ABSTRACT

1) NMR Studies of the Spliced Leader RNA from *Crithidia fasciculata* and *Leptomonas collosoma*.
2) Hydrodynamic Properties of Nucleic Acids by NMR.

Jon Lapham

May, 1998

The first part of this thesis examines the Spliced Leader RNA (SL RNA) from two species of trypanosome, *Crithidia fasciculata* and *Leptomonas collosoma*. Unlike other eukaryotes, trypanosomal genes lack internal introns, rather, they are excised by *trans*-splicing to the SL RNA during pre-mRNA processing. Previous studies have shown that the SL RNA can adopt two alternate secondary structures, form 1 and form 2, and it has been suggested that the RNA may be involved in a conformational switch that could regulate the *trans*-splicing event. Thus, we set out to investigate both the form 1 and form2 secondary structures of the SL RNA. The *in vitro* secondary structure of the *C. fasciculata* SL RNA was found to be in the form 2 and the *L. collosoma* was found to be in the form 1. The form 1 conformation was examined in detail and was found to contain an interesting tri-uridine hairpin loop with the first and third uridine base paired.

The second part of this thesis examines the hydrodynamical properties (translational and rotational diffusion) of nucleic acids using NMR techniques.

The translational diffusion constants for nucleic acids of different sizes and shapes were measured using the pulsed field-gradient NMR technique. The diffusion constants measured in this way were found to be in good agreement with the predicted values using hydrodynamic theory and to the previously published results from other experimental techniques. This technique is shown to be an effective method for solving one of the more common problems in RNA NMR spectroscopy, knowing whether a particular sample is monomeric or not.

The rotational diffusion constants for nucleic acids of different sizes and shapes were examined theoretically and experimentally by NMR via the nuclear Overhauser effect (NOE) and the relaxation matrix. The theory of the hydrodynamics and relaxation matrix calculations are presented in the context of examining molecules that may undergo anisotropic rotation. The results demonstrate that there is a predictable effect on the measured NOEs because of rotational anisotropy of extended shape molecules, such as long DNA fragments.

1) NMR Studies of the Spliced Leader RNA from *Crithidia fasciculata* and *Leptomonas collosoma*.
2) Hydrodynamic Properties of Nucleic Acids by NMR.

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

in Candidacy for the Degree of

Doctor of Philosophy

by

Jon Lapham

Dissertation Director: Dr. Donald M. Crothers

May, 1998

For Renata.

And my family,

Mom, Dad, Brian and Laura.

Acknowledgements

As I look over this thesis, I realize that I am unusually indebted to a large number of people for helping me with my research project. Without the help these people have shown me thorough the years the research presented in this thesis would not have been possible.

Let me begin by thanking my advisor and mentor Prof. Donald M. Crothers. Don is one of those people that you get the feeling knows some truism of life that the rest of us just don't understand. He is one of the only people I have ever met who portrays that most rare of qualities, confidence in others. I really think that is his secret, he *honestly* believes in his students. What a standard for us to emulate!

The other members of my committee are Profs. Peter Moore, Jim Prestegard and Kurt Zilm. Thank you all for reading this manuscript and for your comments and suggestions along the way. I have had the pleasure of writing a paper with Prof. Moore, and I would just like to say that it was a truly rewarding experience. He believes in brevity and clarity above all else in writing, a lesson I hope I have learned. It is difficult, but sometimes (as I like to say) less is more!

I inherited the Spliced Leader RNA project from Karen LeCuyer and Ken Harris and I would like to thank them both for setting the biophysical groundwork for the project. If it were not for Karen, we would not be talking about the form 1 and form 2 secondary structures of these RNAs! Ken is one of the nicest people I have ever had the pleasure to work with, and I hope that one day soon he is a practicing physician in New York City (a dream come true for him).

I would like to thank Jing Xu for collaborating with me on the RNase H cleavage project (Chap. 2). This was the closest thing I ever came to doing "molecular biology", and she helped me to survive unscathed. I also had the pleasure of interacting with Yi Tao Yu in Joan Steitz's lab when we were trying to figure out where RNase H *really* cuts RNA. (PS: It turns out that we were both correct!)

Renata Kover had the original idea and explained the theoretical background for using the isotope selection experiments to distinguish parts of isotope labeled samples (Chap. 3). She has an unusual ability to understand concepts outside her field and to be able to convey them to others to help them accomplish tasks. All the data from the isotope-selection experiments came from samples prepared by Gauri Dhavan. Getting the pulse sequences to work perfectly was an extraordinarily tedious task, and I couldn't have asked for a better partner then Gauri during the process. Oh yeah, I shouldn't forget to mention that Kevin (spelled with a z!) MacKenzie and John Marino were very helpful with ideas concerning the pulse sequence programming of these experiments.

Chapter 4 is chock full of collaborative efforts. I must first say that the idea of performing the diffusion measurements on DNA came to me whilst listening to a talk Mike Andrec gave on his proton exchange experiment, so, thanks Mike! Jason Rife has the dubious honor of being the first person I found that had an RNA sample that existed as both a monomer and dimer. I wanted data on that system! It turns out that I was very lucky to work with him, because I have very much benefited from the many scientific discussions we have had over the years. Thanks Jason. (and yes, I know that I am "preaching to the choir", but I just like to argue…). The D14 sample presented in chapter 4 was kindly provided by Dan Zimmer, thanks!

If I understand anything of the theoretical aspects of NMR, I owe it to a number of people. It should be a requirement of everyone at Yale to take Kurt Zilm's spectroscopy course, I give the class a ten on the Seminole-Head (SH) rating system. I will always remember the rotating frame demonstration… It should NOT be a requirement that everyone at Yale take Pat Vaccaro's Quantum Chemistry course, but DAMN did I learn a lot about quantum mechanics (and indirectly, about NMR), I also give his class a ten SHs. Pat, you will no longer have to fear me starting conversations with you by saying, "I have this equation…" (at this point he would look around like a hunted animal and try to escape). Finally, I would like to thank Don Crothers (once again) who has such an intuitive understanding of all things physical that his non-equations based explanations of NMR theory usually had the effect of me actually understanding things.

The inertia tensor calculations are compliments of Charlie Schmuttenmaer, thanks for the help!

Much of the data analysis presented in this thesis required writing a number of computer programs. When I came to Yale, I had never even used email, let alone programmed a computer. Since then I have, hopefully, picked up some of the rudimentary aspects of programming from a number of people. To the Perl people:

```
%perl_people = (
    "Dr. Jason Kahn" => "scalars are good",
    "Dr. Mike Andrec" => "lists are better",
    "Dr. Dave Schweizguth" => "associative arrays are best",
);

@people = keys %perl_people;

foreach $person ( @people ) {
    print "I learned that, in Perl, ";
    print $perl_people{$person};
    print " from $person\n";
}
```

To the C (C++) people:

```
#include <iostream.h>
#include "myalloc.h"

main () {
    char *Person1 = NEW1D_C( 16 );
    strcpy(Person1, "Dr Klaus Fiebig");
    char *Person2 = NEW1D_C( 16 );
    strcpy(Person2, "Dan Rosewater");

    char *VARptr;
    char VAR[9] = "pointers";
    VARptr = VAR;

    cout << "Thank you " << Person1 << " and " << Person2 << endl;
    cout << "for showing me how to do dynamic memory allocation\n";
    cout << "and working with " << VARptr << endl;
}
```

Finally, but certainly not the least, I would like to thank my friends and family. I was going to rate everyone the SH rating system described before. It turns out, however, that everyone seems to have fallen into two categories, those rated with 10 SHs (Mom, Dad, Brian, Laura, Renata and Chris) and those with 9 SHs (all others). Hmmm, I guess the SH rating system won't be useful after all.

Mom, Dad, thank you for your support over the eons that I have been in graduate school, it meant more than you will know. To Laura "I can't get enough of chemistry" Lapham and Brian "one day I am going to be bigger than you so I can beat you up" Lapham, it seems that under my expert older-brother supervision you have both turned out to be okay. I hope one day we will live nearer each other, because I actually like to hang out with you guys!

To adequately express my gratitude to the friends I have made here at Yale would require writing another volume to this thesis. I cannot mention everyone, but I must mention the following people. When I rotated in the Crothers lab I worked for one John Marino. For my “project” we synthesized some RNA and collected a 1D imino proton spectrum for it on the 490 MHz NMR. This was back in the days when the 490 was accessed using what amounts to a typewriter (no computer screen). Anyway, John, thanks for what ended up being the beginning of a great friendship. Matthew P. Augustine, this is a transcript of the first conversation I had with you, “Hi, my name is Jon”, said I, “Did I ask?!”, replyeth Matt. You haven’t stopped being a bastard ever since, and I thank you for it. You are an oasis of “state school-ness” in an ocean of “private school”. I am looking forward to the first FSU – PSU Rose Bowl game! Who do those Big-Eleven, PAC-10 teams think they are? Of course, I hope it ends in a tie. Rich Roberts, I enjoyed those late nights dog fighting on the SGIs while you collected stopped-flow data. I will (I swear) get out to California to visit you and Maja one day. Dan Zimmer, you have the unfortunate “honor” of being the guy I usually bounce my “great ideas” off of, thanks for listening to them without laughing too much! (I *will* do acrylamide-based NMR, I swear it!). Jason Kahn, thanks for being smart as hell and a nice guy (hi Effie and Evan!). To Anil and the rest of you damn Canadians, I am happy to say that you NEVER "checked me into the boards" (or whatever it is you say when you know how to speak hockey-ese).

To the other members of the Crothers lab whom I have known over the years, thank you (I would list you all, but damn this acknowledgement is getting pretty long as it is). Oh, what the hell, Razmic, Rosa, Karen, Claudia, Bob, Grace, Ernie, Kate, Ramesh (Clawed thanks you too), Ayesha, Mahdu, Andej, Julie (RNA diffusion woman), Susan, Anna, Dafna, Steve, Jing, Ken, Giorgi, Jayshree, Min, Jessica, Camille, Kevin and Sheela (Kanodia.mac).

To the other people in the various labs in the department, you have made my stay here a memorable one. There are WAY too many of you to mention them all, but thank you.

Chris… Yup… Well then.

Renata Kover may be the only person who has (or will ever) read every single word of this thesis. She, more than anyone else, helped me scientifically and personally. Renata, I love you, thanks.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 7

CHAPTER 8

CHAPTER 1 “NMR STUDIES OF THE SPLICED LEADER RNA FROM *CRITHIDIA FASCICULATA* AND *LEPTOMONAS COLLOSOMA*”

1.1 Summary

This chapter presents nuclear magnetic resonance (NMR) studies on the Spliced Leader RNA (SL RNA) from two species of trypanosomes, *Leptomonas collosoma* and *Crithidia fasciculata*. Previous studies showed that the 5' half of the SL RNAs could possibly adopt two secondary structures, denoted form 1 and form 2 (LeCuyer and Crothers, 1993). Using NMR techniques, the *in vitro* secondary structure of the 5' half of both SL RNAs was determined. The *L. collosoma* was found to exist in the form 1 structure as previously proposed (LeCuyer and Crothers, 1993; Cload, et al., 1993). The *C. fasciculata* SL RNA was found to be in the form 2 structure, a surprising result given the high degree of sequence homology between the two RNAs.

The form 1 hairpin of the *L. collosoma* SL RNA was further examined by synthesizing smaller RNA fragments of the parent 55 nucleotide (nt) molecule. A twenty-five nt and a thirteen nt hairpin were studied. They demonstrate a remarkable feature of this class of RNA, the existence of a tri-uridine hairpin loop in which the first and third uridine are basepaired. Furthermore, two possibly monomeric conformations of the form 1 hairpin were found to exist and can be studied independently by appropriately adjusting the buffer salt conditions.

1.2 Introduction and background

The SL RNAs are found in a variety of lower eukaryotic organisms such as nematodes, euglena and trypanosomes. The two parent SL RNA sequences studied in this chapter are derived from two species of trypanosomes. These organisms are dangerous human pathogens that have an interesting molecular biology.

1.2.1 Trypanosome biology

The trypanosomes are flagellated protozoans of the order *kinetoplastida*. The order is so named for the distinctive large mitochondrial kinetoplast found inside each organism. The trypanosomatids include monogenetic insect parasites (such as *Leptomonas* and *Crithidia*, among others) and digenetic parasites that cycle between insects and plants (*Phytomonas*) or insects and vertebrates (*Trypanosoma*, *Leishmania* and *Endotrypanum*).

The trypanosomes are important human pathogens. For example, transmitted via the bite of reduviid ("kissing") bugs in South America, *T. cruzi*, causes Chaga's disease, in which the invading trypanosome burrows into the heart muscle of the victim. Another trypanosome, *T. brucei gambeinse*, is transmitted by the tsetse fly and causes trypanosomiasis (or sleeping sickness), in which the parasite develops in the bloodstream and eventually enters the nervous system. These pathogens can have very severe epidemic consequences; over 4 million people from the African country of Uganda alone were killed by an outbreak of trypanosomiasis that occurred in 1904.

1.2.2 Pre-mRNA processing in trypanosomes

Aside from their interesting pathogenic properties, trypanosomes also exhibit a unique molecular biology. Unlike other eukaryotes, trypanosomal genes that encode for nuclear proteins lack internal introns, rather, they are excised from polycistronic transcription units solely by *trans*-splicing to the SL RNA and poly-adenylation (Clayton, 1992; Ullu et al., 1995). This *trans*-splicing was first discovered in trypanosomes and has subsequently been found to occur in some of the lower eukaryotes such as nematodes, euglena and trematodes (Kraus and Hirsh, 1987; Blumenthal and Thomas, 1988) as well.

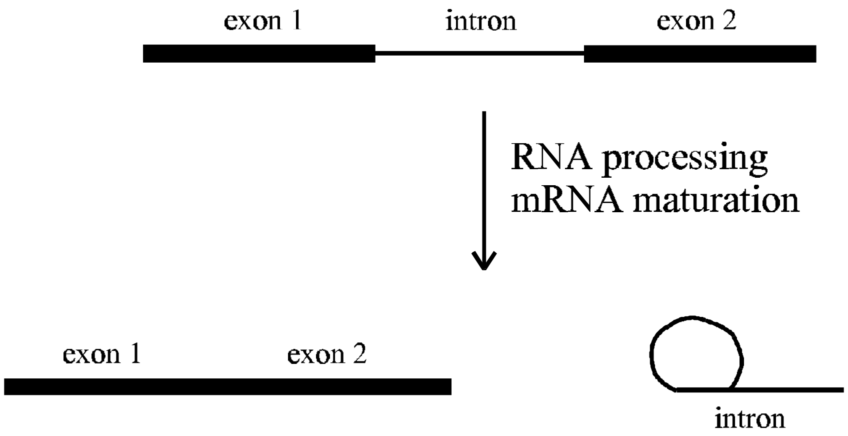


Figure 1. 1 mRNA processing by *cis*-splicing

The molecular mechanism for *trans*-splicing is analogous to that of *cis*-splicing. Most eukaryotic cells, including mammalian, excise introns from mRNA via a *cis*-splicing mechanism to produce a mature mRNA message ready for translation (Fig 1.1) (Padgett et al., 1986; Maniatis and Reed, 1987). In this splicing scheme, the snRNP associates with the intron-exon boundaries of the mRNA and a branch point in the intron. The intron is looped out of the mRNA via a series of *trans*-esterification reactions and the two splice sites are ligated together (Gutherie, 1991). The substrate for this reaction is a single RNA molecule and excision proceeds to give a mature mRNA molecule in which the intron has been removed. An RNA "lariat" is formed when the 5' splice site phosphate is ligated to the 2' hydroxyl of the branch point nucleotide, which leaves a 3' hydroxyl at the 5' splice site nucleotide. The phosphate of the 3' end of the intron is then ligated to this 3' hydroxyl of the 5' exon, excising the intron RNA in the "lariat" shape.

In the *trans*-splicing reaction (Fig 1.2), a spliced leader (SL) exon is joined to the 5' end of a mRNA coding region on a separate transcript (Konarska, et al., 1985; Murphy, et al., 1986; Sutton and Boothroyd, 1986). The SL exon (20-35 nts) is derived from the full length SL RNA (130-220 nts), which exists as a small ribonucleoprotein particle (or snRNP) (Michaeli, et al., 1990; Cross, et al., 1991). The products of the reaction are a mature mRNA transcript "capped" by the SL RNA exon and a Y-branched mRNA/SL RNA intron molecule.

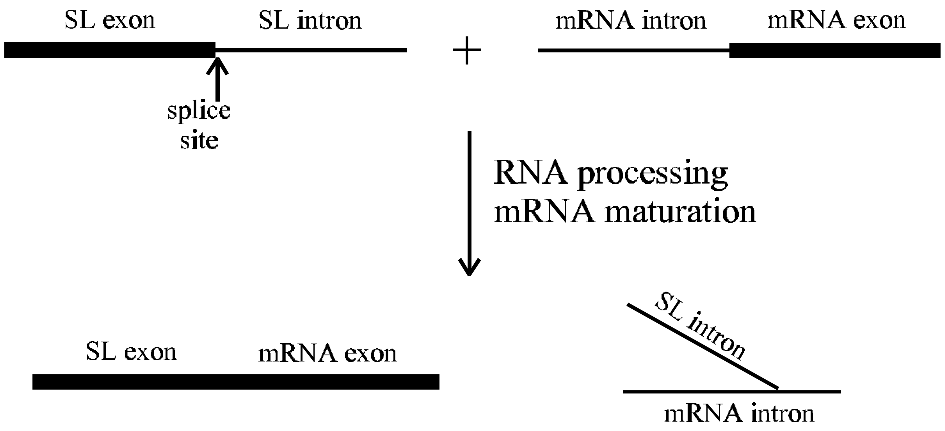


Figure 1. 2 mRNA processing by *trans* splicing

It is not well understood what is the functional role of the *trans*-splicing mechanism in the processing of typanosomal mRNA, or what role the post-spliced SL RNA exon plays. Since the same SL exon is spliced onto the 5' end of all mRNAs, it is speculated that the SL RNA could function to protect the mature transcript from degradation or is involved in signaling the cell to transport the message out of the nucleus. The argument that the SL RNA protects the mRNA from degradation is supported by the fact that the 5' end of SL RNA contains a number of methylated nucleotides, ^{7}Gpppm$_{2}^{6}$ A(2'Om) A(2'Om) C(2'Om)m^{3} U(2'Om) (Perry, et al., 1987;

Freistadt, et al., 1988; Bangs, et al., 1992), which may delay degradation of a mature transcript.

1.2.3 The secondary structure of the SL RNA

Analysis of the primary sequence of the SL RNA from the trypanosome *L. collosoma* originally predicted that the secondary structure of the full length RNA is as shown in figure 1.3 (Bruzik, et al., 1988). This secondary structure was based on calculations of the relative free energies of the base pair formation using the secondary structure prediction program *fold* (Zuker, 1981, 1989), and on the nucleotide sequence conservation between SL RNAs from different species.

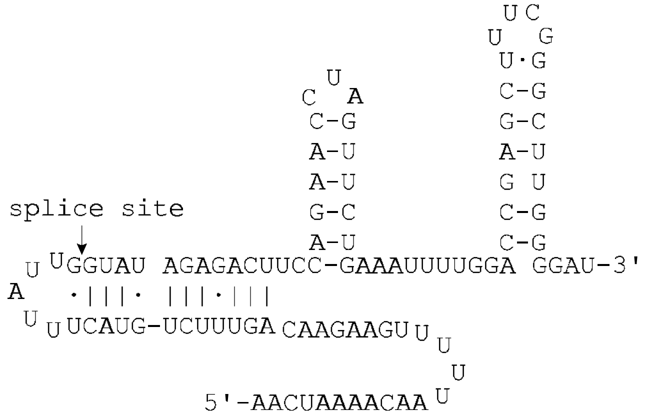


Figure 1. 3 *L. collosoma* Form 2 secondary structure

Further analysis of this RNA, however, revealed that the secondary structure proposed by Bruzik correctly identified the secondary structure of the 3' half of the molecule, but not the 5' half. Using T-jump, native polyacrylamide gel electrophoresis and optical melting experiments, another secondary structure was proposed for the 5' half of the RNA (Fig 1.4) (LeCuyer and Crothers, 1993; Cload et al., 1993). This new secondary structure was named the “form 1” of the molecule as it is the preferred *in vitro* structure, and the originally proposed secondary structure was named the “form 2”. Additionally, the two secondary structures, form 1 and form 2, were found to have nearly the same thermodynamic stability and could interconvert on a fast (<1s) time scale when forced to do so by binding complementary oligonucleotide probes (LeCuyer and Crothers, 1993).

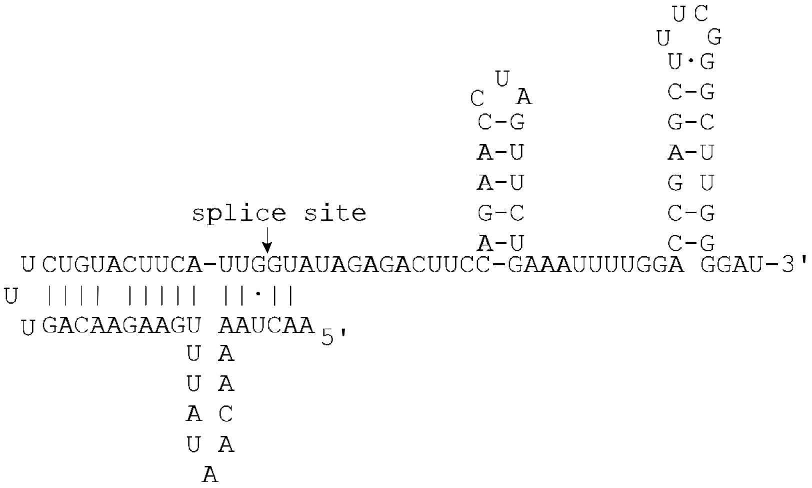


Figure 1. 4 ***L. collosoma*** **Form 1 secondary structure**

Additionally, the form 1 secondary structure has been shown to contain a biphasic UV hypochromic shift melting profile (Fig 1.5), with an anomalous low temperature transition. This early transition in the optical melt has been suggested to be due to some type of higher-order structure (LeCuyer and Crothers, 1993), possibly a tertiary interaction. Both transitions are retained in the melting profile when the 3’ half (the two hairpins) of the molecule is removed. Thus, the structural element responsible for the low temperature transition must be contained in the 60 nt 5’ half of the RNA.

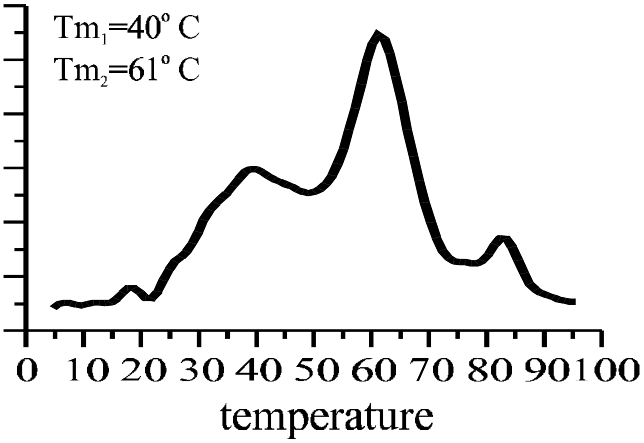


Figure 1. 5 Derivative UV melting curve of the 60 nt 5' half of the *L collosoma* SL RNA

Sequence analysis of the SL RNAs of other trypanosomes (Fig 1.6) shows that this ability to adopt two secondary structures may be a common feature. If the ability to adopt two alternate secondary structures is a common feature of all trypanosomal SL RNAs; this raises the question of why. Steitz (1992) has proposed a model of *trans*-splicing that incorporates components of both structure models in which the structural switch between form 1 and 2 mimics the functions of the U1 and U5 RNAs found in *cis*-splicing. Also, it has been noted (LeCuyer and Crothers, 1993) that the SL exon is extensively basepaired to the intron while in form 2, but there is virtually no base pairing in form 1. Thus, form 1 may be a method of disrupting interactions between the SL RNA intron and a mRNA.

```
AACUAAAACAAUUUUUGAAGAACAGUUUCUGUACUUCAUUGGUAUGGUAUGUAGAGACUUC     L. collosoma

AACUAACGCUAUUAUUAGAA--CAGUUUCUGUACUAUAUUGGUAUGAGAAG------CU       T. brucei
AACUAACGCUAUUAUUGAUA--CAGUUUCUGUACUAUAUUGGUACGCGAAG------CUU      T. cruzi
AACUAACGCUAUUAUUGAUA--CAGUUUCUGUACUAUAUUGGUAUGCAGCG------CUUC     T. rangeli
AACUAAAGCUUUUAUUAGAA--CAGUUUCUGUACUAUAUUGGUAUGAGAAG------CU       T. malayi
AACUAAAGCUUUUAUUAGAA--CAGUUUCUGUACUAUAUUGGUAUGAGAAG------CU       T. vivax
AACUAACGCUAUAUAAGUAU--CAGUUUCUGUACUUUAUUGGUAUGCGAAAC-----CUU      L. enreittii
AACUAACGCUAUAUAAGUAU--CAGUUUCUGUACUUUAUUGGUAUGCGAAA------CUUC     L. mexicana
AACUAACGCUAUAUAAGUAU--CAGUUUCUGUACUUUAUUGGUAUGCGAAA------CUUC     L. donovani
AACUAACGCUAUAUAAGUAU--CAGUUUCUGUACUUUAUUGGUAUGAGAAG------CUUC     L. seymouri
AACUAACGCUAUAUAAGUAU--CAGUUUCUGUACUUUAUUGGUAUAAGAAG------CUUC     C. fasciculata
```

Figure 1. 6 Sequence analysis of trypanosomal SL RNAs

Interestingly, *in vivo* analysis of the *L. collosoma* and *T. brucei* SL RNAs using water-soluble chemical modification probes has shown that the form 2 structure predominates (Harris, et al., 1995). Thus, in the context of the snRNP, form 2 seems to be favored, while the *L. collosoma* RNA alone *in vitro* favors the form 1 structure. Further, it was shown that the methylated nucleotides on the 5’ end of the SL RNA do not play a structural role *in vivo* (Harris, et al., 1995); however, the methyl groups are required for the *trans*-splicing reaction to occur (Ullu and Tschudi, 1991, 1993; McNally and Agabian, 1992).

1.2.4 Project goals

The goal of this project is to investigate the structural features of the SL RNAs from two species of trypanosomes, *L. collosoma* and *C. fasciculata*, by NMR. Two main interests were pursued in the NMR investigations. The first was in determining the *in vitro* secondary structures of the SL RNAs derived from the two species. The second was in finding and characterizing smaller structural fragments derived from the parent RNAs, in the hope that these smaller fragments might prove to be structurally interesting and tractable by NMR methods.

In this report, NMR techniques were used determined the *in vitro* secondary structure of the 5' halves of both SL RNAs. The secondary structure of the *L. collosoma* sequence was confirmed to be in the form 1 secondary structure as predicted by the previous biophysical studies. The SL RNA from *C. fasciculata* was found to be in the form 2. This result is interesting in that most of the SL RNAs have a closer sequence homology to the *C. fasciculata* SL RNA, and possibly the *L. collosoma* SL RNA is the only one found in the form 1 structure *in vitro*.

Further studies were carried out on the form 1 SL RNA from *L. collosoma*, including $^{15}N/^{13}C$ isotope labeling the 55 nt 5' half and characterization of smaller fragments that contain only the form 1 hairpin. The form 1 hairpin was found to contain an unusual feature, a three-uridine loop, with the first and third uridine base paired. This is a surprising result in that it would require the hairpin loop to be spanned by a single nucleotide. Because of this, the possibility of dimerization of the RNA was investigated. The evidence favors that the U=U basepair is found in monomeric RNA, but more work needs to be done to prove that the molecularity is one.

The smallest fragment of the *L. collosoma* SL RNA studied, a thirteen-nucleotide hairpin, was found to exist in two conformations in slow exchange. The ratio of the concentrations of the two conformations was found to be a function of the ionic strength of the solution. Clearly, the possibility exists that the conformational change may be a monomer-dimer exchange, and both biophysical and NMR methods were utilized to investigate this possibility. Both of the two conformations were studied individually and characterized by NMR.

1.3 Results

Five NMR samples were synthesized for study. The names of the samples are derived from the species from where they came, and the length of the RNA. Thus, the sample "rLC55" is an RNA derived from the *L. collosoma* sequence and is 55 nts long. Figure 1.7 shows a complete listing of the samples, their names and the numbering scheme used in identifying the nucleotides

1.3.1 Choice of L. collosoma *experimental samples*

The full length (130 nt) and the 5' half (52 nt) of the SL RNA from the species of trypanosome, *L. collosoma*, (Fig. 1.3 and 1.4) have been extensively studied by biophysical methods. These studies demonstrated that the 5' half of the RNA is structurally independent of the 3' half (LeCuyer and Crothers, 1993) and that the 5' half is still characterized by the biphasic UV melt (Fig. 1.5). Given that the "tertiary" structural elements exist on the 5' half of this SL RNA, the wild type 52 nt 5' half of the *L. collosoma* SL RNA has been selected for studies by NMR. An additional 3 guanine residues were added to the 5' end of the RNA to increase the yield on the transcription reactions, as has been suggested previously (Milligan, et al, 1987).

The parent rLC55 sample is interesting because it represents the SL RNA before the splicing event and spans the splice site. Also of interest is what structure the SL RNA exon (30-40 nts) will adopt after the *trans*-splicing event. The SL exon can only adopt the form 1 hairpin, because the form 2 base pairing occurs on the 3' side of the splice site (Figs 1.3 and 1.4). This form 1 hairpin is presumably the structural element that may be recognized by cellular machinery responsible for transport of the mature mRNA out of

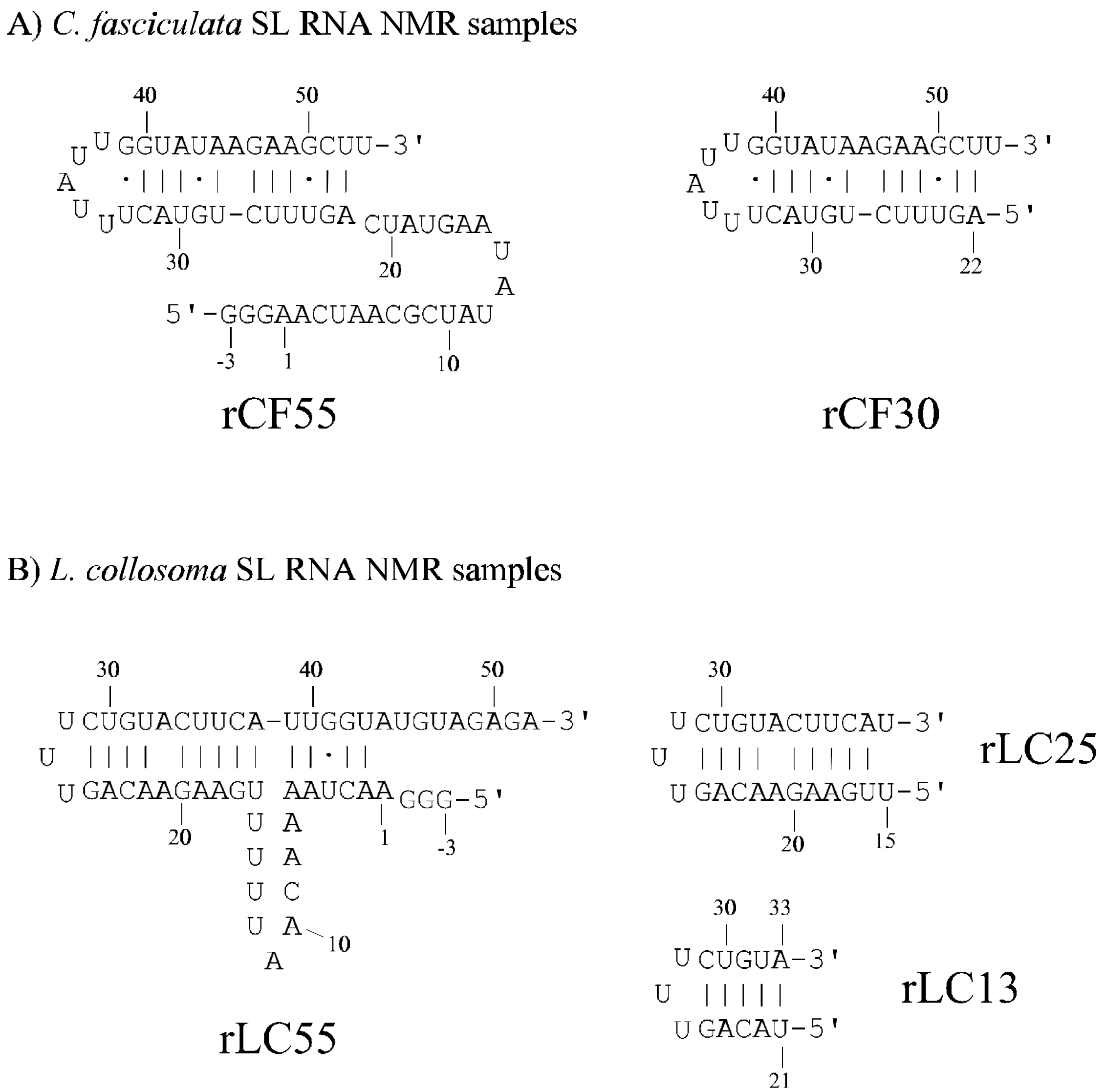


Figure 1.7 NMR samples

A) Two NMR samples were synthesized from the *C. fasciculata* SL RNA sequence, the 55 nt 5' half of the SL RNA (rCF55) and the form 2 hairpin (rCF30). The wild type sequences were used, except that three guanine nucleotides were added to the 5' end of the rCF55 sample to improve transcription yield. B) Three NMR samples were constructed from the L. collosoma SL RNA sequence. The parent 55 nucleotide 5' half SL RNA (rLC55) and two smaller form 1 hairpin fragments (rLC25 and rLC13). The wild type sequences were used except for the addition of three guanine residues on the 5' end of the rLC55 sample and the switching of A21 to U21 in the rLC13 sample to maintain a base-pairing interaction at the terminus of the hairpin.

the nucleus. For this reason, smaller 25 and 13 nt (rLC25 and rLC13) form 1 hairpins were constructed for more detailed spectroscopic study. While these are derived from the *L. collosoma* sequence, they are fairly well representative of all the trypanosome form 1 hairpins given the high level of sequence conservation in this region of the SL RNAs. These samples also have the advantage of being much smaller then the parent SL RNA molecules, which makes them better suited for high resolution NMR characterization.

1.3.2 Choice of C. fasciculata *experimental samples*

When analyzing the sequence homology between the known trypanosome SL RNA sequences (Fig. 1.6) it is clear that two major "sequence classes" exist, one representative of the *L. collosoma* sequence and one representative of all the other species of trypanosomal SL RNAs. For this reason, it seemed appropriate to investigate the SL RNA of the sequences in this latter class. Thus, a second sequence of the SL RNA was chosen for study, the 52 nt of the 5' half of the wild type SL RNA from *C. fasciculata*, rCF55. As with the rLC55 sample, three guanine nucleotides were added to the 5' end of this RNA to increase the transcription yield. While no biophysical studies have been performed on the RNA, it was inferred from the sequence homology with the *L. collosoma* SL RNA that the properties of the two sequences would be similar.

The *C. fasciculata* form 2 hairpin, rCF30, was found serendipitously during the construction of a "segmentally" labeled version of the rCF55 sample. This is discussed in greater detail in the next section of this chapter.

1.3.3 Determining the secondary structure of nucleic acids by NMR

Determination of the secondary structure of nucleic acids by NMR generally involves analysis of the solvent-exchangeable imino proton spectra. These experiments are conducted in H_2O, where the imino is observable only when it is protected from fast exchange with bulk solvent, such as when it is involved in a hydrogen bond in a standard Watson-Crick base pair. Thus, the existence of an imino proton may indicate that there exists some form of a secondary structure for that region of the molecule.

Two NMR experiments are primarily used to analyze the imino protons. The two-dimensional (2D) ^1H-^1H H_2O NOESY experiment is used to give the connectivities from an imino to its nearest neighbors, possibly to the imino in the next basepair. The second experiment is the 2D ^1H-^{15}N HMQC (Szewczak, et al., 1993). This experiment correlates the imino proton to the chemical shift of the nitrogen to which it is directly attached. This is important in the assignment of the imino protons since the nitrogen of the purines (guanine) and pyrimidines (uridine) have unique chemical shifts. Thus, the base-identity of each imino can be established based solely on the distribution of the ^{15}N chemical shifts.

This commonly used approach of analyzing the imino NOESY pattern to determine the secondary structure of nucleic acids failed to work for the SL RNAs studied. The problem lies in the fact that the information derived from the aforementioned NMR experiments is an imino proton pattern such as "GUUGU". If this pattern can exist in more then one region of the RNA, it is difficult to unambiguously make an assignment of the secondary structure. The SL RNAs can possibly adopt either the form 1 or form 2 secondary structures, as mentioned before, both of which share a common stretch of the RNA as shown below (Fig. 1.8). For this reason, depending on what nucleotides are bulged out of the helix, the connectivities of the imino NOESY experiment could not uniquely identify one of the two possible secondary structures.

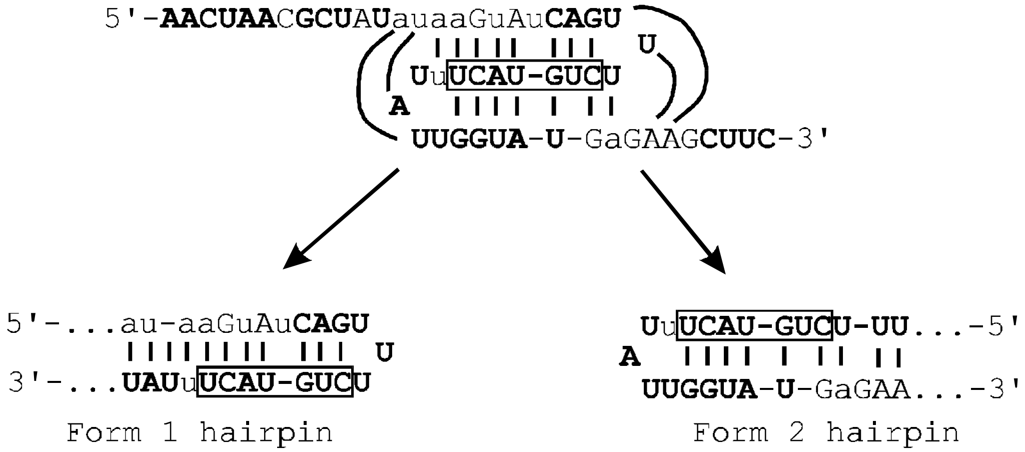


Figure 1.8 Consensus "central core" nucleotides in form 1 and form 2

Because of this problem of assigning the iminos in the "central core" region of the SL RNAs, other methods were used to determine the secondary structures, such as comparison of the spectra of RNA fragments with that of the parent RNA.

1.3.4 The in vitro *secondary structure of the* C. fasciculata *SL RNA*

As was discussed for the *L. collosoma* SL RNA, the *C. fasciculata* SL RNA can adopt both the form 1 and form 2 secondary structure (Fig 1.9).

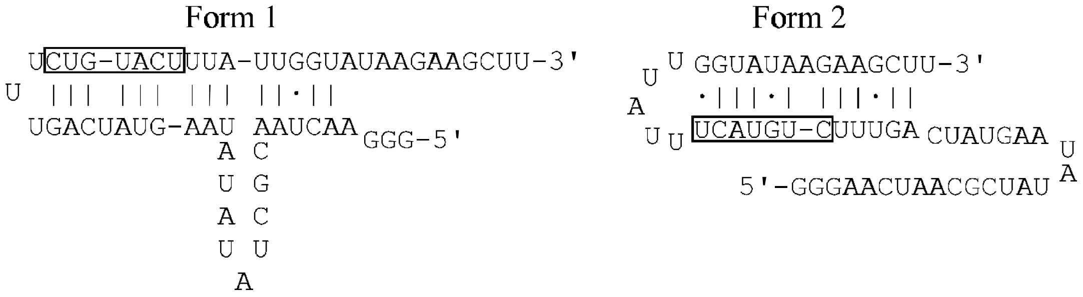

Form 1
UCUG-UACUUUA-UUGGUAUAAGAAGCUU-3'
U
UGACUAUG-AAU AAUCAA GGG-5'
A C
U G
A C
U U
A
Form 2
U U GGUAUAAGAAGCUU-3'
A
U U UCAUGU-CUUUGA CUAUGAA U
A
5'-GGGAACUAACGCUAU

Figure 1.9 *C. fasciculata* 5’ half SL RNA form 1 and form 2 structures

The 2D H_2O NOESY and the 2D ^{1}H-^{15}N HMQC spectra for the rCF55 are shown in figures 1.10 and 1.11 respectively, along with the possible assignments to either the form 1 or form 2 structure. Because of the sequence homology with the *L. collosoma* SL RNA, it was initially assumed that the rCF55 was in the form 1 structure. Upon further analysis of these spectra, it became clear that it was impossible to firmly rule out either the form 1 or form 2 structure based on these imino patterns, thus additional studies were required to elucidate its secondary structure.

One method that could unambiguously determine the secondary structure is the technique of “segmental labeling” in which one section of the RNA is labeled with ^{15}N isotope and the other part contains the natural isotope, ^{14}N. In this manner, a simple isotope selection NOESY experiment could readily distinguish the secondary structure based on the pattern of imino protons which appears in the ^{14}N or ^{15}N subspectra of the experiment (see chapters 2 and 3 of this thesis for further discussion of this approach). The segmental labeling approach has the unique attribute of allowing for the study of a section of an RNA in the context of the full length RNA. This is important for RNAs where interactions between different domains may affect the local environment. This technique for determining RNA secondary structures using segmental labeling was first

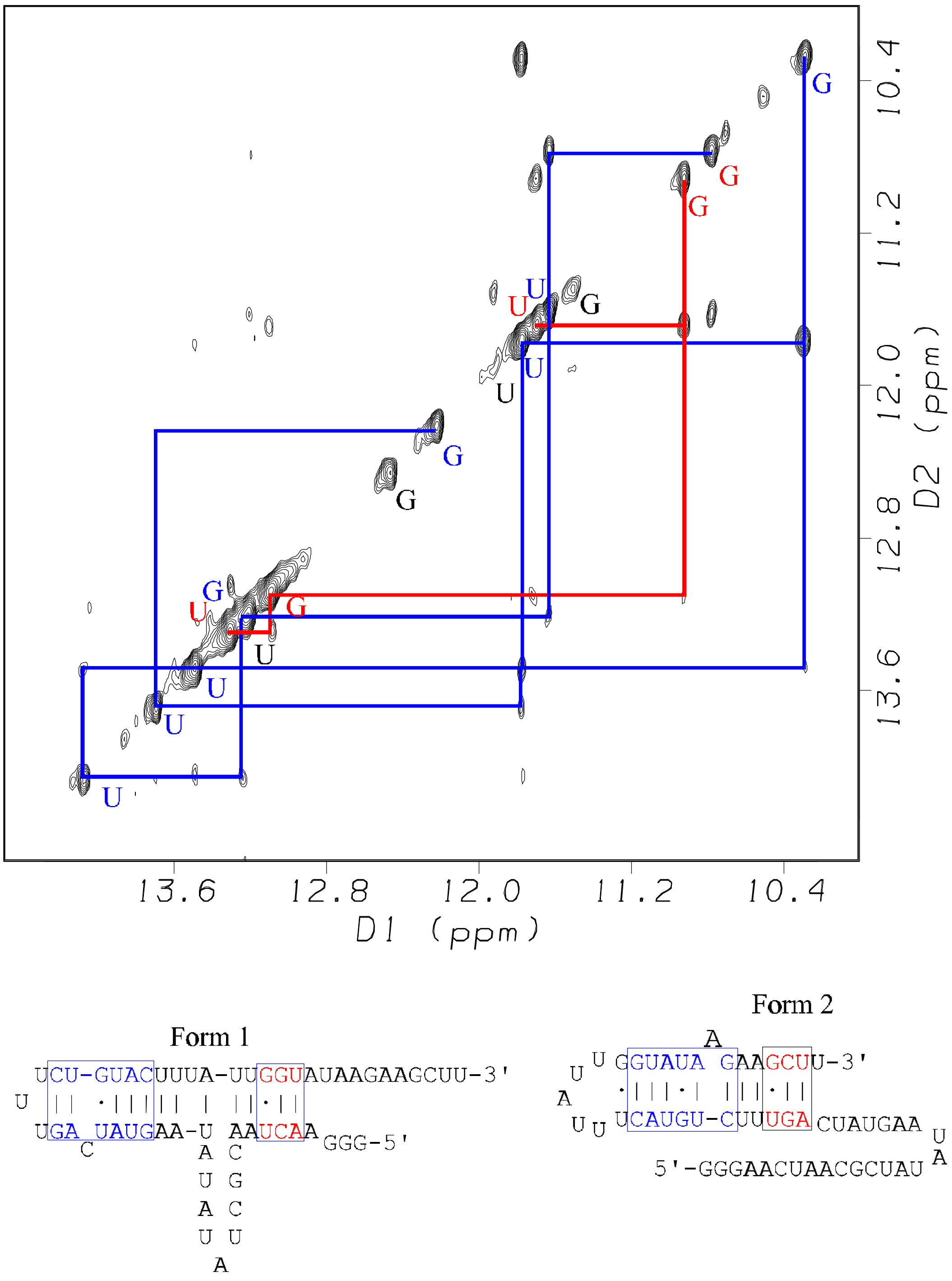


Figure 1.10 H_2O NOESY spectrum of rCF55

The blue and red lines show the two main imino-imino crosspeak connectivity patterns from the JRSE H_2O NOESY experiment. Both the form 1 and the form 2 secondary structure of rCF55 could satisfy these imino crosspeak patterns. One cytosine must be bulged out of the form 1 helix, and one adenine must be bulged out of the form 2 helix to satisfy the connectivities. The experiment was performed at 25°C with a 250 ms mixing time.

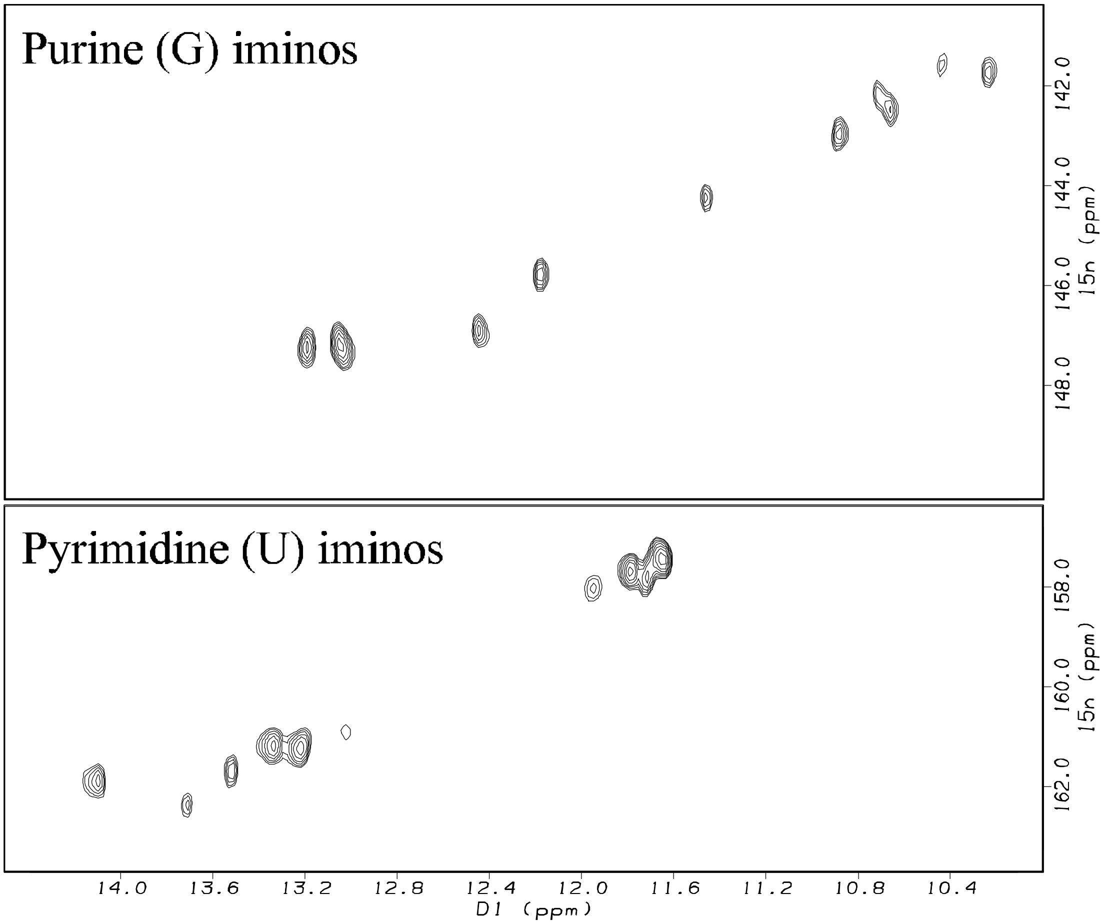


Figure 1.11 rCF55 ^{1}H ^{15}N HMQC

Assignment of the base identity of the imino protons shown in figure 1.10 for rLC55 was based on the imino nitrogen chemical shifts from this HMQC experiment. In total, 5 Watson-Crick base paired uridines, 4 Watson-Crick base paired guanines and 4 guanine-uridine wobble base pairs appear. The rCF55 sample was in 20 mM sodium phosphate buffer (pH 6.5), 150 mM sodium chloride and 1 mM EDTA. The data were collected at 25° C.

utilized and shown to be effective in our lab in the analysis of the secondary structure of the SL RNA from *C. elegens* (Xu, et al., 1996).

It became apparent that it would not be necessary to make the segmentally labeled rCF55 RNA during the process of analyzing one of the segments. The 3' end ^{15}N labeled segment of the RNA, rCF30, contained the entire form 2 hairpin and showed nearly an identical 2D ^{1}H-^{15}N HMQC spectrum to that of the full length RNA. Figure 1.12 shows the comparison of the rCF55 and rCF30 HMQC data. A few iminos found in the rLC55 spectrum are absent in the rLC30 spectrum, the third G imino from the left, the third U imino from the left and one of the G=U base pairs.

Since there is considerable sequence homology between the *C. fasciculata* and the *L. collosoma* SL RNA (Fig. 1.6) we wanted to further confirm the hypothesis that the SL RNA of *C. fasciculata* exists in the form 2 structure *in vitro*. With that goal in mind, constant temperature native gel analysis (see materials and methods) of form 1 and form 2 mutants was performed. The results shown in figure 1.13 confirm that the wild type sequence runs with the same mobility as the form 2 mutant RNA. Oddly, the relative mobility of the form 1 and form 2 at 25° C appears to be the inverse of what is seen for the *L. collosoma* SL RNA (LeCuyer and Crothers, 1993). However, running the gel at 10° C inverts the relative mobilities of the form 1 and form 2 *C. fasciculata* SL RNAs.

1.3.5 The in vitro *secondary structure of the* L. collosoma *SL RNA*

The secondary structure of the *L. collosoma* SL RNA has been well characterized and has been shown to exist *in vitro* as form 1 (LeCuyer and Crothers, 1993; Cload, et al., 1993). The buffer conditions of these studies was typically from pH 6 to 7.5 and between 50 to 200 mM sodium chloride. Choosing optimal NMR buffer conditions is important

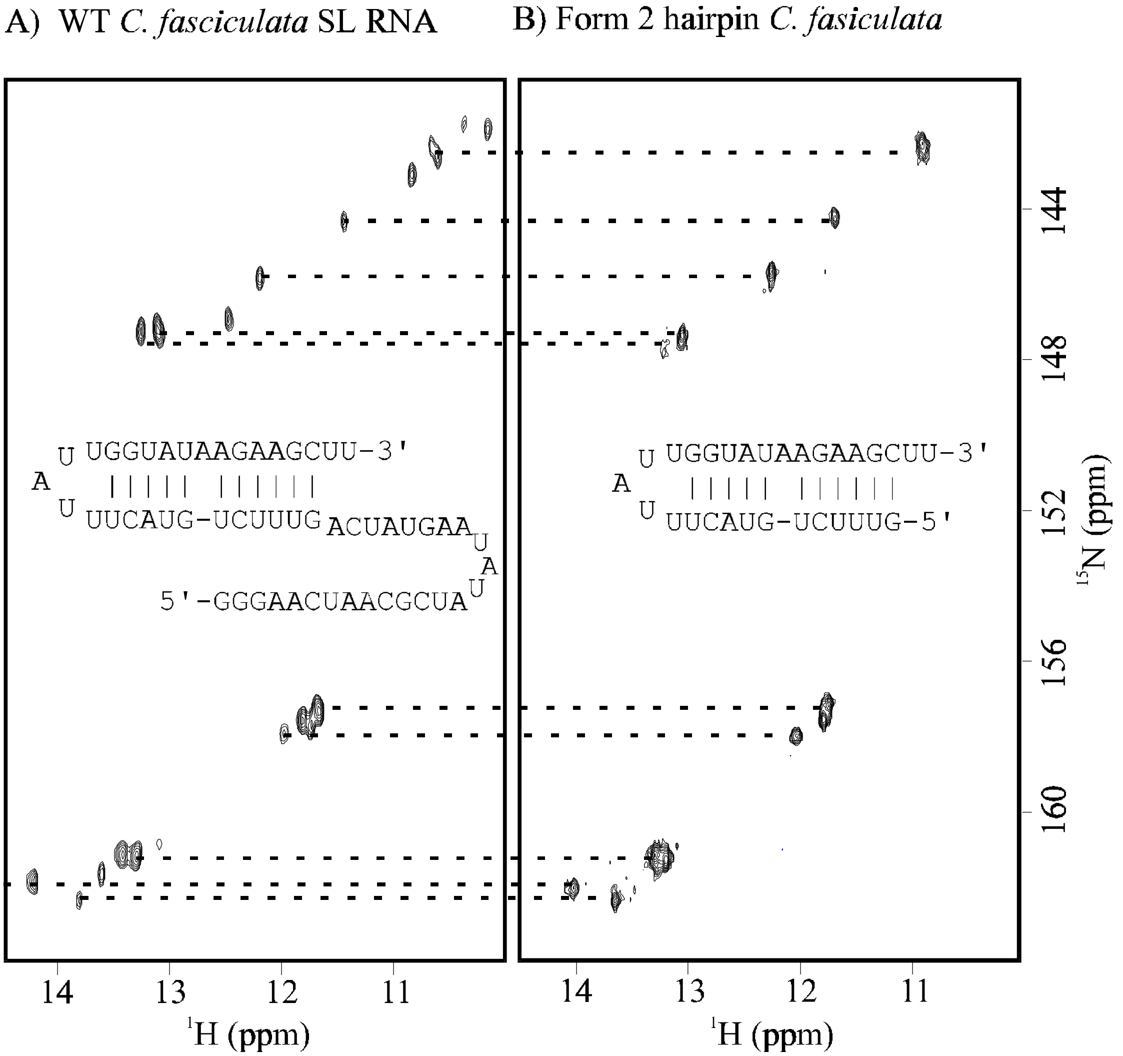


Figure 1.12 HMQC comparison between rCF55 and form 2 hairpin

A) The ^{1}H-^{15}N HMQC spectrum of the rCF55 sample and B) the spectrum of the rCF30 sample (form 2 hairpin). The buffer conditions were identical for both sample, 20 mM sodium phosphate (pH 6.5), 150 mM sodium chloride and 1 mM EDTA. The secondary structures of each sample are shown above. The comparison of the chemical shifts of both the imino protons and the nitrogens in these two spectra allowed for the unambiguous assignment of the *in vitro* secondary structure of the *C. fasciculata* SL RNA to the form 2. Both spectra were collected at 25°C.

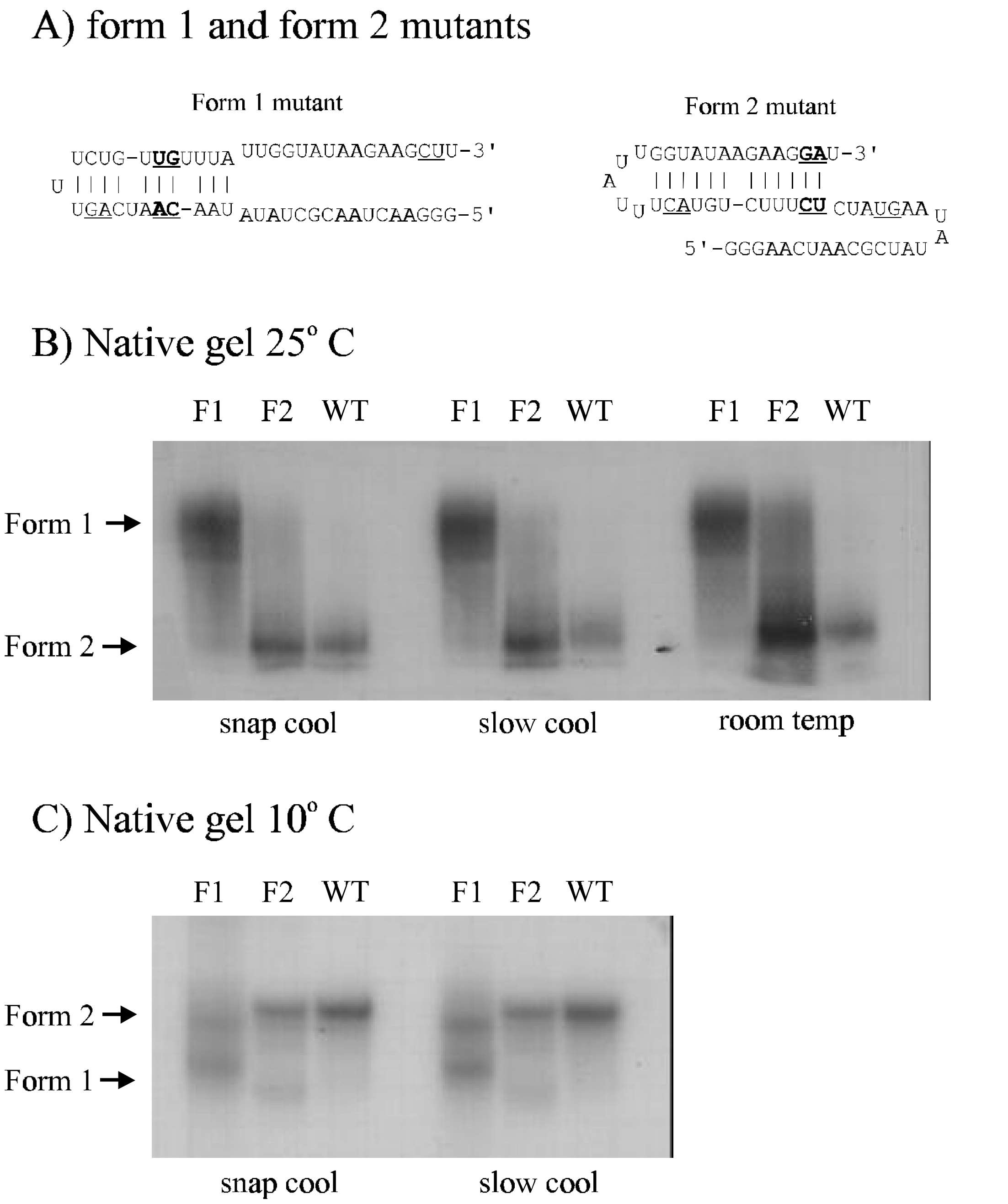


Figure 1.13 *C. fasciculata* SL RNA form 1 and form 2 mutant native gels

A) The form 1 and form 2 mutant sequences used in the native gel mobility study. The bold lettered regions of the sequence represent a position where the basepairs were inverted from their wild type positions. This should have no effect on the desired secondary structure, while inhibiting formation of the other structure. B) Room temperature native gel and C) 10°C native gel. The titles on the lanes represent either the form 1 or form 2 mutant sequence or the wild type rCF55 sequence. Samples were annealed using either snap cooling or slow cooling techniques (see materials and methods), with no effect on the results.

for obtaining well resolved (and meaningful) spectra. The buffer used in these NMR studies was chosen to be sodium phosphate, because it lacks any protons to interfere with the spectra and is slightly acidic, pH 6.5, to favor slower imino proton exchange. Two different sodium chloride salt concentration buffers were studied, a low salt (~30 mM [NaCl]) and a high salt (~130 mM [NaCl]).

The 1D imino proton temperature melt data are shown for both the low and high salt buffers (Figs. 1.14 and 1.15). A number of features can be seen in comparing the two experiments. The iminos in the low salt buffer begin to exchange broaden at approximately 45 degrees, while those in the high salt buffer are still strong at the same temperature. The iminos later identified to be G25, U30, U26 and U28 disappear in the low salt buffer at approximately 30 degrees, while they remain intense in the high salt buffer. The high salt buffer is clearly the better NMR candidate, at least in terms of the spectroscopy of the imino protons, and 130 mM [NaCl] was consequently chosen for further studies.

UV melts were performed on the rLC55 sample using same NMR buffer conditions (Fig. 1.16) to determine whether it displays a similar biphasic melting profile as seen by LeCuyer and Crothers. The concentration of the RNA was varied from 0.6 μ M to 120 μ M to look for signs of concentration dependent aggregation effects. The UV melts appear to be similar and the measured Tm for both transitions are the same at 40° and 61° C.

The 2D H_2O NOESY (Fig 1.17) and the ^1H-^{15}N HMQC (Fig. 1.18) spectra of rLC55 are in agreement with the assignment of the form 1 secondary structure, and the imino proton assignments are shown. Two main regions of imino proton connectivities

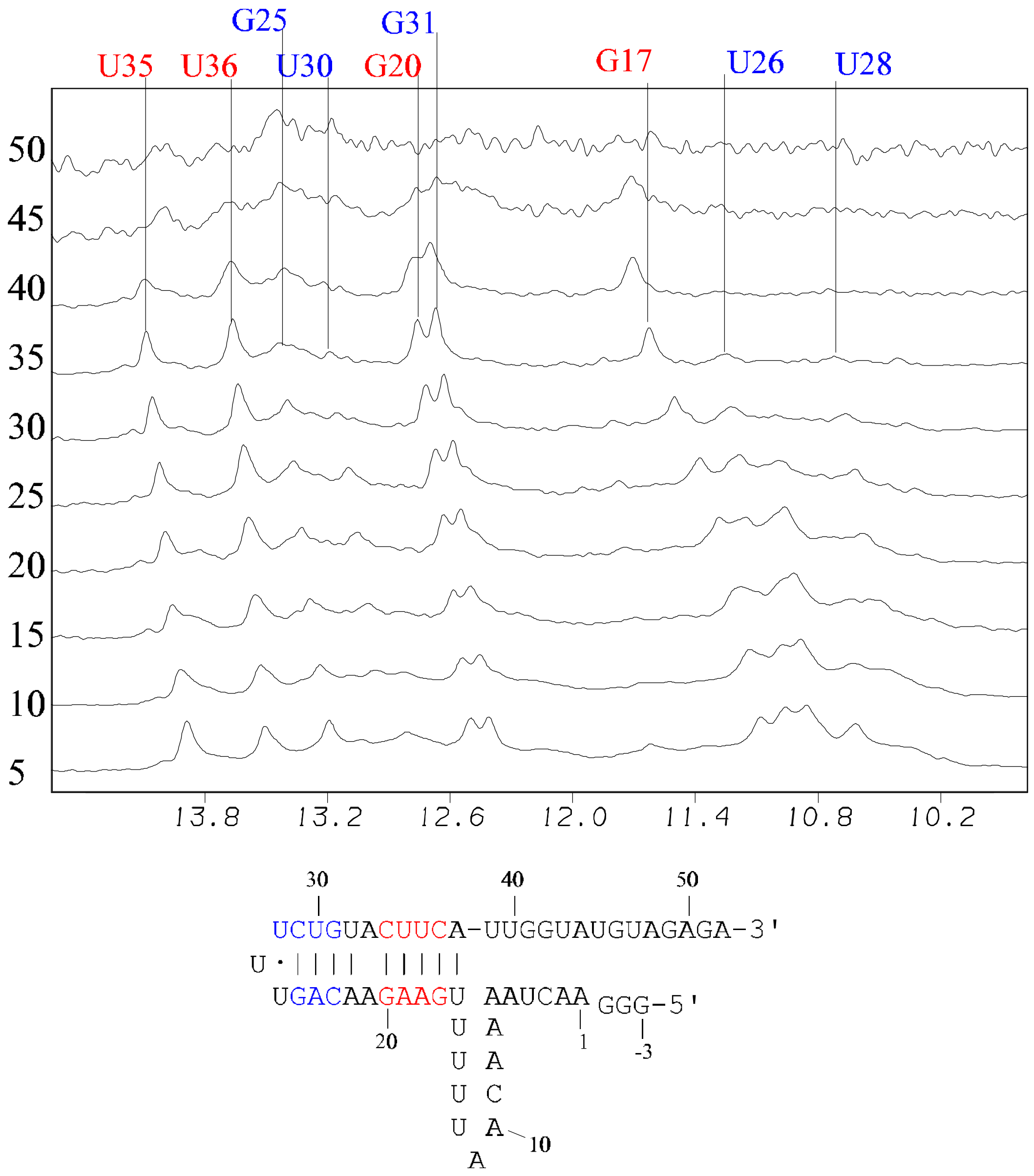


Figure 1.14 Low salt rLC55 imino proton temperature study

The low salt buffer conditions used in this imino proton temperature study were 10 mM sodium phosphate buffer (pH 6.5) and 1 mM EDTA, which is approximately 30 mM in sodium ions. The iminos G25, U30, G17, U26 and U28 appear broad and featureless.

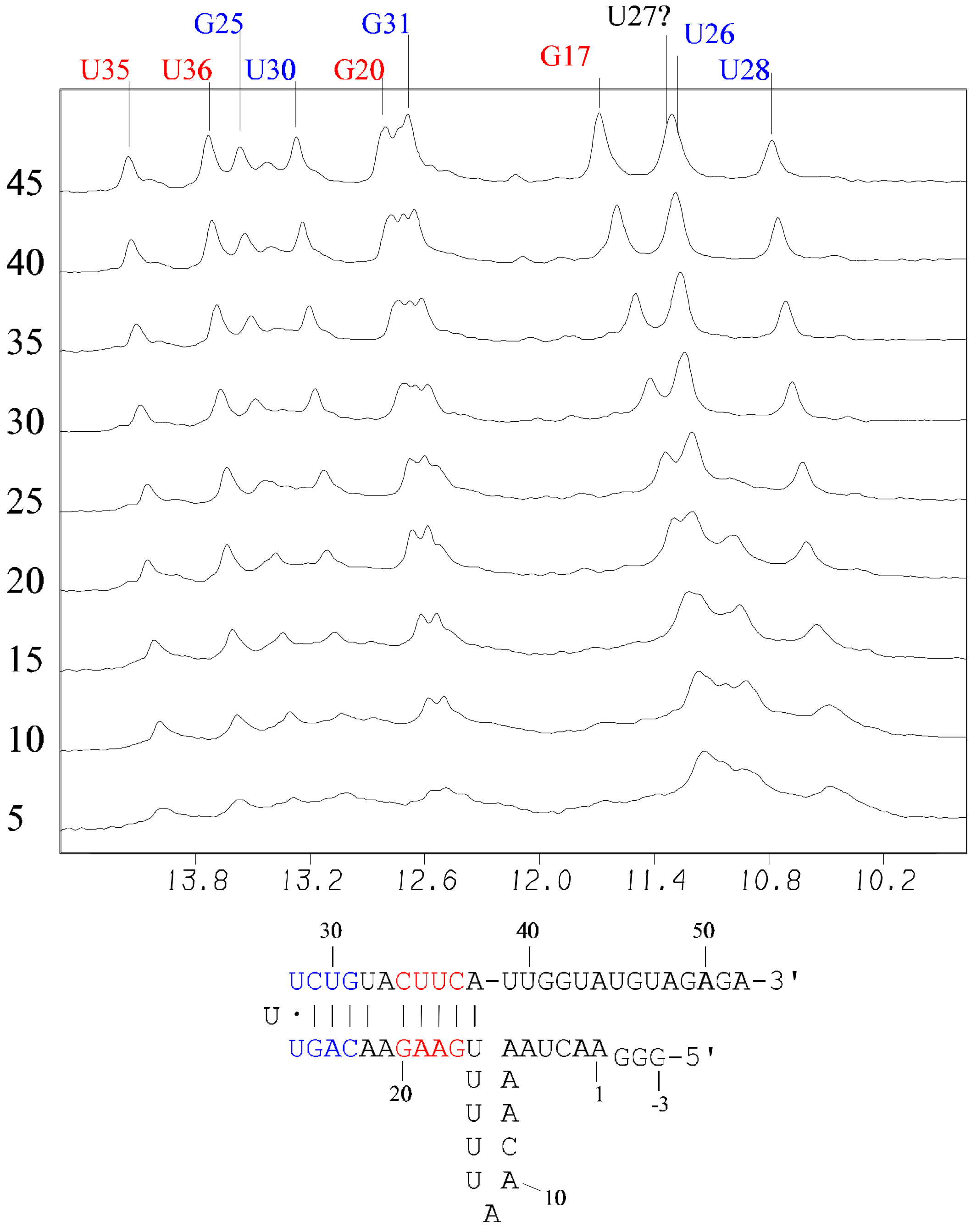


Figure 1.15 High salt rLC55 imino proton temperature study

The high salt buffer conditions used in this imino proton temperature study was 20 mM sodium phosphate buffer (pH 6.5), 100 mM NaCl, and 1 mM EDTA, giving approximately 160 mM in sodium ions. The G25, U30, G17, U26 and U28 iminos appear much more intense and sharp as compared to the iminos found in the low salt buffer.

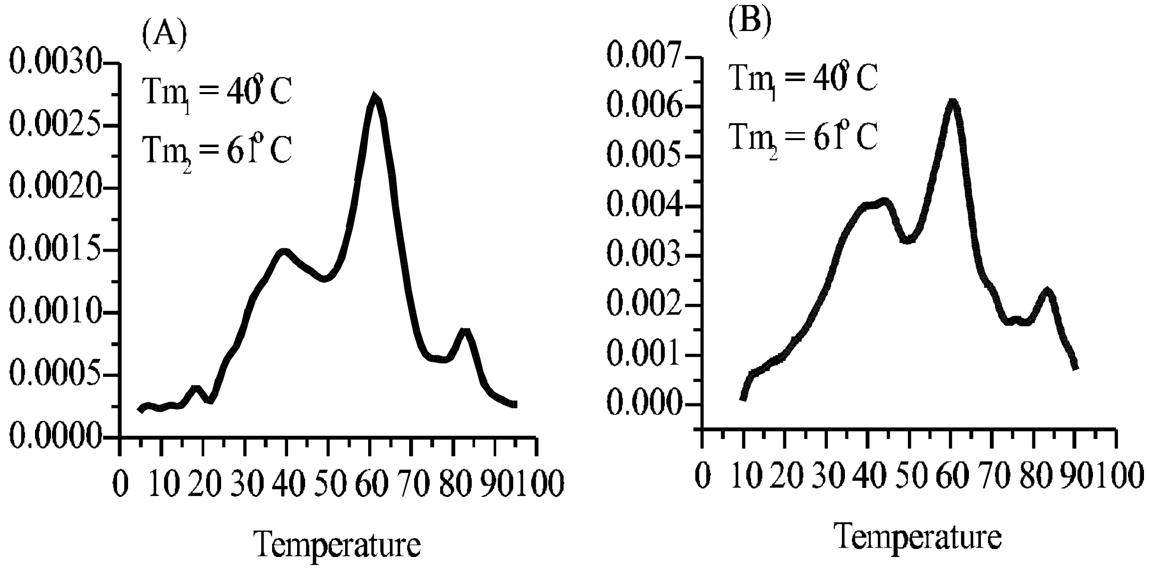


Figure 1.16 High salt rLC55 derivative UV melts

The UV melting curves of rLC55 are presented in the buffer conditions chosen for the NMR studies, 20 mM sodium phosphate (pH 6.5), 100 mM NaCl and 1 mM EDTA. More complete studies, with varied buffer conditions, have been performed (LeCuyer, 1992; Harris, 1994) and the data will not be duplicated here.

A) Low concentration UV melt, [rLC55]=0.6 μM in a 10 mm cuvette. B) High concentration UV melt, [rLC55]=120 μM in a 1 mm “etched quartz” cuvette. There appears to be no appreciable change in the melting profile, indicating that at these concentrations the rLC55 sample is not involved in duplex aggregation.

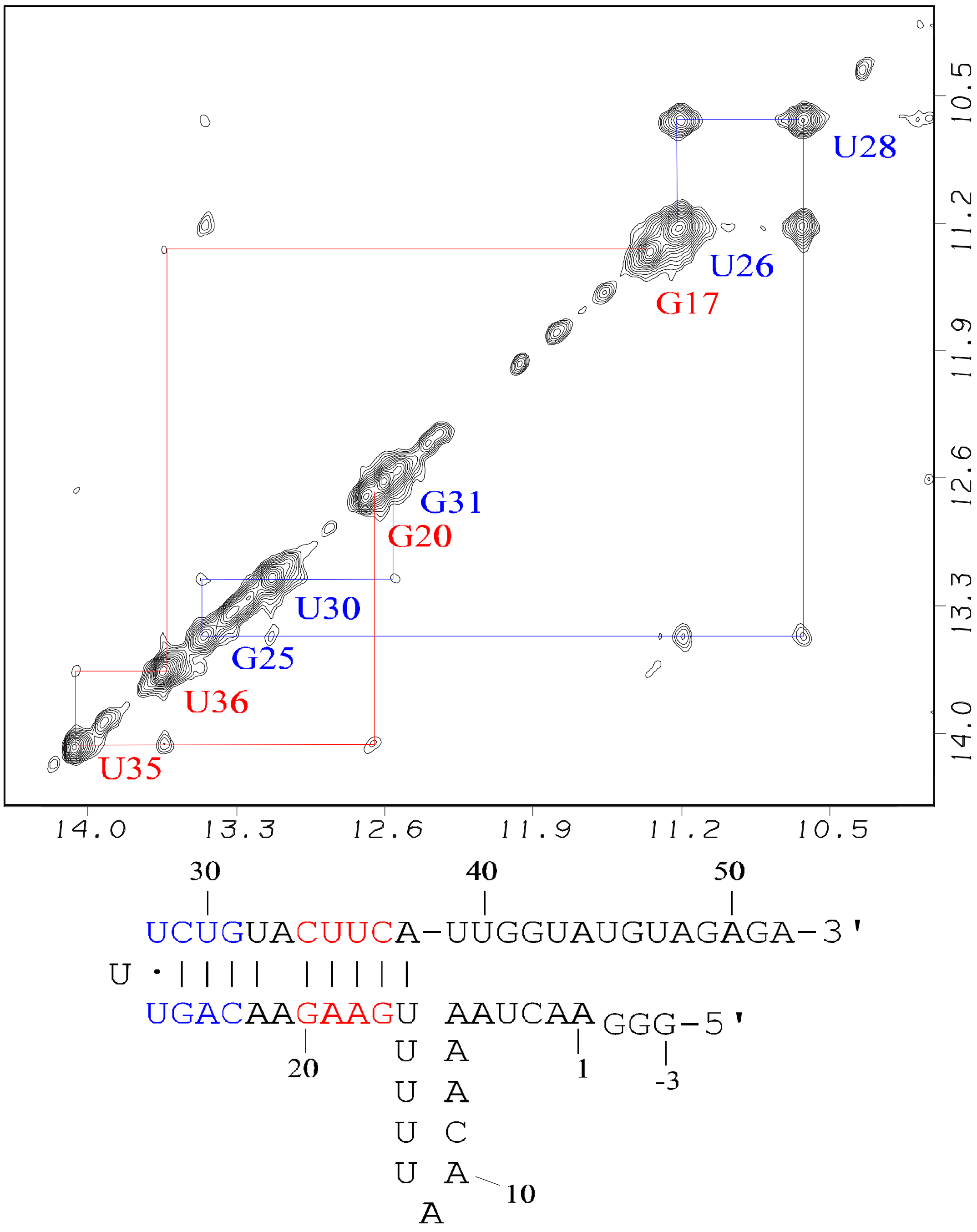


Figure 1.17 rLC55 H_2O NOESY

Jump-Return Spin Echo water suppressed H_2O NOESY spectrum on the rLC55 SL RNA at 25° C. The pattern of iminos fits with the form 1 secondary structure as shown above. The imino protons most stabilized by the higher salt conditions (see Figs. 1.15 and 1.16) are mapped to the loop region of the hairpin.

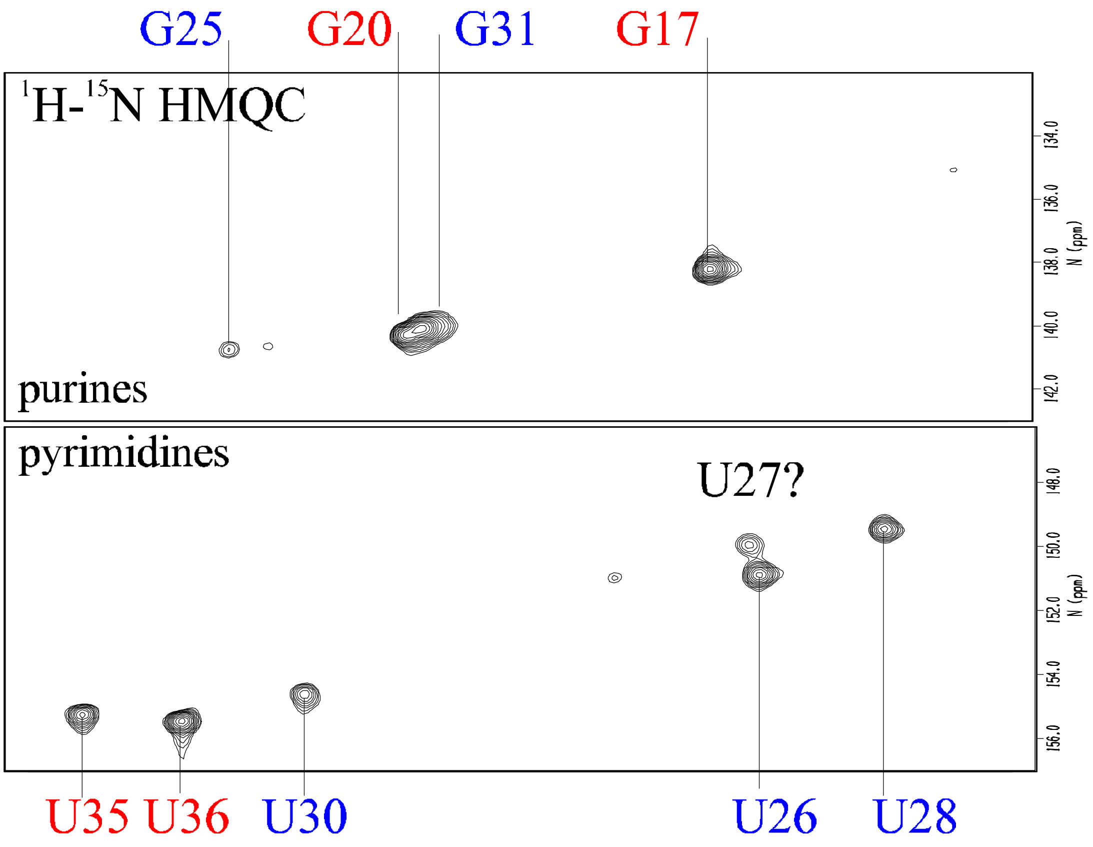


Figure 1.18 rLC55 ^{1}H-^{15}N HMQC

This HMQC was used to identification of the base type of the imino protons. The most striking feature of this spectrum is the U26 and U28 iminos, which have a strong imino-imino NOESY crosspeak signature. They are involved in a U=U wobble. The U27 imino is tentatively assigned as the weak third upfield shifted uridine imino. Notice that the G25 and G17 imino are shifted downfield and upfield, respectively, from the region a "normal" Watson-Crick G:C base-paired imino would appear.

can be identified. The first (shown in blue) is U_{26}=U_{28}, G_{25}, U_{30}, G_{31} and the second (shown in red) is G_{20}, U_{35}, U_{36}, G_{17}.

The orientation of the second stretch of iminos (G_{20} – G_{17}) was identified by observing the NOESY crosspeak between the U_{36} imino and A_{19}H2 proton, but not between the U_{35} imino and A_{18}H2 proton (Fig. 1.19). This pattern can only be explained by the assignment of the iminos in the orientation shown.

Unambiguous confirmation of the secondary structure assignment was accomplished by comparison of the imino NOESY crosspeak patterns of rLC55 with that of the smaller rLC25 RNA (Fig. 1.20). The rLC25 RNA can only adopt the form 1 hairpin because it is missing the nucleotides required for the form 2 base pairing, consequently, comparison of the imino proton spectra between these two samples will prove whether the rLC55 RNA is in the form 1 structure. The spectra are nearly identical, with only the G_{17} imino proton shifting slightly, which can be explained by its proximity to the form 1 hairpin termini where one would expect a slight structural difference between the two samples.

High-resolution studies on the non-exchangeable protons for rLC55 were attempted. The standard experiments, such as the D_2O NOESY and DQFCOSY experiments were conducted with little success (data not shown). The spectral resolution of the data were poor with many overlapped resonances. Qualitatively, the T2 relaxation properties of this large RNA made most of the resonances broad and difficult to assign.

In an attempt to solve the enormous spectral overlap problem, the six cytosines found in rLC55 were selectively $^{15}N/^{13}C$ isotope labeled. This cytosine labeled sample

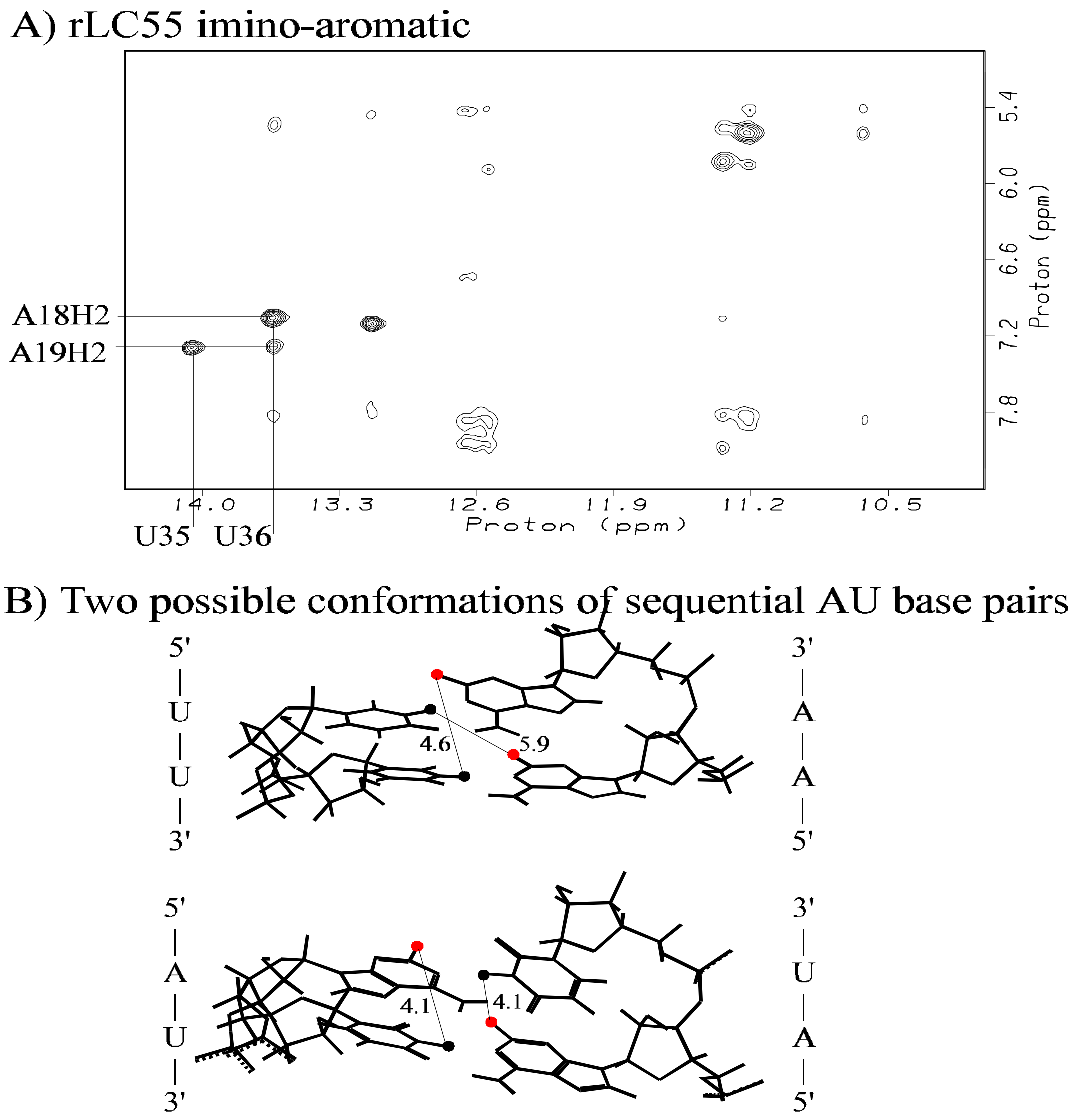


Figure 1.19 Assignment of the rLC55 U35 and U36 iminos

The assignment of the orientation of the AU to AU base pairing was accomplished by observing that the U imino NOESY crosspeak to the neighboring AH2 proton was asymmetric. If one builds a 5'-AU-3' duplex (B), by the nature of the symmetry, the crosspeak intensities will be nearly identical between each U imino to the neighboring AH2 proton. This would appear as a "box" of four crosspeaks in the region of the NOESY spectrum. This symmetric imino-AH2 crosspeak pattern has been observed in other RNAs in which there exists a 5'-AU-3' region of the sequence (personal communication, Dave Schweisguth).

Thus, the assignment of the U35 imino and the U36 imino followed from the observation that the imino-neighboring AH2 proton crosspeak pattern was asymmetric (see A above), and that the 3'-most uridine will have the stronger imino-neighboring AH2 proton crosspeak due to the 1.5 Å closer distance.

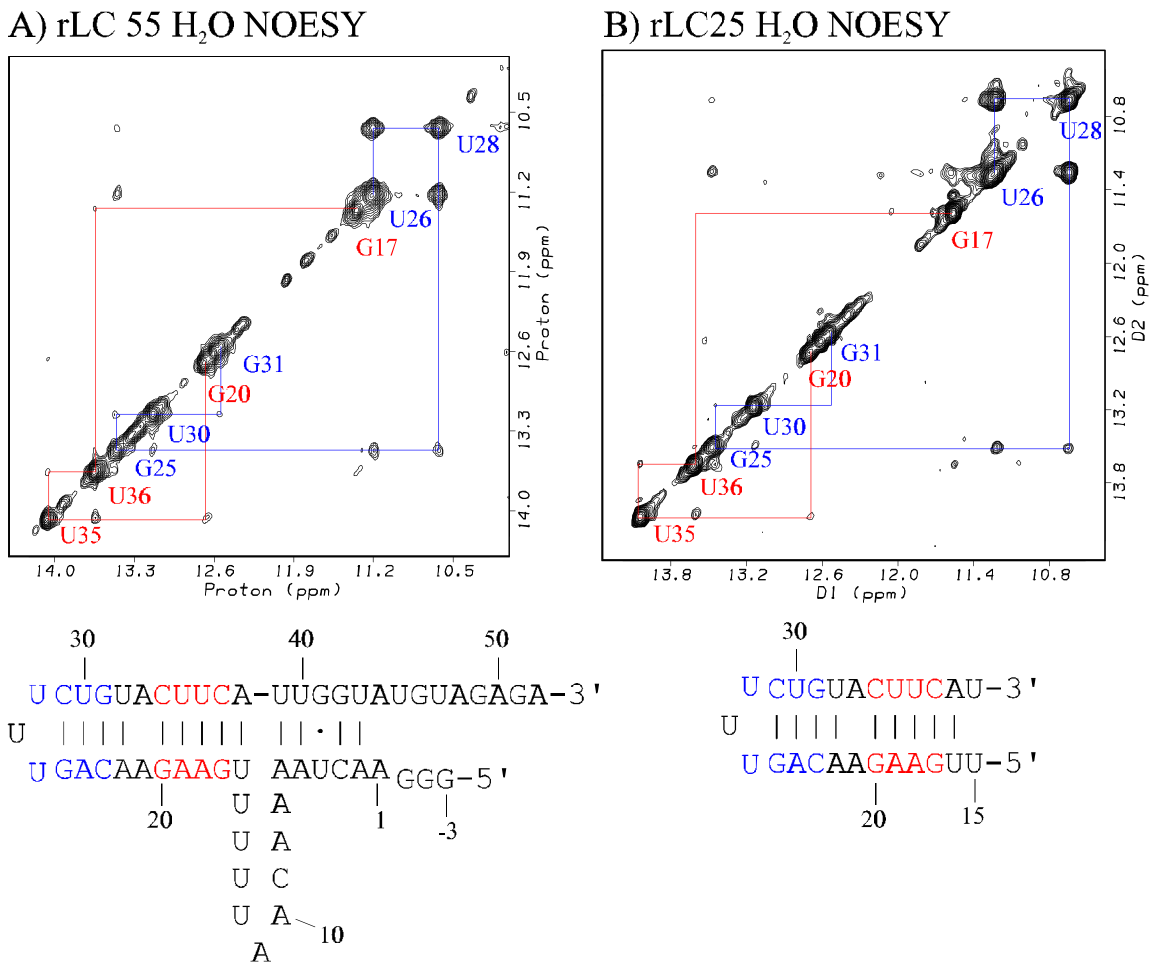


Figure 1.20 Comparison of the H_2O NOESY for rLC55 and rLC25

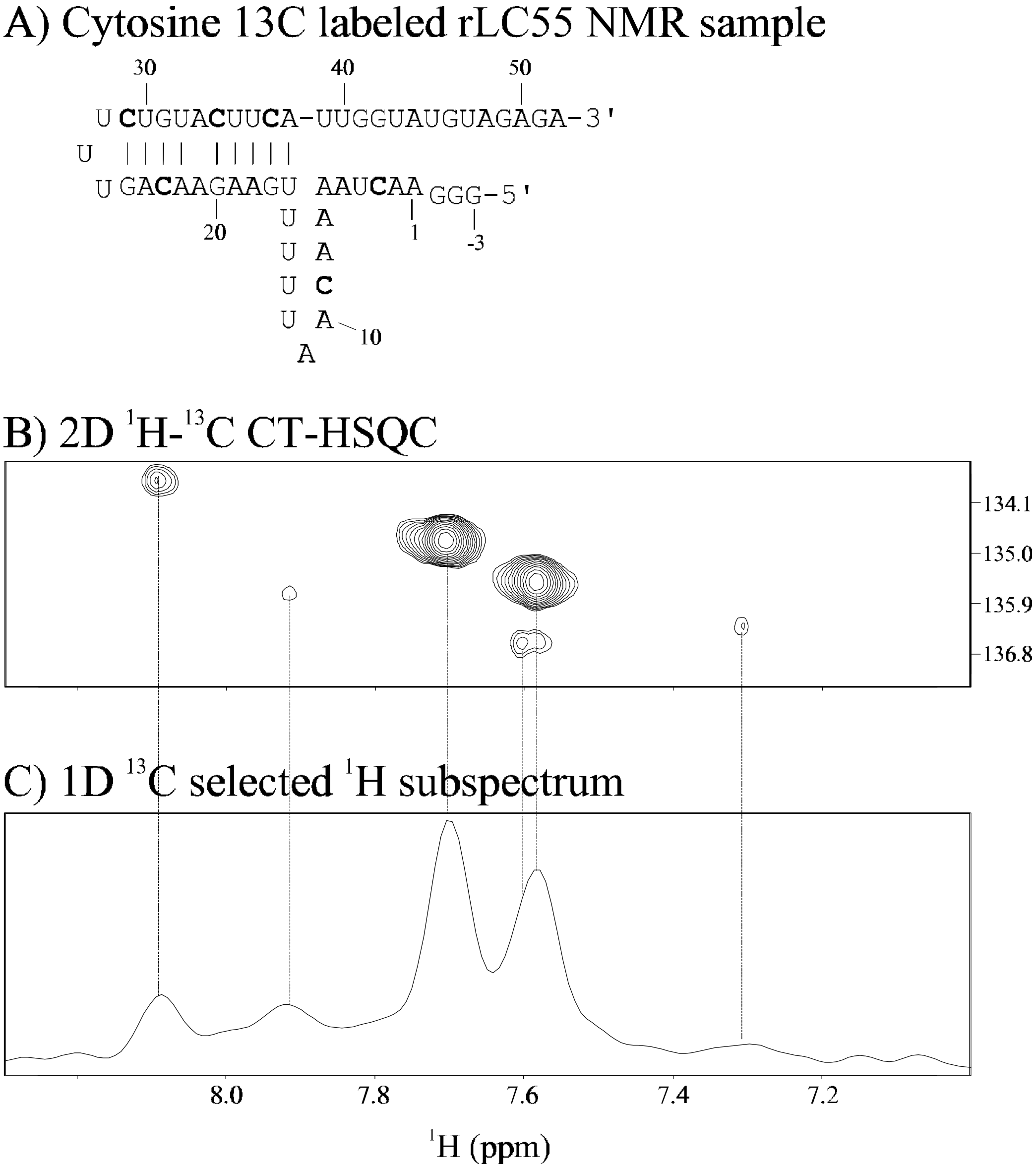
The two spectra were collected under identical conditions, 25°C and 250 ms mixing time. A) the data from the rLC55 sample, B) the rLC25 form 1 hairpin sample The chemical shifts of the iminos are nearly identical between the rLC55 and rLC25 sample, indicating that the secondary structure of the rLC55 RNA is represented well by the form 1 hairpin. The G17 imino does shift slightly, which is not unexpected because it is near the termini of the form 1 hairpin, where one would expect differences between the rLC55 and rLC25 samples.

could then be explored using isotope-filtered experiments (see chapter 3 for a discussion). While this did solve the problem of spectral overlap, the broad linewidths due to short T2 relaxation times became even worse. The 2D constant time ^{1}H ^{13}C HSQC and the 1D filtered spectrum are shown in figure 1.21. The two intense resonances are assigned to the C_3 and C_9, and the broadened resonances (C_{23}, C_{29}, C_{34} and C_{37}) are those involved in the form 1 hairpin base-pairing region. A 2D isotope filtered NOESY experiment (see chapter 3) was also collected on this sample (data not shown) and the ^{12}C subspectrum was characterized by broad overlapped peaks, and the ^{13}C subspectrum had very little signal because the extra proton relaxation by the ^{13}C. High-resolution characterization of the non-exchangeable protons on rLC55 was not successful and any further studies would have to utilize the smaller rLC25 and rLC13 form 1 hairpin fragments.

1.3.6 Salt mediated conformational change in the L. collosoma *form 1 hairpin*

The smaller fragments, rLC25 and rLC13, provided a means to study the form 1 hairpin at higher resolution than was possible with the large rLC55 sample. One intriguing structural feature of these form 1 hairpins is the strong imino-imino crosspeak between the two uridines (Fig. 1.22). This is intriguing because the only position in the sequence where this U=U base pair can form is between U_{26} and U_{28} in the hairpin loop. Requiring that the hairpin loop be spanned by a single nucleotide, U_{27}, unless the RNA is a duplex. Further characterization of this hairpin loop was necessary.

To examine if the rLC25 and rLC13 RNA existed in a single conformation on the NMR time scale, the "double quantum-filtered COSY" (DQFCOSY) NMR experiment was performed. This experiment correlates protons that are three bonds away from each



A) Cytosine 13C labeled rLC55 NMR sample

B) 2D ^{1}H-^{13}C CT-HSQC

C) 1D ^{13}C selected ^{1}H subspectrum

Figure 1.21 rLC55 cytosine ^{13}C labeled spectra

A) The rLC55 sample were synthesized with only the six cytosines ^{13}C isotope labeled, shown in bold. B) The constant time HSQC showing the 6 cytosine correlation between the H6 proton and the C6 carbon. The two strong intensity peaks are probably from the non-base paired C3 and C9 nucleotides that are experiencing a faster local correlation time due to local dynamical movement. C) The 1D ^{13}C subspectrum (see chapter 3) from the same sample.

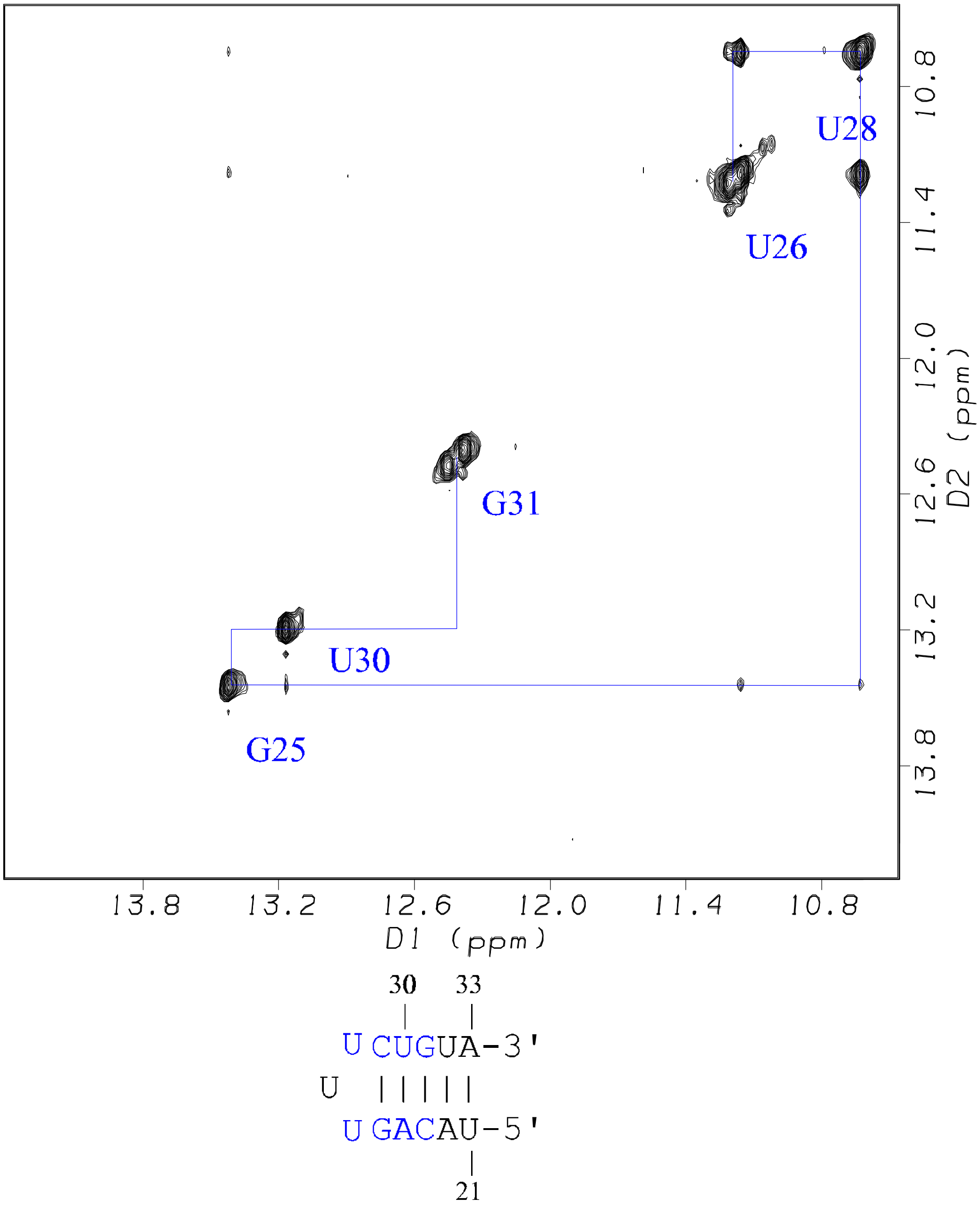


Figure 1.22 2D H2O NOESY of rLC13

The imino protons for the rLC13 RNA are a subset of those for rLC55 and rLC25. The crosspeak between U30 and G31 has never been seen in a H_2O NOESY experiment for rLC13, but the connectivity is drawn in above, by inference from the other NOESY data for rLC25 and rLC55. The absence of the cross peak is probably due to a higher exchange rate of the G31 imino with H_2O since it is near the terminus of the helix and not because of some major structural change.

other. The intensity of the crosspeak is dependent on the magnitude of the vicinal ^{3}J-coupling constant between the protons. Because of the anti-phase nature of the crosspeak quartets, small coupling causes the crosspeaks to cancel out. The ^{3}J-coupling intensity between 3 bond distant protons follows the Karplus relationship (1959), with a maximum coupling at 0 and 180 degrees and a minimum at 90 and 270 degrees. The H5 and H6 protons in the pyrimidine bases of nucleic acids are ideal protons to observe with this experiment because they are fixed in position relative to each other at 0 degrees. Thus they have a large ^{3}J-coupling constant (~10-12 Hz) and are strong crosspeaks in the dqfcosy experiment. For determining if an NMR sample exists in a single conformation, one only has to count the number of H5-H6 DQFCOSY crosspeaks. If they add up the same number as what is expected, then the sample is in a single time-averaged fast-exchange conformation.

DQFCOSY spectra for rLC25 and rLC13 were collected for different buffer salt conditions, and the results are shown in figures 1.23 and 1.24, respectively. We observed more H5-H6 crosspeaks in the spectrum corresponding to the "intermediate" salt conditions (50 mM NaCl) than can be accounted for by the sequence. This could be explained by the existence of two structural conformations in slow-exchange. If the ionic strength of the buffer was lowered (<30 mM NaCl), one of the two conformations was favored (state A) and if the ionic strength of the buffer was raised (>150 mM NaCl), the other conformation was favored (state B). This ability to favor one conformation over the other by adjusting the salt concentration was seen for both the rLC25 and rLC13 samples.

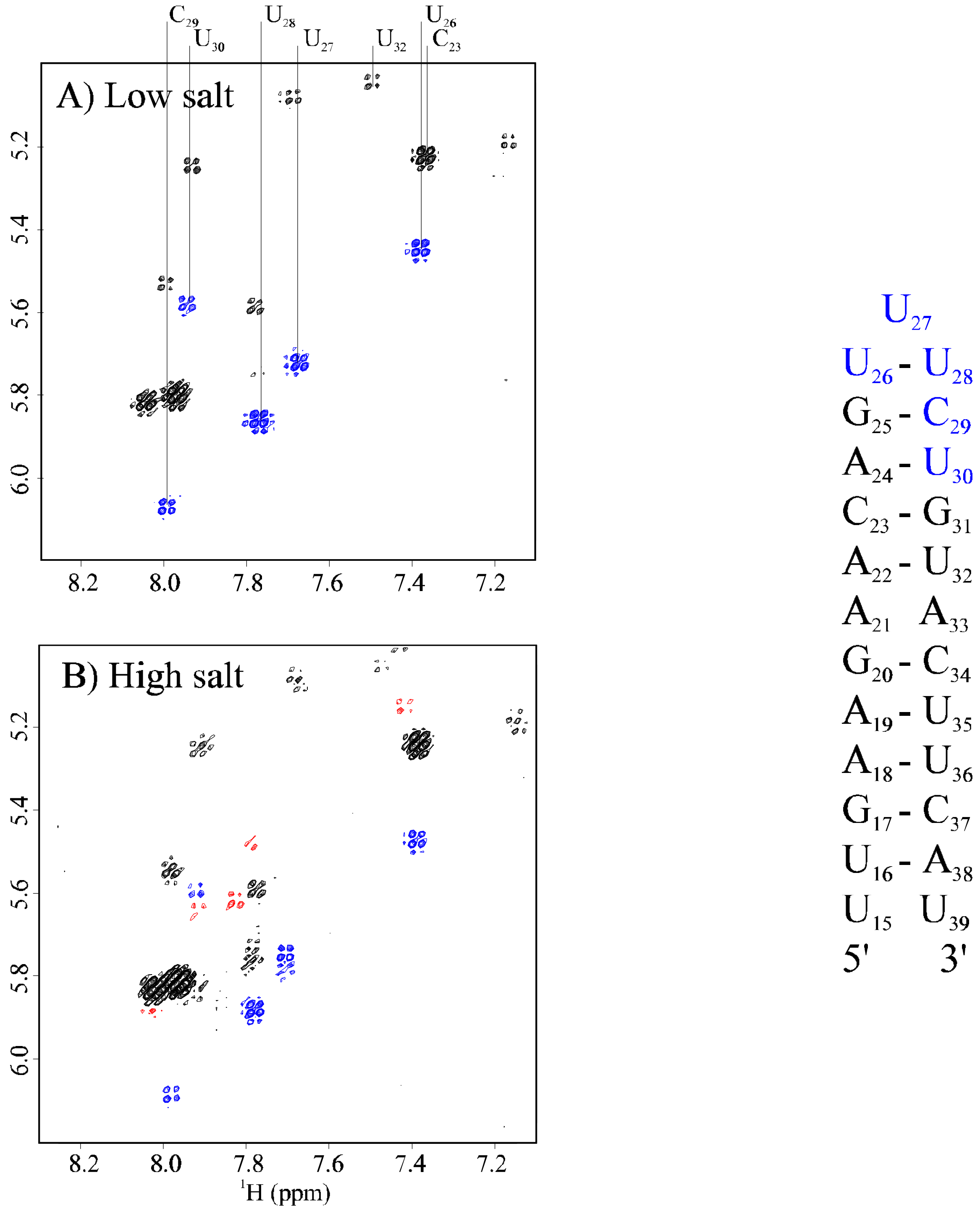


Figure 1.23 rLC25 salt dependence DQFCOSY data

Low salt conditions: 10 mM sodium phosphate buffer (pH 6.5), 1 mM EDTA. B) High salt conditions: 10 mM sodium phosphate (pH 6.5), 1 mM EDTA and 100 mM NaCl.

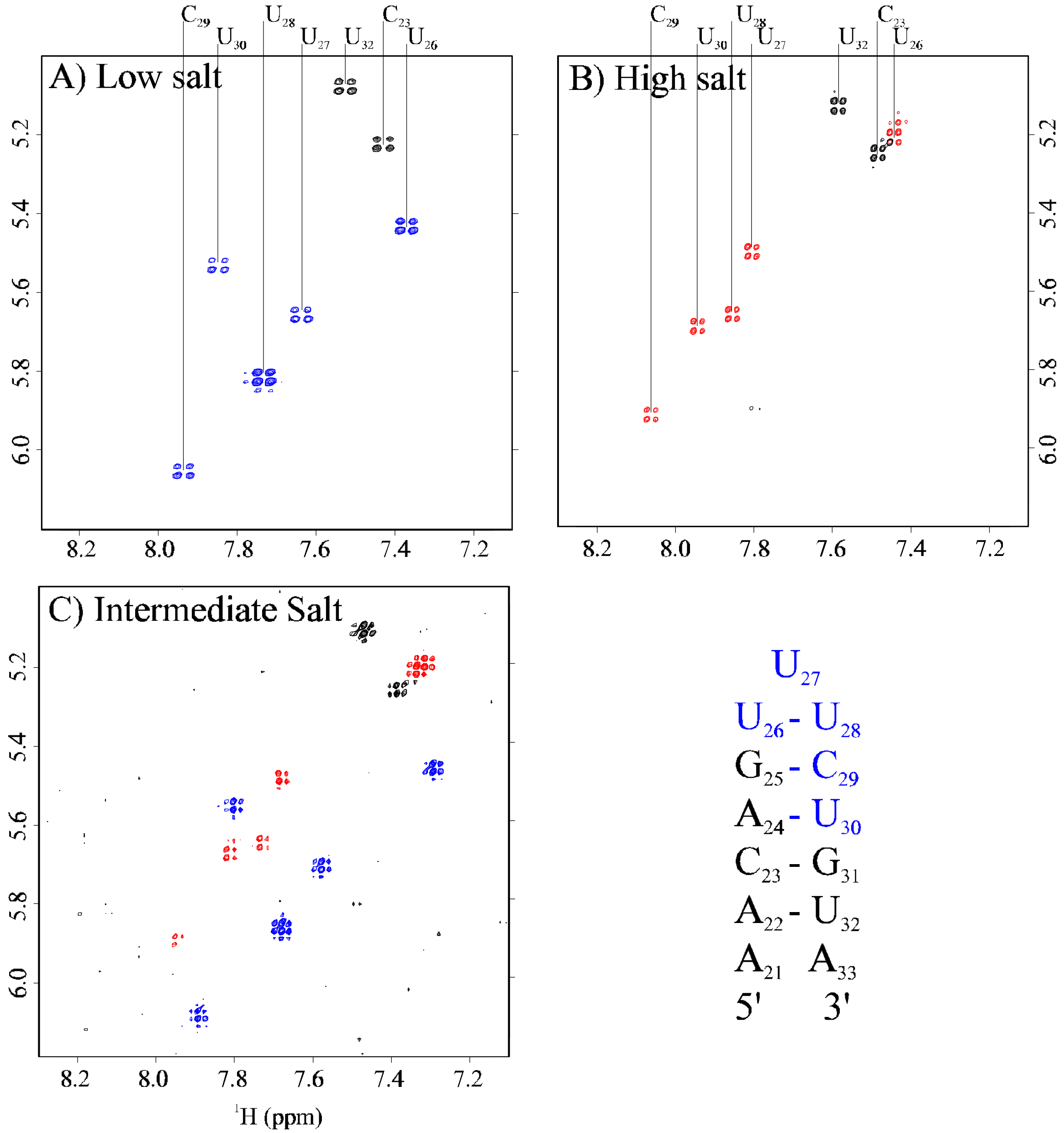


Figure 1.24 rLC13 salt dependence DQFCOSY data

All RNA samples contained the rLC13 sample dialyzed against A) 10 mM phosphate buffer (pH 6.5) and 1 mM EDTA, B) 10 mM phosphate buffer (pH 6.5), 1 mM EDTA and 100 mM NaCl. C) The intermediate buffer condition contained 50 mM NaCl.dimeric duplexes as shown in figure 1.25.

One explanation for the ion strength dependent conformations would be a hairpin-duplex transition. Small RNAs such as these can form either monomeric hairpins or dimeric duplexes as shown in figure 1.25. These monomer-dimer structures are difficult to differentiate by NMR. Because of the inherent dyad symmetry of the dimer they can appear nearly identical to the monomer hairpin spectroscopically. A number of biophysical and NMR techniques can be used to determine if a sample is a monomer or dimer; see chapter 4 for a full discussion of the methods. We used the techniques of optical UV melts and translational diffusion constant measurements to clarify this issue.

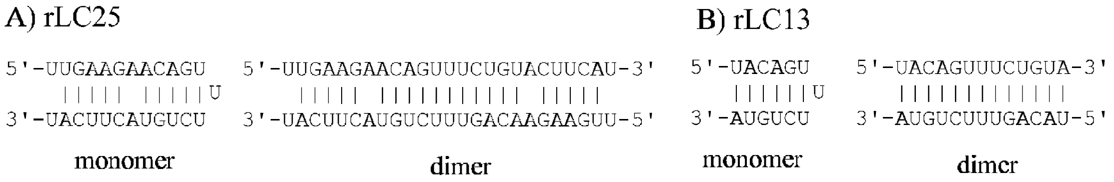


Figure 1.25 rLC13 and rLC25 basepairing possibilities for a monomer or dimer

1.3.6 Evidence that the rLC13 low and high salt samples are monomeric

Optical UV melting curves for the rLC13 samples are shown in figure 1.26 for both the low and high salt conditions, and at two different RNA concentrations. The stability of a hairpin is independent of its concentration in solution, while the stability of a dimer is dependent on its concentration. Thus, UV melting curves will show a concentration dependence to the measured melting temperature (Tm) for a dimer and not for a monomer. Another variable is the ionic strength of the solution. High salt stabilizes the hairpin (or base paired) structures and as expected the Tm of the high salt samples were higher (6 degrees). However, there was no perceptible RNA concentration dependence to the Tm of either the low or high salt sample.

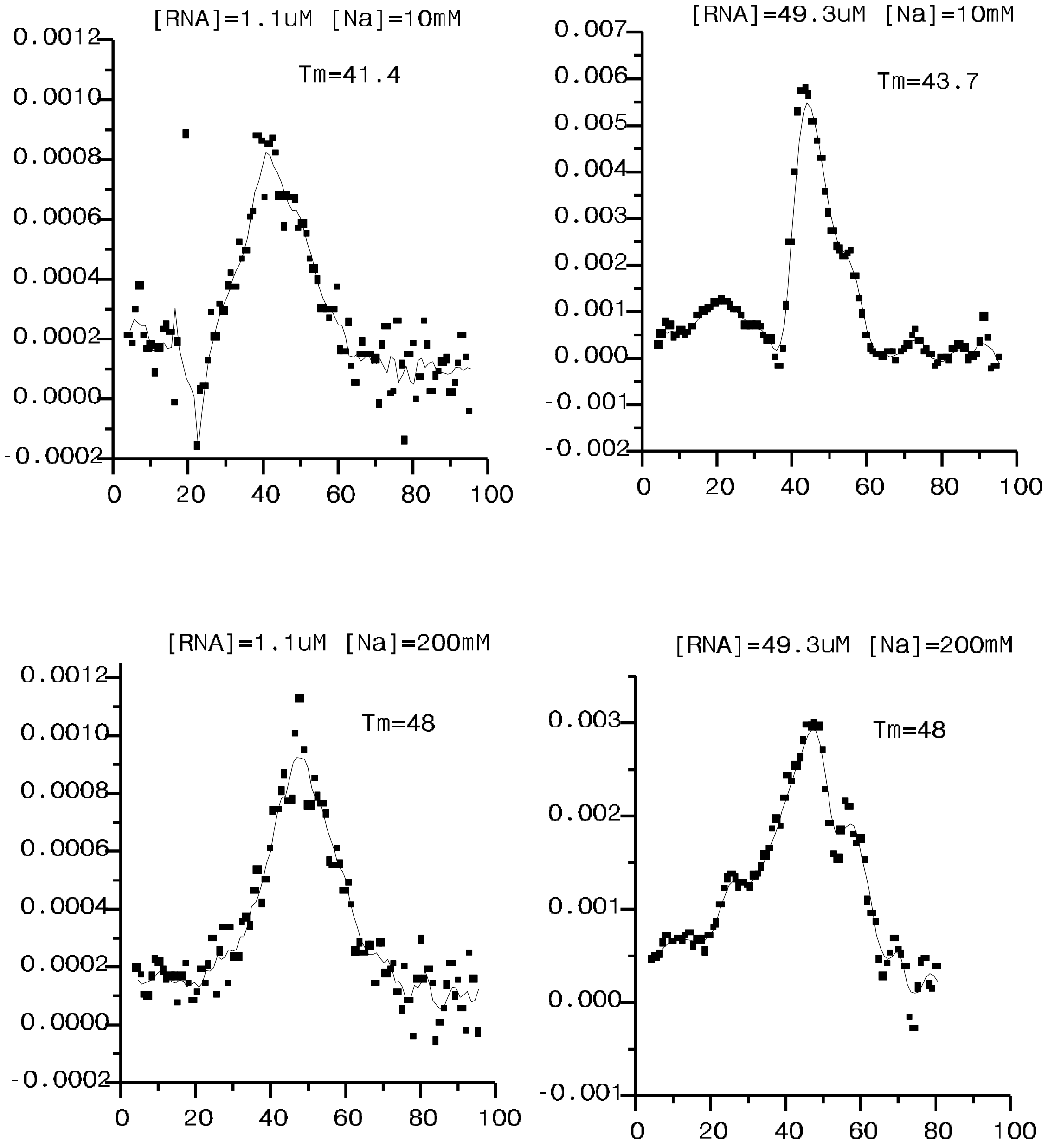


Figure 1.26 rLC13 UV melts

Equilibrium melting curves for rLC13 in the low and high salt buffer used in the NMR experiments. The top two graphs are the low salt buffer (10 mM sodium phosphate pH 6.5, 1 mM EDTA) with [RNA]=1.1 μM and 49.3 μM respectively. The bottom two graphs are the high salt buffer conditions (10 mM sodium phosphate pH 6.5, 200 mM NaCl, 1 mM EDTA) with [RNA]=1.1 μM and 49.3 μM.

The melting temperature is ~42° C for the low salt conditions and ~48° C for the high salt conditions with no appreciable RNA concentration effects.

Unfortunately, this result alone does not guarantee that the same is true for the NMR samples. The major problem with reliance on UV melting curves in this type of analysis is that it is difficult to perform the melts at RNA concentrations high enough to perform NMR experiments (millimolar). The highest concentrations of RNA that can be used for UV melting curves is ~50-200 μM (depending on the size of the RNA) and requires the use of special short path length cuvettes, such as the 1 mm cuvettes used in this study. Thus, even though our results indicate both samples were monomeric, other methods must be employed to secure our conclusion at NMR concentrations. For that, the NMR based method of measuring the translational self-diffusion constants (see chapter 4 for a full discussion) was used.

The translational self-diffusion constants for the low and high salt rLC13 sample were measured as 1.40×10^{-6} cm^2/s and 1.45×10^{-6} cm^2/s. The data are shown in figure 1.27. The results are compared to those from a 14 nt reference RNA which can be examined as either a monomer or a dimer (Lapham, et al., 1997). Hydrodynamics theory predicts that for RNAs of this size, the dimer:monomer ratio of the diffusion constants should be approximately 0.65. This was exactly what was observed when we measured the 14 nt reference RNA in its two conformations. The ratio obtained for the diffusion constant of both the low and high salt rLC13 samples was approximately 1, suggesting that they are similar hydrodynamically. Furthermore, the absolute diffusion rate measured for the rLC13 samples ($\sim 1.4 \times 10^{-6}$ cm^2/s) is what is expected of a 13 nt monomer (see chapter 4 for discussion on predicting diffusion constants). Therefore, the diffusion rate measurements predict that both the low and high salt rLC13 samples are monomeric.

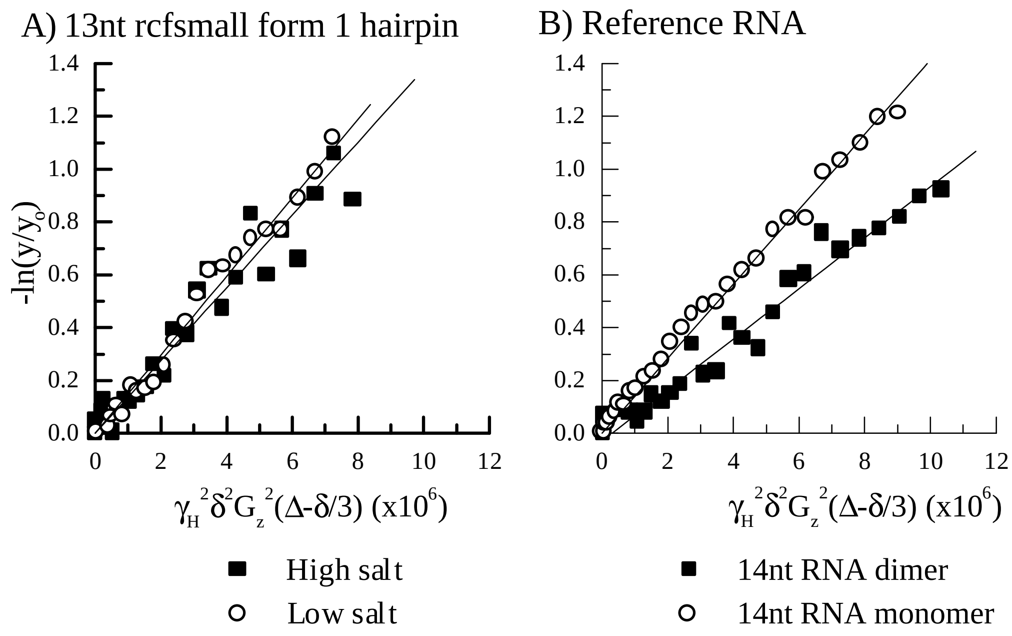


Figure 1.27 Diffusion rate of rLC13 low and high salt conformations

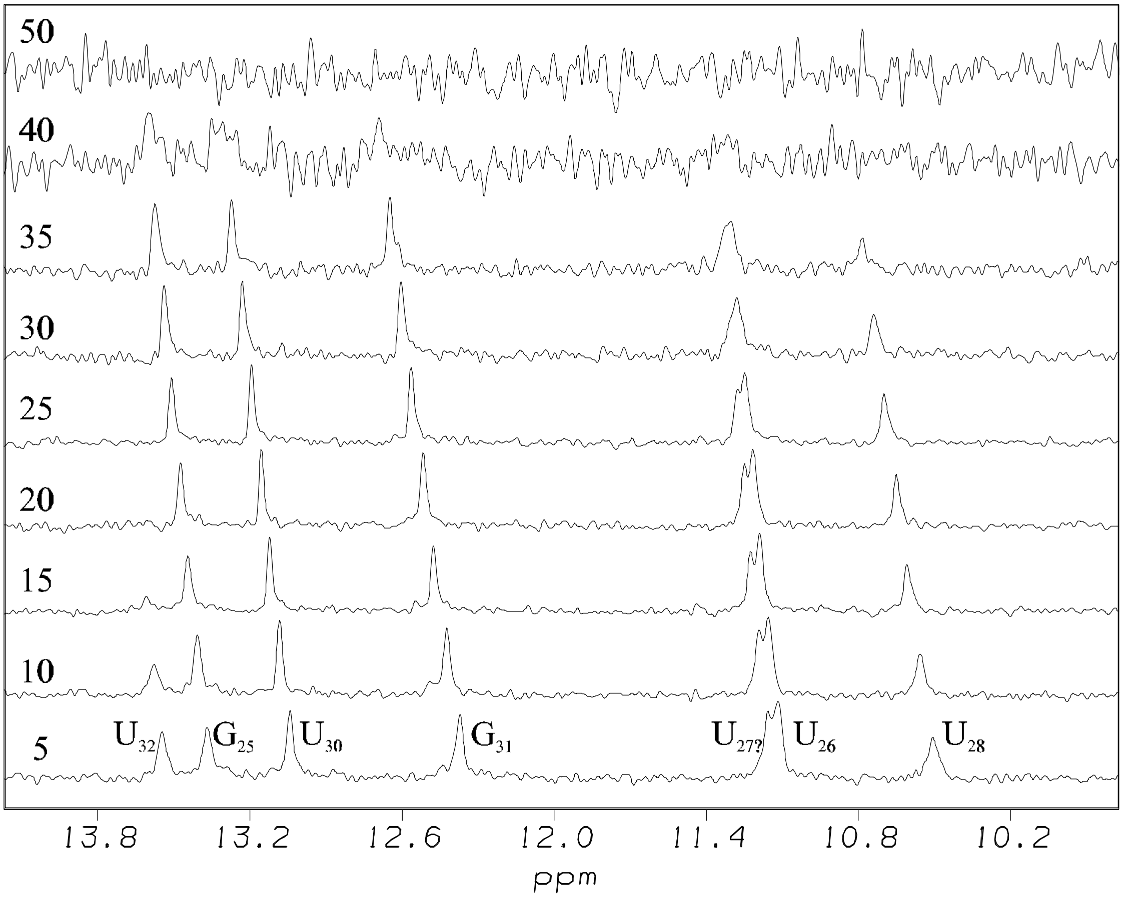
The post-processed (see materials and methods) data from the pulsed field-gradient stimulated echo (pfg-STE) experiment. Δ=0.1s and δ=0.004s, the gradient was varied from 0 to 32 Gauss/cm in steps of 1 Gauss per increment. The experiments were conducted at 25°C. A) The 13 nt rLC13 RNA with a measured diffusion rate of 1.40 x 10^{-6} and 1.45 x 10^{-6} cm^2/s for the low and high salt sample respectively. B) Reference data from a 14 nt RNA (see chapter 4) in either a monomeric hairpin or a duplex form.

The 1D imino proton melts of rLC13 are shown in figure 1.28 for the high salt sample. In the low salt melting experiment (data not shown) all the imino protons disappear by 10° C except for the imino from G31, which melts out at 30° C.

1.3.7 Assignments of the non exchangeable protons for rLC13

The 2D NOESY spectra of the low salt rLC13 sample in D_2O are shown in figures 1.29 and 1.30. The NOESY spectra from the high salt rLC13 sample in D_2O are shown in figures 1.31 and 1.32. The non-exchangeable protons were assigned using the "anomeric-aromatic walk" in which the H6/H8 base proton of a nucleic acid is correlated to its own H1' and the H1' in the 5' direction.

Aside from the anomeric-aromatic walk, additional connectivities confirmed the assignments of the protons. As an example, the A_{24} H2 proton cross-strand and same-strand NOEs confirmed the assignments to the G_{31} H1' and G_{25} H1'. The exchangeable imino protons were correlated to the non-exchangeable protons by the 2D watergate NOESY experiment (Fig. 1.33). This experiment allows for the observation of exchangeable to non-exchangeable NOE crosspeaks that appear close to the water resonance. As an example, the G_{31} imino has a strong NOE connection to the amino protons on C_{23}, which then show a strong crosspeak to the C_{23} H5. In this manner, the assignments of the cytosine H5 protons could be reaffirmed.



$5'\text{-}U_{21}A_{22}C_{23}A_{24}G_{25}U_{26}U_{27}U_{28}C_{29}U_{30}G_{31}U_{32}A_{33}\text{-}3'$

Figure 1. 28 Temperature study of the imino proton from high salt rLC13

The imino protons from the rLC13 RNA sample in a high salt buffer. The profile is very similar in terms of the chemical shifts seen for rLC55, but these imino protons have much sharper line widths. The U32 imino (near the terminus) melts out at a low temperature (15°), while the other iminos melt out at 40° C. The U27?/U26 and U28 iminos seem to exchange broaden before the stem iminos do, which might indicate they are involved in a more solvent accessible conformation. The low salt temperature study on rLC13 showed all the same iminos at 5° C, but only the G31 imino was visibly above 10° C.

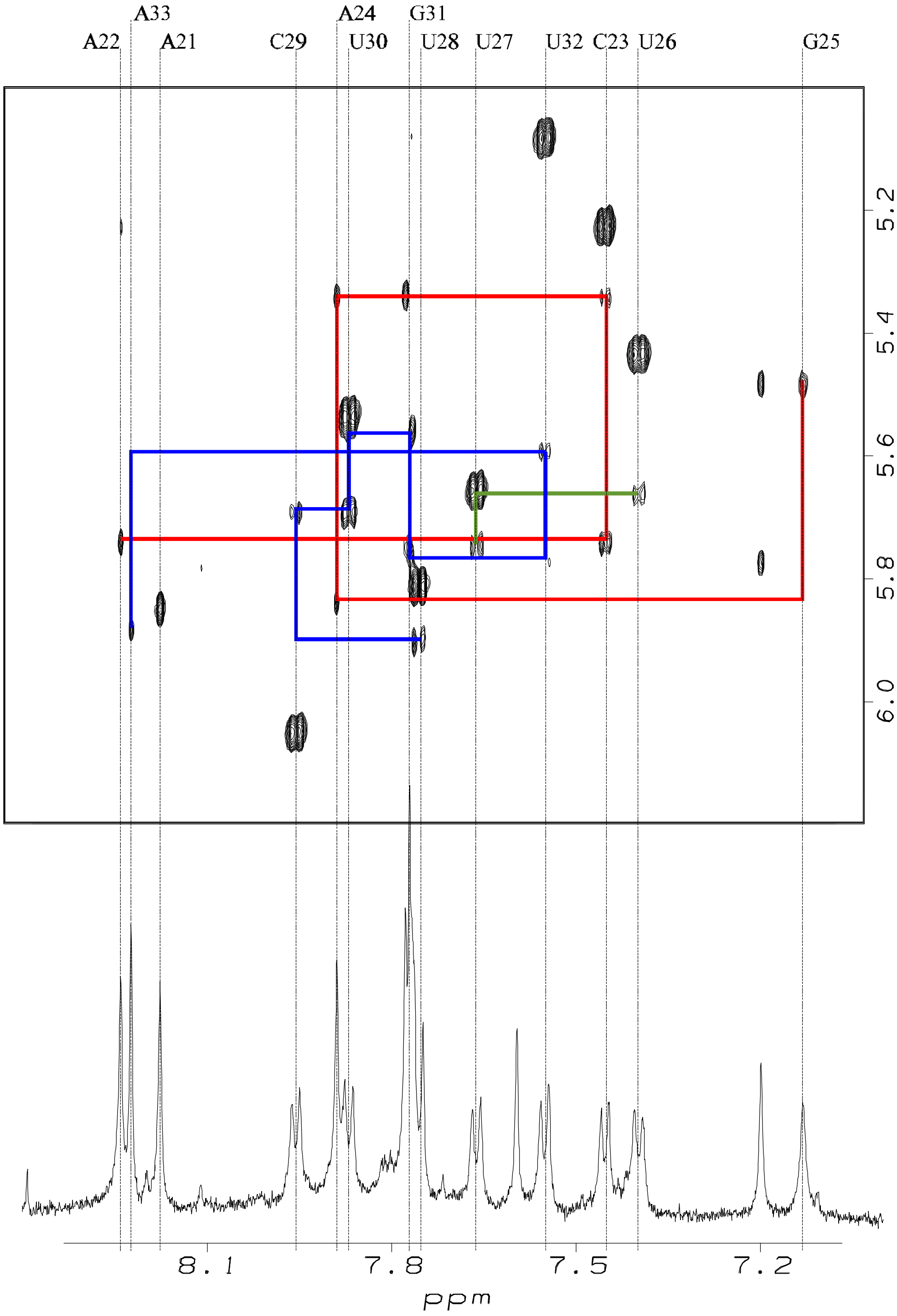


Figure 1.29 NOESY spectrum of the low salt rLC13

The mixing time was 250 and the temperature was 20^{o} C. The anomeric-aromatic walk is demonstrated with the overlay lines. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAA-3'.

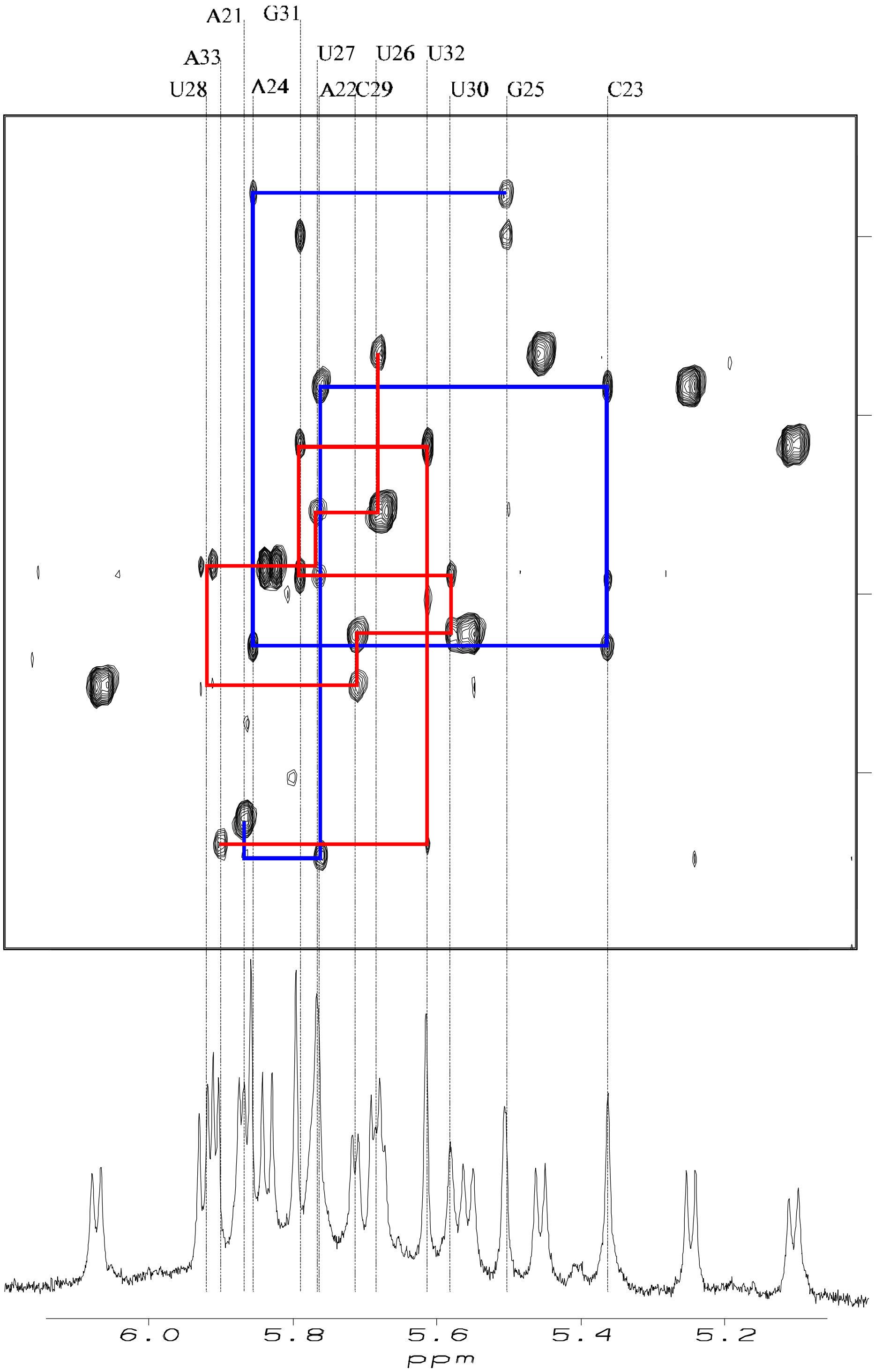


Figure 1.30 NOESY spectrum of the low salt rLC13

The same experiment as shown in figure 1.29, but with the limits transposed. Some of the connectivities are better seen in this region of the spectrum. The mixing time was 250 ms and the temperature was 20° C. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAA-3'.

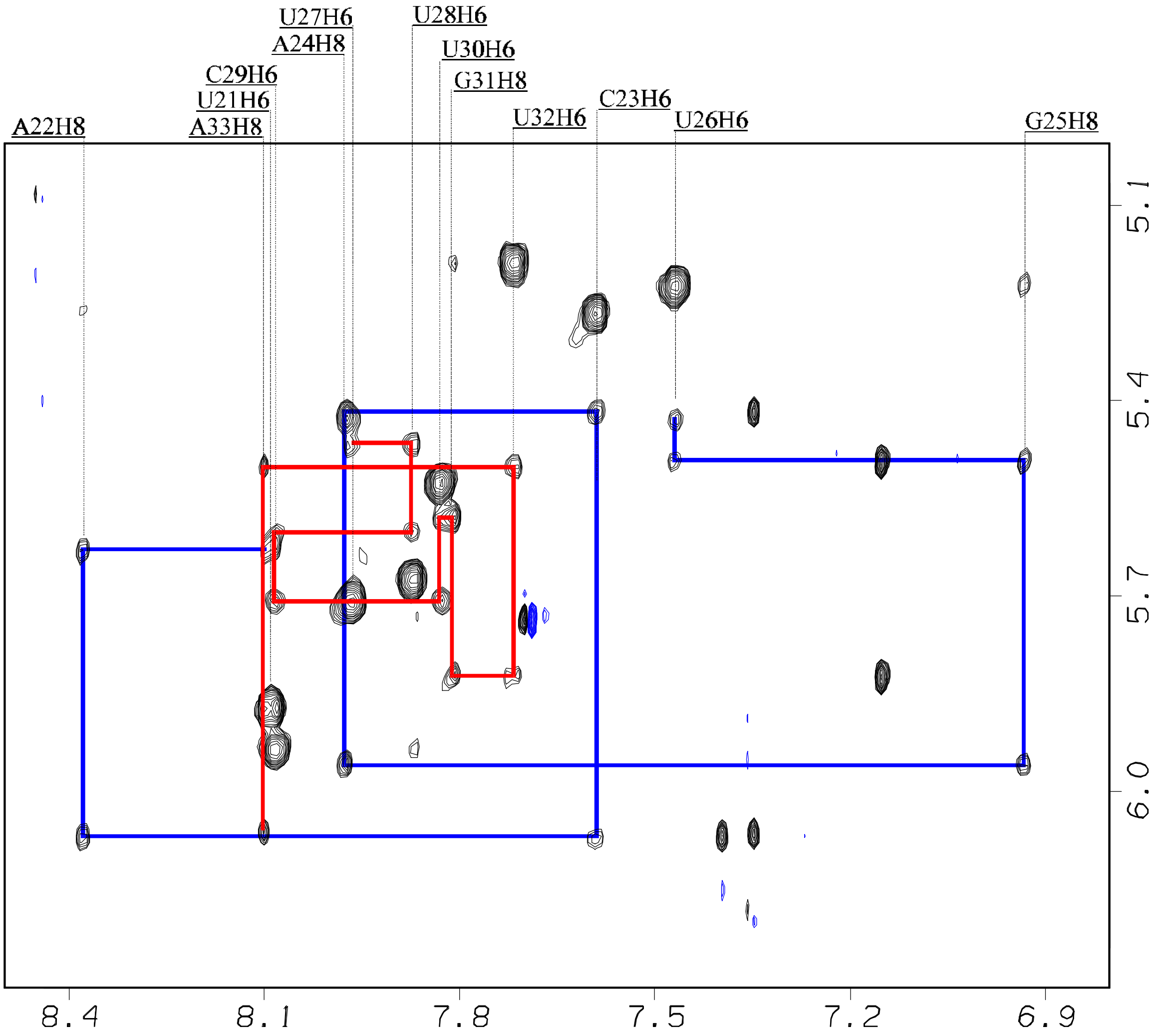


Figure 1.31 NOESY spectrum of the high salt rLC13

The mixing time was 250 ms and the temperature was 25° C. The assignments are shown with the dotted lines and the solid lines represent the anomeric-aromatic walk. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAU-3', there is an extra uridine on the 3' end as compared to the data shown for the low salt sample.

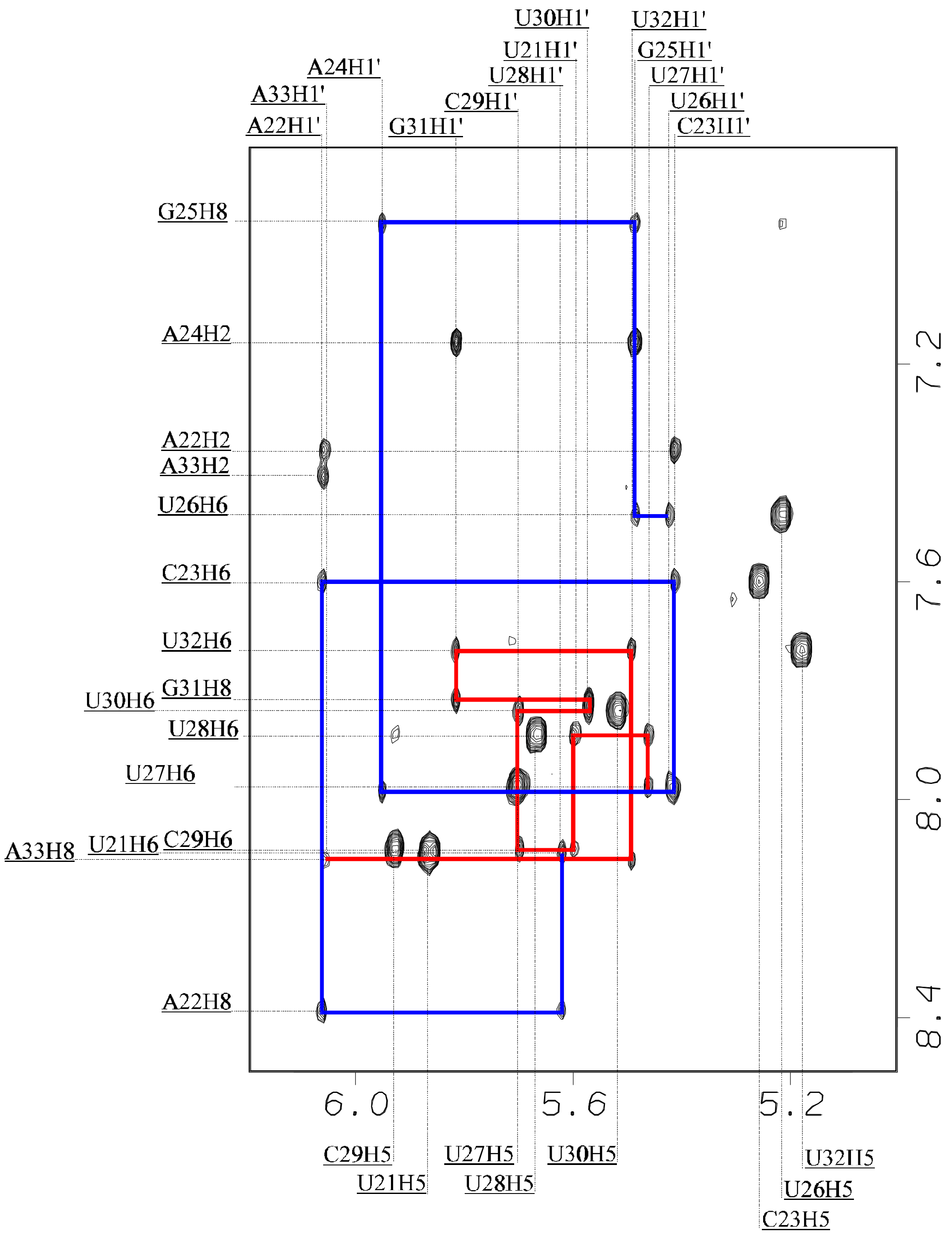


Figure 1.32 NOESY spectrum of the high salt rLC13

The mixing time was 250 and the temperature was 25° C. Assignments are shown with the dotted lines and the anomeric-aromatic walk is shown with the solid lines. Note that the rLC13 sample used in this experiment was 5'-AUGUCUUUGACAU-3', there is an extra uridine on the 3' end as compared to the data shown for the low salt sample.

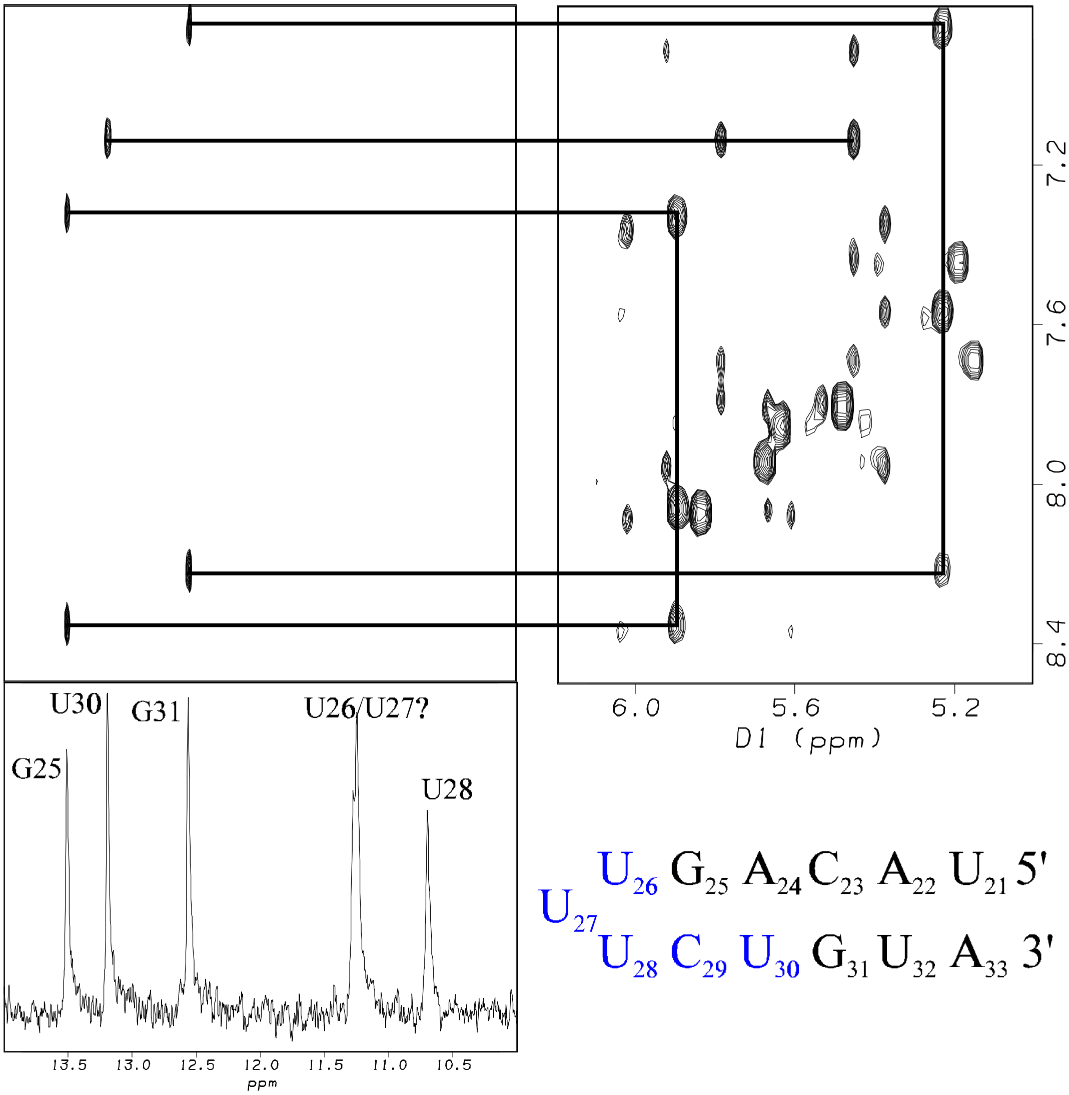


Figure 1.33 Imino to non-exchangeable of high salt rLC13

Watergate NOESY spectrum with 300 ms mixing time at 25° C. Many of the non-exhangeable assignents can be confirmed with this experiment. For instance, the G31 and G25 imino protons have strong NOE cross peaks to the amino protons on their base pair partner cytosines. These aminos then have a strong connectivity to the H5 proton. Additionally, the U30 imino has a strong cross peak to the AH2 proton from A24.

Chemical shift data

	H8/H6	H5/H2	H1'	H2'
U21	8.09	5.88	5.63	4.61
A22	8.38	7.35	6.07	4.59
C23	7.60	5.27	5.42	
A24	7.98	7.16	5.96	4.72
G25		n/a	5.50	4.48
U26	7.47	5.23	5.44	
U27	7.98	5.71	5.47	4.22
U28	7.88	5.68	5.60	4.46
C29	8.03	5.94	5.71	4.43
U30	7.84	5.53	5.59	4.53
G31	7.82	n/a	5.83	4.48
U32	7.72	5.20	5.50	4.27
A33	8.10	7.40	6.07	4.15

Table 1. 1 rLC13 High salt assigments

Chemical shift data

	H8/H6	H5/H2	H1'	H2'
A21	8.17		5.87	
A22	8.24	7.77	5.76	
C23	7.45	5.25	5.36	
A24	7.88	7.19	5.86	
G25	7.12	n/a	5.51	
U26	7.39	5.45	5.68	
U27	7.65	5.67	5.77	
U28	7.75	5.83	5.92	
C29	7.95	6.07	5.71	
U30	7.86	5.55	5.58	
G31	7.77	n/a	5.79	
U32	7.54	5.11	5.61	
A33	8.22		5.90	

Table 1. 2 rLC13 Low salt assignments

1.4 Discussion

The goals of this project were to characterize the *in vitro* secondary structures of the SL RNA from the trypanosomes *L. collosoma* and *C. fasciculata*, and to determine what structural element was responsible for the form 1 low temperature UV melt transition found in the *L. collosoma* SL RNA.

1.4.1 The secondary structures of the C. faciculata *and* L. collosoma *SL RNAs*

The *in vitro* secondary structures of the two SL RNAs have been identified using NMR techniques. The rCF55 RNA was found to exist in the form 2 structure by comparison of its ^{1}H-^{15}N HMQC to the form 2 fragment hairpin, rLC30. The rLC55 RNA was confirmed to exist in the form 1 secondary structure, as previously proposed (Lecuyer and Crothers, 1993) by comparing the H_2O NOESY spectrum of the parent RNA to that of the form 1 fragment hairpin, rLC25.

Because the *C. fasciculata* SL RNA has a closer sequence homology to all the other non-*L. collosoma* SL RNAs, we speculate that most of the trypanosome SL RNAs probably exist, *in vitro*, in the form 2 secondary structure. The biological significance of this is unclear, as it has been shown that the *in vivo* secondary structure of a SL RNA may be different to the *in vitro* secondary structure (Harris, et al., 1995).

1.4.2 The L. collosoma *"tertiary" structure*

The low temperature melting transition has been speculated to be due to a "tertiary" structural element, which may exist between the form 1 hairpin and one of the two RNA "tails" that extends off the terminus of the hairpin. A number of the results presented in this chapter seem to contradict this possibility.

Figure 1.20 demonstrates that the chemical shifts of the imino protons of rLC55 and rLC25 are nearly identical, except for the G_{17} imino proton. This G_{17} imino proton is located on the terminus of the form 1 hairpin, and would be expected to experience a chemical shift change. The correlation of the chemical shifts of all other imino protons indicates that the form 1 hairpins are in similar environments in both samples. If there was some type of tertiary interaction between the hairpin and the rest of the RNA, one might expect the interaction to affect the environment of the exchangeable protons in the rLC55 sample. Thus, the 25 nucleotides of the form 1 hairpin are probably structurally independent of the rest of the SL RNA.

Another indication that the form 1 hairpin is not interacting with the rest of the SL RNA can be inferred by observing the linewidths in the 2D ^{1}H-^{13}C CT-HSQC experiment (Fig. 1.21). The linewidths of the cytosines in the form 1 hairpin (C_{23}, C_{29}, C_{34} and C_{37}) and those in the 5' end of the RNA (C_3 and C_9) suggest that the two regions of the RNA are characterized by different T2 relaxation times. This can be explained if two regions of the RNA experience different effective correlation times, as would be the case if the 5' end of the RNA is unstructured. The cytosines in the form 1 hairpin would then experience the actual correlation time of the molecule, and C_3 and C_9 would experience a faster effective correlation time. As an example, it is often seen that the terminal nucleotides of a nucleic acid duplex have narrow very intense peaks because they experience a faster effective correlation time.

Another explanation for the origin of the low temperature UV melting transition is simply the melting out of the base pairing across the splice site. This theory is based on specific heat calculations using the "rnadraw" computer program (Matzura and

Wennborg, 1996), which was derived from the "rnaheat" program (Hofacker, et al., 1994). The program utilizes the partition function algorithm by McCaskill (McCaskill, et al., 1990) and energy parameters from Turner (Turner, et al., 1988) , Freier (Freier, et al., 1986) and Jaeger (Jaeger, et al., 1989). The calculations for the rLC55 RNA are shown in figure 1.34 below and support the hypothesis that the splice site base pairing is present, but melts earlier than the rest of the RNA.

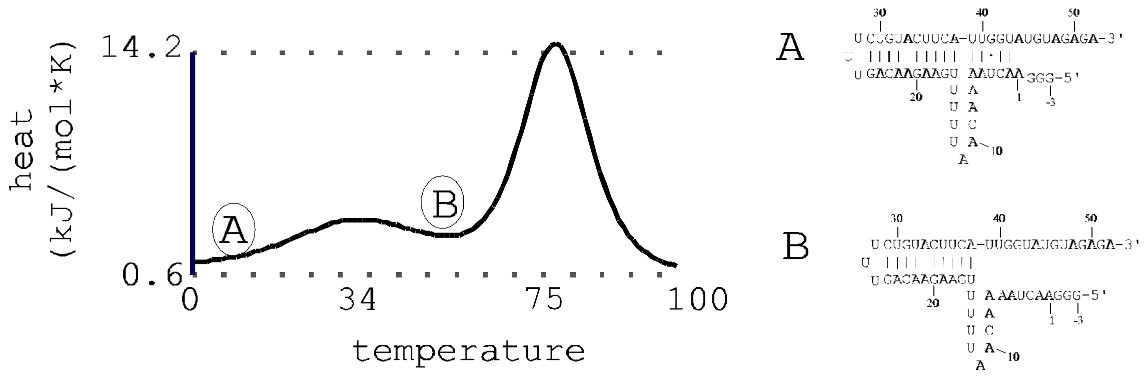


Figure 1.34 Specific heat calculation for rLC55

H_2O NOESY experiments were collected on the rLC55 sample at cold temperatures (data not shown), in an attempt to see the imino proton resonances of the splice site base pairs. No new iminos were observed at the colder temperatures. While this may suggest that the splice site helix never forms, it may also be that the helix forms transiently and the resonances cannot be seen because the imino protons are exchanging rapidly with the solvent.

1.4.3 Form 1 hairpin structure

All three 2D H_2O NOESY spectra from rLC55, rLC25 and rLC13 show a strong uridine to uridine imino proton crosspeak which has been assigned to the imino protons of U_{26} and U_{28}. The ^{1}H ^{15}N HMQC from rLC55 (Fig. 1.18) shows a weak extra uridine imino, which we tentatively assigned to U_{27}. The existence of that imino proton would strongly argue that the rLC55 RNA (and by association, rLC25 and rLC13) is a duplex RNA, rather than a monomeric hairpin. If the RNAs are duplexes, then the appearance of the U_{27} imino is easily accounted for. Due to the symmetry of the duplex, only one of the two U_{27} iminos would be visible.

However, other than this extra imino proton and common sense, all the other evidence suggests that the RNAs are monomeric hairpins. The UV melting curves of rLC55 (Fig. 1.16) and rLC13 (Fig. 1.26) do not show any perceptible RNA concentration dependence for the melting temperature and the translational self-diffusion rates of the RNAs are consistent with monomeric hairpins.

If the form 1 hairpin is shown to be a monomer in the high salt conditions, it is intriguing to imagine how the tri-uridine hairpin loop would form. A uridine-uridine base pair will bring the helix backbone closer together as compared to a Watson-Crick base pair because both pairing partners are pyrimidines. This would make it easier for the middle uridine to extend across the phosphates to close the loop. Some simple model building has shown that it is possible to maintain the U=U base pair in this manner. However, it is difficult to imagine how the U_{27} imino proton would be protected from solvent exchange in this model.

The experimental evidence that would unambiguously answer the question of whether the RNAs are monomers or dimers is to use the NMR method proposed by Pardi and coworkers (Aboul-ela, et al., 1994). In this scheme, ^{15}N labeled rLC13 RNA is mixed at a 1:1 ratio with ^{14}N labeled rLC13. A ½-X-filtered NOESY experiment is used to collect ^{14}N-^{14}N, ^{15}N-^{15}N and ^{14}N-^{15}N NOESY subspectra on the mixture. If crosspeaks are found connecting a ^{14}N labeled imino with a ^{15}N labeled imino, the sample must exist as a dimer. If the only crosspeaks found connect ^{14}N with ^{14}N iminos and ^{15}N with ^{15}N iminos, then the sample must be monomeric.

In conclusion, we have shown that the *in vitro* secondary structures of the *L. collosoma* and the *C. fasciculata* SL RNAs exist in the form 1 and form 2, respectively. The tertiary structure from the *L. collosoma* SL RNA is most likely due to melting out of transiently formed base pairing across the splice site, and does not involve interactions between the form 1 hairpin and the rest of the RNA. The form 1 hairpin from *L. collosoma* contains a U=U base pair, and is well behaved spectroscopically. The assignments of both the exchangeable and non-exchangeable protons for the low and high salt forms of the rLC13 RNA have been determined. Further work needs to be done to determine whether this RNA is a monomer or a dimer.

1.5 Materials and Methods

A number of methods were employed in the synthesis of RNA molecules discussed in this chapter. Each method will be described in this section, and table 1.1 lists each sample and what method was used in its synthesis.

1.5.1 Chemical synthesis of RNA

The HHMI Biopolymer/Keck Foundation Biotechnology Resource Laboratory provided the chemically synthesized RNAs discussed in this chapter. Each 1 μmole RNA synthesis was deprotected by suspension in a 2 mL solution of 1M Tributylammonium flouride in THF for 48 hrs. This solution was concentrated by speed-vac to a volume of less then 0.5 mL and desalted on a size exclusion column. This desalted solution of RNA was then ethanol precipitated and resuspended in a minimum volume of aqueous 8M urea and purified by standard denaturing poly-acrylamide gel electrophoresis (PAGE).

1.5.2 Enzymatic synthesis of RNA

All enzymatically synthesized RNAs were produced from a transcription reaction which utilized a bottom strand DNA template coding for the RNA plus a 5' 17 nucleotide T7 RNA polymerase promoter sequence (Milligan, et al, 1987). The top strand DNA template was complementary to the 17 nucleotide promoter sequence. The T7 RNA polymerase was overexpressed and purified as described previously. All transcription reactions were conducted under identical conditions, except that the magnesium ion concentration was optimized independently for each reaction. The reaction conditions for the transcriptions were typically 40 mM Tris HCl (pH 8.3 @ 20° C), 5 mM DTT, 1mM spermidine, 20 mM $MgCl_2$, 0.01% NP-40, 50 mg/ml PEG 8000, 4 mM in each rNTP (1 mM for $^{15}N/^{13}C$ labeled), 200nM DNA template, and 0.1 mg/ml T7 RNA polymerase. All reactions were carried out at 37° C for 4-8 hours and the products of the transcriptions were purified by 15% denaturing PAGE.

1.5.3 RNaseH cleavage

One of the major disadvantages of the enzymatic synthesis method of producing RNA using T7 RNA polymerase is that the reaction yields are highly dependent on the 5' end sequence. A method for avoiding this problem is to synthesize an RNA with a high yield 5' end sequence, and use RNase H and a 2'-O-methyl RNA/DNA chimera to direct a site-specific cleavage of the RNA. This reaction is described in detail in chapter 2 of this thesis.

1.5.4 List of samples

Table 1.3 lists each of the RNA samples discussed in this chapter and describes which method of synthesis was used to produce it. Molecule names beginning with "rCF" are from the *C fasciculata* SL RNA and those beginning with "rLC" are from the *L. collosoma* SL RNA. Methods of synthesis are abbreviated E=enzymatic, C=chemical and R=RNase H cleavage and are described in other sections of the materials and methods. The sequences of the DNAs used to transcribe the enzymatically synthesized RNA are not given, as they can be inferred from the cDNA sequence to the RNA.

Name	Description	Synthesis Method	Sequence (5’-3’)
rCF55	wild type	E	GGGAACUAACGCUAUAUAAGUAUCAGUUUCUGUACUUUAUUGGUAUAAGAAGCUU
rCF30	F2 hairpin	E, R	GUUUCUGUACUUUAUUGGUAUAAGAAGCUU
rCFf1m	F1 mutant[1]	E	GGGAACUAACGCUAUAUAA**CA**AUCAGUUUCUGU**UG**UUUAUUGGUAUAAGAAGCUU
rCFf2m	F2 mutant[1]	E	GGGAACUAACGCUAUAUAAGUAUC**UC**UUUCUGUACUUUAUUGGUAUAAGAAG**GA**U
rCFf2hp’	F2 hairpin and **rLDR**[2]	E	**GGGAUCACACAAUAC**GUUUCUGUACUUUAUUGGUAUAAGAAGCUU
rLC55	wild type	E	GGGAACUAAAACAAUUUUUGAAGAACAGUUUCUGUACUUCAUUGGUAUGGUAUGUAGAGA
rLC25	F1 hairpin	C	UUGAAGAACAGUUUCUGUACUUCAU
rLC13	F1 hairpin[3]	C	AACAGUUUCUGUA
or			
rLC13	F1 hairpin[3]	C	UACAGUUUCUGUA

[1] Mutations from wild type are shown in bold.
[2] rLDR portion of the sequence shown in bold, see Chapter 2 for a discussion of the rLDR sequence.
[3] Notice that this RNA was synthesized with a "A" and a "U" at the 5' end. The sequence change had no noticeable affect on any of the RNAs characteristics.

Table 1. 3 RNA samples

1.5.5 Optical equilibrium-melting curves

Nucleic acids have a strong UV absorbance at 260 nm because of the conjugated ring structures in the bases. The extinction coefficient for each nucleotide is somewhat dependent on the local environment in which the nucleotide exists. It so happens that the UV absorbance of a nucleotide is lower when it is involved in an RNA double strand helix, because of the tight stacking of the bases. When the nucleic acid is thermally induced to “melt” out of the helix to an unstructured single strand the UV absorbance rises in what is known as a hypochromic shift. The deflection point of the derivative of the change in UV absorbance with respect to temperature is known as the melting temperature (T_m). The T_m and the shape of the melting profile may be used to calculate the thermodynamic parameters of the nucleic acid (Gralla and Crothers, 1973; Puglisi and Tinoco, 1989).

All equilibrium-melting curves were collected on a Varian Cary 1 UV spectrophotometer. The samples were heated to above their melting temperature and either snap-cooled (by placing in ice/water or dry ice/isopropanol) or slow-cooled. The melts were carried out by first cooling the sample to 5 °C, then the temperature was raised by 0.5 to 1 °C per minute. The UV absorbance was collected every 1 °C. Data were processed and analyzed statistically using the software package Origin v4.1 (Microcal Software Inc, USA).

1.5.6 NMR Methods

Homonuclear and heteronuclear NMR data presented in this chapter were collected on either a Varian Unity 500 or Unity+ 600 spectrometer. Most samples were dialyzed extensively against 20 mM sodium phosphate buffer (pH 6.5) and 1 mM EDTA. The high salt buffers typically included 100-200 mM sodium chloride, the low salt buffers typically included 0-50 mM sodium chloride. The D_2O experiments were conducted using 99.996% D_2O and the H_2O experiments used 15% D_2O. Unless otherwise stated, 1024 complex points were collected in the direct dimension, and 300 points in the indirect dimension. Quadrature in the indirect dimension was accomplished using the States method. Data processing was performed using the software package Felix (Biosym Inc.). Unless stated differently a 90 degree shifted sine-bell was used to apodize the FIDs before Fourier transformation.

Spectra from the H_2O NOESY experiments were collected using either the Jump Return Spin-Echo NOESY (JRSE-NOESY) pulse sequence or the Watergate

NOESY (WNOESY) pulse sequence (Piotto, et al., 1992; Lippens, et al., 1995; Sich, et al., 1996). Typically, the sweep width was set to 10,000 hz on a 500 MHz spectrometer to insure complete coverage of the imino protons (14-10 ppm) and the offset frequency was centered on the H_2O line (4.75 ppm).

Spectra from the DQFCOSY experiments were collected using the canned dqfcosy.c pulse sequence supplied with the Varian spectrometers. The sweep width was set to 5000 Hz in each dimension on the 500 MHz spectrometers to insure coverage of the aromatic region of the spectrum (8-7 ppm) and the offset was centered on the residual HDO line (4.75 ppm). Data were processed by apodizing the FID with a zero degree shifted sine-bell.

Spectra from the D_2O NOESY experiments were collected using a modified version of the canned noesy.c pulse sequence supplied with the Varian spectrometers. The modification was to add a gradient pulse to the mixing time, to destroy any transverse magnetization because of single or double-quantum coupling. A low power 0.5 second water presaturation pulse was used to remove the residual HDO line. Typically, 2-10 second recycle delays were utilized. The sweep width was set to 5000 Hz in each dimension on a 500 MHz spectrometer to insure coverage of the aromatic region of the spectrum (8-7 ppm) and the offset was centered on the residual HDO line (4.75 ppm).

The HMQC pulse sequence was derived from that published by Szewczak (1993). The ^{15}N carrier frequency was set to 150 ppm to center on the imino proton nitrogens.

The translational self-diffusion experiments were performed (and the gradients were calibrated) as discussed in chapter 4 of this thesis. The pfg_diffusion.c pulse sequence (see 4.6.1) was used to collect the data, setting Δ=0.1s and δ=0.004s (other values were examined as well, with no effect on the results). 32 experiments were collected arraying the gradient strength from 0 to 31 G/cm. The processing of the data was performed using the Felix95 software package using the macro diffusion.mac (see 4.6.2). The resultant "xy" file was further processed using the xy2xm script (see 4.6.3) using a maximum gradient strength value of 32 G/cm. The final "xm" file was then graphed using the Origin v4.1 statistical software package (Microcal Software Inc, USA). Reported values of the translational self-diffusion rate and the error in the measurement come directly from the built-in linear regression package.

1.6 References

Aboul-ela F, Nikonowicz EP, Pardi A. 1994. Distinguishing between duplex and hairpin forms of RNA by 15N-1H heteronuclear NMR. *FEBS Lett 347*:261-264.

Bangs JD, Crain PF, Hashizume T, McCloskey JA, Boothyroyd JC. 1992. Mass spectrometry of mRNA cap 4 from trypanosomatids reveals two novel nucleosides. *J Biol Chem 267*:9805-9815.

Blumenthal T, Thomas J. 1988. *Cis* and *trans* mRNA splicing in *C. elegans*. *Trends in Genetics 4*:305-308.

Bruccoleri R, Heinrich G. 1988. An improved algorithm for nucleic acid secondary structure display. *Computer Applications in the Biosciences 4*:167-173.

Bruzik JP, Doran KV, Hirsh D, Steitz JA. 1988. *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. *Nature 335*:559-562.

Clayton C. 1992. Developmental regulation of nuclear gene expression in *Trypanosoma brucei*. *Prog Nucleic Acid res Mol Biol 43*:37-66.

Cload ST, Richardson PL, Huang YH, Schepartz A. 1993. Kinetic and thermodynamic analysis of RNA binding by tethered oligonucleotide probes: Alternative structures and conformational changes. *JACS 115*:5005-5014.

Cross M, Günzl A, Palfi Z, Bindereif A. 1991. Analysis of small nuclear ribonucleoprotein (RNPs) in Trypanosoma brucei: Structural organization and protein components of the spliced leader RNP. *Mol Cell Biol 11*:5516-5526.

Freier SM, Kiezerk R, Jaeger JA, Sugimoto N, Caruthers MH, Nelson T, Turner DH. 1986. Improved free-energy parameters for predictions of RNA duplex stability. *PNAS 83*:9373-9377.

Freistadt MS, Cross GAM, Robertson HD. 1988. Discontinuously synthesized mRNA from *Trypanosoma brucei* contains the highly methylated 5' cap structure, m^7GpppA*A*C(2'-O)mU*A. *J Biol Chem 263*:15071-15075.

Gralla J, Crothers DM. 1973. Free Energy of Imperfect Nucleic Acid Helices. III. Small Internal Loops Resulting from Mismatches. *J. Mol. Biol. 78*:301-319.

Harris K. 1995. Biochemical analysis of trypanosomatid SL RNA structure.

Harris KA, Crothers DM, Ullu E. 1995. In Vivo Structural Analysis of Spliced Leader RNAs in Trypanosoma brucei and Leptomonas collosoma: A Flexible Structure that Is Independent of Cap4 Methylations. *RNA 1*:351-362.

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. *Chemical Monthly 125*:167-188.

Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA. *PNAS 86*:7706-7710.

Karplus M. 1959. *J. Chem. Phys. 30*:11.

Konarska MM, Padgett PA, Sharp PA. 1985. *Trans* splicing of mRNA precursors *in vitro*. *Cell 42*:165-171.

Krause M, Hirsh D. 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. *Cell 49*:753-761.

Lapham J, Rife J, Moore PB, Crothers DM. 1997. Measurement of diffusion constants for nucleic acids by NMR. *J. Biomolecular NMR 10*:255-262.

LeCuyer KA. 1992. Conformational dynamics of the *L. collosoma* spliced leader RNA.

LeCuyer KA, Crothers DM. 1993. The *Leptomonas collosoma* Spliced Leader RNA Can Switch between Two Alternate Structural Forms. *Biochemistry 32*:5301-5311.

LeCuyer KA, Crothers DM. 1994. Kinetics of an RNA Conformational Switch. *PNAS 91*:3373-3377.

Lippens G, Dhalluin C, Wieruszeski JM. 1995. Use of the water flip-back pulse in the homonuclear NOESY experiment. *J Biomol NMR 5*:327-331.

Maniatis TM, Reed R. 1987. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature 325*:673-678.

Matzura O, Wennborg A. 1996. RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Computer Applications in the Biosciences 12*:247-249.

McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers 29*:1105-1119.

McNally KP, Agabian N. 1992. *Trypanosoma brucei* spliced leader RNA methylations are required for *trans*-splicing *in vivo*. *Mol Cell Biol 12*:4844-4851.

Michaeli S, Roberts TG, Watkins KP, Agabian N. 1990. Isolation of distinct small ribonucleoprotein particles containing the spliced leader and U2 RNAs of *Trypanosoma brucei*. *J Biol Chem 265*:10582-10588.

Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide Synthesis using T7 RNA Polymerase and Synthetic DNA Templates. *NAR 15*:8783-8798.

Murphy WJ, Watkins KP, Agabian N. 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. *Cell 47*:517-525.

Padgett PA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. 1986. Splicing of messenger RNA precursors. *Ann Rev Biochem 55*:1119-1150.

Perry KL, Watkins KP, Agabian N. 1987. Trypanosome mRNAs have unusual "cap 4" structures acquired by addition of a spliced leader. *PNAS 84*:8190-8194.

Piotto M, Saudek V, Sklenar V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR 2*:661-665.

Puglisi JD, Ignatio Tinoco J. 1989. Absorbance melting curves of RNA. *Meth in Enzymology 180*:304-325.

Sich C, Flemming J, Ramachandran R, Brown LR. 1996. Distinguishing Inter- and Intrastrand NOEs Involving Exchangeable Protons in RNA Duplexes. *J Mag Res Series B 112*:275-281.

Sutton PE, Boothroyd JC. 1986. Evidence for *trans* splicing in trypanosomes. *Cell 47*:527-535.

Szewczak AA, Kellogg GW, Moore PB. 1993. Assignment of NH Resonances in Nucleic Acids Using Natural Abundance ^{15}N-^{1}H Correlation Spectroscopy with Spin-Echo and Gradient Pulses. *FEBS Lett 327*:261-264.

Turner DH, Sugimoto N, Freier SM. 1988. RNA Structure Prediction. *Ann Rev Biophys Chem 17*:167-192.

Turner DH, Sugimoto N, Jaeger JA, Longfellow CE, Freier SM, Kierzek R. 1987. Improved parameters for prediction of RNA structure. *Cold Spring Harb Symp Quant Biol 52*:123-133.

Ullu E, Tschudi C. 1991. *Trans* splicing in trypanosomes requires methylation of the 5' end of the spliced leader RNA. *PNAS 88*:10074-10078.

Ullu E, Tschudi C. 1993. 2'-O-methyl RNA oligonucleotides identify two functional domains in the trypanosome spliced leader ribonucleoprotein particle. *J Biol Chem 268*:13068-13073.

Ullu E, Tschudi C, Günzl A. 1995. *Trans-splicing in trypanosomatid protozoa.* Oxford, UK: Oxford University Press.

Xu J, Lapham J, Crothers DM. 1996. Determining RNA Solution Structure by Segmental Isotopic Labeling and NMR:Applications to *Caenorhabditis elegans* Spliced Leader RNA. *PNAS 93*:44-48.

Zuker M. 1989. On finding all suboptimal foldings of an RNA molecule. *Science 244*:48-52.

Zuker M, Stiegler P. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *NAR 9*:133-148.

CHAPTER 2 “AN RNA ENDONUCLEASE”

2.1 Summary

This chapter presents a method for site-specifically cleaving RNA of any sequence, with high reaction yield. The reaction has been adapted to the cleavage of milligram quantities of RNA, suitable for the sample preparation needs of nuclear magnetic resonance and X-ray crystallographic studies.

2.2 Introduction

Site-specific endonuclease cleavage of DNA is a powerful tool for molecular biologists, making possible procedures such as gene cloning. This reaction is accomplished by means of restriction enzymes that recognize, bind and cleave specific DNA sequences and are usually high yielding. Site-specific restriction enzymes, however, do not exist for RNA. Rather, the world is replete with the bane of the RNA biologist, the non-specific nuclease. One such nuclease, Ribonuclease H (RNase H), has the interesting property in that it only binds to RNA that is base paired with DNA and catalyzes the hydrolysis of the phosphodiester backbone between the nucleotides of the RNA strand. Thus, RNase H'es are biologically important for "cleaning up" during processes which would generate long stretches of RNA/DNA strands, such as during reverse transcription (reverse transcriptase has a built-in RNaseH functionality).

RNaseH is not, however, very site-specific by nature. If one binds a long sequence of complementary DNA to RNA, the cleavage can occur in any position shared by the RNA/DNA duplex. In 1987, Inoue *et al.*, recognized that this could be exploited to cleave "specifically" if one could reduce the number of DNA nucleotides bound to the RNA. In fact, they showed that exactly four DNA nucleotides gave a single specific

RNA cleavage, presumable due to the RNaseH requiring a four base pair binding site. In order to increase the thermodynamic stability of this complex, they surrounded the four DNA nucleotides with stretches of 2'-O-methyl RNA (Fig. 2.1). The 2'-O-methyl RNA is not recognized by the RNase H as a suitable substrate for binding, and thus does not interfere in the reaction, serving to "hold" the DNA in place. A few more papers were published by the Japanese group (Inoue, *et al.*, 1988; Hayase, *et al*, 1990), the later paper demonstrated that a tRNA could be cleaved in different positions using this technique.

Figure 2. 1 RNase H cleavage position

↓

Target RNA: 5′——NNNNNNNNNNNNNNNNN——3′

2'-O-methyl RNA/DNA chimera: 3′–$\underline{\text{NNNNNNN}}$**NNNN**$\underline{\text{NNN}}$–5′

Underlined characters, $\underline{\text{N}}$, represent 2'-O-methylated RNA. Bold characters, **N**, represent DNA. Regular characters, N, represent RNA. The arrow, ↓, indicates the position of cleavage.

This is an important reaction for the RNA biologist that may not have been fully realized. There are a number of very interesting properties of this reaction, the reaction time is short, the efficiency of cleavage is high and the cleavage products of RNase H reactions have a 5'-phosphate and a 3'-hydroxyl (Berkower *et al.*, 1973; Zawadzki & Gross, 1991). These properties can be used to accomplish a number of highly desirable tasks for someone working with RNA. First, this allows for a procedure to synthesize RNAs that have no sequence requirements at their 5' end. The synthesis of large quantities of RNA for biophysical study is often accomplished using the bacteriophage T7 RNA polymerase and a DNA template. One of the unfortunate problems of using this enyzme is that the transcription yield of the target RNA is often highly dependent on the nucleotide sequence of the 5' end. Synthesizing a target RNA with a "high yield" 5' end sequence, followed by a cleavage reaction to produce the final target RNA product can circumvent this problem. The second use of this RNA endonuclease technique is that the RNA products of the cleavage reaction can be used directly in a subsequent religation reaction. This makes it possible, for example, to synthesize "segmentally isotope labeled" RNA for use in NMR (Xu, *et al*., 1996).

The bacteriophage protein T7 RNA polymerase has been used in *in vitro* transcription reactions to generate large quantities of RNA (Milligan *et al*., 1987; Milligan & Uhlenbeck, 1989). While other polymerases have been used to produce RNA, T7 RNA polymerase has been found to be the most amenable to large scale (milligram) RNA synthesis and can be readily obtained in large quantities by over expression and purification techniques (Grodberg & Dunn, 1988; Davanloo *et al*., 1984; Zawadzki & Gross, 1991). It has been shown that the first six nucleotides at the 5' end of the RNA product are important in determining how efficiently the reaction will proceed. Typically, sequences at the 5' end of the RNA must fit a [G(1)G/C(2)N(3)] consensus sequence in order to transcribe well (Milligan & Uhlenbeck, 1989). For this reason, RNAs used in biophysical studies produced by T7 RNA polymerase often contain modifications at their 5' end sequence to maximize transcription yield, a compromise that sometimes must be avoided.

Under the optimum conditions this reaction occurs rapidly, can be scaled up to milligram quantities of RNA, is highly efficient and is absolutely site specific. We have exploited this cleavage reaction to circumvent the problems the T7 RNA polymerase has with transcribing low yielding RNA sequences. We demonstrate that the 2'-O-methyl RNA on the 5' side of the DNA is not a necessary component for the reaction to proceed.

This allows for the same chimeric construct to be used in the production of any RNA sequence, since the base pairing between the chimera and the target RNA occurs only along the 5' side of the cleavage site, as shown below in figure 2.2.

Figure 2. 2 RNase H cleavage occurs without the 5' 2'-O-methyl RNA

↓

Target RNA: 5′——NNNNNNNNNNNNNNNN——3′

2'-O-methyl RNA/DNA chimera: 3′–NNNNNNN**NNNN**–5′

Underlined characters, N, represent 2'-O-methylated RNA. Bold characters, **N**, represent DNA. Regular characters, N, represent RNA. The arrow, ↓, indicates the position of cleavage.

We also demonstrate that this reaction can also be executed off a solid phase media. The chimera can be produced with a 3'-biotin label and adhered to a streptavidin-agarose bead, and the RNase H cleavage reaction can then be performed off this solid phase matrix. The reaction occurs with similar results to those obtained in the solution phase reaction, with the added benefit of an easy route for reusing the 2'-O-methyl RNA/DNA chimera for future reactions.

We demonstrate the practical applicability of this RNA endonuclease reaction by synthesizing an ^{15}N isotope labeled 30 nucleotide RNA hairpin "r3LIG", which contains an inherent poorly transcribing 5' end sequence. The 5' end of this RNA could not be modified since future experiments with the RNA involve ligation of an unlabeled piece of RNA to the 5' end. Any modifications to the sequence would then become internal sequence modifications, which may change the physical properties of the RNA. The RNA was produced by adding 15 nucleotide high-yielding leader sequence "rLDR", as shown below in figure 2.3, and subsequently cleaving rLDR away from the r3LIG RNA.

Figure 2. 3 Synthesis of r3LIG RNA

	5′-GGGAUCACACAAUAC↓GUUCUGUACUUUAUUGGUAUAAGAAGCUU-3′
"rLDR"+"r3LIG":	5′-GGGAUCACACAAUACGUUCUGUACUUUAUUGGUAUAAGAAGCUU-3′
2'-O-methyl chimera:	3′–$\underline{\texttt{CCCUAGUGUGU}}$**TATG**–5′
	↓ RNase H
"rLDR":	5′-GGGAUCACACAAUAC-3′
2'-O-methyl chimera:	3′–$\underline{\texttt{CCCUAGUGUGU}}$**TATG**–5′
	+
"r3LIG":	5′-GUUCUGUACUUUAUUGGUAUAAGAAGCUU-3′

Underlined characters, $\underline{\text{N}}$, represent 2'-O-methylated RNA. Bold characters, **N**, represent DNA. Regular characters, N, represent RNA. The arrow, ↓, indicates the position of cleavage.

A two-dimensional ^{1}H-^{15}N HMQC NMR spectrum is shown from the final product to demonstrate that these reactions can indeed be accomplished on large quantities of RNA.

2.3 Results

2.3.1 Enhancement of transcription yield with a leader sequence

Transcription of the 30 nucleotide hairpin of *C. fasciculata* r3lig with its wild type sequence (5'-rGUUUCUGUACUUUAUUGGUAUAAGAAGCUU-3') using T7 RNA polymerase at best gave a yield of 0.32 nmoles of RNA per 1 ml of reaction after gel purification. Synthesis of an NMR sample of this RNA would require greater then 200 mL of transcription. However, addition of the 15 nt leader sequence rLDR (5' GGGAUCACACAAUAC 3') to the 5' end of the r3lig sequence increased the yield to an average of 10 nmoles of RNA per 1 mL of transcription reaction after gel purification. The yield comparison between these two RNA molecules has further been quantitated by spiking small scale transcription reactions with á-^{32}P UTP and using a phosphorimager to analyze the purification gel (Fig. 2.4). These data show an approximate 13 fold increase in the molar yield of the RNA product of the rLDRr3lig over the r3lig RNA, after taking into account the difference in the number of uridines in the two RNAs.

2.3.2 Yield and site specificity of the cleavage reaction

The rLDRr3lig RNA must be post transcriptionally processed by RNase H to generate the final r3lig RNA. Two chimeras were constructed, 2'SURROUND and 2'LDR, to test for the necessity of having the 2'-O-methyl RNA flanking both sides of the four DNA nucleotides. The RNase H cleavage reaction was attempted with both chimeras in solution (see Fig. 2.5a), and it is clear that both reactions were successful. Scaling the reaction up to NMR quantities of RNA (Fig. 2.5b) shows the typical large scale cleavage yield obtained. More than 90 % of the input target RNA is converted to

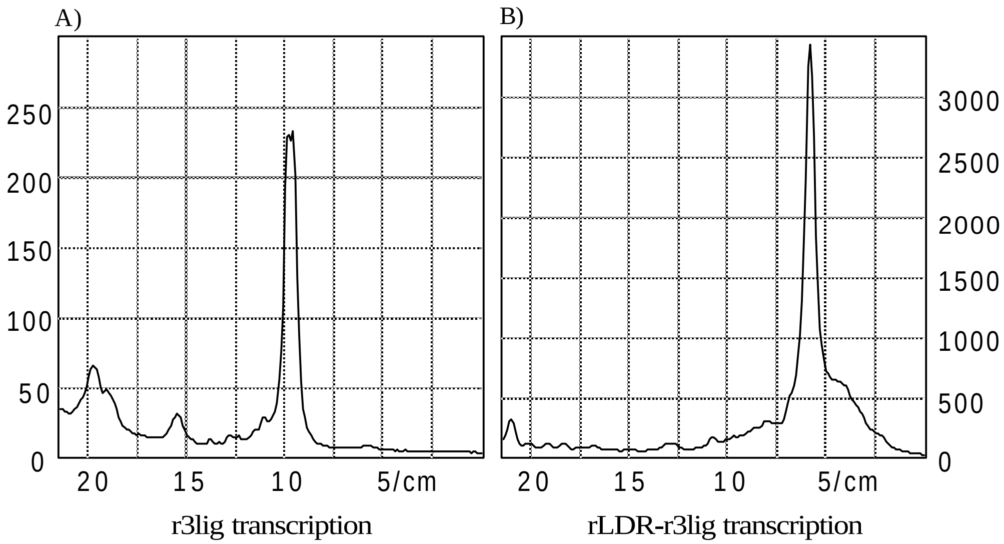


Figure 2. 4 Transcription Comparison: r3lig with rLDRr3lig

Phosphorimage data from 20 uL transcription reactions spiked with á ^{32}P rUTP (40mM Tris-HCl pH 8.3, 20 mM MgCl2, 50 mg/ml PEG 8000, 5 mM DTT, 1 mM Spermidine, 0.01% NP-40, 200 nM DNA template, 4 mM each rNTP, 5 uCi á ^{32}P rUTP and 0.1 mg/ml T7 RNA polymerase at 37 C for 4 hours). Yield comparisons of (A) r3lig and (B) rLDRr3lig demonstrates the poor transcription yield of the r3lig RNA.

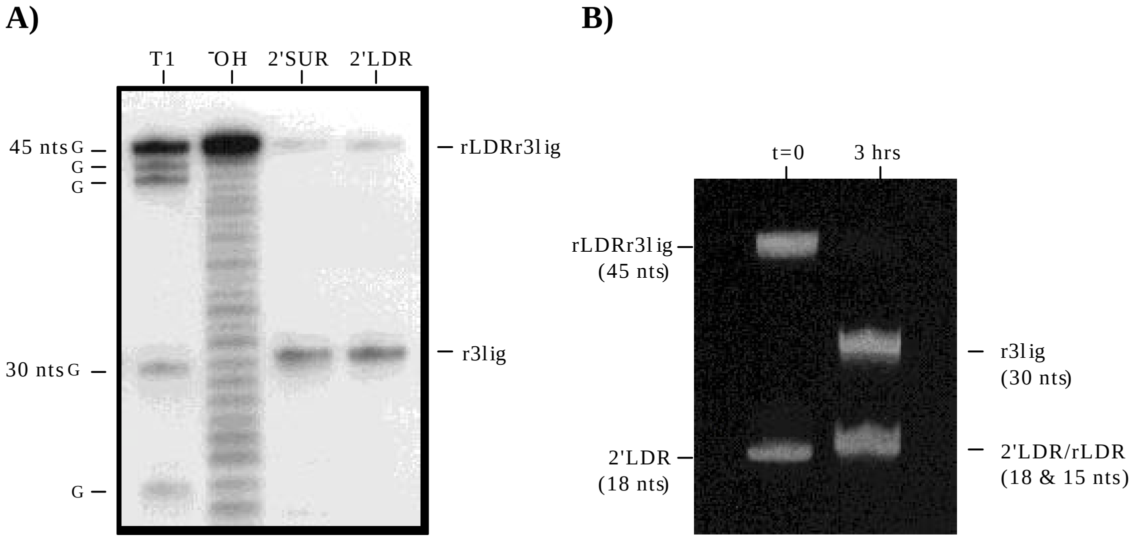


Figure 2. 5 Site specificity of the cleavage and large scale cleavage

A) Site specific cleavage of the 3' end labeled rLDRr3lig by RNase H. Lane 1; $^{-}$OH ladder, lane 2; T1 digestion, lane 3; 2'LDR directed cleavage, lane 4; 2'SURROUND directed cleavage. The product of the cleavage of rLDRr3lig is 3' end labeled 30 nucleotide r3lig. **B)** Ethidium stained 20% PAGE of NMR scale cleavage of rLDRr3lig by 2'LDR.

the desired 30 mer RNA r3LIG, which runs near the expected size of 30 nucleotides on the ethidium stained gel.

For higher precision in identifying the site of cleavage, a 3' labeled sample of rLDRr3lig was purified on a denaturing gel to remove 3' end degeneracy, subjected to cleavage, and the products analyzed on a sequencing gel (Fig. 2.5a). For both the 2'SURROUND and 2'LDR chimera-directed cleavage, there is a barely detectable level of a 31 nucleotide product, in addition to the dominant 30 nucleotide band. However, comparison with the partial T1 ribonuclease digestion land shows a similar level of minor contamination. We conclude that the presence of the n+1 band is due not to lack of specificity in the RNase H cleavage site, but rather to residual n+1 contamination of the starting oligomer. Hence, both chimeras were successful in directing the site-specific cleavage of RNase H. Because the 2'LDR chimera does not base pair to the RNA sequence on the 3' side of the cleavage site, it may be used for production of any RNA sequence. All large scale cleavage reactions were consequently performed with this 2'LDR chimera.

2.3.3 Cleavage on a solid state matrix

A 2'-O-methyl RNA/DNA chimera, "B2'LDR", was synthesized with a 3' end biotin label. This chimera was complexed to a streptavidin-agarose bead matrix and was successfully employed to cleave RNA on this solid phase support. After preparation of the beads and complexing of the B2'LDR chimera to the bead, two reactions were performed. In the first, 5' ^{32}P end labeled rLDRr3LIG was incubated with the beads and cleaved with RNase H. After 3 hours of the reaction, greater then 90% of the counts remained bound to the beads (Fig. 2.6), demonstrating that the rLDR remains bound to the beads after cleavage. In the second reaction, 3' ^{32}P end labeled rLDRr3LIG was produced, bound to the B2'LDR beads and cleaved. After 3 hours of reaction, 70-80% of the counts could be found in the supernatant (Fig. 2.6). As the reaction proceeds, the 3' end of the rLDRr3lig is released into the supernatant. This radiolabeled piece of RNA was analyzed by sequencing and found to indeed be the 30 nucleotide r3LIG RNA (data not shown).

This "cleavage column" was not tested for the ability to scale up to NMR quantities of RNA, because it would require large quantities of bead matrix. However, the method worked quite well in the small-scale reactions, especially when working with radiolabeled RNA. The reaction is followed easily when working with 3' ^{32}P end labeled RNA by observing the increase in counts in the supernatant as the RNA is cleaved from the beads, and no further purification was necessary after cleavage. The B2'LDR beads were also shown to be recyclable. By addition of denaturants at warm temperatures, the post cleavage 5' RNA piece can be removed into the supernatant, and the beads may be used again.

2.3.4 NMR sample preparation

Preparation of a NMR sample of r3LIG required 30 mL of rLDRr3LIG transcription, at an average yield of 5.4 nmoles of RNA per ml of reaction after RNase H cleavage and PAGE purification. The 2'-O-methyl RNA/DNA chimera 2'LDR was used to direct the cleavage the RNA in solution by RNase H and the reaction was followed by

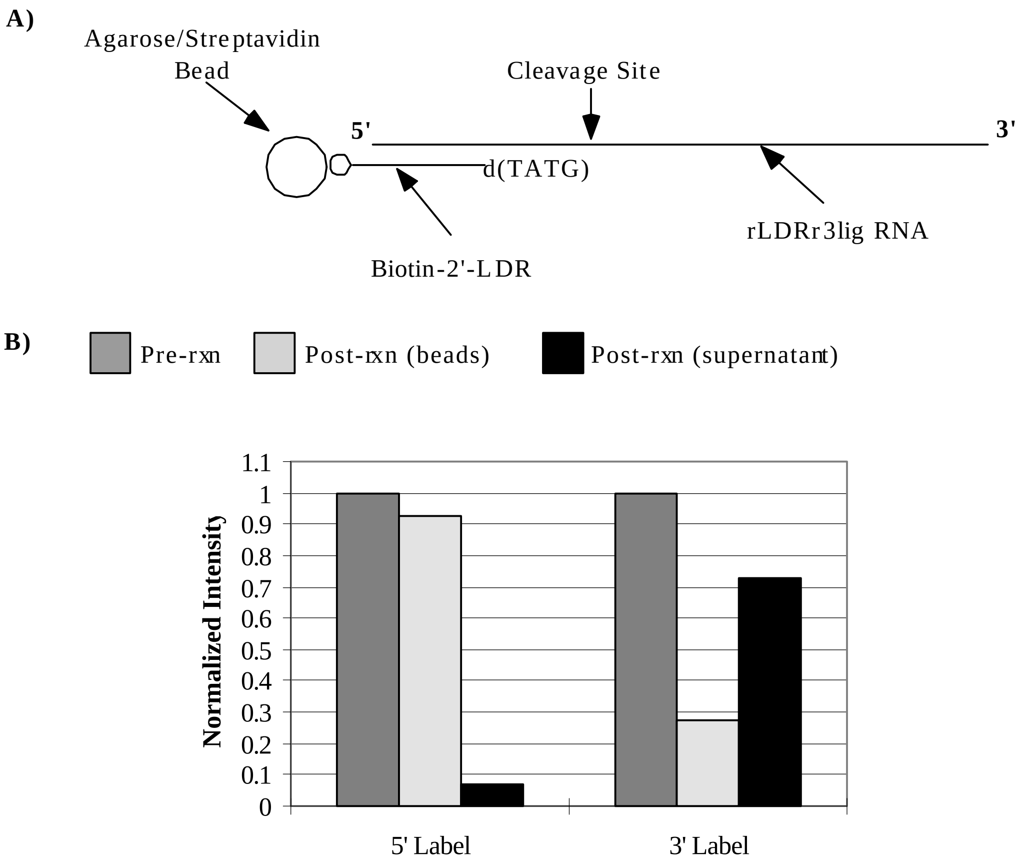


Figure 2. 6 Solid State RNase H Cleavage

A) Diagram of rLDRr3lig RNA bound to B2’LDR column. **B)** Results of RNase H cleavage of rLDRr3lig RNA bound to B2’LDR column. After 3 hours of reaction, the supernatant was removed from the beads by centrifugation and the beads were rinsed. The 5’ end labeled RNA remained bound to the beads, while the 3’ end labeled RNA came off with the supernatant.

denaturing mini-gel until completion, taking an average of 3 hours. After the final gel purification a final yield of 75 nmoles of r3lig RNA was obtained.

2.3.5 NMR spectroscopy

The NMR spectroscopy demonstrates that the isotopically labeled nucleotides were incorporated into the sample and that the sample is adequately concentrated. The 2 dimensional ^{1}H - ^{15}N HMQC (Fig. 2.7) clearly shows 4 A-U base pairs and 2 G-C base pairs. We do not detect the G7-U21 and G27-U2 base pairs at the temperature and buffer used in this experiment, but we have been able to observe corresponding resonances at colder temperatures and higher ionic strength buffers. We have not been able, however, to observe the U11-G17 or G1-C26 base pairs at any condition, probably because of fast solvent exchange due the hairpin loop opening and helical end fraying respectively.

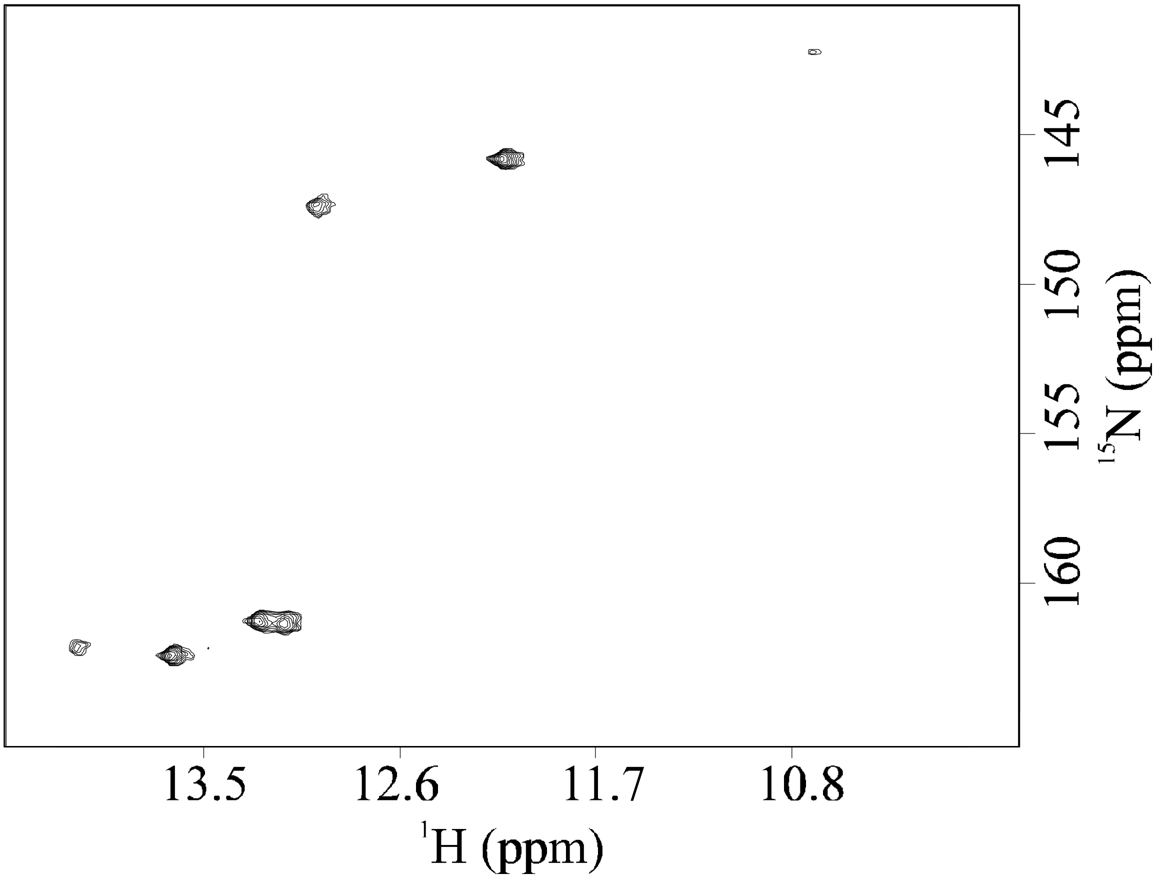


Figure 2. 7 2D ^{1}H-^{15}N HMQC of r3lig product from the RNase H cleavage

2 dimensional ^{15}N-^{1}H HMQC spectra of r3lig obtained from the RNase H cleavage reaction.

2.4 Discussion

In summary, we describe a method that permits the synthesis of large quantities of RNA without regard to the final 5' end sequence. This is accomplished by means of adding a 15 nucleotide leader sequence to the 5' end of the desired RNA, which is subsequently cleaved away from the final product via site-directed RNase H cleavage. The use of the removable 5' leader RNA sequence greatly enhances transcription yield because it can be constructed out of any high transcription yield sequence. Cleavage of the leader sequence from the desired RNA occurs in high yield (>90%), and can be scaled up to large quantities of RNA (milligram).

The same 2'-O-methyl RNA/DNA chimera can be used for any cleavage reaction. The requirement that the RNA portion of the site-directing chimera exist on both sides of the 4 deoxyribonucleotides is not necessary for efficient, site-specific cleavage of RNA. If the chimera is constructed like the 2'LDR sequence, it can be reused for many RNA molecules, since there is no base pairing overhangs between the chimera and the unknown final RNA target sequence. This affords a great advantage in that the chimera can be produced before knowing what RNA sequence is desired. It is also shown that these reactions can be carried out on a solid phase via a biotin-streptavidin linkage between an agarose bead and the chimera. This has interesting possibilities for the construction of a "RNA cleavage column" which could be reused.

2.4.1 The religation of RNA cleavage products

The RNA products from the reaction terminate with a 3' hydroxyl and a 5' phosphate for the 5' and 3' piece respectively. This is intriguing in that these are the required end chemistries for further biochemistry, such as in use with DNA ligase. In fact, this fact has been taken advantage of in a variety of ways. Xu et al. (1996) used this idea to construct a "segmentally" labeled *Caenorhabditis elegens* spliced leader RNA for NMR studies in which sections of the RNA was isotopically labeled. This was accomplished by means of synthesizing of a fully labeled and unlabeled version of the RNA, cleaving them at the same position using this technique, and finally religating a labeled section onto an unlabeled section (and vice versa). In this manner, the secondary structure of the RNA could be unambiguously assigned.

Yu and Steitz (1997b) used this technique to introduce a 4-thiouridine (^{4S}U) nucleotide into a pre-mRNA substrate. The ^{4S}U nucleotide is then used as a structural probe by its propensity to crosslink when exposed to UV light. The pre-mRNA molecule is first cleaved site specificially using this method, then the ^{4S}U nucleotide is added to the 5' half of the RNA with T4 RNA ligase (a template free reaction). The 5' and 3' half RNA are then ligated using a DNA guide template and T7 DNA ligase. The final product contains a single ^{4S}U nucleotide at any desired position within the molecule.

2.4.2 Detection of 2'-O-methyl sites in RNA

Yu *et al*. (1997a) also used this technique as a method of detecting sites of 2'-O-methylation in RNA molecules. Since the cleavage of RNA by RNase H is presumed to go through a 2'-O-P-O-3' intermediate, they assumed that if the 2' hydroxy of the target RNA were blocked with a methyl group, the reaction would not occur. They were indeed able to detect sites of 2'-O-methylation, in both chemically synthesized RNAs with known sites of methylation and in biologically interesting RNA with unknown sites.

2.4.3 The RNase H enzyme source affects the cleavage position

The work of Yu and Steitz (1997a,b), however, did raise one question about the technique. They found that the cleavage position for their chimeric constructs composed of four deoxyribonucleotides was at a position one nucleotide in the 5' direction on the target RNA, as shown below in figure 2.8B.

Figure 2. 8 RNase H cleavage positions for different enzyme sources

A) Pharmacia (cat. # 27-0894), Sigma (cat. # R-6501) or Takarashuzo RNase H

$$\begin{array}{ll} & \phantom{5'\text{——}\mathrm{NNNNNNNNNNNN}}\downarrow \\ \text{RNA:} & 5'\text{——}\mathrm{NNNNNNNNNNNNNNNN}\text{——}3' \\ \text{2'-O-methyl RNA/DNA chimera:} & \quad 3'\text{–}\underline{\mathrm{NNNNNNN}}\mathbf{NNNN}\underline{\mathrm{NNN}}\text{–}5' \end{array}$$

B) Boehringer Mannheim (cat. # 786 349) RNase H four deoxyribonucleotides

$$\begin{array}{ll} & \phantom{5'\text{——}\mathrm{NNNNNNNNNNN}}\downarrow \\ \text{RNA:} & 5'\text{——}\mathrm{NNNNNNNNNNNNNNNN}\text{——}3' \\ \text{2'-O-methyl RNA/DNA chimera:} & \quad 3'\text{–}\underline{\mathrm{NNNNNNN}}\mathbf{NNNN}\underline{\mathrm{NNN}}\text{–}5' \end{array}$$

C) Boehringer Mannheim (cat. # 786 349) RNase H three deoxyribonucleotides

$$\begin{array}{ll} & \phantom{5'\text{——}\mathrm{NNNNNNNNNNN}}\downarrow \\ \text{RNA:} & 5'\text{——}\mathrm{NNNNNNNNNNNNNNNN}\text{——}3' \\ \text{2'-O-methyl RNA/DNA chimera:} & \quad 3'\text{–}\underline{\mathrm{NNNNNNNN}}\mathbf{NNN}\underline{\mathrm{NNN}}\text{–}5' \end{array}$$

Underlined characters, $\underline{\mathrm{N}}$, represent 2'-O-methylated RNA. Bold characters, **N**, represent DNA. Regular characters, N, represent RNA. The arrow, ↓, indicates the position of cleavage.

While the gel data (Fig. 2.5) clearly shows that the cleavage position is as demonstrated above in figure 2.8A, one of Jing Xu's NMR experiments on her segmentally labeled RNAs unequivocally demonstrates that we had correctly assigned the cleavage position (Fig. 2.9). The difference in positioning was finally understood when the source of the enzymes used in each study was examined. In all the studies previous to the Steitz experiments, the RNase H enzyme source was from either Pharmacia, Sigma or the Takarashuzo companies. The Steitz lab had used Boehringer Mannheim enyzme. An

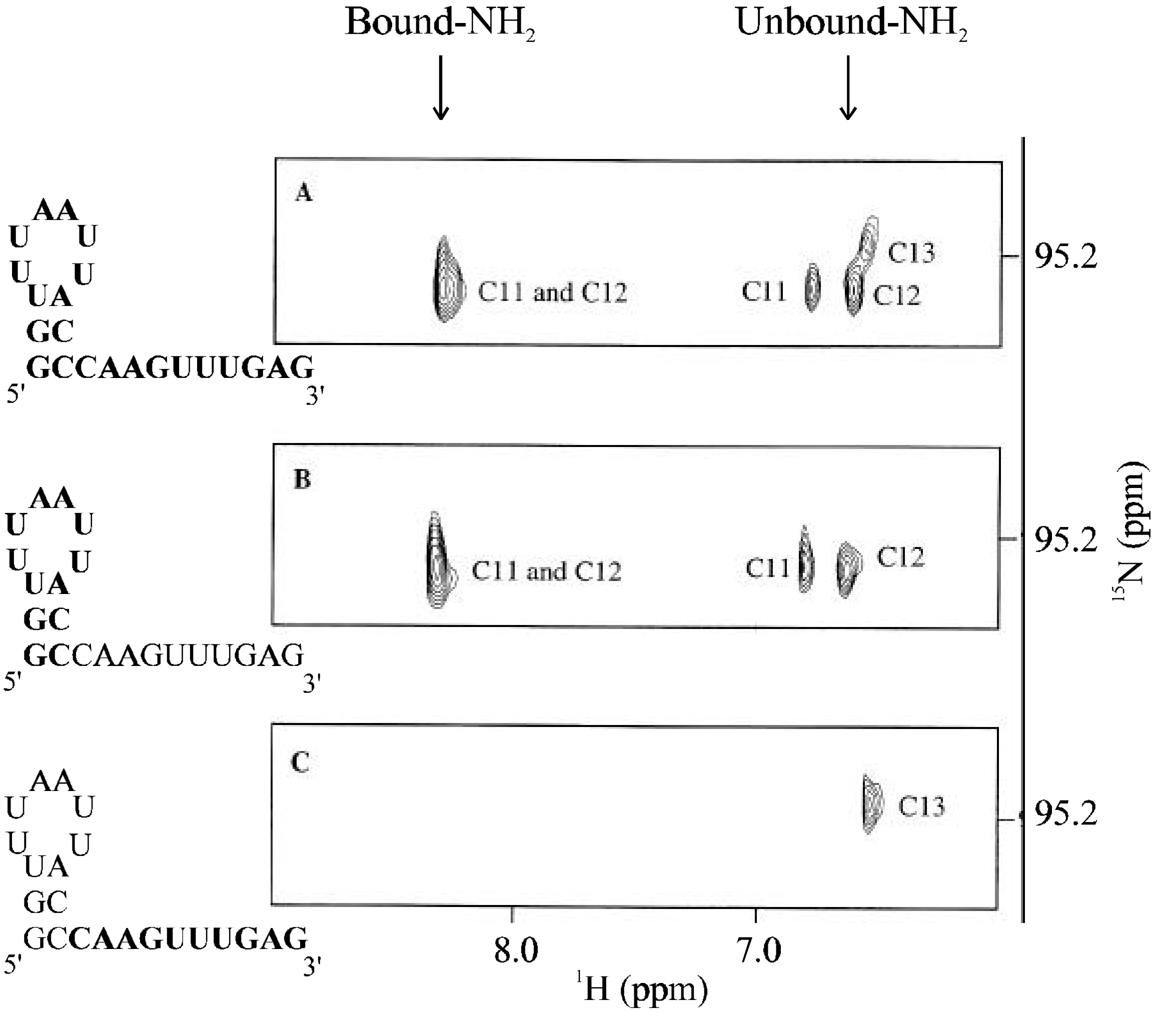


Figure 2. 9 Site specificity of the RNase H cleavage as seen by NMR

Figure kindly provided by Jing Xu (1997). The amino region of ^{1}H-^{15}N HMQC from the 22 nt RNA "CEDONOR" which was synthesized segmentally labeled with ^{15}N (Xu, 1996). **A)** Fully ^{15}N labeled CEDONOR, **B)** the 5' half labeled RNA and **C)** the 3' half labeled RNA. The three amino assignments, C11, C12 and C13 are shown in the figure. It is clear that the two base paired aminos from C11 and C12 are present in the 5' half labeled sample and that the unbase paired C13 is present in the 3' half labeled sample.

experiment in which the same RNase H cleavage reaction was performed side-by-side, except one reaction used Pharmacia RNase H and one used Boehringer Mannheim RNase H. The results (Lapham, *et al*., 1997) (data not shown) were that indeed the position of the cleavage was different by one nucleotide. While it is unknown what the exact reason is for the differences, we do note that the storage buffer for the Boehringer Mannheim RNase H does not contain EDTA and is relatively low in salt concentration compared to the other enzymes. Therefore, when precise cleavage using chimeric oligonucleotides is required, we recommend caution in the construction of the oligonucleotides and in the choice of supplier of enzyme.

2.5 Materials and methods

2.5.1 Oligonucleotide synthesis

All DNA oligonucleotides used as templates for T7 RNA polymerase transcription reactions were synthesized on an Applied Biosystems 380B DNA synthesizer in 1 μmole quantities. The three 2'-O-methyl RNA/DNA chimeras were synthesized by the Keck Foundation Oligonucleotide Synthesis Facility at Yale University in 1 μmole quantities. All oligonucleotides were purified by electrophoresis on denaturing 15% polyacrylamide gels. The sequences and names of these chimeras are as follows:

Abbreviation	Full Name	Sequence (5' - 3') (**RNA in bold is 2'-O-methyl**)
2'SURROUND	2'-O-CH_3-SURROUND	$_r$(**UAGUGUGU**)$_d$(TATG) $_r$(**CAAAG**)
2'LDR	2'-O-CH_3-LEADER	$_r$(**ACGCCCUAGUGUGU**)$_d$(TATG)
B2'LDR	Biotin-2'-O-CH_3-LEADER	Biotin-$_r$(**ACGCCCUAGUGUGU**)$_d$(TATG)

2.5.2 Enzymes

RNase H used in the cleavage reactions was obtained from Pharmacia (27-0894) at 1.9 units/μl where 1 unit is defined as able to catalyze the production of 1 nanomole acid-soluble RNA nucleotide in 20 minutes at 37° C. T4 DNA ligase used in the ligation reactions was obtained from New England Biolabs (202L) at 400 units/μl. T7 RNA polymerase was produced using published techniques (Grodberg & Dunn, 1988; Davanloo *et al*., 1984; Zawadzki & Gross, 1991).

2.5.3 T7 RNA Polymerase Transcriptions

All RNA transcriptions utilized a bottom strand DNA template coding for the RNA plus a 5' 17 nucleotide T7 RNA polymerase promoter sequence. The top strand DNA template was complementary to the 17 nucleotide promoter sequence. All reactions were conducted under identical conditions, except that the magnesium ion concentration was optimized independently for each reaction. ^{15}N isotopically labeled NTPs were obtained using published methods (Nikonowicz *et al.*, 1992; Batey *et al.*, 1992), modified as described below. The reaction conditions for the transcriptions were 40 mM Tris HCl (pH 8.3 @ 20° C), 5mM DTT, 1mM spermidine, 20 mM $MgCl_2$, 0.01% NP-40, 50 mg/ml PEG 8000, 2mM in each rNTP, 200nM DNA template, and 0.1 mg/ml T7 RNA polymerase. All reactions were carried out at 37° C for 4-8 hours. Products of the transcriptions were purified by 15% denaturing PAGE.

Comparisons of transcription yields between r3lig and rLDRr3lig, shown in fig. 2.2, were carried out by analyzing 20 μl transcriptions spiked with 5 μCi of α-^{32}P-UTP, run on 15% denaturing gels, and quantitated by phosphorimager (Fuji Inc., Fujix 2000) analysis. Calculations of transcription yields for the body α-^{32}P-UTP labeled RNAs included a correction factor for the number of uridines in the sequence.

2.5.4 ^{15}N NTP isolation and purification

We used the methods of Batey *et al.* and Nikonowicz *et al.* (1992; 1992), with modification of the method of isolation of nucleic acids from the cell extract. *E. coli* cells were grown on a minimal media containing ^{15}N ammonium chloride as the only nitrogen source. The cells were harvested in the log phase of cell growth by centrifugation. The cell pellet was resuspended in a minimal volume (20 ml per liter growth) of STE buffer (0.1 M NaCl, 10 mM Tris-HCl @ pH 8.0, 1.0 mM EDTA @ pH 8.0) and 0.5% SDS. This whole cell slurry was then sonicated in a Branson Sonifier 450 brand sonicator at its highest power setting for 4 minutes, allowed to cool on ice for 5 minutes, then the procedure was repeated 3 times. This slurry was then extracted once with 25:24:1 equilibrated phenol (pH 8.0) : chloroform : isoamyl alcohol at 60° C for 30 minutes with constant stirring. The mixture was centrifuged, and the aqueous phase removed and saved. The phenol layer was back extracted once with 1/2x volume STE buffer, the aqueous phase removed, and pooled with that from the first extraction. The pooled aqueous phase was extracted 3 times with 1/2x volume chloroform, leaving an aqueous phase essentially free of phenol contamination. The total cellular nucleic acids were precipitated by adding 1/10 volume 3 M sodium acetate and 1x volume isopropyl alcohol and centrifuging.

The pellet was dried and resuspended in P1 nuclease digestion buffer (15 mM sodium acetate @ pH 5.2 and 0.1 mM $ZnSO_4$). The nucleic acids were denatured by heating to 95° C for 1 minute and snap cooled in ice**.** 10 units of P1 nuclease and 100 units of DNase I were added per liter of cell growth and incubation was continued at 37 °C until there were no polymers of nucleic acid left by PAGE analysis, typically 12 hrs. The desalting procedures and conversions to ribonucleotide triphosphates were identical to those published previously (Nikonowicz *et al*., 1992; Batey *et al*., 1992). After complete conversion of the ribonucleotides from the monophosphate to the triphosphate, no further purification was necessary, and the nucleotide triphosphates could be used immediately in transcription reactions.

2.5.5 Cleavage of RNA with the 2'-O-methyl RNA/DNA chimeras in solution

All RNase H cleavage reactions contain 20 mM HEPES-KOH pH 8.0, 50 mM KCl, and 10 mM $MgCl_2$. The chimera was annealed to RNA by heating to 90° C and slowly cooling to room temperature at high concentration, typically in the millimolar range. The chimera was kept at 1.2 times the RNA concentration to insure complete hybridization of the RNA. RNase H was added to a final 20 units per 100 µl reaction. Hoefer Scientific Mini Gels were used to follow the large scale reactions to completion, as shown in fig. 3. The reaction typically takes between 30 minutes to 3 hours and denaturing PAGE was utilized to purify the products.

2.5.6 Cleavage of RNA with an immobilized biotin labeled 2'-O-methyl chimera

B2'LDR was bound to streptavidin beads (Pierce, ImmunoPure immobilized streptavidin, crosslinked, on 6% beaded agarose) using the following procedure. The buffers used are 50 mM wash buffer (20 mM Tris-HCl pH 7.6, 0.01% NP-40, 50 mM NaCl), 250 mM wash buffer (20 mM Tris-HCl pH 7.6, 0.1% NP-40, 250 mM NaCl), and preblock mix (100 µg/ml glycogen, 1 mg/ml BSA, 100 µg/ml tRNA, 33% 50 mM Wash Buffer). 2.0 mL of the 50% bead slurry solution supplied by Pierce was centrifuged to remove the storage solution and washed twice with sterile double distilled (dd) H_2O. 500 µl of preblock mix was added and mixed slowly with the beads for 20 minutes at 4 °C. The preblock mix was removed and the beads were rinsed 3 times with 500 µl of the 50 mM wash buffer. 45 nmoles of the biotinylated chimera B2'LDR (50 µl at 0.9 mM) were added to the beads with 500 µl of the 250 mM wash buffer for 90 minutes at 4° C. The supernatant was removed from the beads and washed 3 times with the 250 mM wash buffer. There was no UV signal at 260 nM for the supernatant or the washings, indicating that all 45 nmoles of B2'LDR was bound completely to the streptavidin beads.

To follow the cleavage of the rLDR3lig RNA on the B2'LDR column, the RNA was prepared 3' end labeled and 5' end labeled in two separate reactions. The 3' end label cleavage reaction (100 μl B2'LDR beads, 40 μl 5x RNase H buffer, 10 μl 20 mM DTT, 70K cpm pCp 3' end labeled rLDRr3lig and 3 μl RNase H at 1.9 U/μl) and the 5' end label cleavage reaction (100 μl B2'LDR beads, 40 μl 5x RNase H buffer, 10 μl 20 mM DTT, 70K cpm 5' end labeled rLDR3lig and 3 μl RNase H at 1.9 U/μl) were heated to 70° C for 1 minute and slow cooled before adding enzyme. Reactions ran for 3 hours at room temperature while mixing slowly to keep the beads in solution. Reactions were harvested by centrifugation and removal of the supernatant.

For the 5' end labeled reaction, greater than 95% of the counts remained on the column beads after removal of the supernatant and repeated washings, as shown in fig. 2.4. For the 3' end labeled reaction, greater than 70% of the counts came off in the supernatant and the PAGE analysis confirmed production of the correct product, r3lig.

2.5.7 Recycling the B2'LDR column

After an RNase H cleavage of an RNA with the rLDR sequence at its 5' end, the B2'LDR column may be regenerated. The rLDR sequence is bound to the column via base pairing to B2'LDR and must be removed before the column may be used again. Two or three washings of an equal volume of denaturing buffer (6M urea, 1mM Tris-HCl pH 7.6, 0.1 mM EDTA, and 20% acetonitrile) to bead material for 30 minutes at 60° C removes the rLDR. The column must then be rinsed several times with sterilized ddH$_2$O to prepare it for the next reaction. This procedure removes 95% of the counts from the 5' end labeled reaction, and the column was able to cleave another batch of RNA successfully.

2.5.8 Analysis of RNA After RNase H Cleavage

To analyze cleavage products, the RNA was 5' end labeled by sequential dephosphorylation with calf intestine phophatase and kinased with polynucleotide kinase and γ-^{32}P-ATP. The radiolabeled products were run on denaturing gels next to RNA sequencing lanes. In addition, a 3' labeled sample prepared as described above was purified on a denaturing polyacrylamide gel to separate polymerization products n and n+1, subjected to the RNase H cleavage reaction, and the product was analyzed on an RNA sequencing gel (see fig. 2.3a). To provide additional proof that the cleavage reaction proceeds site specifically (data not shown), the 3' cleavage product was ligated to another RNA at its 5' end (the site of the cleavage). The ligation reactions were carried out using a buffer of 50 mM Tris-HCl @ pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP and 50 μg/ml BSA. The two pieces of RNA to be ligated are annealed to a complementary strand of DNA which is of a different size than the RNAs or the RNA ligation product (17 nucleotides longer than the product in this case) to facilitate purification of the products. The complex formation can be followed by native PAGE. Typical annealing conditions are to heat to 90° C and slow cool to room temperature over 30 minutes time. All reactions were performed at room temperature and used 1/10 of the total reaction volume as ligase (at 400 U/μl). Yields of the ligations varied from 50 to 80 percent and are consistent with typical RNA ligation yields.

2.5.9 NMR Procedures

NMR samples were dialyzed repeatedly against 20 mM phosphate buffer at pH 6.5, 10% D_2O was added for the lock carrier signal, and the final volume of the sample was 200 μl in a Shigemi NMR tube. NMR spectrum shown (see Fig. 2.5) was collected on a General Electrics Omega 500 spectrometer using a Bruker Instruments ^{1}H, ^{13}C, ^{15}N triple resonance probe with X, Y, Z pulsed field gradient coils. The ^{1}H - ^{15}N HMQC experiment was adapted from Szewczak *et al*., 1993, utilizing GARP decoupling of the nitrogen heteronucleus (Shaka *et al*., 1985). The 0.5 mM r3lig sample required 3 hours of spectrometer time to collect 128 experiments of 64 scans. All NMR data was processed on a Silicon Graphics workstation using Biosym Technologies' Felix v2.3 NMR processing software.

2.6 References

Batey R, Inada M, Kujawinski E, Puglisi J, Williamson J. 1992. Preparation of Isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *NAR 20*:4515-23.

Berkower I, Leis J, Hurwitz J. 1973. Isolation and characterization of an endonuclease from Escherichia coli specific for ribonucleic acid in ribonucleic acid-deoxyribonucleic acid hybrid structures. *J Bio Chem 248*:5914-5921.

Crouch RJ, Dirksen ML. 1982. Cold Spring Harbor, NY: Cold Spring Laboratory.

Davanloo P, Rosenberg AH, Dunn JJ, Studier FW. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *PNAS 81*:2035-2039.

Grodberg J, Dunn JJ. 1988. omp T Encodes the E. coli Outer Membrane Protease That Cleaves T7 RNA Polymerase during Purification. *Journal of Bacteriology 170*:1245-1253.

Hayase Y, Inoue H, Ohtsuka E. 1990. Secondary Structure in Formylmethionine tRNA Influences the Site-Directed Cleavage of Ribonuclease H Using Chimeric 2'-O-Methyl Oligodeoxyribonucleotides. *Biochemistry 29*:8793-8797.

Inoue H, Hayase Y, Imura A, Iwai S, Miura K, Ohtsuka E. 1987. Synthesis and hybridization studies on two complementary nona(2'-O-methyl)ribonucleotides. *NAR 15*:6131-6148.

Inoue H, Hayase Y, Iwai S, Ohtsuka E. 1987. Sequence-dependent hydrolysis of RNA using modified oligonucleotide splints and RNase H. *FEBS Letters 215*:327-330.

Inoue H, Hayase Y, Iwai S, Ohtsuka E. 1988. Sequence-specific cleavage of RNA using chimeric DNA splints and RNase H. *Nucleic Acids Symposium Series 19*:135-138.

Kanaya S, Nakai C, Konishi A, Inoue H, Ohtsuka E, Ikehara M. 1992. A hybrid ribonuclease H. A novel RNA cleaving enzyme with sequence-specific recognition. *J Biol Chem 267*:8492-8.

Koizumi M, Hayase Y, Imura A, Iwai S, Kamiya H, Inoue H, Ohtsuka E. 1989. Design of RNA enzymes distinguishing a single base mutation in RNA. *NAR 17*:7059-7071.

Koizumi M, Hayase Y, Imura A, Iwai S, Kamiya H, Inoue H, Ohtsuka E. 1989. Design of RNA enzymes for sequence-dependent cleavage of RNA. *Nucleic Acids Symposium Series 21*:107-108.

Lapham J, Crothers DM. 1996. RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *RNA 2*:289-296.

Lapham J, Yu Y-T, Shu M-D, Steitz JA, Crothers DM. 1997. The position of site-directed cleavage of RNA using RNase H and 2'-O-methyl oligonucleotides is dependent on the enzyme source. *RNA 3*:950-951.

Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide Synthesis using T7 RNA Polymerase and Synthetic DNA Templates. *NAR 15*:8783-8798.

Milligan JF, Uhlenbeck OC. 1989. Synthesis of Small RNAs Using T7 RNA Polymerase. *Meth Enz 180*:51-62.

Nakai C, Konishi A, Komatsu Y, Inoue H, Ohtsuka E, Kanaya S. 1994. Sequence-specific cleavage of RNA by a hybrid ribonuclease H. *FEBS Letters 339*:67-72.

Nikonowicz E, Sirr A, Legault P, Jucker F, Baer L, Pardi A. 1992. Preparation of ^{13}C and ^{15}N labelled RNAs for heteronuclear multi-dimentional NMR studies. *NAR 20*:4507-13.

Shaka A, Barker P, Freeman R. 1985. Computer-optimized decoupling scheme for wideband application and low-level operation. *Journal of Magnetic Resonance 64*:547-552.

Shibahara S, Mukai S, Nishimura T, Inoue H, Ohtsuka E, Morisawa H. 1987. Site-directed cleavage of RNA. *NAR 15*:4403-4415.

Szewczak AA, Kellogg GW, Moore PB. 1993. Assignment of NH Resonances in Nucleic Acids Using Natural Abundance ^{15}N-^{1}H Correlation Spectroscopy with Spin-Echo and Gradient Pulses. *FEBS Lett 327*:261-264.

Uemura H, Imai M, Ohtsuka E, Ikehara M, Söll D. 1982. E. coli initiator tRNA analogs with different nucleotides in the discriminator base position. *NAR 10*:6531-6539.

Wyatt J, Chastain M, Puglisi J. 1991. Synthesis and purification of large amounts of RNA oligonucleotides. *Biotechniques 11*:764-9.

Xu J, Lapham J, Crothers DM. 1996. Determining RNA Solution Structure by Segmental Isotopic Labeling and NMR:Applications to *Caenorhabditis elegans* Spliced Leader RNA. *PNAS 93*:44-48.

Yu Y-t, Shu M-d, Steitz JA. 1997. A new method for detecting sites of 2'-O-methylation in RNA molecules. *RNA 3*:324-331.

Yu Y-T, Steitz JA. 1997. A new strategy for introducing photoactivable 4-thiouridine (^{4S}U) into specific positions in a long RNA molecule. *RNA 3*:807-810.

Zawadzki V, Gross HJ. 1991. Rapid and simple purification of T7 RNA polymerase. *NAR 19*:1948.

CHAPTER 3 "APPLICATION OF ISOTOPE FILTERED NMR EXPERIMENTS FOR NUCLEIC ACIDS"

3.1 Summary

In this chapter the NMR spectral editing technique of isotope filtering is used to examine nucleic acids that have been partially isotope labeled. The application of an isotope labeled NOESY experiment on a duplex DNA that has been labeled on one strand is demonstrated and is shown to be an effective method of making assignments.

A new pulse sequence is presented that incorporates an isotope-filter with a pulse field-gradient stimulated echo sequence. This new experiment makes it possible to follow the translational self-diffusion of an isotope-labeled species in solution, independent of other solutes. An example is presented in which the diffusion constant of an isotope-labeled DNA is followed before and after binding a protein.

3.2 Introduction

The concept of the isotope filter in NMR is simple. The one-bond J-coupling between a proton and another magnetically active "X" nucleus (^{13}C or ^{15}N for example) is exploited to control the phase of the observable magnetization of the proton. Using some simple phase cycling methods, data can be collected in which only signal arising from a proton covalently attached to this "X" nucleus is observed. If a NMR sample has been synthesized in which only part of the sample is isotopically labeled, it is possible to use isotope filtering to selectively view the signal arising from either the labeled or the unlabeled portion. Some advantages of this technique include spectral simplification and reduction in assignment ambiguity.

3.2.1 Isotope selection by NMR

Otting and Wüthrich (1990) have reviewed the theoretical and practical applications of isotope-filtered techniques for NMR. The utility of these experiments has been shown for a variety of biologically interesting problems, such as in obtaining strand resolved spectra for duplex RNA (SantaLucia, *et al.*, 1995; Cai & Tinoco, 1996), characterization of symmetric protein dimers (Weiss, 1990; Arrowsmith, *et al.*,1990; Folkers, *et al.*, 1993; Burgering, *et al.*, 1993), a protein-DNA complex (Otting, *et al.*, 1990), protein-ligand binding (Fesik, 1988; Fesik *et al.*, 1988), spectral simplification by specific amino acid labeling (Fesik *et al.*, 1987; Torchia *et al.*, 1989) and determination of RNA dimerization (Aboul-ela & Pardi, 1996; Flemming, *et al.*, 1996) among others.

A quick overview of the isotope filtering technique is presented. The pulse sequence elements fundamental to isotope selection are shown below in figure 3.1.

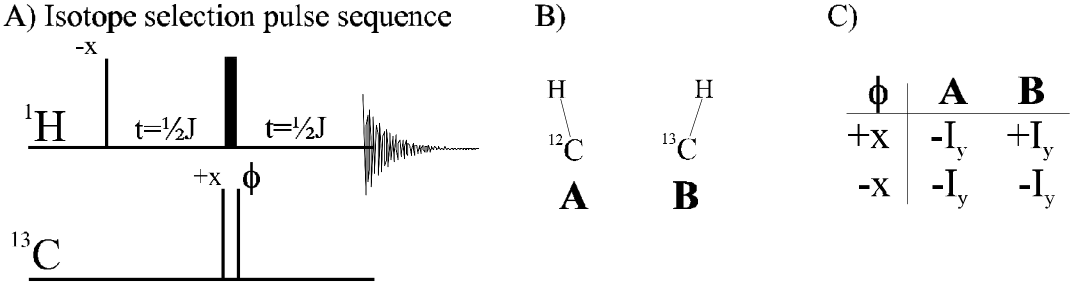


Figure 3. 1 Isotope selection schematic

To illustrate what happens in the isotope selection experiment, follow the magnetization of two protons, as shown in figure 3.1B. The first proton, **A** is attached to a ^{12}C atom, while the second, **B** is attached to a ^{13}C atom. Using the product operator formalism (Sorensen, *et al.*, 1983; Harris, 1985; Howarth, *et al.*, 1986; Shiver, 1992) to follow the evolution of the pulse sequence. (Chemical shift is included only for completeness, it clearly will not affect the final proton magnetization since this is a spin-echo pulse sequence). One observes that,

$$I_z \xrightarrow{-90_x(I)} I_y$$

$$\xrightarrow{t=\frac{1}{2J}} I_y \cos\left({}^{w}\!/_{2J}\right) - I_x \sin\left({}^{w}\!/_{2J}\right)$$

$$\xrightarrow{180_x(I)} -I_y \cos\left({}^{w}\!/_{2J}\right) - I_x \sin\left({}^{w}\!/_{2J}\right)$$

$$\xrightarrow{t=\frac{1}{2J}} -\left[I_y \cos\left({}^{w}\!/_{2J}\right) - I_x \sin\left({}^{w}\!/_{2J}\right)\right]\cos\left({}^{w}\!/_{2J}\right) - \left[I_x \cos\left({}^{w}\!/_{2J}\right) + I_y \sin\left({}^{w}\!/_{2J}\right)\right]\sin\left({}^{w}\!/_{2J}\right)$$

$$= -I_y \cos^2\left({}^{w}\!/_{2J}\right) - I_y \sin^2\left({}^{w}\!/_{2J}\right) = I_y\left[\cos^2(wt) + \sin^2(wt)\right] = -I_y . \qquad 5.1$$

For proton **A** there is no 1-bond J coupling and the pulse sequence acts like a spin-echo. Notice that the chemical shift precession terms will always refocus in this type of pulse sequence.

Atom **B** is covalently attached to a ^{13}C isotope and one-bond J-coupling between the carbon and proton is present. If the phase (ϕ) of the second ^{13}C pulse is set to –x, it “cancels out” the effect of the first ^{13}C pulse with phase +x. Thus, with the phase of ϕ set to –x, both the chemical shift and ^{13}C-^{1}H J-coupling will be refocused by the spin-echo leaving the magnetization state of -I_y for proton **B**, giving **B** the same phase as proton **A**.

However, if the phase of ϕ is set to +x, it works in conjunction with the first ^{13}C π/2 degree pulse to create an “effective” π pulse. With ϕ set to +x, the final magnetization of **B** will be $+I_y$, as shown below (the effects of chemical shift precession have been removed for the sake of brevity),

$$I_z \xrightarrow{-90x(I)} I_y$$

$$\xrightarrow{t=1/2\mathrm{J}} I_y \cos(\mathrm{p}Jt) - 2I_xS_z \sin(\mathrm{p}Jt) = -2I_xS_z$$

$$\xrightarrow{180x(I)} -2I_xS_z \xrightarrow{180x(S)} 2I_xS_y$$

$$\xrightarrow{t=1/2J} 2I_xS_z \cos(\mathrm{p}Jt) + I_y \sin(\mathrm{p}Jt) = +I_y \qquad 5.2$$

The final values obtained are summarized in the table found in figure 5.1C.

Thus, the magnetization of the proton attached to the "X" labeled nucleus can be set to either $+I_y$ or $–I_y$ through the use of the phase ϕ. This can be exploited in an NMR experiment by collecting two sets of data, one in which the phase ϕ is set to +x and one in which the phase ϕ is set to –x. The simulated spectra for the **A** and **B** is shown below in figure 3.2. The ^{13}C and ^{12}C subspectra (figure 3.2 C and D) can then be constructed by respectively subtracting or adding the two original spectra.

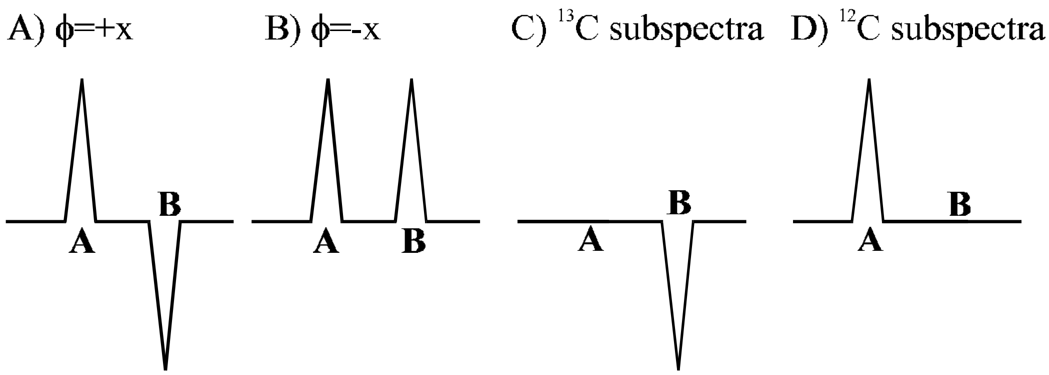


Figure 3. 2 Isotope filtered subspectra

This isotope filter pulse sequence element can be incorporated into some traditional proton NMR experiments.

3.2.2 Isotope filtered NOESY

The "nuclear Overhauser effect spectroscopy" (NOESY) experiment is of fundamental importance in elucidating molecular structure and dynamics information by NMR. The NOESY spectrum contains information on the dipolar relaxation processes occurring in the molecule, and this data can be utilized to calculate proton-proton distances (see Chapter 7). One of the major limitations of the NOESY experiment is finding well resolved cross peaks suitable for volume quantitation. The larger and more homogeneous the molecular structure, the greater this problem can be. For large DNA molecules this can be a formidable obstacle, but the use of isotope selection or filtering experiments can simplify the task. Figure 3.3 demonstrates how the concept of the "isotope subspectrum" presented before can be extended to a two dimensional experiment.

The application of ^{15}N and ^{13}C isotope-filtered NOESY NMR experiments was used for assignment of proton resonances for a DNA molecule in which one strand is uniformly isotope labeled. This procedure utilizes standard isotope-filtered NOESY techniques to assign the exchangeable imino proton spectra and to obtain strand-resolved spectra of the non-exchangeable protons for both the labeled and unlabeled halves of the DNA. Since these experiments can be performed on a single sample, they expedite the process of assigning resonances in large DNA molecules. A comparison between NOESY spectra of an unlabeled sample of the same sequence to those obtained using the filter NOESY experiments on the labeled counterpart will be presented and demonstrates the spectral simplification obtained by this technique.

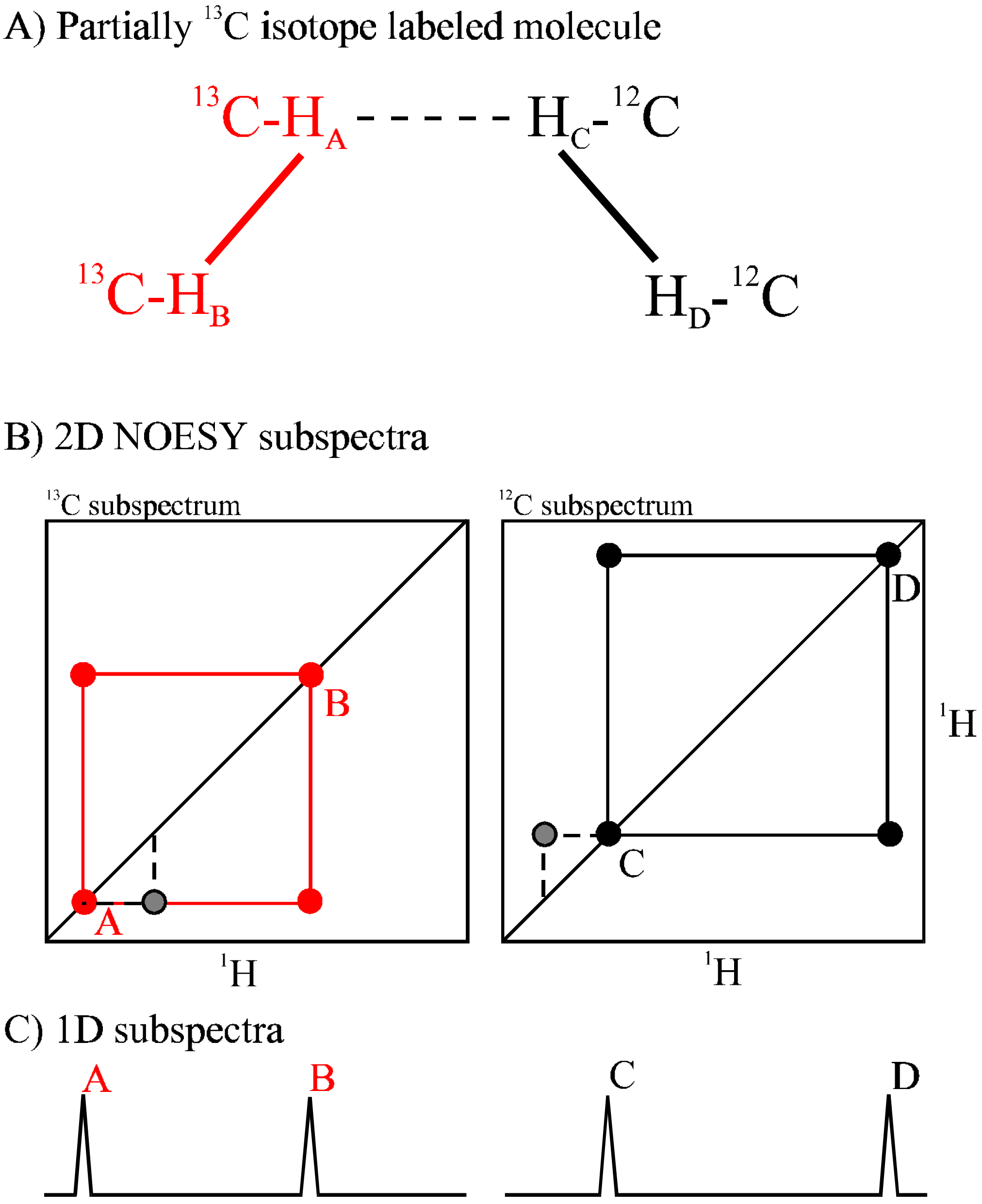


Figure 3. 3 Simulated NOESY subspectra for a partially labeled molecule

A) A partially $^{13}C/^{12}C$ molecule with proton A and B attached to a ^{13}C and proton C and D attached to a ^{12}C. The spacial arrangement of the protons is such that proton A is within an NOE distance from B and C; proton C is within an NOE distance form A and D. The solid and dashed lines represents the NOE connectivities. **B**) The simulated ^{13}C and ^{12}C subspectra from the 2D isotope filtered NOESY experiment. The dashed line represents the connectivity between proton A and C, notice that the crosspeak between A and C is found in BOTH spectra, because A is ^{13}C labeled and C is ^{12}C labeled. **C**) The 1D subspectra for the sample.

3.2.3 Isotope filtered pulsed field-gradient stimulated echo

Determining the translational diffusion rate of a molecule can give important information on the hydrodynamical shape of that molecule and can be used to estimate its approximate molecular size. One of the NMR experiments used for determining the translational diffusion constant of a molecule is known as the "pulsed field-gradient stimulated echo" (PFG-STE) and has been shown to accurately measure the diffusion constants of nucleic acids (Lapham, *et al*., 1997; Chapter 4). This experiment can be modified to include an isotope filter, allowing for the discrimination between the diffusion rate of a labeled and an unlabeled molecule.

The importance of having the ability to observe the translational diffusion constant of a single species in a complex solution is that it avoids the problems that may arise in interpreting diffusion data for complexes which may not be in a 1:1 molar ratio. For instance, if a DNA-protein complex were constructed in a 1:1.2 ratio (an excess of protein), the measured diffusion rate of the complex would be some average of the diffusion rate of the full complex and the 20% free protein. This would, naturally, give rise to an erroneous diffusion constant.

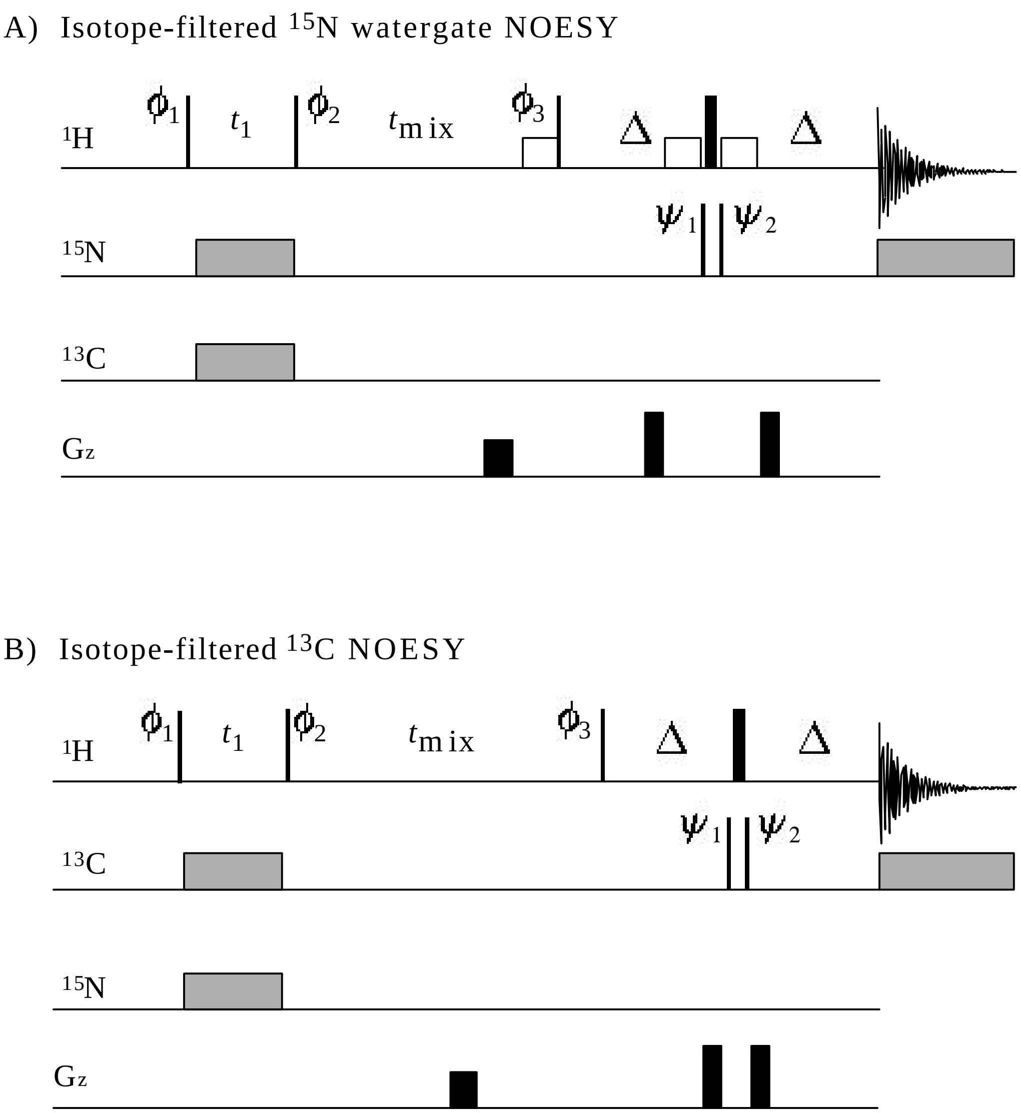
To address this problem, we created a ^{13}C isotope filtered pulsed field-gradient stimulated echo pulse sequence (^{13}C filtered-PFG-STE). It can be used for monitoring protein-DNA binding by NMR, by measuring the diffusion constant of the isotope-labeled strand of the DNA. The experiment is capable of monitoring the diffusion constant of a single component in a complex mixture, and is the only known method for accomplishing this.

3.3 Results

3.3.1 Exchangeable protons

The NOESY NMR spectroscopy of the imino protons of nucleic acids is of critical importance in assigning the secondary structure of a DNA or RNA molecule (Wüthrich, K., 1986). While it is possible to label one strand of a DNA with ^{15}N and perform an ^{15}N-^{1}H HMQC to identify the iminos from the labeled strand, chemical shift degeneracy, common in standard B-form DNA, may make it impossible to resolve every imino proton. This particular problem can be alleviated by observing the crosspeak patterns between adjacent iminos in the isotope filtered NOESY experiment. The crosspeaks of the imino protons from a NOESY spectrum offer a second dimension to resolve such degeneracy. Using these isotope-filtered NOESY techniques on a single strand labeled heteroduplex DNA allows one to assign an orientation to the imino protons based on the pattern of the crosspeaks found in the two subspectra.

The data for the exchangeable proton spectra were collected using a watergate NOESY pulse sequence for the fully unlabeled DNA and an isotope-filtered watergate NOESY for the single strand isotope labeled DNA (Fig. 3.4a). Comparison of the exchangeable imino proton spectrum of the unlabeled D19 and the isotope-filtered NOESY of the single strand labeled D19 is shown in figure 3.5. All crosspeaks found in the unlabeled spectrum (Fig. 3.5b) are clearly visible in either the ^{14}N or ^{15}N subspectrum from the isotope-filtered NOESY (Figs. 3.5c/d). Interpretation of the data from the two subspectra is quite straightforward. If a crosspeak appears on both sides of the diagonal in the ^{14}N subspectra, then the two imino protons which gave rise to the crosspeak belong

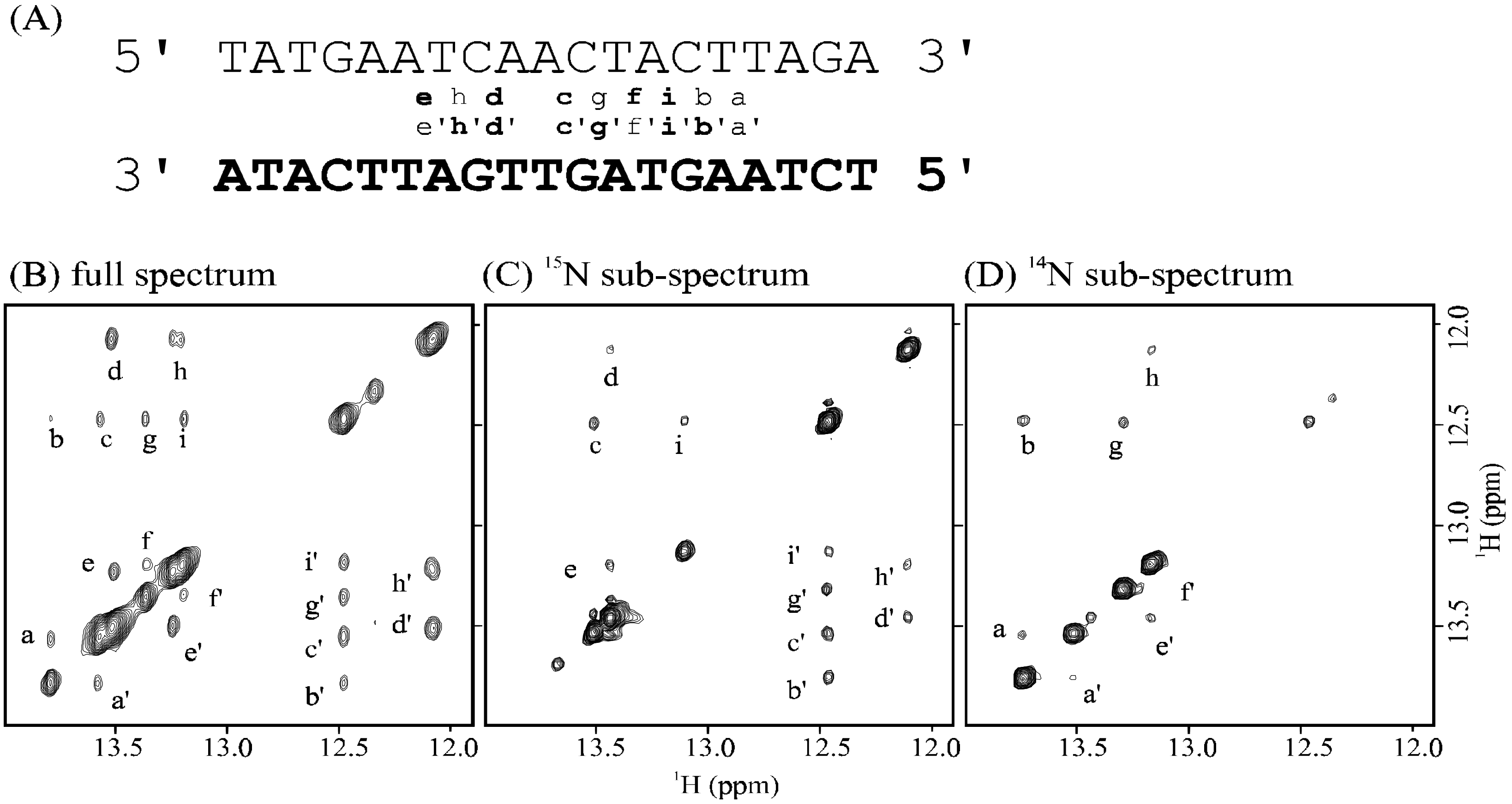


A) Isotope-filtered ^{15}N watergate NOESY

B) Isotope-filtered ^{13}C NOESY

Figure 3. 4 Isotope filtered NOESY pulse sequences

For both the ^{15}N and ^{13}C isotope-filtered experiments, all pulses indicated by the thin lines are $\pi/2$ pulses and the wide lines are π pulses. All hard pulses are phase cycled +x unless otherwise indicated, all the soft pulses are phase cycled -x. Φ_1 is cycled (x, -x) and also includes the States phase cycling for quadature detection (States, et al., 1982). Two experiments are collected for each States cycle, in which the phase of Ψ_1 and Ψ_2 is (+x) for the first experiment, the second experiment is collected with Ψ_1 set to (-x) and Ψ_2 set to (+x). Garp decoupling (Shaka, et al., 1985) was used for both the nitrogen and carbon channels during the t_1 time and acquisition, if indicated. **A)** The isotope filtered pulse sequence used for the exchangeable proton NOESY experiment. **B)** The isotope filtered pulse sequence used for the non-exchangeable proton NOESY experiment.



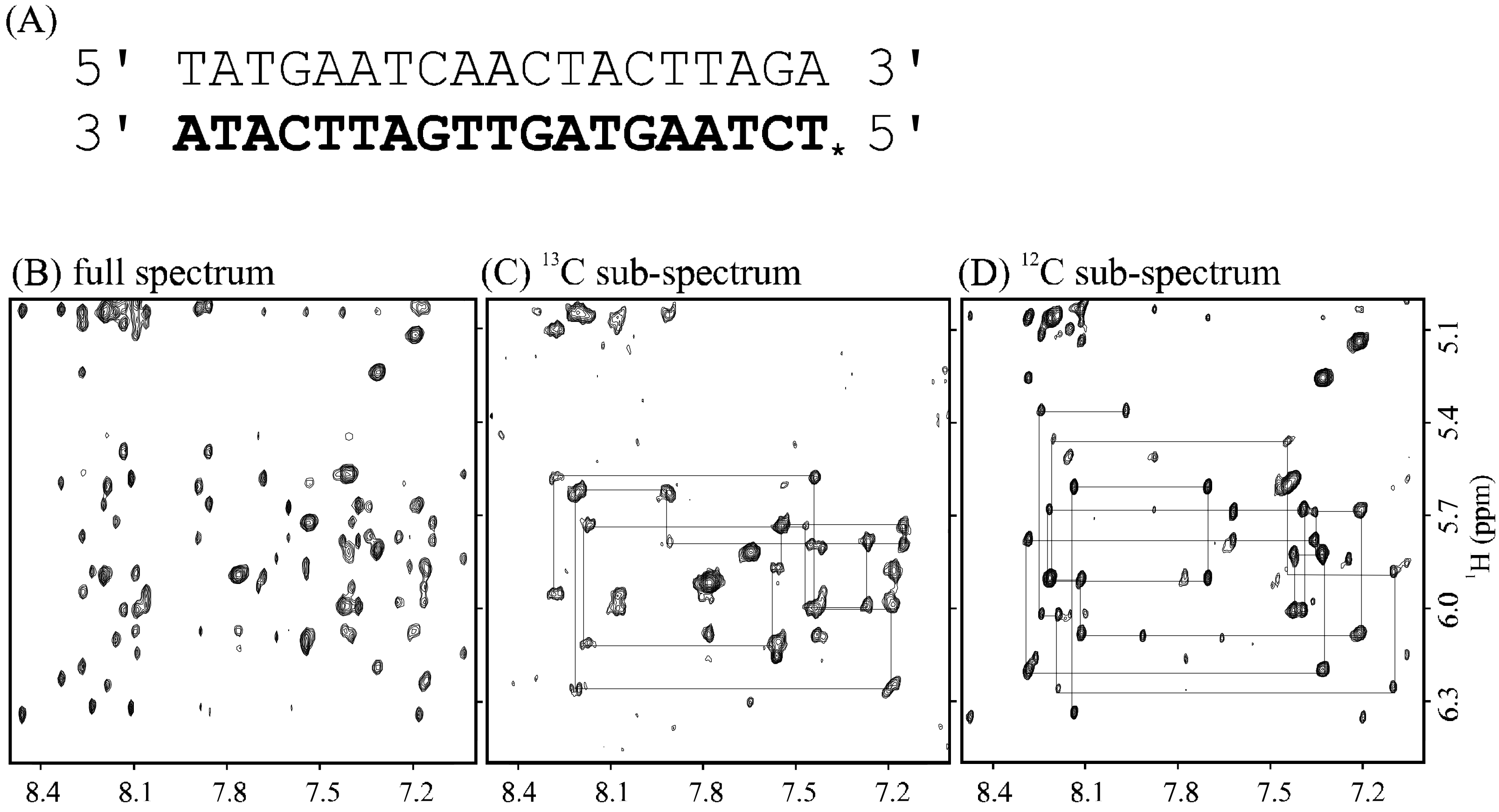
5 ^{14}N/^{15}N isotope filtered watergate NOESY spectra for DNA

The 19 base pair DNA, D19, used in these experiments. The bottom strand, in bold, is the ^{15}N/^{13}C labeled strand. **B)** 2D $_2$O NOESY spectra of the unlabeled D19. The labeling of the crosspeaks begins at the bottom left with “*a*” and *i*”, while each symmetry related crosspeak has the same label with a prime. **C)** The ^{15}N **D)** ^{14}N filtered sub-spectrum of the single strand labeled D19, which were processed using the to the unlabeled strand of DNA. Conversely, if a crosspeak appears on both sides of the diagonal in the ^{15}N subspectra, the two imino protons are located on the labeled strand. Finally, if a crosspeak appears on one side of the diagonal in the ^{14}N subspectra and on the other side of the diagonal in the ^{15}N subspectra, then the two iminos contributing to the crosspeak are on separate strands of the DNA. In this manner, every observable imino proton crosspeak for D19 was assigned, as shown in figure 3.5a.

3.3.2 Non-exchangeable protons

The non-exchangeable protons in DNA are of critical importance in structure determination. For B-form DNA, the sequence specific assignment of these protons can be accomplished by means of the anomeric-aromatic walk found in the 2D NOESY. This connectivity pattern correlates the H6/H8 base proton of a nucleotide to its own H1' sugar proton, and to the H1' sugar proton of the nucleotide in the 5' direction. In a well resolved spectrum, every H6/H8 and H1' can be sequence-specifically assigned in this manner.

The isotope filtered NOESY pulse sequence (figure 3.4b) was used to collect data on the sample with one strand labeled, and the data was compared with that from an unlabeled DNA. The ^{13}C and ^{12}C isotope-filtered subspectra for the single strand labeled sample are shown in figure 3.6c/d respectively. The drawn line in the spectra represents the sequential nucleotide connectivities; the isotope filter clearly separates the two distinct aromatic-anomeric walks. Figure 3.6b demonstrates what the standard 2D NOESY for the fully unlabeled DNA sample looks like.



6 $^{12}C/^{13}C$ isotope filtered NOESY spectra for DNA

The 19 base pair DNA, D19, used in these experiments. The bottom strand, in bold, is the $^{15}N/^{13}C$ labeled strand. **B)** 2D $_2$O NOESY spectrum of the unlabeled D19. **C)** ^{13}C filtered sub-spectrum and the **D)** ^{12}C selected sub-spectrum of the

3.3.3 PFG diffusion measurements

NMR isotope-filtering techniques offers a unique ability to observe a single molecular species in a complex solution. This is an especially powerful tool for the spectroscopist interested in monitoring the physical behavior of a molecule under the influence of another. We demonstrate this by measuring the translational self-diffusion rate of the isotope-labeled strand of D19 both bound and unbound to a protein. The data is simple to interpret in that the resonances of the unlabeled protein do not complicate the spectrum.

The NMR pulsed field-gradient (PFG) spin-echo technique (Hahn, 1950; Stekjskal & Tanner, 1965) has long been used to measure diffusion constants. Applications to biological systems include determination of the aggregation state of proteins (Alteiri, *et al*., 1995; Dingley, *et al*., 1995), measurement of the bulk movement of hemoglobin in human erythrocytes (Kuchel & Chapman, 1991) and quantitation of processes such as amide proton exchange with water (Andrec & Prestegard, 1996). For the NMR spectroscopist, it provides a simple, accurate method for measuring the diffusion constants of the materials they are investigating under the same conditions as all their other NMR experiments. Chapter 4 of this thesis gives a more exhaustive theoretical and experimental discussion of translational self-diffusion.

We present here a new pulse sequence for measuring the diffusion rate of a single isotope-labeled molecule in a complex solution, an isotope-filtered PFG stimulated echo (filtered-PFG-STE, Fig. 3.7). This pulse sequence was adapted from Tanner's (1970) PFG-STE sequence that maximizes the signal of samples with short T_2 relaxation times.

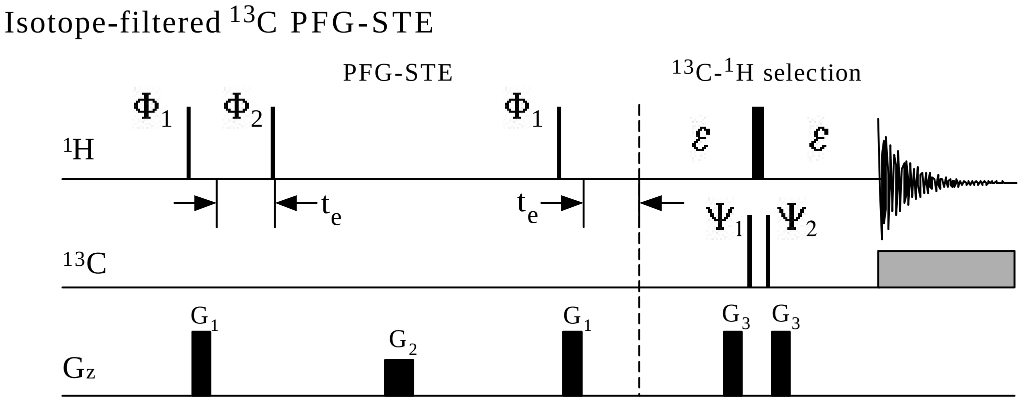


Figure 3. 7 The isotope filtered PFG-STE pulse sequence

Φ_1 was cycled (+x, -x) and Φ_2 was cycled (-x, +x). Two fids were collected for each increment of G_z: the first with the phase Ψ_1 of (+x) and Ψ_2 of (-x), the second with the phase of Ψ_1 of (-x) and Ψ_2 of (-x). ε was set to $1/2J_{H\text{-}C}$ for the methyls in the DNA (1/2•140 hz) and the carbon carrier was centered at 12 ppm for the methyls. The isotope filtered spectra was generated by linear addition of these two fids.

A comparison of the measured self-diffusion rate of the fully unlabeled DNA using the PFG-STE sequence and the single strand labeled DNA using the filtered-PFG-STE sequence (Fig. 3.8) demonstrates that they both give approximately the same values, 1.10(.01)x10^{-6} and 1.12(.01)x10^{-6} cm^2/s respectively. IHF was added to the D19 sample in a 1:1 molar ratio, and the diffusion constant was measured from the ^{13}C PFG-STE subspectra and was found to be 0.76(.034)x10^{-6} cm^2/s for the isotope labeled strand (Fig. 3.8).

This experiment shows protein binding to the DNA by the change in the translational self-diffusion constant of the isotope labeled strand of the DNA. It does not require assignment of any resonances, and the data is not complicated by the additional protein resonances.

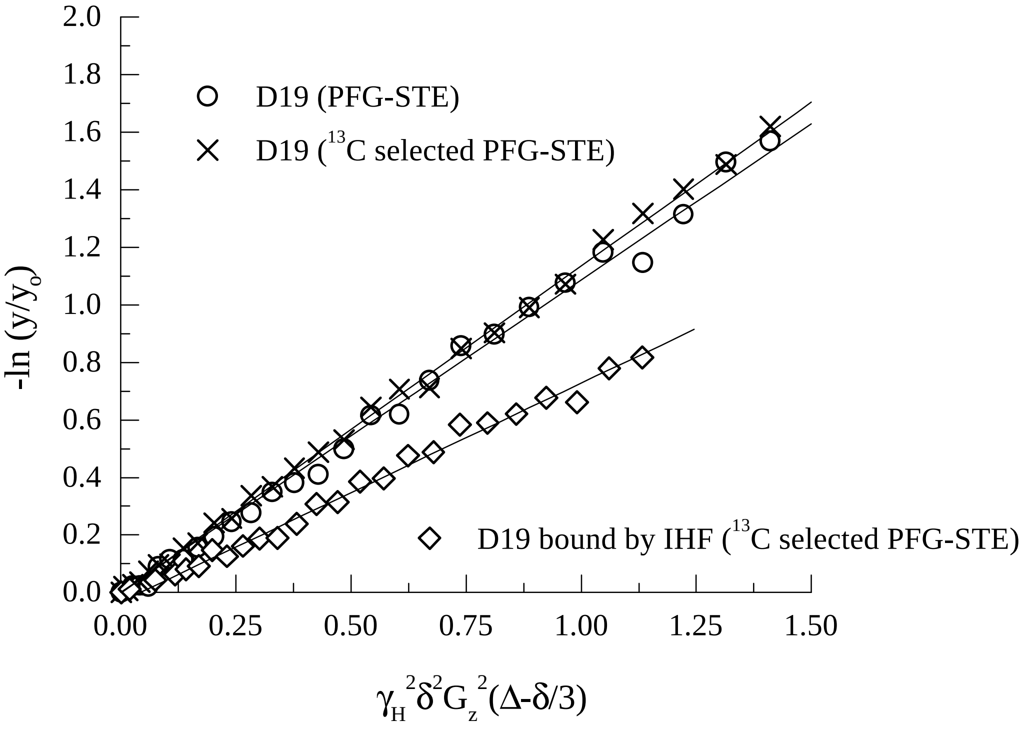


Figure 3. 8 Protein binding DNA as measured by isotope filtered diffusion

Translational diffusion rate for D19 upon binding by IHF. D19 unbound data (circles) was collected using the standard PFG-STE (see Chap. 4) and the isotope-filtered PFG-STE (crosses) as presented in this chapter. The D19 bound by IHF data (diamonds) was collected using the isotope-filtered PFG-STE pulse sequence.

Within the error of the experiment the results from the unbound DNA demonstrate that the isotope-filtered PFG-STE pulse sequence measures the same diffusion rate, as does the standard PFG-STE. The data from the bound D19 demonstrate that the binding of the IHF protein decreases the translational diffusion rate of the isotope-labeled strand of the DNA. This is what would be expected, the protein-DNA complex should have a larger frictional coefficient than the DNA alone.

3.4 Discussion

Traditionally, structural studies of DNA molecules by NMR have been accomplished by means of homonuclear 1D and 2D proton correlation experiments. Our lab has recently published techniques for synthesis of uniformly ^{13}C and ^{15}N isotope labeled DNA molecules (Zimmer & Crothers, 1995) which allows for single strand labeling of any DNA sequence which is not dyad symmetric. Given that most DNAs of biological interest are non-dyad symmetric dimers, we feel that synthesis of these single strand labeled samples, in conjunction with standard isotope filtered NMR experiments, will greatly facilitate the study of larger DNAs. We present this as a general method for obtaining proton assignments for large DNA molecules, while requiring that only one sample be synthesized.

The method of data collection and processing of the isotope filtered NOESY experiments presented allows for obtaining both the labeled and the unlabeled subspectrum at the same time. The same data set is either added together or subtracted to form one subspectrum or the other. In this manner, the experiments are more efficient than the ½-X-filtered type experiments (Otting and Wüthrich, 1990) that require complete data sets be collected for each subspectrum.

In addition to proton assignment, and ultimately structural determination, NMR can be used to measure other physical properties of systems. Recently, pulsed field-gradient (PFG) methods have been employed to measure the translational diffusion constants for nucleic acids (Lapham *et al*., 1997). Use of an isotope-filter in conjunction with these PFG diffusion measurements makes it possible to follow the translational self-diffusion of a single molecular species in a complex solution. This can be used to monitor any process that will change the hydrodynamic properties of the isotope labeled species, such as protein binding. The advantage of being able to filter away the signals due to the unlabeled DNA strand and other ligands (such as a protein in this case) is that it simplifies the interpretation of the data.

3.5 Materials and methods

3.5.1 DNA sample preparation

The unlabeled DNA strands were synthesized on an Applied Biosystems 380B DNA synthesizer. The ^{15}N and ^{13}C uniformly labeled DNA strands were synthesized enzymatically as previously described (Zimmer & Crothers, 1995). Two samples of the 19 base pair D19 sample were produced, one which was composed of two unlabeled strands and one which was composed of an unlabeled top strand and an isotope labeled bottom strand. The top strand sequence for D19 is 5'-TATGAATCAACTACTTAGA-3' and the complementary bottom strand sequence is 5'-TCTAAGTAGTTGATTCATA-3'.

200 nmoles of each DNA strand was combined in a 1:1 molar ratio, concentrated to 160 uL volume, and dialyzed several times against 10 mM sodium phosphate buffer at pH 6.8, 100 mM NaCl and 0.5 mM EDTA. The sample was placed in a Shigemi NMR tube (Shigemi corp., Tokyo, Japan) with a total volume of 160 μl, with a final duplex concentration of 1.25 mM.

The exchangeable data were collected on a 85% H_2O and 15% D_2O sample, while the non-exchangeable data was collected on a 100% D_2O sample. Prior to the NMR experiments the samples were heated to 90° C then allowed to cool slowly to room temperature to insure complete duplex formation.

3.5.2 Protein sample preparation

Aliquots of an IHF protein stock at 2.3 mM were added to the D19 strand labeled duplex at 1.25 mM and dialyzed against a buffer containing D_2O, 100 mM NaCl, 10 mM sodium phosphate (pD 6.8) and 0.5 mM EDTA. Complete protein binding was monitored by native condition gel electrophoresis band shift assay.

3.5.3 NMR spectroscopy: filtered NOESY

Standard isotope-filtered pulse sequences were employed. For the exchangeable proton data, an ^{15}N isotope-filtered NOESY with the watergate (Piotto, *et al.*, 1992; Lippens, *et al.*, 1995; Sich, *et al.*, 1996) water suppression technique was utilized (figure 3.4a). The proton carrier was set to the water resonance and the ^{15}N carrier was set to 150 ppm, centered between the N1 of guanidines and N3 of the thymidines. For the non-exchangeable proton data, a ^{13}C isotope-filtered NOESY was used (figure 3.4b). The ^{13}C carrier frequency was set to 190 ppm, centered between the C1’ and C6/C8 resonances.

All data were acquired on a Varian Unity 500 MHz NMR spectrometer at 30° C. Both the exchangeable H_2O NOESY and non-exchangeable D_2O NOESY experiments were obtained by collecting 2048 complex t_2 points in 32 scans with 300 t_1 time increments with a total experiment time of 24 hrs for each of the data sets. Two FIDS were collected for each states cycle, and were either added together to produce the ^{14}N (or ^{12}C) sub-spectra, or subtracted from each other to produce the ^{15}N (or ^{13}C) sub-spectra as described elsewhere (Otting & Wüthrich, 1990; SantaLucia, *et al.*, 1995). All data shown were apodized using a 90 degree shifted sine bell function. The data were processed on a Silicon Graphics computer using the Felix95 NMR processing program (Biosym Technologies, San Diego, CA).

3.5.4 NMR spectroscopy: filtered PFG-STE

The translation diffusion constant was measured using an isotope filtered PFG-STE pulse sequence (figure 3.7). 32 experiments were collected in which the strength of the gradients G1 were incremented from 1 to 32 gauss/cm. The data were processed and interpreted as previously described (Lapham, *et al*., 1997). The carbon filter was added to the end of the pulse sequence to allow of the observation of only those resonances on the ^{13}C labeled strand of D19. The carbon carrier was set to 12 ppm to center on the DNA methyls, which gave strong signal in the DNA-protein complex.

3.6 Appendix

The following pages contain the processing pulse sequences and felix95 macros used to process the data shown in this chapter.

3.6.1 Isotope filtered jump-return spin-echo 1D pulse sequence

```
#ifndef LINT
static char SCCSid[] = "@(#)GE_hmqc_jrse.c";
#endif

/*     GE_14n_15n variables:

    mix = mixing time.       (50-300ms)
    deltav = imino_v - h2o_v     (3625 hz)
    tau = (1/(4*deltav))         (~69 us)
    tau_corr = tau-(pw+pw2)-rof1     (~45 us)
    tau_corr2=tau*2-(pw+pw2)-rof1    (~114 us)
    post = gradient settling time        (50-200us)
    d1 = relaxation delay        (0.1 - 1.0 s)
    pw = 1H 90            (6 - 8 us)
    grt = gradient time          (1 ms)
    grl = gradient level         (8000)

    phase = 1,2 for States-TPPI

    -J. Lapham 7/25/95 */

#include <standard.h>

/* Define static integers arrays used to create the AP tables */

static int ph1[1] = {0},
    ph2[1] = {2},
    ph3[1] = {0},
    ph4[1] = {2};

pulsesequence()
{

/* Declare Variables */

    /* char charvar; */

    int phase;

    double  post, tau_corr,    tau_corr2,
        djxh2, pw2, jxh,
        grt, grl;

/* Load Variables */
    ni = getval("ni");
    phase = (int) (getval("phase") + 0.5);
    post = getval("post");
    grt = getval("grt");
    grl = getval("grl");
    tau = getval("tau");
    jxh = getval("jxh");
    pw2 = getval ("pw2");
    pwx2 = getval("pwx2");
    pwxlvl2 = getval("pwxlvl2");
    dpwr2 = getval("dpwr2");

/* Initialize variables */
    djxh2 = (1.0 / (2.0 * jxh)) - grt - post;
    tau_corr = tau - (pw + pw2) - rof1;
    tau_corr2 = tau*2 - (pw + pw2) - rof1;

/* check validity of parameter range */

    if((dm[A] == 'y' || dm[B] == 'y' || dm[C] == 'y' || dm[D] == 'y'))
        {
        printf("Decoupler must be set as dm=nnnny or n\n");
        abort(1);
    }

    if((dm2[A] == 'y' || dm2[B] == 'y' || dm2[C] == 'y' || dm2[D] == 'y'))
        {
        printf("Second decoupler must be set as dm2=nnnny or n\n");
        abort(1);
    }

    if( dpwr > 50 )
    {
        printf("dpwr too large (must be less than 51)!\n");
        abort(1);
    }

    if( dpwr2 > 50 )
    {
        printf("dpwr2 too large (must be less than 51)!\n");
        abort(1);
    }

/* Define phase cycling tables */
    settable(t1, 1, ph1);     /* t1 = 0,... */
    settable(t2, 1, ph2);     /* t2 = 2,... */
    settable(t3, 1, ph3);     /* t3 = 0,... */
    settable(t4, 1, ph4);     /* t4 = 2,... */

if (phase == 1)
    {
    assign(zero, v1);
    assign(zero, oph);
    }

if (phase == 2)
    {
    assign(two, oph);
    assign(two, v1);
    }

if (phase == 3)        /* 15N spectrum */
    {
    mod2(ct, v1);  /* v1 = 0,1,... */
    dbl(v1, v1);   /* v1 = 0,2,... */

    mod2(ct, oph);  /* oph = 0,1,... */
    dbl(oph, oph);  /* oph = 0,2, ... */
    }

if (phase == 4)        /* 14N spectrum */
    {
    mod2(ct, v1);  /* v1 = 0,1,... */
    dbl(v1, v1);   /* v1 = 0,2,... */

    assign(two, oph);
    }


/* BEGIN THE ACTUAL PULSE SEQUENCE */
```

```
 status(A);

    rcvroff();
    rlpower(pwxlvl2,DO2DEV);  /* Set decoupler power to pwxlvl */
    rlpower(tpwr,TODEV);      /* Set power for hard pulses  */
    delay(d1);

status(B);
    rgpulse(pw, t1, rof1, 0.0);   /* 90x */
    delay(tau_corr);           /* tau_corr delay */
    rgpulse(pw2, t2, rof1, 0.0);  /* 90-x */
    delay(djxh2);

status(C);
    rgradient('z', grl);      /* apply gradient */
    delay(grt);
    rgradient('z', 0.0);
    delay(post);
    dec2rgpulse(pwx2, v1, rof1, 0.0); /* first dec channel*/
    rgpulse(pw, t3, rof1, 0.0);   /* 90x */
    delay(tau_corr2);     /* tau_corr2 delay */
    rgpulse(pw2, t4, rof1, 0.0);  /* 90-x */

status(D);
    dec2rgpulse(pwx2, zero, rof1, 0.0);   /* first dec channel*/
    rgradient('z', grl);      /* Refocus resonances, remove */
    delay(grt);       /* residual water */
    rgradient('z', 0.0);
    delay(post);
    delay(djxh2);

status(E); /* acquire data */
    rlpower(dpwr2,DO2DEV); /* Set decoupler power to dpwr2 */
}
```

3.6.2 Isotope filtered watergate NOESY 2D pulse sequence

The variable "phase" must be set to 1,2,3,4 (a four step array of 1,2,3,4). Four separate FIDs will be collected for each t1 time increment. The linear combination of FID #1 and #2 will give the ^{14}N subspectrum, while the linear subtraction of the same FIDs will give the ^{15}N subspectrum.

```
/* n_sel_w_noesy.c

    Pulse sequences adapted from the
watergate NOESY pulse sequence
    coded by John Diener.

    Last edited 2/12/97  -JPL
*/

#include <standard.h>

/* Define Phase Tables */

   static int phi1[8] = {0,1,2,3,2,3,0,1},
phi2[8] = {0,1,0,1,2,3,2,3},
    phi3[8] = {2,3,2,3,0,1,0,1},
    rec4[8] = {0,1,2,3,2,3,0,1},
    phi5[2] = {0,2},
    phi6[2] = {0,2},
    phi7[2] = {0,2};

pulsesequence()
{
/* DECLARE VARIABLES */

 double   mix,modmix,tau,modtau,grt,gzlvl1,
sl901,sl902, sl90dif,tpwrsl,stweak,
pshift,djxh1,djxh2,jxh;

 int  phase;

/* LOAD VARIABLES */

  mix = getval("mix");
  gzlvl1 = getval("gzlvl1");
  grt = getval("grt");
  sl901 = getval("sl901");
  sl902 = getval("sl902");
  tau = getval("tau");
  tpwrsl = getval("tpwrsl");
  stweak = getval("stweak");
  phase = (int) (getval("phase") + 0.5);
  pwxlvl = getval("pwxlvl");
  pwxlvl2 = getval("pwxlvl2");
  pwx = getval ("pwx");
  pwx2 = getval ("pwx2");
  jxh=getval("jxh");
/*

/* Set AP Tables */

  settable(t1,8,phi1);
  settable(t2,8,phi2);
  settable(t3,8,phi3);
  settable(t4,8,rec4);
  settable(t5,2,phi5);
  settable(t6,2,phi6);
  settable(t7,2,phi7);


/* Calculate the n_sel phases for the second nitrogen pulse */
    if (phase == 1)
        {
        }
    if (phase == 2)
        {
        tsadd(t7,2,4);
        }
    if (phase == 3)
        {
        tsadd(t1,1,4);
        }
    if (phase == 4)
        {
        tsadd(t1,1,4);
        tsadd(t7,2,4);
        }

/* CHECK VALIDITY OF PARAMETER RANGE */

  if( tpwrsl > 35 )
    {
    printf("TPWRSL too large !!!  ");
    abort(1);
    }

/* Initialize Variables */
```

```
  initval(1.0,v1);
/* required real-time multiplier for
       phase shifts. It is set to 1 so that
       the desired 'pshift' is used as
          determined by 'stweak' */
  pshift = stweak + 360.0;
  modmix = mix - tau - grt - sl901;
  sl90dif = sl901 - sl902;
  modtau = tau + sl90dif;
  djxh1=(1.0 / (2.0 * jxh)) -
2*POWER_DELAY-grt - tau- pwx2-sl901-pw;
  djxh2=(1.0 / (2.0 * jxh)) -
2*POWER_DELAY-grt - modtau- pwx2-sl902-pw;
/* BEGIN ACTUAL PULSE SEQUENCE */

/* Receiver off time */

status(A);
   rcvroff();
   delay(5e-6);
   obsstepsize(pshift);
/* Allows sl90 to be slightly more or
       less than 90 deg. to maximize
       selectivity. On varians this is often
       not necessary so stweak can be set
       to 0.0 */

   rlpower(tpwr, TODEV);
   rlpower(dpwr,DODEV);
   rlpower(dpwr2,DO2DEV);

   delay(d1);
   rgpulse(pw, t1, rof1, 0.0);


status(B);
   delay(d2);


status(C);
   rgpulse(pw, t2, rof1, 0.0);


status(D);
   delay(modmix);
   rlpower(tpwrsl, TODEV);
   rgradient('z', gzlvl1/2);
   delay(grt);
   rgradient('z', 0.0);
   delay(tau);
   xmtrphase(v1);
   rgpulse(sl901, t3, 0.0, 0.0);
   rlpower(tpwr, TODEV);
   xmtrphase(zero);
   rgpulse(pw, t2, 0.0, 0.0);

status(E);
   delay(djxh1);

   delay(2*POWER_DELAY);
   rgradient('z', gzlvl1);
   delay(grt);
   rgradient('z', 0.0);
   delay(tau);

   rlpower(tpwrsl, TODEV);
   rlpower(pwxlvl, DODEV);
   rlpower(pwxlvl2, DO2DEV);
   xmtrphase(v1);
   rgpulse(sl901, t2, 0.0, 0.0);
   rlpower(tpwr, TODEV);
   xmtrphase(zero);

  /* nitrogen pulse for n_sel */
  dec2rgpulse(pwx2, t6, rof1, 0.0);

   rgpulse(2*pw, t3, 0.0, 0.0);
   rlpower(tpwrsl, TODEV);
   xmtrphase(v1);

  /* nitrogen pulse for n_sel */
  dec2rgpulse(pwx2, t7, rof1, 0.0);

   rgpulse(sl902, t2, 0.0, 0.0);
   rlpower(tpwr, TODEV);
   xmtrphase(zero);

   delay(2*POWER_DELAY);
   rgradient('z', gzlvl1);
   delay(grt);
   rgradient('z', 0.0);
   delay(modtau);

    delay(djxh2);

status(F);
   rcvron();
   rlpower(dpwr,DODEV);
   rlpower(dpwr2,DO2DEV);
   setreceiver(t4);
}
```

3.6.3 Isotope filtered ^{13}C 1D pulse sequence

```
#ifndef LINT
static char SCCSid[] = "@(#)GE_hmqc_jrse.c";
#endif

/*      c_sel_1d variables:

post = gradient settling time (50-200us)
d1 = relaxation delay     (0.1 - 1.0 s)
pw = 1H 90         (6 - 8 us)
grt = gradient time       (1 ms)
grl = gradient level      (8000)

phase = 3 for 13C spectrum
phase = 4 for 12C spectrum

-J. Lapham 9/18/95 */

#include <standard.h>

/* Define static integers arrays
   used to create the AP tables */
static int ph10[8] = {1,1,2,2,3,3,0,0};

pulsesequence()
{

/* Declare Variables */

    /* char charvar; */

    int phase;

    double  djxh2, jxh;
```

```
/* Load Variables */
    ni = getval("ni");
    phase = (int) (getval("phase") + 0.5);
    jxh = getval("jxh");
    pwx = getval("pwx");
    pwxlvl = getval("pwxlvl");
    pwx2 = getval("pwx2");
    pwxlvl2 = getval("pwxlvl2");
    dpwr = getval("dpwr");
    dpwr2 = getval("dpwr2");

/* Initialize variables */
    djxh2 = (1.0 / (2.0 * jxh)) - pwx;

    settable(t10, 8, ph10);

/* check validity of parameter range */
    if((dm[A] == 'y' || dm[B] == 'y' ||
 dm[C] == 'y' || dm[D] == 'y'))
    {
    printf("Decoupler must be set as dm=nnnny or n\n");
    abort(1);
    }

    if((dm2[A] == 'y' || dm2[B] == 'y' ||
 dm2[C] == 'y' || dm2[D] == 'y'))
    {
    printf("Second decoupler must be set as dm2=nnnny or n\n");
    abort(1);
    }

    if( dpwr > 50 )
    {
        printf("dpwr too large (must be less than 51)!\n");
        abort(1);
    }

    if( dpwr2 > 50 )
    {
        printf("dpwr2 too large (must be less than 51)!\n");
        abort(1);
    }

/* real time variable calcs */
    mod2(ct,v3);   /* v3=0,1,0,1,.... */
    dbl(v3,v3);/* v3=0,2,... */

if (phase == 1)
    {
    assign(zero, v1);
    assign(zero, oph);
    }

if (phase == 2)
    {
    assign(two, oph);
    assign(two, v1);
    }

if (phase == 3)
    {
    mod2(ct, v1);  /* v1=0,1,... */
    dbl(v1, v1);   /* v1=0,2,... */

    mod2(ct, oph);  /* oph=0,1,...*/
    dbl(oph, oph);  /* oph=0,2, ...*/
    }

if (phase == 4)
    {
```

```
    mod2(ct, v1);  /* v1=0,1,...*/
    dbl(v1, v1);   /* v1=0,2,...*/

    assign(two, oph);
    }


/* BEGIN THE ACTUAL PULSE SEQUENCE */
 status(A);
    rcvroff();
    delay(d1);

    if (satmode[A] == 'y')
    {
    if (fabs(tof-satfrq)>0.0)
offset(satfrq, TODEV);
    rlpower(satpwr,TODEV); txphase(t10);
    rgpulse(satdly, t10, rof1, rof1);
    rlpower(tpwr,TODEV);
    if (fabs(tof-satfrq)>0.0)
    {  offset(tof,TODEV); delay(40.0e-6); }
    }

    rlpower(tpwr,TODEV);
    rlpower(pwxlvl,DODEV);
    rlpower(pwxlvl2,DO2DEV);

status(B);
    rgpulse(pw, zero, rof1, 0.0);

    delay(djxh2);

status(C);
sim3pulse(pw, pwx, pwx2, v3, v1, v1, rof1, 0.0);
sim3pulse(pw, pwx, pwx2, v3, zero, zero, rof1, 0.0);

status(D);
    delay(djxh2);

status(E); /* acquire data */
    rlpower(dpwr,DODEV);
    rlpower(dpwr2,DO2DEV);
}
```

3.6.4 Isotope filtered ^{13}C 2D NOESY pulse sequence

```
#ifndef LINT
static char SCCSid[] = "@(#)c_sel_noesy.c";
#endif

/*      13/12C selected 2D D2O Noesy:

Carbon or Nitrogen (optional) on second
or third channel


Set phase= 1,2,3,4
phase 1: states off, refocus off
phase 2: states off, refocus on
phase 3: states on, refocus off
phase 4: states on, refocus on

Variables:
mix = mixing time.
d1 = relaxation delay
```

```
pw = 90 degree proton pulse width
jxh = proton - carbon 1 bond coupling
pwx = 90 13C
pwxlvl = 13C hard pulse power
pwx2 = 90 15N
pwxlvl2 = 15N hard pulse power

Water Presaturation:
    satmode='ynnnn'
    satfrq = frequency for presat
    satpwr = saturation power (5-8)
    satdly = saturation delay (0.1 - 1.0 s)

t2 processing:
    addition of fid#1 with fid#2 gives c12
    subtraction of fid#2 from fid#1 gives c13

t1 processing:
    normal states processing
    for phasing use phase0 = 90, phase1 = -180

    -- Jon Lapham 7/25/95
    -- G.M. Dhavan 3/1/96
-- Modified by Anna Lee 4/11/97
 */

#include <standard.h>

static int ph3[2] = {2,0},
    ph4[2] = {0,2},
    ph10[8] = {1,1,2,2,3,3,0,0};

pulsesequence()
{
/* Declare Variables */
        int     phase;
    double  mix, djxh2, jxh, t1_delay,
mix_corr, grt, grl, post;

/* Load Variables */
    phase = (int) (getval("phase") + 0.5);
    mix = getval("mix");
        ni = getval("ni");
    jxh = getval("jxh");
    dpwr = getval("dpwr");
    dpwr2 = getval("dpwr2");
    pwxlvl = getval("pwxlvl");
    pwxlvl2 = getval("pwxlvl2");
    pwx = getval("pwx");
    pwx2 = getval("pwx2");
    grt = getval("grt");
    grl = getval("grl");
    post = getval("post");
    sw1 = getval("sw1");

/* initialize variables */
djxh2 = (1.0 / (2.0 * jxh)) - pwx - rof1;

if  ( pwx2 > pwx )
{ t1_delay = (2*pw/3.1415) + pwx2 + rof1; }

if  ( pwx > pwx2 )
{ t1_delay = (2*pw/3.1415) + pwx + rof1; }

    mix_corr = mix - rof1 - grt - post;

/* Set AP tables */
    settable(t3, 2, ph3);
    settable(t4, 2, ph4);
    settable(t10, 8, ph10);
```

```
/* Real time phase cycling calculations */
/* phase = 1,2,3,4 to collect separate
fids for 12C and 13C data */
    mod2(ct,v1);   /* v1 = 0,1 */
    dbl(v1,v1);/* v1 = 0,2 */

    mod2(ct,oph);  /* oph = 0,1 */
    dbl(oph,oph);  /* oph = 0,2 */

    if ((phase == 3) || (phase == 4))
        incr(v1);


/* BEGIN THE ACTUAL PULSE SEQUENCE */

status(A);
    rcvroff();
    delay(d1);

    if (satmode[A] == 'y')
    {
    if (fabs(tof-satfrq)>0.0) offset(satfrq, TODEV);
    rlpower(satpwr,TODEV); txphase(t10);
    rgpulse(satdly, t10, rof1, rof1);
    rlpower(tpwr,TODEV);
    if (fabs(tof-satfrq)>0.0)
    { offset(tof,TODEV); delay(40.0e-6); }
    }

    rlpower(tpwr,TODEV);
    rlpower(pwxlvl,DODEV);
    rlpower(pwxlvl2,DO2DEV);

status(B);

        if (d2 == 0)
        {
         rgpulse(pw, v1, rof1, 0.0);
             delay(d2);
             rgpulse(pw, t3, rof1, 0.0);
        }

        else
        {
         rgpulse(pw, v1, rof1, 0.0);
     delay(d2/2 - t1_delay);
     sim3pulse(0.0,pwx*2,pwx2*2,t3,t4,t4,rof1,0.0);
     delay(d2/2 - t1_delay);
         rgpulse(pw, t3, 0.0, 0.0);

        }

status(C); /* NOE mixing time */
    rgradient('z', grl);
    delay(grt);
    rgradient('z',0.0);
    delay(post);
    delay(mix_corr);


status(D); /* carbon selected HMQC */
    rgpulse(pw, t4, rof1, 0.0);   /* 90x */

    delay(djxh2);

    if ((phase == 1) || (phase == 3))
        rgpulse(2*pw,t4,rof1,rof1);
    if ((phase == 2) || (phase == 4))
        sim3pulse(2*pw,2*pwx,0.0,t4,t4,t4,rof1,rof1);

    delay(djxh2);
```

```
status(E); /* acquire data */
    rlpower(dpwr,DODEV);
    rlpower(dpwr2,DO2DEV);
}
```

3.6.5 Felix macros for processing NOESY subspectra

Notes: Processing the t1 dimension is identical to that of any other States data set. The difference between processing the labeled and unlabeled subspectra is the "mul -1" statement. The "mul –1" is used to subtract the FIDS, because Felix only has a "add to buffer" statement (adb) one of the FIDs must be multiplied by –1, then added to the other.

```
c**14N_NOESYt2 processing

cmx
cl

def phase0 0
def phase1 0
def file
def nrows 500
def wcor 'cnv 0 32'
def wind1 'sb 512 90'
def wind2 'kw 1024 2'

ty Building the matrix
c**bld &filen14.mat 2 1024 1024 0
mat &filen14.mat w
ty Transform t2

for row 1 &nrows
  re &file.dat
  stb 1

  re &file.dat
  mul -1
  adb 1
  ldb 1

c**  &wcor
  &wind1
c**  &wind2
ft

  ph
  red
  sto 0 &row
  esc escape
  if &escape eq 1 escape
  ty Row #&row$
  next
end


c**15N_NOESYt2 processing

cmx
cl
```

```
def phase0 0
def phase1 0
def file imino
def nrows 500
def wcor 'cnv 0 32'
def wind1 'sb 512 90'
def wind2 'kw 400 10'

ty Building the matrix
c**bld &filen15.mat 2 1024 1024 0
mat &filen15.mat w
ty Transform t2

for row 1 &nrows
  re &file.dat
  stb 1

  re &file.dat
  adb 1
  ldb 1

  &wcor
  &wind1
c**  &wind2
  zf 1024
  ft

  ph
  red
  sto 0 &row
  esc escape
  if &escape eq 1 escape
  ty Row #&row$
  next
end
```

3.7 References

Aboul-ela F, Nikonowicz EP, Pardi A. 1994. Distinguishing between duplex and hairpin forms of RNA by ^{15}N-^{1}H heteronuclear NMR. *FEBS Lett 347*:261-264.

Archer SJ, Baldisseri DM, Torcia DA. 1992. Optimization of baseline and folding in spectra obtained using the TPPI format. *J. Mag. Res. 97*:602-606.

Arrowsmith CH, Pachter R, Altman RB, Iyer SB, Jardetzky O. 1990. Sequence-specific ^{1}H NMR assignments and secondary structure in solution of *Escherichia coli* trp repressor. *Biochemistry 29*:6332-6341.

Burgering MJ, Boelens R, Caffrey M, Breg JN, Kaptein R. 1993. Observation of inter-subunit nuclear Overhauser effects in a dimeric protein. Application to the Arc repressor. *FEBS Lett 330*:105-109.

Cai Z, Ignacio Tinoco J. 1996. Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry 35*:6026-6036.

Fesik SW. 1988. Isotope edited NMR spectroscopy. *Nature 332*:865-866.

Fesik SW, Luly JR, Erickson JW, Abad-Zapatero C. 1988. Isotope edited proton NMR study on the structure of a pepsin/inhibitor complex. *Biochemistry 27*:8297-8301.

Folkers PJM, Folmer RHA, Konings RNH, Hilbers CW. 1993. Overcoming the ambiguity problem encountered in the analysis of nuclear overhauser magnetic resonance spectra of symmetric dimer proteins. *JACS 115*:3798-3799.

Harris RK. 1985. *Nuclear Magnetic Resonance Spectroscopy*. New York: John Wiley and Sons.

Howarth MA, Lian LY, Hawkes GE, Sales KD. 1986. Formalisms for the description of multiple-pulse NMR experiments. *J Mag Res 68*:433-452.

John SantaLucia J, Shen LX, Cai Z, Lewis H, Ignacio Tinoco J. 1995. Synthesis and NMR of RNA with selective isotopic enrichment in the bases. *NAR 23*:4913-4921.

Lapham J, Rife JP, Moore PB, Crothers DM. 1997. Measurement of diffusion constants for nucleic acids by NMR. *Journal of Biomolecular NMR 10*:255-262.

Lippens G, Dhalluin C, Wieruszeski JM. 1995. Use of the water flip-back pulse in the homonuclear NOESY experiment. *J Biomol NMR 5*:327-331.

Otting G, Qian YQ, Billeter M, Müller M, Affolter M, Gehring WJ, Wüthrich K. 1990. Protein--DNA contacts in the structure of a homeodomain--DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J 9*:3085-3092.

Otting G, Wüthrich K. 1990. Heteronuclear filters in two-dimensional [1H-1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. *Q. Rev. Biophys 23*:39-96.

Piotto M, Saudek V, Sklenar V. 1992. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR 2*:661-665.

Shaka A, Barker P, Freeman R. 1985. Computer-optimized decoupling scheme for wideband application and low-level operation. *Journal of Magnetic Resonance 64*:547-552.

Shriver J. 1992. Product Operators and Coherence Transfer in Multiple-Pulse NMR Experiments. *Concepts in Mag Res 4*:1-33.

Sich C, Flemming J, Ramachandran R, Brown LR. 1996. Distinguishing Inter- and Intrastrand NOEs Involving Exchangeable Protons in RNA Duplexes. *J Mag Res Series B 112*:275-281.

Simorre J-P, Marion D. 1990. Acquisition schemes and quadrature artifacts in phase-sensitive two-dimensional NMR. *J. Mag. Res. 89*:191-197.

Sørensen OW, Eich GW, Levitt MH, Bodenhausen G, Ernst RR. 1983. Product operator formalism for the description of NMR pulse experiments. *Progress in NMR Spectroscopy 16*:163-192.

States DJ, Haberkorn RA, Ruben DJ. 1982. A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants. *J Mag Res 48*:286-292.

Torchia DA, Sparks SW, Bax A. 1989. Staphylococcal nuclease: sequential assignments and solution structure. *Biochemistry 28*:5589-5524.

Weiss MA. 1990. Distinguishing symmetry-related intramolecular and intermolecular nuclear overhauser effects in a protein by asymmetric isotope labeling. *J Mag Res 86*:626-632.

Zimmer DP, Crothers DM. 1995. NMR of enzymatically synthesized uniformly $^{13}C^{15}N$ labeled DNA oligonucleotides. *PNAS 92*:3091-3095.

CHAPTER 4 "MEASUREMENT OF DIFFUSION CONSTANTS FOR NUCLEIC ACIDS BY NMR"

4.1 Summary

In this chapter the pulsed field-gradient stimulated echo NMR technique is utilized to measure the diffusion rate of a series of standard B-form DNA samples. Effects due to DNA concentration, salt and temperature are addressed. The results are compared to hydrodynamics theory calculations and to the results obtained using non-NMR techniques, and are found to be in good agreement. It is hoped that these results will be used as a yardstick for future diffusion measurements of nucleic acids of unknown shape, which cannot be as easily modeled with hydrodynamic theory, such as bent DNA and DNA-ligand compounds.

The utility of this technique is demonstrated by solving one of the more common problems in RNA NMR spectroscopy, knowing whether a particular sample is monomeric or not. The diffusion measurement technique is shown to be able to solve this problem by measuring the diffusion rates of a 14 nucleotide RNA monomer and a 14 base pair RNA dimer, which were found to be quite different and fairly well predicted by the hydrodynamics theory.

4.2 Introduction

NMR spectroscopy is a powerful tool for studying biomolecular structure and dynamics, and it is in the light of these two goals that many experiments are driven. However, early in the development of this technique (Hahn, 1950) it was noticed that molecular translational diffusion effects could be seen in certain NMR experiments. In fact, Carr and Purcell in 1954 published a paper entitled "Effects of diffusion on free precession in nuclear magnetic resonance experiments", but it is better known to most spectroscopists because of the final paragraph in which they mention that they have also developed the "inversion-recovery" method for measuring longitudinal relaxation.

Measuring the effects of molecular diffusion by NMR requires that there is a gradient of B_o field through the sample; for the early spectroscopists this was provided by the poor homogeneity of their instruments. For later spectroscopists who had the advantage of more homogeneous magnetic fields, this gradient was provided with the advent of magnetic field-gradient coils. As the quality of these inducible gradient-fields has improved, the ability to quantitate molecular diffusion has improved as well.

The rate at which individual DNA and RNA molecules move through solution, the translational self-diffusion rate, is of fundamental importance for many important aspects of nucleic acid biochemistry. Any process that changes the apparent hydrodynamic parameters of a nucleic acid, such as protein or ligand binding, drug intercalation, or bending, can produce a measurable change in this diffusion rate.

The NMR PFG spin-echo technique (Hahn, 1950; Stekjskal and Tanner, 1965) has long been used to measure diffusion constants. Applications to biological systems include determination of the aggregation state of proteins (Alteiri, *et al.*, 1995, Dingley, *et al.*, 1995), measurement of the bulk movement of hemoglobin in human erythrocytes (Kuchel & Chapman, 1991) and quantitation of processes such as amide proton exchange with water (Andrec & Prestegard, 1996). For the NMR spectroscopist, it provides a simple, accurate method for measuring the diffusion constants of the materials they are investigating under the same conditions as other NMR experiments they do. Results of application of this technique to DNA and RNA are presented here, and compared to those obtained by other methods, and the predictions from theory.

The ability to affirm that RNA samples are monomeric is of paramount importance for NMR spectroscopists performing structural studies on short RNA oligonucleotides. The spectrum of a hairpin can often be similar to that of the duplex, formed from the same sequence, due to the inherent symmetry of dimerization. Many experiments have been utilized to investigate this problem: monitoring the hyperchromic UV shift of melting (Marky & Breslauer, 1987; Cheong, *et al.*, 1990; Varani, *et al.*, 1991; Heus & Pardi, 1991), native polyacrylamide gel electrophoresis (Sen & Gilbert, 1992), NMR T_1/T_2 relaxation measurements, and ^{15}N isotope-filtered NOESY experiments (Aboul-ela, *et al*., 1994; Sich, *et al*., 1996). Many of the possible non-NMR experiments must either be done in buffers different from those used for NMR or are incompatible with the high RNA concentrations required for NMR. The T_1/T_2 relaxation measurement can be difficult to implement, especially in the 2D heteronuclear NMR experiments, and may be complicated by dynamics that are independent of the aggregation state of the RNA. The ^{15}N X-filtered NOESY experiment developed by Aboul-ela provides a general solution to the problem, but it requires the labor-intensive synthesis of isotope labeled RNA, and the mixing of precious labeled RNA with unlabeled RNA.

It should be possible to discriminate between an RNA hairpin and the corresponding self-dimer by measuring the translational self-diffusion rates. In the case of short oligonucleotides, it is often possible to drive the hairpin to duplex equilibrium by increasing strand concentration and salt concentration, which makes it possible to compare the two states. Additionally, by selecting the appropriate hydrodynamic model for the RNA, it should be possible to predict the diffusion rates for both states. Further analysis and comparison of the diffusion rate of a variety of RNAs may yield structural insights into their molecular shapes.

4.2.1 Hydrodynamics theory

The translational self diffusion coefficient (D_t) for a molecule in solution is related to its translational frictional coefficient (f_t) by Einstein's equation:

$$D_t = kT / f_t \qquad (4.1)$$

Thus, an accurate calculation of D_t is equivalent to an accurate calculation of a frictional coefficient. Frictional coefficients are usually computed assuming the hydrodynamic shape of a molecule is a sphere, a prolate (or oblate) ellipsoid or a symmetric cylinder. While it seems obvious that the best model for a duplex nucleic acid would be a symmetric cylinder, given that the sizes of the nucleic acids we studied (a 14 nucleotide RNA hairpin to a 24 base pair DNA) we also investigated modeling them as spheres or ellipsoids.

The spherical model for nucleic acids is probably accurate for either short duplexes or short hairpins. In this case: where r is the hydrodynamic radius of the sphere and η is the viscosity of the solvent,

$$f_t = 6\pi\eta r \qquad (4.2)$$

As the length of the nucleic acid duplex increases, prolate ellipsoid models may be more successful. In this case, the Perrin equations (Cantor & Schimmel, 1980) can be used,

$$f_t = 6\pi\eta_o (ab^2)\left[\frac{(1-p^2)^{1/2}}{p^{2/3}}\right]\ln\{(1+(1-p^2)^{0.5}) / p\} \qquad (4.3)$$

Where *a* is defined as half the length of the long axis and *b* as half the length of the short axis for an ellipse. The axial ratio, *p*, is *b*/*a*.

Expressions for the frictional coefficient for a short symmetric cylinder model were developed by Tirado and Garcia de la Torre (1979, 1980) which are appropriate for short rod like molecules with 2 < *q* < 30, where *q=1/p=a/b,*

$$f_t = 6\pi\eta_o \left[\frac{L/2}{\ln q + 0.312 + 0.565 q^{-1} - 0.100 q^{-2}} \right] \qquad (4.4)$$

This expression is known to work well for DNA dimers of moderate size (Eimer, *et al*., 1990).

4.2.2 NMR theory

Stekjskal and Tanner (1965) first proposed a spin-echo experiment to measure the diffusion rate of molecules in solution by NMR (see figures 4.1 and 4.2). Their method relies on two gradient pulses surrounding the 180^{o} pulse in the spin-echo; the first dephases the transverse magnetization in a spatially dependent manner along the z-axis and the second gradient then rephases the magnetization. If the molecule moves along the z-axis during the time between the two gradients, its magnetization will not refocus completely. Thus, if the molecule diffuses rapidly, the attenuation of its resonances will be large; if the molecule diffuses slowly, the attenuation will be relatively small. The following relation exists between translational self-diffusion and the measurable NMR parameters (Stekjskal & Tanner, 1965),

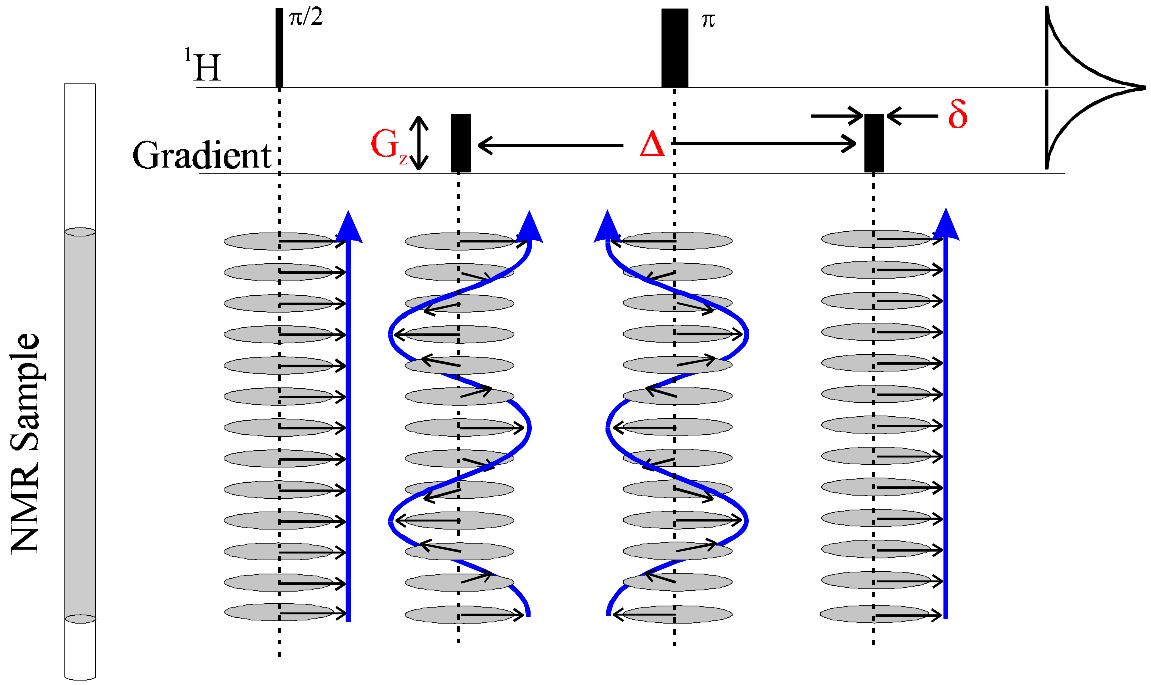


Figure 4. 1 PFG spin-echo without translational diffusion

The concept of measuring translational diffusion can be best explained schematically using the simple PFG spin-echo pulse sequence as shown above. The relationship between the position in the NMR sample and what occurs during the pulse sequence is demonstrated by following the "disks" from left to right. After the first 90^o x pulse, all the magnetization of the sample (in the rotating frame and on-resonance) "points" in the same direction in the transverse plane. The first gradient pulse "encodes" the sample by causing the nuclei of the sample to precess at different frequencies for the gradient duration δ. This has the effect of inducing what appears to be a "spiral staircase" effect through the sample with respect to the z-axis, as demonstrated in the figure. The 180^o pulse inverts the relative position of all the nuclei. The final gradient pulse "decodes" the magnetization and restores the original magnetic vector orientation. If no diffusion has occurred during the time Δ, the resultant signal will be of 100% intensity. The next figure demonstrates this same pulse sequence with translational diffusion.

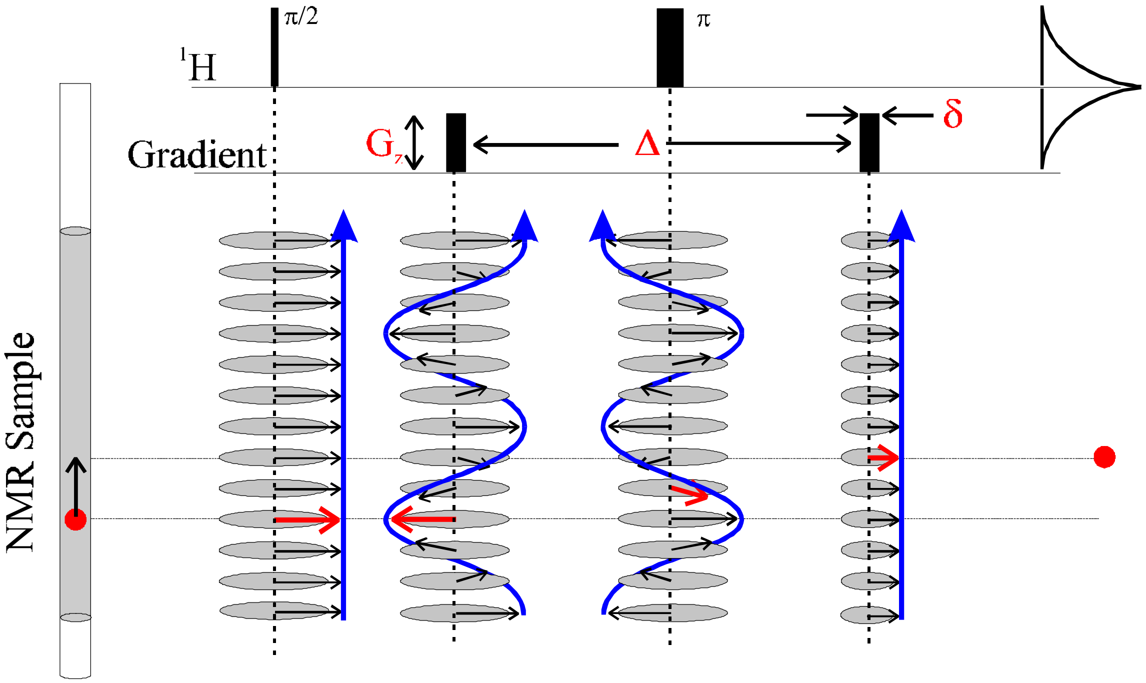


Figure 4. 2 PFG spin-echo with translational diffusion

Similar to figure 4.1, translational diffusion is demonstrated for the same PFG spin-echo pulse sequence. The "molecule" is represented by the red circle, which moves from its original position as shown in the NMR tube on the left to its new position as shown on the right. If this movement occurs between the encoding and decoding gradients (of time duration Δ), this will cause attenuation in the observable signal due to incomplete refocusing. Note that while this is only shown for this one molecule, it is the ensemble average movement of all the molecules in solution that is recorded.

$$A / A_o = -\exp[D_t \mathrm{g}_H{}^2 \mathrm{d}^2 G_z{}^2 (\Delta - \mathrm{d} / 3)] \qquad (4.5)$$

Where A is the measured peak intensity (or volume), A_0 is the maximum peak intensity, D_t is the translational diffusion constant (in cm^2/s), g_H is the gyromagnetic ratio of a proton (2.675197×10^4 $gauss^{-1}$ s^{-1}), d is the duration of the gradient, D is the time between gradients and G_z is the strength of the gradient (in gauss/cm). Data can be plotted as $-\ln(A/A_0)$ vs $\gamma_H{}^2\delta^2 G_z{}^2(\Delta-\delta/3)$. The slope of the line that emerges is D_t.

4.3 Results

4.3.1 NMR Experimental

A number of variants of the original PFG spin-echo pulse sequence have been developed for measuring diffusion rates. A Stimulated Echo (PFG-STE) pulse sequence (see figure 4.3) was developed by Tanner (Tanner, 1970) which makes use of three 90^o pulses and stores magnetization along the z-axis (minimizing T_2 relaxation effects) during a large portion of the experiment. It works well for studying molecules with $T_1 > T_2$, such as large biomolecules. The inductive eddy-currents magnetic field-gradients created in the electronics of probes can affect the line shapes of resonances in PFG experiments. Many variants to the PFG-STE have been developed to minimize these effects. A refocused stimulated echo sequence was developed by Griffiths and Horton (1990) in which a train of refocusing 180^o pulses is applied at the end of the standard PFG-STE as well as a four pulse sequence with a *l*ongitudinal *e*ddy-current *d*elay (PFG-LED) (Gibbs & Johnson, 1991) which allows for an extra delay time before acquisition. Shaped gradient pulses (Price & Kuchel, 1991) have also been used. A water suppression component has been included in the water-suppressed LED (water-sLED) pulse sequence (Altieri, *et al.*, 1995).

We found that for our hardware, the relaxation time required for the gradient induced eddy-currents to decay to zero was short enough so as to not be a factor (see materials and methods, *NMR calibration*). For this reason, we utilized the simpler technique of Tanner's three pulse 'z-storage' pulsed field-gradient Stimulated Echo (PFG-STE) pulse sequence. Many of the more complex eddy current suppression pulse

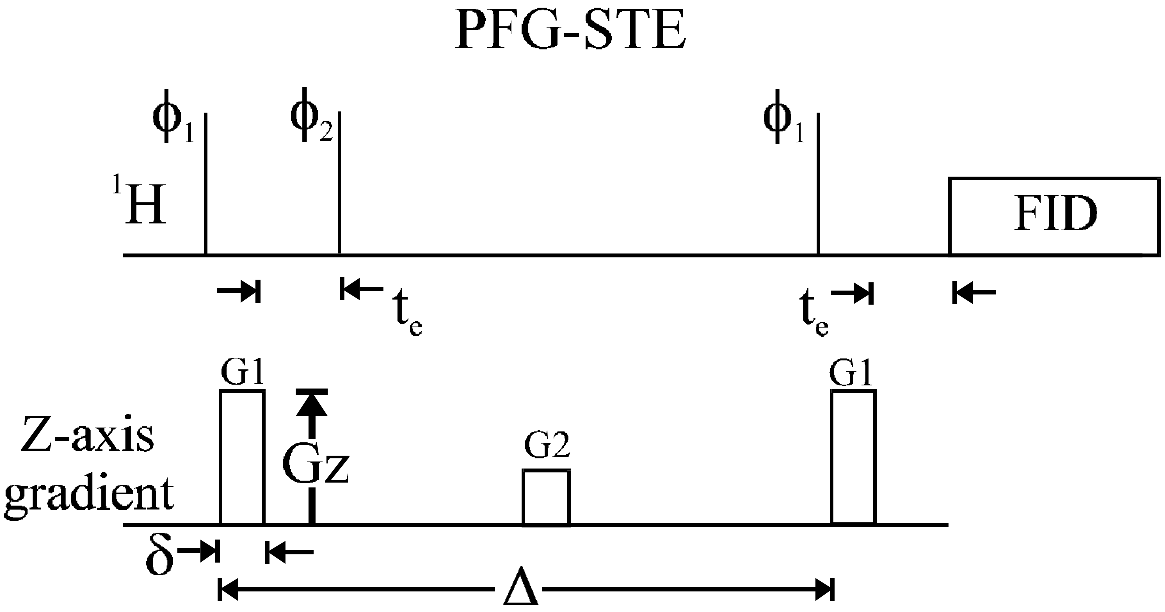


Figure 4. 3 PFG-STE (Tanner, 1970) pulse sequence for the diffusion measurements.

The symbol "δ" refers to the length of the first and third gradient pulse, "Δ" is the time between the first and third gradient pulse and G_z is the strength of the gradient pulse. One experiment would involve choosing a particular δ and Δ value (between 1-5 ms for δ and 25-200 ms for Δ), and collecting 31 1D spectra in which the value of G_z is incremented from 1-31 G/cm. The middle gradient pulse is a spoiler to remove any unwanted transverse magnetization during the z-axis storage. The time t_e is the time for complete eddy-current relaxation, and must be calculated independently for each hardware setup, we used a delay of 2 ms.

sequences just mentioned were also implemented, but they did not affect the quality of the data.

4.3.2 DNA

The three DNA duplexes studied (12, 14 and 24 bps) were prepared in concentrations ranging from 250 μM to 2000 μM to examine the effect of DNA concentration on the translational self-diffusion rate. Figure 4.4 graphically demonstrates the DNA results and Tables 4.1 and 4.2 summarizes the results.

It is clear that there is indeed a concentration dependence, with the apparent diffusion rate being lower for high concentration samples (figure 4.4A). Furthermore, the concentration dependence effect is more pronounced for the longer samples: D24 shows an almost 20% decrease in diffusion rate between the 250 μM and 1500 μM sample, while D12 shows only an ~5% decrease over the same concentration range. Figure 4.4B demonstrates that plots of D_t vs nucleotide concentration gives similar slopes between samples. A simple linear virial correction to the measured self-diffusion rate,

$$D_t(measured) = D_0(1 + kc) \qquad (4.6)$$

describes this concentration dependence quite well, with *c* given in terms of nucleotide concentration (see Table 4.1 for values of *k*). The diffusion constants of DNAs at zero concentration were determined by linear regression of the data plotted in figure 4.4B, and the values are reported in Table 4.2. The theoretical f_t and D_t values calculated for DNAs varying in size from 5 to 35 bps are graphed in figure 2C/D along with the measured D_t (and back calculated f_t) values. Clearly, the Tirado and Garcia de la Torre symmetric cylindrical model fits the DNA data best.

Sample	Complex Conc. (/1 mM)	Nt Conc.[a] (/1 mM)	D_t (/10^{-6} cm^2/s)	Error[b] (/10^{-6} cm^2/s)
HDO[c]	~0	N/A	18.89	0.005
D12	0.25	6	1.241	0.040
D12	0.50	12	1.236	0.029
D12	1.00	24	1.180	0.019
D12	1.50	36	1.188	0.027
D12	2.00	48	1.180	0.023
D14	0.25	7	1.181	0.038
D14	0.50	14	1.163	0.030
D14	1.20	33.6	1.077	0.018
D14	2.00	56	1.034	0.011
D24	0.25	12	0.910	0.015
D24	0.50	24	0.910	0.020
D24	1.00	48	0.854	0.014
D24	1.50	72	0.788	0.013

[a] Nucleotide concentration was calculated by multiplying the number of nucleotides per molecular complex by the molecular complex concentration.

[b] Errors were calculated from the linear graphs of $-\ln(y/y_o)$ vs $\gamma^2\delta^2G_z^2(\Delta-\delta/3)$ using standard linear regression techniques.

[c] The HDO sample was made from Aldrich (cat 26,978-6) "Deuterium oxide 100.0 atom % D".

Table 4. 1 Measured diffusion constants for all samples

Size	Theoretical (/10^{-6} cm^2/s)	Experimental D_t (/10^{-6} cm^2/s)	k (/10^{-3} cm^{2}s^{-1}mM^{-1})
D12	1.247	1.230 (.020)	-1.4(.4)
D14	1.170	1.187 (.015)	-2.7(.2)
D24	0.903	0.954 (.015)	-2.2(.2)
R14ls[b]	1.90	1.41(.014)	
R14hs[c]	1.16	0.918(.024)	

[a] R14ls was modeled as a sphere with a radius of 21Å (as discussed in the text) and the reported D_t values was not corrected for concentration, [R14ls] = 1.8mM.

[b] R14hs was modeled as a rigid cylinder using the hydrodynamic parameters of 2.6Å rise/bp and 24Å diameter and the experimental D_t value was not corrected for concentration, [R14hs] = 2.0mM.

[c] Values were calculated using the rigid cylindrical rod model at 25°C and a 100% D_2O. For DNA the hydrodynamic parameters of 3.4Å rise per bp and 20Å diameter were used. For RNA 2.6Å rise per bp and 24Å diameter were used. Experimental D_t values for the DNA come from extrapolation to zero concentration. *k* is the virial coefficient in equation 8, using concentration units of mM nucleotide (not strand) concentration.

Table 4. 2 Theoretical and experimental self-diffusion constants[a]

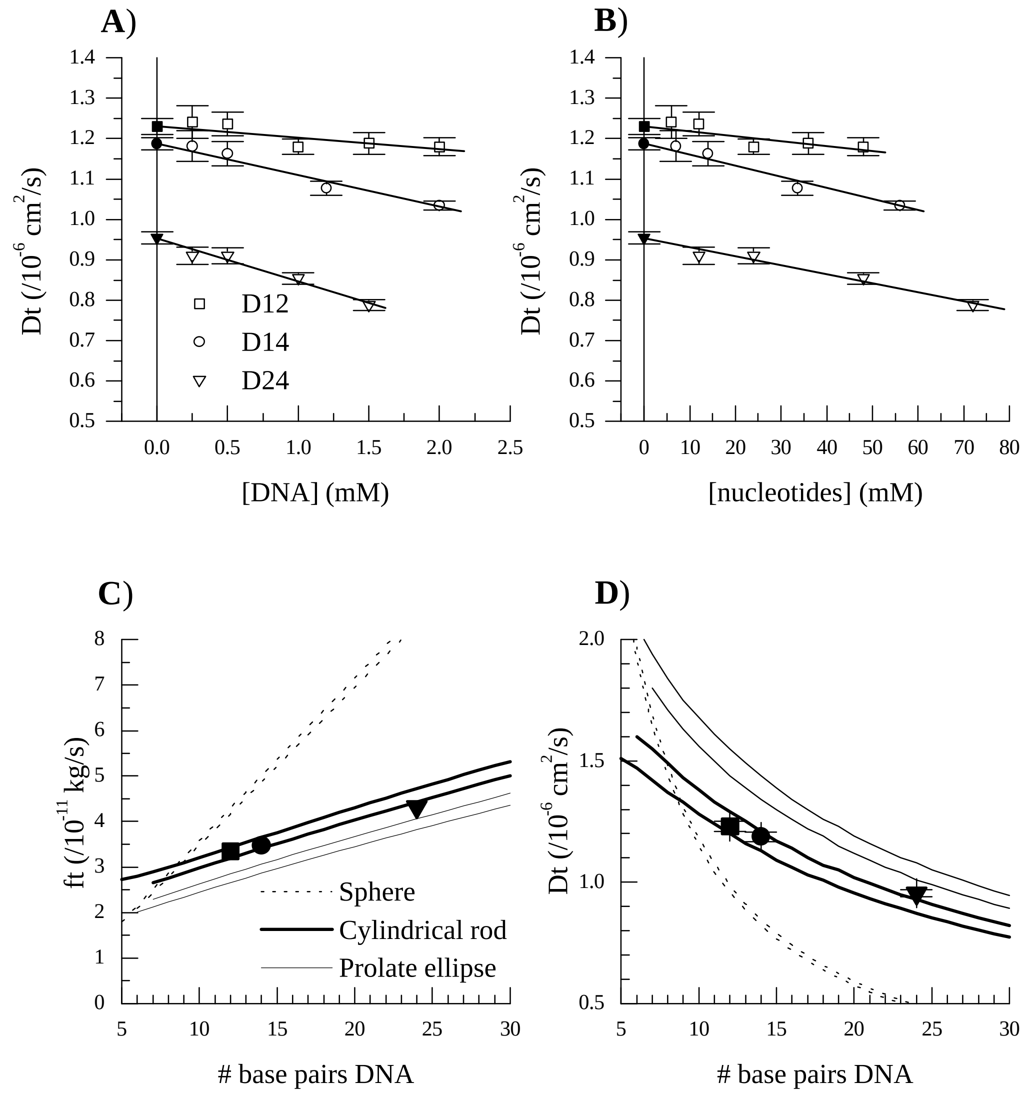


Figure 4. 4 Concentration dependence of D_t and f_t

A) a plot of the concentration dependence (D_t vs [DNA]) of the measured diffusion rate for the D12, D14 and D24 samples. The experimental data are represented by open squares, open circles and open diamonds for each sample respectively. The extrapolated "zero-concentration" values are shown as solid symbols. **B**) The same data as in **A**) but plotting D_t vs nucleotide concentration. **C**) Graph of the theoretically calculated translational friction coefficients for a sphere (between the dotted lines), ellipse (between the thin lines) and cylindrical top (between the thick lines) at 25^o C in 100% D_2O as a function of DNA base pair length, using the hydrodynamic parameter range of 3.4(±0.5)Å rise/bp and a diameter of 20(±1.0)Å. **D**) Graph of the theoretically calculated translational diffusion constant.

The temperature dependence of D_t for the DNA was examined by collecting data on D24 at temperatures ranging from 10-50 ^oC. Equation 4.1 predicts direct proportionality between D_t and temperature; however, the temperature dependence of viscosity must also be calculated (using equation 4.8). Figure 4.5 graphs the theoretically predicted temperature dependence of a 24 bp DNA (using the parameters of 3.4(±0.5)Å rise/bp and 20.0(±1.0)Å diameter), overlayed with the experimentally measured values (corrected for DNA concentration). Data are only shown to 35^oC, because at higher temperatures, the gradients did not give a linear response (see Materials and Methods section 4.5.3 for discussion on examining linear gradient response) and reliable data could not be obtained.

Data were collected on D12 at 3 NaCl ion concentrations (50mM, 100mM and 200mM), to examine the effect this might have on our reported D_t values. There was no appreciable change in the measured D_t values outside experimental error (data not shown). Fujimoto *et al* (1994) have measured the dependence of the hydrodynamic radius (R_H) of a 48 bp DNA on cation concentrations using fluorescence polarization anisotropy (FPA) of intercalated ethidium. They found that NaCl concentration had the smallest effect of any of the cations examined, decreasing R_H by 0.30Å from [NaCl] = 25mM to 100mM. Other cations such as Mn^{2+} and Mg^{2+} gave rise to much larger changes in R_H. Our data on the NaCl effects seem to be in agreement with what they report.

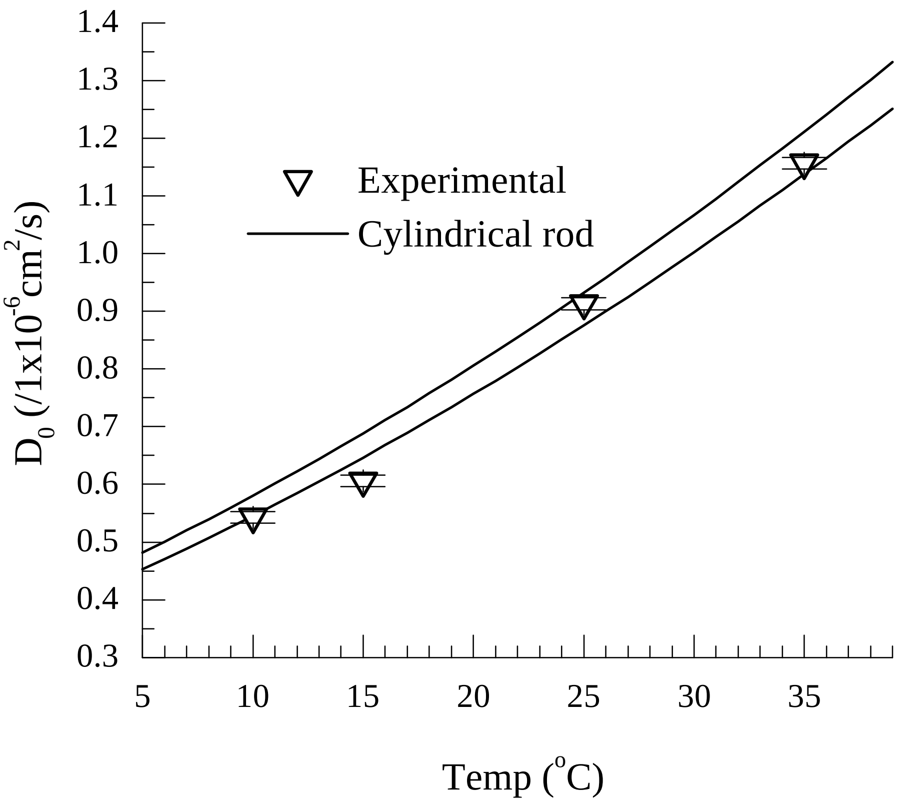


Figure 4. 5 Diffusion constant vs temperature.

Solid lines represent the theoretically calculated diffusion rate using the cylindrical rod method with 3.4(±0.5)Å rise/bp and 20.0(±1.0)Å diameter. Data were collected on D24 at 1.5mM concentration (72 mM nucleotide concentration); the results shown were corrected for concentration using k=-2.2x10^{-3} cm^2 s^{-1} mM^{-1}

4.3.3 RNA

The RNA studied, R14, could be examined either as a hairpin or a duplex because its conformation depends on the NaCl concentration. Under the conditions of low salt (100mM NaCl), the RNA (R14ls) is a hairpin with the approximate hydrodynamic dimensions of L = (2.6 Å rise/pb) * 7 bp = 18.2Å and D = 24Å. Assuming a sphere of radius 18-24Å, the range of D_t predicted is $2.19\text{x}10^{-6}$ to $1.66\text{x}10^{-6}$ cm^2/s using equation 4.2. With an the average radius value of 21Å, the theoretical D_t is $1.90\text{x}10^{-6}$. The rationale for modeling R14ls as a sphere comes from the observation (Eimer, 1990) that a DNA tridecamer which adopted a hairpin structure was nearly spherical in its hydrodynamic dimensions. By analogy the RNA tetradecamer hairpin should adopt a nearly spherical structure. Under the conditions of high salt (400mM NaCl), the duplex RNA (R14hs), can be modeled as a right cylinder of dimensions L = 36.4Å and D = 24Å, which gives a theoretical D_t of $1.16\text{x}10^{-6}$ cm^2/s from equation 4.4. The ratio of the theoretically calculated D_t(duplex) : D_t(monomer) is 0.61.

The data obtained for R14ls and R14hs are shown graphically in figure 4.6. The diffusion constants obtained were $1.41(.014)\text{x}10^{-6}$ and $0.918(.024)\text{x}10^{-6}$ cm^2/s, for the monomer and duplex respectively. These values were not corrected for concentration effects. This gives a experimentally calculated D_t(duplex) : D_t(monomer) of 0.65, in close agreement with the predicted ratio of the diffusion rates for a duplex : monomer.

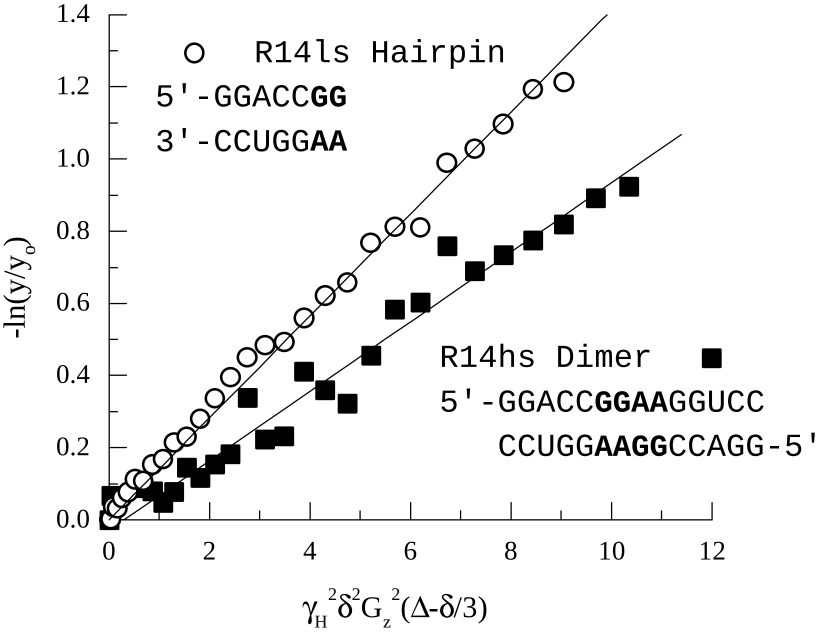


Figure 4. 6 Diffusion constants for RNA

Measurements at 25°C for the low salt hairpin R14ls (1.8 mM strand concentration, 25.2 mM nucleotide concentration) and the high salt duplex R14hs (2.0 mM strand concentration, 28 mM nucleotide concentration). The sequences of the RNA are shown, with the hairpin loop and internal loop regions represented by the bold letters.

4.4 Discussion

4.4.1 DNA: Comparison to other techniques

The hydrodynamic parameters of length and diameter appropriate for double helical DNA have long been debated. Fiber diffraction studies of high humidity B-form DNA suggest a phosphate to phosphate diameter for DNA of 20Å (Arnott & Hukins, 1972; Elias & Eden, 1981). However, the hydrodynamic diameter should include any associated water that moves with the DNA. Our lab has reported a hydrodynamic radius of 22-26Å and 3.34+/-0.1Å rise per base pair for B-form DNA (Mandelkern, *et al*., 1981) based on a combination of quasielastic light scattering and birefringence rise/decay of electric-field oriented molecules in the size range of 64 - 267 base pairs. Measurements of large fragments must be corrected for the bendability of DNA, which was accomplished by Mandelkern *et al.* (1981) by extrapolation to zero bendability with the help of a theoretical model (Hearst, 1963).

Smaller DNA fragments do not require such an extrapolation and should thus be better model compounds for study. Measurements of translational and rotational diffusion rates by dynamic light scattering and NMR relaxation on short fragments (8, 12 and 20 base pairs) of DNA has given values of 20.0(±1.0) Å for the hydrodynamic diameter and a value of 3.4(±0.05) Å rise per base (Eimer, *et al*., 1990; Eimer & Pecora, 1991), and indicate that there may not be a water shell which diffuses with the DNA. These experiments were performed in 50 mM phosphate buffer pH=7, 100 mM NaCl, 2 mM EDTA, 0.1% NaN_3 and in 100% H_2O. The D_t reported for each at 20^oC was 1.52, 1.34 and $1.09x10^{-6}$ cm^2/s for the 8, 12 and 20mer respectively. The only direct comparison we can make with their data is for our 12mer DNA, and our values are in very close agreement, after making the appropriate corrections for both the viscosity differences between H_2O and D_2O and the temperature differences between the two sets of data. We find that the hydrodynamic values they calculate work well for predicting our data as well. A possible reason for the larger hydrodynamic radii (diameter 22-26Å vs 20Å) inferred for DNA molecules of restriction fragment size (Mandelkern, *et al.*, 1981) is the presence of small amounts of intrinsic curvature in such samples.

4.4.2 RNA

In both RNA hairpin and the duplex measurements, our experimentally determined diffusion constants are less then those predicted (see Table 4.1). There are several reasons for this. First, we have not made any concentration correction. Second, the hairpin and a duplex containing an internal loop may be poorly represented using standard A-form helical parameters for diameter and rise/bp. Nevertheless, the similarity between the diffusion constant ratios for the theoretical (0.61) and experimental (0.65) values indicates that hairpin and helical dimers can be clearly distinguished. The analogy is in using diffusion constants to determine the aggregation states of proteins (Alteiri, *et al.*, 1995; Dingley, *et al.*, 1995) when perfect hydrodynamic models are not known.

To summarize, a simple, accurate and quick experiment is presented for determining the translational self-diffusion constants of nucleic acid samples under NMR conditions. These data demonstrate that the PFG-STE technique gives accurate results for double helical standard B-form DNAs, and can be used to determine whether an RNA sample is monomeric.

4.5 Materials and methods

4.5.1 Sample preparation

All the DNA samples were prepared on an Applied Biosystems 380B DNA synthesizer and purified using denaturing PAGE techniques. Concentrations were determined by UV absorbance measurements at 260nm wavelength and calculated using a dinucleotide stacking extinction coefficient formula. The DNA sequences were (5' to 3') D12:CGCGAATTCGCG, D14:GCTATAAAAAGGGA (with the complement TGCCCTTTTTATAGC) and D24:CGCGAATTCGCGCGCGAATTCGCG. Both D12 and D24 were palindromic to alleviate any problems with stoichiometry. Five D12 samples were prepared: 250, 500, 1000, 1500 and 2000 μM. Four D14 samples were prepared: 250, 500, 1200 and 2000 μM. Four D24 samples were prepared: 250, 500, 1000 and 1500 μM. All samples were dialyzed against 20mM sodium phosphate (pH 7.0) and 100mM NaCl for two days, exchanging the dialysis buffer every 12 hours. All samples were placed in a Shigemi (Shigemi Corp., Tokyo Japan) NMR tube in a 170 μl volume, which equated to about a 1 cm sample height. The samples were then lyophylized and resuspended in 100.0 atom % D_2O from Aldrich (cat #26,978-6) to the same final sample volume of 170 μl.

The RNA sequence was (5' to 3') R14:GGACCGGAAGGUCC and was prepared enzymatically using DNA template-directed T7 RNA polymerase (Milligan, et al., 1987), and purified using denaturing PAGE techniques. The RNA was extensively dialyzed against water, concentrated, and exchanged into either a low salt buffer (50 mM NaCl, 5mM cacodylate pH 6.3, 0.1 mM EDTA) or a high salt buffer (400 mM NaCl, 5mM cacodylate pH 6.3, 0.1 mM EDTA) using 1000 MWCO centrifugal concentrators (Filtron

Technology Corp., Northborough, MA). Both samples were heated to 80^{o}C, cooled to room temperature, and placed into a Shigemi NMR tube with a sample volume of 160 μl, lyophylized, and 100.0 atom% D_2O was added to give a final sample volume of 160 μl. The final RNA "strand" concentrations were 1.8 mM and 2.0 mM for the low salt (R14ls) and high salt (R14hs) samples, respectively. The R14ls and R14hs samples were proven to consist of a single species by means of standard homonuclear and heteronuclear experiments. For example, the number of H5-H6 crosspeaks found in a DQFCOSY experiment corresponds to the number of pyrimidines in the sequence. We assume that the difference in the spectra between the two samples are due to a simple hairpin to duplex transition.

4.5.2 Solvent viscosity

All the methods discussed for modeling nucleic acid frictional coefficients require an accurate measure of the solvent viscosity, which was calculated from (Kellomaki, 1975; Natarajan, G, 1989),

$$\log \mathrm{h}_o = a + \left[\frac{b}{c - T} \right] \tag{4.7}$$

where *T* is the temperature in Kelvin. The terms a, b and c are given for a particular D_2O:H_2O ratio. For a 100% D_2O solution, *a* = -4.2911, *b* = -164.97 and *c* = 174.24. This yields a value of h_o at 25^{o} C for a 100% D_2O solution of 1.097 (Kg cm^{-1} s^{-1}) which is what we used in our calculations. For a 100% H_2O solution, *a* = -4.5318, *b* = -220.57 and *c* = 149.39. This yields a value of h_o at 25^{o} C for a 100% H_2O solution of 0.8929 (Kg cm^{-1} s^{-1}).

Corrections for salt effects on viscosity were performed as follows (Harned & Owen, 1958),

$$\mathrm{h} = \mathrm{h}_0[1 + A\sqrt{c} + B(c)] \tag{4.8}$$

$$A = .0067 \;\; B = .0244 \;\; (\,for\ NaCl)$$

where c is molar salt concentration, η_0 is the zero solute solvent viscosity and η is the new viscosity. We found that for the range of NaCl used in this study (50-400 mM) the effect on viscosity was very small, with the largest viscosity correction being $1.014\eta_0$ for the 400 mM NaCl case.

4.5.3 NMR calibration

It is absolutely critical to the interpretation of these experiments that the gradient hardware and probe be calibrated. This was done using a 1 cm high sample of 100% D_2O in a Shigemi NMR tube. Necessary calibrations include: measurement of the maximum strength of the gradient pulse, characterization of the eddy-current recovery time for the probe, and examination of the linear power response of the z-axis gradients. We found that many of our older probes did not behave properly in these tests, and they were not used. This is probably because the electronics of the older probes are not as well shielded from the gradient pulse.

Calibration of the gradient strength was accomplish by two methods. The first, which was previously published (Callaghan, *et al.*, 1983), involves measuring the diffusion rate for the residual proton water line in the calibration sample at 25°C, and back calculating G_z. This procedure assumes that the diffusion rate for HDO in a 100% D_2O sample is $1.90\text{x}10^{-5}$ cm^2/s (Longworth, 1960). The second depended on acquiring a spin-echo FID of the calibration sample with the z-axis gradient on during acquisition.

This yields a spatial profile of the sample, which is a function of the sample height and the gradient strength. Slightly different values for G_z were obtained by these two methods of calibration. The discrepancy was within 3%, and similar to the gradient strength calibration errors reported elsewhere (Doran & Décorps, 1995).

The eddy-current recovery time was examined using a pulse sequence in which a full strength gradient pulse is applied for 10 ms (a longer time than is used in the experiments) followed by an adjustable time delay and finally a 90^o proton observation pulse. Data were collected on the residual proton water line in the calibration sample. It was found that there was complete eddy-current relaxation within less than 1 ms for the triple resonance probe used in these experiments. Because of this, we simply needed to wait longer than 1 ms after applying the gradients in the PFG-STE sequence.

It is absolutely critical for these experiments that the z-axis gradients be linear in the volume occupied by the sample, and respond linearly to the power applied. The region of linearity may only be a little larger than 1 cm in typical gradient-equipped probes, so an accurate measurement requires that the sample height be no larger than this. Measurements were made using the PFG-STE sequence of the residual proton line in the calibration sample over a large range of δ and Δ times. The data gave the same D_t value for each value of δ and Δ, and the plot of $-\ln(y/y_o)$ vs $\gamma^2\delta^2G_z^2(\Delta-\delta/3)$ was a straight line, which demonstrates the linear gradient power response required.

4.5.4 NMR experimental

All the DNA data were collected on a Varian 600 MHz "UnityPlus" spectrometer on a triple resonance (H, C, N) probe. The PFG-STE pulse sequence shown in figure 4.3 was used for all the data reported. However, we also collected data using the simple PFG spin-echo and the PFG-LED pulse sequences, and obtained similar results. A post-gradient eddy-current relaxation delay of 2 ms was used on all experiments. For the 1000-2000 μM samples, 32 scans were collected at each gradient strength reported; however, for the lower concentration samples, more scans were needed to obtain reasonable signal to noise values, up to 256 scans for the most dilute 0.25 mM samples. For each data set, 2048 complex points were collected for each of 32 experiments in which the gradient strength was incremented from 1-31 G/cm in steps of 1 G/cm. A five second recycle delay was used between scans for all data shown. However, data were also collected using a range of recycle delays from 1s - 10s, with no apparent change in the measured diffusion rate. This makes sense because we are fitting the change in the integrated volumes of the molecule, not measuring the absolute volumes, thus full relaxation is not required between experiments.

The region of the spectrum from 8.5-7.0 ppm (which corresponds to the H8/H6/AH2 protons in DNA and RNA) or the region from 6.0-5.0 ppm (corresponding to the H1'/H5 protons in DNA and RNA) was integrated for each data set. Spectra were processed using the Felix95 (Biosym Technologies, San Diego, CA) software package using an automated processing macro which apodized the FID, Fourier transformed the data, applied baseline correction, integrated the peaks (see Figure 4.7 for an example) and saved a volume file for each experiment. These data were then plotted as $-\ln(A/A_0)$ vs

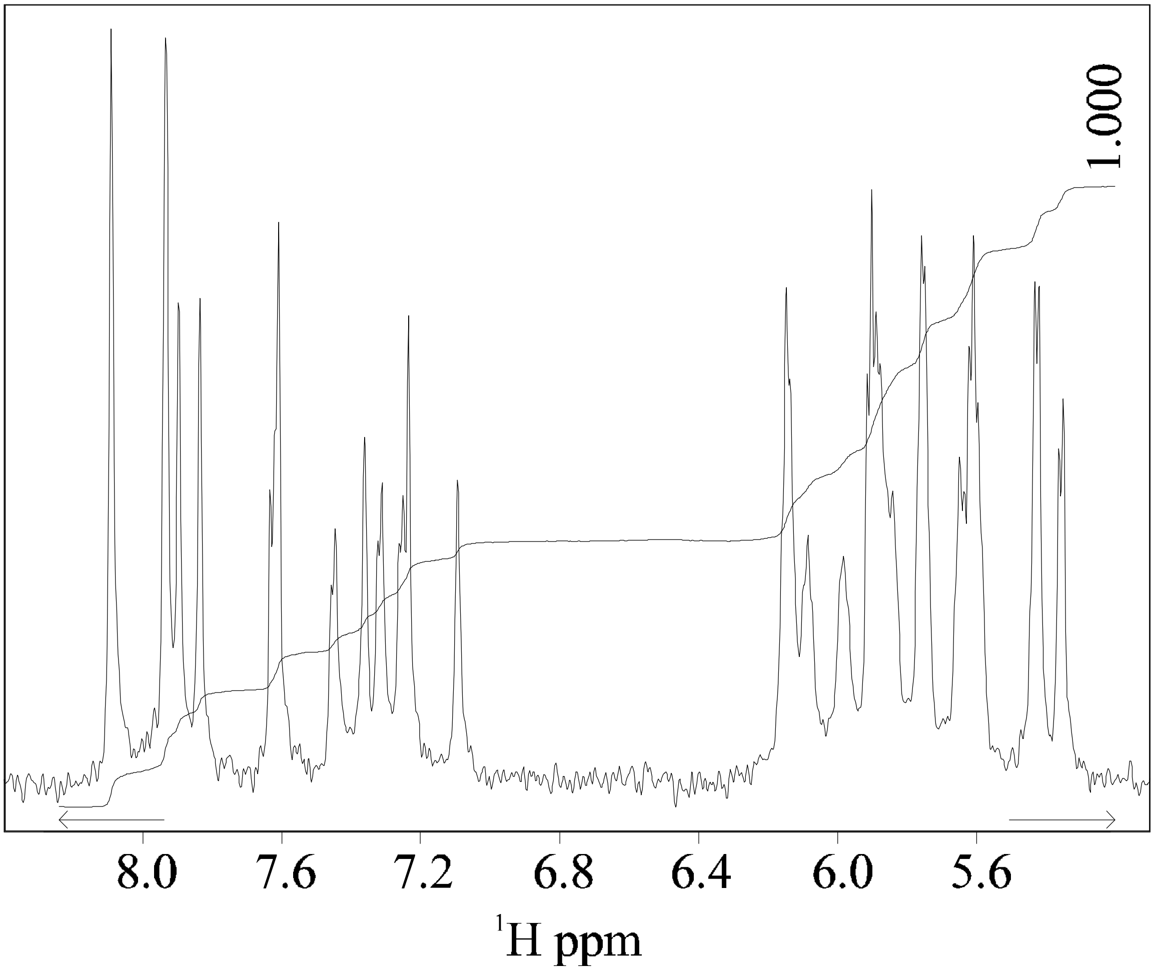


Figure 4. 7 Integration of the D12 1D spectrum

1D spectrum of D12 from the STE-PFG experiment, using the H8/H6/AH2 and H5/H1' region of the spectrum. Δ=5 ms δ=100ms and G_z=2 g/cm. The baseline should not affect the integration value. The peak integration value is measured for each 1D spectrum as the gradient strength value G_z is increased in each experiment.

$\gamma_H{}^2\delta^2G_z{}^2(\Delta\text{-}\delta/3)$ (see Figure 4.8 for an example) in which the slope of the line gives the translational self-diffusion rate of the molecule for a particular concentration

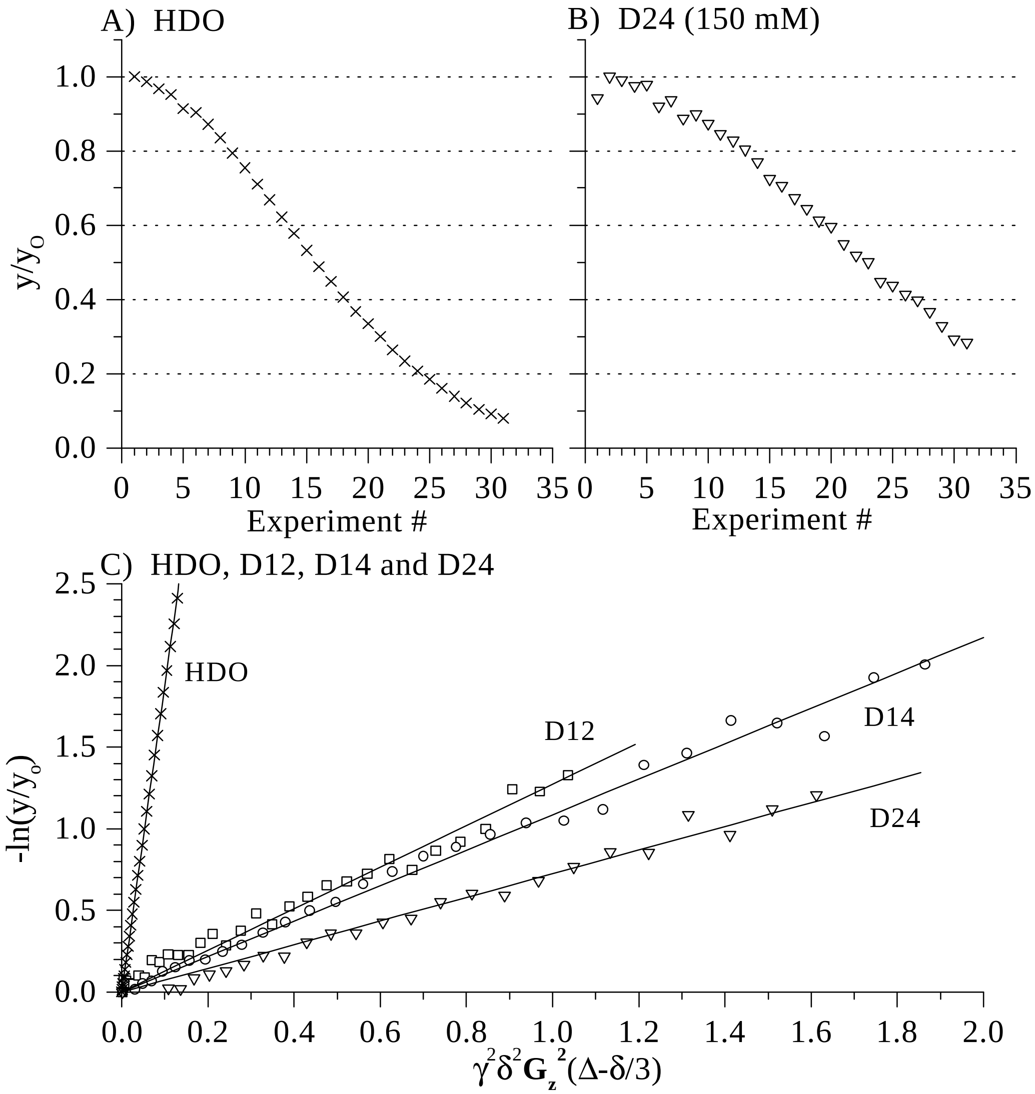


Figure 4. 8 Sample experimental data

All data shown was collected at 25°C on a Varian 600 MHz Unity Plus spectrometer using the STE-PFG (pfg_diffusion pulse sequence) experiment. **A)** Integrated intensity values for the residual HDO line, Δ=1.5 ms, δ=100 ms. The gradient strength G_z was increase from 0 to 31 g/cm in experiment #0 to #31. The sigmoidal (G_z^2) dependence of the data can be clearly seen. **B)** Integrated intensity values for the D24 DNA sample at 150mM concentration, Δ=5 ms, δ=100 ms. The gradient strength G_z was increase from 0 to 31 g/cm in experiment #0 to #31. **C)** The post-processed integrated intensity values for four sample, HDO, D12 (1.50 mM), D14 (1.20 mM) and D24 (1.50 mM). The diffusion constant for each sample comes directly from this plot, 18.89(.005), 1.188(.027), 1.077(.018) and 0.788(.013) cm^2/s respectively.

4.6 Appendix

4.6.1 Varian pulse sequence “pfg_diffusion.c”

This is the pulse sequence code used for all diffusion data collected and presented in this chapter. The graphical representation is shown in figure 4.1. The important variables in this pulse sequence are *grt1* and *dt* (d and D from equation 4.5), which correspond to the width of the encoding gradient pulse and the time between the two gradients respectively. The correct delay times between the various components of the pulse sequence are automatically calculated when setting *dt*, thus *dt* can be set to exactly the value of D needed for the experiment.

There are two additional time delays set in front of either gradient pulse named *tau1* and *tau2*. These were added to allow for ‘tweaking’ the total time of the experiment to get a better baseline, we found that *tau1*=0 and *tau2*=10μs gave a nicer baseline. This is probably due to imperfect chemical shift refocusing during the effective “spin-echo” timing of the experiment, possibly due to the receiver gating delay before FID acquision (see the *alpha* and *beta* variable definitions in the Varian manuals for more information).

```
#ifndef LINT
#endif

/*  Pulsed field gradient diffusion      */
/*  JP Lapham */

#include <standard.h>

/* define phase cycling */
static int ph1[4] = {0,2,3,1},
   ph2[4] = {2,0,1,3},
   ph3[4] = {1,3,0,2},
   ph4[4] = {3,1,3,0};

pulsesequence()
{
double grt1, grl1, post, grt2, grl2, dt, dt_corr, tau1, tau2;
grt1 = getval("grt1");
grl1 = getval("grl1");
grt2 = getval("grt2");
grl2 = getval("grl2");
post = getval("post");
tau1 = getval("tau1");
tau2 = getval("tau2");
dt = getval("dt");

/* variable calculations */
dt_corr = dt-grt1-post-tau-(4*rof1)-(2*pw);

settable(t1, 4, ph1);
settable(t2, 4, ph2);
settable(t3, 4, ph3);
settable(t4, 4, ph4);

/* Begin Pulse Sequence */

status(A);
   delay(d1);

status(B);
   rgpulse(pw, t1, rof1, rof1);
   delay(tau1);
   rgradient('z',grl1);
   delay(grt1);
   rgradient('z',0.0);
   delay(post);
   rgpulse(pw, t2, rof1, rof1);

status(C);

   delay(dt_corr/2-grt2);

   rgradient('z',grl2);
   delay(grt2);
   rgradient('z',0.0);

   delay(dt_corr/2);


status(D);
   rgpulse(pw, t3, rof1, rof1);
   delay(tau2);
   rgradient('z',grl1);
   delay(grt1);
   rgradient('z',0.0);
   delay(post);

status(E);
   setreceiver(t4);
}
```

4.6.2 Felix95 diffusion processing macro "diffusion.mac"

The diffusion data processed using the Felix95 (Biosym Inc.) software package. This Felix95 macro was written to perform the repetitive tasks required for processing the data. The unique feature of this macro is that it outputs to a file a list of the measured integrated areas of 1D peaks, using the "dba element load" statement. This is especially nice because the end user need not actually type in large integration data sets. The output of this macro, called a ".xy" file represents the normalized (all integration data is divided by the first value) integrated values for the experiment. This .xy file is then further processed using the xy2xm script (see 5.6.3).

```
c** This macro can be used to process diffusion data into
c** XMGR able
c** format.  Read the first fid into felix, integrate an area.
c** Then run this macro.  Have fun!
c** -JPL 3/28/96

c** Name of the data file
get 'filename?' file

c** number of experiments
def nexp 31

c** window functions
def wind1 'cnv 0 32'
def wind2 'sb 512 90'

c** phasing
def phase0 118.6
def phase1 0

cl

c** remove any previous .xy files
sys rm &file.xy

c** throw out first data point
c** b/c you have to have some gradient for good data
re &file.dat

for loop 1 &nexp
  re &file.dat
;  bc .1
;  &wind1
;  &wind2
  ft
  ph
  pol 1

  dr
  dba element load seg:segments.1.volume int
  ty Integrated area for exp# &loop: &int $
  sys echo &loop &int >> &file.xy

  esc escape
  if &escape eq 1 escape
```

```
next
```

4.6.3 xy2xm - process diffusion data integration values

This PERL script reads in the output from the Felix95 macro, diffusion.mac, and returns a two column list. The first column is calculated by $\gamma^2\delta^2G_z^2(\Delta-\delta/3)$, where the values of G_z are set by the gradient strength. The second column is calculated by –ln(Y/Y_o) where Y_o comes from the first input integration value and Y is each subsequent integration value, this is a normalization routine. Traditionally the post processed file is given the extension of ".xm", this name comes from the idea that it is ready to be read by the data plotting software xmgr.

Syntax: **xy2xm G_{max} δ Δ input_file** > **output_file**

Example: **xy2xm 32 .002 .1 input.xy** > **output.xm**

In this example, the input file input.xy is being processed for data with a δ=2ms and a Δ=100ms, the maximum gradient possible for the probe was 32 g/cm. Note that the value of the gradient maximum is not necessarily the maximum used in the experiment, it is the theoretical maximum for the instrument hardware.

```
#! /usr/local/bin/perl
# Generates plots of diffusion data for xmgr
# The script reads in the output from the diffusion.mac felix95
# macro
# Usage: xy2xm gmax delta DELTA filename.xy > filename.xm
# where delta is the length of the gradient pulse and
# DELTA is the length of the delay between gradient pulses.

if ($ARGV[0] eq "") {
   print "Usage: xy2xm gmax delta DELTA filename.xy > filename.xm\n";
   exit;
   }

$gmax = $ARGV[0]; shift;
$delta = $ARGV[0]; shift;
$DELTA = $ARGV[0]; shift;

# print header for output file (xmgr will ignore)
print "; gmax = $gmax\n";
print "; delta = $delta\n";
print "; DELTA = $DELTA\n";

foreach (<>) {
   ($x,$y) = (split);
   if ($x eq 1) {
      $y_first = $y;
      }
   $y_new = -log($y/$y_first);

   # x_new = (gyromag H)^2 * (small delta)^2 * (grl1*gmax/32767)^2 *
   #         (big delta-(small delta/3))
   $x_new = (2.675197e4)**2 * ($delta)**2 * ($x*$gmax/32.767)**2 *
       ($DELTA-$delta/3);
   print "$x_new $y_new\n";
}
```

4.6.4 xm2ds – perform a quick linear regression on a ".xm" file

This script is included because it is helpful when processing large numbers of ".xm" files (see 4.6.3 for what a .xm file is). It quickly calculates the slope of the line for a x,y data set. Note, however, that the "error" reported is incorrect. This is not intended to replace using a true data plotting and statistical analysis software package, which should be used for final analysis. The author was Bo-Lu Zhou, his first PERL script, written while doing a rotation project with me.

Syntax: **xm2ds < input_file**

```
#! /usr/local/bin/perl
# This script carries out a regression on two columns of data and
# report the value of Ds in (column2= Ds * column1 + y intercept).
# The standard deviation of the residual errors is also reported.

$mod_x =0;
$tran_yx =0;
$tran_bx =0;
$tran_yxortho =0;
$mod_vec_xortho = 0;
$sum2_error=0;

$i=1;
foreach  (<>) {
   ($x_old,$y_old) = (split);
   $x[$i] = $x_old;
   $y[$i] = $y_old;
   $i++;        }
close (info);
$total= --$i;

for ($i=1; $i<=$total; $i++)
{
  $mod_x = $mod_x+$x[$i]*$x[$i];
  $tran_yx = $tran_yx+$y[$i]*$x[$i];
  $tran_bx = $tran_bx+$x[$i]*1;
}

$yx = $tran_yx / $mod_x;
$bx = $tran_bx / $mod_x;

for ($i=1; $i<=$total; $i++)
{
  $vec_xortho[$i] = 1- $bx*$x[$i];
  $tran_yxortho =$tran_yxortho + $y[$i] * $vec_xortho[$i];
  $mod_vec_xortho = $mod_vec_xortho + $vec_xortho[$i]**2;
}

$yxortho= $tran_yxortho / $mod_vec_xortho;
$A = $yx - $yxortho * $bx;
$B = $yxortho;

print "\n";
for ($i=1; $i<=$total; $i++)
{
  $error[$i] = $y[$i] - ($A * $x[$i] + $B);
}

print "                Ds = $A\n";
print "\n";

for ($i=1; $i<=$total; $i++)
{
  $sum2_error = $sum2_error + $error[$i]**2;
}

$std_deviation_error = ($sum2_error / ($total-1))**(0.5);
print "Standard Deviation = $std_deviation_error\n";
print "\n";
```

4.7 References

Aboul-ela F, Nikonowicz EP, Pardi A. 1994. Distinguishing between duplex and hairpin forms of RNA by 15N-1H heteronuclear NMR. *FEBS Lett 347*:261-264.

Altieri AS, Hinton DP, Byrd RA. 1995. Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *JACS 117*:7566-7567.

Andreasson B, Nordenskiöld L, Braunlin WH, Schultz J, Stilbs P. 1993. Localized interaction of the polyamine methylspermidine with double-helical DNA as monitored by 1H NMR self diffusion measurements. *Biochemistry 32*:961-967.

Andrec M, Prestegard JH. 1996. Quantitation of Chemical Exchange Rates Using Pulsed Field-Gradient Diffusion Measurements. *Journal Biomolecular NMR 9*:136-150.

Arnott S, Hukens DWL, Dover SD, Fuller W, Hodgson AR. 1973. Structures of synthetic polynucleotides in the A-RNA and A'-RNA conformations: X-Ray diffraction analysis of the molecular conformations of polyadenylic acid-polyuridylic acid and polyinosine acid-polycytidylic acid. *J Mol Biol 81*:107-122.

Arnott S, Hukins DWL. 1972. *Biochem Biophys Res Comm 47*:1504.

Berne BJ, Pecora R. 1976. *Dynamic Light Scattering*. Wiley, New York.

Broersma S. 1960. Viscous force constant for a cylindrical particle. *J Chem Phys 32*:1632-1635.

Broersma S. 1981. *J Chem Phys 74*:6989.

Bu Z, Russo PS, Tipton DL, Negulescu II. 1994. Self-Diffusion of Rodlike Polymers in Isotropic Solutions. *Macromolecules 27*:6871-6882.

Callaghan PT, Gros MAL, Pinder DN. 1983. The Measurement of Diffusion Using Deuterium Pulsed Field Gradient Nuclear Magnetic Resonance. *J Chem Phys 79*:6372-6381.

Cantor CR, Schimmel PR. 1980. *Part II: Techniques for the study of biological structure and function*. New York: W. H. Freeman and Co.

Carr HY, Purcell EM. 1954. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review 94*:630-638.

Charles S. Johnson J. 1993. Effects of chemical exchange in diffusion-ordered 2D NMR spectra. *J Mag Res Series A 102*:214-218.

Cheong C, Varani G, Jr IT. 1990. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature 346*:680-682.

Chung J, Prestegard J. 1993. Characterization of field-ordered aqueous liquid crystals by NMR diffusion measurements. *JPC 97*:9837-9843.

Dingley AJ, Mackay JP, Chapman BE, Morris MB, Kechel PW, Hambly BD, King GF. 1995. Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2. *J Biomol NMR 6*:321-328.

Doran SJ, Décorps M. 1995. A robust, single shot method for measuring diffusion coefficients using the "burst" method. *J Mag Res Series A 117*:311-316.

Eimer W, Pecora R. 1991. Rotational and translational diffusion of short rodlike molecules in solution: oligonucleotides. *J Chem Phys 94*:2324-2329.

Eimer W, Williamson JR, Boxer SG, Pecora R. 1990. Characterization of the overall and internal dynamics of short oligonucleotides by depolarized dynamic light scattering and NMR relaxation measurements. *Biochemistry 29*:799-811.

Elias JG, Eden D. 1981. *Biopolymers 20*:2368.

Fujimoto BS, Miller JM, Ribeiro NS, Schurr JM. 1994. Effects of different cations on the hydrodynamic radius of DNA. *Biophysical Journal 67*:304-308.

Gibbs SJ, Charles S. Johnson J. 1991. A PFG NMR Experiment for Accurate Diffusion and Flow Studies in the Presence of Eddy Currents. *J Mag Res 93*:395-402.

Griffiths L, Horton R. 1990. NMR Diffusion Measurements Using Refocused three-pulse stimulated echoes. *J Mag Res 90*:254-263.

Hahn EL. 1950. Spin echoes. *Physical Review 80*:580-594.

Harnet HS, Owen BB. 1958 (pp236-242). *The physical chemistry of electrolytic solutions*. New York: Reinhold Publishing Co.

Hearst JE. 1963. Rotatory diffusion constants for stiff-chain macromolecules. *J Chem Phys 38*:1062-1065.

Hervet H, Leger L, Rondelez F. 1979. Self-diffusion in polymer solutions: a test for scaling and reptation. *Phys Rev Lett 42*:1681-1684.

Heus HA, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science 253*:191-194.

Kellomaki A. 1975. Viscosities of H_2O and D_2O mixtures at various temperatures. *Finn Chem Lett* 51-54.

Klein J, Fletcher D, Fetters LJ. 1983. Diffusional behavior of entangled star polymers. *Nature 304*:526-527.

Kriwacki RW, Hill RB, Flanagan JM, Caradonna JP, Prestegard JH. 1993. New NMR methods for the characterization of bound waters in macromolecules. *JACS 115*:8907-8911.

Kuchel PW, Chapman BE. 1991. Translational diffusion of hemoglobin in human erythrocytes and hemolysates. *JMR 94*:574-580.

Kuchel PW, Chapman BE. 1993. Heteronuclear double-quantum-coherence selection with magnetic-field gradients in diffusion experiments. *JMR 101*:53-59.

Lanni F, Ware BR. 1982. *Rev Sci Instrum 53*:905-908.

Longsworth LG. 1960. The mutual diffusion of light and heavy water. *J Phys Chem 64*:1914-1917.

Mandelkern M, Elias JG, Eden D, Crothers DM. 1981. The dimensions of DNA in solution. *JMB 152*:153-161.

Marky LA, Breslauer KJ. 1987. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers 26*:1601-1620.

Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. 1987. Oligoribonucleotide Synthesis using T7 RNA Polymerase and Synthetic DNA Templates. *NAR 15*:8783-8798.

Natarajan G. 1989. *Data book on the viscosity of liquids*. Hemisphere Publishing Co.

Pecora R. 1991. DNA: a model compound for solution studies of macromolecules. *Science 251*:893-898.

Price WS, Kuchel PW. 1991. Effect of nonrectangular field gradient pulses in the Stejskal and Tanner (diffusion) pulse sequence. *JMR 94*:133-139.

Sen D, Gilbert W. 1992. Novel DNA superstructures formed by telomere-like oligomers. *Biochemistry 31*:65-70.

Sich C, Flemming J, Ramachandran R, Brown LR. 1996. Distinguishing Inter- and Intrastrand NOEs Involving Exchangeable Protons in RNA Duplexes. *J Mag Res Series B 112*:275-281.

Stejskal EO, Tanner JE. 1964. Spin diffusion measurements: spin echoes in the presence of a time dependent field gradient. *J Chem Phys 42*:288-292.

Tanner JE. 1970. Use of Stimulated Echo in NMR Diffusion Studies. *J Chemical Physics 52*:2523-2526.

Tirado MM, Martinez CL, Torre JGdl. 1984. Comparison of theories for the translational and rotational diffusion coefficients of rod-like macromolecules. Application to short DNA fragments. *J Chem Phys 81*:2047-2052.

Tirado MM, Torre JGdl. 1979. Translational friction coefficient of rigid, symmetric top macromolecules. Application to circular cylinders. *J Chem Phys 71*:2581-2587.

Tirado MM, Torre JGdl. 1980. Rotational dynamics of rigid, symmetric top macromolecules. Application to circular cylinders. *J Chem Phys 73*:1986-1993.

Torre JGdl, Martinez MCL, Tirado MM. 1984. Dimensions of short, rodlike macromolecules from translational and rotational diffusion coefficients. Study of the gramicidin dimer. *Biopolymers 23*:611-615.

Tracy MA, Pecora R. 1992. Dynamics of Rigid and Semirigid Rodlike Polymers. *Annu Rev Phys Chem 43*:525-557.

Varani G, Cheong C, Tinoco I. 1991. Structure of an unusually stable RNA hairpin. *Biochemistry 30*:3280-3289.

CHAPTER 5 “NMR HOMONUCLEAR DIPOLAR RELAXATION THEORY: ANISOTROPIC MOLECULAR TUMBLING”

5.1 Summary

This chapter is an introduction for chapters 6-7, both of which utilize the theories discussed here. Presented is a treatment of nuclear magnetic resonance (NMR) dipolar relaxation theory for homonuclear interactions for both isotropic and anisotropic molecular tumbling. Methods are also presented for using these theories to simulate the nuclear Overhauser effect (NOE).

5.2 Introduction

NMR is a powerful spectroscopic technique for studying molecular systems. Data from NMR spectroscopy can give a wealth of information about structure and dynamics. In recent years, this technique has become a useful tool for molecular biochemists in determining the three dimensional structures of biomolecules, such as proteins and nucleic acids.

Measurement of NOE in NMR spectroscopy can be used to calculate the distances between nuclei to determine the three-dimensional structures of molecules. NOE is derived from through-space dipolar relaxation that is induced by time-dependent fluctuating magnetic fields and is dependent on the distance between two dipoles. For solution-state NMR, these fluctuations mainly result from molecular rotational diffusion spinning the nuclear dipole moments that have aligned with a strong external B_0 magnetic field. The current theories and practices in biomolecular structure determination often make the assumption that an isotropic rotation model can adequately describe this molecular rotational diffusion. This assumption is not valid for extended shape biomolecules such as long DNAs, which are better described as having two rotational diffusion rates, one for the long axis and one for the short axis.

The effect of the anisotropic rotation is examined in this chapter theoretically in terms of the effect on the NOE and its interpretation in distance calculations.

5.3 Homonuclear NMR relaxation theory

As mentioned earlier, the NOE is a through-space dipolar relaxation process between magnetically active nuclei. The theories behind NMR relaxation (Abragam & Pound, 1953; Solomon, 1955) will be developed in this section for the special case of two rigid, isolated spins. The concept rate matrix treatment will be decribed, which allows for later application of these theories to multi-spin systems, with coupled relaxation properties.

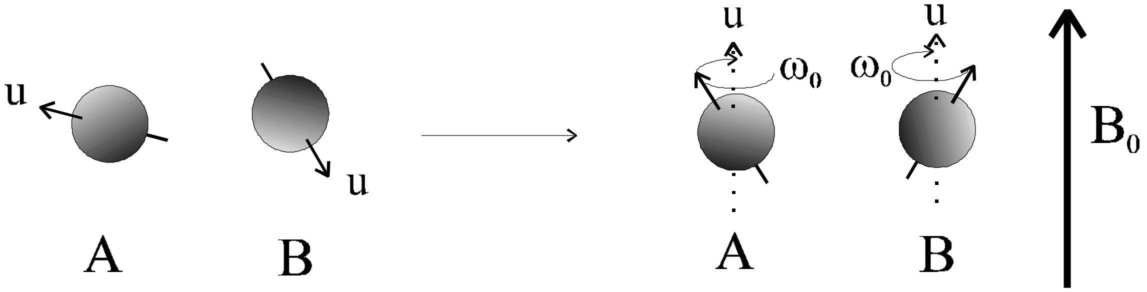


Figure 5. 1 Two magnetic nuclei placed in an external B_0 field

Consider two nuclei (as shown in figure 5.1), A and B, that have an inherent magnetic dipole moment **u**, which have been placed in a strong external magnetic field B_0. The magnetic dipole moment will precess under the torque induced by the B_0 field at the nuclei's characteristic Larmor frequency given by $\omega_0 = -\gamma B_0$ (γ is the gyromagnetic ratio for that spin, with a value of 26.7520×10^7 rad T^{-1} s^{-1} for proton). The net magnetic moment of the precession will lie parallel to the B_0 field (defined as the z-axis) and is represented by the dotted arrow in figure 5.1. The energy for the interaction between the magnetic dipole and the external magnetic field is given by $E(m) = -gB_0\hbar\, m$, where the allowable values for the quantum number *m* are *+I, (+I-1),… , (-I+1), -I*. For spins with a quantum number *I*= ½ (such as the biologically relevant ^{1}H, ^{13}C and ^{15}N nuclei), *m*= +½ or –½ which gives the allowable energies for the spin to be proportional to $\pm\frac{1}{2}\hbar$. These energies are abbreviated as α and β, respectively.

Thus, the energy of the two spin system can be described as one of four possible energy states (αα, αβ, βα or ββ) and the changes in energy of the system are described by the energy diagram shown in figure 5.2 below.

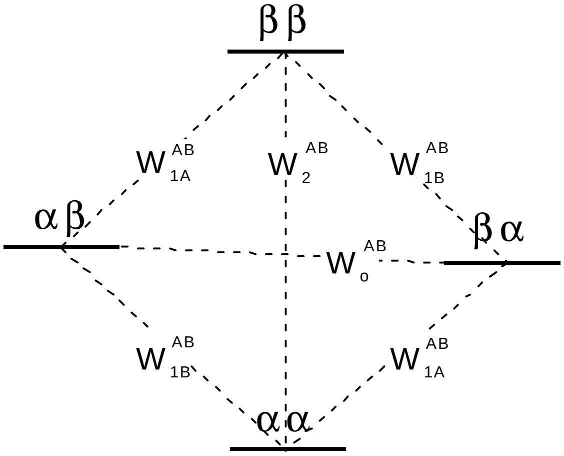


Figure 5. 2 Energy diagram for two nuclei of spin ½

It is these transitions between energy states that give rise to all of NMR relaxation theory. The rate of a transition occurring between any of the above energy states is given by the function *W*, as shown in figure 5.2. The phenomenon of dipolar relaxation occurs because of time-dependent fluctuations in the magnetic field surrounding a nucleus. These fluctuations can arise from a number of molecular properties, such as molecular tumbling in solution, dynamical motions between nuclei or librational atomic motions. For the simple case of a rigid two-spin system, the magnetic field fluctuations are completely described by the molecular tumbling, as shown in the figure below.

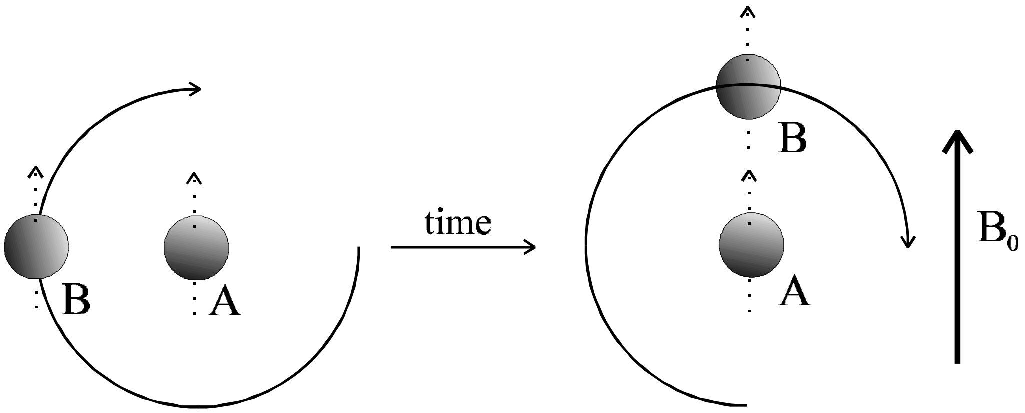


Figure 5. 3 Time-dependent magnetic field fluctuations due to molecular rotation

A complete description of NMR dipolar relaxation thus requires an accurate mathematical description of the frequencies of these magnetic field fluctuations. The frequency domain function used for this purpose is known as the "spectral density function" and can be derived from Brownian motion theory for particles.

5.3.1 The spectral density function for isotropic rotation

Bloembergen, Purcell and Pound (1948) first described the spectral density function for isotropic motion. They used a model of a randomly orienting internuclear vector that is attached to a sphere undergoing isotropic rotational diffusion in a continuous medium. This model is similar to that developed by Debye (1929) for dielectric relaxation. Full mathematical treatments on the derivation of this spectral density function have been well reviewed (Solomon, 1955; Ernst, *et al.*, 1987; Hennel & Klinowski, 1993; Schmidt-Rohr & Spiess, 1994) and will not be presented here. It is the intention of this section to give a qualitative description of the concepts involved in molecular orientation mathematics.

The mathematical description of isotropic molecular reorientation involves three functions, position, *f(t)*, correlation, *g(t)*, and spectral density *J(*W*)*. A graphical representation of these three functions is shown below,

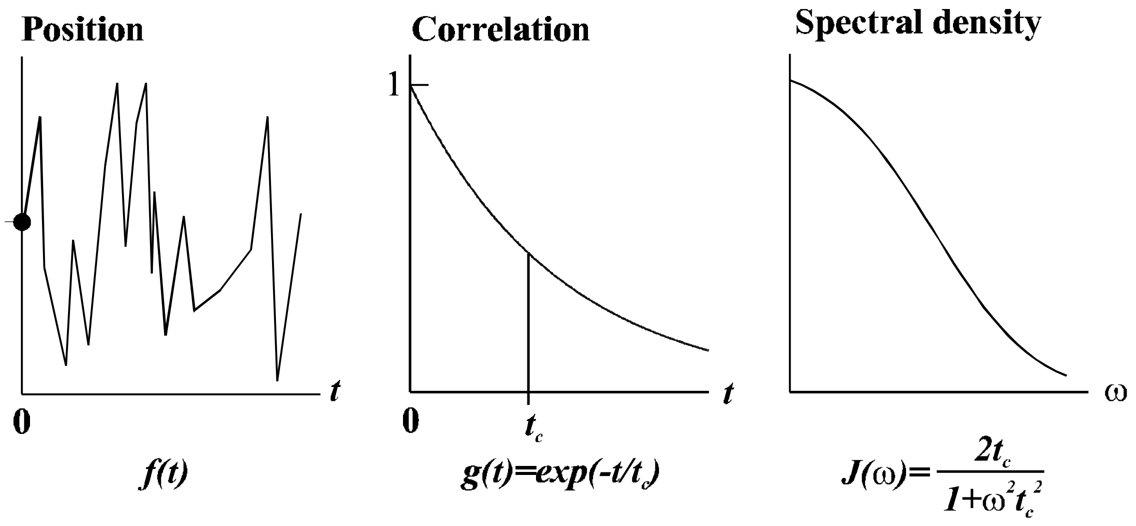


Figure 5. 4 Functions of molecular reorientation

The position function is simply a measure of motional movement as a function of time for a *single* particle in a molecule. If this motion is due to Brownian thermal movements, as shown above, *f(t)* will appear to be random. However, the ensemble average of the position of many particles will be described by an exponential decay,

$$g(t) = \left\langle f(0)\, f(t) \right\rangle = \exp(-t / t_c)\,. \qquad 5.1$$

This ensemble average is called the correlation function, with the time constant for the decay, t_c, defined as the "correlation time". *g(t)* is a probability function that describes the chances of finding a particle near the original position, *f(0)*, in the ensemble, and has been described as a measure of the 'position memory' of a particle. With increasing time this probability diminishes as the positions of particles become less correlated to their starting positions *f(0)*.

The spectral density function is the frequency domain representation of this correlation function, and as such they are mathematically interconvertable by the Fourier transform, *FT*,

$$J(w) = FT\left(g(t)\right) = FT\left(\exp(-t/t_c)\right)$$

$$= \int_0^\infty \exp(-t/t_C)\cos(wt_c)dt = \frac{2t_c}{1+w^2t_c^2} \qquad 5.2$$

5.3.2 Transition rates

With a mathematical definition of the spectral density function, the energy state transition rates can be expressed as functions of J_{AB} by the Solomon equations (Solomon, 1955) (assuming $w_{0A}=w_{0B}\equiv w$, a homonuclear interaction),

$$W_{1A}^{AB} = \frac{3}{4}q_{AB}J_{AB}(w_{0A}) = \frac{3}{4}q_{AB}J_{AB}(w)\,, \qquad 5.3$$

$$W_{1B}^{AB} = \frac{3}{4}q_{AB}J_{AB}(w_{0B}) = \frac{3}{4}q_{AB}J_{AB}(w)\,, \qquad 5.4$$

$$W_{0}^{AB} = \frac{1}{2}q_{AB}J_{AB}(w_{0A}-w_{0B}) = \frac{1}{2}q_{AB}J_{AB}(0)\,, \qquad 5.5$$

$$W_{2}^{AB} = 3q_{AB}J_{AB}(w_{0A}+w_{0B}) = 3q_{AB}J_{AB}(2w)\,. \qquad 5.6$$

The derivation of these expressions comes from perturbation theory and the Hamiltonian of the motion of the particles. Notice that the Solomon equations include a "rate" constant q_{AB}, which is derived from the coulombic interaction of two dipoles, and is defined as,

$$q_{AB} = \frac{1}{10}\gamma_A^2\gamma_B^2 h^2 r_{AB}^{-6}\left[\frac{\mu_0}{4\pi}\right]^2 . \qquad 5.7$$

The r^{-6} term assumes that there are no distance fluctuations between the nuclei **AB**. If fluctuations do exist, then a more complex definition of *r* would be contained in the spectral density function J_{AB}. The rate constant is often conveniently represented as $56.9{\cdot}r^{-6}$ (in units of s^{-1} ns^{-1} $Å^{-6}$) (note the s^{-1} and ns^{-1} component, see appendix 5.7.1 for the derivation and dimensional analysis).

The time dependent change of population (*dN*) of any of the four energy states shown in figure 5.2 can be calculated by multiplying the appropriate population (*N*) by the transition rate, which can be positive or negative depending on whether it is adding or removing magnetization. This is shown by the following equations,

$$\frac{dN_{\beta\beta}}{dt} = -(W_{1A}^{AB} + W_2^{AB} + W_{1B}^{AB})N_{\beta\beta} + W_{1A}^{AB}N_{\alpha\beta} + W_2^{AB}N_{\alpha\alpha} + W_{1B}^{AB}N_{\beta\alpha} , \qquad 5.8$$

$$\frac{dN_{\alpha\beta}}{dt} = -(W_{1A}^{AB} + W_0^{AB} + W_{1B}^{AB})N_{\alpha\beta} + W_{1A}^{AB}N_{\beta\beta} + W_0^{AB}N_{\beta\alpha} + W_{1B}^{AB}N_{\alpha\alpha} , \qquad 5.9$$

$$\frac{dN_{\beta\alpha}}{dt} = -(W_{1B}^{AB} + W_0^{AB} + W_{1A}^{AB})N_{\beta\alpha} + W_{1B}^{AB}N_{\beta\beta} + W_0^{AB}N_{\alpha\beta} + W_{1A}^{AB}N_{\alpha\alpha} , \qquad 5.10$$

$$\frac{dN_{\alpha\alpha}}{dt} = -(W_{1B}^{AB} + W_2^{AB} + W_{1A}^{AB})N_{\alpha\alpha} + W_{1B}^{AB}N_{\alpha\beta} + W_2^{AB}N_{\beta\beta} + W_{1A}^{AB}N_{\beta\alpha} , \qquad 5.11$$

The experimentally observable magnetization, I_z, will be the difference between the populations of the α and β energies for each spin **A** or **B**,

$$I_{z,A}K = (N_{\alpha\alpha} + N_{\alpha\beta}) - (N_{\beta\alpha} + N_{\beta\beta}) , \qquad 5.12$$

$$I_{z,B}K = (N_{\alpha\alpha} + N_{\beta\alpha}) - (N_{\alpha\beta} + N_{\beta\beta}) . \qquad 5.13$$

K is a normalization constant. Substitution of equations 5.8-5.11 into 5.12 and 5.13 gives (see appendix 5.7.4 for the algebra) equations 5.14 and 5.15,

$$K\frac{dI_{z,A}}{dt} = -\left(W_2^{AB} + 2W_{1B}^{AB} + W_0^{AB}\right)I_{z,A} + \left(W_0^{AB} - W_2^{AB}\right)I_{z,B}\,. \tag{5.14}$$

$$K\frac{dI_{z,B}}{dt} = -\left(W_2^{AB} + 2W_{1A}^{AB} + W_0^{AB}\right)I_{z,B} + \left(W_0^{AB} - W_2^{AB}\right)I_{z,A}\,. \tag{5.15}$$

These equations show that for spin A, magnetization is taken away at a rate of -(W_2+2W_{1B}+W_0) and is transferred to spin B at a rate of (W_0-W_2). These values have been given special names and symbols,

$$\mathrm{r}_{A} = -\left(W_0^{AB} + 2W_{1B}^{AB} + W_2^{AB}\right), \tag{5.16}$$

$$\mathrm{r}_{B} = -\left(W_0^{AB} + 2W_{1A}^{AB} + W_2^{AB}\right), \tag{5.17}$$

$$\mathrm{s}_{AB} = \mathrm{s}_{BA} = W_0^{AB} - W_2^{AB}\,. \tag{5.18}$$

Where r is denoted the "spin-lattice relaxation" rate and s is denoted the "cross-relaxation" rate. Mutual energy coupling between the two spins occurs in the cross-relaxation, or "spin-flip" transitions (see figure 5.2). It is this cross-relaxation term that gives rise to the nuclear Overhauser effect. Graphical representation of these relaxation parameters between spins A and B is shown below.

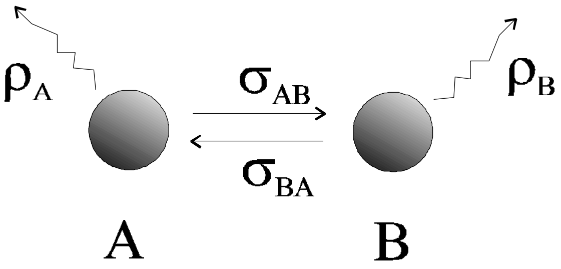


Figure 5. 5 The NMR relaxation parameters s and r

The cross-relaxation rate, s_{AB}, can now be expressed in term of the spectral density function. Notice that $\mathrm{w}_{oA} \approx \mathrm{w}_{oB}$ for nuclei of the same element (protons, for instance) to give,

$$\mathrm{s}_{AB} = \frac{56.9}{r^6}\left[J(0) - 6J(2\mathrm{v}\ )\right]. \qquad 5.19$$

Expansion of this equation with the definition of the isotropic rotation spectral density function, eq. 5.2, gives,

$$\mathrm{s}_{AB} = \frac{56.9}{r^6}\left[2t_c^{AB} - \frac{12t_c^{AB}}{1+(2\mathrm{v}\,t_c^{AB})^2}\right]. \qquad 5.20$$

This is then a complete description of the cross-relaxation rate of any rigid, isotropically rotating spin pair. Often this equation simplified further by making the assumption that we will only consider large, slowly rotating molecules (t_c >> *1/2*w, the slow motion limit), which causes the *6J(2*w*)* term to approach zero, leaving,

$$\mathrm{s}_{AB} = -\frac{56.9}{r^6}\left[2t_c^{AB}\right]. \qquad 5.21$$

However, a plot of the transition probability functions with increasing t_c values (as can be seen in figure 5.6) demonstrates that this assumption may not be completely valid for the size biomolecules (with t_c between 1 and 10 ns) studied here. There may be a significant contribution to the cross-relaxation from the W_2 transition, and thus eq. 5.20 is the preferred definition of s_{AB}.

Similarly, the value of r_A can be expanded in terms of this spectral density function by combining equation 5.16 with 5.3, 5.4, 5.5, 5.6 and 5.7,

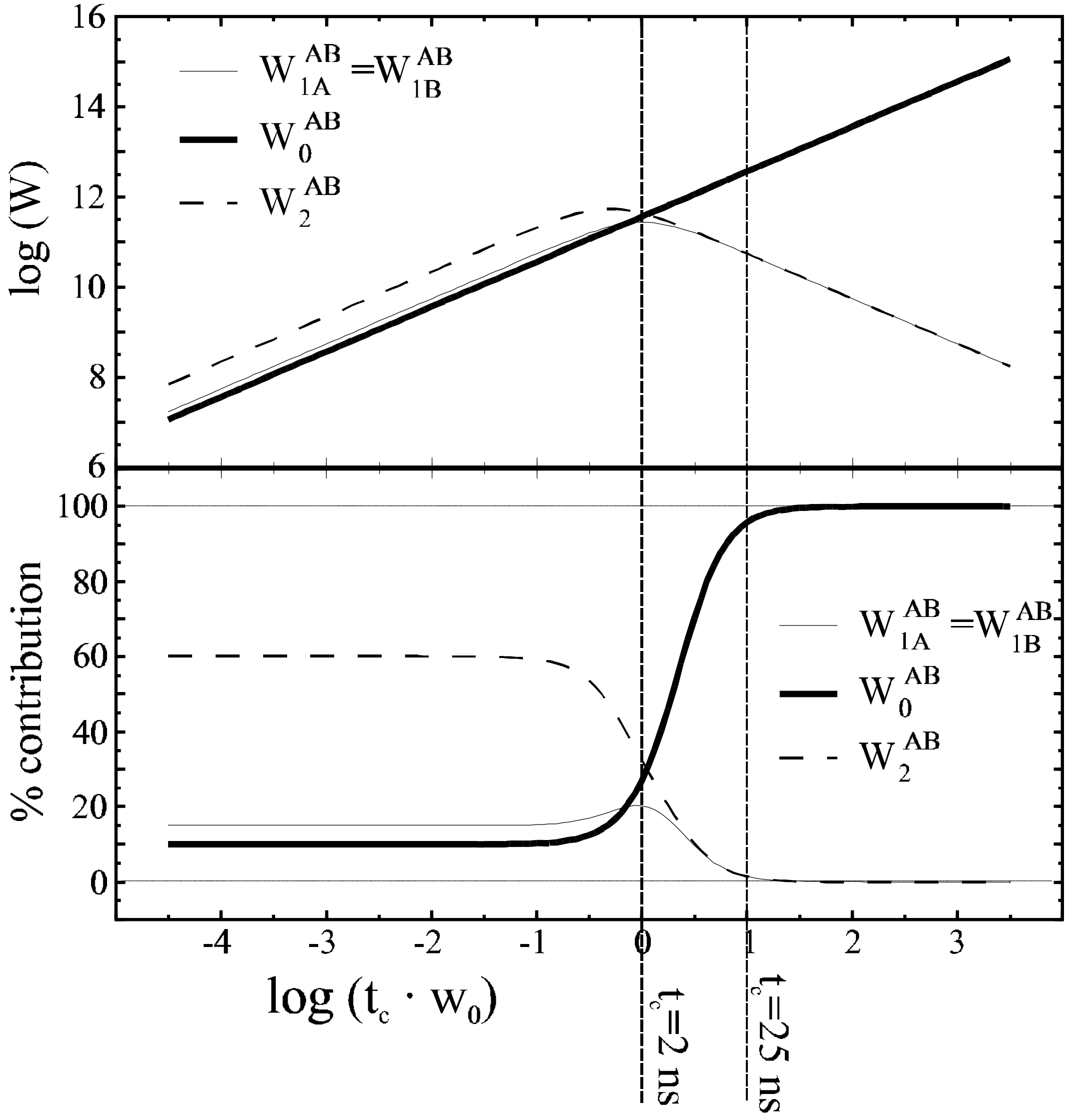


Figure 5.6 Two-spin energy transition rates

Transition rates for the energy diagram (fig 5.2) have been calculated using *W_0=0.5qJ(0), W_1=3qJ(2*w*$_0$)* and *W_2=0.75qJ(*w*$_0$)*, substituting the spectral density function for isotropic rotation. The script "W.pl" (see Chapter 7) was written for this purpose. A range of values for *t_c* were plotted, assuming a 500 MHz NMR spectrometer (w*$_0$* = 500x10^6 s^{-1}). The top graph is the actual values for each *W* function, while the lower graph shows the percentage contribution of each transition rate. For the size DNA molecules used in these studies, the *t_c* values calculated ranged from 2 to 25 ns, and is represented by the vertical dotted lines, notice that for cross-relaxation, (*W_0-W_2*) one cannot make the assumption that the *W_2* term is negligible.

$$
\begin{aligned}
\mathrm{r}_A &= -\left(\frac{1}{2}q_{AB}J_{AB}(0) + 2\left(\frac{3}{4}q_{AB}J_{AB}(\mathrm{w})\right) + 3q_{AB}J_{AB}(2\mathrm{w})\right),\\
&= -5q_{AB}\left[J_{AB}(0) + J_{AB}(\mathrm{w}) + J_{AB}(2\mathrm{w})\right],\\
&= -5\cdot\frac{56.9}{r^6}\left[2t_c + \frac{2t_c}{1+\mathrm{w}^2t_c^2} + \frac{2t_c}{1+(2\mathrm{w})^2t_c^2}\right]. \qquad 5.22
\end{aligned}
$$

5.3.3 The relaxation rate matrix

These two relaxation processes can be followed with the use of a 2x2 "relaxation" or "rate" matrix, **R**, of form (Keepers & James, 1984; Ernst, *et al*., 1987) (see appendix 5.5.2 for the derivation of the rate matrix from chemical exchange theory),

$$
\mathbf{R} = \begin{vmatrix} \mathrm{r}_A & \mathrm{s}_{AB} \\ \mathrm{s}_{BA} & \mathrm{r}_B \end{vmatrix}. \qquad 5.23
$$

The advantage of using the rate matrix for representing the relaxation processes is that it offers a convenient method of multiple (more than two) spin coupled relaxation. Dipolar relaxation in NMR often involve many spins that are in close proximity to each other, as shown below in figure 5.7 for the H6-H2' protons in A-form RNA.

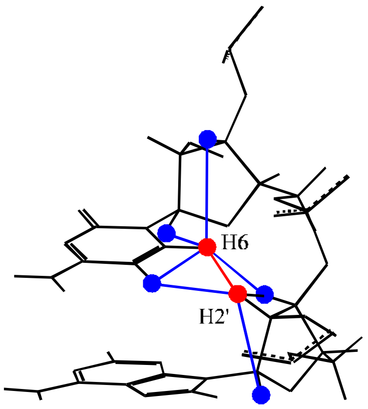


Figure 5.7 Multiple spin coupling in nucleic acids

The H6-H2' distance is important in nucleic acid structure determination because it is one of the few distance restraints which interconnects adjacent nucleotides in standard helical regions of the structure. Besides the two protons of interest, there are five other protons within 3.5Å of the pair. It is important that the distance calculations used to determine RNA structure take into account these multiple spin partners.

The rate matrix is a general method of describing any number of coupled relaxation rate processes, and as such it can be expanded to include more spins. The expanded rate matrix allow for all these additional rate processes to be accounted for simultaneously, and is given by,

$$\mathbf{R}(NxN) = \begin{vmatrix} \mathsf{r}_{1,1} & \mathsf{s}_{2,1} & \dots & \mathsf{s}_{N,1} \\ \mathsf{s}_{1,2} & \mathsf{r}_{2,2} & \dots & \mathsf{s}_{N,2} \\ \dots & \dots & \dots & \dots \\ \mathsf{s}_{1,N} & \mathsf{s}_{2,N} & \dots & \mathsf{r}_{N,N} \end{vmatrix}, \qquad 5.24$$

with the more general definitions for r and s,

$$\mathsf{r}_{i,i} = -\sum_{j(i\neq j)} (W_0^{i,j} + 2W_1^{i,j} + W_2^{i,j}), \qquad 5.25$$

$$\mathsf{s}_{i,j} = W_0^{i,j} - W_2^{i,j}. \qquad 5.26$$

The T1, or longitudinal, relaxation time is a measure of the rate at which the z component of the magnetization returns to the equilibrium state. It can be calculated from these relaxation matrix parameters and goes as the inverse of the sum of the r with all possible s rates.

$$\frac{1}{T1} = \mathsf{r}_i + \sum_{j(i\neq j)} \mathsf{s}_{ij} \qquad 5.27$$

which gives,

$$\frac{1}{T1} = -2\sum_{j(i\neq j)}\left(W_1^{i,j} + W_2^{i,j}\right) \qquad 5.28$$

5.4 Measured NOE volumes and the relaxation matrix

Measurement of the homonuclear relaxation matrix by NMR comes from the interpretation of the volume intensities from NOESY experiments. In fact, the volume matrix (**V**) is fundamentally related to the relaxation matrix in that they are of the same dimension (both are NxN with N equal to the number of protons in the molecule). The diagonal elements of the relaxation matrix correspond to the autopeaks of the volume matrix, and the off-diagonal elements of the relaxation matrix correspond to the crosspeaks of the volume matrix. The two are related by the following equation,

$$\mathbf{V}(t_{mix}) = \mathbf{V}(0)\exp[\mathbf{R}t_{mix}]\,. \qquad 5.29$$

With V(0) being the intensities of the autopeaks at a mixing time of 0.

The power of the relaxation matrix **R** comes from the fact that it offers a way of calculating the intensities of a NOESY spectrum by simultaneously solving all the relaxation rate equations for every nucleus. If one assumes that the rate matrix truly represents all the relaxation properties of the system, it is theoretically possible to back-calculate NOE intensities from a molecular structure model.

5.4.1 Mathematic considerations

As discussed previously, the intensities of all resonances in the NOESY spectrum are represented by an *NxN* matrix $\mathbf{V}(t_m)$, with the intensity of the autopeaks as the $\mathbf{V}_{i,i}$ elements and the crosspeaks as the $\mathbf{V}_{i,j(i\neq j)}$ elements. The zero-time intensity matrix **V**(0) is a matrix with zero value off-diagonal terms and diagonal terms which represent the intensity of the autopeaks at a $t_{mix} = 0$.

This matrix equation can be solved by diagonalizing the rate matrix **R** to determine the eigenvalue matrix **Λ**, and the corresponding eigenvector matrix **X** (see appendix 5.6.3 on solving simultaneous rate equations). This leads to the following,

$$\mathbf{V}(t_{mix}) = \mathbf{V}(0)\mathbf{X} \cdot \exp(-\Lambda t_{mix}) \cdot \mathbf{X}^{-1}, \qquad 5.30$$

that can be used to directly calculate the intensity matrix **V**.

It is apparent that this "relaxation matrix" method of predicting NOE volumes will only be as successful as the model used in building the relaxation rate matrix **R**. It is in this matrix that any and all assumptions made about the relaxation processes of the system are placed. In fact, as discussed previously, any assumptions in the relaxation theory arise from the model used to build the spectral density function.

Thus far it has been assumed that isotropic motion can adequately describe the rotational diffusion of the molecule. This assumption, however, is not true for molecules with extended hydrodynamical shapes such as DNA. The rotational dynamics of these molecules cannot be accurately described using the isotropic definition of the spectral density function.

5.5 Anisotropic molecular tumbling

A molecule undergoing anisotropic molecular tumbling, such as a long thin cylinder, will actually have two correlation times describing its motion; one about the short axis of rotation (t_s) and one about the long axis (t_l) of rotation. Unlike the isotropic dipolar interactions, the effective correlation time any particular pair of nuclei experience will be dependent on the angle they make with respect to the principal axis of rotation.

If, for instance, an isolated pair of nuclei form a vector that lies parallel to the principal axis of rotation, the dipolar interactions they experience will be independent of the rotation about the principal axis and will be described by the short axis rotation. However, an isolated pair of nuclei that form an interaction vector that lies perpendicular to the principal axis of rotation will experience some geometric mean of the long and short axis rotation.

This angular dependence of the correlation time for a pair of dipoles in an anisotropically rotating molecule must be represented in the definition of the spectral density function.

5.5.1 The spectral density function for anisotropic rotation

Woessner (1962) derived the spectral density function for an anisotropically rotating molecule. The derivation will not be presented here, as it is rather lengthy. This is a summation of the results,

$$J(\mathrm{w}) = a_1 J(\mathrm{w},\mathrm{t}_1) + a_2 J(\mathrm{w},\mathrm{t}_2) + a_3 J(\mathrm{w},\mathrm{t}_3), \qquad 5.31$$

where,

$$J(\mathrm{w},\mathrm{t}) = \mathrm{t}/(1 + \mathrm{w}^2\mathrm{t}^2) \qquad 5.32$$

and the amplitudes, a_i are given by,

$$a_1=0.25\ (3\ cos^2\mathrm{b} - 1)^2 \qquad 5.33$$

$$a_2=3\ cos^2\mathrm{b}\ sin^2\mathrm{b} \qquad 5.34$$

$$a_3=0.75\ sin^4\mathrm{b}. \qquad 5.35$$

The angle b is the angle the **AB** vector makes with the principal axis (see appendix 5.7.2 for a discussion of finding this axis vector by calculating the inertia tensor) of the molecule. The correlation times $\tau_{1,2,3}$ are composite correlation times defined by,

$$t_{1} = t_{L} \qquad 5.36$$

$$t_{2} = 6t_{L}t_{S}/(t_{L} + 5t_{S}) \qquad 5.37$$

$$t_{3} = 3t_{L}t_{S}/(2t_{L} + t_{S}) \qquad 5.38$$

This new spectral density function can then be used in place of the isotropic definition to give new equations for the spin-lattice (s_{AB})and cross-relaxation (r_{A}) rates for the elements of the relaxation rate matrix **R**,

$$\begin{aligned} s_{AB} &= 56.9 r_{AB}^{-6}[6a_1 J(2\mathrm{w}, t_1) + 6a_2 J(2\mathrm{w}, t_2) + 6a_3 J(2\mathrm{w}, t_3)] \\ &\quad - [a_1 J(0, t_1) + a_2 J(0, t_2) + a_3 J(0, t_3)]) \,, \qquad 5.39 \end{aligned}$$

$$\begin{aligned} r_{A} &= \sum_{j(i \neq j)} 5 \cdot 56.9 r_{i,j}^{-6} \big(a_1 J(0, t_1) + a_2 J(0, t_2) + a_3 J(0, t_3)\big) + \\ &\quad \big(a_1 J(\mathrm{w}, t_1) + a_2 J(\mathrm{w}, t_2) + a_3 J(\mathrm{w}, t_3)\big) + \\ &\quad \big(a_1 J(2\mathrm{w}, t_1) + a_2 J(2\mathrm{w}, t_2) + a_3 J(2\mathrm{w}, t_3)\big) \,. \qquad 5.40 \end{aligned}$$

5.6 Discussion

The NMR theory for understanding homonuclear dipolar relaxation has been presented in this chapter. The rotational tumbling rate of a molecule is an important component of this dipolar relaxation process, as it is the principal mechanism that induces the fluctuating magnetic fields responsible for dipolar relaxation. An accurate description of the rotational motion of a molecule is thus necessary in order to interpret any experimental manifestations of the dipolar relaxation.

The NOE is an important probe of molecular structure because the intensity of the NOE is related to the spatial proximity between the two nuclei. The NOE arises as a consequence of dipolar relaxation, and as such, it is important for the interpretation of NOE data that the molecular tumbling of the molecule be understood. A measured NOE between two nuclei can only be interpreted as a distance restraint in the context of a rotational dynamics model. For nucleic acids, a description of this rotational dynamics as isotropic may not be adequate, and the spectral density function proposed by Woessner (1962) is preferred.

This chapter is presented as the theoretical basis for the next few chapters, which will discuss the use of these theories in the simulation of NOE intensities from structural and dynamical models as well as their use in methods of structural refinement.

5.7 References

Abragam A. 1961. *Principles of nuclear magnetism*. Oxford: Clarendon Press.

Abragam A, Pound RV. 1953. Influence of Electric and Magnetic Fields on Angular Correlations. *Physical Review 92*:943-962.

Bloembergen N, Purcell EM, Pound RV. 1948. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Physical Review 73*:679-712.

Debye P. 1929. *Polar Molecules*. New York: Dover.

Eigen M, DeMaeyer L. 1963. "Relaxation Methods". In: Friess, Lewis, Weisberger, eds. *Technique of Organic Chemistry*. New York: Interscience. pp. 895-1051.

Ernst RR, Bodenhausen G, Wokaun A. 1987. *Principles of nuclear magnetic resonance in one and two dimensions*. Oxford: Oxford University Press.

Hennel, Klinowski. 1993.

Keepers JW, James TL. 1984. A theoretical study of distance determinations from NMR. Two-dimensional nuclear overhauser effect spectra. *J. Mag. Res. 57*:404-426.

Marion JB, Thornton ST. 1965. *Classical Dynamics of Particles & Systems*. New York: Academic Press.

Schmidt-Rohr K, Spiess HW. 1994. *Multidimensional Solid-State NMR and Polymers*. New York: Harcourt Brace.

Solomon I. 1955. Relaxation Processes in a System of Two Spins. *Physical Review 99*:559-565.

Woessner DE. 1962. Nuclear spin relaxation in ellipsoids undergoing rotational brownian motion. *J. Chem. Phys. 37*:647-654.

5.8 Appendix

5.8.1 Cross relaxation rate constant calculation:

The cross-relaxation rate constant, *q*, is often represented as the value 56.9, I often wondered from where that number came. Relaxation theory gives us the constant q_{AB}, which can be derived from magnetic point charges,

$$q_{AB} = \frac{1}{10}\frac{\gamma_A^2\gamma_B^2\hbar^2}{r^6}\left(\frac{\mu_o}{4\pi}\right)^2 \tag{5.A.1}$$

with,

J = Joules (kg·m^2·s^{-2})
T = tesla (kg·s^{-2}·A^{-1})
γ_H = gyromagnetic ratio for proton
= 26.7520x10^7 (rad·T^{-1} ·s^{-1} or rad·kg^{-1}·s^2·A·s^{-1})
$\hbar$ = planck's constant
= 6.626208x10^{-34} (J·s)/2π
= 1.054593x10^{-34} (kg·m^2·s^{-2}·s)
μ_o = permeability constant
= 4π·1x10^{-7} (kg·m·s^{-2}·A^{-2})
r = distance between the spins (Å or 1x10^{-10} m)

The number 56.9 is derived,

$$\frac{(26.7519x10^7 rad\cdot kg^{-1}s^2A\cdot s^{-1})^4(1.0545938x10^{-34}kg\cdot m^2s^{-2}s)^2(12.56637x10^{-7}kg\cdot m\cdot s^{-2}A^{-2})^2}{10(4\pi)^2}$$

$$= 5.696x10^{-50}(s^{-2}m^6)$$

$$= 5.69x10^{10}(s^{-2}A^6)$$

$$= 56.9(s^{-1}ns^{-1}A^6)$$

5.8.2 Determining the principal axis: the inertia tensor calculation

Determining the principal axis of rotation for a hydrodynamical particle is of fundamental importance for the calculations involving NMR relaxation of anisotropically rotating molecules. Ideally, it is the diffusion tensor that would give the best measure of this axis. However, the diffusion tensor is quite complicated in that it requires knowledge about the frictional components of the solvent and solute. The inertia tensor, on the other hand, is a simpler calculation and is probably very accurate in predicting the principal axis of rotation in most cases. The inertia tensor requires only knowledge of the structure (or structural model) and masses of the atoms of the molecule in question.

The inertia tensor is an important relation in rotational dynamics. For example, angular momentum (**L**) is related to angular velocity (**w**) by means of the inertia tensor,

$$\mathbf{L} = \{\mathbf{I}\} \cdot \mathbf{w}\,, \tag{5.A.2}$$

and torque (**Γ**) is related to angular acceleration (α) by the inertia tensor,

$$\mathbf{\Gamma}=\{\mathbf{I}\}\cdot\alpha. \tag{5.A.3}$$

In a sense, the inertia tensor relates rotational variables much like mass relates non-rotating variables ($\boldsymbol{P}=m\boldsymbol{v}$ and $\mathbf{F}=m\mathbf{a}$). The inertia tensor is a measure of how much “apparent rotational mass” an object has.

The inertia tensor is a 3x3 matrix in which the nine elements are composed of the X,Y or Z cartesian coordinates of a particle α and the distance from that particle to the center of mass of the object, r_a. (Read chapter 10 of Marion and Thornton’s “Classical Dynamics” book (1965) if you are interested in the derivation of the equations for the inertia tensor).

$$\{\mathbf{I}\} = \begin{vmatrix} \sum_a m_a (r_a^2 - x_a^2) & -\sum_a m_a x_a y_a & -\sum_a m_a x_a z_a \\ -\sum_a m_a x_a y_a & \sum_a m_a (r_a^2 - y_a^2) & -\sum_a m_a z_a y_a \\ -\sum_a m_a x_a z_a & -\sum_a m_a z_a y_a & \sum_a m_a (r_a^2 - z_a^2) \end{vmatrix}, \tag{5.A.4}$$

The inertia tensor is characterized by diagonal elements I_{11}, I_{22} and I_{33} that are known as the “moments of inertia” and the 6 independent off-diagonal elements, I_{12}, I_{13}, etc, are termed the "products of inertia" (notice that this matrix is Hermitian, $I_{12} = I_{21}$). The initial Cartesian coordinate system (x,y,z) may be of arbitrary origin as shown in figure 5.14 **A**, meaning the object can be displaced to any position without changing the result of the inertia tensor calculation.

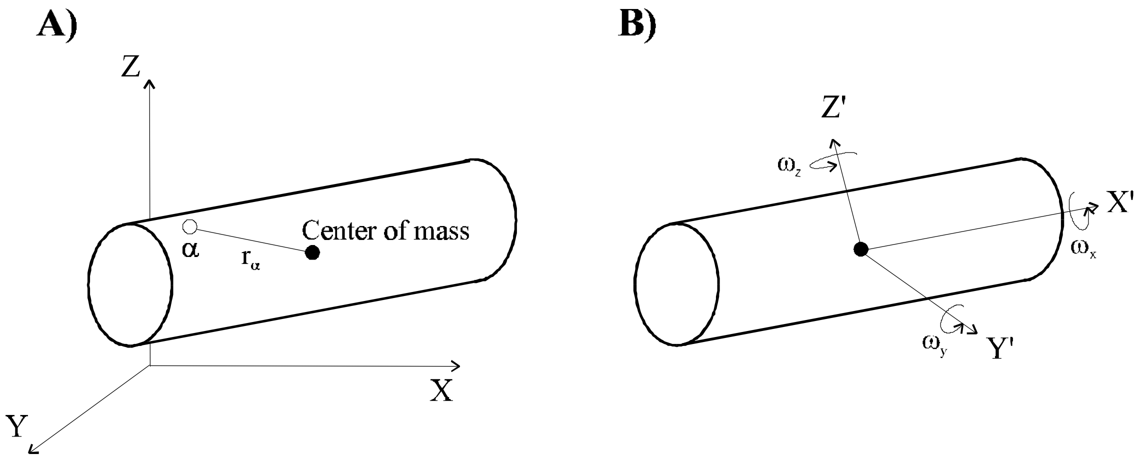


Figure 5.8 The inertia tensor and a symmetrical top

The "principle axes of inertia" is defined as the axis coordinate system (x',y',z') in which the off diagonal terms for {**I**} vanish, $I_{i\neq j}$=0. This diagonalized inertia tensor, **I'**, is calculated from the inertia tensor by finding a transformation matrix, λ, such that,

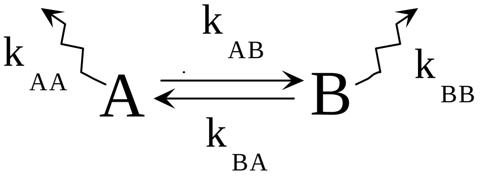
$$\mathbf{I}' = \mathrm{l}\,\mathbf{I}\,\mathrm{l}^{-1}. \qquad 5.A.5$$

When the inertia tensor is transformed in this manner, the three eigenvalues of **I**, I_1, I_2 and I_3 (solved using methods described in appendix 5.6.3) are known as the "principle moments of inertia". Examination of the relative values of I_1, I_2 and I_3 gives much information on the shape of the body. If I_1=I_2=I_3 the body is a "spherical top". If I_1=I_2„ I_3 then the body is termed a "symmetrical top" (DNA and other cylindrical molecules fall into this category). Finally, if I_1=0 and I_2=I_3 then, for instance, the body may be two point masses connected via a weightless shaft, this is known as a "rotor".

5.8.3 Solving coupled rate equations, eigenvalues and eigenvectors

Many mathematical problems deal with solving simultaneous rate equations, such as chemical exchange and NMR relaxation. The concentration (or magnetization) of a species may be dependent on the interactions of many other species which are all undergoing rate processes. This appendix is a short discussion of how to exactly solve the simple case of chemical exchange in a two species system as well as the methods of solving N species problems via approximate methods.

The simplest case of two species in chemical exchange is presented (Eigen & DeMaeyer, 1963; discussions with Pat Vaccaro & Donald Crothers). Assume two chemical species, A and B, which can interconvert at a rate of k_{AB} in which both species also undergo an external decay with rates k_{AA} *and* k_{BB} respectively. This is analogous to the situation of relaxation processes in NMR (think cross-relaxation and T_1).



The kinetic differential equations for the chemical exchange process would be (assuming $k_{AB} = k_{BA}$),

$$\frac{d[A]}{dt} = -k_{AB}[A] - k_{AA}[A] + k_{AB}[B] = -(k_{AB} + k_{AA})[A] + k_{AB}[B], \qquad 5.A.6$$

$$\frac{d[B]}{dt} = -k_{AB}[B] - k_{BB}[B] + k_{AB}[A] = -(k_{AB} + k_{BB})[B] + k_{AB}[A], \qquad 5.A.7$$

which can be rewritten in matrix form,

$$\frac{d}{dt}\begin{vmatrix}[A]\\ [B]\end{vmatrix} = \begin{vmatrix} -(k_{AB}+k_{AA}) & +k_{AB} \\ +k_{AB} & -(k_{AB}+k_{BB}) \end{vmatrix}\begin{vmatrix}[A]\\ [B]\end{vmatrix}. \qquad 5.A.8$$

Symbolically, this matrix equation becomes,

$$\frac{d}{dt}\mathbf{C} = \mathbf{RC}\,, \qquad 5.A.9$$

with a “concentration matrix” **C** and a “rate matrix” **R**. Solving matrix equations of this type requires diagonalization of the matrix **R** to determine a set of eigenvalues, which can then be placed back into equation 5.19 to determine the values for **C**, the eigenvectors.

All matrix operators, such as **R**, can be diagonalized by a similarity transformation **T**,

$$\mathbf{T}^{-1}\mathbf{RT} = \mathrm{l}\ \ , \qquad 5.A.10$$

in which λ is a diagonal matrix (a diagonal matrix is any matrix with off-diagonal term of zero). The transformation matrix is unique in that,

$$\mathbf{T}^{-1}\mathbf{T} = \mathbf{TT}^{-1} = \mathbf{E} \qquad 5.A.11$$

where **E** is the unity matrix (a diagonal matrix with diagonal elements $E_{i,i}$=1 and off-diagonal elements $E_{i,j(i\ „\ j)}$=0).

Define a new matrix, **y**, as,

$$\mathbf{T}^{-1}\mathbf{C} = \mathbf{y}\,. \qquad 5.A.12$$

Multiply both sides of equation 5.19 by $\mathbf{T}^{-1}$ and place the unit matrix between **R** and **C** to give,

$$\mathbf{T}^{-1}\frac{d}{dt}\mathbf{C} = \mathbf{T}^{-1}\mathbf{RTT}^{-1}\mathbf{C} \qquad 5.A.13$$

Now, using the definitions for **λ** and **y** found in equations 5.A.10 and 5.A.12 respectively,

$$\frac{d}{dt}\mathbf{y}_i = \lambda_i \mathbf{y}_i, \tag{5.A.14}$$

which have solutions of,

$$y_i = y_{i,0}\exp(-t\lambda_i) \tag{5.A.15}$$

A new vector, **C’**, is defined as **MCM**$^{-1}$ and is known as the normal coordinate. Which yields,

$$\frac{d}{dt}\mathbf{C}' = \lambda\,\mathbf{C}', \tag{5.A.16}$$

with specific values of,

$$\frac{d}{dt}\begin{vmatrix} C_1' \\ C_2' \\ C_N' \end{vmatrix} = \lambda \begin{vmatrix} C_1' \\ C_2' \\ C_N' \end{vmatrix}. \tag{5.A.17}$$

The matrix matrix can be expanded, term for term, as,

$$\begin{aligned} C_1' &= C_1'\exp(-\lambda t) \\ C_2' &= C_2'\exp(-\lambda t) \\ C_N' &= C_N'\exp(-\lambda t) \end{aligned} \tag{5.A.18}$$

The process of diagonalizing a matrix is accomplished by setting the determinant of the matrix to zero and solving for λ,

$$\det(\mathbf{R}) = \det\begin{vmatrix} -(k_{AB}+k_{AA})-\lambda & k_{AB} \\ k_{AB} & -(k_{AB}+k_{BB})-\lambda \end{vmatrix} = 0 \tag{5.A.19}$$

$$[-(k_{AB}+k_{AA})-\lambda\,][-(k_{AB}+k_{BB})-\lambda\,]-(k_{AB})^2=0 \tag{5.A.20}$$

$$\lambda^2 + (k_{AA}+2k_{AB}+k_{BB})\lambda + k_{AA}k_{AB} + k_{AA}k_{BB} + k_{AB}k_{BB} = 0$$

Which gives (using the quadratic equation to solve for the roots of λ) the eigenvalue (or only non-trivial) solutions of the problem (in this case, two, λ_1 and λ_2),

$$\text{l}_{1and\,2} = \frac{-(k_{AA} + 2k_{AB} + k_{BB}) \pm [k_{AA}^2 + 4k_{AB}^2 - 2k_{AA}k_{BB} + k_{BB}^2]^{1/2}}{2}. \qquad 5.A.21$$

The coefficients from the vector matrix, or the eigenvectors, can now be determined by substitiuting each eigenvalue back into the matrix equation **RT**=**0**,

$$\begin{vmatrix} -(k_{AB} + k_{AA}) - \text{l}_1 & k_{AB} \\ k_{AB} & -(k_{AB} + k_{BB}) - \text{l}_1 \end{vmatrix} \begin{vmatrix} C_{a1} \\ C_{b1} \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \end{vmatrix}, \qquad 5.A.22$$

and solving the equations for C_{a1} and C_{b1} such that $[A](t)=C_{a1}exp(-\text{l}_1 t)$ and $[B](t)=C_{b1}exp(-\text{l}_1 t)$.

This approach gives an exact solution to the problem of two species chemical exchange, however, as the number of coupled equations increase, it becomes exceedingly difficult to solve the matrices exactly and methods of approximation are required. These approximation approaches include the Jacobi rotation-transformation matrix method and those available in the LAPACK software for computers.

5.8.4 Calculation of two-spin state populations

This mathematical transformation is the algebraic substitutions of equations 5.8-5.13. Start from the definitions for the time-dependent change in population (equations 5.8-5.11) and for the NMR observables, $I_{z,A}$ and $I_{z,B}$ (equations 5.12 and 5.13). The time derivative of equation 5.14 gives,

$$\frac{d}{dt}\left(KI_{z,A}\right) = \frac{d}{dt}\left(N_{\text{bb}} + N_{\text{ba}} - N_{\text{ab}} - N_{\text{aa}}\right), \qquad 5.A.23$$

$$K\frac{dI_{z,A}}{dt} = \frac{dN_{\text{bb}}}{dt} + \frac{dN_{\text{ba}}}{dt} - \frac{dN_{\text{ab}}}{dt} - \frac{dN_{\text{aa}}}{dt}. \qquad 5.A.24$$

For which the definitions of dN_{bb}/dt, etc., given in equations 5.8-5.11 are substituted,

$$
\begin{aligned}
K\frac{dI_{z,A}}{dt} &= -(W_{1A}^{AB} + W_{2}^{AB} + W_{1B}^{AB})N_{\mathrm{bb}} + W_{1A}^{AB}N_{\mathrm{ab}} + W_{1B}^{AB}N_{\mathrm{ba}} + W_{2}^{AB}N_{\mathrm{aa}} \\
&\quad + W_{1A}^{AB}N_{\mathrm{bb}} - (W_{1A}^{AB} + W_{0}^{AB} + W_{1B}^{AB})N_{\mathrm{ab}} + W_{0}^{AB}N_{\mathrm{ba}} + W_{1B}^{AB}N_{\mathrm{aa}} \\
&\quad - \left[W_{1B}^{AB}N_{\mathrm{bb}} + W_{0}^{AB}N_{\mathrm{ab}} - (W_{1B}^{AB} + W_{0}^{AB} + W_{1A}^{AB})N_{\mathrm{ba}} + W_{1A}^{AB}N_{\mathrm{aa}}\right] \\
&\quad - \left[W_{2}^{AB}N_{\mathrm{bb}} + W_{1B}^{AB}N_{\mathrm{ab}} + W_{1A}^{AB}N_{\mathrm{ba}} - (W_{1B}^{AB} + W_{2}^{AB} + W_{1A}^{AB})N_{\mathrm{aa}}\right]
\end{aligned}
\tag{5.A.25}
$$

Grouping the N_{bb}, N_{ab}, N_{ba} and N_{aa} terms respectively,

$$
\begin{aligned}
K\frac{dI_{z,A}}{dt} &= \left(-W_{1A}^{AB} - W_{2}^{AB} - W_{1B}^{AB} + W_{1A}^{AB} - W_{1B}^{AB} - W_{2}^{AB}\right)N_{\mathrm{bb}} \\
&\quad + \left(W_{1A}^{AB} - W_{1A}^{AB} - W_{0}^{AB} - W_{1B}^{AB} - W_{0}^{AB} - W_{1B}^{AB}\right)N_{\mathrm{ab}} \\
&\quad + \left(W_{1B}^{AB} + W_{0}^{AB} + W_{1B}^{AB} + W_{0}^{AB} + W_{1A}^{AB} - W_{1A}^{AB}\right)N_{\mathrm{ba}} \\
&\quad + \left(W_{2}^{AB} + W_{1B}^{AB} - W_{1A}^{AB} + W_{1B}^{AB} + W_{2}^{AB} + W_{1A}^{AB}\right)N_{\mathrm{aa}}\ .
\end{aligned}
\tag{5.A.26}
$$

$$
\begin{aligned}
K\frac{dI_{z,A}}{dt} &= \left(-2W_{2}^{AB} - 2W_{1B}^{AB}\right)N_{\mathrm{bb}} + \left(-2W_{0}^{AB} - 2W_{1B}^{AB}\right)N_{\mathrm{ab}} \\
&\quad + \left(2W_{1B}^{AB} + 2W_{0}^{AB}\right)N_{\mathrm{ba}} + \left(2W_{2}^{AB} + 2W_{1B}^{AB}\right)N_{\mathrm{aa}}\ .
\end{aligned}
\tag{5.A.27}
$$

Factoring out the common terms,

$$
K\frac{dI_{z,A}}{dt} = 2\left(W_{2}^{AB} + W_{1B}^{AB}\right)\left(N_{\mathrm{aa}} - N_{\mathrm{bb}}\right) + 2\left(W_{0}^{AB} + W_{1B}^{AB}\right)\left(N_{\mathrm{ba}} - N_{\mathrm{ab}}\right). \tag{5.A.28}
$$

The sum of equations 5.10 and 5.11 gives,

$$
2\left(N_{\mathrm{bb}} - N_{\mathrm{aa}}\right) = I_{z,A} + I_{z,B}\ , \tag{5.A.29}
$$

whilst the difference of the two gives,

$$
2\left(N_{\mathrm{ba}} - N_{\mathrm{ab}}\right) = I_{z,A} - I_{z,B}\ . \tag{5.A.30}
$$

Substitution of equations 5.36 and 5.37 into 5.35 yields,

$$K\frac{dI_{z,A}}{dt} = -\left(W_2^{AB} + W_{1B}^{AB}\right)\left(I_{z,a} + I_{z,B}\right) - \left(W_0^{AB} + W_{1B}^{AB}\right)\left(I_{z,A} - I_{z,B}\right), \tag{5.A.31}$$

$$K\frac{dI_{z,A}}{dt} = -\left(W_2^{AB} + 2W_{1B}^{AB} + W_0^{AB}\right)I_{z,A} + \left(W_0^{AB} - W_2^{AB}\right)I_{z,B}\,. \tag{5.A.32}$$

The same treatment will yield the following for $dI_{z,B}/dt$,

$$K\frac{dI_{z,B}}{dt} = -\left(W_2^{AB} + 2W_{1A}^{AB} + W_0^{AB}\right)I_{z,B} + \left(W_0^{AB} - W_2^{AB}\right)I_{z,A}\,. \tag{5.A.33}$$

CHAPTER 6 “EXPERIMENTAL EVIDENCE OF THE EFFECT OF ANISOTROPIC ROTATION ON NOE INTENSITIES”

6.1 Summary

This chapter addresses the issue of the effect of anisotropic molecular rotation on the nuclear Overhauser effect (NOE) as measured by NMR.

Three samples were examined. The first sample is nearly spherical in its hydrodynamic dimensions and the NOEs reflect this, they show no influence from anisotropic rotation. The second sample is cylindrical in shape with a long to short axis ratio of 2:1. The NOEs from this sample have been influenced by the rotational anisotropy due to the cylindrical shape. The third sample is also cylindrical in shape with a long to short axis ratio of 4:1, and it is with this sample that the strongest influence of the rotational anisotropy on the NOEs are seen.

6.2 Introduction

Methods for structure determination of biomolecules by NMR have historically relied heavily on the use of two primary experiments. The first is the Correlation Spectroscopy (COSY) experiment, and its many derivatives, which gives information relating the energy coupling of spin pairs which are connected via covalent bonds. Most notably, the 3 bond J (or scalar) coupling between two protons is important because the intensity of the coupling relates to the torsion angle formed between the protons. This can be exploited to determine the torsion angle. The most obvious limitation of using COSY data in structure determination is that the information corresponds only to short distances, the torsion angle can only lock down the geometry of a few atoms connected through-bond, which will be necessarily spatially close to one another.

The second major experiment used by NMR spectroscopists for structure determination is to measure the nuclear Overhauser effect. The NOE is a dipole-dipole relaxation rate process in which two magnetic nuclei are coupled by their dipole moments. The intensity of the measured NOE is dependent on the distance separation between the nuclei and is realized experimentally as a crosspeak in a two-dimensional NOESY experiment correlated to the frequencies of the two resonances. NOE information can be a powerful tool for structure elucidation in that it can relate the distances of covalently remote atoms, unlike the COSY data. This is especially important and useful for large biomolecules which may be folded into interesting secondary and tertiary structures.

In practice, however, the crosspeak intensity between two resonances of a NOESY experiment alone does not give the distance between the two nuclei. The NOE is a consequence of dipolar relaxation, and as such, it involves a number of complex processes, all of which must be understood in order to interpret the data correctly. Nuclear spin relaxation derives from the fluctuating magnetic fields surrounding a nucleus. These fields can come from a number of sources, molecular rotational diffusion, global molecular dynamics, localized atomic libration and others. To fully understand the NOE data, an accurate (and verifiable) model of all these motions is necessary. It is the inherent complexity of the underlying theory that has led to assumptions that simplify the interpretation of NOE data.

This chapter attempts to separate the effect of the molecular tumbling component of dipolar relaxation from those of structure and intramolecular dynamics. To do this, the samples chosen for study must meet two criteria: first, the samples must be composed of (or contain a region of) fairly uninteresting “regular” structure; second, the samples must span a range of rotational motions, from spherical isotropic motion to cylindrical anisotropic motion. The first requirement is necessary because we will have to make assumptions about the structure and dynamics of the samples and we feel more confident in making these assumptions on well-defined structural elements. The second requirement amplifies the effect of the rotational dynamics on the NOE.

6.2.1 Hydrodynamics theory for rotational diffusion rates

The theoretical basis for the influence of molecular rotation on homonuclear NMR relaxation has been reviewed in chapter 5 and should be consulted. This section will explore the current hydrodynamics theories for predicting rotational correlation times, given a hydrodynamic particle of defined shape, and other experimental methods used for determining these correlation times.

The rotational diffusion rate, D_r, for a hydrodynamic particle is given by the rotational analog of the Stokes-Einstein equation for translational diffusion,

$$D_r = \frac{k_b T}{f_r}, \qquad 6.1$$

where k_b is Boltzmann’s constant, T is the temperature in Kelvin and f_r is the rotational friction coefficient. It is in the modeling of the rotational friction coefficient that the hydrodynamic shape of the subject is important. For a spheroid,

$$f_r = 8\mathrm{ph}_0 R^3, \qquad 6.2$$

where R is the radius and h_0 is the viscosity of the pure solvent (see the Materials and Methods section of Chapter 4 for a discussion of calculating viscosities for H_2O and D_2O solutions).

For modeling more complex shapes, it is often convenient to express the friction coefficient in terms of a sphere of "equivalent radius", *Re*. This requires the inrtoduction of a new dimensionless frictional coefficient (denoted with a capital F), defined as $F_r = f/f_{sphere,,}$

$$f_r = F_r(8\text{ph}_0 Re^3)\,. \qquad 6.3$$

The *Re* for a cylindrical rod can be calculated using,

$$Re_{(cylinder)} = \left(\frac{2}{2p^2}\right)^{1/3}\left(\frac{L}{2}\right), \qquad 6.4$$

where *p* is the axial ratio of length to diameter ($p=L/d$). F_r must be defined separately for the two axis of rotation for a cylinder, $F_{r,l}$ and $F_{r,s}$ in which the axis labeled *l* is the long axis and the axis labeled *s* is the short axis of a cylinder (Fig. 6.1). Notice that we will define the variables describing the long axis ($D_{r,l}$, and $F_{r,l}$) as that property *about* the long axis. That is, the $D_{r,l}$ of a DNA will be the rotational diffusion rate of the short axis *about* the long axis. It should be noted that in the literature these definitions are sometimes reversed; $D_{r,l}$ may describe the rotational diffusion rate *of* the long axis about the short axis.

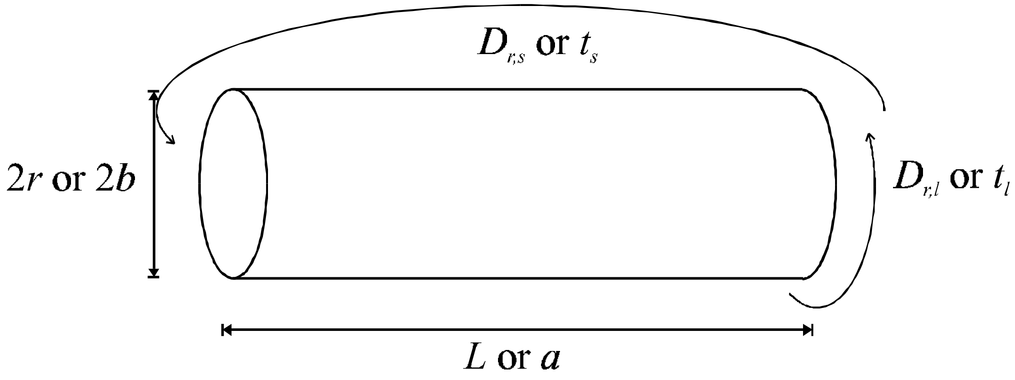


Figure 6.1 Definitions of hydrodynamic variables for a cylinder

$$F_{r,l} = 0.64\left(1 + \frac{0.677}{p} - \frac{0.183}{p^2}\right), \qquad 6.5$$

$$F_{r,s} = \frac{2p^2}{9(\ln p + \mathrm{d}_a)}, \qquad 6.6$$

with the delta function given by the polynomial approximation as described by Tirado and Garcia de la Torre (1979, 1980),

$$\mathrm{d}_a = -0.662 + \frac{0.917}{p} - \frac{0.050}{p^2}. \qquad 6.7$$

Equations 6.2, 6.4, 6.5, 6.6 and 6.7 are combined to give functions for $D_{r,s}$ and $D_{r,l}$ in terms of the axial ratio, *p*, and length,

$$D_{r,l} = \frac{k_b T}{\mathrm{ph}_0 L^3} \cdot \left(\frac{p^2}{0.64 + 0.43328p^{-1} - 0.11712p^{-2}}\right), \qquad 6.8$$

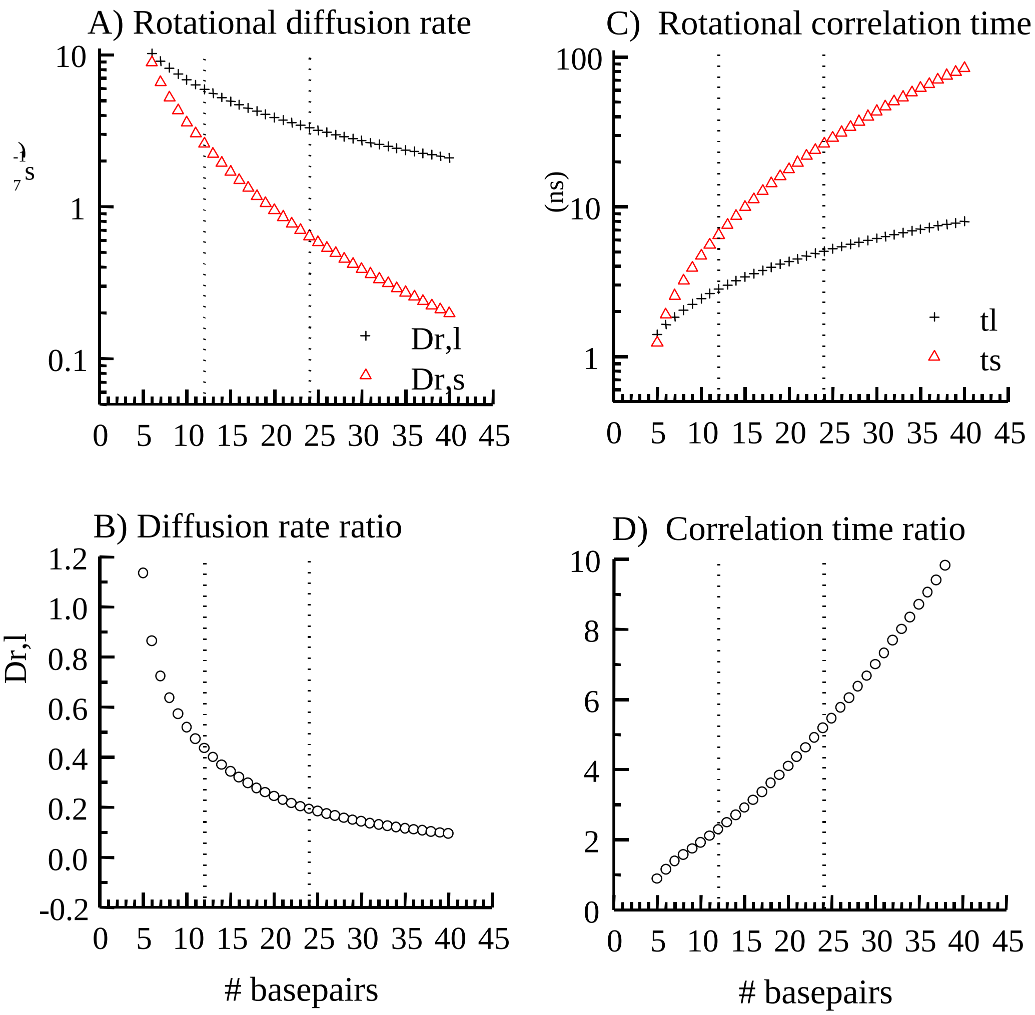
$$D_{r,s} = \frac{3k_b T}{\mathrm{ph}_0 L^3} \cdot (\ln p + \mathrm{d}_a). \qquad 6.9$$

The functions describing the rotational diffusion of a cylinder are shown graphically in figure 6.2 to demonstrates how $D_{r,s}$ and $D_{r,l}$ respond for DNA of size ranging from 5-40 base pairs. Also shown in the graph, are the more NMR-relevant correlation times, t_s and t_l , defined by,

$$t_s = \frac{1}{6D_{r,s}}, \quad t_l = \frac{1}{6D_{r,s}}. \qquad 6.10$$

6.2.2 Experimentally determined correlation times for DNA

Much work has been done on measuring the rotational diffusion rates (D_r) of small and large molecules in solution (Einstein, 1956; Debye, 1929; Perrin, 1934; Perrin, 1936; Alms, *et al*., 1973; Kivelson, 1987; Eimer *et al*., 1990; Eimer & Pecora, 1991). It has been shown that for some systems these measured rates can be calculated accurately for dilute systems by treating the molecule of interest as a hydrodynamic particle. Many



D_r ***t_c***

Graphs were calculated with equations 6.8 and 6.9 (using the script “hydro.pl”, see Chapter 8) at 25 C in 100% D_2
Å rise/bp and a diameter of 20 . **A** **B**) Ratios of $_{r,s}/D$. At 6 base pairs, the length and diameter of DNA is nearly equal, giving a ratio of 1. For larger DNAs, $_{r,s} < D$. **C** t_c $_r)$
D) Correlation time ratios, $_s/t$, with increasing DNA size. For DNAs larger than 6 base-pairs, $_s >$ $_l$.

optical techniques have been developed for measuring the rotational diffusion rates of large molecules (DNAs > 100 bps), such as dichroism and birefringence. However, for accurate measurements of the fast rotational diffusion rates of short oligonucleotides, Pecora *et al*. (1990, 1991) have used a technique known as "Depolarized Dynamic Light Scattering" (DDLS). This technique measures the reorientation relaxation time about the short axis of symmetry ($D_{r,s}$), and is almost completely insensitive to rotation about the long axis ($D_{r,l}$).

Using the DDLS technique, Pecora measured $D_{r,s}$ values of 51.8, 26.1 and 10.3 x 10^6 s^{-1} (t_s values of 3.2, 6.4 and 16.2 ns) for DNAs of sizes 8, 12 and 20 base pairs respectively (Eimer & Pecora, 1991). They showed that these values can be predicted accurately by hydrodynamics theory when modeling DNA as a symmetric top as was presented in section 6.2.1 (Tirado & Garcia de la Torre, 1979, 1980; Tirado, *et al*., 1984; Garcia de la Torre, *et al*., 1984). The hydrodynamic parameters used in their analysis were 3.4 Å rise per base pair and 20 Å diameter. Additionally, it was also found that there was no significant concentration effects for the rotational diffusion measurements for DNA in a concentration range of between 0.1 to ~2.0 mM.

6.2.3 Experimental approach

Analysis of the data presented in this chapter will rely on the technique of back calculating NOE intensities from a structural, rotational and intermolecular dynamic "model" of the molecule in question. A statistical comparison of the simulated NOE intensities and the experimentally measured intensities will evaluate how well the particular model fits the data. The computer program YARM will be used to perform the calculations (see Chapter 7 for a description of the program) and the YARM scripts used

In order to isolate the effect of the rotational motions on the NOE, the structure

motions will be varied. For each rotational motion sampled, the statistical comparison between the simulated and experimental NOEs will be reported. For example, for a truly

and short axis correlation times are identical; indicating that there is no systematic angular dependence to the cross-relaxation rates between the spins.

A variety of molecular hydrodynamic shapes were chosen for analysis, ranging from a small sphere to an elongated cylinder. In the world of nucleic acids, these shapes

duplex (D12 and D24). The approximate hydrodynamic shapes of the samples are shown below (Fig. 6.3).

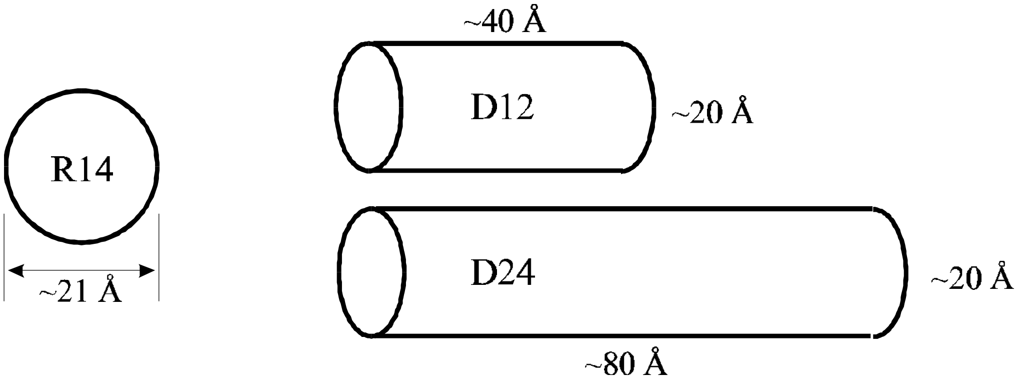


Figure 6.3 Approximate hydrodynamic dimensions of R14, D12 and D24 samples

The R14 RNA is an analog of helix 45 of bacterial 16s rRNA, with a sequence of 5'-GGACCGGAAGGUCC-3'. This RNA has been studied extensively, the hydrodynamical properties of R14 have been examined by measuring the translational diffusion of the RNA (Lapham, et al., 1997) and its solution state structure has also been determined (Rife and Moore, unpublished results).

As was discussed in chapter 4, this RNA can exist in a hairpin conformation in buffers with low salt concentrations and as a dimer in high salt concentration buffers. The hairpin form of the R14 sample was chosen for study because it forms a small compact structure (Rife and Moore, unpublished results), with an approximately spherical hydrodynamic shape, and should closely represent an isotropically rotating molecule. The figure below shows the base pairing for the stem region for R14. The question marks represent, presumably, non standard Watson-Crick A-form RNA structure. Analysis of this RNA involved only the NOEs between protons found in the helix region, specifically G_1, G_2, A_3, C_4, C_5, G_{10}, G_{11}, U_{12}, C_{13} and C_{14} (numbering from 5’ to 3’). The NOE data involving the central 4 nucleotides, G_6, G_7, A_8 and A_9 was ignored, due to the possible existence of “interesting” structure or dynamics.

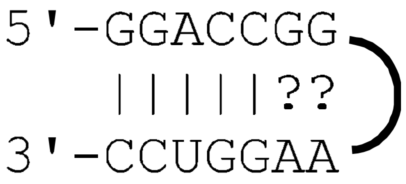


Figure 6.4 14 hairpin

In the analysis of the R14 NOE data a number of assumptions must be made about its molecular model. The helix-stem portion of the molecule will be modeled structurally as A-form RNA and the intramolecular dynamics will be modeled as rigid. Additionally, the hydrodynamic shape of the RNA will be assumed to be a sphere of radius 10.5 Å.

The two DNAs chosen for study, D12 and D24 were both derived from the sequence found at the EcoR1 restriction site. The first sample, D12, is 5’-CGCGAATTCGCG-3’ and is commonly referred to as the “Dickerson dodecamer”. There have been a number of structural studies performed of this palindromic DNA, including an X-Ray crystallography (Drew *et al*., 1981) and NMR spectroscopy (Nerdal, *et al*., 1989). Rotational dynamics studies have been reported as well, including results from dynamic light scattering (Eimer, *et al*., 1990; Eimer & Pecora, 1991). Translational diffusion constants have been measured using NMR techniques and discussed in terms of the hydrodynamical modeling of DNA (Lapham, et al., 1997). This DNA is well understood in terms of its structural, rotational and translational dynamics.

The D24 sample is 5’- **CGCGAATTCGCG**CGCGAATTCGCG -3’ and, as with the D12 sample, is palindromic. The sequence was constructed by simply duplicating the D12 sequence, with the thought that they would exhibit an identical structure and intramolecular dynamics. The rotational motions, however, should be quite different given that this DNA is twice the length of the D12 DNA.

The symmetry of the DNA molecules is shown below (Fig. 6.5). Notice that the D24 sample contains two "pseudo-symmetric" positions. This causes the protons from the nucleotides near the pseudo-symmetric region to have the same chemical shifts as they do in the D12 sample (Figs. 6.6 and 6.7). We shall assume that, aside from the

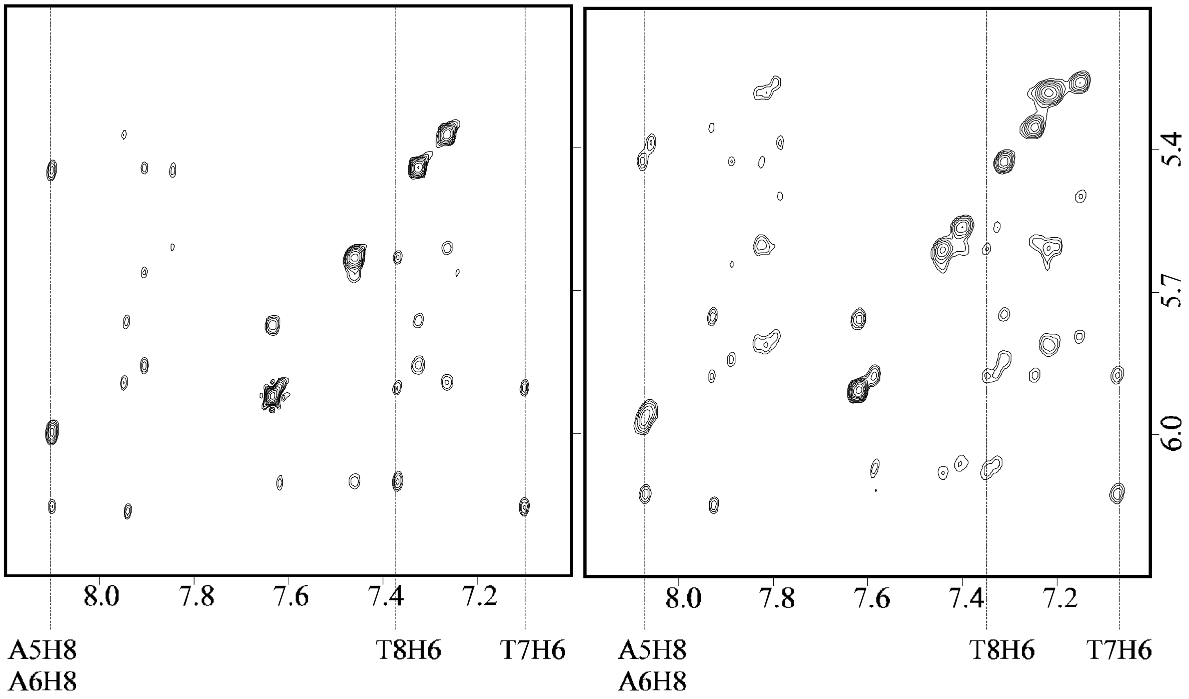


Figure 6.6 D12 and D24 2D NOESY spectra

The 2D NOESY experiments for both the D12 and D24 samples were collected at 25° C using a mixing time of 250 ms and a recycle delay of 30 seconds. Shown is the anomeric-aromatic regions of beth the D12 (left spectra) and D24 (right spectra) samples. The dashed line represent the resonances found in the pseudo-symmetric region of the D24 sample which have identical chemical shifts to the D12 sample.

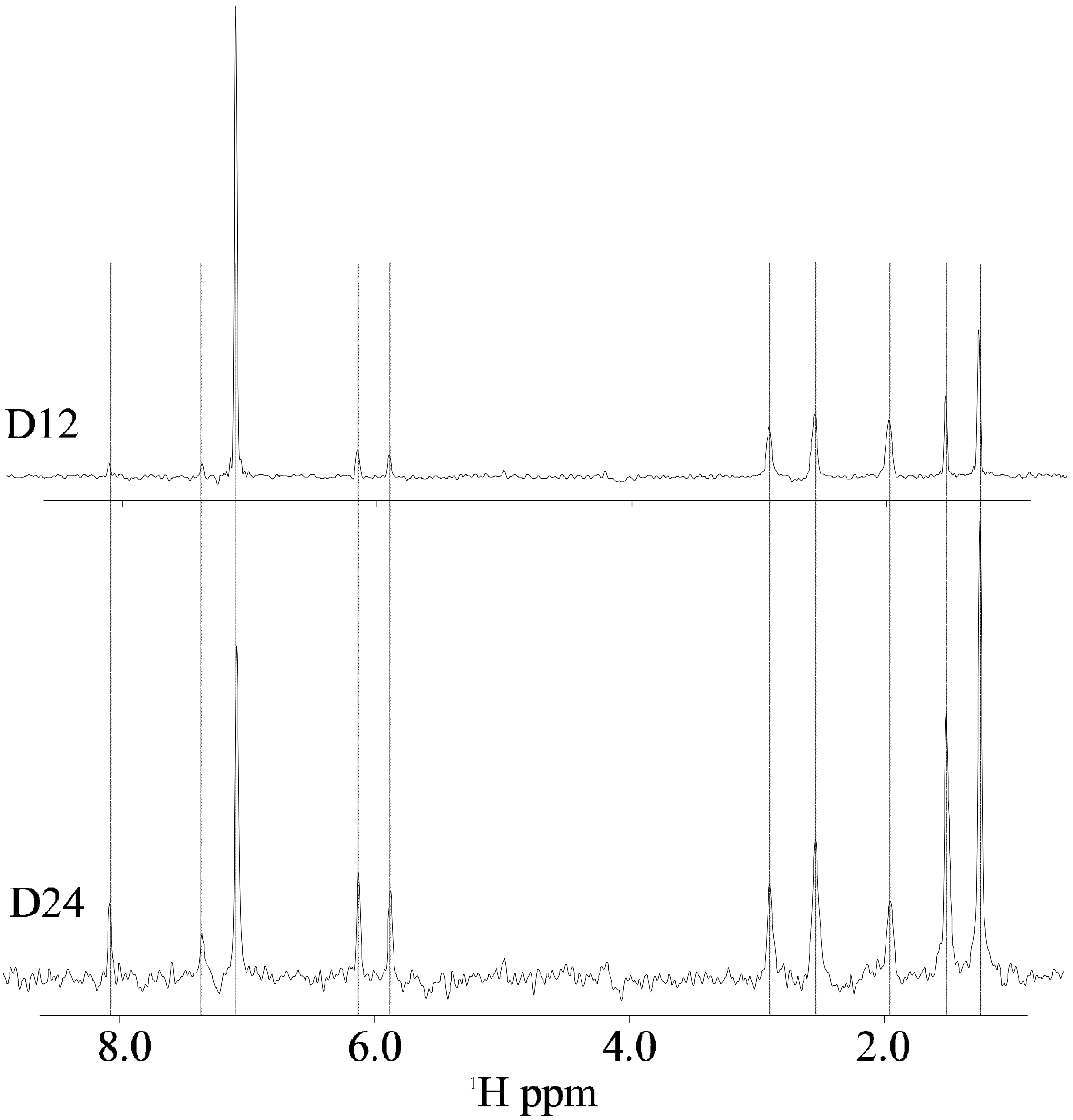


Figure 6.7 T7 H6 1D slice of 2D NOESY

A 1D slice through the thymidine 7 H6 proton from the 2D NOESY spectra for the D12 and D24 samles. The chemical shifts of the crosspeaks are identical between the two samples.

central few G:C base pairs on the D24 sample, the structural and dynamical properties of the D12 and D24 samples are similar.

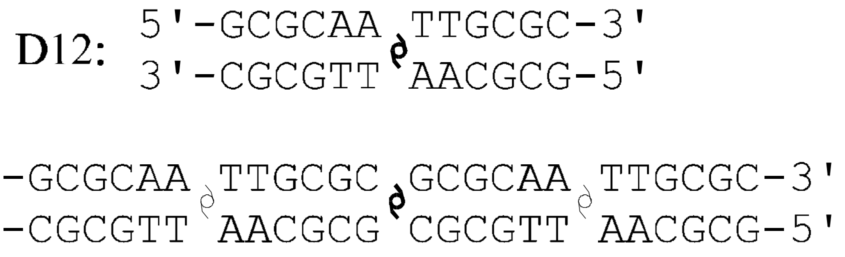

D12:
5'-GCGCAA TTGCGC-3'
3'-CGCGTT AACGCG-5'
D24:
5'-GCGCAA TTGCGC GCGCAA TTGCGC-3'
3'-CGCGTT AACGCG CGCGTT AACGCG-5'

Figure 6.5 Symmetry in the D12 and D24 samples

The assumed hydrodynamic parameters for the two DNAs are 3.4Å rise per base pair and 20Å diameter, as was confirmed from the translational diffusion rate experiments presented in Chapter 4. This gives a length of 40.8Å and 81.6Å for the D12 and D24 samples respectively as shown below.

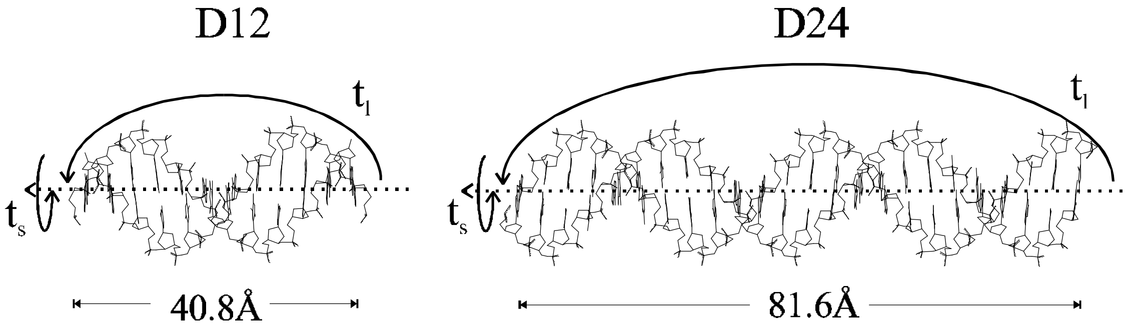


Figure 6.8 Hydrodynamic parameters for the D12 and D24 samples

It should be noted that Jason Rife provided the data for the R14 samples. Any discrepancies in the interpretation of the data, however, lie solely with the author of this thesis.

6.3 Results

This chapter evaluates the effect of different rotational motions on the NOE. This is accomplished by back calculating the simulated NOE intensities and comparing them to experimentally measured values. In order to simulate the NOEs, the computer program YARM will be utilized. Use of the program is described in greater detail in chapter 7, but the theoretical consideration of the calculations are presented in chapter 5. The anisotropic definition of the spectral density function (Woessner, 1962) will be used in these simulations.

6.3.1 Cross-relaxation rate simulations for anisotropic rotation

For an isotropically rotating molecule, all spin pair vectors experience the same rotational diffusion rate and a single “correlation time”, t_c, will accurately describe this motion. The rotational motions of the molecule give rise to dipolar relaxation effects between the dipole pair, which can be measured experimentally as a NOE crosspeak in a NOESY experiment. Thus, for a molecule undergoing isotropic rotation, there will be no coordinated “angular dependence” to the cross-relaxation rates between each dipole pair.

However, for a molecule undergoing anisotropic rotation, the rotational diffusion rate each dipole pair vector experiences will depend on the angle the vector makes with respect to the principal axis of rotation (denoted the b_{ij} angle). There will now be a coordinated “angular dependence” to any dipolar relaxation coupling between spin pairs. In terms of the cross-relaxation rates, this effect can be modeled (Fig. 6.9). By changing the long axis correlation time (t_l) with respect to the short axis correlation time (t_s), the cross-relaxation rate (s_{ij}) is shown to have a strong angular dependence with respect to

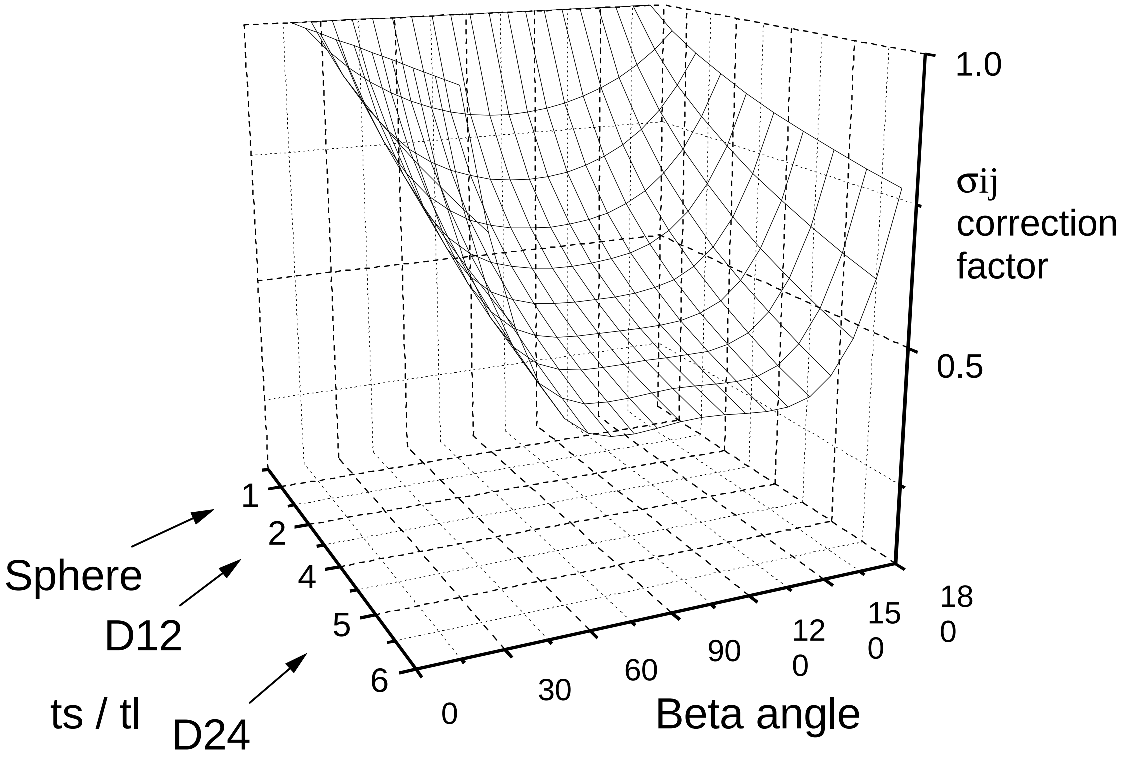


Figure 6.9 Cross-relaxation rate correction factor

As the ratio of the short to long axis correlation time increases, an angular dependence with respect to the principal rotation axis for the cross-relaxation rate is predicted. For the sphere, ts/tl = 1 and there is no angular dependence to sij. For a 12 mer DNA (~40Å x ~20Å) ts/tl ≈ 6.5/2.8 ≈ 2 and for a 24 mer DNA ts/tl ≈ 5. An atom pair vector parallel (0 degrees) to the principal axis of rotation for a 24 mer DNA experiences approximately twice the cross-relaxation rate as a atom pair vector perpendicular (90 degrees) to the principal axis. These values are reported as relative to isotropic rotation with the ts correlation time.

the principal axis of rotation. For a molecule with a $t_s \approx 5t_l$ (such as a 24 base pair DNA), the ratio of the cross-relaxation rates of a spin pair parallel : perpendicular to the principal axis is approximately 2. Since the cross-relaxation rates give rise to the NOE, this effect should be experimentally measurable.

6.3.2 R14 sample

The R14 experimental data was obtained from a 2D NOESY experiment collected at 30° C using a mixing time of 300 ms and a recycle delay of 9 s. A total of 23 well-resolved NOESY crosspeaks (see section 6.5.3 for the volumes list) between protons found in the helix stem region (see Fig. 6.4) of the RNA were used in the analysis. The structure of the helix stem is assumed to be A-form (section 6.5.4).

The rotational hydrodynamics theory (using the program “hydro.pl”, see Chap. 8) can be used to predict the rotational diffusion rate, and thus the correlation time, of the R14 sample. Assuming the RNA is a sphere of diameter 21 Å, equation 6.2 predicts a $D_r = 1.48\text{x}10^8 \text{ s}^{-1}$, which gives a t_c = 1.1 ns (for temp = 30° C, in D_2O).

The results of the analysis of the experimentally measured versus the simulated NOEs are shown in figure 6.10 assuming an isotropic rotation model. As the correlation time is varied from 0.1 to 10 ns, the fit of the simulated to the experimental NOEs is plotted. The best fit occurs for a $t_c \approx 0.7$ ns, which is in good agreement (within the experimental and modeling errors) with the predicted value.

To determine if there was any systematic, coordinated angular dependence to the cross-relaxation rates between the spin pairs in the R14 sample, an anisotropic rotation “surface plot” was calculated (Fig. 6.11). In this analysis, the long and short axis correlation times are varied from 0.1 to 10 ns (the X and Y axis respectively) and the

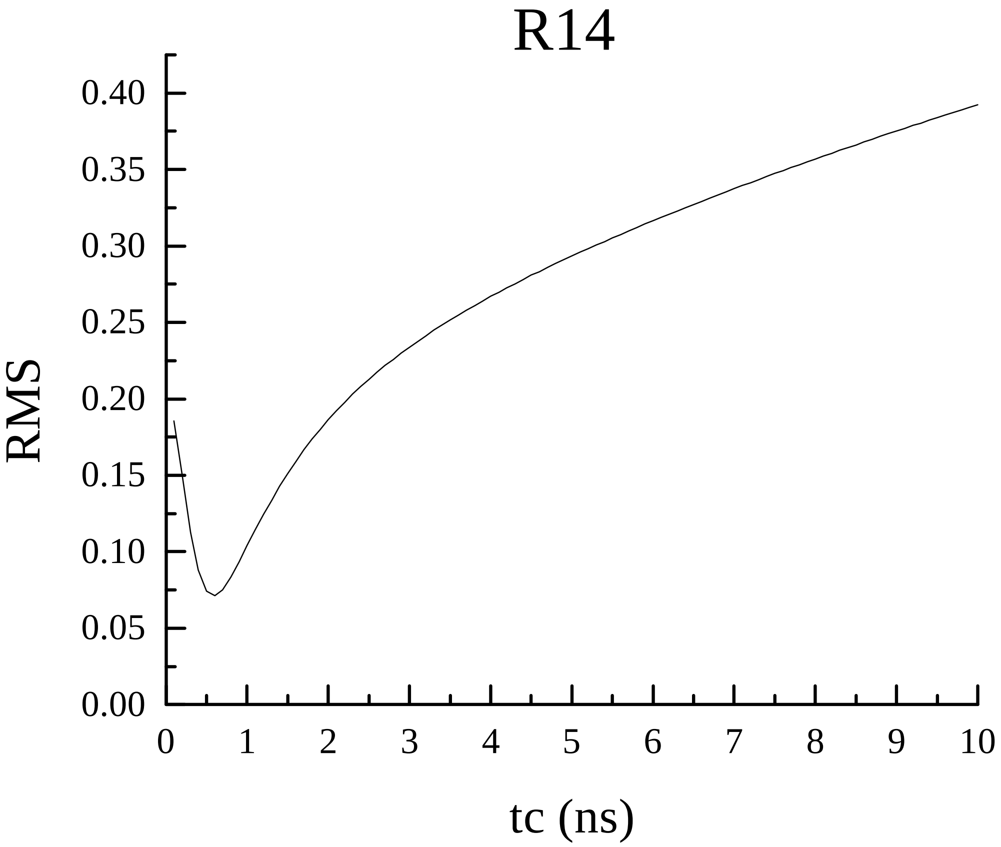


Figure 6.10 R14 isotropic correlation time plot

Graph of the RMS between the experimental NOE data for R14 and the back-calculated NOE data, assuming an A-form RNA with rigid intramolecular dynamics. Using the isotropic definition of the spectral density function, the correlation time of the molecule is varied from 0.1 to 10 ns.

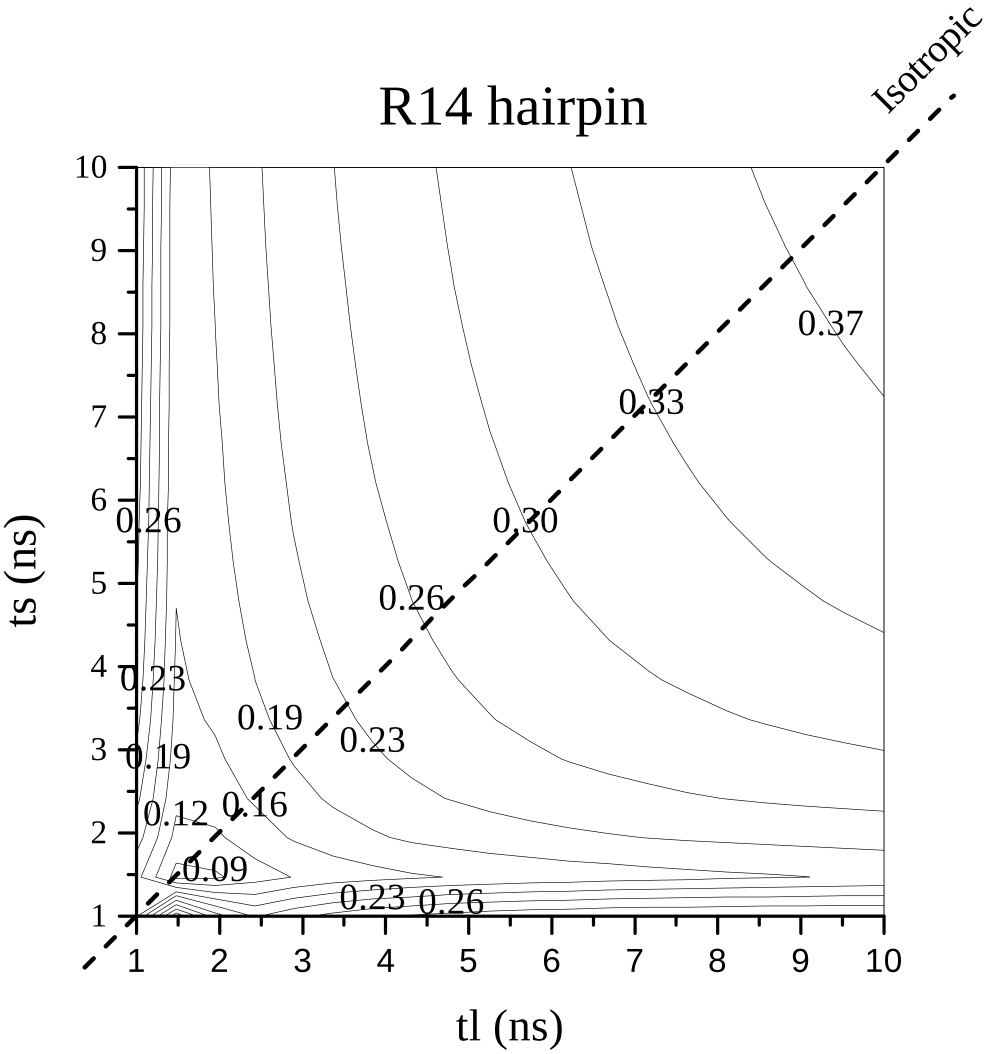


Figure 6.11 R14 anisotropic correlation time "surface plot"

The X and Y axis represent the long and short axis correlation times used in the anisotropic rotation definition of the spectral density function. The Z-axis of the surface plot is the RMS between the experimental and simulated NOESY crosspeak volumes, a minimum in the RMS represents a good fit between the simulated and experimental data. The dashed line represents the position in the graph where $t_s=t_l$, the rotation is isotropic. The simulations were run for a standard A-from RNA at 30° C with a mixing time of 300 ms.

RMS between the experimental and simulated NOEs is graphed on the Z-axis. A "trough" in the surface plot represents a minimum RMS, and a best fit rotational model. Clearly the minimum lies close to the "isotropic line" (where $t_l = t_s$) and modeling the rotational motions as anisotropic gives a worse fit to the data. There is no angular dependence to the measured NOEs for R14, and it is well represented as an isotropically rotating molecule.

6.3.3 D12 sample

A total of 117 well-resolved (234 symmetric) NOESY crosspeaks were used in the analysis of the D12 sample (see section 6.5.3). The structure of the D12 was assumed to be the NMR derived structure (see section 6.5.5) and rigid.

The rotational motions of the D12 sample are predicted using equations 6.5 and 6.6 and assuming the DNA is a cylindrical hydrodynamic particle with dimensions 20 Å by 40.8 Å. This predicts that the D12 DNA has a $D_{r,l} \approx 5.92\text{x}10^{7}$ s^{-1} $D_{r,s} \approx 2.58\text{x}10^{7}$ s^{-1} and a $t_l \approx 2.82$ ns, $t_s \approx 6.46$ ns.

The anisotropic rotation surface plot for the simulated versus experimental data for D12 is shown in figure 6.12. Unlike the R14 sample, the minimum RMS is not found on the isotropic line, rather it is off the diagonal in a region where the $t_s > t_l$. The graph shows that the RMS is not very sensitive to the t_s correlation time, but is very sensitive to the t_l correlation time. The predicted rotational dynamics for the D12 fall in the region of the minimum RMS, indicating that both the theoretical and experimental data indicate that the D12 sample is experiencing anisotropic rotation.

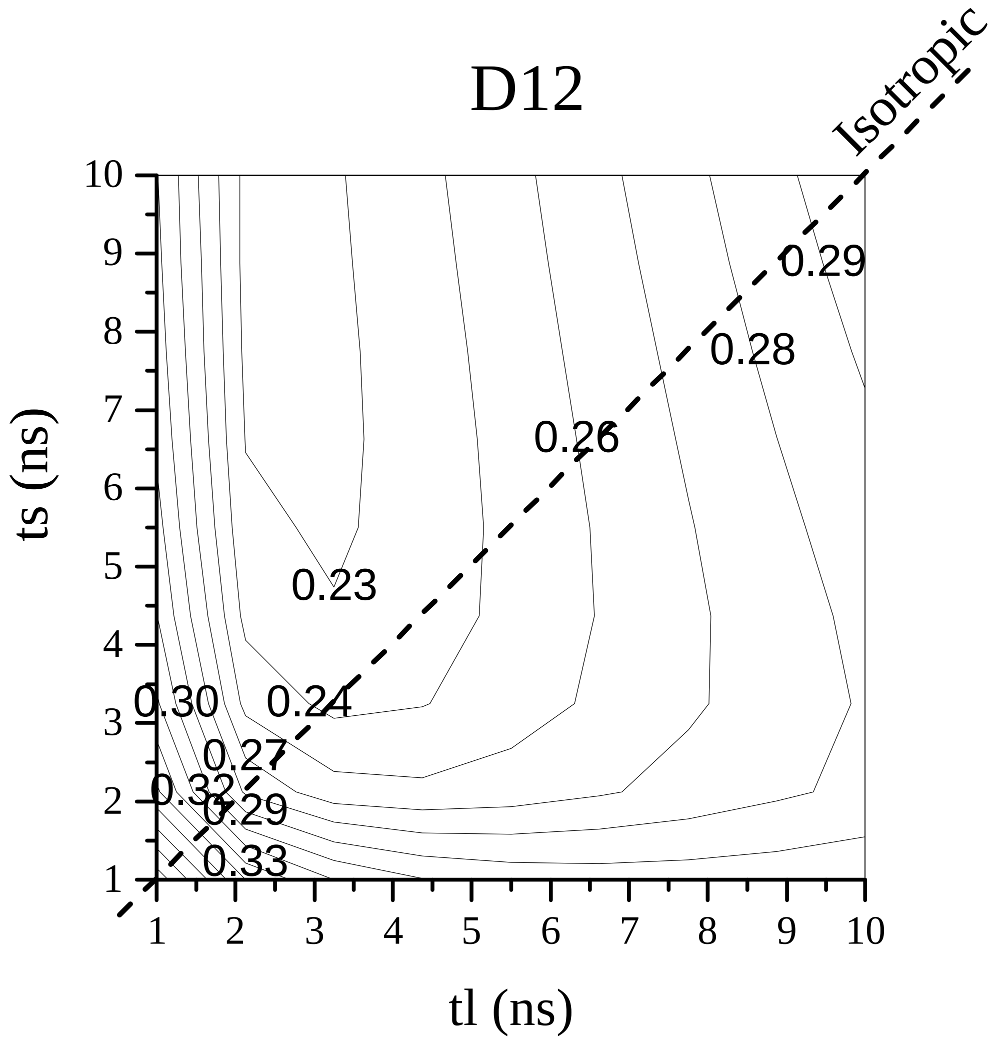


Figure 6.12 D12 anisotropic correlation time "surface plot"

The X and Y axis represent the long and short axis correlation times used in the anisotropic rotation definition of the spectral density function. The Z-axis of the surface plot is the RMS between the experimental and simulated NOESY crosspeak volumes, a minimum in the RMS represents a good fit between the simulated and experimental data. The dashed line represents the position in the graph where t_s=t_l, the rotation is isotropic. The simulations were run at 25° C with a mixing time of 250 ms.

Hydrodynamics theory predicts a correlation time of 2.8 ns about the long axis and 6.5 about the short axis for D12.

6.3.4 D24 sample

A total of 21 well-resolved (42 symmetric) crosspeaks were used in the analysis of the D24 sample (see section 6.5.3), from the pseudo-symmetric region of the DNA (the A-T base pairs). We reasoned that since the crosspeaks from this region of the DNA have exactly the same chemical shifts as for the D12 sample (see Figs. 6.6 and 6.7), the structure and intramolecular dynamics in this region was probably similar.

The predicted rotational correlation times for the D24 samples are $D_{r,l} \approx 3.30\text{x}10^{7}$ s^{-1} $D_{r,s} \approx 0.64\text{x}10^{7}$ s^{-1} and a $t_l \approx 5.05$ ns, $t_s \approx 26.2$ ns (temp = 25° C, D_2O). The hydrodynamics theory suggests that the long axis rotates about 5 times for every short axis rotation.

The anisotropic rotation surface plot for D24 is shown in figure 6.13 (note that the range of the t_l and t_s values was increased to 40 ns as compared to the R14 and D12 surface plots). A tough of minimum RMS is seen where $t_s > t_l$, as would be expected. The minimum is not, however, exactly where the predicted values would suggest. Clearly the D24 sample is undergoing an anisotropic rotation, but the correlation time about the long axis appears to be much bigger than predicted.

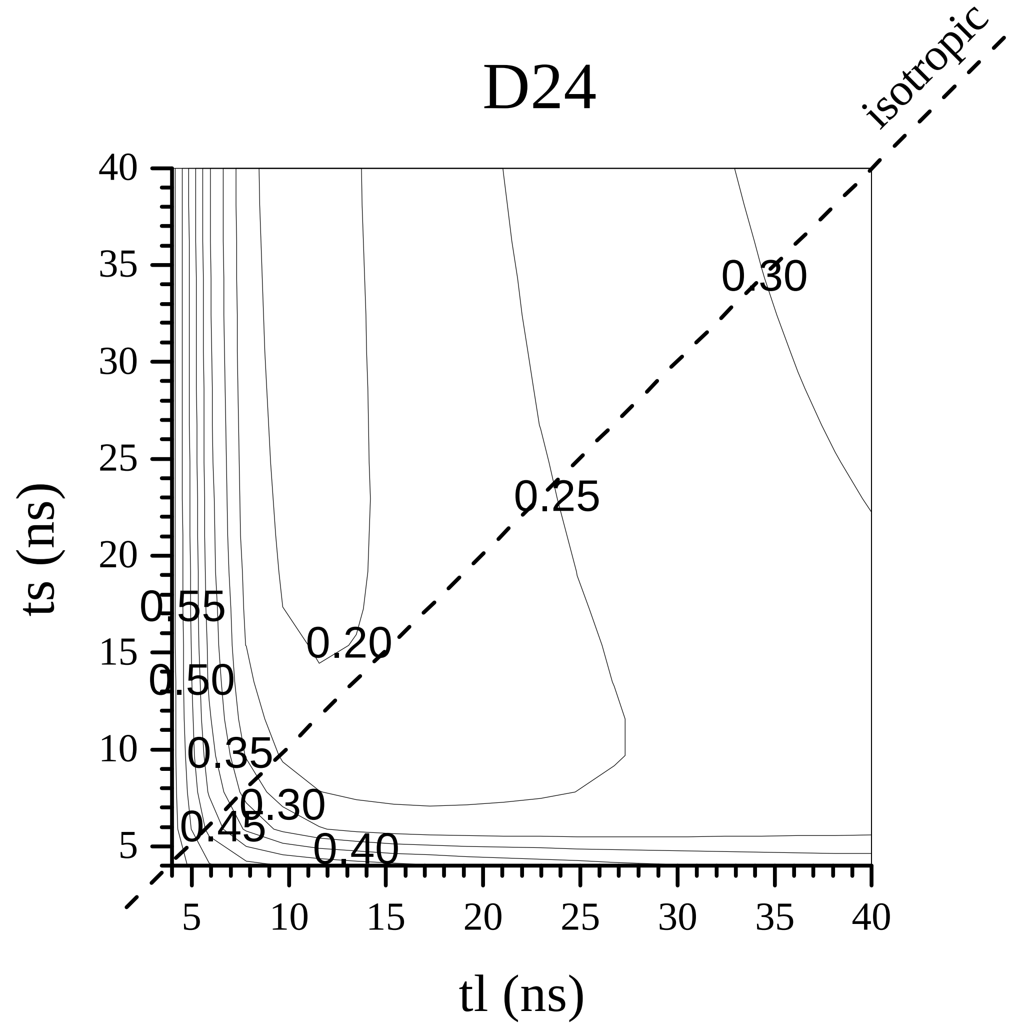


Figure 6.13 D24 anisotropic correlation time surface plot

The X and Y axis represent the long and short axis correlation times used in the anisotropic rotation definition of the spectral density function. The Z-axis of the surface plot is the RMS between the experimental and simulated NOESY crosspeak volumes, a minimum in the RMS represents a good fit between the simulated and experimental data. The dashed line represents the position in the graph where t_s=t_l, the rotation is isotropic. The simulations were run 25° C with a mixing time of 250 ms.

Hydrodynamics theory predicts a t_l=5ns and t_s=26ns. The minimum RMS appears at approximately a long axis correlation time of 10 - 12 ns, and a wide range of short axis correlation times.

6.4 Discussion

The pattern and intensities of NOE crosspeaks in the 2D NOESY experiment represent a wealth of information on the relaxation processes that occur in a molecule. The rates of these dipolar relaxation processes are dependent on three molecular components: the structure of the molecule (r_{ij}), the intramolecular dynamics of the molecule (dr_{ij}/dt) and the rotational motions of the molecule (t_c). These structural and motional components are represented in the relaxation matrix, **R**, as the spectral density function, *J*. Thus, in order to model accurately the relaxation matrix, an accurate definition of the spectral density function is required.

In this chapter, we attempt to deconvolute the effect of molecular rotational dynamics on the relaxation matrix, and ultimately on the NOE intensities. In order to accomplish this, we have quantitated the NOE crosspeak volumes from three NMR samples, which should exhibit different rotational motions. These experimentally measured NOEs are then compared to theoretically simulated NOEs, to ascertain whether the rotational properties of the molecules can be seen in the experimental NMR data itself.

The small RNA hairpin, R14, is predicted to rotate in an isotropic manner. The analysis of the NOEs arising from the helix stem region of the RNA confirms this. The NOE intensities between the measured spin pairs fit well with simulated NOE data using the isotropic definition of the spectral density function. When the anisotropic rotation spectral density function is examined, the back-calculated NOE intensities are a worse fit to the measured experimental data.

The DNA samples, D12 and D24, are predicted to have rotational motions described by two correlation times, one about the long axis and one about the short axis. The data indicate that the predictions are correct, the simulated NOE data is a better fit to the experimental when using the anisotropic rotation spectral density function using a $t_s > t_l$.

The D12 sample shows remarkable agreement between the predicted $t_l \approx 2.8$ ns and $t_s \approx 6.5$ ns and the minimum in the anisotropic rotation surface plot (Fig. 6.12). It can be said that these correlation times probably represent the rotational motions of this DNA well.

The predicted correlation times for the D24 sample, however, do not seem to align with the minimum in the surface plot (Fig. 6.13). The predicted $t_l \approx 5$ ns, while the minimum in the data appears at $t_l \approx 10$-12 ns. One explanation of this discrepancy is that the sample may undergo normal mode bending along the length of the DNA (Zimm, 1956). This bending may cause the long axis rotational diffusion rate to be slower, due to the increased frictional coefficient about the long axis from the bent DNA shape, as shown below in figure 6.14. As the length of the DNA increases, the angular fluctuations due to the normal mode bending in DNA increases. Thus, the D24 DNA would suffer from this more than the D12 sample.

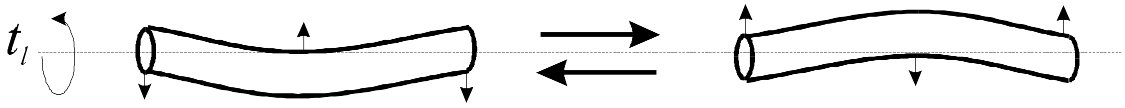


Figure 6.14 Normal mode bending motions of DNA

In conclusion, we are able to observe the effect of molecular tumbling in the experimentally measured NOE data. This is an important consideration for biomolecular NMR spectroscopists as they investigate larger extended shape nucleic acid molecules, as the intensities of their NOESY crosspeaks will have an additional "angular dependence" to them.

6.5 Materials and Methods

6.5.1 Sample preparation

The R14 RNA, sequence 5'- GGACCGGAAGGUCC-3' contained 3 methylated nucleotides in the hairpin loop at positions 6,8 and 9: m^{2}G6, m$_6^2$A8 and m$_2^6$A9. The R14 RNA was prepared by chemical synthesis using methylated phosphoramidites (Rife, J.P., Cheng, C.S., Moore, P.B., Strobel, S.A., submitted manuscript). The methylations should have no appreciable affect on the NOE data for the stem region of the RNA since they are only found in the hairpin loop. The R14 was 2.2 mM in 50 mM NaCl, 5 mM cacodylate (pH 6.3), 1mM EDTA buffer.

The two DNA samples were prepared on an Applied Biosystems 380B DNA synthesizer and purified using denaturing PAGE techniques. Concentrations were determined by UV absorbance measurements at 260nm wavelength and calculated using a dinucleotide stacking extinction coefficient formula. The DNA sequences were (5’ to 3’) D12:CGCGAATTCGCG and D24:CGCGAATTCGCGCGCGAATTCGCG. Both D12 and D24 were palindromic to alleviate any problems with stoichiometry. The samples were dialyzed against 20mM sodium phosphate (pH 7.0) and 100mM NaCl for two days, exchanging the dialysis buffer every 12 hours. Both samples were placed in a Shigemi (Shigemi Corp., Tokyo Japan) NMR tube in a 170 µl volume, which equated to about a 1 cm sample height. The samples were then lyophylized and resuspended in 100.0 atom % D_2O from Aldrich (cat #26,978-6) to the same final sample volume of 170 µl.

6.5.2 NMR experimental

The 2D NOESY data collected for the RNA and DNA samples was performed using a modified version of the canned noesy.c pulse sequence that is supplied with the Varian spectrometers. The modification involved removing the homospoil pulse in the mixing time and replacing it with a z-axis gradient pulse. This insures better removal of COSY-type single and double quantum coupling.

Data for both the D12 and D24 samples were collected at 25º C with a recycle delay of 30 s to insure complete z-magnetization relaxation between scans. Data for the R14 sample was collected at 30º C with w recycle delay of 9.2 s. For all samples, 1024 complex points were collected in the direct dimension, and at least 300 complex points was collected in the indirect dimension. Data processing was accomplished in the direct dimension by applying a 1024 point, 90 degree shifted sine-bell curve to all FIDS. Processing in the indirect dimension was accomplished by applying a 300 point, 90 degree shifted sine-bell curve to all FIDS.

Two-dimensional NOESY experiments were collected for each sample at mixing times ranging from 50-400 ms using very long recycle delays (10–30s) to insure complete Z axis magnetization recovery between scans. Sample spectra are shown in figure 6.6 and a sample one-dimensional slice through the H6 resonance of Thymine #7 for each sample is shown in figure 6.7. Chemical shift assignments for the D12 sample come from those previously published (Nerdal, *et al*., 1989) and D24 assignments followed by simply overlaying the spectra.

6.5.3 Volumes lists

The next few pages contain the volumes experimentally measured for each of the three samples along with the simulated volumes using the best rotational correlation time model.

tc = 0.7 ns

I	atom_j	rij	exp	sim
H8	A 1 GUA H1'	3.73	0.32	0.34
H1'	A 2 GUA H8	4.32	0.38	0.31
H8	A 2 GUA H1'	3.73	0.33	0.25
H1'	A 3 ADE H8	4.32	0.30	0.32
H2	A 4 CYT H1'	3.83	0.48	0.28
H1'	A 3 ADE H8	3.70	0.28	0.25
H1'	A 4 CYT H6	4.30	0.18	0.32
H5	A 4 CYT H6	2.46	2.36	2.36
H6	A 4 CYT H1'	3.52	0.39	0.29
H6	A 5 CYT H1'	3.52	0.31	0.29
H1'	A 6 GUA H8	4.32	0.09	0.31
H1'	A 6 GUA H8	3.73	0.30	0.25
H1'	A 7 GUA H8	0.00	2.60	0.00
H1'	A 9 ADE H8	0.00	0.40	0.00
H1'	A 10 GUA H8	3.73	0.21	0.25
H1'	A 11 GUA H8	4.32	0.27	0.31
H1'	A 11 GUA H8	3.73	0.29	0.25
H1'	A 12 URI H6	4.28	0.20	0.33
H5	A 12 URI H6	2.42	2.31	2.52
H5	A 13 CYT H6	2.46	2.33	2.36
H1'	A 14 CYT H6	4.30	0.29	0.33
H5	A 14 CYT H6	2.46	2.63	2.37
H6	A 14 CYT H1'	3.52	0.36	0.29

ts = 2.8 ns tl = 6 ns

I	atom_j	rij	exp	sim
H6	A 1 CYT H2''	4.27	2.70	1.89
H6	A 1 CYT H1'	3.66	3.50	1.88
H6	A 1 CYT H2'	2.95	9.81	3.00
H1'	A 1 CYT H2''	2.37	10.85	10.36
H1'	A 1 CYT H2'	3.04	3.04	7.87
H2'	A 1 CYT H2''	1.80	22.37	26.30
H2'	A 2 GUA H8	3.45	1.34	1.66
H1'	A 2 GUA H8	3.86	0.19	3.37
H1'	A 2 GUA H2''	3.05	8.71	7.35
H1'	A 2 GUA H2'	2.38	5.73	11.54
H1'	A 3 CYT H5	4.12	1.68	1.46
H1'	A 3 CYT H6	2.95	1.45	3.82
H2'	A 3 CYT H5	3.35	2.14	2.48
H2'	A 3 CYT H6	2.85	0.97	3.97
H8	A 3 CYT H5	4.26	0.65	0.90
H2''	A 3 CYT H5	4.09	2.82	1.51
H2''	A 3 CYT H6	4.35	2.50	2.21
H2''	A 3 CYT H6	3.83	5.12	5.43
H2''	A 3 CYT H1'	2.40	14.22	8.25
H2''	A 3 CYT H2'	1.79	24.57	21.20
H2''	A 4 GUA H8	2.91	1.91	4.20
H1'	A 3 CYT H6	3.72	1.50	1.64
H1'	A 3 CYT H2'	3.07	5.26	5.75
H1'	A 4 GUA H8	4.23	0.41	1.10
H2'	A 3 CYT H5	4.39	2.83	1.87
H2'	A 3 CYT H6	2.35	12.78	9.11
H2'	A 4 GUA H8	3.15	1.43	3.59
H5	A 3 CYT H6	2.46	14.77	14.63
H1'	A 4 GUA H8	3.98	0.68	1.89
H1'	A 4 GUA H2''	3.06	13.61	7.59
H1'	A 4 GUA H2'	2.37	11.52	11.32
H1'	A 5 ADE H8	3.61	2.17	2.07
H8	A 4 GUA H2''	2.45	4.96	10.09
H8	A 4 GUA H2'	3.84	7.93	5.75
H8	A 5 ADE H8	4.74	0.60	0.64
H2'	A 5 ADE H2''	1.77	20.50	26.00
H2'	A 5 ADE H1'	2.37	5.95	10.17
H2	A 6 ADE H2	3.76	2.54	1.74
H2	B 9 CYT H1'	4.18	0.39	0.63
H2''	A 5 ADE H1'	3.05	13.44	7.66
H8	A 6 ADE H1'	3.90	1.17	1.27
H8	A 6 ADE H2'	4.12	8.49	2.43
H8	A 7 THY H6	4.78	0.88	1.15
H8	A 7 THY H7	3.60	7.00	5.36
H2'	A 6 ADE H2''	1.76	19.95	14.86
H2'	A 6 ADE H1'	2.38	7.82	9.18
H2'	A 7 THY H7	4.13	6.61	4.42
H1'	A 6 ADE H2''	3.04	10.19	6.52
H1'	A 7 THY H6	3.84	1.93	2.47
H1'	A 7 THY H7	5.24	0.40	1.42
H2''	A 7 THY H6	3.08	7.16	5.54
H2''	A 7 THY H7	3.39	5.84	5.92
H2	B 8 THY H1'	3.95	0.89	0.87
H1'	A 7 THY H6	3.74	1.26	2.20
H1'	A 7 THY H2''	2.38	12.84	7.83
H1'	A 7 THY H2'	3.05	9.26	5.58
H1'	A 8 THY H6	3.56	1.17	2.42
H2''	A 7 THY H2'	1.78	20.28	17.97
H2''	A 8 THY H7	3.90	6.61	6.10
H6	A 7 THY H7	2.93	13.83	13.54
H6	A 7 THY H2'	2.11	8.76	11.14
H6	A 8 THY H6	4.42	1.08	2.26
H6	A 8 THY H7	3.44	6.56	6.03
H7	A 8 THY H7	3.96	4.01	9.88
H2'	A 8 THY H6	3.27	4.62	6.18
H2'	A 8 THY H7	3.35	6.12	6.64
H6	A 8 THY H7	2.94	12.51	13.84
H6	A 8 THY H1'	3.72	2.38	2.27

```
H6      A 8 THY H2'     2.15 13.06 11.64
H6      A 9 CYT H5      3.56  1.36  1.90
H6      A 9 CYT H6      4.64  2.18  1.55
H7      A 9 CYT H5      4.16  0.51  2.73
H1'     A 8 THY H2''    2.38 11.79  8.42
H1'     A 8 THY H2'     3.05  7.38  5.97
H1'     A 9 CYT H6      3.75  1.78  1.98
H1'     B 6 ADE H2      3.94  0.89  0.89
H2'     A 8 THY H2''    1.78 23.20 20.17
H2'     A 9 CYT H6      3.45  4.15  5.03
H2'     A 9 CYT H5      3.44  3.39  2.51
H2''    A 9 CYT H6      2.39  5.01  7.86
H2''    A 9 CYT H5      3.62  3.36  3.01
H5      A 9 CYT H6      2.43 14.71 13.13
H5      A 9 CYT H2''    5.70  3.92  1.12
H5      A 9 CYT H2'     4.39  2.63  2.33
H6      A 9 CYT H2''    3.74  5.73  6.82
H6      A 9 CYT H1'     3.72  2.67  2.13
H6      A 9 CYT H2'     2.25 13.33 10.96
H1'     A 9 CYT H2''    2.37 10.74  8.86
H1'     A 9 CYT H2'     3.05  4.39  6.18
H1'     A 10 GUA H8     3.60  0.73  2.19
H1'     B 5 ADE H2      4.16  0.39  0.66
H2'     A 10 GUA H8     3.23  2.03  5.21
H2''    A 10 GUA H8     2.50  2.28  7.82
 H2'    A 10 GUA H8     3.68 11.19  8.13
 H2'    A 10 GUA H1'    2.37  7.88  9.69
 H2'    A 11 CYT H6     2.76  1.82  4.37
 H2''   A 10 GUA H8     2.24  6.06 13.06
 H2''   A 10 GUA H1'    3.05  8.26  6.75
 H2''   A 11 CYT H6     4.14  3.03  2.57
 H8     A 10 GUA H1'    3.91  1.07  2.18
 H8     A 11 CYT H5     4.62  0.71  0.44
 H8     A 11 CYT H6     5.14  0.32  0.69
 H1'    A 11 CYT H5     4.86  0.67  0.63
 H1'    A 11 CYT H6     3.37  2.35  2.14
 H1'    A 11 CYT H6     3.70  1.61  2.10
 H1'    A 11 CYT H2''   2.40 14.77  9.49
 H1'    A 11 CYT H2'    3.06  4.91  6.90
 H1'    A 12 GUA H8     4.39  0.69  1.39
 H2'    A 11 CYT H5     4.41  1.66  1.66
 H2'    A 11 CYT H6     2.45 13.88  8.36
 H2'    A 11 CYT H2''   1.79 22.26 25.47
 H2'    A 12 GUA H8     3.54  1.94  4.58
 H2''   A 11 CYT H6     3.91  5.04  5.31
```

```
 H5     A 11 CYT H6     2.46 13.44 15.06
 H6     A 12 GUA H8     5.55  0.13  0.66
 H1'    A 12 GUA H8     3.93  0.92  1.38
 H1'    A 12 GUA H2''   3.00  5.38  9.39
 H1'    A 12 GUA H2'    2.32 15.92 13.46
 H2''   A 12 GUA H2'    1.79 30.03 31.07
H2'     B 1 CYT H6      2.95  9.81  3.00
H2'     B 1 CYT H2''    1.79 22.37 26.32
H2'     B 1 CYT H1'     3.04  3.04  7.86
H2'     B 2 GUA H8      3.45  1.34  1.66
H2''    B 1 CYT H6      4.26  2.70  1.90
H2''    B 1 CYT H1'     2.37 10.85 10.33
H6      B 1 CYT H1'     3.66  3.50  1.88
H1'     B 2 GUA H8      3.86  0.19  3.37
H1'     B 2 GUA H2''    3.05  8.71  7.35
H1'     B 2 GUA H2'     2.38  5.73 11.55
H1'     B 3 CYT H5      4.12  1.68  1.46
H1'     B 3 CYT H6      2.95  1.45  3.83
H2''    B 3 CYT H5      4.09  2.82  1.51
H2''    B 3 CYT H6      4.35  2.50  2.20
H8      B 3 CYT H5      4.26  0.65  0.90
H2'     B 3 CYT H5      3.35  2.14  2.48
H2'     B 3 CYT H6      2.85  0.97  3.97
H5      B 3 CYT H6      2.46 14.77 14.63
H5      B 3 CYT H2'     4.39  2.83  1.87
H6      B 3 CYT H2''    3.83  5.12  5.43
H6      B 3 CYT H1'     3.72  1.50  1.64
H6      B 3 CYT H2'     2.35 12.78  9.11
H2'     B 3 CYT H2''    1.79 24.57 21.22
H2'     B 3 CYT H1'     3.07  5.26  5.76
H2'     B 4 GUA H8      3.15  1.43  3.58
H1'     B 3 CYT H2''    2.40 14.22  8.26
H1'     B 4 GUA H8      4.23  0.41  1.10
H2''    B 4 GUA H8      2.91  1.91  4.19
H2''    B 4 GUA H8      2.46  4.96 10.09
H2''    B 4 GUA H1'     3.06 13.61  7.61
H8      B 4 GUA H1'     3.98  0.68  1.89
H8      B 4 GUA H2'     3.84  7.93  5.74
H8      B 5 ADE H8      4.74  0.60  0.64
H1'     B 4 GUA H2'     2.36 11.52 11.37
H1'     B 5 ADE H8      3.61  2.17  2.07
H1'     B 5 ADE H2''    3.05 13.44  7.67
H1'     B 5 ADE H2'     2.37  5.95 10.18
H2''    B 5 ADE H2'     1.77 20.50 26.00
H2      B 6 ADE H2      3.77  2.54  1.70
```

```
H1'     B 6 ADE H8      3.90  1.17  1.27
H1'     B 6 ADE H2''    3.04 10.19  6.53
H1'     B 6 ADE H2'     2.38  7.82  9.17
H1'     B 7 THY H6      3.84  1.93  2.47
H1'     B 7 THY H7      5.24  0.40  1.42
H8      B 6 ADE H2'     4.12  8.49  2.42
H8      B 7 THY H6      4.78  0.88  1.14
H8      B 7 THY H7      3.60  7.00  5.34
H2'     B 6 ADE H2''    1.76 19.95 14.87
H2'     B 7 THY H7      4.13  6.61  4.42
H2''    B 7 THY H6      3.08  7.16  5.56
H2''    B 7 THY H7      3.39  5.84  5.92
H6      B 7 THY H7      2.93 13.83 13.56
H6      B 7 THY H1'     3.74  1.26  2.19
H6      B 7 THY H2'     2.11  8.76 11.10
H6      B 8 THY H6      4.42  1.08  2.27
H6      B 8 THY H7      3.44  6.56  6.04
H7      B 8 THY H7      3.96  4.01  9.83
H1'     B 7 THY H2''    2.38 12.84  7.81
H1'     B 7 THY H2'     3.05  9.26  5.58
H1'     B 8 THY H6      3.56  1.17  2.43
H2'     B 7 THY H2''    1.77 20.28 17.99
H2'     B 8 THY H6      3.27  4.62  6.22
H2'     B 8 THY H7      3.35  6.12  6.67
H2''    B 8 THY H7      3.90  6.61  6.13
H1'     B 8 THY H2'     3.05  7.38  5.94
H1'     B 8 THY H6      3.72  2.38  2.25
H1'     B 8 THY H2''    2.38 11.79  8.33
H1'     B 9 CYT H6      3.73  1.78  2.08
H2'     B 8 THY H6      2.15 13.06 11.50
H2'     B 8 THY H2''    1.78 23.20 20.02
H2'     B 9 CYT H5      3.45  3.39  2.53
H2'     B 9 CYT H6      3.43  4.15  5.38
H2''    B 9 CYT H5      3.62  3.36  3.07
H2''    B 9 CYT H6      2.35  5.01  8.41
H6      B 8 THY H7      2.94 12.51 13.88
H6      B 9 CYT H5      3.56  1.36  1.91
H6      B 9 CYT H6      4.63  2.18  1.62
H7      B 9 CYT H5      4.16  0.51  2.74
H2'     B 9 CYT H5      4.41  2.63  2.20
H2'     B 9 CYT H6      2.26 13.33 10.51
H2'     B 9 CYT H1'     3.04  4.39  6.32
H2'     B 10 GUA H8     3.23  2.03  5.34
H2''    B 9 CYT H5      5.69  3.92  1.09
H2''    B 9 CYT H6      3.74  5.73  6.67
```

```
H2''    B 9 CYT H1'     2.36 10.74  8.94
H2''    B 10 GUA H8     2.49  2.28  7.92
H5      B 9 CYT H6      2.44 14.71 12.80
H6      B 9 CYT H1'     3.73  2.67  2.10
H1'     B 10 GUA H8     3.59  0.73  2.25
 H2'    B 10 GUA H8     3.70 11.19  7.79
 H2'    B 10 GUA H1'    2.37  7.88  9.68
 H2'    B 11 CYT H6     2.76  1.82  4.37
 H8     B 10 GUA H2''   2.27  6.06 12.29
 H8     B 10 GUA H1'    3.91  1.07  2.11
 H8     B 11 CYT H5     4.63  0.71  0.44
 H8     B 11 CYT H6     5.14  0.32  0.67
 H2''   B 10 GUA H1'    3.04  8.26  6.88
 H2''   B 11 CYT H6     4.13  3.03  2.64
 H1'    B 11 CYT H5     4.86  0.67  0.63
 H1'    B 11 CYT H6     3.37  2.35  2.14
 H2'    B 11 CYT H5     4.43  1.66  1.63
 H2'    B 11 CYT H6     2.47 13.88  8.05
 H2'    B 11 CYT H2''   1.78 22.26 25.28
 H2'    B 11 CYT H1'    3.05  4.91  7.20
 H2'    B 12 GUA H8     3.53  1.94  4.64
 H5     B 11 CYT H6     2.45 13.44 15.38
 H6     B 11 CYT H1'    3.71  1.61  2.09
 H6     B 11 CYT H2''   3.92  5.04  5.14
 H6     B 12 GUA H8     5.54  0.13  0.65
 H2''   B 11 CYT H1'    2.38 14.77  9.93
 H1'    B 12 GUA H8     4.35  0.69  1.46
 H2'    B 12 GUA H2''   1.78 30.03 30.60
 H2'    B 12 GUA H1'    2.32 15.92 13.55
 H2''   B 12 GUA H1'    2.99  5.38  9.48
 H1'    B 12 GUA H8     3.91  0.92  1.41
```

ts = 12 ns tl = 26 ns

```
i          atom_j          rij  exp   sim
   H2'     A 5 ADE H2''    1.77 13.32 11.98
   H2      A 8 ADE H2      0.00  9.89  0.00
   H8      A 6 ADE H1'     3.90  2.90  3.40
   H8      A 7 THY H6      4.78  2.90  4.26
   H2'     A 6 ADE H2''    1.76 10.44  6.80
   H1'     A 7 THY H6      3.84  3.88  5.13
   H2''    A 7 THY H6      3.08  5.42  6.18
   H1'     A 7 THY H6      3.74  2.61  4.63
   H2''    A 7 THY H2'     1.78 12.65  7.74
   H2''    A 8 THY H7      3.90 21.55 15.04
   H6      A 7 THY H7      2.93 21.99 20.68
   H6      A 7 THY H2'     2.11  4.70  7.66
   H6      A 8 THY H6      4.42  2.53  6.07
   H6      A 8 THY H7      3.44 14.92 16.17
   H7      A 8 ADE H8      0.00 17.63  0.00
   H7      A 8 ADE H2''    0.00 18.23  0.00
   H7      A 8 ADE H2'     0.00 19.95  0.00
   H7      A 8 THY H7      3.96 19.45 32.86
   H2'     A 8 THY H7      3.35 15.64 15.56
   H6      A 8 THY H7      2.94 21.66 20.31
   H2'     A 8 THY H2''    1.78 15.58  7.66
   H2'     B 5 ADE H2''    1.77 13.32 11.98
   H2      B 8 ADE H2      0.00  9.89  0.00
   H2'     B 6 ADE H2''    1.76 10.44  6.81
```

```
H2''    B 7 THY H6      3.08  5.42  6.18
H8      B 6 ADE H1'     3.90  2.90  3.40
H8      B 7 THY H6      4.78  2.90  4.26
H1'     B 7 THY H6      3.84  3.88  5.14
H6      B 7 THY H7      2.93 21.99 20.69
H6      B 7 THY H1'     3.74  2.61  4.62
H6      B 7 THY H2'     2.11  4.70  7.66
H6      B 8 THY H6      4.42  2.53  6.09
H6      B 8 THY H7      3.44 14.92 16.17
H7      B 8 ADE H8      0.00 17.63  0.00
H7      B 8 ADE H2''    0.00 18.23  0.00
H7      B 8 ADE H2'     0.00 19.95  0.00
H7      B 8 THY H7      3.96 19.45 32.74
H2'     B 7 THY H2''    1.77 12.65  7.76
H2'     B 8 THY H7      3.35 15.64 15.59
H2''    B 8 THY H7      3.90 21.55 15.08
H2''    B 8 THY H2'     1.78 15.58  7.61
H6      B 8 THY H7      2.94 21.66 20.36
```

6.5.4 R14 structure

```
       rife.pdb

     num res  alpha    beta   gamma     eps    zeta     chi     nu0     nu1     nu2     nu3     nu4     P      numax   pucker

 A    1 GUA    0.00 -179.87   47.43 -151.72  -73.59 -158.02    3.38  -25.82   37.26  -36.18   20.72   13.58  38.34  C3'-endo
 A    2 GUA  -62.10 -179.87   47.44 -151.67  -73.64 -158.01    3.43  -25.85   37.27  -36.15   20.67   13.50  38.33  C3'-endo
 A    3 ADE  -62.09 -179.88   47.44 -151.68  -73.64 -158.97    3.42  -25.84   37.28  -36.18   20.70   13.53  38.34  C3'-endo
 A    4 CYT  -62.07 -179.88   47.41 -151.73  -73.61 -166.06    3.38  -25.84   37.30  -36.20   20.74   13.58  38.38  C3'-endo
 A    5 CYT  -62.07 -179.87   47.43 -151.74  -73.59 -166.06    3.43  -25.87   37.31  -36.20   20.71   13.52  38.37  C3'-endo
 A    6 GUA  -62.12 -179.87   47.44    0.00    0.00 -158.00    3.43  -25.85   37.27  -36.16   20.68   13.51  38.33  C3'-endo
 A    9 CYT  129.63 -179.87   47.43 -151.68  -73.69 -166.06    3.35  -25.80   37.27  -36.21   20.77   13.63  38.35  C3'-endo
 A   10 GUA  -62.06 -179.87   47.44 -151.64  -73.64 -158.03    3.41  -25.83   37.26  -36.16   20.69   13.54  38.33  C3'-endo
 A   11 GUA  -62.07 -179.87   47.44 -151.72  -73.60 -158.01    3.41  -25.85   37.29  -36.20   20.72   13.55  38.36  C3'-endo
 A   12 URI  -62.10 -179.88   47.44 -151.72  -73.64 -165.74    3.35  -25.78   37.25  -36.19   20.76   13.64  38.33  C3'-endo
 A   13 CYT  -62.07 -179.87   47.43 -151.68  -73.66 -166.06    3.35  -25.77   37.24  -36.17   20.74   13.63  38.32  C3'-endo
 A   14 CYT  -62.08 -179.87   47.43    0.00    0.00 -166.06    3.38  -25.82   37.26  -36.18   20.72   13.58  38.34  C3'-endo
```

6.5.5 D12 and D24 structure

```
         dick_nmr.pdb

      num res  alpha    beta   gamma      eps    zeta     chi     nu0     nu1     nu2     nu3     nu4      P     numax   pucker

  A    1 CYT    0.00  -50.84  115.20 -178.52  -86.71 -144.71  -34.63   22.02   -3.03  -16.79   31.89   95.06  34.36  O4'-endo
  A    2 GUA  -62.52 -164.95   56.19  143.08  -93.18  -85.89  -26.97   35.05  -30.43   15.89    6.55  150.65  34.91  C2'-endo
  A    3 CYT  -78.42 -158.07   80.15  176.32  -73.87 -125.61  -28.29   31.64  -24.24    8.79   11.93  139.79  31.74   C1'-exo
  A    4 GUA  -64.41 -153.01   26.73  167.33  -98.36  -92.97  -30.34   35.67  -28.23   11.91   11.10  143.12  35.29   C1'-exo
  A    5 ADE  -73.32 -169.50   61.78  176.47 -107.50 -101.72  -33.46   34.92  -23.94    5.58   17.10  132.68  35.31   C1'-exo
  A    6 ADE  -99.73  167.22  101.00 -167.86  -99.75 -116.22  -31.69   34.06  -24.25    6.81   15.26  135.17  34.19   C1'-exo
  A    7 THY  -21.54 -176.60    2.04 -178.70  -98.46 -109.75  -26.57   33.74  -28.64   14.51    7.38  148.89  33.45  C2'-endo
  A    8 THY  -54.47  172.28   45.60  176.07  -94.16 -114.33  -35.83   35.33  -22.32    2.71   20.23  127.76  36.44   C1'-exo
  A    9 CYT  -70.52 -173.80   54.48  169.58  -82.79 -117.65  -32.84   31.13  -18.89    0.88   19.49  125.14  32.81   C1'-exo
  A   10 GUA  -80.01 -148.25   38.02  162.60  -96.57  -92.14  -28.28   34.91  -29.08   13.79    8.68  147.03  34.67  C2'-endo
  A   11 CYT  -58.95 -173.02   62.42 -170.26 -106.55 -129.53  -34.45   30.93  -16.87   -2.19   22.57  119.95  33.80   C1'-exo
  A   12 GUA  -57.51  162.98   68.25    0.00     0.00 -125.51  -30.59   14.68    4.85  -22.84   33.04   81.61  33.21  O4'-endo

  B    1 CYT    0.00  -60.27  115.24 -178.52  -86.75 -144.65  -34.59   22.00   -3.01  -16.80   31.86   95.04  34.33  O4'-endo
  B    2 GUA  -62.45 -164.96   56.11  143.12  -93.12  -85.86  -26.91   34.99  -30.40   15.89    6.52  150.69  34.86  C2'-endo
  B    3 CYT  -78.46 -158.07   80.17  176.30  -73.86 -125.56  -28.25   31.60  -24.23    8.82   11.88  139.86  31.70   C1'-exo
  B    4 GUA  -64.37 -153.02   26.70  167.34  -98.37  -92.93  -30.33   35.62  -28.20   11.89   11.13  143.10  35.27   C1'-exo
  B    5 ADE  -73.36 -169.50   61.81  176.50 -107.56 -101.74  -33.47   34.89  -23.90    5.51   17.15  132.60  35.31   C1'-exo
  B    6 ADE  -99.78  167.21  101.01 -167.85  -99.75 -116.19  -31.64   33.99  -24.20    6.80   15.25  135.16  34.13   C1'-exo
  B    7 THY  -21.49 -176.61    1.98 -178.70  -98.46 -109.76  -26.51   33.69  -28.60   14.52    7.32  148.95  33.38  C2'-endo
  B    8 THY  -54.44  172.27   45.66  176.09  -94.09 -114.37  -35.91   35.32  -22.22    2.60   20.31  127.55  36.45   C1'-exo
  B    9 CYT  -70.59 -173.77   54.46  169.63  -82.73 -117.63  -32.58   31.04  -18.98    1.08   19.19  125.55  32.64   C1'-exo
  B   10 GUA  -79.94 -148.26   38.09  162.69  -96.73  -92.15  -28.13   34.83  -29.05   13.85    8.57  147.18  34.57  C2'-endo
  B   11 CYT  -58.79 -173.05   62.15 -170.07 -106.43 -129.22  -34.05   30.66  -16.85   -2.03   22.12  120.27  33.42   C1'-exo
  B   12 GUA  -57.54  162.98   68.23    0.00     0.00 -125.42  -30.46   14.68    4.81  -22.76   32.88   81.64  33.09  O4'-endo
```

6.5.6 YARM scripts

The next two pages contain the YARM scripts used to calculate the correlation plots presented in this chapter, one for calculating the isotropic correlation plot (Fig. 6.10) and the other for calculating the anisotropic correlation time plots (Figs. 6.11-6.13). The third YARM script was used to generate the statistical analysis for a specific rotational model and was used to make the volume lists found in section 6.5.3.

```
                perl
 yarm Yet Another Relaxation Matrix program


                   yarm/yarm_lib.pl";




pdb_file = "d12_b.pdb";
f95_vol_file = "d12_30s.vols";
f95_peak_file = "d12_30s.peaks";

sfreq = 600;
vol0 = 100;
tmix = 0.2;

debug = 0;


     yarm modules

       yarm_version\n";
                             rij matrix and principal axis


xyz = &Pdb_Read_All( $pdb_file );
xyz = &Get_Atom_Type( \%xyz, \%nonX_NA );
xyz = &Pseudo_Methyl(\%xyz);

rij = &Rij_Hash( \%xyz, 0, 10 );

      Ay, $Az ) = Principal_Axis( \%xyz );


       $Ax, $Ay, $Az);


vol_exp = &F95_Read_Merge( $f95_vol_file, $f95_peak_file );
vol_exp = &Make_Symm_Molecule( A, B, \%vol_exp );
```

```
               tl_ts_aniso.out" );

    ts = 0.1; $ts < 10; $ts += 0.25) {
    tl = 0.1; $tl < 10; $tl += 0.25) {

      print "Working on tl=$tl and ts=$ts\n";

      # Calculate anisotropic rotation volumes
      %vol_sim = &Sim_Vol( $sfreq, $tmix, $vol0, \%xyz,
rij, $tl, $ts,
             $Ax, $Ay, $Az, \%S );

      # Normalize the experimental volumes by comparing h5-h6

      %vol_exp = &Norm_Hash( \%vol_exp, \%vol_sim);

      # Determine gobs of information out about the  pairwise

      # comparison between the experimental and isotropic

      ( $rms, $r, $q, $q6 ) = &Stats( \%vol_exp, \%vol_sim );

      print "finished tl=$tl ts=$ts RMS=$rms\n\n";
      print REPORT "$rms\n";
```

```
                 perl
 yarm Yet Another Relaxation Matrix program


                    yarm/yarm_lib.pl";




pdb_file = "dick_b.pdb";
f95_vol_file = "d12_30s.vols";
f95_peak_file = "d12_30s.peaks";

sfreq = 600;
vol0 = 100;
tmix = 0.2;




     yarm modules

       yarm_version\n";
                              rij matrix and principal axis


xyz = &Pdb_Read_All( $pdb_file );
xyz = &Get_Atom_Type( \%xyz, \%nonX_NA );
xyz = &Pseudo_Methyl(\%xyz);

rij = &Rij_Hash( \%xyz, 0, 10 );


vol_exp = &F95_Read_Merge( $f95_vol_file, $f95_peak_file );
vol_exp = &Make_Symm_Molecule( A, B, \%vol_exp );

                tc_iso.out" );
     tc = 0.1; $tc < 10; $tc += 0.1) {

   print "Working on tc=$tc\n";
```

```
  # Calculate isotropic rotation volumes
vol_sim = &Sim_Vol( $sfreq, $tmix, $vol0, \%xyz, \%rij, $tl


  # Normalize the experimental volumes by comparing h5-h6

  %vol_exp = &Norm_Hash( \%vol_exp, \%vol_sim );

  # Determine gobs of information out about the  pairwise

  # comparison between the experimental and isotropic volumes
  ( $rms, $r_factor, $q, $q6 ) = &Stats( \%vol_exp,
 vol_sim);

  print "finished tc=$tc  RMS=$rms\n\n";
  print REPORT "$tc $rms\n";
```

```
                perl
                   yarm/yarm_lib.pl";



pdb_file = shift;
sfreq = 600;
vol0 = 100;
tmix = 0.25;

tl = 2.8;
ts = 6.5;


f95_vol_file = "d12_30s.vols";
f95_peak_file = "d12_30s.peaks";

  f95_vol_file = "d24_200_selected_vols.txt";
  f95_peak_file = "d24_200_selected_peaks.txt";



 set to 2 for TONS of debugging messages (lots!)
debug = 0;




       yarm_version\n";


xyz = &Pdb_Read_All( $pdb_file );
xyz = &Get_Atom_Type( \%xyz, \%nonX_NA );
xyz = &Pseudo_Methyl(\%xyz);

                     xyz ) { $S{$atom} = $S; }



rij  = &Rij_Hash( \%xyz );
```

```
vol_exp = &F95_Read_Merge( $f95_vol_file, $f95_peak_file );
                                             segids A and B

vol_exp = &Make_Symm_Molecule( A, B, \%vol_exp );

atoms = keys %vol_exp;
                          atoms\n";




       Ay, $Az ) = Principal_Axis( \%xyz );

                  Ay, $Az);
vol_sim = &Sim_Vol( $sfreq, $tmix, $vol0, \%xyz, \%vol_exp,
   $tl, $ts, $Ax, $Ay, $Az, \%S );


 print "Simulating NOE volumes using isotropic-rigid...\n";
  vol_sim = &Sim_Vol( $sfreq, $tmix, $vol0, \%xyz, \%rij, $tc



vol_exp = &Norm_Hash( \%vol_exp, \%vol_sim );



debug=1;
  rms, $r, $q, $q6 ) = &Stats( \%vol_exp, \%vol_sim );




                          rms value
        Pairwise statistical analysis:\n";
                                    rms);





atoms_i = keys %vol_exp;
atoms_i = Sortme ( \@atoms_i );
```

```
     atom_i ( @atoms_i ) {
@atoms_j = keys %{ $vol_exp{$atom_i} };
@atoms_j = Sortme ( \@atoms_j );
foreach $atom_j ( @atoms_j ) {
    next if ( ($used{$atom_i}{$atom_j}) or
        ($used{$atom_j}{$atom_i}) );
    $used{$atom_i}{$atom_j} = "t";

    $v_e = $vol_exp{$atom_i}{$atom_j};
    $v_s = $vol_sim{$atom_i}{$atom_j};
    $rij = $rij{$atom_i}{$atom_j};
    printf ("%-15s %-15s %4.2f %5.2f %5.2f\n",
        $atom_i, $atom_j, $rij, $v_e, $v_s);
}
```

6.6 References

Alms GR, Bauer DR, Brauman JI, Pecora R. 1973. *J. Chem. Phys* *59*:5310-5321.

Debye P. 1929. *Polar Molecules*. New York: Dover.

Drew HR, Wing RM, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE. 1981. Structure of a B-DNA dodecamer: conformation and dynamics. *PNAS* *78*:2179-2182.

Eimer W, Pecora R. 1991. Rotational and translational diffusion of short rodlike molecules in solution: Oligonucleotides. *J. Chem. Phys.* *94*:2324-2329.

Eimer W, Williamson JR, Boxer SG, Pecora R. 1990. Characterization of the overall and internal dynamics of short oligonucleotides by depolarized dynamic light scattering and NMR relaxation measurements. *Biochemistry* *29*:799-811.

Einstein A. 1956. *Investigations into the Theory of the Brownian Movement*. New York: Dover.

Kivelson D. 1987. *Rotational Dynamics of Small and Macromolecules*. Heidelberg: Springer.

Lapham J, Rife J, Moore PB, Crothers DM. 1997. Measurement of diffusion constants for nucleic acids by NMR. *J. Biomolecular NMR* *10*:255-262.

Nerdal W, Hare DR, Reid BR. 1989. Solution structure of the *Eco*RI DNA sequence: refinement of NMR-derived distance geometry structures by NOESY spectrum back-calculations. *Biochemistry* *28*:10008-10021.

Perrin F. 1934. *J. Phys. Rad.* *5*:497.

Perrin F. 1936. *J. Phys. Rad.* *7*:1.

Tirado MM, Martinez CL, Garcia de la Torre J. 1984. Comparison of theories for the translational and rotational diffusion coefficients of rod-like macromolecules. Application to short DNA fragments. *J Chem Phys* *81*:2047-2052.

Tirado MM, Garcia de la Torre J. 1979. Translational friction coefficient of rigid, symmetric top macromolecules. Application to circular cylinders. *J Chem Phys* *71*:2581-2587.

Tirado MM, Garcia de la Torre J. 1980. Rotational dynamics of rigid, symmetric top macromolecules. Application to circular cylinders. *J Chem Phys* *73*:1986-1993.

Garcia de la Torre J, Martinez MCL, Tirado MM. 1984. Dimensions of short, rodlike macromolecules from translational and rotational diffusion coefficients. Study of the gramicidin dimer. *Biopolymers* *23*:611-615.

Zimm BH. 1956. Dynamics of polymer molecules in dilute solution: viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.* *24*:269-278.

CHAPTER 7 "YARM"

7.1 Summary

This chapter introduces a computer program named Yet Another Relaxation Matrix, or YARM for short. The purpose of YARM is to simplify the analysis of NOESY crosspeak intensity data by creating a common framework around which data analysis programs can be built. Two general uses of YARM will be presented in this chapter; NMR model verification and NMR model refinement.

NMR model verification: Using the relaxation matrix approach, YARM can calculate a set of simulated NOESY crosspeak intensities for a proposed model. These simulated volumes can then be compared to experimentally measured volumes in a quantitative manner. This gives a statistical measure of the "correctness" of the proposed model by directly comparing it to the NMR derived data.

NMR model refinement: YARM provides a mechanism refining a proposed model. Using a least-squares approach, the NMR derived model can be adjusted until the the simulated NOE volumes from a given model match the experimentally measured volumes.

7.2 Introduction and Background

Determination of NMR derived structures relies heavily on the interpretation of the NOE to determine distances between protons. These interproton distance constraints are then used to to calculate a molecular coordinate set that best meets the requirements of the constraints. Of fundamental importance in the process of structure determination is the method by which one calculates the distance constraint from the measured NOE, a number of approaches have been utilized.

Historically, interpretation of experimentally measured NOE intensities has naturally progressed from very simple methods to more complex. One of the earliest studies employing the NOE in a biomolecular structure determination was by Wagner and Wüthrich in the assignment of bovine pancreatic trypsin inhibitor (1982) in which the NOE connectivities were used only to determine sequence information. Kumar et al. (1981) introduced the concept of using the NOE crosspeak intensity as a method of quantifying distances. They proposed the initial concepts of what would be called the "Isolated Spin Pair Approximation" (Reid, 1987; Patel, et al., 1987; Clore and Gronenborn, 1985; Clore and Gronenborn, 1989), which assumes that any two nuclei are relatively isolated from all other nuclei in NOESY experiments at short mixing times. This allows for a simple method of distance determination by "relative intensity comparison" of the NOE between a spin pair of known distance to that of a spin pair of unknown distance.

$$\frac{s_{ref}}{s_{ij}} = \frac{r_{ij}^6}{r_{ref}^6} \qquad 7.1$$

For nucleic acids this is often accomplished by measuring the cross-relaxation rate of the H5-H6 crosspeak in cytosine or (for RNA only) uridine, with a fixed distance interproton distance of 2.46Å. Often the semi-quantitative method of the ISPA is simplified by defining NOE intensities as strong, medium or weak and giving large bounds to the restraint distances.

Spin isolation is rarely observed in biological systems and the ISPA method of distance determination can give incorrect distance values, with errors of up to 1.3 Å (Wemmer, 1991; Reid, 1989; Nerdal, 1989; Schmitz & James, 1995). To account for "spin-diffusion" effects, Keepers and James (1984) realized that dipolar relaxation between nuclei must be accounted for simultaneously for all spin pairs. They use the rate (or relaxation) matrix method for analyzing NOE data in the same manner that chemical exchange processes had been done previously.

The concept of using the relaxation matrix in the analysis of NOE intensities has given rise a number of computer programs that exploit this approach for NMR structure determination: IRMA (Boelens, *et al*., 1988), MARDIGRAS (Borgias & James, 1990) and MORASS (Post, *et al*., 1990). All of these methods, however, have made the assumption that the rotational diffusion of the molecule in question can be adequately described using an isotropic rotation model. The isotropic definition of the spectral density function and originally proposed by Bloembergen, Purcell and Pound (1948) is utilized (see Chapter 5, section 5.3.1).

The problem in this assumption has been pointed out a number of times in the literature (Birchall & Lane, 1990; Schmitz & James, 1995). Nucleic acids may be especially affected by this, as was noted by Withka *et al*. (1990) in which they state "The asymmetry of the duplex DNA complicates the straightforward analysis of NOE data in term of conformational analysis." That is, the rotational diffusion of DNA is asymmetric. The consequence of this is that an internuclear spin-pair vector, parallel to the long axis of a DNA, experiences different fluctuating magnetic fields than a vector perpendicular to the long axis. This will cause differential relaxation effects that would affect the NOE intensity differently.

The computer program YARM was initially created to incorporate the anisotropic rotation definition of the spectral density function into the relaxation matrix calculations.

7.3 YARM

The theoretical basis for these calculations is presented in chapter 5 of this thesis and should be consulted, however, a quick overview of the relaxation matrix method will be given. An example will be shown for evaluating the model of a DNA using a few of the more commonly used structural analysis YARM scripts, along with a description of how each script works. A quick overview will also be given for the structural refinement component of YARM.

Finally, for the programmers (and other interested in how the calculations are performed) the two C++ object definition files and the structural refinement program are included for your perusal. The object definition files are the heart and soul of the mathematical calculations, with all the fodder striped away and should be consulted for a complete understanding of the YARM calculations. It should be noted that the entire source code for YARM is too large to include in this thesis. If interested, see the web page at (http://bass.chem.yale.edu/~lapham/yarm/) where the full code can be downloaded.

7.3.1 Overview of simulating NOE initensities

As mentioned earlier, the full theoretical treatment of using the relaxation matrix to perform NOE intensity simulations is presented in chapter 5. This section is simply a broad overview of the process.

The NOE crosspeak volume matrix, $\mathbf{V}(t_{mix})$, is related to the relaxation matrix, **R**, by the following equation,

$$\mathbf{V}(t_{mix}) = \mathbf{V}(0)\exp[\mathbf{R}t_{mix}] \qquad 7.3$$

Thus, if the relaxation matrix can be accurately constructed, the NOE volumes will be accurately calculated. This is, however, the difficult part of the process. The elements in the relaxation matrix are composed of functions that relate the properties of molecular structure, rotational motion and intramolecular motion to the cross-relaxation rates. The first of these properties is the "molecular structure" of the molecule, or the X, Y and Z Cartesian coordinates of the time-averaged position of each atom. The second is the "rotational motion" of the molecule, also commonly referred to as the correlation time. The third is the "intramolecular motion" of the molecule, a description of the .dynamical movements between atoms on the same molecule.

In YARM, we call these three properties the “model” of the molecule. Thus, a “model” of a biomolecule is not just a description of the coordinate structure, it would also require a description of the tumbling rate and the intramolecular dynamics. Figure 7.1 below demonstrates the overall process of calculating NOE volumes from this “model” and where in the calculations each part of the model is used. For instance, the two motional components of the model are used in converting the relaxation matrix **R** to the distance matrix **r**.

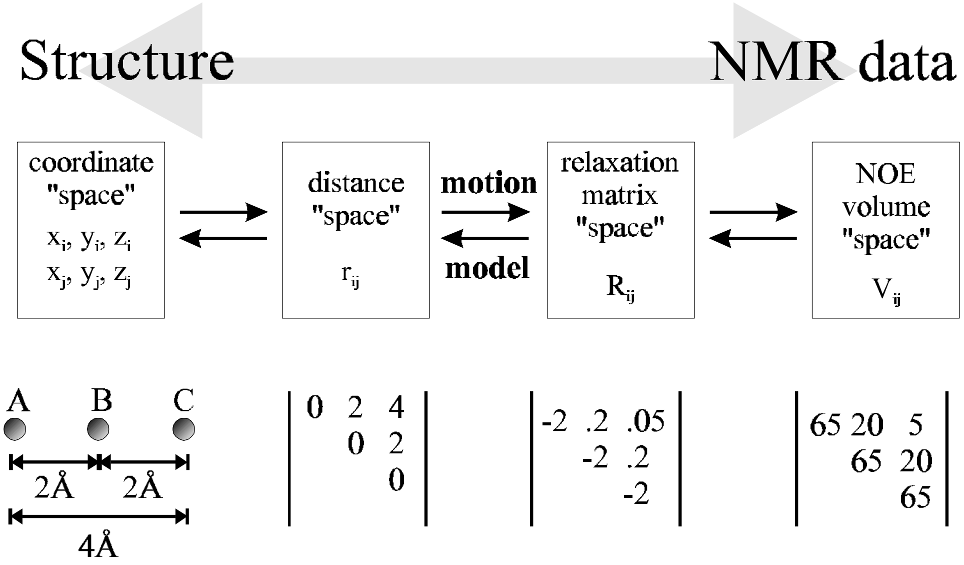


Figure 7. 1 YARM data flowchart

Theoretically, if the three components of the model were perfectly well known, the conversion from coordinate space to NOE volume space would be exactly correct and reversible. This, of course, rarely occurs with experimentally derived data. Often, only a subset of the possible NOE crosspeak volumes are assigned, or are resolved enough for accurate quantitation. Because of this, divining the structure of a molecule based on the NOE intensities is not as straightforward as performing the reverse calculation, shown above.

Often in the world of biomolecular structure determination, assumptions must be made about one or more of the model components. For instance, the rotational correlation time of a molecule is a difficult quantity to measure experimentally, and often it must be estimated. A firm understanding of intramolecular dynamics can be just as elusive; it is often difficult to distinguish between a rigid structure and a two conformation state structure in fast exchange on the NMR time scale. While it may be difficult to exactly determine these quantities, they cannot be ignored without compromising the integrity of the analysis.

It is my opinion that the biomolecular NMR spectroscopist who proposes a molecular "structure" that "best fits their NMR data" must back the statement up with a statistical analysis. The motional components of the model must be proposed, as they are just as important calculations as the atomic coordinates. This can be as simple as stating "we assume an isotropic rotation model with correlation time of 5 ns and a rigid structure". Even if there is no conclusive data to support this, it must be stated to allow for discussion of the structural model. Finally, a statistical analysis of the actual NOE data measured for the molecule can then be presented. Thus, rather than a qualitative "this structural model fits the data" there can be a quantitative report on how *well* it fits the data.

7.3.2 Statistical analysis of volume sets

A number of statistical methods for comparing NOE volumes have been developed. The YARM subroutine "Stats" returns a list of each of these functions in the order shown below:

```
($rms, $r, $q, $q6) = &Stats( \%vol1, \%vol2);
```

Where the definitions of the statistical functions are:

$$RMS = \left[\frac{\sum_{ij} \left\{ VolExp_{ij} - VolSim_{ij} \right\}^2}{\left\{ \sum_{ij} VolExp_{ij}^2 + \sum_{ij} VolSim_{ij}^2 \right\}} \right]^{\frac{1}{2}} \qquad 7.4$$

$$R-\ factor(R)=\frac{\sum_{ij}\left|VolExp_{ij}-VolSim_{ij}\right|}{\sum_{ij}VolExp_{ij}} \qquad 7.5$$

$$Q-\ factor(Q)=\frac{\sum_{ij}\left|VolExp_{ij}-VolSim_{ij}\right|}{\sum_{ij}VolExp_{ij}+\sum_{ij}VolSim_{ij}} \qquad 7.6$$

$$Q^{1/6}-\ factor(Q^{1/6})=\frac{\sum_{ij}\left|VolExp_{ij}^{1/6}-VolSim_{ij}^{1/6}\right|}{\sum_{ij}VolExp_{ij}^{1/6}+\sum_{ij}VolSim_{ij}^{1/6}} \qquad 7.7$$

7.3.3 Model Validation

An example of how the structural analysis works is presented. The Dickerson dodecomer DNA, 5'-CGCGAATTCGCG-3' is a symmetric self-complementary dimer which has been studied extensively by NMR and X-ray crystallography techniques. NMR NOESY data were collected on the DNA (see chapter 6) and the resolved NOE crosspeak volumes were measured quantitatively, a total of 225 volumes in all.

We begin the analysis by arbitrarily proposing the following two models for the DNA (of course, the structural biologists would want to use the structures derived from their XPLOR calculations, and the like).

Property	MODEL #1	MODEL #2
Atom coordinate positions:	A-form DNA	B-form DNA
Molecular tumbling:	isotropic, 5ns	anisotropic, 2 and 6 ns
Intramolecular dynamics:	rigid	S^2=0.9

The "correctness" of the two models can be examined quantitatively by comparison of the back-calculated NOE intensities to the actual experimental data using the YARM scripts, model1.pl and model2.pl (see section 7.3.3). The script model1.pl simulates the NOE intensities using the first proposed model, and model2.pl uses the second proposed model. The following is the output from these programs:

```
bass (lapham): [~/yarm_thesis]> ./model1.pl dick_a.pdb
YARM v0.9 February 22, 1998
Simulating NOE volumes using isotropic-rigid...
Pairwise statistical analysis:
           RMS = 0.5128
      R-factor = 0.7508
      Q-factor = 0.3754
Q^(1/6)-factor = 0.1947

bass (lapham): [~/yarm_thesis]> ./model2.pl dick_b.pdb
YARM v0.9 February 22, 1998
Simulating NOE volumes using anisotropic S=0.9...
Principal axis vector components Ax=0.01 Ay=-0.03 Az=1.00
Pairwise statistical analysis:
           RMS = 0.2888
      R-factor = 0.4162
      Q-factor = 0.2081
Q^(1/6)-factor = 0.0882
```

Clearly the second model is a better fit to our experimental data, but we can be more quantitative than that. The second model fits the experimental data with a rms of 0.29, an R-factor of 0.42, a Q-factor of 0.21 and a $Q^{1/6}$-factor of 0.088.

Additionally, the YARM scripts saved a correlation plot of the simulated data versus the experimental data in a file, so the accuracy of the fit can be viewed graphically, as shown below in figure 7.2.

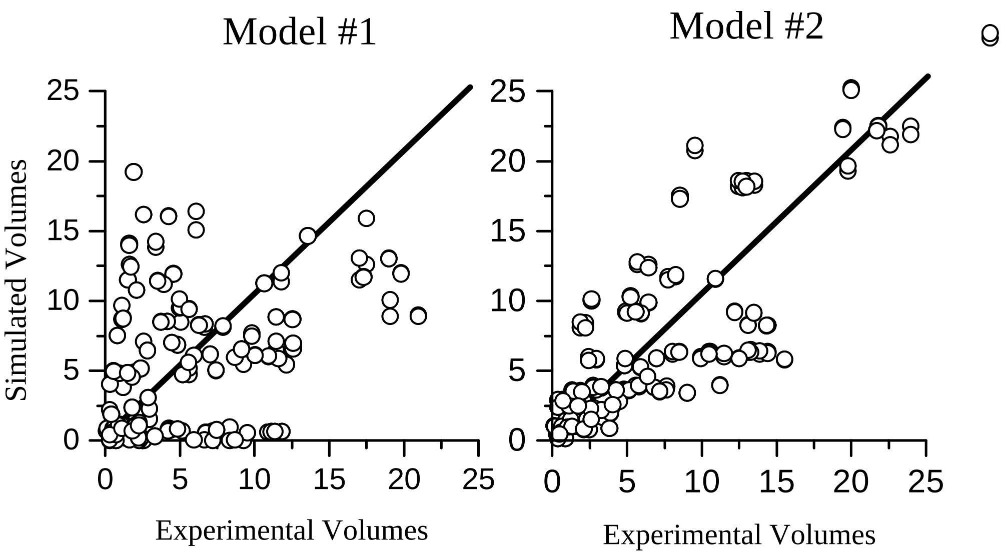


Figure 7. 2 YARM Correlation plots

It is clear that the first model is a better fit to the data, statistically, than the second model. The correlation plots are simply a visual way of coming to the same conclusion, the second model is better correlated to the experimental data.

Incidentally, neither of these models simulate the experimental NOEs very well, the best fit to the data comes from yet another proposed model for this DNA; see chapter 6 for a full discussion.

The model1.pl and model2.pl scripts are presented at the end of this section. Notice that the scripts are written in the Perl scripting language. YARM is actually nothing more then a series of perl subroutines. The first YARM subroutine called in the model1.pl script is:

```
%xyz = &Pdb_Read_All( $pdb_file );
```

The Pdb_Read_All YARM subroutine reads in all the atom names and positions from a PDB formatted structure file. The atom names and coordinates are then stored in the variable %xyz for use in other YARM subroutines. The actual calculation of the simulated NOE volumes comes from the line:

```
%vol_sim = &Sim_Vol( $sfreq, $tmix, $vol0, \%xyz, \%rij, $tc );
```

The YARM subroutine Sim_Vol simulates volumes! It needs to know the spectrometer frequency ($sfreq), mixing time ($tmix), normalized volume ($vol0), atom names and coordinate (\%xyz), which atom pairs to return (\%rij) and the correlation time ($tc). It then returns to the variable %vol_sim the results of the calculation.

The model2.pl script uses the Sim_Vol subroutine in a slightly different manner:

```
%vol_sim = &Sim_Vol( $sfreq, $tmix, $vol0, \%xyz, \%rij, $tl, $ts, $Ax, $Ay, $Az, \%S );
```

Notice that the first five arguments are the same as those in model1.pl, but now two correlation times ($tl and $ts), the vector components of the principal axis of rotation ($Ax, $Ay, $Az), and an Order Parameter (\%S) have been included. That is because the "model" for this script uses anisotropic rotation and includes an order parameter of 0.9.

This is just a short description of the scripts. The web page has many more example scripts and more interesting uses of the program. Additionally, each of the subroutines is explained in much greater detail.

7.3.4 Model refinement

In addition to model verification, YARM contains a structural refinement component. From a given rotational and intramolecular dynamic model, YARM will find the set of Cartesian coordinates that best fit the NOE data.

We ask the simple question: Does the comparison of the simulated and experimental NOEs suggest that an atom pair move closer together, or farther apart? If an atom pair A and B have an experimentally determined NOE volume of 20 and a simulated NOE volume of 10, the two atoms in the model should be moved closer together. The distance they should move will be roughly proportional to the difference in the 1/6 root of the volumes. This is known as the residual function, r:

$$residual_{ij} = VolSim_{ij}^{1/6} - VolExp_{ij}^{1/6} \qquad 7.8$$

The goal of this structure refinement process is to minimizing this function for all atom pairs. The direction each atom should be moved in order to minimize all atom pair interactions is determined by taking the vector sum of all residuals for each individual atom. This overall movement vector is known as the gradient, and is shown below as the thicker line (the vector sum of the thinner lines).

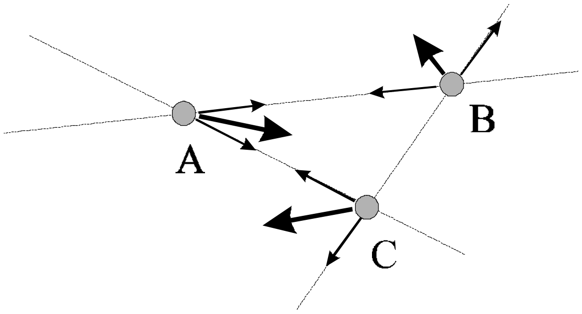


Figure 7. 3 The gradient vector

Movement along the gradient will result in a minimizing of the function defined in equation 7.8.

An example YARM script for model refinement is presented on the next page. The subroutine call that actually performs the calculations is:

```
%xyz2 = &Structure_Refine( \%xyz, \%vol_exp, $num_iter, $sfreq, $tmix, $vol0, $tc );
```

In which the &Structure_Refine subroutine returns a new coordinate hash (in this case, named %xyz2) that can be used as any other coordinate hash in YARM.

7.3.5 Other software packages

As the name of this program implies, there are a number of computer programs available that calculate the relaxation matrix. This begs the question, why write another? I am glad you asked, because I would like to tell you why. This software package was written with the express intention that people can use it to LEARN about calculating the relaxation matrix. It would appear to this author that often the details of HOW calculations are performed are hidden from the end users. Hopefully, it will be clear what parts of the calculation are robust and what parts involve a certain level of assumptions. Great pain have been put forth to separate the code into its constituent parts, for example, if you want to know the mathematics behind calculating a cross relaxation rate member of the relaxation matrix, simply look in the **nmr_relax.c** file under the subroutine "rij2rate_iso". In fact, this is the file in which most of the real calculations are performed. This code is completely removed from the code that manipulates the input and output files, etc.

A quick overview of two other programs available for calculating the relaxation matrix are presented here. It should be stated that it is not with the intent of supplanting the existing rate matrix calculation software that YARM was written. Rather, it is the intention of the author that they are used for the development of new ideas, which require the "relaxation matrix" framework around which to work.

MORASS, Multispin Overhauser Relaxation AnalysiS (Post, et al., 1990; Meadows, et al., 1994) was used initially to understand how one codes these types of programs. The authors kindly release their FORTRAN code with the program at no charge (anonymous ftp dggp12.chem.purdue.edu). This program suffers from the usual problem that all FORTRAN programs suffer from, obscure code. While the authors do make the code publically available, it is nearly impossible to follow the data flow and actually know HOW the calculations are performed. The program also suffers from the “problem” of being a complete software package, it is difficult to integrate it into other calculations or even to modify it.

Another program, Matrix Analysis of Relaxation for DIscerning the Geometry of an Aqueous Structure or MARDIGRAS (Borgias and James, 1990), can be purchased from the regents of the University of California. Professor Thomas James was kind enough to supply the code for this program for the purposes of recompiling it for the LINUX operating system. This is mentioned here because it should be stated that none of the code was examined or used in the creating of the programs written in this section. MARDIGRAS is, once again, not available for free and the FORTRAN source code is not available. Thus, it must be used as a “black box” in which you must trust is performing the calculations correctly. As with MORASS, it is difficult to incorporate into other calculations and impossible to modify.

7.3.6 Source code: nmr_relax.c and nmr_relax.h

These two files define the object “NmrParams”. It is in this object definition that all the NMR relaxation calculations take place.

```
// default constructor and destructor
NmrParams();
~NmrParams();

// set functions
void setNmrParams( float, float, float, float, float, float

void setTc( float );
void setTl( float );
void setTs( float );
void setVol0( float );
void setTmix( float );
void setSfreq( float );

// get functions
float getTc() const;
float getTl() const;
float getTs() const;
float getVol0() const;
float getTmix() const;
float getSfreq() const;
float getW0AB() const;
float getW1AB() const;
float getW2AB() const;
float getW1AA() const;
float getW2AA() const;

// calculation functions
void calcTransIso();                    // Calculate the
```

```
void calcTransAniso( float, float );    // Calculate the

double rij2rho_iso( int, Structure * );
double rij2rho_aniso( int, NmrParams, Structure * );
double rij2sigma( int, int, double );
double sigma2rij( int, int, int, int, double );

// print functions
void printNmrParams() const;
void printTransitionRates() const;


float W0AB, W1AB, W2AB;  // Transition rates
float W1AA;              // Self dipole transition rate

float W2AA;              // Self dipole transition rate

float tc;                // Iso correlation time (ns)
float tl;                // Aniso long axis correlation

float ts;                // Aniso short axis correlation

float sfreq;             // Spectrometer frequency (MHz)
float tmix;              // NOE mixing time (s)
float vol0;              // Normalized autopeak volume at
```

```
NmrParams Class


W0AB = W1AB = W2AB = W1AA = W2AA = tc = tl = ts = 0;
sfreq = tmix = vol0 = 0;


W0AB = W1AB = W2AB = W1AA = W2AA = tc = tl = ts = 0;
sfreq = tmix = vol0 = 0;


tc = tc_in;
tl = tl_in;
ts = ts_in;
sfreq = sfreq_in;
tmix = tmix_in;
```

```
vol0 = vol0_in;


/****
 Calculate the isotropic transition rates, W0, W1 and W2
****/

double sfreq_rad = (sfreq / 1000) * (180 / PI); // GHz

// Use the isotropic definition of the spectral density fns
double J0 = 2 * tc;
double J1 = 2 * tc / (1 + pow((sfreq_rad * tc), 2 ) );
double J2 = 2 * tc / (1 + pow((2*sfreq_rad * tc), 2) );

W0AB =  0.5 * Q * J0;
W1AB = 0.75 * Q * J1;
W2AB =    3 * Q * J2;

// These calculation have not been implemented yet
```

```
W1AA = 0;
W2AA = 0;





/***********************************************
Calculates the W0AB, W1AB and W2AB transition rates using
an anisotropic rotation model for the spectral density
function
beta = angle WRT principal axis of rotation
S2 = order parameter
***********************************************/
// Declare local variables
double a1, a2, a3;
double t1, t2, t3;
double j1, j2, j3;
double J0, J1, J2;

double beta_rad  = beta * (  2 * PI / 360 );
double sfreq_rad = (sfreq / 1000) * ( 2 * PI / 1 ); //


// Calculate the angle dependent coefficients a1, a2 and a3
a1 = 0.25 * pow( (3 * pow(cos(beta_rad), 2) - 1 ), 2);
a2 = 3.00 * pow(cos(beta_rad), 2) * pow(sin(bet a_rad), 2);
a3 = 0.75 * pow(sin(beta_rad), 4);

// Calculate the t1, t2 and t3 values
t1 = tl;
t2 = 6 * tl * ts / (tl + (5 * ts));
t3 = 3 * tl * ts / (ts + (2 * tl));

// Calculate J0
j1 = 2 * t1;
j2 = 2 * t2;
j3 = 2 * t3;
J0 = a1 * j1 + a2 * j2 + a3 * j3;

// Calculate J1
j1 = 2 * t1 / (1 + pow((sfreq_rad * t1), 2) );
j2 = 2 * t2 / (1 + pow((sfreq_rad * t2), 2) );
j3 = 2 * t3 / (1 + pow((sfreq_rad * t3), 2) );
```

```
J1 = a1 * j1 + a2 * j2 + a3 * j3;

// Calculate J2
j1 = 2 * t1 / (1 + pow((2 *  sfreq_rad * t1), 2) );
j2 = 2 * t2 / (1 + pow((2 * sfreq_rad * t2), 2) );
j3 = 2 * t3 / (1 + pow((2 * sfreq_rad * t3), 2) );
J2 = a1 * j1 + a2 * j2 + a3 * j3;

// Transition rates (sans the r^-6 component)
W0AB =  0.5 * Q * J0 * S2;
W1AB = 0.75 * Q * J1 * S2;
W2AB =    3 * Q * J2 * S2;


/*
cout << "Beta=" << beta << " Beta_rad=" << beta_rad << "

cout << "Sfreq=" << sfreq << " Sfreq_rad=" << sfreq_rad <<

cout << "tl=" << tl << " ts=" << ts << endl;
cout << "a1=" << a1 << " a2=" << a2 << " a3=" << a3 <<

cout << "t1=" << t1 << " t2=" << t2 << " t3=" << t3 <<

cout << "J0=" << J0 << " J1=" << J1 << " J2=" << J2 <<

cout << "W0=" << W0 << " W1=" << W1 << " W2=" << W2 << endl

*/






cout << " Current values for the NMR Parameters" << endl;
cout << " tc=" << tc << " tl=" << tl << " ts=" << ts << "

cout << tmix  << " sfreq="  << sfreq << " vol0=" << vol0 <<
cout << " Current values for the Transition Rates" << endl;
cout << " W0AB=" << W0AB << " W1AB=" << W1AB << " W2AB=" <<

cout << " W1AA=" << W1AA << " W2AA=" << W2AA << endl <<




/*******
  Calculate the T1-relaxation rate, rho
    i is the current i atom number
    n[i] and n[j] are the number of equivalent atoms for i,

    rij is the distance between the i,j atom pair
  This calculation is only valid for Rigid Isotropic motion


  rho is defined for two spins, A and B as,
  rho_A = 2(nA - 1)(W1AA - W2AA) + nB(W0AB + 2W1AB + W2AB)

  for more than two spins, sum up all nB(W0AB + 2W1AB +

*****/

// Declare local variables
int j;
int N = XYZptr->getN();
int nj;
double rij;
double rho = 0;
double leakage = 0;  // Not supported yet, what should it


// Initially, calculate the self dipole contribution for
// equivalent atoms and the leakage rate
rho = 2*(XYZptr->getn(i) - 1) * (W1AA + W2AA) + leakage;

// Look at every i j pair, sum up the rho
for (j=0; j<N; j++) {
    // But not at i=j atom pair
    if ( j != i ) {
        nj = XYZptr->getn(j);
        rij = XYZptr->getRij(i,j);
```

```
        // Check to make sure the atoms are not

        if ( rij < 1 ) rij = 1;

        rho += ( -1 * nj * pow( rij, -6 ) *
            (W0AB + 2*W1AB + W2AB) );
    }

}

// Debugging output
// cout << rho << endl;
return (rho);




/*******
  Calculate the T1-relaxation rate, rho
    i is the current i atom number
    n[i] and n[j] are the number of equivalent atoms for i,

    rij is the distance between the i,j atom pair
  This calculation is only valid for Rigid Anisotropic


  rho is defined for two spins, A and B as,
  rho_A = 2(nA - 1)(W1AA - W2AA ) + nB(W0AB + 2W1AB + W2AB)

  for more than two spins, sum up all nB(W0AB + 2W1AB +

*****/

// Declare local variables
int j;
int nj;
double rij;
float S2;    // order parameter
int N = XYZptr->getN();
double rho = 0;
double leakage = 0;  // Not supported yet, what should it
```

```
// Initially, calculate the self dipole contribution for
// equivalent atoms and the leakage rate
rho = 2*(XYZptr->getn(i) - 1) * (NMR.getW1AA() +


// Look at every jth atom, sum up the rho
for (j=0; j<N; j++) {
    // But not at j=i atom pair
    if ( j != i ) {
        // Recalculate the Transition rates at this beta
        S2 = XYZptr->getS(i) * XYZptr->getS(j);
        NMR.calcTransAniso( XYZptr->getBeta( i, j ), S2 );
        nj = XYZptr->getn(j);
        rij = XYZptr->getRij(i,j);
        if (rij < 1) rij = 1;

        rho += ( -1 * nj * pow(rij, -6) *
            (NMR.getW0AB() + 2*NMR.getW1AB() +

    }

}

// Debugging output
// cout << rho << endl;

return (rho);




/*******
  Calculate the cross-relaxation rate, sigma
    i, j are the current atom numbers
    n[i] and n[j] are the number of equivalent atoms for i,

    rij is the distance between the atoms
  This calculation should be valid for all rigid body
    motion models (such as Rigid Isotropic and Rigid
    Anisotropic)
*******/

double sigma;

// Check to make sure the atoms are not UNPHYSICALLY close
```

```
if ( rij < 1 ) rij = 1;

sigma = ( ni * nj * pow(rij, -6) * (W0AB - W2AB) );

// Debugging output
// cout << "sigma=" << sigma << " W0AB=" << W0AB << "

return sigma;





/*******
  Calculate the distance between i,j, rij
    i, j are the current atom numbers
    n[i] and n[j] are the number of equivalent atoms for i,

    sigma is the cross-relaxation rate between the two

  This calculation should be valid for all rigid body
    motion models (such as Rigid Isotropic and Rigid
    Anisotropic)
*******/

double rij6;
double rij;

// Calculate rij, notice that the rate
// is multiplied by the numbers of equivalent atoms
// for each i j atom (ie: for a methyl, n=3)
rij6 = ( ni * nj * (W0AB - W2AB) ) / sigma;
rij = pow(rij6, 1.0/6.0);

// Print debugging
// cout << rij << " " << rij << endl;
return rij;
```

```
int i, j, error;

// We are going to need a bunch of temporary matrices
// Allocate them dynamically
double **VOLEVAL = NEW2D_D( N, N );
double **VOLEVEC = NEW2D_D( N, N );
double **VOLEVECINV = NEW2D_D( N, N );
double **VOLEVAL_LN = NEW2D_D( N, N );
double **MATTEMP = NEW2D_D( N, N );

// Diagonalize the Volume matrix, giving VolumeEvecs and

if ( lapack_eigen_symm( N, VOL, VOLEVEC, VOLEVAL ) != 0 )


/*  DEBUG PRINT EVALS TO FILE NAMED "temp.vol.evals" */
ofstream evals_out;
evals_out.open("temp.vol.evals");
for (i=0; i<N; i++) {
    evals_out << i << " " << VOLEVAL[i][i] << endl;

}
evals_out.close;
/********************************************************/

// Invert the eigenvector matrix, giving VolumeInvEvecs
if ( lapack_inverse( N, VOLEVEC, VOLEVECINV ) != 0 )


// Calculate the ln( V/V0 ) part, placing the result into

for (i=0; i<N; i++) {
    for (j=0; j<N; j++) {
        if (i == j) {
            // These are the diagonal terms
            if ( VOLEVAL[i][j] > 0 ) {
```

```
                VOLEVAL_LN[i][j] = log (VOLEVAL[i][j] /

            }
            else {
                cout << "vol2rate: ERROR: negative

                VOLEVAL_LN[i][j] = 0;
            }
        }
        else {
            // These are off-diagonal terms
            VOLEVAL_LN[i][j] = 0;
        }
    }
}

// Now we must recast the matrix LnVolEvals back to the

// basis set using the eigenvector and inverse eigenvector

// Rate = VolumeEvecs * LnVolEvals * VolumeInvEvecs

if ( lapack_mat_mul( N, N, N, VOLEVEC, VOLEVAL_LN, MATTEMP

if ( lapack_mat_mul( N, N, N, MATTEMP, VOLEVECINV, RATE )


// Free unused memory
DELETE2D_D( VOLEVAL );
DELETE2D_D( MATTEMP );
DELETE2D_D( VOLEVEC );
DELETE2D_D( VOLEVECINV );
DELETE2D_D( VOLEVAL_LN );

// Divide by the mixing time
for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
        RATE[i][j] = RATE[i][j] / NMR.getTmix();
    }
}

// 0=no error
return ( 0 );
```

```
int i, j, error;

// We are going to need a bunch of temporary matrices
// Allocate them dynamically
double **RATEEVAL = NEW2D_D( N, N );
double **RATEEVEC = NEW2D_D( N, N );
double **RATEEVECINV = NEW2D_D( N, N );
double **RATEEVAL_EXP = NEW2D_D( N, N );
double **MATTEMP = NEW2D_D( N, N );

// Diagonalize the RATE matrix, giving RATEEVEC and

if ( lapack_eigen_symm( N, RATE, RATEEVEC, RATEEVAL ) != 0


/*  DEBUG PRINT EVALS TO FILE NAMED "temp.rate.evals" */
ofstream evals_out;
evals_out.open("temp.rate.evals");
for (i=0; i<N; i++) {
    for (j=0; j<N; j++) {
        if (i == j) {
            evals_out << i << " " << RATEEVAL[i][j] <<

        }
    }
}
evals_out.close;
/*********************************************************/

// Invert the eigenvector matrix, giving RATEEVECINV
if ( lapack_inverse( N, RATEEVEC, RATEEVECINV ) != 0 )


// cout << "NMR.tmix = " << NMR.getTmix() << " NMR.vol0 = "


// Calculate the "exp (RATEEVAL * tmix)" manually
for (i=0; i<N; i++) {
    for (j=0; j<N; j++) {
        if (i == j) {
            RATEEVAL_EXP[i][j] = ( exp (RATEEVAL[i][j] *
```

```
        }
        else {
            RATEEVAL_EXP[i][j] = 0;
        }
    }
}

// Volumes = RATEEVEC * RATEEVAL_EXP * RATEEVECINV
if ( lapack_mat_mul( N, N, N, RATEEVEC, RATEEVAL_EXP,

if ( lapack_mat_mul( N, N, N, MATTEMP, RATEEVECINV, VOL )


// Make sure that the Evecs x InvEvecs equals the unity

// if ( lapack_mat_mul( N, N, N, RATEEVEC, RATEEVECINV, VOL


// Free up memory
DELETE2D_D( RATEEVAL );
DELETE2D_D( RATEEVEC );
DELETE2D_D( RATEEVECINV );
DELETE2D_D( RATEEVAL_EXP );
DELETE2D_D( MATTEMP );

// Error checking, 0=no error
return( 0 );





/*****
  Calculates the rij matrix in a structure object
  from a rate matrix
*****/

int i, j;
double rate;  // Temporary storage of RATE[i][j]
int N = XYZptr->getN();

for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
        rate = RATE[i][j];
        if (i == j) {
            // No need to calculate rij
            XYZptr->setRij(i, j, 0);
        }
        else {
            // rij comes from sigma2rij
            XYZptr->setRij(i, j, NMR.sigma2rij( i, j,

        }
    }
}
return( 0 );





/****
 Calculates a rate matrix from a rij matrix using the Rigid

 motion model
****/

int i, j;
int ni, nj;
int N = XYZptr->getN();
double rij;

for (i=0; i<N; i++) {
    // Do a rho calculation, i=j,
    // The NmrParams OBJECT knows how to do this
    RATE[i][i] = NMR.rij2rho_iso( i, XYZptr );

    for (j=i+1; j<N; j++) {
        // Do a sigma calculation
        // The NmrParams OBJECTS know how to do this
        ni = XYZptr->getn(i);
        nj = XYZptr->getn(j);
        rij = XYZptr->getRij(i,j);
        // cout << "rij " << i << " " << j << " =" << rij

        RATE[i][j] = NMR.rij2sigma( ni, nj, rij );
    }
}
```

```
// return 0=no error
return ( 0 );






/****
 Calculates a rate matrix from a rij matrix using the Rigid

 motion model
****/

int i, j;
int N = XYZptr->getN();
double rij;
float S2;   // order parameter

for (i=0; i<N; i++) {
    // Do a rho calculation, i=j,
    // The NmrParams OBJECT knows how to do this
    RATE[i][i] = NMR.rij2rho_aniso( i, NMR, XYZptr );

    for (j=i+1; j<N; j++) {
        // Recalculate the transition rates for each

        S2 = XYZptr->getS( i ) * XYZptr->getS( j );
        NMR.calcTransAniso( XYZptr->getBeta(i, j), S2 );

        // Do a sigma calculation
        // The NmrParams OBJECTS know how to do this
        RATE[i][j] = NMR.rij2sigma( XYZptr->getn(i),

    }
}

// return 0=no error
return ( 0 );
```

7.3.7 Source code: structure.c and structure.h:

The following C++ source code files define the object “Structure” and allow for storage and retrieval of the Cartesian coordinates of a structure, calculation of distances between atoms, calculation of the center of mass, etc.

```
// default constructor and destructor
Structure( int );    // Input the number of atoms
~Structure();

// read functions (read from STDIN)
void readNxyz();     // does not include order parameters
void readNxyzs();    // includes order parameters
void readFull();
void readPair();

// set functions
void setN( int );                // Number of atoms
void setX( int, float );
void setY( int, float );
void setZ( int, float );
void setn( int, int );           // Equivalent atoms
void setS( int, float );
void setRij( int, int, float );
void setRijFix( int, int, float );
void setBeta( int, int, float );

// get functions
int getN() const;                // Number of atoms
float getX( int ) const;
float getY( int ) const;
float getZ( int ) const;
int getn( int ) const;           // Equivalent atoms
float getS( int ) const;
float getRij( int, int ) const;
float getRijFix( int, int ) const;
float getBeta( int, int ) const;
float getCoM_x() const;          // Center of mass, x
```

```
float getCoM_y() const;          // Center of mass, y
float getCoM_z() const;          // Center of mass, z

// print functions (print to STDOUT)
void printFull() const;
void printPair() const;

// write functions (write to a file)
void fileFull( char [] ) const;
void fileNxyzs( char [] ) const;

// calculation functions
void calcRij();    // Builds the rij matrix from current

void calcCoM();    // Calculates current center of mass
void calcBeta( float, float, float );  // Builds beta



float *x, *y, *z;                // Coordinates
float **rij;                     // rij matrix
float *s;                        // Order parameter matrix
float **rij_fix;                 // fixed rij matrix
float **beta;                    // beta matrix
int *n;
int N;                            // Number of atoms
float CoM_x, CoM_y, CoM_z;        // Centers of Mass
int *res_num;                     // Residue number
char **segid;                     // Segment ID (XPLOR)
char **res_type;                  // Residue type
char **atom_type;                 // Atom type
```

```
Structure Class



int i;
N = N_in;
CoM_x = 0;
CoM_y = 0;
CoM_z = 0;

// Dynamically allocate memory for the arrays
x = NEW1D_F( N );
y = NEW1D_F( N );
z = NEW1D_F( N );
n = NEW1D_I( N );
s = NEW1D_F( N );
res_num = NEW1D_I( N );
rij = NEW2D_F( N, N );
rij_fix = NEW2D_F( N, N );
beta = NEW2D_F( N, N );
segid = NEW2D_C( N, 5 );
res_type = NEW2D_C( N, 4 );
atom_type = NEW2D_C( N, 5 );

// Do I have to do this?
```

```
for (i=0; i<N; i++) {
    x[i] = 0;
    y[i] = 0;
    z[i] = 0;
    res_num[i] = 0;
    n[i] = 1;
    s[i] = 1;  // default order parameter is 1 (rigid)
}



// Dynamically deallocate memory for the arrays
DELETE1D_F( x );
DELETE1D_F( y );
DELETE1D_F( z );
DELETE1D_I( n );
DELETE1D_F( s );
DELETE1D_I( res_num );
DELETE2D_F( rij );
DELETE2D_F( rij_fix );
DELETE2D_F( beta );
DELETE2D_C( segid );
DELETE2D_C( res_type );
DELETE2D_C( atom_type );
```

```
int i, j;
// char *firstline = NEW1D_C( 80 );
char firstline[80];

// The first line is a header
cin.getline(firstline, sizeof(firstline));
for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
        cin >> n[i] >> n[j] >> rij[i][j] >> beta[i][j];

        // We set aside the memory, may as well use i t!
        rij[j][i] = rij[i][j];
        beta[j][i] = beta[i][j];
    }
}

// DELETE1D_C( firstline );




int i, j;
char firstline[80];
// char *firstline = NEW1D_C( 80 );

// The first line is a header
cin.getline(firstline, sizeof(firstline));
for (i=0; i<N; i++) {
    cin >> segid[i] >> res_num[i] >> res_type[i] >>

}
// DELETE1D_C( firstline );




int i, j;
// The first line is a header
ofstream coord_out;
coord_out.open( FILE );
```

```
int i, j;
char firstline[80];

// The first line is a header
cin.getline(firstline, sizeof(firstline));
for (i=0; i<N; i++) {
    cin >> n[i] >> x[i] >> y[i] >> z[i];
}




int i, j;
char firstline[80];

// The first line is a header
cin.getline(firstline, sizeof(firstline));
for (i=0; i<N; i++) {
    cin >> n[i] >> x[i] >> y[i] >> z[i] >> s[i];
}
```

```
coord_out << "Full structure file output" << endl;

for (i=0; i<N; i++) {
    coord_out << segid[i] << " " << res_num[i] << " " <<

    coord_out << " " << atom_type[i] << " " << n[i] << " "

    coord_out << " " << y[i] << " " << z[i] << " " << s[i]

}

coord_out.close;





int i, j;
// The first line is a header
ofstream coord_out;
coord_out.open( FILE );

coord_out << "Full structure file output" << endl;

for (i=0; i<N; i++) {
    coord_out << n[i] << " " << x[i] << " " << y[i];
    coord_out << " " << z[i] << " " << s[i] << endl;
}

coord_out.close;






/*****
  Calculates a rij matrix from XYZ coordinates
*****/

int i, j;
double Ax, Ay, Az;

for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
```

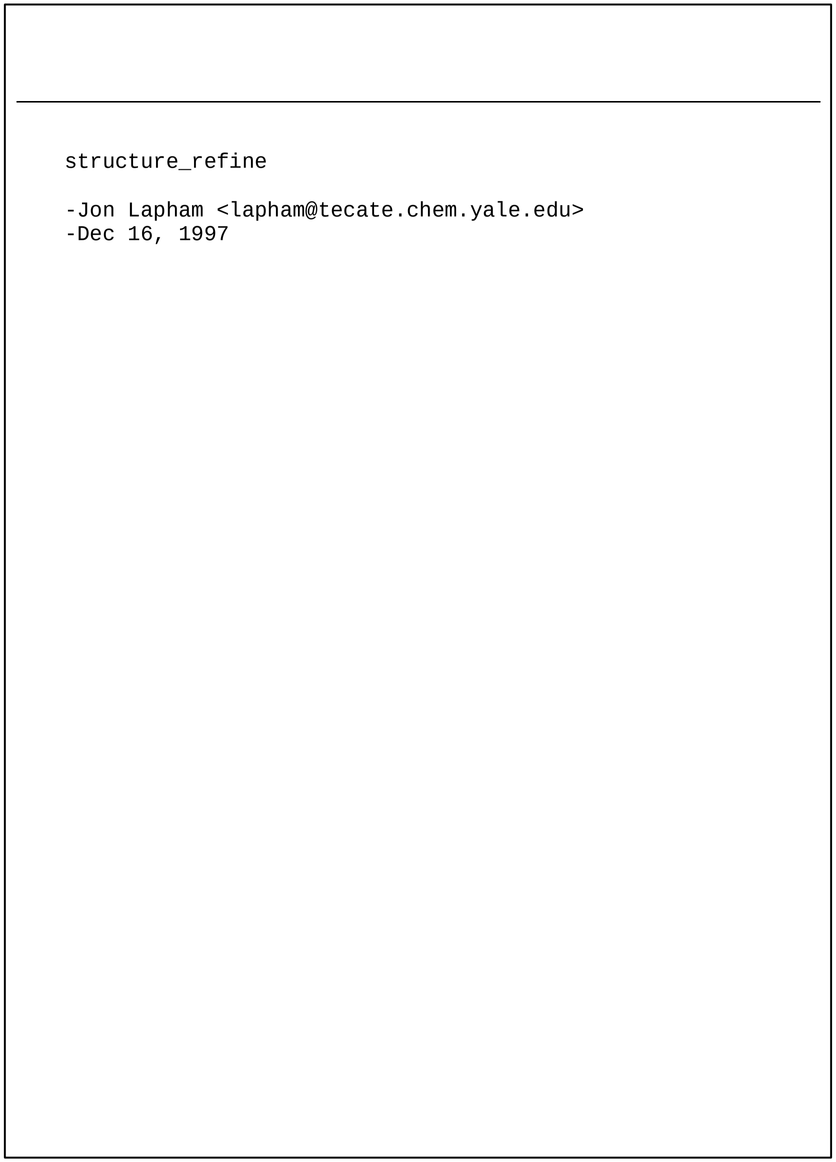
```
        Ax = pow( ( x[i]-x[j] ), 2 );
        Ay = pow( ( y[i]-y[j] ), 2 );
        Az = pow( ( z[i]-z[j] ), 2 );
        rij[i][j] = rij[j][i] = sqrt( Ax + Ay + Az );
    }
}






/*****
  Calculates a beta matrix from XYZ coordinates
*****/

int i, j;
double angle, cos_angle, angle_rad;
double Ax, Ay, Az, magA, magB;

magB = sqrt( pow(Bx, 2) + pow(By, 2) + pow(Bz, 2) );

for (i=0; i<N; i++) {
    for (j=0; j<N; j++) {
        Ax = pow( ( x[i]-x[j] ), 2 );
        Ay = pow( ( y[i]-y[j] ), 2 );
        Az = pow( ( z[i]-z[j] ), 2 );
        magA = sqrt( pow(Ax, 2) + pow(Ay, 2) + pow(Az,  2)


        if (magA*magB != 0) {
            cos_angle = (Ax*Bx + Ay*By + Az*Bz) /

            angle_rad = acos( cos_angle );
            angle = (180/PI) * angle_rad;
            // Check to see if we are over 90 degrees...
            if (angle > 90) { angle = 180 - angle; }
        }
        else {
            angle = 0;
        }
        beta[i][j] = angle;

    }
}
```

```
/*****
  Calculates center of mass from XYZ coordinates

  -need to make this rigorous, use masses...
*****/

int i;
double x_sum=0;
double y_sum=0;
double z_sum=0;

for (i=0; i<N; i++) {
    x_sum += x[i];
    y_sum += y[i];
    z_sum += z[i];
}

CoM_x = x_sum / N;
CoM_y = y_sum / N;
CoM_z = z_sum / N;




int i, j;

// The first line is a header
for (i=0; i<N; i++) {
    cout << "Segid=" << segid[i] << " res_num=" <<

    cout << res_type[i] << " atom_type=" << atom_type[i] <<

    cout << " y=" << y[i] << " z=" << z[i] << endl;
}




/****
```

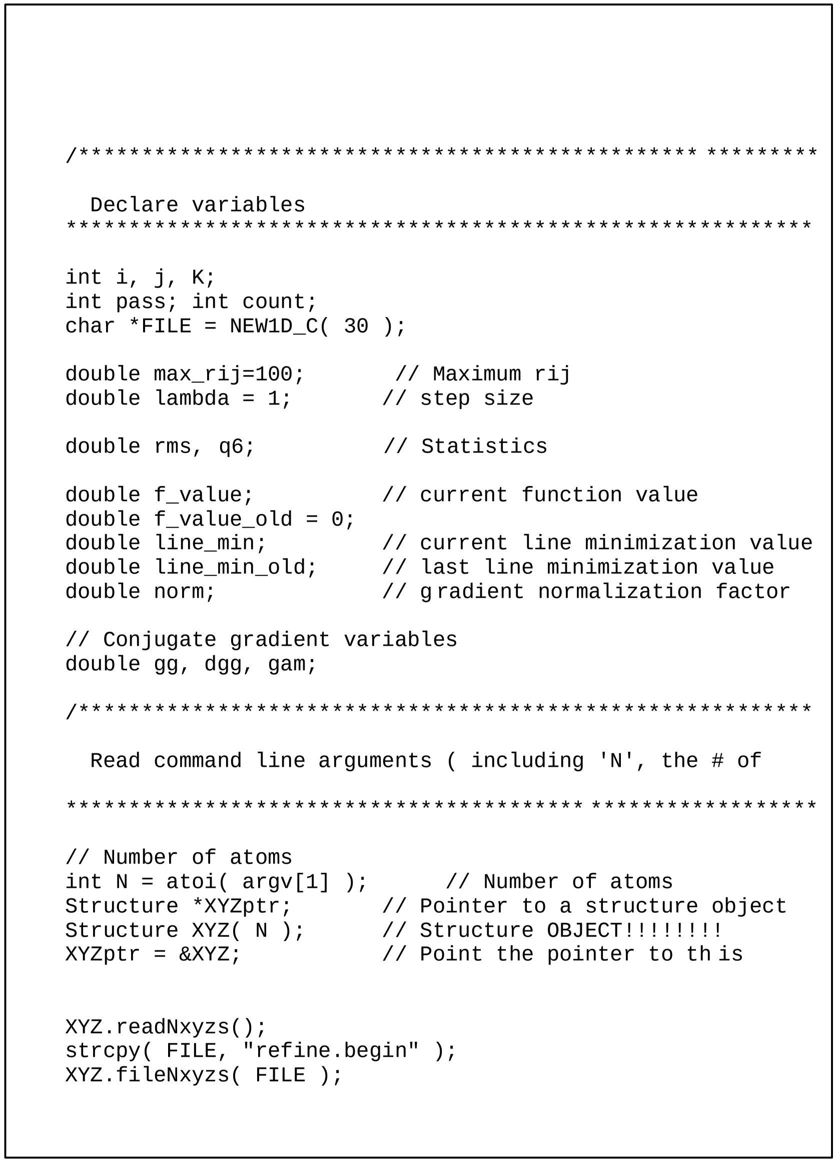
```
   Prints to STDOUT a YARM "Pair" file
****/

int i, j;

// The first line is a header
cout << "Yarm pair file\n";
for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
        cout << n[i] << " ";
        cout << n[j] << " ";
        cout << rij[i][j] << " " << beta[i][j] << endl;
    }
}
```

7.3.8 Source code: structure_refine.c

The following C++ source code is used in the calculations of model refinement (called by the Structure_Refine YARM subroutine).



```
structure_refine

-Jon Lapham <lapham@tecate.chem.yale.edu>
-Dec 16, 1997
```



```
/************************************************** *********

  Declare variables
***********************************************************

int i, j, K;
int pass; int count;
char *FILE = NEW1D_C( 30 );

double max_rij=100;       // Maximum rij
double lambda = 1;       // step size

double rms, q6;          // Statistics

double f_value;          // current function value
double f_value_old = 0;
double line_min;         // current line minimization value
double line_min_old;     // last line minimization value
double norm;             // gradient normalization factor

// Conjugate gradient variables
double gg, dgg, gam;

/**********************************************************

  Read command line arguments ( including 'N', the # of

***************************************** *****************

// Number of atoms
int N = atoi( argv[1] );      // Number of atoms
Structure *XYZptr;       // Pointer to a structure object
Structure XYZ( N );      // Structure OBJECT!!!!!!!!
XYZptr = &XYZ;           // Point the pointer to this


XYZ.readNxyzs();
strcpy( FILE, "refine.begin" );
XYZ.fileNxyzs( FILE );
```

```
NmrParams NMR;            // Create a NMR Parameters OBJECT

// Read in command line arguments
NMR.setTl   ( atof( argv[2] ) );    // Long axis

NMR.setTs   ( atof( argv[3] ) );    // Short axis

NMR.setSfreq( atof( argv[4] ) );    // Spectrometer

NMR.setVol0 ( atof( argv[5] ) );    // tmix=0 volume
NMR.setTmix ( atof( argv[6] ) );    // mixing time
double Ax   = atof( argv[7] );      // x component of

double Ay   = atof( argv[8] );      // y component of

double Az   = atof( argv[9] );      // z component of

int num_pass= atoi( argv[10] );     // number of iterations

/**********************************************************

  Declare other variables that depend on 'N'
***********************************************************


double *gx = NEW1D_D( N );        // gx = X gradient
double *gy = NEW1D_D( N );        // gy = Y gradient
double *gz = NEW1D_D( N );        // gz = Z gradient
double *hx = NEW1D_D( N );        // hx = X descent
double *hy = NEW1D_D( N );        // hy = Y descent
double *hz = NEW1D_D( N );        // hz = Z descent
double *xix = NEW1D_D( N );       // next gradient
double *xiy = NEW1D_D( N );       // next gradient
double *xiz = NEW1D_D( N );       // next gradient

double *R = NEW1D_D( N*N );        // R = residuals
double **ExpVol = NEW2D_D( N, N );  // experimental NOE

double **ExpRijFix = NEW2D_D( N, N );  // experimental

double **ExpRij = NEW2D_D( N, N );  // experimental rijs
double **SimVol = NEW2D_D( N, N );  // simulated NOE

double **SimRate = NEW2D_D( N, N ); // simulated rate
```

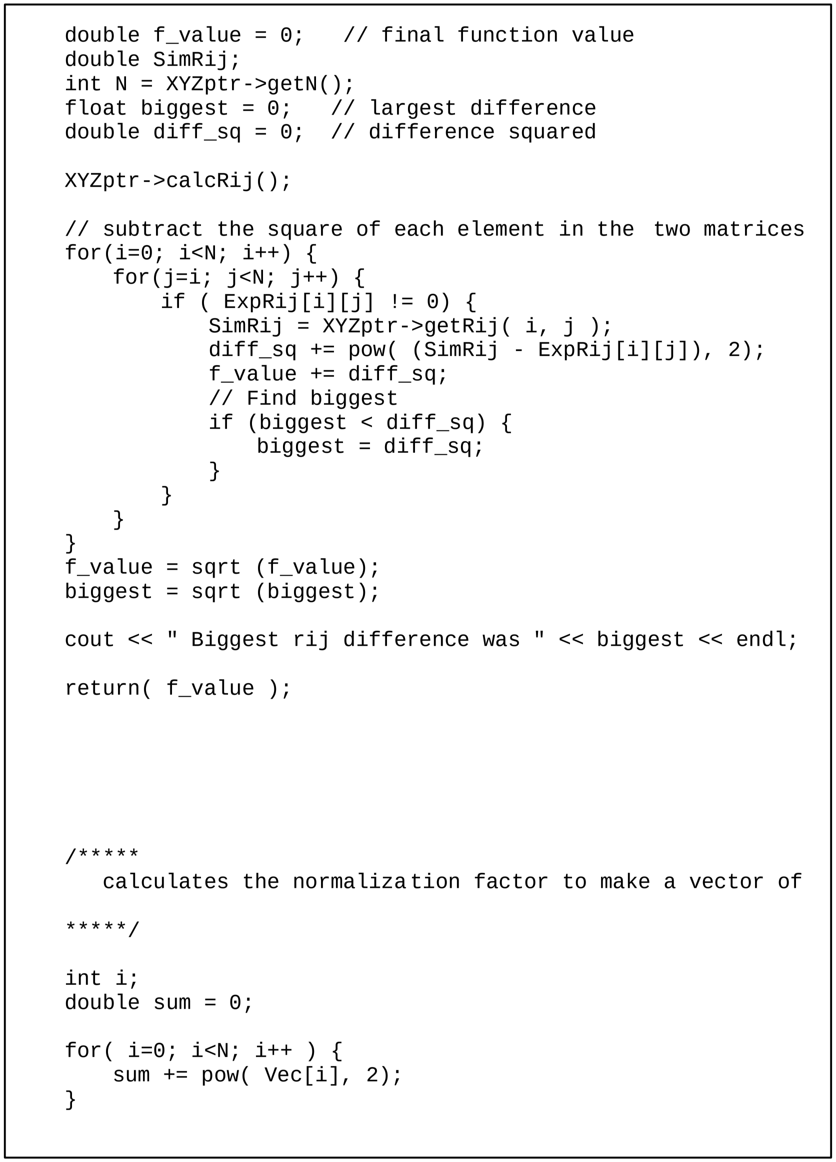
```
/**********************************************************

   Read in the experimental volumes from STDIN
***********************************************************


for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
        cin >> ExpVol[i][j] >> ExpRijFix[i][j];
        ExpVol[j][i] = ExpVol[i][j];
        ExpRijFix[j][i] = ExpRijFix[i][j];
    }
}

// XYZ.printFull();

/**********************************************************

  BEGIN COORDINATE REFINEMENT PROGRAM
******************************************************* ****


XYZ.calcBeta( Ax, Ay, Az );

NMR.printNmrParams();

// Calculate the experimental rijs
if ( func_real( NMR, XYZptr, ExpVol, ExpRij, ExpRijFix,

    cout << "ERROR in func_real function!\n";
    return(0);
}

// strcpy(FILE, "simvol.out");
// print_mat( N, SimVol, FILE );

// Calculate the root-square difference of the ExpVol and

f_value = mat_diff( XYZptr, ExpRij );
cout << " FUNCTION VALUE (ExpRij - SimRij) = " << f_value


// Calculate the gradient
if ( calc_gradient( XYZptr, max_rij, ExpRij, xix, xiy, xiz

    cout << "error in calc_gradient function\n";
```

```
        return(0);
    }

    // initialize arrays
    for (i=0; i<N; i++) {
        gx[i] = -xix[i];
        gy[i] = -xiy[i];
        gz[i] = -xiz[i];

        xix[i]=hx[i]=gx[i];
        xiy[i]=hy[i]=gy[i];
        xiz[i]=hz[i]=gz[i];
    }

    /***********************************************************

      Iteratively determine the merged volume matrix
    ************************************************************

    for ( pass=0; pass<num_pass; pass++ ) {
        // Header for the beginning of an iteration
        cout <<

        cout << " Iterative pass number " << pass << " of " <<



        // Positions of the first two atoms
        cout << " atom 0 x=" << XYZ.getX(0) << " y=" <<

        cout << " atom 1 x=" << XYZ.getX(1) << " y=" <<


        // Calculate the step size, lambda
        if ( f_value < f_value_old*0.999999999 9 ) {
            cout << " lambda raised from " << lambda;
            lambda *= 1.2;
            cout << " to " << lambda << endl;
        } else if ( f_value_old != 0 ) {
            cout << " lambda lowered from " << lambda;
            lambda *= 0.5;
            cout << " to " << lambda << endl;
        } else {
            cout << " first time through, lambda remains " <<
```

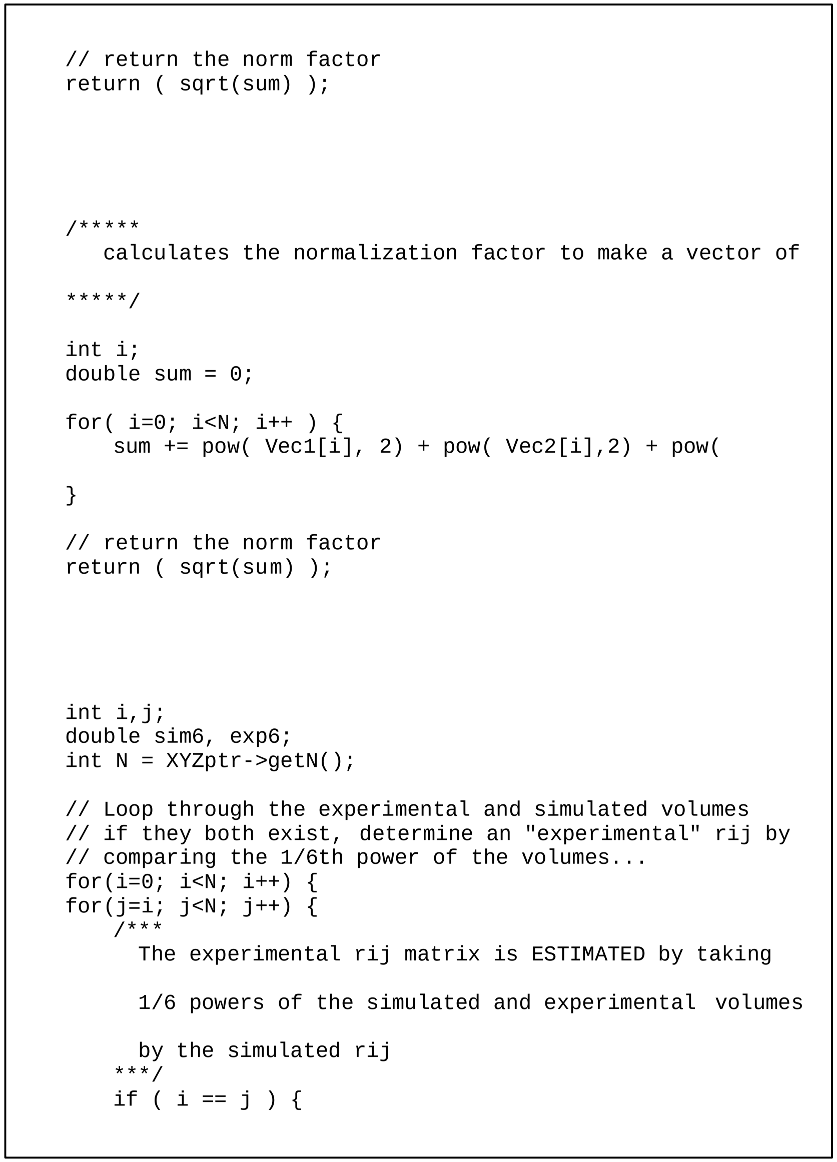
```
        }

        // Check to see if things aren't moving anymore
        if ( lambda < 1e-4 ) {
            cout << " EARLY TERMINATION, lambda=" << lambda <<

            strcpy( FILE, "refine.done" );
            XYZ.fileNxyzs( FILE );
            return (0);
        }

        // Maximum step size allowed is 5 angstroms
        if ( lambda > 5 ) lambda = 5;
        f_value_old = f_value;

        // We have to do this so we don't change the f_value

        line_min = f_value;
        line_min_old = 2*line_min;
        count = 0;

        // Move the atoms along the gradient until they settle

        // or make 5 steps at the current lambda2 value
        while( ( line_min < line_min_old ) && ( ++count < 6 ) )

            line_min_old = line_min;

            // move the atoms a lambda * (xix, xiy, xiz)

            move_atoms( XYZptr, lambda, xix, xiy, xiz );

            // Calculate the root-square difference of the

            line_min = mat_diff( XYZptr, ExpRij );
            cout << " " << line_min;
        }
        // Back up one step if necessary
        if ( line_min > line_min_old ) {
            // move the atoms a lambda * (xix, xiy, xiz)

            lambda *= -1.0;
            move_atoms( XYZptr, lambda, xix, xiy, xiz );
            lambda *= -1.0;
            // Calculate the root-square difference of the

            line_min = mat_diff( XYZptr, ExpRij );
            cout << " backup to " << line_min;
        }
        cout << endl;

        XYZ.calcBeta( Ax, Ay, Az );

        // Recalculate the experimental rijs using these new

        if ( func_real( NMR, XYZptr, ExpVol,  ExpRij, ExpRijFix,

            cout << "ERROR in func_real function!\n";
            return(0);
        }

        // Recalculate the root-square difference of the ExpVol

        f_value = mat_diff( XYZptr, ExpRij );
        cout << " FUNCTION VALUE (ExpRij - SimRij) = " <<


        // Recalculate the gradient at this new position
        if ( calc_gradient( XYZptr, max_rij, ExpRij, xix, xiy,

            cout << "error in calc_gradient function\n";
            return(0);
        }

        /* BEGIN CONJUGATE GRADIENT CODE */
        dgg = gg = 0.0;

        for (i=0; i<N; i++) {
            gg += gx[i]*gx[i];
            gg += gy[i]*gy[i];
            gg += gz[i]*gz[i];

            // dgg += xix[i] * xix[i];   // This is Fletcher-

            // dgg += xiy[i] * xiy[i];   // This is Fletcher-

            // dgg += xiz[i] * xiz[i];   // This is Fletcher-
```

```
            dgg += (xix[i]+gx[i]) * xix[i];   // This is Polak-

            dgg += (xiy[i]+gy[i]) * xiy[i];   // This is Polak-

            dgg += (xiz[i]+gz[i]) * xiz[i];   // This is Polak-

        }

        // Gamma definition. never let it get above 1!
        if ( gg == 0 ) { gam = 0; }
        else { gam=dgg/gg; }

        cout << " gamma=" << gam << endl;

        // Use this line to bypass conjugate gradient, called

        // gam=0;

        for (i=0; i<N; i++) {
            gx[i] = -xix[i];
            gy[i] = -xiy[i];
            gz[i] = -xiz[i];

            xix[i]=hx[i]=gx[i]+gam*hx[i];
            xiy[i]=hy[i]=gy[i]+gam*hy[i];
            xiz[i]=hz[i]=gz[i]+gam*hz[i];
        }

        /* END CONJUGATE GRADIENT CODE */

    }

    strcpy( FILE, "refine.done" );
    XYZ.fileNxyzs( FILE );

    // free up memory, do I have to do this?
    DELETE1D_D( gx );
    DELETE1D_D( gy );
    DELETE1D_D( gz );
    DELETE1D_D( hx );
    DELETE1D_D( hy );
    DELETE1D_D( hz );
    DELETE1D_D( xix );
    DELETE1D_D( xiy );
    DELETE1D_D( xiz );
    DELETE1D_D( R );
    DELETE2D_D( ExpVol );
    DELETE2D_D( ExpRij );
    DELETE2D_D( ExpRijFix );
    DELETE2D_D( SimVol );
    DELETE2D_D( SimRate );

    return(0);






    int i;
    int N = XYZptr->getN();
    double oldx, oldy, oldz;
    double x, y, z;

    for (i=0; i<N; i++) {
        oldx = XYZptr->getX(i);
        oldy = XYZptr->getY(i);
        oldz = XYZptr->getZ(i);
        x = oldx + (gx[i] * size);
        y = oldy + (gy[i] * size);
        z = oldz + (gz[i] * size);
        XYZptr->setX(i, x);
        XYZptr->setY(i, y);
        XYZptr->setZ(i, z);
    }

    return(0);







    int i,j;
    int N = XYZptr->getN();
    double q6, rms;
    double SimRij;

    // Simulate the NOE spectrum from the current coordinates
```

```
    cout << " Converting current xyz coordinates into rij

    XYZptr->calcRij();

    cout << " Converting current rij matrix into Rate matrix"

    if ( rij2rate_aniso( NMR, XYZptr, SimRate ) != 0 )


    cout << " Converting Rate matrix into Volume matrix" <<

    if ( rate2vol( N, NMR, SimRate, SimVol ) != 0 ) return(1);
    cout << " Finished building Volume matrix" << endl;

    // Calculate the "experimental rijs" by comparing the
    // experimental volumes to the simulated
    if ( calc_exprij( XYZptr, ExpVol, SimVol, ExpRijFix, ExpRij

    cout << " ExpVol[0][0]=" << ExpVol[0][0] << "

    cout << " SimRate[0][0]=" << SimRate[0][0] << endl;
    cout << " ExpRij[0][0]=" << ExpRij[0][0] << "

    cout << " ExpVol[0][1]=" << ExpVol[0][1] << "

    cout << " SimRate[0][1]=" << SimRate[0][1] << endl;
    cout << " ExpRij[0][1]=" << ExpRij[0][1] << "

    cout << " ExpVol[1][1]=" << ExpVol[1][1] << "

    cout << " SimRate[1][1]=" << SimRate[1][1] << endl;
    cout << " ExpRij[1][1]=" << ExpRij[1][1] << "


    rms = calc_mat_rms( N, SimVol, ExpVol );
    q6 = calc_mat_q6( N, SimVol, ExpVol );

    cout << " RMS=" << rms << " q^(1/6)=" << q6 << endl;

    return(0);




    int i,j;
```



```
double f_value = 0;   // final function value
double SimRij;
int N = XYZptr->getN();
float biggest = 0;   // largest difference
double diff_sq = 0;  // difference squared

XYZptr->calcRij();

// subtract the square of each element in the  two matrices
for(i=0; i<N; i++) {
    for(j=i; j<N; j++) {
        if ( ExpRij[i][j] != 0) {
            SimRij = XYZptr->getRij( i, j );
            diff_sq += pow( (SimRij - ExpRij[i][j]), 2);
            f_value += diff_sq;
            // Find biggest
            if (biggest < diff_sq) {
                biggest = diff_sq;
            }
        }
    }
}
f_value = sqrt (f_value);
biggest = sqrt (biggest);

cout << " Biggest rij difference was " << biggest << endl;

return( f_value );






/*****
   calculates the normalization factor to make a vector of

*****/

int i;
double sum = 0;

for( i=0; i<N; i++ ) {
    sum += pow( Vec[i], 2);
}
```



```

// return the norm factor
return ( sqrt(sum) );





/*****
   calculates the normalization factor to make a vector of

*****/

int i;
double sum = 0;

for( i=0; i<N; i++ ) {
    sum += pow( Vec1[i], 2) + pow( Vec2[i],2) + pow(

}

// return the norm factor
return ( sqrt(sum) );





int i,j;
double sim6, exp6;
int N = XYZptr->getN();

// Loop through the experimental and simulated volumes
// if they both exist, determine an "experimental" rij by
// comparing the 1/6th power of the volumes...
for(i=0; i<N; i++) {
for(j=i; j<N; j++) {
    /***
      The experimental rij matrix is ESTIMATED by taking

      1/6 powers of the simulated and experimental  volumes

      by the simulated rij
    ***/
    if ( i == j ) {
            ExpRij[i][j] = 0;
        }
        else if ( ExpRijFix[i][j] != 0 ) {
            // If this is a fixed distance, set it
            ExpRij[i][j] = ExpRij[j][i] = ExpRijFix[i][j];
        }
        else if ( ExpVol[i][j] > 0 ) {
            sim6 = pow( SimVol[i][j], 1.0/6.0 );
            exp6 = pow( ExpVol[i][j], 1.0/6.0 );
            ExpRij[i][j] = ExpRij[j][i] = ( sim6 / exp6 ) *

        }
        else {
            // Use the SimRij for ExpRij if a ExpVol doesn't

            // ExpRij[i][j] = ExpRij[j][i] = XYZptr->getRij(i,

            ExpRij[i][j] = ExpRij[j][i] = 0;
        }
    }
    }

    // 0=no error
    return( 0 );





    /******
      calculate gradient

      g = nabla V(r)
      nabla = d/dx + d/dy + d/dz
      V(r) = delta_r = sRij - eRij
      dV/dx = xi-xj / sqrt( pow((xi-xj),2) + pow((yi-yj),2) +

    ******/

    // Define variables
    int i,j;
    double dV_dx, dV_dy, dV_dz;   // Temp storage of  ij

    double norm;
```

```
    double eRij, sRij;            // Temporary storage of exp

    double delta_r;               // delta_r = sRij-eRij
    int N = XYZptr->getN();       // Number of atoms

    // Loop through each i atom
    for(i=0; i<N; i++) {

        // Initialize the gradient to zero for each i atom
        dx[i] = 0; dy[i] = 0; dz[i] = 0;

        // Sum ith gradient WRT all j atoms
        for(j=0; j<N; j++) {

            // Don't use i=j atom pairs or eRij bigger than

            eRij = ExpRij[i][j];
            if ( (i != j ) && (eRij < max_rij) ) {

            sRij = XYZptr->getRij(i, j);

            // This is important!  If we DON'T have

            // then there should not be any gradient!
            if (eRij != 0) delta_r = sRij - eRij;
            else delta_r = 0;

            // cout << " calc_gradient: delta_r=" << delta_r <<


            // The "potential energy" is delta_r, which goes to

            // we don't have to calculate this...
            // V = delta_r;
            dV_dx = ( XYZptr->getX(i) - XYZptr->getX(j) ) /

            dV_dy = ( XYZptr->getY(i) - XYZptr->getY(j) ) /

            dV_dz = ( XYZptr->getZ(i) - XYZptr->getZ(j) ) /


            // Sum up the gradient on each atom i

            dx[i] += dV_dx * delta_r;
            dy[i] += dV_dy * delta_r;
        dz[i] += dV_dz * delta_r;
        }
    }
}

// Normalize the gradient vector
norm = calc_3vec_norm( N, dx, dy, dz );
// cout << " calc_gradient: norm_factor=" << norm << endl;

// divide each gradient vector by the norm factor
for (i=0; i<N; i++) {
    if (norm != 0) {
        dx[i] *= 1 / norm;
        dy[i] *= 1 / norm;
        dz[i] *= 1 / norm;
    }
}

// debugging printout
cout << " N=" << N << endl;
cout << " calc_gradient: atom 0 dx[0]=" << dx[0] << "

cout << " calc_gradient: atom 1 dx[1]=" << dx[1] << "

cout << " calc_gradient: atom 2 dx[2]=" << dx[2] << "


// 0 = no error
return( 0 );





int i, j;
double mat;   // Temporary storage variable for Rate**

cout << "Writing MATRIX to file named " << FILE << endl;
ofstream mat_out;
mat_out.open(FILE);

// Write out the matrix to file
// Only send out the lower triangle of data
for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
```

```
        mat = MAT[i][j];

        mat_out << "i=" << i << " j=" << j << " " << mat <<

    }
}
mat_out.close();




/****
  Calculate the rms between common elements between two

****/

int i, j;
double RMS = 0;
double RMS_top = 0;
double RMS_bottom = 0;
double RMS_div = 0;
double mat1, mat2;
int count=0;

// Do RMS statistics on the volume sets
for (i=0; i<N; i++) {
    for (j=0; j<N; j++) {

        // Use temp vars to make the equations more

        mat1 = MAT1[i][j];
        mat2 = MAT2[i][j];

        if ( ( mat1 != 0.0 ) && ( mat2 != 0.0 ) ) {
            RMS_top    += pow( ( mat1 - mat2 ), 2 );
            RMS_bottom += pow( mat1, 2) + pow( mat2, 2 );
            ++count;
        }
    }
}
RMS_div = RMS_top / RMS_bottom;

if ( RMS_div < 0 ) {
    RMS = 0;
```

```
}
else {
    RMS = sqrt( RMS_div );
}

// cout << "RMS = " << RMS << endl;
// cout << " Used " << count << " elements in RMS


return (RMS);




/****
  Calculate the Q6-factor between common elements between

****/

int i, j;
double Q6 = 0;
double Q6_top = 0;
double Q6_bottom = 0;
double mat1, mat2;

// Do RMS statistics on the volume sets
for (i=0; i<N; i++) {
    for (j=0; j<N; j++) {

        // Use temp vars to make the equations more

        mat1 = MAT1[i][j];
        mat2 = MAT2[i][j];

        if ( ( mat1 != 0.0 ) && ( mat2 != 0.0 ) ) {
            Q6_top    += fabs( pow( mat1, 1.0/6.0 ) - pow(

            Q6_bottom += pow( mat1, 1.0/6.0 ) + pow( mat2,

        }
    }
}

if ( Q6_bottom == 0 ) {
    Q6 = 0;
```

```
} else {
    Q6 = Q6_top / (0.5 * Q6_bottom);
}

// cout << "Q6 = " << Q6 << endl;
return (Q6);




/*****
  Print out the sum total of all shared elements between

*****/

int i, j, k;
double mat1, mat2;  // temporary storage or each Matrix

double Sum_Mat1 = 0;
double Sum_Mat2 = 0;

for (i=0; i<N; i++) {
    for (k=0; k<i; k++) {
        mat1 = MAT1[k][i];
        mat2 = MAT2[k][i];

        // Sum up each non-zero shared element
        if ( (mat1 !=0) && (mat2 != 0) ) {
            Sum_Mat1 += mat1;
            Sum_Mat2 += mat2;
        }
    }
    for (j=i; j<N; j++) {
        mat1 = MAT1[i][j];
        mat2 = MAT2[i][j];

        // Sum up each non-zero shared element
        if ( (mat1 !=0) && (mat2 != 0) ) {
            Sum_Mat1 += mat1;
            Sum_Mat2 += mat2;
        }

    }
}
```

```
// Print a little report
cout << " print_mat_sum: First matrix total  = " << Sum_Mat1

cout << " print_mat_sum: Second matrix total = " <<




/*****
  normalize two matrices
*****/

int i, j;
double mat1, mat2;  // temporary storage or e ach Matrix

double Sum_Mat1=0;
double Sum_Mat2=0;

for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
        mat1 = MAT1[i][j];
        mat2 = MAT2[i][j];

        // Sum up each non-zero shared element
        if ( (mat1 !=0) && (mat2 != 0) ) {
            Sum_Mat1 += mat1;
            Sum_Mat2 += mat2;
        }

    }
}

double norm_factor = Sum_Mat2/Sum_Mat1;

for (i=0; i<N; i++) {
    for (j=i; j<N; j++) {
        MAT1[i][j] = MAT1[i][j] * norm_factor;
    }
}
```

7.4 References

Bloembergen N, Purcell EM, Pound RV. 1948. Relaxation Effects in Nuclear Magnetic Resonance Absorption. *Physical Review 73*:679-712.

Boelens R, Koning TMG, Kaptein R. 1988. *J. Mol. Struct. 173*:299.

Borgias BA, James TL. 1990. MARDIGRAS-A procedure for matrix analysis of relaxation for discerning geometry of an agueous structure. *J Mag Res 87*:475-487.

Clore GM, Gronenborn AM. 1985. *J. Mag. Res. 61*:158.

Clore GM, Gronenborn AM. 1989. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Crit. Rev. Biochem. Mol. Biol. 24*:479-564.

Lipari G, Szabo A. 1980. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. *Journal of the American Chemical Society 104*:4546-4559.

Nerdal W, Hare DR, Reid BR. 1989. Solution structure of the *Eco*RI DNA sequence: refinement of NMR-derived distance geometry structures by NOESY spectrum back-calculations. *Biochemistry 28*:10008-10021.

Patel DJ, Shapiro L, Hare D. 1987. DNA and RNA: NMR studies of conformations and dynamics in solution. *Q. Rev. Biophys. 20*:35-112.

Post CW, Meadows RP, Gorenstein DG. 1990. *JACS 112*:6796.

Reid BR. 1987. Sequence-specific assignments and their use in NMR studies of DNA structure. *Q. Rev. Biophys. 20*:1-34.

Schmitz U, James TL. 1995. How to generate accurate solution structures of double-helical nucleic acid fragments using nuclear magnetic resonance and restrained dynamics. *Methods in Enzymology 261*:3-44.

Tropp J. 1980. Dipolar relaxation and nuclear Overhauser effects in non-rigid molecules: the effect of fluctuating internuclear distances. *J Chem Phys 72*:6035-6043.

Wagner G, Wüthrich K. 1982. Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J. Mol. Bio. 155*:347-366.

CHAPTER 8 “COMPUTER PROGRAMS”

This chapter is intended to be a resource guide for anyone interested in performing any of the calculations presented in chapters 1 through 6 of this thesis. These programs were written to give the biomolecular NMR spectroscopist the tools necessary to work with their NMR data and structural models.

The X-PLOR utilities are a series of scripts written to facilitate the use of the restrained molecular dynamics package, X-PLOR, with nucleic acids. These scripts use a common input file that allows one to easily generate and manipulate torsion angle, base pairing and planarity restraint files.

The hydrodynamic calculation scripts were developed for calculating both the rotational and translational diffusion rates of a hydrodynamical particle using spherical, elliptical and cylindrical models. This is useful for predicting the correlation time of a given biomolecule.

The inertia tensor program is used to predict the possible rotational anisotropy of a biomolecule by calculating the moment of inertia about the rotation reference frame. This is useful in deciding if a biomolecule is best described using one, two or three rotational correlation times.

Almost all programs in this chapter were written using the scripting language "PERL", and a quick discussion of why this language was used is in order. First, as an interpreted *scripting* language, PERL enjoys the enormous advantage of being almost completely portable between many computer types and operating systems. All the machine specific coding is done in the PERL interpreter, which is available for most major operating systems. The second advantage PERL has over other languages is that it excels at text manipulations. PERL scripts can easily be a fraction of the size of their FORTRAN counterparts due to the many text manipulation tools built into the language.

The author wrote most of the computer programs presented in this chapter. The few exceptions (dm, noe_in and planar_make) are duly noted and are included only for the sake of completeness. Most of the programs were written specifically for the purposes needed by the author and may not be a general solution to for all situations. People are encouraged to use, and improve on, these programs.

8.1 X-PLOR utilities

X-PLOR is a software package that performs restrained molecular dynamics on biomolecules, and is the mechanism by which the distance and angular restraints determined by NMR are incorporated into a structural model. There are a number of restraint files that X-PLOR needs to successfully perform these calculations on nucleic acids.

Torsion angle restraints are used to adjust any of the seven torsion angles that define a nucleotide (cdih.dat). Distance restraints can be distilled into two types of files, the first is distances between heavy atoms to force hydrogen bonding between base pairs (hbond.dat) and the second is distances between protons (noe.dat). Finally, planarity restraints are used to force two nucleic acid bases to remain planar, typically between two base-paired nucleotides (planar.dat).

Necessarily, these input files are large, tedious and complex. A 20 nucleotide DNA, for instance would have 7x20 = 140 entries in the cdih.dat file that might look like that shown below (this entry defines the alpha torsion angle to have an angle of -46.8 degrees +/- 20). That would be nearly 5x140 = 700 lines of text.

```
assign (segid a and resid 1 and name O3')
      (segid a and resid 2 and name P  )
       (segid a and resid 2 and name O5')
      (segid a and resid 2 and name C5')
       1 -46.8  20  2
```

To encourage experimentation with the X-PLOR restraints, a number of programs have been written that perform the arduous task of building these input files. cdih_make build the torsion angle file, noe_hbond_make builds the hydrogen bonding file, dm/noe_in builds the noe.dat file and planar_make builds the planarity restraint file.

Additionally, cdih_measure has been written to examine the torsion angles of nucleic acid PDB files, useful for determining these angles after X-PLOR has created a structural model.

8.1.1 "seq" file format

Many of the scripts in this section utilize an input file called the "seq" (sequence) file. This file is intended to be a simple, yet powerful method of concisely defining the various parameters needed in order to generate XPLOR restraint files.

Two simple examples of this format are shown on the next page. In the first example, a standard B-form duplex DNA with sequence 5'-ATGC-3' is represented. Notice that the defaults (as defined by the Insight95 software package) for the B-form torsion angles and the default range of motion are defined at the beginning of this file.

In the next example, an A-form RNA monomer hairpin is being represented, 5'-UACAGUUUGUCUA-3'. Notice that the "ADDTONUM 20" line causes the numbering of the nucleotide to begin with the number 21, otherwise the numbering will begin with the number 1. Thus, the first uridine nucleotide will be named "U21". Some of the nucleotide letters are upper case, and some are not. If an upper-case nucleotide letter is found, the "default" torsion angles, as defined earlier in the file, are used for generating the "cdih.dat" X-PLOR restraint file. If a lower-case nucleotide letter is found, followed by the word 'none', then the X-PLOR restraint files will be built with no restraints for that nucleotide. Finally, if a lower-case letter is found followed by a list of the torsion angles as shown in the example for U26, that nucleotide will be given those specific torsion angle restraints (in this case, C2' endo sugar pucker, and more relaxed backbone angles).

8.1.2 cdih_make - create XPLOR dihedral files

This script creates XPLOR dihedral restraint files from the “sequence file” from section 7.2.1.

Syntax: **cdih_make < seq_file > cdih.dat**

In this example, the sample seq file for building the DNA 5’-ATGC-3’ will be used and the first 30 lines of the XPLOR cdih.dat restaint file will be shown. Note, however, that the full length of this restaint file is ~400 lines. Also notice that the header information the cdih_make generates includes information on the length of the DNA, and the sequence. Because the input seq file contained the “segment a” and “segment b” lines, XPLOR segids are used in the atom definitions.

```
bass (lapham): [~/xplor/thesis]> cdih_make < atgc > cdih.dat
bass (lapham): [~/xplor/thesis]> head -30 cdih.dat
! file cdih_std.dat

! Backbone restraints are from Arnott fiber
! coordinates as reported by Altona (1982).
! Sequence length (per strand): 4
! There are 2 segments in this structure
! Segment #1 is named a and has sequence:ATGC
! Segment #2 is named b and has sequence:GCAT
!

!------------------------------------------------
! Torsional angle alpha
! defined: O3'(n-1)-P-O5'-C5'
!------------------------------------------------
! T2 of segment: a
assign (segid a and resid 1 and name O3') (segid a and resid 2 and name P  )
       (segid a and resid 2 and name O5') (segid a and resid 2 and name C5')
       1 -46.8  20  2

! G3 of segment: a
assign (segid a and resid 2 and name O3') (segid a and resid 3 and name P  )
       (segid a and resid 3 and name O5') (segid a and resid 3 and name C5')
       1 -46.8  20  2

! C4 of segment: a
assign (segid a and resid 3 and name O3') (segid a and resid 4 and name P  )
       (segid a and resid 4 and name O5') (segid a and resid 4 and name C5')
       1 -46.8  20  2
```

```
"alpha",0,"beta",1,"gamma",2,"epsilon",3,"zeta",4,
"chi",5,"nu0",6,"nu1",7,"nu2",8,"nu3",9,"nu4",10);




($first) = (split)[0];

# Is the first word a remark character?
if (/^!/) {# print "Remark: $_";
    next;}

# Is the first word a angle name?  If so, define the angle.
if ($angle_names{$first} ne '') {

($angle_default[$angle_names{$first}],$flex_default[$angle_

    # print "$first angle:

    next;}

# Is the first word a segment name?  If so, define the

if ($first eq "segment") {
    ++$segid_num;
    $segid = (split)[1];

    # $segment[] holds all the segment names for use later
    $segment[$segid_num] = $segid;
    $seq_num=0;
    next;}
```

```
# Is the first word NOT a legit nucleotide letter?
if (/^[^agcutAGCUT]/) {next};

# Is the first nucleotide letter an a, g, c, u, or t?
# if so, the the user wants to define new angles!
if (/^[agcut]/) {

    # first, break the line into the nucleotide  letter and

    ($nuc,$temp) = split(/\s+/);

    # If the person puts the word "none" for the second

    # then there are no angle restrictions for that

    if ($temp eq "none") {
        $list[$counter] = "$segid:$seq_num:$nuc:none:non e";
        }
    else {
    (@temp) = split(/:|,/,$temp);
    (@angle) = @temp[0,2,4,6,8,10,12,14,16,18,20];
    (@flex) = @temp[1,3,5,7,9,11,13,15,17,19,21];

    # set new angle and flex to appropriate variable names

    # set Master list
    $list[$counter] = "$segid:$s eq_num:$nuc:@angle:@flex";
    }
}

# The angles are of stardard form
else {
    $nuc = $first;
    $list[$counter] =

    }

# All variables have been set, now we just have to print

# print $list[$counter],"\n";
++$counter;
++$seq_num;
```

```
print "! Segment #",$i," is named $segment[$i] and has

foreach $j (($i-1)*$seq_num .. ($i*$seq_num-1)) {
    (@temp) = split(/:/,$list[$j]);
    print $temp[2];
}
```

```
# Call $segid the name of whichever segment name we are

# $segment[] holds all the segment names
$segid = $segment[$i];

# Count (0 .. x-1) for 1st segment, (x .. 2x-1) for 2nd

# This way we can use $list[] to retrieve all the info for

foreach $j (($i-1)*$seq_num .. ($i*$seq_num-1)) {
    (@temp) = split(/:/,$list[$j]);

    # Remember to skip over anything which has "none" for
```

```
if ($temp[3] eq "none") { next;}

$nuc_num = @temp[1]+1;
$nuc = @temp[2];
$temp_angle = @temp[3];
$temp_flex = @temp[4];
(@temp_angle) = split(/\s+/,$temp_angle);
(@temp_flex) = split(/\s+/,$temp_flex);
$angle = @temp_angle[$angle_names{$position}];
$flex = @temp_flex[$angle_names{$position}];

# if you are on alpha, you must subtract one from

if ($position eq "alpha") {$resid1=$nuc_num-1;


# if you are on epsilon, you must add one onto $resid4
elsif ($position eq "epsilon") {$res id1=$nuc_num;


# if you are on zeta, you must add one to both $resid3

elsif ($position eq "zeta") {$resid1=$nuc_num;


# All others can use $nuc_num for $resid
else {$resid1=$nuc_num; $resid2=$nuc_num;


# Don't print out angles which extend 5' of the 5' end,

if (($resid1 eq "0") || ($resid3 > $seq_num) ||


# Okay, now we have to take care of that DAMNABLE chi:
if ($position eq "chi") {
    if ($nuc =~ /[AGag]/) {$atom1="O4'"; $atom2="C1'";

    else {$atom1="O4'"; $atom2="C1'"; $atom3="N1 ";

    }

# For single strand with no segid defined
if ($segid eq "none") {
    print "!",$nuc,$nuc_num,"\n";
```

```
        print "assign (resid $resid1 and name $atom1)";
        print " (resid $resid2 and name $atom2) \n";
        print "       (resid $resid3 and name $atom3) ";
        print " (resid $resid4 and name $atom4) \n";
        print "       1 $angle  $flex  2\n";
        }

    # For multiple strands or single strand with segid

    else {
        print "! ",$nuc,$nuc_num," of segment: $segid\n";
        print "assign (segid $segid and r esid $resid1 and

        print " (segid $segid and resid $resid2 and name

        print "       (segid $segid and resid $resid3 and

        print " (segid $segid and resid $resid4 and name

        print "       1 $angle  $flex  2\n\n";
    }
}
```

8.1.3 planar_make – create XPLOR planar restraint files

Dan Zimmer wrote the original version of this script. It creates XPLOR planar restraint files from an input seq file.

Syntax: **planar_make < seq_file > planar_restraint_file**

In this example, the DNA seq file from section 7.2.1 is used to create a planarity restraint file, and the first 30 lines of the output planar.dat file are shown.

```
bass (lapham): [~/xplor/thesis]> planar_make < atgc > planar.dat
bass (lapham): [~/xplor/thesis]> head -30 planar.dat
! file: planar.dat
! DNA
! Planar restraints to maintain base pair planarity
! This file was created by planar_make.pl
! Update October 22, 1996
! NOTE: $PSCALE MUST BE DEFINED WITHIN THIS FILE OR IN EACH PROTOCOL!
! OTHERWISE THE DEFAULT IS: 300kcal/mol/A^2

evaluate ($pscale = 50)

! A1-T4 Watson-Crick
!--------------------------------------------------------------------
group
selection= ((segid a and resid  1 and name n1) or (segid a and resid  1 and name n3) or
        (segid a and resid  1 and name c5) or (segid b and resid  4 and name n1) or
        (segid b and resid  4 and name n3) or (segid b and resid  4 and name c5))
weight = $pscale end

! T2-A3 Watson-Crick
!--------------------------------------------------------------------
group
selection= ((segid a and resid  2 and name n1) or (segid a and resid  2 and name n3) or
        (segid a and resid  2 and name c5) or (segid b and resid  3 and name n1) or
        (segid b and resid  3 and name n3) or (segid b and resid  3 and name c5))
weight = $pscale end

! G3-C2 Watson-Crick
!--------------------------------------------------------------------
group
selection= ((segid a and resid  3 and name n1) or (segid a and resid  3 and name n3) or
        ($first) = (split)[0];

    # Is the first word a remark character?
    if (/^!/) {# print "Remark: $_";
        next;}

    # Is the first word a segment name?  If so, define the

    if ($first eq "segment") {
        ++$segid_num;
        $i=1;
        $segid = (split)[1];

        # $segment[] holds all the segment names for use later
        $segment[$segid_num] = $segid;
        $seq_num=0;
        next;}


    # Is the first word NOT a legit nucleotide letter?
    if (/^[^agcutAGCUT]/) {next};

    #
    if ($segid_num == 1) {
        $res[$i] = $first;
        $i++;
        $m=$i;
    }
```

```
    $j=$m-$i;
    if ($res[$i] =~ /[aA]/) {
        print "! A$i-T$j Watson-Crick\n";
        print "!-----------------------------------------------

        print "group\n";
        print "selection=\t((segid a and resid  $i and name n1)

        print "\t(segid a and resid  $i and name c5) or (segid

        print "\t(segid b and resid  $j and name n3) or (segid

        print "weight = \$pscale end\n";
    }
    elsif ($res[$i] =~ /[gG]/) {
        print "! G$i-C$j Watson-Crick\n";
        print "!-----------------------------------------------

        print "group\n";
        print "selection=\t((segid a and resid  $i and name n1)

        print "\t(segid a and resid  $i and name c5) or (segid

        print "\t(segid b and resid  $j and name n3) or (segid

        print "weight = \$pscale end\n";
    }
    elsif ($res[$i] =~ /[cC]/) {
        print "! C$i-G$j Watson-Crick\n";
        print "!-----------------------------------------------
        print "group\n";
        print "selection=\t((segid a and resid  $i and name n1)

        print "\t(segid a and resid  $i and name c5) or (segid

        print "\t(segid b and resid  $j an d name n3) or (segid

        print "weight = \$pscale end\n";
    }
    elsif ($res[$i] =~ /[tT]/) {
        print "! T$i-A$j Watson-Crick\n";
        print "!-----------------------------------------------

        print "group\n";
        print "selection=\t((segid a and resid  $i and name n1)

        print "\t(segid a and resid  $i and name c5) or (segid

        print "\t(segid b and resid  $j and name n3) or (se gid

        print "weight = \$pscale end\n";
    }
    print "\n";
    $i++;
```

8.1.4 dm – measures the distances between protons in pdb files

Dave Schweisguth originally wrote this script. It measures the distances between protons in a pdb structure file.

Syntax: **dm < pdb_file > distance_output_file**

In this example, the distances between all the protons within 5Å of each other for the pdb file “dickerson.pdb” will be saved to a file, the first 20 lines of the file will then be examined.

```
bass (lapham): [~/xplor/thesis]> dm < dickerson.pdb > distances
bass (lapham): [~/xplor/thesis]> head -20 distances
A CYT 1 H1'     A CYT 1 H2''       2.372
A CYT 1 H1'     A CYT 1 H2'        3.032
A CYT 1 H1'     A CYT 1 H3'        3.933
A CYT 1 H1'     A CYT 1 H4'        3.646
A CYT 1 H1'     A CYT 1 H5'        4.461
A CYT 1 H1'     A CYT 1 H6         3.697
A CYT 1 H1'     A GUA 2 H1'        4.921
A CYT 1 H1'     A GUA 2 H2'        4.144
A CYT 1 H1'     A GUA 2 H3'        4.965
A CYT 1 H1'     A GUA 2 H4'        4.240
A CYT 1 H1'     A GUA 2 H5'        1.805
A CYT 1 H1'     A GUA 2 H5''       3.362
A CYT 1 H1'     A GUA 2 H8         2.828
A CYT 1 H2''    A CYT 1 H2'        1.758
A CYT 1 H2''    A CYT 1 H3'        2.703
A CYT 1 H2''    A CYT 1 H4'        4.095
A CYT 1 H2''    A CYT 1 H5'        4.964
A CYT 1 H2''    A CYT 1 H5''       4.939
A CYT 1 H2''    A CYT 1 H6         3.424
A CYT 1 H2''    A GUA 2 H2'        3.835
```

```
              =~ s|.*/||;    # `basename $0`

      = 1;
      = 5;

                                   # Switches with arguments

                    { $d_diff = $arg; }

              { &usage("$whatami: -$first is not an

/^(?:ATOM  |HETATM).{5} (.{5})(.{3}) (.)(.{ 4}).

# Column 17 ("alternate location") appended to columns 13-

#     (atom type), mostly to compensate for bad Insight PDB
# Columns 77-80 (element and charge fields) ignored
```

```
$i =~ s/^\s*(\S*)\s*/$1/;

# Chain and segment IDs not always present, so not required

"$res_type $res_num $atom_type");

next if abs($res_num[$i] - $res_num[$j]) > $res_diff;

    ($z[$i] - $z[$j]) ** 2);
printf("$tag[$i]\t$tag[$j]\t%8.3f\n", $d)  unless $d >

    Print distances less than # (default 5)
    Print distances between residue numbers differing by

    (default 1)
This message
```

8.1.5 noe_in – converts the output of dm to an XPLOR input format file

This is a simple script to convert the output of the dm script to a XPLOR readable restrain file. Jason Rife wrote the original version of the program.

Syntax: **noe_in < dm_output_file > distance_restraint_file**

In this example, the 'distances' file generated from the 'dm' script will be used to build an XPLOR distance restraint file. The first 20 line of the resultant restraint file will be examined.

```
bass (lapham): [~/xplor/thesis]> noe_in < distances > noe.dat
bass (lapham): [~/xplor/thesis]> head -20 noe.dat
assign (resid CYT and name 1)
         (resid A and name CYT)  1 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name CYT)  1 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name CYT)  1 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name CYT)  1 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name CYT)  1 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name CYT)  1 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name GUA)  2 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name GUA)  2 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name GUA)  2 0.1 0.1
assign (resid CYT and name 1)
         (resid A and name GUA)  2 0.1 0.1
```

The source code for the noe_in script:

```
#!/usr/local/bin/perl
# noe_in
# Creates noe constraint input file from 'dm' output file.
# Jason P. Rife 11/24/95
# Jon Lapham edit 3/14/96 to automatically remove HO2'
# usage: noe_in dm_file > out_file

while(<>) {
        ($junkA,$resIDA,$nameA,$junkB,$resIDB,$nameB,$dist) = split(/\s+/,$_);
    if (($nameA eq "H5'") && ($nameB eq "H5'")) {next;}
    if (($nameA eq "HO2'") || ($nameB eq "HO2'")) {next;}
    print "assign (resid $resIDA and name $nameA) \n  \t (resid $resIDB and name $nameB)
$dist 0.1 0.1 \n";
}
```

8.1.6 noe_hbond_make – builds an XPLOR hydrogen bonding restraint file

This script generates the hbond.dat restraint file which forces standard base pairing distances between two nucleotides. This is accomplished by defining the distances between a few heavy atoms as found in a standard Watson-Crick type base pair.

USAGE: **noe_hbond_make < seq_file > hbond.dat**

In the example below, the h-bond restraint file is generated from the input seq file as shown for the DNA ATGC in section 7.2.1. The first 2 base pairs of the resultant h-bonding restraint file will be examined.

```
bass (lapham): [~/xplor/thesis]> noe_hbond_make < atgc > noe_hbond.dat
bass (lapham): [~/xplor/thesis]> head -20 noe_hbond.dat
! base pairing constraint file
! created by noe_hbond_make.pl
!
! A1-T4 Watson-Crick (B-form DNA)
assign (segid A and resid  1 and name N1 )
       (segid B and resid  4 and name H3 )  1.92 0.20 0.20
assign (segid A and resid  1 and name N1 )
       (segid B and resid  4 and name N3 )  2.95 0.20 0.20
assign (segid A and resid  1 and name N6 )
       (segid B and resid  4 and name O4 )  2.81 0.20 0.20
assign (segid A and resid  1 and name H62)
       (segid B and resid  4 and name O4 )  1.78 0.20 0.20

! T2-A3 Watson-Crick (B-form DNA)
assign (segid A and resid  2 and name H3 )
       (segid B and resid  3 and name N1 )  1.92 0.20 0.20
assign (segid A and resid  2 and name N3 )
       (segid B and resid  3 and name N1 )  2.95 0.20 0.20
assign (segid A and resid  2 and name O4 )
       (segid B and resid  3 and name N6 )  2.81 0.20 0.20
assign (segid A and resid  2 and name O4 )
       (segid B and resid  3 and name H62)  1.78 0.20 0.20
```

```
    ($first) = (split)[0];

# Is the first word a remark character?
if (/^!/) {# print "Remark: $_";
    next;}

# Is the first word a segment name?  If so, define the

if ($first eq "segment") {
    ++$segid_num;
    $i=1;
    $segid = (split)[1];
    $start_num = (split)[2];
    if ($start_num eq "") {$start_num =0;}
    $start_num[$segid_num]=$start_num;

    # $segment[] holds all the segment names for use later
    $segment[$segid_num] = $segid;
    $seq_num=0;
    next;}


# Is the first word NOT a legit nucleotide letter?
if (/^[^agcutAGCUT]/) {next};

#
if ($segid_num == 1) {
    $res[$i] = $first;
    $i++;
    $m=$i;
}
```

```
$i=$index + $start_num[1];
$j=$m-$i + $start_num[2];
if ($res[$index] =~ /[aA]/) {
    print "! A$i-T$j Watson-Crick (B-form DNA)\n";
    print "assign (segid A and resid  $i and name N1 )

    print "assign (segid A and resid  $i and name N1 )

    print "assign (segid A and resid  $i and name N6 )

    print "assign (segid A and resid  $i and name H62)

}
elsif ($res[$index] =~ /[gG]/) {
    print "! G$i-C$j Watson-Crick (B-form DNA)\n";
    print "assign (segid A and resid  $i and name H1 )

    print "assign (segid A and resid  $i and name N1 )

    print "assign (segid A and resid  $i and name H22)

    print "assign (segid A and resid  $i and name N2 )

    print "assign (segid A and resid  $i and name O6 )

    print "assign (segid A and resid  $i and  name O6 )


}
elsif ($res[$index] =~ /[cC]/) {
    print "! C$i-G$j Watson-Crick (B-form DNA)\n";
    print "assign (segid A and resid  $i and name N3 )

    print "assign (segid A and resid  $i and name N3 )

    print "assign (segid A and resid  $i and name O2 )
```

```
        print "assign (segid A and resid  $i and name O2 )

        print "assign (segid A and resid  $i and name H42)

        print "assign (segid A and resid  $i and name N4 )

    }
    elsif ($res[$index] =~ /[tT]/) {
        print "! T$i-A$j Watson-Crick (B-form DNA)\n";
        print "assign (segid A and resid  $i and name H3 )

        print "assign (segid A and resid  $i and name N3 )

        print "assign (segid A and resid  $i and name O4 )

        print "assign (segid A and resid  $i and name  O4 )

    }
    $index++;
```

8.1.7 cdih_measure – measures the dihedral angles of nucleic acid pdb files

This script is useful for measuring the heavy atom torsion angles from a nucleic acid PDB file. For nucleic acids, 11 torsion angles completely describe the conformation of a nucleotide monomer, named α, β, γ, ε, ζ, X, nu0, nu1, nu2, nu3 and nu4 with definitions as given below

```
     alpha = O3'(n-1)-P-O5'-C5'         nu0 = C4'-O4'-C1'-
C2'
     beta = P-O5'-C5'-C4'               nu1 = O4'-C1'-C2'-
C3'
     gamma = O5'-C5'-C4'-C3'       nu2 = C1'-C2'-C3'-C4'
     epsilon = C4'-C3'-O3'-P(n+1)       nu3 = C2'-C3'-C4'-
O4'
     zeta = C3'-O3'-P(n+1)-O5'(n+1)     nu4 = C3'-C4'-O4'-
C1'
     chi(pur) = O4'-C1'-N9-C4
     chi(pyr) = O4'-C1'-N1-C2
```

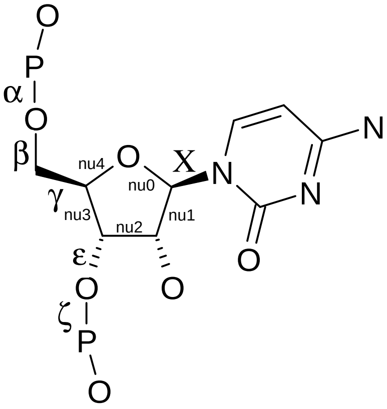


Figure 8. 4 Definition of torsion angles in nucleic acids

In order to perform this calculation, a general method for calculating torsion angles between two vectors that share a common third vector must be developed. Credit for this program must be given to discussions with Dan Zimmer and Charlie Schmuttenmaer.

This is a quick overview of how the torsion angles are calculated. Given that vectors **A** and **B** share are intersected by a common third vector **C**.

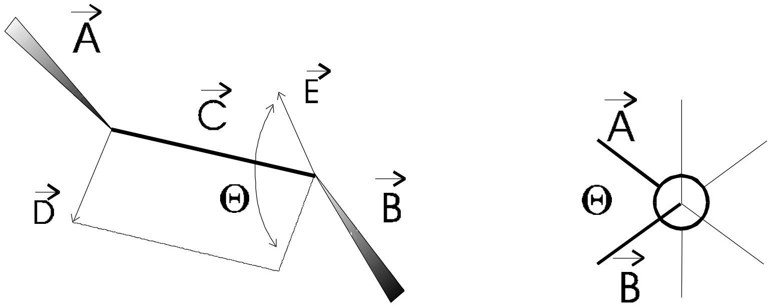


Figure 8. 5 Vector representation of the torsion angles

The cross product of **A** with **C** gives **D**, a vector orthogonal to both **A** and **C**. Likewise, the cross product of **B** with **C** gives **E**, a vector orthogonal to both **B** and **C**. Given that **A** and **B** both intersect **C**, the two planes that describe the possible positions of both **D** and **E** must be parallel. This requires that the angle that **D** and **E** make (as shown in the figure above with the two headed arrow) will truly represent θ, the torsion angle between **A** and **B**. The magnitude of the dot product between **D** and **E** gives the torsion angle.

$$\mathbf{A} \times \mathbf{C} = \mathbf{D}$$
$$\mathbf{B} \times \mathbf{C} = \mathbf{E}$$
$$\mathbf{D} \bullet \mathbf{E} = DE\cos\Theta = |D||E|\cos\Theta$$
$$\Theta = \arccos(|D||E|)$$

The next two pages show an example of the output from cdih_measure for the dickerson dodecamer 12 base pair DNA. The DNA was generated using the program Insight95 as standard B-form DNA. Notice that along with the individual torsion angles, the "P" (pseudo-rotation angle), numax and sugar pucker are calculated, as well. These sugar conformation calculations were an addition to the program by Dan Zimmer, thanks!

This example was run with the argument “define”, which caused all the header information (definitions of angles, etc..) to be displayed. For more concise output, omit the word “define”.

```
  C3'-exo
  C4'-endo
  O4'-exo
  C1'-endo
  C2'-exo
36   C3'-endo    A
  C4'-exo
  O4'-endo
  C1'-exo
  C2'-endo    B
```

```
15      38      C3'-endo
        36      C3'-exo
        38      C3'-endo
```

```
"alpha", ["O3'","P","O5'","C5'"],
"beta", ["P","O5'","C5'","C4'"],
"gamma",["O5'","C5'","C4'","C3'"],
"chi", ["O4'","C1'","xx","xx"],
"eps", ["C4'","C3'","O3'","P"],
"zeta", ["C3'","O3'","P","O5'"],
"nu0", ["C4'","O4'","C1'","C2'"],
"nu1", ["O4'","C1'","C2'","C3'"],
"nu2", ["C1'","C2'","C3'","C4'"],
"nu3", ["C2'","C3'","C4'","O4'"],
"nu4", ["C3'","C4'","O4'","C1'"]


print "\n    Name = Definition\n";


print "\nStandard Values:";
```

```
print "


shift (@ARGV);
}


($segid) = ($1);
# print "Segid: $segid\n";


$i =~ s/^\s*([^\s]*)\s*$/$1/;


die "cdih_measure: Bad PDB record on input line ",
    if $i eq '';



$segid="";
$seg_num=0;}

if ($segid ne $segid[$seg_num]) {


    }
















$segid = $segid[$h];




if ($j eq "chi") {
    ($atom_type1, $res_type1, $res_num1, $xyz1) =
    ($types{$j}[0], $res[$i]{$segid}, $i,

    ($atom_type2, $res_type2, $res_num2, $xyz2) =
    ($types{$j}[1], $res[$i]{$segid}, $i,

    if (($res[$i]{$segid} eq 'A') || ($res[$i]{$segid} eq

    ($atom_type3, $res_type3, $res_num3, $xyz3) =
        ("N9", $res[$i]{$segid}, $i,
```

```
    ($atom_type4, $res_type4, $res_num4, $xyz4) =
        ("C4", $res[$i]{$segid}, $i,

    } else {
    ($atom_type3, $res_type3, $res_num3, $xyz3) =
        ("N1", $res[$i]{$segid}, $i,

    ($atom_type4, $res_type4, $res_num4, $xyz4) =
        ("C2", $res[$i]{$segid}, $i,

    }
} else {
    if ($j eq "alpha") {
    if ($i == 1) {
        $angle{'alpha'} = 0; next;
    } else {
        ($atom_type1, $res_type1, $res_num1, $xyz1) =
            ($types{$j}[0], $res[$i]{$segid}, $i-1,

    }
    } else {
    ($atom_type1, $res_type1, $res_num1, $xyz1) =
        ($types{$j}[0], $res[$i]{$segid}, $i,

    }

    ($atom_type2, $res_type2, $res_num2, $xyz2) =
    ($types{$j}[1], $res[$i]{$segid}, $i,


    if ($j eq "zeta") {
    if ($res[$i+1]{$segid}) {
        ($atom_type3, $res_type3, $res_num3, $xyz3) =
        ($types{$j}[2], $res[$i+1]{$segid}, $i+1,

    } else {
        $angle{'zeta'} = 0;
        next;
    }
    } else {
    ($atom_type3, $res_type3, $res_num3, $xyz3) =
        ($types{$j}[2], $res[$i], $i,

    }
    if (($j eq "zeta") && $res[$i+1]{$segid}) {
    ($atom_type4, $res_type4, $res_num4, $xyz4) =
        ($types{$j}[3], $res[$i+1]{$segid}, $i+1,

    } elsif ($j eq "eps") {
    if ($res[$i+1]{$segid}) {
         ($atom_type4, $res_type4, $res_num4, $xyz4) =
        ($types{$j}[3], $res[$i+1]{$segid}, $i+1,

    } else {
        $angle{'eps'} = 0;
        next;
    }
    } else {
    ($atom_type4, $res_type4, $res_num4, $ xyz4) =
        ($types{$j}[3], $res[$i]{$segid}, $i,

    }
}

($x1, $y1, $z1) = @$xyz1;
($x2, $y2, $z2) = @$xyz2;
($x3, $y3, $z3) = @$xyz3;
($x4, $y4, $z4) = @$xyz4;

$r21x=$x2-$x1; $r21y=$y2-$y1; $r21z=$z2-$z1;
$r23x=$x2-$x3; $r23y=$y2-$y3; $r23z=$z2-$z3;
$r34x=$x3-$x4; $r34y=$y3-$y4; $r34z=$z3-$z4;
$Ax=($r21y*$r23z)-($r21z*$r23y);
$Ay=($r21z*$r23x)-($r21x*$r23z);
$Az=($r21x*$r23y)-($r21y*$r23x);
$Bx=($r34y*$r23z)-($r34z*$r23y);
$By=($r34z*$r23x)-($r34x*$r23z);
$Bz=($r34x*$r23y)-($r34y*$r23x);
$magA=sqrt($Ax**2+$Ay**2+$Az**2);
$magB=sqrt($Bx**2+$By**2+$Bz**2);
$dot=($Ax*$Bx+$Ay*$By+$Az*$Bz);

#Avoid a "divide by zero error"
if ($magA*$magB ne 0) {
$arccos=$dot/($magA*$magB);

$bot=$arccos;
$top=sqrt(1-$arccos**2);
```

```
$theta=(180/3.1415927)*atan2($top,$bot);

# new shit to calculate sign
# calc theta2 between A and r34 because if <90, theta
# should be positive, if >90 theta should be negative
$mag34=sqrt($r34x**2+$r34y**2+$r34z**2);
$dot2=($r34x*$Ax+$r34y*$Ay+$r34z*$Az);
$arccos2=$dot2/($magA*$mag34);
$theta2=(180/3.1415927)*atan2(sqrt(1-

if ($theta2 > 90) {$theta=$theta*-1;}

# Set the output variables
$angle{$j} = $theta;
}
else {
$angle{$j} = 0;}
```

8.2 Hydrodynamics

Calculation of the rotational and translational diffusion properties of regularly shaped hydrodynamic particles is presented in this section. These values are important for the evaluation of a number of NMR experiments. The rotational diffusion rate of a molecule (which can be expressed as a correlation time) is intimately related to the dipolar relaxation parameters for the molecule (see Chapter 5 of this thesis). The translational diffusion rate can be measured experimentally using the experiments presented in Chapter 4 of this thesis, and the ability to calculate a theoretic value is important in the interpretation of the results of the experiments.

Four hydrodynamic particle shapes are supported in this program; a sphere, a prolate ellipse, an oblate ellipse and a right cylinder. The equations for the calculations can be found in the references from Chapter 4 and 5. Below is an example of running the program for calculating first the rotational properties of a 12 mer DNA using the standard rise/base pair and diameter values in D2O at 25° C:

```
bass (lapham): [~]> hydro.pl
hydro.pl
    A program for simulating rotational and translational
    diffusion constants for nucleic acids from model
    hydrodynamic systems.

Would you like to enter the hydrodynamic parameters of a/b explicitly (y/[n])?n
Enter hydrodynamic diameter ([bdna], arna or angs): bdna
   using 20A diameter
Enter hydrodynamic rise/bp ([bdna], arna or angs): bdna
   using 3.4 rise/bp
Enter number of basepairs [12]: 12
   using 12 base pairs
Enter temperature (celcius)[25]: 25
   using 25 C
Enter viscosity ([d2o], h2o or user_specified): d2o
   using d2o for viscosity
Translational or rotational calculations ([trans] or rot): rot
   calculating rotational values
      units of Dr are (s-1)

 T  #bp Nu        Dr_s     Dr_pe_a  Dr_pe_b  Dr_cr_a  Dr_cr_b
 25  12 1.097e-03 1.76e+07 1.41e+08 2.03e+07 5.92e+07 2.58e+07
```

The results from this calculation are the the rotaional diffusion rate for the DNA is 1.76x10^{7} for the spherical model, 1.41x10^{8} and 2.03x10^{7} for the two axis of the prolate ellisoid model and 5.92x10^{7} and 2.58x10^{7} for the two axis of the cylindrical rod model. All the rates are, naurally, in units of s^{-1}. To convert the rotational diffusion rates (D_r) to correlation times (t_c), the equation is:

$$t_c = \frac{1}{6 \cdot D_r}$$

For example, the correlation time of the long axis of the DNA using the cylindrical rod model is:

$$t_l = \frac{1}{6 \cdot 5.92x10^7 s^{-1}} = 2.81ns$$

```
$pregunta = <STDIN>; chop $pregunta;
```

```
$h_diameter = <STDIN>; chop $h_diameter;




$h_rpb = <STDIN>; chop $h_rpb;




$nbp=<STDIN>; chop $nbp;


$temp=<STDIN>; chop $temp;

$nu_type=<STDIN>; chop $nu_type;


$calc_type = <STDIN>; chop $calc_type;
```

```
$temp_c, $bp, $nu, $Dt_s, $Dt_pe, $Dt_oe, $Dt_cr);


$temp_c, $bp, $nu, $Dr_s, $Dr_pe_a, $Dr_pe_b, $Dr_cr_a,
# $a is bigger than $b, i.e.:DNA shorter than 6 bp...
# This keeps us from getting 'nan' (not a number) outputs
$Ft_pe = 0; $Fr_pe_a = 0; $Fr_pe_b = 0; $Dt_pe = 0;
```

8.3 Moment of Inertia

Many of the calculations presented in this thesis require knowledge of the “principal axis of rotation” for a given molecule. For instance, the definition of the spectral density function for anisotropic rotation as formulated in section 5.5.3, has a b term, defined as the angle the ij atom pair makes with respect to the principal axis.

The rigorously correct method for determining the principal axis of rotation for a molecule in a solvent, would be to exactly determine the rotational diffusion tensor. However, this calculation is extremely difficult to perform, as it requires knowledge of the frictional coefficient. Approximation methods have been developed for certain hydrodynamic “regular” shapes, such as spheres, ellipses and cylinders (see the previous section, 7.3). But, what if you, the biomolecular spectroscopist, have a uniquely unusual shaped biomolecule and you want a rough estimate of the propensity of the structure to rotate in an anisotropic manner? And you want a quantitative method for determining the principal axis of rotation (read: so computer programs can take over the process).

The only method I have found for accomplishing these goals is to determine the “moments of inertia” for the molecule. This was derived from discussions with Profs. Kurt Zilm and Charlie Schmuttenmaer and the mathematics comes directly from the text “Classical Dynamics of Particles and Systems”. The theory is developed in chapter 5 in section 5.8.2, if you are interested.

This is a program that calculates the inertia tensor of an arbitrary molecular structure and returns information related to the moments of inertia for that molecule.

8.3.1 Examples

Below are a few examples of how to use this program in everyday life, and how to incorporate it into other programs. The text shown below is the actual output from the program.

Running the program 'principal_axis' with no command line arguments prints out a short usage listing:

```
bass (lapham): [~/xplor/dickerson]> principal_axis
USAGE:
principle_axis <-options> <pdb file>
options:
    -segid  : Calculate segments seperately
    -nomass : Use mass 1 for every atom
    -midas  : Automatically start midas
    -report : Print a report to STDOUT
    -base   : Use only nucleic acid base heavy atoms
    -range  : Prompt for valid residues
    -xy     : Print small X and Y axis
    -short  : Print only the 3 principle axis XYZ components
```

The simplest way to use the program (and probably all most people will need) is to use the '-short' qualifier:

```
bass (lapham): [~/xplor/dickerson]> principal_axis -short dick_b.pdb
0.159361767996386 0.0325789844992612 0.986682540977625
```

The numbers that are returned above are the normalized unit length vector that lies parallel to the principal axis of the molecule (dick_b.pdb in this case).

The most information rich method of using the program is to use the '-report' qualifier. This prints to standard output a report that includes the actual numbers used in the initial inertia tensor matrix, the eigenvalues and eigenvectors of the inertia tensor, the molecular center of mass, the principal axis vector components, the number of segments (XPLOR definition of segments) and number of atoms used in the calculation.

```
bass (lapham): [~/xplor/dickerson]> principal_axis -report dick_b.pdb

=========================================================
Calculation for pdbfile:dick_b.pdb segment: all

Starting matrix:
               x                 y                 z
```

```
       ---------         ---------         ---------
593543.056251672  3262.04898805153  -72434.171595150
3262.04898805153  577973.569530177  -14798.775561448
-72434.171595150  -14798.775561448  157924.089998264

Diagonalized matrix (eigenvalues):
               x                 y                 z
       ---------         ---------         ---------
  6.06386485E+05                 0                 0
               0    5.77317817E+05                 0
               0                 0    1.45736414E+05

Transformation matrix (eigenvectors):
               x                 y                 z
       ---------         ---------         ---------
  9.67070506E-01   -1.98440075E-01    1.59361768E-01
  1.95747721E-01    9.80112973E-01    3.25789845E-02
 -1.62657512E-01   -3.11472103E-04    9.86682541E-01

The resultant MAJOR principle axis has:
X components: 1.59361768E-01
Y components: 3.25789845E-02
Z components: 9.86682541E-01
with:
actual X:7.87716043957749
actual Y:2.42422749530196
actual Z:68.6825969478827
and is centered at:
X component: -0.0909279604225049
Y component: 0.795278270301956
Z component: 19.3484698978827
There are 1 segments, segid_num is 0
with 380 valid atoms in this segment
=========================================================
```

Notice the three eigenvalue numbers above, X = 6.06x10^5, Y = 5.77x10^5 and Z = 1.46x10^5. These numbers are proportional to the actual "moments of inertia" about the principal axis vector, the smaller the number, the smaller the inertial moment about that axis. Thus, we can learn much about this molecule based on these numbers. If the three moments of inertia were the same, the molecule is described as a "spherical top". This means that the molecule is described by a single moment of inertia. Examples would be a sphere, a perfect cube or any other shape symmetric about all three axis.

In the example we used above, the Z component of the inertial moment is smaller than the other two. This is called a "symmetrical top" and it means that the shape is described by two moments of inertia, one about the principal axis and one about each of the other two axis.

The '-range' qualifier allows one to input a specific range of nucleic acid residues to use in the calculation. In the example below, only residue numbers 1, 2, 3, 4 and 5 will be used in the calculation:

```
bass (lapham): [~/xplor/dickerson]> principal_axis -range -short dick_b.pdb
Enter valid residue numbers, separated by spaces: 1 2 3 4 5

okay, calculating only for residue(s) 1 2 3 4 5
-0.469957055070988 0.501848901075805 0.726146023109685
```

The '-nomass' qualifier causes the program to arbitrarily weight all atoms with an atomic mass unit of 1. Normally, the correct mass of the atoms is used, consequently the heavier atoms (C, O, N, etc..) are weighted much more heavily (as they should be) in the calculation than the hydrogens.

Two nucleic acid specific qualifier were created. The '-base' qualifier will force the program to use only the base atoms in the calculation. This is useful for short segments of nucleic acid when you want to obtain a principal axis vector that runs through the center of the helix.

The '-segid' qualifier will perform the calculations on each segment (using the XPLOR definition of segid) separately.

A number of additional qualifiers were created for visualizing the results. The '-midas' qualifier causes midas to launch, graphically displaying the pdb file overlayed with either just the principal axis vector, or (with the '–xy' qualifier) with all three axis of the moments of inertia, as shown below in figure 7.5.

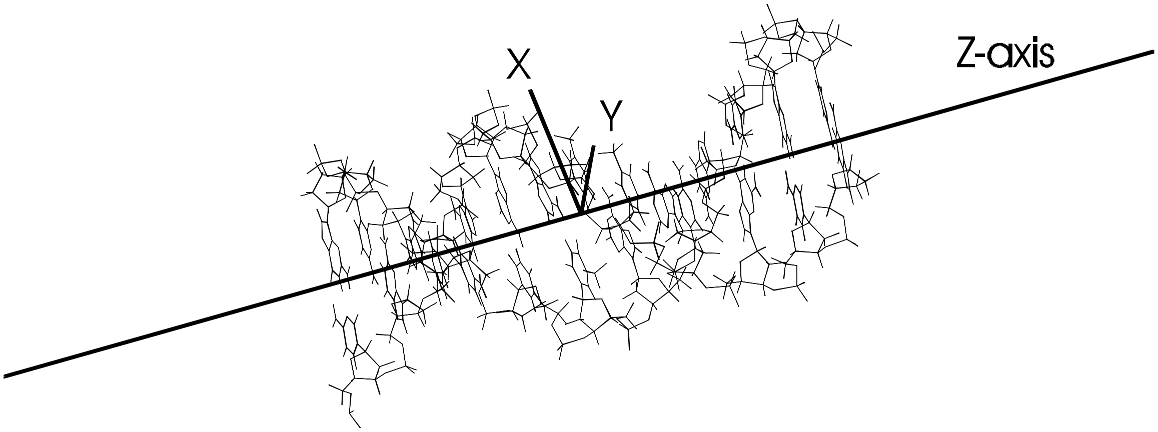


Figure 8. 6 principal_axis -xy -midas dickerson.pdb

```
lapham@tecate.chem.yale.edu



                                    # `basename $0`








                        { $analysis_type = "segid"; }
                                { $use_mass = "false"; }
                            { $launch_midas = "true";}
                                { $report = "true";}
                            { $base = "true";}
                            { $range = "true";}
                        { $xy = "true";}
                            { $short = "true";}


  $ARGV[0] = "-$rest";
```

```
  &center_mass_calc;



  &inertia_matrix_prep;
```

```
print "Launching midas...\n";
`midas $pdb_file $pdb_file.vector &`;









&reset_inertia_prep;
&center_mass_calc;
&inertia_matrix_prep;
&inertia_calc;



$ext="vector.$segid_num";



print "Launching midas...\n";
`midas $pdb_file $pdb_file.vector.* &`;
```

```
        $atom_num[$segid_num]=0;
```

```
next unless ($atom =~ /C2|C4|C5|C6|C8|N1|N3|N7|N9/);


next unless (grep(/$num/,@valid_res));






    $bogus_segid = "true";
    $segid_num=0;
    $segid[$segid_num] = $segid;
    $segid[$segid_num] = $segid;
    $atom_num[$segid_num] = 0;
    }




















    $bogus_segid = "true";
    $segid_num=0;
    $segid[$segid_num] = $segid;
    $segid[$segid_num] = $segid;

}
```

```
# Using exact atomic AMUs
# Ideally we want to match 0 or more numbers first...
$atom =~ s/^[0-9]*(\w)\w+/$1/;
if ($atom =~ /C/) {$mass=12.011;}
elsif ($atom =~ /H/) {$mass=1.00794;}
elsif ($atom =~ /N/) {$mass=14.00674;}
elsif ($atom =~ /O/) {$mass=15.9997;}
elsif ($atom =~ /P/) {$mass=30.973762;}
elsif ($atom =~ /S/) {$mass=32.066;}
else {print "ERROR:\n";
    print "I don't know the mass of $atom\n" ;
}
```

```
# Using exact atomic AMUs
# Ideally we want to match 0 or more numbers first...
$atom =~ s/^[0-9]*(\w)\w+/$1/;
if ($atom =~ /C/) {$mass=12.011;}
elsif ($atom =~ /H/) {$mass=1.00794;}
elsif ($atom =~ /N/) {$mass=14.00674;}
elsif ($atom =~ /O/) {$mass=15.9997;}
elsif ($atom =~ /P/) {$mass=30.973 762;}
elsif ($atom =~ /S/) {$mass=32.066;}
else {print "ERROR:\n";
    print "I don't know the mass of $atom\n";
}
```

```
++$count;
$U[$i][$j] = $temp_result[$count+2];
chop $U[$i][$j];
$U[$i][$j] =~ s/^\s*([^\s]*)\s*$/$1/;
```