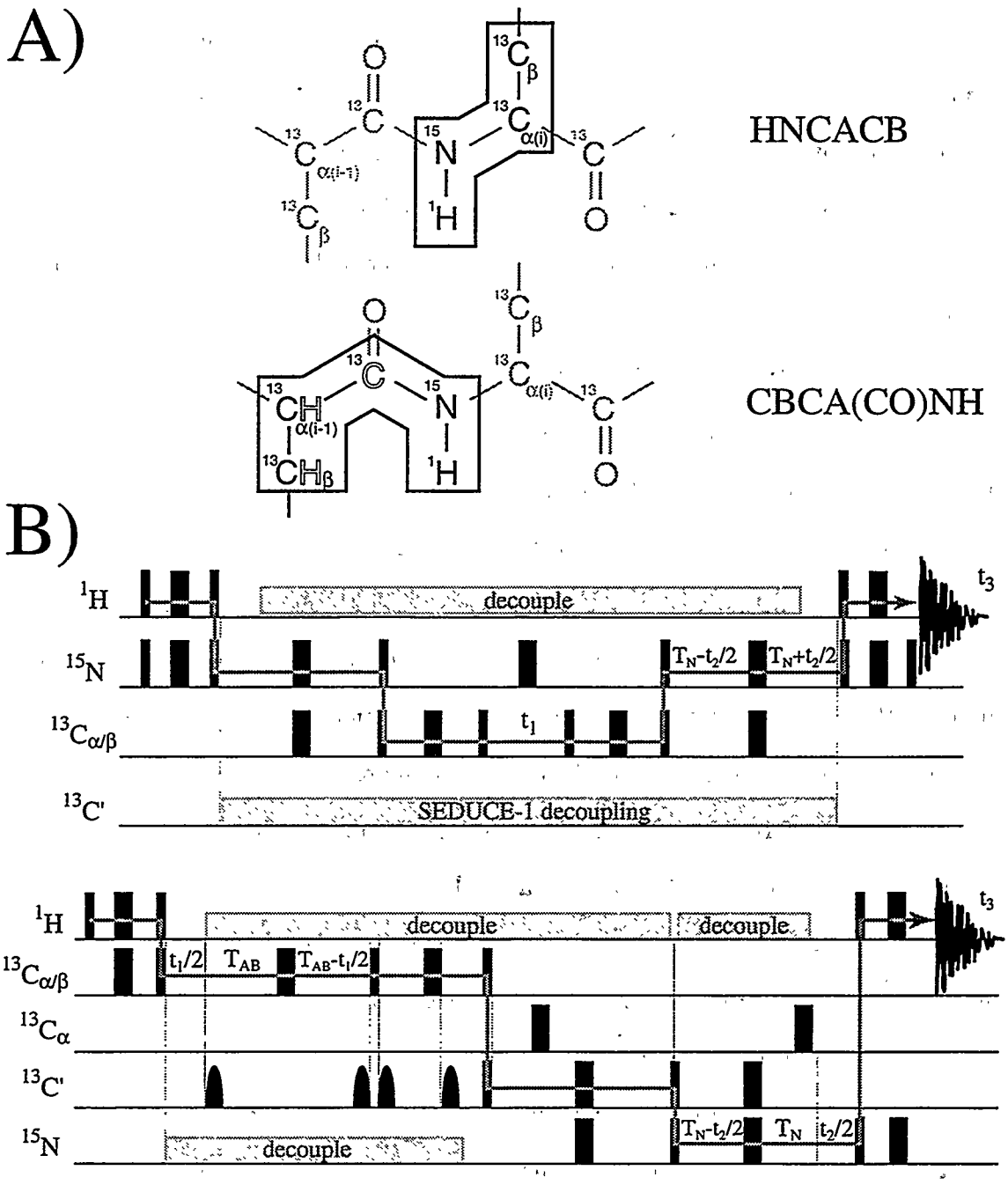


Figure 1.8. A) Schematic representation of the magnetization transfer pathways for HNCACB and CBCA(CO)NH experiments (see note in text on HNCACB additional pathway) Nuclei filled in as black are labeled by chemical shift in these experiments. B) Respective pulse sequences for HNCACB and CBCA(CO)NH. The gray line shows the path of magnetization during the course of the pulse sequences. Thin, thick, and shaped bars correspond to 90°, 180°, and selective 180° pulses, respectively. "T" corresponds to "constant-time" periods (T_N=12 msec).

3D HCCH-TOCSY and 4D HMQC-NOESY-HMQC (HCCH-NOESY)

Two more multidimensional experiments are worth mentioning. The first is the 3D HCCH-TOCSY experiment (Bax et al., 1990). This experiment is widely used to obtain complete ^{1}H and ^{13}C sidechain assignments after the backbone has been assigned via ^{15}N-separated TOCSY/NOESY or triple resonance methods. In the HCCH-TOCSY experiment, protons and carbons of sidechain resonances are correlated in a 3D experiment, where there are two ^{1}H dimensions and one ^{13}C dimension. Entire spin-systems are easily observed in ^{13}C planes since a "^{13}C TOCSY", or isotropic mixing period on ^{13}C, distributes magnetization throughout the carbons in a given sidechain (Figure 4.3). This works spectacularly well compared to the original ^{1}H TOCSY since ^{13}C-^{13}C one-bond couplings are larger, 35-40 Hz. Furthermore, the information in the HCCH-TOCSY is highly redundant since magnetization can start on any proton along the sidechain. Spin-system identification is aided by the characteristic ^{13}C chemical shifts found in amino acids sidechains.

Finally, a very important experiment for structure determination of proteins is a 4D NOESY experiment, in which two extra ^{13}C dimensions are used to disperse the large number of crowded crosspeaks between ^{13}C-bound protons (Figure 1.5). The version of this experiment that we use in the lab is technically called 4D ^{13}C/^{13}C-separated HMQC-NOESY-HMQC (Clore et al., 1991). It is simple in the sense that it is a regular NOESY experiment with extra ^{13}C chemical shift labeling for both starting and destination ^{13}C-bound protons of an NOE. By separating out NOESY peaks into these four dimensions, there is a dramatic reduction in the number of overlapped NOEs. This makes it possible to tabulate many more NOEs (as well as their intensities) between specific protons so that they may be used as qualitative distance restraints in structure calculations. Alternatively a 3D ^{13}C-separated NOESY-HMQC (or NOESY-HSQC) could be collected instead of the 4D experiment. Certainly, higher resolution could be attained in ^{13}C for this 3D experiment, not to mention higher resolution in the indirect ^{1}H dimension. However this experiment

would still have crowded regions in larger proteins since it is only a 3D experiment. The 3D and 4D experiments both have advantages and disadvantages.

Structure calculations based on distance restraints

After complete assignments have been obtained for a protein and higher dimensional NOESY data has been collected, structures that are consistent with the set of interproton distance restraints can be calculated. This can be carried out using two different types of approaches. The first approach, distance geometry, essentially reduces the problem to a geometric one (Havel & Wüthrich, 1985). Groups of atoms and distance "bounds" are processed through use of the triangle inequality, and the coordinates are "embedded" back into Cartesian coordinate space and then refined. The philosophy of the second approach is to 1) ascribe energetic attributes to the molecule including a distance restraint energy function 2) define an energy-based penalty function 3) solve Newton's equations of motion to allow the protein to "fold" into a structure. This is called molecular dynamics/simulated annealing (MDSA) (Brunger & Karplus, 1991). It is also possible to create a hybrid structure calculation protocol that involves both distance geometry and MDSA (Nilges et al., 1988), and it turns out that this method is quite robust. This is the method that Fred Damberger, Jeff Pelton, Brian Volkman, Mike Nohaile, and I have used to calculate families of protein structures based on NMR-derived distance restraints. More details about the structure calculations on Sxl-RBD2 will be given in Chapter 4.

Chemical shifts—what can they tell us?

As a final note in this introduction to protein NMR, I would like to acknowledge that there have been significant advances in understanding chemical shifts in the last 5-10 years (Wishart et al., 1991; Case & Osapay, 1994; Wishart et al., 1995; Oldfield, 1995). Because chemical shifts are exquisitely sensitive to the surrounding chemical environment, they contain valuable structural (and dynamic) information. This information has greater

accuracy than that typically obtained from the NOE, and therefore much research has been devoted to understanding the underlying mechanisms which give rise to chemical shift dispersion. The goal of these forward-looking researchers is to make possible structure determination from chemical shift data alone.

Already, chemical shifts are widely being used as a qualitative and quantitative means of determining secondary structure in proteins. The information is often used as the difference between a nucleus's actual chemical shift and it's "random coil" chemical shift (Wishart et al., 1995). $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$, and $^{1}H_{\alpha}$ "secondary shifts" in are now routinely used as a reliable method for determining secondary structure type in peptides and proteins (Wishart & Sykes, 1994). This approach has been employed here for Sxl-RBD2 (Chapter 4) and the complex of Sxl-RBD1+2 with a 10-mer RNA (Chapter 8). As an initial step with regard to overall structure deterimination, Clore and coworkers (Kuszewski et al., 1995a; Kuszewski et al., 1995b) as well as Montelione and coworkers (Celda et al., 1995) have begun to include ^{13}C and/or ^{1}H chemical shifts into the structure refinement process in protein structure calculations. Time will tell how rigorously we can use chemical shift information.

Concluding remarks

The content of this chapter is the result of an awesome series of advances in NMR over the last 20 years or so. I have only attempted to communicate a general appreciation for the NMR concepts and how the data can be used for obtaining assignments and calculating structures. I certainly do not want to trivialize the methods described, for they are by no means trivial. I have left out many of the details which are so important in pulse sequences, for example. More of these details can be found in Appendix D as well as throughout the chapters.

Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229-2246.

Bax, A., Clore, G. M. & Gronenborn, A. M. (1990) *J. Magn. Reson. 88*, 425-431.

Bax, A. & Grzesiek, S. (1993) *Acc. Chem. Res. 26*, 131-138.

Blum, K. (1981) in Density Matrix Theory ans Its Applications, Plenum Press, New York.

Bodenhausen, G. & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185-189.

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

Brunger, A. T. & Karplus, M. (1991) *Acc. Chem. Res. 24*, 54-61.

Case, D. A. & Osapay, K. (1994) *J. Biomol. NMR 4*, 215-230.

Celda, B., C., B., M.J., A., R., T. & Montelione, G. T. (1995) *J. Biomol. NMR 5*, 161-172.

Clore, G. M. & Gronenborn, A. M. (1991) *Science 252*, 1390-1399.

Clore, G. M., Kay, L. E., Bax, A. & Gronenborn, A. M. (1991) *Biochemistry 30*, 12-18.

Ernst, R. R., Bodenhausen, G. & Wokaun, A. (1987) in Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Oxford University Press, New York.

Goldman, M. (1988) in Quantum Description of High Resolution NMR in Liquids, Clarendon Press, New York.

Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988) *J. Magn. Reson. 110*, 7870-7820.

Grzesiek, S. & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291-6293.

Harris, R. K. (1986) in Nuclear Magnetic Resonance Spectroscopy, Longman Scientific & Technical, New York.

Havel, T. F. & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281-294.

Hibler, D. W., Harpold, L., Acqua, M. D., Pourmotadbbed, T., Gerlt, J. A., Wilde, J. A. & Bolton, P. H. (1989) *Methods Enzymol. 177*, 74-87.

Ikura, M., Kay, L. E. & Bax, A. (1990) *Biochemistry 29*, 4659-4667.

Kay, L. E., Marion, D. & Bax, A. (1989) *J. Magn. Reson. 84*, 72-84.

Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990) *J. Magn. Reson. 89*, 496-514.

Kumar, A., Ernst, R. R. & Wüthrich, K. (1980) *Biochem. Biophys. Res. Com. 95*, 1-6.

Kuszewski, J., Qin, J., Clore, G. M. & Gronenborn, A. M. (1995a) *Journal of Magnetic Resonance (Series B) 106*, 92-96.

Kuszewski, J., Gronenborn, A. M. & Clore, G. M. (1995b) *Journal of Magnetic Resonance (Series B) 107*, 293-297.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989b) *J. Am. Chem. Soc. 111*, 1515-1517.

McIntosh, L. P. & Dahlquist, F. W. (1990) *Q. Rev. Biophys. 23*, 1-38.

Mueller, L. (1979) *J. Am. Chem. Soc. 101*, 4481-4484.

Muhandiram, D. R. & Kay, L. E. (1994) *Journal of Magnetic Resonance (B) 103*, 203-216.

Neuhaus, D. & Williamson, M. P. (1989) in The Nuclear Overhauser Effect in Structural and Conformational Analysis, VCH Publishers, Inc., New York.

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

Oldfield, E. (1995) *J. Biomol. NMR 5*, 217-225.

Sanders, J. K. M. & Hunter, B. K. (1987) in Modern NMR Spectroscopy - A Guide for Chemists, Oxford University Press, Oxford.

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

Sorenson, O. W., Eich, G. W., Levitt, M. H., Bodenhausen, G. & Ernst, R. R. (1983) *Progress in NMR Spectroscopy 16*, 163-192.

Vuister, G. W. & Bax, A. (1992) *J. Magn. Reson. 98*, 428-435.

Wemmer, D. E. (1989) *Concepts in Magnetic Resonance 1*, 59-72.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295-315.

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

Wishart, D. S. & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171-180.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67-81.

Wittekind, M. & Mueller, L. (1993) *J. Magn. Reson. 101*, 201-205.

Wüthrich, K. (1986) in NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York.

Chapter 2

Biophysical Characterization of an Activation Domain From the Eukaryotic Transcription Factor c-jun

Eukaryotic transcription factors & activation domains

Transcription in eukaryotes is initiated and regulated by a complex array of proteins which assemble at specific gene promoters (Conaway & Conaway, 1993; Zawel & Reinberg, 1993). Ultimately, transcription is catalyzed by RNA polymerase II, a multi-subunit enzyme which is the target of these transcription regulatory proteins or "transcription factors". One of the major goals of research in this area is to reveal the three-dimensional architecture of the transcription initiation nucleoprotein complex to the point where critical interactions between regulatory proteins and their effects on RNA polymerase II (pol II) can be understood at a mechanistic level.

In order to help simplify the complicated problem of the combined effect of more than 20 proteins (Buratowski, 1994), these proteins involved in transcriptional regulation have been categorized according to their role in the initiation process. Before describing these categories, it should be mentioned that a minimal collection of proteins can be assembled at the core promoter to drive transcription. Transcriptional activity from this minimal set of proteins is known as basal transcription (Conaway & Conaway, 1993; Buratowski, 1994). Basal-type transcription is constitutive at low levels of activity. It is therefore important to make a distinction between basal or "necessary" transcription factors and the "extra" proteins which confer a higher degree of control through more sophisticated interactions.

In higher eukaryotes, the general transcription factors that comprise the basal machinery consist of the TATA-binding protein (TBP) and factors TFIIB, TFIIF, TFIIE,

and TFIIH. TBP is one of many factors which comprise the TFIID activity, and it is essential for recognizing the TATA element. TFIID has always been recognized as a central player in the initiation machinery, although it turns out that this collection of proteins (TAFs to be discussed below), aside from TBP, is not required for basal transcription. TBP is the first transcription factor to assemble at the TATA site of the promotor, located approximately at position -25 to -30 , and the three-dimensional structure of this protein dimer bound to DNA has been determined (Kim et al., 1993). Binding of TBP induces a dramatic bending in the DNA which presumably helps to facilitate the assembly of more general factors. Next, TFIIB binds to the TBP/DNA complex, which allows for further recruitment of factors TFIIF, TFIIE, TFIIH, and RNA polII itself (Buratowski, 1994). The ternary TBP/TFIIB/DNA(TATA) complex has also been solved by X-ray crystallographic methods (Nikolov et al., 1995). Although TBP and TFIIB are small, single proteins, most of the general factors are multi-protein complexes. In the end, a mass of ~20 proteins assembled together at the core promoter comprises the general transcription machinery which is capable of non-activated or basal transcription.

This categorization of proteins involved in transcription provides a relatively simple model for constitutive initiation of transcription. However, because of the cell's requirement for precise and responsive control of gene expression for countless sets of genes there is a need for high variability in the level of transcriptional activation. Nature has largely solved this problem by adding another level of complexity to how transcription can be activated. Transactivating transcription factors are proteins that bind sequence specifically to upstream enhancer sequences on the DNA. These "activators" have the ability to interact with the general transcriptional machinery at the promoter and modulate the rate of RNA polymerase II transcription (Tjian & Maniatis, 1994). In the context of activation from distal enhancers, TFIID consists of TBP plus at least 8 other TAF (TBP Associated Factors) proteins. When TFIID replete with TAFs is assembled at the core promoter, the general transcription machinery is no longer a basal apparatus. It is a higher

order initiation complex capable of greater responsiveness due to interactions with distal enhancers through interactions with coactivators (TAFs) or other general factors. It has been demonstrated that the upstream activators often interact with specific TAFs or coactivators in a way which triggers increased levels of transcription (Pugh & Tjian, 1990; Hoey et al., 1993; Goodrich et al., 1993; Gill et al., 1994), although the mechanisms for such processes are not yet understood. It appears that there are many novel, gene-specific mechanisms for activators or sets of activators to modulate transcription levels. With these activators, the cell can produce varying amounts of gene products as is required, and with a wide variety of activator proteins and combinations thereof, the cell has fine, responsive control of transcription levels. This degree of precision has come at the expense of complexity, so unraveling the mechanisms of these proteins has become a difficult problem indeed. Nevertheless, the importance of this problem is appreciated and has thus attracted much attention to all facets of the "puzzle".

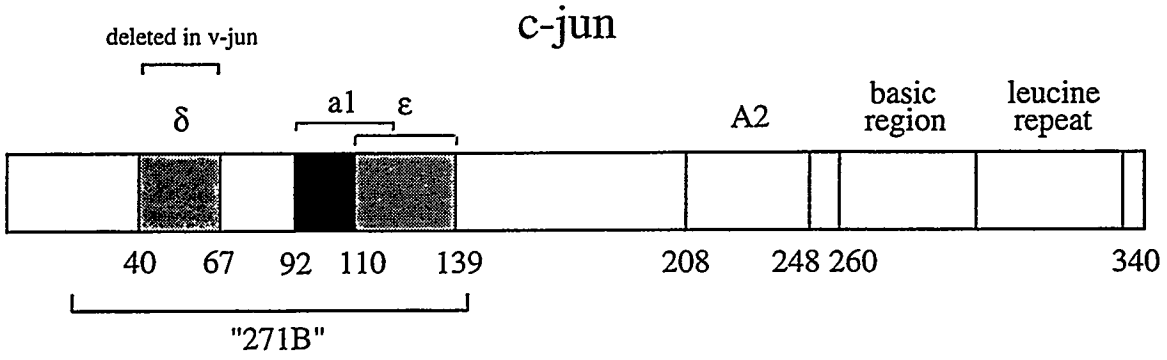
In the last 5-10 years, much attention has been payed to the upstream activators since they are varied in nature and modular in structure (Mitchell & Tjian, 1989; Tjian & Maniatis, 1994). All of these proteins contain a domain which allows them to bind to DNA in a sequence-specific manner. In many cases, this activity is combined with an ability to homo or heterodimerize. They must by definition also have at least one "activation domain". Through deletion analysis it has been demonstrated that these domains are absolutely required for modulating transcriptional activity (Kadonaga et al., 1988; Bohmann & Tjian, 1989). Therefore, the DNA and dimerization domains are required to bind the enhancer sequences; once the factor is on the DNA, the activation domains transactivate transcription through interactions with TAFs or other components of the general transcription machinery. Researchers have found activation domains to be particularly intriguing because a) unlike DNA-binding or multimerization domains, activation domain function has not yet been understood mechanistically b) activation domains often have unusual amino acid sequence properties such as clustering of proline,

glutamine, or acidic residues (Mitchell & Tjian, 1989). Curious as these properties are, they have not lent any insight into how transcriptional activity is boosted, and thus despite many efforts, activation domains remain enigmatic.

c-jun (AP-1)

The AP-1 family of regulatory proteins are involved in the earliest nuclear response of cells stimulated by growth-promoting agents and mitogens (Curran & Franza, 1988). Two members of this family are the proto-oncogene products c-fos and c-jun. Because of their early identification and subsequent cloning and purification (Bohmann et al., 1987), they among others (Sp1) have served as prototypical transcription factors for more controlled biochemical and structural studies (O'Shea et al., 1991; O'Shea et al., 1992).

The gene product of the human proto-oncogene *c-jun* differs from the gene product of the oncogene *v-jun* by a stretch of 27 amino acids near the N-terminus of c-jun (Bohmann et al., 1987). This 27 residue difference between these two gene products leads to v-jun having a much greater enhancement of transcription compared to c-jun. Individual, independently folded domains within these proteins have been mapped by making deletion mutants followed by activity assays (Bohmann & Tjian, 1989; Baichwal & Tjian, 1990; Baichwal et al., 1992). This domain map is presented in Figure 2.1. Both c-jun and v-jun contain a leucine-rich dimerization domain at the C-terminus and an adjacent domain rich in basic amino acids. Just as this type of DNA-binding domain in GCN-4 recognizes the core sequence TGACTCA, the jun "B-Zip" DNA-binding motif recognizes the DNA sequence ATGACTCTT in a sequence-specific manner. Two separate activation domains, a1 and A2, are located N-terminal to the DNA-binding domain. The a1 activation domain is flanked by two functionally separable regulatory subdomains, δ and ε. Analysis of transcriptional activation in different cell types has implicated the δ and ε domains as a cell type specific inhibitor-binding region which serves to attenuate transcriptional activities in different cell types (Baichwal & Tjian, 1990; Baichwal et al., 1992). It is the δ domain



271B

M E T T F Y D D A L N A S F L P S E S G P Y G Y S N P K I L K Q S M

"δ"

T L N L A D P V G S L K P H L R A K N S D L L T S P D V G L L K L A

"a1"

S P E L E R L I I Q S S N G H I T T T P T P T Q F L C P K N V T D E

"ε"

Q E G F A E G F V R A L A E L H S Q N T L P V P P S L D

Figure 2.1. Domain map of c-jun and amino acid sequence of the 271B construct.

that v-jun lacks which accounts for the 27 residue difference and greater transcriptional activation properties of v-jun over c-jun. Neither the a1 or A2 activation domain of c-jun falls into one of the above-mentioned categories (acidic, Q-rich, P-rich).

Due to the relatively small size of the δ-a1-ε series of domains as well as the novel features of the δ and ε sub-domains, this fragment was sub-cloned into an expression vector and overexpressed in *E. coli* for the purposes of biophysical studies. It was hoped that structural properties inherent in these domains would lead to an increased understanding in transactivated transciptional activity of RNA polymerase II.

Expression and purification of c-jun activation domain

Expression. From the recombinant *c-jun* used in the Tjian laboratory, two constructs corresponding to the δ-a1-ε activation domain were sub-cloned into overexpression vectors. Originally, the construct 271 was sub-cloned by Vijay Baichwal. 271 encodes for the δ-a1-ε activation domain with a 6-histidine tail at the N-terminus (Figure 2.1) and is under control of a T7 promoter (Studier et al., 1990). 271 was overexpressed in the BL21-(DE3) strain of *E. coli* at a relatively high level of ~30-60 mg/liter. After much work with the 271 protein fragment it was decided that it formed aggregates too easily for NMR work. Because the protein expressed in inclusion bodies and the hisitine tag was not employed as a purification scheme, the histidines were superfluous. The construct 271B was re-cloned by Ho Cho and Tae Ahn (undergraduate in the Wemmer group) with the histidine tag removed, yielding a protein with seven fewer residues (Figure 2.1). In BL21-(DE3) cells, 271B overexpressed at an approximately 3-fold lower level. However, the purified protein had superior solution characteristics compared to 271. By CD, the extent of helix in the two constructs was indistinguishable, and 271B is more soluble than 271. Therefore, all subsequent work (including all data below) was done with 271B.

Purification. The following purification protocol has been slightly modified from Dirk Bohman's (Robert Tjian's lab) original inclusion body purification protocol and utilizes the following solutions:

A
10 mM Tris-HCl (pH 7.9)
25% sucrose
100 mM KCl
10 mM DTT
2 mM PMSF
2 mM Na-metabisulfite

B
300 mM Tris-HCl (pH 7.9)
100 mM EDTA
4 mg/ml lysozyme

C
1 M LiCl
20 mM EDTA

D
10 mM Tris-HCl (pH 7.9)
0.1 mM EDTA
0.5 M LiCl
5 mM DTT
1 mM PMSF
1 mM Na-metabisulfite

E
10 mM Tris-HCl (pH 7.9)
0.1 mM EDTA
5 mM DTT
1 mM PMSF
1 mM Na-metabisulfite

* add PMSF, lysozyme, and DTT just prior to use
* 1/2 teaspoon methionine per 1 liter solution A, B, C, D, and E should be use to prevent oxidation of cystines or methionines in the protein.

271B was overexpressed in the BL21(DE3) strain of *E. coli* at 37 °C. Typically, 1 liter of cells were grown to OD_{595}=0.6-0.7 at which point 30 ug/ml IPTG was added to induce T7-controlled expression. The cells were placed on the shaker for an additional 2-4 hours. Both 271 and 271B express in the insoluble fraction and must be purified as inclusion bodies. The cells were spun for 10 minutes at 4,500 rpm in 500 mL centrifugation bottles (GS-3) to generate 2 pellets. The pellets were resuspended in 18 ml of solution A per 250 ml of culture and transferred into four polyethylene SS-34 tubes (50

ml tubes are the best, since spillage can occur during sonication steps). 4 ml of solution B was added to each tube and incubated for 10 minutes on ice. 20 ml of solution C was then added to each tube and vortexed for a few seconds to allow for mixing. The bacterial suspension of cells was lysed using the microtip of a Branson sonicator. Three pulses of 10 seconds at setting 5 was given under ice-cooling. At the end of sonication, the pellets should be completely dispersed and off the sides of the tube. Care must be exercised at this step since improper placement of the tip can cause the solution to "jump" out of the SS-34 tube. The inclusion bodies containing the protein were pelleted by 10 minutes of centrifugation at 13,000 rpm in a SS-34 rotor at 4 °C. To wash the pellet, 25 ml of solution D was added, the pellet was broken up by a short sonication pulse (or a few short pulses of sonication, if necessary) and the desired insoluble material was recovered by a 10 minute SS-34 spin at 13,000 rpm. This washing step was repeated once with solution D and twice with solution E. The remaining pellet could be stored indefinitely at -20 °C or -70 °C.

The purified inclusion body pellet was then solubilized and HPLC-purified. To solubilize the pellet, about 1.5-2.0 ml of 50:50 H_2O/Buffer B (given below) was added to the pellet in the SS-34 tube. Mixing with a magnetic stir bar at room temperature overnight would disperse the pellet and solubilize the majority of it. Usually a few flakes of methionine was added to insure that the protein remained fully reduced. The next morning, this cloudy concoction was spun at 10,000 rpm for 10 minutes. Just prior to the HPLC injection, the supernatent was filtered through a 0.45 micron filter. HPLC uses three buffers, A, B, and C:

A: 0.1% TFA in H_2O
B: 0.1% TFA
60% acetonitrile
40% H_2O
C: 100% acetonitrile

The column used was a preparative C-18 Waters Delta-Pak column. The gradient employed was:

time (min)	%A	%B	%C
0	50	50	0
2	50	50	0
45	14	86	0
55	0	0	100%

flow rate = 12 ml/min
wavelength = 220 nm
AUFS = 2.0

A sample HPLC trace using this gradient is shown in Figure 2.2. Usually, about 1.3 ml of the filtered solution of 271B was injected. 271B elutes at approximately 28 minutes. Fractions were collected and lyophilized to yield the fully purified protein. For NMR samples, the lyophilized powder was dissolved directly into the appropriate buffer. Previous studies on 271 (minus His-tag) demonstrated that refolding from denaturants did not produce protein with a larger extent of structure. This was not attempted with 271B.

Mass spectrometric analysis (thank you once again, David King) of NMR samples of 271B showed that dimerization would occur at high concentrations via a disulfide bridge at C95, the only cystine in 271B. This dimer was shown to have the same percentage of helicity as the monomer. In order to prevent this dimerization, the cystine sulfhydryl was alkylated with iodoacetamide as the final step of the protein prep. Typically, 5.0 mg of 271B was reacted with 20 mg iodoacetamide in ~45 ml of 100 mM tris pH=8 for ~45 minutes - 1 hour. The protein in this reaction mixture was finally dialyzed against H_2O with enough TFA to buffer at pH=2.

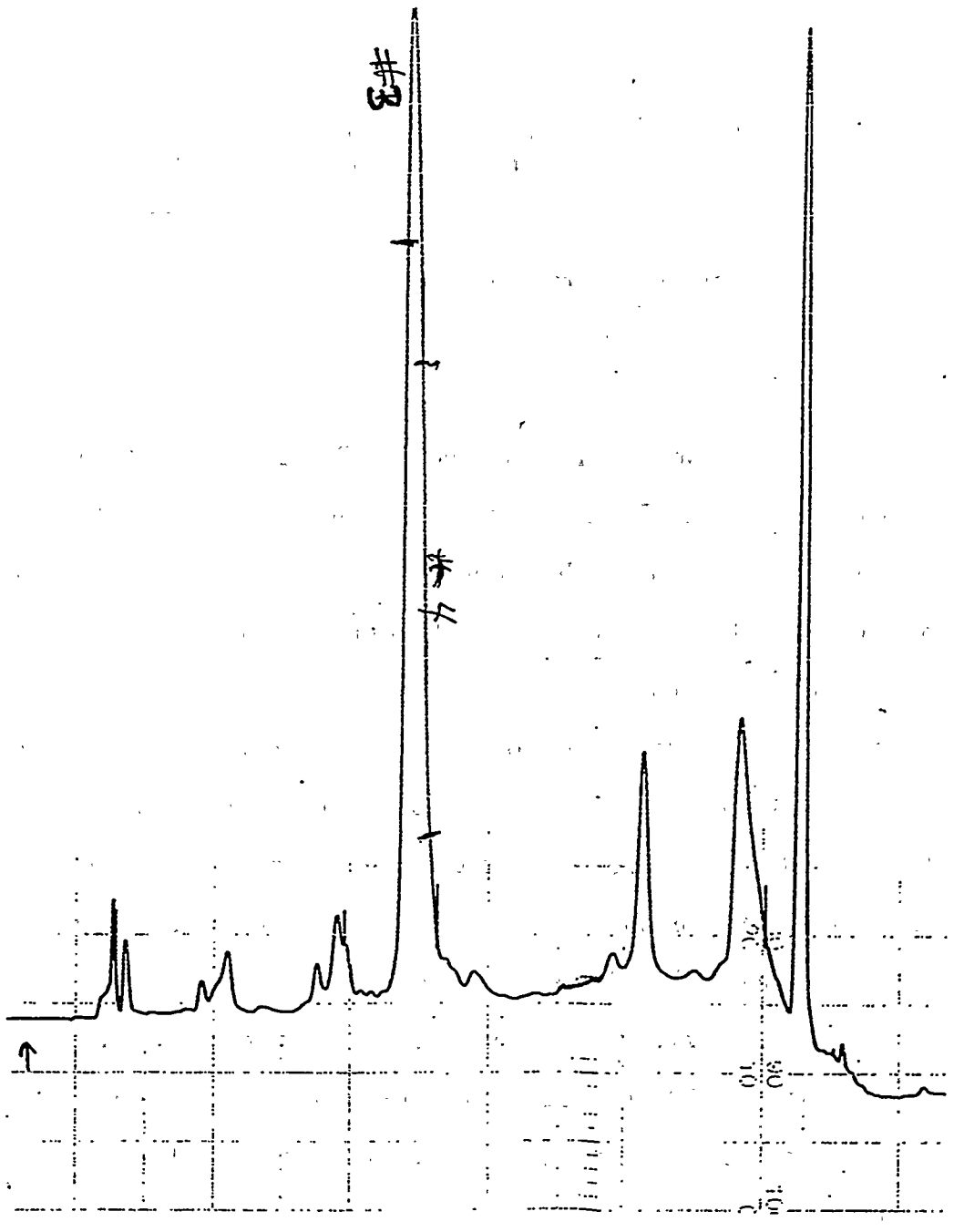


Figure 2.2. HPLC trace of the 271B purification over a reversed-phase C-18 preparative column. Approximately 7 mg was loaded for the injection shown. The run proceeds left to right. 271B elutes at the 28 minute mark, and the fractions collected are shown.

Biophysical characterization

Circular dichroism. Circular dicroism (CD) is a useful technique for determining the type and approximate extent of secondary structure in polypeptides (Cantor & Schimmel, 1980). It is convenient because both measurement and interpretation are fast, and only a small amount of material (~100 ug of 271B, for example) is required. Ellipticity (proportional to helicity), θ, as determined by CD was used as a method for optimizing conditions for NMR studies. Ellipticity is related to differential absorption of circularly polarized light (given in mdeg) through the following expression:

$$[\theta] = \frac{(m\,\mathrm{deg})(100)}{(nres)(mM)(cm)}$$

where nres is the number of amino acids in the polypeptide, mM is the polypeptide concentration, and cm is the cell path length. In the literature, θ is usually given in deg/(cm^2 dmole) with θ_{222}=34,000 corresponding to 100% helix. Ellipticity was measured as a function of pH, % organic solvent miscible in water, ionic strength, temperature, and chemical denaturant (urea). All CD spectra were collected at room temperature on a Jasco J-600 spectropolarimeter with 271B concentrations in the 5-10 uM range.

Ellipticity is defined as the differential absorption between left- and right-circularly polarized light. For polypeptides, ellipticity is typically measured in the 190-300 nm range. Random coil polypeptides have characteristic ellipticities at 208 nm. Helices and beta-sheets have characteristic ellipticities at 222 and 230 nm, respectively. Circular dichroism cannot differentiate between 3_{10} and α-helices nor between parallel and anti-parallel β-sheets. Aromatic residues such as Phe and Tyr can exhibit non-zero ellipticities in the near-UV range if their sidechains are packed into the hydrophobic interior of a globular protein. Such CD signals are therefore indicative of stable tertiary structure. For the case of 271B, conditions were sought which yielded a relatively large degree of secondary structure, i.e.

helicity in 271B. There is no evidence for the presence of "extended" or β-structure in the δ-a1-ε domain of c-jun, and there is also no evidence for the presence of stable tertiary structure.

Under low-salt, aqueous conditions, 271B is approximately 10-15% helical, determined from θ_{222}, at 4 different pHs ranging from 2 to 6.5. Inspection of Figure 2.3 reveals that the amount of helicity at these different pHs is essentially invariant. Nevertheless, conditions with variable amounts of organic solvents or ions were explored at pH=2, 3.0, 4.5 and 6.5 in order to find any conditions which might confer stability to the residual secondary structure in 271B.

Figure 2.4 shows CD spectra of 271B at pH=2 or 3.0 (i.e. low pH) without and with 20% (by volume) organic solvents. Examination of these titration spectra for helicity content should take into account the dilution of the protein solution upon sequential 20% volume additions of methanol, ethanol, trifluoroethanol (TFE), 1-propanol, 2-propanol, or acetonitrile (spectra A-E). Qualitatively, this can be accomplished by calculating $\theta_{222}/\theta_{208}$ (read as "mdeg" off of the ordinate) to factor out the dilution. At low pH, all organic solvents increased the amount of helicity in 271B, and the greatest increases are observed in 1-propanol and TFE (data not shown, see pH=6.5 data for a comparable spectrum). TFE is well-known for its helix-stabilizing properties (Nelson & Kallenbach, 1986; Nelson & Kallenbach, 1989), although its mechanism is not understood, but the 5 other organic solvents examined here are typically not used as helix-stabilizing agents. For all mixed-phase solutions up to 30% in organic composition, increasing the % of organic solvent increases the magnitude of θ_{222}, or helicity by CD. At 20% 1-propanol, for example, helicity is calculated to be approximately 20%. At 20% acetonitrile, helicity is calculated to be approximately 16%.

Figure 2.5 shows more titrations of 271B at pH=6.5, 50 mM phosphate buffer with organic solvents in a manner similar to the low-pH studies. The results are quite similar, as can be seen clearly in a comparison of the 1-propanol or methanol titrations at

Figure Legends for Figures 2.3-2.5

Figure 2.3. Circular dichroism (CD) spectra of 271B (5-10 uM) at four pH values, all with no salt. Each spectrum is scanned from 260 to 200 nm. A) pH=2 was buffered by trace levels of trifluoroacetic acid (TFA). B) pH=3 was buffered with 50 mM phosphate buffer. C) pH=4.5 was buffered with 30 mM acetate buffer. D) pH=6.5 was buffered with 50 mM phosphate buffer. Spectrum "E" is a superposition of the pH=2 spectrum (A) and the pH=6.5 spectrum (D). All spectra were recorded on a Jasco J-600 spectropolarimeter at room temperature.

Figure 2.4. Circular dichroism (CD) spectra of 271B (5-10 uM) at various mixed-phase solvent compositions, low pH (= 2 or 3, buffered with 50 mM phosphate), and no salt. Organic solvents were added to "0%" solutions to yield the "20%" spectra. A) +/- 20% methanol, pH=2. B) +/- 20% ethanol, pH=2. C) +/- 20% 1-propanol, pH=2. D) +/- 2-propanol, pH=2. E) 0% acetonitrile spectrum is at pH=2. 20% acetonitrile spectrum is closer to pH=3. All spectra were recorded on a Jasco J-600 spectropolarimeter at room temperature.

Figure 2.5. Circular dichroism (CD) spectra of 271B (5-10 uM) at various mixed-phase solvent compositions, pH=6.5-7.0 (buffered with 50 mM phosphate), and no salt. Organic solvents were added to "0%" solutions in 10% volume increments. In all titration spectra shown, the spectrum with lowest absorbance (small negative) at θ_{222} corresponds to the start point of the titration, and the spectrum with greatest absorbance (large negative) at θ_{222} corresponds to the end point of the titration. A) Methanol series, pH=6.5. B) Ethanol series, pH=6.5. C) +/- 10% trifluoroethanol (TFE), pH=7.0. D) 1-propanol series, pH=6.5. E) Acetonitrile series, pH=6.5.

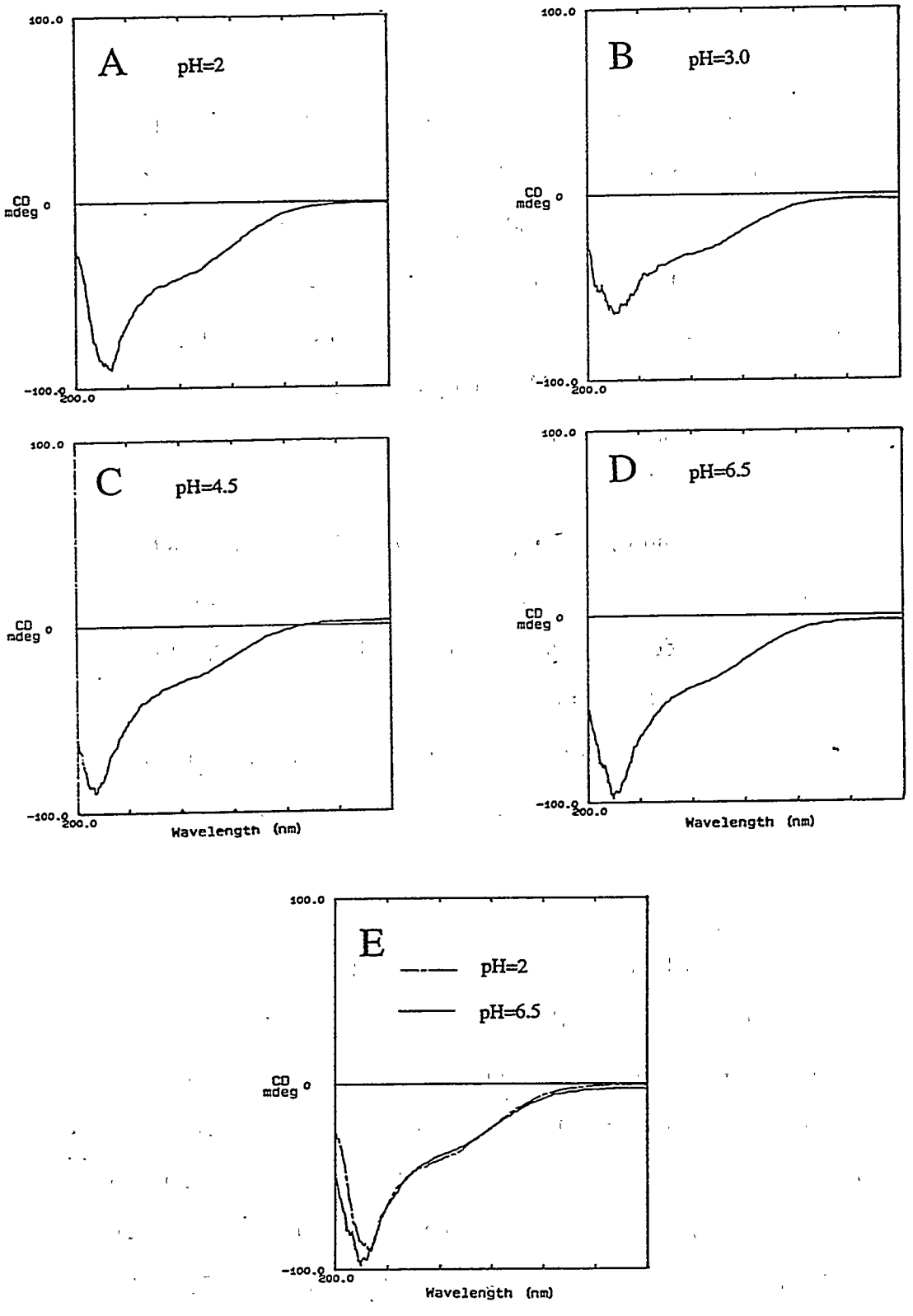

A
pH=2
B
pH=3.0
C
pH=4.5
D
pH=6.5
E
pH=2
pH=6.5
100.0
-100.0
200.0
CD
mdeg
0
Wavelength (nm)

Figure 2.3

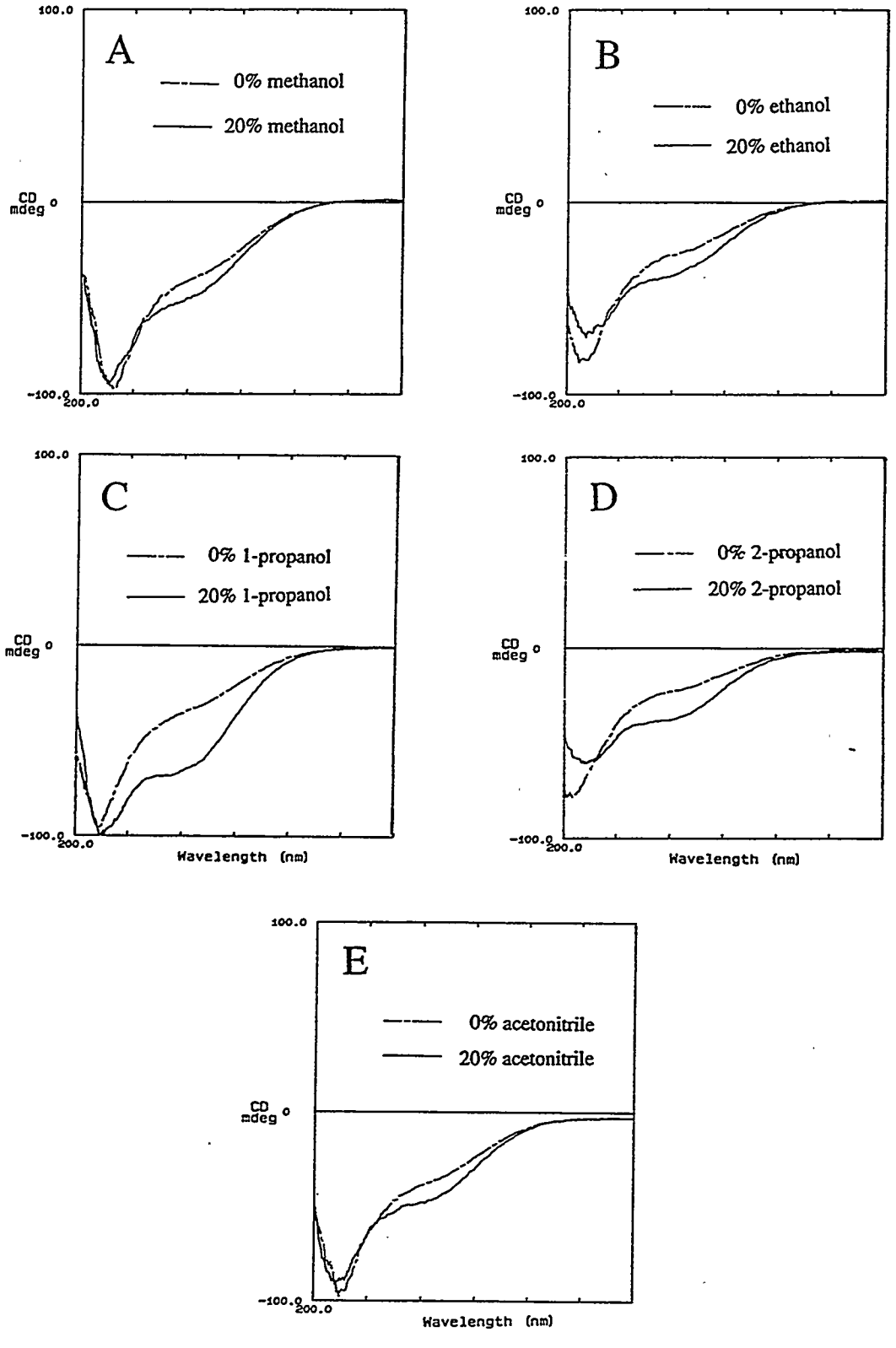

A
0% methanol
20% methanol
B
0% ethanol
20% ethanol
C
0% 1-propanol
20% 1-propanol
D
0% 2-propanol
20% 2-propanol
E
0% acetonitrile
20% acetonitrile
CD mdeg
100.0
0
-100.0
200.0
Wavelength (nm)

Figure 2.4

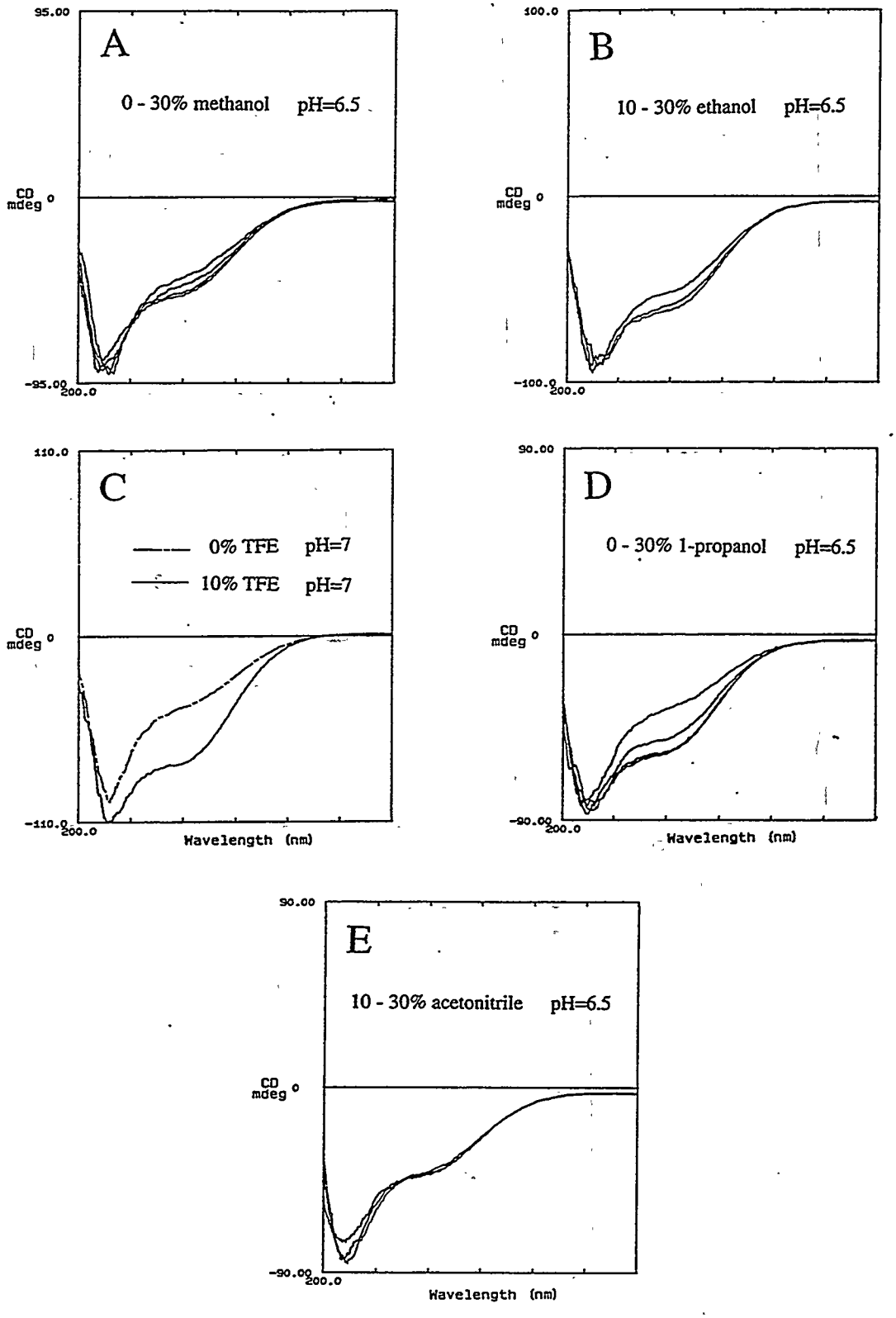

A
0 - 30% methanol pH=6.5
95.00
CD mdeg 0
-95.00
200.0
B
10 - 30% ethanol pH=6.5
100.0
CD mdeg 0
-100.0
200.0
C
0% TFE pH=7
10% TFE pH=7
110.0
CD mdeg 0
-110.0
200.0
Wavelength (nm)
D
0 - 30% 1-propanol pH=6.5
90.00
CD mdeg 0
-90.00
200.0
Wavelength (nm)
E
10 - 30% acetonitrile pH=6.5
90.00
CD mdeg 0
-90.00
200.0
Wavelength (nm)

Figure 2.5

pH=2 and 6.5. In the far-UV region (around 200 nm), there are some subtle differences in ellipticities between pH=2 and 6.5 as can be seen in the methanol and 1-propanol series, but such effects are difficult to interpret as absorbances in this region are less easily characterized and the region below 200 nm tends to give less accurate readings due to lamp performance/stability in addition to absorption by contaminating air gases. Also, it is possible that slight differences in sample preparation could lead to such subtle changes in this region of the CD spectra. Overall, it appears that by CD, 271B behaves quite similarly at pH=2-3 and pH=6.5. Thus, the organic solvent-induced structural changes observed, as well as the residual helical structure in purely aqueous conditions, are not likely to be mediated by ionic interactions involving titrateable groups such as Glu, Asp, and His sidechains.

In addition to various mixed-phase solvents, it was thought that ionic strength or certain divalent cations could play an important role in forming stable structure in 271B. CD measurements of 271B in the presence of mM concentrations of Ni^{2+}, Zi^{2+}, and Mg^{2+} were made, but these cations did not have any effect on the CD spectra. Similarly, adding NaCl up to 500 mM did not have an appreciable effect on the spectra at any pH.

These ellipticities observed in 271B, I believe, are reflective of a significant proportion of helix. The next step is to determine how this helicity is distributed throughout the polypeptide. CD cannot tell us this, but NMR, in principle, can. The NMR characterization of 271B will follow in a few sections. However, first I would like to describe some CD-based unfolding studies of 271B.

"Unfolding" of 271B. Another useful aspect of circular dichroism is that it can be used as a means of characterizing unfolding transitions in polypeptides (Cantor & Schimmel, 1980). A canonical protein will unfold upon increasing the temperature or the concentration of a chemical denaturant such as urea or guanidinium hydrochloride. It is common to assume a two-state system (folded/unfolded) which gives rise to a cooperative

unfolding transition. The midpoint of this transition is referred to as the melting temperature, T_m, and it is a measure of the stability of the folded state of the protein.

Although there are no indications that 271B has any true tertiary structure and it is clearly not a canonical protein, it does contain helical structure, especially in mixed-phase solutions, which should disappear with either increasing temperature or denaturant. It was thought that the "melting curve" would give us an idea of the free energy stabilization of the helical structure in 271B. Figure 2.6A shows the temperature "unfolding" curves monitored at θ_{222} for 271B in aqueous solution (pH=2.9) and in 20% acetonitrile. Surprisingly in aqueous solution, the amount of helicity *increases* with increasing temperature up to 90 °C. In 20% acetonitrile, the amount of helicity does not increase, but loses only < 5% of its helicity at 90 °C relative to 25 °C. This was and is quite perplexing data. It should be noted that these results have been reproduced several times with different protein preparations, different spectropolarimeters, and also by a different person (Larry Grace). Figure 2.6B shows θ_{222} in 271B as a function of urea concentration up to 6M. In contrast to temperature, urea reduces the amount of helicity in both solvent compositions, although the transition is linear, not a sigmoidal curve indicative of a cooperative transition. 271B is clearly not a canonical globular protein.

Why is structure removed by 6M urea but not by high temperature? A related question is: Is this curious behavior an artifact of studying an isolated fragment of a transcription factor or is this behavior actually related to function? If this behaviour *is* related to function, then I would propose that the c-jun activation properties present a different paradigm for protein function compared to existing models such as "lock and key" which goes hand in hand with stable three-dimensional structure of proteins. Even the "induced fit" model relies on the context of some pre-existing tertiary or quarternary structure (Watson et al., 1987). In truth, we do not yet know if studying this isolated fragment is biologically relevant.

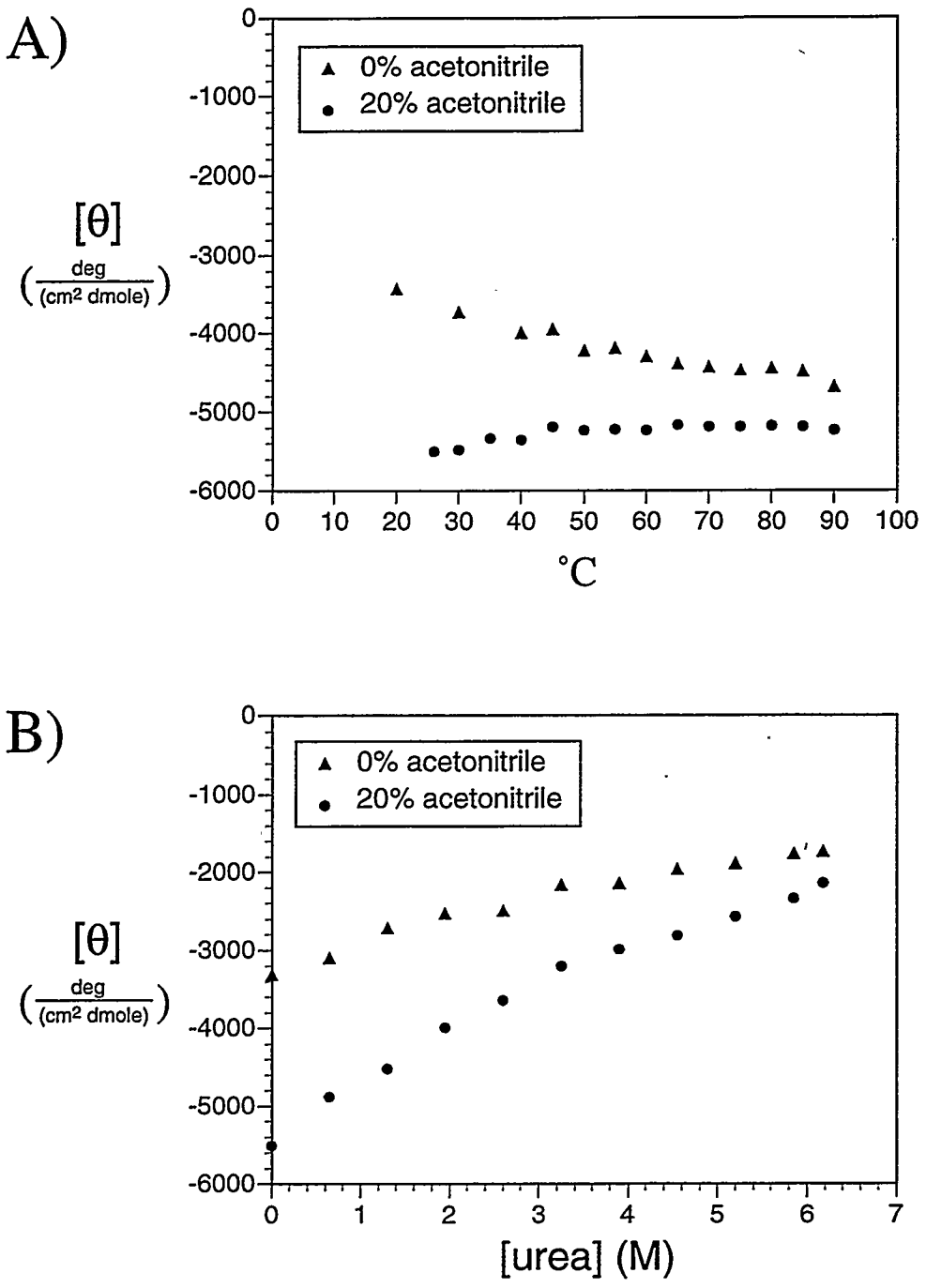


Figure 2.6. 271B ellipticity [θ] as a function of increasing temperature (A) or urea concentration at room temperature (B). Solution conditions are 5 uM 271B, 50 mM phosphate, and no salt. Temperature series was taken at pH=2. Urea series was taken at pH=2.9.

Solubility under various solution conditions. One of the motivations for the CD studies was to test which solution conditions would be appropriate for characterizing structure in 271B by NMR. From the CD studies, TFE and 1-propanol induce the greatest amount of helical structure in 271B. However, because TFE has extraordinary helix-inducing properties in general, further work in TFE was avoided. Therefore, 1-propanol would seem to be the natural choice for further biophysical characterization of 271B, namely solution NMR. One of the stringent requirements of 2D solution NMR is that sample concentrations of > 1 mM are desired for sensitivity considerations. This turned out to be partially problematic because 271B is rather insoluble in 20% 1-propanol at pH=6.5. It is more soluble (~0.5 mM) at pH=2, however. Unfortunately, perdeuterated 1-propanol was not commercially available at the time, which precluded any NMR work in an appreciable amount of 1-propanol. This perdeuterated reagent is now commercially available.

Because the two greatest "helix-inducers" as determined by CD could not be used, the remaining organic solvent/water mixtures were also tested for solubility at either low pH (2 or 3.0) or pH=6.5. It appeared that 271B is universally marginally soluble (< 0.5 mM) at pH=4, so this pH was not considered further. 20% mixtures of methanol, ethanol, and acetonitrile were assayed for their ability to dissolve lyophilized 271B. DMSO was also assayed even though no CD work was done in DMSO since it absorbs strongly in the UV range, precluding any CD measurements. The following solution conditions can solubilize 271B to > 0.5 mM:

20% acetonitrile, pH=3, 50 mM phosphate buffer
20% DMSO, pH=3, 50 mM phosphate buffer
20% DMSO, pH=2, 50 mM phosphate buffer
20% methanol, pH=2, 50 mM phosphate buffer
20% 1-propanol, pH=2, 50 mM phosphate buffer

No solution conditions were found where 271B reached 1 mM. In addition, from these limited tests it appears that 271B is generally most soluble at pH ≤ 3.0. In conclusion from

the CD and solubility test results, acetonitrile, DMSO, and methanol were candidate mixed-phase solvents for NMR studies.

1D and 2D ^1H NMR. Ideally, it would have been nice to obtain NMR spectra of 271B in all five of the conditions just given. Use of these solvents, however, requires using perdeuterated solvents due to dynamic range issues. Because perdeuterated acetonitrile, DMSO, and TFE were readily available, only these solvents were tested. In Figure 2.7, 1D NMR spectra at 500 MHz of 271B in pure aqueous, 10% TFE, and 20% acetonitrile are given. The data for 20% DMSO was lost, but from my notebook, this condition appears to be intermediate between TFE and acetonitrile, at least in terms of 271B linewidths. The general envelope of the amide and aromatic region is retained when going from aqueous conditions to mixed-phase conditions. All spectra are more dispersed than what one would expect from a random coil polypeptide, a feature consistent with the observation from CD that 271B has secondary structure. The most striking NMR spectrum of 271B is at 20% acetonitrile (spectrum C). Under these conditions, the linewidths are narrow compared to the other solution conditions. Sharp lines are desireable for 2D NMR because coherence transfer steps are more efficient when there are sharp lines. In order to assign the resonances to specific residues, successful coherence transfer steps in COSY or TOCSY type experiments are a necessity for larger polypeptides (> 50 residues). Therefore, 2D NMR of 271B was pursued in 20% acetonitrile at pH=2-3.

Figure 2.8 shows regions of a ^1H-^1H 2D NOESY spectrum at 600 MHz of 0.6 mM 271B in 70% H_2O, 10% D_2O, 20% CD_3CN, 50 mM phosphate pH=2, and a trace amount of TFA from lyophilization out of HPLC buffer. There are several features in this spectrum which argues for the presence of structure. Approximately 30 HN-HN crosspeaks, symmetrically opposed around the diagonal, are observed in the downfield region (8-9 ppm) in Figure 2.8A. HN-HN crosspeaks are characteristic of helices in proteins, where HN(residue i) is < 5 Å from HN (residue i+1). This type of NOE is not observed in unstructured regions of polypeptides. However, weak cross-strand HN-HN

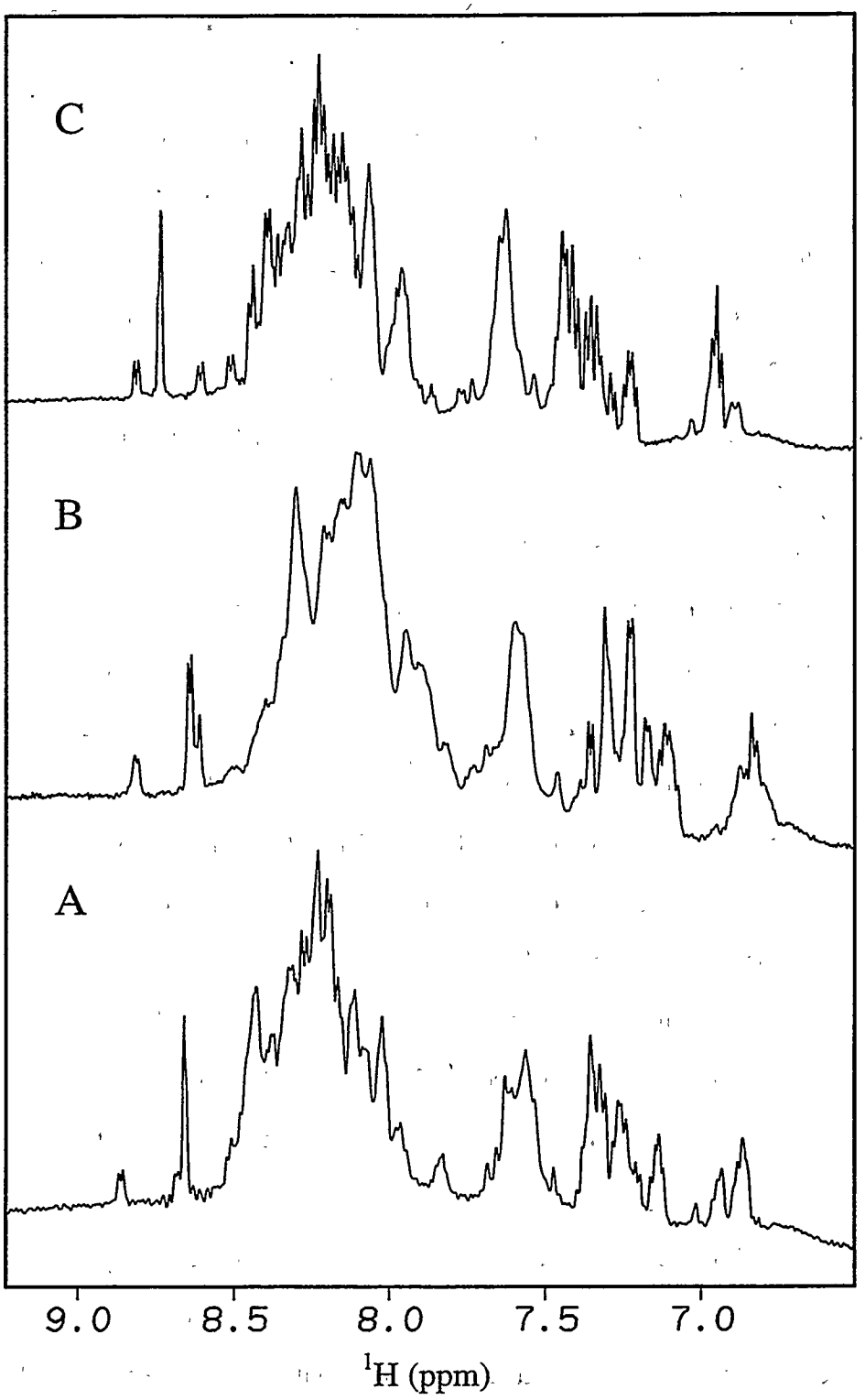


Figure 2.7. Amide and aromatic region of 1D NMR spectra at 500 MHz of 271B in various organic mixed-phase solution conditions. A) aqueous, pH=2 B) 10% trifluoroethanol (TFE), pH=2 C) 20% acetonitrile, pH=2. All spectra were acquired at 25 °C.

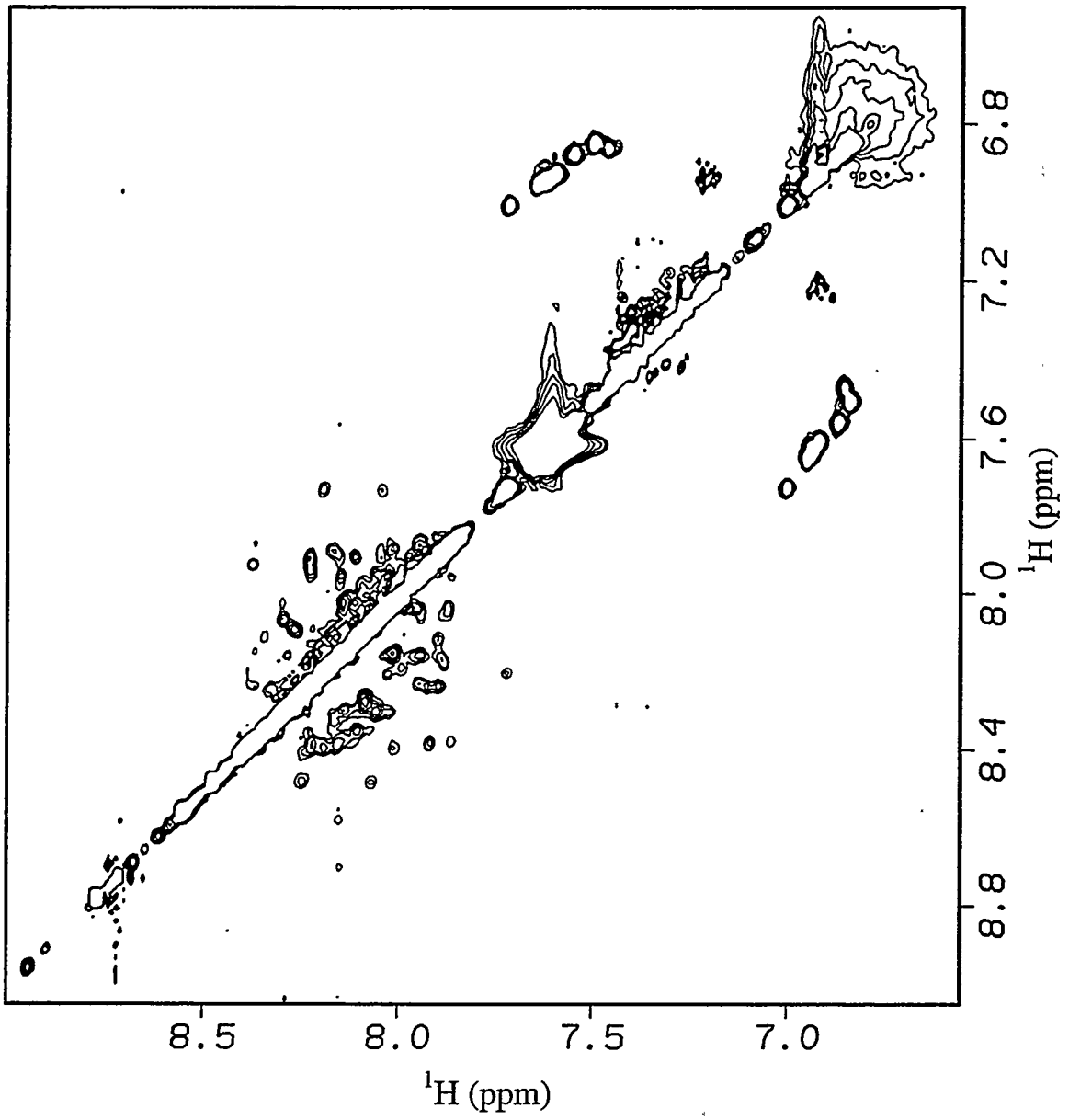


Figure 2.8A. HN-HN and aromatic region of a 2D NOESY experiment at 600 MHz on 271B (c-jun δ-a1-ε activation domain) at 25 °C. Protein concentration is ~0.6 mM in 70% H_2O, 10% D_2O, 20% CD_3CN, 50 mM phosphate (pH=2). The NOE mixing time is 125 ms.

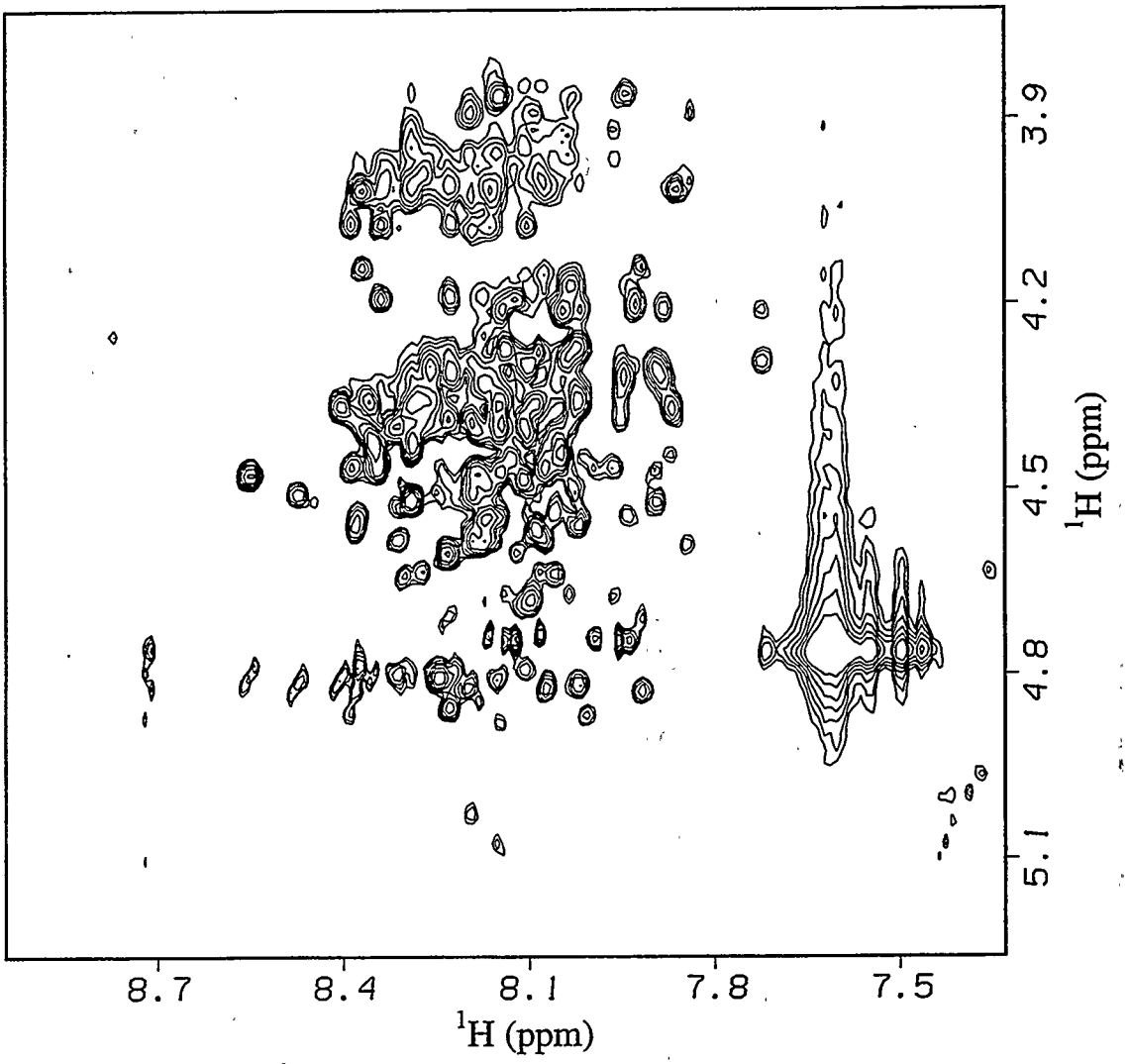


Figure 2.8B. H_α-HN ("Fingerprint") region of a 2D NOESY experiment at 600 MHz on 271B (c-jun δ-a1-ε activation domain) at 25 °C. Protein concentration is ~0.6 mM in 70% H_2O, 10% D_2O, 20% CD_3CN, 50 mM phosphate (pH=2). The NOE mixing time is 125 ms.

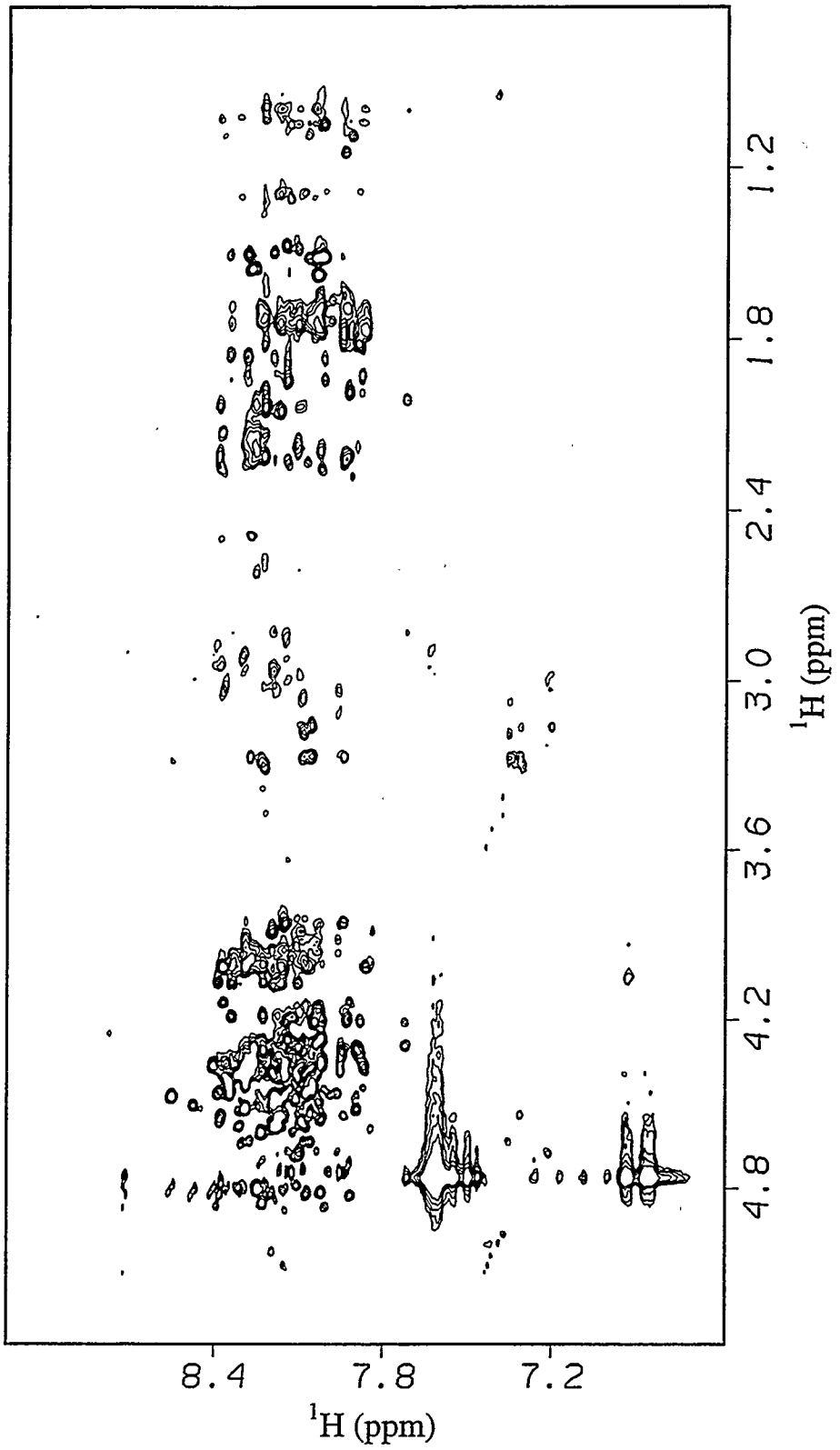


Figure 2.8C. Downfield region of a 2D NOESY experiment at 600 MHz on 271B (c-jun δ-a1-ε activation domain) at 25 °C. Protein concentration is ~0.6 mM in 70% H_2O, 10% D_2O, 20% CD_3CN, 50 mM phosphate (pH=2). The NOE mixing time is 125 ms.

crosspeaks are observed in β-sheets, and HN-HN crosspeaks can also be observed in turns. Because it is known that helical structure exists in 271B from CD, this set of HN-HN crosspeaks is most likely arising from helical structure. Next, the region shown in Figure 2.8B is known as the "fingerprint" region in polypeptide NOESY spectra since a crosspeak from every residue is normally observed and the arrangements of peaks in this region is unique to a given peptide or protein. Crosspeaks in this region correspond to H_α-HN NOEs, and an intraresidue NOE from each residue is normally observed. This region in 271B exhibits surprisingly good chemical shift dispersion. In fact, it could be considered native-like dispersion. The dispersion is such that H_α chemical shifts downfield of the H_2O solvent (~4.8 ppm) are observed. Such shifts are a sure sign of protein structure and are often associated with β- or extended structure, although this cannot be regarded as proof of β-structure. From the "fingerprint" region of 271B alone, it can be concluded that 271B cannot be regarded simply as an unstructured protein. Figure 2.8C shows a more distant view of the "fingerprint" region and the upfield-HN region of the NOESY spectrum. Because there are many more crosspeaks in the "fingerprint" region compared to the upfield region (0 - 3.5 ppm) it is probable that most NOEs are of the intraresidue type, consistent with a lack of stable tertiary structure in 271B. Tertiary structure in proteins leads to a large number of long-range (i, >i+4) NOEs which should give rise to more crosspeaks in the upfield-HN region of the NOESY. Long-range NOEs are also normally observed in the upfield-aromatic region of the NOESY for globular proteins. This would correpond to packing in the hydrophobic core. Such NOEs are not present in Figure 2.8C. Finally, all NOEs are negative, i.e. all crosspeaks have the same sign as the diagonal. This is what one would expect for a molecule tumbling slowly in solution (Neuhaus & Williamson, 1989). Conversely, it is feasible that a random coil polypeptide of similar length would have positive NOEs, and the observation of such NOEs would exclude the possibility of structure in a protein of this size. This type of argument was used by Cho et al. with regard to the N-terminal activation domain from heat

shock transcription factor (Cho et al., 1996). NOESY spectra of 271B were also collected at pH=6.5; fewer NOESY crosspeaks were observed at this pH.

$^{15}N/^{1}H$ NMR. For a 14 kD protein, it is common knowledge that ^{1}H NMR alone is not the most efficient way of getting resonance assignments and determining three-dimensional structure. The last 5-10 years has seen a revolution in heteronuclear multidimensional NMR of proteins in particular (Chapter 1). Using heteronuclei such as ^{15}N and ^{13}C, NOESY experiments can be extended so that the normal crosspeaks can be spread out over 3 dimensions, reducing spectral overlap, with the heteronucleus providing the 3rd (or 4th in some cases) dimesion (Figure 1.7). In addition, assignments can be obtained from using strictly through-bond correlations so that connectivities are largely independent of structure. These types of experiments are referred to as triple resonance (^{1}H, ^{13}C, ^{15}N) experiments, and they have only been available for only the last 6 years. This heteronuclear approach is the most attractive, and thus I would have liked to uniformly label 271B with ^{15}N or both ^{13}C and ^{15}N. Unfortunately, the cells (BL21-DE3) which best overexpressed 271B in rich media (~ 7-10 mg purified/liter) could not sufficiently overexpress 271B on minimal media (M9). However, these cells would sufficiently overexpress 271B when grown on a defined media of M9 plus 125 ug/ml of each of the 20 amino acids. A 1-liter prep of 271B was done with this growth media where ^{15}N alanine and leucine amino acids were substituted for the normal isotopic residues. Alanine and leucine were chosen for ^{15}N incorporation because they are distributed reasonably well throughout the 271B sequence and also because there should be relatively little "leakage" of these amino acids into other amino acid biosynthetic pathways in *E. coli.* Therefore, there should be nearly 100% incorporation of ^{15}N into these residues.

A 0.3-0.4 mM 271B NMR sample with specific ^{15}N-labeling of alanines and leucines was successfully prepared. The $^{15}N/^{1}H$ HSQC correlation spectrum at 600 MHz is shown in Figure 2.9. Because 271B has 8 alanines and 18 leucines, 26 peaks were expected in this spectrum. There are approximately 26 intense peaks in addition to a

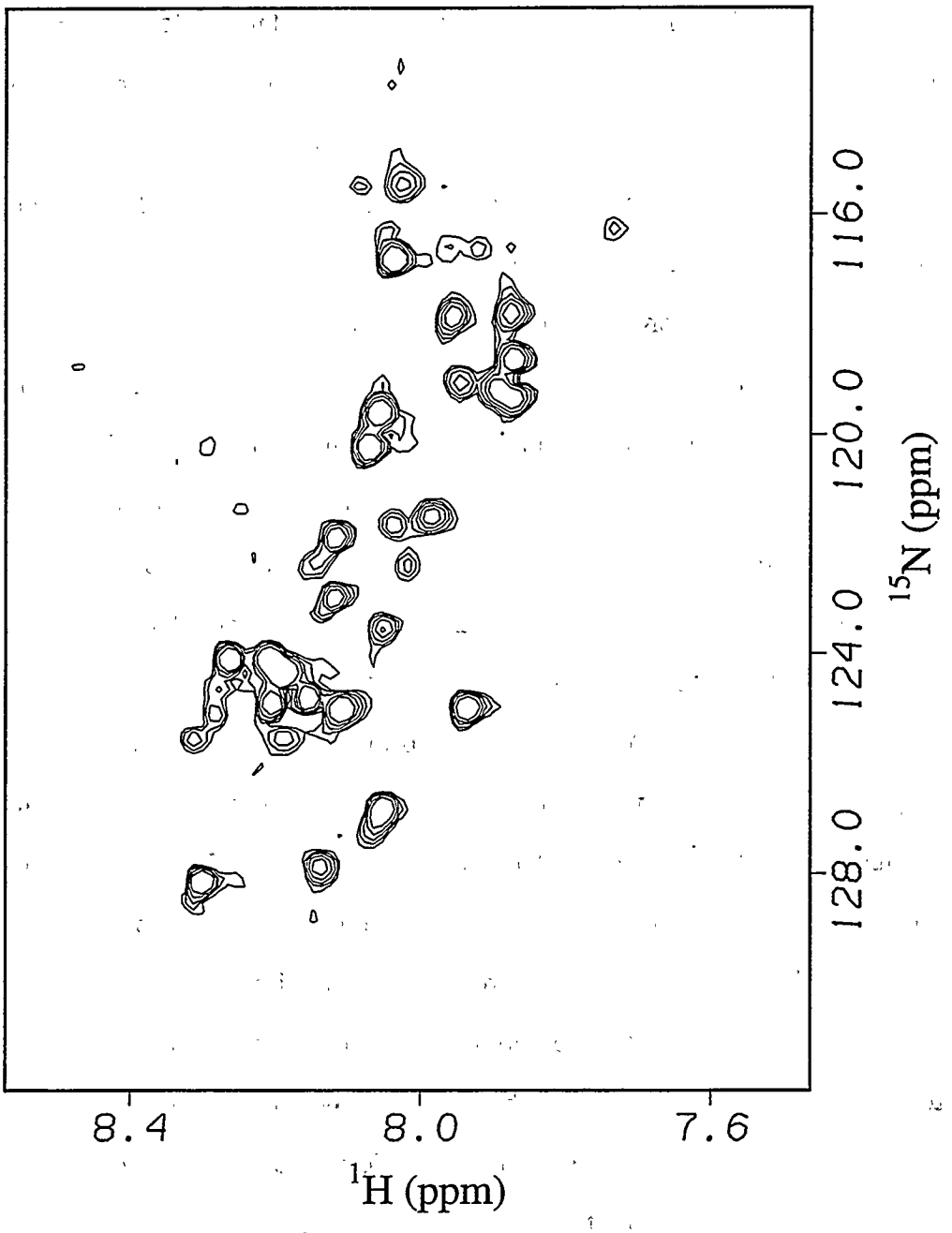


Figure 2.9. ^{15}N/^{1}H HSQC at 600 MHz of 0.3-0.4 mM 271B (pH=2.5) specifically labeled with ^{15}N alanine and leucine residues. 50* (t_1) x 512* (t_2) points were acquired over approximately 80-90 minutes (32 scans/block). The ^{15}N spectral width is 1200 Hz. The contour level multiplier for the plot shown is set at 1.5.

number of weaker peaks, most likely due to "leakage" in the amino acid biosynthetic pathways. As in the proton 1D and 2D spectra, the HSQC spectrum has relatively well dispersed chemical shifts, again consistent with the existence of structure. Looking closer, there does appear to be a more degenerate patch of peaks at approximately 125 ppm in the ^{15}N dimension. Interestingly, random coil values of amide ^{15}N in alanines are centered around 125 ppm, whereas leucine random coil values are centered around 122 ppm (Wishart et al., 1995). One possible way of looking at the dispersion of peaks in Figure 2.9 is that the 8 alanine NH peaks are relatively degenerate and form the "patch" centered at ^{1}H=8.2 ppm and ^{15}N=125 ppm. If this were the case, the leucine NH peaks would have much larger chemical shift dispersion than the alanines. Moreover, the leucine dispersion would be centered around 122 ppm in ^{15}N. This non-rigorous interpretation of the data provocatively supports the hypothesis that hydrophobic residues in activation domains are important determinants of target (i.e. TAFs) interactions (Tjian & Maniatis, 1994). If the leucines are indeed well dispersed in contrast to alanines, it would appear that leucines, not alanines, are involved in some sort of organization, even in the absence of a target molecule.

With confirmation of reasonably successful specific labeling, the next experiment to try was an ^{15}N-edited ^{1}H-^{1}H NOESY experiment (Figure 2.10, Bruker AMX pulse program in Appendix D1). Using half-X filtering techniques (Otting & Wuthrich, 1990), a ^{1}H-^{1}H NOESY data set with $\omega 2$ refocused ^{15}N half-filter was collected. The data was collected such that 2 sub-experiments were stored separately on the disk drive. Adding the 2 sub-experiments yielded a 2D NOESY spectrum with only non-^{15}N-bound protons detected in $\omega 2$. Subtracting the 2 sub-experiments yielded a 2D NOESY spectrum with only ^{15}N-bound protons detected in $\omega 2$. Unfortunately, the sensitivity of the resultant data set was too low for interpretive use. The "add" spectrum had only a few crosspeaks in the "fingerprint" region (compare to Figure 8B) and no detectable HN-HN crosspeaks. The "subtract" spectrum had no signals above the noise level. This poor quality data set is the

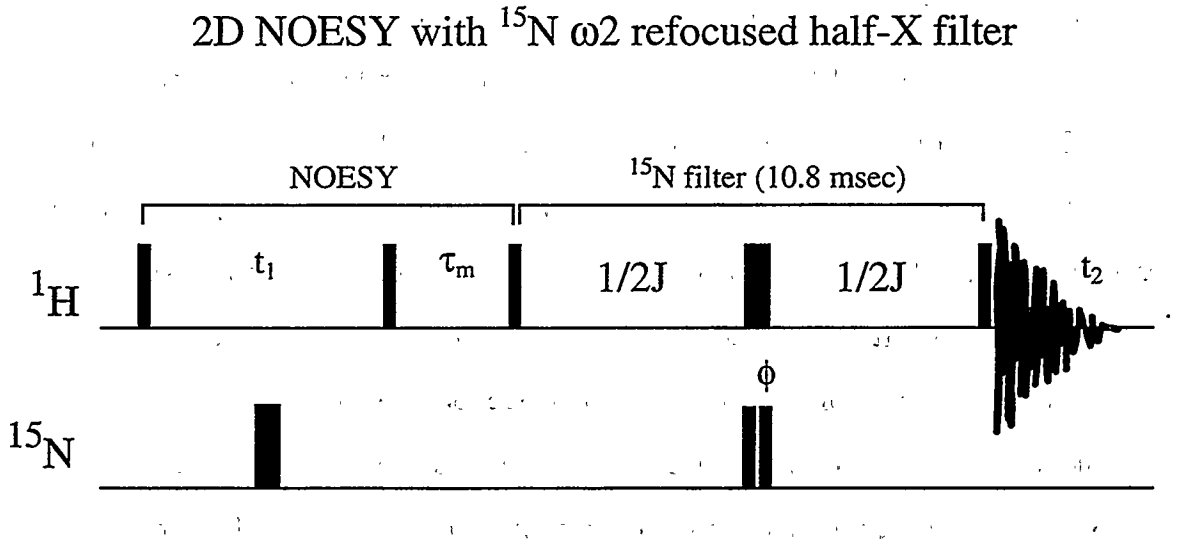


Figure 2.10. Pulse sequence for a 2D ^{1}H-^{1}H NOESY with ^{15}N ω2 refocused half-X filter. "J" refers to the one-bond ^{15}N-^{1}H coupling constant (= 92 Hz). ϕ refers to the ^{15}N "editing" pulse which is phase cycled to give 2 sub-experiments. Thin and thick bars correspond to 90° and 180° pulses, respectively.

result from a combination of 1) the extra 10.8 msec of delays in the pulse sequence (Figure 2.10) that are required for the refocused half-filter 2) the sample being more dilute than the previous unlabeled sample. During delays in the pulse sequence where magnetization is spent in the transverse (x-y) plane, T_2 relaxation reduces the net amount of observable magnetization in a time-exponentially decaying fashion. This has potentially very large effects in the ^{15}N half-filtered NOESY experiment compared to regular NOESY since in a non-edited NOESY experiment, essentially no time is spent in the transverse plane for initial t_1 points. Of course, the effect is linewidth dependent. It can be argued that linewidth problems are also evident in the HSQC (Figure 2.9)—where ^{15}N-^{1}H transfer efficiencies appear to be low for the concentration of this sample—because the peaks are approximately 2-fold less intense if 1024* t_2 points are apodized over, instead of the 512* t_2 points used for the plot shown. There is no visible difference in $f1$ resolution between the two processing schemes. In other words, the HSQC peaks are weaker than what one would expect for a well-behaved protein with sharp ^{1}H lines (< 15-20 Hz).

^{1}H linewidths are somewhat cumbersome to measure in a large protein, but ^{15}N linewidths are convenient to measure, provided that high-quality HSQC spectra can be acquired in several hours (Barbato et al., 1992). This can be accomplished by strategically inserting a Carr-Purcell-Meiboom-Gill pulse train (normally used for 1D T_2 experiments) into an HSQC pulse sequence since T_2 is inversely proportional to linewidth. A series of 2D HSQCs recorded with different T_2 relaxation periods should yield correlation peaks whose intensities decay with a characteristic time constant (T_2) as a function of the relaxation period length. A more complete discussion of these experiments can be found in Chapter 6 as applied to Sxl-RBD2. For the purposes of this project, ^{15}N T_1 and T_2 time constants, as well as the $^{15}N/^{1}H$ NOE can be measured using this approach. The T_1 experiment was done, but I had difficulty obtaining good fits to the data. In retrospect, a significant portion of this difficulty could have resulted from a combination of 1) not using signal-enhancing processing schemes as described above 2) using crosspeak volumes

instead of intensities, which in practice appear to have larger associated errors. Even so, it is not clear whether reliable T_1s can be obtained from this data. At the time I stopped working on 271B, I was forced to conclude that for most of the heteronuclear experiments, the combination of low protein concentration (at least for the ^{15}N sample I made) and relatively broad lines of 271B renders these experiments too insensitive for the extraction of meaningful data. In retrospect, this may not be true for the ^{15}N T_1 and T_2 experiments. This is important because information on internal dynamics and tumbling correlation times can be inferred from ^{15}N T_1, T_2 , and $^{15}N/^1H$ NOE values (Chapter 6).

Analytical ultricentrifugation, fluorescence, and light scattering. By this point we had formulated more specific questions about 271B: a) how compact is it? b) is it a molten globule? c) what is the hydrodynamic radius? d) is there a significant population of multimerized states? For many protein systems, these questions may be addressed through NMR or crystallography, but for reasons mentioned above, these preferred techniques were not a viable option. Therefore, we looked at some simpler techniques which could potentially answer these questions with relatively little effort.

Analytical ultracentrifugation is an excellent technique for determining the molecular weight, sedimentation coefficient, and subsequently a hydrodynamic radius for proteins > 10 kD. There are two main types of experiments: sedimentation veclocity ultracentrifugation and equilibrium ultracentrifugation (Cantor & Schimmel, 1980; Hansen et al., 1994). Sedimentation velocity runs will yield a measurement for the sedimentation coefficient, s, given in Svedberg units (S). This coefficient can be interpreted in the context of the overall molecular shape of the protein and has the potential to describe the compactness or hydrodynamic radius of the protein. Equilibrium runs allow for a calculation of the molecular weight of the sedimenting species. Unfortunately, equilibrium runs take much longer (2-3 days) than velocity runs, and because 271B was being studied under the condition of 20% acetonitrile (potentially damaging to the ultracentrifugation cells) equilibrium runs were not performed.

A single sedimentation velocity ultracentrifugation run of 271B with 7 equally spaced time points (4500 to 11,700 seconds) was carried out by Dr. Russell Heath in Howard Schachman's lab. The conditions were identical to the NMR conditions, except that 271B was at 100 uM concentration. Presumably from the rate of migration of the protein/solvent boundary, a computerized analysis routine calculated a sedimentation coefficient of 1.48 S. Because the rate of the sedimenting "particle" depends on the viscosity of the actual solution, an interpretation of this number requires a knowledge of the effects from the viscosity as well as the partial specific volume of the solute (protein) molecule, $\overline{V}_2$, which can be either calculated from theory or estimated with a reasonably small error (Cantor & Schimmel, 1980). As it turned out from subsequent measurements by Larry Grace, the obtained value for s was non-reproducible, and becuase I made only one measurement myself, I do not have high confidence in the value obtained. Nevertheless, I attempted to interepret my value of 1.48 S under 20% acetonitrile conditions.

The following relationship allows one to compare sedimentation coefficients for proteins investigated under various solution conditions, namely different solution viscosities:

$$\frac{s_{20,w}}{s} = \frac{1-0.9982(\overline{V}_2)_{20,w}}{1-\overline{V}_2\rho}\left(\frac{\eta_{T,w}}{\eta_{20,w}}\right)\left(\frac{\eta_{T,soln}}{\eta_{T,w}}\right)$$

where $\overline{V}_2$ is the partial specific volume, ρ is solution density, and η is viscosity. Essentially, protein sedimentation coefficients can be converted into a "standard" condition of 20 °C in pure H_2O. This was done for 271B as well as for sedimentation coefficients of random coil polypeptides (Tanford et al., 1967). For ribonuclease A denatured in 6M Gu-HCl, a converted s value of 1.43 S was obtained for standard conditions. This value is reasonable in that it is significantly less than 1.85 S which corresponds to native RNase A

(Cantor & Schimmel, 1980). The s value for 271B under standard conditions is 1.07 S. This value for 271B is difficult to interpret because there may be a significant error in the viscosities used since the viscosity effects from 271B protein are unknown. Disregarding this problem for a moment, the value of 1.07 S intuitively seems unreasonable since it is much less than the value of 1.43 S for a random coil of similar molecular weight. Therefore, I do not put much stock in this one piece of sedimentation velocity data. A more careful series of ultracentrifugation studies and a deeper grasp of the associated theory than I have may lead to meaningful conclusions from this kind of experiment.

Another question was whether 271B is a molten globule under the condtions employed. A popular method which was initially used to define the molten globule state in proteins is the differential fluorescence of 1-anilino-naphthalene-8-sulfonate (ANS) in the presence and absence of protein (Semisotnov et al., 1991). It has been observed in numerous cases that proteins in the molten globule state bind ANS, presumably by allowing ANS to insert into the loosely packed hydrophobic interior of the protein. This in turn causes a dramatic increase of intensity (approximately 10-fold or greater) in the fluorescence emission spectrum around 500 nm. This simple experiment was carried out with the assistance of Jana Steiger, a graduate student in Prof. Ken Sauer's group. The fluorescence spectrum of ANS with and without 271B (pH=3, 20% acetonitrile or 0% acetonitrile) did not change dramatically. Therefore, it appears that 271B does not bind ANS and should not be seriously considered as a molten globule. This result is not inconsistent with the other data. It does appear to be consistent with our model picture of 271B as a protein with residual helical secondary structure and a negligible degree of tertiary structure.

Finally, in an attempt to obtain a semi-quantitative measurement of the hydrodynamic radius of 271B dynamic light scattering was attempted. The value obtained was unreasonably high for a 14 kD polypeptide and the problem was attributed (not conclusively) to the presence of dust particles, even though the sample was filtered through

a 0.45 um filter. Light scattering is very sensitive to such particles since the light reflected from the large dust particles tend to dominate the signals from the smaller proteins. I did not try this measurement again, but Larry Grace did and obtained a somewhat smaller value.

Proteolysis experiments. Because 271B was shown to lack any stable tertiary structure, a nagging question was if we simply were looking at an "incorrect" fragment from c-jun. This issue was addressed by subjecting full-length c-jun to a series of protease digests. Any individually folded domains should have less proteolytic susceptibility and should be observable as a stable band using SDS-PAGE. The non-sequence-specific proteases papain, subtilisin, chymotrypsin, and elastase were all incubated separately with full-length c-jun. Full-length c-jun was generously supplied by Arie Admon from Dr. Robert Tjian's laboratory. Unfortunately, although the gels were preserved, they developed extensive cracking over time and at the time of this writing were useless. From my notebook, it can be said that when digested with papain there was a very stable band which ran at about 10 kD. Because of a combination of my lack of expertise with regard to peptide sequencing and a certain amount of resistance by the few capable of carrying out peptide sequencing, this proteolytically stable 10 kD band was never mapped to a specific region of c-jun, although it is likely that this band could correspond to the B-zip DNA-binding region of c-jun.

Concluding remarks

Biophysical characterization of 271B (δ-a1-ε activation domain from c-jun) was pursued primarily in 20% acetonitrile, pH=2, and 50 mM phosphate. This solvent composition (along with other mixed-phase solvents) exhibited helix stabilizing properties for 271B and NMR line-narrowing properties, without decreasing solubility. Although 271B lacks globular tertiary structure and does not appear to be a molten globule using standard definitions, 271B has a significant helix-forming propensity, with an estimated

average helix content of 16% as determined by CD. Helical HN-HN NOEs and chemical shift dispersion support this conclusion. Temperature and urea concentration dependences of helicity were determined. The structure in 271B does not appear to have large thermodynamic stability, and the temperature dependence is non-canonical.

A number of other activation domains have previously been characterized by CD and/or NMR. VP16 contains a small, C-terminal acidic activation domain which has been the focus of many biochemical and biophysical studies. Using NMR, this domain was shown to have no signs of secondary structure, even in mixed-phase solvent compositions of up to 80% methanol (O'Hare & Williams, 1992). The N-terminal activation domain from heat shock transcription factor (HSF) has also been shown to be completely unstructured using CD and NMR techniques (Cho et al., 1996). Specifically, it was shown to exist as a dynamically disordered polypeptide through use of the $^{15}N/^{1}H$ heteronuclear NOE experiment. This has been the most compelling data for the unstructured nature of an activation domain. Other activation domains, such as the Tat protein from equine infectious anemia virus (EIAV) and the C-terminal domain from NF-kB, have been shown to have helical propensity in H_2O/TFE solutions (Willbold et al., 1993; Schmitz et al., 1994). The Tat protein even exhibits helical properties in aqueous solution as determined by NMR. 271B appears to resemble the Tat protein, with its helical propensity and dispersed NMR spectrum, more than any other activation domain in the literature. Nevertheless, in all cases, there is a lack of stable globular structure. This now seems to be a general property of isolated eukaryotic activation domains.

Baichwal, V. R. & Tjian, R. (1990) *Cell 63*, 815-825.

Baichwal, V. R., Park, A. & Tjian, R. (1992) *Genes & Development 6*, 1493-1502.

Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992) *Biochemistry 31*, 5269-5278.

Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K. & Tjian, R. (1987) *Science 238*, 1386-1392.

Bohmann, D. & Tjian, R. (1989) *Cell 59*, 709-717.

Buratowski, S. (1994) *Cell 77*, 1-3.

Cantor, C. R. & Schimmel, P. R. (1980) in Biophysical Chemistry, (P. C. Vapnek), W.H. Freeman and Company, New York.

Cho, H. S., Liu, C. W., Damberger, F. F., Pelton, J. G., Nelson, H. C. M. & Wemmer, D. E. (1996) *Protein Science 5*, 262-269.

Conaway, R. C. & Conaway, J. W. (1993) *Annual Reviews in Biochemistry 62*, 161-190.

Curran, T. & Franza, B. R. (1988) *Cell 55*, 395-397.

Gill, G., Pascal, E., Tseng, Z. & Tjian, R. (1994) *Proceedings of the National Acadamy of Sciences 91*, 192-196.

Goodrich, J. A., Hoey, T., Thut, C. J., Admon, A. & Tjian, R. (1993) *Cell 75*, 519-530.

Hansen, J. C., Lebowitz, J. & Demeler, B. (1994) *Biochemistry 33*, 13155-13163.

Hoey, T., Winzierl, R. O. J., Gill, G., Chen, J. L., Dynlacht, B. D. & Tjian, R. (1993) *Cell 72*, 247-260.

Kadonaga, J. T., Courey, A. J., Ladika, J. & Tjian, R. (1988) *Science 242*, 1566-1570.

Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993) *Nature 365*, 520-527.

Mitchell, P. J. & Tjian, R. (1989) *Science 245*, 371-378.

Nelson, J. W. & Kallenbach, N. R. (1986) *Proteins 1*, 211-217.

Nelson, J. W. & Kallenbach, N. R. (1989) *Biochemistry 28*, 5256-5261.

Neuhaus, D. & Williamson, M. P. (1989) in The Nuclear Overhauser Effect in Structural and Conformational Analysis, VCH Publishers, Inc., New York.

Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G. & Burley, S. K. (1995) *Nature 377*, 119-128.

O'Hare, P. & Williams, G. (1992) *Biochemistry 31*, 4150-4156.

O'Shea, E. K., Klemm, J. D., Kim, P. S. & Alber, T. (1991) *Science 254*, 539-544.

O'Shea, E. K., Rutkowski, R. & Kim, P. S. (1992) *Cell 68*, 699-708.

Otting, G. & Wuthrich, K. (1990) *Q. Rev. Biophys. 23*, 39-96.

Pugh, B. F. & Tjian, R. (1990) *Cell 61*, 1187-1197.

Schmitz, M. L., dos Santos Silva, M. A., Altmann, H., Czisch, M., Holak, T. A. & Baeuerle, P. A. (1994) *J. Bio. Chem. 41*, 25613-25620.

Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991) *Biopolymers 31*, 119-128.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60-89.

Tanford, C., Kawahara, K. & Lapanje, S. (1967) *J. Am. Chem. Soc. 89*, 729-736.

Tjian, R. & Maniatis, T. (1994) *Cell 77*, 5-8.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. & Weiner, A. M. (1987) in Molecular Biology of the Gene, (J. R. Gillen), Benjamin/Cummings Publishing Company, Inc., Menlo Park.

Willbold, D., Kruger, U., Frank, R., Rosin-Arbesfeld, R., Gazit, A., Yaniv, A. & Rosch, P. (1993) *Biochemistry 32*, 8439-8445.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67-81.

Zawel, L. & Reinberg, D. (1993) *Progress in Nucleic Acids Research 44*, 67-108.

Chapter 3

Sex-lethal and Sex Determination in *Drosophila*

Sex determination in *Drosophila*

The *Drosophila* binary switch gene, *Sex-lethal (Sxl)*, plays a central role in directing events during somatic and germline sexual development. Its expression mode, initiated by upstream events involving the X:A ratio early in development, is dependent upon the sexual state of the organism (Baker, 1989; Cline, 1993; Parkhurst & Meneely, 1994). The functional *Sxl* gene product is the female-specific protein, Sxl. The male-specific product is non-functional because it is a truncated protein generated by alternative splicing (Bell et al., 1988; Samuels et al., 1993). Sxl protein actively maintains the female-specific expression mode through an autoregulatory feedback loop (Keyes et al., 1992; Cline, 1993). In addition, Sxl is the master regulator of a cascade of downstream events required to maintain female cell fate. This dual regulation is accomplished by a mechanism initially involving transcriptional control (Keyes et al., 1992) and is maintained at the level of pre-mRNA splicing (Bell et al., 1988).

Sxl maintains female cell fate because it acts as a positive regulator of the feminizing gene, *transformer* (*tra*) (Cline, 1979). The *tra* pre-mRNA can be alternatively spliced to yield a non-sex-specific or a female-specific mRNA (Sosnowski et al., 1989). The Sxl protein defines which of these *tra* splicing pathways is taken because the full-length female Sxl protein binds specifically to the pyrimidine-rich tract on the *tra* pre-mRNA just upstream of the 3' splice site of exon 2 (Inoue et al., 1990), as shown in Figure 3.1. The 20 bases preceding the 3' splice site are identical across 5 species of flies given, thereby making the splicing-regulated events a general phenomenon for *Drosophila*. Furthermore, as pre-mRNA splicing in general usually involves a pyrimidine-rich tract at the 3' splice

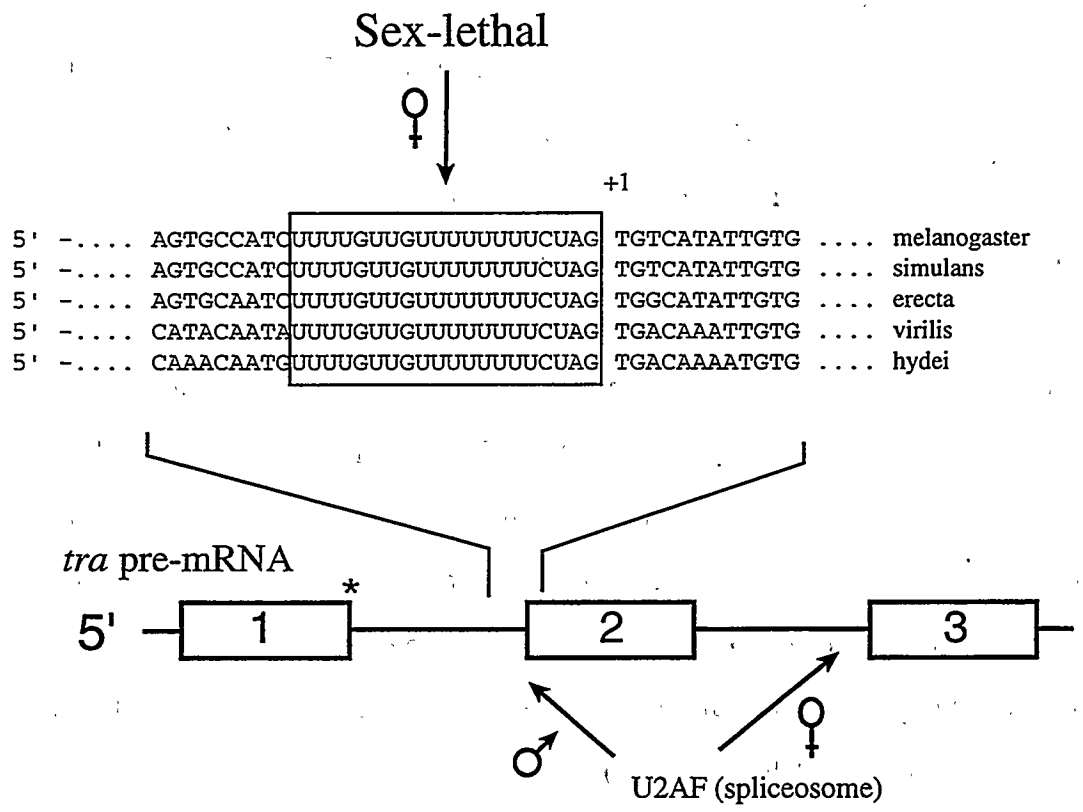


Figure 3.1. Schematic diagram of the *transformer* pre-mRNA. Exons are boxed and numbered. The polypyrimidine tract at the 3' splice site before exon 2 is expanded and highlighted. The asterisk represents the activated 5' splice site. Functional female sex-lethal binds to the U-rich tract. The differential gender pathways for the spliceosomal factor U2AF illustrate the alternative tra splicing pathways in drosophila.

site, it is possible that the Sxl/*tra* recognition could be relatable to general splicing mechanisms in eukaryotes. Sxl accomplishes this binding through its two conserved RNA binding domains (RBDs) of the RNP consensus type (Bell et al., 1988; Valcarcel et al., 1993; Burd & Dreyfuss, 1994). Sxl binding prevents splicing to the 3' splice site of exon 2. Instead, a distal alternative 3' splice site (exon 3) is activated. It is believed that a mechanism involving Sxl-RNA interactions is also responsible for Sxl autoregulation (Sakamoto et al., 1992). In parallel processes, Sxl is a negative regulator of dosage compensation genes in females and is responsible for increasing the transcriptional activity of the X chromosome in males (Parkhurst & Meneely, 1994). The name Sex-lethal arises from the observation that introduction of a mutant version of the protein in males or females leads quickly to death due to an imbalance of dosage compensation of X chromosomal genes.

Sex-lethal and other RBD-containing proteins

It is currently understood that sexual differentiation and maintainance of female cell fate in *Drosophila* is directly linked to the ability of Sxl protein to bind RNA. In fact, the two RBDs are the only well-recognized regions of the 354 residue full-length protein. Studies on other RBD(RNP consensus type)-containing proteins have suggested that specific aromatic sidechains are involved in RNA binding (Scherly et al., 1990; Jessen et al., 1991). Most recently, Nagai and coworkers have solved the structure of the U1A snRNP RBD/RNA hairpin complex by X-ray crystallography (Oubridge et al., 1994), and they delineated many if not all of the specific contacts of this kind. The amino acid residues carrying these sidechains are in the highly conserved octamer (RNP-1) and hexamer (RNP-2) consensus motifs found in all RBDs (Kenan et al., 1991).

The RNP-1 and RNP-2 consensus motifs within RBDs of several selected proteins can be seen in the sequence alignments in Figure 3.2. The hexamer (RNP-2) and octamer (RNP-1) consensus motifs contain the aromatic residues which are believed to form critical

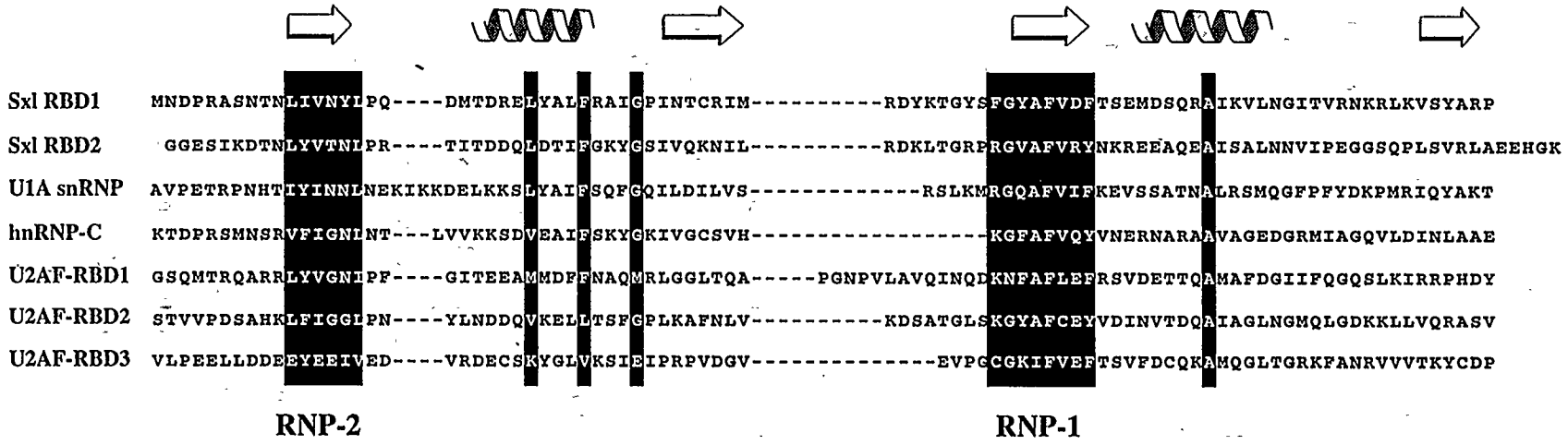


Figure 3.2. Sequence alignment for seven selected RNP consensus type RBDs. RNP-1, RNP-2, and highly conserved regions are highlighted.

contacts with RNA (Oubridge et al., 1994). Over 100 proteins from eukaryotes and prokaryotes have been identified which have homology to the RBD consensus, perhaps revealing that this is an ancient protein fold (Birney et al., 1993; Burd & Dreyfuss, 1994). To date, three-dimensional structures of only U1A snRNP (Nagai et al., 1990), hnRNP-C (Wittekind et al., 1992), and Sxl-RBD2 (Chapter 5) RBDs have been solved, the latter two by NMR. They all share the same overall fold, as expected, but there are some interesting subtle differences in specific regions of the structures that may reflect the differences in base sequence specificities. Perhaps a more interesting feature of RBD-containing proteins is that they often contain two or more RBDs, yet all of the domains are not always involved in binding the same RNA. It is in this context of multiple RNA-binding domains that the Sxl protein has become a system of particular interest. As extraordinary as the complex of U1A snRNP with its RNA substrate is, it is only a single RBD complexed with RNA. How do double or triple RBDs bind RNA compared to single RBD proteins? Or even more fundamental, why is there a need for multiple RBDs in a protein? Answers are not yet known in either structural or energetic detail, but the Sxl protein with its two RBDs appears to be a promising system for addressing questions regarding multiple domains.

The general approach to this project was to work on the individual domains initially; in theory these would be easier to work on simply because they are smaller. The various protein constructs are shown schematically in Figure 3.3. As it turned out, RBD2 had good solubility properties and excellent features for NMR while RBD1 was less soluble and unstable over several weeks. Therefore RBD2 was the initial focus of attention. It was decided later that it would be preferable to get information on RBD1 in the context of RBD1+2 rather than work with RBD1 alone. Stephanie Robertson would eventually take on this aspect of the project. RBD1+2 is less soluble than RBD2 but is at least stable indefinitely. RBD1+2 is a tractable protein for NMR, although it has significantly larger NMR linewidths than RBD2, and is soluble up to only about 0.8 mM before the onset of severe aggregation problems. Ultimately we would like to determine the structure of

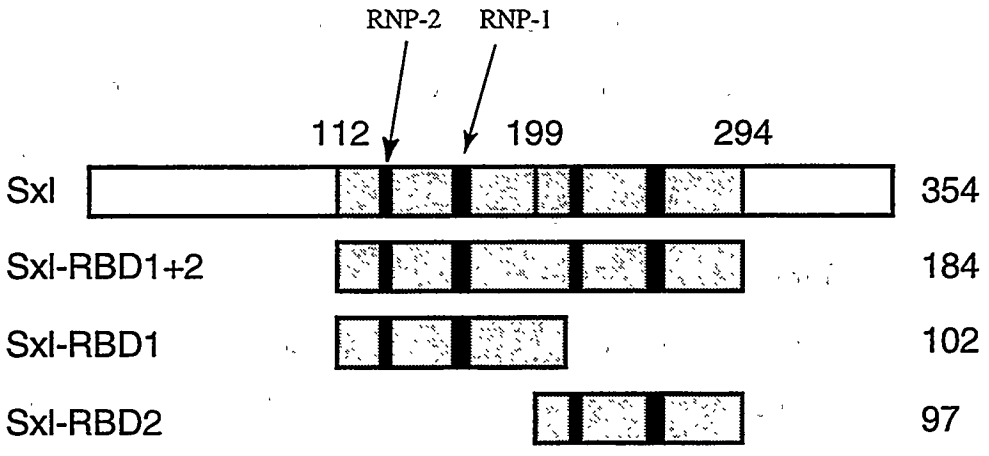


Figure 3.3. The different Sxl protein contructs that will be the subject of discussion in the ensuing chapters. The number of amino acids is given to the right of each construct. The black areas correspond to the consensus RNP-1 or RNP-2 motifs.

RBD1+2 complexed with the RNA polypyrimidine tract. This research has been initiated and will be discussed in chapters 7 and 8.

Sxl RNA-binding domain sample preparations

A total of 3 Sxl RNA-Binding Domain constructs have been overexpressed and purified (lower three constructs in Figure 3.3). The degree to which these various protein fragments were purified varied significantly depending on the purpose of study and will be discussed in the following chapters. Nevertheless, the initial stages of the purification schemes were nearly identical. This "baseline" purification was simply extended as needed. Following are the overexpression and purification protocols used for Sxl-RBD2, Sxl-RBD1, and Sxl-RBD1+2, which each yield >95% purity. The latter purification was to be the most extended because Sxl-RBD1+2 was most frequently used for making complexes with RNA.

Sxl-RBD2. Expression vector pAR-Sxl-RNP-2 (constructed in the lab of Prof. Donald C. Rio) including a T7 promoter (Studier et al., 1990) and cDNA fragment encoding for the C-terminal RNA-binding domain (residues 199-294, hereafter referred to as residues 2-97 in the context of RBD2 alone) of Sex-lethal was transformed into *E. coli* BL21(DE3)pLysS cells. Cells were adapted and grown on M9 minimal media at 37 °C with $^{15}NH_4Cl$ and [U-$^{13}C_6$]-D-glucose as the sole sources of nitrogen and carbon, respectively. When the cells reached an OD_{595} of 0.6-0.8 AU, transcription from the T7 promoter was induced by the addition of isopropyl β-D-thiogalactopyranoside to 0.1 mM. After 2-4 hours of additional shaking, the cells were spun at 5000 rpm for 10 minutes and the pellet was resuspended in approximately 12 ml of lysis buffer (50 mM tris-HCl [pH=7.5], 1 mM EDTA, 1 mM DTT). At this point the resuspended cells were transferred into a 60Ti Beckman ultracentrifugation tube, frozen in liquid nitrogen, and stored indefinitely at -70 °C. The cells were lysed by 2 freeze/thaw cycles. Thawing was done in a bath of water at RT. Before the second freezing in liquid nitrogen, NaCl was added to a

final concentration of 1 M in order to prevent protein/nucleic acid interactions during the nucleic acid precipitation step. The crude extract was prepared by centrifugation of the lysed cells for 1 hour at 40,000 rpm in a 60Ti rotor (Beckman; Fullerton, CA). Nucleic acids were removed from the crude extract by polyethyleneimine precipitation (final concentration 0.5%) in SS-34 tubes followed by a 10 minute spin at 10,000 rpm. $(NH_4)_2SO_4$ was added to the supernatent to a final concentration of 65%. The resulting precipitate was collected by centrifugation in SS-34 tubes at 15,000 rpm for 12 minutes and resuspended in 2-5 ml phosphate buffer (50 mM KPO_4 [pH=7.0], 25 mM KCl, 1 mM EDTA, 1 mM DTT, and 10% glycerol). 1 mM PMSF was included in the phosphate buffer for this initial resuspension as a protease inhibitor.

This solution was dialyzed at 4 °C against phosphate buffer and run through a Q-Sepharose (Pharmacia; Piscataway, NJ) column. The flow-through containing RBD2 was collected and loaded onto a Blue-Trisacryl (IBF Biotechnics; France) column equilibrated in phosphate buffer. Sxl-RBD2 protein was eluted with a KCl concentration gradient of 50mM to 2M. Gradient programs for all three Sxl constructs are given in Table 3.1. Sxl-RBD2 elutes as a broad peak around 0.8 M KCl. Collected fractions (total volume approximately 80 ml) were assayed for protein purity by SDS-PAGE (>95% Sxl-RBD2), concentrated with a Centriprep device (Amicon; 10,000 MW cutoff), dialyzed against H_2O with trace levels of trifluoroacetic acid (TFA), and lyophilized. The resultant powder was stored at -20 °C or dissolved in either 0.5 mL D_2O (99.996%) containing 25 mM d_4-acetic acid pH=5.0, 25 mM NaCl, and 0.01% NaN_3 or 0.5 mL of an identical mixture with 90%/10% H_2O/D_2O. Approximately 25 mg of purified protein was obtained per liter of culture. The final protein concentration was approximately 1.5-2.0 mM. NMR samples were stored at +4 or -20 °C.

Sample purity was assayed with 15% SDS-PAGE and by electrospray-ionization mass spectrometry. Mass spectrometry showed that the first Met residue had been

processed off for approximately 70% of the protein molecules. The total mass observed was in agreement with the calculated mass to within 1 mass unit.

If Sxl-RBD2 was prepared for studies with RNA, a further purification was necessary in order to remove RNases which co-purify on the blue-trisacryl column. This was achieved with HPLC (Waters C-18 delta-pak column) using standard buffers (A: H_2O in 0.01% TFA; B: 40% H_2O, 60% acetonitrile, 0.01% TFA). A gradient from 60% A/40% B to 30% A/70% B over 43 minutes was used, and Sxl-RBD2 elutes as a sharp peak at 32-33 minutes. The resulting protein was RNase-free.

Sxl-RBD1. The expression vector is pAR-Sxl-RNP-1A (also from the Rio lab). The steps up to the blue-trisacryl column are identical to Sxl-RBD2 except that the phosphate buffer should contain 100 mM KCl instead of 25 mM KCl. See Table 3.1 for the blue-trisacryl column gradient program.

Sxl-RBD1+2. The expression vector is pAR-Sxl-RNP-1+2 (again from the Rio lab). The steps through the blue-trisacryl column are identical to Sxl-RBD2 except that RBD1+2 elutes slightly later. RBD1+2 also elutes in a larger volume of approximately 100-200 ml. Therefore, it was found to be more convenient to concentrate protein by a 65% ammonium sulfate precipitation followed by resuspension and dialysis into NMR buffer (25-50 mM phosphate buffer, pH=6.0-6.4, 25-50 mM NaCl). NaN_3 (0.02%) and D_2O (5-10%) were always added after dialysis. It should be noted that Sxl-RBD1+2 can be lyophilized but only in the presence of salt (communication with Stephanie Robertson). Approximately 25-50 mg of purified protein was obtained per liter of culture. RBD1+2 was studied extensively as a complex with 5'-GUUUUUUUUC-3' single stranded RNA, and for these studies, trace levels of RNases (which co-purify off the blue-trisacryl column) had to be removed. Therefore, 2 additional columns were added to the purification. This extended purification, as well as RNase assays and purification of RNA will be discussed in chapter 7.

Table 3.1: Blue Trisacryl FPLC on RBD1, RBD2, and RBD1+2

For all three proteins: column bed = 25 mL
flow rate = 1.0 mL/min

RBD-1:

Buffer A: 0 M KCl
50 mM potassium phosphate, pH=7.0
1 mM DTT
1 mM EDTA
10% glycerol

Buffer B: 2 M KCl
rest same as Buffer A

Gradient:

mL	%B	
0	2.5	(50 mM KCl)
50	2.5	
75	25	(0.5 M KCl)
100	25	
125	100	(2 M KCl)
150	100	wash

RBD-1 elutes between 0.5 and 2 M KCl

RBD-2:

Buffer A: 0 M KCl
50 mM potassium phosphate, pH=7.0
1 mM DTT
1 mM EDTA
10% glycerol

Buffer B: 1 M KCl
rest same as Buffer A

Gradient:

mL	%B	
0	2.5	
50	2.5	
425	70	
475	100	wash

RBD-1+2: Use same buffers and gradient as for RBD-2; RBD-1+2 elutes slightly later than RBD-2.

Baker, B. S. (1989) *Nature 340*, 521-524.

Bell, L. R., Maine, E. M., Schedl, P. & Cline, T. W. (1988) *Cell 55*, 1037-1046.

Birney, E., Kumar, S. & Krainer, A. R. (1993) *Nucl. Acids Res. 21*, 5803-5816.

Burd, C. G. & Dreyfuss, G. (1994) *Science 265*, 615-621.

Cline, T. W. (1979) *Developmental Biology 72*, 266-275.

Cline, T. W. (1993) *Trends in Genetics 9*, 385-390.

Inoue, K., Hoshijima, K., Sakamoto, H. & Shimura, Y. (1990) *Nature 1990*, 461-463.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. (1991) *Embo Journal 10*, 3447-3456.

Kenan, D. J., Query, C. C. & Keene, J. D. (1991) *Trends in Biochemical Sciences 16*, 214-220.

Keyes, L. N., Cline, T. W. & Schedl, P. (1992) *Cell 68*, 933-943.

Nagai, K., Oubridge, C., Jessen, T., Li, J. & Evans, P. R. (1990) *Nature 348*, 515-520.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994) *Nature 372*, 432-438.

Parkhurst, S. M. & Meneely, P. M. (1994) *Science 264*, 924-932.

Sakamoto, H., Inoue, K., Higuchi, I., Ono, Y. & Shimura, Y. (1992) *Nucl. Acids Res. 20*, 5533-5540.

Samuels, M. L., Weber, M. J., Bishop, J. M. & McMahon, M. (1993) *Mol. Cell. Biol. 13*, 6241-6252.

Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J. & Mattaj, I. W. (1990) *Nature 345*, 502-506.

Sosnowski, B. A., Belote, J. M. & McKeown, M. (1989) *Cell 58*, 449-459.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60-89.

Valcarcel, J., Singh, R., Zamore, P. D. & Green, M. R. (1993) *Nature 362*, 171-175.

Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G. & Mueller, L. (1992) *Biochemistry 31*, 6254-6265.

Chapter 4

Resonance Assignments and Secondary Structure in Sxl-RBD2

As mentioned in the last chapter, Sxl-RBD2 has very favorable NMR properties, and therefore structural work on RBD2 was the initial direction that was taken. Before NOESY spectra can be analyzed for distance restraints, however, all (or nearly all) ^{1}H, ^{13}C, and ^{15}N resonances must be assigned. This chapter presents the assignment process for RBD2, the assignments themselves (Table 4.1), the secondary structure in RBD2, and the experimental methods used for obtaining the data used for assignment as well as for structure determination (Chapter 5).

Resonance assignments

Figure 4.1 shows the 2D ^{15}N-^{1}H HSQC correlation spectrum of uniformly ^{15}N-labeled Sxl RBD-2. The wide dispersion of peaks and the absence of additional peaks for individual residues indicates that the protein has a well-defined tertiary fold. Labeled peaks were assigned using the methodology described below.

Assignments for N, HN, and H_α, as well as unambiguous sidechain proton resonances were made from analysis of ^{15}N-separated 3D TOCSY-HMQC and NOESY-HMQC spectra (Kay et al., 1989; Marion et al., 1989b; Driscoll et al., 1990). After identifying spin systems in the TOCSY-HMQC, these spin systems could be specifically identified by residue number if there were sequential NOEs to the preceding or following amino acid in the known sequence (Wüthrich, 1986). HN(i)-HN(i+1), H_α(i)-HN(i+1), and H_β(i)-HN(i+1) connectivities were found in helices, while for residues in an anti-parallel β-sheet, H_α(i)-HN(i+1) and H_β(i)-HN(i+1) were generally found. This analysis was carried out interactively using locally written macros for the NMR processing and

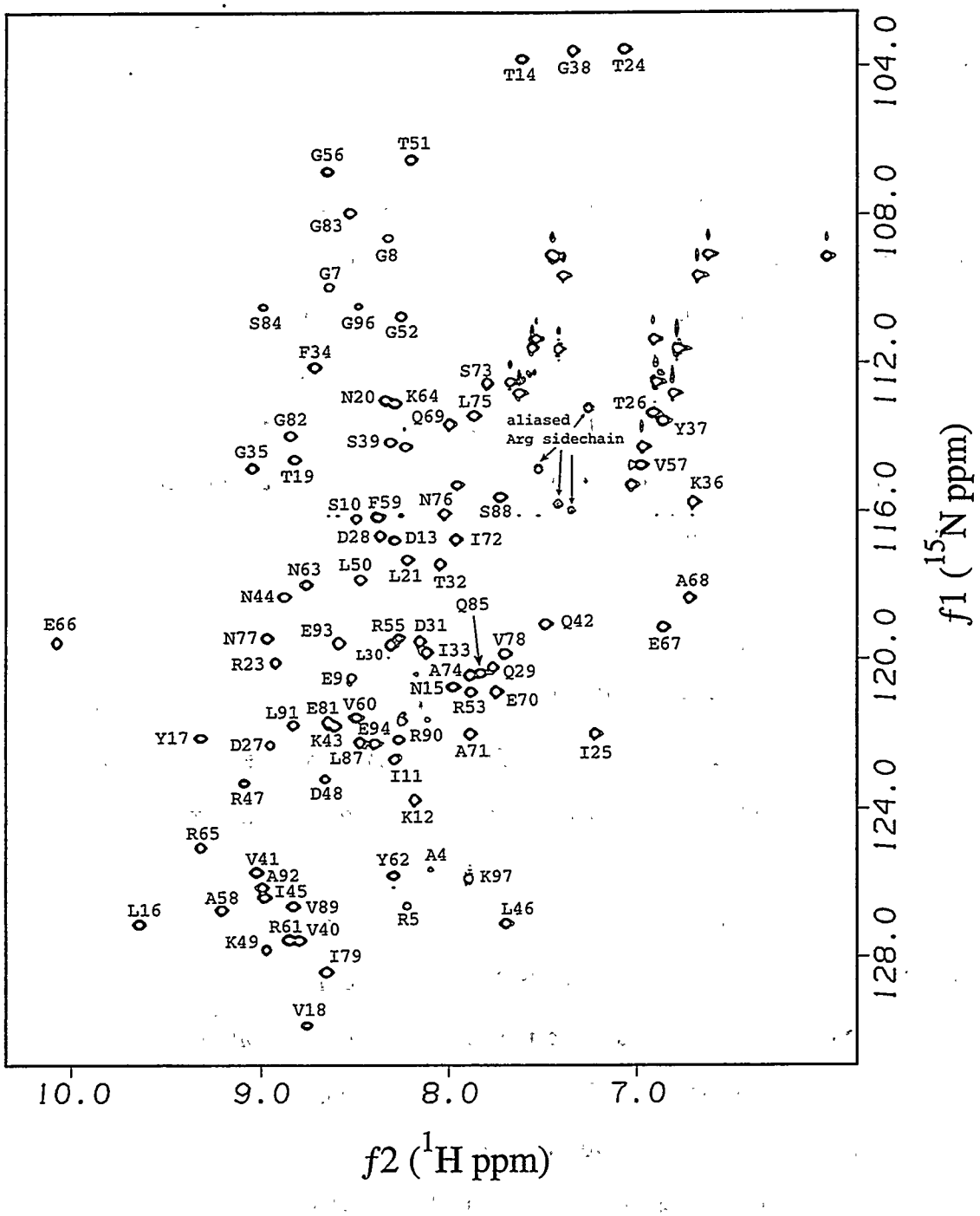


Figure 4.1. 2D HSQC ^{15}N-^{1}H correlation spectrum of uniformly ^{15}N-labeled Sxl RBD-2 and assignments. Unlabeled crosspeaks between 108 and 116 ppm are NH2 groups from Asn and Gln sidechains.

analysis software package, Felix (version 2.14β or 2.10), and they are supplied in Appendix E2. In order to illustrate the quality of the data, Figure 4.2A shows a series of sequential assignments in the form of plotted 2D strips from the 3D NOESY/TOCSY pair. Figures 4.2B-4.2D form a specific example of a sequential assignment from F59 to A58 using Felix macros which are now described for the general case. With the TOCSY-HMQC matrix in one window frame and the NOESY-HMQC matrix in a second frame, $^{1}H(f2)/^{15}N(f3)$ 2D slices at the same $^{1}H(f1)$ frequency were displayed from which a sequential NOESY crosspeak or intraresidue TOCSY crosspeak could be considered (Figure 4.2B, 4.2C). In the case of considering a H_{α}(i-1) sequential NOESY crosspeak at N(i) and HN(i) frequencies, the corresponding intraresidue crosspeak in the TOCSY could be identified by screening all TOCSY spin systems that had peak intensities in the displayed $^{1}H(f2)/^{15}N(f3)$ 2D slice. This screen was performed by displaying all candidate $^{1}H(f1)$ TOCSY 1D vectors in a third window frame (Figure 4.2D) as a stacked plot. These TOCSY spin system candidates could then be lined up with the NOESY ^{1}H(f1) vector which contained the H_{α}(i-1) sequential crosspeak, and the correct i-1 spin system could be confirmed if its chemical shift pattern matched the expected pattern based on the known amino acid sequence. Once this was confirmed, this procedure could be continued by displaying new 2D planes which corresponded to a H_{α}(i-2) sequential NOESY crosspeak at N(i-1) and HN(i-1) frequencies. Following this procedure allows one to "walk" backward through the sequence until a proline or the amino-terminus is encountered. In a similar fashion, a second Felix macro allows moving forward through the sequence by starting from an intraresidue TOCSY crosspeak and screening H_{α}(i+1) sequential NOESY crosspeak candidates. The search can also be conducted from H_{β} or HN resonances instead of H_{α} crosspeaks. This approach to the sequential resonance assignments proved to be convenient since all sequential spin system candidates in three dimensions could be reduced, visualized, and evaluated as a compact spectral subset in the 1D stacked plot

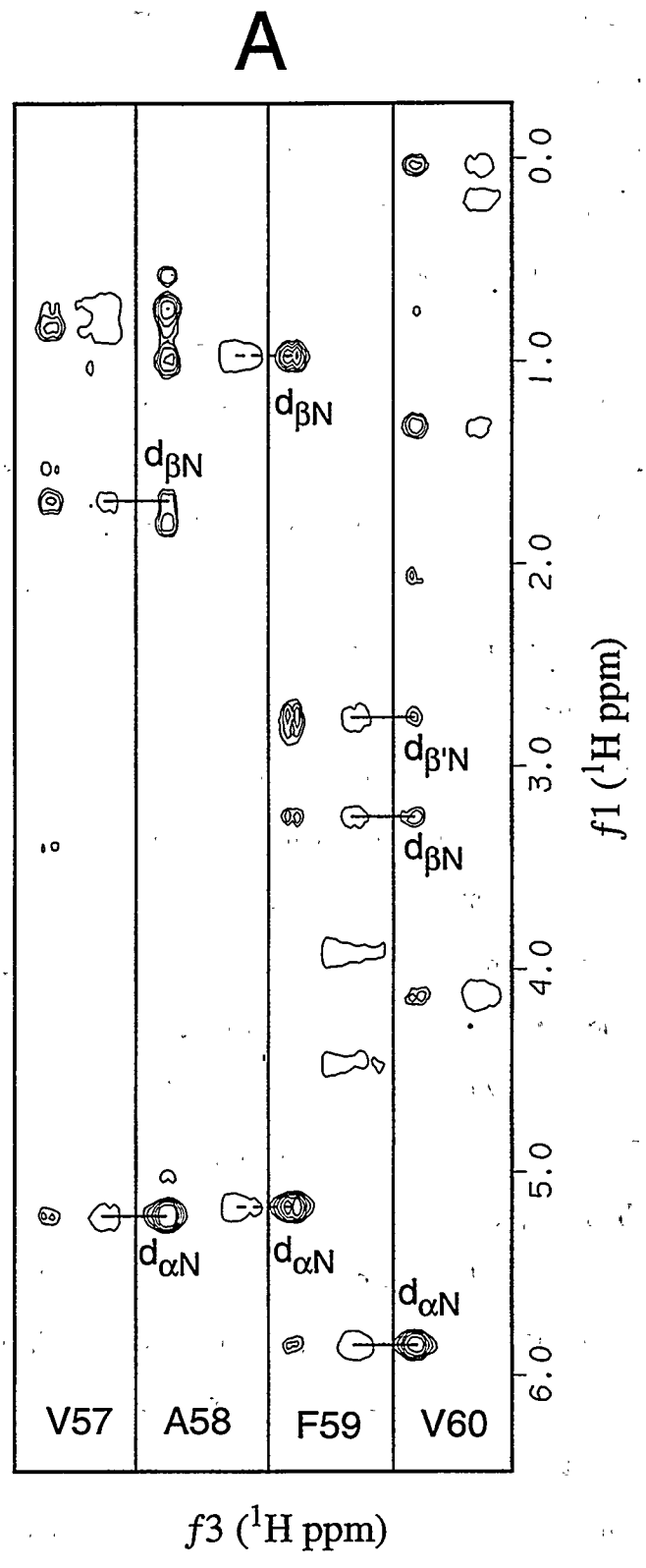


Figure 4.2A. NOESY/TOCSY strip plot. See Figure 4.2 legend on next page for a full description.

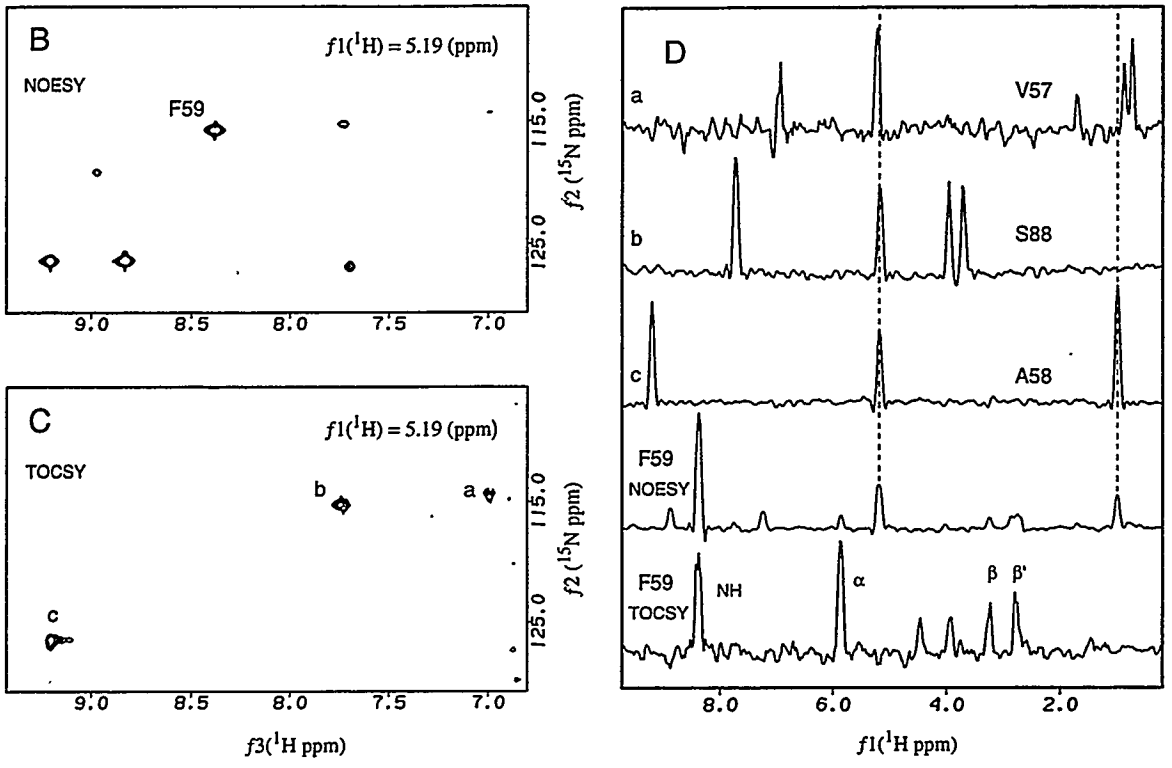


Figure 4.2. Various 2D and 1D slices from ^{15}N-separated 3D NOESY-HMQC and TOCSY-HMQC to illustrate the resonance assignments strategy used. (A) On the previous page, 2D strips are plotted for four different $f_3(^1H)$ and $f_2(^{15}N)$ values, which correspond to residues V57-V60. Within each strip, NOESY peaks are on the left with multiple contour levels, and TOCSY peaks are on the right with a single contour level. Horizontal lines represent sequential NOE connectivities. Figures 4.2B-4.2D display the three window frames viewed while running the Felix assignment macros (see text). Figures 4.2B and 4.2C show a 2D slice at $f_1(^1H) = H_\alpha(i-1)$ from NOESY and TOCSY, respectively. (B) 2D slice from 3D NOESY-HMQC. The peak labeled F59 is a sequential crosspeak from HN(i) of F59 to $H_\alpha(i-1)$ of A58. (C) 2D slice from 3D TOCSY-HMQC. Candidate $H_\alpha(i-1)$ peaks for A58 are labeled a-c. (D) 1D TOCSY vectors along $f_1(^1H)$ corresponding to a-c are shown on the top. The bottom two 1D vectors correspond to $f_2(^{15}N)/f_3(^1H)$ of F59, with H_α, H_β, and H_β' labeled for the TOCSY spin-system. Dashed lines in panels A-D represent searches based on sequential NOESY peaks from F59 HN to A58 H_α and H_β. Because these dashed lines line up with TOCSY spin-system c, c must correspond to A58.

display. In addition, the visualization of very weak resonances is enhanced with this method because there is no need to define a contour threshold as required for 2D plots.

^{13}C assignments and all remaining unassigned proton side-chain resonances (except aromatic rings) were assigned from the ^{13}C-separated 3D HCCH-TOCSY experiment (Bax et al., 1990; Clore et al., 1990). Figure 4.3 displays the high efficiency of magnetization transfer observed for various spin systems. The more complete spin systems here were matched with incomplete proton spin-system data from the ^{15}N-separated 3D TOCSY-HMQC, thereby allowing unambiguous sequence-specific assignments to be made for nearly all sidechain resonances. Exact placement of correlated $^{1}H/^{13}C$ pairs in the sidechains was relatively unambiguous due to the characteristic ^{13}C chemical shifts for amino acid sidechain resonances in proteins (Clore et al., 1990; Ikura et al., 1991). Aromatic resonances were assigned by analysis of the $^{13}C/^{13}C$-separated 4D NOESY (Clore et al., 1991) spectrum. NOEs were generally observed from H_β (and/or H_α) to H_δ. In addition, strong NOEs were generally observed between adjacent ring protons, allowing assignments to be made for most aromatic resonances.

In order to verify and complete the resonance assignments, a triple resonance CBCA(CO)NH spectrum (Grzesiek & Bax, 1992) was acquired. This experiment correlates backbone N(i) and HN(i) with C_α(i-1) and C_β(i-1), using one bond scalar couplings via the backbone carbonyl carbon to bridge sequential resonances. The resultant 3D data proved to be an excellent complement for NOE-based sequential assignments, identifying erroneous assignments derived from proton chemical shift degenericies, as well as allowing the specific assignment of C_α and C_β resonances of Pro residues. RBD2 resonance assignments are given in Table 4.1. Sometime after publishing (Lee et al., 1994) these assignments, I realized that the NH of R90 had been mis-assigned to a very close (~30 Hz difference) peak in the HSQC, and that the NHs in the published table of L91 and A92 had been accidently swapped. Table 4.1 has these corrections incorporated into it.

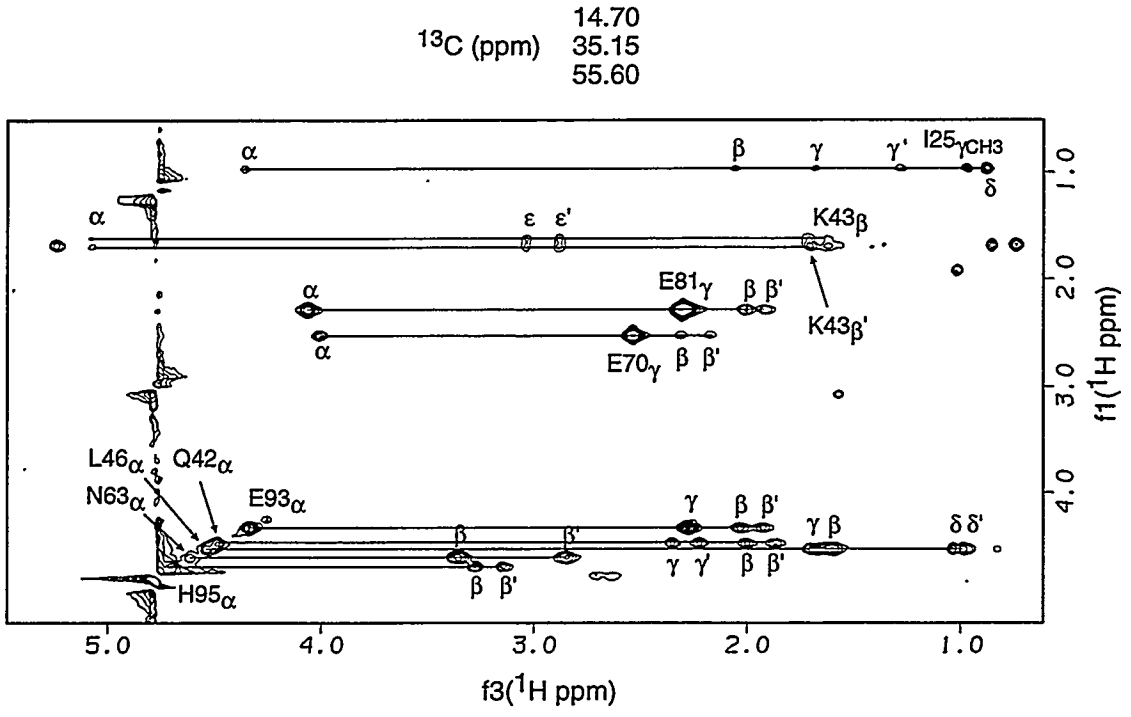


Figure 4.3. 2D slice of ^{13}C-separated 3D HCCH-TOCSY at f_2(^{13}C) = 14.70, 35.15, 55.60 ppm. Three ^{13}C frequences are represented due to aliasing in f_2. All resonances centered in this plane are labeled with their assignments. Lines represent individual amino acid spin-systems.

Table 4.1: ^{15}N, ^{13}C, and ^{1}H Resonance Assignments (in ppm) for Sxl RBD-2 at pH=5.0 and 25° C[a]

Residue	N	C_α	C_β	other
M1	* (*)	* (*)	* (*, *)	C_γ, *(*, *)
S2	* (*)	* (*)	* (*, *)	
Y3	* (*)	* (*)	* (*, *)	C_δ, * (*); C_ε, * (*)
A4	125.6 (8.09)	51.8 (4.29)	19.4 (1.30)	
R5	126.6 (8.22)	* (*)	* (*, *)	C_γ, *(*, *); C_δ, *(*, *)
P6	- (-)	* (4.45)	* (*, *)	C_γ, * (*, *); C_δ, 50.7 (3.67, 3.84)
G7	109.9 (8.63)	* (*, *)		
G8	108.6 (8.32)	* (*, *)		
E9	120.4 (8.50)	56.5 (4.34)	* (1.97, 2.11)	C_γ, 35.9 (2.32, 2.37)
S10	116.2 (8.48)	58.3 (4.48)	63.6 (3.90, 3.95)	
I11	122.6 (8.29)	63.0 (4.20)	38.2 (1.94)	$C_{\gamma1}$, 27.6 (1.20, 1.45); $C_{\gamma2}$, 17.4 (0.94); C_δ, 13.2 (0.86)
K12	123.7 (8.18)	57.4 (4.28)	32.8 (1.81, 1.81)	C_γ, 24.8 (1.43, 1.43); C_δ, 29.1 (1.66, 1.66); C_ε, 41.9 (3.00, 3.00)
D13	116.8 (8.29)	55.1 (4.75)	* (2.88, 2.88)	
T14	103.7 (7.61)	61.2 (4.40)	* (4.75)	C_γ, * (0.87)
N15	120.7 (7.98)	52.2 (5.36)	41.0 (2.54, 3.12)	N_δ, 109.2 (6.00, 7.43)
L16	127.1 (9.64)	53.7 (5.06)	44.1 (1.17, 2.07)	C_γ, 27.0 (1.96); $C_{\delta1}$, 24.8 (0.84); $C_{\delta2}$, 27.3 (1.11)
Y17	122.1 (9.31)	55.9 (4.79)	40.8 (2.63, 2.71)	C_δ, 132.6 (6.34); C_ε, 117.7 (6.37)
V18	129.8 (8.75)	60.0 (5.05)	34.3 (1.81)	$C_{\gamma1}$, 21.2 (0.77); $C_{\gamma2}$, 22.1 (1.06)
T19	114.6 (8.82)	59.3 (5.04)	71.5 (4.21)	C_γ, 22.5 (0.79)
N20	113.0 (8.33)	53.6 (4.32)	37.1 (3.17, 3.17)	
L21	117.3 (8.22)	53.7 (4.14)	41.0 (1.18, 1.23)	C_γ, 26.8 (1.46); $C_{\delta1}$, 24.1 (0.89); $C_{\delta2}$, 27.2 (0.60)
P22	- (-)	62.5 (4.54)	32.2 (1.91, 2.44)	C_γ, * (2.04, 2.04); C_δ, 49.4 (3.40, 3.82)
R23	120.1 (8.92)	57.8 (3.94)	28.9 (1.82, 2.08)	C_γ, 27.9 (1.60, 1.82); C_δ, 42.8 (3.17, 3.17); N_ε, 113.2 (7.28)
T24	103.5 (7.06)	60.1 (4.27)	68.5 (4.65)	C_γ, 21.5 (1.18)
I25	121.9 (7.22)	59.1 (4.35)	39.2 (2.06)	$C_{\gamma1}$, 28.5 (1.28, 1.68); $C_{\gamma2}$, 16.6 (0.88); C_δ, 14.8 (0.97)
T26	113.3 (6.91)	58.4 (4.71)	73.1 (4.67)	C_γ, 21.8 (1.35)
D27	122.2 (8.95)	58.5 (4.17)	40.6 (2.64, 2.64)	

Table 4.1 (cont.)

Residue	N	C_α	C_β	other
D28	116.6 (8.37)	57.0 (4.45)	40.4 (2.60, 2.70)	
Q29	120.2 (7.77)	58.4 (4.18)	29.0 (2.06, 2.23)	C_γ, 33.8 (2.41, 2.52); N_ε, 109.1 (6.62, 7.45)
L30	119.6 (8.31)	58.5 (4.22)	42.0 (1.57, 1.97)	C_γ, 26.6 (1.76); $C_{\delta 1}$, 24.6 (0.74); $C_{\delta 2}$, 26.3 (0.75)
D31	119.5 (8.16)	57.7 (4.37)	40.3 (2.75, 2.95)	
T32	117.4 (8.05)	66.6 (3.94)	68.7 (4.38)	C_γ, 21.8 (1.24)
I33	119.8 (8.12)	65.0 (3.74)	40.0 (1.75)	$C_{\gamma 1}$, 28.5 (1.02, 1.95); $C_{\gamma 2}$, 18.0 (0.40); C_δ, 13.5 (0.83)
F34	112.1 (8.71)	61.0 (4.46)	38.4 (3.01, 3.38)	C_δ, 132.3 (7.80); C_ε, 130.4 (7.01); C_ζ, *(*)
G35	114.8 (9.05)	46.6 (4.29, 4.37)		
K36	115.7 (6.70)	57.2 (3.95)	32.1 (1.08, 1.18)	C_γ, 24.1 (0.93, 0.95); C_δ, 28.8 (1.54, 1.54); C_ε, 41.6 (2.92, 2.92)
Y37	113.5 (6.86)	58.4 (4.41)	38.7 (2.59, 3.36)	C_δ, 132.4 (7.15); C_ε, 118.7 (6.77)
G38	103.5 (7.34)	44.7 (3.97, 4.23)		
S39	114.1 (8.30)	57.6 (4.57)	63.3 (3.94, 3.94)	
I40	127.5 (8.79)	61.5 (3.82)	38.2 (1.68)	$C_{\gamma 1}$, * (*, *); $C_{\gamma 2}$, 17.9 (0.57); C_δ, 14.2 (0.72)
V41	125.7 (9.02)	61.9 (4.21)	32.7 (1.92)	$C_{\gamma 1}$, 19.7 (0.74); $C_{\gamma 2}$, 20.9 (0.88)
Q42	119.0 (7.48)	55.8 (4.49)	32.5 (1.87, 1.99)	C_γ, 33.8 (2.22, 2.34); N_ε, 111.6 (6.79, 7.55)
K43	121.7 (8.61)	54.8 (5.08)	35.2 (1.62, 1.70)	C_γ, 23.6 (1.34, 1.40); C_δ, 29.6 (1.57, 1.57); C_ε, 41.1 (2.88, 3.04)
N44	118.3 (8.87)	52.8 (4.98)	41.2 (2.72, 2.87)	N_δ, 109.7 (6.68, 7.39)
I45	126.4 (8.97)	61.8 (4.03)	38.9 (1.80)	$C_{\gamma 1}$, 29.8 (0.63, 1.64); $C_{\gamma 2}$, 17.4 (0.83); C_δ, 13.5 (0.75)
L46	127.1 (7.69)	55.6 (4.54)	41.2 (1.59, 1.59)	C_γ, 28.4 (1.69); $C_{\delta 1}$, 23.1 (0.98); $C_{\delta 2}$, 25.0 (1.03)
R47	123.3 (9.09)	* (*)	33.4 (*, *)	C_γ, *(*, *); C_δ, *(*, *)
D48	123.2 (8.66)	54.3 (4.54)	42.9 (2.61, 2.97)	
K49	127.8 (8.97)	58.5 (4.04)	32.5 (1.90, *)	C_γ, 24.9 (1.51, 1.51); C_δ, 29.1 (1.73, 1.73); C_ε, 41.9 (*, 3.04)
L50	117.8 (8.47)	56.9 (4.39)	42.1 (1.69, 1.93)	C_γ, 27.3 (1.64); $C_{\delta 1}$, 23.4 (0.91); $C_{\delta 2}$, 24.4 (0.98)
T51	106.4 (8.20)	61.7 (4.45)	71.3 (4.38)	C_γ, 21.0 (1.25)
G52	110.7 (8.26)	45.6 (3.86, 4.25)		
R53	120.8 (7.88)	54.5 (4.66)	30.2 (1.71, 1.91)	C_γ, 27.2 (1.73, 1.73); C_δ, 43.2 (3.24, 3.24)
P54	- (-)	63.5 (4.45)	32.1 (1.96, 2.33)	C_γ, 27.4 (2.04, 2.08); C_δ, * (3.67, 3.85)
R55	119.4 (8.27)	56.2 (4.42)	32.0 (1.53, 2.05)	C_γ, 27.5 (1.53, 1.71); C_δ, 43.6 (3.01, 3.21)

Table 4.1 (cont.)

Residue	N	C_α	C_β	other
G56	106.8 (8.64)	45.6 (3.40, 4.37)		
V57	114.7 (6.98)	58.8 (5.24)	35.4 (1.69)	$C_{\gamma1}$, 20.8 (0.85); $C_{\gamma2}$, 21.9 (0.74)
A58	126.7 (9.21)	50.7 (5.19)	23.8 (0.98)	
F59	116.1 (8.38)	56.3 (5.86)	41.9 (2.78, 3.25)	C_δ, 132.2 (7.23); C_ε, 131.1 (7.28); C_ζ, *(*)
V60	121.5 (8.49)	61.9 (4.13)	36.0 (1.33)	$C_{\gamma1}$, 20.3 (0.05); $C_{\gamma2}$, 20.8 (0.22)
R61	127.5 (8.85)	54.3 (5.59)	35.9 (*, *)	C_γ, * (*, *); C_δ, 44.2 (3.03, 3.03)
Y62	125.7 (8.29)	58.4 (5.28)	42.1 (2.58, 4.10)	C_δ, 132.7 (6.94); C_ε, 117.2 (6.64)
N63	118.0 (8.75)	55.6 (4.61)	40.3 (2.85, 3.36)	N_δ, 114.3 (6.98, 8.24)
K64	113.0 (8.27)	53.9 (4.88)	36.1 (1.60, 2.26)	C_γ, 24.9 (1.48, 1.48); C_δ, 29.1 (1.74, 1.74); C_ε, 41.8 (3.03, 3.03)
R65	125.0 (9.31)	59.3 (3.98)	29.8 (*, *)	C_γ, * (*, *); C_δ, 43.4 (*, 3.27); N_ε, 114.9 (7.51)
E66	119.5 (10.05)	60.1 (4.04)	28.5 (*, *)	C_γ, 36.7 (2.35, 2.53)
E67	119.1 (6.85)	58.1 (3.92)	29.3 (*, *)	C_γ, * (*, *)
A68	118.3 (6.72)	54.9 (3.85)	17.7 (1.59)	
Q69	113.6 (8.00)	58.4 (3.92)	28.4 (2.10, 2.10)	C_γ, 33.7 (2.45, 2.45)
E70	120.8 (7.76)	58.9 (4.01)	28.3 (2.18, 2.31)	C_γ, 35.0 (2.54, 2.54)
A71	121.9 (7.89)	55.2 (3.09)	19.0 (1.57)	
I72	116.7 (7.96)	66.4 (3.29)	38.6 (*)	$C_{\gamma1}$, 28.2 (1.67, *); $C_{\gamma2}$ 17.3 (0.92); C_δ, 14.0 (0.81)
S73	112.5 (7.80)	61.2 (4.11)	63.0 (3.90, *)	
A74	120.4 (7.88)	53.8 (4.19)	20.4 (1.10)	
L75	113.4 (7.87)	54.7 (4.47)	44.9 (*, *)	C_γ, * (*); $C_{\delta1}$, 22.4 (0.66); $C_{\delta2}$, 27.3 (0.21)
N76	116.0 (8.02)	56.2 (4.34)	38.9 (2.92, 3.03)	N_δ, 115.3 (7.03, 7.96)
N77	119.4 (8.97)	55.1 (4.39)	37.7 (2.95, 3.17)	N_δ, 112.8 (6.81, 7.62)
V78	119.8 (7.70)	61.4 (4.22)	34.3 (2.03)	$C_{\gamma1}$, 21.1 (0.85); $C_{\gamma2}$, 21.1 (0.85)
I79	128.4 (8.64)	58.9 (4.41)	* (1.69)	$C_{\gamma1}$, 27.9 (0.30, 1.57); $C_{\gamma2}$, 16.6 (0.67); C_δ, 13.2 (0.78)
P80	- (-)	62.2 (4.45)	32.1 (*)	C_γ, *(*, *); C_δ, *(*, *)
E81	121.6 (8.64)	58.2 (4.06)	28.9 (1.91, 2.01)	C_γ, 35.5 (2.30, 2.30)
G82	113.9 (8.84)	45.1 (3.73, 4.21)		
G83	107.9 (8.52)	43.6 (3.56, 4.56)		

Table 4.1 (cont.)

Residue	N	C_α	C_β	other
S84	110.4 (8.99)	58.1 (4.72)	65.0 (3.79, 3.93)	
Q85	120.3 (7.83)	52.6 (4.92)	* (1.90, 2.11)	C_γ, 33.2 (2.35, 2.35); N_ε, 111.6 (6.79, 7.41)
P86	- (-)		62.0 (4.14)	31.3 (1.78, 1.78) C_γ, *(*, *); C_δ, *(*, *)
L87	122.2 (8.47)	55.1 (4.39)	43.4 (1.29, 2.23)	C_γ, 21.9 (0.86); $C_{\delta 1}$, 24.2 (0.83); $C_{\delta 2}$, 26.1 (0.85)
S88	115.6 (7.73)	56.9 (5.17)	63.1 (3.71, 3.96)	
V89	126.6 (8.82)	61.5 (4.66)	34.9 (1.93)	$C_{\gamma 1}$, 21.8 (1.02); $C_{\gamma 2}$, 24.3 (1.02)
R90	122.1 (8.27)	53.7 (4.55)	30.5 (1.89, 1.97)	C_γ, 26.8 (1.67, 1.67); C_δ, 43.3 (*, 3.23)
L91	121.7 (8.82)	56.1 (4.54)	41.7 (1.68, 1.68)	C_γ, * (*, *); $C_{\delta 1}$, 23.1 (0.82); $C_{\delta 2}$, 25.4 (0.95)
A92	126.1 (8.99)	53.2 (4.24)	19.4 (1.48)	
E93	119.5 (8.58)	55.9 (4.33)	30.2 (1.92, 2.03)	C_γ, 35.7 (2.27, 2.27)
E94	122.2 (8.39)	56.3 (4.25)	30.5 (1.88, 1.98)	C_γ, 35.8 (2.19, 2.23)
H95	119.1 (8.61)	55.2 (4.70)	29.1 (3.14, 3.28)	C_δ, 119.9 (7.27); C_ε, 136.4 (8.53)
G96	110.4 (8.49)			
K97	125.8 (7.90)	57.5 (4.17)	33.6 (1.70, 1.82)	C_γ, 24.8 (1.37, 1.37); C_δ, 29.1 (1.64, 1.64); C_ε, 41.8 (2.98, 2.98)

[a] In each column, ^{15}N or ^{13}C shift precedes the corresponding ^{1}H shift given in parentheses. An asterisk indicates that the chemical shift was not determined. Methyl group resonances for Val and Leu residues have not been stereospecifically assigned.

Secondary structure

The secondary structure in Sxl RBD-2 consists of two α-helices and one anti-parallel β-sheet. The characteristic NOE patterns for an α-helix include the following: HN(i)-HN(i+1), H_β(i)-HN(i+1), H_α(i)-HN(i+3), H_α(i)-H_β(i+3), H_α(i)-HN(i+4). The characteristic NOE pattern for an anti-parallel β-sheet is a strong H_α(i)-HN(i+1) along with long-range NOEs to opposite strands, such as strong H_α-H_α and weak H_α-HN, HN-H_α, and HN-HN NOEs (Figure 4.5) (Wüthrich, 1986). Backbone amide protons involved in secondary structure are often hydrogen bonded in a regular fashion, giving rise to slow exchange of these protons. Based on these criteria, the secondary structure in Sxl RBD-2 was mapped and is summarized in Figure 4.4. Recently, carbon chemical shifts have been observed to correlate with secondary structural elements (Spera & Bax, 1991; Wishart et al., 1991). It has been observed that helices give rise to upfield shifts from random coil values of a few ppm for C_α, as well as a small downfield shift for C_β. An opposite trend has been observed for β-sheets. Figure 4.4 also reports C_α and C_β deviations from random coil values.

The first α-helix begins with T26 and ends at G35. T26 is the N-Cap residue in a capping box, an α-helix initiating signal comprised of four amino acids which recently has been recognized in both peptides and proteins (Harper & Rose, 1993). In a capping box, the sidechain of the first (N-Cap) of these four residues forms a hydrogen bond with the backbone amide of the fourth (N3) residue, while in turn a reciprocal hydrogen bond from the N3 residue sidechain back to the backbone amide of the N-Cap residue completes the "box". With such an interaction, two net hydrogen bonds are gained, which has been shown to be energetically favorable. Hence the N-Cap and N3 residues are preferably Thr, Ser, Asn, Gln, Asp, or Glu, since the sidechain carbonyl or hydroxyl oxygen serves as a suitable hydrogen bond acceptor. A capping box has previously been proposed for hnRNP-A1 based on sequence analysis and ^{13}C chemical shift data (Garrett et al., 1994). In the case of Sxl RBD-2, characteristic NOEs, chemical shifts, and amide protection

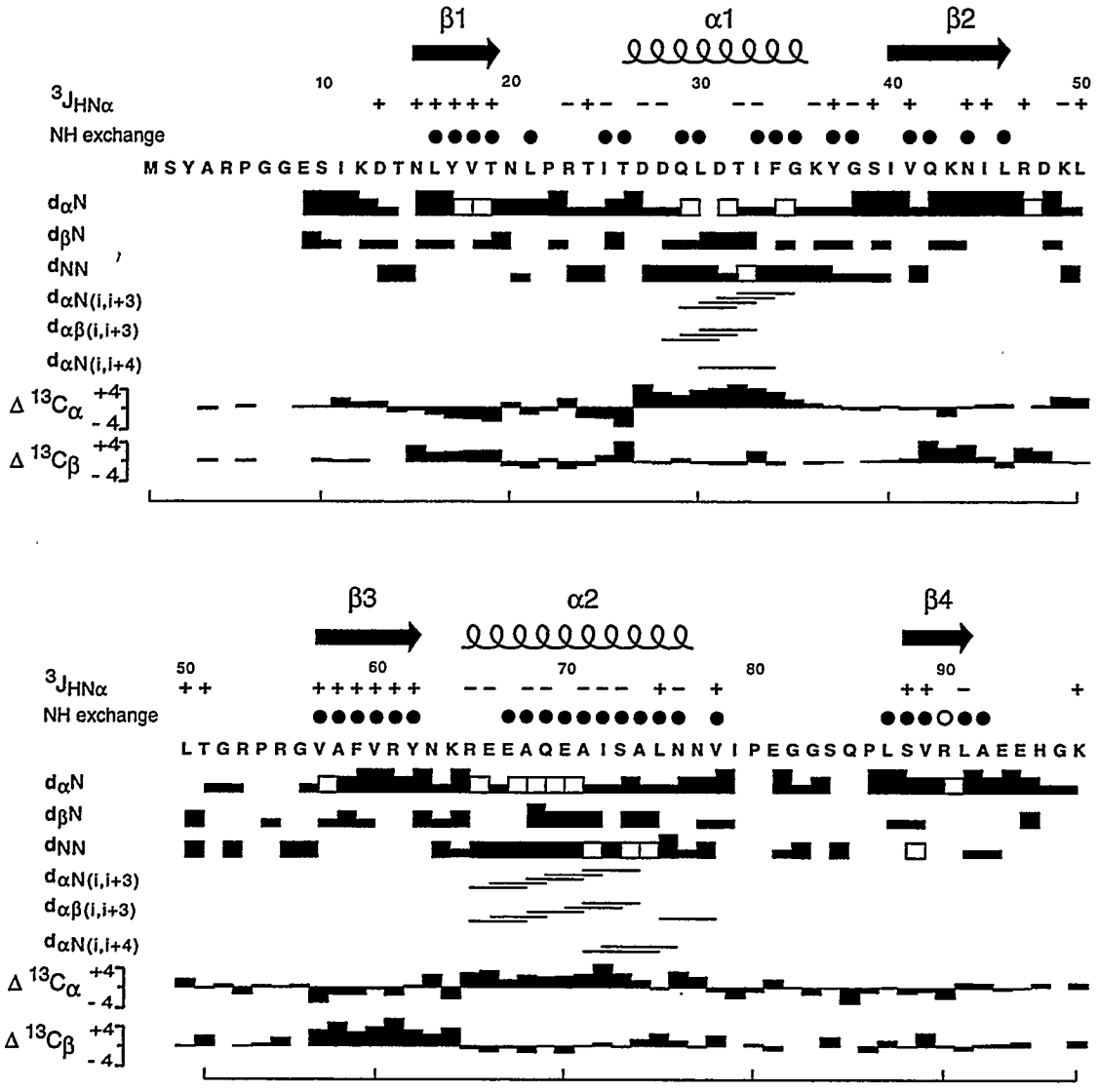


Figure 4.4. Summary of sequential and medium-range NOE connectivities, amide hydrogen exchange rates, ^{3}JHNa coupling constants, and deviations of $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts from random coil values. NOE intensities are denoted by bar height as weak, medium, or strong. Open boxes indicate ambiguity due to chemical shift degeneracy or solvent presaturation. For Pro residues, NOE intensities to their H_δ protons were substituted for NOE intensities to HN protons. Filled circles indicate residues with slowed amide exchange (HN resonances observable three hours after dissolving lyophilized protein into 99.9% D_2O). The open circle for R90 indicates uncertainty since this residue has been re-assigned, but its exchange rate has not been re-checked. Pluses and minuses represent ^{3}JHNa coupling constants greater than 8 Hz or less than 6 Hz, respectively. Secondary structure as determined by NOESY data is indicated above the sequence numbering.

patterns are observed as expected for the capping box. There is a medium intensity NOE from T26 HN to Q29 H_β, a strong NOE from T26 H_α to D27 HN, and NOEs from the sidechain protons of T26 to D28 HN. This NOE pattern is consistent with the amide of T26 being oriented nearly anti-parallel to all other amides in helix 1, allowing the sidechain (Q29 O_ε)-to-mainchain (T26 NH) hydrogen bond. A similar NOE pattern was observed for a capping box in a short peptide (Zhou et al., 1994). Chemical shift values for C_α and C_β of T26 have shifts from random coil values of -4 and +3 ppm, respectively. These shifts are characteristic for capping boxes (Gronenborn & Clore, 1994). Finally, the amides of T26 and Q29 are protected from exchange with solvent (Figure 4.4), while the amides of D27 and D28 exchange rapidly. This pattern of amide protection is consistent with the previous model of a capping box (Zhou et al., 1994). Along with the first α-helical hydrogen bond from L30 NH to T26 CO, T26 is completely fulfilled with three hydrogen bonding interactions. From the calculated structures, the values for ϕ and ψ of T26 are $-105 \pm 21°$ and $150 \pm 10°$, respectively, which is consistent with other observed capping boxes (Harper & Rose, 1993). The C-terminal end of helix 1 will be discussed in Chapter 5. The second α-helix begins at R65 and ends with N76, running slightly longer than helix 1. It is very well-defined due to a continuous stretch of helical-type NOEs (Figure 4.4) and hydrogen bonds (as detected by slowed amide exchange).

For the previously structurally characterized snRNP and hnRNP RBD proteins (Nagai et al., 1990; Wittekind et al., 1992), it has been suggested that a 4-stranded anti-parallel β-sheet provides a "platform" for RNA binding. This could be the case for Sxl RBD-2 because it also contains a 4-stranded anti-parallel β-sheet. The topology as well as NOEs and hydrogen bonds for this sheet are shown in Figure 4.5. Strands β1 and β3 contain the highly conserved RNP-2 and RNP-1 consensus sequences, respectively. The two aromatic sidechains of Y17 and F59 are expected to interact directly with *tra* pre-mRNA (Merrill et al., 1988; Hoffman et al., 1991), and they are situated directly cross-strand from one another in the center of the sheet, which forms a regular pleated pattern.

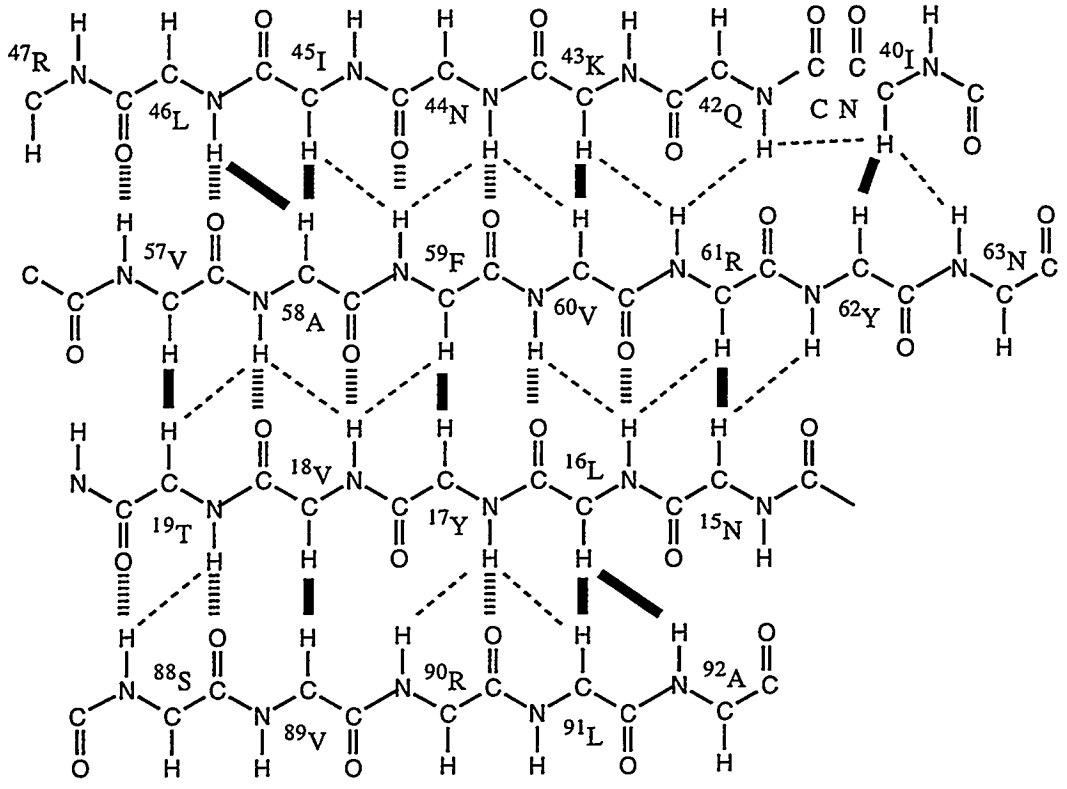


Figure 4.5. Anti-parallel β-sheet of Sxl RBD-2. Residues are labeled at their own C_α. Observed weak NOEs are represented by dashed lines. Observed medium or strong NOEs are represented by thick lines. Hydrogen bonds consistent with NOE and exchange data are represented by hatched lines.

There is a β-bulge involving I40 and V41, which appear similar to residues I40 and L41 in the snRNP-U1A X-ray crystal structure (Nagai et al., 1990). Figure 4.5 shows the unusual pattern of NOEs resulting from the bulge. The bulge can also be visualized in a ribbon diagram of Sxl RBD-2 shown in Figure 5.2B. Aside from this β-bulge, the only other unusual feature of the sheet structure is the fourth strand, S88-L91. V89 HN is oriented down towards helix 2, giving rise to numerous NOEs between it and helix 2, and a $^3J_{HN\alpha}$ value < 6 Hz is observed for L91, where one might expect a value > 8 Hz for an extended β conformation. In the calculated structures (Chapter 5), residues S88-L91 appear to be in an extended conformation, but the strand as a whole is tilted away from the rest of the sheet. This feature is also observed in the crystal structure of snRNP-U1A.

NMR spectroscopy methods

All NMR experiments were performed at 25° C on a Bruker AMX 600 spectrometer equipped with an external ENI 50 W linear amplifier and triple-resonance (^1H, ^{13}C, ^{15}N) probehead. Chemical shifts were referenced to 3-(2,2,3,3-$^2H_4$) trimethylsilyl propionate (TSP) (^1H), liquid ammonia (^{15}N) (Live et al., 1984), and TSP (^{13}C) (Bax & Subramanian, 1986). Quadrature detection using TPPI-States (Marion et al., 1989) was employed for all experiments except for 3D NOESY-HMQC and 3D TOCSY-HMQC, in which case TPPI (Marion & Wüthrich, 1983) was used. All data manipulations as well as assignment analysis were accomplished with the NMR processing program Felix (2.14β or 2.30β). Linear prediction was carried out using the algorithm supplied with the Felix software package (Barhuijsen et al., 1985; Biosym Technologies, 1993).

^{15}N-separated 3D NOESY-HMQC (Kay et al., 1989; Marion et al., 1989b) and 3D TOCSY-HMQC (Driscoll et al., 1990) with mixing times of 100 ms each were acquired with spectral widths of 7246 Hz for ^1H dimensions (t_1 and t_3) and 1861 Hz for the ^{15}N (t_2) dimension. In the case of TOCSY-HMQC, a compensation delay during the MLEV-17 spin-lock was used to suppress NOESY-type cross peaks (Griesinger et al., 1988). The

number of real data points collected for t_1, t_2, and t_3 were 220, 64, and 2048, respectively. The time domain data were processed with a 70° phase shifted sine-bell apodization function in each dimension, with linear prediction used in t_2 to double the number of data points prior to apodization. Because the ^1H carrier in t_3 was placed at 4.78 ppm (H_2O resonance), the upfield data points in this dimension were discarded following Fourier transformation. Zero-filling was included for t_1 to yield a 512 x 64 x 512 real data matrix.

The triple resonance 3D CBCA(CO)NH (Grzesiek & Bax, 1992) experiment was acquired with spectral widths of 8446, 1861, and 7246 Hz for $t_1(^{13}C)$, $t_2(^{15}N)$, and $t_3(^1H)$, respectively. Total acquisition time was 68 hours. Using processing similar to 3D NOESY and TOCSY above, a 128 x 64 x 512 real data matrix resulted from 48, 32, and 512 complex data points, respectively. Zero-filling was used in t_1 only.

^{13}C-separated 3D HCCH-TOCSY was used to assign sidechain ^1H and ^{13}C resonances (Bax et al., 1990). Carrier frequencies were centered at 3.00 ppm and 42.94 ppm for proton and carbon, respectively. Spectral widths of 4348 Hz (t_1 and t_3) and 3086 Hz (t_2) resulted in a spectrum aliased in t_2 only. Signal aliasing from upfield and downfield is desirable for the ^{13}C dimension because it increases digital resolution up to 3-fold without introducing ambiguities (Clore et al., 1990). 128(^1H), 32(^{13}C), and 256(^1H) complex data points were processed into a 256 x 64 x 512 real data matrix. As usual, linear prediction was used for the dimension with the lowest digital resolution (t_2 in this experiment), and zero-filling was used whenever possible.

The $^{13}C/^{13}C$-separated 4D HMQC-NOESY-HMQC (^{13}C-^1H-^{13}C-^1H) experiment (Clore et al., 1991) was acquired over 88 hours with a mixing time of 100 ms. Carrier frequencies were centered at 3.81 ppm and 42.94 ppm for proton and carbon, respectively, and aliasing similar to the 3D HCCH-TOCSY was employed for both ^{13}C dimensions. 8, 64, 8, and 256 complex points for each dimension were acquired with corresponding spectral widths of 3086, 4808, 3086, and 5208 Hz. After data acquisition, t_4 was processed with a 70° phase shifted sine-bell window into a 32 x 128 x 32 x 256 real data

matrix followed by similar processing in t_2. Next, t_3 was only transformed (not apodized) in order to spread out the signals to insure accurate linear prediction in t_1 (Clore et al., 1991). Finally, t_3 was inverse transformed, linear predicted, and processed to yield the fully processed frequency domain matrix. In both t_1 and t_3, linear prediction was used to extend the data by 50%, and zero-filling was used in t_1, t_2, and t_3.

2D HSQC experiments (Bodenhausen & Ruben, 1980) were used to initially judge the quality of samples, as well as to accurately define resonances due to the high digital resolution available. In the case of ^{15}N-^{1}H HSQC, amide ^{1}H exchange rate information was also obtained. For the ^{13}C-^{1}H correlation spectra, a constant-time evolution period (2T equal to $1/J_{CC} = 26$ ms) was used to eliminate ^{13}C-^{13}C couplings in t_1 (Vuister & Bax, 1992). In the case of aliphatic ^{13}C-^{1}H resonances, the carbon carrier frequency was placed at 46.13 ppm, and spectral widths of 5000 Hz (128 complex points) in t_1 and 7246 Hz (2048 real points) in t_2 were used. In the case of aromatic ^{13}C-^{1}H resonances, a second HSQC was acquired with the carbon carrier placed at 125.13 ppm, using spectral widths of 3968 Hz (56 complex) in t_1 and 7246 Hz (2048 real) in t_2. The aromatic region of the resultant spectrum has a small number of resonances. In addition, the aromatic rings have no carbonyl groups and very few nitrogens. For these reasons, carbonyl and nitrogen decoupling pulses were removed from the pulse sequence. For the ^{15}N-^{1}H HSQC spectra, the ^{15}N carrier was placed at 116.2 ppm, with spectral widths of 1861 Hz (256 complex) in t_1 and 7246 Hz (2048 real) in t_2. For the purpose of estimating amide exchange rates, a rapid-acquisition pulse-program (Marion et al., 1989) was used to collect early time points for a series of ^{15}N-^{1}H HSQC experiments on Sxl RBD-2 immediately following dissolution in 99.9% D_2O, where an entire 2D experiment with 128 complex t_1 points is collected in 6 minutes.

The 2D HMQC-J experiment (Kay & Bax, 1990) was used to obtain qualitative $^{3}J_{HN\alpha}$ coupling constants, which in turn could be used to define ϕ dihedral angle ranges via the Karplus relation. For J values > 8 Hz, a ϕ dihedral restraint of $-120 \pm 40°$ was

implemented in structure calculations. For J values < 6 Hz, a ϕ dihedral restraint of -60 ± 30° was implemented if that residue was known to occur in a helical region. Data aquisition parameters are identical to the ^{15}N-^{1}H HSQC experiments, except that aquisition in t_1 was extended to 200 msec. An exponential line broadening of 7 Hz was applied in the directly detected dimension as described (Kay & Bax, 1990).

Barhuijsen, H., de Beer, R., Bouee, W. M. M. J. & van Ormondt, D. (1985) *J. Magn. Reson. 61*, 465-480.

Bax, A. & Subramanian, S. (1986) *J. Magn. Reson. 67*, 565-569.

Bax, A., Clore, G. M. & Gronenborn, A. M. (1990) *J. Magn. Reson. 88*, 425-431.

Biosym Technologies, I. (1993) *Felix User Guide Version 2.3*, San Diego, CA.

Bodenhausen, G. & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185-189.

Clore, G. M., Bax, A., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. (1990) *Biochemistry 29*, 8172-8184.

Clore, G. M., Kay, L. E., Bax, A. & Gronenborn, A. M. (1991) *Biochemistry 30*, 12-18.

Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T. & Gronenborn, A. M. (1990) *Biochemistry 29*, 3542-3556.

Garrett, D. S., Lodi, P. J., Shamoo, Y., Williams, K. R., Clore, G. M. & Gronenborn, A. M. (1994) *Biochemistry 33*, 2852-2858.

Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988) *J. Magn. Reson. 110*, 7870-7820.

Gronenborn, A. M. & Clore, G. M. (1994) *J. Biomol. NMR 4*, 455-458.

Grzesiek, S. & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 6291-6293.

Harper, E. T. & Rose, G. D. (1993) *Biochemistry 32*, 7605-7609.

Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W. & Keene, J. D. (1991) *Proceedings of the National Academy of Sciences 88*, 2495-2499.

Ikura, M., Spera, S., Barbato, G., Kay, L. E., Krinks, M. & Bax, A. (1991) *Biochemistry 30*, 9216-9228.

Kay, L. E., Marion, D. & Bax, A. (1989) *J. Magn. Reson. 84*, 72-84.

Kay, L. E. & Bax, A. (1990) *J. Magn. Reson. 86*, 110-126.

Lee, A.L., Kanaar, R., Rio, D.C. & Wemmer, D.E. (1994) *Biochemistry 33*, 13775-13786.

Live, D. H., Davis, D. G., Agosta, W. C. & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1934-1941.

Marion, D. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Comm. 113*, 967-974.

Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989) *J. Magn. Reson. 85*, 393-399.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989b) *J. Am. Chem. Soc. 111*, 1515-1517.

Merrill, B. M., Stone, K. L., Cobianch, F., Wilson, S. H. & Williams, K. R. (1988) *J. Biol. Chem. 263*, 3307-3313.

Nagai, K., Oubridge, C., Jessen, T., Li, J. & Evans, P. R. (1990) *Nature 348*, 515-520.

Spera, S. & Bax, A. (1991) *J. Am. Chem. Soc. 113*, 5490-5492.

Vuister, G. W. & Bax, A. (1992) *J. Magn. Reson. 98*, 428-435.

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

Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G. & Mueller, L. (1992) *Biochemistry 31*, 6254-6265.

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

Zhou, H. X., Lyu, P. C., Wemmer, D. E. & Kallenbach, N. R. (1994) *PROTEINS: Structure, Function, and Genetics 18*, 1-7.

Chapter 5

Structure Determination of Sxl-RBD2

Expressed, purified, and uniformly isotopically labeled (^{15}N or $^{13}C/^{15}N$) Sxl-RBD2 protein has been 90% assigned using a combination of ^{15}N-separated 3D TOCSY-HMQC and NOESY-HMQC, ^{13}C-separated 3D HCCH-TOCSY, and triple resonance 3D CBCA(CO)NH spectra (Chapter 4). In the present chapter, the structure determination process will be described, and the results of the RBD2 structure calculations will be presented.

NOE-derived restraints

In order to carry out structure calculations, interproton NOEs were assigned from the ^{15}N-separated 3D HMQC-NOESY and $^{13}C/^{13}C$-separated 4D HMQC-NOESY-HMQC. A large number of intra-residue and short-range NOEs had been previously identified during the assignment of the backbone. In order to expand this distance restraint list to contain as many long-range NOEs as possible, a systematic approach was adopted for the reliable and consistent extraction of unambiguous NOE restraints. The process was begun by attempting to manually pick all crosspeaks in the 3D and 4D NOESY spectra. In the case of the 3D NOESY, picking peaks is relatively straightforward using Felix 2.30. However, picking 4D peaks was not explicitly supported by this software. Fortunately, Felix has a relatively powerful macro language which allowed a few macros to be written which facilitated accurate picking and tabulation of 4D NOESY crosspeaks. These macros are provided as documentation in Appendix E2. Each NOE was given an intensity of weak, medium, or strong based on the number of contour levels for that peak. Medium and long-range (as well as more intra-residue and short-range) NOEs were assigned by

matching NOE crosspeak frequencies against existing resonance assignments. As this task is quite arduous, I wrote macros to do this repetitive matching process in Microsoft Excel 4.0, and these macros are also provided as documentation in Appendix F. In the end, a total of 241 unambiguous NOEs were obtained from the ^{15}N-separated 3D NOESY-HMQC and 253 unambiguous NOEs were obtained from the $^{13}C/^{13}C$-separated 4D HMQC-NOESY-HMQC.

Structure calculations methods

The program X-PLOR 3.1 (Brünger, 1992) was used for all structure calculations on a Silicon Graphics (Mountain View, CA) workstation. The calculation protocol consisted of distance geometry on substructures comprised of backbone as well as C_β and C_γ atoms, followed by high temperature molecular dynamics, simulated annealing, and Powell energy minimization as described (Nilges et al., 1988; Brünger, 1992). More details of the protocol are presented in the caption for Table 5.1. The calculations were carried out in an iterative fashion, initially using completely unambiguously assigned NOEs. The resultant structures were used to evaluate additional NOEs which were then included to increase the number of restraints. After the global fold had been established, torsional and hydrogen bond restraints were added. NOEs were classified qualitatively as strong (1.8-2.7 Å), medium (1.8-3.5 Å), or weak (1.8-5.0 Å) (Williamson et al., 1985; Clore et al., 1986). As a reference, NOEs between H_δ and H_ε in Tyr residues were considered strong, while NOEs between $H_\alpha(i)$ and HN(i+3, i+4) in α-helices were considered weak. A correction was added to the upper limit for constraints involving methyl protons and non-stereospecifically assigned protons (Wüthrich et al., 1983; Clore et al., 1986). Hydrogen bonds involving slowly exchanging amide protons were included if a unique acceptor could be identified from the structures and the pattern of NOEs. A total of 26 hydrogen bonds, included as 26 HN$\leftrightarrow$O (1.8-2.3 Å) and 26 N$\leftrightarrow$O (2.8-3.3 Å) NOE-type distance restraints, were identified.

Table 5.1: Structural Statistics and RMSDs for 17 Sxl RBD-2 Structures[a]

structural statistics	<SA>
X-PLOR energies (kcal mol^{-1})	
E_{tot}	293 ± 43
E_{bond}	13 ± 3
E_{angle}	170 ± 14
E_{impr}	24 ± 3
E_{vdw}[b]	46 ± 15
E_{noe}[c]	37 ± 13
E_{cdih}[d]	2.1 ± 1.1
RMSD from idealized covalent geometry	
bonds (Å)	0.0030 ± 0.0003
angles (deg)	0.63 ± 0.02
impropers (deg)	0.45 ± 0.03
rms deviations from experimental distance restraints (Å)[e]	0.036 ± 0.006
rms deviations from experimental dihedral restraints (deg)[f]	0.91 ± 0.26

Atomic RMSDs (Å)	N, C_α, C', O	all non-H
<SA> vs. $< \overline{SA} >_{sxl}$	1.55 ± 0.30	2.20 ± 0.32
<SA> vs. $< \overline{SA} >_{str}$	0.76 ± 0.14	1.35 ± 0.13

[a] Notation is as follows: <SA> is the ensemble of 17 final X-PLOR structures. $< \overline{SA} >_{sxl}$ is the mean coordinate set for residues 11-93 obtained from a least-squares superposition to backbone or non-H atoms. $< \overline{SA} >_{str}$ is the mean coordinate set for residues involved in secondary structure (15-19, 26-35, 40-46, 57-62, 65-77, 88-91) which was obtained from a least-squares superposition to backbone or non-H atoms. [b] The X-PLOR F_{repel} function was used to simulate the van der Waals potential with atomic radii ranging from 0.9 times their CHARMM (Brooks et al., 1983) values at high temperatures to 0.75 times their CHARMM values at low temperatures (Brünger, 1992). [c] NOE-derived distance restraints were applied with a square-well potential with force constants of 50 kcal·mol^{-1}·Å^{-2}. [d] Dihedral angles were given force constants of 200 kcal·mol^{-1}·rad^{-2} which were applied at the beginning of the annealing/refinement stage. [e] The single NOE violation > 0.5 Å in the family of 17 structures is 0.61 Å. [f] There is one dihedral violation > 5° in the family of 17 structures, and it is 5.40°.

Calculated structures of RBD2

A total of 587 experimental restraints were included in the final round of structure calculations. This total consisted of 50 intraresidue, 185 sequential, 90 medium-range ($|i-j| \leq 4$), and 169 long-range NOEs, as well as 41 ϕ dihedral and 52 hydrogen bond restraints (from 26 hydrogen bonds). Hydrogen bonds were included for non-exchangeable amides in stretches of regular secondary structure where hydrogen-bond acceptors were unambiguous. In the case of the capping box, two additional hydrogen bonds were included. Justification for these is given below. The corresponding restraint list (round 22) is provided in Appendix C. Of the 30 structures calculated, 17 converged to reasonably low energies. This sub-family had a single NOE violation greater than 0.5 Å. Table 5.1 and Figure 5.1 contain calculation statistics for the 17 accepted structures. A superposition to the average coordinates for these 17 structures is shown in Figure 5.2A and exhibits the well defined backbone atoms involved in secondary structure. The RMSD to the average coordinates for backbone atoms (of the accepted family) involved in secondary structure is 0.76 ± 0.14 Å. The RMSD for backbone atoms of residues 11-93 is 1.55 ± 0.30 Å. This difference is primarily due to the relatively unrestrained loop regions (47-56, 79-86) connecting elements of secondary structure. The RMSD for all non-hydrogen atoms for residues 11-93 is higher at 2.20 ± 0.32 Å. However, Figure 5.1C shows that residues involved in secondary structure frequently have non-hydrogen RMSDs below 1 Å, especially residues in the hydrophobic core (see below). Usually this is a consequence of a relatively large number of long-range NOEs for these particular residues, as shown in Figure 5.1A.

The folding topology of Sxl RBD-2 is similar to previously described RNA-binding proteins, consisting of a $\beta\alpha\beta$-$\beta\alpha\beta$ folding pattern (Figure 5.2B). However, Sxl RBD-2 has some significant differences, such as the ten residues between β2 and β3 as opposed to five, four, and eight residues in snRNP-U1A, hnRNP-C, and hnRNP-A1, respectively. This turn or loop has been implicated in conferring specificity to binding RNA for the

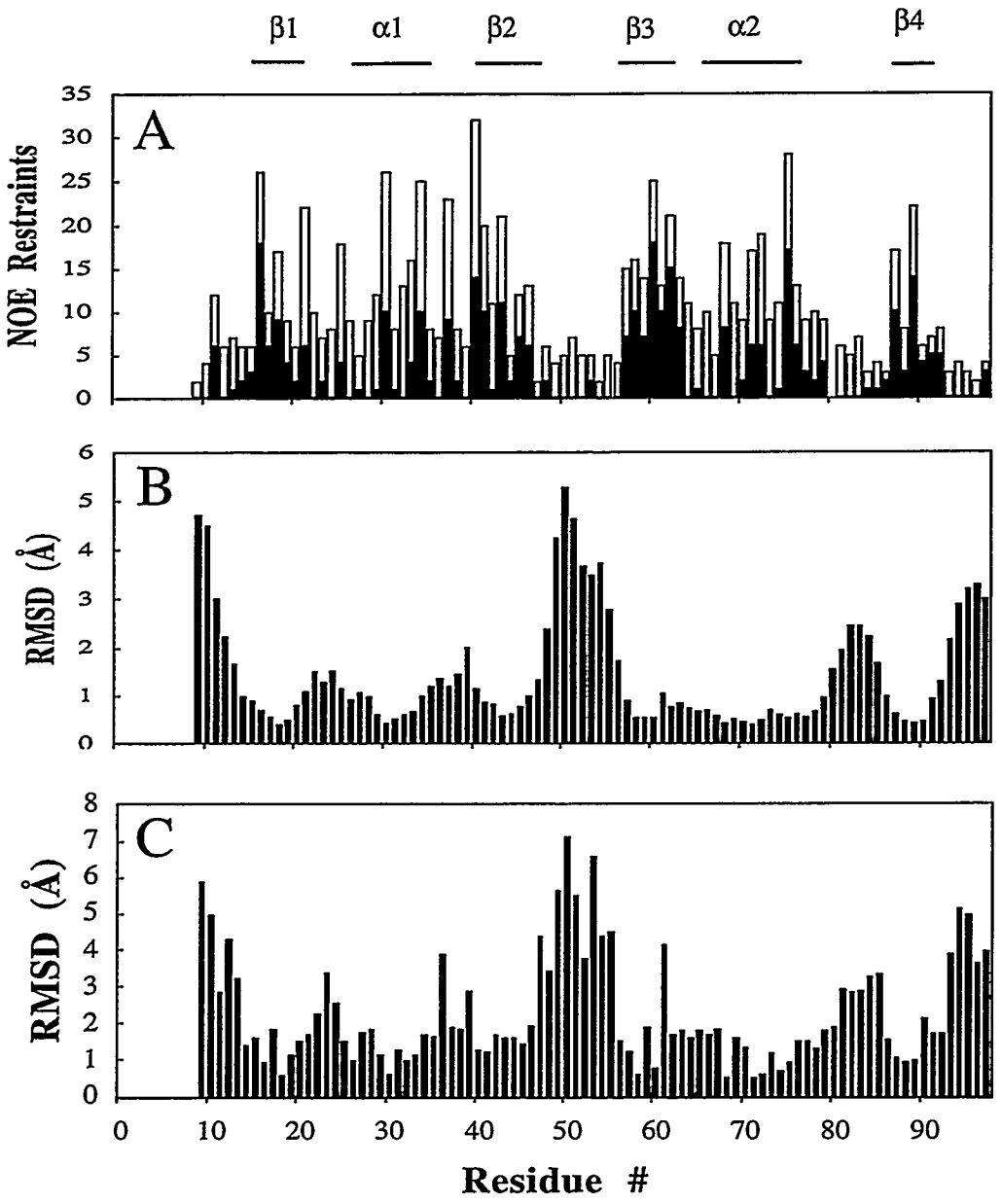
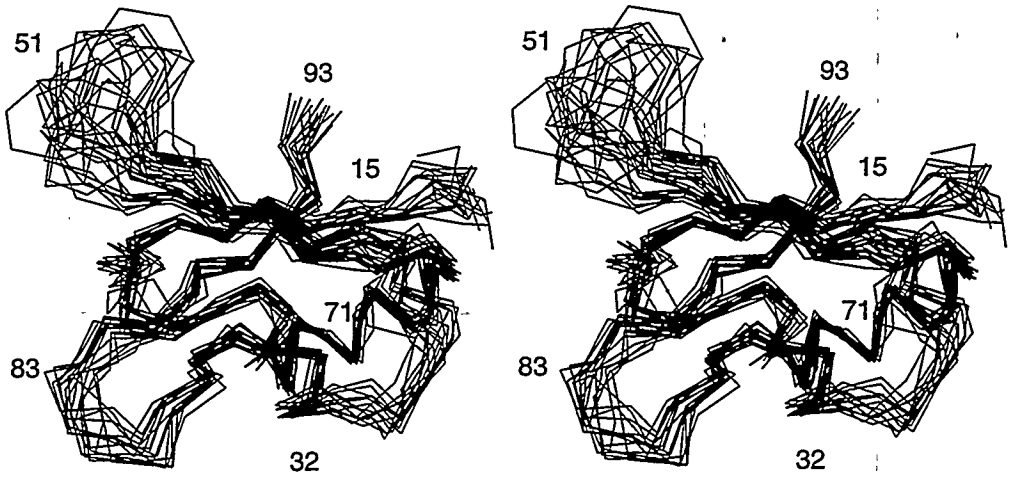


Figure 5.1. (A) Distribution of NOE restraints used in structure calculations. The bar height reflects the total number of NOEs for a given residue. Filled portions represent long-range NOEs. Secondary structure is denoted at the top of the figure. (B) Residue-based root mean square deviation to the mean coordinates for backbone atoms. (C) Residue-based root mean square deviation to the mean coordinates (backbone-superimposed) for non-hydrogen atoms.

A)



B)

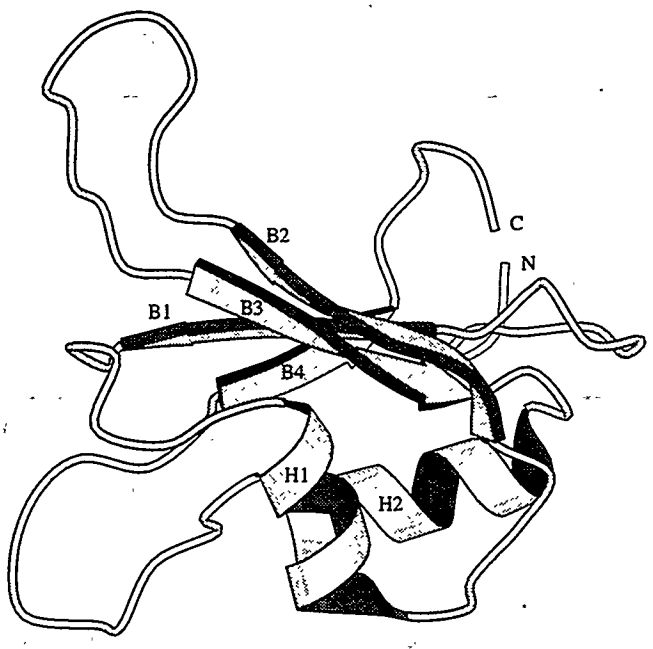


Figure 5.2. A) Stereoview of C_α traces for a family of 17 structures superimposed to mean backbone atoms involved in secondary structure. B) Ribbon diagram of Sxl RBD-2. Secondary structural elements are labeled accordingly and shown as ribbons. All other regions are shown as tubes. The figure was created with MOLSCRIPT software (Kraulis, 1991) .

snRNP-U1A protein (Scherly et al., 1990; Mattaj, 1993) and is expected to be important for all proteins of this class. Sxl RBD-2 has one Lys and three Arg residues in this 10 residue stretch alone, and in the absence of RNA, the lack of long-range NOEs suggests that this loop is flexible (see relaxation studies in Chapter 6). A second loop connecting α2 and β4 also has few medium and long-range NOEs, leaving it poorly defined in the calculated structures. It should be noted that NMR studies of hnRNP-C (Burd & Dreyfuss, 1994) have revealed a short stretch of anti-parallel beta-sheet in between α2 and β4. It appears that this is an inherent difference between these two proteins.

The two α-helices are anchored to the bottom of the sheet by a network of helix-helix and helix-sheet NOEs, which define the relative orientations of these elements (Figure 5.2) and reflect the contacts that define the hydrophobic core of the protein. At this stage of refinement, the overall conformation of the protein resembles hnRNP-C more closely than snRNP-U1A. This is primarily because the helices in both hnRNP-C and Sxl RBD-2 are nearly perpendicular to each other. In snRNP-U1A, helix 1 extends for an extra turn, and the two helices form a more acute interhelical angle (Nagai et al., 1990).

The backbone atoms involved in the RNP-1 and RNP-2 conserved sequences are especially well defined. Consequently, this portion of the ensemble of 17 structures presents an accurate representation of what is believed to be the RNA binding surface. Very few long-range NOEs involving the sidechains of F59 and Y17, aromatic groups implicated in RNA binding affinity, were observed, suggesting that these sidechains are mobile in the absence of RNA. In the calculated structures, these sidechains are found in quite different orientations. Similarly, other sidechains (N15, T19, R47, K49, R53, R55, R61, H95, K97) which might be presumed to be involved in RNA recognition sample a wide range of orientations in the structures. These observations are not surprising since all of these sidechains are either directed away from the hydrophobic core (sidechains of L16, L30, I33, F34, Y37, I40, V60, Y62, A68, A71, L75) or in flexible regions.

Comparison with other RBD structures

There are many proteins which contain single or multiple RBDs. Many of these are known to be involved in RNA processing as well as regulation of gene expression (Kenan et al., 1991; Mattaj, 1993). In addition to an overall sequence similarity, these proteins have highly conserved octamer (RNP-1) and hexamer (RNP-2) sequences, which occur in the central two strands of the β-sheet. Of this class of proteins, a few have been structurally characterized to date. The first is snRNP-U1A, for which a 2.8 Å resolution X-ray crystal structure was solved (Nagai et al., 1990). The second is a solution structure of the hnRNP-C RBD solved by NMR methods (Görlach et al., 1992; Wittekind et al., 1992). Most recently, NMR resonance assignments and secondary structure analysis have been given for hnRNP-A1 (Garrett et al., 1994). It is clear that all four proteins have the same global folding pattern, two helices packed up against a 4-stranded anti-parallel β-sheet. This fold was originally predicted through secondary structure prediction algorithms of published RBD sequences followed by model building based on acylphosphatase (Ghetti et al., 1990). The primary difference among these proteins is the variability in the length of loop β2-β3: five amino acids in snRNP-U1A, four in hnRNP-C, eight in hnRNP-A1, and ten in Sxl RBD-2. In the case of snRNP-U1A, this loop has been implicated in conferring specificity for binding of RNA (Scherly et al., 1990; Mattaj, 1993), and it is quite likely that it plays a major role in Sxl RBD-2 RNA binding. However, the RNAs recognized by snRNP-U1A and Sxl have dramatically different base sequences and may adopt different conformations when bound to protein. The RNA recognized by snRNP-U1A is known to have a stem-loop structure (Jessen et al., 1991), but Sxl recognizes a site comprised of U-tracts (Chapter 7), as do some hnRNP proteins. It is difficult to imagine these U-tracts forming continuous base pairs as is required for stem-loop structures. Therefore, it is not unreasonable to think that different RNAs are recognized by different β2-β3 loop structures. However, one cannot expect the β2-β3 loop to solely determine specificity of RNA binding in all RBDs (Scherly et al., 1990). In the absence of RNA this loop in Sxl-

RBD2 has greater mobility than regions involved in regular secondary structure (Chapter 6), suggesting that structural characterization may be possible only after RNA has been complexed with the protein (Chapter 8).

Görlach et al. (Görlach et al., 1992) successfully complexed the hnRNP-C RBD with a poly-U octamer. They observed that resonances for terminal residues of the protein which originally appeared to be unstructured in solution shifted significantly upon addition of $r(U)_8$. Moreover, deletion of only five amino-terminal residues from the hnRNP-C1 protein abolishes its RNA-binding activity (Görlach et al., 1992). The calculated structures here are consistent with the notion that the termini interact with RNA substrate. The residues entering β-strand 1 and exiting β-strand 4 are above the surface of the face of the sheet (Figure 5.2). Thus, it is likely that these residues will contact the RNA upon binding, and as a result they may adopt a more rigid conformation (Chapter 8).

The β-sheet in Sxl RBD-2 has a right-handed twist as in the other RBD proteins. However, there is a bulge at the beginning of β-strand 2 (I40-V41), which results in V41 popping above the plane of the sheet. Bulges have been observed in β-strands 1 and 4 in hnRNP-C, β-strand 2 of snRNP-U1A, and not at all in hnRNP-A1. It may well be that these perturbations in local geometry contribute to the specificty of binding RNAs of differing sequences. Another interesting feature of these proteins is the relative orientation of the two helices to each other and the sheet. In this regard, snRNP-U1A appears distinct from hnRNP-C and Sxl RBD-2, where the two helices are nearly perpendicular to each other. It is possible that this is due to the extended length and C-terminal structure of helix 1 in snRNP-U1A, which leaves fewer residues to connect with the beginning of β-strand 2. Helix 1 is reported to terminate near the first conserved aromatic residue (residue preceding position A in Table 5.2) for both Sxl RBD-2 and hnRNP-C and near the second conserved aromatic residue (position C in Table 5.2) for snRNP-U1A and hnRNP-A1.

Table 5.2: Chemical Shift Values Near C-terminus of Helix 1 for 3 RBD-containing proteins[a]

Position[b]	A	B	C	D
NH exchange[c]	mixed	fast	slow	slow
hnRNP-C	^{38}S (8.95)	^{39}K (7.10)	^{40}Y (7.01)	^{41}G (7.47)
hnRNP-A1	^{34}E (9.30)	^{35}Q (6.98)	^{36}W (7.05)	^{37}G (7.34)
Sxl RBD-2	^{35}G (9.05)	^{36}K (6.70)	^{37}Y (6.86)	^{38}G (7.34)

[a] ^{1}H chemical shifts in ppm are given in parentheses. [b] Positions A-D are a result from aligning residues near the C-terminus of helix 1 for the three proteins. [c] Fast and slow refers to rapidly and slowly exchanging amide protons, respectively, observed for the corresponding position in all three RBDs. Mixed refers to both fast and slow exchange rates observed. Although all authors have categorized amide exchange rates into these two regimes, their respective criteria and measurements differ slightly.

It is intriguing to compare the ^{1}H chemical shift data near the C-terminal region of helix 1 for the three RBDs with published chemical shifts (hnRNP-C, hnRNP-A1, Sxl RBD-2). In Table 5.2, A-D correspond to aligned amino acids from these three RBDs near the C-terminus of helix 1. For all three proteins, B amide protons exchange freely, while C and D amide protons are protected from exchange. In addition, all A amide protons are shifted downfield from random coil values (Wishart et al., 1991), and all B, C, and D amide protons are shifted significantly upfield. This evidence argues that there should be a common structural feature shared by these proteins, despite the inconsistencies in the reported termination sites of helix 1. For Sxl RBD-2, if restraints could be included that

extend helix 1 beyond G35 and up to G38, there is little doubt that the C-terminus would move towards the sheet, reducing the interhelical angle in the process. In this case, the helix orientations would appear more like those in snRNP-U1A, in which the C-terminus extends for a few more residues. However, presently there is not sufficient experimental evidence to suggest helix 1 extends beyond G35. This issue of helix 1 termination can be more accurately discussed when structures of higher resolution are attained.

Helix capping

As described in the Results, a capping box α-helix initiation signal (Harper & Rose, 1993) has been identified in helix 1 (Figure 5.3). This structural motif enables two extra intramolecular hydrogen bonds to be formed, further stabilizing the helix. Sidechain-to-mainchain hydrogen bonds as observed by Harper & Rose were included in the structure calculations with T26 as the N-Cap residue and Q29 as the N3 residue. Inclusion of these two hydrogen bonds was further justified from structures calculated without capping box hydrogen bond restraints, where the relative orientation of residues 26-29 were retained. In these structures, Q29 O_ε was the closest potential acceptor for HN of T26. Also, the only potential acceptors for HN of Q29 were T26 O_γ and the backbone carbonyl oxygen of T26, as HN of Q29 was nearly equidistant from these two atoms. As a result, the N-terminus of helix 1 is well-defined, shown in Figure 5.3 as a superposition of all atoms from T26 to Q29. Interestingly, HN of Q29 does not adopt a linear orientation with respect to its hydrogen bond acceptor, O_γ of T26, but instead occupies the space between O_γ of T26 and the backbone carbonyl of T26. Hence it appears from the structure calculations that the amide of Q29 is potentially involved in a bifurcated hydrogen bond. The carbonyl oxygen of T26 was included as a hydrogen bond acceptor for the amide proton of L30 (a position in the middle of the α-helix), and this donor/acceptor pair adopts a linear orientation. It should be noted that NOE and torsional restraints alone do not exclude the possibility of a 3_{10}-helix for the first turn of helix 1. However, these data are sufficient

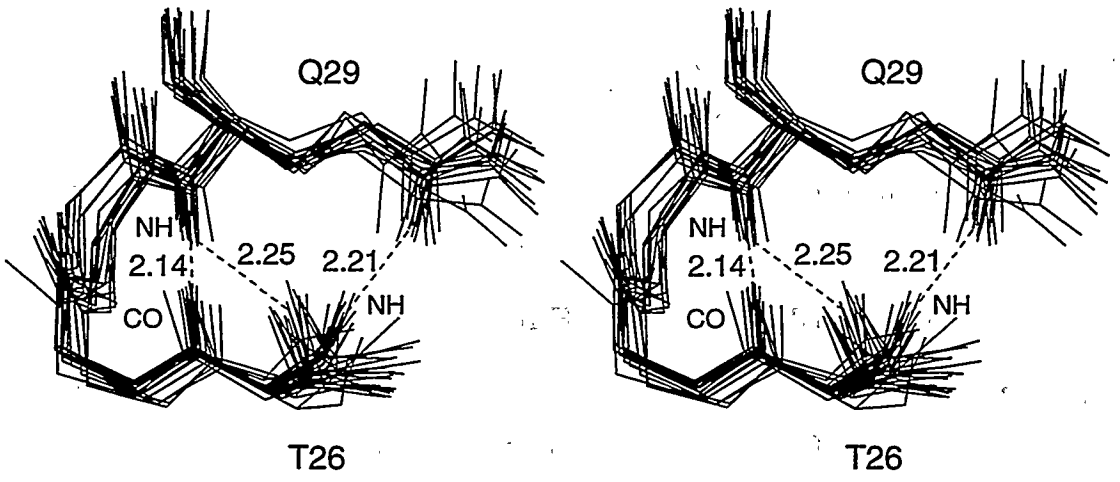


Figure 5.3. Stereoview of the "capping box". Backbone heavy atoms for T26-Q29, as well as sidechains for T26 and Q29 are shown in a superposition of 17 calculated structures. All atoms for T26-Q29 were used in the superpostion to mean coordinates. Labels for residues 26 and 29 occur near their respective sidechains. Backbone amide protons are denoted by NH for T26 and Q29, and the backbone carbonyl oxygen of T26 is denoted by CO. Average distances in Å are shown (left to right) as dashed lines for $Q29_{HN}$-$T26_{O'}$, $Q29_{HN}$-$T26_{O\gamma}$, and $T26_{HN}$-$Q29_{O\varepsilon}$. Standard deviations for these distances (left to right) are 0.09, 0.08, and 0.10 Å.

such that the accompanying shift in backbone geometry would not change the sidechain positions of the capping box.

As discussed above, there is some uncertainty as to where helix 1 is terminated. In the light of a recent survey (Aurora et al., 1994), it is tempting to speculate that G35 and/or G38 might be involved in a C-terminal glycine helix-capping interaction. Unfortunately, the sequence LDTIFGKYGSI (residues 30-40) is not consistent with the rules provided, although the amide protection data is suggestive of a Schellman-like motif (Schellman, 1980). It is probable that there is a distinct pattern of hydrogen bonds since residues 37 and 38 (C and D in Table 5.2) have slowly exchanging amide protons. Unfortunately, no NOEs could be identified to support this type of structural motif. Again, higher-resolution structures should be able to resolve this matter.

RBD2 structures including additional restraints from J_{NC} and J_{CC} experiments

At some point after the family of 17 structures was calculated, some extra information on coupling constants and torsional angles was obtained out of a pair of "quantitative J-correlation experiments", known as the J_{NC} (Vuister et al., 1993) and J_{CC} (Grzesiek et al., 1993) . These two experiments allow for the deterimination of the 3-bond J-coupling between amide ^{15}N and $^{13}C_\gamma$ or $^{13}C'$ and $^{13}C_\gamma$, hence J_{NC} and J_{CC}. They are most effectively applied to amino acid residues which contain gamma methyl groups (Val, Ile, Thr) since these resonances are often well resolved in a $^{13}C/^{1}H$ correlation spectrum and have intense signals. In the case of Val, stereospecific assignment of the gamma methyl carbons may be obtained as long as there is not a mixture of χ_1 rotamer states, or "rotamer averaging". The coupling constants are measured by quantitating volume or intensity differences of 2D $^{13}C/^{1}H$ correlation peaks between a reference experiment, S_a, and the difference of the reference experiment with a second experiment, S_a-S_b. Both experiments are conducted as "constant-time" experiments (Chapter 1) which remove the effects of the one-bond ^{13}C-^{13}C coupling during t_1. Experiment B differs from the

reference experiment by the placement of the 180° ^{15}N (J_{NC}) or $^{13}C'$ (J_{CC}) pulse inside of the constant-time period, and this allows for the 3-bond coupling to evolve and attenuate the observed signal of interest in experiment B relative to the reference experiment, according to the relation

$$J_{NC}(J_{CC}) = \frac{\sin^{-1}\left[\left(\frac{x}{2}\right)^{-2}\right]}{\pi T}$$

$$x = \frac{S_a - S_b}{S_a} \quad , \qquad 2T = 28.6 \text{ msec } (1/J_{CC}).$$

Just as in the standard constant-time HSQC (Vuister & Bax, 1992), a value of 28.6 msec for the ^{13}C constant-time t_1 period, 2T, is chosen because it allows for complete refocusing of the 35-40 Hz J_{CC} one-bond coupling, maximizing the amount of observed signal under the assumption that T_2 relaxation is not too severe during 2T. Residues in Sxl-RBD2 for which these coupling constants could be measured are given in Table 5.3.

Table 5.3: 3-bond Coupling Constants Determined From the J_{NC} and J_{CC} Experiments

Residue	$^3J_{NC}$ (Hz)	$^3J_{CC}$ (Hz)
V18 (γ1)	-	-
V18 (γ2)	0.90	-
V41 (γ1)	0.85	-
V41 (γ2)	1.23	-
V57 (γ2)	1.09	-
V60 (γ1)	0.97	-
V60 (γ2)	1.99	-
V89 (γ1)	-	-
V89 (γ2)	1.70	-
I33	1.96	-

I40	1.97	-
I45	1.93	-
I79	2.42	-
T19	-	1.48-1.74
T24	-	2.22
T26	-	2.26
T32	1.68	-
T51	-	2.65

Errors for the coupling constant values in Table 5.3 were not estimated, but the values are probably good to 0.1-0.2 Hz. When a value is missing from the table, a reasonable value could not be computed because the corresponding peak was either missing or too weak to measure reliably from the difference spectrum. These values were interpreted qualitatively. In the case of the J_{CC} experiment, there is a distinct reason why the values should only be interepreted qualitatively. This is because the published pulse sequence for the J_{CC} experiment has z-pulsed field gradients. On Sxl-RBD2 pulsed field gradients were not employed and the experiment was performed in perfect analogy to the J_{NC} experiment, for which the published pulse sequence does not have gradients. It was not clear if this would adversely affect the quality of the data, but since the difference spectrum had many negative peaks (only positive peaks in J_{NC}), I believe that the gradients are mostly likely necessary for proper acquisition. For this reason, the data was interpreted conservatively.

Essentially, the 3-bond coupling constants were categorized as either small or large for the purpose of determining χ_1 angles in these residues. The χ_1 angles could in turn be used as loose torsional restraints in further structure calculations. A large ^3J (> 1.6 Hz) was interpreted as arising from an *anti* rotamer state with respect to the two relevant atoms (NC_γ or $C'C_\gamma$) and a small ^3J (no measurable intensity in the difference spectrum) was interpreted as arising from 1 of 2 possible *gauche* rotamer states. Therefore, the presence

of a "large" ^3J unambiguously defines the rotamer state about the C_α-C_β axis, i.e. χ_1. In the event that no "large" ^3J values were observed from either J_{NC} or J_{CC} experiment, it was concluded that there must be rotamer averaging about the C_α-C_β axis. Table 5.4 summarizes the χ_1 data for Val, Ile, and Thr residues and also shows the corresponding restraints used for round 24 of structure calculations of RBD2.

Table 5.4: χ_1 Values Determined From the J_{NC} and J_{CC} Experiments

Residue	χ_1
V18	r. a.
V41	r. a.
V57	r. a.
V60*	180 ± 20°
V89*	180 ± 20°
I33	-60 ± 20°
I40	-60 ± 20°
I45	-60 ± 20°
I79	-60 ± 20°
T19	60 ± 20°
T24	60 ± 20°
T26	60 ± 20°
T32	-60 ± 20°
T51	60 ± 20°

*Stereoassignments for γCH_3 groups of V60 and V89 using standard nomenclature:

V60γ1 - 20.8 (0.22) V60γ2 - 20.3 (0.05)
V89γ1 - 24.3 (1.02) V89γ2 - 21.8 (1.02)

Certainly, inclusion of these χ_1 restraints is desireable in structure calculations. The accuracy and precision of the structures should improve. These additional torsional restraints were included in the last round of structure calculations to date, and several long-range NOE distance restraints were tightened for V60 γCH_3 groups since the extremely large pseudoatom used for the 2 methyls was no longer necessary. Only 15 structures were calculated in this round (round 24), but 12 of the structures had low energies comparable to the previous round of 17 structures. This most recent restraint file is given in Appendix C. This is a convergence rate of 80% vs. 57%, a significant improvement. Although not accurately characterized, it appears that some of the elements of secondary structure, namely the helices, shifted slightly relative to the previous round. Specifically, the C-terminal end of helix 1 seems to move slightly closer to the sheet, making the interhelical angle slightly more acute. The specific cause(s) for this was not investigated. It is also interesting to note that the χ_1 angle for T26 confirms the presence of the N-terminal capping box in helix 1, as any other rotamer state besides $\chi_1 = 60°$ would have been inconsistent with the geometry of the capping box.

Aurora, R., Srinivasan, R. & Rose, G. D. (1994) *Science 264*, 1126-1130.

Brünger, A. T. (1992) *X-PLOR Manual Version 3.1*, (xxx), Yale University Press, New Haven, CT.

Burd, C. G. & Dreyfuss, G. (1994) *Science 265*, 615-621.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M. & Gronenborn, A. M. (1986) *Embo Journal 5*, 2729-2735.

Garrett, D. S., Lodi, P. J., Shamoo, Y., Williams, K. R., Clore, G. M. & Gronenborn, A. M. (1994) *Biochemistry 33*, 2852-2858.

Ghetti, A., Bolognesi, M., Cobianchi, F. & Morandi, C. (1990) *FEBS Lett. 277*, 272-276.

Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L. & Dreyfuss, G. (1992) *Embo Journal 11*, 3289-3295.

Grzesiek, S. G., Vuister, G. W. & Bax, A. (1993) *J. Biomol. NMR 3*, 487-493.

Harper, E. T. & Rose, G. D. (1993) *Biochemistry 32*, 7605-7609.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C. & Nagai, K. (1991) *Embo Journal 10*, 3447-3456.

Kenan, D. J., Query, C. C. & Keene, J. D. (1991) *Trends in Biochemical Sciences 16*, 214-220.

Mattaj, I. W. (1993) *Cell 73*, 837-840.

Nagai, K., Oubridge, C., Jessen, T., Li, J. & Evans, P. R. (1990) *Nature 348*, 515-520.

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

Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J. & Mattaj, I. W. (1990) *Nature 345*, 502-506.

Vuister, G. W. & Bax, A. (1992) *J. Magn. Reson. 98*, 428-435.

Vuister, G. W., Wang, A. C. & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 5334-5335.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295-315.

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

Wittekind, M., Gorlach, M., Friedrichs, M., Dreyfuss, G. & Mueller, L. (1992) *Biochemistry 31*, 6254-6265.

Wüthrich, K., Billeter, M. & Braun, W. (1983) *J. Mol. Biol. 169*, 949-961.

Chapter 6

Backbone Dynamics of Sxl-RBD2

Introductory remarks

Having calculated structures of RBD2 to intermediate resolution, it was decided that determination of higher resolution structures was of secondary interest. The interaction of the Sxl-RBD proteins with *tra* RNA was of primary interest. However, a few obstacles to studying any of these complexes (Chapters 7 and 8) slowed progress in terms of NMR studies. This delay allowed for some quick RBD2 ^{15}N relaxation studies which probe fast (< 1 ns) motional dynamics of the backbone amide NH moieties. Of course, ultimately we are interested in complete characterization of structure and dynamics of free Sxl as well as Sxl bound to RNA. If the Sxl RBDs undergo a structural or dynamic transition upon binding RNA, it is likely that the changes in disordered or flexible regions of the protein will be reflected in the relaxation behavior of spins in those regions, and it is hoped that such effects could eventually be related to RNA binding activity. In this light, the relaxation/dynamics described here shall serve as both a study of primary interest as well as an introduction to ^{15}N relaxation studies for the greater Sxl project.

NMR has been recognized as a tool for investigation of dynamic motions in molecules since the earliest days of NMR. In principle, NMR is advantageous in that individual signals from each nuclear spin are observables, which allows one to probe properties of the protein in exquisite detail. Specifically, the relaxation rates of perturbed spins toward equilibrium are sensitive to molecular fluctuations which contribute to the spectral density function near the Larmor frequencies of the spins involved in relaxation. An illustration of this is provided in Figure 6.1. Relaxation of protein resonances is sensitive to motions on the timescale of 10^{-12} - 10^{-8} and 10^{-5} - 1 seconds, a range as broad

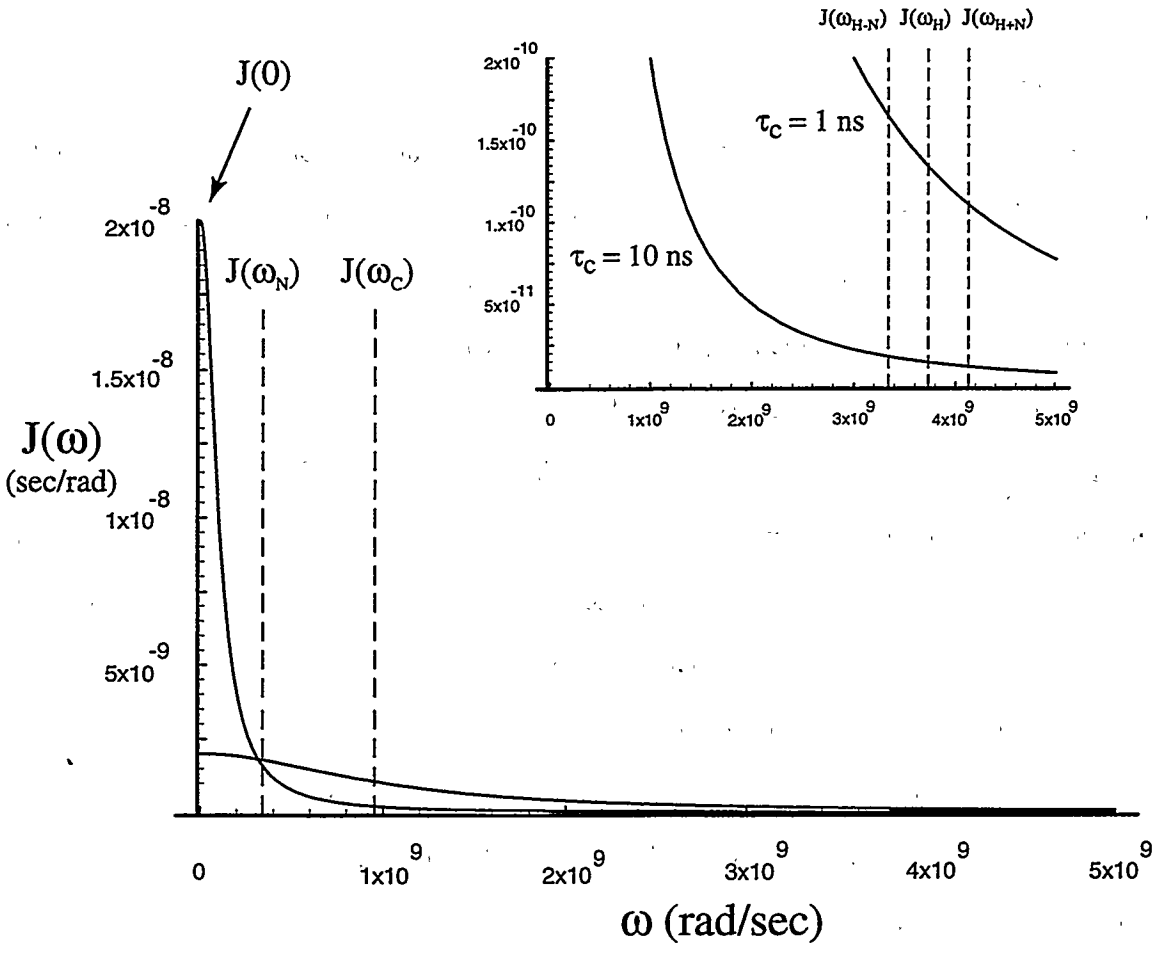


Figure 6.1. Lorentzian spectral density functions for two correlation times (τ_c), 1 ns and 10 ns, are plotted. Dotted lines correspond to the frequencies found in equations 1-3 for a B_o field strength of 14 tesla. $J(\omega_c)$ is also given. The inner panel is the same plot with the y-axis expanded so that differences at higher frequencies can be seen.

as any other specific technique. These rates can be interpreted in the framework of specific motional models or alternatively in a "model-free" approach. The latter approach has become the most popular in recent years. Twenty-five years ago, relaxation studies could only be applied to resonances which were resolved in a one-dimensional spectrum. With the advent of two-dimensional NMR, the problem of spectral overlap was solved to a great degree. Progress in experiment development, such as HSQC, in combination with specific and uniform isotope labeling techniques of proteins has made relaxation experiments on protonated nitrogen and carbon spins quite attractive. In the last 7 years, examination of relaxation rates of ^{15}N spins along the polypeptide backbone has become the most widespread approach to probing protein dynamics due to the relative simplicity of the data collection and analysis.

Relaxation theory

Introduction to relaxation. In liquids, the relationship between molecular motions and nuclear spin relaxation can be most concisely depicted in a plot of the spectral density function, $J(\omega)$, for a vector between two spins (Figure 6.1). For our present purpose, a discussion of protein backbone dynamics, consider a $^{15}N/^{1}H$ pair from a backbone amide group of an amino acid. The spectral density shows the distribution of motional frequencies experienced by the NH vector, usually a Lorentzian distribution or a sum of Lorentzian terms. For a protein in solution, fluctuations of motions can arise from multiple sources: overall tumbling, internal conformational exchange processes, and fast vibrational or librational motions. All of these contribute to the spectral density and therefore any of these motions can affect relaxation rates of ^{1}H or ^{15}N spins. The basic mechanism for relaxation of longitudinal or transverse magnetization is that fluctuations in the local magnetic field which occur at or near the Larmor frequencies of the spins can induce spin transitions which bring the system closer to equilibrium. Transverse relaxation (T_2) can occur through an additional low frequency [J(0)] mechanism. Because the specific shape

of the spectral density function is defined by the nature of *all* of the motions of the NH vector, knowing the details of this shape is the ultimate goal in characterizing dynamic motions in proteins or other molecules. Obtaining detailed information on the shape of the spectral density function is presently a difficult problem, but extracting very simple properties from it is both possible and informative.

Although relaxation of longitudinal (T_1) or transverse (T_2) magnetization towards equilibrium can occur through many different types of mechanisms, it is often the case that a single mechanism dominates the relaxation. When this is the case, the relaxation can be described in a few simple analytical expressions to a good approximation. This desirable situation is created for a protein when it is uniformly labeled with ^{15}N. The backbone nitrogens are relaxed primarily by their directly attached protons through the dipole-dipole mechanism. At fields up to 600 MHz the ^{15}N relaxation is primarily dipolar, and the expressions for T_1, T_2, and the heteronuclear ($^{15}N/^{1}H$) NOE are given by (Abragam, 1961):

$$T_1^{-1} = \frac{h^2 \gamma_N^{\,2} \gamma_H^{\,2}}{4 r_{NH}^{\,6}} \left(J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) \right) \qquad (1)$$

$$T_2^{-1} = \frac{h^2 \gamma_N^{\,2} \gamma_H^{\,2}}{8 r_{NH}^{\,6}} \left(4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N) \right) \qquad (2)$$

$$\mathrm{NOE} = 1 + \frac{\gamma_H \left(6J(\omega_N + \omega_H) - J(\omega_H - \omega_N) \right)}{\gamma_N \left(J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_N + \omega_H) \right)} \qquad (3)$$

The T_1, T_2, and NOE for each backbone amide ^{15}N in the protein can be measured in a convenient manner (see below), and depending on the level of sophistication of the analysis of these relaxation parameters, information on the spectral density, $J(\omega)$, or motional dynamics can be extracted using the above expressions.

Generally speaking, large relaxation rates (short T_1 or T_2, negative NOE) result from large amplitude motions of the NH bond vector at the frequencies evaluated in the spectral density functions in equations 1-3. The advantage of interpretation of relaxation data in this very simple manner is that it requires no assumptions about the form of the spectral density function, $J(\omega)$. An approach used frequently in the past is to interpret the relaxation rates within the framework of a specific motional model, such as diffusion in a cone (Lipari & Szabo, 1982). Given a specific model, the relaxation data can be fit to a particular range of motions within the model. This allows for a visual representation of the motion which is appealing, but this approach can be dangerous since usually there may be no evidence to warrant the assumption of a specific model.

Lipari and Szabo's "model free" approach. Today, the most widely used method for interpretation of NMR relaxation data is the "model free" approach pioneered by Lipari and Szabo (Lipari & Szabo, 1982). No specific motional models are assumed. Instead, two model-independent parameters, S^2 and τ_e, are defined in the context of a simplified internal correlation function which collectively describes all internal motions independent from overall tumbling. This internal correlation function, $C_I(t)$, is shown to be exact at t=0 and t=∞. This is useful since S^2 (the generalized order parameter) $\equiv C_I(\infty)$. It follows that the value of S^2 gives a "rigorous model-independent measure of the degree of spatial restriction of the internal motion." For example, S^2=0 corresponds to isotropic motion of the NH bond vector, and S^2=1 corresponds to a perfectly rigid, single internal orientation. On the other hand, τ_e is known as the effective correlation time, which is related to the timescale of rapid (<0.3 ns) internal fluctuations. Formally, τ_e is defined to be the integrated area of $C_I(t)$. An important condition of this approach is that $\tau_e << \tau_M$, where τ_M is the overall correlation time of tumbling in solution. For internal motions in the extreme narrowing limit, S^2 and τ_e are free of error. For motions on a slower timescale (> 0.3 ns), the calculated values can contain a significant amount of error (Lipari & Szabo, 1982). Using

the proposed internal correlation function, $C_I(t)$, a spectral density function can be constructed in terms of the two model-free parameters,

$$J(\omega) = \frac{2}{5}\left(\frac{S^2 \tau_M}{1+(\omega\tau_M)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2} \right) \qquad (4)$$

$$\text{where} \qquad \tau^{-1} = \tau_M^{-1} + \tau_e^{-1}$$

Using equations 1-3 for relaxation rates and the NOE enhancement, this form of the spectral density can be least-squares fit to the relaxation data to obtain values for S^2 and τ_e. It should be noted that relaxation contributions from chemical shift anisotropy can easily be integrated into the formalism. This is accomplished by adding an additional $J(\omega_N)$ term with a CSA coefficient into the expressions for T_1 and T_2. It should also be noted that Clore et al. (Clore et al., 1990) extended the "two parameter" model-free approach to allow for motions on different timescales. Briefly, if it is assumed that the fast internal motions are independent of the slow motions, S^2 can be decomposed into a fast and a slow component,

$$S^2 = S_f^2 S_s^2 \qquad (5)$$

where τ_f and τ_s are the fast and slow effective internal correlation times, respectively.

All of the previous discussion was addressed for the case of isotropic overall tumbling of the protein. Clearly, all proteins do not tumble perfectly isotropically. It is relatively straightforward to apply the model-free analysis to cases with anisotropic tumbling as long as the anisotropy can be separated into a parallel and a perpendicular component (Lipari & Szabo, 1982; Schneider et al., 1992). The calculations are slightly more involved, but the underlying ideas are the same. An excellent example of this is the work on calmodulin by Bax and coworkers where they demonstrated that the central helix in calmodulin is actually flexible in solution (Barbato et al., 1992), contrary to crystalline

calmodulin. There are many other examples in the literature demonstrating the simple and effective description afforded by the model-free approach. To summarize, the model-free approach requires no *a priori* knowledge about the types of motions experienced by the NH bond vectors, and it yields an accurate and physically significant description of rapid (<0.3 ns) internal dynamic motions in proteins.

NMR relaxation experiment and analysis methods

Collection of NMR data. ^{15}N T_1 rate constants were measured using a standard two-dimensional $^{15}N/^{1}H$ inverse-detected correlation experiment with a parametrically varied delay to allow for longitudinal ^{15}N relaxation to occur (Barbato et al., 1992). A series of 8 experiments were collected at 600 MHz and 25 °C using the pulse program hetT1y.jh. 2D spectra were collected with T_1 time delay values of 26.0, 152.2, 320.5, 530.8, 811.3, 1063.7, 1274.0, and 1624.6 ms, each consisting of 256 real x 2048 real data points. A recycle delay of 2.0 s and 8 scans per real t_1 point were collected, making the total acquisition time for the entire series 14 hours. Quadrature detection using the TPPI method (Marion & Wüthrich, 1983) was used. Spectral widths for t_1 and t_2 were 7246 and 1861.5 Hz, respectively.

^{15}N T_2 rate constants were measured using a standard two-dimensional $^{15}N/^{1}H$ inverse-detected correlation experiment with a parametrically varied CPMG pulse train to allow for in-phase transverse ^{15}N relaxation to occur (Barbato et al., 1992). A series of 7 experiments were collected at 600 MHz and 25 °C using the pulse program hetT2y.cl. T_2 time delay values were set to 7.9, 39.6, 95.1, 174.4, 301.2, 475.6, and 713.4 ms, each consisting of 256 real x 2048 real data points. A recycle delay of 1.95 s and 16 scans per real t_1 point were collected, making the total acquisition time for the entire series 21 hours. Quadrature detection using the TPPI method was used. Spectral widths for t_1 and t_2 were 7246 and 1861.5 Hz, respectively.

The $^{15}N/^{1}H$ NOEs were measured using a sensitivity-enhanced two-dimensional $^{15}N/^{1}H$ inverse-detected correlation experiment (Skelton et al., 1993; Palmer et al., 1991a). The measurement is made from the ratio of intensities or volumes between a reference experiment and an experiment with the steady-state $^{15}N/^{1}H$ NOE produced by ^{1}H saturation of backbone amide resonances:

$$NOE = \frac{I_{sat}}{I_{unsat}}$$

Saturation of amide protons was acheived with a broadband decoupling scheme (DIPSI-2) (Skelton et al., 1993). A relatively short recycle delay of 3.1 s was employed, and therefore it is possible that the NOE values obtained are systematically high due to incomplete recovery of ^{1}H magnetization at the beginning of the pulse sequence. A short recycle delay of ~ 3 s was also used by Mandel et al. (Mandel, 1995), and the values obtained here on RBD2 are reasonable, so the NOE was not re-measured. It should be kept in mind, however, that perhaps artificially high NOE values employed in the full analysis (below) may be responsible for systematic errors in the 5-10% range. The data were collected as 80 complex x 1024 complex data matrices with spectral widths identical to T_1 and T_2 experiments mentioned above. 16 scans/block were used to make the total experiment time 10 hours.

Analysis of NMR data. For the T_1 and T_2 experiments, crosspeak intensities were measured in Felix using xpk_hgt_vol.mac, a Felix macro written by Dr. Michael Akke in Dr. Arthur G. Palmer's (Professor, Columbia University) research group which finds the highest intensity point of each crosspeak and adds it to the database. These intensities were then written as ASCII files and converted to a specified format using Art's awk script felix2relax. The fitting algorithms run.r1 or run.r2 were then used to fit monoexponential decay curves to the T_1 and T_2 data. Finally, r1x2comp (T_1) and r2bestfit (T_2) were used to

summarize the T_1 or T_2 values as well as the quality of the fit via a χ^2 residual. These values as well as the heteronuclear NOE values for RBD2 are given in Appendix A.

The extended Lipari and Szabo model free analysis was carried out with Art's FORTRAN programs mfgrid and modelfree (versions 3.1). In order to obtain input files of the required format, the awk script makemfgin was executed. This script takes plot files for the T_1, T_2, and NOE data and creates input files for mfgrid or modelfree. I found that it was essential to do the parameter grid search (mfgrid) first in order to find reasonable starting input parameters for the full model free analysis (modelfree). This was particularly true for residues which were eventually characterized by slower ($\tau_e > 300$ ps) motions or had relatively large contributions from exchange (R_{ex}).

RBD2 backbone dynamics

RBD2 relaxation data. ^{15}N T_1, ^{15}N T_2, and the $^{15}N/^1H$ NOE were measured for each resolvable 2D $^{15}N/^1H$ correlation peak in RBD2. Relaxation parameters for nearly all residues are plotted in Figure 6.2 and given in tabular form in Appendix A. From the T_2 and NOE data it is immediately apparent that both termini of RBD2 have dramatically different relaxation properties than the rest of the protein. From the 3D spectra and structure calculations we know (and have suspected all along) that the termini are unstructured and have sharp (long T_2) ^1H lines, although they had not been measured accurately. The ^{15}N T_2 and NOE data show definitively that the ends of RBD2 are truly experiencing fast ($<< \tau_M$), random reorientations. The heteronucelar NOE is actually quite a good probe for fast motions since it is most sensitive to high frequency motions (equation 3). To first order, protein T_2 values are generally long for less structured regions and comparatively short for structured regions. This is not rigorously so, and exceptions are commonly observed when there are effects from chemical exchange processes or slower motions ($> \tau_M$), for example. Nevertheless, all 6 elements of secondary structure in RBD2 appear to have similar T_2 values as might be expected for a roughly spherical protein. The

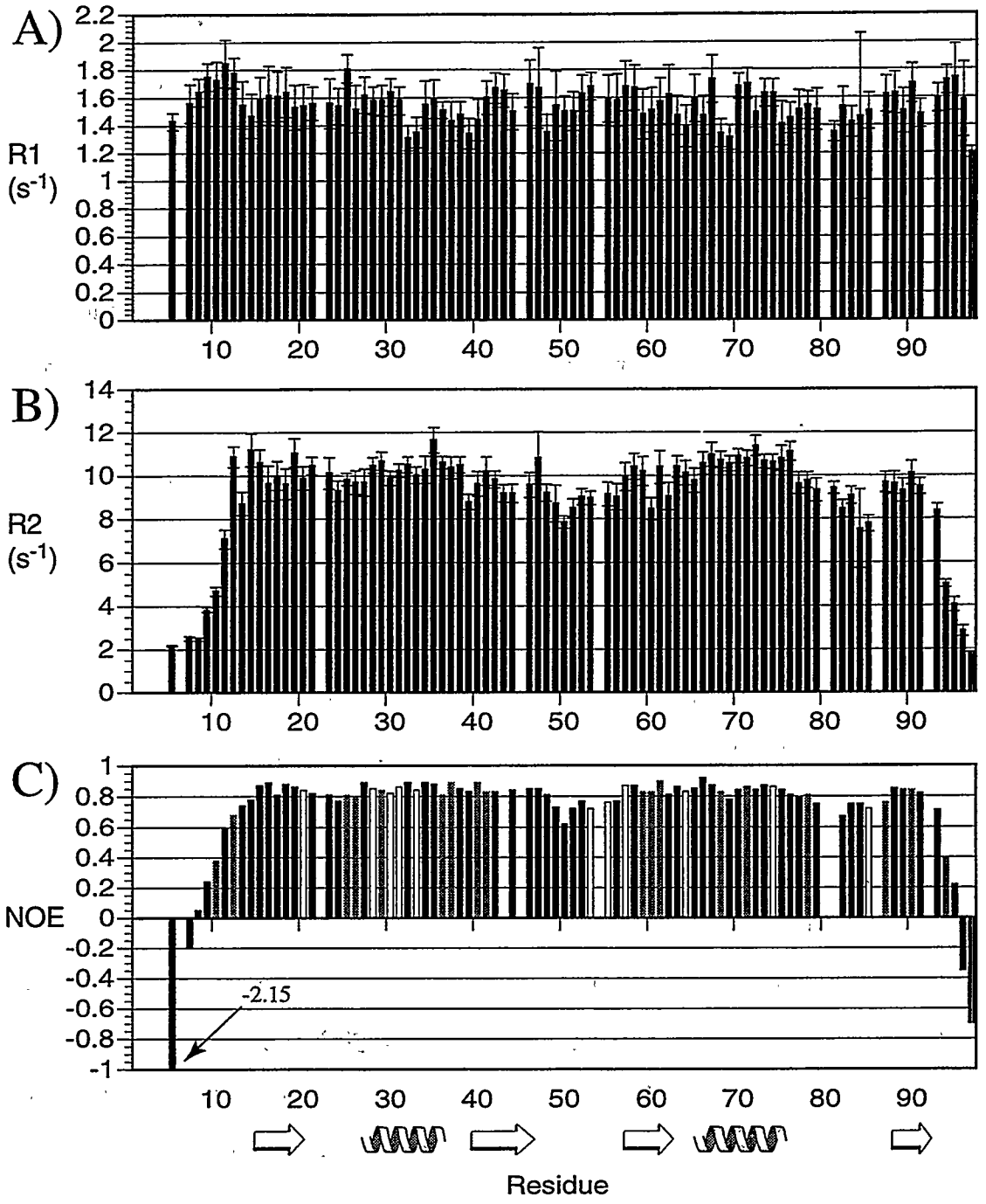


Figure 6.2. Measured ^{15}N relaxation parameters as a function of residue position for RBD2 at 25 °C, pH=5.0. A) R1 ($1/T_1$) B) R2 ($1/T_2$) C) $^{1}H/^{15}N$ NOE. Because the NOE experiment was collected with limited digital resolution in t_1 (80* points), there was more spectral overlap than in the T_1 or T_2 experiments (128* points). Black bars correspond to peaks with no overlap. Grey bars correspond to peaks with a slight degree of overlap. White bars correspond to peaks with a significant degree of overlap. Error bars for the NOE experiment were estimated to be 0.05 for the model free analysis.

β2-β3 and α2-β4 loops on the other hand have comparatively long T_2 values, suggesting that these loops have a significant degree of dynamic flexibility. This observation is consistent with the lack of mid to long range NOESY crosspeaks involving these residues. Small (< 5-10%) differences in T_2 values usually represent a true difference in relaxation rates, but for the same reasons mentioned above it would be dangerous to simply attribute this small difference to a proportional degree of dynamic motion. Therefore, simple inspection of T_2 or heteronuclear NOE values are reliable only for dramatic differences, such as the termini of RBD2. The β2-β3 and α2-β4 loops probably have more dynamic flexibility than the 6 elements of secondary structure, but a more sophisticated analysis (using all 3 relaxation measurements) of the relaxation data is advisable.

Model free analysis of RBD2. A more rigorous analysis of the relaxation data was carried out using the formalism pioneered by Lipari and Szabo (Lipari & Szabo, 1982) as discussed in the theory section above. Dr. Art Palmer has made available a series of programs (Mandel, 1995) which were used for all calculation-intensive steps of this analysis. His Modelfree3.1 program actually carries out the analysis using the Lipari and Szabo formalism. After the T_1, T_2, and heteronuclear NOE were determined (see methods section), an estimate of the overall correlation time, τ_M, was determined using "grun.tmest". First, an average T_1/T_2 for the highly structured regions of the protein backbone was calculated. For large macromolecules (MW > 2000-4000) the T_1/T_2 ratio increases with increasing τ_M (Harris, 1986), so this ratio can be used to estimate τ_M. More rigorously, because T_1/T_2 in ordered regions of the protein is essentially independent of S^2, simple Lipari and Szabo theory can be used to estimate τ_M by inserting T_1/T_2 into the fundamental expressions such as equations 1, 2, and 4 (Palmer et al., 1991b). An average T_1/T_2 ratio of 6.56 ± 0.76 was calculated for the ordered backbone regions of RBD2 and it was iteratively fed to "grun.tmest", which yielded a τ_M value of 7.6 ns. This "overall tumbling" correlation time was used in all subsequent steps requiring this parameter.

The original Lipari and Szabo formalism implemented a simplified spectral density defined in terms of two parameters, τ_e, and S^2, which can be fit to the relaxation data using equations 1-4. More recently, researchers have discovered that the spectral density can be modeled with different sets of "model free" parameters or "models": 1) S^2 2) S^2, τ_e 3) S^2, R_{ex} (line broadening from chemical exchange or other effects) 4) S^2, τ_e, R_{ex} 5) S_f^2, $S_s^2, \tau_e = \tau_s$. Different researchers have developed various methods for selecting a specific "model" for a given NH. I have adopted Art Palmer's method of independently using all 5 models for each residue, followed by selection of the model which fits the data best and uses the smallest number of parameters (i.e. degrees of freedom). This method is appealing since models are selected in an unbiased fashion. In this approach, model 1 is statistically the most desirable since it uses 1 parameter, and models 4 and 5 are the least statistically desirable since they use 3 independent parameters. The specific criteria he uses for selecting from the 5 models can be put into the form of a flow chart (Mandel, 1995). Within this flow chart, the path taken is mainly dependent upon the value of the sum-squared error, Γ, relative to the α=0.05 critical value of the distribution of Γ determined from Monte Carlo simulations:

$$\Gamma = \sum_{j=1}^{M} \frac{\left(R_j - \hat{R}_j\right)^2}{\sigma_j^2} \qquad (6)$$

where Γ is the sum-squared error, M is the number of experimental relaxation parameters, R_j is the jth experimental relaxation parameter, $\hat{R}_j$ is the jth theoretical relaxation parameter, and σ_j is the experimental uncertainty in the jth relaxation parameter. Instead of using Palmer's explicit recommendations involving an F-statistic in certain steps, I simply compared Γ values for different competing models. For example, if two models differing by one degree of freedom (e.g. model 1 vs. model 2) could both be considered, I chose the model with more degrees of freedom only if Γ was at least 2 or 3 times smaller than the model with fewer degrees of freedom. Also, $\Gamma < \Gamma(0.05)$ was a strict requirement for

models 2-4. $\Gamma = 0.00$ was a requirement for model 5. Finally, models 4 or 5 were generally used when $\Gamma > 20$ for all other models, a conservative guideline since I only had data at a single field strength. Table 6.1 contains the model selections for all residues of RBD2 where all three relaxation parameters could be measured reliably. The vast majority of amide NHs were fit satisfactorily using model 1, meaning that the amplitudes of motion reflected in S^2 are occuring on a very fast timescale relative to τ_M (= 7.6 ns). In contrast, the terminal ends (up to I11, E94-K97) are fit best using model 5 which allows for motion on two timescales.

The outcome of the "model-free" analysis is given in Figure 6.3. The order parameters (Figure 6.3A) at the termini approach zero in a smooth manner, confirming that the termini are indeed unstructured and dynamically disordered on a sub-nanosecond timescale. Other regions of RBD2 with relatively low order parameters are the β2-β3 and α2-β4 loops. As mentioned earlier, these loops have no long-range ^1H-^1H NOEs and appear to lack well-defined structure. The β2-β3 loop has the lowest order parameter at L50, which is approximately in the middle of the loop. The values of $S^2 = 0.65$-0.80 for K49-T51 suggests that the outlying residues in the loop have a greater degree of flexibility than NHs involved in secondary structure but have smaller amplitudes of motion than truly disordered regions such as the termini.

It is intriguing to look at parameter values for τ_e and R_{ex} which result from models 2 and 3. These are shown in Figures 6.3B and 6.3C, respectively. For the order parameters shown in Figure 6.3A, model 3 was officially used for T32, G35, A68, and Q69 only (black bars in Figure 6.3C, also Table 6.1). It is interesting, however, to see where significantly large values of R_{ex} were fit to model 3, even if model 3 was not officially selected for a given residue. In fact, many of these residues (white bars in Figure 6.3C) did not have poor fits to model 3. Therefore, the presence of a significant R_{ex} value suggests that these residues might have some additional exchange or motional properties which might add to the ^{15}N line broadening. In this light, it is interesting to note that the

Table 6.1: Selection of Models in the Extended Lipari and Szabo Analysis of RBD2 ^{15}N Relaxation Data

Residue	Model	Γ (SSE)	Residue	Model	Γ (SSE)
R5	5	-	G52	1	4.54
P6	-	-	R53	1	11.22
G7	5	0.00	P54	-	
G8	5	0.00	R55	1	2.27
E9	5	0.00	G56	1	1.89
S10	5	0.00	V57	1	1.75
I11	5	0.00	A58	1	1.28
K12	2	1.25	F59	1	0.22
D13	1	3.49	V60	1	1.65
T14	1	1.76	R61	1	2.75
N15	1	1.19	Y62	1	1.20
L16	1	1.19	N63	1	1.26
Y17	1	0.21	K64	1	0.81
V18	1	2.56	R65	1	0.73
T19	1	1.33	E66	1	5.48
N20	1	0.25	E67	1	1.26
L21	1	0.06	A68	3	0.06
P22	1	-	Q69	3	0.64
R23	1	0.02	E70	1	0.28
T24	1	1.50	A71	1	1.11
I25	1	9.00	I72	1	3.25
T26	1	0.31	S73	1	1.13
D27	1	3.16	A74	1	0.90
D28	1	0.58	L75	1	1.63
Q29	1	0.40	N76	1	3.35
L30	1	1.69	N77	1	0.34
D31	1	0.94	V78	1	0.30
T32	3	2.45	I79	1	2.23
I33	1	2.03	P80	-	
F34	1	2.43	E81	1	0.83
G35	3	1.67	G82	2	1.85
K36	1	1.33	G83	1	1.74
Y37	1	2.75	S84	1	2.20
G38	1	1.51	Q85	1	8.45
S39	1	0.07	P86	-	
V40	1	2.12	L87	1	2.55
I41	1	0.27	S88	1	1.55
Q42	1	2.16	V89	1	0.41
K43	no NOE		R90	1	1.24
N44	1	0.74	L91	1	0.15
I45	-		A92	no NOE	
L46	1	2.16	E93	1	13.10
R47	1	0.37	E94	5	0.00
D48	1	0.12	H95	5	0.00
K49	1	3.32	G96	5	0.00
L50	5	0.00	K97	5	0.00
T51	2	1.44			

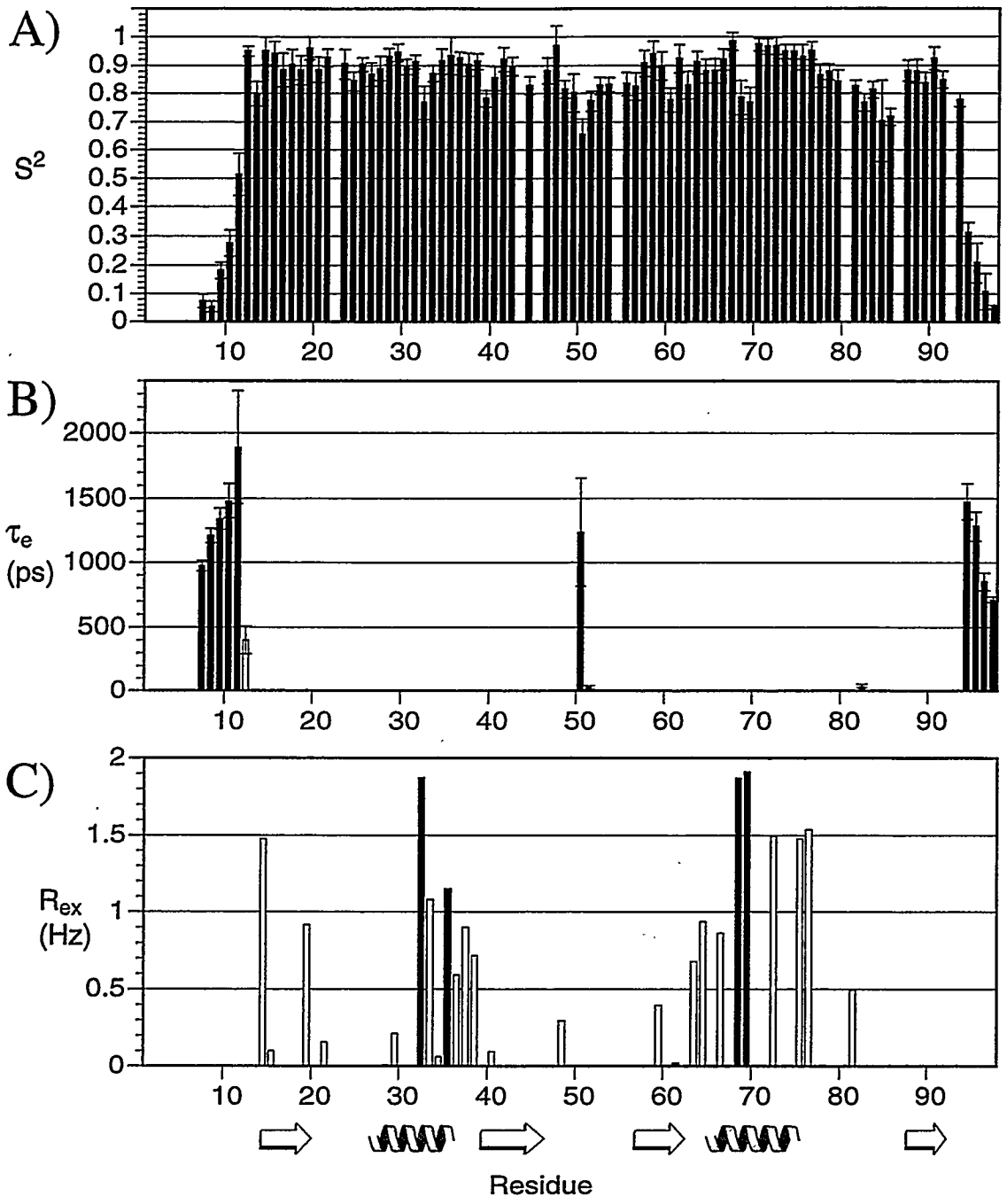


Figure 6.3. Extended Lipari and Szabo "model free" analysis of ^{15}N relaxation data in Figure 2. A) Generalized order parameter, S^2. B) Effective or internal correlation time, τ_e. This parameter comes from either model 2 or model 5 and is deemed significant if it is > 20 ps. Black bars correspond to τ_e derived from model 5. White bars correspond to τ_e derived from model 2. C) Contribution to linewidth (= $1/\pi T_2$), R_{ex}, from chemical exchange or other processes. Since model 4 was not used at all, all of these values are taken from the model 3 fits. Only the residues with black bars were officially ascribed an R_{ex} value for the order parameters calculated in panel A. Error bars are on the order of the values of R_{ex}.

amide protons for D31 and T32 are not protected from exchange even though these residues are in the middle of helix 1. In addition, from the J_{NC} experiment (Chapter 5) T32 has the unusual χ_1 value of -60°, all of which suggest that this portion of helix 1 has some unusual properties, perhaps a breaking or opening process on the microsecond to millisecond timescale.

Concluding remarks

The fast internal backbone dynamics of Sxl-RBD2 has been characterized using ^{15}N NMR relaxation methods. The majority of the protein backbone is rigid, with order parameters in the range of 0.80-0.95. The exceptions are the two termini, which exhibit high dynamic disorder (S^2 approaching zero), and the β2-β3 and α2-β4 loops, which exhibit moderate dynamic disorder (S^2 = 0.6-0.7) in the outlying regions of the loops. Although there were no major surprises based on what is known about the structure of RBD2 (Chapter 4), we now have an accurate, quantitative, and direct description for the amplitudes of motions in the loops, termini, and structured regions of RBD2. This is in contrast to the indirect recognition of "flexible" or unstructured regions based on the absence of NOEs. In the future, it is hoped that complete relaxation experiments can be carried out on an RBD/RNA complex. The dynamic motions in this complex can be compared with those in the free protein, and the role of entropy in RNA binding can then begin to be explored. Because two RBDs in Sxl are necessary for high affinity binding of RNA (Chapter 7), it will be of special interest to observe the motional changes, if any, in the linker residues which connect RBD1 and RBD2.

Abragam, A. (1961) in Principles of Nuclear Magnetism, Clarendon Press, Oxford.

Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992) *Biochemistry 31*, 5269-5278.

Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C. & Gronenborn, A. M. (1990) *J. Am. Chem. Soc. 112*, 4989-4991.

Harris, R. K. (1986) in Nuclear Magnetic Resonance Spectroscopy, Longman Scientific & Technical, New York.

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

Mandel, A. M. (1995) *J. Mol. Biol. 246*, 143-163.

Marion, D. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967-974.

Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991a) *J. Magn. Reson. 93*, 151-170.

Palmer, A. G., Rance, M. & Wright, P. E. (1991b) *J. Am. Chem. Soc. 113*, 4371-4380.

Schneider, D. M., Dellwo, M. J. & Wand, A. J. (1992) *Biochemistry 31*, 3645-3652.

Skelton, N. J., Palmer, A. G., Akke, M., Kordel, J., Rance, M. & Chazin, W. J. (1993) *J. Magn. Reson. 102*, 253-264.

Chapter 7

Complexing Sxl RBDs with *transformer* Polypyrimidine Tracts (PPTs)

Introduction

Because sex determination in *Drosophila* is directly related to the RNA binding activity of Sxl, I was most intererested in using NMR to characterize the interactions between the Sxl RBDs and *transformer* (*tra*) RNA sequences. Previously, Sxl had been shown to bind an RNA 20-mer which contains the polypyrimidine tract (PPT) located just upstream of the 3' splice site between exons 1 and 2 in *tra* pre-mRNA (Inoue et al., 1990). In that paper, they also demonstrated that Sxl binds the 20-mer sequence specifically; a substitution of three cytidines for uridines (...GUUUUUUUUC... → ...GUUCUCUCUC...) abolished RNA binding affinity. Kanaar et al. expanded on these results by characterizing Sxl binding to many different types of PPTs using a quantitative electrophoretic bandshift assay (Kanaar et al., 1995). Finally, Sakashita et al. (Sakashita & Sakamoto, 1994) and Singh et al. (Singh et al., 1995) have used *in vitro* selection methods to determine the RNA sequence requirements for binding to Sxl. As expected, primarily RNAs containing U-stretches were selected for. There was high variability in the placement of the U-stretches within the selected sequences. Although relative binding efficiencies were reported, dissociation constants were not accurately measured. From these selection experiments, the common conclusion is that Sxl prefers to bind RNAs which are rich in uridines. The authors of these two papers each found additional consensus sequence elements which were selected for, but they were not in agreement.

Some of the results from experiments presented in Kanaar et al. (Kanaar et al., 1995) are summarized in Table 7.1. RBD1+2 has the highest affinity for the wild-type Tra-PPT as well as a pronounced specificity for uridines over cytidines in positions 11-15

(Table 7.1). RBD1 and RBD2 both have much lower affinity for the WT Tra-PPT than RBD1+2. Moreover, the individual domains no longer have specificity for uridines in positions 11-15.

Table 7.1: Dissociation Constants for Various RBD/PPT Complexes as Determined by Quantitative Bandshift Assay (Kanaar et al., 1995)

Protein	RNA	K_d(M)
His-Sxl[a]	WT Tra-PPT[b]	1.4×10^{-9}
His-Sxl	Mutant Tra-PPT[c]	8.4×10^{-9}
Sxl-RBD1+2	WT Tra-PPT	6.7×10^{-11}
Sxl-RBD1+2	Mutant Tra-PPT	4.8×10^{-8}
Sxl-RBD1+2	WT Tra-PPT (10-mer)[d]	7.5×10^{-11}
Sxl-RBD1+2	DNA-PPT[e]	5×10^{-9}
Sxl-RBD1+2	DNA-PPT (uracil)[f]	$\sim 5 \times 10^{-7}$
Sxl-RBD1	WT Tra-PPT	4.8×10^{-8}
Sxl-RBD1	Mutant Tra-PPT	5.1×10^{-8}
Sxl-RBD2	WT Tra-PPT	4.0×10^{-6}
Sxl-RBD2	Mutant Tra-PPT	3.8×10^{-6}
Sxl-RBD2	DNA-PPT	$> 3 \times 10^{-6}$

[a] His-Sxl is full-length (354 a.a.) protein which has been expressed with a poly-Histidine tail for purification.
[b] WT Tra-PPT: 5'-UUUUGUUGUUUUUUUUCUAG-3' (positions 1-20)
[c] Mutant Tra-PPT: 5'-UUUUGUUGUU**CUCUCUCUAG**-3' (mutated nucleotides in bold)
[d] WT Tra-PPT (10-mer): 5'-GUUUUUUUUC-3'
[e] DNA-PPT: 5'-d(TTTTGTTGTTTTTTTTCTAG)-3'
[f] DNA-PPT (uracil): 5'-d(UUUUGUUGUUUUUUUUCUAG)-3'

At the time that the results shown in Table 7.1 were being produced by Dr. Roland Kanaar (Rio lab, UC Berkeley), I was characterizing some of these complexes by 1D

NMR. I wanted to study the structural aspects of a specific Sxl-RBD/RNA complex, but I did not know which Sxl RBD construct would be best suited for this type of study. There were three relevant issues: 1) What sequence of RNA should be used? 2) How should milligram quantities of pure RNA be generated? 3) Which Sxl-RBD construct should be complexed with RNA in the NMR tube so that the complex gives good spectra? The first issue was soon to be resolved by Roland—the 10-mer 5'-GUUUUUUUUC-3' retains all binding affinity for RBD1+2 (Table 7.1), but this was not known at the time. The second issue involved a simple choice between *in vitro* transcription of RNA and automated chemical sythesis of RNA. The third issue could only be resolved by systematically making different protein/RNA complexes in the NMR tube, since favorable NMR characterisitics of proteins and protein/nucleic acid complexes cannot be predicted reliably. This chapter will focus on addressing these three issues, and Chapter 8 will focus on the most promising complex identified through the work described in this chapter, the specific complex of RBD1+2 with 5'-GUUUUUUUUC-3'.

Sample Preparations

In vitro transcription of RNA

Basic principles. Milligram quantities of purified RNA can generally be made in two ways: chemical synthesis and *in vitro*, "run-off" transcription using RNA polymerase. I was originally partial to using *in vitro* transcription because it should generally be cheaper and it is relatively simple and inexpensive to incorporate ^{15}N and ^{13}C into the transcribed RNAs (Nikonowicz et al., 1992). In order to transcribe RNA *in vitro*, one needs a DNA template to transcribe from, RNA polymerase to do the transcribing, and NTPs. The DNA template can be in the form of a linearized plasmid where the desired sequence to be transcribed is just upstream of the restriction site. Alternatively, the template can be in the form of synthesized and annealed oligomers which contain the double-stranded polymerase promoter sequence followed by the sequence to be transcribed on the coding strand only.

The RNA polymerase used was T7 RNA polymerase, and therefore all promoters in the plasmid and oligo templates contained a T7 promoter in front of the sequence to be transcribed. Having ^{15}N and ^{13}C isotopic labels would make assignment of RNA resonances much easier when it came time to assign the RNA spectrum. Unfortunately, that time did not come for me. Nevertheless, I expended a lot of time and effort trying to work out some way of generating milligram quantities of *in vitro* transcribed *tra* RNA. It proved to be difficult because of the high occurance of uridines (see below). I never generated milligram quantities of RNA this way, but I believe that I made significant progress in this direction, and in my opinion it should not require vast improvements to make *in vitro* transcription a viable method for generating Tra-PPT RNA sequences.

T7 RNA polymerase and conditions for in vitro transcription. At first, T7 polymerase which was previously purified in the lab by Wemmer graduate students (working on the hammerhead RNA project) was used. This T7 polymerase was prepared using a protocol which had been handed down from student to student over the years. Because they were having a difficult time preparing high quality T7 polymerase using this protocol, I looked for an alternative protocol. Prof. Don Rio suggested trying two different published T7 polymerase purification protocols (Grodberg & Dunn, 1988; Zawadzki & Gross, 1991), and so Elena Rodriguez, an undergraduate working with me at the time, and I each tried one of these "new" protocols. Elena's T7 prep was the best (Figure 7.2). We obtained large quantities (50-100 mg/liter) of T7 polymerase which had perhaps an order of magnitude or more activity compared to the "old" T7 we already had in the lab. We concluded that this purification protocol (Zawadzki & Gross, 1991) is vastly superior. We also concluded that care must be taken with the induction of overexpression of T7 polymerase. It appears that the overexpression level can be quite sensitive to shaker speeds.

Another way of improving transcription yields was in the choice of buffer used for the transcription reactions. Previously, people in our group used either a phosphate or a

tris buffer in their transcription reactions. On one occasion, I tried using Hepes buffer, and this dramatically improved transcription yields for all sequences tried. I settled on the following conditions for my new "standard" *in vitro* transcription recipe:

50 mM Hepes KOH (pH=8)
2 mM spermidine
0.01% triton X-100
5 mM NTPs
10 mM DTT
20 mM $MgCl_2$

The amount of DNA template that was used depended upon the particular sequence and construct used.

Transcription of the Tra-PPT. The first two *tra* RNA sequences that I attempted to transcribe were placed in a plasmid-type DNA template that Roland had previously constructed. pBS-SMP4 encoded for a 53 nucleotide run-off product containing the *tra* PPT and pBS-SB27 encoded for a 33 nucleotide run-off product containing the Tra-PPT. RNA from these plasmids was transcribed with a very high failure rate, resulting in a "ladder" on the gel. These RNAs were longer than what was desired for NMR studies anyway, so I did not spend much time on them. I wanted to use the shortest *tra* sequence possible which retained full binding affinity and specificity because extra nucleotides would give unwanted NMR signals which would only make analysis of the interesting signals more difficult. At the time, the shortest sequence we knew of that retained affinity and specificity was the WT Tra-PPT, 5'-UUUUGUUGUUUUUUUUCUAG-3'. I decided that it would be worthwhile to investigate whether oligo-based run-off transcription would work better than plasmid-based transcription on these sequences, so I prepared a DNA-oligo which would serve as the coding strand for T7 polymerase and transcribe the above 20-mer. In order to maximize transcription yields, nucleotides corresponding to 3 guanosines were inserted at the beginning of the *tra* sequence (Brown et al., 1986). Ho Cho, a Wemmer graduate student, was kind enough to give me plenty of T7 promoter "top

strand". Top and coding strands were mixed at concentrations ranging from 2 to 40 uM. The mixtures were heated to 90 °C for 1 minute and then cooled to 37 °C, prior to addition of T7 polymerase. Despite switching to oligo-based transcription and using excellent T7 polymerase (determined by the transcription yields of hammerhead RNAs as a positive control), the 23-mer transcription product yield was miserably low.

Long stretches of uridines are perceived as notoriously difficult to transcribe *in vitro* (Dave Wemmer communication with Art Pardi). If this is indeed true, then transcribing the Tra-PPT is inherently a difficult problem. However, it may be that the problems that T7 polymerase has with transcribing long U-stretches may not be due to polymerase processivity problems, but to problems with initiation of transcription. Since the above transcription sequences are all relatively short and have long U-stretches, the poor yields may result from T7 polymerase not attaining full processivity before it hits the U-stretches. This idea suggested that there could be a solution to the Tra-PPT transcription difficulties.

Hammerhead-based strategy. One possible way of getting the Tra-PPT transcribed with higher efficiency is to ensure that T7 polymerase has attained full processivity before it hits the U-stretches. In our lab we have always had excellent transcription yields from hammerhead ribozyme RNAs. In particular, RNA3 hammerhead (Anne Caviani's Ph.D. dissertation), a self-cleaving RNA, is a bimolecular ribozyme consisting of a "3/4" ribozyme RNA and a "1/4" ribozyme substrate RNA. Both RNAs transcribe well *in vitro*. Figure 7.1 depicts the strategy I proposed for a potential increase in transcription efficiency for the *tra* PPT. The basic idea is to substitute the PPT sequence for the non-conserved RNA3 "1/4" sequence at positions which are 3' of the cleavage site. In this scheme, the PPT at the 3' end of the "1/4" strand will be cleaved off by the "3/4" strand. Since the RNA "1/4" strand is known to transcribe well, T7 polymerase is likely to have fairly high processivity after the cleavage site, which should allow for the PPT to be transcribed with a lower failure rate. Because only the boxed regions in the RNAs are conserved, all other nucleotides can be changed as long as standard Watson-Crick base pairing is preserved.

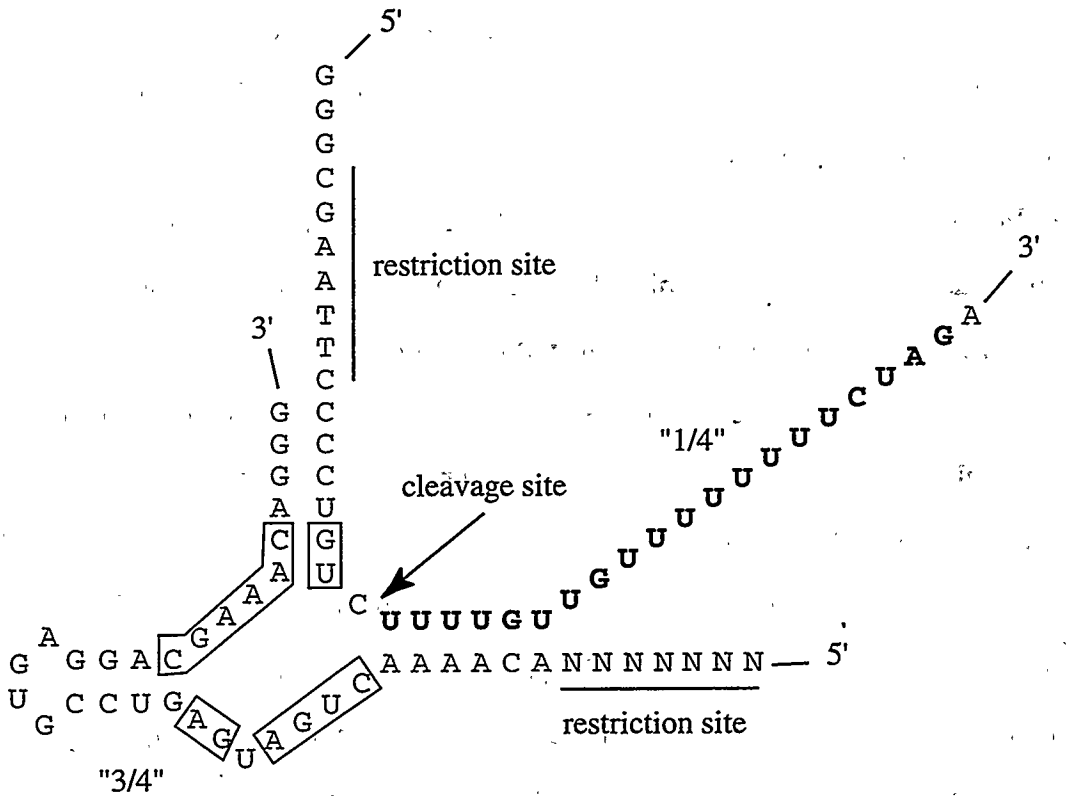


Figure 7.1. "Hammerhead 1/4 - 3/4" strategy proposed to both increase *in vitro* transcription yields and result in an RNA with the undesired 5' nucleotides cleaved off by the "3/4" ribozyme. The "1/4" strand is encoded in the pSMP025 plasmid. The "3/4" strand is encoded in the pSMP075 plasmid. The TRA-PPT is in bold, and boxed nucleotides are those which are conserved for ribozyme activity.

The desired fragment can then be purified from the other longer RNA fragments by either slab-gel or anion exchange HPLC.

Roland sub-cloned the two required plasmids, pSMP025 and pSMP075, which contained the "1/4" and "3/4" sequences after a T7 promoter sequence. In both cases the plasmid could be linearized by use of the unique HindIII restriction site located at the end of the "1/4" and "3/4" sequences. Figure 7.2 shows the transcription yields obtained for pSMP025 and pSMP075 relative to some other *in vitro* transcribed RNAs. The "3/4" strand transcribes quite well, not a surprise since it is quite similar to the corresponding strand in RNA3. The U-rich "1/4" strand does not transcribe as well as the "3/4" strand, but the yield is considerably better than any other RNA containing the *tra* U-stretches. This was quite an encouraging result. With both pieces transcribed, it was now possible to test if the ribozyme strategy would work. After several attempts, it became clear that the "3/4" strand would not (at least in my hands) cleave the "1/4" strand if I tried to transcribe the two strands in the same tube or if they were transcribed separately and mixed afterward. Figure 7.3 shows that the cleavage appears to have worked after cleaning up the transcribed RNAs. The band which appears in lane 4 ("1/4") is converted into two smaller bands in lanes 6-8. Apparently it is necessary to remove the protein from the transcription reactions by extracting the reactions with phenol/chloroform/isoamyl alcohol. I also desalted the extracted solution using a NAP-5 column (Pharmacia). At the end of this procedure, 80 uL of pSMP025 transcription reaction yielded approximately 1 ug of 21-mer *tra* RNA (Figure 7.1), as estimated from UV shadowing band intensities. In Figure 7.3, I also attempted to show a Mg^{2+} dependence of the cleavage reaction (lanes 6-8), but clearly I failed in this regard. The hammerhead cleavage reaction should have a Mg^{2+} dependence, and I most likely did not remove all residual Mg^{2+} from the "desalted" RNA mixtures.

Although the hammerhead strategy for generating *tra*-PPT RNA appears to have worked, I felt that the yield was not quite high enough to warrant a preparative-scale transcription reaction. Also, by this time Roland had shown that the 10-mer, 5'-

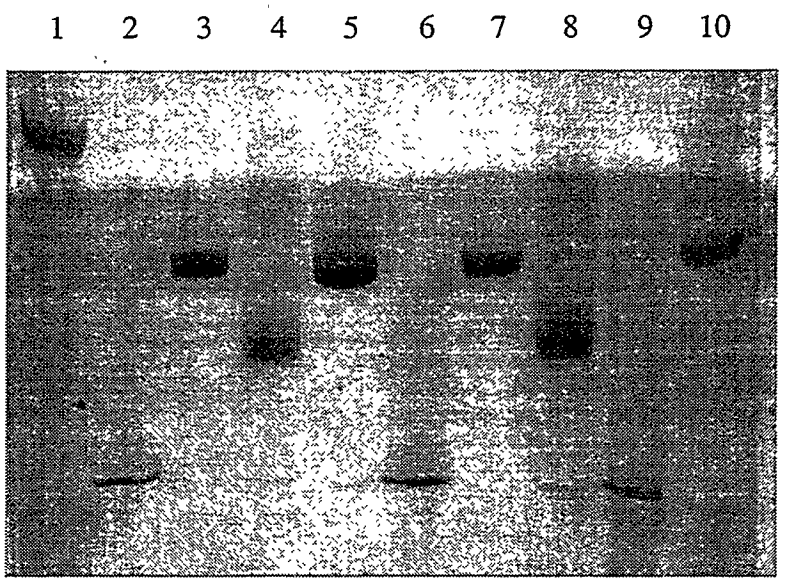


Figure 7.2. Run-off transcription of pSMP025 "1/4" and pSMP075 "3/4" RNAs from Figure 7.1 using three different batches of T7 RNA Polymerase. Nine 50 ul reactions were done at 37 °C, and 8 ul from each reaction were run out on an 18% polyacrylamide gel with 6 M urea. The volumes of ~1-2 mg/ml DNA template plasmid used in the 50 ul reactions is given for each lane. Lane 1) 5 ug t-RNA. Lane 2) 0.02 ul of Elena's T7 + 7 ul FUTZ plasmid. Lane 3) 0.06 ul of Elena's T7 + 2 ul RNA8 plasmid. Lane 4) 0.06 ul Elena's T7 + 4 ul pSMP025 ("1/4"). Lane 5) 0.06 ul Elena's T7 + 4 ul pSMP075 ("3/4"). Lane 6) 1 ul Andrew's T7 + 7 ul FUTZ plasmid. Lane 7) 1 ul Andrew's T7 + 2 ul RNA8 plasmid. Lane 8) 1 ul Andrew's T7 + 4 ul pSMP025 ("1/4"). Lane 9) 1 ul Don's T7 + 7 ul FUTZ plasmid. Lane 10) 1 ul Don's T7 + 2 ul RNA8 plasmid.

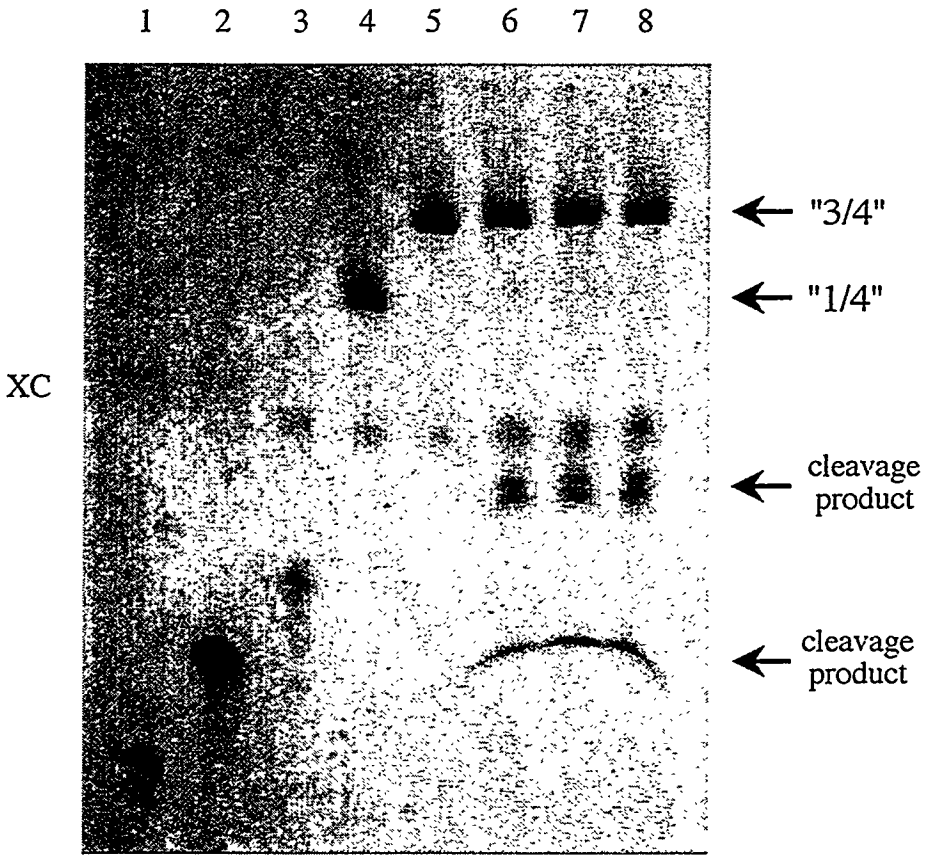


Figure 7.3. Demonstration of the hammerhead intermolecular cleavage of the "1/4" RNA containing the Tra-PPT by the "3/4" ribozyme. RNAs were run out on a 23% polyacrylamide gel with 6 M urea. Lane 1) RNA 10-mer (GUUUUUUUUC) control. Lane 2) RNA 15-mer control. Lane 3) RNA 20-mer (UUUUGUUGUUUUUUUUCUAG) control. Lane 4) purified (see text) "1/4" RNA 38-mer transcribed from pSMP025. Lane 5) purified (see text) "3/4" RNA 41-mer transcribed from pSMP075. Lane 6) "1/4" + "3/4" incubated for 90' at 37 °C in 0 mM $MgCl_2$. Lane 7) "1/4" + "3/4" incubated for 90' at 37 °C in 10 mM $MgCl_2$. Lane 8) "1/4" + "3/4" incubated for 90' at 37 °C in 20 mM $MgCl_2$.

GUUUUUUUUC-3', was sufficient to retain all binding affinity to Sxl-RBD1+2 (Table 7.1), and Prof. Don Rio had already used automated chemical synthesis to make 20-30 mg of this 10-mer. Finally, I would like to mention that soon after I demonstrated the success of the hammerhead strategy (Figure 7.3), Price et al. published a similar strategy for improving the quality and quantity of *in vitro* transcription products (Price et al., 1995).

Materials and methods

DNA synthesis, deprotection, and purification. 5'-d(TTTTGTTGTTTTTTTTCT AG)-3' (10 umole) was synthesized with DMT on, deprotected with 30% NH_4OH at 70 °C, and HPLC purified with standard 0.1 M TEAA (triethylammonium acetate) buffers on a preparative C-18 reverse phase column.

RNA synthesis, deprotection, and purification. All RNA used in NMR experiments was synthesized by Prof. Don Rio at UC Berkeley on an ABI 392 DNA/RNA synthesizer. Initially, he did 30 1-umole syntheses of 5'-GUUUUUUUUC-3'. The RNA from this first batch was used up in the initial complex studies when I was learning about the severity of the RNase problem. In August of 1995 Don completed 21 additional 1-umole syntheses of the same sequence.

Deprotection of the 2'-OH t-butyldimethylsilyl protecting group was carried out as follows. First, the 21 syntheses were pooled into 3 groups of 7 1-umole syntheses. The RNA was cleaved off of the support resin by addition of 4 mL 30% NH_4OH:EtOH (3:1) to each of the 3 vials, and these were incubated at 55 °C for 5 hours. The solutions were then transferred to Eppendorf tubes and dried down in a speed-vac apparatus without heat. The dried RNA was then resuspended in 14 mL of triethylamine trihydrofluoride (10 uL per ODU) for deprotection and stirred in a 50 mL conical tube with a magnetic stir bar for 24 hours. The deprotection reaction was quenched with 2.8 mL H_2O, and 140 mL of 1-butanol was added. This volume was spread over 16 2059 Falcon tubes and incubated at -20 °C overnight. The tubes with RNA were then spun at 3000 rpm for 35 minutes,

although it would have been better to spin them at 5000 rpm. The supernatent was removed and the pellets dried at 65 °C for ~10 minutes. The "dry" pellets at this point looked more like a gooey oil, probably due to the t-butyldimethylsilyl protecting group. Pellets were stored at -80 °C. The pellets, although oily-looking, were water miscible and were resuspended in ~200-300 uL of H_2O prior to HPLC purification.

Deprotected RNA was purified on a Dionex NucleoPac PA-100 (9 x 250) HPLC anion exchange column. The following 2 buffers and elution gradient were used:

Buffer A: 20 mM $LiClO_4$
20 mM NaOAc
pH to 6.5 with dilute acetic acid

make this solution in 9:1 H_2O/CH_3CN

Buffer B: 600 mM $LiClO_4$
20 mM NaOAc
pH to 6.5 with dilute acetic acid

make this solution in 9:1 H_2O/CH_3CN

Gradient:

time	flow (ml/min)	%A	%B
0	1.0	100	0
16	1.0	100	0
56	1.0	90	10

The 10-mer elutes at the 45 minute mark. The RNA eluate was precipitated at -20 °C overnight with 4 volumes n-propanol. Only the RNA precipitates, not the $LiClO_4$. The RNA was stored in this precipitated form. RNA was eventually collected by spinning at 8000 rpm for 15 minutes, removing the supernatent, drying the pellet at 65 °C for ~10 minutes, and resuspending in H_2O.

Removal of RNases from RBD2 and RBD1+2

RBD1, RBD2, and RBD1+2, when purified as described in Chapter 3, have problematically high levels of RNase activity which co-purify with the RBDs. Perhaps the

level of RNase activity is acceptable for low concentration work at low temperatures, but for the NMR studies intended to be carried out at millimolar protein and RNA concentrations, room temperature, and for up to 80-90 hours of continuous data collection, this level of RNase activity would undoubtedly destroy the complex in the early stages of these long NMR experiments. This egregious RNase problem was still more painful since even though the RNase activity was relatively high, the concentrations of RNases were sufficiently low such that RNase bands could not be observed on an SDS-PAGE gel. Therefore, the most convenient method for assaying RNase activity was to mix small quantities of "purified" protein and RNA at near-NMR concentrations and let them incubate at approximately room temperature for about a week. This RNase assay will be described in greater detail below. To reduce the level of RNase activity in purified RBD2 and RBD1+2, additional purification steps were appended to the purifications described in Chapter 3. An additional purification of RBD1 was not developed, as RBD1 was generally more difficult to work with and it does not form a specific complex with RNA (Table 7.1). Nevertheless, initial NMR characterization of RBD1 complexed with PPTs was still possible since 1D monitored titrations can generally be carried out significantly faster than the rate of RNA degradation . Only once was there so much RNase activity that it actually made a titration difficult (this was with RBD1+2). The additional purification steps for reducing the level of RNase activity will now be described for RBD2 and RBD1+2. The RBD1+2 extended purification scheme was developed by Roland Kanaar.

RBD2. I found that a single HPLC run using a C-18 reverse phase column was sufficient to completely remove all RNase activity from RBD2. After the fractions of RBD2 from the blue trisacryl column were pooled (~80 mL), RBD2 was concentrated with a Centriprep-10 device made by Amicon Inc. Just prior to the HPLC injection, the supernatent was filtered through a 0.45 micron filter. HPLC uses three buffers, A, B, and C; buffer A is 0.1% TFA in H_2O, buffer B is 60% acetonitrile/40% H_2O with the mixture

at 0.1% TFA, and buffer C is 100% acetonitrile. The column used was a preparative reverse-phase C-18 Waters Delta-Pak column. The gradient employed was:

time (min)	%A	%B	%C
0	60	40	0
2	60	40	0
52	30	70	0
60	0	0	100%

flow rate = 12 ml/min
wavelength = 220 nm
AUFS = 2.0

RBD2 eluted at ~31-33 minutes and was completely RNase free. At this point, RBD2 was either lyophilized or dialyzed against H_2O and then lyophilized.

RBD1+2. Roland found that 2 additional columns, Poros HS and Superdex-75, were sufficient to remove excess RNase activity from RBD1+2. After the fractions of RBD1+2 from the blue trisacryl column were pooled (~100-200 mL), RBD1+2 was concentrated by a 65% ammonium sulfate precipitation. The cloudy precipitate solution was transferred into clean GSA bottles and centrifuged at 10,000 rpm for 25 minutes. The pellet(s) was resuspended in < (5 mL/10 mg RBD1+2) (see below) of Hepes buffer with 50 mM NaCl and dialyzed overnight against this buffer with 1 buffer change.

Hepes buffer:
20 mM Hepes (pH=8.0 for 1 M solution)
1 mM EDTA
1 mM DTT
10% glycerol

50 mM NaCl

The first "extra" column is a Poros HS FPLC cation exchange column. In order to remove as much RNase as possible, it is best to do several small injections (7-15 mg RBD1+2 in less than 5 mL) onto the HS column instead of one large injection (50 mg). Therefore, the ammonium sulfate pellet in the previous step can be resuspended in a volume of buffer that

results in an injection volume of < 5 mL for each run. The protein was then filtered with a 0.2 micron low protein-binding filter. 10 mg per injection of RBD1+2 was loaded onto the HS column and the run was carried out according to the following specifications:

column bed = 1 mL
flow rate = 1.0 mL/min
chart recorder = 0.5 cm/mL

Buffer A: Hepes buffer with 0 M NaCl
Buffer B: Hepes buffer with 1 M NaCl

Gradient:	mL	%B	
	0	5	
	5	5	
	10	5	
	22	28	
	23	100	
	25	100	wash
	30	5	
	35	5	re-equilibrate for next run

The elution profile for this HS column is shown in Figure 7.4A. Normally 1.0 mL fractions were collected. At this point of the purification, approximately 200 ug of the individual fractions or of the pooled fractions of RBD1+2 was saved for assaying of RNase. The desired fractions (#2-#6 in Figure 7.4A) from the HS column were then pooled and concentrated by another 65% ammonium sulfate precipitation carried out in an RNase-free Beckman SS-34 tube. The pellet was spun down for 15 minutes at 12,000 rpm and resuspended in 1-2 mL Hepes buffer with 200 mM KCl (instead of 50 mM NaCl). This solution was then dialyzed against this same buffer overnight with 1 buffer change. It is important to keep the volume low in order to have high resolution on the next column, a Superdex-75 sizing column. Just prior to running this column, the sample was filtered again with a 0.2 micron low protein-binding filter. Less than 5.0 mL RBD1+2 was injected onto the Superdex-75 FPLC column and the run was carried out isocratically:

column bed = 120 mL
flow rate = 0.5 mL/min (1.0 mL/min is OK too)
chart recorder = 0.2 cm/mL

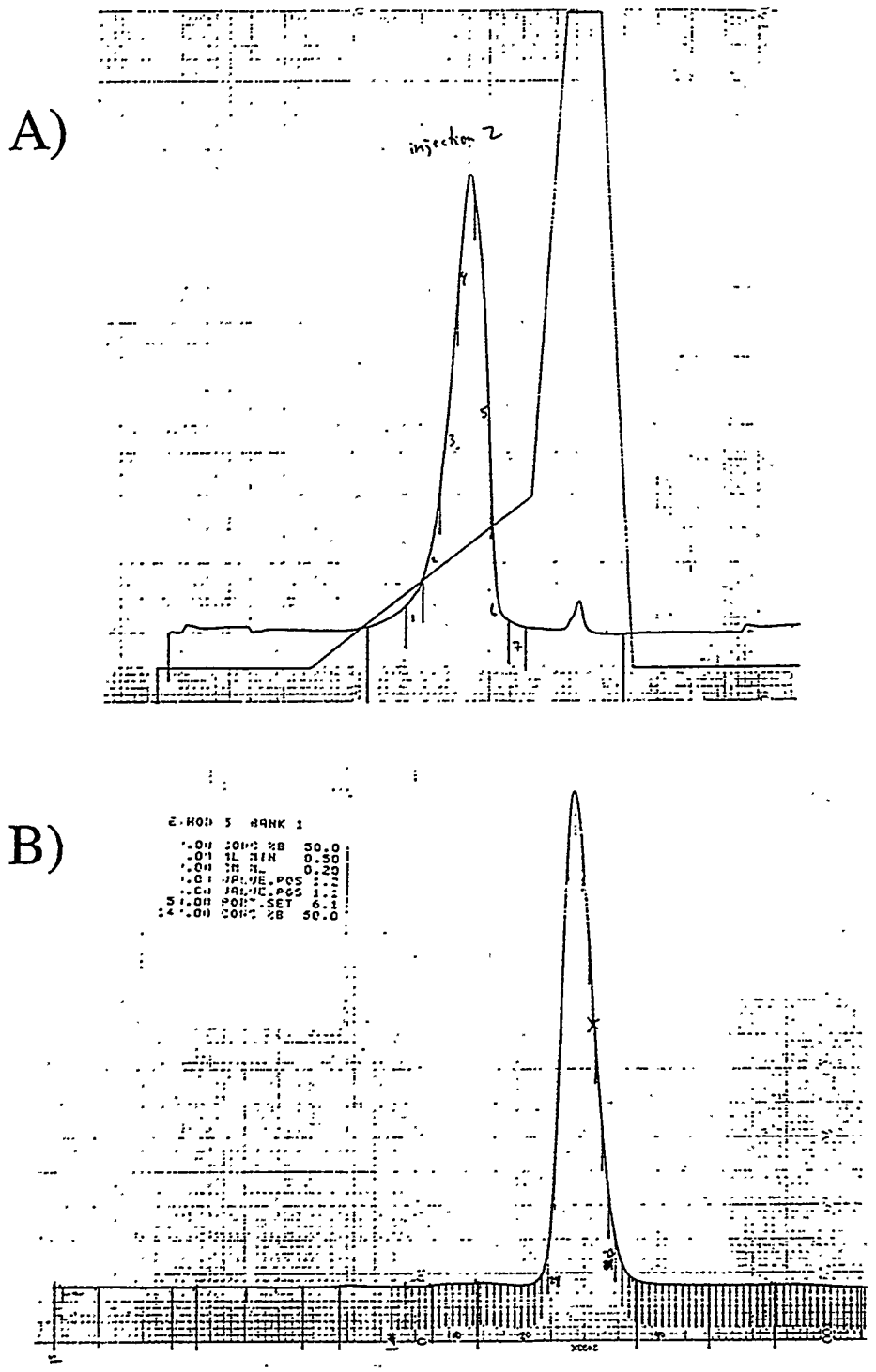


Figure 7.4. A) FPLC trace of RBD1+2 run over a Poros HS cation exchange column. The NaCl gradient is superimposed over the run. The major peak was collected in 1 mL fractions (fractions 2-6). B) FPLC trace of RBD1+2 run over a Superdex-75 gel-filtration column. The peak was collected in 1 mL fractions (fractions 24-33). The run program printout from the Pharmacia FPLC is also given.

buffer = Hepes buffer with 200 mM KCl (this column is run isocratically)

Run for 1 column volume, and then equilibrate for 2 column volumes.

In my hands, RBD1+2 eluted in approximately 10-12 mL. Figure 7.4B shows a representative Superdex-75 run. At this point RBD1+2 had significantly lower levels of RNase activity than after the blue trisacryl column, and this was demonstrated by RNase assays as described in the next section.

My experience, mostly with RBD1+2, was that the levels of RNase activity *before and after* the purification was very prep- and growth-dependent. In all cases, however, the extra two columns reduced the amount of RNase activity. This will be demonstrated by a comparison of "RNase-free" sample preparations between uniformly ^{15}N-labeled and uniformly ^{13}C- and ^{15}N-labeled RBD1+2 (next section). Therefore I believe that if this approach to removing RNases from "purified" protein is employed, great care must be taken to ensure that *all* surfaces that the protein (or *E. coli*) contacts are RNase-free.

RNase Assays. In order to monitor the levels of RNase activity at any stage of the protein purifications, RNase assays were performed. Briefly, protein and RNA (5'-GUUUUUUUUC-3') were combined at a 1:1 ratio at near-NMR concentrations. Typically, two concentrations at 50 uM and 250 uM were used in the hopes that a concentration-dependent activity could be observed. These were incubated at room temperature (normally around 22 °C) for up to 1 or 2 weeks. It was necessary to use such long incubation times since the NMR experiments would last up to 3 or 4 days and would require that the complex remain intact for that duration. At the end of the incubation period, the protein/RNA mixture was extracted with 1 volume of phenol/chloroform/isoamyl alcohol, followed by a second extraction with chloroform containing 4% isoamyl alcohol. This procedure removed protein from the mixture and left the RNA (or nucleotides) in the aqueous phase. I found that these extractions were most easily done if I started with 100 ul of protein/RNA incubation mixture (water was added at the end of the incubation period to

bring the volume up to 100 ul in most cases). The RNA was then precipitated by adding 1/20 volume of 3 M NaOAc, 4 volumes of 1-propanol, and letting it sit at -20 °C overnight. The precipitate (in Eppendorf tubes) was then spun down at 4 °C at approximately 12,000 rpm for 15 minutes. The supernatent was discarded and the pellets were quickly dried in a speed-vac apparatus. These resultant pellets were then resuspended in loading buffer (80% formamide, bromophenol blue, xylene cyanol) and ran out on a 23% denaturing polyacrylamide gel at approximately 300 volts.

Figure 7.5 shows the RNase assay on the uniformly ^{15}N-labeled RBD1+2 prep for various stages of the purification. The reactions were set up so that a total of ~21 ug of RNA per tube was used. This would allow for 3 separate lane loadings on a gel for each assay tube. Because 5 ug of RNA shows up well on a UV-shadowed gel, and because approximately 30% of the RNA gets lost during the extractions and precipitation step, it is optimal to start the assays with ~7 ug of RNA for each lane on a gel. In Figure 7.5, only 2 lanes per assay tube were needed (24 hours, 6 days), leaving enough assay in the tubes left over for another lane if something were to go wrong or if a third time point longer than 6 days was desired. From Figure 7.5, it is clear that the HS column removes approximately 50% of the RNase activity present after the dye column, and the sizing column removes more than 50% of the RNase activity present after the HS column. At concentrations of up to 250 uM protein/RNA complex the RNase activity in the 4-column purified protein appears to be under control for up to 6 days. This is quite acceptable. However, it was my experience that the RNase activity was significantly higher after dialysis into NMR buffer, concentration to 0.7 mM, and transfer into the NMR tube (treated with 50% HNO_3 for at least 3 days). Certainly there is the possibility that I introduced RNase during one of these steps. It is also possible that concentration effects are at play or that conditions in the NMR tube are different than in the assays. Indeed, these two environments have different levels of glycerol, buffers, and salts. Slight differences in temperature could also be a source of discrepancy. Room temperature assays ranged in temperature from approximately 21 °C to

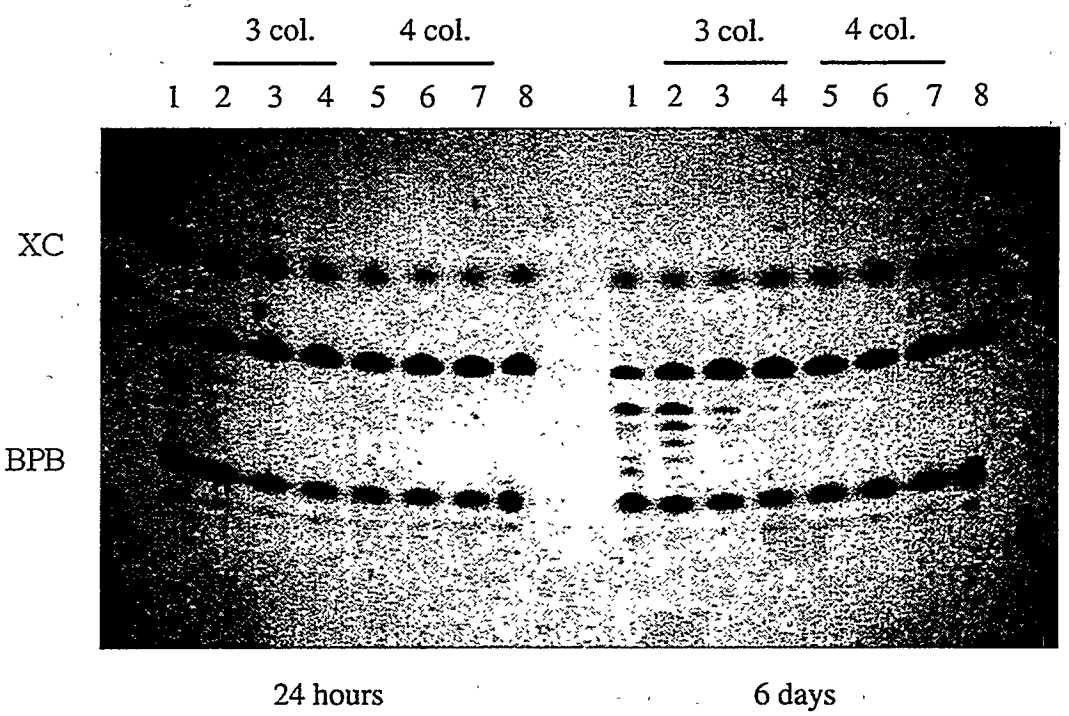


Figure 7.5. RNase Assay on uniformly labeled ^{15}N RBD1+2 as a function of purifcation steps, concentration, and time. 1.5 ul of 5 mM RNA was used in each reaction (protein/RNA = 1:1). Lanes 1-8 are the same for the 24-hour incubation and the 6-day incubation. Incubations were carried out at room temperature (~21-24 °C)

Lane	Description	Inhibitor *	[RBD1+2]	Total Vol.
1	**50 ul** 1+2, 2-column pure (BTA)	-	50 uM	150 ul
2	**19 ul** 1+2, 3-column pure (HS)	-	50 uM	150 ul
3	**19 ul** 1+2, 3-column pure (HS)	-	250 uM	30 ul
4	**19 ul** 1+2, 3-column pure (HS)	10 U rRNasin	250 uM	30 ul
5	**22 ul** 1+2, 4-column pure (S-75)	-	50 uM	150 ul
6	**22 ul** 1+2, 4-column pure (S-75)	-	250 uM	30 ul
7	**22 ul** 1+2, 4-column pure (S-75)	10 U rRNasin	250 uM	30 ul
8	**148.5** ul phosphate buffer, pH=6.3	-	-	150 ul

* Recombinant RNasin obtained from Promega.

24 °C, as opposed to the NMR which was done at either 20 °C or 25 °C. The RNase activity at 25 °C was always noticeably higher than at 20 °C, so temperature may play an important role. In the end, the source of this differential activity (which I have observed repeatedly) between assays and finished samples remains unknown. Nevertheless, this particular ^{15}N RBD1+2/10-mer complex was sufficiently stable in the NMR tube. At 25 °C the RNA degraded approximately 15-20% over 4 days. At 20 °C the amount of RNA degradation over 4 days was negligible.

Lanes 4 (24-hour time point) and 7 (6-day time point) in Figure 7.5 show that RNasin (recombinant RNasin obtained from Promega) does indeed inhibit RNase activity. Further experimentation with RNasin convinced me that measures would have to be taken to prevent the RNasin from petering out, and that associated inconveniences with RNasin made this option less attractive than the 4-column purification. The reader should refer to my notebook if more information on inhibitors is desired.

Figure 7.6 shows the RNase assay on the uniformly ^{13}C-, ^{15}N-labeled RBD1+2 prep for various stages of the purification for comparative purposes with Figure 7.5. Unfortunately, because I was rushing to get the sample shipped off to Madison, WI (see Chapter 8), I had to stop this assay after only 4 days. A comparison between Figures 7.5 and 7.6 reveals the fact that different sample preparations have differently behaving RNase activities, at least in my hands. After the 4-column purification, this sample had more RNase activity than in the ^{15}N-labeled RBD1+2, a disappointment. Looking closely, each column removed RNase activity but not as dramatically as with the previous sample. It is possible that this activity comes from a different type of RNase than in the previous sample. We (Dr. Brian Volkman and I, see Chapter 8) were forced to work at 20 °C for this ^{13}C-, ^{15}N-labeled sample since the rate of RNA degradation at 25 °C was crippling. After 40 hours at 20 °C, the RNA in this sample was approximately 30% degraded, a barely acceptable rate.

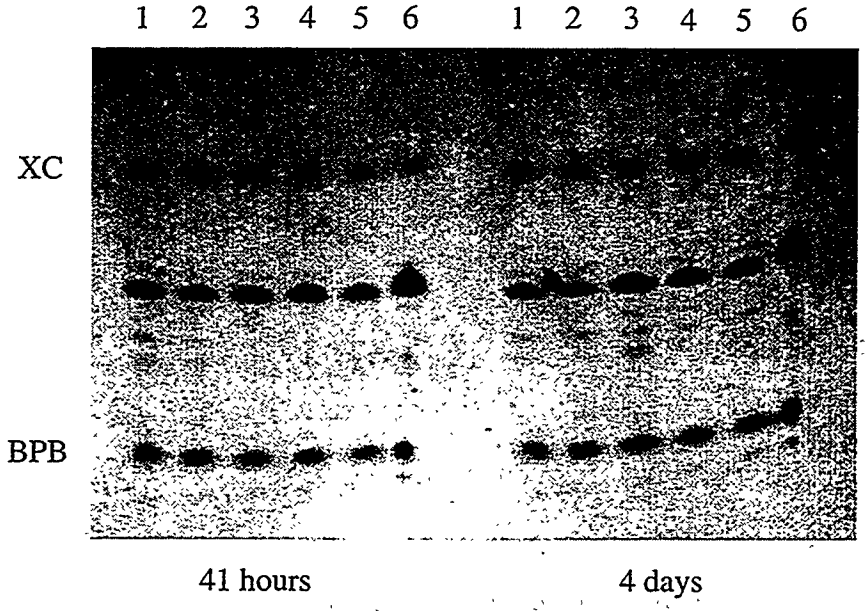


Figure 7.6. RNase Assay on uniformly labeled ^{13}C, ^{15}N RBD1+2 as a function of purifcation steps, concentration, and time. 2.9 ul of 1.6 mM RNA was used for each reaction (protein/RNA = 1:1). Lanes 1-6 are the same for the 41-hour incubation and the 4-day incubation. Incubations were carried out at room temperature (~21-24 °C).

Lane	Description	[RBD1+2]	Total Vol.
1	**31 ul** 1+2, 2-column pure (BTA)	47 uM	100 ul
2	**18 ul** 1+2, 3-column pure (HS)	47 uM	100 ul
3	**18 ul** 1+2, 3-column pure (HS)	197 uM	24 ul
4	**22 ul** 1+2, 4-column pure (S-75)	47 uM	100 ul
5	**22 ul** 1+2, 4-column pure (S-75)	197 uM	24 ul
6	**97.1 ul** phosphate buffer, pH=6.3	-	100 ul

NMR Characterization of RBD/Tra-PPT Complexes

A note on the "NMR timescale"

For intermolecular complexes, a major concern in NMR is the kinetics of exchange between states. In this particular instance, the exchange is between the free and bound states of both the protein and the RNA. Different exchange regimes exist because NMR spectra are not instantaneous "snapshots" of molecules. For protein FT-NMR the shortest data collection time period typically lasts for approximately 100 ms. And because all signals are simultaneously detected as oscillating sinusoids and these sinusoids can only be differentiated from one another after approximately ($1/\Delta\nu$) seconds (where $\Delta\nu$ is the frequency difference between two sinusoidal signals to be differentiated), exchange processes can affect the appearance of the individual sinusoidal signals. For example, if a molecule is in equilibrium between 2 states (free and bound), individual spins can have 2 frequencies, one for the free and one for the bound state. Therefore each spin has its own $\Delta\nu$ (ν_{free}-ν_{bound}), independent of all other spins. If the molecule converts (exchanges) from one form to the other during the data collection period, the sinusoids will change frequency during this period and complicate the resultant Fourier transformed spectrum. Of course, in reality these effects are stochastic for an ensemble of molecules in solution, and therefore a smooth distribution of these effects are observed.

Depending on the exact rates of exchange between states (as well as the values of $\Delta\nu$), there are 3 regimes of exchange effects on a frequency domain NMR spectrum:

1) "Fast Exchange" — The molecule is undergoing exchange between states at a rate much faster than the differences in chemical shifts between the states ($\Delta\nu$). In this regime, the observed sinusoids for each spin are well defined as single apparent frequencies. However, because of the fast exchange, these apparent frequencies correspond to the population weighted averages of the individual frequencies in the populated states. Lines are not broadened by the exchange process(es).

2) "Intermediate Exchange" — The molecule is undergoing exchange between states at a rate comparable to $\Delta\nu$. This regime is considered undesirable by just about everyone. Stochastic effects give rise to severe line-broadening. Under particularly bad circumstances, entire regions of the spectrum can seem to disappear entirely.

3) "Slow Exchange" — The molecule is undergoing exchange between states at a rate much slower than $\Delta\nu$. If two (or more) states are significantly populated, the individual signals (or spectra) from both states become superimposed in the resultant spectrum. Lines are not broadened by the exchange process(es).

The relationship between $\Delta\nu$ and exchange rates and the various exchange regimes that arise from this relationship is known as the "NMR timescale". These regimes will be frequently referred to in following sections.

RBD/PPT complexes studied by NMR

To further characterize the interaction between the Sxl RBDs and the Tra-PPT, solution NMR studies on Sxl-RBD/Tra-PPT complexes were initiated. Because RNA was not generated early on, I did some initial protein/Tra-PPT NMR experiments with a DNA-PPT (Table 7.1) since DNA is inexpensive and easier to handle in general than RNA. The DNA sequence employed was 5'-d(TTTTGTTGTTTTTTTTCTAG)-3'. This DNA-PPT was studied in a complex with RBD2, the RBD construct with the most favorable solution properties. After RNA was generated in milligram quantities (Tra-PPT 10-mer), each of the three RBD constructs was complexed with this RNA in order to evaluate which complex would be the most interesting to study. All 1D NMR spectra collected on these complexes are of the "1-1" type to ensure that saturation transfer effects would not attenuate signals (via exchange with solvent) from imino base protons on the DNA or RNA.

RBD2/DNA-PPT. I first studied a complex between uniformly ^{15}N-labeled Sxl-RBD2 and the DNA-PPT. The K_d for this complex, as estimated from electrophoretic mobility shift experiments, is > 3 μM (Table 7.1). This K_d is similar to the K_d for the interaction between RBD2 and the RNA Tra-PPT. Although the details of the binding modes of the DNA- and RNA-PPTs will not be identical, NMR analysis of the complex between RBD2 and the DNA-PPT should allow qualitative mapping of the protein surface involved in nucleic acid binding. As expected from the lower limit on the value of the K_d, the protein resonances of the free and bound forms were found to be primarily in fast exchange on the NMR timescale (from HSQC titration). The 1D titration is shown in Figure 7.7. Notice that a broad resonance from the imino protons on the DNA grows in. There may be some protection of these protons from exchange with solvent, but because the signal is so broad, there is most likely not any hydrogen bonding involving the imino protons. Fast exchange behavior allows the transfer of protein assignments from free to bound states during the course of a titration. As the DNA-PPT was titrated into a solution containing a constant amount of RBD2, the titration was monitored by taking 2D ^{15}N/^{1}H HSQC spectra. Individual backbone amide resonances could be tracked during the course of the titration for nearly all previously assigned residues (Chapter 4). For chemical shift differences between the two forms (of RBD2) of > 100 Hz, significant line-broadening was observed. The final protein to DNA ratio was estimated to be approximately 1:1.5.

The differences in ^{1}H and ^{15}N chemical shifts between free and bound states are plotted as a function of residue position in Figure 7.8. Residues that are interacting with the DNA or that are positioned near the binding interface can be expected to have relatively large chemical shift changes. The largest chemical shift changes were observed for the RNP-1 and RNP-2 consensus sequences (β-strands 3 and 1, respectively). Other regions of significant perturbations are in β-strand 2, the first few residues of the β2-β3 loop, and the carboxy-terminal residues after β-strand 4. It is interesting that β-strand 2 appears to interact with the DNA-PPT yet β-strand 4 has minimal interaction. The significant

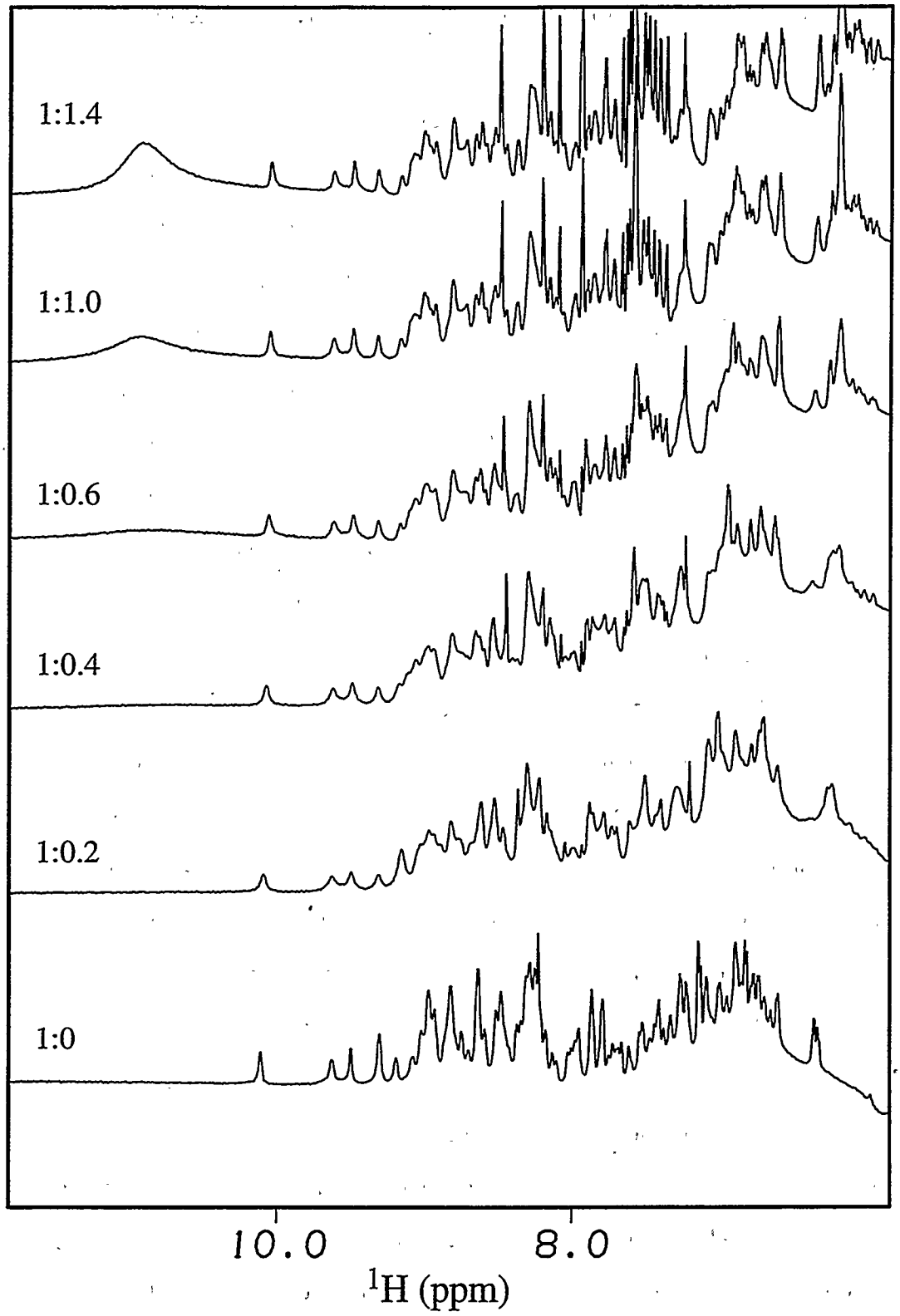


Figure 7.7. 1D Titration of RBD2 with 5'-d(TTTTGTTGTTTTTTTTCTAG)-3', pH = 6.0, RBD2 concentration = 1 mM. RBD2/DNA stoichiometries are shown for each titration step.

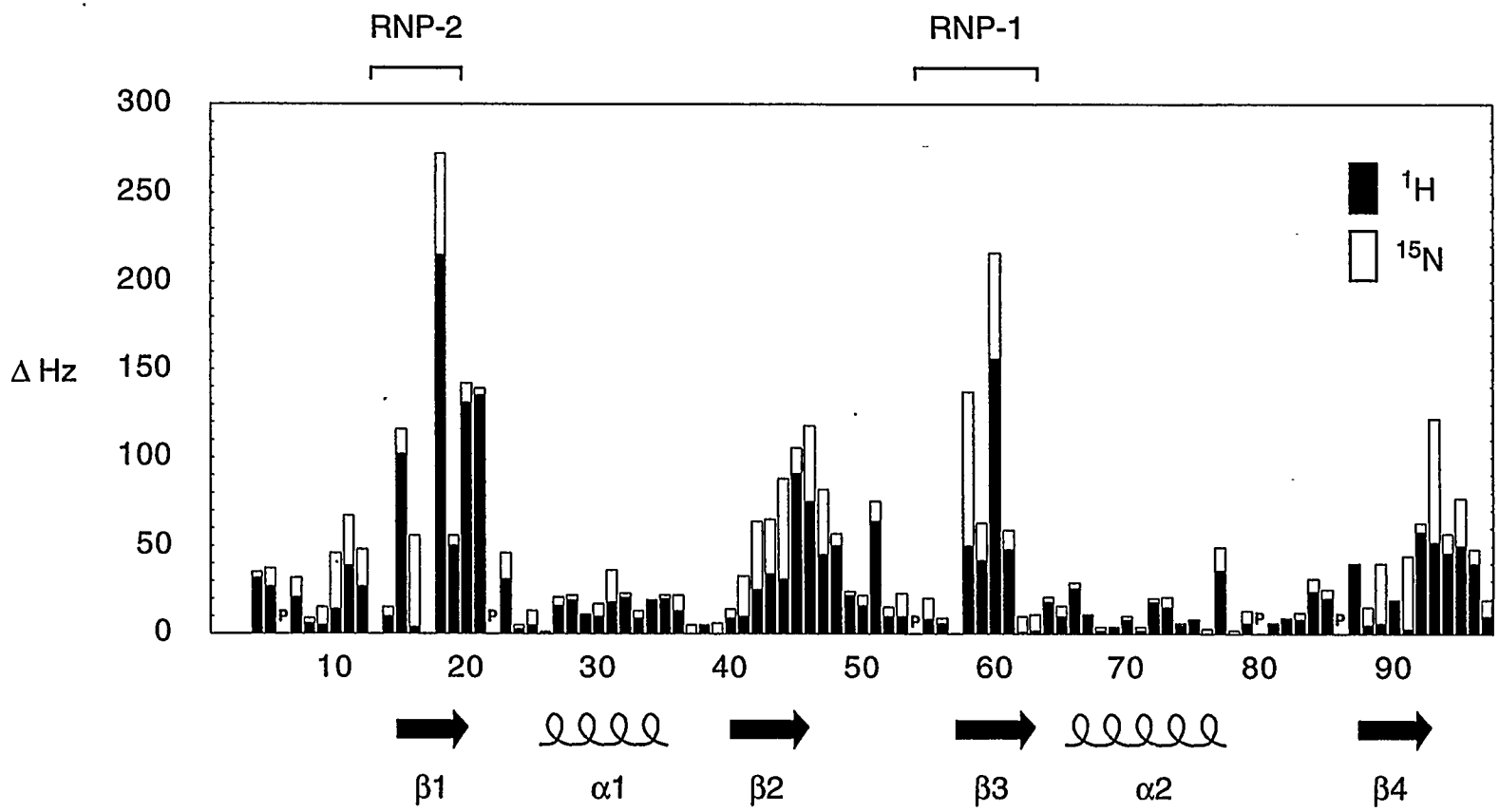


Figure 7.8. Chemical shift changes (in Hz) between RBD2 in the presence and absence of d(5' - TTTTGTTGTTTTTTTTCTAG), plotted as a function of amino acid residue position. Positions labeled with P correspond to proline residues. Residue positions without chemical shift change data correspond to untrackable resonances due to intermediate exchange rates.

perturbations of residues carboxy-terminal to β-strand 4 are consistent with the observed chemical shift changes in the C-terminus of hnRNP-C upon addition of RNA (Görlach et al., 1992). In the case of snRNP-U1A, these carboxy-terminal residues adopt a helical conformation upon binding RNA (Howe et al., 1994; Oubridge et al., 1994). The β2-β3 loop region of Sxl-RBD2 is interesting in that there appear to be significant changes at the transition from the β-strand 2 into this region, but a relatively large number of residues in this region do not have significant chemical shift changes for this complex. Nevertheless, threonine 51 in the β2-β3 loop of Sxl-RBD2 does have a notably large ^{1}H chemical shift change. This observation suggests that the β2-β3 loop could be involved in RNA binding. This mapping of the binding surface of Sxl-RBD2 is consistent with what is known about binding surfaces of other RBDs (Burd & Dreyfuss, 1994; Oubridge et al., 1994). Finally, it should be noted that 2D NOESY spectra of this complex were acquired, but no obvious intermolecular NOEs could be identified.

RBD1/RNA-PPT. Figure 7.9 shows a 1D titration of RBD1 with the Tra-PPT 10-mer, 5'-GUUUUUUUUC-3'. RBD1 was initially at pH=6.0, but because precipitate formed upon addition of RNA, the pH was raised to 7.5. The rest of the titration was carried out at pH=7.5. The titration endpoint was approximately 1:1.15 protein/RNA. It is clear that the protein resonances are shifting as RNA is added, presumably from conformational shifts and/or RNA contacts. It also appears that the protein resonances at the last titration point are broadened relative to the free protein, perhaps due to an increase in the overall tumbling correlation time of the complex which may be 30% greater in molecular weight. The K_d for this complex should be ~5 x 10^{-8} M (Table 7.1). With this value it is difficult to accurately predict what exchange regime this complex should be in, since the on-rate is not known to high accuracy and the K_d for the 10-mer was not specifically measured. Assuming a diffusion-limited on-rate of between 10^8-10^9 $M^{-1}s^{-1}$, this complex would be expected to be near the intermediate exchange regime since the corresponding off-rate would be 5-50 Hz. In actuality, from Figure 7.9 this complex

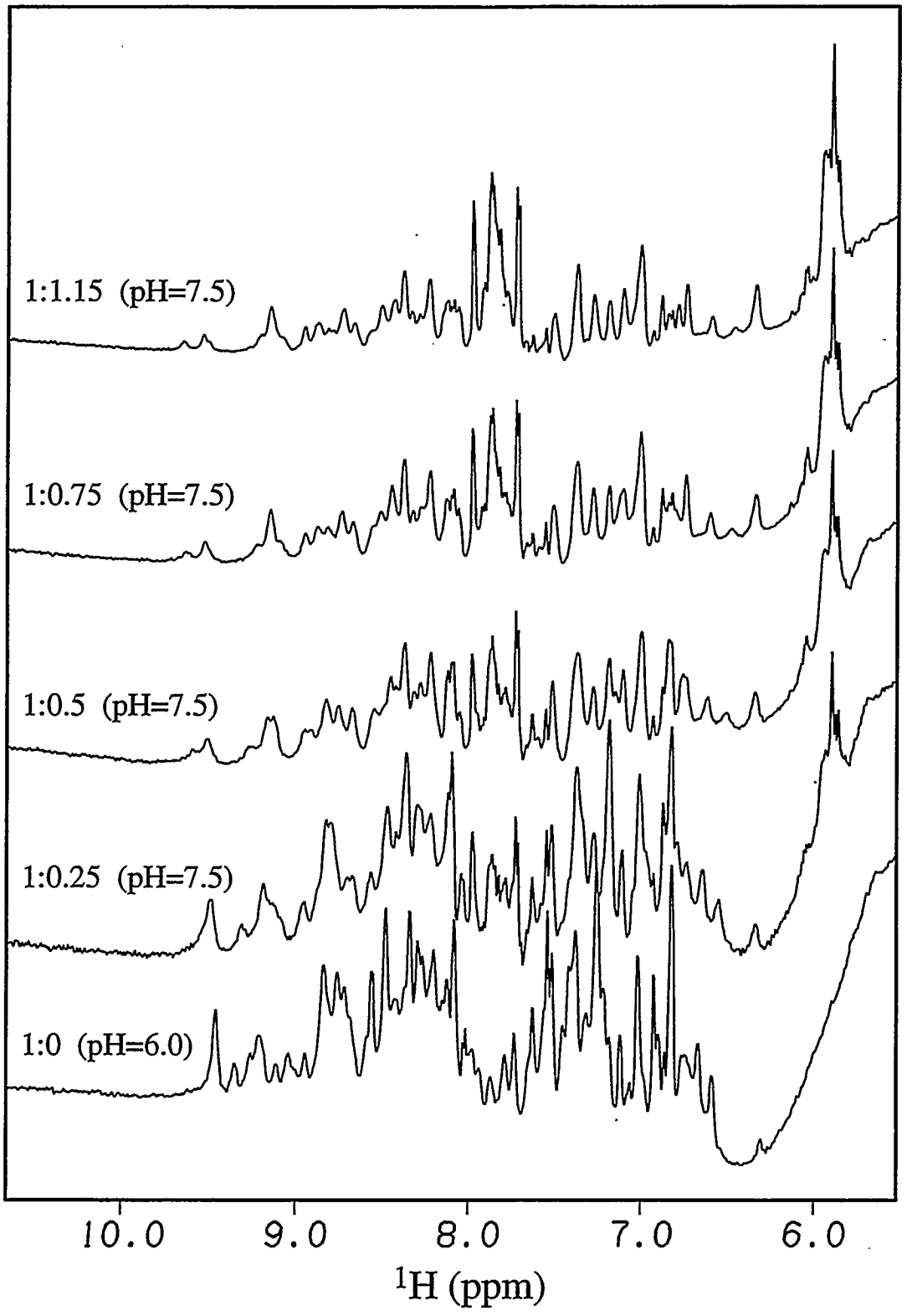


Figure 7.9. 1D titration of RBD1 with 5'-GUUUUUUUUC-3' RNA, pH = 6.0, RBD1 concentration = ~0.8 mM. RBD1/DNA stoichiometries are shown for each titration step. The pH was increased to 7.5 for the second titration point.

displays primarily fast exchange characteristics, although line-broadening from intermediate-type exchange is observed. This complex was not pursued further.

RBD2/RNA-PPT. Figure 7.10 shows a 1D titration of RBD2 with 5'-GUUUU UUUUC-3'. This titration was carried out at pH=6.0, and the titration endpoint was approximately 1:1.1 protein/RNA. From the measured K_d of ~4 x 10^{-6} M (Table 7.1), this complex is expected to be primarily in fast exchange, and indeed this appears to be the case. Protein resonances at the endpoint of the titration are shifted relative to the free protein resonances, and the lines are broadened from an increase in the overall tumbling correlation time. Line-broadening may also result from exchange contributions. Unfortunately, 2D HSQC spectra were not collected for this titration, so more specific information about this complex is not currently available.

RBD1+2/RNA-PPT. I next analyzed the complex between uniformly ^{15}N-labeled Sxl-RBD1+2 and 5'-GUUUUUUUUC-3'. This RNA binds to Sxl-RBD1+2 with high affinity (Table 7.1, Figure 7.11) and therefore this complex is expected to be in the slow exchange regime on the NMR timescale. This was confirmed by titration, shown in Figure 7.11A. Normally unobservable imino proton resonances from the RNA were also observed growing in with slow-exchange behavior. These resonances were not split by a large heteronuclear coupling in spectra collected without ^{15}N-decoupling (data not shown), indicating that they are from imino protons on the RNA. In the fully bound state (1:1 protein to RNA), the number of resonances in the HMQC spectrum was consistent with a single form in the bound state, as can be seen in Figure 7.11B. The relatively favorable chemical shift dispersion in the ^{15}N/^{1}H correlation spectrum of Sxl-RBD1+2 should facilitate further structural characterization of this complex.

Evaluation of the various RBD/PPT complexes

The third issue mentioned at the beginning of this chapter was which specific RBD/PPT complex would be the most favorable candidate for detailed structural NMR

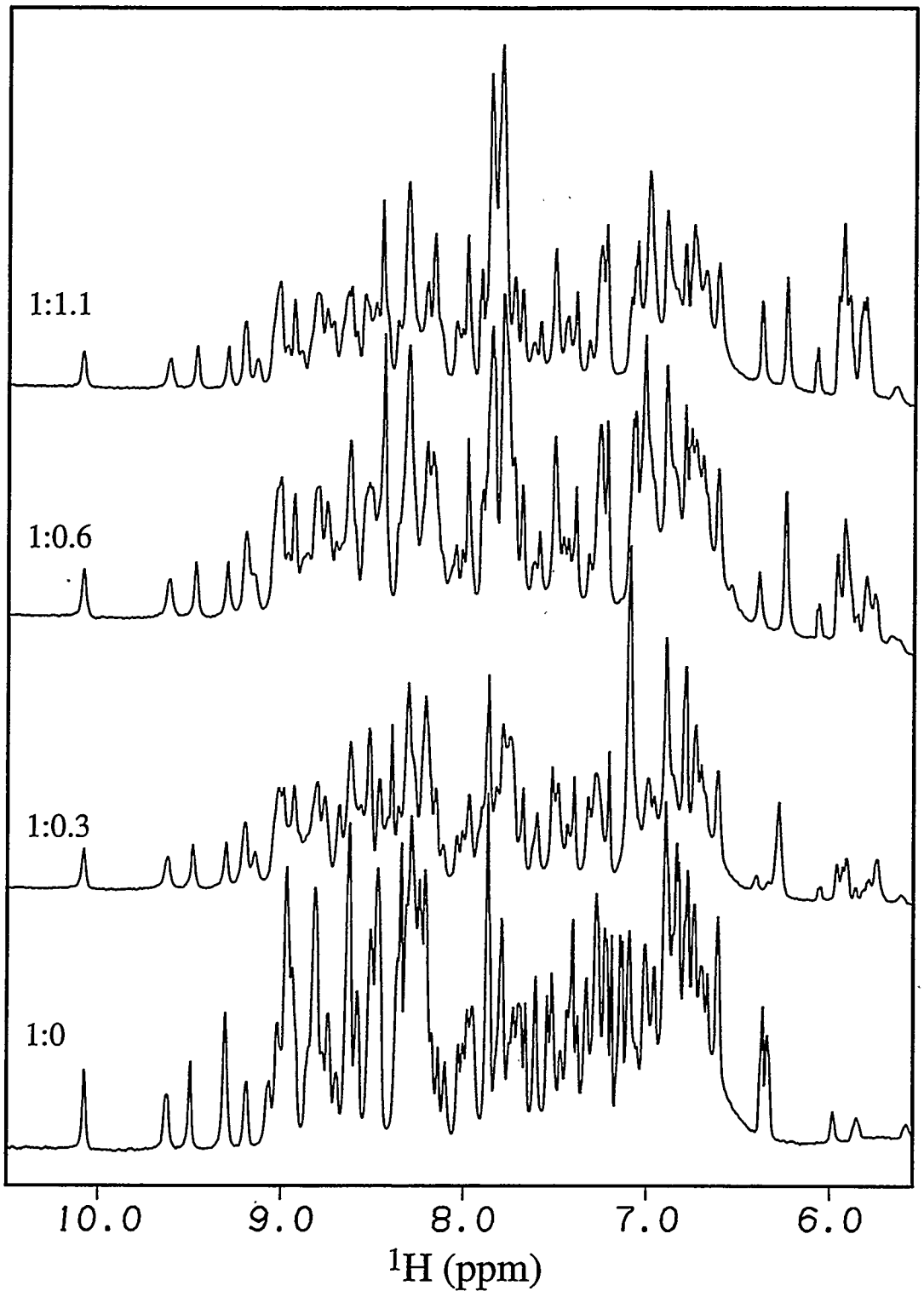


Figure 7.10. 1D titration of RBD2 with 5'-GUUUUUUUUC-3' RNA, pH = 6.0, RBD2 concentration = 1 mM. RBD2/RNA stoichiometries are shown for each titration step.

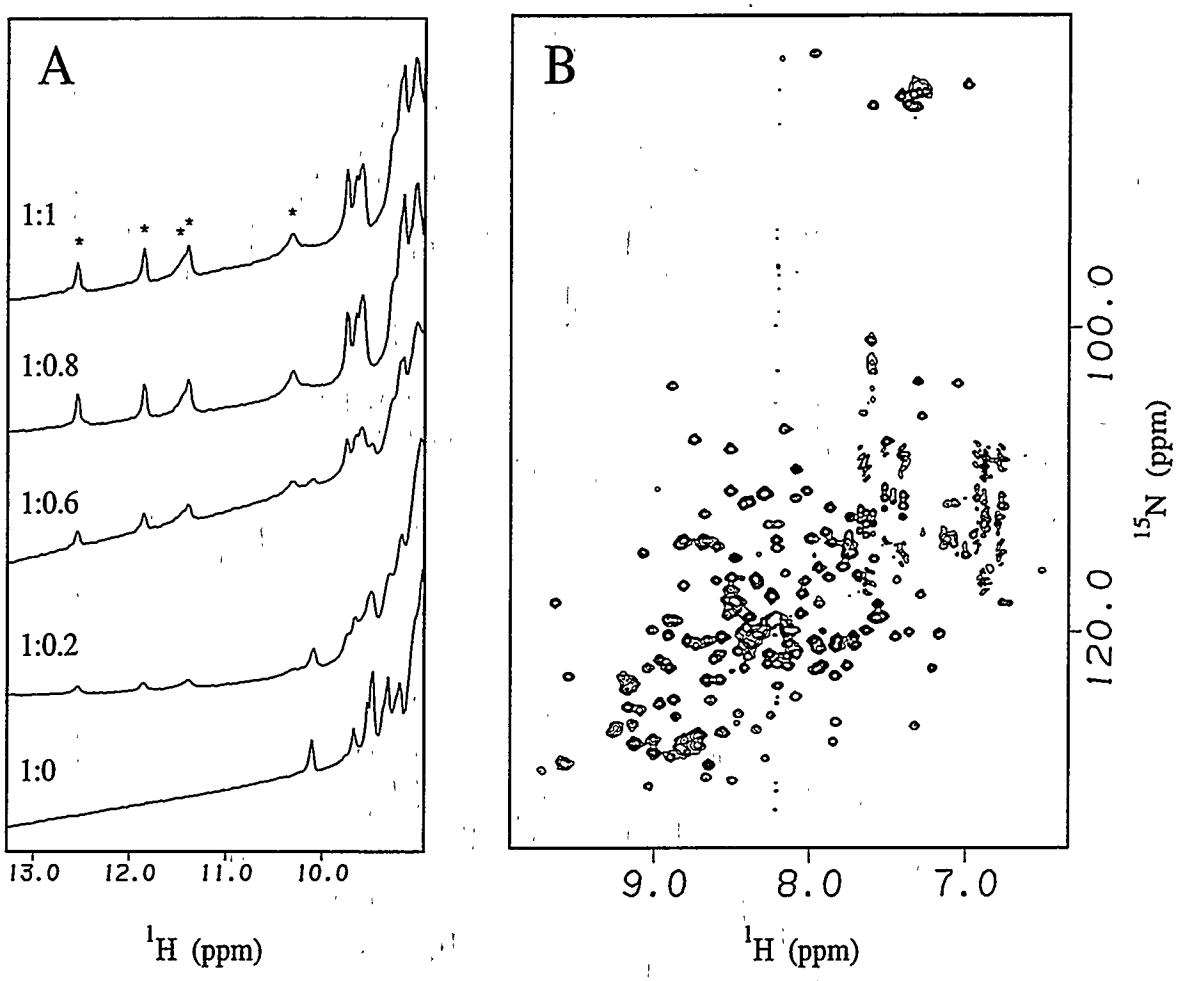


Figure 7.11. In Figure 7.11A, the 1D titration of RBD1+2 by 5'-GUUUU UUUUC-3' is shown. Imino protons are indicated by asterisks, and protein/RNA ratios are indicated for each step. In Figure 7.11B, the 600 MHz ^{15}N/^{1}H HMQC of RBD1+2 complexed with 1 equivalent of 5'-GUUUUUUUUC-3' is shown. Protein concentration is ~0.3 mM. Conditions are pH=6.5, 25 °C.

studies. Four different RBD/PPT complexes have been initially characterized (Figures 7.7-7.11). From these four, only the RBD1+2 / 5'-GUUUUUUUUC-3' complex (Figure 7.11) is a high affinity complex with slow exchange NMR kinetics. Slow exchange kinetics is generally considered to be favorable for the study of intermolecular complexes by NMR. When a 1:1 bimolecular complex is in this regime, the NMR data collected arises only from the bound state, as opposed to the fast exchange regime where the data is a weighted average of all populated states. In addition, this is the only complex where sharp imino protons from the RNA have been observed, which reinforces the idea that this is the only specific complex as originally determined from Roland's bandshift data (Table 7.1). For these reasons, I have chosen to focus on the RBD1+2 / 5'-GUUUUUUUUC-3' complex as the subject for a series of heteronuclear multidimensional NMR structural studies. This will be presented in the next chapter.

Brown, J. E., Klement, J. F. & McAllister, W. T. (1986) *Nucl. Acids Res. 14*, 3521-3526.

Burd, C. G. & Dreyfuss, G. (1994) *Science 265*, 615-621.

Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L. & Dreyfuss, G. (1992) *Embo Journal 11*, 3289-3295.

Grodberg, J. & Dunn, J. J. (1988) *J. Bacteriol. 170*, 1245-1253.

Howe, P. W. A., Nagai, K., Neuhaus, D. & Varani, G. (1994) *The EMBO Journal 13*, 3873-3881.

Inoue, K., Hoshijima, K., Sakamoto, H. & Shimura, Y. (1990) *Nature 1990*, 461-463.

Kanaar, R., Lee, A. L., Rudner, D. Z., Wemmer, D. E. & Rio, D. C. (1995) *The EMBO Journal 14*, 4530-4539.

Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M. & Pardi, A. (1992) *Nucl. Acids Res. 20*, 4507-4513.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994) *Nature 372*, 432-438.

Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. (1995) *J. Mol. Biol. 249*, 398-408.

Sakashita, E. & Sakamoto, H. (1994) *Nucl. Acids Res. 22*, 4082-4986.

Singh, R., Valcarcel, J. & Green, M. R. (1995) *Science 268*, 1173-1176.

Zawadzki, V. & Gross, H. J. (1991) *Nucl. Acids Res. 19*, 1948.

Chapter 8

Assignments and Secondary Structure of Sxl-RBD1+2 Complexed With the Tra-PPT 10-mer

Introduction

Having characterized the exchange behavior of a number of RBD/PPT complexes, the 24 kD complex of RBD1+2 with the Tra-PPT 10-mer (5'-GUUUUUUUUC-3'), stood out as the most promising candidate for NMR structural studies (Figure 7.11). There was concern, however, over the broad lines and relatively low solubility of this complex. The HMQC in Figure 7.11B was taken at only 0.3 mM, for example. When more concentrated samples were made, significant line broadening was observed. I eventually settled at approximately 0.7 mM as the optimal concentration for the complex, as the line broadening is not too severe at this concentration. Still, linewidths were a major concern since triple resonance experiments would need to be collected in order to assign the protein backbone. It was not clear if this class of experiments would work at all for this complex. The triple resonance experiments (described below) generally have long magnetization transfer steps, and with broad lines (short T_2) there simply may not be any net magnetization remaining after these transfer steps are completed.

Dr. Ad Bax (National Institutes of Health) has a convenient "rule of thumb" for determining whether a protein is a good candidate for structural determination by heteronuclear multidimensional NMR: if the protein's amide ^1H T_2 values are longer than 15 ms, then the protein is a good candidate. This can be checked quickly with a simple spin-echo 1D experiment with "jump and return" water suppression (to remove scalar effects from the 3-bond ^1HN-$^1H_\alpha$ coupling constant). The spin-echo delay is adjusted to a short (200 us) and a long (10 ms) value, and the envelope of the ^1HN spectrum is

compared at both values to get an estimate of T_2. The complex failed Ad Bax's test—the ^1H T_2 values appeared to be mostly in the 10-15 ms range. Nevertheless, with Dave's prodding, I made the ^{13}C-, ^{15}N-labeled sample of RBD1+2. I was not completely confident that the sample would be sufficiently RNase-free, and it was also not clear whether any triple resonance experiments would work on this complex. I would like to thank Dave for his encouragement at this stage.

Around the same time (fall of '95), Dr. Brian Volkman, a recent Wemmer group graduate and presently a postdoc with Professor John Markley at the Nuclear Magnetic Resonance Facility at Madison (NMRFAM), offered to look at the RBD1+2/Tra-PPT complex at a higher field strength than is available at Berkeley. At Madison is a relatively new 750 MHz Bruker DMX NMR spectrometer which has superior sensitivity due to the field strength as well as its hardware components. It has triple resonance capability and is fully equipped with 3-axis pulsed field gradients (PFGs). The facility also has a DMX 600 MHz and two DMX 500 MHz Bruker spectrometers, all with triple resonance capability and most with 3-axis pulsed field gradients. As sensitivity was clearly going to be an issue, Brian's generous offer to try a few tests on the complex sample (with ^{15}N RBD1+2) at 750 MHz was an exciting prospect. This was the beginning of a fruitful collaboration which culminated in the resonance assignments and ^{15}N relaxation results presented in this chapter. I don't think that these results could have been obtained as rapidly at any other accessible facility. The superior equipment performance was matched by Brian's skill and determination in implementing the experiment versions that we eventually settled on.

The doubly labeled complex sample (^{13}C, ^{15}N RBD1+2) was made (Figure 7.6) under time contraints because I had signed up for spectrometer time at Madison in early December (1995). This is the reason why the RNase assay could only be run for 4 days (Figure 7.6). The level of RNase was such that we could run an experiment at 20 °C for ~40 hours, accompanied by ~20-30% degradation of the RNA. Going to lower temperature was not a good option because significant line-broadening was observed below

20 °C. The rate of degradation at 20 °C, although disappointing, was acceptable because at this stoichiometry free RBD1+2 resonances are absolutely not visible (data not shown). This is most likely due to the existence of multiple "free" protein states at sub-stoichiometric protein/RNA ratios, although this has not been well characterized. Before starting each 3D experiment on the doubly labeled complex, Brian would add enough RNA to bring the complex back to 1:1 stoichiometry. This diluted the sample for each successive experiment by approximately 6-10%. We began this series of triple resonance 3D experiments with the sample at 0.7 mM.

Brian and I worked in a complementary fashion. He physically did all of the NMR, but since we had a challenging sensitivity problem (not to mention Brian's other projects and responsibilities), it helped having two of us to think about potential solutions. For example, if we wanted to try out a new idea or NMR pulse sequence, I would write out the pulse program and then send it to Brian, who would make it work on the spectrometer. This was a very efficient process. I didn't have to spend long hours on the spectrometer, and therefore I had time to read the pulse sequence literature, evaluate our possibilities, and consult with Brian so that decisions could be made about how a particular NMR experiment should be implemented. We iterated over a number of experiments and techniques, and in the end we had a nice set of double and triple resonance NMR experiments optimized for a large protein at intermediate pH (6.3) with broad lines (^{1}HN linewidths up to 50 Hz; average ^{1}HN linewidth is probably ~25-35 Hz). Since these experiments yielded reasonably complete data sets (Table 8.1), we were able to obtain nearly complete protein backbone resonance assignments for the complex.

NMR experiments and pulse sequence building blocks

One trick we picked up early on with the ^{15}N RBD1+2 complex was keeping the large water solvent signal along +Z for the recycle delay (Grzesiek & Bax, 1993). Because water protons can exchange or cross-relax with protein amide protons, saturating or

dephasing the water signal can result in partially saturated amide protons at the beginning of each scan. This is due to the long water T_1 compared to the T_1 for an amide proton. The gains from this "water flip-back" trick become smaller as the pH and intrinsic amide exchange rates are reduced. We first implemented this technique into simple $^{15}N/^{1}H$ HSQCs (Zhang et al., 1994) using sensitivity enhancement (recovery of the two orthogonal t_1 components) with gradients for coherence selection (Palmer et al., 1991; Kay et al., 1992a). In fact, this implementation was identical for all the experiments we were to implement which start and end on amide protons. A 2 ms water-selective pulse is normally applied after the first INEPT from ^{1}H to ^{15}N (Figure 8.3). This returns water to +Z (in some cases -Z), and the remaining non-selective ^{1}H pulses are designed to have phases which will return water to +Z prior to acquisition. Complications from radiation damping are removed by temporarily dephasing the water with gradients when water magnetization is in the transverse plane. Therefore, the beauty of this method is that one always knows exactly where all of the water magnetization is. The gains obtainable from using the "water flip-back" technique are illustrated in Figure 8.1 for flavodoxin at pH=6.5. Brian also started using "magic angle" gradients for the coherence selection gradients, resulting in improved water suppression over normal z-gradients. $^{15}N/^{1}H$ HSQCs collected at 750 MHz using these methods are presented in Figure 8.2 for free RBD1+2 (A) and RBD1+2 bound to the Tra-PPT 10-mer (B). The overall improvement in both sensitivity and resolution relative to non-gradient HSQCs collected at 600 MHz was very encouraging.

For the triple resonance experiments, we also added ^{1}H decoupling during transfer steps involving only the heteronuclei (Grzesiek & Bax, 1992). Turning on ^{1}H decoupling at the point when ^{15}N magnetization is no longer antiphase with respect to ^{1}H prevents ^{1}H coherences from reappearing and hence adding to relaxation rates of transverse heteronuclear coherences (Bax et al., 1990). The phases of the decoupling pulses are designed so that the net water magnetization becomes spin-locked, not dephased or partially saturated. ^{1}H decoupling is then turned off when it is time to generate antiphase ^{15}N

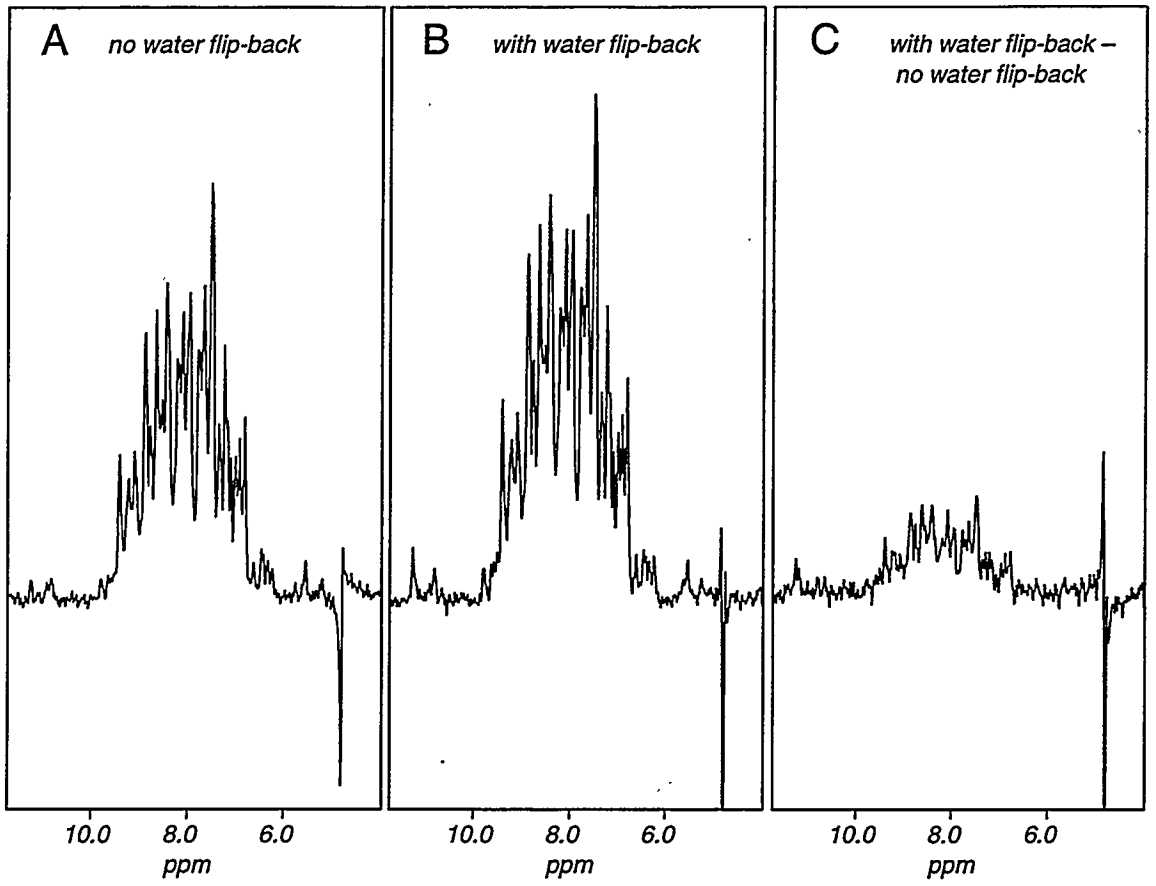


Figure 8.1. Demonstration of enhancement realized from minimal saturation of water (returning water to +Z for the recycle delay). Panel A shows the first t_1 point from an HNCO of flavodoxin (19 kD, pH=6.5) acquired with the pulse sequence in Figure 3 minus the selective water flip-back pulse. Panel B shows the corresponding "first block" obtained after inclusion of the selective water flip-back pulse immediately following the second ^{1}H 90° pulse. Panel C illustrates the net gain in signal by subtraction of Panel A from Panel B. A gain of ~20% is observed on average, although some signals have gains of up to 100% or more.

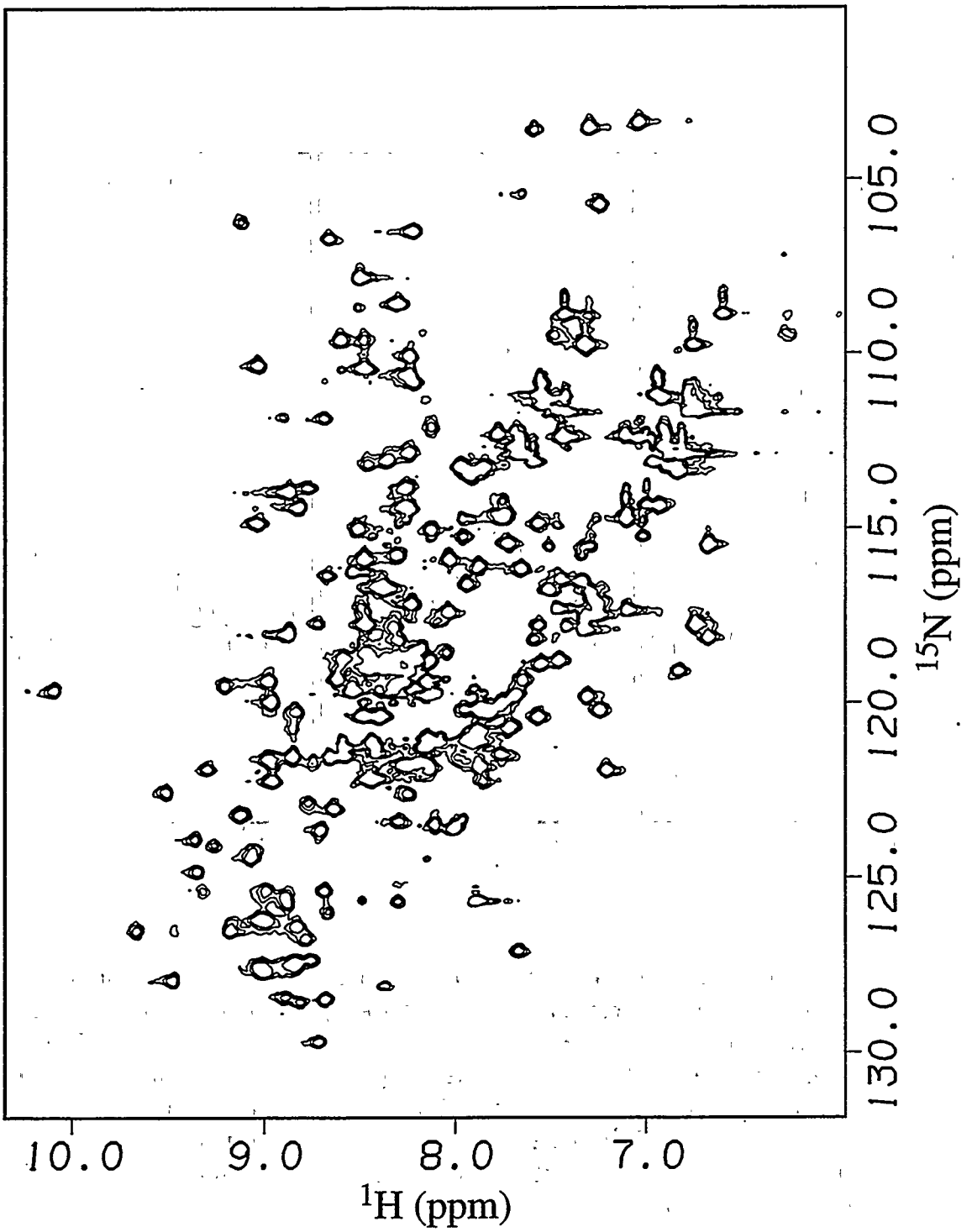


Figure 8.2A. ^{15}N/^{1}H HSQC (water "flip-back", sensitivity enhanced with gradient coherence selection) of RBD1+2 collected in ~2 hours at 750 MHz, 20 °C.

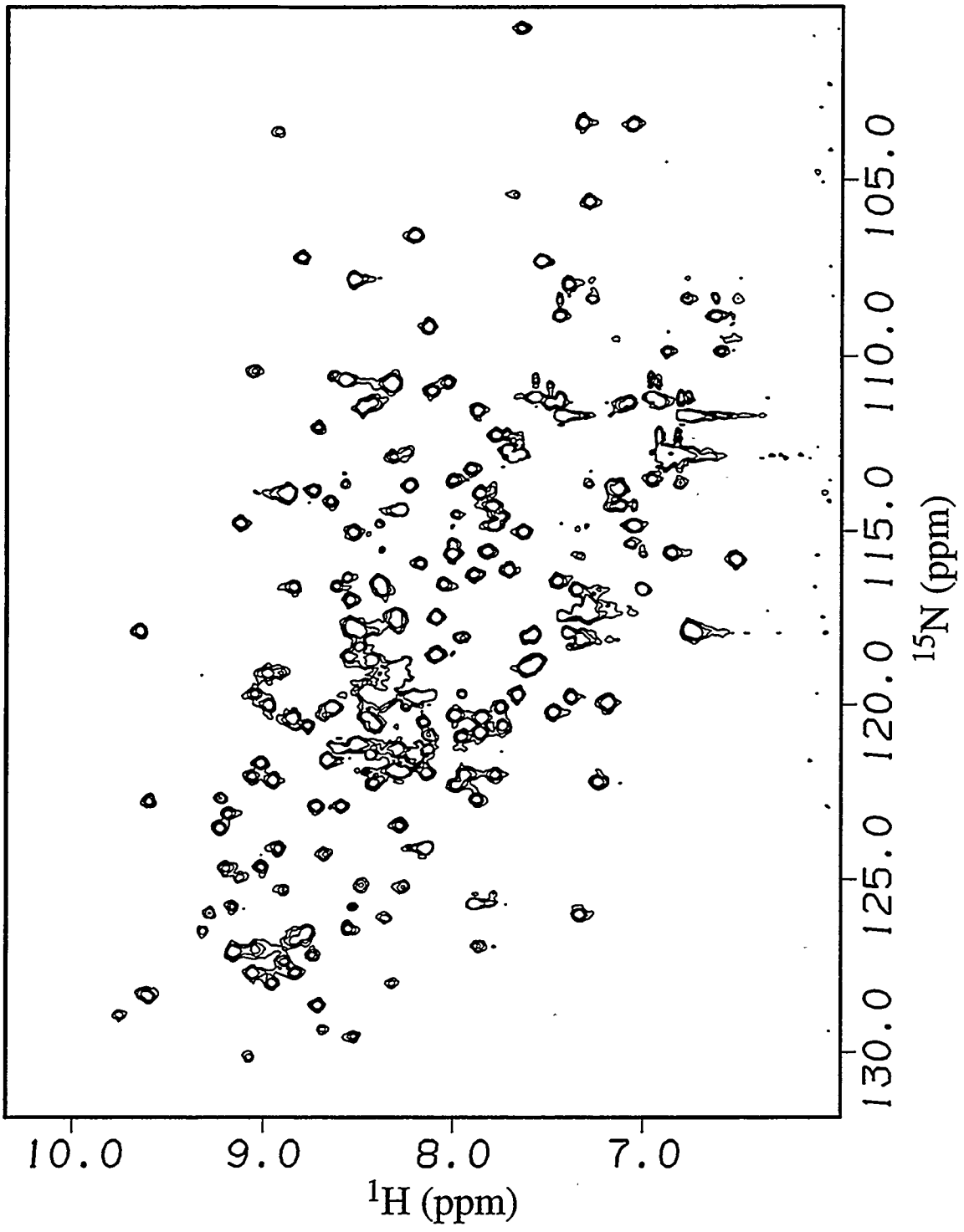


Figure 8.2B. ^{15}N/^{1}H HSQC (water "flip-back", sensitivity enhanced with gradient coherence selection) of RBD1+2 bound to 5'-GUUUUUUUUC-3' collected in ~2 hours at 750 MHz, 20 °C.

magnitization with respect to ^1H, so that coherences can be transferred back to ^1H for detection. According to Grzesiek and Bax, (Grzesiek & Bax, 1992), this decoupling should result in an enhancement of approximately 50% for a protein with 25 Hz ^1HN linewidths.

All of these "bells and whistles", sensitivity enhancement (SE), "water flip-back" pulses, ^1H decoupling, and magic angle gradients, can be viewed in the Figure 8.3 pulse sequences which we implemented in Dec('95)/Jan('96). This particular version of the HNCO-SE (Figure 8.3A) is a published pulse sequence (Kay et al., 1994), and it provided the starting point for the HNCA-SE (Figure 8.3B) and HN(CO)CA-SE (Figure 8.3C) which to the best of my knowledge are not in the literature with these particular enhancement features. The HNCA-SE, however, is very similar to the HNCACB-SE (Muhandiram & Kay, 1994). For convenience, the "SE" will be dropped from the experiment names from here on.

Originally, we had hoped to have success with CBCA(CO)NH and HNCACB. Apparently, the lines in our complex are too broad for these experiments with their many transfer steps. The HNCACB was < 10% complete while the CBCA(CO)NH was approximately 15-20% complete. Both experiements were attempted at 600 MHz. The latter experiment was used in the analysis and proved to be quite useful for assigning the unstructured or highly mobile terminal residues which had sharp lines. We quickly realized that only the most sensitive triple resonance experiments with the fewest and most efficient transfer steps were going to work on this complex, and even some of these did not work as well as we had hoped. HNCO, HNCA, HN(CO)CA, and HCACO were the only experiments which yielded more than 50% of the possible correlation peaks. With these experiments, our assignment strategy would be very close to the original strategy used by Ikura et al. (Ikura et al., 1990) for the first protein assigned with triple resonance experiments. Table 8.1 reports the "completeness" for each experiment attempted on the RBD1+2//Tra-PPT 10-mer complex.

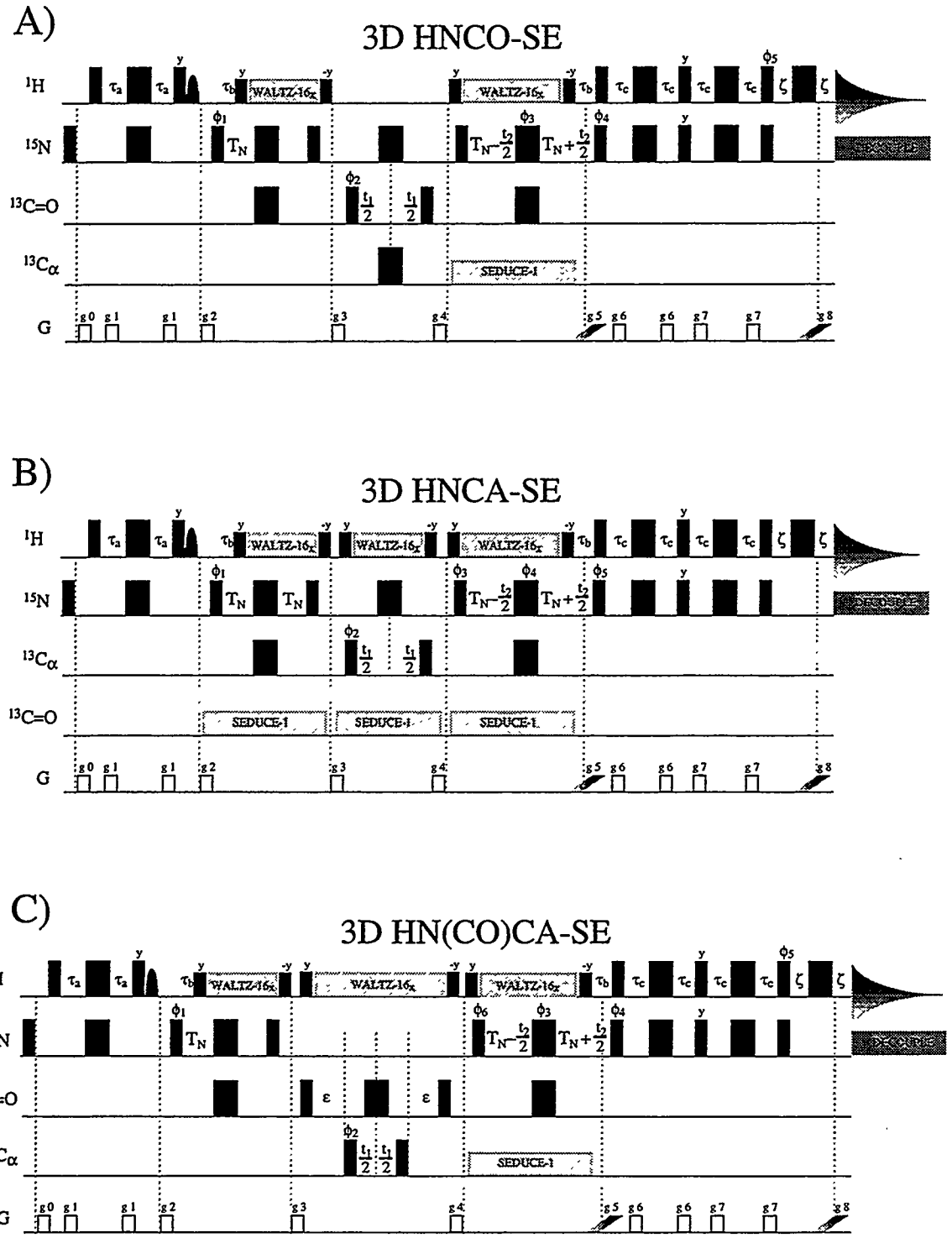


Figure 8.3. Triple resonance pulse sequences used for the assignment of RBD1+2 complexed with the Tra-PPT 10-mer. All three pulse sequences employ elements which increase overall sensitivity as discussed in the text, namely sensitivity enhancement, water flip-back pulses, and proton decoupling during remote transfer steps. A) HNCO-SE B) HNCA-SE C) HN(CO)CA-SE.

Table 8.1 - Spectral Parameters for NMR datasets collected on Sex-lethal/Tra-PPT (10-mer) complex

	^1H[a]		d2			d3			matrix size[b]	
Experiment	SF (MHz)	SW(Hz)	nucleus	SW(Hz)	N*	nucleus	SW(Hz)	N*	(d1 × d2 × d3)	completeness[c]
HNCA	750.13	11062.00	$^{13}C_\alpha$	5000.00	48	^{15}N	2500.00	40	512 × 128 × 128	85%
HN(CO)CA	600.13	8333.32	$^{13}C_\alpha$	4000.00	48	^{15}N	2000.00	44	512 × 128 × 128	80%
CBCA(CO)NH	600.13	8333.34	$^{13}C_{\alpha/\beta}$	8333.33	40	^{15}N	2000.00	32	512 × 128 × 128	15-20%
HNCO	600.13	8333.32	$^{13}C'$	2000.00	60	^{15}N	2000.00	32	512 × 128 × 128	95%
HCACO	600.13	6250.00	$^{13}C'$	2000.00	72	$^{13}C_\alpha$	4000.00	28	512 × 256 × 128	65-70%
^{15}N-TOCSY	600.13	8333.34	^1H	5000.00	80	^{15}N	2000.00	36	512 × 256 × 128	45-50%[d]
^{15}N-TOCSY	750.13	10000.00	^1H	8333.33	96	^{15}N	2272.73	35	512 × 256 × 128	45-50%[d]
^{15}N-NOESY	750.13	8333.34	^1H	8333.33	120	^{15}N	2272.73	35	512 × 256 × 128	100%
2D ^{15}N-^1H HSQC	750.13	10000.00	^{15}N	2500.00	200				512 × 512	100%
^{15}N T_2 relaxation	750.13	10000.00	^{15}N	2500.00	150				512 × 512	

[a] A total of 1024 complex points were collected in the ^1H dimension (d1) of all experiments. In all experiments except the HCACO, the right half of the spectrum (upfield of H_2O) in the ^1H dimension contained no signals and was discarded.

[b] Data were zero-filled to final sizes indicated. Linear prediction was used to extend data by 50% in the d3 dimension only.

[c] Completeness indicates the percentage of residues for which expected signals in triple-resonance experiments were observed.

[d] Indicates completeness for combined total of both 3D ^{15}N-separated TOCSY experiments.

Materials and methods

All NMR experiments were performed at 600.13 MHz or 750.13 MHz on Bruker DMX spectrometers at 20° C. Both spectrometers were equipped with Bruker ^{1}H, ^{13}C, ^{15}N triple-resonance probeheads with three-axis self-shielded gradient coils. Spectra were acquired with Xwin-NMR software, v1.1 (Bruker). All spectra were collected in 90% H_2O/10% D_2O solution, pH 6.3, 30 mM sodium phosphate, 30 mM NaCl, and 0.02% NaN_3. Protein concentrations ranged from 0.4 - 0.7 mM, depending on the order in which the experiments were collected. This was a result of having to add ~30-50 ul (<15 ul for the ^{15}N RBD1+2 sample, see Chapter 7) of 5 mM RNA at the start of each 3D experiment to bring the complex stoichiometry up to 1:1. The complex was initally made by titration of 5 mM RNA directly into a solution of 0.7-0.9 mM RBD1+2. Mixing of protein and RNA under more dilute conditions (~20 uM) followed by concentration did not reduce linewidths or improve the quality of the spectra of the complex.

A number of three-dimensional triple-resonance and ^{15}N-edited spectra were aquired on the RBD1+2/Tra-PPT 10-mer (5'-GUUUUUUUUC-3') complex. Table 8.1 contains a list of experiments and spectral parameters which were used for the assignment of backbone protein resonances of the complex. All pulse schemes utilized the signal enhancing elements (when applicable) described in the previous section. The CBCA(CO)NH experiment was implemented as previously described (Muhandiram & Kay, 1994), with the subsititution of WALTZ-16 . The HCACO experiment was modified from previous descriptions (Powers et al., 1991) to include gradient coherence selection and sensitivity enhancement, as this provides superior solvent suppression. These modifications were implemented by Dr. Frits Abildgaard at NMRFAM. Three-dimensional ^{15}N-separated NOESY (75 ms NOE mixing time) and TOCSY experiments (33 ms mixing time at 600 MHz, 41.5 ms mixing time at 750 MHz), as well as 2D ^{15}N-^{1}H HSQC experiments were also acquired as described previously (Zhang et al., 1994).

Gradient coherence selection was achieved as previously described, but with the substitution of magic angle gradients for the Z-gradients used for selection of coherence (g5 and g8 in Figure 8.3). This was accomplished by applying both X- and Z- gradients simultaneously at a ratio which is optimized to result in minimal residual water. The ratio of X:Z was found to be 1.8:1 on both the DMX-600 and DMX-750.

Chemical shift values were externally referenced to DSS (Jeener, 1971). All data were processed with Felix version 95.0β (Biosym), including linear prediction calculations. All data analysis was carried out with Felix 2.30. In the acquired ^{1}H dimension (d1) a shifted sine-bell function was used for apodization of the free induction decays, followed by left circular shifting of the time-domain data by 61 or 70 points to remove phasing artifacts of proprietary oversampling technology. Quadrature was achieved in the d2 dimension of all experiments by the States-TPPI method (Marion et al., 1989), while quadrature in d3 was obtained by deconvolution of echo and anti-echo signals resulting from the gradient coherence selection scheme (Kay et al., 1992a).

Assignment strategy

The assignment strategy employed was centered around the sequence-specific identification of backbone amide $^{1}H/^{15}N$ pair resonances. This could be done when enough sequence-adjacent $^{1}H/^{15}N$ pairs were linked so that $^{13}C_{\alpha}$ chemical shifts, $^{13}C_{\beta}$ chemical shifts [when present in CBCA(CO)NH], and NOE patterns would unambiguously place the stretch into the amino acid sequence. Links between residues were accomplished using 3 possible methods:

1) When signals from all basic triple resonance experiments [HNCO, HNCA, HN(CO)CA, HCACO] and a $^{1}HN/^{1}H_{\alpha}$ correlation from either ^{15}N-separated TOCSY-HSQC or ^{15}N-separated NOESY-HSQC were present, the HCACO experiment could be used as a "connector" between sequential $^{1}H/^{15}N$ pairs. This was the most desirable

method for linking sequence-adjacent $^1H/^{15}N$ pairs because all correlations are extended by the matching of *two* independent chemical shifts (Ikura et al., 1990). Additionally, this method makes connectivities between sequence-adjacent $^1H/^{15}N$ pairs by using "through-bond" correlations exclusively. This method for using the HCACO as a "pivot" between $^1H/^{15}N$ pairs is represented schematically in Figure 8.4.

2) Candidates for sequential $^1H/^{15}N$ pairs were generated by matching $^{13}C_\alpha$ chemical shifts from HNCA and HN(CO)CA, and these candidates were evaluated by analysis of NOE patterns from ^{15}N-separated NOESY-HSQC (Chapter 4). This approach was useful when corresponding peaks in the HCACO and/or ^{15}N-separated TOCSY-HSQC were either not present or ambiguous due to chemical shift degeneracy. It helped that the HNCA and HN(CO)CA experiments were reasonably complete (Table 8.1).

3) When the triple resonance data was too incomplete to use, connectivities could in some cases be made through NOESY/TOCSY analysis (Chapter 4) or from NOEs exclusively. This worked reasonably well in regions of the sequence where sequential HN-HN NOEs were present, since connectivities could be confirmed by the "back crosspeak".

Methods 1 and 2 were carried out with the aid from some simple perl (Wall & Schwartz, 1991) scripts I wrote to do the repetitive matchings of chemical shifts. These scripts are presented in Appendix F, and they will be briefly described now.

The scripts used in method 1 (Figure 8.4) work only when all of the triple resonance data (and ^{15}N-separated TOCSY) is complete for a pair of adjacent residues. All scripts which will be mentioned here rely on the existence of peak pick list (in ppm) text files from individual 3D experiments. The perl script "hnhacag" correlates $^1H_\alpha(i)$ and $^{13}C_\alpha(i)$ with the amide $^1H/^{15}N$ pair of residue i using the ^{15}N-separated TOCSY and HNCA, respectively, for these intraresidue correlations. Similarly, "hncocag" correlates

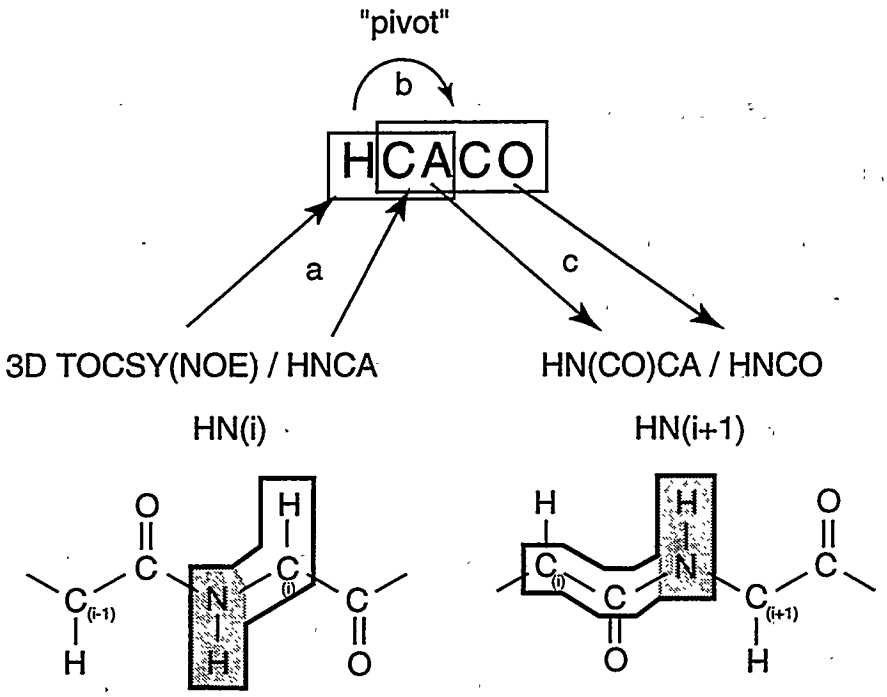


Figure 8.4. Schematic representation of the assignment strategy employed (see text). The arrows represent assigning in a forward direction through the sequence. HN(i) correlations to intraresidue $^{1}H_{\alpha}(i)$ and $^{13}C_{\alpha}(i)$ were aided by the perl script "hnhacag", and this step is denoted along the arrows as "a". The HCACO experiment is used to "pivot" between [$^{1}H_{\alpha}(i)$ and $^{13}C_{\alpha}(i)$] and [$^{13}C_{\alpha}(i)$ and $^{13}C'(i)$]. Correlating $^{13}C_{\alpha}(i)$ and $^{13}C'(i)$ to HN(i+1) was aided by the perl script "cocahng", and this step is denoted along the arrows as "c". Assigning in the reverse direction (i.e. reversing the arrows) was aided by the scripts "hncocag" and "hacahng", respectively.

$^{13}C'(i-1)$ and $^{13}C_\alpha(i-1)$ of the previous residue with the amide $^1H/^{15}N$ pair of residue i using the HNCO and HN(CO)CA, respectively. Using the HCACO experiment, $^1H_\alpha$, $^{13}C_\alpha$, and $^{13}C'$ can be correlated with the "hcacog" or "cocahg" scripts, depending on the assignment direction through the sequence (see below). Finally, in order to make the vital "link" to the sequential $^1H/^{15}N$ pair, two different scripts could be used depending on whether I was moving forward or backward through the sequence. For the forward case (using "hnhacag" initially and "hcacog" to correlate with $^{13}C'$), "cocahng" correlates $^{13}C'$ and $^{13}C_\alpha$ of residue i with the $^1H/^{15}N$ pair of residue i+1 via HNCO and HN(CO)CA. For the backwards case (using "hncocag" initially and "cocahg" to correlate with $^1H_\alpha(i-1)$), "hacahng" correlates $^1H_\alpha(i-1)$ and $^{13}C_\alpha(i-1)$ with the intraresidue $^1H/^{15}N$ (i-1) pair.

If a peak was missing in any of the 5 experiments (often TOCSY or HCACO), the above approach could not be used, since matching a single frequency instead of two frequencies simultaneously rarely leads to unambiguous sequential links. In this event, candidates for the sequential $^1H/^{15}N$ pair could be generated by matching $^{13}C_\alpha$ chemical shifts between HNCA and HN(CO)CA peaks. This is carried out using the perl script, "ca_stripsearch", a script I found to be extremely useful. The candidate $^1H/^{15}N$ pairs could then be screened by analyzing HN "strips" from the 3D ^{15}N-separated NOESY. This was accomplished by creating a "scratch" or temporary strip matrix using the Felix macro 67_tempstrip.mac, derived from Fred Damberger's original Felix "strip" macros. The strip (synonymous with $^1H/^{15}N$ pair) candidates were simply typed into a file and read into the Felix database. Then 67_tempstrip.mac would take care of the rest, yielding a strip matrix of only the candidates which are to be evaluated.

It should be noted that all connectivites made, no matter what method used, were further strengthened by observation of sequential NOE patterns (Chapter 4). Figure 8.5 summarizes all of the triple resonance and NOE sequential connectivities used in the assignment process. Approximately 85-90% of the backbone was assigned using these methods. These assignments are given in Table 8.2.

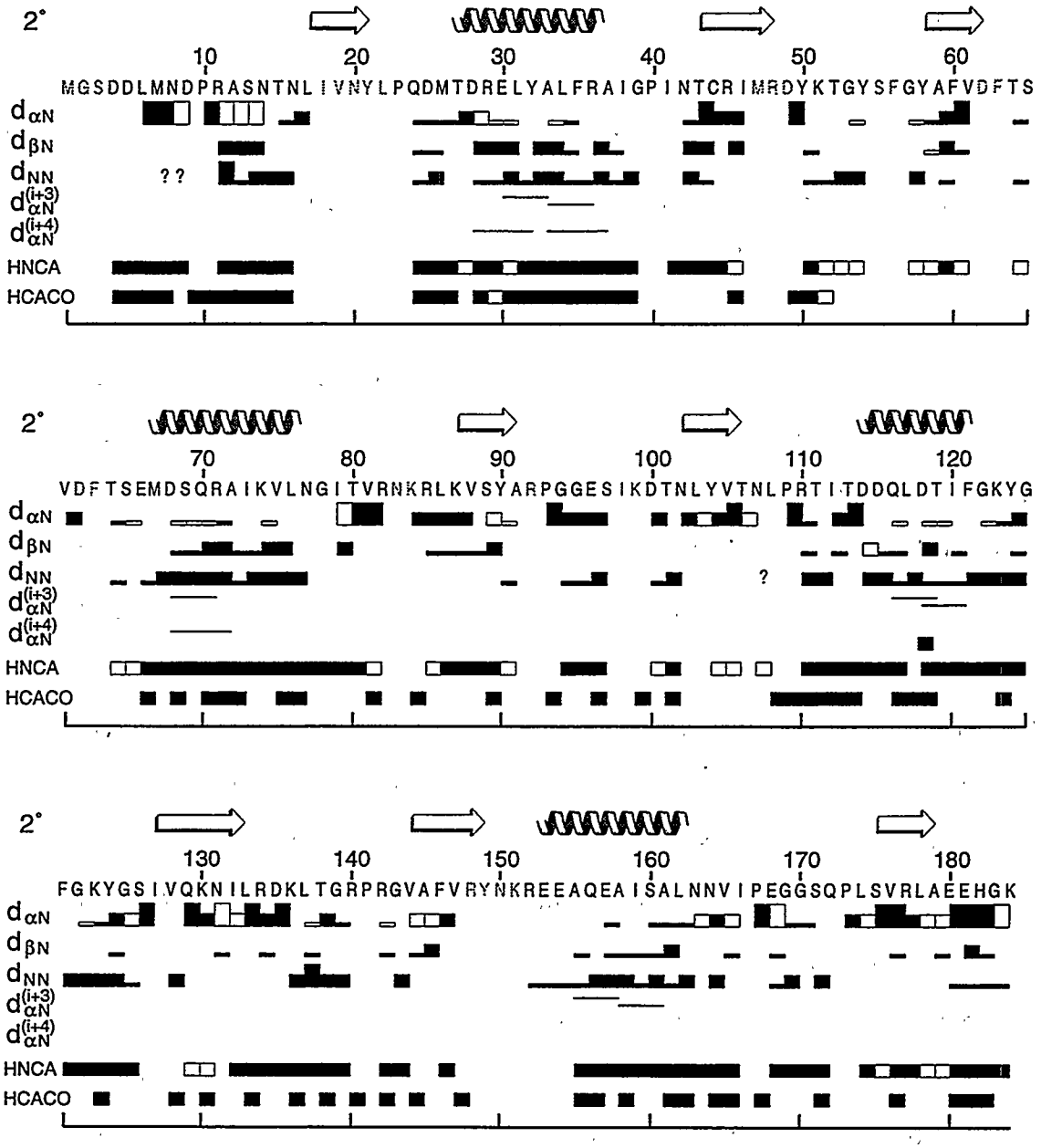


Figure 8.5. Summary of the triple resonance and NOE connectivities used to assign the backbone of bound RBD1+2. Gray letters indicate residues which were not assigned. "HNCA" is shorthand notation for sequential links made by correlating $^{13}C_\alpha$ chemical shifts from HNCA and HN(CO)CA. "HCACO" is shorthand notation for sequential links made from the full suite of triple resonance experiments (see text and Figure 8.4), where $^{13}C_\alpha$ and $^{13}C'$ chemical shifts were correlated simultaneously. NOE intensities are denoted by bar height as weak, medium, or strong. "$d_{\beta N}$" represents sequential sidechain-NH NOEs which are likely to be from beta protons. Open boxes indicate ambiguity due to chemical shift degeneracy in the case of NOEs, very weak peaks in the case HNCA/HN(CO)CA correlations, and both degeneracy and very weak peaks in the case of HCACO based correlations. Question marks denotes degeneracy in ^{1}HN chemical shifts.

Table 8.2.[a] Backbone Resonance Assignments of RBD1+2 When Complexed With the Tra-PPT 10-mer (5'-GUUUUUUUUC-3') at pH=6.3, 20 °C.

Res #	Type	^{1}HN	^{15}N	$^{13}C_{\alpha}$	$^{1}H_{\alpha}$	^{13}C'
1	M					
2	G					
3	S					
4a	D	8.61	121.58	54.5	4.64	174.61
4b	D	8.43	121.76	54.7	4.59	175.88
5a	D	8.12	120.06	54.7	4.54	175.97
5b	D	8.25	120.40	54.6	4.54	176.29
6a	L	8.20	121.85	55.5	4.29	176.35
6b	L	8.13	121.60	55.6	4.28	177.54
7	M	8.29	119.78	55.6	4.43	175.89
8	N	8.28	119.26	53.1	4.68	174.29
9	D	8.28	122.38	52.1		
10	P	-		63.5	4.43	177.05
11	R	8.44	120.06	56.0	4.27	176.05
12	A	8.14	124.42	52.4	4.31	177.16
13	S	8.28	114.74	57.8	4.29	173.99
14	N	8.51	121.44	53.1		175.07
15	T	8.32	111.01	62.7	4.56	174.66
16	N	7.98	122.63	53.6	5.30?	174.32
17	L	9.62	128.63	53.4		
18	I					
19	V					
20	N					
21	Y					
22	L					
23	P	-				177.94
24	Q	9.04	127.34	58.0	3.95	174.51
25	D	7.88	111.87	52.3	4.49	176.73
26	M	7.38	120.11	56.5	4.22	175.66
27	T	7.95	118.39	60.0	4.43	174.71
28	D	9.01	122.00	58.3	4.06	178.45
29	R	8.30	117.96	58.8		178.42
30	E	7.88	123.03	59.1	4.03	179.04
31	L	7.95	121.25	58.2	4.05	177.68
32	Y	8.49	118.67	62.1	3.93	176.67
33	A	8.09	118.88	55.2	3.97	180.55
34	L	7.58	118.31	57.7	4.01	177.63
35	F	7.75	114.90	61.3	4.11	178.02
36	R	9.60	123.09	55.0	4.38	176.76
37	A	6.75	118.17	53.8	4.04	178.92
38	I	7.89	116.57	61.0	3.86	175.99
39	G	7.29	105.91	45.6		
40	P	-		63.8		175.79
41	I	8.28	123.79	60.4		175.51

Table 8.2 (cont.)

Res #	Type	^{1}HN	^{15}N	$^{13}C_{\alpha}$	$^{1}H_{\alpha}$	$^{13}C'$
42	N	9.19	125.01	54.8	4.80	174.74
43	T	7.39	108.27	60.0	4.58	172.19
44	C	8.72	123.23	58.3	5.00	172.00
45	R	8.55	126.75	55.9	4.70	173.77
46	I	8.48	125.50	61.9?		
47	M					
48	R					
49	D			53.6	4.50	177.74
50	Y	8.95	128.35	60.1	4.24	177.18
51	K	8.49	118.20	58.7	4.14	178.57
52	T	7.68	105.73	61.6	4.38?	176.28
53	G	8.43	111.65	45.5		173.56
54	Y	7.78	122.32	58.0		
55	S					
56	F			55.6?		176.06
57	G	8.93	103.92	45.6		173.28
58	Y	7.13	114.08	54.9	5.26	170.96
59	A	8.68	120.63	49.6	4.75	174.07
60	F	7.99	113.88	56.9	5.58	176.23
61	V ?	9.01	125.00	63.5	4.22?	
62	D					
63	F			60.1		174.02
64	T	8.03	111.05	65.3	4.14	174.80
65	S	8.74	114.17	55.8	4.91	175.55
66	E	9.60	128.75	58.4	4.33	177.96
67	M	8.29	119.48	58.3		178.82?
68	D	7.18	120.28	56.6	4.29	176.83
69	S	7.71	116.43	62.0	3.77	175.66
70	Q	7.58	119.22	59.1	3.99	178.90
71	R	7.47	120.52	59.2	3.98	177.13
72	A	8.41	120.94	55.5		179.40
73	I	7.45	116.73	66.3		177.69
74	K	7.29	117.73	59.2	4.00	178.88
75	V	8.53	115.36	64.7	3.92	177.37
76	L	8.37	117.02	55.2	4.43	177.26
77	N	7.63	115.39	56.8		176.73
78	G	8.61	116.88	45.5		174.01
79	I	7.62	119.31	62.1		176.23
80	T	8.20	120.07	62.2	5.17	175.30
81	V	8.83	128.04	62.1	4.00	173.73
82	R	9.18	123.46	59.2		
83	N					
84	K				4.69	175.30
85	R	8.35	120.15	54.1	5.02	176.42
86	L	8.93	124.45	56.4		178.56
87	K	8.68	124.62	54.8		174.20

Table 8.2 (cont.)

Res #	Type	^{1}HN	^{15}N	$^{13}C_{\alpha}$	$^{1}H_{\alpha}$	^{13}C'
88	V	8.74	127.50	61.4		173.40
89	S	8.55	118.96	56.6	4.63	172.48
90	Y	8.43	119.06	60.7		177.20
91	A	9.28	126.30	52.8?		177.67?
92	R					
93	P	-		64.0	4.34	176.63
94	G	8.13	109.48	45.0		174.76
95	G	7.54	107.61	44.7		175.33
96	E	8.84	120.71	59.2	3.87	178.48
97	S	8.48	111.76	61.3	4.29	
98	I					
99	K			56.4?		174.37
100	D	8.65	114.48	55.3	4.42	174.75
101	T	7.64	100.97	60.1	4.20	174.38
102	N	8.15	120.81	52.2	5.23	174.28
103	L	9.75	129.26			
104	Y	9.06	122.38	56.9	4.42	174.08
105	V	8.53	129.90	60.1	4.93	177.73
106	T	8.87	114.35	59.6		173.35
107	N	8.31	113.20	53.8		175.89
108	L	8.31	117.69	53.8	4.10	177.72
109	P	-		62.5	4.48	177.76
110	R	8.97	120.32	57.8	3.92	175.54
111	T	7.06	103.74	60.3	4.23	174.37
112	I	7.24	122.54	59.4	4.33	172.2
113	T	6.97	113.79	58.7	4.71	175.15
114	D	8.95	122.50	58.6		177.79
115	D	8.40	116.80	57.0	4.43	178.88
116	Q	7.75	120.39	58.6	4.13	178.84
117	L	8.46	120.67	58.8	4.13	179.09
118	D	8.28	119.48	57.9	4.32	179.65
119	T	8.09	117.81	66.8		176.16
120	I	8.18	120.18	65.1		177.34
121	F	8.71	112.35	61.2		180.64
122	G	9.13	115.11	46.5		173.82
123	K	6.52	116.16	57.4	3.92	176.85
124	Y	6.88	113.40	58.6	4.37	173.79
125	G	7.33	103.71	45.4		171.2
126	S	8.23	114.02	57.8	4.59	172.95
127	I ?	8.77	126.84			176.9
128	V	9.08	130.46	63.7	3.83	174.81
129	Q	7.55	119.08	55.0	4.44	171.51
130	K	8.39	120.22	54.7	4.97	174.78
131	N	8.84	116.91	53.7	4.60?	171.93
132	I	8.78	127.02	61.8		174.81
133	L	7.86	127.26	56.5	4.50	177.93

Table 8.2 (cont.)

Res #	Type	^{1}HN	^{15}N	$^{13}C_{\alpha}$	$^{1}H_{\alpha}$	^{13}C'
134	R	9.23	123.86	54.8	4.74	175.64
135	D	8.59	123.22	54.4		177.79
136	K	9.06	128.04	58.8	4.01	177.46
137	L	8.53	118.16	56.9	4.40	179.12
138	T	8.20	106.86	61.8	4.42	176.63
139	G	8.33	111.01	45.6	4.23	173.55
140	R	7.86	121.12	54.6	4.63	175.89
141	P	-		61.1		176.8
142	R	7.33	126.32	56.4	4.34	176.52
143	G	8.80	107.50	45.6		171.35
144	V	6.85	115.91	58.9	5.14	175.16
145	A	9.16	127.41	51.3?	5.16	173.92
146	F	8.18	116.24	57.0	5.49	176.19
147	V ?	8.46	120.05	62.3	4.10	
148	R					
149	Y					
150	N					
151	K					178.25
152	R ?	8.90	125.64	60.0		178.88
153	E ?	9.65	118.20	59.9	4.10	
154	E ?	7.17	120.36	58.2		176.28
155	A	6.73	118.26	54.7	3.79	178.92
156	Q	7.86	114.27	58.6	3.88	178.72
157	E	7.74	120.97	59.1	3.96	178.45
158	A	7.94	122.32	55.4	3.01	179.28
159	I	8.05	116.84	66.5		178.36
160	S	7.77	112.57	61.4		176.11
161	A	7.99	120.63	53.8	4.15	179.15
162	L	7.90	113.55	54.8	4.40	177.58
163	N	8.01	116.02	56.3		175.90
164	N	8.98	119.42	55.2	4.36	173.92
165	V	7.66	120.03	61.6	4.21	175.56
166	I	8.71	128.98	59.1		
167	P	-		62.1	4.53	175.76
168	E	8.66	121.90	58.3		174.39
169	G	8.87	114.27	45.3		174.67
170	G	8.52	108.12	43.9	4.54	172.43
171	S	9.05	110.74	58.3	4.08	173.40
172	Q	7.85	120.69		4.30	
173	P	-		62.1	4.08	176.60
174	L	8.42	122.59	55.3		177.62
175	S	7.82	115.89	56.9	5.17	173.59
176	V	8.83	127.16	61.8	4.49	172.89
177	R	8.14	122.28	53.4	4.78	174.46
178	L	8.77	120.92	55.9		178.47?
179	A	9.16	126.11	53.8		177.70

Table 8.2 (cont.)

Res #	Type	^{1}HN	^{15}N	$^{13}C_{\alpha}$	$^{1}H_{\alpha}$	$^{13}C'$
180	E	8.54	117.29	56.2	4.36	175.82
181	E	8.31	122.29	56.6	4.25	176.16
182	H	8.64	120.41	55.4	4.68	174.93
183	G	8.57	110.99	45.2	3.89	
184	K	7.89	126.02	57.5		

[a] All assignments are given in ppm. ^{13}C values are referenced to DSS. ^{1}H values are referenced to H_2O. Question marks represent uncertainty in an assignment. Residues D4-L6 have two sets ("a" and "b") of assignments. These assignments are as of 4/23/96.

Secondary structure

Placement of secondary structure was inferred from ^{13}C chemical shifts (Wishart & Sykes, 1994) and NOE patterns (Chapter 4). The relevant quantities when using ^{13}C chemical shifts are the differences from the random coil chemical shifts of the relevant amino acids (Wishart et al., 1995). These differences for $^{13}C_\alpha$, $^{13}C'$, and $^{1}H_\alpha$ are shown in Figure 8.6, with the consensus chemical shift index (CSI) given in the bottom panel. The consensus index in panel D is not strictly the same as that presented originally (Wishart & Sykes, 1994) since I had only 3 index types instead of 4 [most $^{13}C_\beta$s are missing from CBCA(CO)NH]. It differs in that only an agreement of 2 out of 3 (instead of 3 out of 4) are required for a consensus CSI value. To try to increase the rigorousness of this "compromised" CSI, I required that if 2 out of 3 had an index of "1", the third index of "-1" would break the consensus while "0" would not. This improved the results slightly (data not shown) and this protocol was used for the consensus plot in Figure 8.6D. Basically, residue stretches rich in +1 values in the consensus plot are helical, and residue stretches rich in -1 values in the consensus plot are β-sheet, although the β-sheet consensus is not as strong of an indicator as in helices.

In both RBD domains the βαβ-βαβ pattern of secondary structure seen in all RBDs is observed, and the secondary structure in the second domain of RBD1+2 matches that previously observed in RBD2 (Chapter 4). Therefore, it appears that no major rearrangement of secondary structure accompanies RNA binding, at least not for the second RBD. However, such comments should be reserved until the structure of free RBD1+2 has been fully mapped. N-terminal capping boxes (Harper & Rose, 1993) were observed from the NOE patterns (Chapter 4) in the first helices of both RBD domains. The linker between the two domains does not have any stand-out characteristics in terms of chemical shifts or NOE patterns, although the stretch of medium-intensity H_α(i-1)-HN(i) NOEs (Figure 8.5) is slightly suggestive of an extended conformation from P93 to S97. The C-terminal regions after the 4th strand of β-sheet in the U1A snRNP and hnRNP-C proteins

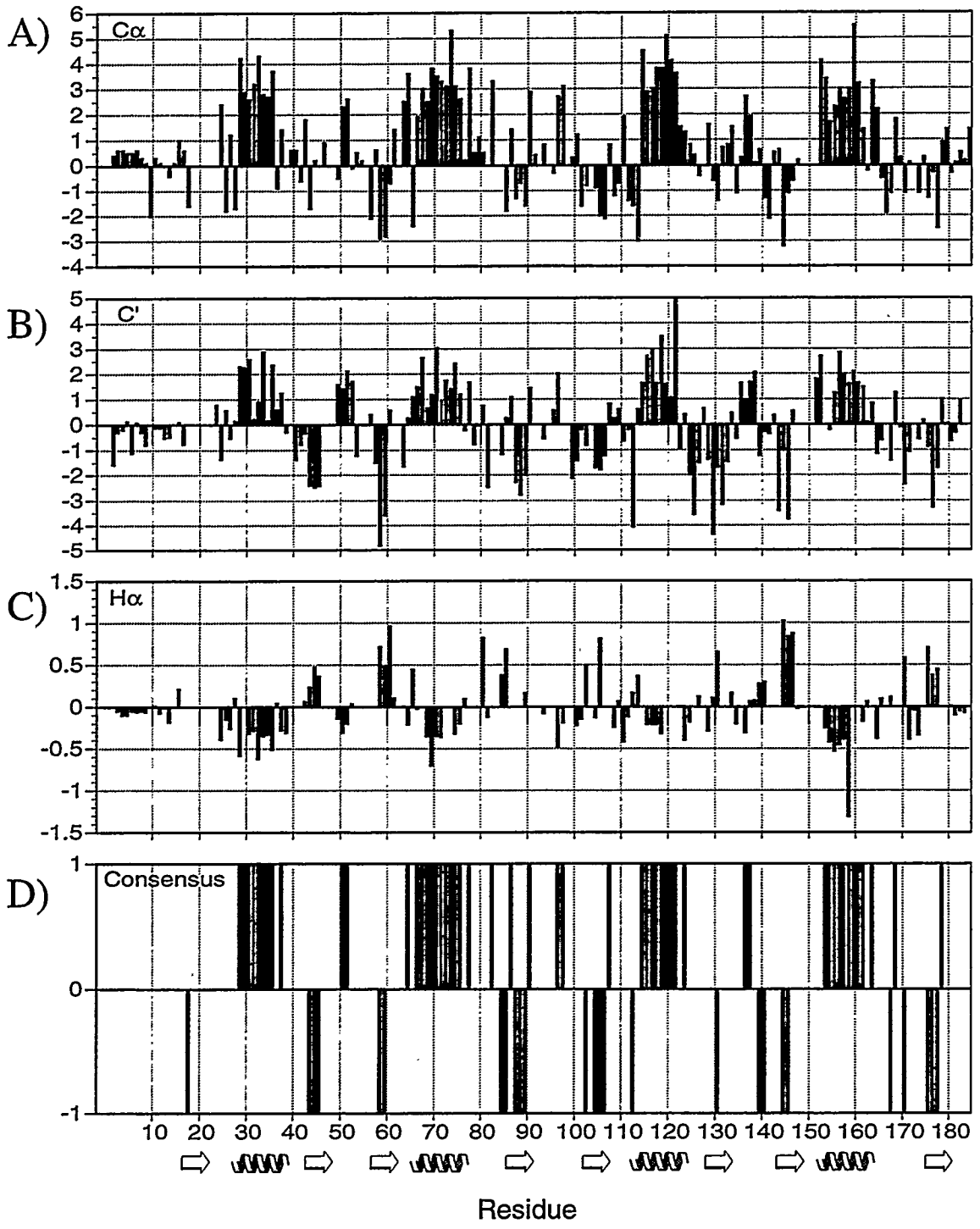


Figure 8.6. A) $^{13}C_{\alpha}$ chemical shift differences from random coil values. B) $^{13}C'$ chemical shift differences from random coil values. C) $^{1}H_{\alpha}$ chemical shift differences from random coil values. D) "Consensus" chemical shift index (CSI) as discussed in the text.

appear to be important for RNA-binding (Görlach et al., 1992; Oubridge et al., 1994). Moreover, a third helix is present at the C-terminus in the free and bound states of U1A snRNP (Howe et al., 1994; Oubridge et al., 1994). Therefore it is interesting that in Sxl-RBD1+2 no additional helices or elements of secondary structure are observed upon RNA binding. It is not clear from our present results how important the C-terminal and linker residues in RBD1+2 are for RNA binding, although there is certainly no helix formation.

Comparison with RBD2 chemical shifts

Even though the assignments for free RBD1+2 are not yet completed, a "free" vs. "bound" comparison of chemical shifts is possible, albeit only for the second RBD, which should highlight any residues that either contact RNA or undergo some kind of conformational change. A comparison between chemical shifts of RNA-bound RBD1+2 and free RBD2 should reveal similar features to bound RBD1+2 vs. free RBD1+2. This is because the $^{15}N/^{1}H$ HSQC spectrum of RBD2 superimposes quite well onto the HSQC spectrum of free RBD1+2 (data not shown). Figure 8.7A shows the chemical shift differences in Hz between RBD1+2 complexed with RNA and RBD2 for amide ^{1}H and ^{15}N resonances. Figure 8.7B shows an analogous plot with $^{13}C_{\alpha}$ and $^{13}C'$ differences shown. Carbonyl assignments for free RBD2 were made from 3D HNCO by Brian Volkman. Major changes in chemical shifts occur in and around the β-sheet of RBD2 except for the 4th strand, which has only moderate shifts. This is consistent with the β-sheet forming the majority of the RNA-binding interface. The largest shift is clearly in the amide of R142, the first residue of the conserved RNP-1 octamer in RBD2. In the U1A-snRNP complex (Oubridge et al., 1994), the corresponding arginine (R52) backbone amide is involved in extensive hydrogen bonding interactions with the RNA. Perhaps the recognition processes in these two systems share some important features even though they differ in their number of RBD domains and in their cognate RNAs (stem loop vs. single stranded). It is interesting that there are large shifts on one edge of the β-sheet (strand 2)

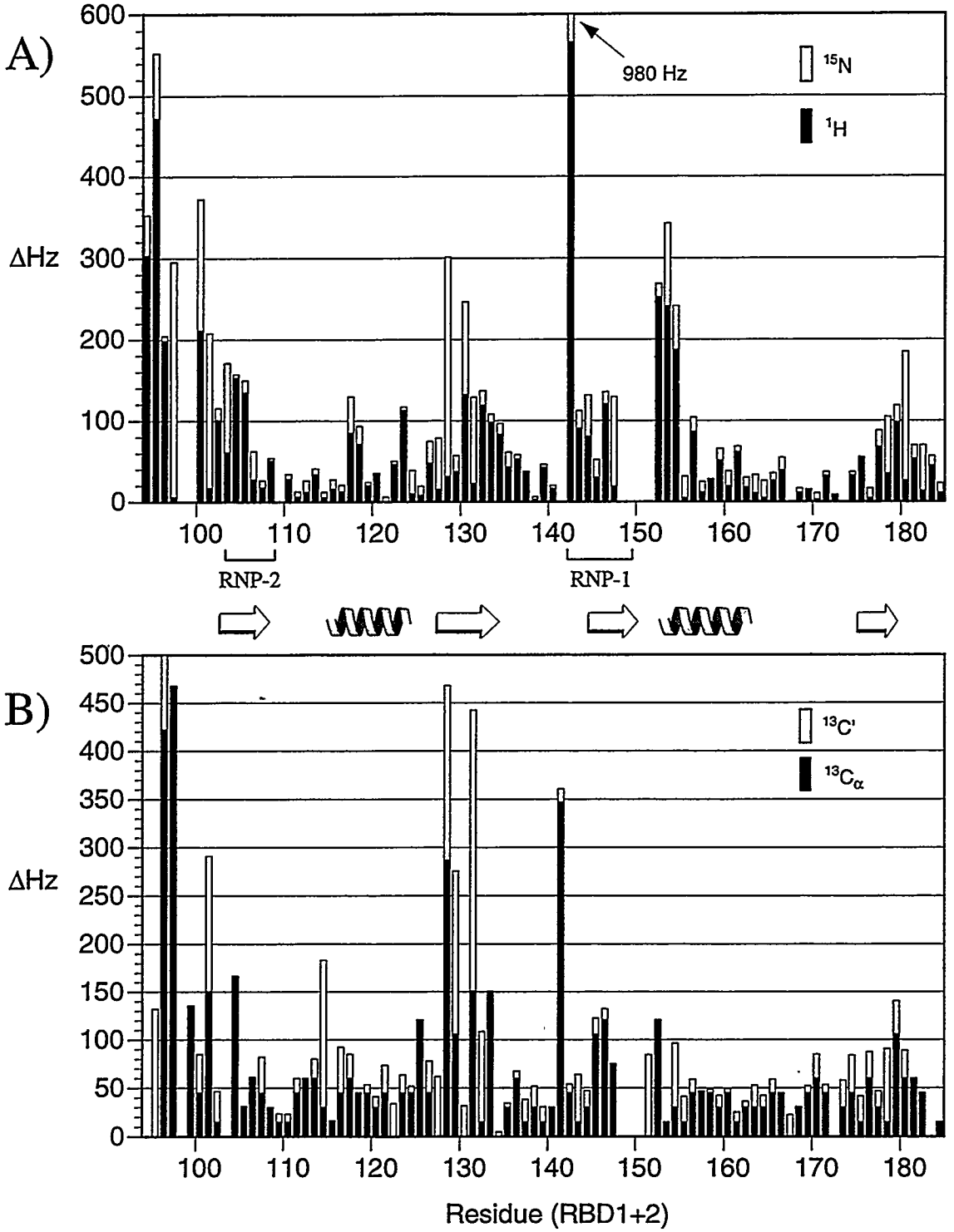


Figure 8.7. A) ^{1}HN (black bars) and ^{15}N (white bars) chemical shift differences between RBD1+2 bound to 5'-GUUUUUUUUC-3' and RBD2 (Chapter 4). B) $^{13}C_{\alpha}$ (black bars) and ^{13}C' (white bars) chemical shift differences between RBD1+2 bound to 5'-GUUUUUUUUC-3' and RBD2 (Chapter 4). Differences in Hz are calculated using 600 Hz/ppm. Shifts are stacked to show total shiftsin Hz.

and not on the other (strand 4). It is also interesting that the largest changes in this region occur near the β-bulge involving I127-V128. Perhaps the bulge plays a direct role in making specific interactions with the RNA. It follows that this might be an interesting target for mutagenesis with regard to the effects on binding affinity and specificity. Finally, there appear to be significant shifts at the N-terminus of helix 2 in RBD2. As there are two consecutive positively charged residues here (K151-R152), these residues may be making contacts with the RNA as well, and therefore the N-terminus of helix 2 of RBD2 would have to be considered as an important recognition scaffolding along with the sheet. K151 and R152 are on the surface of the protein in the RBD2 structures, and the sidechain of R152 in particular is positioned at the edge of the sheet near the N- and C-termini. This sidechain is oriented towards the space in which the RNA is expected to interact with the protein.

Unfortunately, because comparisons cannot yet be made with free RBD1+2, the changes in chemical shifts for the linker residues between the two domains cannot be taken seriously, since they are terminal residues in RBD2. It will be of great interest to see if the linker plays an active role in RNA recognition. One main question with regard to RNA binding in this two-RBD-domain system is how the two domains are oriented relative to one another in the free and bound states. Do the sheets align to form one long, continuous binding surface, or do they somehow "clamp" down on the RNA? At this stage, it has been difficult to find evidence in the NOE data for cross-strand, cross-domain NOEs which could confirm a formation of continuous sheet; this may be something to watch for in the future. No interdomain NOEs could be identified either, although it will probably require more data to unambiguously assign such NOEs.

Finally, and very importantly, no NOEs between the protein and the RNA have been sequence-specifically assigned in any of the data so far collected. Such NOEs have been identified in 2D NOESY data (Figure 8.8), but the corresponding NOEs in the ^{15}N-separated 3D NOESY-HSQC are either so weak as to become unobservable (NOEs to

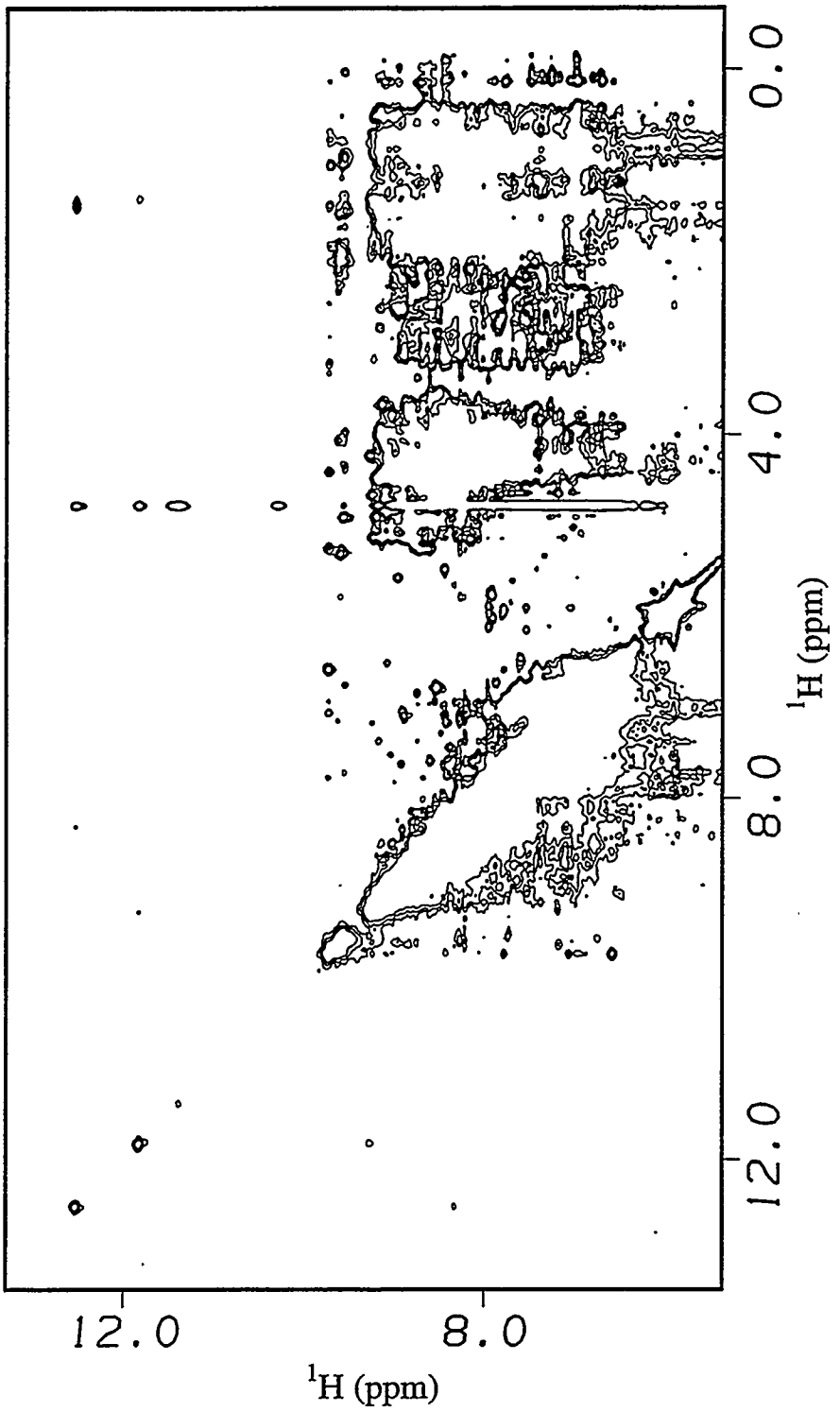


Figure 8.8. Downfield half of a 2D NOESY specturm of Sxl-RBD1+2 bound to the Tra-PPT 10-mer (5'-GUUUUUUUUC-3') at 750 MHz and 25 °C. The NOE mixing time was 75 ms, and water suppression was obtained with WATERGATE. Intermolecular NOEs can be seen to the imino protons of the RNA.

imino protons) or cannot be easily differentiated from intramolecular protein NOEs. I believe that ^{13}C isotope edited experiments (Ikura & Bax, 1992; Lee et al., 1994; Vuister et al., 1994) will eventually allow these intermolecular NOEs to be identified and assigned.

Relaxation measurements

One of the most intriguing aspects of Sxl RNA-binding activity is that it requires two RBD domains for high affinity binding. Specifically, it appears that these domains must be covalently linked (Roland Kanaar, personal communication). Mixing RBD1, RBD2, and the Tra-PPT does not yield an observable complex on a native gel. Why is it so important that these two domains be covalently linked together? What is the role of the linker itself with respect to RNA binding? Perhaps one way to address such questions is to directly probe the motional dynamics of the linker region using spin relaxation as was done for RBD2 (Chapter 6).

Brian and I have initiated such an undertaking and have started with the measurement of protein ^{15}N T_2 relaxation rates in the RBD1+2/Tra-PPT 10-mer complex. With only T_2 values, a rigorous characterization of the dynamics using a Lipari and Szabo type analysis (Chapter 6) is not possible. But since T_2 trends mirrored the order parameter (S^2) trends in RBD2, the T_2 values presented here for the complex may qualitatively reflect relative order and disorder in the protein backbone. We also have preliminary data on ^{15}N T_1 time constants, which will not be reported here because these will soon be replaced with data of higher quality. Nevertheless, an estimate for the average T_1/T_2 ratio for structured regions made possible an estimate for the overall tumbling correlation time (see below). Ultimately, it will be of interest to compare the backbone dynamics of the bound RBD1+2 to free RBD1+2, where any changes in dynamics upon binding RNA will be detectable. We are very interested in observing any potential rigidifications that might occur in either the linker or in loop regions upon RNA-binding.

The ^{15}N T_2 (and T_1) data was acquired at 750 MHz using a sensitivity enhanced, water flip-back, gradient coherence selected HSQC-type experiment with a parametrically variable CPMG transverse relaxation period (Farrow et al., 1994). A slight alteration was made in the CPMG pulse train to allow for more rapid sampling of the exponential decays as well as more efficient cancelation of the effects from cross-correlation between dipolar and CSA relaxation mechanisms (Kay et al., 1992b; Palmer et al., 1992). In Kay's "water flip-back" CPMG sequence, ^{1}H 180° pulses (for cancelation of cross-correlation effects) are placed after every eight ^{15}N 180° pulses. Because the ^{15}N pulses are spaced by 0.9 ms, the ^{1}H pulses occur every ~8 ms. Therefore, the shortest possible CPMG period must be ~16 ms, as a second ^{1}H 180° pulse with opposite phase must occur to return water to +Z at the end of the CPMG cycle. Cancelation of cross-correlation effects is accomplished by flipping protons at a rate fast compared to the decay of transverse ^{15}N magnetization. Since the relaxation rates for RBD1+2 in this complex are considerably faster than other proteins which have been studied, doubling the frequency of proton 180° pulses in the CPMG should result in more efficient cancelation of cross-correlation effects on ^{15}N transverse relaxation rates. This was accomplished by simply placing ^{1}H pulses after every four (instead of eight) ^{15}N pulses, making the minimum CPMG period ~8 ms. In addition, this allowed for better sampling of the decay since many ^{15}N resonances decayed completely within the first 50-60 ms. Nine T_2 time points and three duplicate points (for error estimates) were acquired, each with 64 scans/block. Transverse relaxation was sampled at 0.0, 8.1, 16.2, 24.3, 32.4, 40.5, 48.6, 64.8, and 129.5 ms. Duplicate points were acquired for 0.0, 8.1, and 16.2 ms time points. The data were analyzed in a manner similar to the descriptions in Chapter 6.

Figure 8.9 contains the ^{15}N R2 ($1/T_2$) relaxation rates for RBD1+2 bound to the Tra-PPT 10-mer. These rates are also given in tabular form in Appendix B. The standard deviations are larger than desired, but small enough so that dramatic changes in relaxation rates can be detected easily. The reason for these large error bars is the relatively low

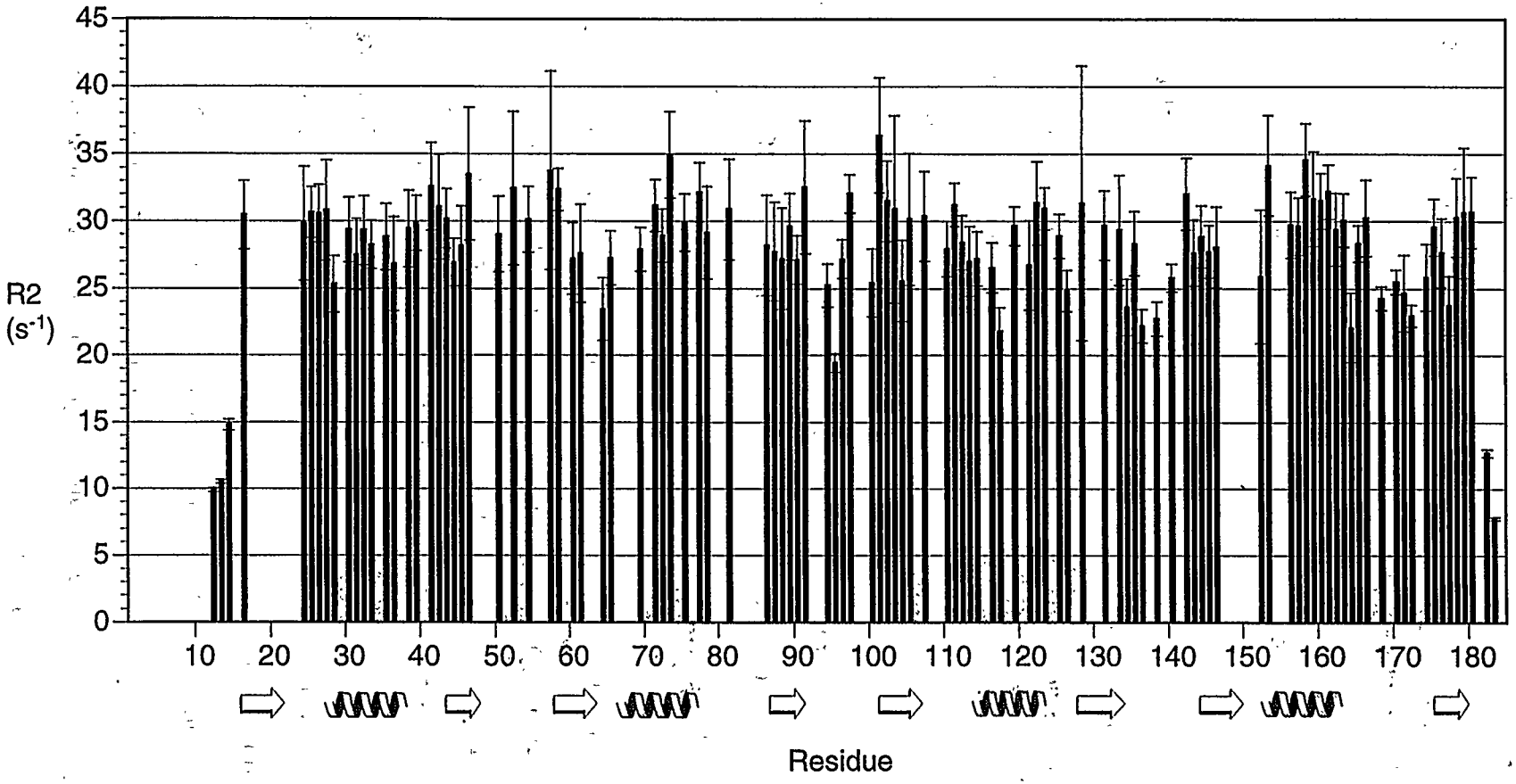


Figure 8.9. Backbone amide ^{15}N R2 ($1/T_2$) relaxation rates of RBD1+2 bound to the Tra-PPT 10-mer measured at 76 Mhz. Values are tabulated in Appendix B.

signal/noise ratio of the crosspeaks. These weak crosspeaks result from broad lines which have a significant amount of relaxation during the INEPT and RINEPT sequeneces of the HSQC. This is why the larger R2 values generally have larger percentage errors than smaller R2 values. It is worth noting here that the transverse relaxation rates in this complex are approximately 3-fold larger than those observed in RBD2 (Chapter 6). Clearly, the N-terminus has very sharp lines (small R2). Unfortunately, many of these residues could not be fit since they occur in crowded regions of the HSQC. The 3 residues at the C-terminus are also quite sharp, even in the bound state. It was suspected that these C-terminal residues might be involved in RNA recognition as was mentioned above, but at least in the bound complex these residues seem to "fray" as in free RBD2 (Chapters 5, 6). Residues involved in secondary structure have relaxation rates of around 28 s^{-1}. Since the corresponding R1 rates were estimated to be 0.6 s^{-1}, the average R2/R1 ratio is estimated to be 47 ± 8. This was used (as in Chapter 6) to calculate an overall tumbling correlation time of 17.8 ± 1.5 ns for the complex.

With the exception of perhaps T64, the backbone of RBD1 appears to lack any obvious regions with relatively slow transverse relaxation rates, typically associated with dynamic disorder. In RBD2 there are two regions which appear to have relatively slow relaxation rates: the β2-β3 loop and the α2-β4 loop. Interestingly, these are the same regions that have long T_2 values and low order parameters (S^2) in RBD2 (Chapter 6). Therefore, at first glance it appears that these loops, especially the first 2/3 of the β2-β3 loop, do *not* rigidify upon binding RNA, contrary to U1A-snRNP (Oubridge et al., 1994). This loop in RBD proteins is often considered to be a "specificity loop" intimately involved in RNA-recognition. It is possible that the basic residues in this loop are important for the initial recognition but do not form stable contacts with the RNA in the bound state. Nevertheless, the *non-rigification* implied by this relaxation data is supported by the chemical shift data presented in Figure 8.6. L117, which is in the middle of the helix1 of RBD2, also shows a significantly slower relaxation rate. It is difficult to say anything

definite with regard to this residue, although interesting dynamic behavior was observed in the middle of this helix in free RBD2 as evidenced by ^{15}N relaxation (Chapter 6) and amide exchange rates with solvent (Chapter 4). It should be remembered, however, that the conclusions just drawn must eventually be rigorously drawn from a complete set of relaxation data, not simply from transverse relaxation rates. Hopefully this will be completed in the future.

Perhaps one of the most interesting features in this relaxation data is the different relaxation patterns observed in RBD1 vs. RBD2. The above-mentioned loops in RBD2 with some flexibility have no counterparts in RBD1, even though the corresponding loops exist and have similar lengths (Figure 3.2). This appears to suggest that the types of interactions in which RBD1 and RBD2 participate are of a different nature. It implies that the RNA recognition of the Tra-PPT by Sxl is not a simple sum of two analogous or generic protein/RNA interactions, but an interaction of greater complexity where the two domains take on asymmetric roles in specific RNA recognition. I eagerly await the completion of the free RBD1+2 assignments. The changes in chemical shifts upon RNA binding can then be observed for RBD1 and the linker, and this information should help to clarify this matter of domain heterogeneity in Sex-lethal. Ultimately, structural characterization of the complex will provide the answer in satisfying, unequivocal detail.

The question which initiated this section had to do with the role of the linker between RBD1 and RBD2. From the T_2 data alone, it is difficult to unequivocally state that the linker is either rigid or flexible in the RNA-bound state. This difficulty stems from the range of rates observed as well as the conspicuous absences of T_2 data for residues in the linker, a result of incomplete assignments as well as the presence of a proline at position 93. G95, however, has a significantly slower rate ($< 20\ s^{-1}$) than the rest of the protein. Therefore, at least one part of the linker appears to be flexible in the bound state. It should be very interesting to characterize the relaxation behavior of the linker in the absence of RNA, allowing a comparison of backbone dynamics between the bound and free forms of

RBD1+2. Stephanie Robertson, a graduate student in the Wemmer group, is presently making progress on the resonance assignments and relaxation/dynamics studies of free RBD1+2. I hope that the knowledge ultimately gained from this simplest of multi-RBD/RNA complexes can be applied to predict RNA-binding properties of other systems with two or more RBD domains.

Bax, A., Ikura, M., Kay, L. E., Torchia, D. A. & Tschudin, R. (1990) *J. Magn. Reson. 86*, 304-318.

Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994) *Biochemistry 33*, 5984-6003.

Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L. & Dreyfuss, G. (1992) *Embo Journal 11*, 3289-3295.

Grzesiek, S. & Bax, A. (1992) *J. Magn. Reson. 96*, 432-440.

Grzesiek, S. & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 12593.

Harper, E. T. & Rose, G. D. (1993) *Biochemistry 32*, 7605-7609.

Howe, P. W. A., Nagai, K., Neuhaus, D. & Varani, G. (1994) *The EMBO Journal 13*, 3873-3881.

Ikura, M., Kay, L. E. & Bax, A. (1990) *Biochemistry 29*, 4659-4667.

Ikura, M. & Bax, A. (1992) *J. Am. Chem. Soc. 114*, 2433-2440.

Jeener, J. (1971), Ampere International Summer School, Basko Polje, Yugoslavia.

Kay, L. E., Keifer, P. & Saarinen, T. (1992a) *Journal of the American Chemical Socity 114*, 10663-10665.

Kay, L. E., Nicholson, L. K., Delaglio, F., Bax, A. & Torchia, D. A. (1992b) *J. Magn. Reson. 97*, 359-375.

Kay, L. E., Xu, G. Y. & Yamazaki, R. (1994) *Journal of Magnetic Resonance, Series A 109*, 129-133.

Lee, W., Revington, M. J., Arrowsmith, C. & Kay, L. E. (1994) *FEBS Lett. 350*, 87-90.

Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989) *J. Magn. Reson. 85*, 393-399.

Muhandiram, D. R. & Kay, L. E. (1994) *Journal of Magnetic Resonance (B) 103*, 203-216.

Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. (1994) *Nature 372*, 432-438.

Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991) *J. Magn. Reson. 93*, 151-170.

Palmer, A. G., Skelton, N., Chazin, W. J., Wright, P. E. & Rance, M. (1992) *Mol. Phys. 75*, 699-711.

Powers, R., Gronenborn, A. M., Clore, G. M. & Bax, A. (1991) *J. Magn. Reson. 94*, 209-213.

Vuister, G. W., Kim, S. J., Wu, C. & Bax, A. (1994) *J. Am. Chem. Soc. 116*, 9206-9210.

Wall, L. & Schwartz, R. L. (1991) *Programming perl*, (T. O'Reilly), O'Reilly & Associates, Inc., Sebastopol.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67-81.

Wishart, D. S. & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171-180.

Zhang, O., Kay, L. E., Olivier, J. P. & Forman-Kay, J. D. (1994) *J. Biomol. NMR 4*, 845-858.

Appendix A

Sxl-RBD2 ^{15}N R1 ($1/T_1$), ^{15}N R2 ($1/T_2$), and $^{15}N/^{1}H$ NOE Values at 60.8 MHz, pH=5.0, 25 °C

Residue	R1 (sec^{-1})	error	χ^2	R2 (sec^{-1})	error	χ^2	$^{15}N/^{1}H$ NOE
arg-5	1.43	0.06	1,118.53	2.17	0.04	1,722.33	-2.12
pro-6	-	-	-	-	-	-	-
gly-7	1.56	0.14	4.51	2.52	0.10	1.93	-0.20
gly-8	1.64	0.10	4.77	2.48	0.08	5.41	0.05
glu-9	1.75	0.10	4.12	3.84	0.10	1.49	0.24
ser-10	1.73	0.13	2.85	4.69	0.20	1.80	0.38
ile-11	1.85	0.17	1.33	7.09	0.43	0.89	0.60
lys-12	1.78	0.11	0.78	10.87	0.48	7.20	0.68
asp-13	1.55	0.17	0.10	8.71	0.54	0.76	0.74
thr-14	1.47	0.16	0.36	11.18	0.76	1.23	0.78
asn-15	1.59	0.16	0.60	10.60	0.61	1.17	0.87
leu-16	1.62	0.21	0.06	9.65	0.81	1.32	0.89
tyr-17	1.61	0.18	0.19	10.00	0.67	0.92	0.81
val-18	1.64	0.18	0.47	9.62	0.71	2.87	0.88
thr-19	1.53	0.16	0.70	11.05	0.66	0.99	0.86
asn-20	1.54	0.16	0.51	9.88	0.54	0.50	0.84
leu-21	1.56	0.12	0.66	10.45	0.41	3.06	0.82
pro-22	-	-	-	-	-	-	-
arg-23	1.56	0.18	0.64	10.11	0.72	1.69	0.81
thr-24	1.54	0.13	1.09	9.29	0.46	2.60	0.77
ile-25	1.81	0.10	1.40	9.82	0.28	0.78	0.81
thr-26	1.52	0.17	1.76	9.69	0.53	1.82	0.79
asp-27	1.62	0.13	0.72	9.68	0.61	2.30	0.89
asp-28	1.58	0.10	1.63	10.47	0.37	1.40	0.85
gln-29	1.58	0.11	0.96	10.67	0.42	1.14	0.84
leu-30	1.64	0.10	0.51	9.87	0.33	8.03	0.82
asp-31	1.58	0.10	2.52	10.22	0.30	4.25	0.86
thr-32	1.31	0.09	1.01	10.50	0.36	4.23	0.89
ile-33	1.35	0.11	1.63	10.01	0.37	2.38	0.84
phe-34	1.55	0.17	0.47	10.28	0.63	1.28	0.89
gly-35	1.59	0.14	0.64	11.64	0.60	0.96	0.88
lys-36	1.51	0.06	0.73	10.59	0.29	6.22	0.81
tyr-37	1.43	0.14	0.87	10.36	0.50	4.60	0.89
gly-38	1.48	0.09	0.24	10.48	0.38	7.49	0.85
ser-39	1.34	0.11	0.30	8.78	0.35	1.32	0.83

Sxl-RBD2 ^{15}N R1 ($1/T_1$), ^{15}N R2 ($1/T_2$), and $^{15}N/^{1}H$ NOE Values (cont.)

Residue	R1 (sec^{-1})	error	χ^2	R2 (sec^{-1})	error	χ^2	$^{15}N/^{1}H$ NOE
val-40	1.44	0.15	0.21	9.63	0.56	0.85	0.89
val-41	1.60	0.12	0.97	10.20	0.66	5.43	0.83
gln-42	1.67	0.11	0.72	9.82	0.38	0.89	0.83
lys-43	1.65	0.12	0.19	9.18	0.39	1.65	
asn-44	1.50	0.13	0.18	9.18	0.42	0.49	0.84
ile-45							
leu-46	1.70	0.17	1.18	9.57	0.57	4.17	0.85
arg-47	1.67	0.29	0.53	10.82	1.20	1.18	0.85
asp-48	1.35	0.13	1.12	9.21	0.39	2.04	0.81
lys-49	1.54	0.25	1.08	8.71	0.83	2.03	0.73
leu-50	1.50	0.11	0.52	7.84	0.28	0.54	0.62
thr-51	1.50	0.14	0.42	8.49	0.44	1.98	0.72
gly-52	1.63	0.13	0.81	9.01	0.37	1.05	0.77
arg-53	1.68	0.10	0.79	8.96	0.32	2.03	0.72
pro-54	-	-	-	-	-	-	-
arg-55	1.59	0.17	0.62	9.14	0.54	2.78	0.76
gly-56	1.58	0.19	0.47	9.01	0.60	2.34	0.77
val-57	1.68	0.18	0.55	9.94	0.64	1.37	0.87
ala-58	1.67	0.16	0.12	10.40	0.59	0.83	0.87
phe-59	1.48	0.18	0.64	10.19	0.66	0.99	0.83
val-60	1.51	0.16	0.30	8.46	0.51	0.99	0.83
arg-61	1.57	0.17	0.40	10.38	0.76	1.01	0.90
tyr-62	1.62	0.21	0.36	9.03	0.63	1.97	0.81
asn-63	1.47	0.14	0.28	10.39	0.50	2.00	0.86
lys-64	1.39	0.15	0.07	10.11	0.55	1.50	0.83
arg-65	1.59	0.17	2.68	9.75	0.56	3.12	0.85
glu-66	1.47	0.14	0.90	10.56	0.51	3.78	0.92
glu-67	1.73	0.17	0.86	10.95	0.56	3.62	0.87
ala-68	1.34	0.10	0.75	10.70	0.38	10.53	0.83
gln-69	1.31	0.09	2.36	10.53	0.35	4.97	0.78
glu-70	1.68	0.09	0.94	10.88	0.33	1.86	0.84
ala-71	1.70	0.11	1.59	10.78	0.36	2.01	0.86
ile-72	1.49	0.11	0.61	11.35	0.48	3.65	0.84
ser-73	1.63	0.10	2.54	10.64	0.31	0.44	0.87
ala-74	1.63	0.10	1.68	10.61	0.33	0.79	0.86
leu-75	1.41	0.16	0.07	10.78	0.59	1.53	0.84
asn-76	1.45	0.11	1.00	11.10	0.44	3.01	0.81
asn-77	1.51	0.14	0.54	9.63	0.51	0.84	0.79
val-78	1.54	0.10	0.29	9.77	0.37	1.44	0.81
ile-79	1.51	0.15	1.32	9.31	0.54	0.55	0.75
pro-80	-	-	-	-	-	-	-
glu-81	1.35	0.07	0.31	9.41	0.28	1.10	
gly-82	1.53	0.14	0.79	8.46	0.40	0.67	0.67

Sxl-RBD2 ^{15}N R1 ($1/T_1$), ^{15}N R2 ($1/T_2$), and $^{15}N/^{1}H$ NOE Values (cont.)

Residue	R1 (sec^{-1})	error	χ^2	R2 (sec^{-1})	error	χ^2	$^{15}N/^{1}H$ NOE
gly-83	1.42	0.11	0.12	9.06	0.40	1.16	0.75
ser-84	1.46	0.60	0.17	7.53	1.80	0.64	0.75
gln-85	1.50	0.13	0.28	7.77	0.36	1.45	0.72
pro-86	-	-	-	-	-	-	-
leu-87	1.62	0.13	0.64	9.67	0.46	6.37	0.76
ser-88	1.63	0.15	0.69	9.63	0.53	0.34	0.85
val-89	1.50	0.16	0.21	9.29	0.53	0.92	0.84
arg-90	1.70	0.14	0.61	10.09	0.55	3.30	0.84
leu-91	1.48	0.10	0.07	9.48	0.38	0.86	0.82
ala-92							
glu-93	1.59	0.10	0.50	8.35	0.31	0.47	0.71
glu-94	1.72	0.10	1.67	5.00	0.16	2.78	0.40
his-95	1.74	0.24	1.28	4.04	0.32	3.93	0.22
gly-96	1.58	0.27	1.20	2.82	0.24	1.28	-0.35
lys-97	1.20	0.04	10.35	1.85	0.02	27.02	-0.70

$^{15}/^{1}H$ NOE error values are estimated to be approximately $\pm$ 0.05.

Appendix B

Sxl-RBD1+2 ^{15}N R2 ($1/T_2$) Values When Bound to the Tra-PPT 10-mer (5'-GUUUUUUUUC-3') at 76.0 MHz, pH=6.3, 25 °C

Residue	R2 (sec^{-1})	error	χ^2
asp-4a	12.12	0.23	306,535.90
asp-4b	16.34	0.89	13,322.12
asp-5a			
asp-5b			
leu-6a	16.44	0.26	159,060.40
leu-6b	12.13	1.01	7,153.24
met-7			
asn-8			
asp-9			
pro-10			
arg-11			
ala-12	9.89	0.15	113.22
ser-13	10.53	0.17	15.16
asn-14	14.81	0.40	12.05
thr-15			
asn-16	30.45	2.55	2.20
leu-17			
ile-18			
val-19			
asn-20			
tyr-21			
leu-22			
pro-23			
gln-24	29.84	4.24	4.75
asp-25	30.64	1.89	1.87
met-26	30.56	2.15	0.42
thr-27	30.80	3.73	0.62
asp-28	25.32	2.13	0.80
arg-29			
glu-30	29.36	2.41	1.13
leu-31	27.52	2.64	1.65
tyr-32	29.30	2.53	1.13
ala-33	28.24	1.79	5.35
leu-34			
phe-35	28.82	2.47	3.16
arg-36	26.83	3.49	2.12
ala-37			
ile-38	29.44	2.84	3.31
gly-39	29.85	2.02	1.74

Sxl-RBD1+2 ^{15}N R2 ($1/T_2$) Values When Bound to the Tra-PPT 10-mer (5'-GUUUUUUUUC-3') (cont.)

Residue	R2 (sec^{-1})	error	χ^2
pro-40			
ile-41	32.57	3.24	12.63
asn-42	31.06	3.89	0.41
thr-43	30.19	2.21	0.90
cys-44	26.94	1.78	1.98
arg-45	28.16	2.95	2.60
ile-46	33.51	4.94	0.76
met-47			
arg-48			
asp-49			
tyr-50	29.03	2.82	2.74
lys-51			
thr-52	32.46	5.70	1.36
gly-53			
tyr-54	30.14	2.45	2.37
ser-55			
phe-56			
gly-57	33.79	7.38	3.31
tyr-58	32.36	1.58	7.09
ala-59			
phe-60	27.24	2.65	0.71
val-61	27.61	3.63	3.01
asp-62			
phe-63			
thr-64	23.47	2.35	0.54
ser-65	27.27	2.01	2.09
glu-66			
met-67			
asp-68			
ser-69	27.91	1.62	2.82
gln-70			
arg-71	31.16	1.95	1.14
ala-72	28.92	1.96	2.25
ile-73	34.94	3.20	1.95
lys-74			
val-75	29.89	2.12	3.82
leu-76			
asn-77	32.17	2.20	2.92
gly-78	29.15	3.45	4.54
ile-79			
thr-80			
val-81	30.89	3.74	5.02
arg-82			
asn-83			
lys-84			
arg-85			

Sxl-RBD1+2 ^{15}N R2 ($1/T_2$) Values When Bound to the Tra-PPT 10-mer (5'-GUUUUUUUUC-3') (cont.)

Residue	R2 (sec^{-1})	error	χ^2
leu-86	28.19	3.73	2.41
lys-87	27.74	3.67	2.66
val-88	27.21	3.79	2.72
ser-89	29.60	2.48	2.28
tyr-90	27.14	1.80	17.74
ala-91	32.52	4.94	1.25
arg-92			
pro-93			
gly-94	25.24	1.62	4.46
gly-95	19.45	0.71	5.95
glu-96	27.18	1.44	4.74
ser-97	32.05	1.42	3.80
ile-98			
lys-99			
asp-100	25.43	2.53	1.32
thr-101	36.39	4.27	1.55
asn-102	31.50	3.00	1.31
leu-103	30.88	6.96	0.52
tyr-104	25.56	3.01	1.11
val-105	30.15	4.91	1.28
thr-106			
asn-107	30.39	3.33	1.91
leu-108			
pro-109			
arg-110	27.94	2.05	1.66
thr-111	31.22	1.61	3.55
ile-112	28.42	2.00	5.28
thr-113	27.02	2.57	0.96
asp-114	27.21	2.02	3.05
asp-115			
gln-116	26.54	1.86	0.41
leu-117	21.83	1.76	14.18
asp-118			
thr-119	29.65	1.43	2.06
ile-120			
phe-121	26.73	3.31	2.07
gly-122	31.36	3.09	1.08
lys-123	30.93	1.57	6.59
tyr-124			
gly-125	28.89	1.62	3.39
ser-126	24.82	1.53	3.20
ile-127			
val-128	31.34	10.20	2.95
gln-129			
lys-130			
asn-131	29.70	2.58	5.44

Sxl-RBD1+2 ^{15}N R2 ($1/T_2$) Values When Bound to the Tra-PPT 10-mer (5'-GUUUUUUUUC-3') (cont.)

Residue	R2 (sec^{-1})	error	χ^2
ile-132			
leu-133	29.35	4.08	3.38
arg-134	23.61	2.10	1.77
asp-135	28.35	2.38	3.56
lys-136	22.19	1.25	0.95
leu-137			
thr-138	22.74	1.25	7.69
gly-139			
arg-140	25.81	1.04	2.16
pro-141			
arg-142	32.03	2.69	3.83
gly-143	27.65	2.46	3.77
val-144	28.84	2.30	1.69
ala-145	27.74	1.95	3.37
phe-146	28.06	3.00	0.54
val-147			
arg-148			
tyr-149			
asn-150			
lys-151			
arg-152	25.90	4.97	3.53
glu-153	34.15	3.72	2.58
glu-154			
ala-155			
gln-156	29.75	2.44	0.71
glu-157	29.64	2.13	2.65
ala-158	34.59	2.68	1.69
ile-159	31.66	3.51	2.84
ser-160	31.53	2.09	2.44
ala-161	32.22	2.05	0.68
leu-162	29.39	2.70	1.36
asn-163	30.08	1.98	4.91
asn-164	22.09	2.60	2.30
val-165	28.36	1.36	3.21
ile-166	30.26	2.83	4.91
pro-167			
glu-168	24.27	0.88	0.34
gly-169			
gly-170	25.50	0.89	1.56
ser-171	24.67	2.87	1.14
gln-172	22.99	0.82	5.55
pro-173			
leu-174	25.85	2.50	3.77
ser-175	29.56	2.10	2.18
val-176	27.67	2.55	2.80
arg-177	23.74	2.19	4.28

Sxl-RBD1+2 ^{15}N R2 ($1/T_2$) Values When Bound to the Tra-PPT 10-mer (5'-GUUUUUUUUC-3') (cont.)

Residue	R2 (sec^{-1})	error	χ^2
leu-178	30.30	2.93	1.52
ala-179	30.64	4.83	0.26
glu-180	30.68	2.63	3.31
glu-181			
his-182	12.71	0.27	33.64
gly-183	7.77	0.12	78.80
lys-184	13.77	0.19	143,591.20

Appendix C

Distance Restraints (NOEs and hydrogen bonds):

X-PLOR distance restraint file (noe.tbl) used for round 24 of structure calculations of Sxl-RBD2. Changes made for round 24 involve stereo-selective NOEs to val-60 hg1* or hg2* methyl protons. Other changes are annotated in braces. The distance restraint file used for round 22 is obtained by removing all restraint additions for round24. The explicit restraint file for round 22 can be obtained from the supplementary floppy disk.

```
! residue  4 ala-4                round 24   dg

! residue  5 arg-5

! residue  6 pro-6
! assign (resid  6  and name hd*) (resid  54  and name hg*) 7.0 5.2 0.0
 assign (resid  6  and name hd*) (resid  90  and name ha) 4.3 2.5 0.2

! residue  7 gly-7

! residue  8 gly-8

! residue  9 glu-9
 assign (resid  9  and name ha) (resid  10  and name hn) 2.4 0.6 0.3
 assign (resid  9  and name hb*) (resid  10  and name hn) 4.3 2.5 0.2

! residue  10 ser-10
 assign (resid  10  and name ha) (resid  11  and name hn) 2.4 0.6 0.3
 assign (resid  10  and name hb*) (resid  11  and name hn) 6.0 4.2 0.0

! residue  11 ile-11
 assign (resid  11  and name hb) (resid  11  and name hd*) 4.8 3.0 0.2
 assign (resid  11  and name hn) (resid  11  and name hg1*) 6.0 4.2 0.0 {3d #9}
 assign (resid  11  and name ha) (resid  12  and name hn) 2.4 0.6 0.3
 assign (resid  11  and name hg2*) (resid  12  and name hn) 6.5 4.7 0.0 {3d #14}
 assign (resid  11  and name hd*) (resid  38  and name ha*) 7.5 5.7 0.0
! assign (resid  11  and name hb) (resid  41  and name ha) 3.3 1.5 0.2 {removed rnd 20e}
 assign (resid  11  and name hg1*) (resid  41  and name ha) 6.0 4.2 0.0  {4d #431,432}
 assign (resid  11  and name hg1*) (resid  41  and name hg*) 8.9 7.1 0.0
 assign (resid  11  and name hg2*) (resid  61  and name hd*) 7.0 5.2 0.0  {4d #794}
 assign (resid  11  and name hd*) (resid  61  and name hd*) 7.5 5.7 0.0  {4d $812}
 assign (resid  11  and name hg1*) (resid  97  and name hb*) 7.0 5.2 0.0

! residue  12 lys-12
 assign (resid  12  and name ha) (resid  12  and name hg*) 4.3 2.5 0.2
 assign (resid  12  and name ha) (resid  13  and name hn) 3.3 1.5 0.2
 assign (resid  12  and name hb*) (resid  13  and name hn) 6.0 4.2 0.0
 assign (resid  12  and name hg*) (resid  13  and name hn) 6.0 4.2 0.0 {3d #20}

! residue  13 asp-13
 assign (resid  13  and name ha) (resid  14  and name hn) 5.0 3.2 0.0  {included rnd 18}
```

```
assign (resid  13  and name hn) (resid  14  and name hn) 3.3 1.5 0.2
assign (resid  13  and name hb*) (resid  14  and name hn) 6.0 4.2 0.0
assign (resid  13  and name ha) (resid  63  and name hb*) 4.3 2.5 0.2  {4d #1066}

! residue  14 thr-14
assign (resid  14  and name hn) (resid  15  and name hn) 3.3 1.5 0.2
assign (resid  14  and name ha) (resid  65  and name ha) 5.0 3.2 0.0  {4d #1026}
assign (resid  14  and name ha) (resid  68  and name hb*) 4.8 3.0 0.2 {round1}

! residue  15 asn-15
assign (resid  15  and name ha) (resid  16  and name hn) 2.4 0.6 0.3
assign (resid  15  and name hb*) (resid  16  and name hn) 6. 4.2 0.0
assign (resid  15  and name ha) (resid  61  and name ha) 3.3 1.5 0.2
assign (resid  15  and name ha) (resid  62  and name hn) 5.0 3.2 0.0
assign (resid  15  and name hb*) (resid  92  and name hb*) 7.5 5.7 0.0  {4d #1021, rnd
       20}

! residue  16 leu-16
assign (resid  16  and name ha) (resid  16  and name hd*) 4.8 3.0 0.2
assign (resid  16  and name hb*) (resid  16  and name hd*) 8.9 7.1 0.0
assign (resid  16  and name ha) (resid  17  and name hn) 2.4 0.6 0.3
assign (resid  16  and name hb*) (resid  17  and name hn) 6.0 4.2 0.0
assign (resid  16  and name hb*) (resid  18  and name hg*) 8.9 7.1 0.0
assign (resid  16  and name hd*) (resid  18  and name hg*) 10.3 8.5 0.0
assign (resid  16  and name hd1*) (resid  34  and name he*) 8.5 6.7 0.0
assign (resid  16  and name hd2*) (resid  34  and name he*) 8.5 6.7 0.0 {4d #399,#417;
       rnd 20d}
assign (resid  16  and name hn) (resid  60  and name hn) 5.0 3.2 0.0
assign (resid  16  and name hb*) (resid  60  and name hn) 6.0 4.2 0.0  {3d #341}
assign (resid  16  and name hb*) (resid  60  and name hb) 6.0 4.2 0.0
assign (resid  16  and name hd*) (resid  60  and name hg1*) 8.9 7.1 0.0  {4d #527; rnd
       24}
assign (resid  16  and name hn) (resid  61  and name ha) 5.0 3.2 0.0
assign (resid  16  and name hd*) (resid  62  and name hd*) 9.9 8.1 0.0 {4d #222, rnd
       20c}
assign (resid  16  and name hd*) (resid  62  and name he*) 7.3 5.5 0.3 {4d #460, rnd
       20c}
assign (resid  16  and name hd*) (resid  68  and name ha) 6.2 4.4 0.2
assign (resid  16  and name hd1*) (resid  68  and name hb*) 5.8 4.0 0.2
assign (resid  16  and name hd2*) (resid  68  and name hb*) 5.8 4.0 0.2 {4d #546,#795;
       rnd 20d}
assign (resid  16  and name hd*) (resid  71  and name hb*) 6.3 4.5 0.3
assign (resid  16  and name hb*) (resid  87  and name hd*) 8.9 7.1 0.0
assign (resid  16  and name hg) (resid  87  and name hb*) 6.0 4.2 0.0
assign (resid  16  and name hg) (resid  87  and name hd*) 6.2 4.4 0.2
assign (resid  16  and name ha) (resid  91  and name ha) 3.3 1.5 0.2
assign (resid  16  and name ha) (resid  92  and name hn) 5.0 3.2 0.0

! residue  17 tyr-17
assign (resid  17  and name hb*) (resid  17  and name hd*) 8.0 6.2 0.0
assign (resid  17  and name hd*) (resid  19  and name hg*) 8.5 6.7 0.0  {4d #188}
assign (resid  17  and name he*) (resid  57  and name hg*) 7.3 5.5 0.3
assign (resid  17  and name ha) (resid  59  and name ha) 3.3 1.5 0.2
```

```
! assign (resid  17  and name hn) (resid  90  and name hn) 5.0 3.2 0.0 {removed round 14}
 assign (resid  17  and name hb*) (resid  90  and name hd*) 7.0 5.2 0.0  {4d #862}
 assign (resid  17  and name hn) (resid  91  and name ha) 5.0 3.2 0.0
 assign (resid  17  and name hb*) (resid  92  and name ha) 4.3 2.5 0.2 {4d #70,71; rnd 20}
 assign (resid  17  and name hb*) (resid  92  and name hb*) 5.8 4.0 0.2 {4d #1020, rnd
        20}

! residue  18 val-18
 assign (resid  18  and name ha) (resid  18  and name hg1*) 4.8 3.0 0.2
 assign (resid  18  and name ha) (resid  18  and name hg2*) 4.8 3.0 0.2 {4d #909,981; rnd
        20d}
 assign (resid  18  and name hb) (resid  19  and name hn) 5.0 3.2 0.0
 assign (resid  18  and name hb) (resid  21  and name hd*) 7.9 6.1 0.0  {4d #556}
 assign (resid  18  and name hg*) (resid  21  and name hg) 6.2 4.4 0.2  {4d #512}
 assign (resid  18  and name hg*) (resid  21  and name hd*) 8.6 6.8 0.2  {4d #509}
 assign (resid  18  and name hg*) (resid  34  and name he*) 9.9 8.1 0.0
 assign (resid  18  and name hn) (resid  58  and name hn) 5.0 3.2 0.0
 assign (resid  18  and name hb) (resid  58  and name hb*) 6.5 4.7 0.0  {4d #813}
 assign (resid  18  and name hn) (resid  59  and name ha) 5.0 3.2 0.0
 assign (resid  18  and name hg*) (resid  60  and name hn) 7.9 6.1 0.0  {3d #343}
 assign (resid  18  and name hg*) (resid  60  and name hg2*) 7.2 5.4 0.2  {4d #940; rnd
        24}
 assign (resid  18  and name hb) (resid  87  and name hn) 5.0 3.2 0.0  {3d #560}
 assign (resid  18  and name ha) (resid  89  and name ha) 3.3 1.5 0.2
 assign (resid  18  and name hg*) (resid  89  and name ha) 7.9 6.1 0.0

! residue  19 thr-19
 assign (resid  19  and name ha) (resid  20  and name hn) 3.3 1.5 0.2
 assign (resid  19  and name hb) (resid  20  and name hn) 3.3 1.5 0.2
 assign (resid  19  and name ha) (resid  21  and name hn) 5.0 3.2 0.0 {3d#51}
 assign (resid  19  and name ha) (resid  57  and name ha) 2.4 0.6 0.3
 assign (resid  19  and name hb) (resid  57  and name hg*) 7.9 6.1 0.0  {4d #191}
 assign (resid  19  and name ha) (resid  58  and name hn) 5.0 3.2 0.0
 assign (resid  19  and name hn) (resid  88  and name hn) 5.0 3.2 0.0
! assign (resid  19  and name ha) (resid  89  and name ha) 3.3 1.5 0.2 {round 14}

! residue  20 asn-20
 assign (resid  20  and name ha) (resid  21  and name hn) 3.3 1.5 0.2
 assign (resid  20  and name hn) (resid  21  and name hn) 5.0 3.2 0.0
 assign (resid  20  and name hb*) (resid  87  and name hd*) 8.9 7.1 0.0  {4d #736}
 assign (resid  20  and name hb*) (resid  88  and name hn) 4.3 2.5 0.2  {3d #568}

! residue  21 leu-21
 assign (resid  21  and name ha) (resid  21  and name hg) 5.0 3.2 0.0
 assign (resid  21  and name hb*) (resid  21  and name hd1*) 7.5 5.7 0.0
 assign (resid  21  and name hb*) (resid  21  and name hd2*) 7.5 5.7 0.0 {4d #532,782;
        rnd 20d}
 assign (resid  21  and name ha) (resid  22  and name hd*) 4.3 2.5 0.2
 assign (resid  21  and name hb*) (resid  22  and name hd*) 7.0 5.2 0.0
 assign (resid  21  and name hd*) (resid  22  and name hd1) 6.2 4.4 0.2  {4d #288,340}
 assign (resid  21  and name hd*) (resid  22  and name hd2) 6.2 4.4 0.2  {4d #288,340;
        rnd 20d}
 assign (resid  21  and name hd*) (resid  25  and name hb) 7.9 6.1 0.0
```

```
 assign (resid  21  and name hd*) (resid  25  and name hg1*) 8.9 7.1 0.0
 assign (resid  21  and name hd*) (resid  25  and name hd*) 8.9 7.1 0.0
 assign (resid  21  and name hd*) (resid  30  and name hd*) 8.6 6.8 0.2
 assign (resid  21  and name hd1*) (resid  57  and name ha) 6.5 4.7 0.0  {4d #539}
 assign (resid  21  and name hd2*) (resid  57  and name ha) 6.5 4.7 0.0  {4d #539,833; rnd
        20d}
 assign (resid  21  and name hb*) (resid  58  and name hb*) 7.5 5.7 0.0 {4d #868}
 assign (resid  21  and name hd*) (resid  58  and name hn) 7.9 6.1 0.0  {3d #328}
 assign (resid  21  and name hd*) (resid  58  and name hb*) 6.3 4.5 0.3  {4d #416}

! residue  22 pro-22
 assign (resid  22  and name ha) (resid  23  and name hn) 2.4 0.6 0.3
 assign (resid  22  and name ha) (resid  23  and name hg*) 6.0 4.2 0.0
 assign (resid  22  and name hb*) (resid  23  and name hn) 6.0 4.2 0.0
 assign (resid  22  and name hb*) (resid  24  and name hn) 6.0 4.2 0.0 {3d #70}
 assign (resid  22  and name hg*) (resid  24  and name hn) 6.0 4.2 0.0 {3d #71}
 assign (resid  22  and name hd*) (resid  25  and name hd*) 7.5 5.7 0.0
! assign (resid  22  and name ha) (resid  54  and name hg*) 4.3 2.5 0.2
! assign (resid  22  and name hb*) (resid  54  and name hg*) 5.3 3.5 0.2

! residue  23 arg-23
 assign (resid  23  and name ha) (resid  24  and name hn) 5.0 3.2 0.0
 assign (resid  23  and name hn) (resid  24  and name hn) 3.3 1.5 0.2
 assign (resid  23  and name hb*) (resid  45  and name hg2*) 7.5 5.7 0.0
 assign (resid  23  and name hg*) (resid  45  and name hd*) 5.8 4.0 0.2  {4d #468}
! assign (resid  23  and name hb*) (resid  70  and name hb*) 4.4 2.6 0.3 {round 8}

! residue  24 thr-24
 assign (resid  24  and name ha) (resid  24  and name hg*) 3.9 2.1 0.3
 assign (resid  24  and name ha) (resid  25  and name hn) 5.0 3.2 0.0  {med to weak rnd
        18}
 assign (resid  24  and name hn) (resid  25  and name hn) 3.3 1.5 0.2
 assign (resid  24  and name hg*) (resid  25  and name hn) 6.5 4.7 0.0 {3d #83}

! residue  25 ile-25
 assign (resid  25  and name ha) (resid  25  and name hg1*) 6.0 4.2 0.0
 assign (resid  25  and name hb) (resid  25  and name hd*) 4.8 3.0 0.2
 assign (resid  25  and name ha) (resid  26  and name hn) 3.3 1.5 0.2  {str to med rnd 18}
 assign (resid  25  and name hb) (resid  26  and name hn) 3.3 1.5 0.2
 assign (resid  25  and name hg1*) (resid  26  and name hn) 6.0 4.2 0.0  {3d #89,90}
 assign (resid  25  and name ha) (resid  29  and name hb*) 6.0 4.2 0.0
 assign (resid  25  and name hd*) (resid  29  and name hb*) 5.8 4.0 0.2
 assign (resid  25  and name hg1*) (resid  30  and name hd*) 7.5 5.7 0.0
 assign (resid  25  and name hd*) (resid  30  and name hg) 6.5 4.7 0.0
 assign (resid  25  and name hd*) (resid  30  and name hd*) 7.2 5.4 0.2
 assign (resid  25  and name hg1*) (resid  45  and name hd*) 7.5 5.7 0.0
! assign (resid  25  and name hd*) (resid  54  and name hg*) 7.5 5.7 0.0

! residue  26 thr-26
 assign (resid  26  and name ha) (resid  27  and name hn) 2.4 0.6 0.3
 assign (resid  26  and name hg*) (resid  27  and name hn) 4.8 3.0 0.2  {3d #96} {included
        rnd 18}
 assign (resid  26  and name hg*) (resid  27  and name hn) 4.8 3.0 0.2
```

```
! assign (resid  26  and name ha) (resid  28  and name hn) 3.3 1.5 0.2  {3d #100}
        {removed rnd 18}
 assign (resid  26  and name ha) (resid  28  and name hn) 5.0 3.2 0.0  {3d #100} {included
        rnd 18}
 assign (resid  26  and name hb) (resid  28  and name hn) 5.0 3.2 0.0  {3d #100} {included
        rnd 18}
 assign (resid  26  and name hg*) (resid  28  and name hn) 6.5 4.7 0.0  {3d #103}
        {included rnd 18}
 assign (resid  26  and name hn) (resid  29  and name hb*) 4.3 2.5 0.2  {3d #87}

! residue  27 asp-27
 assign (resid  27  and name ha) (resid  28  and name hn) 5.0 3.2 0.0
 assign (resid  27  and name hn) (resid  28  and name hn) 3.3 1.5 0.2
 assign (resid  27  and name hb*) (resid  45  and name hd*) 7.5 5.7 0.0  {4d #1069}
! assign (resid  27  and name hn) (resid  60  and name hb) 3.3 1.5 0.2  {3d #96} {removed
        rnd 18}

! residue  28 asp-28
 assign (resid  28  and name ha) (resid  29  and name hn) 5.0 3.2 0.0
 assign (resid  28  and name hn) (resid  29  and name hn) 3.3 1.5 0.2
 assign (resid  28  and name hb*) (resid  29  and name hn) 6.0 4.2 0.0
 assign (resid  28  and name ha) (resid  31  and name hb*) 4.3 2.5 0.2
! assign (resid  28  and name hn) (resid  60  and name hb) 5.0 3.2 0.0  {3d #103}
        {removed rnd 18}

! residue  29 gln-29
 assign (resid  29  and name hn) (resid  30  and name hn) 3.3 1.5 0.2
 assign (resid  29  and name hb*) (resid  30  and name hn) 6.0 4.2 0.0
 assign (resid  29  and name ha) (resid  32  and name hn) 3.3 1.5 0.2
 assign (resid  29  and name ha) (resid  32  and name hb) 3.3 1.5 0.2  {4d #447}
 assign (resid  29  and name ha) (resid  32  and name hg*) 6.5 4.7 0.0  {4d #917}
 assign (resid  29  and name ha) (resid  43  and name hd*) 6.0 4.2 0.0

! residue  30 leu-30
 assign (resid  30  and name hb1) (resid  30  and name hd1*) 6.5 4.7 0.0
 assign (resid  30  and name hb2) (resid  30  and name hd1*) 6.5 4.7 0.0
 assign (resid  30  and name hb1) (resid  30  and name hd2*) 6.5 4.7 0.0
 assign (resid  30  and name hb2) (resid  30  and name hd2*) 6.5 4.7 0.0 {4d
        #715,572,712,571; rnd 20d}
 assign (resid  30  and name hn) (resid  31  and name hn) 3.3 1.5 0.2
 assign (resid  30  and name hd*) (resid  31  and name hn) 7.9 6.1 0.0  {3d #127}
 assign (resid  30  and name hb*) (resid  31  and name hn) 4.3 2.5 0.2
 assign (resid  30  and name ha) (resid  33  and name hn) 3.3 1.5 0.2
 assign (resid  30  and name ha) (resid  33  and name hb) 5.0 3.2 0.0
 assign (resid  30  and name ha) (resid  33  and name hd*) 6.5 4.7 0.0  {4d #1117}
 assign (resid  30  and name ha) (resid  34  and name hn) 5.0 3.2 0.0
 assign (resid  30  and name ha) (resid  34  and name hd*) 5.3 3.5 0.2  {4d #264}
 assign (resid  30  and name hd*) (resid  34  and name hd*) 8.2 6.4 0.2
 assign (resid  30  and name hd*) (resid  34  and name he*) 9.9 8.1 0.0
 assign (resid  30  and name hd*) (resid  43  and name hb1) 7.9 6.1 0.0
 assign (resid  30  and name hd*) (resid  43  and name hb2) 7.9 6.1 0.0 {4d #747,748; rnd
        20d}
 assign (resid  30  and name hd1*) (resid  43  and name hd*) 5.8 4.0 0.2  {4d #224,229}
```

```
 assign (resid  30  and name hd2*) (resid  43  and name hd*) 5.8 4.0 0.2  {4d #224,229;
        rnd 20d}
 assign (resid  30  and name hd*) (resid  60  and name hg2*) 8.9 7.1 0.0  {4d #720; rnd
        24}
 assign (resid  30  and name hg) (resid  78  and name hb) 3.3 1.5 0.2

! residue  31 asp-31
 assign (resid  31  and name hn) (resid  32  and name hn) 5.0 3.2 0.0
 assign (resid  31  and name hb*) (resid  32  and name hn) 4.3 2.5 0.2
 assign (resid  31  and name ha) (resid  34  and name hn) 5.0 3.2 0.0  {1H-1H; included
        rnd 20b}
 assign (resid  31  and name ha) (resid  40  and name hd*) 4.8 3.0 0.2

! residue  32 thr-32
 assign (resid  32  and name ha) (resid  32  and name hg*) 4.8 3.0 0.2
 assign (resid  32  and name hn) (resid  32  and name hg*) 4.8 3.0 0.2  {3d #135}
 assign (resid  32  and name ha) (resid  33  and name hn) 5.0 3.2 0.0
 assign (resid  32  and name hb) (resid  33  and name hn) 3.3 1.5 0.2  {3d #137}
 assign (resid  32  and name hg*) (resid  33  and name ha) 6.5 4.7 0.0  {4d #772}
 assign (resid  32  and name hg*) (resid  33  and name hn) 6.5 4.7 0.0
 assign (resid  32  and name ha) (resid  35  and name hn) 5.0 3.2 0.0  {1H-1H; included
        rnd 20b}
 assign (resid  32  and name hg*) (resid  35  and name ha*) 7.5 5.7 0.0

! residue  33 ile-33
 assign (resid  33  and name hg1*) (resid  33  and name hg2*) 5.8 4.0 0.2
 assign (resid  33  and name ha) (resid  34  and name hn) 5.0 3.2 0.0
 assign (resid  33  and name hn) (resid  34  and name hn) 3.3 1.5 0.2
 assign (resid  33  and name hb) (resid  34  and name hd*) 7.0 5.2 0.0
 assign (resid  33  and name hg2*) (resid  34  and name hn) 6.5 4.7 0.0  {3d #156}
 assign (resid  33  and name hg2*) (resid  75  and name hd1*) 5.8 4.0 0.2
 assign (resid  33  and name hg2*) (resid  75  and name hd2*) 5.8 4.0 0.2 {4d #540,879;
        rnd 20d}
 assign (resid  33  and name hg2*) (resid  87  and name hd1*) 7.5 5.7 0.0
 assign (resid  33  and name hg2*) (resid  87  and name hd2*) 7.5 5.7 0.0 {4d #586,792;
        rnd 20d}

! residue  34 phe-34
 assign (resid  34  and name ha) (resid  34  and name hd*) 7.0 5.2 0.0
 assign (resid  34  and name hb*) (resid  34  and name hd*) 8.0 6.2 0.0
 assign (resid  34  and name hn) (resid  35  and name hn) 3.3 1.5 0.2
 assign (resid  34  and name hb*) (resid  35  and name hn) 6.0 4.2 0.0
 assign (resid  34  and name ha) (resid  37  and name hd*) 7.0 5.2 0.0  {4d # 246, rnd
        20c}
 assign (resid  34  and name hb*) (resid  37  and name hh) 6.0 4.2 0.0  {included rd. 17}
 assign (resid  34  and name hb*) (resid  40  and name hd*) 7.5 5.7 0.0  {4d #1110}
 assign (resid  34  and name hb1) (resid  60  and name hg1*) 6.5 4.7 0.0  {4d #998,999}
 assign (resid  34  and name hb2) (resid  60  and name hg1*) 6.5 4.7 0.0  {4d #998,999;
        rnd 20d}
 assign (resid  34  and name hd*) (resid  62  and name he*) 9.0 7.2 0.0  {4d #269, rnd
        20c}
 assign (resid  34  and name hd*) (resid  71  and name hb*) 6.8 5.0 0.2
 assign (resid  34  and name ha) (resid  75  and name hd*) 7.9 6.1 0.0
```

```
 assign (resid  34  and name hd*) (resid  75  and name hd*) 9.9 8.1 0.0

! residue  35 gly-35
 assign (resid  35  and name ha*) (resid  36  and name hn) 5.0 3.2 0.0
 assign (resid  35  and name hn) (resid  36  and name hn) 3.3 1.5 0.2
 assign (resid  35  and name ha*) (resid  40  and name hd*) 7.5 5.7 0.0
 assign (resid  35  and name hn) (resid  40  and name hd*) 6.5 4.7 0.0

! residue  36 lys-36
 assign (resid  36  and name ha) (resid  37  and name hn) 5.0 3.2 0.0
 assign (resid  36  and name hn) (resid  37  and name hn) 3.3 1.5 0.2
 assign (resid  36  and name hb*) (resid  37  and name hn) 6.0 4.2 0.0
 assign (resid  36  and name hg*) (resid  37  and name he*) 8.0 6.2 0.0 {4d #798, rnd
        20c}
 assign (resid  36  and name hd*) (resid  37  and name he*) 8.0 6.2 0.0 {4d #382, rnd
        20c}

! residue  37 tyr-37
 assign (resid  37  and name hn) (resid  37  and name hh) 5.0 3.2 0.0  {included rnd 20b}
 assign (resid  37  and name ha) (resid  37  and name hd*) 5.3 3.5 0.2 {4d #202, rnd 20c}
 assign (resid  37  and name hb*) (resid  37  and name hd*) 8.0 6.2 0.0 {4d #253, rnd
        20c}
 assign (resid  37  and name ha) (resid  38  and name hn) 3.3 1.5 0.2
 assign (resid  37  and name hn) (resid  38  and name hn) 3.3 1.5 0.0
 assign (resid  37  and name hb*) (resid  38  and name hn) 6.0 4.2 0.0
 assign (resid  37  and name hh) (resid  40  and name hd*) 6.5 4.7 0.0  {included rnd 20b}
 assign (resid  37  and name hd*) (resid  70  and name hg*) 8.0 6.2 0.0 {4d #268, rnd20c}
 assign (resid  37  and name hb*) (resid  71  and name hb*) 7.5 5.7 0.0
 assign (resid  37  and name hd*) (resid  71  and name hb*) 8.5 6.7 0.0 {4d #250, rnd
        20c}
 assign (resid  37  and name he*) (resid  71  and name hb*) 8.5 6.7 0.0 {4d #1083, rnd
        20c}
 assign (resid  37  and name he*) (resid  74  and name hb*) 6.8 5.0 0.2 {4d #1001, rnd
        20c}
 assign (resid  37  and name hd*) (resid  75  and name hd*) 9.9 8.1 0.0  {4d #223, rnd
        20c}
 assign (resid  37  and name he*) (resid  75  and name hd1*) 6.8 5.0 0.2 {4d #889, rnd
        20c}
 assign (resid  37  and name he*) (resid  75  and name hd2*) 6.8 5.0 0.2 {4d #458, rnd
        20c}
 assign (resid  37  and name hh) (resid  75  and name hd*) 7.0 5.2 0.0  {included rnd 20}

! residue  38 gly-38
 assign (resid  38  and name ha*) (resid  39  and name hn) 3.4 1.6 0.3
 assign (resid  38  and name hn) (resid  39  and name hn) 5.0 3.2 0.0
 assign (resid  38  and name ha*) (resid  41  and name hg*) 8.9 7.1 0.0
 assign (resid  38  and name hn) (resid  62  and name he*) 5.3 3.5 0.2 {3d #181 included
        rnd 20e}
! assign (resid  38  and name ha*) (resid  67  and name hg*) 7.0 5.2 0.0  {included rnd 20,
        removed rnd 21}

! residue  39 ser-39
 assign (resid  39  and name ha) (resid  40  and name hn) 2.4 0.6 0.3
```

```
assign (resid  39  and name ha) (resid  40  and name hd*) 6.5 4.7 0.0  {4d #1172}
assign (resid  39  and name hn) (resid  40  and name hn) 5.0 3.2 0.0
assign (resid  39  and name hb*) (resid  40  and name hn) 6.0 4.2 0.0

! residue  40 ile-40
assign (resid  40  and name ha) (resid  40  and name hg2*) 3.9 2.1 0.3
assign (resid  40  and name ha) (resid  40  and name hd*) 6.5 4.7 0.0
assign (resid  40  and name ha) (resid  41  and name hn) 2.4 0.6 0.3
assign (resid  40  and name hg2*) (resid  41  and name hn) 6.5 4.7 0.0  {3d #207}
assign (resid  40  and name hg2*) (resid  41  and name ha) 6.5 4.7 0.0
assign (resid  40  and name hg2*) (resid  41  and name hg*) 8.9 7.1 0.0  {4d #994}
assign (resid  40  and name ha) (resid  42  and name hn) 5.0 3.2 0.0  {3d #212}
assign (resid  40  and name hb) (resid  42  and name hn) 3.3 1.5 0.2  {3d #213}
assign (resid  40  and name hg2*) (resid  42  and name hn) 4.8 3.0 0.2  {3d #216}
assign (resid  40  and name hg2*) (resid  43  and name hb*) 7.5 5.7 0.0
assign (resid  40  and name hg2*) (resid  43  and name hg*) 5.8 4.0 0.2
assign (resid  40  and name hg2*) (resid  43  and name hd*) 5.8 4.0 0.2
assign (resid  40  and name hd*) (resid  43  and name hg*) 7.5 5.7 0.0  {4d #852}
assign (resid  40  and name hg2*) (resid  60  and name hg1*) 4.9 3.1 0.3  {4d #957; rmd
      24}
assign (resid  40  and name hg2*) (resid  60  and name hg2*) 7.5 5.7 0.0  {4d #1031; rmd
      24}
assign (resid  40  and name hd*) (resid  60  and name hg*) 7.2 5.4 0.2  {4d #1011}
assign (resid  40  and name hg2*) (resid  61  and name hn) 6.5 4.7 0.0  {3d #352}
assign (resid  40  and name ha) (resid  62  and name ha) 3.3 1.5 0.2
assign (resid  40  and name hg2*) (resid  62  and name ha) 6.5 4.7 0.0
assign (resid  40  and name ha) (resid  63  and name hn) 5.0 3.2 0.0  {3d #370}
assign (resid  40  and name ha) (resid  75  and name hd*) 7.9 6.1 0.0
assign (resid  40  and name hb) (resid  75  and name hd*) 7.9 6.1 0.0
assign (resid  40  and name hg2*) (resid  75  and name hd*) 7.2 5.4 0.2
assign (resid  40  and name hd*) (resid  75  and name hd*) 7.2 5.4 0.2  {4d #467}
! assign (resid  40  and name hb) (resid  79  and name hg2*) 3.9 2.1 0.3 {round 8}

! residue  41 val-41
assign (resid  41  and name ha) (resid  41  and name hg*) 6.2 4.4 0.2
assign (resid  41  and name hn) (resid  41  and name hg*) 6.2 4.4 0.2  {3d #206}
assign (resid  41  and name ha) (resid  42  and name hn) 5.0 3.2 0.0
assign (resid  41  and name hn) (resid  42  and name hn) 3.3 1.5 0.2
assign (resid  41  and name hg*) (resid  42  and name hn) 6.2 4.4 0.2
assign (resid  41  and name hg*) (resid  61  and name hd*) 7.2 5.4 0.2
assign (resid  41  and name hn) (resid  62  and name ha) 5.0 3.2 0.0  {3d #200}
assign (resid  41  and name hg*) (resid  62  and name ha) 7.9 6.1 0.0
assign (resid  41  and name hn) (resid  63  and name hn) 5.0 3.2 0.0
assign (resid  41  and name hg*) (resid  63  and name hn) 7.9 6.1 0.0
assign (resid  41  and name hg1*) (resid  63  and name ha) 4.8 3.0 0.2
assign (resid  41  and name hg2*) (resid  63  and name ha) 4.8 3.0 0.2 {4d #967,1044;
      rmd 20d}
assign (resid  41  and name hg*) (resid  63  and name hb*) 7.2 5.4 0.2

! residue  42 gln-42
assign (resid  42  and name ha) (resid  43  and name hn) 2.4 0.6 0.3
assign (resid  42  and name ha) (resid  43  and name hg*) 6.0 4.2 0.0  {4d #841}
assign (resid  42  and name hb*) (resid  43  and name hn) 6.0 4.2 0.0
```

```
 assign (resid  42  and name hg*) (resid  43  and name hn) 6.0 4.2 0.0
 assign (resid  42  and name hn) (resid  61  and name hn) 5.0 3.2 0.0

! residue  43 lys-43
 assign (resid  43  and name ha) (resid  44  and name hn) 2.4 0.6 0.3
 assign (resid  43  and name hb*) (resid  44  and name hn) 6.0 4.2 0.0
 assign (resid  43  and name ha) (resid  60  and name ha) 2.4 0.6 0.3
 assign (resid  43  and name ha) (resid  60  and name hg1*) 6.5 4.7 0.0
 assign (resid  43  and name ha) (resid  60  and name hg2*) 6.5 4.7 0.0 {4d #1051,1050
        rnd 20d}
 assign (resid  43  and name hb*) (resid  60  and name hg2*) 7.5 5.7 0.0  {4d #1049}
 assign (resid  43  and name hg*) (resid  60  and name hg1*) 7.5 5.7 0.0  {4d #824}
 assign (resid  43  and name hg*) (resid  60  and name hg2*) 7.5 5.7 0.0  {4d #867}
 assign (resid  43  and name ha) (resid  61  and name hn) 5.0 3.2 0.0

! residue  44 asn-44
 assign (resid  44  and name ha) (resid  45  and name hn) 2.4 0.6 0.3
 assign (resid  44  and name hn) (resid  59  and name hn) 5.0 3.2 0.0
 assign (resid  44  and name hn) (resid  60  and name ha) 5.0 3.2 0.0

! residue  45 ile-45
 assign (resid  45  and name ha) (resid  45  and name hg2*) 6.0 4.2 0.0
 assign (resid  45  and name hg2*) (resid  45  and name hg1*) 5.8 4.0 0.2
 assign (resid  45  and name ha) (resid  46  and name hn) 2.4 0.6 0.3
 assign (resid  45  and name hg2*) (resid  46  and name hn) 4.8 3.0 0.2  {3d #248}
 assign (resid  45  and name ha) (resid  59  and name hn) 5.0 3.2 0.0
 assign (resid  45  and name ha) (resid  58  and name ha) 3.3 1.5 0.2
 assign (resid  45  and name hg1*) (resid  58  and name hb*) 7.5 5.7 0.0  {4d #183,290}
! assign (resid  45  and name hb) (resid  69  and name ha) 3.3 1.5 0.2 {round 7}
! assign (resid  45  and name hb) (resid  72  and name ha) 5.0 3.2 0.0 {round 7}

! residue  46 leu-46
 assign (resid  46  and name ha) (resid  46  and name hd*) 6.2 4.4 0.2
 assign (resid  46  and name hb*) (resid  46  and name hd1*) 5.8 4.0 0.2
 assign (resid  46  and name hb*) (resid  46  and name hd2*) 5.8 4.0 0.2 {4d #665,829
        rnd 20d}
 assign (resid  46  and name ha) (resid  47  and name hn) 2.4 0.6 0.3
 assign (resid  46  and name hd*) (resid  47  and name hn) 7.9 6.1 0.0  {3d #251}
 assign (resid  46  and name hd1*) (resid  57  and name hg*) 7.2 5.4 0.2  {4d #658}
 assign (resid  46  and name hd2*) (resid  57  and name hg*) 7.2 5.4 0.2  {4d #658,666
        rnd 20d}
 assign (resid  46  and name hn) (resid  58  and name ha) 3.3 1.5 0.2
 assign (resid  46  and name hg) (resid  58  and name ha) 5.0 3.2 0.0  {4d #220}
 assign (resid  46  and name hd*) (resid  59  and name hd*) 8.2 6.4 0.2
 assign (resid  46  and name hd*) (resid  59  and name he*) 9.9 8.1 0.0

! residue  47 arg-47

! residue  48 asp-48
 assign (resid  48  and name ha) (resid  49  and name hn) 2.4 0.6 0.3
 assign (resid  48  and name hb*) (resid  49  and name hn) 6.0 4.2 0.0
 assign (resid  48  and name hb*) (resid  51  and name hn) 6.0 4.2 0.0  {3d #276}
 assign (resid  48  and name hb*) (resid  52  and name hn) 6.0 4.2 0.0  {3d #285}
```

```
assign (resid  48  and name ha) (resid  53  and name hb*) 6.0 4.2 0.0  {4d #14}
assign (resid  48  and name hb*) (resid  53  and name hn) 4.3 2.5 0.2

! residue  49 lys-49
assign (resid  49  and name ha) (resid  50  and name hn) 5.0 3.2 0.0
assign (resid  49  and name hn) (resid  50  and name hn) 3.3 1.5 0.2

! residue  50 leu-50
! assign (resid  50  and name ha) (resid  51  and name hn) 5.0 3.2 0.0 {round 14}
assign (resid  50  and name hn) (resid  51  and name hn) 3.3 1.5 0.2
assign (resid  50  and name hn) (resid  51  and name hg*) 6.5 4.7 0.0  {3d #272}
assign (resid  50  and name hb*) (resid  51  and name hn) 4.3 2.5 0.2
! assign (resid  50  and name hb*) (resid  91  and name hd*) 7.2 5.4 0.2 {removed rnd
      20d}

! residue  51 thr-51
assign (resid  51  and name ha) (resid  52  and name hn) 5.0 3.2 0.0
assign (resid  51  and name hg*) (resid  52  and name hn) 6.5 4.7 0.0  {3d #287}
assign (resid  51  and name hb) (resid  53  and name hg*) 6.0 4.2 0.0  {4d #170}

! residue  52 gly-52
assign (resid  52  and name ha*) (resid  53  and name hn) 6.0 4.2 0.0
assign (resid  52  and name hn) (resid  53  and name hn) 3.3 1.5 0.2

! residue  53 arg-53

! residue  54 pro-54
assign (resid  54  and name ha) (resid  54  and name hg*) 6.0 4.2 0.0
assign (resid  54  and name hb*) (resid  55  and name hn) 6.0 4.2 0.0

! residue  55 arg-55
assign (resid  55  and name hn) (resid  56  and name hn) 3.3 1.5 0.2
assign (resid  55  and name hn) (resid  57  and name hg*) 7.9 6.1 0.0  {3d #306}
assign (resid  55  and name hb1) (resid  57  and name hg*) 7.9 6.1 0.0  {4d #126}
assign (resid  55  and name hb2) (resid  57  and name hg*) 7.9 6.1 0.0  {4d #126,77; rnd
      20d}

! residue  56 gly-56
assign (resid  56  and name ha*) (resid  57  and name hn) 5.0 3.2 0.0
assign (resid  56  and name hn) (resid  57  and name hn) 3.3 1.5 0.2
assign (resid  56  and name hn) (resid  57  and name hg*) 7.9 6.1 0.0  {3d #312}

! residue  57 val-57
assign (resid  57  and name hb) (resid  58  and name hn) 5.0 3.2 0.0
assign (resid  57  and name hg*) (resid  58  and name hn) 7.9 6.1 0.0

! residue  58 ala-58
assign (resid  58  and name ha) (resid  59  and name hn) 3.3 1.5 0.2
assign (resid  58  and name hb*) (resid  59  and name hn) 4.8 3.0 0.2
assign (resid  58  and name hb*) (resid  60  and name hb) 6.5 4.7 0.0  {4d #806}
assign (resid  58  and name hb*) (resid  60  and name hg1*) 7.5 5.7 0.0  {4d #780}
assign (resid  58  and name hb*) (resid  60  and name hg2*) 5.8 4.0 0.2  {4d #778}
```

```
! residue  59 phe-59
 assign (resid  59  and name ha) (resid  59  and name hd*) 7.0 5.2 0.0
 assign (resid  59  and name hn) (resid  59  and name hd*) 7.0 5.2 0.0  {3d #330}
 assign (resid  59  and name hb*) (resid  59  and name hd*) 8.0 6.2 0.0
 assign (resid  59  and name ha) (resid  60  and name hn) 2.4 0.6 0.3
 assign (resid  59  and name hb*) (resid  60  and name hn) 6.0 4.2 0.0
 assign (resid  59  and name hd*) (resid  92  and name hb*) 8.5 6.7 0.0  {4d #248, rnd
        20}

! residue  60 val-60
 assign (resid  60  and name ha) (resid  61  and name hn) 2.4 0.6 0.3
 assign (resid  60  and name hg1*) (resid  61  and name hn) 6.5 4.7 0.0  {3d #353}
 assign (resid  60  and name hg1*) (resid  62  and name he*) 8.5 6.7 0.0 {4d #1008, rnd
        20c}
! assign (resid  60  and name hg*) (resid  70  and name hb*) 8.9 7.1 0.0  {removed rnd
        24}

! residue  61 arg-61
 assign (resid  61  and name ha) (resid  62  and name hn) 3.3 1.5 0.2
 assign (resid  61  and name ha) (resid  97  and name hb*) 6.0 4.2 0.0
 assign (resid  61  and name hd*) (resid  97  and name hb*) 7.0 5.2 0.0

! residue  62 tyr-62
 assign (resid  62  and name hd*) (resid  62  and name hb1) 7.0 5.2 0.0 {4d #236, rnd
        20c}
 assign (resid  62  and name hd*) (resid  62  and name hb2) 7.0 5.2 0.0 {4d #239, rnd
        20c}
 assign (resid  62  and name ha) (resid  63  and name hn) 2.4 0.6 0.3
 assign (resid  62  and name hb*) (resid  63  and name hn) 4.3 2.5 0.2
 assign (resid  62  and name hd*) (resid  68  and name hb*) 8.5 6.7 0.0 {4d #196, rnd
        20c}
 assign (resid  62  and name hd*) (resid  68  and name ha) 7.0 5.2 0.0 {4d #271, rnd 20c}
 assign (resid  62  and name he*) (resid  68  and name ha) 7.0 5.2 0.0 {4d #1178, rnd
        20c}
 assign (resid  62  and name he*) (resid  71  and name hb*) 6.8 5.0 0.2 {4d #1062, rnd
        20c}
 assign (resid  62  and name he*) (resid  72  and name hg1*) 8.0 6.2 0.0 {4d #551, rnd
        20c}
 assign (resid  62  and name he*) (resid  75  and name hd*) 9.9 8.1 0.0 {4d #592, rnd20c}

! residue  63 asn-63
 assign (resid  63  and name ha) (resid  64  and name hn) 5.0 3.2 0.0
 assign (resid  63  and name hn) (resid  64  and name hn) 3.3 1.5 0.2
 assign (resid  63  and name hb*) (resid  64  and name hn) 6.0 4.2 0.0
 assign (resid  63  and name hb*) (resid  65  and name hd*) 7.0 5.2 0.0  {4d #826}
 assign (resid  63  and name hb*) (resid  91  and name hd*) 8.9 7.1 0.0

! residue  64 lys-64
 assign (resid  64  and name ha) (resid  65  and name hn) 2.4 0.6 0.3
 assign (resid  64  and name hn) (resid  65  and name hn) 5.0 3.2 0.0
 assign (resid  64  and name hb*) (resid  65  and name hn) 4.3 2.5 0.2  {3d #390,393}
 assign (resid  64  and name ha) (resid  66  and name hn) 5.0 3.2 0.0  {3d #397}
 assign (resid  64  and name hb*) (resid  66  and name hn) 4.3 2.5 0.2  {3d #400,403}
```

```
assign (resid  64  and name hb*) (resid  67  and name hn) 4.3 2.5 0.2  {3d #411,413}
assign (resid  64  and name hg*) (resid  67  and name hn) 4.3 2.5 0.2  {3d #414}
assign (resid  64  and name hb*) (resid  68  and name hn) 6.0 4.2 0.0  {3d #421}
! assign (resid  64  and name ha) (resid  87  and name hg) 5.0 3.2 0.0 {round1}

! residue  65 arg-65
assign (resid  65  and name hn) (resid  66  and name hn) 3.3 1.5 0.2
assign (resid  65  and name ha) (resid  68  and name hn) 3.3 1.5 0.2  {3d #417}
assign (resid  65  and name ha) (resid  68  and name hb*) 4.8 3.0 0.2  {4d #1075}

! residue  66 glu-66
assign (resid  66  and name ha) (resid  66  and name hg*) 6.0 4.2 0.0
assign (resid  66  and name ha) (resid  67  and name hn) 5.0 3.2 0.0  {included rnd 18}
assign (resid  66  and name hn) (resid  67  and name hn) 3.3 1.5 0.2
assign (resid  66  and name ha) (resid  69  and name hn) 5.0 3.2 0.0
assign (resid  66  and name ha) (resid  69  and name hb*) 3.4 1.6 0.3
assign (resid  66  and name ha) (resid  69  and name hg*) 6.0 4.2 0.0
assign (resid  66  and name hg*) (resid  69  and name hb*) 5.3 3.5 0.2

! residue  67 glu-67
assign (resid  67  and name hn) (resid  68  and name hn) 3.3 1.5 0.2

! residue  68 ala-68
assign (resid  68  and name hn) (resid  69  and name hn) 3.3 1.5 0.2
assign (resid  68  and name hb*) (resid  69  and name hn) 3.9 2.1 0.3
assign (resid  68  and name hb*) (resid  70  and name hn) 6.5 4.7 0.0  {3d #437}
assign (resid  68  and name ha) (resid  71  and name hn) 5.0 3.2 0.0
assign (resid  68  and name ha) (resid  71  and name hb*) 4.8 3.0 0.2
assign (resid  68  and name hb*) (resid  72  and name hd*) 7.5 5.7 0.0
assign (resid  68  and name hb*) (resid  91  and name hd*) 8.9 7.1 0.0  {4d #881}

! residue  69 gln-69
! assign (resid  69  and name ha) (resid  70  and name hb*) 3.3 1.5 0.2  {4d
      #386,removed, rd. 17}
assign (resid  69  and name hn) (resid  70  and name hn) 3.3 1.5 0.2
assign (resid  69  and name hb*) (resid  70  and name hn) 4.3 2.5 0.2
assign (resid  69  and name ha) (resid  72  and name hn) 5.0 3.2 0.0
assign (resid  69  and name ha) (resid  72  and name hd*) 4.8 3.0 0.2  {4d #1142}
assign (resid  69  and name hn) (resid  72  and name hd*) 6.5 4.7 0.0  {3d #430}

! residue  70 glu-70
assign (resid  70  and name hn) (resid  71  and name hn) 3.3 1.5 0.2
assign (resid  70  and name hb*) (resid  71  and name hn) 4.3 2.5 0.2
assign (resid  70  and name ha) (resid  73  and name hb*) 4.3 2.5 0.2
assign (resid  70  and name hg*) (resid  74  and name hb*) 7.5 5.7 0.0

! residue  71 ala-71
assign (resid  71  and name ha) (resid  72  and name hn) 5.0 3.2 0.0
assign (resid  71  and name hb*) (resid  72  and name hn) 4.8 3.0 0.2
assign (resid  71  and name ha) (resid  74  and name hn) 5.0 3.2 0.0
assign (resid  71  and name ha) (resid  74  and name hb*) 4.8 3.0 0.2
assign (resid  71  and name hb*) (resid  74  and name hb*) 7.5 5.7 0.0
```

```
assign (resid 71 and name ha) (resid 75 and name hn) 5.0 3.2 0.0 {3d, included rnd
      20b}
assign (resid 71 and name hb*) (resid 75 and name hd*) 7.2 5.4 0.2

! residue 72 ile-72
assign (resid 72 and name ha) (resid 73 and name hn) 5.0 3.2 0.0
assign (resid 72 and name hn) (resid 73 and name hn) 3.3 1.5 0.2
assign (resid 72 and name hg2*) (resid 73 and name hn) 6.5 4.7 0.0 {3d #464}
assign (resid 72 and name hg2*) (resid 73 and name ha) 6.0 4.2 0.0 {4d #1040}
assign (resid 72 and name hd*) (resid 73 and name hn) 6.5 4.7 0.0 {3d #465}
assign (resid 72 and name ha) (resid 76 and name hn) 5.0 3.2 0.0
assign (resid 72 and name hg2*) (resid 76 and name hb*) 7.5 5.7 0.0
assign (resid 72 and name ha) (resid 89 and name hb) 5.0 3.2 0.0
assign (resid 72 and name ha) (resid 89 and name hg1*) 4.8 3.0 0.2 {4d #608}
assign (resid 72 and name ha) (resid 89 and name hg2*) 6.5 4.7 0.0 {4d #661}
assign (resid 72 and name hd*) (resid 90 and name hb*) 5.8 4.0 0.2 {4d #158}
assign (resid 72 and name hd*) (resid 90 and name hg*) 5.8 4.0 0.2
assign (resid 72 and name hd*) (resid 91 and name hb*) 7.5 5.7 0.0 {4d #972}

! residue 73 ser-73
assign (resid 73 and name ha) (resid 74 and name hn) 3.3 1.5 0.2
assign (resid 73 and name hb*) (resid 74 and name hn) 4.3 2.5 0.2
assign (resid 73 and name hb*) (resid 74 and name hb*) 7.5 5.7 0.0

! residue 74 ala-74
assign (resid 74 and name ha) (resid 75 and name hn) 5.0 3.2 0.0
assign (resid 74 and name hb*) (resid 75 and name hn) 4.8 3.0 0.2
assign (resid 74 and name hb*) (resid 75 and name hd*) 7.2 5.4 0.2

! residue 75 leu-75
assign (resid 75 and name ha) (resid 75 and name hd*) 5.3 3.5 0.3
assign (resid 75 and name ha) (resid 76 and name hn) 5.0 3.2 0.0
assign (resid 75 and name hn) (resid 76 and name hn) 2.4 0.6 0.3
! assign (resid 75 and name hn) (resid 76 and name hb*) 6.0 4.2 0.0 {3d #477;
      removed rnd 23}
assign (resid 75 and name ha) (resid 78 and name hb) 3.3 1.5 0.2
assign (resid 75 and name hd*) (resid 78 and name hb) 7.9 6.1 0.0
assign (resid 75 and name hd1*) (resid 87 and name hd*) 8.9 7.1 0.0
assign (resid 75 and name hd2*) (resid 87 and name hd*) 8.9 7.1 0.0 {4d #484,615;
      rnd 20d}
assign (resid 75 and name hd*) (resid 89 and name hg1*) 8.9 7.1 0.0 {4d #500,513}
assign (resid 75 and name hd*) (resid 89 and name hg2*) 8.9 7.1 0.0 {4d #500,513;
      rnd 20d}

! residue 76 asn-76
assign (resid 76 and name ha) (resid 77 and name hn) 3.3 1.5 0.2
assign (resid 76 and name hn) (resid 77 and name hn) 5.0 3.2 0.0
assign (resid 76 and name ha) (resid 89 and name hn) 3.3 1.5 0.2
assign (resid 76 and name ha) (resid 89 and name hg2*) 4.8 3.0 0.2 {4d #921}
assign (resid 76 and name hn) (resid 89 and name hb) 5.0 3.2 0.0 {3d #491}
assign (resid 76 and name hn) (resid 89 and name hg*) 6.2 4.4 0.2 {3d #494}
assign (resid 76 and name hb*) (resid 89 and name hn) 6.0 4.2 0.0
assign (resid 76 and name hb*) (resid 89 and name hb) 6.0 4.2 0.0 {4d #1108}
```

```
! residue  77 asn-77
 assign (resid  77  and name ha) (resid  78  and name hn) 3.3 1.5 0.2
 assign (resid  77  and name hn) (resid  78  and name hn) 3.3 1.5 0.2
 assign (resid  77  and name hb*) (resid  78  and name hn) 6.0 4.2 0.0
 assign (resid  77  and name hb*) (resid  81  and name ha) 6.0 4.2 0.0
 assign (resid  77  and name hn) (resid  88  and name ha) 5.0 3.2 0.0
 assign (resid  77  and name hn) (resid  89  and name hn) 5.0 3.2 0.0
 assign (resid  77  and name hn) (resid  89  and name hg*) 7.9 6.1 0.0  {3d #503}

! residue  78 val-78
 assign (resid  78  and name ha) (resid  79  and name hn) 2.4 0.6 0.3
 assign (resid  78  and name ha) (resid  79  and name hg1*) 6.0 4.2 0.2  {4d #530}
 assign (resid  78  and name hb*) (resid  79  and name hn) 5.0 3.2 0.0
 assign (resid  78  and name hn) (resid  87  and name hn) 5.0 3.2 0.0

! residue  79 ile-79
 assign (resid  79  and name hg2*) (resid  79  and name hg1*) 5.8 4.0 0.2
 assign (resid  79  and name hg2*) (resid  83  and name ha*) 7.5 5.7 0.0
 assign (resid  79  and name hg2*) (resid  84  and name ha) 4.8 3.0 0.2
 assign (resid  79  and name hd*) (resid  85  and name ha) 6.5 4.7 0.0  {4d #136}
 assign (resid  79  and name hg1*) (resid  86  and name ha) 6.0 4.2 0.0
 assign (resid  79  and name hg2*) (resid  86  and name ha) 6.5 4.7 0.0

! residue  80 pro-80

! residue  81 glu-81
 assign (resid  81  and name ha) (resid  82  and name hn) 2.4 0.6 0.3
 assign (resid  81  and name hn) (resid  82  and name hn) 5.0 3.2 0.0
 assign (resid  81  and name hg*) (resid  82  and name hn) 6.0 4.2 0.0  {3d #531}
 assign (resid  81  and name ha) (resid  83  and name hn) 5.0 3.2 0.0
 assign (resid  81  and name hb*) (resid  83  and name hn) 6.0 4.2 0.0  {3d #539}

! residue  82 gly-82
 assign (resid  82  and name ha*) (resid  83  and name hn) 6.0 4.2 0.0
 assign (resid  82  and name hn) (resid  83  and name hn) 3.3 1.5 0.2

! residue  83 gly-83
 assign (resid  83  and name ha*) (resid  84  and name hn) 4.3 2.5 0.2
 assign (resid  83  and name ha*) (resid  85  and name hg*) 5.3 3.5 0.2

! residue  84 ser-84
 assign (resid  84  and name hn) (resid  85  and name hn) 3.3 1.5 0.2

! residue  85 gln-85
 assign (resid  85  and name ha) (resid  85  and name hg*) 6.0 4.2 0.0

! residue  86 pro-86
 assign (resid  86  and name ha) (resid  87  and name hn) 2.4 0.6 0.3

! residue  87 leu-87
 assign (resid  87  and name hb1) (resid  87  and name hd*) 6.2 4.4 0.2
```

```
 assign (resid  87  and name hb2) (resid  87  and name hd*) 6.2 4.4 0.2 {4d #771,769; rnd
        20d}
 assign (resid  87  and name ha) (resid  88  and name hn) 2.4 0.6 0.3
 assign (resid  87  and name hb*) (resid  88  and name hn) 6.0 4.2 0.0
 assign (resid  87  and name hb1) (resid  89  and name hg2*) 4.8 3.0 0.2  {4d #817,864}
 assign (resid  87  and name hb2) (resid  89  and name hg2*) 4.8 3.0 0.2  {4d #864,817;
        rnd 20d}

! residue  88 ser-88
 assign (resid  88  and name ha) (resid  89  and name hn) 3.3 1.5 0.2
 assign (resid  88  and name hn) (resid  89  and name hg*) 7.9 6.1 0.0  {3d #571}
 assign (resid  88  and name hb*) (resid  89  and name hn) 6.0 4.2 0.0

! residue  89 val-89
 assign (resid  89  and name ha) (resid  89  and name hg1*) 4.8 3.0 0.2
 assign (resid  89  and name ha) (resid  89  and name hg2*) 4.8 3.0 0.2 {4d #718,901; rnd
        20d}
 assign (resid  89  and name ha) (resid  90  and name hn) 3.3 1.5 0.2  {3d #582}

! residue  90 arg-90
 assign (resid  90  and name hg*) (resid  91  and name ha) 6.0 4.2 0.0  {4d #591}

! residue  91 leu-91
 assign (resid  91  and name ha) (resid  92  and name hn) 2.4 0.6 0.3
! assign (resid  91  and name hn) (resid  92  and name hn) 5.0 3.2 0.0  {removed rnd 20}

! residue  92 ala-92
 assign (resid  92  and name ha) (resid  93  and name hn) 3.3 1.5 0.2  {included rnd 18}
 assign (resid  92  and name hn) (resid  93  and name hn) 5.0 3.2 0.0

! residue  93 glu-93
 assign (resid  93  and name ha) (resid  94  and name hn) 2.4 0.6 0.3

! residue  94 glu-94
 assign (resid  94  and name ha) (resid  94  and name hg*) 4.3 2.5 0.2
 assign (resid  94  and name ha) (resid  95  and name hn) 3.3 1.5 0.2
 assign (resid  94  and name hb*) (resid  95  and name hn) 4.3 2.5 0.2

! residue  95 his-95
 assign (resid  95  and name ha) (resid  96  and name hn) 5.0 3.2 0.0

! residue  96 gly-96
 assign (resid  96  and name ha*) (resid  97  and name hn) 5.0 3.2 0.0

! residue  97 lys-97

! residue  103 pro-103
! assign (resid  103  and name hd*) (resid  23  and name hb*) 7.0 5.2 0.0
! assign (resid  103  and name hd*) (resid  23  and name hg*) 7.0 5.2 0.0
! assign (resid  103  and name hd*) (resid  53  and name ha) 4.3 2.5 0.2
! assign (resid  103  and name hd*) (resid  70  and name hb*) 7.0 5.2 0.0

! residue  105 pro-105
```

```
! assign (resid  105  and name hb*) (resid  85  and name ha) 4.3 2.5 0.2

! hydrogen bond restraints:
 assign (resid 16 and name hn) (resid 60 and name o) 2.1 0.3 0.2
 assign (resid 16 and name n) (resid 60 and name o) 3.1 0.3 0.2
 assign (resid 17 and name hn) (resid 90 and name o) 2.1 0.3 0.2
 assign (resid 17 and name n) (resid 90 and name o) 3.1 0.3 0.2
 assign (resid 18 and name hn) (resid 58 and name o) 2.1 0.3 0.2
 assign (resid 18 and name n) (resid 58 and name o) 3.1 0.3 0.2
 assign (resid 19 and name hn) (resid 88 and name o) 2.1 0.3 0.2
 assign (resid 19 and name n) (resid 88 and name o) 3.1 0.3 0.2
 assign (resid 26 and name hn) (resid 29 and name oe1) 2.1 0.3 0.2  {N-capping box}
 assign (resid 26 and name n) (resid 29 and name oe1) 3.1 0.3 0.2  {N-capping box}
 assign (resid 29 and name hn) (resid 26 and name og1) 2.1 0.3 0.2  {N-capping box}
 assign (resid 29 and name n) (resid 26 and name og1) 3.1 0.3 0.2  {N-capping box}
 assign (resid 30 and name hn) (resid 26 and name o) 2.1 0.3 0.2
 assign (resid 30 and name n) (resid 26 and name o) 3.1 0.3 0.2
 assign (resid 33 and name hn) (resid 29 and name o) 2.1 0.3 0.2
 assign (resid 33 and name n) (resid 29 and name o) 3.1 0.3 0.2
 assign (resid 34 and name hn) (resid 30 and name o) 2.1 0.3 0.2
 assign (resid 34 and name n) (resid 30 and name o) 3.1 0.3 0.2
 assign (resid 35 and name hn) (resid 31 and name o) 2.1 0.3 0.2
 assign (resid 35 and name n) (resid 31 and name o) 3.1 0.3 0.2
! assign (resid 37 and name hn) (resid 34 and name o) 2.1 0.3 0.2  {C-cap, rnd 21}
! assign (resid 37 and name n) (resid 34 and name o) 3.1 0.3 0.2  {C-cap, rnd 21}
! assign (resid 38 and name hn) (resid 33 and name o) 2.1 0.3 0.2  {C-cap, rnd 21}
! assign (resid 38 and name n) (resid 33 and name o) 3.1 0.3 0.2  {C-cap, rnd 21}
! assign (resid 42 and name hn) (resid 61 and name o) 2.1 0.3 0.2
! assign (resid 42 and name n) (resid 61 and name o) 3.1 0.3 0.2
 assign (resid 44 and name hn) (resid 59 and name o) 2.1 0.3 0.2
 assign (resid 44 and name n) (resid 59 and name o) 3.1 0.3 0.2
 assign (resid 46 and name hn) (resid 57 and name o) 2.1 0.3 0.2
 assign (resid 46 and name n) (resid 57 and name o) 3.1 0.3 0.2
 assign (resid 57 and name hn) (resid 46 and name o) 2.1 0.3 0.2
 assign (resid 57 and name n) (resid 46 and name o) 3.1 0.3 0.2
 assign (resid 58 and name hn) (resid 18 and name o) 2.1 0.3 0.2
 assign (resid 58 and name n) (resid 18 and name o) 3.1 0.3 0.2
 assign (resid 59 and name hn) (resid 44 and name o) 2.1 0.3 0.2
 assign (resid 59 and name n) (resid 44 and name o) 3.1 0.3 0.2
 assign (resid 60 and name hn) (resid 16 and name o) 2.1 0.3 0.2
 assign (resid 60 and name n) (resid 16 and name o) 3.1 0.3 0.2
! assign (resid 61 and name hn) (resid 42 and name o) 2.1 0.3 0.2  {removed rnd 22}
! assign (resid 61 and name n) (resid 42 and name o) 3.1 0.3 0.2  {removed rnd 22}
! assign (resid 62 and name hn) (resid 14 and name o) 2.1 0.3 0.2
! assign (resid 62 and name n) (resid 14 and name o) 3.1 0.3 0.2
! assign (resid 67 and name hn) (resid 63 and name o) 2.1 0.3 0.2
! assign (resid 67 and name n) (resid 63 and name o) 3.1 0.3 0.2
 assign (resid 68 and name hn) (resid 64 and name o) 2.1 0.3 0.2
 assign (resid 68 and name n) (resid 64 and name o) 3.1 0.3 0.2
 assign (resid 69 and name hn) (resid 65 and name o) 2.1 0.3 0.2
 assign (resid 69 and name n) (resid 65 and name o) 3.1 0.3 0.2
 assign (resid 70 and name hn) (resid 66 and name o) 2.1 0.3 0.2
 assign (resid 70 and name n) (resid 66 and name o) 3.1 0.3 0.2
```

```
 assign (resid 71 and name hn) (resid 67 and name o) 2.1 0.3 0.2
 assign (resid 71 and name n) (resid 67 and name o) 3.1 0.3 0.2
 assign (resid 72 and name hn) (resid 68 and name o) 2.1 0.3 0.2
 assign (resid 72 and name n) (resid 68 and name o) 3.1 0.3 0.2
 assign (resid 73 and name hn) (resid 69 and name o) 2.1 0.3 0.2
 assign (resid 73 and name n) (resid 69 and name o) 3.1 0.3 0.2
 assign (resid 74 and name hn) (resid 70 and name o) 2.1 0.3 0.2
 assign (resid 74 and name n) (resid 70 and name o) 3.1 0.3 0.2
 assign (resid 75 and name hn) (resid 71 and name o) 2.1 0.3 0.2
 assign (resid 75 and name n) (resid 71 and name o) 3.1 0.3 0.2
 assign (resid 76 and name hn) (resid 72 and name o) 2.1 0.3 0.2
 assign (resid 76 and name n) (resid 72 and name o) 3.1 0.3 0.2
! assign (resid 78 and name hn) (resid 74 and name o) 2.1 0.3 0.2
! assign (resid 78 and name n) (resid 74 and name o) 3.1 0.3 0.2
 assign (resid 88 and name hn) (resid 19 and name o) 2.1 0.3 0.2
 assign (resid 88 and name n) (resid 19 and name o) 3.1 0.3 0.2
! assign (resid 92 and name hn) (resid 15 and name o) 2.1 0.3 0.2 {removed rnd 20}
! assign (resid 92 and name n) (resid 15 and name o) 3.1 0.3 0.2  {removed rnd 20}
```

Torsional Restraints:

X-PLOR dihedral restraint file (dihe.tbl) used for round 24 of structure calculations of Sxl-RBD2 All changes made for round 24 are from the jnc and jcc experiments as indicated below. All other dihedral restraints were used in round 22.

```
restraints dihedral reset
! Torsion psi constraints
! Torsion chi1 constraints started round23

!residue 13 asp-13
 assign (resid 12 and name c) (resid 13 and name n)
        (resid 13 and name ca) (resid 13 and name c) 1.0 -120.0 40.0 2
!residue 14 thr-14
 assign (resid 13 and name c) (resid 14 and name n)
        (resid 14 and name ca) (resid 14 and name c) 1.0 -120.0 40.0 2
!residue 15 asn-15
 assign (resid 14 and name c) (resid 15 and name n)
        (resid 15 and name ca) (resid 15 and name c) 1.0 -120.0 40.0 2
!residue 16 leu-16
 assign (resid 15 and name c) (resid 16 and name n)
        (resid 16 and name ca) (resid 16 and name c) 1.0 -120.0 40.0 2
!residue 17 tyr-17
 assign (resid 16 and name c) (resid 17 and name n)
        (resid 17 and name ca) (resid 17 and name c) 1.0 -120.0 40.0 2
!residue 18 val-18
 assign (resid 17 and name c) (resid 18 and name n)
        (resid 18 and name ca) (resid 18 and name c) 1.0 -120.0 40.0 2
!residue 19 thr-19
 assign (resid 18 and name c) (resid 19 and name n)
        (resid 19 and name ca) (resid 19 and name c) 1.0 -120.0 40.0 2

!residue 23 arg-23
! assign (resid 22 and name c) (resid 23 and name n)
!        (resid 23 and name ca) (resid 23 and name c) 1.0 -60.0 30.0 2 {removed rnd 22}
!residue 24 thr-24
 assign (resid 23 and name c) (resid 24 and name n)
        (resid 24 and name ca) (resid 24 and name c) 1.0 -120.0 40.0 2
!residue 25 ile-25
! assign (resid 24 and name c) (resid 25 and name n)
!        (resid 25 and name ca) (resid 25 and name c) 1.0 -60.0 30.0 2 {removed rnd 22}

!residue 27 asp-27
 assign (resid 26 and name c) (resid 27 and name n)
        (resid 27 and name ca) (resid 27 and name c) 1.0 -60.0 30.0 2
!residue 28 asp-28
 assign (resid 27 and name c) (resid 28 and name n)
        (resid 28 and name ca) (resid 28 and name c) 1.0 -60.0 30.0 2

!residue 32 thr-32
 assign (resid 31 and name c) (resid 32 and name n)
        (resid 32 and name ca) (resid 32 and name c) 1.0 -60.0 30.0 2
```

```
!residue 33 ile-33
 assign (resid 32 and name c) (resid 33 and name n)
        (resid 33 and name ca) (resid 33 and name c) 1.0 -60.0 30.0 2

!residue 36 lys-36
! assign (resid 35 and name c) (resid 36 and name n)
!       (resid 36 and name ca) (resid 36 and name c) 1.0 -60.0 30.0 2 {removed rnd 22}
!residue 37 tyr-37
 assign (resid 36 and name c) (resid 37 and name n)
        (resid 37 and name ca) (resid 37 and name c) 1.0 -120.0 40.0 2
!residue 38 gly-38
! assign (resid 37 and name c) (resid 38 and name n)
!       (resid 38 and name ca) (resid 38 and name c) 1.0 -60.0 30.0 2 {removed rnd 21}
!residue 39 ser-39
 assign (resid 38 and name c) (resid 39 and name n)
        (resid 39 and name ca) (resid 39 and name c) 1.0 -120.0 40.0 2

!residue 41 ile-41
 assign (resid 40 and name c) (resid 41 and name n)
        (resid 41 and name ca) (resid 41 and name c) 1.0 -120.0 40.0 2

!residue 44 asn-44
 assign (resid 43 and name c) (resid 44 and name n)
        (resid 44 and name ca) (resid 44 and name c) 1.0 -120.0 40.0 2
!residue 45 ile-45
 assign (resid 44 and name c) (resid 45 and name n)
        (resid 45 and name ca) (resid 45 and name c) 1.0 -120.0 40.0 2

!residue 47 arg-47
 assign (resid 46 and name c) (resid 47 and name n)
        (resid 47 and name ca) (resid 47 and name c) 1.0 -120.0 40.0 2

!residue 49 lys-49
! assign (resid 48 and name c) (resid 49 and name n)
!       (resid 49 and name ca) (resid 49 and name c) 1.0 -60.0 30.0 2 {removed rnd 22}
!residue 50 leu-50
 assign (resid 49 and name c) (resid 50 and name n)
        (resid 50 and name ca) (resid 50 and name c) 1.0 -120.0 40.0 2
!residue 51 thr-51
 assign (resid 50 and name c) (resid 51 and name n)
        (resid 51 and name ca) (resid 51 and name c) 1.0 -120.0 40.0 2

!residue 57 val-57
 assign (resid 56 and name c) (resid 57 and name n)
        (resid 57 and name ca) (resid 57 and name c) 1.0 -120.0 40.0 2
!residue 58 ala-58
 assign (resid 57 and name c) (resid 58 and name n)
        (resid 58 and name ca) (resid 58 and name c) 1.0 -120.0 40.0 2
!residue 59 phe-59
 assign (resid 58 and name c) (resid 59 and name n)
        (resid 59 and name ca) (resid 59 and name c) 1.0 -120.0 40.0 2
!residue 60 val-60
 assign (resid 59 and name c) (resid 60 and name n)
```

```
        (resid 60 and name ca) (resid 60 and name c) 1.0 -120.0 40.0 2
!residue 61 arg-61
 assign (resid 60 and name c) (resid 61 and name n)
        (resid 61 and name ca) (resid 61 and name c) 1.0 -120.0 40.0 2
!residue 62 tyr-62
 assign (resid 61 and name c) (resid 62 and name n)
        (resid 62 and name ca) (resid 62 and name c) 1.0 -120.0 40.0 2

!residue 65 arg-65
 assign (resid 64 and name c) (resid 65 and name n)
        (resid 65 and name ca) (resid 65 and name c) 1.0 -60.0 30.0 2
!residue 66 glu-66
 assign (resid 65 and name c) (resid 66 and name n)
        (resid 66 and name ca) (resid 66 and name c) 1.0 -60.0 30.0 2

!residue 68 ala-68
 assign (resid 67 and name c) (resid 68 and name n)
        (resid 68 and name ca) (resid 68 and name c) 1.0 -60.0 30.0 2
!residue 69 gln-69
 assign (resid 68 and name c) (resid 69 and name n)
        (resid 69 and name ca) (resid 69 and name c) 1.0 -60.0 30.0 2

!residue 71 ala-71
 assign (resid 70 and name c) (resid 71 and name n)
        (resid 71 and name ca) (resid 71 and name c) 1.0 -60.0 30.0 2
!residue 72 ile-72
 assign (resid 71 and name c) (resid 72 and name n)
        (resid 72 and name ca) (resid 72 and name c) 1.0 -60.0 30.0 2
!residue 73 ser-73
 assign (resid 72 and name c) (resid 73 and name n)
        (resid 73 and name ca) (resid 73 and name c) 1.0 -60.0 30.0 2

!residue 75 leu-75
 assign (resid 74 and name c) (resid 75 and name n)
        (resid 75 and name ca) (resid 75 and name c) 1.0 -120.0 40.0 2
!residue 76 asn-76
 assign (resid 75 and name c) (resid 76 and name n)
        (resid 76 and name ca) (resid 76 and name c) 1.0 -60.0 30.0 2

!residue 78 val-78
 assign (resid 77 and name c) (resid 78 and name n)
        (resid 78 and name ca) (resid 78 and name c) 1.0 -120.0 40.0 2

!residue 88 ser-88
 assign (resid 87 and name c) (resid 88 and name n)
        (resid 88 and name ca) (resid 88 and name c) 1.0 -120.0 40.0 2
!residue 89 val-89
 assign (resid 88 and name c) (resid 89 and name n)
        (resid 89 and name ca) (resid 89 and name c) 1.0 -120.0 40.0 2

!residue 91 leu-91
! assign (resid 90 and name c) (resid 91 and name n)
!       (resid 91 and name ca) (resid 91 and name c) 1.0 -60.0 30.0 2 {removed rnd 22}
```

```
!residue 97 lys-97
 assign (resid 96 and name c) (resid 97 and name n)
        (resid 97 and name ca) (resid 97 and name c) 1.0 -120.0 40.0 2

!
*****************************************************************************
! chi1 restraints from jnc and jcc experiments

!residue 19 thr-19
 assign (resid 19 and name n) (resid 19 and name ca)
        (resid 19 and name cb) (resid 19 and name og1) 1.0 60.0 20.0 2
!residue 24 thr-24
 assign (resid 24 and name n) (resid 24 and name ca)
        (resid 24 and name cb) (resid 24 and name og1) 1.0 60.0 20.0 2
!residue 26 thr-26
 assign (resid 26 and name n) (resid 26 and name ca)
        (resid 26 and name cb) (resid 26 and name og1) 1.0 60.0 20.0 2
!residue 32 thr-32
 assign (resid 32 and name n) (resid 32 and name ca)
        (resid 32 and name cb) (resid 32 and name og1) 1.0 -60.0 20.0 2
!residue 33 ile-33
 assign (resid 33 and name n) (resid 33 and name ca)
        (resid 33 and name cb) (resid 33 and name cg1) 1.0 -60.0 20.0 2
!residue 40 ile-40
 assign (resid 40 and name n) (resid 40 and name ca)
        (resid 40 and name cb) (resid 40 and name cg1) 1.0 -60.0 20.0 2
!residue 45 ile-45
 assign (resid 45 and name n) (resid 45 and name ca)
        (resid 45 and name cb) (resid 45 and name cg1) 1.0 -60.0 20.0 2
!residue 51 thr-51
 assign (resid 51 and name n) (resid 51 and name ca)
        (resid 51 and name cb) (resid 51 and name og1) 1.0 60.0 20.0 2
!residue 60 val-60
 assign (resid 60 and name n) (resid 60 and name ca)
        (resid 60 and name cb) (resid 60 and name cg1) 1.0 180.0 20.0 2
!residue 79 ile-79
 assign (resid 79 and name n) (resid 79 and name ca)
        (resid 79 and name cb) (resid 79 and name cg1) 1.0 -60.0 20.0 2
!residue 89 val-89
 assign (resid 89 and name n) (resid 89 and name ca)
        (resid 89 and name cb) (resid 89 and name cg1) 1.0 180.0 20.0 2



end
```

Appendix D1

Bruker AMX (on 600 MHz at MCL) pulse programs

These are pulse programs that I used or implemented on the AMX-600 at the Melvin Calvin Laboratory. They are stored on the supplementary floppy disk.

HSQC (HMQC) pulse programs:

hmqcjrY.bv	HMQC with "jump return" water suppression
hsqcY.jgp	basic presat HSQC
fasthsqc.al	fast aquisition HSQC (2-step phase cycle) with shortened recycle delay
se_hsqc.al	original sensitivity enhanced HSQC (presat)
pep_z_hsqc.bv	PEP-Z-HSQC (no presat necessary)
hmqc15n.jgp	HMQC-J experiment used for determination of $^{1}H_{a}$-^{1}HN ^{3}J coupling constants.
ct-hsqc.jgp	constant-time HSQC

Pulse programs used for 15N relaxation/dynamics:

hetT1y.jh	^{15}N T_1 experiment
hetT2y.cl	^{15}N T_2 experiment
hetNOEy.jh	$^{15}N/^{1}H$ NOE experiment
hetnoNOEy.jh	$^{15}N/^{1}H$ reference experiment (for determination of heteronuclear NOE)
se_hetnoe.al	sensitivity enhanced $^{15}N/^{1}H$ NOE experiment

15N-edited pulse programs:

hsqcnoe.al	2D HSQC-NOESY (presat)
noeprtpw2hf.al	2D ^{1}H-^{1}H NOESY with ^{15}N w2 refocused half-X filter
noei4pr3d.al	3D ^{15}N-separated NOESY-HMQC
tocinv3d.al	3D ^{15}N-separated TOCSY-HMQC

13C-edited pulse programs:

hcchtoc.al	3D HCCH-TOCSY
hcchnoe4d.al	4D HMQC-NOESY-HMQC

Triple-resonance pulse programs:

cbcaconh.sg	CBCA(CO)NH
hnco_ct.al	constant-time HNCO
hcaco_ct.al	constant-time HCACO
cthsqcjnc.jgp	J_{NC} quantitative J-correlation experiment
cthsqcjcc.jgp	J_{CC} quantitative J-correlation experiment

Appendix D2

Bruker DMX pulse programs

These are pulse programs (and pulsed field gradient programs) that Dr. Brian Volkman and I implemented and used at NMRFAM (Nuclear Magnetic Resonance Facility at Madison). They are stored on the supplementary floppy disk.

HSQC pulse programs:

sehsf2_fb.bv	$^{15}N/^{1}H$ HSQC with gradient coherence selection, sensitivity enhancement, and water along +Z — gradient program = bvhsqcse

Pulse programs used for 15N relaxation/dynamics:

t1_15n_sefb.bv	^{15}N T_1 experiment with gradient coherence selection, sensitivity enhancement, and water along +Z — gradient program = bv_t1_15n
t2_15n_sefb.al	^{15}N T_2 experiment with gradient coherence selection, sensitivity enhancement, and water along +Z — gradient program = al_t2_15n
t2_15n_sefb0.al	for "zero time point" (i.e. no CPMG) — gradient program = al_t2_15n

15N-edited pulse programs:

noehsqcse_fb.bv	3D ^{15}N-separated NOESY-HSQC with gradient coherence selection, sensitivity enhancement, and water along +Z — gradient program = noehsqcsefa
tochsqc_fb2.bv	3D ^{15}N-separated TOCSY-HSQC with gradient coherence selection, sensitivity enhancement, and water along +Z — gradient program = bvtochsqc_fb

Triple-resonance pulse programs:

hnco_sefb.al	HNCO with gradient coherence selection, sensitivity enhancement, water along +Z, and ^{1}H decoupling. (http:/www.bmrb.wisc.edu/pulse_seq/files/hnco_sefb.al)
hnca_sefb.al	HNCA with gradient coherence selection, sensitivity enhancement, water along +Z, and ^{1}H decoupling. (http:/www.bmrb.wisc.edu/pulse_seq/files/hnca_sefb.al)
hncoca_sefb.al	HN(CO)CA with gradient coherence selection, sensitivity enhancement, water along +Z, and ^{1}H decoupling. (http:/www.bmrb.wisc.edu/pulse_seq/files/hncoca_sefb.al)

(gradient programs for triple-resonance pulse programs are at the end of each file)

Appendix E

Non-standard Felix data processing macros*:

se_s2d.mac	for original sensitivity enhanced HSQC with States-TPPI quadrature in t_1
se_hetnoe.mac	for sensitivity enhanced "reference" and "$^{15}N/^{1}H$ hetNOE" experiments
hf2d.mac	for 2D NOESY with ^{15}N $\omega 2$ refocused half-X filter (Chapter 2)
jnc_jcc.mac	for J_{NC} and J_{CC} experiment data processing
dalls4d.mac	for 4D $^{13}C/^{13}C$-separated HMQC-NOESY-HMQC

Felix analysis macros*:

These three macros were used to assign the backbone of Sxl-RBD2 from 3D NOESY-HMQC and 3D TOCSY-HMQC in Felix 2.14 as described (Chapter 4).

pul3dvec3.mac
feditvec3.mac
reditvec3.mac

These four macros were used to view and pick peaks from 4D data in Felix 2.30. They are also contained (along with a description of how they are used) in the UNIX "tar" file felix_4dmacs.tar. These particular macros are written such that they are integrated into the Felix menus.

2dfrom4dmnu.mac
manpick4dmnu.mac
show4dpkmnu.mac
extractdbamnu.mac

These four macros were used to generate "strip" matrices. The first two are Fred Damberger's original strip macros; the second two were edited by me from Fred's macros to make "scratch" or temporary strip matrices from a dba entity instead of from a dba list.

stripa5.mac
stripb5.mac
67_tstrip.mac
67_tstripb5.mac

* All of these macros are given on the supplementary floppy disk.

Appendix F

Excel 4.0 macros:

extract2.mac	macro "worksheet" used to direct the matching of 4D HMQC-NOESY-HMQC crosspeaks with the existing Sxl-RBD2 assignments (Chapter 5)
n15_extract.mac	edited from extract2.mac to do the same task for the 3D ^{15}N-separated NOESY-HMQC Sxl-RBD2 data
convert.mac	a macro which converts the output format from extract2.mac to standard nomenclature

perl scripts:

dist_fam	used to obtain specified distances from a family of pdb structures
dihe_fam	used to obtain specified dihedral angles from a family of pdb structures
get_accept	used to summarize important information from the X-PLOR accept.log file (log file for "accepted" structures)
grep4d	used to extract specific information on 4D NOEs
grep4dpeak	used to extract specific information on 4D NOEs
match3d	
pdbcat	used to concatenate a family of pdb protein structures for submission to the Protein Data Bank

[the following perl scripts were used in the analysis of Sxl-RBD1+2 complexed with the Tra-PPT 10-mer (Chapter 8)]:

ca_stripsearch	given below, p. 236 (described in Chapter 8)
cocahg	given below, p. 238 (described in Chapter 8)
cocahng	given below, p. 239 (described in Chapter 8)
ha_stripsearch	used to search 3D NOESY-HSQC for sequential H_α-HN crosspeaks
hacahng	given below, p. 242 (described in Chapter 8)
hcacog	given below, p. 246 (described in Chapter 8)
hn2hng	used to extract NH-NH crosspeaks from 3D NOESY-HSQC
hncocag	given below, p. 247 (described in Chapter 8)
hnhacag	given below, p. 249 (described in Chapter 8)
reduce_strip	used to generate a list of 15N/1H correlation peaks which have not been assigned
toc_stripsearch	used to search 3D TOCSY-HSQC for intraresidue H_α-HN crosspeaks

```
#!/usr/bin/perl
# ca_stripsearch                A. Lee  1/29/96
# This perl script "greps" the 3D HN(CO)CA or HNCA for
# sequential 13Ca chemical shifts.  Feed it an 13Ca shift
# and it will produce candidate HN strips for the next
# (i+1) residue.
#
# syntax:  perl ca_stripsearch <1Ha ppm> <hncoca or hnca>

$coff = 0.20;           # 13Ca error
$file1 = "/usr3/people/alee/perl/69/69_$ARGV[1]_ppm2.txt";
$file2 = "/usr3/people/alee/perl/69/67_hsqc_ppm2.txt";
$file3 = "/usr3/people/alee/perl/69/69_assigned2.txt";

### load 2D array with 13Ca matches ###
open (FILE1, "$file1") || die "Couldn't open file: $!\n";

$linecount = 1;
print "\n13Ca matches from 3D $ARGV[1]:\n\n";
while (<FILE1>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if($fld[5]>$ARGV[0]-$coff && $fld[5]<$ARGV[0]+$coff) {
      $ca_err = $ARGV[0]-$fld[5];
      print "$linecount   $fld[1]  $fld[4] / $fld[6]    $fld[5] ";
      printf " (%2.3f)\n", $ca_err;
      $hn_list{$linecount, 1} = $fld[4];  # 1H ppm
      $hn_list{$linecount, 2} = $fld[6];  # 15N ppm
      $hn_list{$linecount, 3} = $fld[1];  # peak number
      $linecount++;
   }
}
close(FILE1);

### match each HN pair with HSQC strip number ###
print "\n\nPut the following into tempstrip.txt in felix230 dir.\n";
print "You will want to add in as item #1 the strip which you would \n";
print "like to compare to.\n\n";
print "c**/usr3/people/alee/felix230/text/tempstrip.txt\n";
print "number   1\n";

$poff = 0.03;
$noff = 0.3;
$ca_keys = keys(%hn_list);
$ca_match = $ca_keys/3;

#++++++++ read 69_assigned.txt into array ++++++++
open (FILE3, "$file3") || die "Couldn't open file: $!\n";
$num = 0;
while (<FILE3>) {
   @fld = split(/\s+/, $_);       # delimits by whitespace
   push(@used,$fld[0]);           # add "assigned residue to array
   $num++;
}
```

```
close(FILE3);
$assigned = @used;
#++++++++++++++++++++++++++++++++++++++++++++++++++++++++
$counter = 1;
for ($i = 1; $i <= $ca_match; $i++) {
   $item = $i+1;
   $in_hsqc = 0;                    # flag to see if hsqc peak exists
   open (FILE2, "$file2") || die "Couldn't open file: $!\n";
   while (<FILE2>) {
      @fld = split(/\s+/, $_);    #delimits by whitespace
      if(($fld[4]>$hn_list{$i,1}-$poff && $fld[4]<$hn_list{$i,1}+$poff) &&
      ($fld[6]>$hn_list{$i,2}-$noff && $fld[6]<$hn_list{$i,2}+$noff)) {
         $in_hsqc = $in_hsqc+1;
         #++++++++++++ check to see if already assigned ++++++++
         $flag = 0;                    # "0" means not assigned
         for ($j = 0; $j < $assigned; $j++) {
            if($fld[1]==$used[$j]) {
            $flag = 1;    # "1" means strip #$fld[1] has been assigned
            }
         }
           #++++++++++++++++++++++++++++++++++++++++++++++++++++++++
         if($flag == 0) {
            $counter++;
            print "$counter    $fld[1]\n";
            #print "from 3D $ARGV[1]: $hn_list{$i,3}\n";
         }
      }
   }
   close(FILE2);
   if($in_hsqc == 0) {
   print "No HSQC peak found at $hn_list{$i,1}/$hn_list{$i,2}\n";
   }
}
```

```
#!/usr/bin/perl
# cocahg
# This script "greps" or matches 13C'/13Ca values to HCACO.
#
# syntax: perl cocahg <13C' ppm> <13Ca ppm>

$hoff = 0.05;           # 13C' error
$noff = 0.2;            # 13Ca error
$filename = "/usr3/people/alee/perl/69/69_hcaco_ppm2.txt";

open (FILE, "$filename") ||
die "Couldn't open file: $!\n";

print "\n", "HCACO correlation matches:  ", "$ARGV[0]/$ARGV[1]\n", "\n";
print "       13C'          13Ca          1H\n", "\n";
while (<FILE>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if(($fld[5]>$ARGV[0]-$hoff && $fld[5]<$ARGV[0]+$hoff) &&
   ($fld[6]>$ARGV[1]-$noff && $fld[6]<$ARGV[1]+$noff)) {
      $h_err = $ARGV[0]-$fld[5];
      $n_err = $ARGV[1]-$fld[6];
      print "$fld[1]\t", "$fld[5]", " (";
      printf "%2.3f", $h_err;
      print ")   ", "$fld[6]", " (";
      printf "%2.3f", $n_err;
      print ")   ", "$fld[4]\n";
   }
}

close(FILE);
```

```
#!/usr/bin/perl
# cocahng       A. Lee  1/29/96
# This script "greps" or matches 13C'/13Ca values to HNCO, HN(CO)CA,
# and CBCA(CO)NH which share HN chemical shifts.
#
# syntax:  perl cocahng <13C' ppm> <13Ca>

$co_off = 0.05;                 # 13C' error
$ca_off = 0.2;                  # 13Ca error
$file1 = "/usr3/people/alee/perl/69/69_hnco_ppm2.txt";
$file2 = "/usr3/people/alee/perl/69/69_hncoca_ppm2.txt";
$file3 = "/usr3/people/alee/perl/69/69_cbcaconh_ppm2.txt";
$file4 = "/usr3/people/alee/perl/69/67_hsqc_ppm2.txt";

print "\n15N HNCO, HN(CO)CA, and CBCA(CO)NH correlation matches:  ",
    "$ARGV[0] / $ARGV[1]\n", "\n";

#### get 13C' matches from HNCO first .... ####
open (FILE1, "$file1") || die "Couldn't open file1: $!\n";

print "HNCO 13C' matches ... \n\n";
$linecount = 1;
while (<FILE1>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if($fld[5]>$ARGV[0]-$co_off && $fld[5]<$ARGV[0]+$co_off) {
      $co_err = $ARGV[0]-$fld[5];
      $hn_hnco{$linecount, 1} = $fld[1];       # add item# to "match" table
      $hn_hnco{$linecount, 2} = $fld[4];       # add 1HN to "match" table
      $hn_hnco{$linecount, 3} = $fld[6];       # add 15N to "match" table
      $hn_hnco{$linecount, 4} = $fld[5];       # add 13C' to "match" table
      $hn_hnco{$linecount, 5} = $co_err;       # add error to "match" table
      $linecount++;
   }
}

close(FILE1);

# print out matches ...
$hncokeys = keys(%hn_hnco);
$hncomatch = $hncokeys/5;
for ($i = 1; $i <= $hncomatch; $i++) {
   print "$hn_hnco{$i,1} \t $hn_hnco{$i,2}  $hn_hnco{$i,3}  $hn_hnco{$i,4} ",
        "(13C' error = ";
   printf "%2.3f)\n", $hn_hnco{$i,5};
}

#### get 13Ca matches from HN(CO)CA second .... ####
open (FILE2, "$file2") || die "Couldn't open file1: $!\n";

print "\nHN(CO)CA 13Ca matches ... \n\n";
$linecount = 1;
while (<FILE2>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
```

```
   if($fld[5]>$ARGV[1]-$ca_off && $fld[5]<$ARGV[1]+$ca_off) {
      $ca_err = $ARGV[1]-$fld[5];
      $hn_hncoca{$linecount, 1} = $fld[1];     # add item# to "match" table
      $hn_hncoca{$linecount, 2} = $fld[4];     # add 1HN to "match" table
      $hn_hncoca{$linecount, 3} = $fld[6];     # add 15N to "match" table
      $hn_hncoca{$linecount, 4} = $fld[5];     # add 13Ca to "match" table
      $hn_hncoca{$linecount, 5} = $ca_err;     # add error to "match" table
      $linecount++;
   }
}

close(FILE2);

# print out matches ...
$hncocakeys = keys(%hn_hncoca);
$hncocamatch = $hncocakeys/5;
for ($i = 1; $i <= $hncocamatch; $i++) {
   print "$hn_hncoca{$i,1} \t $hn_hncoca{$i,2}  $hn_hncoca{$i,3} ",
        "$hn_hncoca{$i,4} (13Ca error = ";
   printf "%2.3f)\n", $hn_hncoca{$i,5};
}


#########################################################################
###################    match HNCO and HN(CO)CA    ######################
#########################################################################
print "\n\n\nHNCO / HN(CO)CA Agreement ..... \n";

$poff = 0.03;
$noff = 0.3;
$counter = 1;
for ($i = 1; $i <= $hncomatch; $i++) {
  for ($j = 1; $j <= $hncocamatch; $j++) {
    if(($hn_hnco{$i,2}>$hn_hncoca{$j,2}-$poff &&
      $hn_hnco{$i,2}<$hn_hncoca{$j,2}+$poff) &&
      ($hn_hnco{$i,3}>$hn_hncoca{$j,3}-$noff &&
      $hn_hnco{$i,3}<$hn_hncoca{$j,3}+$noff)) {          #test if HNs match

      ## print out match results
      print "* * * * * * * * * * * * * * * * * * * * * * * * * * * * *\n";
      print "$counter\n";
      print "HNCO:    $hn_hnco{$i,1} \t $hn_hnco{$i, 2}",
           "    $hn_hnco{$i,3}  $hn_hnco{$i,4} ";
      printf "(%2.3f)\n", $hn_hnco{$i,5};
      print "HNCOCA:  $hn_hncoca{$j,1} \t $hn_hncoca{$j, 2}",
           "    $hn_hncoca{$j,3}  $hn_hncoca{$j,4} ";
      printf "(%2.3f)\n\n", $hn_hncoca{$j,5};

      ## get some statistics
      $h_ave = ($hn_hnco{$i,2}+$hn_hncoca{$j,2})/2;
      $n_ave = ($hn_hnco{$i,3}+$hn_hncoca{$j,3})/2;
      $hdiff = $hn_hnco{$i,2}-$hn_hncoca{$j,2};
      $ndiff = $hn_hnco{$i,3}-$hn_hncoca{$j,3};
```

```
    printf "averages:       %2.3f     %2.3f\n", $h_ave, $n_ave;
    printf "differences:    (%2.3f)   (%2.3f)\n", $hdiff, $ndiff;

    ## print out HSQC peak identities
    open (FILE4, "$file4") || die "Couldn't open file: $!\n";
    print "\n      HSQC correlation matches:\n";
       while (<FILE4>) {
         @fld2 = split(/\s+/, $_);      #delimits by whitespace
      if(($fld2[5]>$h_ave-$poff && $fld2[5]<$h_ave+$poff) &&
      ($fld2[6]>$n_ave-$noff && $fld2[6]<$n_ave+$noff)) {
        $h_err = $h_ave-$fld2[5];
        $n_err = $n_ave-$fld2[6];
        print "      # $fld2[1]    ", "$fld2[5] (";
        printf "%2.3f", $h_err;
        print ")\t", "$fld2[6] (";
        printf "%2.3f", $n_err;
        print ")\n";
      }
    }
    close(FILE4);

    ## print out CBCA(CO)NH peaks
    open (FILE3, "$file3") || die "Couldn't open file: $!\n";
    print "\n      CBCA(CO)NH correlation matches:\n";
    while (<FILE3>) {
         @fld2 = split(/\s+/, $_);      #delimits by whitespace
      if(($fld2[4]>$h_ave-$poff && $fld2[4]<$h_ave+$poff) &&
      ($fld2[6]>$n_ave-$noff && $fld2[6]<$n_ave+$noff)) {
        $h_err = $h_ave-$fld2[4];
        $n_err = $n_ave-$fld2[6];
        print "      # $fld2[1]    ", "$fld2[4] (";
        printf "%2.3f", $h_err;
        print ")\t", "$fld2[6] (";
        printf "%2.3f", $n_err;
        print ")    $fld2[5]\n";
      }
    }
    close(FILE3);

   $counter = $counter +1;
   }
  }
}
print "* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *\n\n";
```

```
#!/usr/bin/perl
# hacahng       A. Lee  1/26/96
# This script "greps" or matches 1Ha/13Ca values to HNCA and 15N 3D TOCSY
# and/or 3D NOESY which share HN chemical shifts.
#
# syntax: perl hacahng <1Ha ppm> <13Ca ppm>

$hoff = 0.05;           # 1Ha error
$coff = 0.20;           # 13Ca error
$file1 = "/usr3/people/alee/perl/69/67_tocsy_ppm3.txt";
$file2 = "/usr3/people/alee/perl/69/69_hnca_ppm2.txt";
$file3 = "/usr3/people/alee/perl/69/67_hsqc_ppm2.txt";
$file4 = "/usr3/people/alee/perl/69/67_3dnoesy_ppm2.txt";

open (FILE1, "$file1") || die "Couldn't open file1: $!\n";

print "\n", "15N 3D TOCSY and HNCA correlation matches:  ",
    "$ARGV[0]/$ARGV[1]\n", "\n";
print "TOCSY 1Ha matches ... \n\n";
$linecount = 1;
#
# get 1Ha matches from 3D TOCSY first ....
while (<FILE1>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if($fld[5]>$ARGV[0]-$hoff && $fld[5]<$ARGV[0]+$hoff) {
       $h_err = $ARGV[0]-$fld[5];
       $hn_toc{$linecount, 1} = $fld[1];          # add item# to "match" table
       $hn_toc{$linecount, 2} = $fld[4];          # add 1HN to "match" table
       $hn_toc{$linecount, 3} = $fld[6];          # add 15N to "match" table
       $hn_toc{$linecount, 4} = $fld[5];          # add 1Ha to "match" table
       $hn_toc{$linecount, 5} = $h_err;           # add error to "match" table
       $linecount++;
   }
}

close(FILE1);

# print out matches ...
$tockeys = keys(%hn_toc);
$tocmatch = $tockeys/5;
for ($i = 1; $i <= $tocmatch; $i++) {
   print "$hn_toc{$i,1} \t $hn_toc{$i,2}  $hn_toc{$i,3}  $hn_toc{$i,4} ",
       "(1Ha error = ";
   printf "%2.3f)\n", $hn_toc{$i,5};
}

# get 13Ca matches from HNCA second ....
open (FILE2, "$file2") || die "Couldn't open file2: $!\n";

print "\nHNCA 13Ca matches ... \n\n";
$linecount = 1;
#
while (<FILE2>) {
```

```
    @fld = split(/\s+/, $_);       #delimits by whitespace
    if($fld[5]>$ARGV[1]-$coff && $fld[5]<$ARGV[1]+$coff) {
        $c_err = $ARGV[1]-$fld[5];
        $hn_hnca{$linecount, 1} = $fld[1];        # add item# to "match" table
        $hn_hnca{$linecount, 2} = $fld[4];        # add 1HN to "match" table
        $hn_hnca{$linecount, 3} = $fld[6];        # add 15N to "match" table
        $hn_hnca{$linecount, 4} = $fld[5];        # add 13Ca to "match" table
        $hn_hnca{$linecount, 5} = $c_err;         # add error to "match" table
        $linecount++;
    }
}

close(FILE2);

# print out matches ...
$hncakeys = keys(%hn_hnca);
$hncamatch = $hncakeys/5;
for ($i = 1; $i <= $hncamatch; $i++) {
    print "$hn_hnca{$i,1} \t $hn_hnca{$i, 2}  $hn_hnca{$i,3}  $hn_hnca{$i,4} ",
        "(13Ca error = ";
    printf "%2.3f)\n", $hn_hnca{$i,5};
}


############ match TOCSY and HNCA ############
print "\nTOCSY/HNCA Agreement .... \n";
$poff = 0.03;
$noff = 0.3;
$counter = 1;
for ($i = 1; $i <= $tocmatch; $i++) {
  for ($j = 1; $j <= $hncamatch; $j++) {
    if(($hn_toc{$i,2}>$hn_hnca{$j,2}-$poff &&
      $hn_toc{$i,2}<$hn_hnca{$j,2}+$poff) &&
      ($hn_toc{$i,3}>$hn_hnca{$j,3}-$noff &&
      $hn_toc{$i,3}<$hn_hnca{$j,3}+$noff)) {              #test if HNs match
      ## print out match results
      print "* * * * * * * TOCSY * * * * * * * * * * TOCSY * * * * * * * *\n";
      print "$counter TOCSY\n";
      print "TOCSY:  $hn_toc{$i,1} \t $hn_toc{$i, 2}",
          "    $hn_toc{$i,3}  $hn_toc{$i,4} ";
      printf "(%2.3f)\n", $hn_toc{$i,5};
      print "HNCA:   $hn_hnca{$j,1} \t $hn_hnca{$j, 2}",
          "    $hn_hnca{$j,3}  $hn_hnca{$j,4} ";
      printf "(%2.3f)\n\n", $hn_hnca{$j,5};
      ## get some statistics
      $h_ave = ($hn_toc{$i,2}+$hn_hnca{$j,2})/2;
      $n_ave = ($hn_toc{$i,3}+$hn_hnca{$j,3})/2;
      $hdiff = $hn_toc{$i,2}-$hn_hnca{$j,2};
      $ndiff = $hn_toc{$i,3}-$hn_hnca{$j,3};
      printf "averages:       %2.3f    %2.3f\n", $h_ave, $n_ave;
      printf "differences:    (%2.3f)   (%2.3f)\n", $hdiff, $ndiff;
      ## print out HSQC peak identities
      open (FILE3, "$file3") || die "Couldn't open file: $!\n";
```

```
        print "\n      HSQC correlation matches:\n";
        #print "          1H              15N\n\n";
           while (<FILE3>) {
             @fld2 = split(/\s+/, $_);     #delimits by whitespace
           if(($fld2[5]>$h_ave-$poff && $fld2[5]<$h_ave+$poff) &&
           ($fld2[6]>$n_ave-$noff && $fld2[6]<$n_ave+$noff)) {
             $h_err = $h_ave-$fld2[5];
             $n_err = $n_ave-$fld2[6];
             print "      # $fld2[1]    ", "$fld2[5] (";
             printf "%2.3f", $h_err;
             print ")\t", "$fld2[6] (";
             printf "%2.3f", $n_err;
             print ")\n";
           }
        }
        close(FILE3);
     $counter = $counter +1;
     }
   }
}
print "* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *\n\n";

#
# get 1Ha matches from 3D NOESY third ....
open (FILE4, "$file4") || die "Couldn't open file1: $!\n";
$linecount = 1;
while (<FILE4>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if($fld[5]>$ARGV[0]-$hoff && $fld[5]<$ARGV[0]+$hoff) {
       $h_err = $ARGV[0]-$fld[5];
       $hn_noe{$linecount, 1} = $fld[1];          # add item# to "match" table
       $hn_noe{$linecount, 2} = $fld[4];          # add 1HN to "match" table
       $hn_noe{$linecount, 3} = $fld[6];          # add 15N to "match" table
       $hn_noe{$linecount, 4} = $fld[5];          # add 1Ha to "match" table
       $hn_noe{$linecount, 5} = $h_err;           # add error to "match" table
       $linecount++;
   }
}

close(FILE4);

$noekeys = keys(%hn_noe);
$noematch = $noekeys/5;

############## match NOESY and HNCA ##############
print "\nNOESY/HNCA Agreement ...\n";
$counter = 1;
for ($i = 1; $i <= $noematch; $i++) {
  for ($j = 1; $j <= $hncamatch; $j++) {
    if(($hn_noe{$i,2}>$hn_hnca{$j,2}-$poff &&
      $hn_noe{$i,2}<$hn_hnca{$j,2}+$poff) &&
      ($hn_noe{$i,3}>$hn_hnca{$j,3}-$noff &&
      $hn_noe{$i,3}<$hn_hnca{$j,3}+$noff)) {          #test if HNs match
```

```
        ## print out match results
        print "* * * * * * * NOESY * * * * * * * * * * NOESY * * * * * * *\n";
        print "$counter NOESY\n";
        print "NOESY:  $hn_noe{$i,1} \t $hn_noe{$i, 2}",
              "    $hn_noe{$i,3}  $hn_noe{$i,4} ";
        printf "(%2.3f)\n", $hn_noe{$i,5};
        print "HNCA:   $hn_hnca{$j,1} \t $hn_hnca{$j, 2}",
              "    $hn_hnca{$j,3}  $hn_hnca{$j,4} ";
        printf "(%2.3f)\n\n", $hn_hnca{$j,5};
        ## get some statistics
        $h_ave = ($hn_noe{$i,2}+$hn_hnca{$j,2})/2;
        $n_ave = ($hn_noe{$i,3}+$hn_hnca{$j,3})/2;
        $hdiff = $hn_noe{$i,2}-$hn_hnca{$j,2};
        $ndiff = $hn_noe{$i,3}-$hn_hnca{$j,3};
        printf "averages:       %2.3f     %2.3f\n", $h_ave, $n_ave;
        printf "differences:    (%2.3f)   (%2.3f)\n", $hdiff, $ndiff;
        ## print out HSQC peak identities
        open (FILE3, "$file3") || die "Couldn't open file: $!\n";
        print "\n      HSQC correlation matches:\n";
        #print "          1H              15N\n\n";
          while (<FILE3>) {
            @fld2 = split(/\s+/, $_);      #delimits by whitespace
          if(($fld2[5]>$h_ave-$poff && $fld2[5]<$h_ave+$poff) &&
          ($fld2[6]>$n_ave-$noff && $fld2[6]<$n_ave+$noff)) {
            $h_err = $h_ave-$fld2[5];
            $n_err = $n_ave-$fld2[6];
            print "      # $fld2[1]    ", "$fld2[5] (";
            printf "%2.3f", $h_err;
            print ")\t", "$fld2[6] (";
            printf "%2.3f", $n_err;
            print ")\n";
          }
        }
        close(FILE3);
      $counter = $counter +1;
      }
    }
  }
print "* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *\n\n";
```

```
#!/usr/bin/perl
# hcacog        A. Lee   1/29/96
# This script "greps" or matches 1Ha/13Ca values to HCACO.
#
# syntax: perl hcacog <1Ha ppm> <13Ca ppm>

$hoff = 0.04;             # 13C' error
$coff = 0.2;              # 13Ca error
$filename = "/usr3/people/alee/perl/69/69_hcaco_ppm2.txt";

open (FILE, "$filename") ||
die "Couldn't open file: $!\n";

print "\n", "HCACO correlation matches:  ", "$ARGV[0] / $ARGV[1]\n", "\n";
print "       1H            13Ca          13C'\n", "\n";
while (<FILE>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if(($fld[4]>$ARGV[0]-$hoff && $fld[4]<$ARGV[0]+$hoff) &&
   ($fld[6]>$ARGV[1]-$coff && $fld[6]<$ARGV[1]+$coff)) {
      $h_err = $ARGV[0]-$fld[4];
      $c_err = $ARGV[1]-$fld[6];
      print "$fld[1]\t", "$fld[4]", " (";
      printf "%2.3f", $h_err;
      print ")   ", "$fld[6]", " (";
      printf "%2.3f", $c_err;
      print ")   ", "$fld[5]\n";
   }
}

close(FILE);
```

```
#!/usr/bin/perl
# hncocag    A. Lee    1/22/96
# This script "greps" or matches HN values to 3D HNCO, HN(CO)CA,
# and CBCA(CO)NH and prints out correlated 13CO, 13Ca,
# and 13Cb ppm values.
#
# syntax: perl hncocag <1H ppm> <15N ppm>

$hoff = 0.03;              # 1H error
$noff = 0.3;               # 15N error
$file1 = "/usr3/people/alee/perl/69/69_hnco_ppm2.txt";
$file2 = "/usr3/people/alee/perl/69/69_hncoca_ppm2.txt";
$file3 = "/usr3/people/alee/perl/69/69_cbcaconh_ppm2.txt";

#### HNCO ####
open (FILE1, "$file1") || die "Couldn't open file: $!\n";

print "\n\n1H/15N search ppm values: $ARGV[0]/$ARGV[1]\n\n";
print "      1H            15N            13C\n";
print "HNCO:\n";
while (<FILE1>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if(($fld[4]>$ARGV[0]-$hoff && $fld[4]<$ARGV[0]+$hoff) &&
   ($fld[6]>$ARGV[1]-$noff && $fld[6]<$ARGV[1]+$noff)) {
      $h_err = $ARGV[0]-$fld[4];
      $n_err = $ARGV[1]-$fld[6];
      print "$fld[1]\t", "$fld[4]", " (";
      printf "%2.3f", $h_err;
      print ")   ", "$fld[6]", " (";
      printf "%2.3f", $n_err;
      print ")   ", "$fld[5]\n";
   }
}

close(FILE1);

####  HN(CO)CA ####
open (FILE2, "$file2") || die "Couldn't open file: $!\n";

print "\n\n", "HN(CO)CA:\n";
while (<FILE2>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if(($fld[4]>$ARGV[0]-$hoff && $fld[4]<$ARGV[0]+$hoff) &&
   ($fld[6]>$ARGV[1]-$noff && $fld[6]<$ARGV[1]+$noff)) {
      $h_err = $ARGV[0]-$fld[4];
      $n_err = $ARGV[1]-$fld[6];
      print "$fld[1]\t", "$fld[4]", " (";
      printf "%2.3f", $h_err;
      print ")   ", "$fld[6]", " (";
      printf "%2.3f", $n_err;
      print ")   ", "$fld[5]\n";
   }
}
```

```
close(FILE2);

#### CBCA(CO)NH ####
open (FILE3, "$file3") || die "Couldn't open file: $!\n";

print "\n\n", "CBCA(CO)NH:\n";
while (<FILE3>) {
   @fld = split(/\s+/, $_);        #delimits by whitespace
   if(($fld[4]>$ARGV[0]-$hoff && $fld[4]<$ARGV[0]+$hoff) &&
   ($fld[6]>$ARGV[1]-$noff && $fld[6]<$ARGV[1]+$noff)) {
      $h_err = $ARGV[0]-$fld[4];
      $n_err = $ARGV[1]-$fld[6];
      print "$fld[1]\t", "$fld[4]", " (";
      printf "%2.3f", $h_err;
      print ")   ", "$fld[6]", " (";
      printf "%2.3f", $n_err;
      print ")   ", "$fld[5]\n";
   }
}

close(FILE3);
```

```
#!/usr/bin/perl
# hnhacag       A. Lee  1/29/96
# This script "greps" correlates 1Ha/13Ca values to intraresidue HN values
# from HNCA, 3D 15N TOCSY, and/or 3D 15N NOESY.
#
# syntax: perl hnhacag <1H ppm> <15N ppm>

$hoff = 0.03;           # 1Ha error
$noff = 0.30;           # 15N error
$file1 = "/usr3/people/alee/perl/69/69_hnca_ppm2.txt";
$file2 = "/usr3/people/alee/perl/69/67_tocsy_ppm3.txt";
$file3 = "/usr3/people/alee/perl/69/67_3dnoesy_ppm2.txt";

###### HNCA ######
open (FILE1, "$file1") || die "Couldn't open file: $!\n";

print "\n\n1H/15N search ppm values: $ARGV[0] / $ARGV[1]\n\n";
print "-------------------------------------------------\n";
print "      1H            15N           13Ca\n";
print "HNCA:\n";
while (<FILE1>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if(($fld[4]>$ARGV[0]-$hoff && $fld[4]<$ARGV[0]+$hoff) &&
   ($fld[6]>$ARGV[1]-$noff && $fld[6]<$ARGV[1]+$noff)) {
      $h_err = $ARGV[0]-$fld[4];
      $n_err = $ARGV[1]-$fld[6];
      print "$fld[1]\t", "$fld[4]  (";
      printf "%2.3f", $h_err;
      print ")   $fld[6] (";
      printf "%2.3f", $n_err;
      print ")   $fld[5]\n";
   }
}

close(FILE1);

###### TOCSY ######
open (FILE2, "$file2") || die "Couldn't open file: $!\n";

print "-------------------------------------------------\n";
print "      1H            15N           1Ha\n";
print "TOCSY:\n";
while (<FILE2>) {
   @fld = split(/\s+/, $_);       #delimits by whitespace
   if(($fld[4]>$ARGV[0]-$hoff && $fld[4]<$ARGV[0]+$hoff) &&
   ($fld[6]>$ARGV[1]-$noff && $fld[6]<$ARGV[1]+$noff)) {
      $h_err = $ARGV[0]-$fld[4];
      $n_err = $ARGV[1]-$fld[6];
      print "$fld[1]\t", "$fld[4]  (";
      printf "%2.3f", $h_err;
      print ")   $fld[6] (";
      printf "%2.3f", $n_err;
      print ")   $fld[5]\n";
```

```
    }
}

close(FILE2);

###### NOESY ######
open (FILE3, "$file3") || die "Couldn't open file: $!\n";

print "\nNOESY:\n";
while (<FILE3>) {
    @fld = split(/\s+/, $_);        #delimits by whitespace
    if(($fld[4]>$ARGV[0]-$hoff && $fld[4]<$ARGV[0]+$hoff) &&
    ($fld[6]>$ARGV[1]-$noff && $fld[6]<$ARGV[1]+$noff)) {
        $h_err = $ARGV[0]-$fld[4];
        $n_err = $ARGV[1]-$fld[6];
        print "$fld[1]\t", "$fld[4]  (";
        printf "%2.3f", $h_err;
        print ")   $fld[6] (";
        printf "%2.3f", $n_err;
        print ")   $fld[5]\n";
    }
}
print "-------------------------------------------------\n";

close(FILE3);
```