Structure and Dynamics of the HIV-2 TAR RNA-Argininamide Complex by NMR Spectroscopy

by

Alexander S. Brodsky

B.A. in Biochemistry
University of Pennsylvania, 1992

M.S. in Chemistry
University of Pennsylvania, 1993

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

DOCTOR OF PHILOSOPHY
IN BIOCHEMISTRY

at the
Massachusetts Institute of Technology

February, 1998

Signature of Author ....
Department of Chemistry
December 8, 1997

Certified by ....
James R. Williamson
Associate Professor of Chemistry
Thesis Supervisor

Accepted by ....
Dietmar Seyferth
Chair, Departmental Committee on Graduate Studies

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

Professor Robert G. Griffin ....

Chair

Professor James R. Williamson...

Thesis Supervisor

Professor Larry Stern ....

ABSTRACT

Structure and Dynamics of the HIV-2 TAR RNA-Argininamide Complex by NMR Spectroscopy

by

Alexander S. Brodsky

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

ABSTRACT

The Trans Activating Region (TAR) RNA - Tat protein interaction is important for activation of transcription in the human immunodeficiency virus (HIV). A model complex for this interaction composed of the two base bulge HIV-2 TAR and the amide derivative of arginine was studied by multidimensional heteronuclear NMR. Because of the improved spectral properties of the HIV-2 TAR complex, a larger number of NOEs in the bulge region were observed than in previous studies of the HIV-1 TAR-argininamide complex. 681 NOE distance restraints were collected and used to determine the solution structure of the HIV-2 TAR-argininamide complex. As observed in the previously proposed model, the two A-form stems co-axially stack and the critical U23 and the argininamide are located in the major groove. Model calculations including non-experimental restraints indicate that U23 is within hydrogen bonding distance to A27 consistent with the formation of a U-A•U base triple. Base triple formation helps open the major groove to increase the accessibility of G26 to hydrogen bond donors from the guanidinium group of argininamide. Argininamide binding is stabilized by stacking of the guanidinium group between the bases of A22 and U23, forming an argininamide sandwich. A C-G•C$^+$ isomorphic base triple mutant was designed and demonstrated that base triple formation is required for argininamide binding. This base triple mutant provides direct NMR evidence for base triple formation. Relaxation rate measurements and spectral density calculations show that the base triple nucleotides are not very flexible compared to A-form residues, consistent with their involvement in a specific interaction. These studies represent some of the first detailed RNA dynamics measurements on isotopically labeled samples.

Thesis supervisor: James R. Williamson

Title: Associate Professor of Chemistry

Acknowledgments

Graduate school is not only a time to learn how to do science but also a time to question yourself and your goals. Many people have helped me through the many tough times during the course of this thesis, however, the limited space available allows me to mention only a few of these special people.

My advisor, James R. Williamson, not only believes in skiing and the occasional trips to the bar, but also pursuing excellent science along with learning other practical skills inherent to surviving at MIT including spelunking, roach eradication, painting, plumbing, magnet quenching, and gardening, which as we now know is an essential part of honing the free food and beer radar. Jamie sets the tone for a cooperative lab atmosphere along with the will to pursue many exciting, yet, challenging projects, even through the thin times of disposable glove conservation. Good work Jamie! Thanks!!

This atmosphere would not be possible without the many courteous and helpful people in the lab. I was fortunate to learn NMR from the Chris Turner and the ever-patient John Battiste, and more recently Kwaku Dayie, without whom Chapter 5 would not exist. No peak would have been assigned without the smooth and elegant working computers maintained by our resident UNIX deity, Christopher Cilley, who also was proficient in providing peptide cloning and purification expertise. I often commiserated with fellow spectroscopists, Hong-yuan Mao and Heidi Erlacher as well as classmate Martha Rook. Dan Treiber and Jeff Orr have been around for many entertaining scientific and non-scientific discussions and help put a little reality into graduate school life. Robert Batey helped make the isotopically labeled samples and shared in many, intense scientific discussions. Pat Zarrinkar was not only a walking library of useful information but also made sure the lab hummed along with minimal disruptions. Many thanks to all the Williamsonites who continue the tradition of long-winded discussion on any topic, Sharon Barker-Parker, Phil Zamore, Ivan Baxter, and the unforgettable Funke. All these people provided significant hands-on help and discussion to help me study all kinds of RNAs from $tRNA^{not}$ to TAR.

Life away from the stresses of graduate school is critical for mental health. Jen, Chris, and John were all part of many good times on and off the slopes. I released excess energy and aggression on the soccer field with many soccer teammates over the years including the ever-present Phil, Wilfred, and Tim. The continuing lifelong friendship of the man with many names (Mike, Cambo, Camby etc.) and all the NY home folks and the whole 3922 Sansom gang is immeasurable.

My family also deserves many thanks. My parents have always been extremely supportive and optimistic of my endeavors and provided many opportunities to pursue my interests. The many parental visits are appreciated more than I usually acknowledge. They were perhaps most excited when I told them that I had met a wonderful woman, Natasha Kablaoui, who recently became my wife. I will always cherish her love and support as a best friend.

Table of Contents

Chapter 1: Introduction

1.1 RNA Structure and RNA Complexes

RNA gets lots of respect. For many, this respect is warranted because RNA plays a central role in many critical biological processes and RNA has become the focus of research and life itself. In the NMR community, almost every solution and solid state NMR spectroscopist, is experiencing RNA envy and wants to study this macromolecule that is not a protein but unlike DNA also folds into biologically relevant complicated three dimensional structures. RNA is extensively studied because of its ability to help make proteins (Moore, 1993; Soll, 1993), to catalyze a wide variety of chemical reactions (Cech, 1993), to bind specifically to small molecules (Gold *et al.*, 1995), to regulate transcription and translation (Tan *et al.*, 1997), to help maintain telomeres (Blackburn, 1993), to help splice introns (Moore *et al.*, 1993), to regulate nuclear transport (Ullman *et al.*, 1997), and to regulate embryonic development (Gavis & Lehmann, 1992), among some of its fundamental functions (Gesteland & Atkins, 1993). Clearly, RNA is far more than just a transporter of the genetic message. In all these guises, RNA folds into complicated structures that enable it to perform its function much like any protein.

1.1.1 RNA structure

For many years, the only RNA structural information, other than from helical duplexes, was from the tRNA crystal structures. These remarkable structures provided a wealth of information and a hint of the diverse features available to RNA including base triples, metal binding, small molecule binding, loop structures, base-phosphate interactions, modified nucleotides, and base intercalation (Roberts *et al.*, 1974; Rould *et al.*, 1989; Ruff *et al.*, 1991). In many ways, these form the basic tertiary interactions that are found in RNA structure.

Secondary structure predictions have identified a number of basic RNA structural motifs such as loops, bulges, internal loops, three and four way junctions. These secondary structure maps often show large regions which are apparently unstructured. Recently, a number of RNA structures have been solved highlighting how these apparently unstructured regions often take advantage of non-Watson-Crick base pairing and base triples to organize into three-dimensional structures. These structures have expanded our knowledge of the versatility of RNA structure, as almost every new structure has presented new and exciting surprises. Some recurring motifs have been identified, including the GNRA tetraloop appearing in a most unexpected place as part of the ATP aptamer (Dieckmann *et al.*, 1996; Jiang *et al.*, 1996). Recently, larger, more complicated RNA structures have been solved by X-ray crystallography. The Hammerhead and P4/P5/P6 structures revealed new and recurring motifs of tertiary structure such as the tetraloop receptor, A-platforms, and anticodon loops (Cate *et al.*, 1996; Pley *et al.*, 1994a). Most larger RNAs require Mg^{2+} to fold and the P4/P5/P6 structure also showed how a large RNA specifically binds and folds around a Mg^{2+} core (Cate & Doudna, 1996; Cate *et al.*, 1996; Cate *et al.*, 1997).

1.1.2 RNA-protein interactions

RNA protein complexes are found at all levels of gene expression in biology. However, these complexes are often part of larger ribonucleoprotein complexes, like the spliceosome and the ribosome, making their dissection into isolated complexes difficult. In addition to the tRNA-tRNA synthetase structures (Biou *et al.*, 1994; Rould et al., 1989; Ruff et al., 1991), three RNA-protein complexes have recently been solved: the MS2 coat protein (Valegard *et al.*, 1994), U1A bound to a hairpin loop (Oubridge *et al.*, 1994), and a NMR structure of U1A bound to its mRNA regulation site (Allain *et al.*, 1996). These RNA protein complexes showed how the loop residues are splayed towards the protein allowing sequence specific recognition with no specific RNA conformation contributing to

the interaction. The next few years are expected to yield a number of new NMR and X-ray protein-RNA complex structures as the technology to purify and characterize them improves. Meanwhile, a number of RNA-protein interactions from retroviruses have been minimized to peptide-RNA complexes and their structures have been solved by NMR including the Rev peptide-RRE RNA (Battiste *et al.*, 1996; Ye *et al.*, 1996) and the BIV Tat - TAR complexes (Puglisi *et al.*, 1995; Ye *et al.*, 1995). These structures further emphasize themes originally outlined in the TAR-arginine model, that the direct amino acid-nucleotide specific contacts do not provide all the specificity for the interactions but rather specific non-canonical base pairing plays a critical role in specific molecular recognition, often by opening the major groove (Puglisi *et al.*, 1993; Puglisi *et al.*, 1992). In the major groove, these peptides' arginine guanidium groups form specific hydrogen bonds, often to guanines for molecular recognition.

1.1.3 RNA-ligand complexes

With the advent of *in vitro* Systematic Evolution of Ligands by Exponential enrichment (SELEX), RNAs have been found to bind tightly and specifically to a number of different small molecules (Gold *et al.*, 1995). Solutions structures of a number of aptamers has revealed the ability of RNA to organize into specific binding sites. The ATP aptamer revealed the surprising result that the ATP fit into a binding pocket as the A of a GNRA tetraloop structure (Dieckmann *et al.*, 1996; Ye *et al.*, 1996). The theophyllin aptamer explains how the aptamer discriminates between theophyllin and caffeine which differ by a single methyl group (Jenison *et al.*, 1994; Zimmerman *et al.*, 1997). This structure also found the new A-platform and S-turn motifs originally identified in the P4/P5/P6 (Cate et al., 1996) and Rev-RRE structures (Battiste et al., 1996; Battiste *et al.*, 1995; Ye et al., 1996), respectively. Other recent NMR structures include the arginine (Yang *et al.*, 1996), FMN (Fan *et al.*, 1996a), and tobramycin aptamers (Jiang *et al.*, 1997) which highlight RNA's ability to organize purine rich regions into distinct binding pockets.

1.2 RNA Structure Determination by NMR

1.2.1 General features of RNA NMR

Four important types of information accessible by NMR: ^{1}H-^{1}H NOEs, scalar couplings, chemical shifts, and dynamics. Of these four major NMR observables, the nuclear Overhauser effect (NOE) and the scalar coupling constant (J) are the most important to define a macromolecular structures. The third observable, the chemical shift, has been successfully applied to help refine protein secondary structure (Wishart & Sykes, 1994; Wishart *et al.*, 1991), but the database is not yet large enough for RNA NMR to utilize the chemical shift for structure refinements reliably. Recently, relaxation measurements have been introduced to measure long range structural restraints, but these difficult experiments are only in their infancy and have not yet been widely applied to protein structure determinations (Reif *et al.*, 1997; Tjandra *et al.*, 1997). The more common relaxation experiments help verify the observed structural data, but their correlations to the structure are often difficult to determine (Dayie *et al.*, 1996; Kay *et al.*, 1996).

The NOE is proportional to $1/r^6$ of the distance between any two protons so that every pair of protons less than 5 Å apart should be observable. The goal of NMR structure determinations is to try to assign as many of these proton-proton distances as possible to define the structure to the highest possible resolution. A number of studies have shown that not only the precision but also the accuracy is directly proportional with the number of distance restraints. The coupling constant, J, is sinusoidally proportional to torsion angles such that any three or more bond network, with NMR active nuclei at the ends, e.g. a H-C-C-H or H-C-O-P spin system, will have a J-coupling which is conformationally sensitive, as described by the Karplus equation (Karplus, 1969).

$$^3J = A\cos^2\Theta + B\cos\Theta + C \qquad (1)$$

The caveat of the Karplus equation is that the constants A, B, and C must be empirically parameterized for each three bond network. For nucleic acids this has been

A) Imino protons indicate base-pairing

C   G   U   A
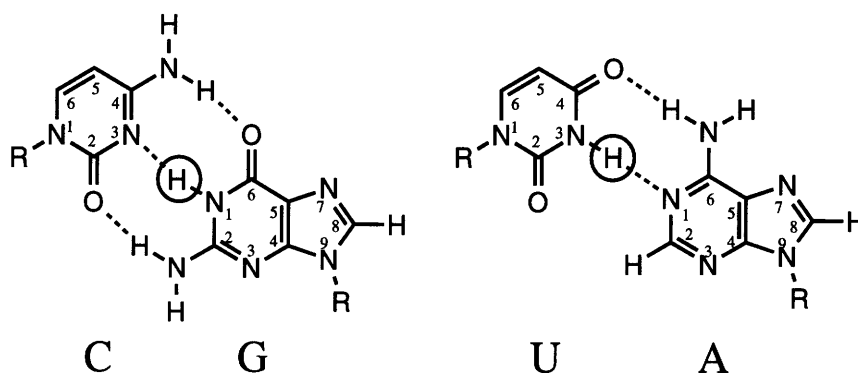
B) Sugar pucker determined from $J_{H1'-H2'}$

C3'-endo
$J_{H1'-H2'}$ ~2Hz

C2'-endo
$J_{H1'-H2'}$ ~8Hz
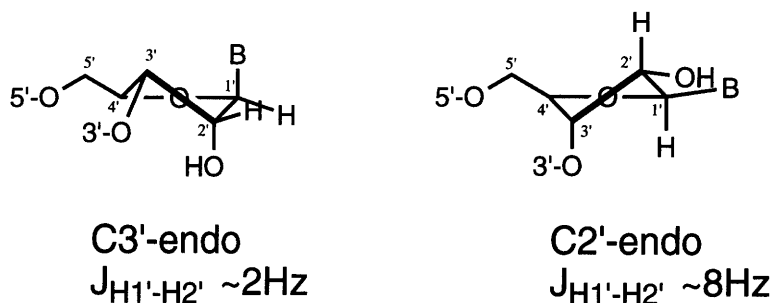
C) Glycosidic conformation from the H1'-H8 distance

anti   *syn*
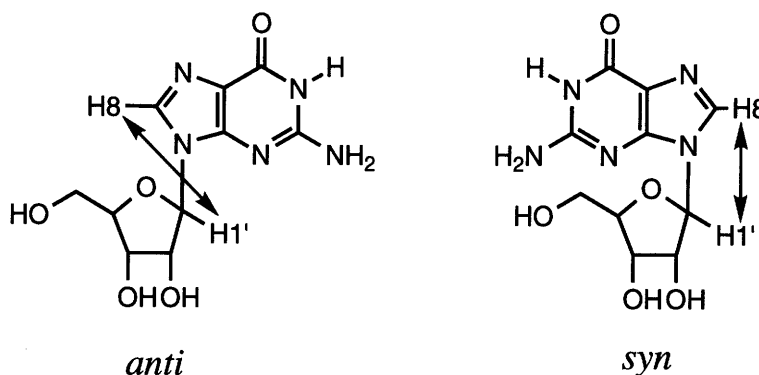
Figure 1.1: Basic RNA structural features that are easily characterized by NMR. A) Imino proton resonances, highlighted by the circles, are only observed when they are in a base pair. B) The two major sugar puckers are determined by coupling constants. C) The two glycosidic conformations can be differentiated by the H8-H1' NOE.

done for the major measurable torsions , but the empirical nature of this equation is another source of error. This error can be minimized by following a semi-quantitative approach where if the J is in a certain regime a broad torsion restraint is applied.

A number of excellent reviews are available describing RNA NMR are available (Pardi, 1995; Varani *et al.*, 1996; Varani & Tinoco, 1991). This discussion will emphasize some of the more modern multidimensional heteronuclear techniques and only briefly review the basic assignment procedure that has not changed significantly in years. After a brief review of homonuclear spectroscopy, a synopsis of some of the recent heteronuclear multidimensional experiments is presented. The central information accessible from RNA NMR is summarized in Figure 1.1.

Exchangeable protons. Exchangeable protons are protons that are bonded to oxygen or nitrogen and are thus labile with water. Typically, the protons attached to oxygen such as 2'OH protons exchange very quickly with water such that most of the time they are part of a water molecule and are not observed at a hydroxyl chemical shift. In a standard A-form helix, the 2'OHs are directed towards the solvent allowing fast exchange with water. Only a few cases of observable 2'OH proton resonances in RNA NMR spectra have been reported and these are generally involved in a specific interaction (Allain & Varani, 1995; Dieckmann et al., 1996; Jiang et al., 1996; Jiang et al., 1997). In these cases, the 2'OH is making specific hydrogen bonds and/or is buried in the structure, protecting the proton from water exchange.

The imino protons (~11-16 ppm) of guanine and uridine nucleotides resonate far downfield from most of the other RNA protons when they are involved in a hydrogen bond (Figure 1.2). These imino protons are generally only observable when they are involved in Watson-Crick or non-Watson-Crick base pairs and therefore are a good indicator of RNA secondary structure. Imino-imino NOEs between stacked base pairs help identify helical regions and make initial assignments. Of the different types of amino protons, the cytidine

²H₂O 1D spectrum

H2'/H3'/H4'/H5'/H5"

H1'/H5

H8/H6/H2

9 8 7 6 5 4 3 2 1 0

H₂O 1D spectrum

aminos

iminos

14 12 10 8 6 4 2 0
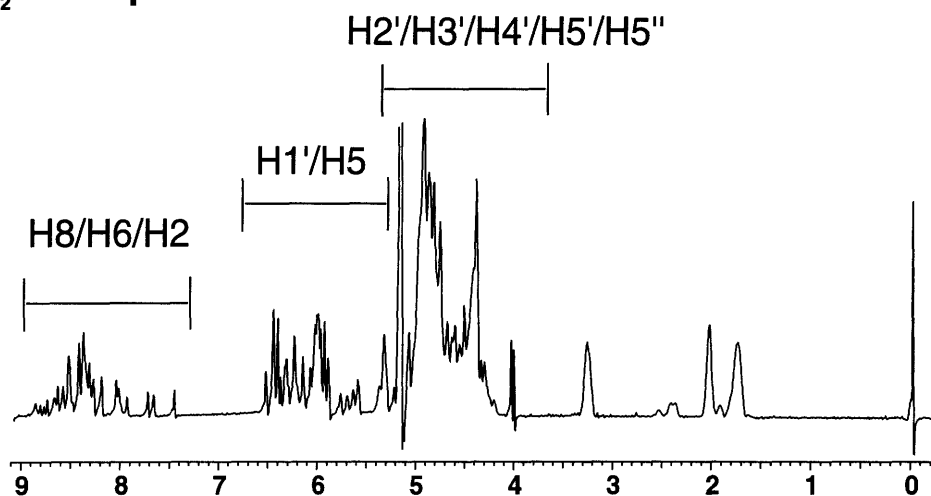
¹H (ppm)
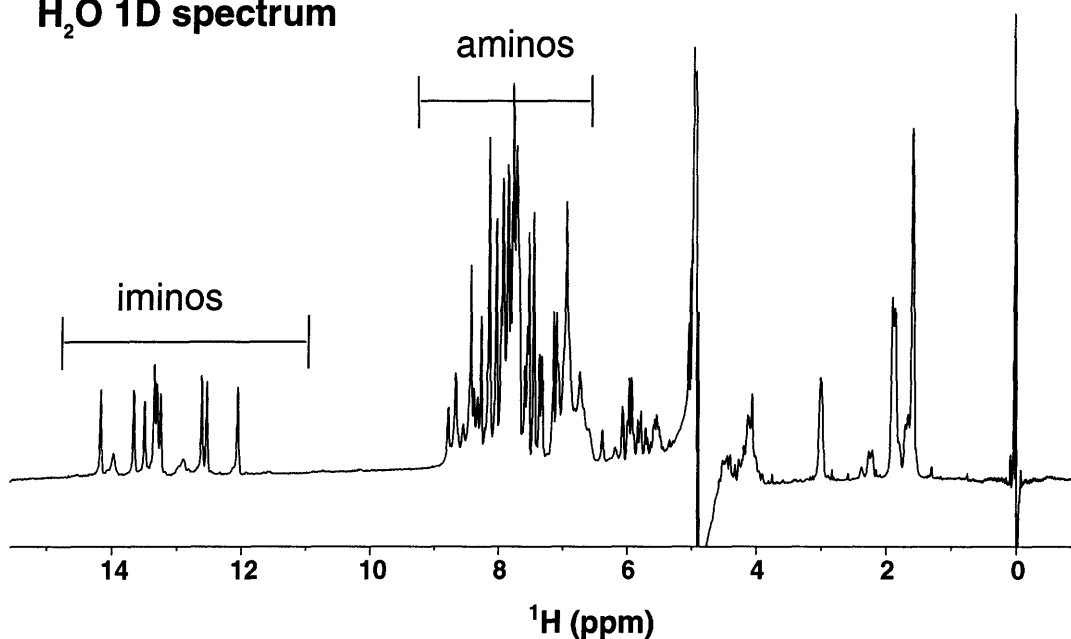
Figure 1.2: 1D spectra in ^2H_2O and H_2O of a TAR mutant show the typical proton chemical shifts found in RNA. Notice how more than half the protons are between ~3.8 - 5 ppm. The exchangeable H_2O spectrum shows how the imino protons are resolved from the rest of the resonances allowing straightforward monitoring of titrations and secondary structure from 1D spectra. Because there are no resonances above 9 ppm in the nonexchangeable spectrum, the sweep widths differ between the two spectra illustrated.

aminos are the most easily observed as two distinct peaks, one for each amino proton. They can be identified in homonuclear spectra by observing strong NOEs to guanine iminos in a G-C Watson-Crick base pair and by strong NOEs to the nearby cytidine H5 from the same nucleotide. Guanine and adenine aminos are notoriously difficult to observe and assign because upon base pairing these resonances often fall are in the intermediate exchange regime where they are unobservable. These protons, especially the guanine aminos, provide cross-strand NOEs that significantly help define an A-form helix (Mueller *et al.*, 1995). Furthermore, these resonances are often involved in interesting non-canonical base pairing, making these NOEs critical to identify.

Sugar pucker and glycosidic torsion One of the hallmarks of A-form helices is the conformation of the sugar pucker. The pucker can be qualitatively measured by measuring the $^3J_{H1'-H2'}$ from the H1'-H2' crosspeak in a ^1H-^1H COSY experiment. The C3'-endo sugar pucker found in A-form helices has a small $^3J_{H1'-H2'}$ of <3 Hz making these crosspeaks unobservable for most RNA molecules. However, the $^3J_{H1'-H2'}$ is >8 Hz for the C2'-endo sugar pucker which is often observed in non-helical RNA structure.

The glycosidic torsion angle (χ) is generally found in one of two different conformations, syn and anti. The χ torsion is measured by monitoring the H8/H6 to H1' distance which is ~2.5 Å the syn conformation and ~3.8 Å for the anti conformation. This difference can be effectively determined in short (<50 ms) mixing time NOESY experiments. Other base to ribose distances are also dependent on the χ torsion, but they generally require multidimensional heteronuclear experiments and/or specifically labeling to assign.

Helical sequential assignments Along with imino-imino NOEs, the NOE anomeric-aromatic sequential assignment pathway helps define the RNA secondary structure and make sequential assignments. Each H8/H6 proton gives a NOE to its own H1' and to the H1' from its 5' neighbor as illustrated in Figure 1.3. These resonances are particularly
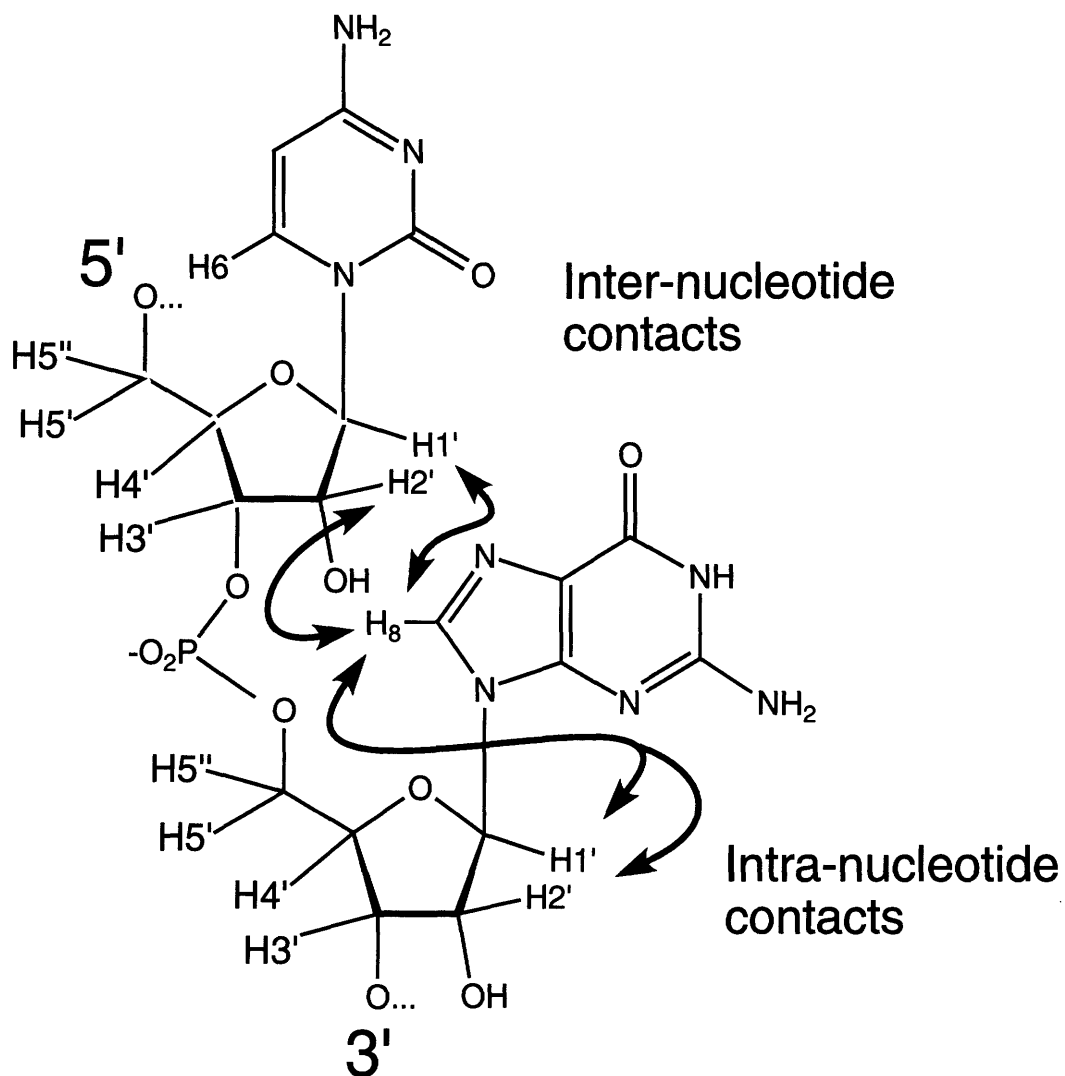
Figure 1.3: The standard sequential assignment pathway for RNA A-form helices between the H8/H6 and H1'/H2' resonances. Note that the internucleotide H8/H6-H2' NOEs are very strong (~2.4 Å) while the other inter-and intranucleotide distances illustrated are weaker (>4.0 Å).

useful because they generally have well-resolved chemical shifts. Although it is critical to try to assign as many of these "sequential walk" NOEs as possible to define the A-form regions, the absence of these NOEs usually suggest that a more interesting non-A-form structure is present. When beginning a RNA NMR project, it is imperative to try to maximize the information from the basic ^{1}H-^{1}H NOESY experiment as this will help identify any unusual connectivities and also make the interpretation of more complicated 3D and 4D experiments easier. Although unambiguous assignments require heteronuclear spectroscopy (see below), as larger RNAs are investigated, more dependence on nonexchangeable homonuclear spectra may be necessary due to broad carbon lines preventing quality heteronuclear experiments.

Spectral crowding In principle, these basic NMR measurements can quickly and easily define an A-form helix as the H1', imino, and aromatic resonances are some of the best resolved resonances as shown in Figure 1.2. However, to obtain high resolution solution structures, the other extremely overlapped proton resonances need to be assigned as well. These protons give a large number of internucleotide NOEs to each other and to aromatic protons that help define RNA structure as summarized in Table 1.1. Most of the internucleotide NOEs are sequential in nature as opposed to the very useful cross-strand NOEs to the adenine H2 and exchangeable protons.

1.2.2 Exchange in NMR spectroscopy

Exchange is a broad term that encompasses many different mechanisms including lability with water, conformational motions, binding kinetics, etc. The net effect of exchange is an effect on the observed linewidths. When the different sites are magnetically distinct, these processes will effect the observed signal. Exchange phenomena are generally characterized into three broad categories as illustrated in Figure 1.4. Slow exchange is frequency difference, Δv, of the two sites is much larger than the rate constant,
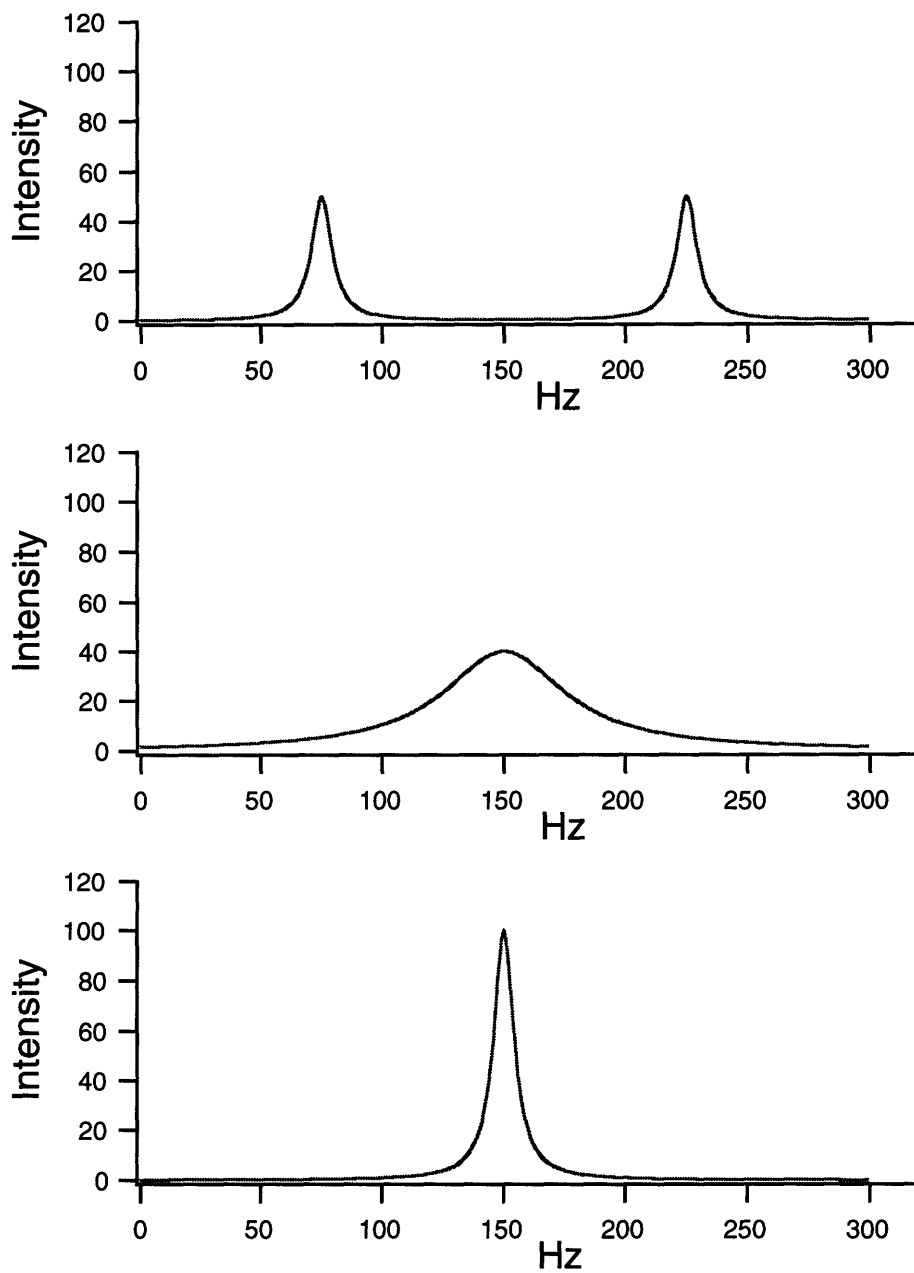
Figure 1.4: Exchange for a two-site system. Two peaks are observed in the slow exchange regime. As the exchange rate between the two sites increases, the two peaks becomes one broad peak until the rate is significantly greater than Δ and the system enters the fast exchange regime.

Table 1.1: Short interproton distances (Å) in an A-form helix.*

Intranucleotide Nonexchangeable

Aromatic-ribose		Ribose-ribose C3'-endo	
H6 ↔ H5	2.4	H1' ↔ H2'	2.4
H6/H8 ↔ H1'	3.5	H2' ↔ H3'	2.4
H6/H8 ↔ H2'	4.0	H3' ↔ H4'	2.9
H6/H8 ↔ H3'	2.7-3.0	H4' ↔ H5'	2.4
H6/H8 ↔ H4'	4.2	H5' ↔ H5"	1.8
H6/H8 ↔ H5'	3.3	H1' ↔ H3'	3.7
H6/H8 ↔ H5"	4.0	H1' ↔ H4'	3.3
		H2' ↔ H4'	3.6
		H3' ↔ H5'	3.0

Internucleotide Nonexchangeable 3' ↔ 5'

Aromatic-ribose		Ribose-ribose C3'-endo	
H6/H8 ↔ H1'	4.5	H2' ↔ H1'	4.0
H6/H8 ↔ H2'	2.0	H2' ↔ H3'	4.1
H6/H8 ↔ H3'	3.0	H4' ↔ H2'	4.0-4.6
H6/H8 ↔ H6/H8	4.5	H5' ↔ H2'	3.0-3.4
H5 ↔ H2'	3.5	H5" ↔ H2'	4.3-4.5
H5 ↔ H3'	3.3	H5' ↔ H3'	3.8-4.2
H5 ↔ H8/H6	3.75		
H5 ↔ H5	3.7		
H1' ↔ H2	4.0		

Internucleotide Exchangeable Stacking 3' ↔ 5'

G(H3) ↔ G(H3)	3.5-3.7	H62 ↔ H1/H3	3.7-4.2
U(H1) ↔ U(H1)	3.5-3.7	H42 ↔ H1/H3	3.8-4.1
U(H1) ↔ G(H3)	3.5-3.7	H22/H21 ↔ H1	4.4-4.6
H5 ↔ H1/H3	4.7-4.9		
H42 ↔ H5	4.1-4.3		

* Distances were measured in A-form regions of a tRNA structure (Westhof *et al.*, 1988) and a duplex (Holbrook *et al.*, 1991).

k, describing the rate of exchange between the two sites. The other extreme is fast exchange, when $k \gg v$, and only a single resonance is observed at the population-weighted average of the frequencies of the two sites. Intermediate exchange describes everything between the two extremes and is a major problem when the lines broaden to the point where they are unobservable. The frequency of the spin in each site can only be observed for a time of approximately $1/k$ before the spin jumps to the other site. In macromolecular complexes, the affinity is often good indicator of the potential to observe sharp lines. Assuming a diffusion limiting on rate, a complex with dissociation constant of one nanomolar will have a very slow off rate of ~ 1 s^{-1} and a long lifetime allowing observation of the bound complex. However, if the affinity is one millimolar, the lifetime of the complex would be approximately one microsecond, which is faster than any typical frequency difference which is approximately 0.1-3 milliseconds on currently available spectrometers.

1.2.3 Isotopic labeling of RNA

In order to perform a wide variety of heteronuclear NMR experiments, isotopic enrichment of RNA samples with NMR active ^{13}C- and ^{15}N-nuclei is required. The standard method for preparation of isotopically labeled RNA is *in vitro* transcription from synthetic DNA oligonucleotide templates using T7 RNA polymerase and isotopically labeled nucleotide triphosphates (NTPs). Labeled NTPs can be isolated from bacteria grown on ^{13}C and/or ^{15}N-enriched media (Batey *et al.*, 1995). The advantage of enzymatic RNA synthesis is that uniform isotopic enrichment is relatively inexpensive and easy to accomplish. However, the disadvantage is the lack of selectivity in labeling. Chemical synthesis of RNA, on the other hand, has more potential for applications where selective labeling is desired. Methods for chemical synthesis of isotopically labeled RNA, however, are not as straightforward and are relatively expensive. Nevertheless, there has

been recent progress in chemical synthesis methods, and they may become more utilized in the near future (Foldesi *et al.*, 1992; Foldesi *et al.*, 1996; Quant *et al.*, 1994; SantaLucia *et al.*, 1995; Tolbert & Williamson, 1996; Tolbert & Williamson, 1997). Recent work has demonstrated the cost effectiveness of using metabolic enzymes to make selectively labeled nucleotides on both the ribose and base from glucose in only a few steps (Tolbert & Williamson, 1997).

1.2.4 Assignment of RNA resonances

Assignment of RNA resonances has historically been achieved through identification of sequential base to ribose NOE patterns seen in helical regions of nucleic acid structure that were originally utilized for DNA studies in the 1980s (Wuthrich, 1986). With the advent of isotopic labeling for RNA, the basic NOE assignment approach was initially expanded to include multi-dimensional (3D and 4D) versions of the standard NOESY, COSY, and TOCSY 2D experiments, which simplified assignment and identification of NOEs (Nikonowicz & Pardi, 1992b; Nikonowicz & Pardi, 1993; Pardi & Nikonowicz, 1992). The NOE-based approach, however, is susceptible to errors from structural bias, and methodology that achieves sequential assignment via through-bond correlation experiments, as is the case for proteins, would be more ideal. Towards this aim, there has been a proliferation of experiments in the past two-three years that correlate RNA proton resonances through the heteronuclear spin-systems of the base and phosphate backbone. The new NMR experiments for resonance assignment can be conceptually organized into three classes (Figure 1.5 and Table 1.2). One HCN class correlates the protons of the bases (both exchangeable and non-exchangeable) to one another via J_{CC} and J_{CN} couplings (Farmer *et al.*, 1994; Farmer *et al.*, 1993; Skélnar *et al.*, 1993a; Skélnar *et al.*, 1993b). Another HCN class correlates the base resonances (H8/H6) to the ribose H1' proton through the intervening carbon (C8/C6/C1') and nitrogen (N9/N1) resonances (Farmer et al., 1994; Farmer et al., 1993;
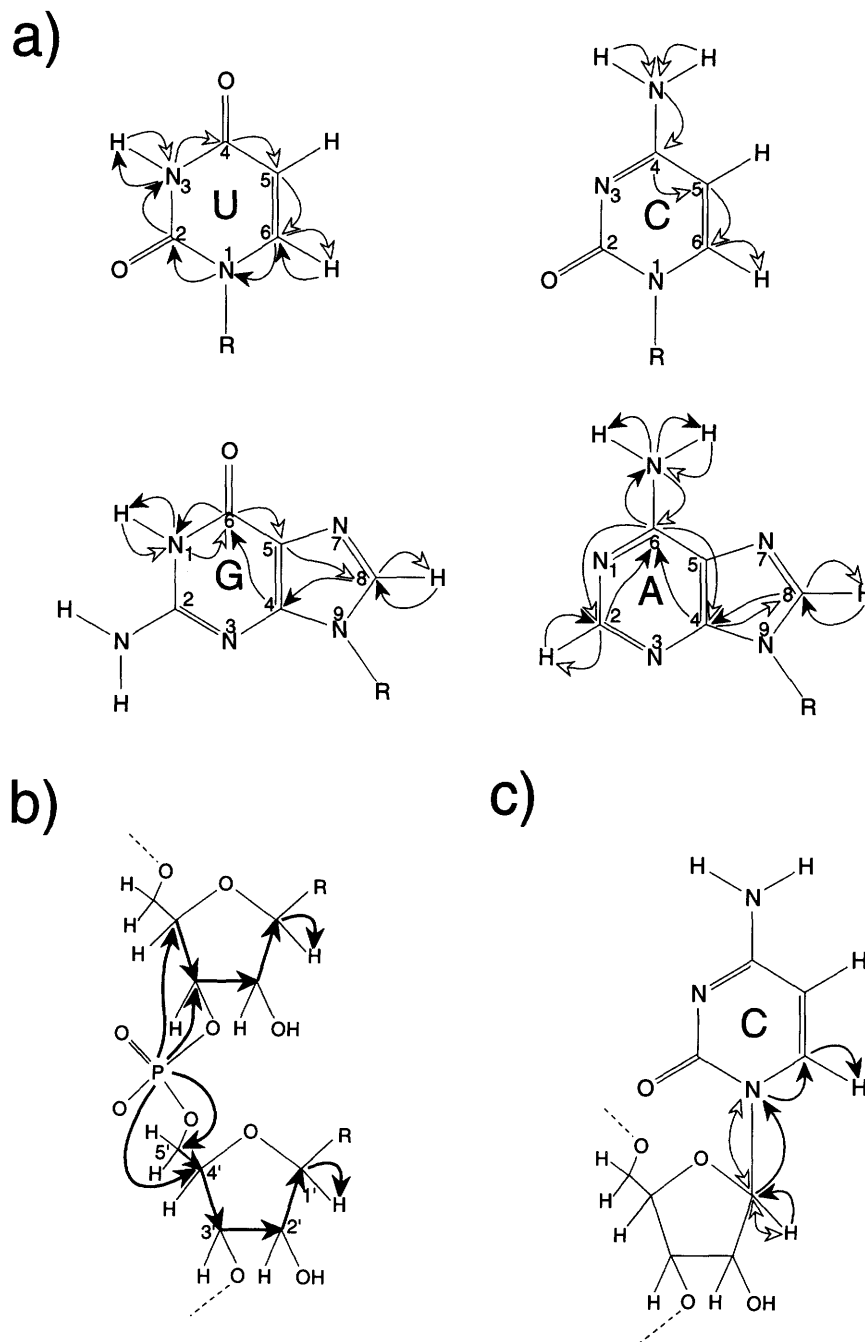
Figure 1.5: Schematic of coherence transfer for through-bond assignment experiments. **(A)** HCN correlation experiments for assignment of the base resonances. Closed and open arrows show HC-TOCSY-CNH and HNC-TOCSY-CH type experiments, respectively. **(B)** HCP experiments for sequential ribose-ribose assignment. **(C)** HCN experiments for intranucleotide correlation

Table 1.2: Through-Bond Correlation Experiments for Assignment of RNA Resonances

Experiment	Nucleotide	Coherence Pathway[a]
Base-Base		
H(NC)-TOCSY-(C)H[c]	guanosine	H1(t1)$\xrightarrow{I}$N1$\xrightarrow{HT}$C6/C2$\xrightarrow{T}$C8$\xrightarrow{I}$H8(t2)
(H)N(C)-TOCSY-(C)H[d]	adenosine	H6$\xrightarrow{HT}$N6(t1)$\xrightarrow{HT}$C6$\xrightarrow{T}$C8/C2$\xrightarrow{I}$H8/H2(t2)
H(NCCC)H[e]	uridine	H3(t1)$\xrightarrow{I}$N3$\xrightarrow{HT}$C4$\xrightarrow{I}$C5$\xrightarrow{I}$C6$\xrightarrow{I}$H6(t2)
	cytosine	H4(t1)$\xrightarrow{HT}$N4$\xrightarrow{HT}$C4$\xrightarrow{I}$C5$\xrightarrow{I}$C6$\xrightarrow{I}$H6(t2)
HC-TOCSY-(CN)H[b,f]	purine	H8/H2(t1)$\xrightarrow{I}$C8/C2(t2)$\xrightarrow{T}$C6$\xrightarrow{I}$N6/N1$\xrightarrow{I}$H6/H1(t3)
H(C)-TOCSY-(CN)H[b,g]	guanosine	H8(t1)$\xrightarrow{I}$C8$\xrightarrow{T}$C6$\xrightarrow{HT}$N1$\xrightarrow{I}$H1(t2)
	uridine	H6(t1)$\xrightarrow{I}$C6$\xrightarrow{HT}$N3$\xrightarrow{I}$H3(t2)
Base-Sugar		
$H_s(C_sN_bC_b)H_b$[b,h]	all	H1'(t1)$\xrightarrow{I}$C1'$\xrightarrow{I}$N9/N1$\xrightarrow{I}$C8/C6$\xrightarrow{I}$H8/H6(t2)
$H_sC_sN_b$[b,i]	all	H1'$\xrightarrow{I}$C1'(t1)$\xrightarrow{I}$N9/N1(t2)$\xrightarrow{I}$C1'$\xrightarrow{I}$H1'(t3)
$H_bC_bN_b$[b,i]	all	H8/H6$\xrightarrow{I}$C8/C6(t1)$\xrightarrow{I}$N9/N1(t2)$\xrightarrow{I}$C8/C6$\xrightarrow{I}$H8/H6(t3)
$H_sC_s(N_bC_b)H_b$[j]	all	H1'(t1)$\xrightarrow{I}$C1'(t2)$\xrightarrow{I}$N9/N1$\xrightarrow{I}$C8/C6$\xrightarrow{I}$H8/H6(t3)
$H_sC_sN_b$[k]	purine	H1'$\xrightarrow{I}$C1'(t1)$\xrightarrow{I}$N9(t2)$\xrightarrow{I}$C1'$\xrightarrow{I}$H1'(t3)
$H_sC_s(N_b)C_b$[k]	purine	H1'$\xrightarrow{I}$C1'$\xrightarrow{I}$N9$\xrightarrow{I}$C8(t1)$\xrightarrow{I}$N9$\xrightarrow{I}$C1'(t2)$\xrightarrow{I}$H1'(t3)
$H_bN_bC_b$[k]	purine	H8$\xrightarrow{I}$N9(t1)$\xrightarrow{I}$C8(t2)$\xrightarrow{I}$H8(t3)
Sugar-Sugar		
HCP[l]	all	H4'$\xrightarrow{I}$C4'$\xrightarrow{I}$P(t1)$\xrightarrow{I}$C4(i+1)(t2)$\xrightarrow{I}$H4(i+1)'(t3)
HCP-CCH-TOCSY[m]	all	H4'$\xrightarrow{I}$C4'$\xrightarrow{I}$P(t1)$\xrightarrow{I}$C4(i+1)$\xrightarrow{T}$C1'(i+1)(t2)$\xrightarrow{I}$H1'(i+1)'(t3)
P(CC)H-TOCSY[n]	all	P(t1)$\xrightarrow{I}$C4(i,i+1)$\xrightarrow{T}$C1'(i,i+1)$\xrightarrow{HT}$H1'(i,i+1)'(t2)

[a]The letter above the arrow indicates the type of coherence transfer method used: I - INEPT, T-TOCSY, HT-Hetero-TOCSY. In () are the dimensions for which chemical shift is detected. [b]Experiment titles have been slightly altered in order to be presented in a uniform fashion. () indicate a nucleus that coherence is transfered through, but the chemical shift is not detected. [c] [Simorre, 1996 #79]; [d] [Simorre, 1996 #80]; [e] [Simorre, 1995 #78]; [f] [Fiala, 1996 #21]; [g][Skélnar, 1996 #325]; [h] [Skélnar, 1993 #84]; [i] [Skélnar, 1993 #83; Marino, 1997 #248]; [j] [Farmer, 1993 #19]; [k] [Farmer, 1994 #20]; [l] [Heus, 1994 #114; Marino, 1994 #117]; [m] [Marino, 1995 #50]; [n][Wijmenga, 1995 #90].

Skélnar et al., 1993a; Skélnar et al., 1993b). Finally, an HCP class of experiments

correlates the sugar resonances of sequential nucleotides through the phosphate backbone

(Heus *et al.*, 1994; Marino *et al.*, 1994; Marino *et al.*, 1995; Varani *et al.*, 1995; Wijmenga

et al., 1995).

Base HCN correlation experiments

A recent set of HNCCH- and HCCNH-TOCSY experiments have been developed

that correlate the exchangeable imino and amino proton resonances with the non-

exchangeable base resonances for the complicated spin systems of all four nucleotides

(Fiala *et al.*, 1996; Simorre *et al.*, 1996a; Simorre *et al.*, 1995; Simorre *et al.*, 1996b;

Skélnar *et al.*, 1987). These experiments differ in the exact mechanism of transfer, but

predominantly rely on one-bond heteronuclear coupling constants to transfer magnetization

from protons to the heteronuclei of the ring, and homonuclear ^{13}C-TOCSYs to transfer

magnetization across the ring. There are two important issues which impinge upon the

sensitivity of these experiments. One is the chemical shift differences between the

heteronuclei in the ring. Since the C5 atom of both purines and pyrimidines resonate

around 20-60 ppm upfield from the other carbon resonances in the base, ^{13}C-DIPSI spin

locks cannot effectively excite all resonances without excessive sample heating. Therefore,

magnetization is usually transferred via smaller two-bond carbon-carbon ($^2J_{CC}$) coupling

constants (C8-C4-C6 for purines and C6-C4 for pyrimidines), which require long spin-

lock mixing times for transfer and loss of sensitivity due to transverse relaxation. There

have been two approaches to circumvent this problem. For pyrimidines, sequential

homonuclear INEPT transfers (C4-C5-C6) can be used rather than a ^{13}C-TOCSY

(H(NCCC)H) (Simorre et al., 1995), since short, high power π pulses excite a wider

bandwidth. For purines, the C5 resonance is only 20-40 ppm away from the other

aromatic carbon resonances (Table 1.3) and a FLOPSY-8 spin locking scheme (Mohebbi &

Shaka, 1991) can be used to increase the efficiency of one-bond carbon-carbon transfer and

Table 1.3: Chemical Shifts and Coupling Constants in Nucleotides

Atom(s)	Chemical Shift Range
Carbon	
Purine C5	~120 ppm
Purine C8	~140 ppm
Purine C2,C4,C6	~150-160 ppm
Pyrimidine C5	~100 ppm
Pyrimidine C6	~145 ppm
Pyrimidine C2,C4	~155-170 ppm
Ribose carbons	~70-90 ppm
Nitrogen	
Purine N9	~170 ppm
Purine N3,N7	~220-240
Guanosine N1	~145 ppm
Guanosine N2	~70-80 ppm
Adenosine N1	~215 ppm
Adenosine N6	~80-90 ppm
Pyrimidine N1	~145 ppm
Uridine N3	~155 ppm
Cytosine N4	~95 ppm

Atom(s)	Coupling Constant
Carbon-Carbon	
$^1J(C,C)$	~40-60 Hz
$^2J(C8,C4)$ (purine)	~8-11 Hz
$^2J(C6,C4)$ (pyrimidine)	~8-11 Hz? double check
Carbon-Nitrogen	
$^1J(C1',N9/N1)$	~11-12 Hz
$^1J(C8,N9)$	~4-8 Hz
$^1J(C6, N1)$	~13 Hz
$^1J(C8,N7)$	~10 Hz
Carbon-Phosphorous (include?)	
$^2J(C3'/C5',P)$	~3-5 Hz
$^3J(C4',P)$	~8-10 Hz

decrease mixing times (Simorre et al., 1996a; Simorre et al., 1996b). Another factor which

decreases the sensitivity of some base resonances is rotational exchange broadening of exo-

cyclic amino protons on adenosine, guanosine, or cytosine. It has been shown from

studies on NOE transfer from amino protons that cross-polarization (hetero-TOCSY) or

CPMG-INEPT transfer is more efficient than the standard INEPT procedure when

resonances are broadened by rotational exchange (Mueller et al., 1995). In addition, the

nitrogen linewidths are narrower than protons undergoing chemical exchange, and

detection of nitrogen chemical shift rather than proton is more efficient. Simorre et. al have

utilized nitrogen chemical shift detection and cross-polarization transfer from proton to

nitrogen to optimize sensitivity for the amino protons of adenosine and cytosine (Simorre et

al., 1996a; Simorre et al., 1995). The extra sensitivity gained by optimization is offset by

increased experiment time for acquiring a separate data set for each nucleotide.

Nevertheless, the added sensitivity will probably be essential for assigning the resonances

of many larger RNAs.

Base to ribose correlation experiments

An important application of $^{13}C/^{15}N$-labeling of RNA is through-bond

intranucleotide correlation of base to ribose resonances, which historically has been

achieved through NOE transfer. There are now many published experiments for achieving

this through bond transfer (Farmer et al., 1994; Farmer et al., 1993; Skélnar et al., 1993a;

Skélnar et al., 1993b) (reviewed in (Dieckmann & Feigon, 1994; Pardi, 1995)). Some of

these experiments transfer magnetization sequentially from ribose to base protons (Farmer

et al., 1993; Skélnar et al., 1993b), while others use an "out-and-back" approach (Farmer

et al., 1994; Skélnar et al., 1993a). A simple sequential INEPT transfer procedure with

non-selective pulses can be used to provide sugar to base correlations. Branching

coherence transfer pathways, however, reduce the sensitivity of the basic approach.

Therefore, selective pulses have been cleverly utilized in the above sequences to direct

coherence predominantly along the desired pathways. For example, during transfer from N9/N1 to C8/C6, selective excitation of the C8/C6 resonances refocuses magnetization proceeding from N9/N1 to C4/C2 resonances in purines and pyrimidines, respectively. A recent demonstration of multiple-quantum transfer to correlate the C1' and N9/N1 resonances increases the sensitivity of these type of experiments (Marino *et al.*, 1997). Furthermore, the multiple quantum transfer is predicted to improve as the RNA increases in size. In principle, any one of the above experiments could provide complete assignment of all nucleotides (though separate optimized sequences for purines are often required because of sensitivity problems (Farmer et al., 1994)). In practice, however, spectral overlap will probably prevent complete assignment from only one experiment. Acquiring several experiments detecting different heteronuclei (C8/C6, N9/N1, or C1') can potentially resolve overlap problems for larger RNA molecules. The experiments which will provide the best resolution may vary for different RNA molecules and can be best determined empirically from 2D ^{1}H-^{13}C or ^{1}H-^{15}N HSQCs of the base and C1'/H1' resonances.

ribose to ribose HCP experiments

Sugar proton resonances of ^{13}C-labeled RNA can be correlated via HCCH-type (Kay *et al.*, 1990) experiments that utilize the large one-bond carbon-carbon coupling constants to transfer magnetization throughout the ribose ring (Nikonowicz & Pardi, 1993; Pardi, 1995; Pardi & Nikonowicz, 1992). In order to link the resonances of each ribose nucleotide sequentially without the use of NOEs, magnetization must be transferred through the phosphate backbone. A straightforward approach is HCP correlation through an "out-and-back" procedure via sequential INEPT transfers (^{1}H -> ^{13}C -> ^{31}P -> ^{13}C -> ^{1}H) (Heus et al., 1994; Marino et al., 1994). This approach has two limitations for complete assignment of RNA molecules over ~20 nucleotides. One is that the C4'(i)-P(i)-C4'(i+1) connectivity is severely overlapped due to poor spectral dispersion of C4' and P resonances, particularly in A-form helical regions. Second, the better resolved C3'(i)-P(i)-

C5'(i+1) connectivity requires a separate HCCH-TOCSY experiment to connect the C3'

and C5' resonances on the same nucleotide. Subsequent experiments (HCP-CCH-TOCSY

(Marino et al., 1995) and P(CC)H-TOCSY (Wijmenga et al., 1995)) resolve both problems

by combining the HCP and HCCH-TOCSY experiments. Therefore, sequential

correlations can be observed between C1'/H1' resonances, which are the best resolved

ribose resonances in both carbon and proton. The drawback to this experiment is long

delays and TOCSY mixing times where significant relaxation can occur for larger

molecules. It is not clear whether these experiments will be generally useful for RNA

molecules larger than ~30 nucleotides (~10 kDa), particularly with the severe line

broadening of ^{31}P resonances at higher molecular weights.

Unlike triple resonance experiments used for proteins, most of the new triple

resonance assignment experiments for RNA have been utilized only in "proof of principle"

situations and have yet to be performed in a published structure determination. The only

exceptions are the use of HCCNH-TOCSY experiments for assignment of base resonances

in a couple of cases where a complete structure determination was achieved (Dieckmann et

al., 1996; Fan *et al.*, 1996b; Jiang et al., 1996; Zimmerman et al., 1997). The experiments

for correlation of the base resonances are the most likely to succeed with medium-size RNA

molecules (30-40 nucleotides, ~10-15 kDa) due to less spectral overlap and relatively long

relaxation times of the base resonances. Nevertheless, the standard NOE method for

sequential assignment was used in these structures, and it is not yet clear how generally

useful the through-bond experiments will be for complete non-NOE based sequential

assignment. Complete through-bond assignment will probably not be feasible for many

medium sized RNAs, since the HCP experiments are most susceptible to relaxation

problems. A hybrid approach with HCN and NOESY experiments may be a temporary

substitute. The HCN experiments can unambiguously determine the intranucleotide

correlations within and between the base and ribose resonances, which will significantly reduce the ambiguity present in the complete NOESY-based assignment procedure.

Even assuming no problems due to relaxation, it does not appear that there will be a simple and quick procedure for assignment of RNA molecules. Neglecting the problems with sensitivity or overlap, complete assignment with only through bond methods still requires a large number of experiments, if all of the optimized sequences are required (~4 experiments for the bases, ~3 experiments to correlate the base resonances to the ribose, and ~2-3 experiments to correlate the ribose resonances). Nevertheless, this is around the same number of experiments that have been required for many RNA structures solved recently by the NOE-based method, since multiple samples with different specific labeling patterns (see below) were required to resolve ambiguities in the sequential NOEs (Battiste et al., 1995; Cai & Tinoco, 1996; Dieckmann et al., 1996). Despite roughly the same number of experiments, analysis of the spectra for the through-bond experiments should hopefully be quicker in addition to lowering the potential for misassignment with the NOE-method.

1.2.4 NOE measurement

The end goal of NMR analysis is usually a structure determination. Despite the method used for assignment of resonances, the main source of structural data is obtained from NOEs, which provide distance restraints for pairs of hydrogen atoms in the RNA molecule. Identification of NOEs can be aided by resolving the standard NOESY experiment into 3 and 4 dimensions through detection of the heteronuclear ($^{13}C/^{15}N$) chemical shifts of the proton-attached nuclei (Nikonowicz & Pardi, 1992a; Nikonowicz & Pardi, 1992b; Nikonowicz & Pardi, 1993). Even with the extra dimensions in 3D NOESY-HSQC or 4D HMQC-NOESY-HSQC experiments, overlap can still be severe for medium size RNAs (~30-40 nucleotides). An approach to aid in the identification of NOEs that has been very successful is specific isotopic labeling of RNA with either heteronuclei

(^{13}C/^{15}N) or deuterium (^{2}H). ^{13}C/^{15}N-labeled nucleotides can be separated and used individually in vitro transcriptions with other unlabeled nucleotides (e.g. ^{13}C-GTP plus unlabeled ATP, CTP, and UTP) to produce RNA molecules selectively labeled at only one nucleotide (Batey et al., 1995; Batey *et al.*, 1992; Nikonowicz *et al.*, 1992). There are two advantages to selective labeling for identification of NOEs. One is simply reduced overlap in 3D NOESY experiments, which permits identification of more NOEs. Secondly, isotopic filtering experiments (Otting & Wuthrich, 1989a; Otting & Wuthrich, 1990) can be performed to help resolve intra-nucleotide from inter-nucleotide NOEs (Battiste et al., 1995). This is a particularly powerful approach that can also be utilized for sequential assignment of RNA where the through-bond methods described above are not successful. A similar approach has also been used with RNAs containing site-specific labels produced synthetically at the C8/C6 positions of the base (Cai & Tinoco, 1996; SantaLucia et al., 1995). In addition to NOE identification, this labeling pattern has application for measurement of ^{13}C dynamics, since the removal of ^{13}C-^{13}C interactions simplifies interpretation. One additional labeling methodology that has been developed is specific labeling of a short segment of nucleotides within a larger RNA sequence (Xu *et al.*, 1996). Depending on the RNA being studied and the structural questions addressed, each of these specific heteronuclear labeling methods will greatly assist in the assignment and structure determination of medium sized RNA molecules.

Another way to remove or filter proton resonances is to specifically substitute them with deuterium. An advantage not present with specific ^{13}C/^{15}N-labeling is decreased relaxation of protons not labeled with deuterium. In addition, any ^{13}C atoms attached to deuterium will have decreased relaxation, and ^{13}C-TOCSY transfer will be more efficient. Therefore, specific deuterium labeling should greatly aid in the study of larger RNA molecules. Two protocols have been published for specific deuterium labeling. One uses a combination chemical/enzymatic synthesis approach (Tolbert & Williamson, 1996), while

the other uses all chemical synthesis (Foldesi et al., 1992; Foldesi et al., 1996; Glemarec *et al.*, 1996). In addition to the reduction in overlap, the relaxation times of the non-deuterated protons were increased about two-fold. The disadvantage to deuterium labeling is the complete loss of structural information at the deuterated positions. It is possible, however, to make multiple samples with different deuteration patterns to circumvent this problem (Tolbert & Williamson, 1996). Alternatively, since many of the NOEs between the base and sugar protons that define the conformation of the helices are redundant, the reduced NOE data set from one specific deuteration pattern may still be useful for determining qualitative structural models of larger RNAs.

An experimental approach to improving the sensitivity of NOEs to amino protons in RNA has been developed (Mueller et al., 1995). Amino protons are often significantly broadened due to rotational exchange about the C-N bond. It has been shown that either hetero-TOCSY or CPMG-INEPT transfer is more efficient than INEPT transfer when there is conformational exchange broadening (Krishnan & Rance, 1995; Mueller et al., 1995). As was noted above, this concept has also been utilized to improve the sensitivity of through-bond experiments for assignment of base amino protons (Simorre et al., 1995; Simorre et al., 1996b). Easier assignment and identification of amino proton NOEs, which are often difficult to observe, has the potential to greatly increase the quality of structure determinations. This is particularly true for RNA-protein interactions, since amino protons line the grooves where proteins often make specific contacts. RNA structure tends to be very dynamic and flexible, and it is very common to observe NMR resonances that are broadened by conformational exchange. Hetero-TOCSY transfer, therefore, might be generally applicable in 3D NOESY-HSQC experiments to improve the sensitivity of any conformational exchange broadened resonances.

1.2.5 Measurement of backbone torsions

Many of the interesting RNA structures studied exhibit a wide variety of backbone
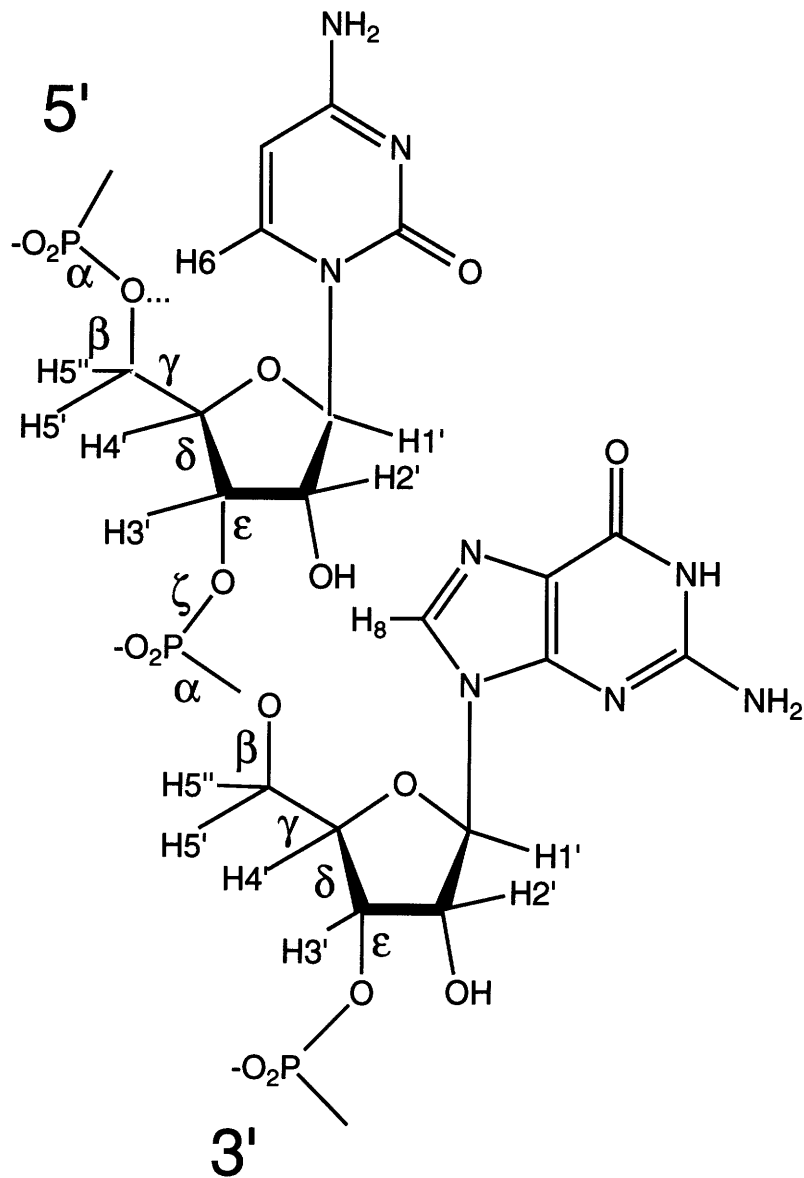
Figure 1.6: The torsion angles which define RNA structures. Notice that the α and β torsions are not near any protons and are thus not directly measureable.

conformations suggesting that torsion restraints will be important to define RNA structure. RNA conformations are characterized by six backbone torsion angles as illustrated in Figure 1.6. NMR analysis of the backbone conformation is complicated by the lack of useful ^{1}H -^{1}H NOE distance restraints available that define the backbone torsions. With the introduction of isotopic labeling methods, a number of approaches have been developed that take advantage of both ^{1}H-^{31}P and ^{13}C-^{31}P couplings to measure these backbone torsions.

sugar pucker and the δ torsion

A number of ^{1}H-^{1}H and ^{1}H-^{13}C scalar couplings are available to measure the sugar pucker but the most useful is the H1'-H2' coupling constant. This coupling is >8 Hz for C2'-endo puckers and ~1 Hz for the C3'-endo puckers, typically found in A-form helices. The most direct method to measure the H1'-H2' coupling constant is through the use of 2D ^{1}H-^{1}H COSY experiments (Varani & Tinoco, 1991). The absence of a H1'-H2' COSY cross-peak is often interpreted as the C3'-endo conformation, however, this approach may be misleading where broad lines lead to the cancellation of the antiphase crosspeaks. This pitfall can be avoided with the use of the HCCH E. COSY experiment (Schwalbe *et al.*, 1995), where even small couplings can be measured directly. This experiment will also be adversely affected by broad lines but because the measurement is not based on the absence of a peak a more definitive measurement is possible. In the presence of severe spectral crowding, the HCC- E. COSY CCH TOCSY experiment, a variant which contains a TOCSY period to transfer magnetization to the well-resolved C1', improves the spectral dispersion (Schwalbe et al., 1995). An alternative method involves a collection of undecoupled HMQC-NOESY and HMQC-TOCSY spectra that permit are collected such that the sign of a number of $^2J_{CH}$ and $^3J_{CH}$ couplings to be determined and used to define the sugar pucker as shown in Table 1.4 (Hines *et al.*, 1994). Often riboses

Table 1.4: Backbone J-Couplings in RNA

Angle	Rotamer	Active bond	J (Hz)*
β	trans	H5' - P(i)	1-4
		H5" - P(i)	1-3
		H4' - P(i)	8-10
		C4' - P(i)	8-10
	gauche⁻	H5' - P(i)	1-3
		H5" - P(i)	< 5
		C4' - P(i)	< 5
	gauche⁺	H5' - P(i)	> 5
		H5" - P(i)	1-3
		C4' - P(i)	< 5
ϵ	trans	C3'/H3' - P(i+1)	8-10
		C2' - P(i+1)	<5
		C4' - P(i+1)	>5
	gauche⁻	C3'/H3' - P(i+1)	8-10
		C2' - P(i+1)	>5
		C4' - P(i+1)	<5
	gauche⁺	C3'/H3'-P(i+1)	8-10
		C2' - P(i+1)	<5
		C4' - P(i+1)	<5
γ	gauche⁺	H4' - H5'	1-3
		H4' - H5"	3
		C4' - H5'	-
		C4 - H5"	+
		C5' - H4'	+
	trans	H4' - H5'	3
		H4' - H5"	9-11
		C4' - H5'	+
		C4 - H5"	-
		C5' - H4'	-
	gauche⁻	H4' - H5'	9-11
		H4' - H5"	3
		C4' - H5'	-
		C4 - H5"	-
		C5' - H4'	-
δ	C3' endo	H1' - H2'	> 8
		H3' - H4'	< 5
	C2' endo	H1' - H2'	~ 1
		H3' - H4'	> 5

* A + or - indicates that the coupling is positive or negative.

are found with coupling constants in the 3-6 Hz range indicative of a conformation in between the C2'-endo and C3'-endo puckers. This mixed conformation is often left unrestrained as it is averaging between the two major conformations (Aboul-ela *et al.*, 1995; Brodsky & Williamson, 1997).

ε *and* β *torsions*

In a standard A-form helix, β is in the *trans* conformation for which the $^3J_{PH5'}$ and $^3J_{PH5''}$ are small and the $^3J_{PC4'}$ and $^3J_{PH4'}$ are large as summarized in Table 1.4. Therefore, when one of the $^3J_{H5'}$ couplings is large, an unusual conformation is present. This analysis can also easily indicate the presence of conformational averaging as found in the HIV-2 TAR-argininamide complex where one of the nucleotides exhibited large couplings for both the $^3J_{PH5'}$ and $^3J_{PH5''}$ [Brodsky, 1997 #10; this thesis]. A similar analysis exists for the ε torsion which can be measured by monitoring the $^3J_{C4'P}$, $^3J_{C2'P}$, and $^3J_{H3'P}$ couplings. However, the $^3J_{H3'P}$ coupling does not change significantly for the different conformations and the $^3J_{C4'P}$ and $^3J_{C2'P}$ need to be analyzed. The ε torsion is in the *trans* conformation in a standard A-form helix where the $^3J_{C2'P}$ coupling is small and the $^3J_{C4'P}$ coupling is large as shown in Table 1.4.

The ε and β torsions can be determined by measuring a variety of $^{13}C-^{31}P$ and $^1H-^{31}P$ scalar couplings. Some of these torsions may be measured directly in 2D HP-HETCOR (Skélnar *et al.*, 1986) and HSQC if the phosphorus and proton resonances are sufficiently resolved. Accurate measurements for H5'/H5''-P and H3'-P couplings can be obtained from spin echo difference experiments as well (Legault *et al.*, 1995). However, both the proton and phosphorus resonances are generally overlapped for even moderate size RNAs. A more extensive strategy has been introduced based on the intensity of cross-peaks in 3D HCP and HPCH experiments (Varani et al., 1995). The intensity of crosspeaks in these spectra are proportional to the $^3J_{PH5'}$, $^3J_{PH5''}$, and $^3J_{PC4'}$ couplings. This

approach suffers from the potential drawback that peak intensities may also be affected by other factors, such as dynamics. However, at this time, this scheme is the most comprehensive method to gather a large number of loose restraints in crowded spectra.

γ torsion

Measurement of the γ torsion is difficult due to the need for stereospecific assignments of the H5' and H5". The C4'-H5'/H5" couplings can be used in conjunction with the H4'-H5'/H5" couplings to define γ (Hines et al., 1994; Schwalbe *et al.*, 1994). Some couplings may be obtained from a collection of experiments undecoupled 3D ^{13}C edited TOCSY and NOESY spectra (Hines et al., 1994; Hines *et al.*, 1993). Unfortunately, both the C4' and the H4' in A-helical RNA's are extremely crowded making correlations from these to the crowded C5'/H5' region difficult. Therefore, experiments similar to the HCC- E. COSY CCH TOCSY, where the overlapped H4'-H5' crosspeaks to the C1'/H1' are correlated to the C1'/H1', may be particularly useful if sufficient TOCSY transfer can be achieved (Schwalbe et al., 1994). In A-form helices, both the $^3J_{H4'H5'}$ and $^3J_{H4'H5''}$ scalar couplings are small in the gauche$^+$ conformation. Table 1.4 indicates that when γ changes to an unusual conformation either the $^3J_{H4'H5'}$ or $^3J_{H4'H5''}$ coupling becomes larger. However, additional information is required to determine the stereospecific assignment. The sign of the $^2J_{C4'H5'}$, $^2J_{C4'H5''}$, and $^2J_{C5'H4'}$ couplings are very useful to help determine not only the stereo specific assignments but also the conformation as summarized in Table 1.4.

α and ζ torsions

Unfortunately, the α and ζ torsions are not accessible by direct J-coupling measurements. Some groups have used ^{31}P chemical shifts as a guide for loose constraints

on these torsions, however, the correlation between ^{31}P chemical shifts and the phosphodiester backbone conformation is not well understood in RNA. For example, in tRNA a wide range of phosphorus chemical shifts are observed that are outside the ranges expected for variation of the α or ζ torsions, and other factors may be major contributors to phosphorus chemical shifts (Gorenstein & Luxon, 1979; Gueron & Shulman, 1975; Salemink *et al.*, 1979). Furthermore, quantum calculations have shown that counterions also effect the chemical shift, further indicating that ^{31}P chemical shifts may not be a reliable indicator of the conformation (Giessner-Prettre & Pullman, 1987). Until a more complete chemical shift data base is created, restraints based on the phosphorus chemical shift alone should probably be used with caution.

1.3 RNA Structures by NMR

To generate a family of structures consistent with the NMR data, structures are refined against the distance and torsion restraints along with geometric and non-bonded terms using restrained molecular dynamics calculations (Brunger & Karplus, 1991; Nilges, 1996). A number of RNA structures have been recently determined utilizing many of the heteronuclear techniques discussed in this review. Many of the problems in RNA NMR studies revolve around the difficulty of gathering a large data restraint set due to chemical shift overlap along with dynamics issues. These issues also need to be considered during the molecular modeling process and for interpreting the resulting family of structures. Furthermore, the paucity of protons along the backbone leave few distance restraints to define the many degrees of freedom of the RNA backbone.

1.3.1 Overall quality of RNA NMR structures

In evaluating the quality of a family of RNA NMR structures, a number of statistics can be evaluated: rmsd, number of NOE and torsion restraints, residual distance and torsion violations, and the largest distance and torsion violations. The distance constraints are further dissected into the number of interresidue, intraresidue, and intermolecular

NOEs. A listing of the recent RNA NMR structures is shown in Table 1.5 with the rmsd's shown in the context of the number of NOE and torsion restraints. The local rmsd is given because the overall global rmsd is usually in the 2.0 - 2.6 Å range indicative of poor convergence. This is because almost every RNA structure studied includes a region which is poorly defined like a disordered loop, terminal base pair, or a nucleotide without any internucleotide NOEs. This situation is comparable to protein NMR and X-ray studies which often neglect the N and C terminal ends of proteins because of the lack of structural data from these regions. Therefore, the more useful rmsd to consider includes only the region of interest and is usually a more accurate description of the quality of the structure than the overall global rmsd. However, for RNA NMR studies, the region that is defined as the structured "core" can be a subjective decision as the tendency is to find a combination of nucleotides which will give the lowest possible rmsd. One alternative is to define the well-ordered region by using an average standard deviation matrix which will calculate the ordered and disordered regions (Kundrot, 1996).

All the structures listed in Table 1.5 have low rmsd's indicating that a family of structures converged to a very similar structure, but the crucial question is whether the structures are accurate. As in all structural biology techniques, correlations with other biochemical and structural data are used to help determine the accuracy of the structure. For RNA NMR studies, it is possible to make mutations of the RNA to determine whether any specific nucleotide alters the structure or complex as has been done in a few cases (Dieckmann et al., 1996; Puglisi et al., 1995; Ye et al., 1995) to help support the proposed interactions. In addition, the GNRA tetraloops were first determined by NMR (Heus & Pardi, 1991) and have now been seen in both the Hammerhead crystal and P4/P5/P6 IVS crystal structures (Cate et al., 1996; Pley et al., 1994a; Pley *et al.*, 1994b). The GNRA tetraloops determined by NMR and X-ray crystallography are extremely similar and

Table 1.5: Recent Structures and Structural Statisics

Structure	Number of Structures	local rmsd (Å)[a]	Total NOEs per nucleotide	Inter-nucleotide NOEs per nucleotide	Torsion restraints per nucleotide	Total Peptide NOEs per amino acid	Inter-molecular NOEs	Initial Structures	Reference
SMALL RNAS									
GNRA tetraloops									
GAAA	10	0.96 P	7.1	4.4	1.6			R.T.	(Jucker et al., 1996)
GAGA	10	0.89 P	7.3	5.5	1.3			R.T.	(Jucker et al., 1996)
GAAA	10	0.57 P	7.2	4.6	2.8			R.T.	(Jucker et al., 1996)
CUUG tetraloop	13	1.3 A	6.8	4.2				D.G.	(Jucker & Pardi, 1995)
P1 Helix	20	0.59 A	24	12.5	4.6			R. T.	(Allain & Varani, 1995)
P loop	33	0.68 Å						R.C.	(Puglisi et al., 1997)
SMALL LIGAND COMPLEXES									
AMP aptamer	8	1.48 P	11	5.9	0		45	D. G.	(Jiang et al., 1996)
	10	1.52 P	15.6[b]	11.56[b]	2.1[b]		40	D. G.	(Dieckmann et al., 1996)
FMN aptamer	5/4	1.24/ 1.31 P	12.8	5	0		18	D. G.	(Fan et al., 1996)
HIV TAR - Arg	20	1.22 A	26.9	8	4.8		8	R.T.	(Aboul-ela et al., 1995)
	20	0.84 A	22	9.7	1.9		19	R.C.	(Brodsky & Williamson, 1997)
Paromomycin -16S RNA	20	0.61 A	10.7	7.3	5.4		47	R.C.	(Fourmy et al., 1996)
Theophyllin- aptamer	10	1.32 P	7.56	6.65	2.82		41	R.T.	(Zimmerman et al., 1997)
Tobramycin aptamer	7	1.13 P	15.41	9.89			45	A-form	(Jiang et al., 1997)
PEPTIDE/PROTEIN COMPLEXES									
BIV Tat-TAR	20	1.49 A	10.8	6.9	2.43	15.2	26	R.C.	(Puglisi et al., 1995)
	4	1.16 P	24.7	9.8	4.7	4.1	102	D.G.	(Ye et al., 1995)
HIV Rev-RRE	7	0.91- 1.18 P	15	7.8	5.8	8.1	92	D.G.	(Ye et al., 1996)
	19	1.05- 1.39 A	11.6	5.4	2.8	14.7	61	D.G.	(Battiste et al., 1996)
U1A	25	1.0- 1.36 A	23.7	12.5	3.5	17.4	96	R.T.	(Allain et al., 1996)

[a] When the reported rmsd is the average pairwise rmsd a P is indicated. An A indicates the rmsd was calculated by comparing the structures to the average structure.

[b] Restraints were only given for the 16 nucleotides defining the "core."

incorporate the same essential features. Furthermore, RNA NMR studies have been found to be very self-consistent as all the general features and many of the specifics were very similar for the four structures that have been independently solved by two research groups.

1.3.2 Choosing a molecular modeling protocol

The goal of the molecular modeling calculations is to start from a sufficiently random initial model and then to explore as much conformational space as possible. It is difficult to assess whether the protocols currently used are sampling all conformational space consistent with the data and this is an area of much study (Brunger & Karplus, 1991). Three major types of initial models are predominantly used as shown in Table 1.5: random coordinates, random torsions, and distance geometry. The major difference between these approaches is the computational expense required to generate a family of structures. Generally, random torsions offers high convergence rates in the 40-50% range (Aboul-ela et al., 1995; Allain et al., 1996; Allain & Varani, 1995; Jucker *et al.*, 1996) while random coordinates requires significant sorting of the atoms and generally generates lower convergence rates in the 10-20% range (Brodsky & Williamson, 1997; Puglisi et al., 1995). Distance geometry is an algorithm routinely used to sample conformational space in protein NMR studies with great success (Vlieg & Gunsteren, 1991). In RNA NMR studies, distance geometry has been found to be especially useful when many long-range, global fold type of NOEs are found as in peptide complexes and the highly folded ATP aptamer (Battiste et al., 1996; Dieckmann et al., 1996; Fan et al., 1996a; Ye et al., 1996).

1.3.3 Determining bounds for the distance restraints

A large difference in the number of NOEs per nucleotide is seen in the structures summarized in Table 1.5 in part because not all NOEs are always included in the distance restraint data set for molecular modeling calculations. These apparent discrepancies occur because although some intranucleotide NOEs are important in defining RNA structure, there are a number of intraribose and intrabase NOEs that do not contain any useful

information either because the range of sugar puckers do not change these distances significantly within the precision that the constraints can be accurately defined or the covalent structure of the nucleotide itself defines the structure. The HIV-2 TAR argininamide NMR study did not use intrabase H5-H6 and H5-H4 NOEs (Brodsky & Williamson, 1997). Also, one of the BIV Tat-TAR and one of the ATP aptamer studies excluded many "conformationally unimportant" NOEs from the NOE restraint list, thus accounting for the small difference between the total number of NOEs and internucleotide NOEs shown in Table 1.5 (Dieckmann et al., 1996; Puglisi et al., 1995).

Deciding what bounds to use for NOE restraints is a critical issue for RNA NMR structures. Modeling studies of protein structures using both synthetic and real data have demonstrated a correlation between the precision of a family of structures and the accuracy to a target structure (Liu *et al.*, 1992). This correlation is critically dependent on the accuracy of the bounds used and the proper identification of NOEs. However, it is unclear whether this same correlation between rmsd and accuracy pertains to RNA structures which have fewer inter-residue NOEs and lack the global fold, long-range NOEs often found in protein NMR studies. However, as discussed extensively in this review, the dynamic character of many RNA molecules on the NMR time scale prevents the accurate determination of tight bounds. For RNA NMR studies, NOE constraints are often determined semi-quantitatively and placed into four categories strong, medium, weak and very weak NOEs. Protein NMR studies have shown that with a sufficiently large restraint data set, a precise and accurate family of structures can be determined even when the loose bounds are used (Liu et al., 1992). The RNA NMR studies listed in Table 1.5 have used different upper and lower bounds to define these categories and these differences lead to slightly different residual distance violation statistics which may or may not reflect the quality of the structure. Each research group has used a slightly different methodology to determine the NOE bounds which also changes on a case by case basis depending on the

quality of the spectra. The P1 helix, HIV TAR and U1A studies set all the lower bounds to 1.8 Å with upper bounds ranging from 3.0 Å for the most intense NOEs to 7.0 Å for the weakest NOEs found in H_2O experiments (Aboul-ela et al., 1995; Allain et al., 1996; Allain & Varani, 1995; Brodsky & Williamson, 1997). One of the reasons for this very conservative approach is demonstrated by the extreme case of the HIV TAR studies where a weak binding ligand which caused a variety of conformational exchange issues that precluded the use of accurate tight bounds such that both groups used very conservative bounds (Aboul-ela et al., 1995; Allain et al., 1996; Allain & Varani, 1995; Brodsky & Williamson, 1997). However, this approach may be too conservative in some cases as much of the information in the NOE data is not being used. An alternative approach, often used, is to use slightly tighter bounds where the lower bounds are set to 1.8, 2.5, and 3.5 Å with upper bounds of 2.5, 3.5, and 5 Å for the strong, medium, and weak categories respectively (Fan et al., 1996a; Fourmy *et al.*, 1996; Jucker et al., 1996; Puglisi et al., 1995; Ye et al., 1996; Ye et al., 1995). This strategy has been used in a number of cases with water NOESY spectra treated differently to account for possible spin diffusion effects of close amino protons. Both these approaches lead to a family of converged structures as shown in Table 1.5. However, when looser bounds were used, a significant number of restraints were used and thus presumably overcame the larger conformational space allowed by each bound as illustrated by the P1 helix, U1A, and HIV TAR studies (Aboul-ela et al., 1995; Allain et al., 1996; Allain & Varani, 1995; Brodsky & Williamson, 1997). The danger of using tight bounds is that errors could be made some of which may lead to the wrong structure or can limit the rate of convergence making the structure calculations difficult. Some techniques have been developed to help limit the amount of spin diffusion and therefore increase the accuracy of tight bounds by limiting the effects of spin diffusion thus increasing the precision and presumably the accuracy of the resulting structures (Hoogstraten & Markley, 1996), but these have yet to be applied to RNA.

1.3.4 Hydrogen bonds in RNA NMR structures

Some of the most important interactions are hydrogen bonding interactions which can be very difficult to determine unambiguously by NMR especially when exploring non-canonical base pairing interactions which are the most biologically interesting and novel aspects of RNA structure. Evidence for Watson-Crick base pairing in the A-form helix usually consists of the observation of upfield shifted imino resonances along with a network of cross-strand NOEs indicative of a typical base pair. Interestingly, with a sufficiently large NOE restraint data set, the global fold of an RNA can be determined in the absence of any explicit hydrogen bonding restraints (Brodsky & Williamson, 1997); this thesis. In this study, the initial stages of the molecular modeling calculations were performed in the absence of any hydrogen bonding constraints, but because of the large number of cross strand NOEs in the base-paired regions, the correct global fold was obtained, and hydrogen bonding partners could be unambiguously assigned. This approach explicitly demonstrated that base pair hydrogen bond restraints could be safely added to the data restraint set for the final refinement stages of the molecular modeling calculations.

The situation is more complicated for non-canonical base pairing and 2'OH hydrogen bonding interactions. There is no general consensus describing the rules for identifying a hydrogen bonding interaction in RNA. However, it makes sense to use the statistical nature of the structure determination to report the ranges and average distances for the heavy atoms involved in the proposed hydrogen bonding interaction as has been done in a few studies (Brodsky & Williamson, 1997; Fourmy et al., 1996; Jiang et al., 1996; Jucker et al., 1996).

Recently, crystal structures of two different ribozymes have highlighted the important structural role of 2'OH's (Cate et al., 1996; Pley et al., 1994a; Scott *et al.*, 1995). Only a handful of 2'OH have been observed by NMR thus making any direct evidence for hydrogen bonding difficult (Allain & Varani, 1995). However, hydrogen

bonding interactions have been proposed based on a large number of NOEs in the region which localize possible hydrogen bond acceptor and donor groups near the 2'OH as in the GNRA structures (Jucker et al., 1996). One additional consideration is suggested by the paromomycin-16 S RNA structure where some potential interactions appeared with heavy atom distances in the 4.0 - 4.5 Å range (Fourmy et al., 1996). This distance is generally considered too long for hydrogen bond, but could possibly be due to a water mediated hydrogen bond. These type of interactions are very difficult to determine even in protein NMR where large hydrophobic cores help protect single water molecules from exchanging with the bulk solvent (Otting *et al.*, 1991; Otting & Wuthrich, 1989b). The structural role of 2'OH's and water molecules in RNA structure will be an important area of future investigation.

1.4 NMR Dynamics Measurements

NMR relaxation parameters provide a powerful method to study the internal dynamics of macromolecules (Palmer *et al.*, 1996). The relaxation properties of protonated ^{15}N or ^{13}C heteronuclei are dominated by dipole-dipole interactions, permitting the measurements to be correlated with the interspin bond vectors, which in turn are related to the internal dynamics and the overall molecular tumbling. The relationship between the internal dynamics of macromolecule and their biological function has been extensively investigated. To date, the role of dynamics for biological functions has not been conclusively determined, though some studies have found correlations with elements of specific binding (Kay et al., 1996). Recent studies have also used dynamics to probe protein stability and folding (Buck *et al.*, 1995; Farrow *et al.*, 1997). RNA offers a number of unique possibilities for NMR dynamics measurements because of the isolated ^{1}H-^{15}N and ^{1}H-^{13}C spin pairs in the aromatic bases. In proteins, most ^{1}H-^{13}C spin pairs are next to other carbons and ^{13}C-^{13}C homonuclear scalar relaxation mechanisms complicate the measurements. Few studies of RNA dynamics have been pursued as isotopically

labeled samples and high resolution structures have only recently been available (Akke *et al.*, 1997; King *et al.*, 1995). Previous studies of RNA dynamics have often encountered conformational exchange dominating the measurements, and transverse relaxation rates are often used to determine whether the residues are structured (Address *et al.*, 1997; Legault & Pardi, 1997). RNA dynamics may be more complicated to study as different motional models and analysis of the relaxation rates may need to be developed because many RNAs probably tumble anisotropically. Also, the spin pairs in the heteroaromatic rings of the bases have large Chemical Shift Anisotropies (CSA) contributing to the relaxation permitting the measurement of cross-correlation effects to determine the intrinsic dynamics without adverse effects of conformational exchange (Dayie *et al.*, 1996).

A number of excellent reviews have been written describing dynamics measurements (Dayie et al., 1996; Peng & Wagner, 1994), therefore, this discussion will focus on briefly summarizing the measurements and analysis performed in Chapter 5. Experiments to measure six basic relaxation rate constants with high sensitivity have been devised (Peng & Wagner, 1994). For a ^{1}H-^{15}N spin pair, these relaxation rates can be written in terms of the spectral densities as follows:

$$R_N(N_z) = (3D + C)J(\omega_N) + DJ(|\omega_H|+|\omega_N|) + 6DJ(|\omega_H|-|\omega_N|) \tag{2}$$

$$R_N(H_z \leftrightarrow N_z) = D[6J(|\omega_H|-|\omega_N|) - J(|\omega_H|+|\omega_N|)] \tag{3}$$

$$R_N(N_x) = (2D + 2C/3)J(0) + (3D/2 + C/2) J(\omega_N) + (D/2)J(|\omega_H|+|\omega_N|) + 3DJ(\omega_H) + 3DJ(|\omega_H|-|\omega_N|) + R_{ex} \tag{4}$$

$$R_{HN}(2H_zN_x) = (2D + 2C/3)J(0) + (3D/2 + C/2)J(\omega_N) + (D/2)J(|\omega_H|+|\omega_N|) + 3DJ(|\omega_H|-|\omega_N|) + \rho_{HH} + R_{ex} \tag{5}$$

$$R_{HN}(2H_zN_z) = (3D + C)J(\omega_N) + 3DJ(\omega_H) + \rho_{HH} \tag{6}$$

$$R_H(H_z) = D[J(|\omega_H|+|\omega_N|) + 3J(\omega_H) + 6J(|\omega_H|-|\omega_N|)] + \rho_{HH} \tag{7}$$

where $R_N(N_z)$ is the ^{15}N longitudinal rate constant, $R_N(H_z \leftrightarrow N_z)$ is the heteronuclear NOE

(longitudinal cross-relaxation), $R_N(N_x)$ is the ^1H longitudinal, $R_{HN}(2H_zN_x)$ is the antiphase

^{15}N single quantum coherence, $R_{HN}(2H_zN_z)$ is the longitudinal heteronuclear two-spin

order, and $R_H(H_z)$ is the ^1H longitudinal relaxation rate constant. The constant D is

proportional to the mean squared value of the heteronuclear DD interaction and is given by

$$D = \frac{h^2 \gamma_H^2 \gamma_c^2}{4 r_{HC}^6} \qquad (8)$$

The constant C is proportional to the mean square value of the anisotropic component of the

^{15}N-B_o interaction and is defined as

$$C = (\gamma^2 B_o^2)(\Delta^2) / 3 \qquad (9)$$

and Δ is the chemical shift anisotropy for which some values of been reported for both

carbon (Williamson & Boxer, 1989) and nitrogen (Akke et al., 1997). P_{HH} describes the

sum of all proton-proton dipole-dipole interactions for the proton attached to nitrogen. The

R_{ex} term describes any exchange contributions to the relaxation rates and can be measured

by rotating field experiments (Akke & Palmer, 1996; Palmer et al., 1996; Szyperski *et al.*,

1993).

These relaxation rate constants are sensitive to a range of time scales allowing NMR

spectroscopists to monitor molecular motions from the picosecond to hour time scale as

summarized in Table 1.6. To measure these relaxation rate constants, various pulse

sequences have been designed to isolate just the relaxation process desired.

The spectral density function J(ω) describes the contribution of orientational

dynamics of the molecule from motions in a particular frequency range. Once the

frequencies and magnitudes of the different orientational processes are determined,

Table 1.6: NMR explores dynamics on a variety of time-scales

Time Scale (τ)	NMR Methods	Molecular Process
s < τ <hrs	Temp, pH, ^{2}H Exchange HSQC	•Protein/RNA Folding •Amide hydrogen Exchange
ms < τ < s	Line shape analysis, EXSY, Pulse Labeling	•Protein/RNA folding •Proline Isomerization •Imino proton exchange
ns < τ < ms	Variable RF field on/off resonance Spectral Density Mapping at diff B_o	•H-bond formation/breaking •Isomerization processes •Change in metal coordination •Helix-coil transition
ps < τ < ns	R_1, R_2, R_{1p} NOE, ROE	•Tightly bound water •Overall tumbling •Local conformational motion
τ < ps	Averaging of relaxation rates	•Fast atomic motions •Fast side chain dynamics •Dihedral Angle Fluctuations

motional models can be fit to $J(\omega)$ to determine how the behavior of individual spin

systems and the overall tumbling of the molecule. Peng and Wagner introduced the full

spectral density mapping approach in which six relaxation rates are measured at multiple

fields to determine the spectral density values at several frequencies (Peng & Wagner,

1994; Peng & Wagner, 1995). The full mapping approach measures the six rate constants

listed in equations 2-7, and then solves for the five spectral density terms at different

frequencies. However, the full mapping approach suffers from a few practical problems

originating

from experimental difficulties. Because the difference of some of the relaxation rate

constants is required and these rate constants are of the same magnitude, the method

required a high level of precision that is not always possible and thus, it is sensitive to

experimental error. An alternative approach, promoted by several groups, is to use a

reduced spectral density mapping approach (Farrow *et al.*, 1995; Ishima & Nagayama,

1995; Lefevre *et al.*, 1996). Since the high frequency terms are the same magnitude, it can

be assumed that $J(\omega_H \pm \omega_N)=J(\omega_H)$ which then allows the spectral densities to be calculated

from just three relaxation measurements, ^{15}N longitudinal and ^{15}N transverse relaxation

rates, and the heteronuclear NOE. Instead of calculating the spectral densities from the

relaxation rates, a functional form of $J(\omega)$ is assumed containing a limited number of

parameters that are fit to the data. One of these is the commonly applied Lipari-Szabo

model-free approach which describes the dynamics in terms of internal correlation times

and order parameters which describe the flexibility of the measured spin system (Lipari &

Szabo, 1982a; Lipari & Szabo, 1982b).

1.5 HIV Tat-TAR Background

1.5.1 HIV life cycle and gene expression

HIV contains ~10-20 viral proteins that are produced from overlapping exons on a 10 kilobase (kB) RNA transcript. Early in the viral life cycle the small regulatory proteins (Rev, Tat, Vif, Nef, etc.) are translated from mRNAs generated by multiple splicing events on the full 10 kB mRNA. Tat acts during these early stages of the life cycle to increase the expression of the viral mRNAs by enhancing the efficiency of transcription elongation and perhaps also by increasing the rate of transcription initiation. To function, Tat binds to the Transactivation Response Element (TAR), an RNA hairpin that resides in both the 3' and 5' LTR of the HIV genome as shown in Figure 1.7. However, only the 5' copy is active and binds Tat. Using both *in vivo* and *in vitro* biochemical approaches, the Tat binding site has been localized to a small RNA element within TAR containing a hexanucleotide loop and a three base bulge, shown in Figure 1.8 (Cordingly *et al.*, 1990; Dingwall *et al.*, 1989; Dingwall *et al.*, 1990; Roy *et al.*, 1990b; Weeks *et al.*, 1990). The loop sequence is critical for transcriptional activation but does not affect Tat binding to TAR, and the residues critical for Tat binding are localized to the bulge region (Cordingly et al., 1990; Roy *et al.*, 1990a).

1.5.2 Mechanism of Tat action in vivo

In vivo, Tat requires additional cellular factors to activate transcription. It is clear that TAR provides the binding site for Tat and is formed on the nascent transcript, but it is less certain whether the target of activation is the elongating transcription complex or subsequently initiating complexes. To recognize TAR during transcription, it appears that at least one additional protein must assist Tat in binding. It is believed that his protein is encoded on human chromosome 12 and helps tether Tat to TAR by interacting with the hexanucleotide loop (Alonso *et al.*, 1992; Hart *et al.*, 1995; Newstein *et al.*, 1993). Rodent
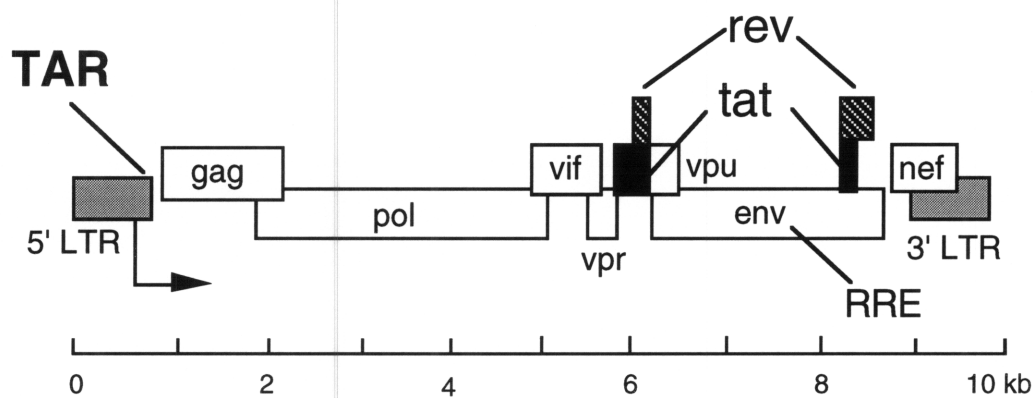
Figure 1.7: Open reading frames in HIV. The boxes represent the position of genes on the genomic RNA. Single and multiple splicing events occur to produce the different mRNAs to express the various gene products. Tat is expressed in two exons, highlighted in black. The active TAR is located in the 5'-LTR.

cells appear to lack the factor, supporting only low levels of Tat activity unless complemented by chromosome 12. The loop is clearly important for Tat activation as most nucleotides within the loop are required for transactivation and the sequence is highly conserved among both HIV-1 and HIV-2 isolates. However, since Tat can also activate transcription when tethered to a heterologous RNA-binding protein and delivered to a corresponding RNA-binding site placed at the 5' end of the transcript (Selby & Peterlin, 1990; Southgate *et al.*, 1990), or even when bound to DNA (Southgate & Green, 1991).

Although the identity of the loop-binding protein is not yet clear, several candidates have emerged from in vitro binding specificities (Jones & Peterlin, 1994). One of the these, is TRP-185 (or TRP-1) binds poorly to loop mutants or to mutants where the distance between the loop and the bulge has been increased, correlating with decreased Tat transactivation (Sheline *et al.*, 1991; Wu *et al.*, 1991). A cDNA encoding TRP-185 reveals a protein of 1621 amino acids contains a leucine zipper and may be a novel RNA-binding motif (Wu-Baer *et al.*, 1995). Three cofactors have been identified which help stimulate TRP-185 in vitro TAR binding including elongation factor 1α (EF-1α), the

polypyrimidine tract-binding protein (PTB), and a novel protein aptly named the stimulator of TAR RNA-binding (SRB). Recombinant EF-1a, PTB, and SRB have been shown to stimulate binding to TAR in vitro (Wu-Baer *et al.*, 1996). TRP-185 and RNA polymerase II bind in a mutually exclusive manner, and it has been proposed that TRP-185 may disengage RNA polymerases stalled by binding to TAR. TRP-185 is located on human chromosome 1p and thus is not the factor which resides on chromosome 12. A TAR loop-binding protein of 83 kDa has been observed in human cells and in rodent cells containing human chromosome 12, but not in the parental rodent cells.

Recently, some progress has been made towards dissecting the mechanism of Tat function. Tat can be positioned relatively far from the 5' end (~500 nucleotides) and Tat remains associated with the elongating RNA polymerase II (Keen *et al.*, 1996). Evidence

A)

```
        G  G
    U          G
    C          A
      C • G
      29   36
      G • C
      A • U
      G • C
        26   39
   C U
    U
    23
      A • U
      22   40
      G • C
      A • U
      C • G
      C • G
      G • C
      G • C
      16   46
```

```
        G  G
    U          G
    C          A
      C • G
      29   36
      G • C
      A • U
      G • C
        26   39
   U
   U
   23
      A • U
      22   40
      G • C
      A • U
      C • G
      C • G
      G • C
      G • C
      16   46
```
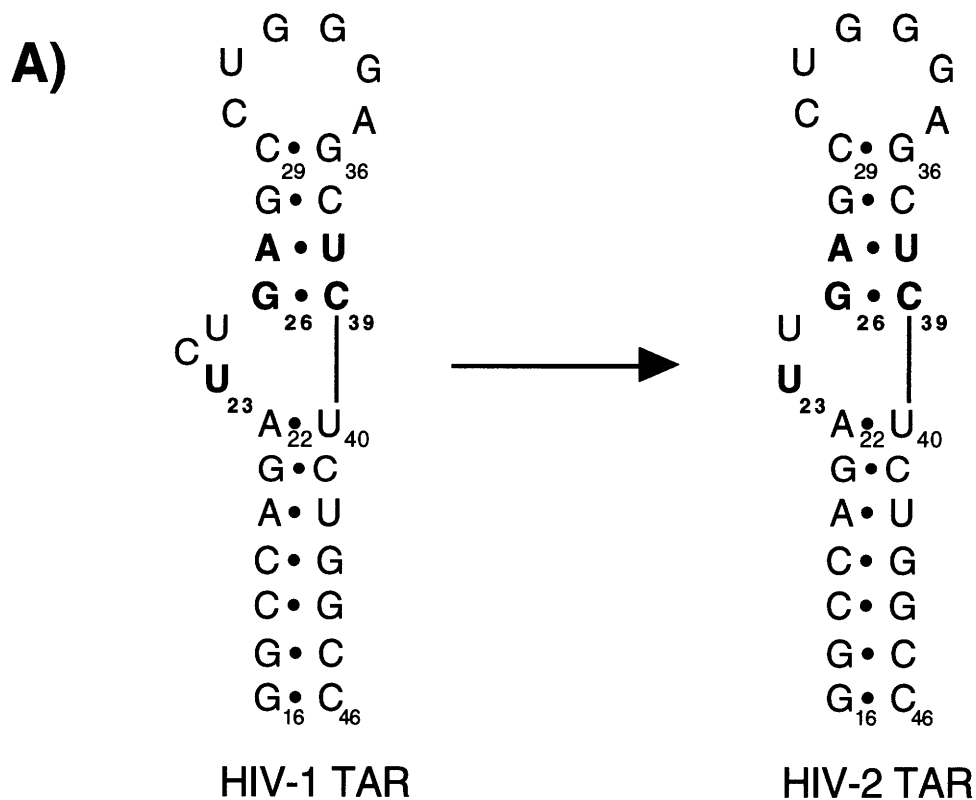
HIV-1 TAR   HIV-2 TAR

B)
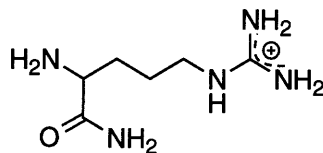
Figure 1.8: A) The sequence and secondary structure of HIV-2 TAR is identical to HIV-1 TAR except for the deletion of a bulge nucleotide that is not required for Tat or argininamide binding. The critical nucleotides required in the bulge region are highlighted in bold. **B)** The molecular structure of argininamide.

has pointed to a role for kinases in Tat function. A Tat stimulatory factor, Tat-SF1, co-purifies with a cellular kinase and stimulates Tat activity in transfection experiments. Tat-SF1 us a 754 amino acid protein which contains two RNA recognition motifs and a highly acidic carboxyl-terminal half. A second kinase, the Tat-associated kinase (TAK), can specifically bind to the Tat activation domain and phosphorylate the C-terminal domain (CTD) of RNA polymerase II. A third kinase, the TFIIH-associated kinase which also phosphorylates the CTD of RNA polymerase II is stimulated by Tat in vitro. Phosphorylation of the CTD is believed to trigger the transition from initiating to elongating transcription complexes and also to influence later stages during elongation (O'Brien *et al.*, 1994). Thus, an intriguing model for Tat function emerges from these kinase studies which link Tat to regulation of the transcription machinery. However, it remains unclear which, if any, of these kinases is biologically important for HIV Tat function.

1.5.3 Biochemical analysis of the Tat-TAR complex

Features of RNA recognition Mutational studies have shown that U23 is the only essential nucleotide in the three nucleotide HIV-1 TAR bulge (Churcher *et al.*, 1993; Weeks & Crothers, 1991). The other two bulge residues can be freely substituted, acting as spacers that can even be replaced with a non-nucleotide (Churcher et al., 1993; Sumner-Smith *et al.*, 1991). The bulge can vary in size from two to four residues without significant loss of Tat binding affinity (Weeks & Crothers, 1991). Two base pairs, G26-C39 and A27-U38, in the upper stem have also been shown to be important for specific Tat binding (Churcher et al., 1993; Weeks & Crothers, 1991). The same nucleotides, U23, G26, and A27, are also critical for transactivation assays *in vivo* (Churcher et al., 1993). Modification of G26 and A27 at the N7 position with DMS or DEPC interferes with Tat binding, consistent with the importance of major groove recognition (Churcher et al., 1993; Hamy *et al.*, 1993; Weeks & Crothers, 1991). Any change to the base of U23 severely inhibits Tat binding, although the U23 2'OH is not essential for Tat binding (Hamy et al., 1993; Roy et al.,

1990a; Weeks & Crothers, 1991). The mutation and modification interference data indicate that just a few nucleotides in and around the bulge region of TAR are required for specific Tat protein binding. Except for the G26-C39 and A27-U38 base pairs, the sequence of the two stems surrounding the bulge region is not critical for Tat binding as long as Watson-Crick base pairing is maintained (Weeks & Crothers, 1991).

Features of Tat protein required for TAR binding The HIV-1 Tat protein is 86 aminos acids long and is encoded by two exons, but only the first exon (1-72 aa) is sufficient to activate the HIV-1 LTR (Kuppuswamy *et al.*, 1989). Tat has a domain-like structure identified by sequence analysis as shown in Figure 1.9. It can be functionally divided into an N-terminal activation domain (1-48) and a C-terminal RNA-binding domain (49-72). The activation domain consists of an acidic and proline-rich amino-terminal region (1-21), a conserved cysteine-rich region (22-37), and a hydrophobic core region (38-48) found in all lentiviral Tat proteins (Kuppuswamy et al., 1989). The RNA-binding domain (49-57) is rich in arginines also acts as a nuclear localization signal (Hauber *et al.*, 1989; Ruben *et al.*, 1989). A 25 amino acid Tat fragment containing only the core and arginine-rich domains can weakly activate the HIV-1 LTR promoter (Derse *et al.*, 1991), suggesting that these regions are sufficient for a minimal TAR response.

Short peptides spanning the arginine rich region of the Tat protein bind specifically to TAR, recognizing the same U23, A27, and G26 nucleotides important for Tat binding and transactivation (Churcher et al., 1993; Cordingly et al., 1990; Long & Crothers, 1995; Weeks et al., 1990; Weeks & Crothers, 1991). The same modification interference patterns observed for Tat were also observed for these peptides suggesting that TAR binds to the peptides in a similar manner as Tat protein (Churcher et al., 1993). In addition, modification of phosphates P22 and P23 with ENU interferes with Tat peptide binding (Calnan *et al.*, 1991; Churcher et al., 1993; Pritchard *et al.*, 1994; Tao & Frankel, 1992). Although peptide-TAR complexes have been very useful for performing careful
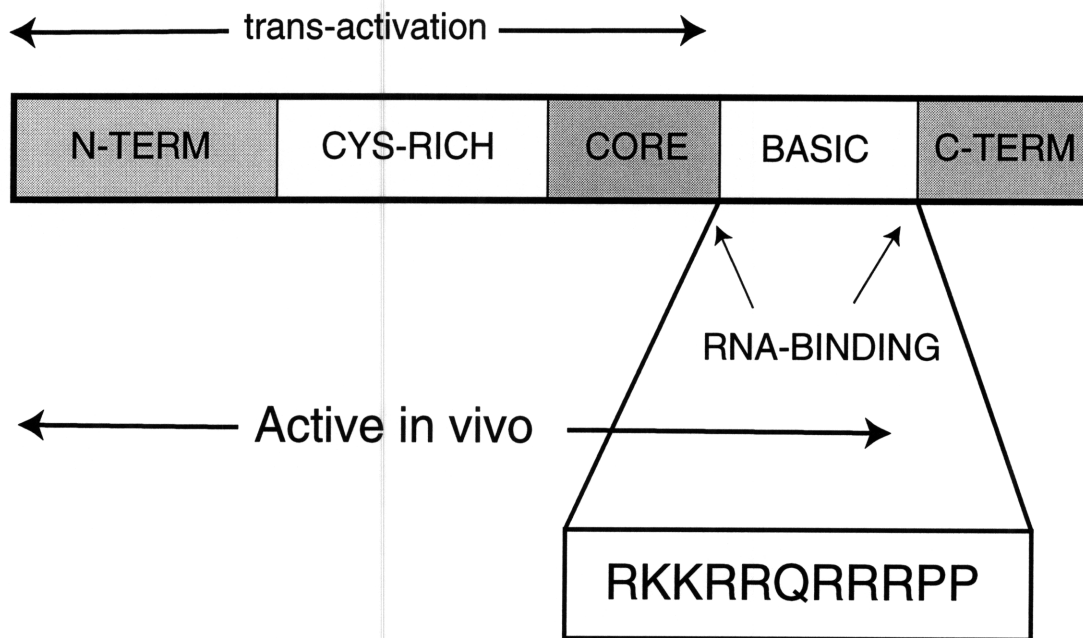
Figure 1.9: HIV Tat proteins have a domain structure. The C-terminal region is not required for in vivo activity. The N-terminus is highly acidic and the core region is hydrophobic. Other than the basic RNA binding domain, very little is known about the function of these domains.

biochemical studies, they have not yielded high quality NMR spectra suitable for detailed

NMR structural studies (Aboul-ela et al., 1995; Puglisi et al., 1992).

1.5.4 Previous NMR studies of TAR

Two NMR studies have been performed on the HIV-1 TAR-argininamide complex

(Aboul-ela et al., 1995; Puglisi et al., 1992). The first study by Puglisi *et al.* showed that

a specific complex forms between the amide derivative of arginine, argininamide, and HIV-

1 TAR forms a complex suitable for study by NMR. Similar binding modes of the peptide

and argininamide were indicated by the very similar spectral characteristics in both the

peptide and argininamide complexes. This similarity of conformations has also been seen

in transient electric birefringence studies (Zacharias & Hagerman, 1995). The NMR data

showed that the argininamide guanidinium group is binding near G26 and that U23 is in the

major groove near A27, suggesting the formation of a U38-A27•U23 base triple upon

argininamide binding. When non-experimental base triple and argininamide hydrogen

bonding restraints were added, the NMR data were found to be consistent with the

resulting model of the bound conformation of TAR. The major strength of the model is

that it provides a structural role for every functional group on the RNA identified as

important by binding studies. However, no direct evidence for the proposed base triple

hydrogen bonds was obtained. To test the base triple proposal, an isomorphic $C-G•C^+$

base triple was designed and found to bind argininamide in a pH dependent manner, further

suggesting that base triple formation is important for argininamide binding (Puglisi et al.,

1993). The Puglisi *et al.* NMR studies did not address the structure of the hexanucleotide

TAR loop. Two studies focusing on the TAR loop showed how this loop has little ordered

structure defined by relatively few internucleotide (Colvin *et al.*, 1993; Jaeger & Tinoco,

1993).

The second NMR study by Aboul-ela *et al.* of the HIV-1 TAR-argininamide

complex used heteronuclear multidimensional NMR to assign a much larger number of

NOEs and torsion restraints for the structure determination than the previous NMR studies. Although the U23 was located in the major groove near the argininamide ligand, the structure determination by Aboul-ela *et al.* did not find evidence for the proposed base triple (Aboul-ela et al., 1995). The global conformation of the TAR bulge was similar to that of the low resolution structure of Puglisi *et al.*, but the details of the structure differ. Both studies positioned the U23 and the argininamide in the major groove but Aboul-ela *et al.* found that a U38-A27•U23 base triple was inconsistent with their NOE data.

Chapter 2: HIV-2 TAR-Argininamide NMR

2.1 Formation of the TAR-Argininamide Complex

In previous NMR studies of the HIV-1 TAR-argininamide complex, the U40 imino proton resonance was not observed despite numerous other NOEs indicating that this base pair was formed. Since only U23 in the bulge is important for binding, and reasoning that a shorter bulge might have more favorable dynamic properties, the HIV-2 TAR-argininamide complex was examined. The improved spectral properties due to the differences in dynamics of the HIV-2 TAR RNA were immediately evident from the imino proton spectrum of the argininamide complex, which was very similar to the HIV-1 TAR-argininamide complex, except that an additional resonance at 14.06 ppm was observed from the imino proton of U40. Figure 2.1 shows how the HIV-2 TAR imino spectrum changes upon the addition of argininamide. The U40 imino resonances appears at 5 mM argininamide and sharpens significantly at 5°C. The apparent binding constant of the HIV-2 TAR-argininamide complex is 2-3 mM, which is the same as observed previously for the HIV-1 complex. The total argininamide concentration in the sample is ~6 mM, which corresponds to 76 - 88% of the TAR molecules bound to argininamide in a rapid equilibrium. Higher concentrations of argininamide produced no further spectral changes, and weak NOEs were observed to other non-specific argininamide binding sites. Therefore, approximately 6 mM argininamide concentrations were used for all structural studies.

2.2 Materials and Methods

2.2.1 Sample preparation

Three samples were prepared for this study: unlabeled, uniformly ^{13}C-labeled, and uniformly ^{15}N-labeled RNA, by *in vitro* transcription using T7 RNA polymerase from
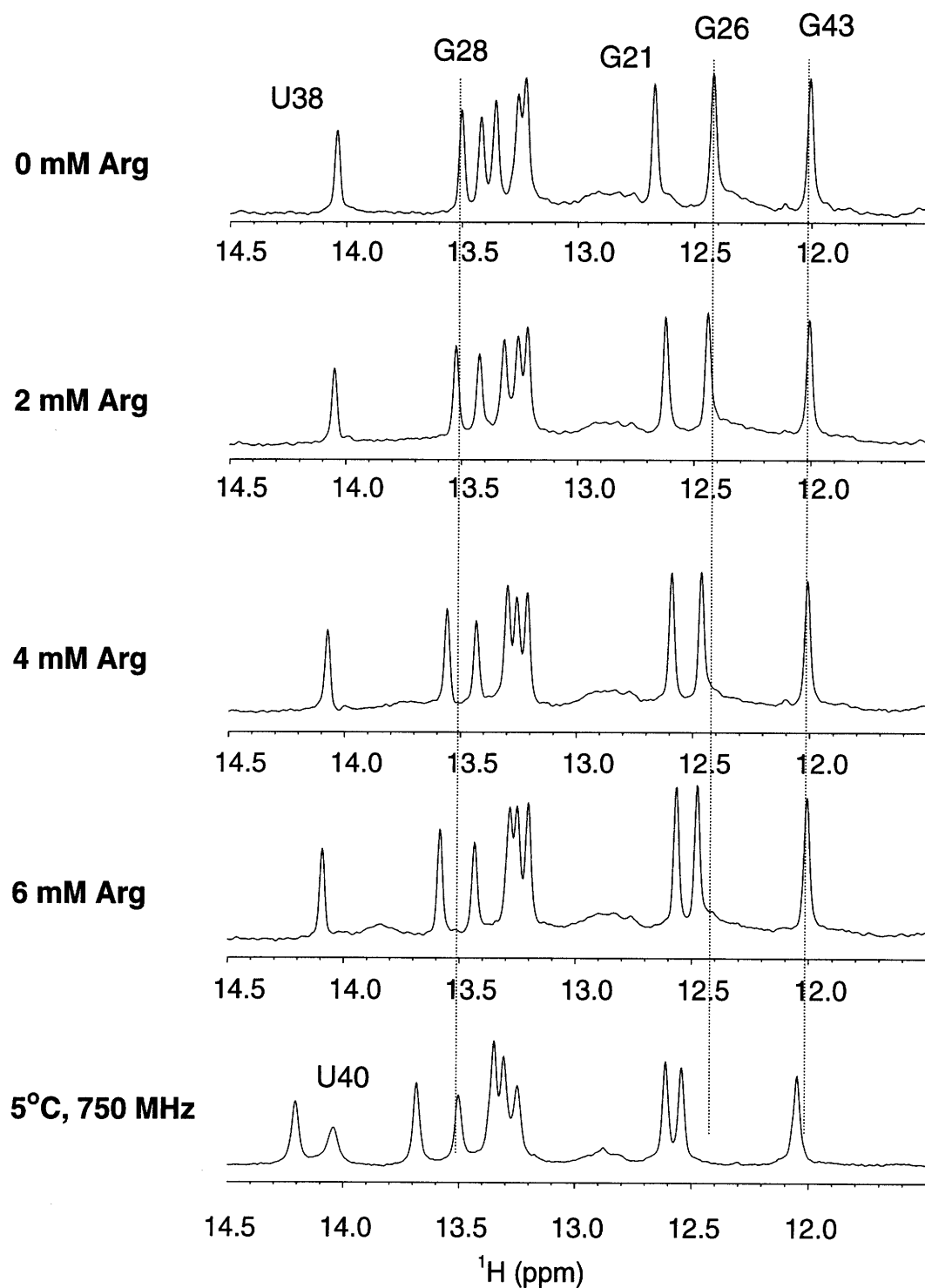
Figure 2.1: The imino spectrum of HIV-2 TAR changes upon addition of argininamide. In particular the resonances around the bulge region, G26, G21, and U38 imino resonances shift and the U40 imino proton appears when the temperature is lowered to 5°C. The HIV-2 TAR RNA concentration is ~2.5 mM.

oligonucleotide templates (Milligan *et al.*, 1987). Isotopically labeled samples were prepared using the procedure reviewed in (Batey et al., 1995). Slightly different transcriptions conditions were used for the isotopically labeled nucleotide triphosphates prepared from cells. For the labeled samples, 20 μl transcription trials were performed to maximize the molar yield per nucleotide triphosphate. The transcription conditions used for each sample are summarized in Table 2.1. Each T7 polymerase preparation was optimized as the activity per milligram varies significantly. Recently, two new tricks were introduced and found to increase transcription yields significantly. The first addition is to add an extra G-C base pair in the promoter region which is reported to increase yields by approximately two to ten-fold. Also, the addition of a methylindole to the 5' end of the template helps prevent additional nucleotides from being added to the 3' end, e.g. N+1. These two advances yielded three micromoles of pure dialyzed HIV-2 TAR RNA from a 80 ml transcription reaction. Without these new additions, 80 ml of transcription typically yielded only about one micromole. Radioactive small scale trial transcriptions indicated that an increase of between 3-5 fold in yield is expected.

After transcription, the RNAs were purified by electrophoresis on 20% (w/v) polyacrylamide gels, electroeluted from the gel slices, and ethanol precipitated. The samples were dialyzed for >48 hours against NMR buffer (10 mM sodium phosphate, pH 6.4, 50 mM sodium chloride and 0.1 mM EDTA). The final concentrations in 500 μl were 2.5 mM, 1.5 mM, and 1.5 mM for the unlabeled, ^{13}C-labeled and ^{15}N-labeled TAR RNA samples, respectively.

The labeled RNA's were titrated in the NMR tube to a final total argininamide (Sigma) concentration of 6 mM, while the unlabeled TAR RNA was titrated to 6.5 mM. No changes in the spectra are observed above this concentration indicating that the binding site is nearly saturated. As in the previous NMR studies of HIV 1 TAR, argininamide is in fast exchange between bound and unbound forms.

Table 2.1: Transcription Conditions for HIV-2 TAR.

Sample	Total NTP Concentration (mM)	MgCl$_2$ (mM)	Yield (μmoles)	Total Volume of Reaction
unlabelled	16	32	0.6	60
unlabelled with methyl-indole and extra G-C	16	32	3	80
^{13}C TAR	8	16	0.75	120
^{15}N TAR	8	16	0.6	100

All reactions were performed in 40 mM Tris pH 8.1, 1 mM spermidine, 0.01% (v/v) triton X-100, and 5 mM dithiothreotol.

HIV-2 DNA Template:
5' - X GGC CAG AGA GCT CCC AGG CTC AAT CTG GCC TAT AGT GAG TCG TAT TA(G)- 3'

DNA T7 Top Strand:
5' - (C)TA ATA CGA CTC ACT ATA

The nucleotides in parantheses represent the added G-C base pair which increases transcription yields.

2.2.2 NMR spectroscopy

Spectra were collected on 500 MHz or 591 MHz spectrometers constructed at the Francis Bitter Magnet Lab, equipped with Nalorac triple resonance gradient probes. 750 MHz spectra were recorded on a Varian Unity Plus spectrometer equipped with a Varian triple resonance gradient probe. Some 500 MHz spectra were collected on a Varian VXR-500 spectrometer with a Varian inverse detection probe. The 4D HMQC-NOESY-HSQC spectrum was collected on a Bruker DMX 600 MHz spectrometer. All spectra were processed and analyzed using NMRPipe (Delaglio *et al.*, 1995) and PIPP (Garrett *et al.*, 1991) on a Silicon Graphics Indy workstation.

Proton NMR spectra in H_2O were recorded at a variety of temperatures and mixing times, including 40, 50, 100, 150, and 200 ms. NOESY spectra used a WATERGATE 3-9-19 water suppression scheme with a selective E-BURP-1 flip-back pulse (Green & Freeman, 1991; Lippens *et al.*, 1995; Piotto *et al.*, 1992) . The flip-back pulse significantly helped the observation of NOEs to the argininamide guanidinium protons. Sweep widths were 12000, 14000, 16000 Hz at 500, 600, and 750 MHz, respectively, with 4096 x 512 real points collected. Data were zero filled to 4K x 2K real points and apodized using Gaussian-Lorentzian functions in both dimensions before Fourier transformations.

NOESY, DQF-COSY, and TOCSY experiments were recorded in 99.996% ^2H_2O at 25°C (Varani & Tinoco, 1991). Data sets with 4096 real points in *t2* and 512 real points in *t1* were acquired for the NOESY and TOCSY. The DQF-COSY was collected with 1024 points in *t2* with WALTZ decoupling of ^{31}P during acquisition. To monitor NOE buildups, spectra were recorded at 500, 600, and 750 MHz with mixing times of 50, 100, 200, 300, and 400 ms. In addition, a 35 ms NOESY experiment at 500 MHz was also collected. Sweep widths were 5500 Hz, 6000 and 8000 Hz at 500, 600, and 750 MHz, respectively.

Acquisition parameters for all the heteronuclear spectra are provided in Table 2.2.

Table 2.2: Data acquisition parameters for heteronuclear NMR experiments.

Molecule	Experiment	Field (MHz)	$\omega_4{}^a$	$\omega_3{}^a$	$\omega_2{}^a$	$\omega_1{}^a$	J_{XH} (Hz)[b]	Other	No. of scans	Ref.
Unlabelled TAR RNA	HP-HSQC	600			550 Hz 512 pts ^{31}P	6000 Hz 4096 pts ^{1}H		13 ms[c]	16	
	HP-COSY	600			5000 Hz 512 pts ^{31}P	2000 Hz 4096 pts ^{1}H			16	(Skélnar et al., 1986)
Uniformly ^{13}C-labeled TAR RNA	3D NOESY-HMQC	500,600		5050 Hz 32 pts ^{13}C	5000 Hz 128 pts ^{1}H	5000 Hz 512 pts ^{1}H	160	200 ms[d]	16	(Clore et al., 1990)
	3D HCCH-DIPSI	500		5000 Hz 37 pts ^{13}C	2500 Hz 64 pts ^{1}H	2500 Hz 512 pts ^{1}H	142	20 ms[e]	16	(Clore et al., 1990))
	3D HCCH-COSY	500		5000 Hz 32 pts ^{13}C	3000 Hz 64 pts ^{1}H	3000 Hz 512 pts ^{1}H	142		16	(Clore et al., 1990)
	4D HMQC-NOESY-HSQC	600	3333 Hz 20 pts ^{13}C	3333 Hz 20 pts ^{13}C	3003 Hz 50 pts ^{1}H	6024 Hz 384 pts ^{1}H	156	90 ms[d,f]	2	(Vuister et al., 1993)
	HSQC-CT	500,600			5000 Hz 256 pts ^{13}C	5000 Hz 512 pts ^{1}H	142	25 ms[g]	16	(Santoro & King, 1992)

Table 2.2 (cont): Data acquisition parameters for heteronuclear NMR experiments.

Molecule	Experiment	Field (MHz)	$\omega_4{}^a$	$\omega_3{}^a$	$\omega_2{}^a$	$\omega_1{}^a$	J_{XH} (Hz)^b	Other	No. of scans	Ref.
	3D-HCP	600		550 Hz 32 pts ^{31}P	6000 Hz 150 pts ^{13}C	6000 Hz 512 pts ^{1}H	165 Hz 8 Hz		16	(Heus et al., 1994) (Marino et al., 1994)
Uniformly ^{15}N-labeled TAR RNA	3D NOESY-HSQC	600,750		5000 Hz 31 pts ^{15}N	7500 Hz 128 pts ^{1}H	16500 Hz 1024 pts ^{1}H	90 Hz	150 ms^d	8	(Mori et al., 1995)
	HSQC	600,750			6800 Hz 64 pts ^{15}N	18000 Hz 2048 pts ^{1}H	90 Hz		8	(Mori et al., 1995)
	2 bond HSQC	500			10000 Hz 128 pts ^{15}N	6000 Hz 1024 pts ^{1}H	18 Hz		16	(Skélnar et al., 1994)

^{a} The number of real data points acquired, the spectral width, and the nucleus detected are listed for each dimension.

^{b} Given is the value of the of the one-bond X-nucleus-hydrogen coupling constant used for determination of the delay for transfer of magnetization between the proton and the X-nucleus. For the carbon NOESY experiments, the value used was an average of the aromatic and ribose coupling constants.

^{c} Although the H-P coupling constant is usually between 8-10 Hz, due to fast ^{31}P relaxation, reducing the transfer time from ~25 ms to ~13 ms significantly increases the sensitivity.

^{d} The mixing time of the period for NOE transfer.

^{e} The mixing time of the carbon spin- lock pulse for transfer of magnetization through the ribose ring.

^{f} The ^{13}C transmitter in the 4D experiment was at 80.9 ppm.

^{g} Value of the constant-time interval for evolution of the carbon chemical shift in ω_1.

In all ^{13}C experiments, ^{13}C decoupling was achieved with a GARP (Shaka *et al.*, 1983) scheme during acquisition and a composite 180° ^{13}C pulse in the middle of the t_1 proton evolution times. This composite pulse is required to effectively decouple enough bandwidth covering both the ribose and aromatic carbon regions. Spectra were apodized with shifted squared sine bell functions. For some spectra, the ω_1 and ω_2 dimensions were Fourier transformed first. The ω_3 dimension was linear predicted by a factor of 2 and then zero-filled to 128 points. The ω_2 dimension was then inverse Fourier transformed and linear predicted by a factor of two and zero filled by a factor of two. This processing strategy did not significantly help resolve peaks compared to the more straightforward and easier to implement method where the ω_2 dimension is not linear predicted.

The ^{15}N NOESY-HSQC at 5°C used a 3-9-19 WATERGATE sequence for water suppression at 750 and 600 MHz with the transmitter set to 124.7 ppm (Mori *et al.*, 1995). GARP was used to decouple ^{15}N during acquisition (Shaka et al., 1983), and a composite 180° ^{15}N pulse in the middle of the *t1* proton evolution time.

2.3 Assignment of Exchangeable Resonances

2.3.1 Secondary structure of HIV-2 TAR

Imino proton resonances were assigned from NOESY spectra in H_2O and confirmed in a 3D ^{15}N-NOESY-HSQC spectrum (Varani et al., 1996; Varani & Tinoco, 1991). Figure 2.2 shows the imino-imino proton region of a 75 ms mixing time 750 MHz ^{1}H-^{1}H NOESY spectrum in H_2O at 5°C. The U40 imino peak is sufficiently sharp that NOEs to the G21 and G26 imino protons are observable, as well as a NOE to C39(H42) and a strong NOE to A22(H2) (data not shown). These NOEs are consistent with stacking between the A22(U40) base pair and the G26-C39 base pairs at the junction
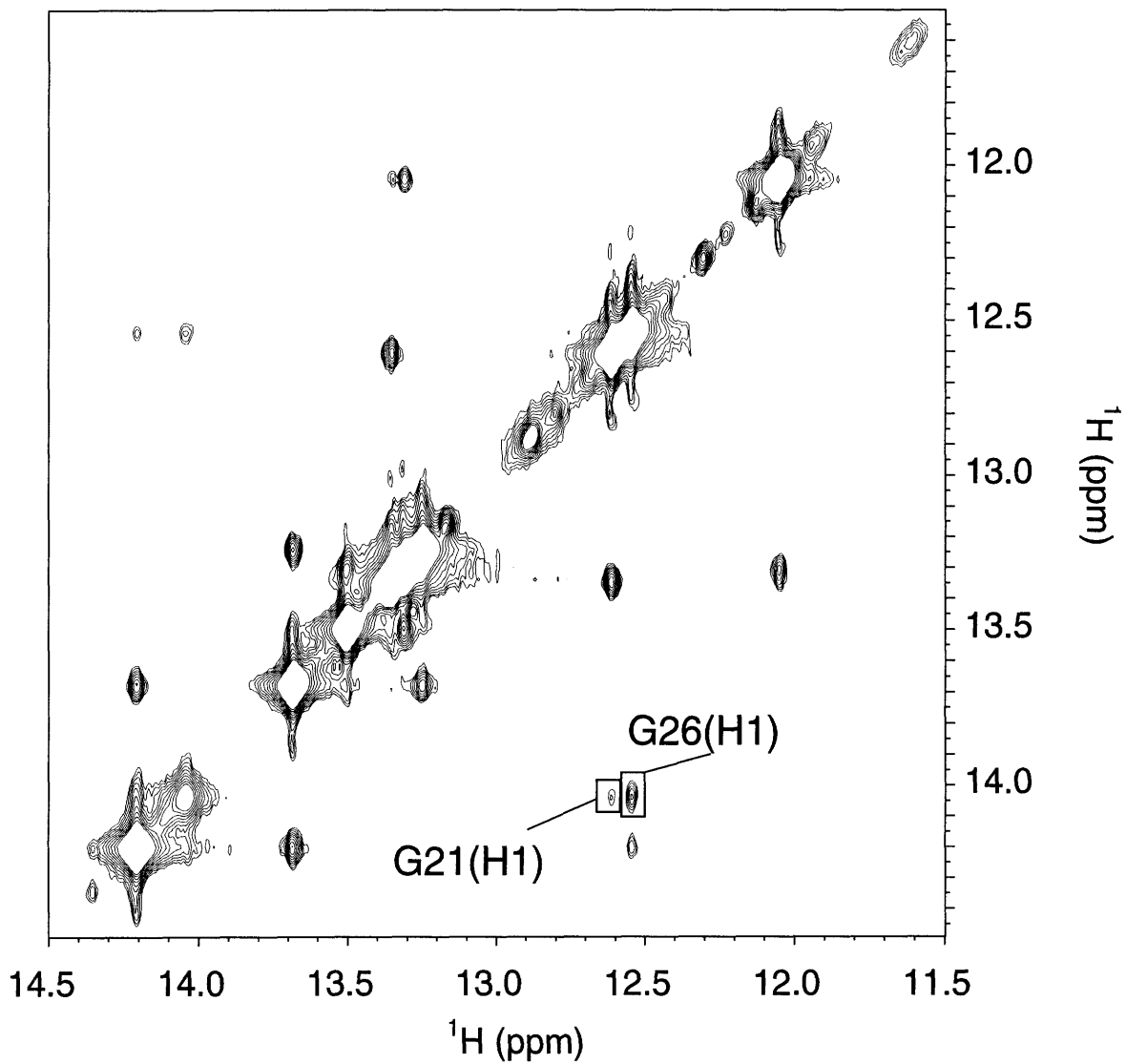
Figure 2.2: The imino-imino region of a 75ms 750 MHz ^{1}H-^{1}H NOESY in 90% H_2O/10%^2H_2O at 5°C indicates that the U40 imino proton is observable and gives stacking NOEs to the G26 and G21 imino protons.
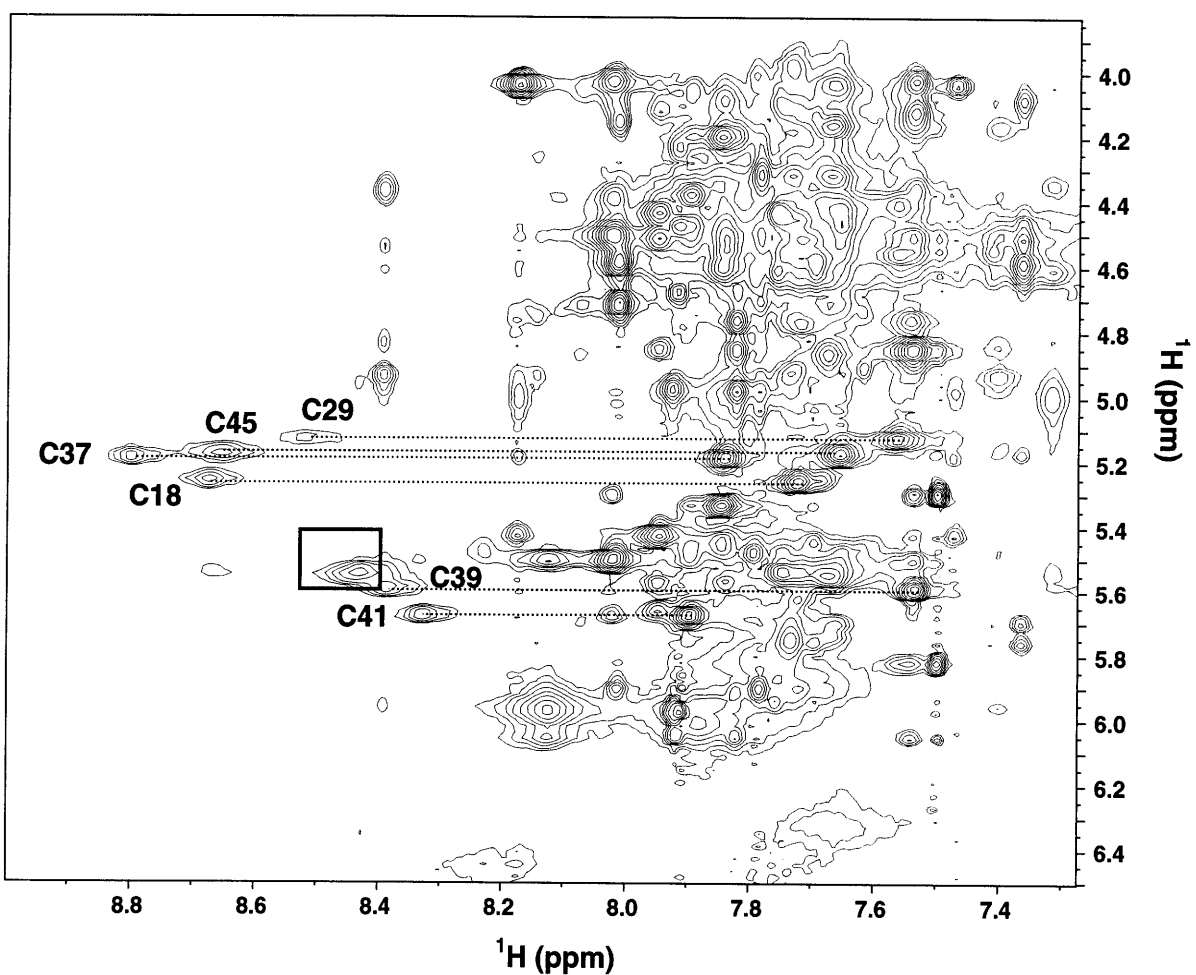
Figure 2.3: A 40 ms 750 MHz 5°C NOESY in H_2O highlights an alternative connection between the exchangeable and non-exchangeable spectra. The lines connect cytidine H42 amino proton resonances with the H5-H6 crosspeaks from the same nucleotide. Every base paired cytidine is observed and resolved except for the C19 and the terminal C46. The C19 H42 is highlighted by the box but the H5-H6 crosspeak is overlapped.

between the upper and lower stems.

The standard approach to begin correlating the exchangeable and non-exchangeable spectra is to identify the strong U(H3)-A(H2) NOE crosspeaks. A second pathway is available by looking for NOEs between the cytidine amino protons to their own H5s and H6s as illustrated in Figure 2.3. The good spectral dispersion of both the H5 and downfield amino proton resonances helps the identification of many of these crosspeaks. Because these NOEs are observable even at really short mix time NOESY experiments, fewer intervening and deceiving crosspeaks are observed. Each additional assignment pathway helps solve the puzzle as the different correlations will have different dispersion and give additional information.

The proposed base triple model of Puglisi et al. predicts that the U23 imino proton is hydrogen bonded and thus should be observable (Puglisi et al., 1992). However, the HIV-1 TAR-argininamide NMR studies found no evidence for this resonance. For these HIV-2 TAR studies, temperatures ranging from 0°C to 35°C, NaCl concentrations from 5 - 100 mM, and magnetic field strengths from 11.4 to 17.5 T were tested to try to find this missing imino resonance without any success. The 5°C 750 MHz ^{15}N-FHSQC of Figure 2.4 shows only the three Watson-Crick base paired uridine iminos. Thus, no overlapping hidden uridine imino resonances appear to be present that could not be observed in the 1D imino spectrum.

Of the four adenines in HIV-2 TAR, a single adenine amino crosspeak is observed in the ^{15}N-HSQC even at 750 MHz as shown in Figure 2.5. This adenine amino resonance is only observable at temperatures less than 10 °C. Assignment of the A27 amino protons was obtained from a 150 ms mixing time 750 MHz 3D NOESY-HSQC experiment at 5°C, shown in Figure 2.6. A strong NOE to U38(H3) and NOEs consistent with stacking to the C39(H42/H41) and G26(H1) are observed. However, the two A27
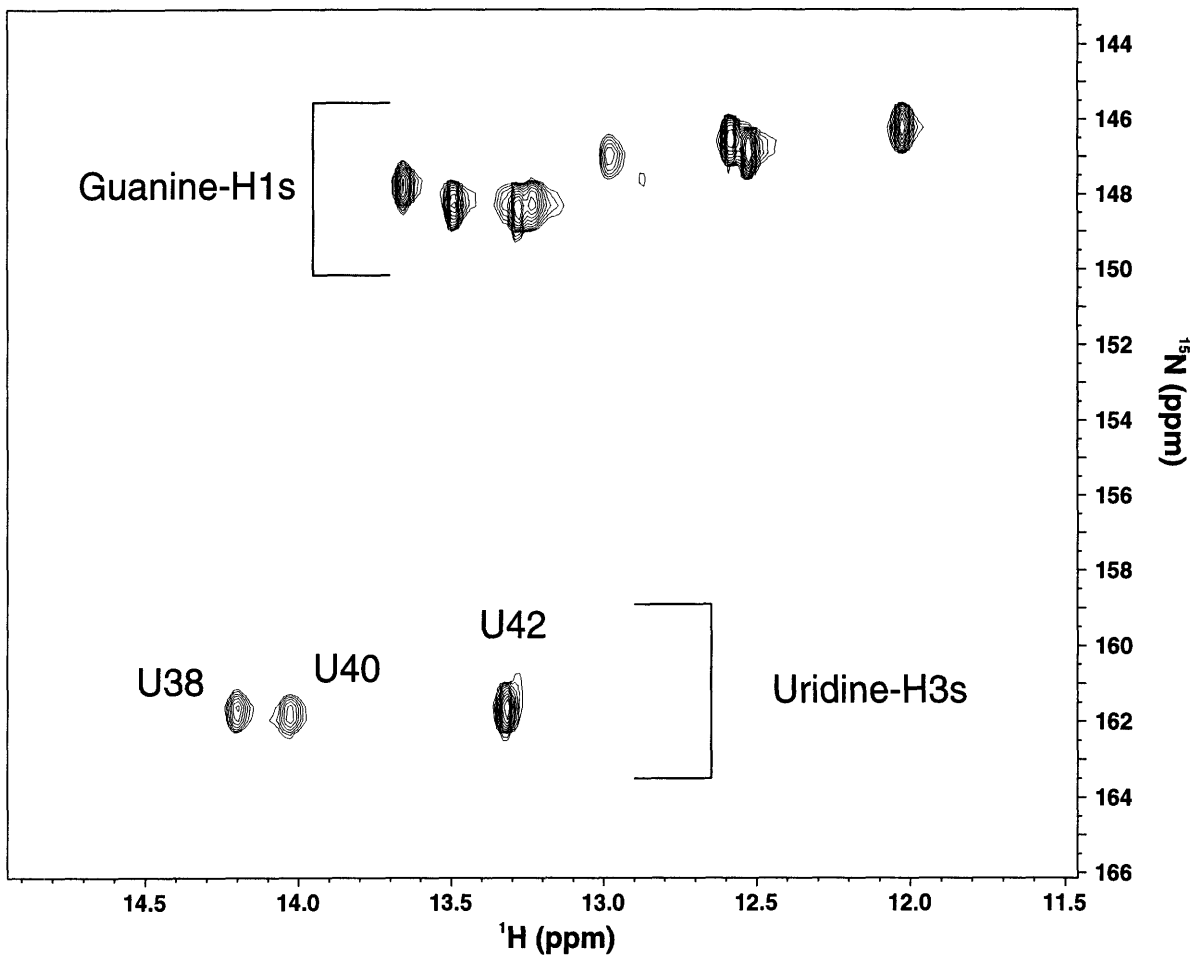
Figure 2.4: The imino region of a 750 MHz FHSQC at 5°C. No evidence of the U23 imino resonance is observed in the Uridine-H3 region. The nitrogen imino resonances are distinctly resolved by nucleotide type.
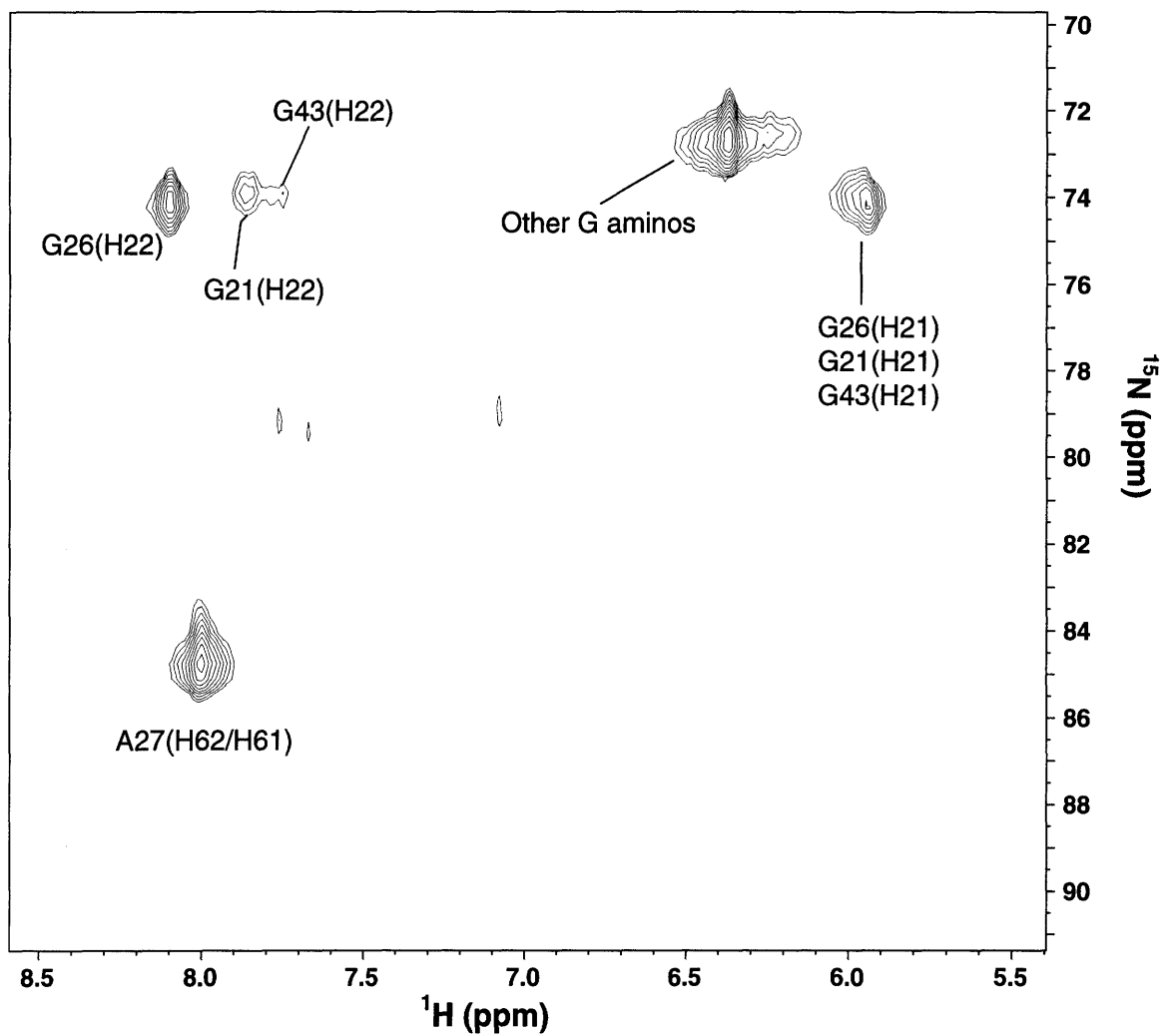
Figure 2.5: ^{15}N FHSQC at 5°C at 750 MHz of the HIV-2 TAR argininamide complex. Only one adenine amino nitrogen resonance is observable and the proton resonances have degenerate chemical shifts. This is the A27 amino which is part of the proposed base triple. Three guanine amino resonances have two distinct proton chemical shifts with the intensity becoming weaker as the difference between the two proton chemical shifts decreases.

amino protons are apparently coincident, perhaps because they are in exchange due to rotation about the C-N bond. The observation of the A27 amino resonances suggests that the A27 amino group may be involved in additional interactions beyond Watson-Crick base pairing.

Three guanine amino nitrogen resonances were also observable as shown in Figure 2.5. In the absence of $^{13}C/^{15}N$ double labeled samples, the only way to assign these resonances is by NOE analysis in the 750 MHz 3D NOESY-HSQC. Why the G26, G21, and G43 are the only guanosine amino protons which are exchanging in the right regime to be observed as distinct proton resonances remains a mystery. Because the argininamide interacts in the major groove of TAR, it is not expected to interact and possibly effect the observation of the guanine amino proton resonances. Interestingly, as the difference in chemical shift between the H22 and H21 amino proton resonances decreases, the intensity of the peaks is also reduced. This is expected as the two proton peaks move from slow exchange and enter the intermediate exchange regime. The other eight guanine amino resonances remain in fast exchange in one large crosspeak at ~6.3-6.4 ppm. As discussed in Chapter 1, guanine amino protons give a number of extremely useful long-range cross-strand NOEs, similar to adenine H2 protons. These NOEs are used to assign the guanine amino resonances as shown in the 750 MHz 3D ^{15}N-NOESY-HSQC in Figure 2.6. Because a $^{15}N/^{13}C$ sample was not prepared, the only way to assign these exchangeable ^{15}N and ^1H resonances is by NOESY based assignments. The danger of depending on the structure as a guide to assign the NOEs was minimized as other information suggested that the upper and lower stems are A-form. In addition, the observed NOEs made sense with this structure and as Chapter 3 discusses were consistent with all the other observed NOEs.

2.4 Assignment of Non-Exchangeable Resonances

2.4.1 Homonuclear spectroscopy
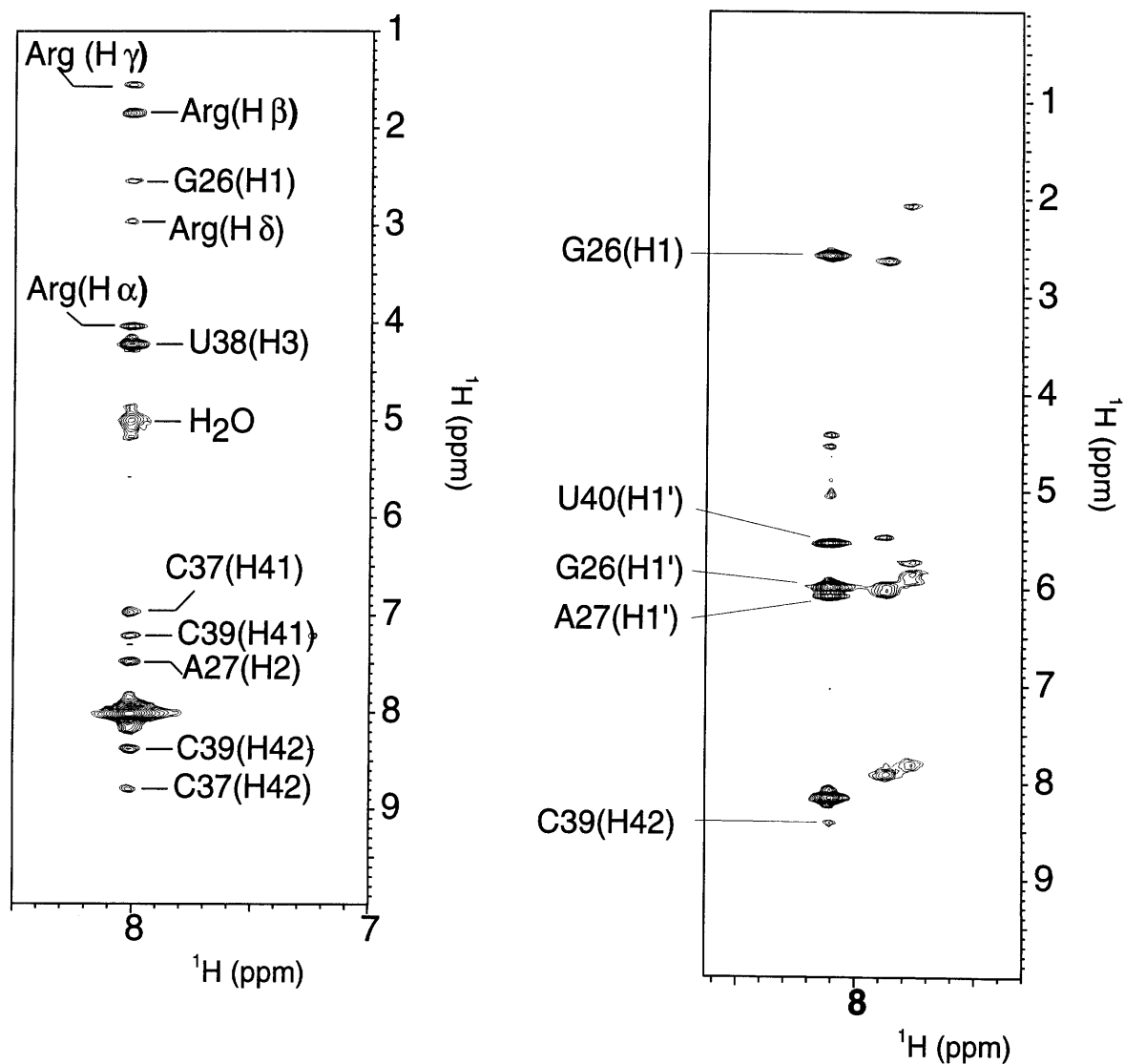
Figure 2.6: Two slices from a 150 ms ^{15}N-NOESY-HSQC at 750 MHz in 90% H$_2$O/10% ^{2}H$_2$O at 5°C. These purine amino resonances were assigned by these NOEs. The G26(H22) proton gives a number of useful cross-strand NOEs.

Non-exchangeable proton resonances were assigned from ^{1}H-^{1}H 2D NOESY, COSY, and TOCSY spectra, their 3D ^{13}C-resolved counterparts, and a 4D HMQC-NOESY-HSQC experiment (Nikonowicz & Pardi, 1992b). These multidimensional experiments were absolutely necessary to assign all the resonances because of the extreme overlap that the HIV-2 TAR argininamide complex demonstrated. They also allowed the identification of a very large distance restraint data set.

The sequential assignment pathway in the aromatic-anomeric regions is highlighted in the NOESY in 99.996% ^{2}H$_2$O of Figure 2.7. Although some of the A-helical regions can be assigned in this spectrum, the key resonances in the bulge region cannot be identified. (Note that this high quality 750 MHz spectra was not even available when this project began.) In particular, the U23(H1') - G26(H8) NOE, which was a key restraint observed in the HIV 1 TAR - argininamide complex, cannot be identified in the ^{1}H-^{1}H NOESY spectrum due to spectral overlap. For additional sequential assignments, the aromatic-aromatic region displays excellent dispersion over a 1.6 ppm range. Figure 2.8 shows a number of interesting aromatic-aromatic NOEs in the bulge region and in the TAR loop, including a stacking NOE between the G26(H8) and A22(H2). These types of crosspeaks are generally observed at longer mixing times and may be influenced by spin diffusion because of the strong H8/H6-H2' internucleotide NOEs. Thus, even though these NOEs are expected for A-form RNA (see Table 1.1), careful analysis of shorter mixing times is required to determine whether any of the NOEs can be converted to restraints.

^{1}H-^{1}H DQF-COSY and TOCSY experiments provide through bond assignments of some pyrimidine H5-H6 and H1'-H2' spin systems. These experiments are limited to riboses with C2'-endo as the couplings are too small for the C3'-endo case. Longer TOCSY mix times allow ^{1}H-^{1}H magnetization transfer from the H1' to the H5'/H5". In conjunction with the H$_2$O NOESY data, significant progress towards complete assignments is usually possible with homonuclear data, especially with the excellent resolution and
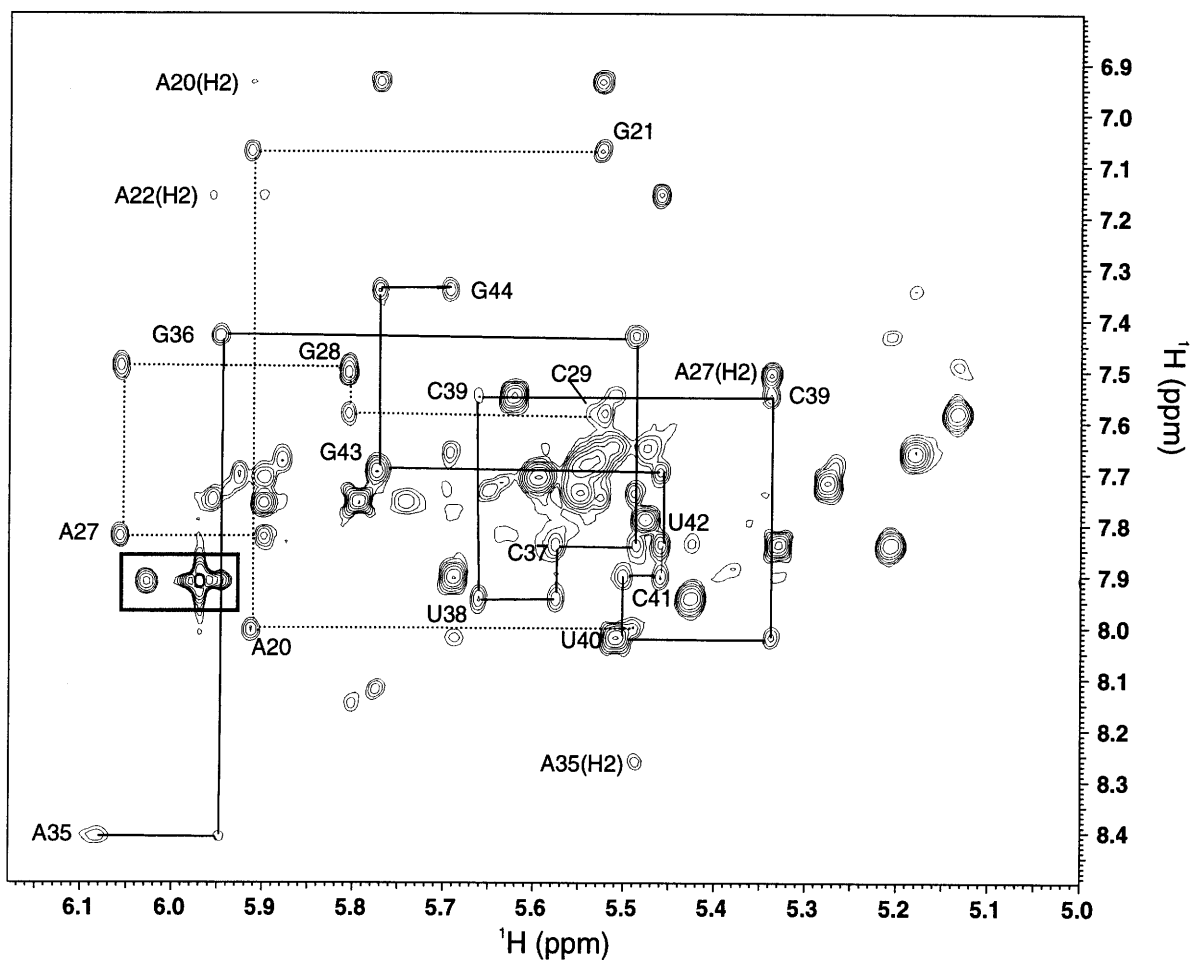
Figure 2.7: The aromatic-anomeric region of a 750 MHz NOESY spectrum (mix=100 ms) of the HIV-2 TAR-argininamide complex in 99.996% ^2H_2O. The sequential assignment pathway is highlighted by dashed and solid lines for nucleotides G21-C29 and A35-G44 respectively. The numbers indicate intranucleotide H8/H6 to H1' NOEs. The U23 resonances are overlapped and are not indicated along with the C17-C18, A22, G26, C30-G34, and C46 resonances. The U25 H6-H5 and H6-H1' crosspeaks are marked by the rectangle.
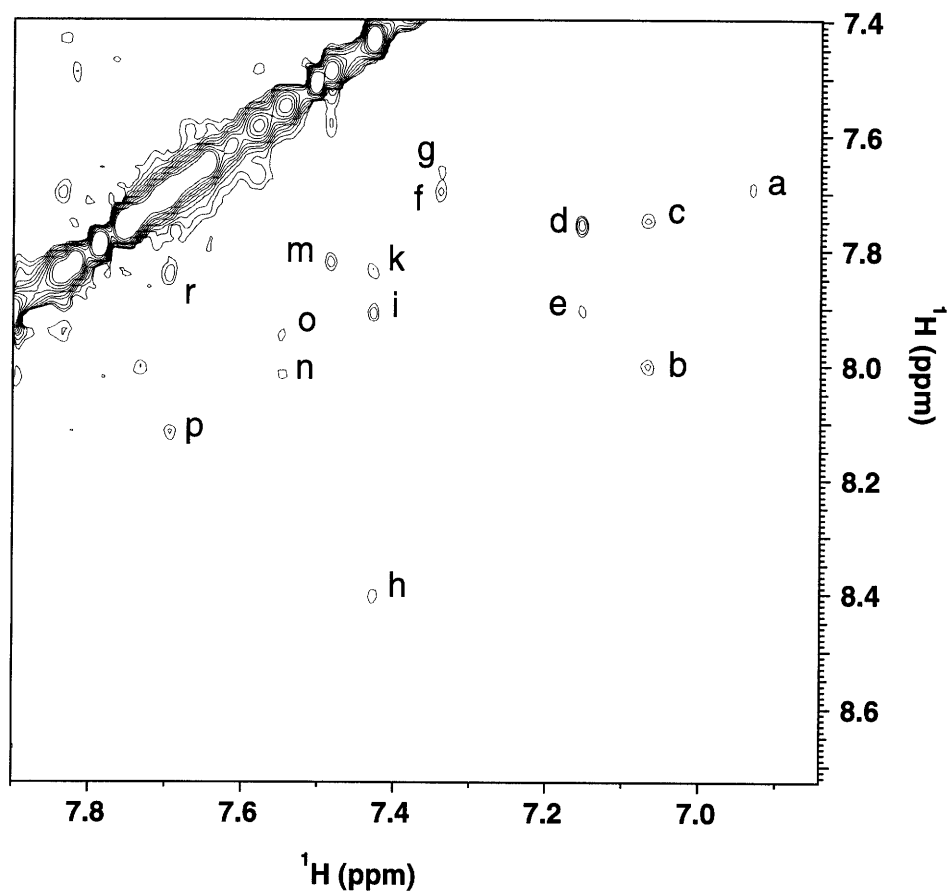
Figure 2.8: Aromatic-aromatic NOEs in a 750 MHz 100 ms ^2H_2O NOESY at 25°C.

a) A20(H2)-C19(H6)   b) G21(H8)-A20(H8)   c) G21(H8)-A22(H8)

d) A22(H2)-G26(H8)   e) A22(H2)-C41(H6)   f) G44(H8)-G43(H8)

g) G44(H8)-C45(H6)   h) G36(H8)-A35(H8)   i) G36(H8)-G34(H8)

k) G36(H8)-C37(H6)   m) G28(H8)-A27(H8)   n) C39(H6)-U40(H6)

o) C39(H6)-U38(H6)   p) G17(H8)-G16(H8)   r) G43(H8)-U432(H6)

signal to noise available at 750 MHz. This analysis does not preclude the use of ^{13}C

spectroscopy to make unambiguous assignments, but some interesting information about

the RNA structure may be deduced long before the first ^{13}C HSQC is acquired.

2.4.2 ^{13}C NMR assignments

To make unambiguous assignments, a ^{13}C isotopically labeled sample proved to be

extremely useful and necessary for HIV-2 TAR. The ribose region of a constant time

HSQC with the constant time interval set to $1/J_{CC}$ which differentiates the C1' and C5'

from the C4', C2', and C3' is shown in Figure 2.9. The J-coupling modulation during the

constant time period causes a sign alteration due to the number of carbon-carbon bonds.

Thus, because the C1' and C5' have only one carbon-carbon bond they will have the

opposite sign of the other ribose carbons which have two carbon-carbon bonds. RNA

carbons usually resonate in very distinct regions allowing for initial assignments to be made

by simply monitoring the carbon chemical shifts, but significantly shifted resonances have

been found in a number of cases, most notably the tetraloop structures (Heus & Pardi,

1991). In HIV-2 TAR, only the G26(C5') and the C46(C3') is significantly shifted as

highlighted in Figure 2.9 and no H1' or C1' resonances are significantly shifted into any

unusual regions. A similar analysis was performed in the aromatic region by a ^{13}C-CT-

HSQC spectrum to differentiate the pyrimidine C6's from the purine C8's and adenine

C2's. A separate CT-HSQC experiment is usually required to observe the aromatic and

ribose regions because the aromatic carbons are far (>40 ppm) from the ribose region

preventing carbon 180's from exciting both regions and thus limiting the obtainable signal.

The resolution of constant time experiments is limited resolution as points can only

be collected up to the total constant time period. A non-constant HSQC can provide more

resolution in the carbon dimension and the pyrmidine and purine resonances can be

distinguished as only the C5 and C6 are split by a single carbon-carbon coupling. Figure

2.10 shows how almost every C6 and C8 is observable when 512 points are collected.

3D HCCH-COSY and HCCH-TOCSY experiments correlate the usually well-
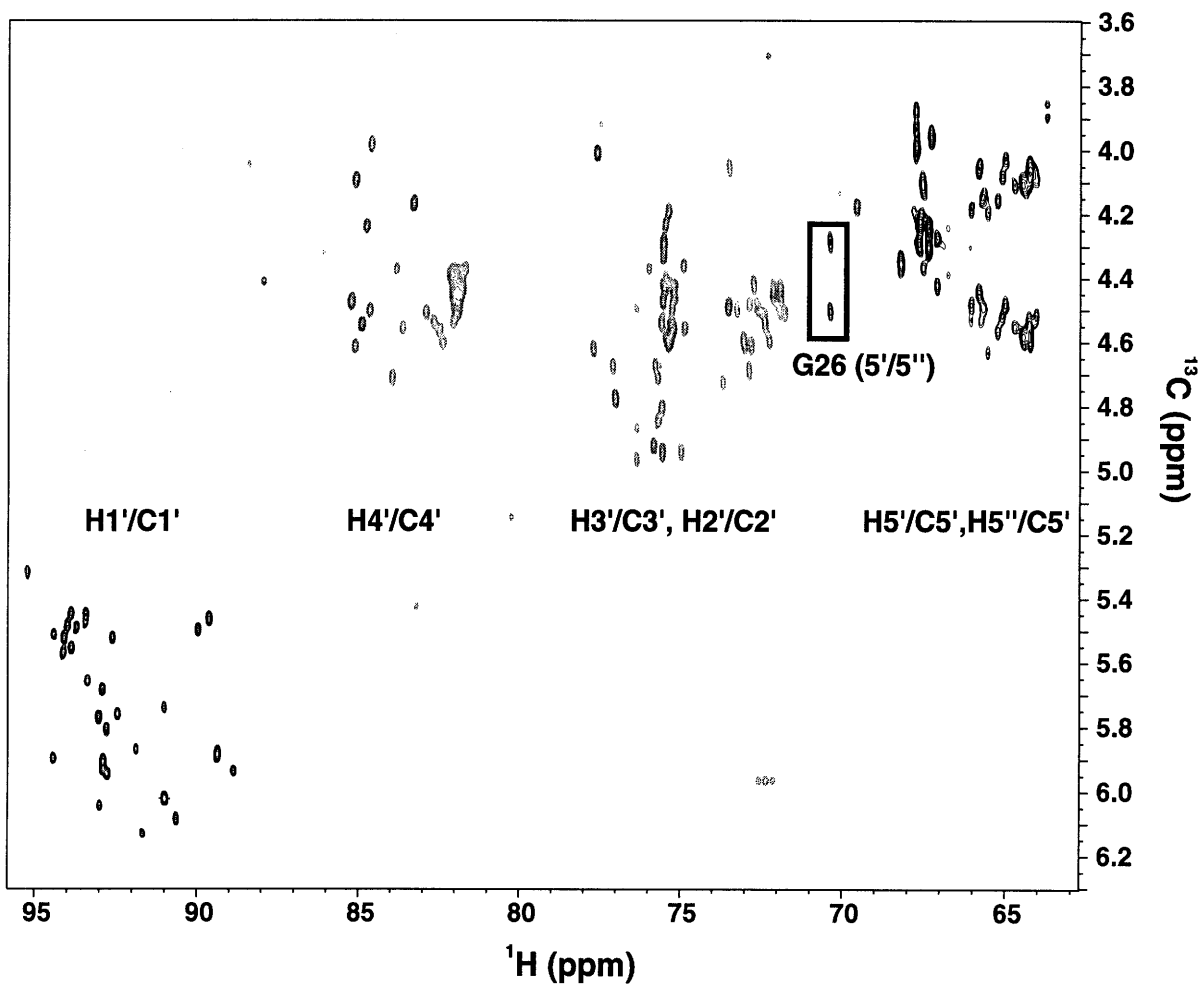
Figure 2.9: Constant time HSQC of the ribose region of HIV-2 TAR in the argininamide complex at 600 MHz. RNA ribose carbon resonances are generally in distinct chemical shift regions. The C4' and C3'/C2' regions are negative in this experiment because these carbons are attached to two other carbons. Notice the unusually shifted G26 C5' resonance marked by the box.
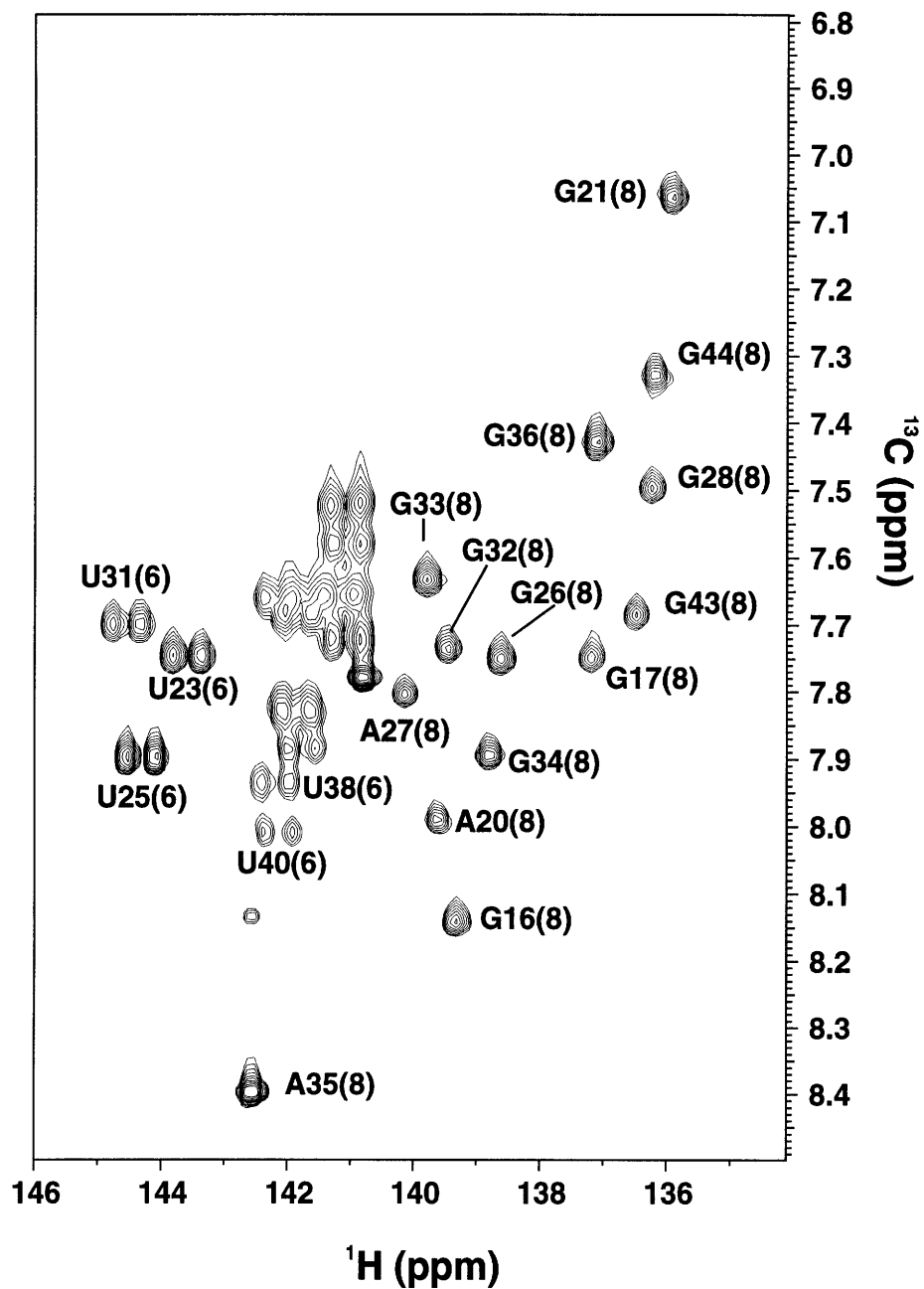
Figure 2.10: The aromatic region of a ^{13}C-HSQC shows the good dispersion of the purine resonances and the generally crowded pyrmidines except for a few non-helical resonances.

resolved H1'/C1' resonances with the rest of the ribose spin system (H2'->H5'/H5").

These experiments transfer magnetization by using the large one-bond, conformationally

independent, J_{HC} and J_{CC} couplings instead of the very small $^3J_{HH}$ couplings of the ribose

spin system. For HIV-2 TAR, these were the only unambiguous through bond

assignments that were made. A 20 ms, 3000 Hz spin lock time was sufficient to transfer

magnetization from the C1' to the C5' allowing almost all 150 resonances in the 30 riboses

to be unambiguously assigned. The HCCH-COSY identifies the various H1'-C1'-C2'-

H2' and H5-C5-C6-H6 spin systems. The other ribose HCCH spin systems (e.g. H5'-

C5'-C4'-H4') are difficult to resolve because the proton resonances do not have sufficient

spectral dispersion and thus will be severely overlapped. Once the different ribose spin

systems were identified, they can be sequentially linked by NOESY experiments.

2.4.3 ^{13}C NOESY experiments

In the HIV-2 TAR-argininamide complex, chemical shift overlap was a particularly

severe problem. NOEs between H1' protons in the 5.90-5.95 ppm range to H8/H6

protons at 7.70-7.75 ppm were very crowded and unambiguous assignments of all the

observed crosspeaks were not possible in 3D ^{13}C-NOESY-HSQC experiments at either

500 or 591 MHz. Figure 2.11a shows the crowded H1'/H5 region of the 2D ^{1}H-^{1}H

NOESY where many overlapping NOEs were observed in the 4.0-4.8 and the 7.70-7.75

ppm regions. The corresponding region from the U23 C1' plane of a 3D NOESY-HMQC

spectrum is shown in Figure 2.11b. Although the U23 C1' resonates at a distinct chemical

shift, resonance overlap still precludes identification of some of the NOEs in the 4.0-4.8

and 7.70-7.75 ppm regions. However, a 4D HMQC-NOESY-HSQC experiment (Vuister

et al., 1993) proved to be extremely valuable for resolving these peaks. One example of a

severe overlap problem is seen for peak **a** in the 3D plane of Figure 2.11b. Although this

may appear to be one NOE, the 4D spectrum showed that this cross peak is actually two

distinct NOEs, including one important long-range internucleotide NOE between U23(H1')

A) 2D NOESY

B) 3D NOESY-HMQC
U23(C1')
ω_1=89.45 ppm

C) 4D HMQC-NOESY-HSQC
U23(C1'/H1')
ω_4=89.45 ppm, ω_2=5.90 ppm

U23(5'/5")
G26(5',5")
U23(4')
U23(3')
U23(2')
G26(2'/3')
U23(1')
U23(6)
G26(8)
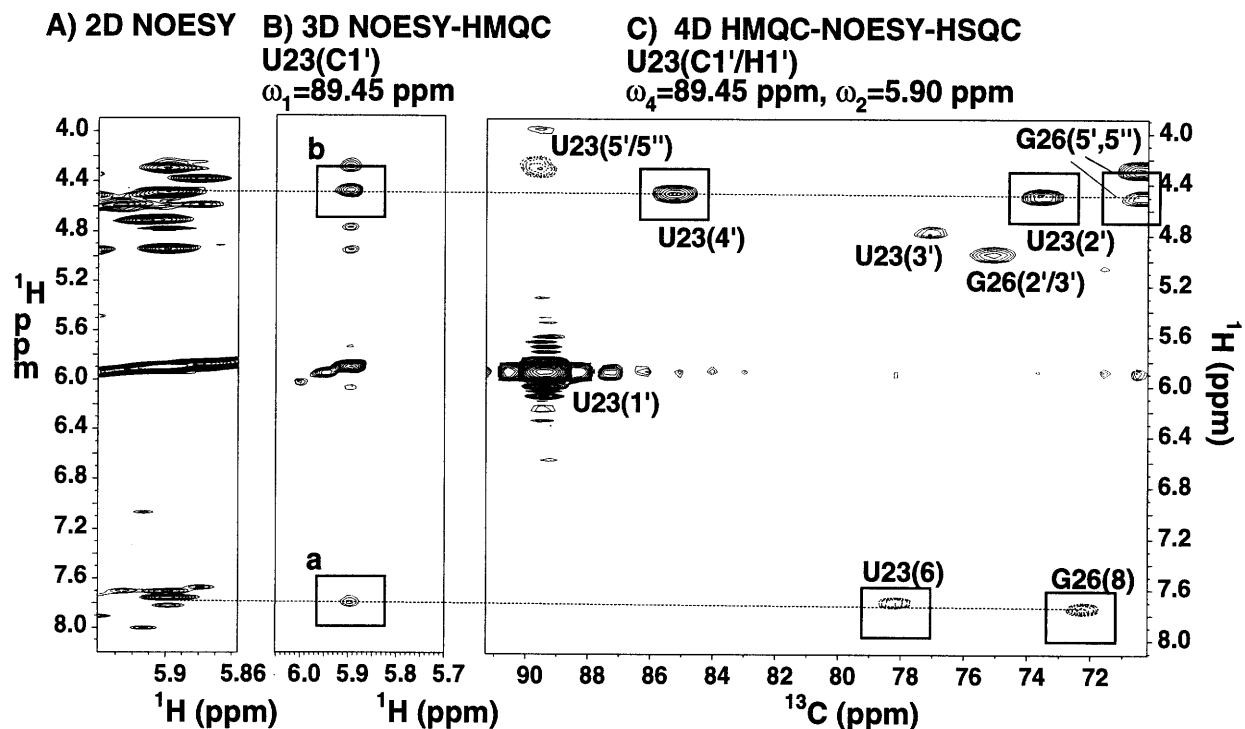
Figure 2.11: A. The spectral crowding in the H1' region is shown in a 100 ms 750 MHz 2D NOESYspectrum in ^2H_2O at 25°C. **B.** At the U23(C1') frequency in a slice from a 3D NOESY-HMQC shows significant clarification of the NOEs. **C.** However, some of these crosspeaks still contain more than one NOE as shown by a plane of a 600 MHz 90 ms 4D HMQC-NOESY-HSQC at the U23(H1'/C1') resonance. These are critical NOEs to the U23.

and G26(H8). A second example is illustrated by peak **b** in Figure 2.11b, which is resolved into three NOEs in the 4D plane. The U23(H4'), U23(H2') and one of the G26(H5'/H5") protons all have about the same chemical shift, but very distinct carbon chemical shifts allowing the unambiguous interpretation of these useful NOEs in the 4D spectrum.

The NOEs highlighted in Figure 2.11 could be not be assigned from the 3D NOESY-HMQC because of the extreme overlap of the H1' resonances. This is explicitly illustrated in Figure 2.12 which shows the G26(C8), G26(C1'), U23(C6), and U23(C1') ^{13}C slices from a 3D NOESY-HMQC. The carbon resonances resolved the overlapped aromatic proton resonances but not the anomeric proton crowding. Examination of the U23(C1') and the G26(C1') slices shows the same problem as the crosspeak now suffers from the aromatic proton resonances overlapping. In none of the slices, can the crosspeak at approximately 5.90 - 7.73 ppm be unambiguously identified. Without the 4D spectrum, one may think that theses NOEs are simply the intranucleotide aromatic-anomeric NOEs and no long range NOEs between G26 and U23 exist. It is even difficult to determine that these crosspeaks are particularly bigger than other observed peaks which would suggest an additional resonance is overlapped with the expected intranucleotide NOE. The 3D spectrum still relies on proton chemical shift dispersion in one of the slices to identify the crosspeaks. However, we had suspicions from the previous HIV-1 TAR NMR studies that an NOE between the U23(H1') and the G26(H8) exists and therefore performed the 4D experiment to unambiguously identify this critical internucleotide NOE. This incredible overlap presents a serious warning flag for RNA NMR; severe chemical shift overlap can prevent the identification of key NOEs. This scenario places more emphasis on performing more selective labeling strategies and/or 4D NMR to assign every possible NOE. In addition, it may be prudent to build initial models and determine what other NOEs should be observed and then try to find them in the spectra.

The potential of using a 4D NOESY spectrum for stereospecfic H5'/H5"
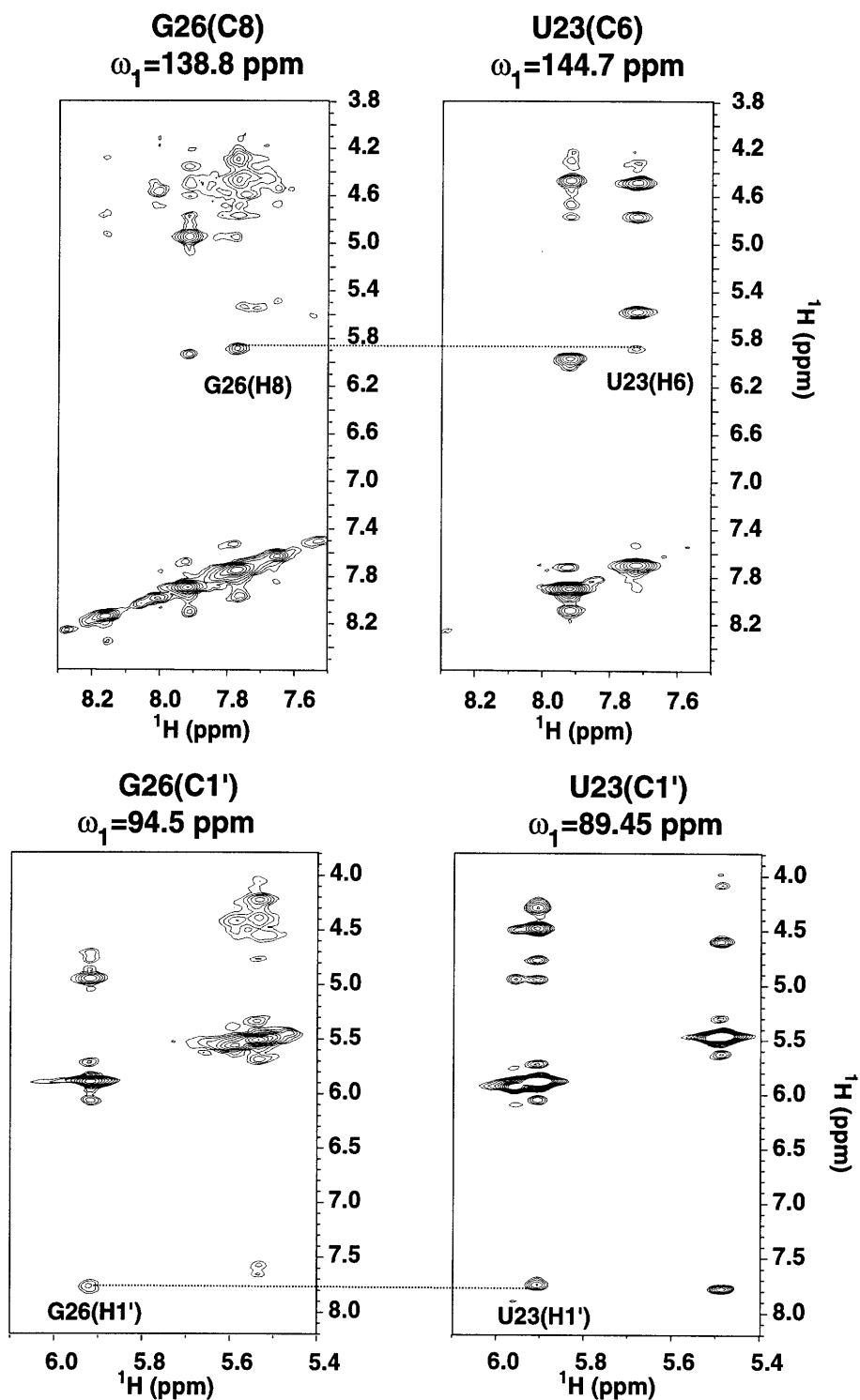
Figure 2.12: The proton spectral crowding in 3D the NOESY-HMQC prevents assignment of the crosspeak at ~7.73 - 5.91 ppm in any of these carbon slices. Looking at either the aromatic slices or the C1' slices suffers from severe proton chemical shift overlap problems preventing assignment of these NOEs. Is there a NOE between G26(H8) and U23(H1') or not?

A) G25(H5'/C5') ω_4=70.92 ppm ω_2=4.50 ppm
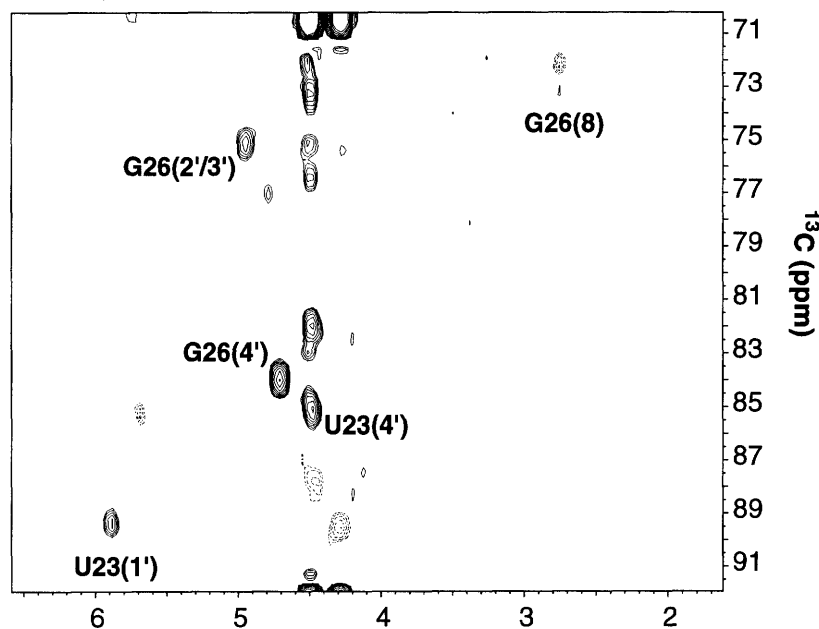
B) G25(H5"/C5') ω_4=70.92 ppm ω_2=4.28 ppm
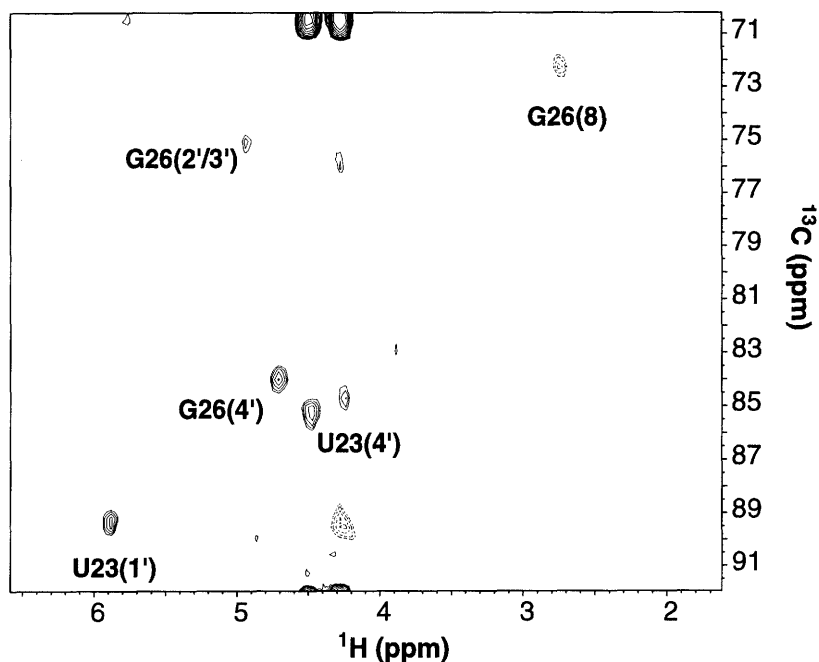
Figure 2.13: Two slices from the 4D HMQC-NOESY-HMQC.
A) The G26(H5'/C5') slice. **B)** The G26(H5"/C5') slice. At 90 ms, a slight difference in intensity between the G26(4') and the G26(5') and the G26(5") is apparent, suggesting the possiblity of stereo assignments, especially with shorter mixing times. The dashed crosspeaks are negative.

assignments is shown in Figure 2.13. The H4'-H5'/H5" NOEs are resolved and their intensities in a shorter mixing time experiment should vary with the γ torsion. The problem with this strategy is that by going through two proton - carbon magnetization transfers, the magnetization could be differentially affected by dynamics. Nonetheless, in this 90 ms experiment, many additional NOEs to the H5' could be assigned. The 4D NOESY spectrum permitted the identification of 10 of 19 possible 5'/5"-2' and 9 of 19 3'-2' internucleotide NOEs in the A-helical regions. A number of additional interesting internucleotide NOEs were identified in the bulge region, including G26(H4')-U23(H1') and G26(H8)-U23(H4'). The NOEs in the bulge region observed in previous NMR studies of the HIV-1 TAR-argininamide complex (Aboul-ela et al., 1995; Puglisi et al., 1992) were also observed here, indicating that the HIV-1 and HIV-2 TARs are binding in similar conformations. Table 2.3 lists all the NOEs observed in the bulge region of the HIV-2 TAR - argininamide complex.

2.4.4 ^{31}P NMR spectroscopy

Phosphor is an intriguing nucleus as it is in a prime position to make ribose-ribose sequential assignments. However, because ^{31}P relaxes very fast in macromolecules along with the small C-P and H-P J couplings ($\leq$10-11 Hz), acquiring sufficient signal becomes very difficult for these experiments. In addition, ^{31}P nuclei are very sensitive to motions such as conformational exchange which is very common in RNA and may cause the signals to decay even more rapidly than predicted for the size of the RNA. Luckily, the ^{31}P spectra had fairly strong signal for the HIV-2 TAR - argininamide complex allowing for some sequential assignments to be verified by reliable through-bond experiments, as shown in Figure 2.14. The HP-HSQC in Figure 2.15 shows that many of the non-A-form ^{31}P resonances are resolved but that the A-form crosspeaks are packed together in one big happy family.

The H5'/H5"-P and H3'-P J-couplings can be measured in a HP-COSY or HP-

Table 2.3: NOEs in the bulge region.

A22	H8	-G21	H1'	w	G26	H1'	-A22	H2	w
		-G21	H2'	s					
		-G21	H3'	w					
		-G21	H8	w	G26	H2'/H3'	-U23	H1'	m
		-A22	H1'	w			-G26	H1'	m
		-A22	H2'	w					
		-A22	H3'	m	G26	H4'	-U23	H1'	w
		-A22	H4'	w			-G26	H1'	m
							-G26	H2'/H3'	m
A22	H2	-G21	H1	w	G26	H5'/H5"	-U23	H1'	w
		-G21	H22	v.w.			-U23	H4'	w
		-A22	H1'	w			-G26	H1'	w
		-A22	H2'	w			-G26	H2'/H3'	w
							-G26	H4'	s
A22	H1'	-G21	H2'	w					
					A27	H8	-G26	H1'	w
U23	H6	-U23	H1'	m			-G26	H2'/H3'	s
		-U23	H2'	m			-G26	H8	w
		-U23	H3'	m			-A27	H1'	w
		-U23	H4'	m			-A27	H2'	w
							-A27	H3'	m
U23	H5	-U23	H2'	w			-A27	H5'/H5"	w
U23	H5'/5"	-A22	H2'	w	A27	H2	-G26	H1	v.w.
U25	H6	-U25	H1'	m					
			H2'	s	A27	H61/H62	-G26	H1	v.w.
			H3'	m					
			H4'	w	A27	H1'	-G26	H2'/H3'	v.w.
			H5'/5"	w			-G26	H22	v.w.
							-G26	H21	v.w.
U25	H5	-U25	H2'	w					
U25	H3'	-U23	H4'	m	A27	H2'	-A27	H1'	s
U25	H1'	-U42	H1'	w					
					A27	H3'	-G26	H2'/H3'	v.w.
G26	H8	-A22	H2'	w			-A27	H1'	m
		-A22	H2	w			-A27	H2'	s
		-U23	H1'	w					
		-U23	H4'	w	A27	H4'	-A27	H1'	m
		-G26	H1'	w			-A27	H2'	w
		-G26	H2'/H3'	w			-A27	H3'	s
		-G26	H4'	w					
		-G26	H5'*	w	A27	H5'/H5"	-G26	H2'/H3'	v.w.
							-A27	H3'	w
G26	H1	-A22	H2	v.w.			-A27	H4'	m

s = strong NOE, m = medium NOE, w = weak NOE, v.w. = very weak NOE

A)

6.4   6.0   5.6   5.2   4.8   4.4   4.0   3.6   3.2

^{1}H (ppm)

B)

7   6   5   4   3

^{1}H (ppm)

Figure 2.14: The first fid from the A) HP-HSQC and B) HCP experiments shows the signal obtained for the ~10 kDa HIV-2 TAR-argininamide complex.
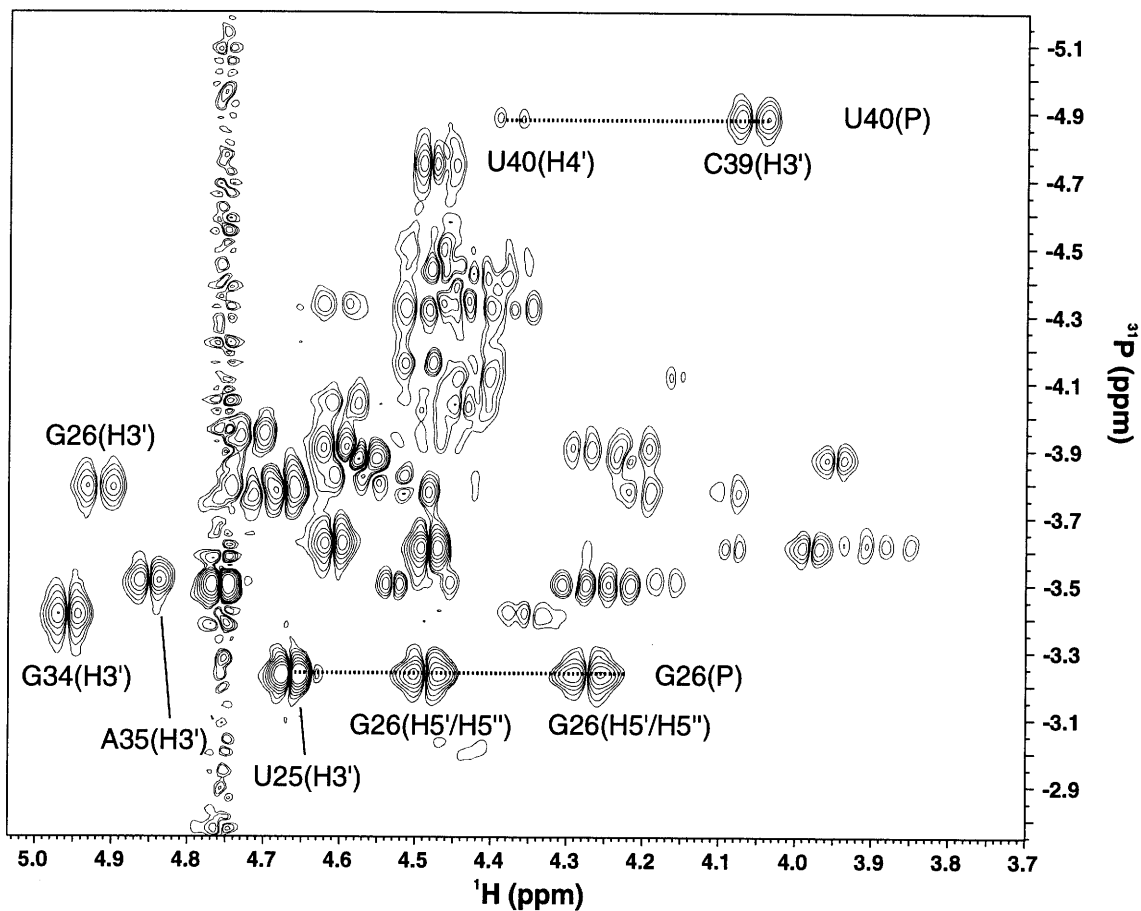
Figure 2.15: ^{1}H-^{31}P HSQC of HIV-2 TAR-argininamide complex at 591 MHz proton frequency at 25°C. Notice the two G26(H5'/H5") crosspeaks to the G26(P) whose appearance is not consistent with a single rotamer suggesting conformational averaging. Notice the non-helical resonances from U25, G26, G34, A35, and U40 are resolved.

HSQC experiment and provide restraints for the β and ε torsions, respectively. The HSQC

is a more sensitive experiment and suffers from less overlap because there are only two

antiphase peaks per HP spin system as opposed to four in a HP-COSY spectrum. For

HIV-2 TAR, the HP HSQC displayed evidence of mixed populations of backbone

conformers in the bulge region. For example, the G26 H5' and H5" both gave medium

couplings to the G26 phosphorus indicative of a mixture of gauche⁻ and gauche⁺

conformations as highlighted in Figure 2.15. As seen in Table 1.4, none of the three major

rotamers of the β torsion are consistent with the observation of cross peaks for both the

H5' and the H5" - P. Evidence for mixed conformers was also observed for the backbone

torsions of U23 and U25. Thus, no α, β, γ, δ, or ε torsion restraints were used in the

structure determination.

The HCP experiment is a series of INEPT transfers from ^{1}H to ^{13}C to ^{31}P and back

to ^{1}H (Heus et al., 1994; Marino et al., 1994). The HCP experiment links sequential

riboses as each C4' has a 8-10 Hz coupling to both its own and neighboring phosphorous

(Heus et al., 1994; Marino et al., 1994). The HCP spectrum's signal was sufficient to see

most of the expected crosspeaks (Figure 2.15). Figure 2.16 shows two representative ^{31}P

slices from the HCP experiment for some of the better resolved resonances. These slices

provide straightforward identification of C5', C3', and the two sequential C4' resonances

for all nucleotide conformations. On the other hand, the observation of the C2' crosspeak

suggests that the ε torsion is gauche⁻ for the U25 (Varani et al., 1995). The sequential

assignment pathway from the HCP experiment is illustrated by following the ^{13}C slices as

shown in Figure 2.17. This pathway depends on the poor C4' chemical shift resolution

(see the CT-HSQC of Figure 2.9). By combining the information from both the ^{31}P and

^{13}C slices, through-bond sequential assignments are possible for many residues. A more

powerful assignment experiment, the HCP- CCH TOCSY (Marino et al., 1995),
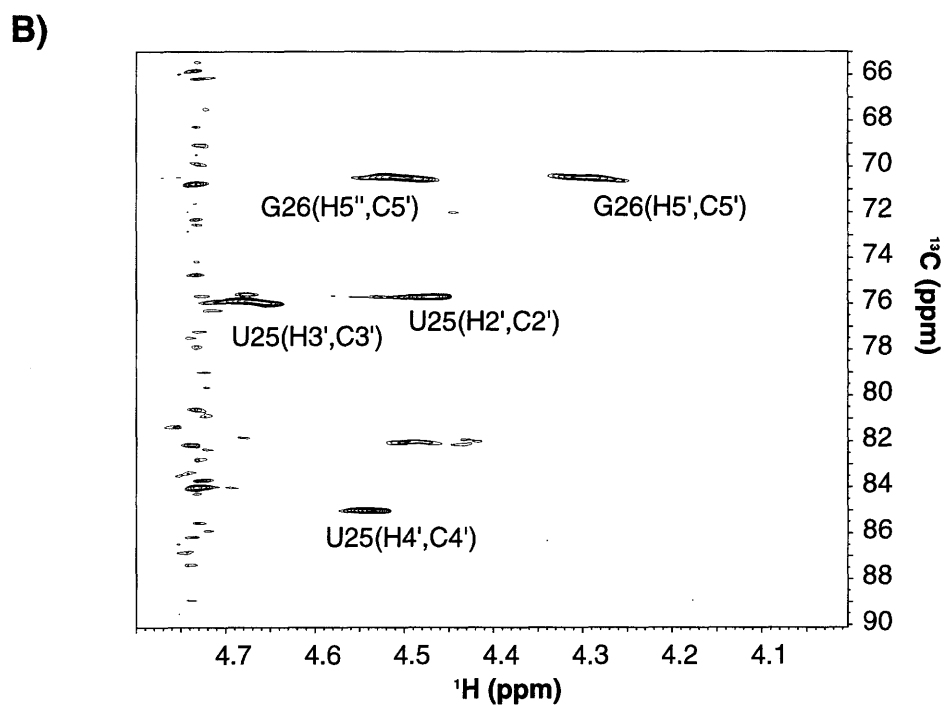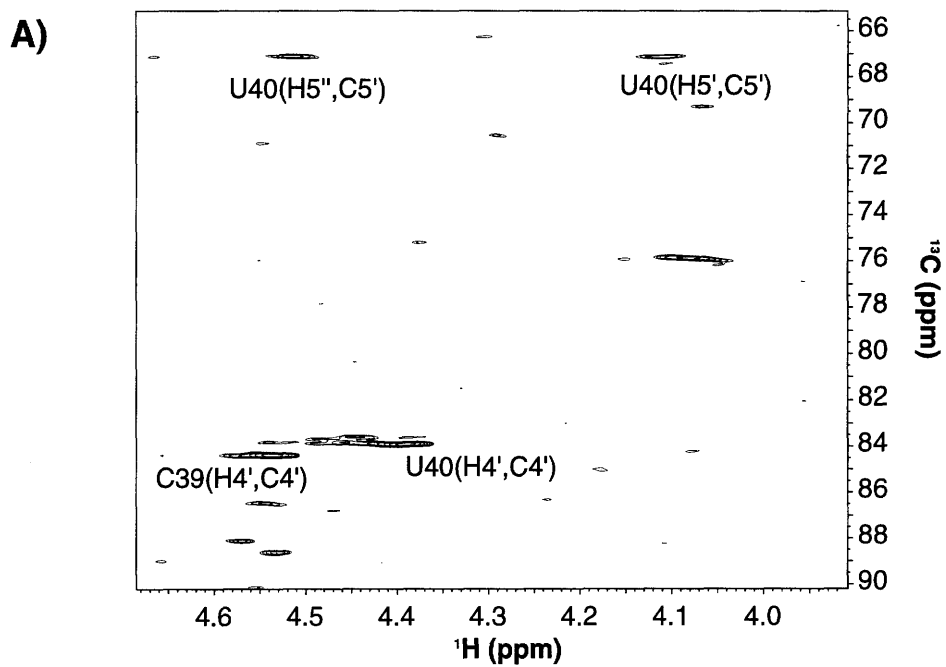
Figure 2.16: ^{31}P slices from the HCP experiment at 25°C. **A)** U40(P) at -4.92 ppm **B)** G26(P) at -3.25 ppm. The G26(H4',C4') crosspeak is weak and overlapped with the water peak at 4.75 ppm.
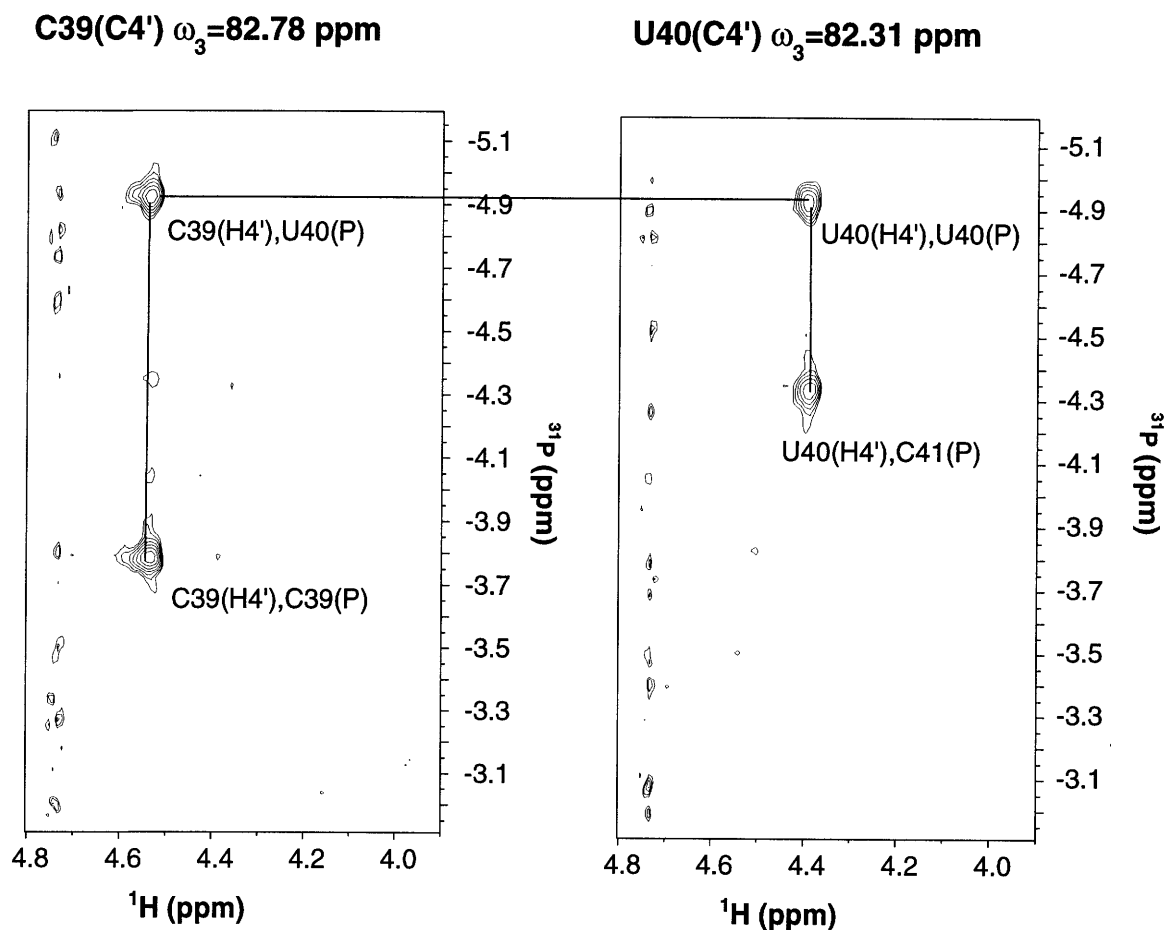
Figure 2.17: ^{13}C slices from the same HCP experiment as in Figure 2.16 showing the sequential assignment pathway linking the C39 and U40 residues. These resonances are well-resolved unlike most of the A-form residues.

which combines the HCP and HCCH TOCSY experiments failed to provide significant signal for the HIV-2 TAR - argininamide complex, probably because the magnetization cannot survive the demands of a ~20 ms DIPSI delay along with the reliance on the fast relaxing ^{31}P resonances .

Nearly complete assignments were obtained for the HIV-2 TAR - argininamide complex as shown in Table 2.4. Of the ^{13}C resonances, only the C19(H5'/H5'') and C37(H5'/H5'') protons could not be unambiguously assigned due to spectral overlap. Because of the limited ^{31}P chemical shift dispersion, only about half of the ^{31}P resonances could be assigned. In addition to the fiendish overlap of some of the bulge and loop resonances, many of the resonances from the upper and lower A-form stems are overlapped making their assignments difficult as well. Many of the A-form pyrimidines have degenerate chemical shifts even for all the different types of carbon and proton resonances, making assignments very difficult. These difficulties highlight the potential problems for NMR studies of even larger RNAs where chemical shift overlap, especially of the A-form regions will be poor at best.

2.4.5   Non-exchangeable nitrogen resonances

A number of nitrogens that do not have directly attached protons are also observable by monitoring their two bond coupling to protons in ^{15}N-HSQC experiments. Figure 2.18 shows the various purine signals accessible from a two bond HSQC. The N7, N1, and N9 regions give the most signal and are well-resolved as all 15 N7's can be counted in the 2D spectra. However, three N7-H8 crosspeaks could be assigned because of the overlapping H8 chemical shifts as listed in Table 2.5. A ^{13}C/^{15}N labeled sample could alleviate this problem by correlating the C1' and/or C8 resonances to the N7 and N9. For larger ^{13}C/^{15}N labeled RNAs, where chemical shift overlap becomes a major problem, using some of these other nitrogen resonances to gain dispersion may be useful.

An upfield shift of a nitrogen chemical shift may suggest a hydrogen bonding interaction at that position (Gaffney *et al.*, 1993). The A27(N7) is of particular interest in

Table 2.4. Proton, Nitrogen, Carbon, and Phosphor chemical shifts (ppm) for HIV-2 TAR RNA bound to argininamide.

Residue	1/3	2/4	8/6	2/5	1'	2'	3'	4'	5'/5"	31_P
G16	12.99		8.12	--	5.80	4.92	4.74	4.56	4.42,4.27	
	146.92		139.51		92.52	75.22	75.05	83.65	67.10	
G17	13.50		7.70	--	5.93	4.59	4.61	4.55	4.55,4.29	
	147.92		137.37			75.35	72.97	82.58	66.09	
C18	--	8.68,6.91	7.72	5.28	5.56	4.43	4.47	4.44	4.57,4.12	
		99.41	141.25	97.23	93.91	75.53	72.06	81.95	64.42	
C19	--	8.44,6.97	7.69	5.56	5.47	4.58	4.54		4.16	-4.06
		97.66	141.25	98.14	93.88	75.30	73.32			
A20	--		8.00	6.94	5.92	4.72	4.67	4.51	4.57,4.15	
			139.80	152.70	93.39	75.72	72.98	81.95	65.27	
G21	12.61	7.88,6.0	7.07	--	5.53	4.43	4.43	4.48	4.45,4.06	-3.81
	146.57	73.90	136.01		92.61	75.22	72.70	81.76	65.80	
A22	--		7.75	7.16	5.95	4.35	4.63	4.45	4.59,4.09	
			139.65	154.24	92.77	75.55	72.90	82.03	64.35	
U23		--	7.71	5.60	5.90	4.49	4.79	4.48	4.29,4.22	
			144.75	103.85	89.45	73.46	77.12	85.33	67.64	
U25		--	7.91	5.97	6.03	4.49	4.69	4.55	4.30,4.24	-3.95
			144.52	105.60	91.09	75.39	75.74	84.93	67.37	
G26	12.54	8.13,5.94	7.75	--	5.90	4.95	4.95	4.75	4.50,4.29	-3.25
	146.92	74.25	138.80		94.51	75.57	75.39	84.30	70.39	
A27	--	8.03	7.82	7.51	6.06	4.86	4.74	4.62	4.63,4.19	-3.82
		84.73	140.32	153.90	93.07	75.70	73.67	82.40	65.55	
G28	13.69		7.49	--	5.81	4.55	4.51	4.52	4.55,4.16	
	147.62		136.44		92.58	75.39	72.98	81.79	65.78	
C29	--	8.52,6.98	7.58	5.14	5.52	4.25	4.44	4.40	4.55,4.06	
		97.31	140.80	97.23	94.06	75.49	72.01	81.98	64.32	
C30	--		7.68	5.55	5.88	4.37	4.59	4.38	4.45,4.06	
			141.90	98.33	91.94	76.00	75.39	83.68	65.91	
U31		--	7.75	5.80	5.75	4.31	4.51	4.25	3.96	-4.13
			143.80	105.40	91.09	75.50	76.43	84.62	67.29	
G32			7.79	--	5.48	5.59	4.63	4.13	3.99	-3.90
			140.98		89.69	75.39	77.81	85.11	67.74	
G33			7.65	--	5.51	4.53	4.67	4.01	3.88	-3.64
			139.92		90.05	75.63	77.12	84.63	67.77	
G34			7.91	--	5.95	4.96	4.98	4.50	4.23,4.11	-3.64
			138.99		88.93	75.74	75.74	84.68	67.81	
A35	--		8.41	8.25	6.09	4.94	4.88	4.63	4.35	-3.79
			142.74	155.66	90.73	75.84	76.37	85.13	68.22	
G36	13.25		7.43	--	5.49	4.57	4.51	4.47	4.48,4.19	-3.45
	148.32		137.29		93.38	74.90	73.17	82.71	66.06	
C37	--	8.81,6.98	7.84	5.21	5.58	4.43	4.48	4.45	4.09	-3.54
		98.71	142.00	96.90	94.03	75.44	72.00	81.98		
U38	14.21		7.95	5.43	5.66	4.80	4.52	4.43	4.58,4.10	
	161.59		142.37	102.15	93.44	75.59	72.45	81.97	64.47	
C39		8.40,7.23	7.55	5.63	5.34	4.50	4.08	4.54	4.37,4.11	-3.77
		96.26	141.20	98.90	95.27	75.39	73.67	82.44	67.53	
U40	14.05		8.02	5.51	5.49	4.38	4.52	4.39	4.52, 4.11	-4.92
	161.94		142.34	103.25	93.64	74.93	71.74	82.03	64.08	
C41	--	8.34,7.12	7.90	5.69	5.44	4.20	4.47	4.40	4.55,4.09	-4.34
		98.36	142.30	97.35	93.84	75.40	71.95	81.85	64.28	
U42	13.35	--	7.84	5.33	5.46	4.56	4.52	4.40	4.57,4.08	
	161.59		142.2	103.5	92.45	75.28	72.39	82.07	64.34	
G43	12.05	7.75,5.81	7.70	--	5.77	4.59	4.59	4.49	4.51,4.14	
	146.22	73.90	136.67		93.05	75.45	73.04	82.11	65.75	
G44	13.31		7.34	--	5.69	4.47	4.52	4.46	4.51,4.08	
	148.32		136.38		92.76	75.05	72.63	81.76	65.05	
C45	--	8.67,7.00	7.66	5.18	5.53	4.30	5.53	4.39	4.56,4.09	-3.77
		99.41	141.20	97.10	93.98	75.50	71.94	81.80	64.39	
C46	--	8.56,7.04	7.69	5.55	5.77	4.02	4.17	4.20	4.48,4.03	-4.13
			141.90	98.33	93.01	77.61	69.58	83.31	65.02	

The errors in chemical are ±0.01 for proton and ±0.1 ppm for the other nuclei. The exchangeable protons and nitrogen chemical shifts are given for 5°C while all the non-exchangeable data were assigned at 25°C.
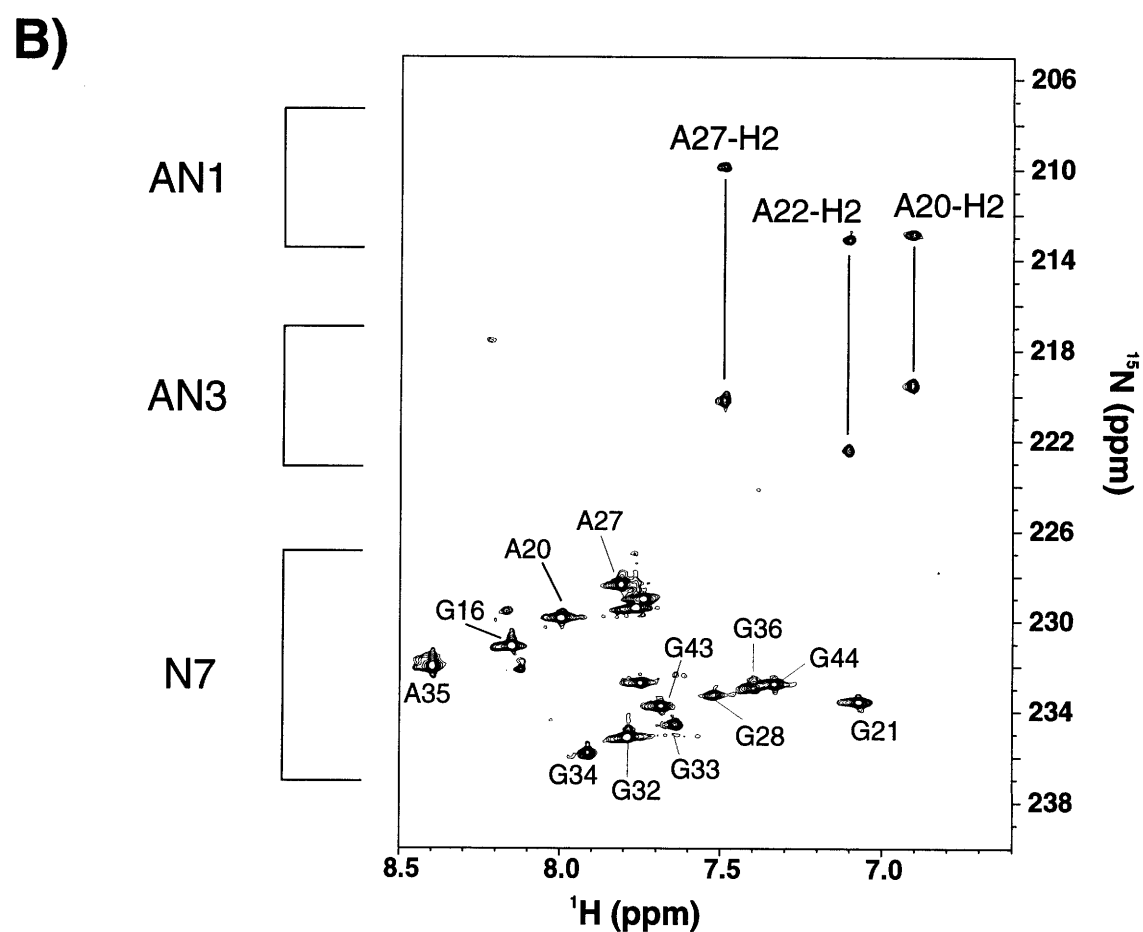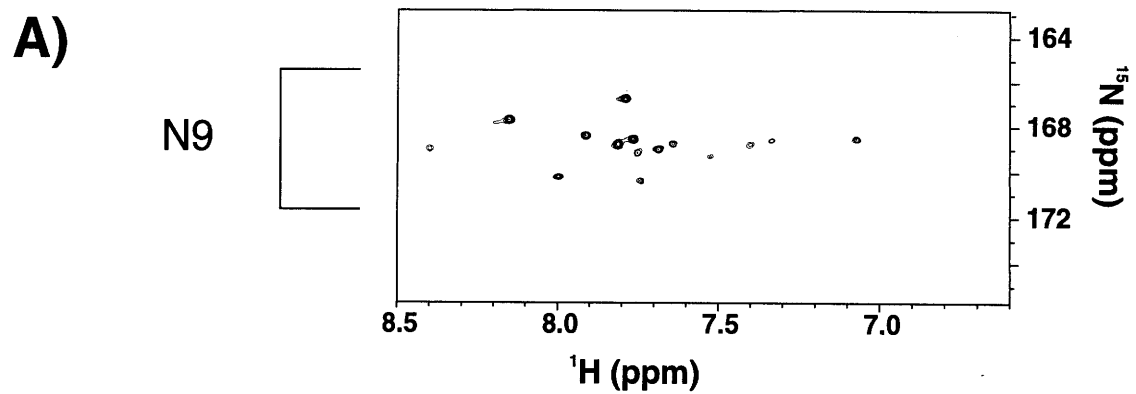
Figure 2.18: Portions of a 600 MHz 2-bond ^{1}H-^{15}N FHSQC of ^{15}N labeled HIV-2 TAR-argininamide at 15°C, showing A) H8-N9, B) AH2-N1, AH2-N3, and H8-N7 crosspeaks.

Table 2.5: Chemical Shifts of the purine non-exchangeable nitrogens in the free and bound forms of HIV-2 TAR at 15°C. The G17, A22, and G26 resonances have overlapped proton chemical shifts prohibiting their N7 and N9 assignment.

Purine	Free N7	Bound N7	Free N9	Bound N9
G16	231.42	231.08	167.86	167.80
G17	234.70			168.95
A20	229.64	229.93	170.16	170.16
G21	233.44	233.61	168.09	168.49
A22	230.45			
G26				
A27	228.38	228.32		168.78
G28	232.98	233.38		169.30
G32		235.16	166.71	166.83
G33		234.65	168.55	168.72
G34	235.74	235.80	168.78	168.38
A35	232	232		
G36	232.75	232.98		168.78
G43		233.84		
G44	232.46	232.75	168.55	168.61

Adenine	Free N1	Bound N1	Free N3	Bound N3
A20	212.79	212.84	219.34	219.52
A22	213.13	213.02	222.16	222.39
A27	211.00	209.91	220.44	220.21
A35*				

* One weak peak at the A35(H8) frequency was observed between the N1 and N3 chemical shift regions as seen in Figure 2.18 but it is unclear which nitrogen this peak represents.

this complex because of its predicted role in the A27-U38•U23 base triple. Table 2.5 lists the assignments in the free and bound form of HIV-2 TAR. However, even though the A27(H8) resonance changes upon addition of argininamide ($\Delta\delta$=-0.05 ppm), the N7 chemical shift does not shift significantly (Table 2.5). Since the complex is in fast exchange, different conformational exchange processes may be occurring that would effect the ^{1}H and ^{15}N chemical shifts differently because they are effected by motions on slightly different time scales (see Chapters 1 and 5). An alternative explanation, is that the base triple may not exist, though the evidence for it is very significant, as this thesis discusses extensively everywhere.

2.5 Ribose Torsion Measurements

To generate sugar pucker torsion restraints, J-couplings were estimated from H1'-H2' crosspeaks in a DQF-COSY spectrum. Strong crosspeaks were observed for U23 and U25 with J-coupling ~8 Hz, indicative of a C2'-endo conformation. A weak DQF-COSY crosspeak for A35 suggested a mixed conformation and was left unconstrained. The other loop nucleotides, C30-G34, were also constrained to C2'-endo while the remaining nucleotides were constrained to the C3'-endo A-form sugar pucker based on the absence of a DQF-COSY crosspeak and the presence of NOEs indicative of A-form geometry.

Various other schemes were attempted to try to measure the γ torsion including a variation of the protein backbone HAHB experiment. The problem with this strategy was that the H4' and the H5'/H5'' proton resonances are typically < 0.3 ppm from each other causing extensive overlap and preventing the measurement of any γ torsions. However, this experiment worked well for the H1'-H2' couplings where the protons are usually >0.5 ppm apart. Thus, this experiment may be a viable alternative to the DQF-COSY experiment for larger RNAs when the lines are very broad and limit the observable signal from the antiphase DQ-COSY crosspeaks.

2.6 Intermolecular NOEs

^{1}H-^{1}H NOESY spectra in ^{2}H$_2$O and H$_2$O of the HIV-2 TAR-argininamide complex permitted assignment of a large number of intermolecular NOEs including intermolecular NOEs to A22, U23, G26, A27 and C39 as listed in Table 2.6. Of particular note is that the same strong U23(H5) to Arg(Hδ) and Arg(Hγ) that was observed in the HIV-1 TAR-argininamide and TAR-peptide studies is observed indicating that the HIV-2 TAR is binding argininamide in a similar manner. Even in a 500 MHz 35 ms ^{2}H$_2$O NOESY spectrum both the Arg(Hδ) and Arg(Hγ) crosspeaks to the U23(H5) were observed suggesting that both of these observed crosspeaks are real and not due to spin diffusion. Because the complex is in fast exchange, the intermolecular crosspeaks were carefully analyzed to determine which are real NOEs and which may be due to spin diffusion and/or the conformational mobility of the complex. Only the strongest intermolecular crosspeaks in a short 40 ms 750 MHz ^{1}H-^{1}H H$_2$O and/or a 50 ms ^{2}H$_2$O NOESY spectra were used as restraints in the structure determination. Figure 2.17 shows traces from the 40 ms 750 MHz ^{1}H-^{1}H H$_2$O NOESY spectrum illustrating the crosspeaks observed. The Hε showed a number of crosspeaks of which only a few are unambiguous NOEs. The RNA crosspeaks **e** and **c** are near very strong crosspeaks in the 2D spectrum while **f** could be spin diffusion through Hδ. **o** and **q** are assigned to A22(H3') and A22(H2') NOEs. Hδ and Hγ give crosspeaks to these A22 ribose protons only very weakly in a 100ms 750 MHz ^{2}H$_2$O NOESY spectrum. The C39(H42) shows an interesting pattern of NOEs by giving crosspeaks to the Hβ and Hδ and a weaker crosspeak to the Hγ suggesting that the Hβ and Hδ crosspeaks are real and that the Hγ crosspeak may be spin diffusion influenced.

Peak **m** is a weak peak that could be A22(H1') but cannot be unambiguously assigned in the H$_2$O. The C37 amino gives a number of crosspeaks but they are all very weak and the
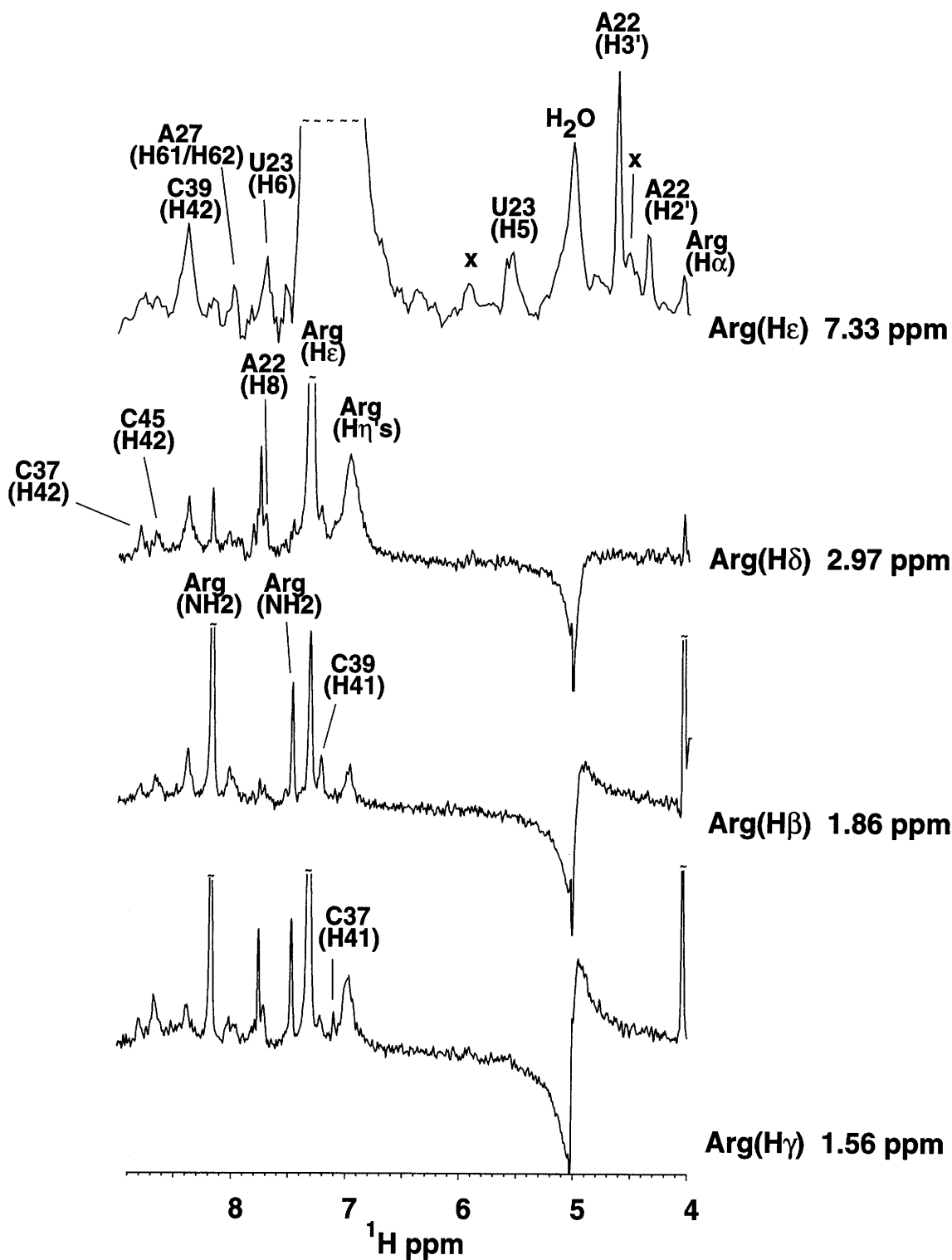
Figure 2.19: Intermolecular NOEs between the HIV-2 TAR and argininamide are shown in a series of traces from a 40ms 750 MHz ^{1}H-^{1}H NOESY in 90% $H_2O/^2H_2O$ at 5°C. Not all of the observed intermolecular crosspeaks were used as restraints. The Hε trace is from the F1 dimension while the other three are from the F2 dimension. The peaks marked with an **x** cannot be unambiguously assigned. Peaks that appear in multiple traces are labeled in only one trace.

Table 2.6. Intermolecular NOEs.

Arg(NH$_2$)	-U38	H5	v.w.
	-C37	H5	v.w.
Arg(Hβ)	-A27	H61/H62	v.w.
	-C39	H42	v.w.
		H41	v.w.
Arg(Hγ)	-A22	H8	w
	-U23	H5	s
		H6	v.w
Arg(Hδ)	-A22	H8	w
	-U23	H5	s
	-C39	H42	v.w.
		H41	v.w.
Arg(Hε)	-A22	H2'	v.w.
		H3'	w
		H8	v.w.
	-U23	H5	v.w.
		H6	v.w.
Arg(Hη)	-A22	H8	w
	-G26	H8	w

s = strong NOE, m = medium NOE,
w = weak NOE, v.w. = very weak NOE

C37 amino is far from the majority of the RNA NOEs. The C39 amino has NOEs to the

Hδ and Hβ. In the trace, a crosspeak is observed to Hε but this peak is next to and may be

a part of the very large H41-H42 crosspeak and the height of the peak may be influenced this

large intraresidue crosspeak. The C39(H41) crosspeak **j** is included in the restraint set

because the C39(H42) gives strong NOEs to the argininamide and spin diffusion is

formidable for the amino protons. The A27(H61/H62) - Hβ cross peak is a weak NOE

but was used as a restraint because it was consistent with all the other restraints and no spin

diffusion pathway is available. Peak **b** is one of two crosspeaks observed from a non-

specific argininamide binding site in the lower stem providing further reasoning for

ignoring many of the weaker intermolecular crosspeaks. The other is a G43(H3') to Hδ

NOE observed in the 3D ^{13}C NOESY-HSQC spectra. This conservative approach yielded

19 intermolecular NOE which are all observed in short mixing time experiments at a variety

of magnetic fields (Table 2.6).

Chapter 3: The HIV-2 TAR-Argininamide Structure

3.1 Assignment of Bounds

The improved spectral properties of the HIV-2 TAR-argininamide complex compared to the three base bulge HIV-1 TAR-argininamide complex are apparent from the appearance of the U40 imino proton resonance, and NOEs from the U40(H3) to the G26(H1) and G21(H1). In addition, two internucleotide NOEs to U25 were observed in the 4D NOESY spectrum suggesting that U25 may have a more unique position in the HIV-2 TAR-argininamide complex than in the three base bulge TAR complexes where no internucleotide NOEs to U25 were observed. In the HIV-2 TAR complex, 19 intermolecular NOEs were assigned, suggesting that the argininamide is well-constrained relative to TAR. Using restrained molecular dynamics calculations, a set of 20 converged structures of the 30 nucleotide TAR RNA complex with argininamide was obtained.

Although the spectral properties of the bulge region were improved in the two base bulge TAR, there is still significant evidence for dynamics. In ^{13}C-CT-HSQC experiments, the intensities of the peaks were different for various nucleotides indicating that they were being differentially affected by motions on the NMR time scale. The linewidths for the A-helical nucleotides were similar, while the loop and bulge residues gave a variety of broader and sharper lines (See Chapter 5 for more details). J-coupling data were also indicative of the averaging of multiple conformations. These observations are not surprising since the complex is in fast exchange with an estimated lifetime of about 5-10 μs. In addition, a number of nucleotides do not give many inter-residue NOEs indicating that they are unstructured. As a result, the NOEs and J-coupling data are time-averaged between the bound and free conformations. Therefore, conservative loose bounds were used in the bulge region to allow for possible effects on NOE intensities due

III *HIV-2 TAR-Argininamide Structure*

to averaging. A more detailed explanation of the dynamics and its implications for the structure is discussed in Chapter 5.

Many of the NOEs observed in the bulge region may also be affected by dynamics, but it is difficult to determine the range of conformations that the complex may be sampling. Because the complex is nearly saturated, the NOEs are expected to be primarily diagnostic of the bound conformation. However, due to their fast exchange of the argininamide ligand it is possible that the NOE intensities are affected by sampling other conformations, especially for the intermolecular NOEs. In particular, the methylene side chain of the argininamide is very likely sampling multiple conformations in fast exchange, and few NOEs are observed to the side chain distal to the guanidinium group. Because a number of NOEs are observed to the Hε, Hδ, Hγ, and Hη protons, the position of the guanidinium group of the argininamide is likely representative of the bound conformation.

3.2 Materials and Methods

3.2.1 NMR-derived restraints

A total of 681 NOEs were obtained that define the RNA structure (See Appendix B.1 for a complete list). This restraint set does not include covalently fixed distances such as pyrimidine H5-H6, cytosine amino H42 to H41, cytosine amino to H5, or adenosine H2 to H8 NOEs. All lower bounds were set to 1.8 Å. NOEs to prochiral hydrogens such as the H5'/H5'' and the argininamide backbone protons were accounted for by using $1/r^6$ averaging option for NOE restraints in X-PLOR. For exchangeable protons, restraints were generally assigned in two categories with upper bounds of 4.0 or 6.0 Å. The unusually strong NOEs between uridine-H3 and adenine-H2 protons were assigned an upper bound of 3.0 Å.

NOEs observed in D_2O NOESY experiments were placed into four categories: strong, medium, weak, and very weak, with upper bounds of 3.0, 4.0, 5.0 and 6.0 Å. A

101

101

6.0 Å upper bound was used for weak NOEs to H8/H6/H5 resonances where a strong spin diffusion pathway is available. Intermolecular NOEs between the U23-H5 and the Arg-Hγ and Arg-Hδ were observed at equal intensity in a very short 35 ms mixing time NOESY experiment. No intramolecular argininamide NOEs were used for distance restraints.

Hydrogen bond restraints were only used for the base pairs where an imino peak was observed and characteristic internucleotide NOEs were observed such as NOEs between the base paired nucleotides and other internucleotide NOEs indicating that the two hydrogen bonded nucleotides are close and that the base pair is stacked. Two restraints were used to define each hydrogen bond: one between the proton and the hydrogen bond acceptor and a second between the two heavy atoms centered one angstrom longer than the proton-heavy atom restraint. The following restraints were used for a G-C base pair, the G(O6)-C(H42) is 1.91 ± 0.2 Å, the G(H1)-C(N3) is 1.95 ± 0.2 Å, and the G(H22)-C(O2) is 1.86 ± 0.2 Å. For an A-U base pair, A(N1) - U(H3) is 1.70 ± 0.2 Å and the A(H62) - U(O4) is 1.95 ± 0.2 Å.

The sugar pucker was estimated from analysis of the H1'-H2' coupling constants in the ^{31}P-decoupled DQF-COSY. Nucleotides with a H1'-H2' coupling constant of >8 Hz in the COSY spectrum were classified as C2'-endo. Nucleotides with no COSY and TOCSY cross peaks were classified as C3'-endo. Weak COSY cross-peaks indicative of a mixed population of conformers were left unconstrained. Since the correlation of ^{31}P chemical shifts and backbone torsions is poorly understood in RNA structures, no torsion restraints based on chemical shifts were used (Jucker & Pardi, 1995). This is particularly important in the bulge region where evidence for conformational averaging of backbone coupling constants is observed and the ^{31}P chemical shifts are not very far from A-form.

3.2.2 Structure determination

Structure determination was performed with X-PLOR 3.1 (Brunger, 1992) using a modified force field. Bonds, angles, impropers, and non-bond parameters were taken from the updated parameters from the Nucleic Acid database (Berman *et al.*, 1992; Gelbin *et al.*, 1996). Proton parameters were added adapting those from the standard X-PLOR forcefield for nucleic acids. Dihedral parameters were adapted from the AMBER force field (Weiner *et al.*, 1986). Electrostatic terms were not used at any time during the calculations.

The following is a description with explanations for the molecular modeling protocols. Appendix B contains the particular XPLOR scripts in all their glorious detail.

Initial Dynamics

Random coordinates uniformly distributed in a 20 Å cube were used as the starting structures. Force constant for the NOE terms were scaled to 2% of their final value, and the bonds and angles were scaled 0.01% at the beginning of the calculations. No hydrogen bond restraints were used during this first phase. The repulsive quartic potential was used for the non-bonding terms and was scaled to 10% of its final value. The improper, dihedral and the torsion restraint terms were set to zero during these initial dynamics calculations. Dynamics at 1000 K were performed for 100 steps of 40 fs. During this time, the NOEs were scaled up to their full value. With the NOEs now at their full value, 1000 K dynamics was continued with the bond, angle, non-bond and improper terms scaled to 1% of their final value for 500 steps of 3 fs. The bonds, angles, and impropers were ramped to their full value during another 500 steps of 3 fs of 1000 K dynamics. A third round of 500 steps of 3 fs 1000 K dynamics is performed while the quartic non-bonding term is ramped to its full value. During this third round, the dihedral terms were included in addition to the torsion restraints which were introduced at 10% of their full value, which helped the overall convergence rate. At this stage, 41 of 142 structures were chosen for further refinement. Structures were chosen based on both their total and violation energies. The rmsd of this initial set to the average is 3.82 Å.

Simulated annealing

A procedure similar to Jucker and Pardi (Jucker & Pardi, 1995) was performed in the next phase. Hydrogen bond restraints were introduced for the first time for three rounds of 10000 steps of 1 fs dynamics at 1000 K, 600 K, and 400 K with 700 steps of energy minimization between the dynamics and 1000 steps at the end. All restraint terms were now set to their full value: the NOE force constant was 50 kcal/mol•Å^2 and the torsion restraint force constant was 1000 kcal/mol•deg^2. The dihedral force field term was turned off to avoid being locked into incorrect conformers in these lower temperatures. Structures were again chosen based on the total energy and violation energies with the converged structures having 3-11 violations > 0.2 Å.

Refinement

At this stage, the dihedral force field term was turned off to lower the barrier to dihedral conformational changes. Dynamics was performed for 2 ps in 2 fs steps at 500 K, 300 K, 150 K, 50 K and 10 K followed by 1000 steps of energy minimization. At this stage, each converged structure had only 3-10 NOE violations > 0.2 Å. For the last stage, the Lennard-Jones non-bonding term was used for 2000 one fs steps of dynamics at 300 K, 150 K, 50 K, 10 K. The calculations ended with 1000 steps of energy minimization. The 20 structures with the lowest total energy and violation energy were chosen. The average structure was calculated from this set of 20 and refined with 1000 steps of minimization to arrive at the final minimized average structure.

Modeling calculations

Two aspects of the structure were tested with non-experimental hydrogen bonding and planarity restraints: the argininamide binding to the major groove face of G26 and the base triple (U38-A27•U23). The set of 41 structures created from random coordinates was used as the starting point for all non-experimental restraint calculations. The same restrained molecular dynamics protocol was performed from this point. The base triple hydrogen bonds were restrained in the same way as the A-form base pairs: U23(N3)-

Table 3.1: Restraint and Structure Statistics.

Distance Restraint	Number of restraints	Restraints per Nucleotide
Sequential NOEs:		
strong	19	
medium	19	
weak	77	
very weak	97	
Long Range NOEs		
strong	3	
medium	19	
weak	20	
very weak	58	
Total Internucleotide NOEs:	292	9.73
strong	96	
medium	96	
weak	153	
very weak	25	
Total Intranucleotide NOEs:	370	12.33
Intermolecular NOEs:	19	
Total Number of NOEs	681	
Hydrogen Bonds	60	
Total distance restraints	741	

Dihedral restraints

C3'-endo	44
C2'-endo	14

Superposition statistics over 20 structures

All nucleotides and arg (Å)	2.63
16-29,36-46,arg (Å)	1.32
21-28,37-41,arg (Å)	0.84

RMSD of experimental restraints

All distance restraints (Å)	0.01 ± 0.001
All torsion restraints (°)	0.10 ± 0.02
Maximum distance restraint violation (Å)	0.15
Maximum torsion restraint violation (°)	0.61
Average number of distance violations > 0.1 Å	1
Average number of torsion violations > 0.5 (°)	0.15

RMSD from idealized geometry

Bonds (Å)	$0.003 \pm .0002$
Angles (°)	0.54 ± 0.02
Impropers (°)	0.46 ± 0.02

A27(N7), 2.95 ± 0.2 Å, U23(H3)-A27(N7), 1.95 ± 0.2 Å and U23(O4)-A27(N6), 2.95 ± 0.2 Å and U23(O4)-A27(H61), 1.95 ± 0.2 Å. The U23 and A27 were held planar to each other by using the planar restraint option of X-PLOR with a force constant of 250.0 kcal/mol•deg^2. For the calculations including torsion restraints, the α, β, γ, ϵ, and ζ A-form conformers were restrained into ±30° ranges. The 3' side of A22, the 5' side of G26, the two bulge nucleotides and the hexanucleotide loop were left unrestrained. During the 1000, 600, 400 K 30 ps dynamics phase of the calculations, the torsion restraints were slowly scaled to full value which helped the structures fit both the large number of distance and torsion restraints. Good convergence was obtained for each set of calculations: 17 for the argininamide hydrogen bonds to G26, 10 for the argininamide hydrogen bonds and planarity, 23 for the base triple hydrogen bonds, 11 for the base triple planarity, and 11 structures were obtained when all hydrogen bonds and planarity were included. 15 structures were obtained for the calculations without sugar pucker restraints in the bulge and 20 structures were obtained when all the intermolecular NOEs were used. Rmsd's were measured using X-PLOR and the structures were viewed and displayed in Biosym's InsightII Release 95.0 software package (Biosym/MSI, San Diego, CA).

3.3 Structure Determination

The NOEs and coupling constant data, summarized in Table 3.1, were converted to distance and torsion restraints for restrained molecular dynamics calculations. Non-exchangeable data were placed into one of three categories with upper bounds of 3.0 Å, 4.0 Å, and 5.0 Å while some exchangeable NOEs had upper bounds of 6.0 Å for very weak resonances and when a clear spin diffusion pathway was evident. No non-experimental restraints were used in the structure calculations, and electrostatic interactions were neglected. In the first of three phases for the calculations, random coordinates within a 20 Å cube were generated and refined against the experimental restraints to obtain the global

106
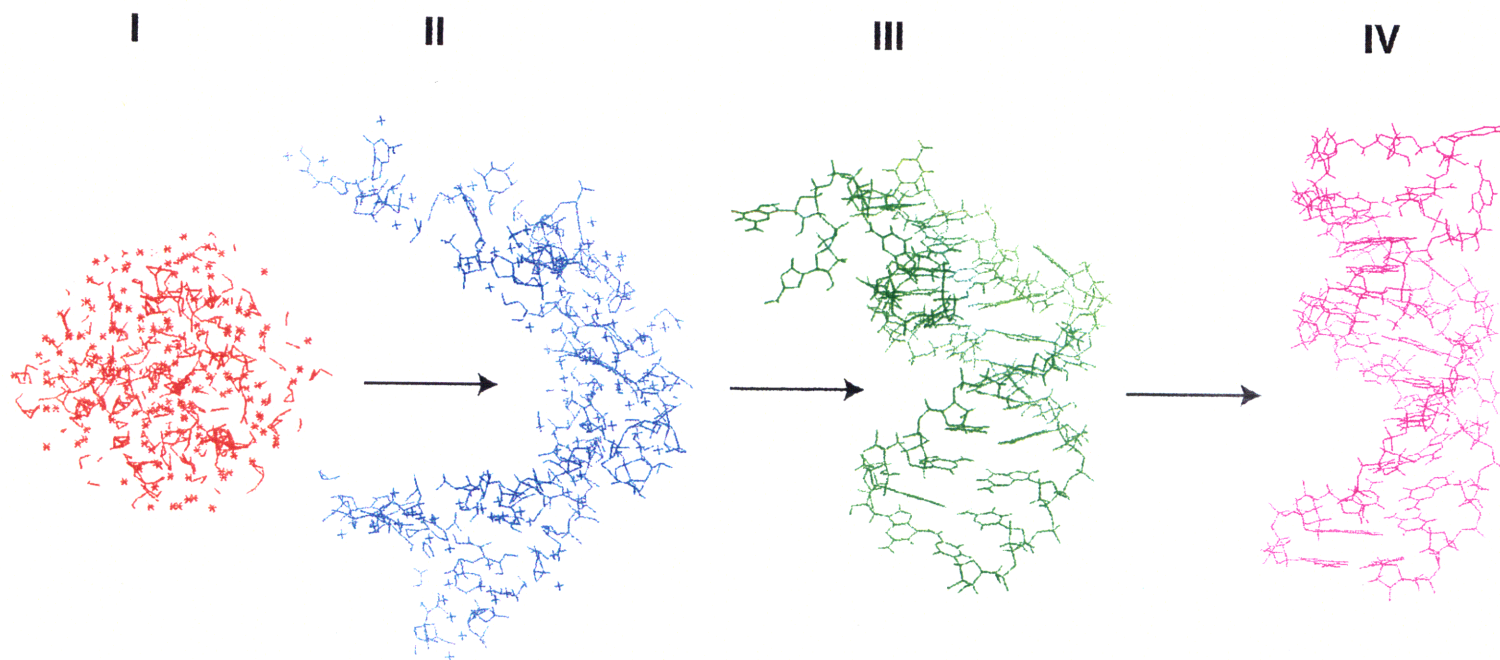
Figure 3.1: Pictorial representation of the structure calculations. **I.** Random coordinates **II.** The NOE restraints have guided the structures before covalent terms are scaled up. **III.** Before annealing, the RNA is defined in the absence of hydrogen bond restraints. **IV.** The final cooled structure.

fold of the complex. In the second and third phases, the structures were refined by slowly increasing the scale factors of the various force field terms and then lowering the temperature. A pictorial representation of the structure calculations is shown in Figure 3.1. The 142 random starting structures were refined with 40 ps of initial restrained molecular dynamics where the NOE restraints dominate, and the covalent and the repulsive non-bonding terms are greatly reduced. 41 structures with violation energies <600 kcal/mol-Å^2 were chosen for further refinement. Surprisingly, the global fold of the complex is defined at this stage and an RNA A-form helix is clearly recognizable despite the lack of hydrogen bonding restraints for the base pairs, as seen in Figure 3.1c. During the second phase of the calculations, refinement was continued with the introduction of hydrogen bond restraints during 10 ps of restrained molecular dynamics at 1000 K, 600 K, and 400 K. The last stage of the calculations consisted of cooling from 500 to 10 K in 2 ps steps. The dihedral force field term was turned off at this stage to allow sampling of all torsion conformations. The Lennard-Jones potential was introduced for a second annealing starting from 300 K and cooled in 2 ps steps.

At each stage of the refinement, NOE violations were reduced, indicating that the calculations were refining against the restraints to define the structure. Using this protocol, 20 converged structures were obtained starting from 142 sets of random coordinates. Examination of the family of structures revealed that an RNA A-form helix was defined in the absence of backbone torsion restraints. A complete listing of the rmsd values for the restraint and idealized geometry violations is given in Table 3.1.

3.4 Structure of the TAR-Argininamide Complex

The superposition of the 20 structures to the average structure is shown in Figure 3.2. The overall structure is well-defined except for the six nucleotide loop which is not shown. The rmsd to the average for this set of structures is 1.32 Å when all residues of the complex except the loop nucleotides (30-35) are superimposed. The loop region gives few
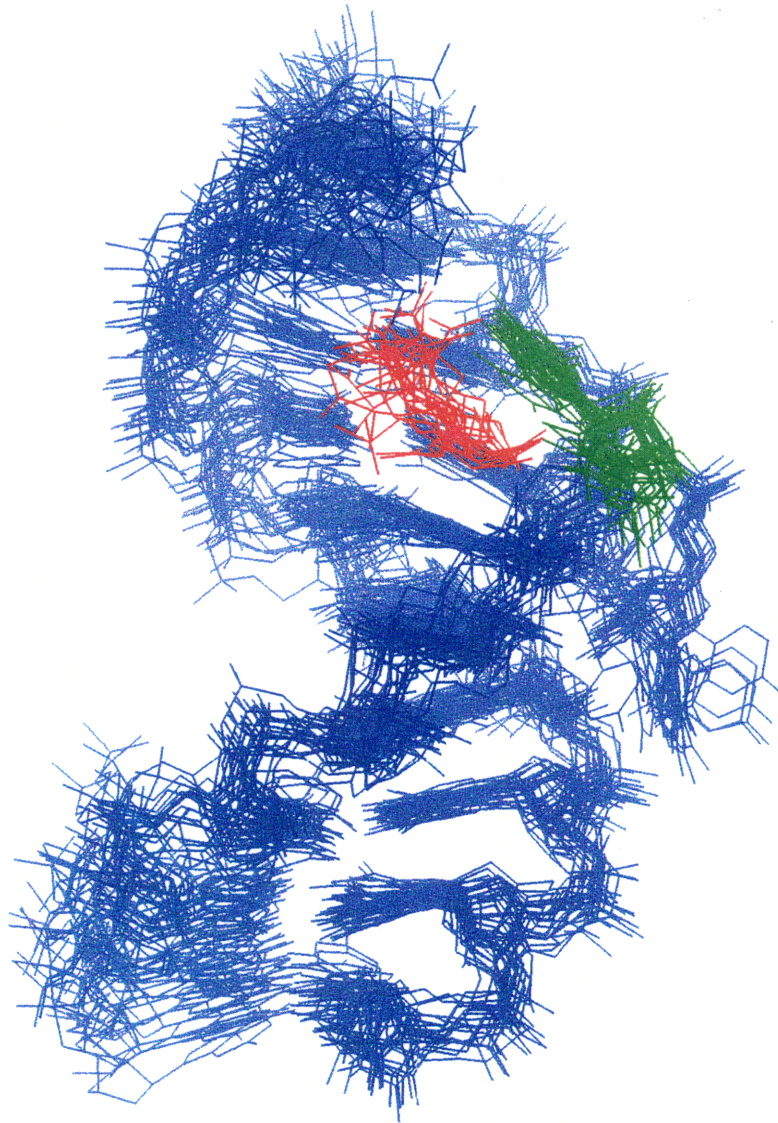
Figure 3.2: The best fit superpositions to the average structure of all 20 structures of all residues except for the poorly defined loop. The U23 is in green and the argininamide is in red. Notice that the guanidium group of the argininamide is well-positioned especially compared to the distal amide side of the argininamide which is poorly positioned.

internucleotide NOEs leading to very poor definition, consistent with previous work (Aboul-ela et al., 1995; Colvin et al., 1993; Jaeger & Tinoco, 1993). The bulge region is well defined with a rmsd of 0.84 Å to the average structure (nucleotides 21-28,37-41 and the argininamide) as shown in Figure 3.3. The position of the argininamide guanidinium group is well-defined, while the position of the remainder of the side chain is poorly defined.

A stereo view of the average structure with the phosphate backbone highlighted by a ribbon representation is shown in Figure 3.4. U23 and the argininamide are in the major groove while U25 is in the minor groove. The upper and lower stems are continuously stacked between C39 and U40 opposite the bulge, but there is an over twisting of the two stems relative to each other, and G26 does not stack directly on A22. The argininamide is located in a pocket formed by A22, U23, and G26. Figure 3.5 illustrates how well the structure is defined by the NOEs as all the internucleotide and intermolecular NOEs are overlayed on the average structure. Only three internucleotide NOEs are in the hexanucleotide loop preventing any conclusions from being drawn about this part of the structure. In the lower and upper stems, most of the NOEs are aligned with the helix and only a few connect the two strands. A much more complex, denser network of NOEs define the structure of the argininamide binding site which is the best defined region of the structure. This pattern of NOEs is explicitly tabulated in Table 3.2 where the number of long range distance restraints to G26 is clearly seen.

The intermolecular NOEs are highlighted on the average structure in Figure 3.6. Most of the 19 intermolecular NOEs to the argininamide are to protons near the guanidinium group, while few NOEs were observed to the amide and Hα protons of the argininamide. The position of the guanidinium group of the argininamide is defined by intermolecular NOEs to nucleotides on all sides of the binding pocket in the major groove. This localization of the intermolecular NOEs explains why the guanidium group is well-
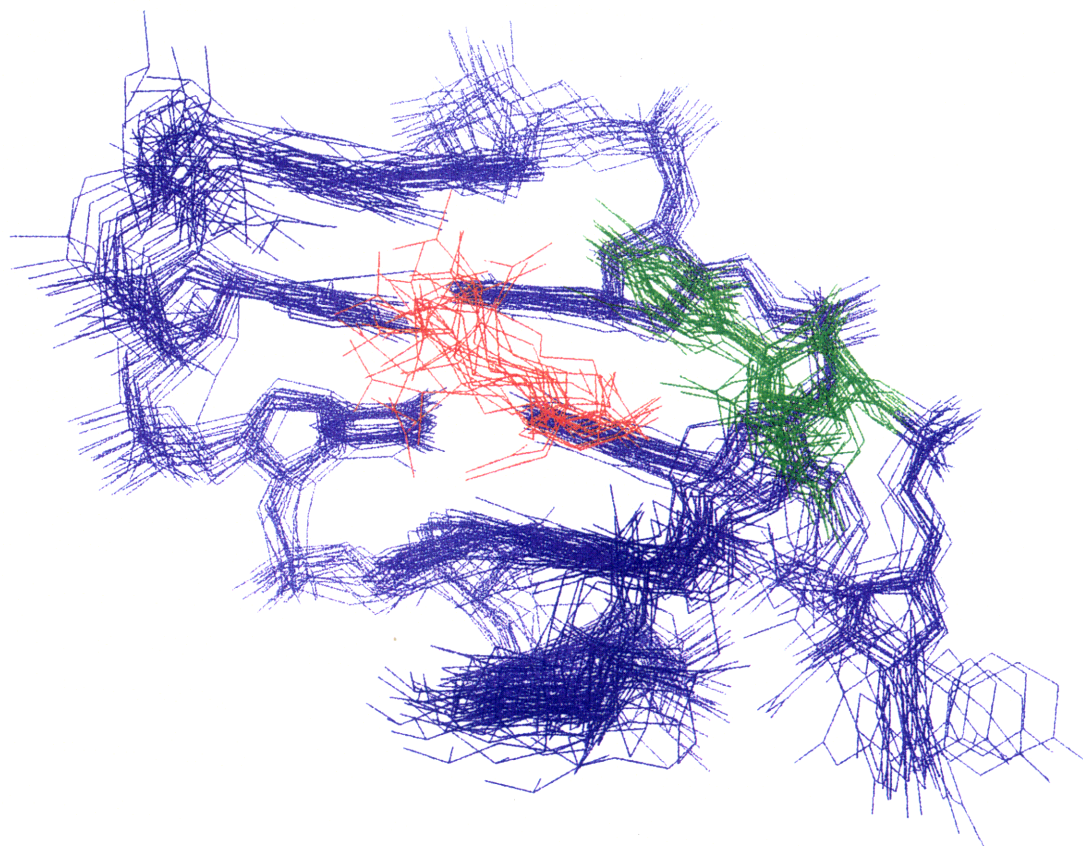
Figure 3.3: The best fit superpositions to the average of all 20 structures for the argininamide binding site region. The U23 is in green and the argininamide is in red. Notice the excellent definition of the U23 and the surrounding base pairs.
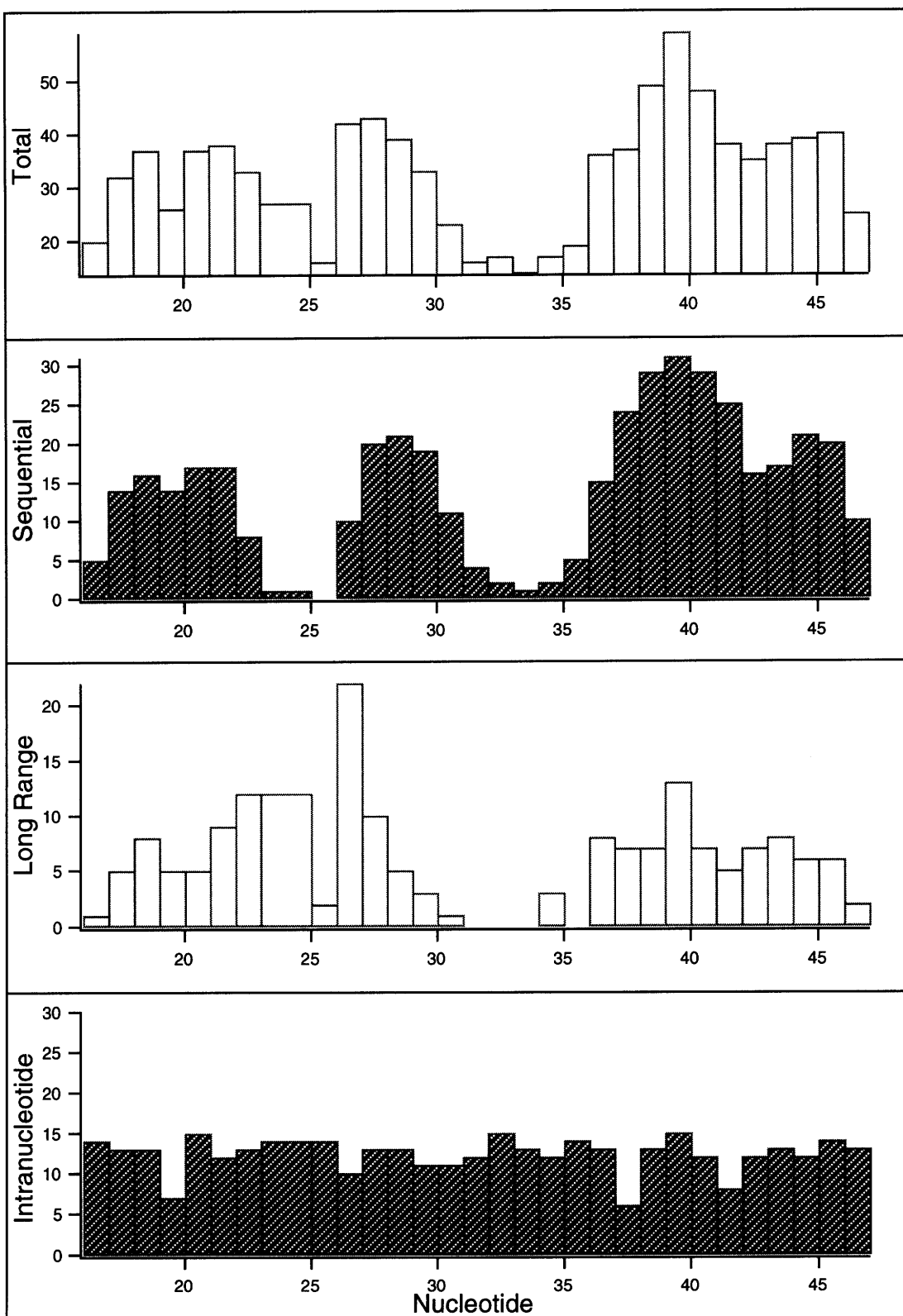
Table 3.2: NOEs by residue. Note that inter-residue NOEs count as one for both residues; therefore, the total number of NOEs will be larger than listed in Table 3.1.
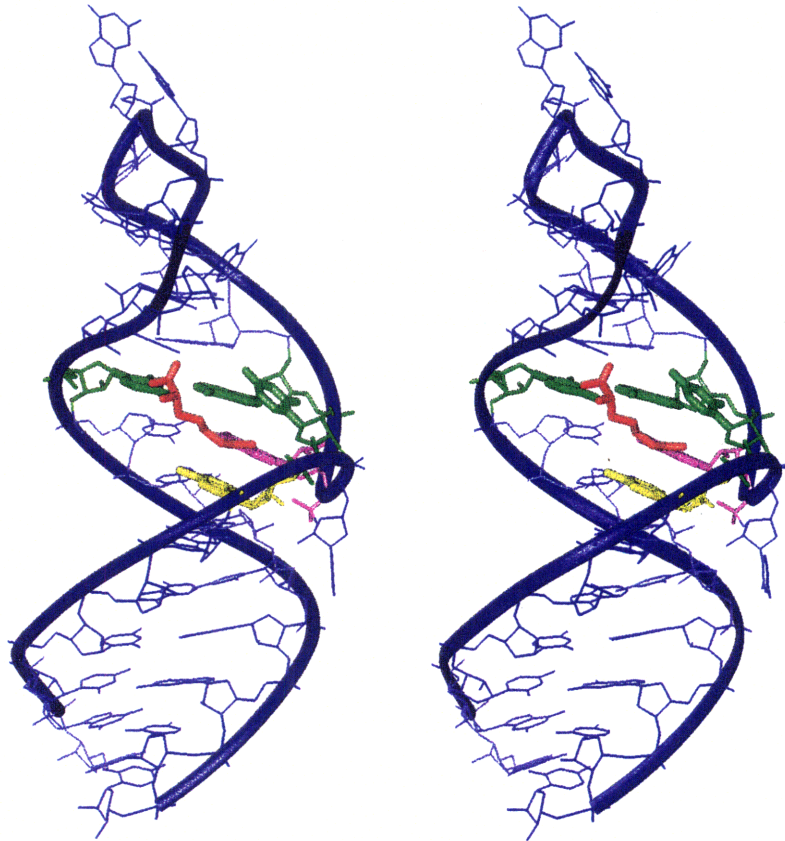
Figure 3.4: A stereo view of the average structure of the TAR-argininamide complex is shown with U23 highlighted in green, G26 in purple, A22 in yellow and the argininamide in red. The ribbon traces the backbone showing the distortion in the bulge region.
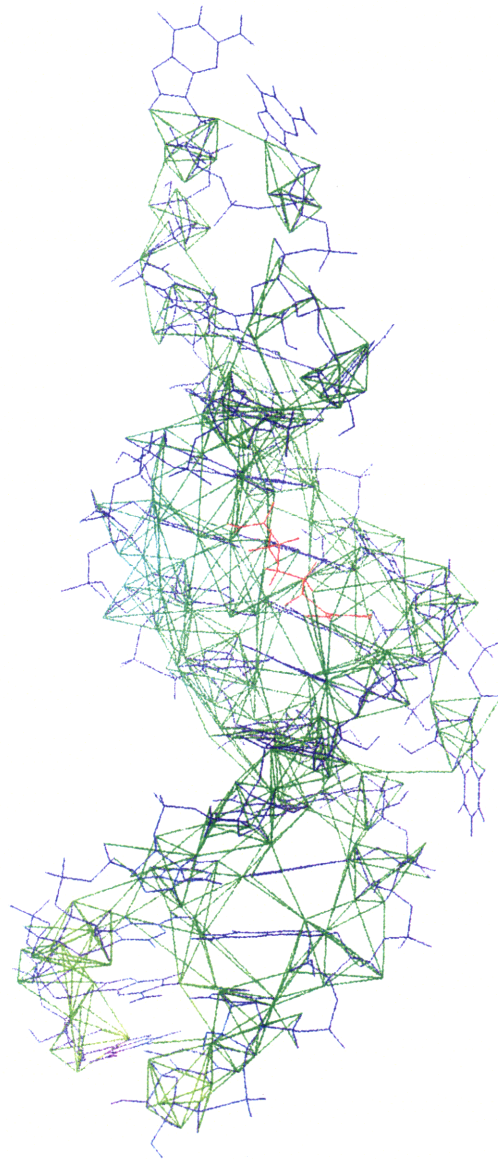
Figure 3.5: All the internucleotide and intermolecular NOEs are overlayed in green on the average structure. The argininamide is in red. Most of the NOEs are aligned along the helix axis except in the bulge region where a much more complex network of NOEs is found. Also, notice the very few number of NOEs in the loop.
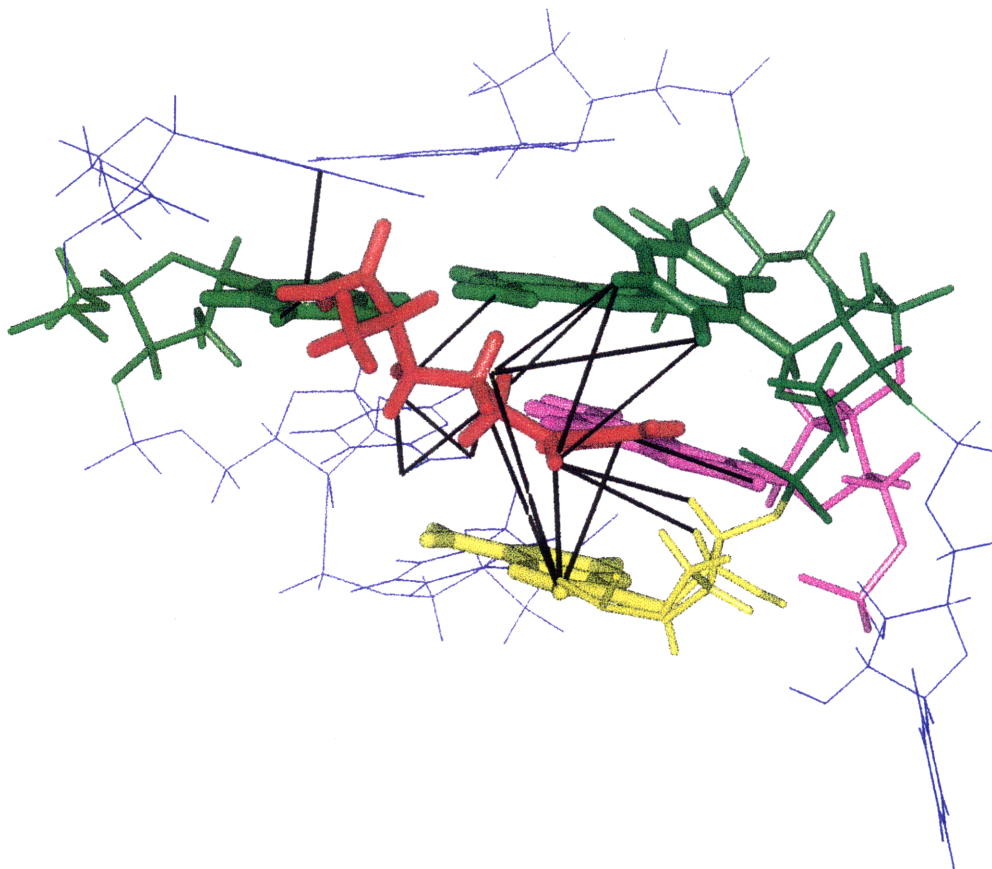
Figure 3.6: The intermolecular NOEs to the argininamide are highlighted in black with the U23, A27, and U38 in green, A22 in yellow, G26 in purple and the argininamide in red. Most of the intermolecular NOEs are to the guanidium group side of the argininamide.

defined while the amide side is not. One of the guanidinium group η-nitrogens is within

hydrogen bond distance to the G26(N7) and the G26(O6) in all 20 structures suggesting

that the argininamide is hydrogen bonding to G26. In addition, the guanidinium group is

making van der Waals contacts to the U23 and the A22 bases, as shown in the two views

of the bulge region in Figure 3.7. Thus, the argininamide guanidinium group is lying

between the planes of the bases, U23 and A22, making contacts analogous to stacking

interactions, to form an argininamide sandwich.

In 17 of the 20 structures the O4 and N3 of U23 are within hydrogen bonding

distance (3.5 Å) to the N6 and N7 of A27, respectively, and in all 20 structures the

distances are less than 4.08 Å. The average U23(N3) to A27(N7) distance is 2.94 ± 0.15

Å and the average U23(O4) to A27(N6) distance is 3.56 ± 0.33 Å. The U23 is very close

to the major groove face of the A27 in the orientation of a typical U-A•U base triple as

shown in Figure 3.7. However, the view into the major groove in Figure 3.7a, shows that

the U23 is not in the same plane as the A27-U38 base pair as might be expected for an ideal

base triple. The conformation of this bulge region is well-defined by a large number of

conservatively assigned and loosely constrained distance restraints. Therefore, this

structure represents a static view of the critical features of the dynamic complex between

HIV-2 TAR and argininamide.

3.4.1 The argininamide binding site

The Puglisi *et al.* model of the bound conformation of the HIV-1 TAR suggested

that the argininamide is hydrogen bonding to the major groove of G26. The high

resolution structure of the HIV-2 TAR reveals that one of the guanidinium group nitrogens

is less than 3.5 Å from the G26(N7) in all 20 structures and is less than 3.5 Å from

G26(O6) in 18 of 20 structures. Arginine guanidinium groups have been frequently

observed in hydrogen bonding interactions with guanines in the major groove in many

DNA-protein and RNA-peptide complexes. The structure of HIV-2 TAR suggests an

116
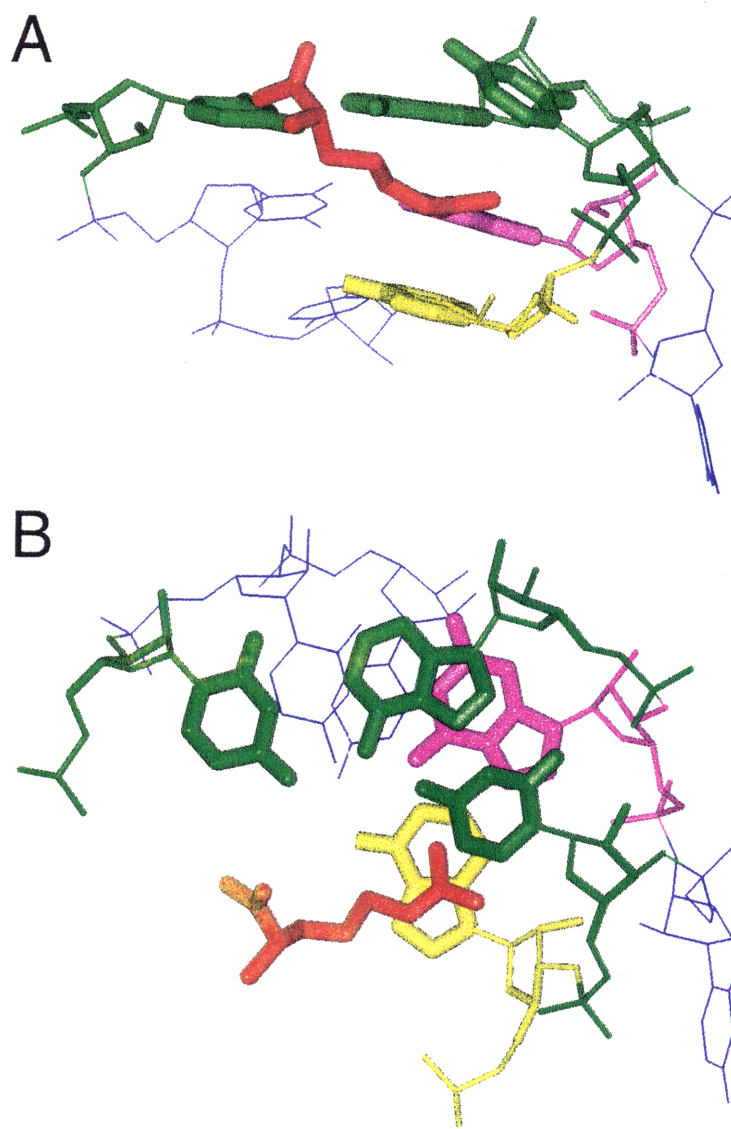
Figure 3.7: Two views of the bulge region of the minimized average structure are shown. **A)** Looking into the major groove, the argininamide, in red, is seen lying next to the G26, in purple. The argininamide is stacking between the U23, in green, and the A22, in yellow. The A27 and U38 are also highlighted in green. **B)** Looking down the helix axis, the orientation of the U38, A27, and U23, highlighted in green, in a U-A·U base triple is seen.

additional mode of interaction with the RNA by stacking of the guanidinium group between the bases of U23 and A22. A number of NOEs position the guanidinium group between U23 and A22 including NOEs from the A22(H8) and G26(H8) to guanidinium protons, NOEs from the A22 ribose, A22(H8) and the U23(H5) to the argininamide Hε. These argininamide Hε NOEs orient the argininamide in the groove by drawing the Hε proton towards A22.

The guanidinium group of the argininamide is stacking on A22, lying where G26 would be positioned if it continued the A-form helix by coaxial stacking on A22. On the opposite strand, the C39 and U40 nucleotides are stacking much like they would in an A-form helix. G26 is pushed aside by over twisting at the junction of the coaxially stacked helices to be juxtaposed to the argininamide while the U23 is stacking on top of the argininamide guanidinium group. This local rearrangement serves to widen the major groove making the G26 more accessible for hydrogen bond contacts to the argininamide.

The two backbone phosphates, P22 and P23, identified in ethylation interference studies are near the argininamide in the 20 structures (Calnan et al., 1991; Churcher et al., 1993; Pritchard et al., 1994; Tao & Frankel, 1992). The two oxygens of these phosphates were not consistently <4.0 Å from the guanidinium group nitrogens in the calculated structures, and direct hydrogen bonds to these phosphates are not consistent with the NMR data. However, these phosphates may help the argininamide bind to TAR through favorable electrostatic interactions.

3.4.2 The base triple

The proposed base triple plays a critical role in forming the argininamide binding site by opening the major groove to provide access to G26. In the HIV-2 TAR-argininamide family of structures, all 20 structures show that the U23(O4) and U23(N3) are within hydrogen bonding distance to the A27(N6) and A27(N7), respectively. Although hydrogen bond formation does not require the U23 and the A27 to be co-planar,

previously observed base triples have been co-planar. The reason for the non-planarity in the structure may simply be that there are no extensive geometric constraints as observed in the triplex and tRNA structures where the base triples are part of a stacking network (Macaya *et al.*, 1992; Roberts et al., 1974).

Of the three adenine amino protons involved in base pairing, only the A27 amino resonances were observed in a ^{15}N-HSQC spectrum even at 750 MHz suggesting that the A27 amino is different than the other adenosine aminos. The A27 amino proton resonances were sufficiently sharp to allow the observation of a number of NOEs including a weak NOE to Arg(Hβ) and along with NOEs to the G26(H1), G28(H1), and C39 amino group. However, one long (>4.0 Å) NOE, A27(H62/H61)-U23(H5), predicted from the base triple model, is unobserved. We believe that the base triple is formed, but dynamic, and that some weak NOEs are not observed because of a combination of conformational and solvent exchange. The base triple may be simply fraying, analogous to a terminal Watson-Crick base pair in a helix, where imino protons are often rapidly exchanging. Despite the lack of any direct NOEs between the U23 base and the A27 base which would unambiguously define a base triple, the U23 base is well-positioned by NOEs to its own ribose and to the argininamide, which in turn give many NOEs to the surrounding nucleotides in the binding pocket.

The base triple model has been tested by designing a C-G•C$^+$ isomorphic base triple mutant (Puglisi et al., 1993). This base triple mutant binds argininamide in a pH dependent manner as predicted suggesting the importance of base triple formation for argininamide binding. Chapter 4 discusses studies of a two base bulge C-G•C$^+$ base triple mutant which found direct NOE and hydrogen bonding evidence for the existence of a base triple.

3.5 Modeling Calculations

The structures based solely on experimental NMR data are consistent with hydrogen bonding interactions in the bulge region. However, we did not obtain direct evidence of hydrogen bonding between the U23 and the A27, or the G26 and the argininamide. To test various possible hydrogen bonding interactions, modeling calculations were performed using non-experimental hydrogen bonding restraints in addition to the NOE experimental restraints. All modeling calculations were performed beginning with the set of 41 structures resulting from the initial phase of the restrained molecular dynamic calculations. The second and third phases of the restrained molecular dynamics refinement protocol was performed for each set of non-experimental restraints exactly as was done for the structure determination. The distance restraint violations and the total energy of the sets of structures were monitored to examine the effect of the inclusion of the non-experimental restraints.

The first calculations examined the potential hydrogen bonds of the U-A•U base triple between the A27-U38 base pair and U23. Refined structures including two hydrogen bond restraints between the U23 and the A27 produced a set of structures with the same range of total and violation energies as the structure determination, as summarized in Table 3.3. Although the hydrogen bond restraints are satisfied, the U23 is not co-planar with the A27-U38 base pair in the resulting set of structures. A second set of calculations were performed including a planarity restraint between the A27 and U23. No violations greater than 0.1 Å were induced and the U23(N3) and U23(O4) were within hydrogen bonding distance of the A27(N7) and A27(N6) as expected for base triple formation. These calculations clearly show that our experimental restraint set is consistent with base triple formation, but that the experimental constraints alone do not specify the ideal base triple geometry.

Modeling calculations were also done to test argininamide hydrogen bonding to the G26. Hydrogen bond restraints between the guanidinium group and the G26 were added

Table 3.3: Summary of total and distance violation energies.

Structure Determination

Avg. total energy (kcal/mol)	-278 ± 18
Avg. distance violation energy (kcal/mol)	4.4 ± 1.0

Calculations with non-experimental restraints

Arg H-bonds to G26

Avg. total energy (kcal/mol)	-275 ± 21
Avg. distance violation energy (kcal/mol)	4.4 ± 0.5

Arg H-bonds and planarity to G26

Avg. total energy (kcal/mol)	-277 ± 19
Avg. distance violation energy (kcal/mol)	4.4 ± 0.6

Base triple H-bonds

Avg. total energy (kcal/mol)	-277 ± 15
Avg. distance violation energy (kcal/mol)	5.0 ± 1.7

Base triple planarity

Avg. total energy (kcal/mol)	-279 ± 16
Avg. distance violation energy (kcal/mol)	3.9 ± 0.5

All non-exp. restraints

Avg. total energy (kcal/mol)	-268 ± 15
Avg. distance violation energy (kcal/mol)	4.9 ± 0.6

Additional (27) intermolecular restraints

Avg. total energy (kcal/mol)	-271.91 ± 8.14
Avg. distance violation energy (kcal/mol)	3.95 ± 0.15

Alternative Arg(Hε) hydrogen bonding restraints

Avg. total energy (kcal/mol)	-261.91 ± 19.04
Avg. distance violation energy (kcal/mol)	8.84 ± 1.61

For each value, the mean is given with the standard deviation.

and found to be consistent with the NMR data as monitored by both the total and violation energies, although the argininamide guanidinium group was not co-planar with G26. A planarity restraint between the guanidinium group and the G26 was added, and the calculations were repeated. The co-planarity of G26 and the argininamide was also found to be consistent with all the NMR restraints.

An alternative hydrogen bonding between the argininamide Hε hydrogen hydrogen bonding to the G26(N7) and one of the guanidium protons hydrogen bonding to the G26(O6) was tested. Although 15 structures with a rmsd of 1.36 Å were obtained, these structures have higher violation energies with an average of 4.47 NOE violations > 0.1 Å per structure as seen in Table 3.3. In addition, some hydrogen bonding restraints were violated in every structure and the base triple is not as well defined as the average U23(N3) to A27(N7) distance is 3.69 Å ± 0.21 Å and the average U23(O4) to A27(N6) distance is 4.15 ± 0.29 Å. These results suggest that a non-experimental restraints are able to induce violations indicating that the NMR restraints sufficiently define a particular argininamide binding orientation.

The last set of modeling calculations was performed to test if all the hydrogen bond and planarity restraints for both the argininamide and the base triple could be satisfied simultaneously. An idealized family of structures was obtained consistent with all the experimental NMR restraints. The average structure of this idealized TAR-argininamide complex has an rmsd to the experimental NMR average structure for all the residues, except the loop, of 0.73 Å. This difference is smaller than the precision of the experimental structure, indicating that the idealized average structure is within the range of the family of 20 experimental structures. A close-up view of the minimized average from this family of structures is shown in Figure 3.8 illustrating the argininamide stacking between the A22 and the U23. These calculations demonstrate that all the NMR data are consistent with both base triple formation and argininamide hydrogen bonding to G26. In the family of 20 structures derived solely from experimental restraints, no planar base triple is observed.
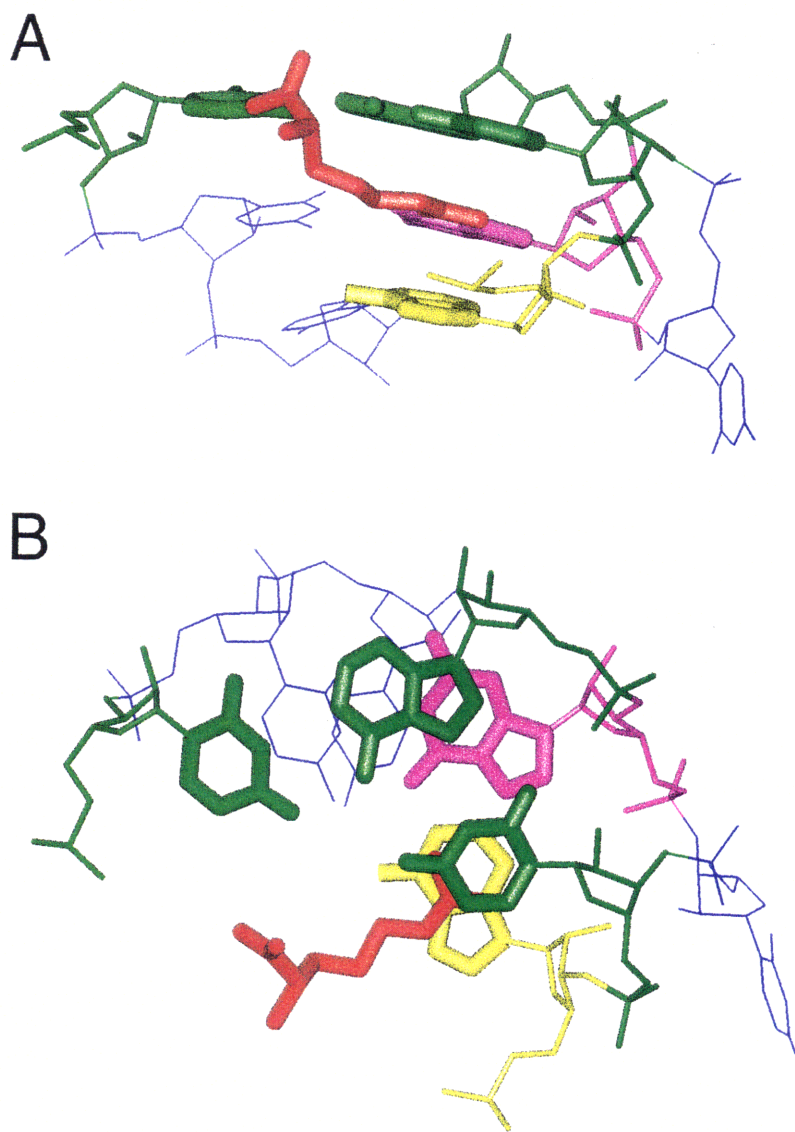
Figure 3.8: The same two views of Figure 3.7 of the bulge region of the minimized average structure with non-experimental hydrogen bond and planarity restraints are shown. **A)** Looking into the major groove, the argininamide, in red, is seen lying next to the G26, in purple. The argininamide is stacking between the U23, in green, and the A22, in yellow. The A27 and U38 are also highlighted in green. **B)** Looking down the helix axis, the U38-A27·U23 base triple is highlighted in green.

This may be a sampling problem of the molecular dynamics protocol since all the calculations with non-experimental restraints are within the same range of both total and violation energies as shown in Table 3.3.

To test whether the stems are truly A-form, a family of structures was generated including non-experimental restraints for A-form torsions with a rmsd of 1.22 Å. No significant differences in the structure of the bulge region are induced, and U23 and A27 are still within hydrogen bonding distance. The slightly lower rmsd in this family of structures is consistent with the expected higher precision that the A-form torsion restraints induce in the structures. However, a 0.1 Å increase in precision is a modest gain, suggesting that a large number of NOEs, as obtained for this complex, can define the structure in the absence of backbone torsion restraints.

Effect of additional intermolecular NOEs Some intermolecular NOEs were not used for the structure determination because they may be influenced by spin diffusion and/or dynamic motions. Additional modeling calculations were performed to test the range of intermolecular NOEs that are required to define the interaction. A total of 27 intermolecular NOEs that include some of the weaker NOEs that were observed in the spectra and not used for the conservative structure determination were used. These restraints (see Appendix B) include some restraints between the argininamide Hε and the C39 amino protons, seven structures with a rmsd of 1.29 Å were obtained. No large violations > 0.1 Å were induced by these extra restraints suggesting that these loose constraints are consistent with all the other experimental restraints.

Effect of no sugar pucker restraints Because of the conformational averaging observed for some of the backbone torsions, it would not be surprising if some of the sugar puckers were also sampling multiple conformations. Only the A35 sugar pucker showed evidence of conformational averaging and was left unconstrained in the original structure

calculations. Nevertheless, to ensure that the sugar pucker constraints did not unduly influence the final structure, calculations were performed with no torsion restraints. 15 structures with a 1.39 Å rmsd were obtained from the same molecular dynamics protocol without any sugar pucker or backbone torsion restraints. The average U23(N3) to A27(N7) distance is 2.95 ± 0.04 Å and the average U23(O4) to A27(N6) distance is 3.59 ± 0.10 Å. These calculations show that enforcing the sugar puckers does adversely effect the base triple nor the argininamide binding structures.

No intraribose restraints Although the intraribose NOEs appear to be redundant and unnecessary for the structure determination, calculations without any intraribose distance restraints suffered from extremely very poor convergence. This is likely a function of the molecular dynamics protocol utilized in these studies where the random coordinates require the restraints to guide the structures, even before the geometric covalent terms are defined. Other protocols which use other random starting structures such as random torsions which have intact covalent structure from the beginning may work better when no or very few intraribose distance restraints are implemented.

H1' and H2' restraints only Calculations were also performed with all the restraints involving the exchangeable and aromatic protons along with the hydrogen bond restraints but with only the distance restraints to the H1' and H2' protons from the ribose for a total of 403 distance restraints. These calculations suffered from poor convergence as only four structures from 75 different random coordinates were obtained. However, the salient features of the TAR-argininamide structure remain well-defined as the average U23(N3) to A27(N7) distance is 2.83 Å ± 0.63 Å and the average U23(O4) to A27(N6) distance is 3.39 $\pm$ 0.75 Å. Surprisingly, the rmsd of these four structures is a very low 1.17 Å. This low rmsd could be because the influence of the statistics of small numbers as previous similar calculations found that the rmsd for structures from limited restraint sets is higher

than with all the restraints (Allain & Varani, 1997). Alternatively, it may reflect the

stringency required to converge with only a limited number of distance restraints. These

structures may be a subset of the structures calculated with all the restraints included.

3.6 Backbone Torsion Analysis

Inclusion of experimental backbone torsion restraints was limited due to the extreme

chemical shift overlap of A-form helix proton and phosphorous resonances in HCP

experiments. Furthermore, conformational averaging of J-couplings provided evidence

that no unique conformer was assignable for the bulge region nucleotides U23, U25, and

the G26 γ and β. Despite the lack of backbone torsion restraints, the structures had well-

defined phosphodiester backbones as illustrated in Table 3.4 which shows that many of the

backbone torsions were well-defined by the NMR data.

Table 3.4: Torsion Analysis

Torsion	Number in A-form Rotamer
α	151/368
β	252/368
γ	149/368
ϵ	354/388
ζ	322/388

The backbone torsions, ϵ and β, fall within the expected range 93% and 84%,

respectively, of the time in the A-form helical regions. The γ torsion is not as well defined

as it falls into the expected gauche[+] conformer only 40% of the time in the A-helical

regions. Because α and ζ cannot be directly measured by NMR, it is interesting to see

how well they are defined in the A-form regions in this high resolution structure. The ζ

torsion falls into the expected A-form gauche[-] conformer 68% of the expected A-form

Table 3.5: Average torsions with rms deviations for each residue from the 20 structures.

RESIDUE	α	β	γ	δ	ε	ζ
16	N.M.†	N.M.	201.30 ± 13.26	86.37 ± 0.43	178.80 ± 9.74	275.01 ± 12.28
17	167.06 ± 15.72	194.66 ± 4.73	152.84 ± 9.87	84.90 ± 0.75	163.25 ± 13.66	232.14 ± 22.31
18	180.42 ± 23.57	187.97 ± 1.87	128.80 ± 15.22	84.70 ± 0.36	192.23 ± 7.86	255.92 ± 19.34
19	172.80 ± 14.43	185.23 ± 9.74	161.69 ± 10.27	87.24 ± 0.62	193.56 ± 9.64	276.36 ± 14.12
20	254.00 ± 16.83	179.73 ± 4.71	80.47 ± 10.45	85.01 ± 0.28	199.95 ± 2.34	277.72 ± 13.51
21	233.53 ± 15.72	175.17 ± 5.65	108.62 ± 13.13	83.74 ± 0.36	162.08 ± 10.74	233.96 ± 21.67
22	176.01 ± 18.28	197.43 ± 3.78	141.71 ± 12.45	85.06 ± 0.67	138.73 ± 22.98	202.36 ± 16.96
23	192.12 ± 29.03	211.88 ± 13.87	176.86 ± 17.90	144.45 ± 0.56	277.06 ± 1.46	68.47 ± 5.21
25	181.35 ± 3.86	237.01 ± 6.65	296.64 ± 15.22	139.66 ± 0.17	296.54 ± 1.52	200.77 ± 37.91
26	209.14 ± 10.83	242.48 ± 15.13	169.64 ± 6.36	87.12 ± 0.11	189.62 ± 1.61	301.59 ± 1.60
27	161.71 ± 6.97	181.55 ± 2.49	169.53 ± 6.44	86.68 ± 0.33	193.95 ± 6.71	276.71 ± 10.20
28	226.05 ± 15.92	189.18 ± 4.76	109.09 ± 10.78	85.92 ± 0.32	210.63 ± 6.29	266.96 ± 12.84
29	210.72 ± 15.63	191.14 ± 8.80	113.68 ± 12.21	85.73 ± 0.43	186.42 ± 9.56	230.17 ± 24.24
30	211.56 ± 19.19	187.39 ± 5.05	107.42 ± 13.76	143.87 ± 0.69	233.47 ± 10.97	209.92 ± 25.86
31	166.88 ± 22.10	196.21 ± 10.82	136.72 ± 13.37	144.00 ± 0.51	258.16 ± 5.44	183.02 ± 23.32
32	165.53 ± 22.35	184.37 ± 12.32	243.76 ± 18.28	148.20 ± 0.88	261.34 ± 5.90	159.75 ± 11.57
33	134.52 ± 19.18	211.37 ± 9.13	124.46 ± 14.98	144.40 ± 0.34	245.21 ± 6.31	211.69 ± 21.82
34	181.26 ± 19.75	175.63 ± 14.53	190.30 ± 22.32	144.66 ± 0.54	244.58 ± 6.02	223.22 ± 12.91
35	163.92 ± 18.72	179.31 ± 17.23	158.75 ± 17.10	128.88 ± 6.48	234.52 ± 14.88	151.94 ± 21.38
36	140.94 ± 18.81	187.06 ± 8.16	145.46 ± 14.36	85.79 ± 0.47	173.65 ± 13.76	242.38 ± 20.73
37	193.60 ± 22.21	189.15 ± 3.38	122.81 ± 14.23	85.89 ± 0.38	199.10 ± 1.63	295.59 ± 1.42
38	217.41 ± 15.81	182.66 ± 2.17	119.36 ± 13.66	84.75 ± 0.49	187.20 ± 6.55	278.79 ± 10.08
39	191.52 ± 14.99	189.49 ± 5.07	139.83 ± 10.37	85.89 ± 0.43	211.80 ± 3.44	279.68 ± 1.92
40	187.13 ± 13.91	204.94 ± 7.51	140.37 ± 9.19	85.53 ± 0.41	201.92 ± 7.92	263.16 ± 15.76
41	233.71 ± 16.17	177.54 ± 7.67	99.72 ± 10.19	83.96 ± 0.57	146.45 ± 14.47	208.98 ± 22.74
42	177.59 ± 23.40	191.44 ± 2.82	123.80 ± 12.51	85.54 ± 0.41	199.13 ± 1.89	296.52 ± 1.13
43	238.22 ± 13.97	185.13 ± 1.64	97.25 ± 12.41	84.23 ± 0.46	196.83 ± 7.43	278.30 ± 10.78
44	200.56 ± 17.94	183.60 ± 5.91	132.35 ± 13.08	85.63 ± 0.45	205.23 ± 3.26	290.37 ± 1.79
45	216.20 ± 16.07	186.91 ± 2.74	117.73 ± 14.04	85.63 ± 0.38	214.21 ± 6.63	259.84 ± 12.71
46	217.31 ± 17.23	185.44 ± 9.53	112.02 ± 12.10	85.63 ± 0.39	N.M.	N.M.
A-form	292	178	60	88*	205	292

* Except for A35 all the riboses were restrained to be C3'-endo or C2'-endo.
† The terminal torsions were not measured as there are no restraints at the termini.

torsions but α is only in the A-form gauche$^-$ rotamer 40% of the time. The average

torsions at each position for the 20 structures is shown in Table 3.5. The ε and β torsions

are reasonably well-defined for many of the A-form residues as the standard deviations are

fairly small with the average value close to the expected A-form values. Surprisingly, the

5' side of A27 and 3' side of G26 are fairly well-defined to non-A-form conformations.

When calculations restraining every all the A-form backbone torsions except without the

restrained except the 5' side of A27, nine structures with a very low rmsd of 0.99 Å were

obtained. Furthermore, the A27 and G26 backbone torsions were again well-defined in the

same range as in Table 3.5. Thus, these backbone torsions are non-A-form to widen the

major groove and allow the argininamide to bind to G26.

Correlations among the backbone torsions have been observed in polynucleotides

with correlation coefficients ranging from 0.65 - 0.71 (Saenger, 1983). Interestingly, in

the A-form regions, the backbone torsions are found in distinct conformations depending

on the rotamer of the other surrounding backbone torsions. For example, when γ is trans β

is either trans or gauche$^+$ and only rarely gauche$^-$ (Figure 3.9). But when γ is in the A-

form gauche$^+$, the other backbone torsions are almost always in their correct rotamer. On

the other hand, when α is in the typical A-form gauche$^-$ the other backbone torsions are

almost always in only a single rotamer (Figure 3.10). These correlations are in sharp

contrast to the non-A-form regions which are sampling many possible torsions, and show

no distinct correlations.

These correlations suggest that restraining either the α or γ torsion may force all the

other backbone torsions into their correct A-form rotamers. Unlike the γ torsion, the α

torsion is not directly measurable by NMR and can only be roughly estimated from the ^{31}P
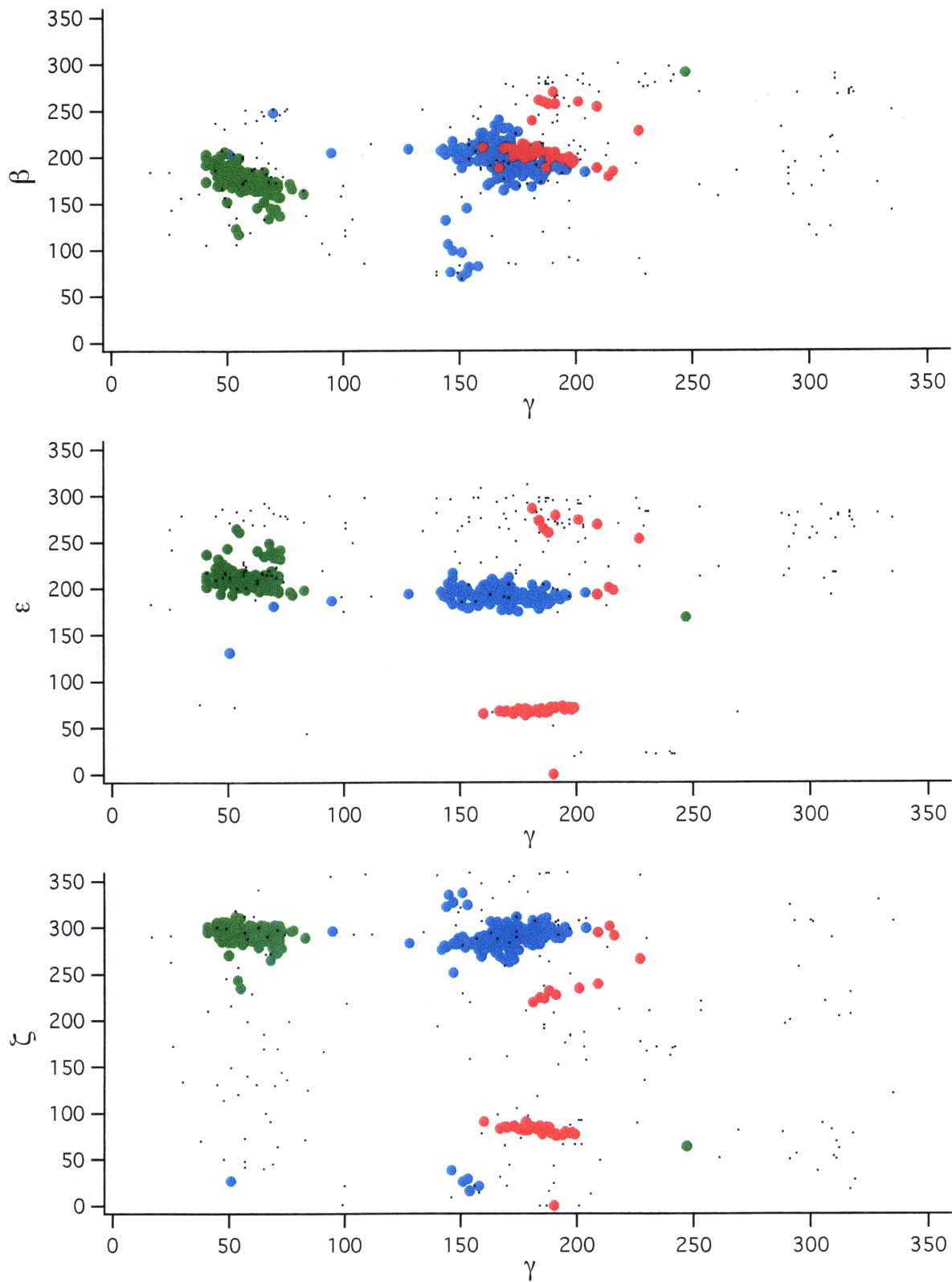
Figure 3.9: Correlations to the γ torsion. When γ is gauche$^+$ the other backbone torsions are A-form. Black dots represent non-A-form residues. Notice that γ is never gauche$^-$ for the expected A-form residues.
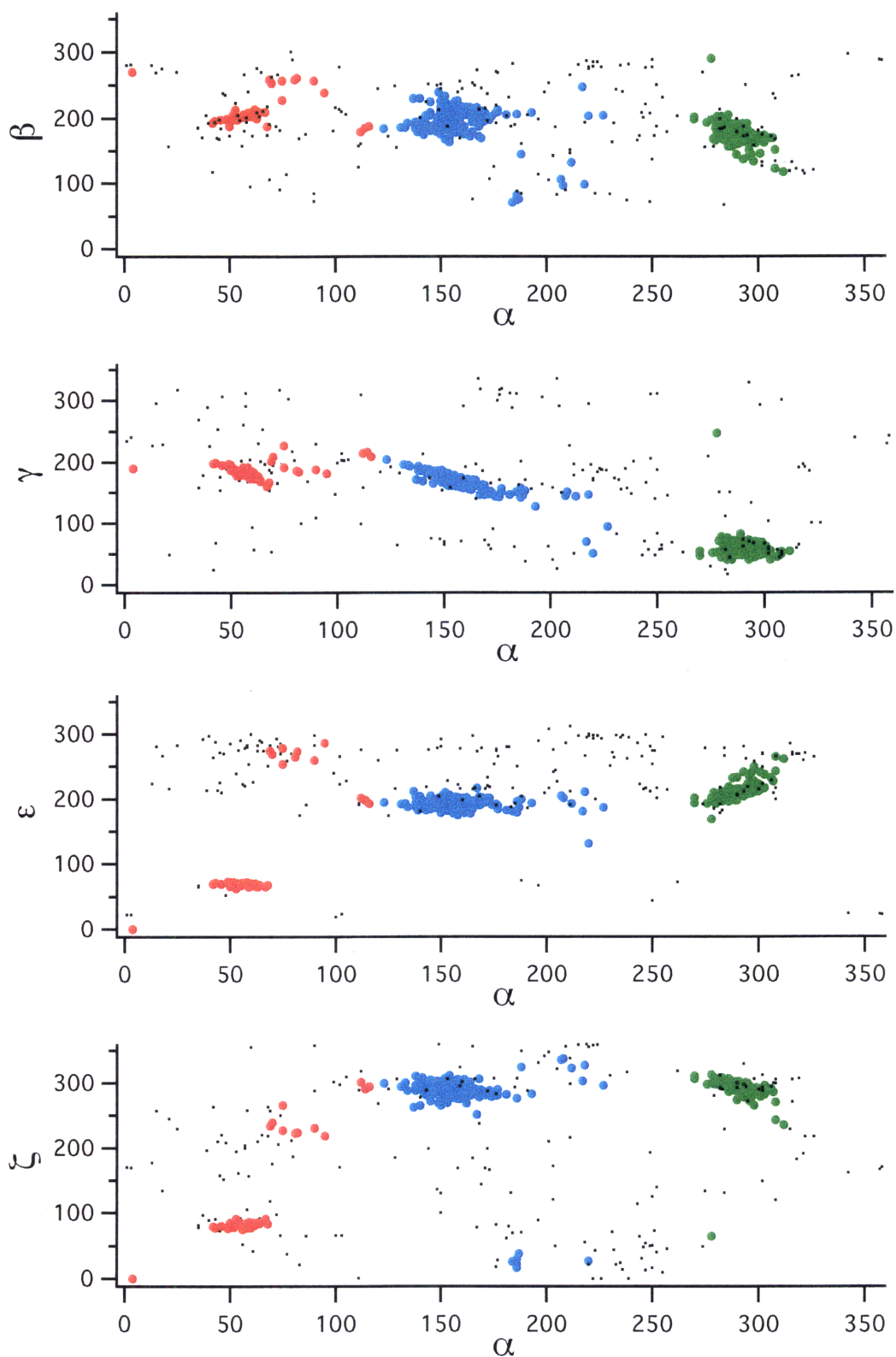
Figure 3.10: Correlations to the α torsion. Each color represents a different α rotamer: red is gauche⁺, blue is trans, and green is gauche⁻. The backbone torsions are in a single rotamer at each α conformation. Black dots represent non-A-form residues.

Table 3.6: Average torsions for each residue with rms deviations for 8 structures with α restrained.

RESIDUE	α	β	γ	δ	ε	ζ
16	N.M.	N.M.	237.92 ± 20.68	86.30 ± 0.71	186.35 ± 22.26	278.16 ± 14.26
17	287.85 ± 4.68*	179.78 ± 15.52	85.57 ± 14.65	83.54 ± 0.56	203.47 ± 1.74	293.05 ± 1.04
18	288.37 ± 1.35*	178.80 ± 2.70	56.89 ± 3.28	81.69 ± 0.30	211.94 ± 10.57	256.53 ± 26.97
19	282.95 ± 4.08*	168.79 ± 20.68	92.63 ± 20.83	82.00 ± 1.29	181.74 ± 20.50	248.74 ± 28.05
20	288.25 ± 5.17*	202.97 ± 18.71	98.65 ± 22.60	83.03 ± 0.90	198.58 ± 3.78	265.57 ± 27.68
21	282.39 ± 2.23*	185.18 ± 14.69	89.80 ± 19.91	80.62 ± 1.20	146.68 ± 21.13	214.35 ± 37.71
22	188.57 ± 41.43	179.11 ± 6.42	133.20 ± 21.64	84.35 ± 1.14	194.22 ± 32.73	223.58 ± 24.34
23	191.61 ± 42.78	233.02 ± 14.50	181.00 ± 32.13	145.19 ± 0.35	250.08 ± 10.11	101.19 ± 14.88
25	217.71 ± 33.69	160.85 ± 22.74	298.76 ± 5.99	138.79 ± 0.18	285.55 ± 3.64	104.33 ± 34.87
26	229.86 ± 17.77	234.55 ± 31.67	181.01 ± 10.36	86.81 ± 0.29	205.96 ± 2.20	294.41 ± 2.35
27	284.18 ± 4.03*	177.48 ± 2.53	66.40 ± 4.12	83.05 ± 0.73	194.02 ± 4.24	234.11 ± 35.80
28	282.07 ± 2.35*	202.21 ± 19.56	109.97 ± 25.34	84.21 ± 0.71	196.08 ± 4.85	238.24 ± 34.49
29	280.91 ± 3.03*	208.19 ± 18.30	100.43 ± 25.85	84.41 ± 0.90	187.01 ± 22.79	277.29 ± 16.15
30	204.23 ± 36.56	170.91 ± 18.72	135.56 ± 17.27	144.28 ± 1.18	255.71 ± 8.77	247.05 ± 20.56
31	205.97 ± 40.06	210.52 ± 15.82	136.78 ± 22.76	145.15 ± 0.87	254.87 ± 8.47	195.46 ± 34.23
32	197.84 ± 34.03	169.86 ± 20.68	249.16 ± 19.62	148.85 ± 1.33	251.65 ± 8.85	213.64 ± 14.26
33	167.57 ± 34.31	203.22 ± 15.25	137.27 ± 28.30	145.84 ± 0.40	260.37 ± 7.75	150.56 ± 23.74
34	210.93 ± 36.85	215.98 ± 19.14	224.97 ± 29.96	145.41 ± 0.24	239.01 ± 9.03	233.89 ± 26.14
35	152.33 ± 18.64	179.66 ± 21.95	131.81 ± 23.43	121.06 ± 9.59	257.62 ± 12.58	125.03 ± 25.25
36	194.68 ± 35.39	188.06 ± 15.37	127.15 ± 22.86	86.48 ± 0.58	204.24 ± 21.98	259.33 ± 19.24
37	300.09 ± 2.68*	155.89 ± 7.21	93.79 ± 33.32	83.06 ± 0.66	200.08 ± 3.76	258.68 ± 26.21
38	286.46 ± 2.66*	180.95 ± 15.35	91.13 ± 19.37	82.94 ± 0.62	190.49 ± 6.71	207.12 ± 38.59
39	277.19 ± 2.50*	211.35 ± 23.80	139.33 ± 26.65	84.56 ± 0.83	200.31 ± 1.97	277.97 ± 4.49
40	151.15 ± 3.49	223.61 ± 2.50	163.59 ± 2.32	87.20 ± 0.03	208.25 ± 8.62	232.02 ± 31.61
41	290.57 ± 2.88*	193.36 ± 21.40	105.73 ± 24.83	86.53 ± 0.23	204.05 ± 4.23	259.32 ± 26.88
42	287.15 ± 2.69*	183.69 ± 14.69	89.51 ± 19.88	85.27 ± 0.45	192.24 ± 6.02	213.03 ± 39.16
43	279.41 ± 2.49*	219.42 ± 21.13	128.28 ± 28.39	85.29 ± 0.77	193.39 ± 9.89	178.85 ± 36.57
44	282.18 ± 3.71*	225.51 ± 24.90	147.09 ± 28.26	85.40 ± 0.78	190.44 ± 5.62	212.70 ± 37.18
45	281.68 ± 1.62*	227.65 ± 18.05	118.77 ± 30.96	84.04 ± 1.10	144.30 ± 25.22	219.38 ± 30.86
46	164.58 ± 37.68	196.37 ± 10.79	128.26 ± 18.88	85.82 ± 0.44	N.M.	N.M.
A-form	292	178	60	88	205	292

* These residues have the α torsion restrained to the A-form gauche⁻

chemical shifts, which is not a reliable method at this time. To test whether constraining

the α torsion is sufficient to define the entire phosphodiester backbone, a series of

calculations defining the α torsion to its A-form gauche$^-$ rotamer 292 ± 30° were

performed. Eight structures with a rmsd of 1.14 Å were obtained, but the other torsions

did not consistently fall into A-form rotamers as shown in Table 3.6. Careful analysis

revealed that four of the eight structures had well-defined torsions for all the A-form

nucleotides but the other four could not be distinguished from these structures as they all

had about the same total, violation, and angle energies.

On the other hand, when the A-form γ torsions were restrained to 60 ± 30°, all the

backbone torsions consistently fell into their A-form rotamers for almost all the expected

positions as shown in Table 3.7. The six converged structures had a rmsd of 1.01 Å in

the range obtained when all the backbone torsions were restrained. Remarkably, even

though the γ torsion is broadly defined to a 60° range, many of the backbone torsions,

including γ, are defined very precisely with rmsd's less than one. Furthermore, the

torsions are centered at about their A-form rotamer further suggesting that all the backbone

torsions are correlated and fall into the lowest energy state together. Calculations were

performed to test whether the γ torsion in the context of a more limited data set with only

H1' and H2' NOEs from the ribose is sufficient to define an A-form helix. This data set

may be realistic for larger RNAs where the extreme chemical shift overlap along with line

broadening problems may inhibit full assignments of all the ribose protons (see Chapter 1).

The phosphodiester backbone is well defined for most of the constrained residues in the

five structures that converged except for residues 17-20. Although the 1.14 Å rmsd is only

slightly lower than without the γ torsion restrained, the backbone is significantly better

defined. Thus, restraining the γ torsion has profound effects on the phosphodiester

Table 3.7: Average torsions for each residue with rms deviations for 6 structures with γ restrained.

RESIDUE	α	β	γ	δ	ε	ζ
16	N.M.	N.M.	177.78 ± 38.14	87.08 ± 0.03	196.69 ± 12.93	246.95 ± 40.07
17	270.84 ± 13.27	165.67 ± 5.43	71.08 ± 4.23*	82.44 ± 0.66	200.66 ± 1.33	290.84 ± 1.00
18	288.05 ± 0.97	173.26 ± 1.69	61.71 ± 2.17*	81.11 ± 0.23	211.06 ± 7.33	286.30 ± 4.02
19	285.30 ± 1.03	160.33 ± 7.47	67.98 ± 4.19*	80.00 ± 0.90	177.30 ± 19.51	254.27 ± 31.48
20	268.36 ± 18.17	164.67 ± 1.00	77.61 ± 2.59*	83.16 ± 0.37	202.30 ± 1.39	300.00 ± 2.48
21	283.41 ± 2.44	180.45 ± 4.95	59.13 ± 2.70*	82.99 ± 0.35	185.31 ± 2.14	296.22 ± 10.73
22	178.51 ± 20.78	183.91 ± 17.70	160.77 ± 16.88	85.34 ± 1.08	135.16 ± 34.42	193.95 ± 36.67
23	215.00 ± 49.53	209.06 ± 21.70	118.87 ± 32.18	143.23 ± 0.73	227.10 ± 36.98	111.63 ± 36.61
25	227.88 ± 22.01	205.60 ± 20.93	257.29 ± 45.37	138.78 ± 0.34	287.32 ± 4.15	165.88 ± 56.20
26	197.02 ± 21.99	210.43 ± 36.08	164.35 ± 11.60	86.94 ± 0.33	188.37 ± 2.88	254.48 ± 46.21
27	163.99 ± 7.95	162.48 ± 12.00	170.59 ± 5.42	87.33 ± 0.06	214.86 ± 6.98	286.70 ± 3.60
28	287.97 ± 1.71	159.92 ± 6.97	72.82 ± 2.67*	85.18 ± 0.51	205.99 ± 8.48	292.46 ± 4.55
29	265.95 ± 9.80	196.45 ± 12.37	60.07 ± 5.41*	84.24 ± 0.66	196.54 ± 11.18	207.65 ± 50.38
30	228.60 ± 28.96	185.73 ± 17.31	161.12 ± 37.38	143.16 ± 1.10	220.30 ± 32.47	233.94 ± 20.55
31	109.82 ± 36.91	232.51 ± 11.77	167.13 ± 21.31	146.29 ± 0.36	245.65 ± 11.63	176.55 ± 43.84
32	217.16 ± 36.20	198.15 ± 22.64	280.36 ± 14.88	145.88 ± 1.69	276.52 ± 1.95	193.18 ± 14.21
33	212.38 ± 47.89	225.82 ± 22.17	188.65 ± 28.35	145.25 ± 0.71	263.66 ± 7.50	151.87 ± 49.01
34	164.69 ± 26.28	216.99 ± 20.38	199.49 ± 36.55	145.50 ± 0.78	242.02 ± 11.55	169.17 ± 34.04
35	164.49 ± 22.66	155.19 ± 21.95	104.03 ± 20.52	113.27 ± 12.37	223.38 ± 18.95	43.97 ± 13.56
36	179.29 ± 25.49	154.98 ± 27.03	146.74 ± 28.04	87.20 ± 0.10	222.18 ± 3.65	290.22 ± 3.31
37	297.06 ± 2.82	178.60 ± 4.82	45.33 ± 3.06*	84.57 ± 0.45	210.09 ± 3.18	283.00 ± 2.88
38	293.59 ± 2.32	164.68 ± 1.95	64.18 ± 2.25*	82.39 ± 0.28	198.09 ± 0.90	292.17 ± 1.22
39	282.95 ± 2.58	162.02 ± 1.17	79.90 ± 2.48*	82.69 ± 0.32	204.58 ± 6.00	237.60 ± 38.04
40	163.25 ± 13.20	185.58 ± 24.38	160.88 ± 5.50	87.38 ± 0.09	228.99 ± 4.22	277.85 ± 2.70
41	294.33 ± 1.93	150.27 ± 6.11	67.71 ± 3.98*	83.97 ± 0.58	205.36 ± 3.86	294.60 ± 1.86
42	278.29 ± 10.78	185.96 ± 8.30	61.42 ± 5.85*	82.94 ± 0.37	201.10 ± 1.07	296.75 ± 1.21
43	283.86 ± 0.72	175.51 ± 0.91	62.56 ± 1.63*	81.42 ± 0.18	204.25 ± 1.15	286.01 ± 2.41
44	287.75 ± 1.09	170.43 ± 1.69	62.73 ± 1.18*	82.56 ± 0.20	198.99 ± 0.90	289.64 ± 0.69
45	285.75 ± 0.80	176.44 ± 0.88	59.18 ± 1.05*	82.65 ± 0.10	219.64 ± 10.34	271.07 ± 8.64
46	290.42 ± 4.86	154.47 ± 9.92	60.89 ± 1.06*	83.99 ± 0.41	N.M.	N.M.
A-form	292	178	60	88	205	292

* These residues have the γ torsion restrained to the A-form gauche⁺

backbone and improves the quality of RNA NMR structures. Measuring the γ torsion is difficult because of spectral overlap, but stereospecific selective labeling techniques may solve this problem. These results suggest that this effort is worthwhile because this restraint will greatly improve the quality of structures.

The total and violation energies were used to determine which structures converged, but the importance of the angle energy term became clear when the backbone torsions were restrained. When the ζ and α torsions around a particular phosphor were not in their A-form rotamer, the C3'-O3'-P and C2'-C3'-O3' angles were significantly violated by 4-5°, resulting in angle energy terms a few kcal/mol higher than the lowest energy structures, even though the total and violation energies were in the same range as the lowest energy converged structures.

3.7 Does the Structure Make Sense?

All the important nucleotides identified in binding studies play a critical role in forming the argininamide binding pocket in the HIV-2 TAR-argininamide structure. The argininamide binds to the major groove side of G26 where a strong N7 modification interference is observed (Churcher et al., 1993; Hamy et al., 1993; Weeks & Crothers, 1991). The A27 N7, which gives a very strong modification interference with DEPC, interacts with U23 in the proposed base triple, forming the scaffold for the argininamide binding site (Churcher et al., 1993; Hamy et al., 1993; Weeks & Crothers, 1991). Chemical modification of the N6 position of A27 with a methyl group reduces Tat peptide binding by a factor of five, which represents half of the specific free energy of binding (Hamy et al., 1993). All the modification interference and mutational binding data of this interaction are consistent with the structure presented here. Thus, the U23 and A27 are part of a critical RNA structural element required for argininamide binding while the G26 interacts directly with the argininamide. Furthermore, this structure also explains why the

binding data are so similar for both the small ligand, argininamide, and Tat protein binding

to TAR. Considerable evidence suggests that the TAR conformation in the argininamide

complex is relevant to Tat binding, and one of the arginines in Tat, most likely R52, should

hydrogen bond directly to G26. Thus, a small ligand is able to interact specifically with

TAR RNA by recognizing a specific conformation with multiple contacts. Undoubtedly,

other contacts to the protein exist, but these bulge nucleotides, which are all involved in

argininamide binding, constitute the most critical determinants for both specific Tat and

argininamide binding.

Chapter 4: HIV-2 TAR CGC Base Triple Mutant

4.1 Design of an Isomorphic Base Triple Mutant

Chapters 2 and 3 discuss the structure determination of the two base bulge HIV-2 TAR-argininamide complex, which exhibited higher quality spectra than the previously studied three base bulge HIV-1 TAR, allowing for an even larger NOE data set to be obtained. Even though the NMR spectra of the two base bulge HIV-2 TAR-argininamide complex are superior to those of HIV-1 TAR, the effects of dynamics are still present, limiting the number of weak NOEs observed and preventing the unambiguous identification of the proposed base triple. The expected U23 imino proton may simply exchange too rapidly to be observed, as is frequently observed for terminal base pairs.

To determine if the base triple is required for argininamide binding, an isomorphic C-G•C$^+$ base triple mutant was introduced in the three base bulge HIV-1 TAR and was found to bind argininamide in the expected pH dependent manner (Puglisi et al., 1993). Binding only occurred when the pH was lowered to ~5.5 where the cytidine is protonated and thus able to form a base triple. This mutant helped demonstrate that the base triple is required for specific argininamide binding. However, the proton resonances at the interface of the base triple were unobservable and no NOEs between the C23 and the G27-C38 base pair were observed due to the dynamic nature of the complex. Previous studies showed that an isomorphic base triple mutant TAR can be designed that binds argininamide.

Chapter 2 describes some of the improved dynamic properties of the TAR - argininamide complex allowing for a high resolution structure determination of this complex. Although this work provided significant indirect evidence for base triple formation upon argininamide binding, no direct NOE evidence indicative of the presence of a base triple in the HIV-2 TAR argininamide complex was observed. We reasoned that

```
        G  G              G  G
    U         G       U         G
    C         A       C         A
      C29•G36            C29•G36
      G • C              G • C
    ┌─────────┐        ┌─────────┐
    │ A • U   │        │ G • C   │
    └─────────┘        └─────────┘
      G • C              G • C
  U    26  39       U    26  39
 ┌───┐             ┌───┐
 │ U │             │ C │
 │ 23│ A •U        │ 23│ A •U
 └───┘  22  40     └───┘  22  40
      G • C              G • C
      A • U              A • U
      C • G              C • G
      C • G              C • G
      G • C              G • C
      G•C                G•C
       16  46             16  46

      HIV-2               Triple
      TAR                 Mutant
```

Figure 4.1: Sequence and secondary structure of the wild-type HIV-2 TAR and the HIV-2 TAR CGC triple mutant. The critical nucleotides required for arginine and Tat binding in the bulge region are highlighted by the rectangles.

replacement of the U38-A27•U23 base triple with an isomorphic C38-G27•C23$^+$ base triple

in the context of the two base bulge HIV-2 TAR may allow us to observe critical

resonances indicative of base triple formation in the argininamide complex. Typically, the

cytidine amino proton resonances are sharper than uridine imino proton resonances and,

therefore, may be observable upon hydrogen bonding in the base triple of TAR. Also, the

extra G-C base pair in the upper stem may help the overall stability of the complex. Thus,

we constructed the two base bulge HIV-2 TAR CGC triple mutant shown in Figure 4.1.

The same pH dependent argininamide binding was observed for the HIV-2 TAR

CGC as was observed for the HIV-1 TAR base triple mutant. However, additional proton

resonances due to hydrogen bonding at the interface of the proposed base triple were

observed for the first time. These protons exhibit a number of inter-residue NOEs that

position C23 and G27 near each other, consistent with base triple formation. This work

provides direct NMR evidence of the base triple structure in the HIV-2 TAR-argininamide

complex.

4.2 Materials and Methods

Sample Preparation   The HIV-2 TAR CGC RNA was synthesized by *in vitro*

transcription by T7 RNA polymerase and purified by denaturing gel electrophoresis

(Milligan et al., 1987). The DNA template for T7 transcription was: 5' - GGC CAG AGG

GCT CCC AGG CCC AGT CTG GCC TAT AGT GAG TCG TAT TA- 3' and

transcriptions were primed with 10 mM GMP. Samples were dialyzed for >48 hours

against NMR buffer (10 mM sodium phosphate, pH 6.4 at 25°C, 50 mM sodium chloride

and 0.1 mM EDTA). The final concentration in 600 μl was approximately 1.5 mM. The

sample was then titrated with argininamide (Sigma) in the NMR tube to a final total

concentration of 6 mM, and the pH was lowered to 5.4 with HCl.

NMR spectroscopy NMR experiments were recorded on a Varian INOVA 600 MHz spectrometer, Varian Unity Plus 750 MHz spectrometer or a 500 MHz spectrometer constructed at the Francis Bitter Magnet Lab. All spectra were processed and analyzed using NMRPipe and PIPP on a Silicon Graphics Indy workstation (Delaglio et al., 1995; Garrett et al., 1991). Proton NMR spectra in H_2O were recorded at a variety of temperatures and mixing times including 50, 100, and 200 ms. NOESY spectra used a WATERGATE 3-9-19 water suppression scheme with a selective E-BURP-1 flip-back pulse (Green & Freeman, 1991; Lippens et al., 1995; Piotto et al., 1992). Sweep widths were 12000, 14000, and 16000 Hz at 500, 600, and 750 MHz, respectively, with 4096 x 512 points collected. Some spectra were recorded with sweep widths of 14000 x 8000 Hz which increased the digital resolution. Data were zero filled to 4K x 2K real points and apodized using Gaussian-Lorentzian functions in both dimensions before Fourier transformation.

Standard NOESY, TOWNY, and DQF-COSY spectra in 99.996% ^2H_2O were recorded at 25°C. To monitor NOE buildups, spectra were recorded at 500, 600, and 750 MHz with mixing times of 50, 100, 200 and 400 ms. Sweep widths were 5500 Hz, 6000 and 8000 Hz at 500, 600, and 750 MHz, respectively. Some spectra were obtained with 4096 x 1024 real points which aided the assignment of some crowded resonances.

4.3 Identification of a Base Triple

As previously observed in the studies of the HIV-1 TAR CGC, argininamide does not bind significantly at pH 6.4. However, at pH 5.4, the U40 imino proton resonance significantly sharpens and the G26 and G21 imino resonances shift as is observed for argininamide binding to wild-type TAR. In addition, a number of new peaks appear in the imino and amino proton regions, as shown in Figure 4.2. These exchangeable resonances are generally not observable under these conditions unless they are involved in a hydrogen bonding interaction. The peak at 15.55 ppm is too broad to be observed in 2D ^1H-^1H
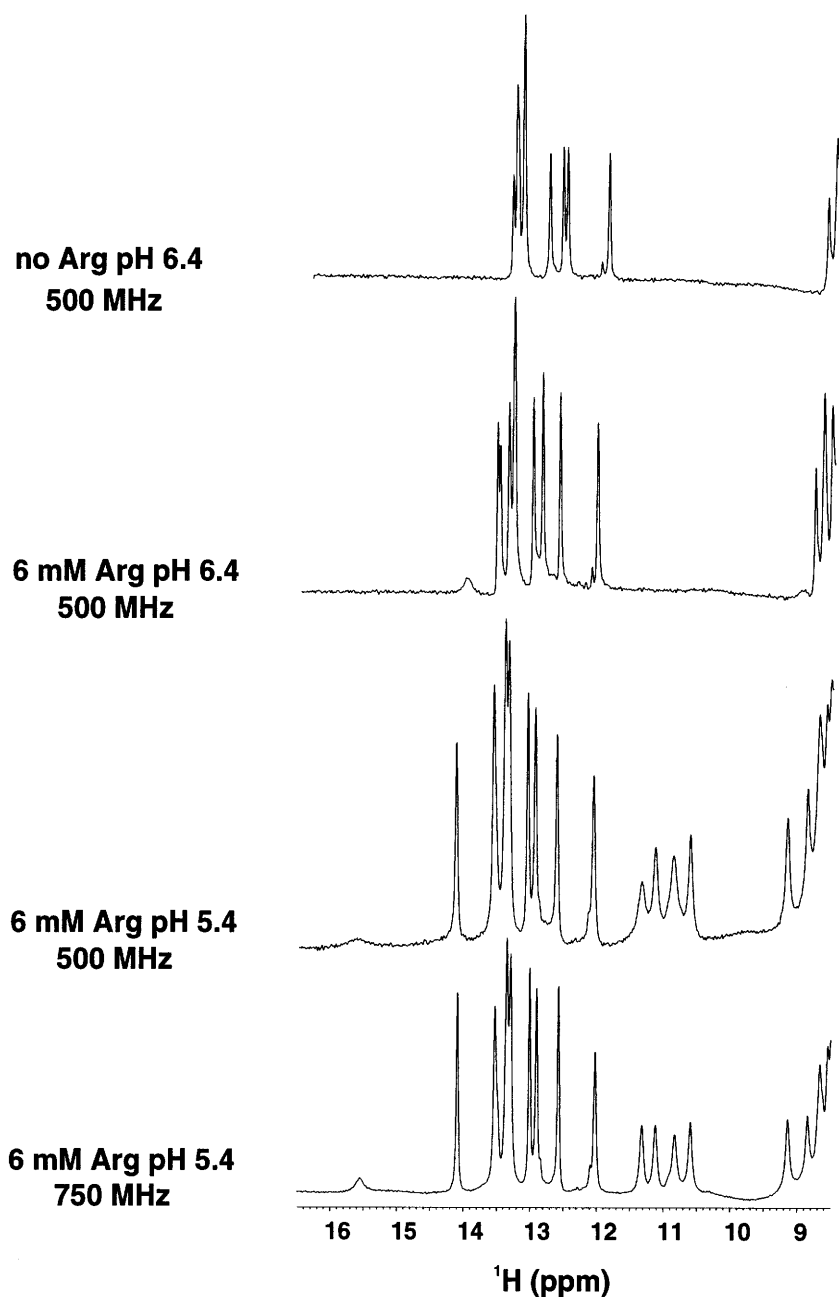
Figure 4.2: The imino-amino proton region of HIV-2 TAR CGC shows a number of new resonances upon addition of argininamide and the lowering of the pH to 5.4. At pH 6.4, a few changes are observed upon addition of argininamide. At pH 5.4, the U40(H3) at 14.11 ppm sharpens significantly and the C23(H3) resonance appears at 15.55 ppm. In addition, the C23(H42/H41) resonances are also observable at 10.61 and 9.17 ppm. The chemical shifts of the C23 amino and imino protons are in the range for protonated cytidines found in C-G·C⁺ base triples.

NOESY experiments but its chemical shift is in the exact range observed for protonated cytidine imino protons found in base triple structures in DNA triplex NMR studies (Dongen *et al.*, 1996; Macaya et al., 1992; Santos *et al.*, 1989). The remaining new resonances are easily identified as cytidine amino protons in 750 MHz NOESY spectra. The cytidine amino protons are also observed at chemical shifts typical for a base triple structure (Dongen et al., 1996; Macaya et al., 1992; Santos et al., 1989).

The NOEs observed to the C23 amino protons in a H_2O NOESY are shown in Figure 4.3. Since the chemical shifts for most of the molecule are very similar to the wild-type HIV-2 TAR, resonance assignments were straightforward. Most of the exchangeable imino protons and amino protons that were observed in the high resolution HIV-2 TAR-argininamide structure were assigned, along with most of the H8/H6/H5 and H1' resonances, as shown in Table 4.1. Unlike the wild type HIV-2 TAR spectra, the critical bulge nucleotides could be identified along with most of the upper and lower stems. This chemical shift dispersion allowed the key resonances in the bulge region to be assigned as many of the A-form resonances have almost identical chemical shifts to the wild type HIV-2 TAR-argininamide complex and the anomeric-aromatic sequential walk assignable by homonuclear spectroscopy as shown in Figure 4.5. Complete assignments were not possible without isotopic labeling, but the salient features of the structure could be identified. The C38 and G27 resonances were readily assigned by the observation of standard A-form sequential NOEs. The C23 amino protons were assigned based on very strong NOEs to the C23 H5, which in turn gives NOEs to the C23 base and ribose. The C23 sugar pucker is C2'-endo, as in the wild-type HIV-2 TAR, which allows assignment via ^1H-^1H TOCSY and DQF-COSY experiments (data not shown) (Varani & Tinoco, 1991).

Some weaker NOEs that were observed in the HIV-2 TAR-argininamide complex were not observed in the triple mutant spectra. Most of these differences can be attributed

Table 4.1: Proton chemical shifts (ppm) for HIV-2 TAR CGC RNA bound to argininamide.

Residue	H1/H3	H42/H41	H8/H6	H5/H2	H1'
G16		---	8.09	---	5.78
G17	13.51	---	7.69	---	5.93
C18		8.70*	7.72	5.27	
C19		8.44,6.95	7.74	5.19	5.55
A20		---	8.00	6.91	5.92
G21	12.60	---	7.10	----	5.56
A22	---	---	7.76	7.15	5.99
C23	15.55	10.61,9.17	7.76	5.77	5.85
U25		---	7.89	5.99	6.05
G26	12.93	---	7.79	---	5.94
G27	13.02	---	7.58	---	5.95
G28	13.56	---	7.43	---	5.87
C29		8.37	7.59	5.17	5.54
C30			7.69	5.55	
U31		---	7.74	5.84	
G32		---		---	
G33		---		---	
G34		---	7.94	---	
A35		---	8.42	8.20	5.98
G36	13.25	---	7.45	---	5.55
C37		8.88,7.03	7.89	5.23	5.55
C38		8.56,7.01	7.88	5.45	5.63
C39		8.51,7.42	7.41	5.58	5.45
U40	14.11	---	8.02	5.52	5.57
C41		8.36,7.10*	7.94	5.67	5.50
U42	13.35	---	7.86	5.34	5.46
G43	12.05	---	7.86	---	5.77
G44	13.31	---	7.35	---	5.70
C45		8.66,7.00	7.66	5.18	
C46					

The errors for the chemical shift are ±0.01. The exchangeable proton chemical shifts were assigned at 2°C while all the non-exchangeable data were assigned at 25°C.

* indicates the chemical shift was assigned at 15°C as the peaks were overlapped at 2°C.
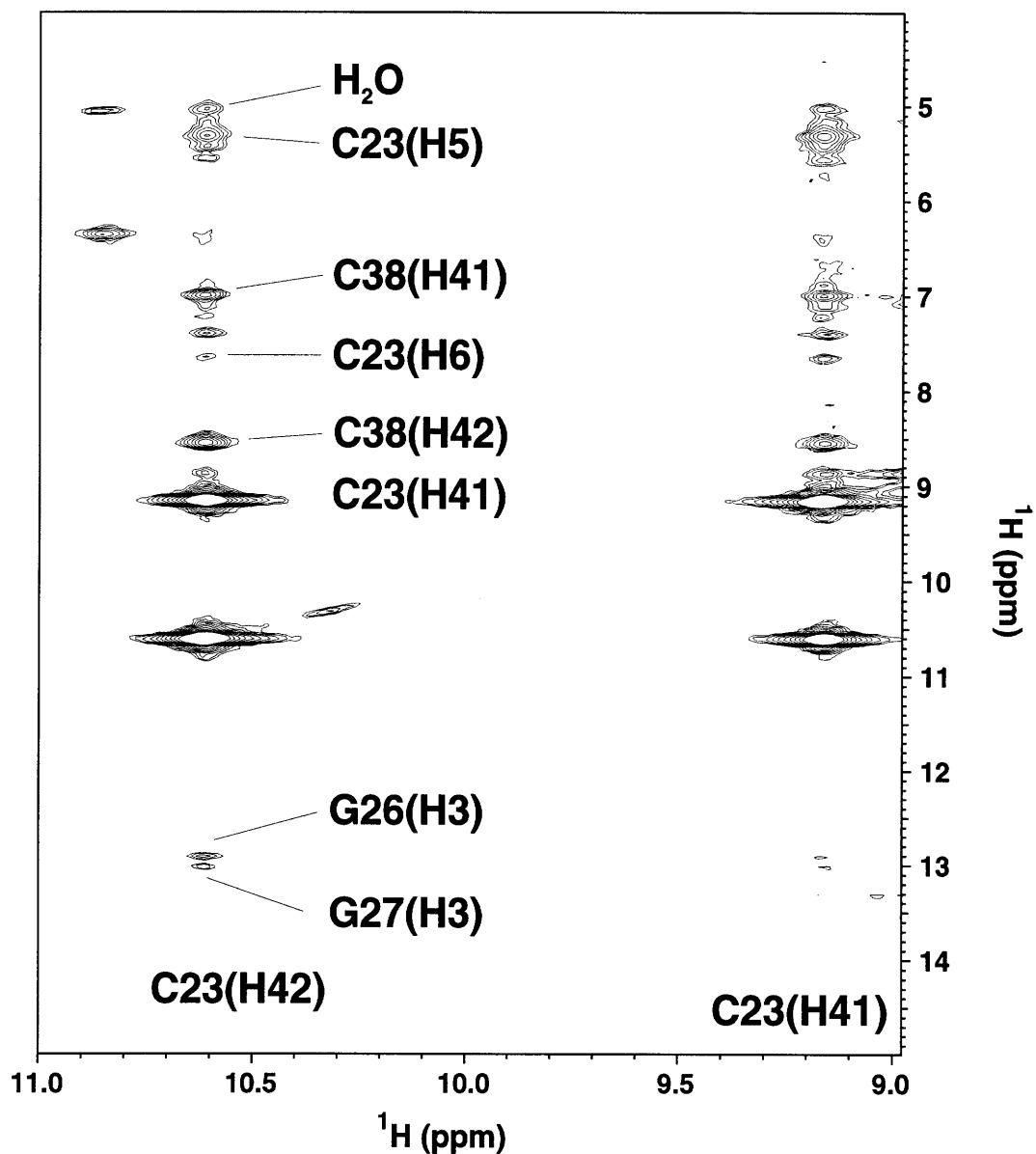
Figure 4.3: The internucleotide NOEs to the C23 amino protons are shown in a 750 MHz H_2O NOESY at 2°C. A number of NOEs between the C23 and the G26, G27, and U38 position the C23 near the G27, consistent with a base triple structure.
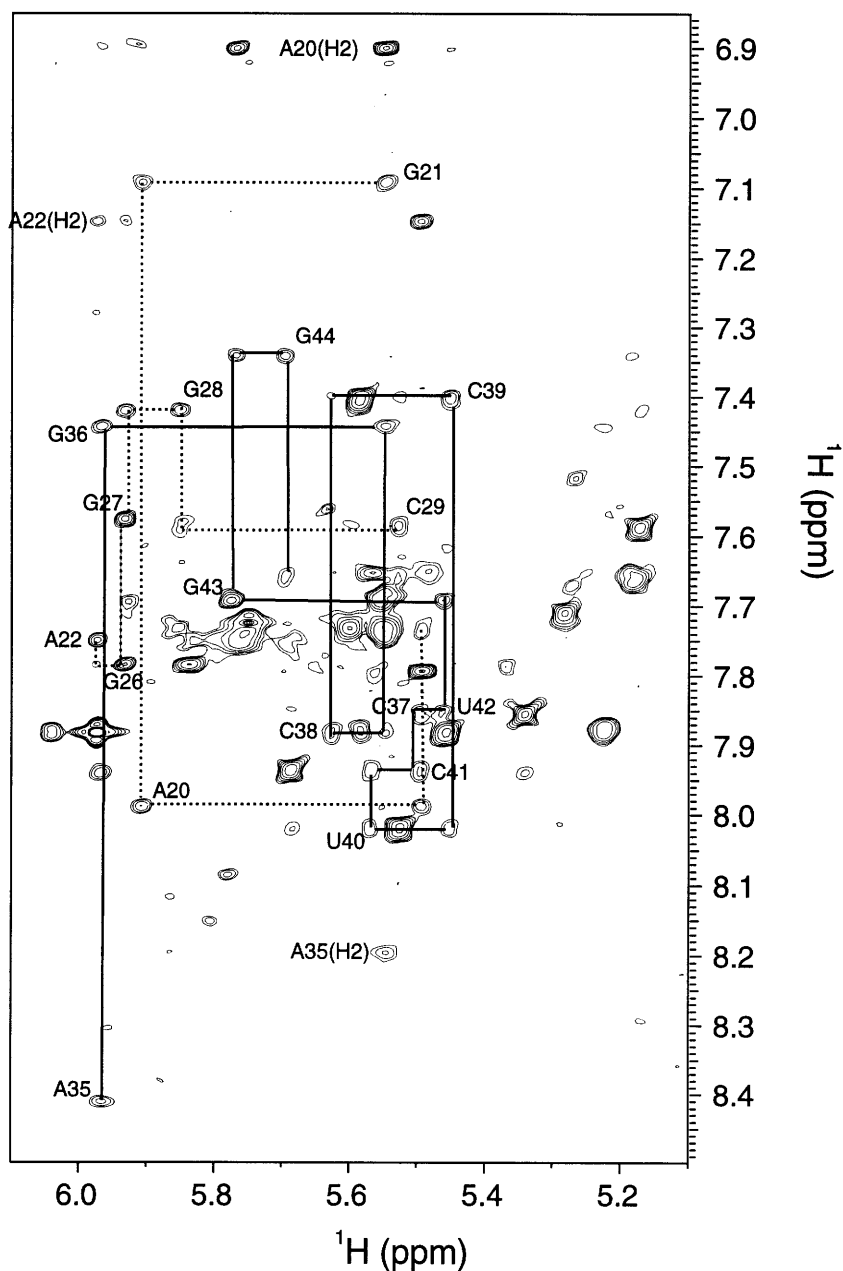
Figure 4.4: Aromatic-anomeric region of a 600 MHz NOESY (mix=100ms) of the HIV-2 TAR CGC - argininamide complex in 99.996% ^2H_2O. The sequential assignment pathway is highlighted by dashed and solid lines for the C19-G21, A22-C29 and the A35-C45 nucleotide pathways, respectively. The numbers indicate intranucleotide H8/H6 to H1' NOEs.

to the use of multidimensional heteronuclear NMR for the HIV-2 TAR assignments, which was not replicated here for the base triple mutant. Intermolecular NOEs to the argininamide guanidium protons were not detected because these resonances were broader in the HIV-2 TAR CGC spectra. Also, the weak C39(H42/H41)-Arg(Hδ) intermolecular NOEs were not observed. Spectral crowding inhibited the unambiguous assignment of intermolecular NOEs between the A22(H2') and A22(H3') and the Arg(Hε) even though candidate peaks are observed.

All the observed NOEs for the triple mutant are consistent with the base triple structure of the wild type HIV-2 TAR-argininamide complex. In fact, a similar set of intermolecular and internucleotide NOEs are observed in both the HIV-2 TAR CGC and the wild-type HIV-2 TAR complexes indicating that argininamide is binding to both molecules in the same manner. Table 2 lists the critical NOEs in the bulge region that are observed in the HIV-2 TAR CGC argininamide complex. Important intermolecular NOEs that are found in the wild-type HIV-2 TAR-argininamide complex including NOEs between the C23-H5 and the argininamide Hγ and Hδ are also observed in the triple mutant.

Significantly, NOEs from the C23(H42/H41) to the G27(H1) and C38(H42/H41) are observed positioning C23 near G27, as shown schematically in Figure 4.5.

Summary We have investigated an isomorphic C38-G27•C23$^+$ base triple mutant TAR which provides direct NMR evidence for base triple formation in the TAR - argininamide complex. The C23 imino and amino protons were observed upon the addition of argininamide at pH 5.4, suggesting that these C23 amino and imino protons are involved in hydrogen bonding interactions. The chemical shifts of the C23 amino and imino are in the exact range found for base triple structures in numerous nucleic acid triplex studies. Furthermore, a number of NOEs were identified, positioning the C23 near the G27 consistent with base triple formation. Finally, a similar set of intermolecular and internucleotide NOEs are observed in both the HIV-2 TAR CGC and the wild type HIV-2

145

TAR complexes indicating that argininamide is binding to both molecules in the same manner.

Additional evidence for a base triple structure has emerged from the Rev aptamer NMR structure. The Rev aptamer has a TAR bulge on one end that increased the binding affinity to the Rev peptide compared to the wild type RRE sequence. A TAR-like set of mutants and chemical modification interference patterns were also observed suggesting that the TAR bulge may also be recognizing an arginine like the HIV TAT-TAR interaction. The ensuing solution structure of the Rev aptamer found direct evidence of a U-A•U base triple that promoted arginine binding to a G-C base pair just like TAR. The studies presented here along with the building evidence in other systems, strongly suggests that a base triple is forming in the HIV TAR argininamide complex.
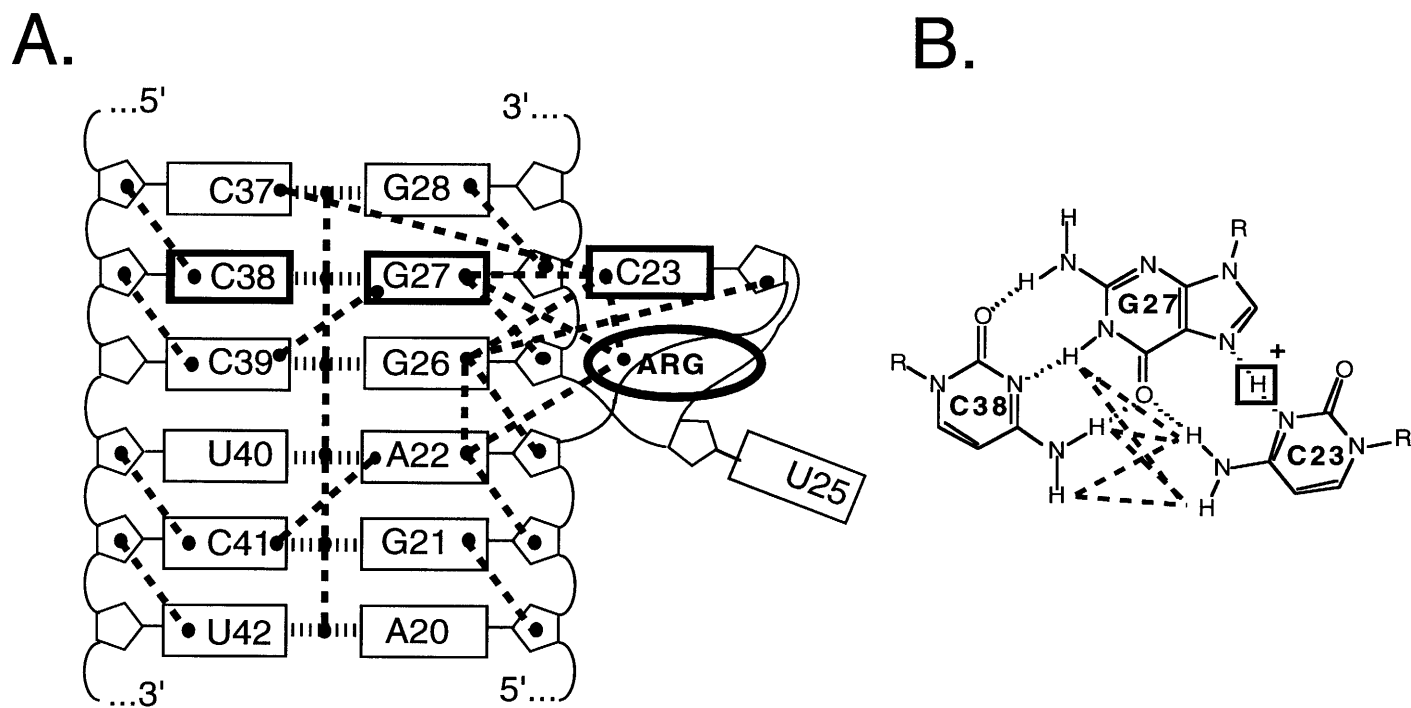
Figure 4.5: Summary of NMR data for the HIV-2 TAR-CGC triple mutant bound to argininamide. **A.** Bases are indicated by rectangles and ribose sugars by pentagons. Ribose, base H8/H6, and imino protons are represented by dots within pentagons or on the outside of the base, respectively. Base-pair hydrogen bonding is shown by dashed lines between bases. Observed internucleotide NOEs are indicated by thin dashed lines; NOEs that indicate base-triple formation are shown by thicker dashed lines. The nucleotides that form the base triple are highlighted. **B.** The C38-G27•C23$^+$ is shown with the key internucleotide NOEs that are observed that position the C23 next to the G27 shown by the dashed lines. The C23 imino proton is highlighted by the box.

Table 4.2: Key intermolecular and internucleotide NOEs in the HIV-2 TAR CGC bound to argininamide.

Arg(Hβ)	-C39	H42	C23	H42/H41*	-G26	H1
		H41			-C37	H42
					-C38	H42
						H41
Arg(Hγ)	-A22	H8			-C39	H42
	-U23	H5				H41
		H6	G26	H8	-A22	H1'
Arg(Hδ)	-A22	H8				H2
	-U23	H5			-C23	H1'
	-G27	H8	G26	H1'	-A22	H2
Arg(Hε)	-U23	H5				
		H6	G27	H1	-C23	H42
						H41

* All these NOEs were also observed for the wild-type HIV 2 TAR-argininamide complex except those to the C23 H42/H41 protons, which have no counterparts in the wildtype RNA sequences.

Chapter 5: Dynamics of TAR-Argininamide

5.1 Motivation for Dynamics Studies

In the previous chapters, the improved NMR spectra of the two base bulge HIV-2 TAR were used to solve the high resolution structure. Structures, however, are not static and it is important to understand how the intrinsic internal dynamics may play a role in the molecule's function. Furthermore, conformational changes are important parts of many macromolecular complexes, especially RNA, and local internal motions may play a role in these changes. Also, recent advances have calculated thermodynamic parameters from internal dynamics measurements, which likely play a role in the stability of the macromolecule. In addition, knowledge of how the different spin systems are relaxing may help design improved pulse sequences for future structure determinations. Only a few dynamics studies have been performed on isotopically labeled RNA. One study focused on ^{13}C measurements of HIV-1 TAR in the absence of argininamide and interpreted the data in term of the Lipari-Szabo model. The other study focused on ^{15}N imino measurements of a RNA tetraloop where a slightly different behavior for the non-canonical base pair compared to the A-form Watson-Crick base pair residues was observed.

This study examines both the ^{15}N and ^{13}C dynamics behavior in isotopically labeled RNA samples of the HIV-2 TAR-argininamide complex. TAR is an excellent model system because of its structure has been solved and it has a variety of basic RNA structural elements with a hexanucleotide loop, base triple, internal bulge, and A-form regions. The excellent proton spectral dispersion of the imino resonances allows collection of one-dimensional versions of the relaxation measurements to be collected, permitting all the experiments to be collected in a few hours with a large number of relaxation times. Unfortunately, the imino resonances do not sample a variety of RNA structure, because they are generally only observable when they are hydrogen bonding as part of a base pair.

The amino resonances are not examined because of the more complicated interpretation of the rotating amino proton. On the other hand, there are proton-carbon spin systems available on every ribose and base of each nucleotide. The purines are of particular interest because the H8-C8 and H2-C2 spin systems are not next to any other carbons, presenting a unique opportunity to study isolated proton-carbon spin systems in macromolecules, in the absence of ^{13}C-^{13}C dipolar coupling.

Full spectral density mapping allows the spectral densities to be calculated in the absence of any mechanical motional models or analytical models from the measurement of six relaxation rate constants (Peng & Wagner, 1992; Peng & Wagner, 1992b; Peng & Wagner, 1994; Peng & Wagner, 1995). One of the initial motivations for this work was the observation that adenine H2-C2 spin systems are very isolated as it gives only a few NOEs in an A-form RNA helix reducing the proton-proton dipole-dipole interactions that interfere with accurate measurements of the $R_H(H_Z)$ and $R_{HX}(2H_ZX_Z)$ relaxation rate constants. One of the central problems of the full spectral density mapping approach is the complication introduced by the surrounding proton spins around each H-X spin system, invalidating the two-spin assumption implicit in all the measurements. An alternative method to determine the spectral densities is a reduced spectral density approach which avoids measurements of the spin assumption, where $J(\omega_H)$ is assumed to be equal to

$J(\omega_H \pm \omega_C)$ (Farrow et al., 1995; Ishima & Nagayama, 1995), which is a reasonable assumption for nitrogen but may not work as well for carbon as ω_C is 25% of ω_H compared to 10% for ω_N.

The full spectral density method did not work well for the adenine(2), as physically impossible negative values of $J(\omega_H)$ were calculated. The reduced mapping approach gave $J(0)$ values that were very similar to the values obtained from full spectral density mapping. Both the ^{15}N and ^{13}C data reveal that the relaxation rates and spectral densities have about

the same values for all the A-form residues. Surprisingly, only the U40 imino resonance showed evidence of conformational exchange. Some flexibility is observed in the hexanucleotide loop and U25 of the bulge. Surprisingly, the bulge nucleotide, U23, is not experiencing fast internal motions, which the previous NMR studies suggested.

5.2 Materials and Methods

The two isotopically labeled samples used for the structure determination were used for these measurements and all assignments have been made as discussed in Chapter 2. ^{15}N NMR experiments were performed at a variety of temperatures as discussed in the text on either a 600 MHz or 750 MHz Varian INOVA spectrometer equipped with a triple resonance shielded z-gradient probe. Pulse sequences for the measurement of ^{15}N longitudinal (R$_1$), transverse (R$_2$) relaxation rates and the heteronuclear NOE in NH spin systems were adopted from Dayie and Wagner (Dayie & Wagner, 1994). Suppression of the solvent resonance was achieved by 3-9-19 WATERGATE. Both one and two dimensional versions were performed with a recycle delay of 4.0 seconds for the R$_1$ and R$_{1\rho}$ experiments. One dimensional experiments were collected with 1024 real points at a sweep width of 6000 Hz with the transmitter set to the middle of the imino proton and nitrogen regions. For the 1D experiments, the longitudinal R$_1$ experiments used relaxation delays of 10, 40, 80, 120, 240, 320, 600, and 1200 ms while the R$_2$ experiments used relaxation delays of 4, 8, 12, 16, 24, 32, 48, 64, 80, 88 ms with a spin lock radio-frequency field of 3 kHz with the transmitter set to 155.27 ppm, which is between the U and G nitrogen imino chemical shifts. For the 2D experiments, longitudinal R$_1$ experiments used relaxation delays of 10, 40, 80, 320, 640, 1280 ms and while the R$_2$ experiments used relaxation delays of 2, 5, 10, 20, 40, 60, 80 ms. Heteronuclear cross-relaxation rates require recording of steady-state NOE intensities. Two pairs of spectra were recorded for the steady-state NOE intensities. One pair involved proton saturation to achieve the steady-state intensity for four seconds, while the other pair consisted of the

control spectra with no saturation to obtain the Zeeman intensity. The rate constants

$(R_C(H_Z \leftrightarrow N_Z)$ were then calculated from the ratio $NOE=(\gamma_H/\gamma_N)[R_N(H_Z \leftrightarrow C_Z)/R_N(N_Z)]$.

Two dimensional experiments were acquired with 1024 real points and 64 real

points for the 2D experiments. The one-dimensional data was processed and analyzed

using VNMR software while 2D spectra were processed in nmrPipe (Delaglio et al., 1995)

and analyzed in NMRView (Johnson & Blevins, 1994). Data sets were zero filled once

and apodized using a 90 shifted squared sine bell. Rates were fitted by the Levenburg-

Marquardt algorithm assuming a monoexponential decay. Error estimates were obtained

using a Monte Carlo procedure described by Peng and Wagner (Peng & Wagner, 1992b).

All ^{13}C measurements were collected at 25°C on a 600 MHz Varian INOVA

spectrometer equipped with a triple resonance shielded z-gradient probe. Six relaxation rate

constants were measured, as outlined by Peng and Wagner (Peng & Wagner, 1992), on the

aromatic the H-C spin systems. Longitudinal ^{13}C relaxation rate, $R_C(C_Z)$, used relaxation

delays of 10, 40, 80, 160, 320, 640, and 1280 ms. In-phase ^{13}C transverse relaxation

rate, $R_C(C_X)$, measurements used relaxation delays of 4, 8, 16, 24, 32, and 48 ms.

Measurements of transverse antiphase, $R_{HC}(2H_ZC_X)$, used delays of 8, 16, 24, 32, 48, and

64 ms. Longitudinal two-spin order, $R_{HC}(2H_Z C_Z)$, used delays of 4, 16, 32, 64, 128, and

400 ms. Measurement of proton longitudinal relaxation rate measurements, $R_H(H_Z)$, used

delays of 8, 16, 32, 64, 100, 200, and 400 ms. All these spectra were collected with a

recycle delay of 5 seconds. Heteronuclear NOE spectra recorded with proton saturation

utilized a relaxation delay of 3s, followed by a 3 s period of saturation, while spectra

recorded in the absence of saturation employed a recycle delay of 6 s. Constant time data

were collected for the $R_C(C_C)$, $R_c(C_x)$, $R_{HC}(2H_ZC_Z)$, and NOE relaxation rate constants.

Representative fits are shown in Figure 5.1.

The spectral densities can be calculated from the longitudinal, transverse, and

heteronuclear cross-relaxation NOE relaxation rate constants by assuming that
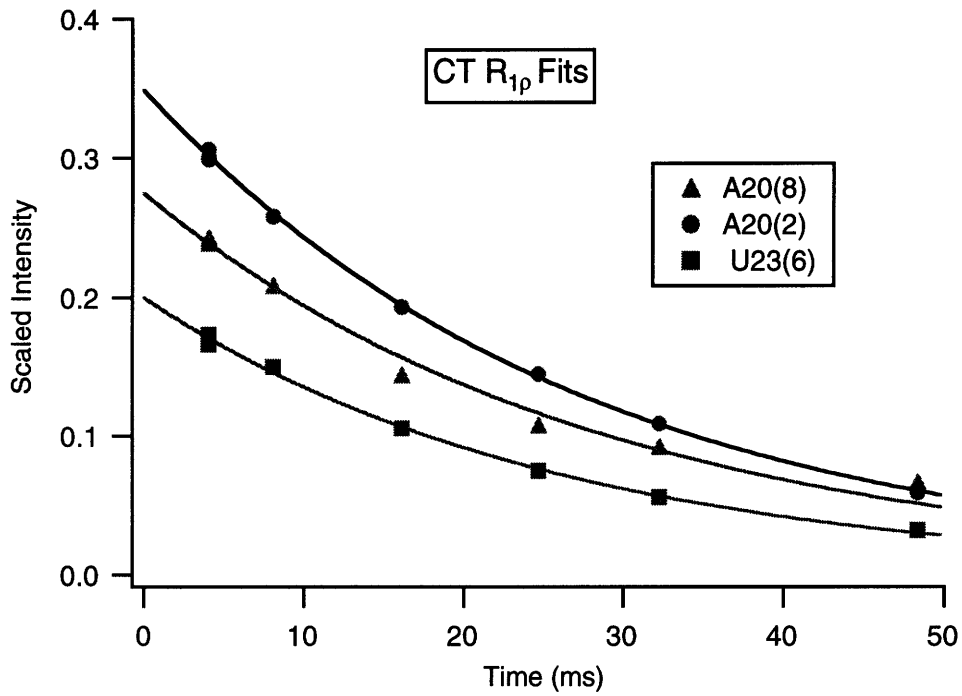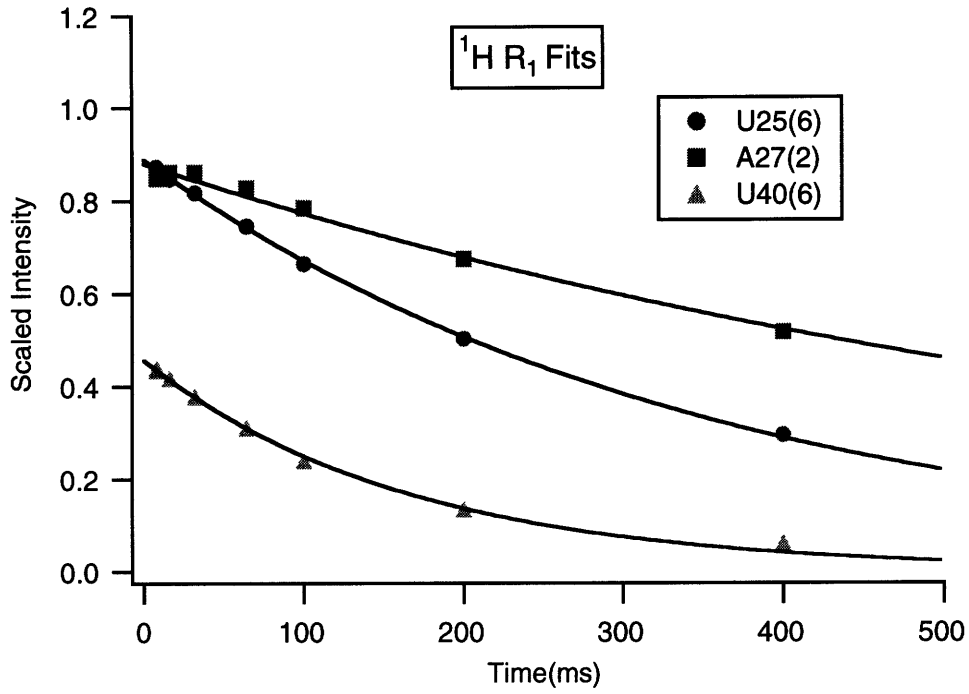
Figure 5.1: Representative data illustrating the quality of the single exponential fits for the ^{1}H R$_1$ and ^{13}C R$_{1\rho}$.

$J(\omega_H)=J(\omega_H\pm\omega_N)$. Spectral densities from both ^{15}N and ^{13}C data are calculated as

previously described (Farrow et al., 1995; Ishima & Nagayama, 1995; Lefevre et al.,

1996) with chemical shift anisotropy (CSA) values of 180 ppm for all the carbons

(Williamson & Boxer, 1989) and 100 ppm for uridine and 130 ppm for guanine iminos

(Akke et al., 1997) as previously reported .

5.3 ^{15}N Base Dynamics

5.3.1 Rate constants

The ability to measure relaxation rate constants of the imino H-N spin systems by

one dimensional spectra allows fast collection of spectra at many different temperatures

with a larger number of points than can be reasonably collected by two-dimensional

experiments. The longitudinal and transverse rate constants were measured by fitting peak

intensities at the different relaxation delays to a monoexponential decay.

Figure 5.2 shows the R_1 and $R_{1\rho}$ relaxation rates and heteronuclear NOE values for

the six imino resonances that are observable at all temperatures. Little variation among the

different Watson-Crick base paired imino resonances is observed and the relative

magnitude of the rates do not change at the different temperatures. Uridine and guanine

imino nitrogens have different CSA values which differentially affect the relaxation. Thus,

the U38 has lower R_1 and $R_{1\rho}$ rate constants than the guanine iminos but is not necessarily

more mobile. The $R_{1\rho}$ rate constant is approximately proportional to the rotational

correlation time. As expected $R_{1\rho}$ increases with decreasing temperatures as the molecule

tumbles more slowly as shown in Figure 5.2. On the other hand, the R_1 rate constant is

inversely proportional to the correlation time for this size molecule and decreases at lower

temperatures. The cross-relaxation rate is sensitive to high frequency motions but the

measurements suffer from poor signal to noise causing larger errors than the other rate
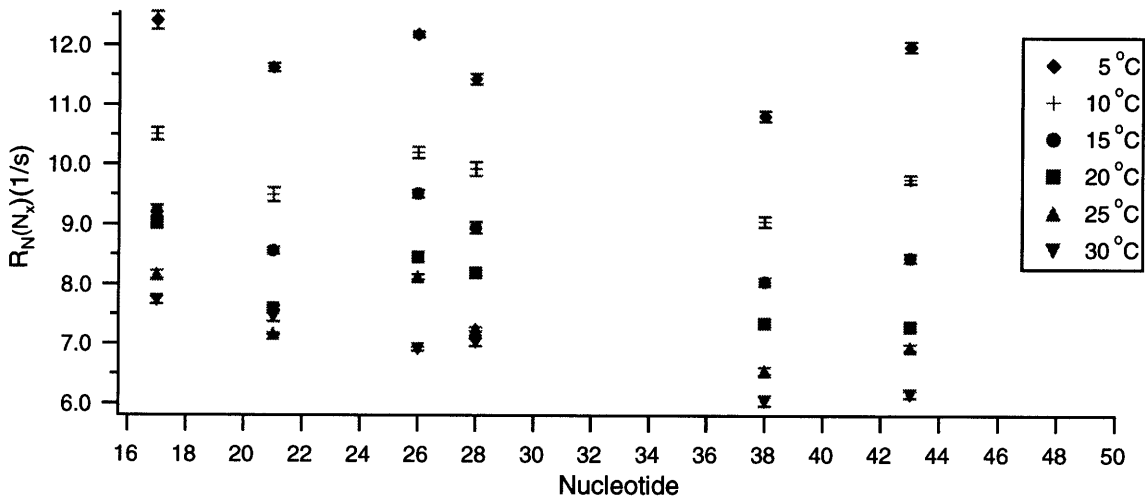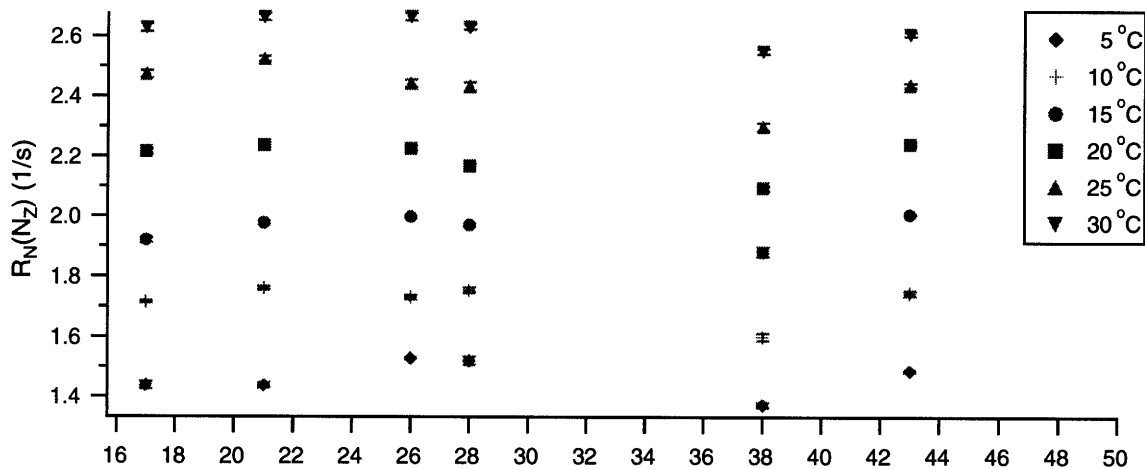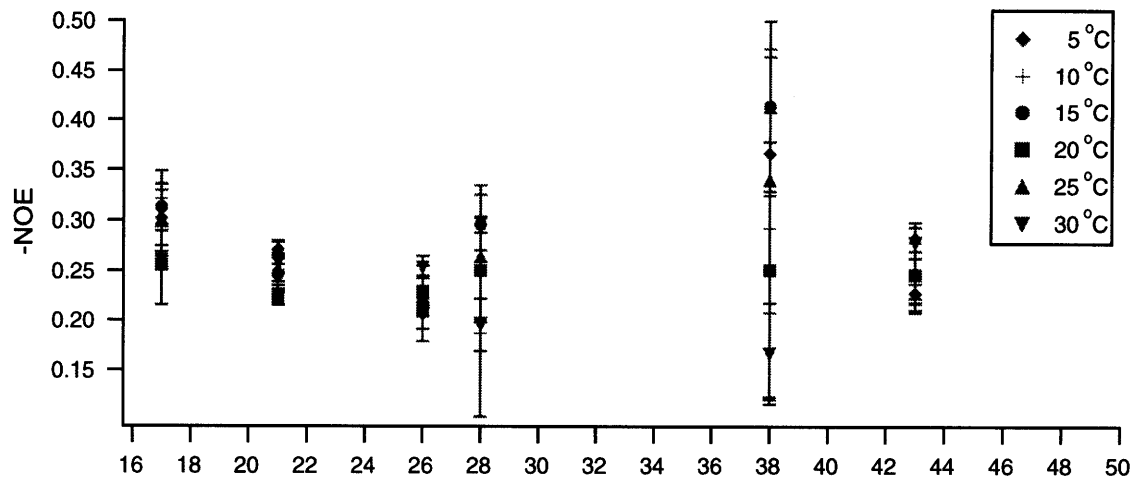
Figure 5.2: ^{15}N relaxation data for the G and U iminos collected by 1D experiments at a ^{15}N frequency of 60.8 MHz.

constant measurements. In particular, the G26 and U38 have very large NOE errors from the 1D data probably because of poor signal to noise. The G26 imino is partially overlapped with the G21 imino while the intensity of U38 decreases at higher temperatures. At lower temperatures, both these problems are alleviated and the errors are smaller.

Relaxation rate constants calculated from 1D and 2D experiments agree well with each other as shown in Figure 5.2. With the 2D data at 5°C, accurate measurements of the U40 and U42 imino resonances are possible, as shown in Figure 5.2. At higher magnetic fields, $R_{1\rho}$ is should increase because of the increasing CSA contribution while the $J(\omega_C)$ dependence of R_1 should cause it to decrease. These trends are observed as shown in Figure 5.3 suggesting that the experiments are measuring the expected effects. For both the R_1 and $R_{1\rho}$ rate constants, the U40 imino is different than the other two uridine imino nitrogens. The U40 imino proton resonances is only clearly observable at temperatures <10°C and is part of the argininamide binding site (see Chapters 2 and 3). The relaxation rates suggest that this residue may be more rigid than the other uridine iminos. As the R_1 is not very different than the other two uridines, a more plausible interpretation is that the U40 may be experiencing conformational exchange processes that increase the measured $R_{1\rho}$ relaxation rates. Additional rotating frame experiments are required to measure directly any conformational exchange effects (Akke & Palmer, 1996; Palmer et al., 1996; Szyperski et al., 1993). The different relaxation of the U40 imino correlates with the U40 being part of the argininamide binding site where argininamide binding and/or stacking of the guanidium group on U40 might be influencing the rate.

5.3.2 Reduced spectral density mapping

Protein relaxation studies have observed that the $J(0)$, $J(\omega_N)$, and $J(\omega_H)$ values and the $R_{1\rho}$, R_1, and (1-NOE) rate constants, respectively, are correlated (Buck et al., 1995;
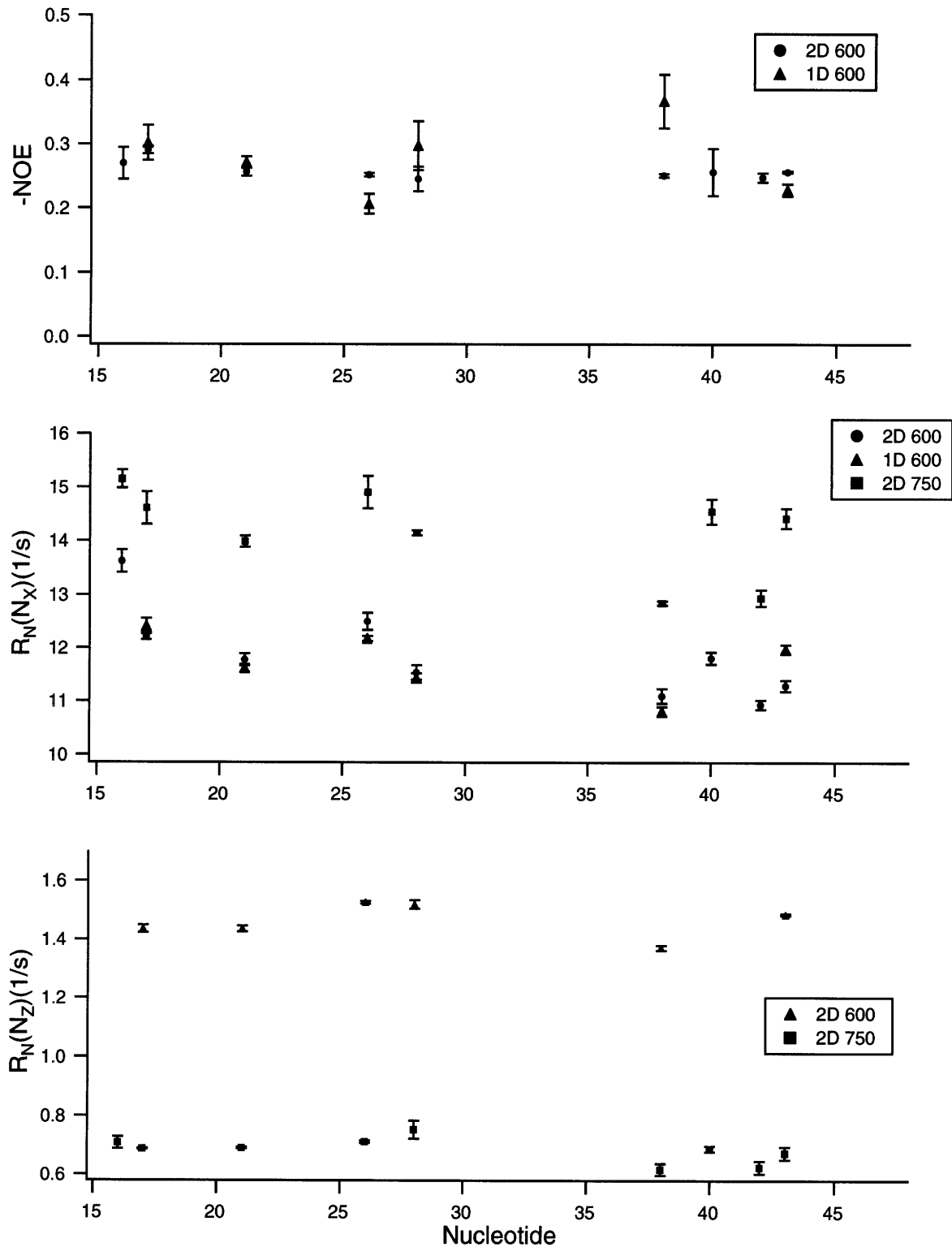
Figure 5.3: Relaxation rates at 5°C shows the reproducability of the 1D and 2D data for the $R_{1\rho}$ and NOE rates. These plots also reveal the slight variablity of the rates with sequence. The 2D data is able to quantitate the U42 which is overlapped in the 1D data. Also, U40 has a higher $R_{1\rho}$ and R_1 than the other two uridines, U42 and U38. The rates change as expected at higher magnetic fields.
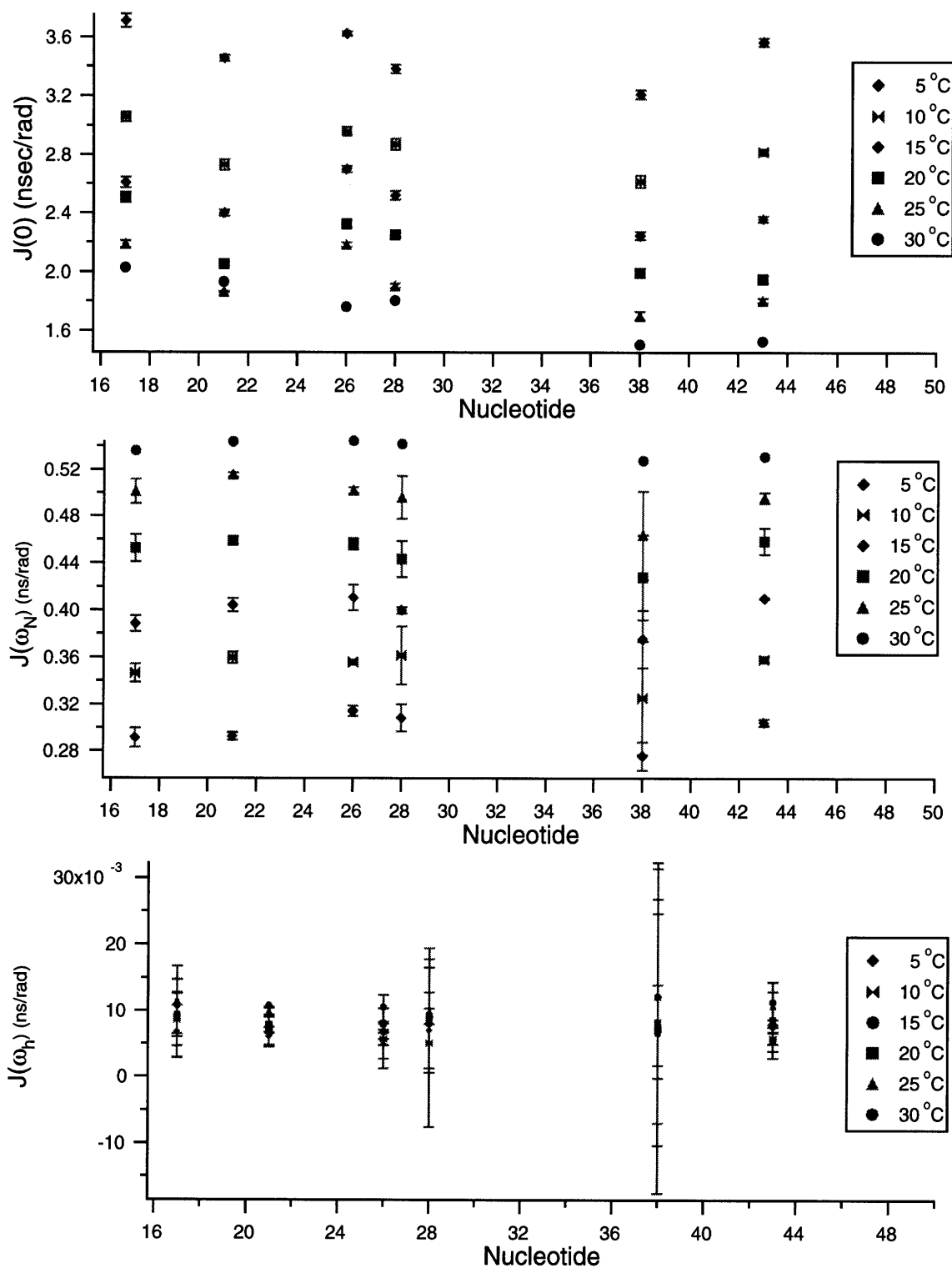
Figure 5.4: Spectral density values versus HIV-2 TAR sequence at different temperatures. Notice how J(0) values are higher while higher frequency J(ω) terms are lower at lower temperatures. The U38 and G28 J(ω_H) values suffer from large errors translated from the NOE measurements because of poor signal to noise.

Ishima & Nagayama, 1995; Palmer et al., 1996). For the motionally restricted protein backbone amides, J(0) makes the most significant contribution to J(ω) and the higher frequency components are almost negligible. When more mobility is observed, the J(ω_H) and J(ω_N) increase while the J(0) decreases. Reflecting the relaxation rates, the imino resonances do not show significant variation of their spectral densities compared to sequence as shown in Figure 5.4. As the temperature is lowered, J(0) increases and J(ω_N) decreases as the rotational correlation time increases by approximately a factor of 1.5. Taking into account the slightly different CSA, uridine J(0) values are slightly smaller than guanine and the same pattern of correlations between the relaxation rate constants and the spectral densities is observed for this RNA.

5.4 ^{13}C Base Dynamics

5.4.1 Relaxation rate constants

Relaxation measurements were made on the H8-C8, H6-C6, and H2-C2 spin pairs at 25°C. For RNA, the advantage for carbon relaxation measurements is that the purine H-C spin systems are not affected by carbon-carbon relaxation mechanisms allowing for a unique opportunity to measure H-C dynamics. The pyrimidine C6 is next to the C5 carbon and carbon-carbon relaxation will interfere with the measurements. Measurements were made with and without a constant time delay that helps the observation of the C6 resonances. However, most of the pyrmidine resonances are severely overlapped, preventing accurate measurement of intensities. Only the U23, U25, and C39 could be consistently measured while the U40 and U38 were accessible in the better signal to noise experiments.

As observed for the ^{15}N dynamics, the A-form helical residues exhibit fairly uniform dynamics behavior as shown in Figures 5.5 and 5.6. The A-form helical H2-C2
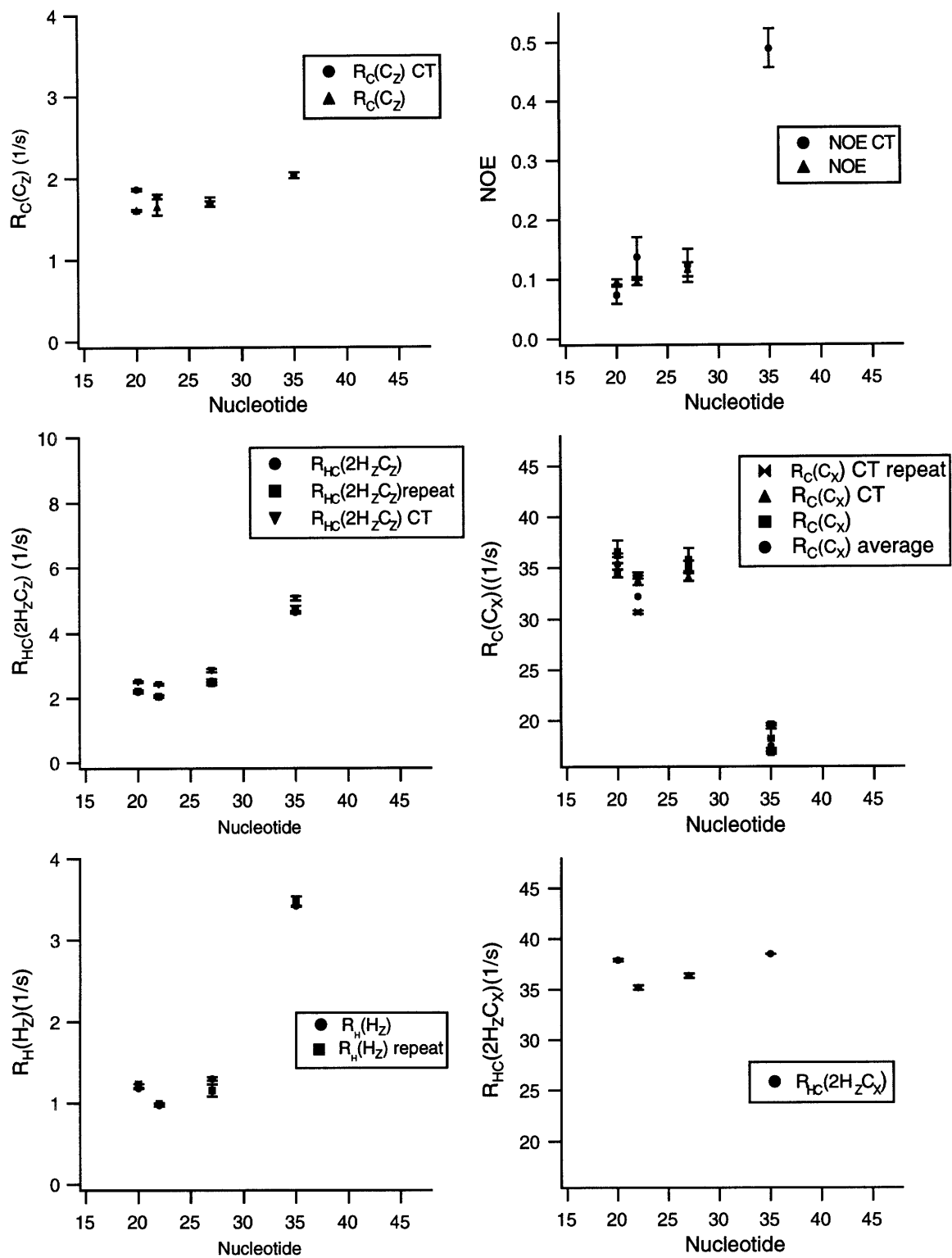
Figure 5.5: ^{13}C relaxation rates versus RNA sequence for the adenine H2-C2 spin systems of HIV-2 TAR argininamide complex. All data were collected at a ^{13}C frequency of 150 MHz. Some data were collected with a constant time delay as indicated.
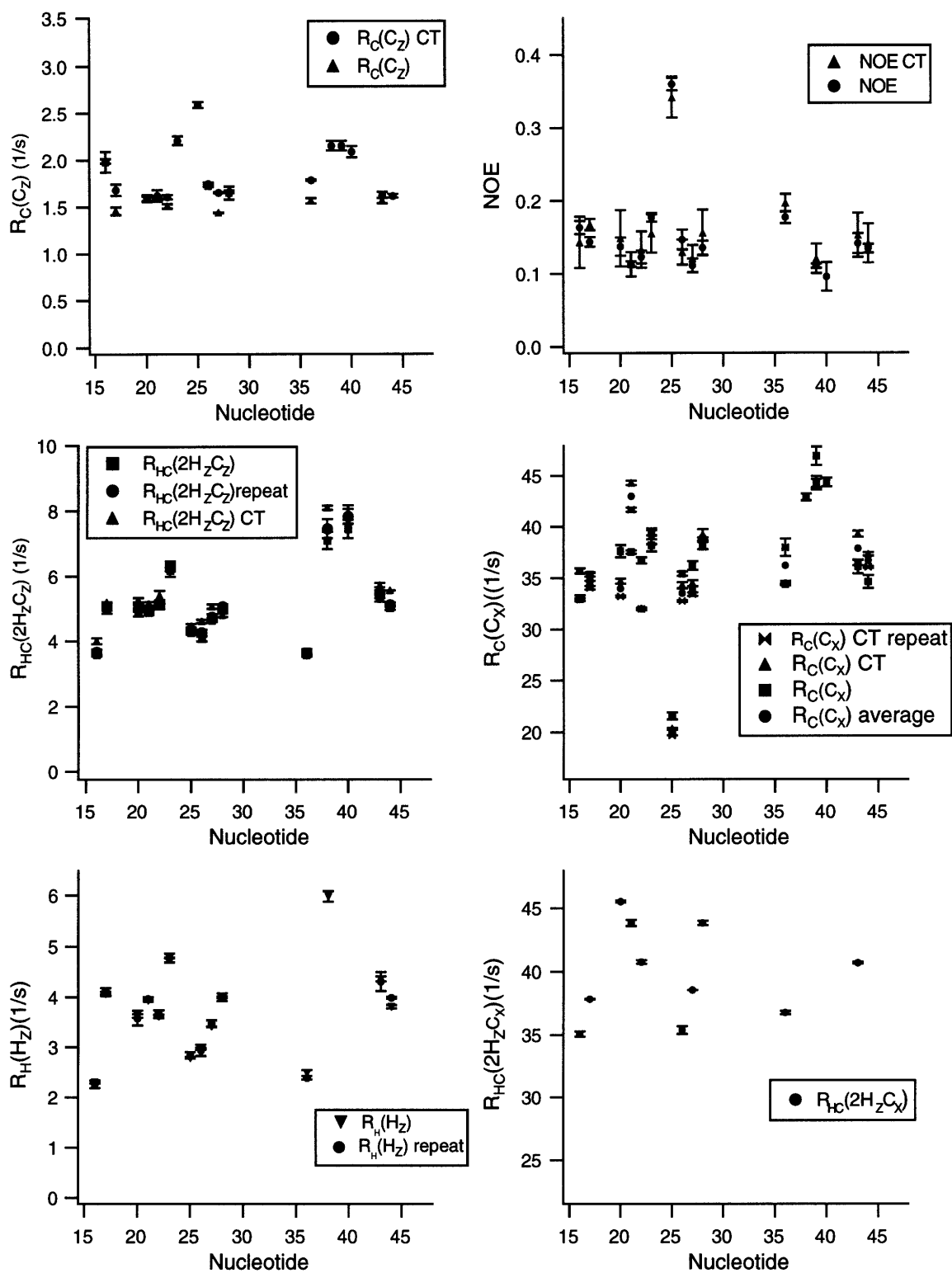
Figure 5.6: ^{13}C relaxation data for the C8 and C6 carbons of HIV-2 TAR-argininamide complex versus RNA sequence. All data were collected at a ^{13}C frequency of 150 MHz.

spin systems are just as rigid as the other A-form aromatic spin systems as seen by the similar ^{13}C R_1 and $R_{1\rho}$ rate constants. However, the isolated nature of the H2-C2 spin systems is illustrated by the shorter ^{1}H R_1 rate constants. The non-helical residues, U25(6) and A35(2), exhibit different ^{1}H R_1 and ^{13}C $R_{1\rho}$ rate constants than the other A-form pyrmidine(6)'s and adenine(2)'s, respectively. The large R_1 and NOE suggest that these resonances have large high frequency motions contributing to their relaxation. As observed for its imino, the terminal G16(8) has very similar relaxation rate constants as the other A-form residues, suggesting that even though H_2O exchange measurements suggest that terminal base pairs are fraying, the bases, themselves, are not very mobile.

5.4.2 Full spectral density mapping

The common problem with full mapping is the proton dipole-dipole term affecting the $R_{HN}(2H_zN_z)$, $R_{HN}(2H_zN_x)$, and $R_H(H_z)$ which cannot be easily measured and distorts these measurements. This term is proportional to molecular weight and also to the nearby proton density with which the proton of interest can cross-relax. The advantage of full spectral density mapping is that the spectral densities are calculated from six independent relaxation measurements in the absence of any motional models or other major approximations. The H2-C2 spin system presents an excellent candidate for the full mapping approach since it is a very isolated spin system with only a few NOEs to each H2 proton resonance in a RNA A-form helix when the sample is in ^2H_2O. To test whether full spectral density mapping is applicable to adenine H2-C2 spin system in medium sized RNAs, all six relaxation rate constants were measured as shown in Figure 5.6. When the spectral densities were calculated, some $J(\omega_H)$'s had negative values which are physically impossible, as summarized in Table 5.1. Perhaps by eliminating all the NOEs to the adenine H2s eliminating almost every proton by fully deuterating the ribose, the full mapping strategy may be possible.

Table 5.1: Spectral densities from ^{13}C data: Full vs. Reduced Mapping

Position	J(0) full	J(0) full rmsd	J(0) reduced	J(0) reduced rmsd	$J(\omega_c)$ full	$J(\omega_c)$ full rmsd	$J(\omega_c)$ reduced	$J(\omega_c)$ reduced rmsd	$J(\omega_H)$ full	$J(\omega_H)$ full rmsd	$J(\omega_H)$ reduced	$J(\omega_H)$ reduced rmsd
Adenine(2)												
A20(2)	2.11	0.02	2.0213	0.0322	0.0531	0.001	0.0712	0.0010	0.013	0.016	0.0005	0.0005
A22(2)	1.97	0.01	1.8578	0.0176	0.0554	0.003	0.0673	0.0021	-0.016	0.015	0.0009	0.0012
A27(2)	2.00	0.02	1.9821	0.0405	0.0550	0.001	0.0641	0.0019	-0.016	0.017	0.0008	0.0010
A35(2)			0.9920	0.0233			0.0732	0.0022			0.0037	0.0012
Purine(8)												
G16(8)	1.94	0.03	1.9103	0.0132	0.0625	0.002	0.0754	0.0020	0.113	0.027	0.0011	0.0013
G17(8)	1.95	0.02	1.9745	0.0152	0.0499	0.003	0.0636	0.0023	0.041	0.026	0.0011	0.0003
A20(8)	2.25	0.03	1.9964	0.0174	0.0581	0.001	0.0611	0.0016	-0.156	0.029	0.0009	0.0014
G21(8)	2.19	0.03	2.3436	0.0118	0.0534	0.001	0.0618	0.0025	-0.109	0.031	0.0007	0.0006
A22(8)	2.07	0.03	1.9030	0.0102	0.0550	0.002	0.0611	0.0016	-0.011	0.029	0.0008	0.0009
G26(8)	1.93	0.02	1.9319	0.0114	0.0527	0.001	0.0665	0.0011	0.125	0.025	0.0009	0.0006
A27(8)	2.02	0.02	1.9677	0.0177	0.0580	0.001	0.0630	0.0010	0.069	0.016	0.0008	0.0007
G28(8)	2.23	0.03	2.1993	0.0224	0.0520	0.003	0.0630	0.0017	-0.084	0.034	0.0010	0.0011
G36(8)	1.98	0.04	2.0143	0.0215	0.0541	0.001	0.0672	0.0007	0.011	0.038	0.0013	0.0004
G43(8)	2.01	0.03	2.1183	0.0233	0.0568	0.001	0.0617	0.0020	-0.063	0.032	0.0009	0.0011
G44(8)	2.04	0.02	2.0436	0.0170	0.0582	0.002	0.0614	0.0016	-0.095	0.019	0.0009	0.0010
Pyrimidine(6)												
U23(6)			2.1940	0.0310			0.0835	0.0022			0.0013	0.0010
U25(6)			1.1148	0.0085			0.0956	0.0019			0.0034	0.0010
U38(6)			2.5272	0.0193			0.0841	0.0026			0.0006	0.0012
C39(6)			2.5585	0.0377			0.0819	0.0023			0.0010	0.0007
U40(6)			2.5720	0.0242			0.0824	0.0023			0.0010	0.0005

5.4.3 Reduced spectral density mapping

Because full spectral mapping did not accurately determine the $J(\omega_H)$ values, the reduced spectral density approach was applied to calculate the spectral densities. As discussed above, these calculations use only the longitudinal, transverse, and NOE relaxation rates to calculate the spectral densities by assuming that $J(\omega_H \pm \omega_c) = J(\omega_H)$. This assumption may be suspect because the carbon Larmor frequency is about 25% of the proton Larmor frequency, which is not as insignificant a difference as for nitrogen. The high frequency components still suffer from large errors translated from the NOE experiments as seen in Table 5.1. The two spectral density mapping methods give very similar $J(0)$ values, but the reduced mapping $J(\omega_C)$ values are systematically higher than those from the full mapping calculations. This over-estimation originates from the $J(\omega_H \pm \omega_c) = J(\omega_H)$ assumption.

The spectral densities indicate that the A-form residues exhibit uniform dynamics as they all have very similar $J(0)$ values. The spectral densities indicate that the A35 and U25 residues are very flexible. This increased mobility for the U25 and A35 correlates with the structure and previous NMR studies, as these residues have very few internucleotide NOEs and are solvent exposed. Despite not being imbedded in helix, the U23(6) is fairly rigid, as it has a slightly lower $J(0)$ than the other pyrmidines, but higher than the U25(6) as seen in Table 5.1.

5.5 Summary

These results present some of the first detailed ^{15}N and ^{13}C relaxation measurements made on isotopically labeled RNA. All the A-form residues seem to behave similarly, while the non-helical regions exhibit a variety of dynamic behavior. The U25 and A35 are very flexible while the U23 is not as rigid as other A-form pyrmidines but also is not as

flexible as the unstructured residues. The HIV-2 TAR-argininamide structure suggests that the U23 may behave like a terminal base because it is only stacking on one face similar to a terminal base pair. The U23 and terminal G16 both have fairly rigid J(0) values that are very similar to the A-form residues and do not show any signs of conformational exchange. Thus, the U23 is truly behaving like a terminal base as the imino proton is difficult to observe, while the aromatic H-C spins system is not very flexible. Although the G16 imino proton resonance is observable as a very broad peak at 5°C, the U23 is not.

Perhaps the slightly more flexible dynamics of the U23 compared to the terminal G16 are sufficient to prevent the observation of its imino proton. This dynamics behavior may be reflecting the partial base stacking and protection from solvent that these residues may be experiencing.

This work explores how the various residues are behaving in the saturated argininamide complex of argininamide. To understand the dynamics of the argininamide binding site and the changes that occur upon binding, measurements of both the free HIV-2 TAR can be collected and at various argininamide concentrations. Because the dissociation constant of argininamide is ~1-3 mM, the dynamics of TAR can be monitored as it undergoes the conformational changes required for specific binding. In addition, TAR's behavior in peptide complexes can also be monitored. It may also be interesting to explore the differences between the three and two base bulges in the argininamide complexes. The power of NMR to study weakly binding complexes can be fully applied to the arginine binding site of TAR permitting the dissection of the role of dynamics in specific binding.

Chapter 6: Tat Peptide - HIV-2 TAR NMR Studies

6.1 Choosing a Tat Peptide

Chapters 2-4 focused on the HIV-2 TAR-argininamide complex and showed how TAR specifically recognizes a single arginine. Even though most of the specificity for the Tat-TAR interaction resides in a single arginine, some minor specificity elements may exist in other residues of the RNA binding domain. Also, the conformation of the Tat peptide remains unknown even though some unconvincing evidence has emerged suggesting that the Tat peptide may be α-helical. The structure of the Tat peptide-TAR complex would be valuable towards understanding how the arginine is tethered to the rest of the peptide and how other residues may be specifically recognizing TAR. As discussed in Chapter 1, previous NMR studies demonstrated that the structure of TAR has a similar conformation when in complex with either argininamide or Tat peptides (Aboul-ela et al., 1995; Puglisi et al., 1992). The improved spectral properties of the two base bulge HIV-2 TAR helped determine a high resolution structure and identify the base triple. This chapter discusses the Tat peptide complexes in the context of the two base bulge HIV-2 TAR. Although HIV-2 TAR may help improve the RNA spectra, the choice of which peptide to study remains unclear. A number of different peptides have been tested but the spectral quality of these complexes with HIV-1 TAR have been very poor and not amenable to high resolution structural studies (Aboul-ela et al., 1995). Peptides containing a single arginine with 10 lysines, the R52 peptide with the HIV-1 TAR were tested. By having only one arginine residue instead of the six found in the Tat basic domain, the key arginine residue contacting the RNA would be assignable in homonuclear spectra. However, the peptide proton resonances were broad and essentially uninterpretable. These studies were performed before the technology to produce isotopically labeled peptides became available.

Recently, *in vivo* peptide expression as a fusion protein making isotopic labeling

```
ADP-6  FHCQVCFTTKALGISYGRKKRRQRRRPPQGSQTHQVSLSKQ
ADP-1          CFTTKALGISYGRKKRRQRRRPPQGSQTHQVSLSKQ

tfr14                  RKKRRQRRRPPQGS
tfr24                  RKKRRQRRRPPQGSQTHQVSLSKQ
tfr38                  RKKRRQRRRPPQGSQTHQVSLSKQPTSQSRGDPTGPKE


t2fr27 from HIV 2      RKGRRRRSPKKTKAHSSPASDTRTGNS
```

Figure 6.1: A number of different peptides have been tested for specific TAR binding. However, it is not clear what other Tat residues, if any, in addition to the critical arginine is required for specific binding. The arginine-rich RNA binding domain is highlighted by the shaded rectangle.

possible became available, thus creating an opportunity to revisit Tat peptide-TAR complexes with some hope of making full assignments. Because of a continuing controversy on whether there are minor elements of specificity and which residues may be involved, the choice of which peptide to study remains non-trivial. Figure 6.1 highlights the major candidates which have been extensively studied (Churcher et al., 1993; Hamy et al., 1993; Long & Crothers, 1995; Weeks & Crothers, 1991; Weeks & Crothers, 1992). Besides the key arginine, the other residues only induce very small specificity effects of factors of two to four, making these measurements difficult. Because another research group has been studying different ADP peptides with HIV-1 TAR in the NMR (Aboul-ela et al., 1995), we chose to start with the TFR series of peptides which have been extensively investigated by Crothers and co-workers (Long & Crothers, 1995; Weeks et al., 1990; Weeks & Crothers, 1991; Weeks & Crothers, 1992). Also, the ADP peptides contain a stretch of hydrophobic residues which may explain the slightly steeper binding curves and may cause aggregation problems at the high concentrations required for NMR studies. The tfr24 peptide is a good first candidate because of its kinetic stability compared to shorter peptides and the additional residues of the tfr38 fragment are probably not critical for specific RNA binding as the last 14 residues are unimportant for *in vivo* activity.

6.2 Materials and Methods

6.2.1 Peptide expression and preparation

To obtain isotopically labeled Tat peptides, peptide sequences were expressed as a C-terminal fusion with a modified, His-tagged TrpE leader polypeptide (Figure 6.2) (Schumacher *et al.*, 1996; Staley & Kim, 1994). This construct allows a straightforward purification because the trpE readily forms inclusion bodies. The peptide is cleaved from the fusion protein by cyanogen bromide at the methionine immediately N-terminal to the Tat peptide.
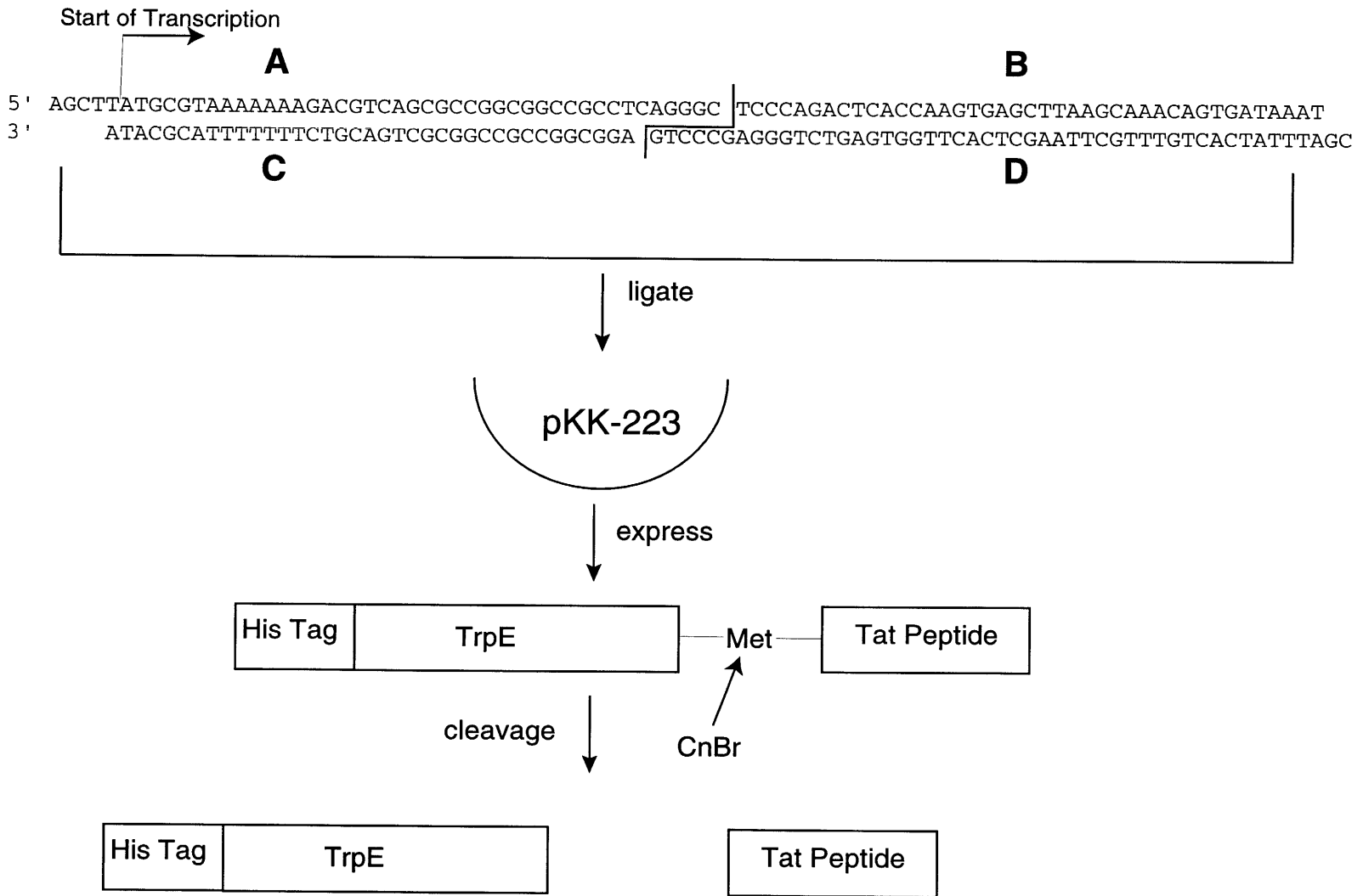
Figure 6.2: Oligo A is annealed with C and B with D before being ligated together and to the plasmid. The Tat peptide is expressed as a fusion with TrpE and separated from the fusion by cyanogen bromide cleavage.

Plasmid construction The genes to express the peptides were constructed by direct three piece ligation into the pKK223-3-λN expression vector (Pharmacia Biotech), which was altered by Christopher Cilley for expression of λN peptides (personal communication). This vector contains the TrpE protein under the control of a strong tac promoter and an ampicillin resistance gene marker. Four oligos were synthesized, kinased, and annealed before being ligated between ClaI and HindIII restriction sites as shown schematically in Figure 6.2. The plasmids were transformed into *E. Coli* strain JM109 by the Hannehan method. Attempts to use frozen competent JM109 cells for the transformation failed probably because JM109 does not transform with high yields in combination with the low transformation efficiencies obtained from three piece ligations. SacI was added to linearize any uncut plasmid and reduce the number of background colonies, since there was an unique SacI in the original pKK223-3-λN plasmid and none in the desired product. Transformants were selected by growing on ampicillin LB plates, tested with restriction digests, and sequenced.

Peptide purification For isotopic labeling, cells were grown on minimal media consisting of: 21.5 g/L glycerol (17.2 ml), 0.7 g/L ^{15}N-$(NH_4)_2SO_4$, 10.01 g/L $Na_2HPO_4 \cdot 7\ H_2O$, 1.6 g/L KH_2PO_4, 0.5 g/L sodium citrate, 0.3 g/L $MgSO_4$, 5 mg/L Vitamin B1 (thiamine), 50 mg/L ampicillin, and 1.5 ml of a trace elements solution consisting of 20 g/L Na_2EDTA, 0.5 g/L $CaCl_2 \cdot H_2O$, 16.7 g/L $FeCl_2 \cdot 6\ H_2O$, 0.18 g/L $ZnSO_4 \cdot 7H_2O$, 0.16 g/L $CuSO_4 \cdot 5\ H_2O$, and 0.18 g/L $CoCl_2 \cdot 6\ H_2O$. Expression was induced with isopropyl-β-D-thiogalactopyranoside (IPTG) when the cells reached an optical density between 0.6 and 0.8 at 600 nanometers. Typically the optical density increased by a factor of 2-2.5 during the 4-8 hours of induction and 2-3 grams of wet cells per liter were obtained.

An inclusion body preparation was performed by initiating lysis with lysozyme and sonicating the cells for three 20 minutes periods. Per liter, the cells were resuspended in 30

ml of 50 mM Tris-Acetate, pH 8.0, 10 mM $MgCl_2$, 100 mM NaCl, and 0.1 mM EDTA

(SB1) with 5 mg lysozyme and ~50 μg of DNAse I. After gently agitating the solution at

room temperature for 15 minutes, the solution was cooled on ice before being sonicated for

five 30 second periods with two minutes between the pulses. The pellet was collected by

centrifugation at 6000 g for 30 minutes and resuspended in 45 ml of 20 mM Tris-Acetate,

pH 8.0, 0.5 M NaCl, and 1 mM EDTA (SB2) with 2.25 ml of 20% Triton X-100 before a

second round of sonication. After centrifugation, the pellet was dissolved in 45 ml of H_2O

and the sonication was repeated for a third and final time. This procedure solubilizes most

of the cellular proteins, except the inclusion bodies containing the fusion protein, resulting

in a simple, large-scale purification. The remaining pellet was dissolved in 30-40 ml of

100 mM HCl or 10 ml of 70% formic acid and incubating with 100 mg cyanogen bromide

was added for one hour at room temperature. Cyanogen bromide cleaves the methionine

immediately N-terminal to the Tat peptide, releasing the peptide from the fusion protein

(Schumacher et al., 1996; Staley & Kim, 1994). A second 100 mg aliquot of cyanogen

bromide was added for an additional hour. After lyophilization to remove the cyanide, the

pellets were dissolved in water and neutralized with sodium hydroxide which caused the

trpE to precipitate. The supernatant, which contained the cleaved peptide, was further

purified by two rounds of reverse phase HPLC with 0.1% TFA in H_2O/acetonitrile and

concentrated by lyophilization before dialysis for >48 hours into water using a 1000

molecular weight nitrocellulose membrane (Spectrum). To check the peptide composition

and estimate the concentration, amino acid analysis was performed. For concentration

measurements, two amino acid analyses were averaged because no aromatic residues are in

these sequences for UV spectrometric determination of the concentration and the peptide

backbone is not very accurate because many small impurities have large extinction

coefficients in the low UV region. From two liters of minimal media growth,

approximately 830 nanomoles were obtained, which was split into two ^{15}N tfr24 samples:

one 1.2 mM peptide-TAR complex and one 0.45 mM tfr24 free peptide sample. Synthetic

tfr24 peptide was synthesized and purified by Andrew Bennett and Rob Tycko of the NIH.
Samples were dialyzed and concentrations determined by amino acid analysis. To assay a
number of different sample conditions, the peptide-TAR complex sample was divided into
thirds. Temperatures from 5-30°C, pHs from 5-7, and NaCl concentrations from 5-100
mM were tested.

6.2.2 NMR spectroscopy

Spectra were collected on either a 600 Varian INOVA spectrometer or 750 MHz
Varian Unity Plus spectrometer equipped with a Varian triple resonance gradient probe. All
spectra were processed and analyzed using NMRPipe (Delaglio et al., 1995) on a Silicon
Graphics Indy workstation.

Proton NMR spectra in H_2O were recorded at a variety of temperatures and mixing
times, including 50 and 100 ms. NOESY spectra used a WATERGATE 3-9-19 water
suppression scheme with a selective E-BURP-1 flip-back pulse (Green & Freeman, 1991;
Lippens et al., 1995; Piotto et al., 1992) . Sweep widths were 14000 or 16000 Hz at 600
and 750 MHz, respectively, and 4096 x 512 real points collected. Data were zero filled to
4K x 2K real points and apodized using Gaussian-Lorentzian functions in both dimensions
before Fourier transformations.

^{15}N FHSQC spectra were collected at a variety of temperatures and peptide:RNA
concentrations using 3-9-19 WATERGATE for water suppression with the transmitter set
to 98 ppm (Mori et al., 1995; Piotto et al., 1992). Sweep widths were 7000 x 6000 Hz
and 512 x 128 real points collected. The ^{15}N NOESY-HSQC at 25°C used a 3-9-19
WATERGATE sequence for water suppression at 600 MHz with the transmitter set to 90
ppm with a 100ms mixing time (Mori et al., 1995). Sweep widths of 6000 x 13,500 x
4500 were used and 512 x 128 x 32 real points for the ω_3, ω_2, and ω_1 dimensions,

respectively. GARP was used to decouple ^{15}N during acquisition (Shaka et al., 1983), and a composite 180° ^{15}N pulse decoupled the protons during the *t1* proton evolution time.

6.3 Formation of the tfr24 Peptide-TAR Complex

The titration of the tfr24 peptide with TAR was carefully monitored as the peptide concentrations determined from amino acid analysis are susceptible to errors by as much as a factor of two. We expected to add peptide until the TAR imino spectrum looked like the argininamide complex which should indicate that most of the TAR was bound. Before reaching this endpoint, the RNA resonances almost completely broadened significantly before returning to observable levels as shown in Figure 6.3. The imino lines are significantly broader than in the TAR-argininamide complex as illustrated in Figure 6.4. Some broadening is expected because of the larger molecular weight of the peptide complex. The peptide complex imino spectra look strikingly similar to the argininamide complex suggesting that the RNA secondary structure is the same in the two complexes. Further tests verified that the complex is approximately equimolar peptide and RNA. The titration was performed twice: once with ^{15}N-labeled tfr24 and once with synthetic tfr24 in a 2 mM sample. The synthetic tfr24 complex was prepared for solid state NMR studies where the two labeled ^{13}C carbonyl carbons could be integrated and compared to all the natural abundance methyl and methylene peaks. This calculation demonstrated that the complexes were at a 1:1 peptide:RNA stoicheiometry (data not shown). Thus, more confidence that the observed NMR spectra are 1:1 complexes was obtained.

6.3.1 ^{15}N-NMR spectroscopy of tfr24

To determine whether the peptide spectra were of sufficient quality for high resolution structural studies, the peptide backbone and arginine sidechains were monitored by ^{15}N-FHSQC spectra. In a structured peptide or protein, the amide ^{15}N/^{1}H and the arginine Nε/Hε crosspeaks are sharp and well-resolved as observed for the Rev (Battiste et
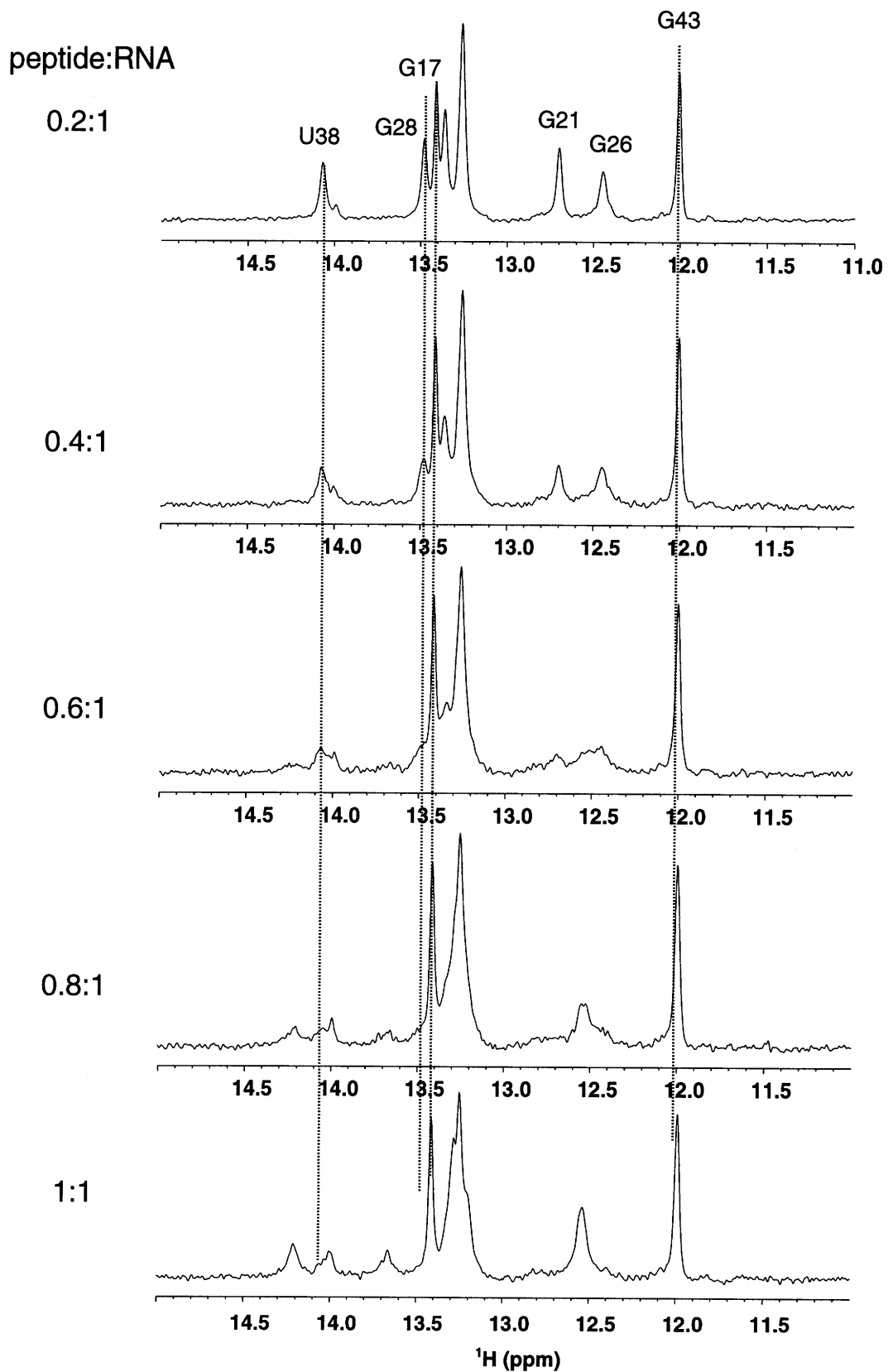
Figure 6.3: The imino spectrum of HIV-2 TAR broadens upon addition of ^{15}N-tfr24 peptide. The resonances that broaden are in the bulge region and are likely part of the peptide binding site.

free

6.5 mM Arg

1:1 tfr24

14.5   14.0   13.5   13.0   12.5   12.0

^{1}H (ppm)
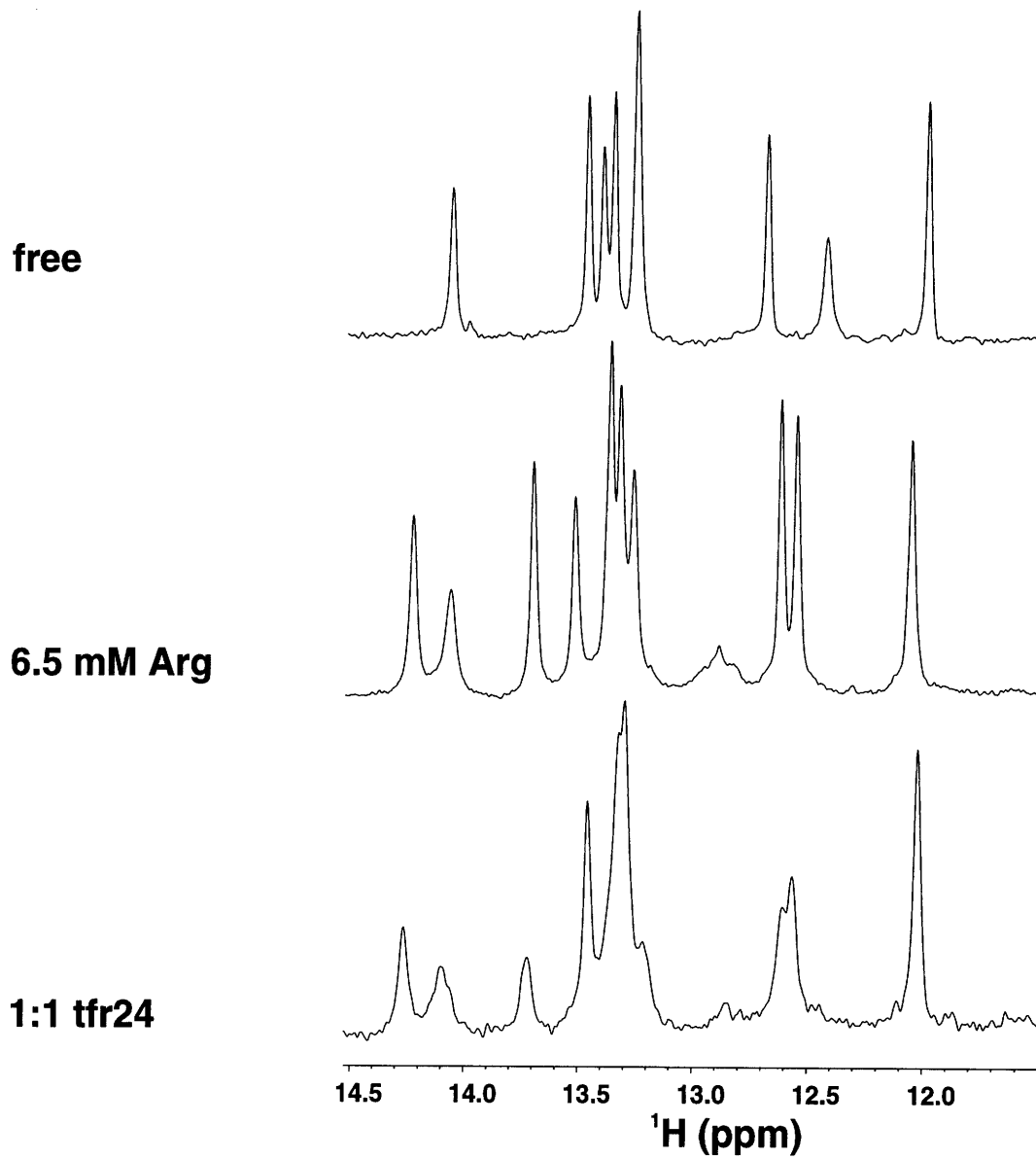
Figure 6.4: The imino region of HIV-2 TAR shows that the secondary structure of the RNA in the argininamide and the tfr24 Tat peptide complexes are similar.

al., 1996) and λ peptide RNA complexes (Cilley and Williamson, unpublished results).

Figure 6.5 compares the ^{15}N FHSQC of the tfr24 peptide before and after the addition of

HIV-2 TAR. The amide and Hε resonances are very poorly resolved and the proton peaks

are broadened in the RNA complex. All five of the critical arginine Hε/Nε crosspeaks are

severely overlapped and only a single backbone amide significantly shifts upon addition of

TAR. The 3D ^{15}N NOESY-HSQC experiment suggest that this amide is an arginine as

shown in Figure 6.6. This result is consistent with all the previous NMR and biochemical

results which suggest that only a single arginine is required for specific TAR binding. The

other residues of the peptide increase the affinity of the complex presumably by making

non-specific phosphate backbone contacts and increasing the contact surface area. The

spectra are certainly not of high enough quality to pursue high resolution structural studies.

It would be very difficult to assign the backbone where only 11 of the 24 amide resonances

can be clearly observed. Furthermore, the critical arginine intermolecular contacts would

be impossible to resolve due to the lack of dispersion of the arginine Hε resonances, which

proved extremely valuable in solving the Rev-RRE structure. Because of the lack of

dispersion, it is unlikely that the peptide is forming an α-helix. Furthermore, base on the

chemical shift index, the peptide is likely in a random coil or extended strand conformation

and not a α-helix as has been previously proposed (Mujeeb *et al.*, 1994). These results

are consistent with careful difference CD measurements showing that the argininamide

complex CD spectrum is very similar to the wild type peptide TAR complex suggesting that

the peptide does not significantly change conformation upon binding (Long & Crothers,

1995). To try to improve the quality of the spectra, a wide range of pH, salt and sample

concentrations were tested but no significant improvements in the chemical shift dispersion

or linewidths were observed. More peptide was also added to one of the diluted samples
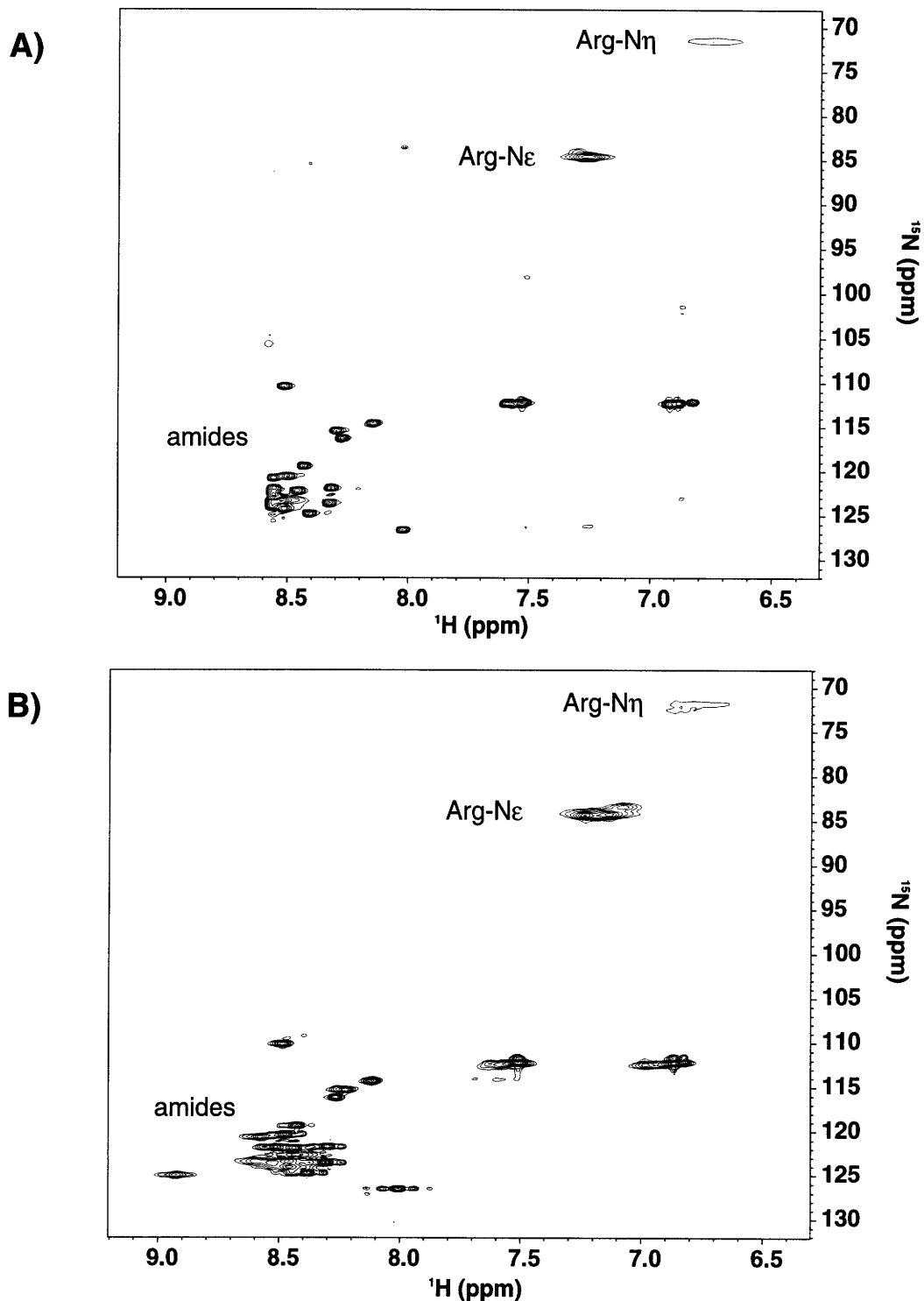
Figure 6.5: ^{15}N-FHSQC spectra of the Tat tfr24 peptide **A)** in the absence of TAR and **B)** in complex with HIV-2 TAR at 25°C at 600 MHz. Some chemical shift changes occur but the lines broaden and only about half of the amides can be observed. Also, the Hε/Nε crosspeaks are not resolved, preventing assignment of these critical resonances.
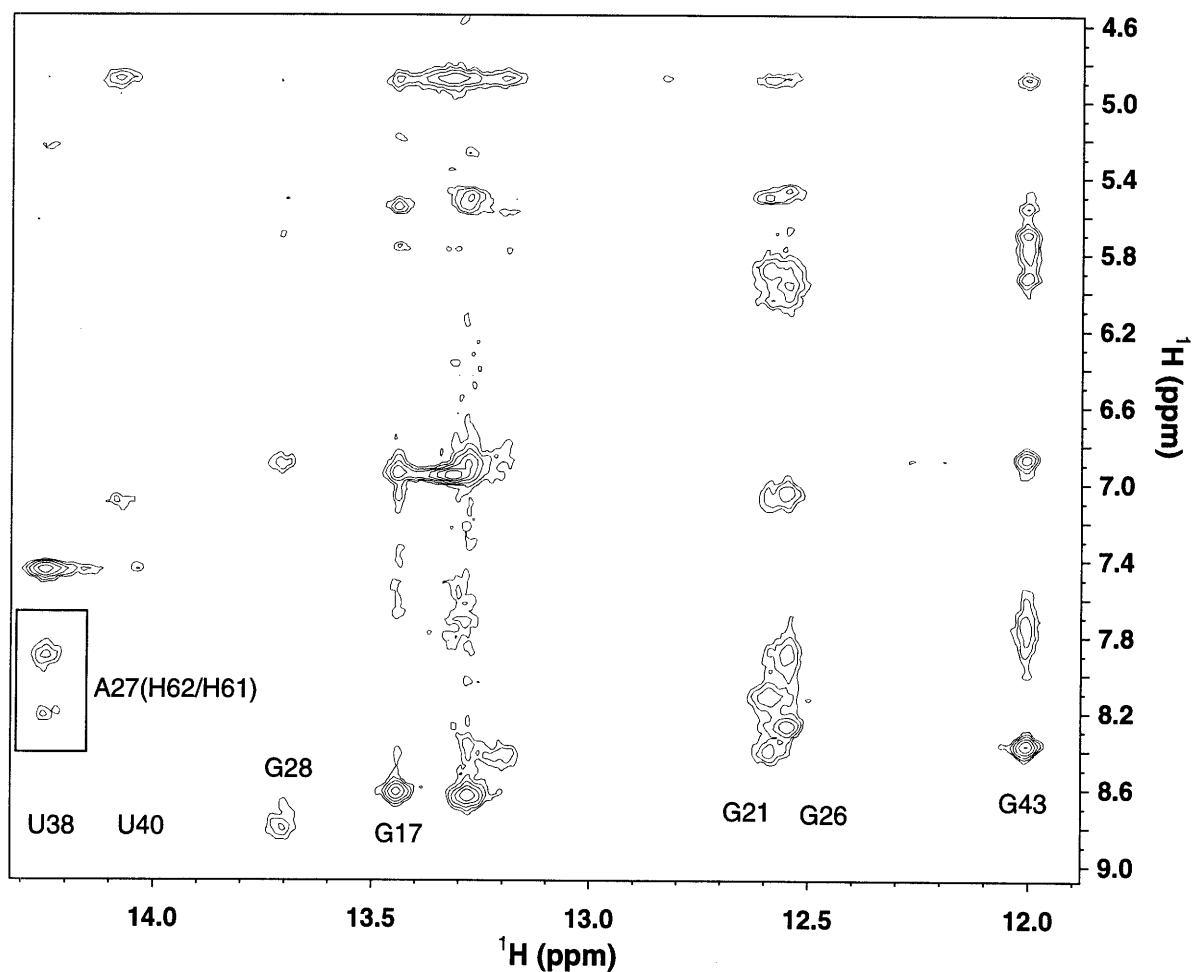
Figure 6.6: Watergate NOESY (mix=100 ms) at 15°C at 600 MHz of the tfr24 peptide-HIV-2 TAR RNA complex. The lines are very broad. However, the A27 amino proton resonances, highlighted by the rectangle, are split indicating that the base triple structure is more stable. These resonances are exchanging more slowly than in the argininamide complex where only a single proton peak was observed at the average of these two resonances.

beyond the 1:1 equimolar ratio, but no significant changes in the ^{15}N-HSQC or imino spectra were observed.

6.3.2 Conformation of TAR in the tfr24 complex

The RNA spectrum is very similar to the TAR - argininamide complex spectra, suggesting that the RNA conformation is very similar in the two complexes. The 25°C imino proton chemical shifts of the imino spectrum is almost identical to the 5°C spectra of the TAR - argininamide spectra. Many of the cytidine aminos and H5 chemical shifts also are very similar to the TAR-argininamide spectra. One interesting difference is that both A27 amino protons are observed as distinct peaks, while those in the argininamide complex are degenerate (Figure 6.7). This observation is consistent with a base triple interaction where one of the A27 amino protons is hydrogen bonding to the U23. Disappointingly, the already broad spectra continued to broaden at lower temperatures where the base triple imino is generally more easily observable as seen in the BIV Tat-TAR and HIV Rev-aptamer peptide complexes (Ye et al., 1996; Ye et al., 1995). This broadening probably contributed to obscuring the U23 imino resonance which is not clearly observable in the homonuclear spectra in any of the conditions surveyed.

6.4 Future Prospects and Research

6.4.1 The TAR arginine binding motif

The arginine binding motif originally identified in the HIV Tat-TAR interaction has recently been observed in other contexts. A TAR bulge motif was identified in the selection of high affinity Rev-binding aptamers, which gave the same chemical modification interference signature as TAR (Giver *et al.*, 1993; Jensen *et al.*, 1994). The ensuing solution structure of the Rev peptide-aptamer complex provided unambiguous evidence for a U-A•U base triple forming an arginine binding site (Ye et al., 1996), very similar to the HIV-2 TAR structure. The rmsd between the Rev-aptamer and the HIV-2 TAR for the
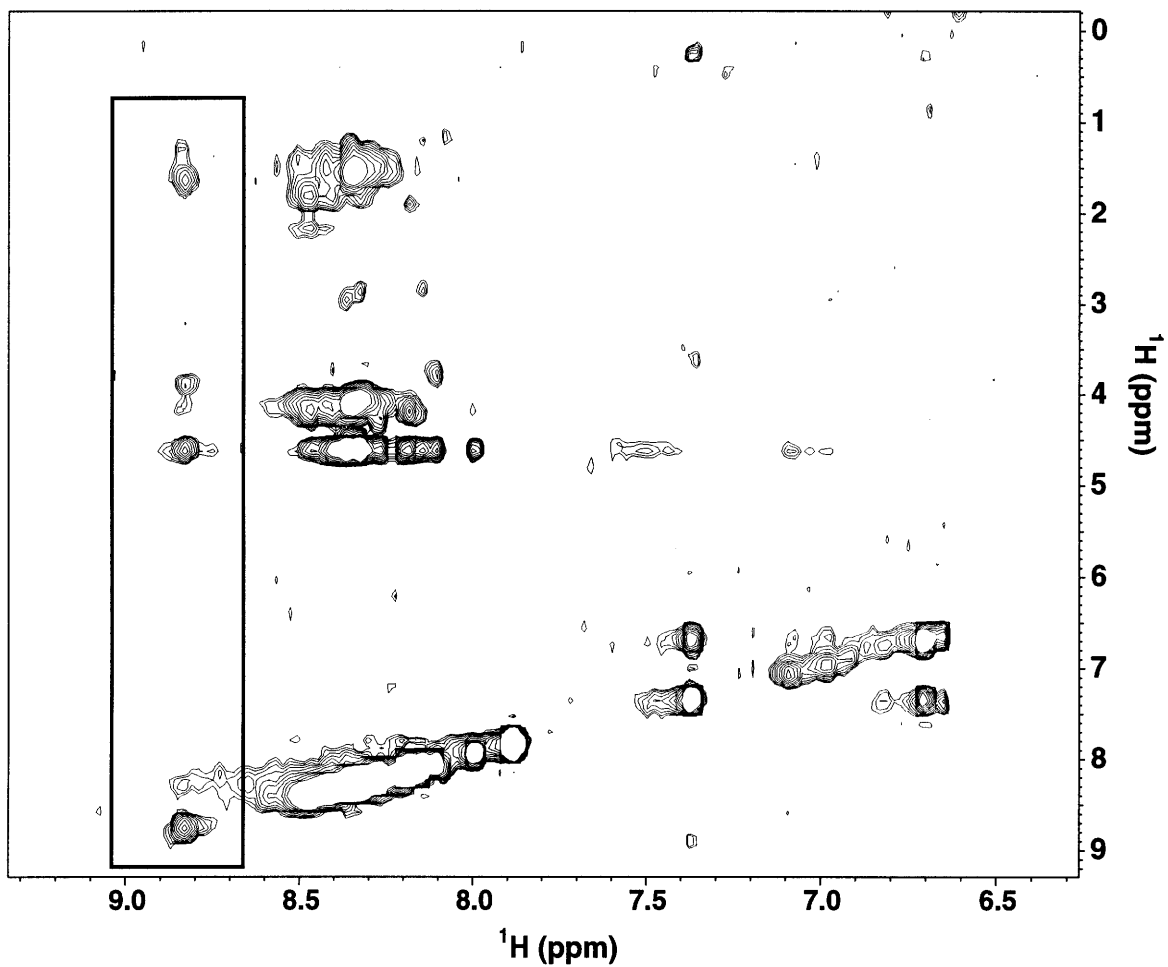
Figure 6.7: Slice from a 3D 100 ms mixing time ^{15}N NOESY-HSQC of tfr24-HIV-2 TAR complex at 600 MHz. The amide that shifts significantly upon binding TAR,and appears to be an arginine residue, as highlighted by the box.

bases involved in the base triple, U-A•U, the arginine, and the G-C base pair contacting the arginine is 1.12 Å. The architecture of the arginine binding site in these two different contexts is the same. An U-A•U base triple also forms an arginine binding site in the BIV Tat-TAR complex (Ye et al., 1995). The TAR bulge motif has also been identified in the Rex Response Element (RexRE) RNA-Rex peptide interaction by SELEX (Baskerville *et al.*, 1995). Because of the sequence similarity, we expect this structure to resemble the HIV TAR-arginine interaction. Finally, small molecules that bind to the HIV TAR bulge mimic the arginine guanidium group and bind in a very similar manner as arginine-rich peptides (Hamy *et al.*, 1997). Thus, the base triple in the TAR bulge may prove to be a common structural motif for arginine guanidium group recognition.

6.4.2 Other Tat-TAR complexes

Because of the many subtle kinetic effects involved in the Tat peptide-TAR RNA interaction, a wide variety of peptides may have to be tested to find one that would give high quality spectra for a high resolution structure determination. Perhaps the next best candidate after tfr24 is the HIV-2 tfr24 equivalent, t2fr27. Comparing the HIV-1 and HIV-2 Tat protein sequences reveals significant variation of the other amino acids in the basic RNA binding domain (Figure 6.1). Some subtle specificity differences may exist between the two and three base bulge TARs, so that it is worth investigating the HIV-2 equivalent of tfr24 (Chang & Jeang, 1992). Another possibility is to examine the C38-G27•C23$^+$ base triple mutant, discussed in Chapter 4, as the peptide does not seem to be adversely affected at lower pHs. Because of the extra G-C base pair in the upper stem, this mutant may be more stable and help improve the quality of the peptide complex spectra. Chapter 5 presents evidence suggesting that U40 may be experiencing conformational exchange, thus, mutating the A22-U40 to a G-C base pair may stabilize the peptide complex. A C22-G40 mutant binds the tfr24 peptide with a factor of two higher affinity than wild type TAR.

One of these mutants or a combination may help improve the quality of the peptide complex spectra.

Perhaps the future of structural studies of the Tat-TAR system depends on the identification of the loop binding protein and the ensuing formation of a tripartite complex. These three molecules probably bind in a highly cooperative manner providing some hope that a complex suitable for NMR and/or X-ray studies is possible.

6.4.3 Drug discovery

Extensive efforts over the last decade have focused on finding inhibitors of the HIV Tat-TAR interaction. A number of different candidates have been found that specifically bind TAR and inhibit Tat binding *in vitro*. One of the problems of HIV pharmaceuticals is HIV's remarkable ability to mutate rapidly and avoid the drug even with the new triple drug therapy approaches. The Tat-TAR interaction may be an excellent target not only because it is critical for the viral life cycle but to avoid an inhibitor, a large number of concomitant mutations would be required compared to mutating a single amino acid in a protein. Inhibitors that are based on peptide or peptoid chemistry have arginine guanidium groups that bind to TAR which even has the same conformation when bound to the drug as it does in arginine complexes (Hamy et al., 1997). This candidate inhibits the Tat-TAR interaction in both *in vivo* transactivation and in live virus assays. Perhaps finding a small molecule that can specifically bind to both the TAR loop and bulge regions would provide more specificity. One recently introduced method, Structure-Activity Relationships, SAR, to identify weak binding molecules is to test libraries of small molecules by NMR and then linking these weak binding molecules together to make a high affinity inhibitor (Shuker *et al.*, 1996). Perhaps weak binding molecules that specifically bind to different parts of TAR including the hexanucleotide loop can be linked together to form a highly specific Tat inhibitor. SAR requires large quantities RNA and more importantly a large and diverse

library of molecules to screen. However, it offers the ability to find weak binding molecules which can be used as candidates to start making high affinity inhibitors.

Much effort and many neurons have been applied to the Tat-TAR interaction and we still do not know how this critical step in the viral life cycle works. These proposed studies may not lead to the holy grail of a therapeutic agent but they may help discover more about the Tat-TAR interaction and identify specific inhibitors.

Appendix A: 2' Hydroxyls in RNA

This appendix discusses the observation of 2'hydroxyl protons in homonuclear NOESY experiments. These protons are rarely observed in RNA spectra. The few examples that have been identified are involved in unusual structure like in tetraloops, tobramycin or ATP binding. In typical A-form helices, the 2'OH is directed out into solution and does not make any hydrogen bonding contacts with other parts of the structure. Thus, the 2'OH proton is exchanging rapidly with water making the observation and assignment of 2'OH protons very difficult. However, at higher fields and lower temperatures, 2'OH protons were observed in homonuclear watergate NOESY experiments. They were identified by their unique pattern of NOEs to ribose protons, summarized in Table A.1, and through spin-diffusion cross peaks to aromatic H8 and H6 protons. Figure A.1 shows that some of the 2'OH protons that were observed in the HIV 2 TAR - argininamide complex. Of course NOESY assignments are unreliable, so numerous attempts were made to devise experiments to make through bond correlations to the 2'OH protons. These experiments would need to overcome both the fast water exchange rate and the small two bond coupling constants to either the C2' or >3 bond coupling to a nearby proton. Neither CPMG type experiments nor direct INEPT and TOCSY transfers succeeded. Perhaps, higher field and better signal to noise spectrometers will help make unambiguous assignments of 2'OH protons. Although there does not seem to be many particularly useful NOEs to 2'OH protons in a standard A-form helix, their unambiguous through-bond assignment in other RNA structures may be imperative.
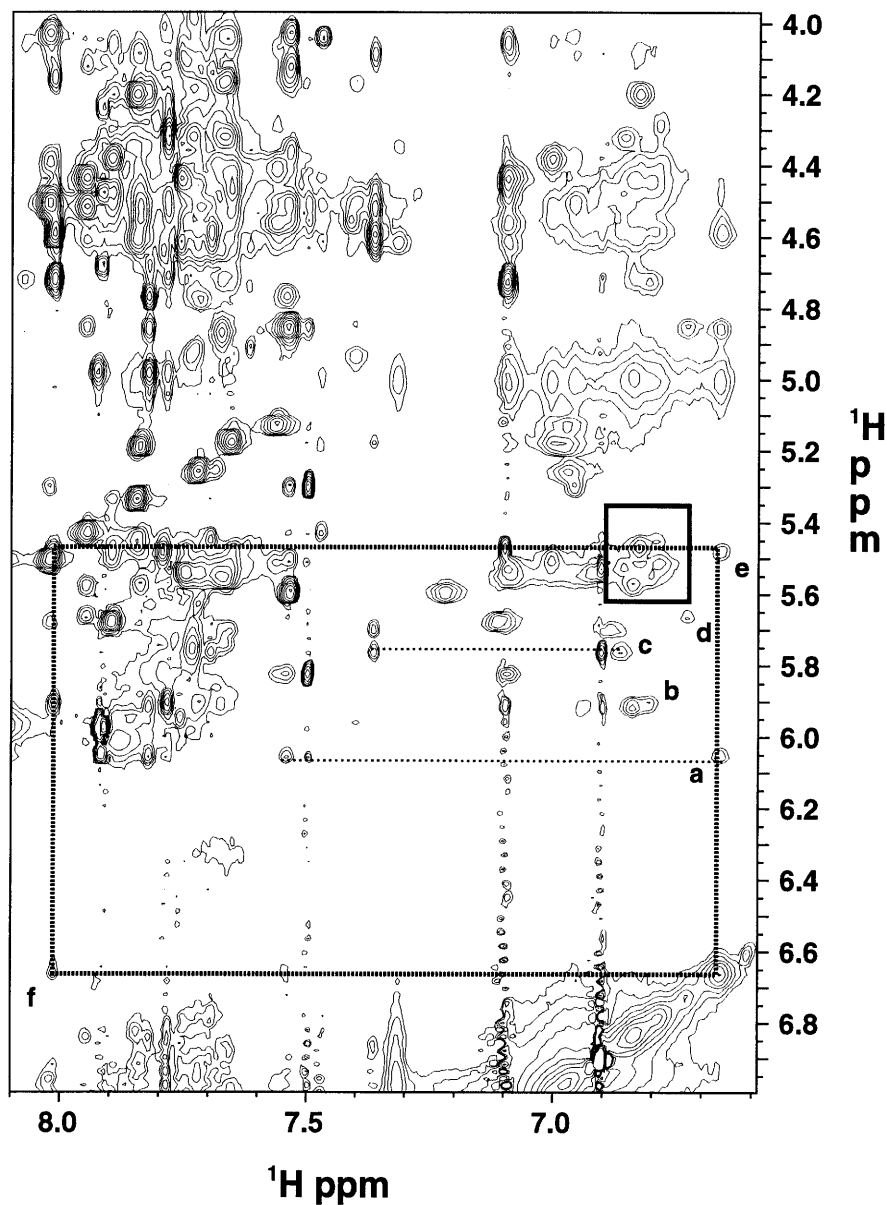
Figure A.1: A 40 ms 750 MHz watergate NOESY at 5°C reveals a number of assignable 2'OH proton resonances. **a)** A27(2'OH),A27(H1') **b)** A20(2'OH),A20(H1') **c)** G43(2'OH),G43(H1') **d)** U38(2'OH),U38(H1') **e)** G21(2'OH),G21(H1') **f)** A20(H8),G21(2'OH) The solid rectangle highlights other non-assigned 2'OH resonances. The dashed rectangle illustrates how the G21(2'OH) was assigned by identifying peak **f**. The dashed lines correlate the H1' chemical shifts to the internucleotide aromatic crosspeak used for assignment.

Appendix B: Molecular Modeling Details

This appendix is a compilation of the XPLOR scripts and NMR-derived restraints used for the structure determination. These files provide all the details of exactly how the structure was calculated, as described in the materials and methods section of Chapter 3. In addition the non-experimental files are also included showing how the argininamide and base triple hydrogen bonds were restrained along with A-form torsions. The first section contains the three .tbl files which contain all the NMR restraints along with the .tbl files used to define the various hydrogen bonding and planarity restraints for the modeling calculations discussed in section 3.6. Appendix B.2 lists the four XPLOR scripts that constitute the molecular dynamics protocol used for structure determination. Numerous additional XPLOR and PERL scripts were written to analyze the structures and some of the more useful and generally applicable ones are compiled in appendix B.3. These scripts include measuring torsions from XPLOR .pdb files, measuring particular distances, like the U23-N3 to A27-N7. Also, a PERL script which converts the .tbl files to INSIGHT .rstrnt files is provided. These scripts were necessary to make the NOE figures of Chapter 3. Note that for many of the PERL scripts, master csh scripts are used to direct input/output and run logic loops through multiple structures/files. Finally, in section B.4, a series of csh scripts, a c program, and a PERL script that work together to convert XPLOR .pdb files to standard .pdb (c.a. 1997) for submission to the Brookhaven PDB are presented. For further explanations of the XPLOR scripts, consult the XPLOR manual (Brunger, 1992) or the excellent on-line manual at the XPLOR web page. The final 20 structures have been deposited into the Brookhaven Protein Data Bank.

Appendix B.1 NMR restraints

NOE Restraints:

These are all the NOE restraints used for the structure determination. Residue number 47 is the argininamide. The three columns define a distance range in the following order: the center distance, center - this column, center + this column.

! noes in order residue number order
! for square well potential, the format is: d dminus dplus
! adjusted ARG HH restraints to 4.0 A upper bound and two epsilon noes where
! spin diffusion is unlikely to 5.0 A upper bound
!removed intranuc H5 - H4* and H5-H6 noes
! made H5 noes upper bound 6.0 A except for the strong internuc H5-H3'

set echo = off end
set message = off end

! LOWER STEM

assign (resid 16 and name H2')(resid 16 and name H1')	2.40	0.60	0.60
assign (resid 16 and name H3')(resid 16 and name H1')	2.90	1.10	1.10
assign (resid 16 and name H3')(resid 16 and name H2')	2.40	0.60	0.60
assign (resid 16 and name H4')(resid 16 and name H1')	2.90	1.10	1.10
assign (resid 16 and name H4')(resid 16 and name H2')	3.40	1.60	1.60
assign (resid 16 and name H4')(resid 16 and name H3')	2.40	0.60	0.60
assign (resid 16 and name H5' or resid 16 and name H5")(resid 16 and name H4')	2.90	1.10	0.10
assign (resid 16 and name H5' or resid 16 and name H5")(resid 16 and name H1')	3.90	2.10	1.10
assign (resid 16 and name H5' or resid 16 and name H5")(resid 16 and name H2')	3.90	2.10	1.10
assign (resid 16 and name H5' or resid 16 and name H5")(resid 16 and name H3')	3.40	1.60	1.60
assign (resid 16 and name H8)(resid 16 and name H1')	3.40	1.60	1.60
assign (resid 16 and name H8)(resid 16 and name H2')	2.90	1.10	1.10
assign (resid 16 and name H8)(resid 16 and name H3')	3.40	1.60	1.60
assign (resid 16 and name H8)(resid 16 and name H4')	3.40	1.60	1.60
assign (resid 16 and name H8)(resid 16 and name H5' or resid 16 and name H5")	3.90	2.10	1.10
assign (resid 17 and name H1')(resid 16 and name H2')	3.40	1.60	1.60
assign (resid 17 and name H2')(resid 17 and name H1')	2.40	0.60	0.60
assign (resid 17 and name H3')(resid 17 and name H1')	2.90	1.10	1.10
assign (resid 17 and name H3')(resid 17 and name H2')	2.40	0.60	0.60
assign (resid 17 and name H4')(resid 17 and name H1')	2.90	1.10	1.10
assign (resid 17 and name H4')(resid 17 and name H2')	2.90	1.10	1.10
assign (resid 17 and name H4')(resid 17 and name H3')	2.40	0.60	0.60
assign (resid 17 and name H5' or resid 17 and name H5")(resid 17 and name H3')	3.90	2.10	1.10
assign (resid 17 and name H5' or resid 17 and name H5")(resid 17 and name H4')	2.90	1.10	1.10
assign (resid 17 and name H8)(resid 16 and name H1')	3.40	1.60	1.60
assign (resid 17 and name H8)(resid 16 and name H2')	2.40	0.60	0.60
assign (resid 17 and name H8)(resid 16 and name H3')	2.90	1.10	1.10
assign (resid 17 and name H8)(resid 16 and name H8)	3.40	1.60	2.60
assign (resid 17 and name H8)(resid 17 and name H1')	2.90	1.10	1.10
assign (resid 17 and name H8)(resid 17 and name H2')	3.40	1.60	1.60
assign (resid 17 and name H8)(resid 17 and name H3')	2.90	1.10	1.10
assign (resid 17 and name H8)(resid 17 and name H4')	3.40	1.60	1.60

```
assign (resid 17 and name H8  )(resid 17 and name H5' or resid 17 and name H5")        3.90  2.10  1.10

assign (resid 18 and name H1' )(resid 17 and name H2' )                                3.40  1.60  1.60
assign (resid 18 and name H2' )(resid 18 and name H1' )                                2.40  0.60  0.60
assign (resid 18 and name H3' )(resid 18 and name H1' )                                2.90  1.10  1.10
assign (resid 18 and name H4' )(resid 18 and name H1' )                                2.90  1.10  1.10
assign (resid 18 and name H4' )(resid 18 and name H3' )                                2.40  0.60  0.60
assign (resid 18 and name H5' or resid 18 and name H5" )(resid 17 and name H2' )       3.90  2.10  1.10
assign (resid 18 and name H5' or resid 18 and name H5" )(resid 18 and name H1' )       3.90  2.10  1.10
assign (resid 18 and name H5' or resid 18 and name H5" )(resid 18 and name H3' )       3.40  1.60  0.60
assign (resid 18 and name H5' or resid 18 and name H5" )(resid 18 and name H4' )       2.90  1.10  0.10
assign (resid 18 and name H41 )(resid 17 and name H1  )                                3.90  2.10  2.10
assign (resid 18 and name H42 )(resid 17 and name H1  )                                3.90  2.10  2.10
assign (resid 18 and name H6  )(resid 17 and name H1' )                                3.40  1.60  1.60
assign (resid 18 and name H6  )(resid 17 and name H2' )                                2.90  1.10  1.10
assign (resid 18 and name H6  )(resid 17 and name H3' )                                3.40  1.60  1.60
assign (resid 18 and name H6  )(resid 18 and name H1' )                                3.40  1.60  1.60
assign (resid 18 and name H6  )(resid 18 and name H2' )                                3.40  1.60  1.60
assign (resid 18 and name H6  )(resid 18 and name H3' )                                2.40  0.60  0.60
assign (resid 18 and name H6  )(resid 18 and name H4' )                                3.40  1.60  1.60
assign (resid 18 and name H6  )(resid 18 and name H5' or resid 18 and name H5" )       3.90  2.10  1.10
assign (resid 18 and name H5  )(resid 17 and name H2' )                                3.65  1.85  2.35
assign (resid 18 and name H5  )(resid 17 and name H3' )                                3.65  1.85  1.35
assign (resid 18 and name H5  )(resid 18 and name H3' )                                3.40  1.60  2.60

assign (resid 19 and name H2' )(resid 19 and name H1' )                                2.40  0.60  0.60
assign (resid 19 and name H3' )(resid 19 and name H1' )                                2.90  1.10  1.10
assign (resid 19 and name H3' )(resid 19 and name H2' )                                2.40  0.60  0.60
assign (resid 19 and name H4' )(resid 19 and name H2' )                                3.40  1.60  1.60
assign (resid 19 and name H42 )(resid 18 and name H42 )                                3.90  2.10  2.10
assign (resid 19 and name H42 )(resid 18 and name H5  )                                3.90  2.10  2.10
assign (resid 19 and name H6  )(resid 18 and name H2' )                                2.40  0.60  0.60
assign (resid 19 and name H6  )(resid 18 and name H3' )                                2.90  1.10  1.10
assign (resid 19 and name H6  )(resid 19 and name H1' )                                3.40  1.60  1.60
assign (resid 19 and name H6  )(resid 19 and name H3' )                                2.90  1.10  1.10
assign (resid 19 and name H6  )(resid 19 and name H5' or resid 19 and name H5" )       3.90  2.10  1.10
assign (resid 19 and name H5  )(resid 18 and name H41 )                                3.90  2.10  2.10
assign (resid 19 and name H5  )(resid 18 and name H42 )                                3.90  2.10  2.10
assign (resid 19 and name H5  )(resid 18 and name H5  )                                3.40  1.60  2.60

assign (resid 20 and name H1' )(resid 19 and name H2' )                                3.40  1.60  1.60
assign (resid 20 and name H2' )(resid 20 and name H1' )                                2.40  0.60  0.60
assign (resid 20 and name H3' )(resid 19 and name H2' )                                3.40  1.60  1.60
assign (resid 20 and name H3' )(resid 20 and name H1' )                                2.90  1.10  1.10
assign (resid 20 and name H3' )(resid 20 and name H2' )                                2.40  0.60  0.60
assign (resid 20 and name H4' )(resid 20 and name H1' )                                2.90  1.10  1.10
assign (resid 20 and name H4' )(resid 20 and name H2' )                                2.90  1.10  1.10
assign (resid 20 and name H4' )(resid 20 and name H3' )                                2.40  0.60  0.60
assign (resid 20 and name H5' or resid 20 and name H5")(resid 20 and name H3' )        3.40  1.60  1.60
assign (resid 20 and name H5' or resid 20 and name H5")(resid 20 and name H4' )        2.90  1.10  1.10
assign (resid 20 and name H8  )(resid 19 and name H1' )                                3.40  1.60  1.60
assign (resid 20 and name H8  )(resid 19 and name H2' )                                2.40  0.60  0.60
assign (resid 20 and name H8  )(resid 19 and name H3' )                                2.90  1.10  1.10
assign (resid 20 and name H8  )(resid 19 and name H6  )                                3.40  1.60  2.60
```

```
assign (resid 20 and name H8  )(resid 20 and name H1' )                              3.40  1.60  1.60
assign (resid 20 and name H8  )(resid 20 and name H2' )                              3.40  1.60  1.60
assign (resid 20 and name H8  )(resid 20 and name H3' )                              2.90  1.10  1.10
assign (resid 20 and name H8  )(resid 20 and name H4' )                              3.40  1.60  1.60
assign (resid 20 and name H8  )(resid 20 and name H5' or resid 20 and name H5")      3.90  2.10  1.10
assign (resid 20 and name H2  )(resid 20 and name H1' )                              3.40  1.60  1.60
assign (resid 20 and name H2  )(resid 20 and name H2' )                              3.40  1.60  1.60

assign (resid 21 and name H1' )(resid 20 and name H2' )                              3.40  1.60  1.60
assign (resid 21 and name H1' )(resid 20 and name H2 )                               2.90  1.10  1.10
assign (resid 21 and name H2' )(resid 21 and name H1' )                              2.40  0.60  0.60
assign (resid 21 and name H3' )(resid 20 and name H2' )                              3.40  1.60  1.60
assign (resid 21 and name H3' )(resid 21 and name H1' )                              3.40  1.60  1.60
assign (resid 21 and name H3' )(resid 21 and name H2' )                              2.40  0.60  0.60
assign (resid 21 and name H4' )(resid 21 and name H1' )                              2.90  1.10  1.10
assign (resid 21 and name H4' )(resid 21 and name H3' )                              2.40  0.60  0.60
assign (resid 21 and name H5' or resid 21 and name H5")(resid 20 and name H2' )      3.90  2.10  1.10
assign (resid 21 and name H5' or resid 21 and name H5")(resid 21 and name H3' )      3.90  2.10  1.10
assign (resid 21 and name H5' or resid 21 and name H5")(resid 21 and name H4' )      2.90  1.10  0.10
assign (resid 21 and name H8  )(resid 20 and name H1' )                              3.40  1.60  1.60
assign (resid 21 and name H8  )(resid 20 and name H2' )                              2.40  0.60  0.60
assign (resid 21 and name H8  )(resid 20 and name H3' )                              2.90  1.10  1.10
assign (resid 21 and name H8  )(resid 20 and name H8 )                               3.40  1.60  2.60
assign (resid 21 and name H8  )(resid 20 and name H2 )                               3.40  1.60  1.60
assign (resid 21 and name H8  )(resid 21 and name H1' )                              3.40  1.60  1.60
assign (resid 21 and name H8  )(resid 21 and name H2' )                              3.40  1.60  1.60
assign (resid 21 and name H8  )(resid 21 and name H3' )                              2.40  0.60  0.60
assign (resid 21 and name H8  )(resid 21 and name H4' )                              3.40  1.60  1.60
assign (resid 21 and name H8  )(resid 21 and name H5' or resid 21 and name H5")      3.90  2.10  1.10
assign (resid 21 and name H1  )(resid 20 and name H2 )                               3.90  2.10  2.10

assign (resid 22 and name H1' )(resid 21 and name H2' )                              3.40  1.60  1.60
assign (resid 22 and name H2' )(resid 22 and name H1' )                              2.40  0.60  0.60
assign (resid 22 and name H3' )(resid 22 and name H1' )                              2.90  1.10  1.10
assign (resid 22 and name H3' )(resid 22 and name H2' )                              2.40  0.60  0.60
assign (resid 22 and name H4' )(resid 22 and name H1' )                              2.90  1.10  1.10
assign (resid 22 and name H4' )(resid 22 and name H2' )                              2.90  1.10  1.10
assign (resid 22 and name H4' )(resid 22 and name H3' )                              2.40  0.60  0.60
assign (resid 22 and name H5' or resid 22 and name H5")(resid 22 and name H3' )      3.90  2.10  1.10
assign (resid 22 and name H8  )(resid 21 and name H1' )                              3.40  1.60  1.60
assign (resid 22 and name H8  )(resid 21 and name H2' )                              2.40  0.60  0.60
assign (resid 22 and name H8  )(resid 21 and name H3' )                              3.40  1.60  1.60
assign (resid 22 and name H8  )(resid 21 and name H8 )                               3.40  1.60  2.60
assign (resid 22 and name H8  )(resid 22 and name H1' )                              3.40  1.60  1.60
assign (resid 22 and name H8  )(resid 22 and name H2' )                              3.40  1.60  1.60
assign (resid 22 and name H8  )(resid 22 and name H3' )                              2.90  1.10  1.10
assign (resid 22 and name H8  )(resid 22 and name H4' )                              3.40  1.60  1.60
assign (resid 22 and name H2  )(resid 21 and name H1 )                               3.40  1.60  1.60
assign (resid 22 and name H2  )(resid 21 and name H22 )                              3.90  2.10  2.10
assign (resid 22 and name H2  )(resid 22 and name H1' )                              3.40  1.60  1.60
assign (resid 22 and name H2  )(resid 22 and name H2' )                              3.90  2.10  2.10

! INTERNUCLEOTIDE NOES TO U23
assign (resid 23 and name H5' or resid 23 and name H5")(resid 22 and name H2' )      2.90  1.10  2.10
```

189

```
assign (resid 23 and name H1' )(resid 26 and name H2' or resid 26 and name H3' )      2.90  1.10  1.10
assign (resid 23 and name H1' )(resid 26 and name H4' )                                3.40  1.60  1.60
assign (resid 23 and name H1' )(resid 26 and name H5' or resid 26 and name H5")        2.90  1.10  1.10
assign (resid 23 and name H1' )(resid 26 and name H8  )                                2.90  1.10  1.10
assign (resid 23 and name H4' )(resid 26 and name H5' or resid 26 and name H5")        3.40  1.60  1.60
assign (resid 23 and name H4' )(resid 26 and name H8  )                                3.40  1.60  1.60
```

! INTRANUCLEOTIDE NOES TO U23
```
assign (resid 23 and name H2' )(resid 23 and name H1' )                                2.40  0.60  0.60
assign (resid 23 and name H3' )(resid 23 and name H1' )                                3.40  1.60  1.60
assign (resid 23 and name H3' )(resid 23 and name H2' )                                2.40  0.60  0.60
assign (resid 23 and name H3' )(resid 23 and name H5' or resid 23 and name H5")        3.40  1.60  1.60
assign (resid 23 and name H4' )(resid 23 and name H1' )                                2.40  0.60  1.60
assign (resid 23 and name H4' )(resid 23 and name H2' )                                2.90  1.10  1.10
assign (resid 23 and name H4' )(resid 23 and name H3' )                                2.40  0.60  0.60
assign (resid 23 and name H5' or resid 23 and name H5")(resid 23 and name H1' )        3.65  1.85  1.35
assign (resid 23 and name H5' or resid 23 and name H5")(resid 23 and name H4' )        2.65  0.85  0.35
assign (resid 23 and name H6  )(resid 23 and name H1' )                                2.90  1.10  1.10
assign (resid 23 and name H6  )(resid 23 and name H2' )                                2.90  1.10  1.10
assign (resid 23 and name H6  )(resid 23 and name H3' )                                2.90  1.10  2.10
assign (resid 23 and name H6  )(resid 23 and name H4' )                                2.90  1.10  2.10
assign (resid 23 and name H5  )(resid 23 and name H2' )                                3.65  1.85  2.35
```

!internucleotide NOEs to U25
```
assign (resid 25 and name H3' )(resid 23 and name H4' )                                2.90  1.10  1.10
assign (resid 25 and name H1' )(resid 42 and name H1' )                                3.40  1.60  1.60
```

! intranucleotide NOEs to U25
```
assign (resid 25 and name H2' )(resid 25 and name H1' )                                2.40  0.60  0.60
assign (resid 25 and name H3' )(resid 25 and name H1' )                                2.90  1.10  1.10
assign (resid 25 and name H3' )(resid 25 and name H2' )                                2.40  0.60  0.60
assign (resid 25 and name H4' )(resid 25 and name H1' )                                2.40  0.60  1.60
assign (resid 25 and name H4' )(resid 25 and name H2' )                                3.40  1.60  1.60
assign (resid 25 and name H4' )(resid 25 and name H3' )                                2.40  0.60  0.60
assign (resid 25 and name H5' or resid 25 and name H5")(resid 25 and name H3' )        3.40  1.60  1.60
assign (resid 25 and name H5' or resid 25 and name H5")(resid 25 and name H4' )        2.90  1.10  0.10
assign (resid 25 and name H6  )(resid 25 and name H1' )                                2.90  1.10  1.10
assign (resid 25 and name H6  )(resid 25 and name H2' )                                2.40  0.60  0.60
assign (resid 25 and name H6  )(resid 25 and name H3' )                                2.90  1.10  2.10
assign (resid 25 and name H6  )(resid 25 and name H4' )                                3.40  1.60  1.60
assign (resid 25 and name H6  )(resid 25 and name H5' or resid 25 and name H5")        3.65  1.85  1.35
assign (resid 25 and name H5  )(resid 25 and name H2' )                                3.65  1.85  2.35
```

```
assign (resid 26 and name H1' )(resid 22 and name H2  )                                3.40  1.60  1.60
assign (resid 26 and name H2' or resid 26 and name H3' )(resid 26 and name H1' )       2.50  0.70  0.50
assign (resid 26 and name H4' )(resid 26 and name H1' )                                2.40  0.60  1.60
assign (resid 26 and name H4' )(resid 26 and name H2' or resid 26 and name H3' )       3.40  1.60  1.60
assign (resid 26 and name H5' or resid 26 and name H5")(resid 26 and name H1' )        3.90  2.10  1.10
assign (resid 26 and name H5' or resid 26 and name H5")(resid 26 and name H2' or resid 26 and name H3' )
                                                                                       3.90  2.10  2.10
assign (resid 26 and name H5' or resid 26 and name H5")(resid 26 and name H4' )        2.90  1.10  0.10
assign (resid 26 and name H8  )(resid 22 and name H2' )                                3.65  1.85  1.85
assign (resid 26 and name H8  )(resid 22 and name H2  )                                3.40  1.60  1.60
assign (resid 26 and name H8  )(resid 26 and name H1' )                                3.40  1.60  1.60
```

190

```
assign (resid 26 and name H8  )(resid 26 and name H2' or resid 26 and name H3' )      3.90  2.10  1.10
assign (resid 26 and name H8  )(resid 26 and name H4' )                               3.40  1.60  1.60
assign (resid 26 and name H8  )(resid 26 and name H5' or resid 26 and name H5")       3.90  2.10  2.10
assign (resid 26 and name H1  )(resid 22 and name H2  )                               3.40  1.60  1.60

assign (resid 27 and name H1' )(resid 26 and name H2' or resid 26 and name H3' )      3.90  2.10  2.10
assign (resid 27 and name H1' )(resid 26 and name H22 )                               3.90  2.10  2.10
assign (resid 27 and name H1' )(resid 26 and name H21 )                               3.90  2.10  2.10
assign (resid 27 and name H2' )(resid 27 and name H1' )                               2.40  0.60  0.60
assign (resid 27 and name H3' )(resid 26 and name H2' or resid 26 and name H3' )      3.90  2.10  2.10
assign (resid 27 and name H3' )(resid 27 and name H1' )                               2.90  1.10  1.10
assign (resid 27 and name H3' )(resid 27 and name H2' )                               2.40  0.60  0.60
assign (resid 27 and name H4' )(resid 27 and name H1' )                               2.90  1.10  1.10
assign (resid 27 and name H4' )(resid 27 and name H2' )                               3.40  1.60  1.60
assign (resid 27 and name H4' )(resid 27 and name H3' )                               2.40  0.60  0.60
assign (resid 27 and name H5' or resid 27 and name H5" )(resid 26 and name H2' or resid 26 and name H3' )
                                                                                     4.40  2.60  2.60
assign (resid 27 and name H5' or resid 27 and name H5" )(resid 27 and name H3' )      3.90  2.10  1.10
assign (resid 27 and name H5' or resid 27 and name H5" )(resid 27 and name H4' )      2.90  1.10  0.10
assign (resid 27 and name H8  )(resid 26 and name H1' )                               3.40  1.60  1.60
assign (resid 27 and name H8  )(resid 26 and name H2' or resid 26 and name H3' )      3.90  2.10  2.10
assign (resid 27 and name H8  )(resid 26 and name H8  )                               3.40  1.60  2.60
assign (resid 27 and name H8  )(resid 27 and name H1' )                               3.40  1.60  1.60
assign (resid 27 and name H8  )(resid 27 and name H2' )                               3.40  1.60  1.60
assign (resid 27 and name H8  )(resid 27 and name H3' )                               2.90  1.10  1.10
assign (resid 27 and name H8  )(resid 27 and name H4' )                               3.40  1.60  1.60
assign (resid 27 and name H8  )(resid 27 and name H5' or resid 27 and name H5" )      3.90  2.10  1.10
assign (resid 27 and name H2  )(resid 26 and name H1  )                               3.90  2.10  2.10
assign (resid 27 and name H61 or resid 27 and name H62)(resid 26 and name H1  )       3.90  2.10  2.10

assign (resid 28 and name H1' )(resid 27 and name H2' )                               3.40  1.60  1.60
assign (resid 28 and name H1' )(resid 27 and name H2  )                               2.90  1.10  1.10
assign (resid 28 and name H2' )(resid 28 and name H1' )                               2.40  0.60  0.60
assign (resid 28 and name H3' )(resid 27 and name H2' )                               3.40  1.60  1.60
assign (resid 28 and name H3' )(resid 28 and name H1' )                               3.40  1.60  1.60
assign (resid 28 and name H3' )(resid 28 and name H2' )                               2.40  0.60  0.60
assign (resid 28 and name H4' )(resid 28 and name H1' )                               2.90  1.10  1.10
assign (resid 28 and name H4' )(resid 28 and name H2' )                               2.90  1.10  1.10
assign (resid 28 and name H4' )(resid 28 and name H3' )                               2.40  0.60  0.60
assign (resid 28 and name H5' or resid 28 and name H5" )(resid 27 and name H2' )      3.90  2.10  1.10
assign (resid 28 and name H5' or resid 28 and name H5" )(resid 28 and name H3' )      3.40  1.60  0.60
assign (resid 28 and name H5' or resid 28 and name H5" )(resid 28 and name H4' )      2.90  1.10  0.10
assign (resid 28 and name H8  )(resid 27 and name H1' )                               2.90  1.10  2.10
assign (resid 28 and name H8  )(resid 27 and name H2' )                               2.40  0.60  0.60
assign (resid 28 and name H8  )(resid 27 and name H3' )                               2.90  1.10  1.10
assign (resid 28 and name H8  )(resid 27 and name H8  )                               3.40  1.60  2.60
assign (resid 28 and name H8  )(resid 28 and name H1' )                               3.40  1.60  1.60
assign (resid 28 and name H8  )(resid 28 and name H2' )                               3.40  1.60  1.60
assign (resid 28 and name H8  )(resid 28 and name H3' )                               2.90  1.10  1.10
assign (resid 28 and name H8  )(resid 28 and name H4' )                               3.40  1.60  1.60
assign (resid 28 and name H8  )(resid 28 and name H5' or resid 28 and name H5" )      3.65  1.85  1.35
assign (resid 28 and name H1  )(resid 27 and name H2  )                               3.90  2.10  2.10
assign (resid 28 and name H1  )(resid 27 and name H61 or resid 27 and name H62)       3.90  2.10  2.10
```

```
assign (resid 29 and name H1' )(resid 28 and name H2' )                            3.40  1.60  1.60
assign (resid 29 and name H2' )(resid 29 and name H1' )                            2.40  0.60  0.60
assign (resid 29 and name H3' )(resid 29 and name H1' )                            2.90  1.10  1.10
assign (resid 29 and name H3' )(resid 29 and name H2' )                            2.40  0.60  0.60
assign (resid 29 and name H4' )(resid 29 and name H1' )                            2.90  1.10  1.10
assign (resid 29 and name H4' )(resid 29 and name H2' )                            2.90  1.10  1.10
assign (resid 29 and name H41 )(resid 28 and name H1  )                            3.90  2.10  2.10
assign (resid 29 and name H42 )(resid 28 and name H1  )                            3.90  2.10  2.10
assign (resid 29 and name H6  )(resid 28 and name H1' )                            3.40  1.60  1.60
assign (resid 29 and name H6  )(resid 28 and name H2' )                            2.40  0.60  0.60
assign (resid 29 and name H6  )(resid 28 and name H3' )                            3.40  1.60  1.60
assign (resid 29 and name H6  )(resid 28 and name H8  )                            3.40  1.60  2.60
assign (resid 29 and name H6  )(resid 29 and name H1' )                            3.40  1.60  1.60
assign (resid 29 and name H6  )(resid 29 and name H2' )                            3.40  1.60  1.60
assign (resid 29 and name H6  )(resid 29 and name H3' )                            3.40  1.60  1.60
assign (resid 29 and name H6  )(resid 29 and name H4' )                            3.40  1.60  1.60
assign (resid 29 and name H6  )(resid 29 and name H5' or resid 29 and name H5" )   3.90  2.10  1.10
assign (resid 29 and name H5  )(resid 28 and name H2' )                            3.90  2.10  2.10
assign (resid 29 and name H5  )(resid 28 and name H3' )                            3.65  1.85  1.35
assign (resid 29 and name H5  )(resid 28 and name H8  )                            3.65  1.85  2.35
assign (resid 29 and name H5  )(resid 28 and name H1  )                            3.90  2.10  2.10
assign (resid 29 and name H5  )(resid 29 and name H3' )                            3.40  1.60  2.60

! TAR   LOOP
assign (resid 30 and name H1' ) (resid 29 and name H2')                            2.40  0.60  2.60
assign (resid 30 and name H2' )(resid 30 and name H1' )                            2.40  0.60  0.60
assign (resid 30 and name H3' )(resid 30 and name H1' )                            3.40  1.60  1.60
assign (resid 30 and name H3' )(resid 30 and name H2' )                            2.40  0.60  0.60
assign (resid 30 and name H4' )(resid 29 and name H2' )                            2.90  1.10  2.10
assign (resid 30 and name H4' )(resid 30 and name H1' )                            2.90  1.10  1.10
assign (resid 30 and name H4' )(resid 30 and name H2' )                            3.40  1.60  1.60
assign (resid 30 and name H4' )(resid 30 and name H3' )                            2.40  0.60  0.60
assign (resid 30 and name H5' or resid 30 and name H5" )(resid 30 and name H3' )   3.90  2.10  1.10
assign (resid 30 and name H5' or resid 30 and name H5" )(resid 30 and name H4' )   3.40  1.60  0.60
assign (resid 30 and name H6  )(resid 29 and name H1' )                            3.40  1.60  1.60
assign (resid 30 and name H6  )(resid 29 and name H2' )                            2.40  0.60  0.60
assign (resid 30 and name H6  )(resid 29 and name H6  )                            3.90  2.10  2.10
assign (resid 30 and name H6  )(resid 30 and name H1' )                            3.40  1.60  1.60
assign (resid 30 and name H6  )(resid 30 and name H2' )                            2.40  0.60  1.60
assign (resid 30 and name H6  )(resid 30 and name H3' )                            3.40  1.60  1.60
assign (resid 30 and name H6  )(resid 30 and name H4' )                            3.40  1.60  2.60
assign (resid 30 and name H5  )(resid 29 and name H41 )                            3.90  2.10  2.10
assign (resid 30 and name H5  )(resid 29 and name H42 )                            3.90  2.10  2.10
assign (resid 30 and name H5  )(resid 29 and name H5  )                            3.65  1.85  2.35

assign (resid 31 and name H2' )(resid 31 and name H1' )                            2.40  0.60  0.60
assign (resid 31 and name H3' )(resid 31 and name H1' )                            3.40  1.60  1.60
assign (resid 31 and name H4' )(resid 31 and name H1' )                            2.90  1.10  1.10
assign (resid 31 and name H4' )(resid 31 and name H2' )                            3.40  1.60  1.60
assign (resid 31 and name H4' )(resid 31 and name H3' )                            2.40  0.60  0.60
assign (resid 31 and name H5' or resid 31 and name H5" )(resid 30 and name H2' )   3.90  2.10  2.10
assign (resid 31 and name H5' or resid 31 and name H5" )(resid 31 and name H2' )   3.90  2.10  2.10
assign (resid 31 and name H5' or resid 31 and name H5" )(resid 31 and name H3' )   3.90  2.10  2.10
assign (resid 31 and name H5' or resid 31 and name H5" )(resid 31 and name H4' )   2.90  1.10  1.10
```

192

```
assign (resid 31 and name H6  )(resid 30 and name H2' )                                3.40  1.60  1.60
assign (resid 31 and name H6  )(resid 30 and name H4' )                                3.40  1.60  1.60
assign (resid 31 and name H6  )(resid 31 and name H1' )                                2.90  1.10  1.10
assign (resid 31 and name H6  )(resid 31 and name H3' )                                2.90  1.10  2.10
assign (resid 31 and name H6  )(resid 31 and name H4' )                                2.90  1.10  2.10
assign (resid 31 and name H6  )(resid 31 and name H5' or resid 31 and name H5" )       2.90  1.10  2.10

assign (resid 32 and name H2' )(resid 32 and name H1' )                                2.40  0.60  0.60
assign (resid 32 and name H3' )(resid 32 and name H1' )                                3.40  1.60  1.60
assign (resid 32 and name H3' )(resid 32 and name H2' )                                2.40  0.60  0.60
assign (resid 32 and name H4' )(resid 32 and name H1' )                                2.90  1.10  1.10
assign (resid 32 and name H4' )(resid 32 and name H2' )                                3.40  1.60  1.60
assign (resid 32 and name H4' )(resid 32 and name H3' )                                2.40  0.60  0.60
assign (resid 32 and name H5' or resid 32 and name H5" )(resid 31 and name H2' )       3.90  2.10  2.10
assign (resid 32 and name H5' or resid 32 and name H5" )(resid 32 and name H1' )       3.90  2.10  2.10
assign (resid 32 and name H5' or resid 32 and name H5" )(resid 32 and name H2' )       3.90  2.10  2.10
assign (resid 32 and name H5' or resid 32 and name H5" )(resid 32 and name H3' )       3.40  1.60  1.60
assign (resid 32 and name H5' or resid 32 and name H5" )(resid 32 and name H4' )       2.90  1.10  1.10
assign (resid 32 and name H8  )(resid 32 and name H1' )                                2.90  1.10  1.10
assign (resid 32 and name H8  )(resid 32 and name H2' )                                2.90  1.10  1.10
assign (resid 32 and name H8  )(resid 32 and name H3' )                                3.40  1.60  1.60
assign (resid 32 and name H8  )(resid 32 and name H4' )                                3.40  1.60  1.60
assign (resid 32 and name H8  )(resid 32 and name H5' or resid 32 and name H5" )       3.90  2.10  2.10

assign (resid 33 and name H2' )(resid 33 and name H1' )                                2.40  0.60  0.60
assign (resid 33 and name H3' )(resid 33 and name H1' )                                3.40  1.60  1.60
assign (resid 33 and name H3' )(resid 33 and name H2' )                                2.90  1.10  1.10
assign (resid 33 and name H4' )(resid 33 and name H2' )                                3.40  1.60  1.60
assign (resid 33 and name H4' )(resid 33 and name H3' )                                2.40  0.60  0.60
assign (resid 33 and name H5' or resid 33 and name H5" )(resid 33 and name H1' )       3.90  2.10  1.10
assign (resid 33 and name H5' or resid 33 and name H5" )(resid 33 and name H2' )       3.90  2.10  1.10
assign (resid 33 and name H5' or resid 33 and name H5" )(resid 33 and name H3' )       3.40  1.60  1.60
assign (resid 33 and name H5' or resid 33 and name H5" )(resid 33 and name H4' )       2.90  1.10  0.10
assign (resid 33 and name H8  )(resid 32 and name H1' )                                2.90  1.10  1.10
assign (resid 33 and name H8  )(resid 33 and name H1' )                                3.40  1.60  1.60
assign (resid 33 and name H8  )(resid 33 and name H2' )                                2.90  1.10  1.10
assign (resid 33 and name H8  )(resid 33 and name H3' )                                3.40  1.60  1.60
assign (resid 33 and name H8  )(resid 33 and name H4' )                                2.90  1.10  2.10

assign (resid 34 and name H2' )(resid 34 and name H1' )                                2.40  0.60  0.60
assign (resid 34 and name H3' )(resid 34 and name H1' )                                2.90  1.10  1.10
assign (resid 34 and name H3' )(resid 34 and name H2' )                                2.90  1.10  0.10
assign (resid 34 and name H4' )(resid 34 and name H1' )                                2.90  1.10  1.10
assign (resid 34 and name H4' )(resid 34 and name H3' )                                2.40  0.60  0.60
assign (resid 34 and name H5' or resid 34 and name H5" )(resid 34 and name H1' )       3.90  2.10  1.10
assign (resid 34 and name H5' or resid 34 and name H5" )(resid 34 and name H3' )       3.90  2.10  1.10
assign (resid 34 and name H5' or resid 34 and name H5" )(resid 34 and name H4' )       2.90  1.10  0.10
assign (resid 34 and name H8  )(resid 34 and name H1' )                                2.90  1.10  1.10
assign (resid 34 and name H8  )(resid 34 and name H2' )                                2.40  0.60  0.60
assign (resid 34 and name H8  )(resid 34 and name H3' )                                3.40  1.60  1.60
assign (resid 34 and name H8  )(resid 34 and name H5' or resid 34 and name H5" )       3.90  2.10  2.10

assign (resid 35 and name H2' )(resid 35 and name H1' )                                2.40  0.60  0.60
assign (resid 35 and name H3' )(resid 35 and name H1' )                                3.40  1.60  1.60
```

193

assign (resid 35 and name H4')(resid 35 and name H1')	2.90	1.10	1.10
assign (resid 35 and name H4')(resid 35 and name H2')	2.90	1.10	1.10
assign (resid 35 and name H4')(resid 35 and name H3')	2.40	0.60	0.60
assign (resid 35 and name H5' or resid 35 and name H5")(resid 34 and name H3')	3.90	2.10	2.10
assign (resid 35 and name H5' or resid 35 and name H5")(resid 35 and name H1')	3.90	2.10	2.10
assign (resid 35 and name H5' or resid 35 and name H5")(resid 35 and name H2')	3.90	2.10	2.10
assign (resid 35 and name H5' or resid 35 and name H5")(resid 35 and name H3')	3.90	2.10	2.10
assign (resid 35 and name H5' or resid 35 and name H5")(resid 35 and name H4')	2.90	1.10	1.10
assign (resid 35 and name H8)(resid 34 and name H3')	3.40	1.60	1.60
assign (resid 35 and name H8)(resid 35 and name H1')	3.40	1.60	1.60
assign (resid 35 and name H8)(resid 35 and name H2')	2.90	1.10	2.10
assign (resid 35 and name H8)(resid 35 and name H3')	2.90	1.10	2.10
assign (resid 35 and name H8)(resid 35 and name H4')	3.40	1.60	1.60
assign (resid 35 and name H8)(resid 35 and name H5' or resid 35 and name H5")	3.90	2.10	1.10

! UPPER STEM

assign (resid 36 and name H1')(resid 34 and name H1')	3.40	1.60	1.60
assign (resid 36 and name H2')(resid 36 and name H1')	2.40	0.60	0.60
assign (resid 36 and name H3')(resid 36 and name H1')	3.40	1.60	1.60
assgn (resid 36 and name H3')(resid 36 and name H2')	2.40	0.60	0.60
assign (resid 36 and name H4')(resid 36 and name H1')	3.40	1.60	1.60
assign (resid 36 and name H4')(resid 36 and name H2')	3.40	1.60	1.60
assign (resid 36 and name H4')(resid 36 and name H3')	2.40	0.60	0.60
assign (resid 36 and name H5' or resid 36 and name H5")(resid 36 and name H3')	3.90	2.10	1.10
assign (resid 36 and name H5' or resid 36 and name H5")(resid 36 and name H4')	2.90	1.10	0.10
assign (resid 36 and name H8)(resid 34 and name H1')	3.40	1.60	2.60
assign (resid 36 and name H8)(resid 35 and name H2')	3.40	1.60	2.60
assign (resid 36 and name H8)(resid 35 and name H3')	3.40	1.60	2.60
assign (resid 36 and name H8)(resid 34 and name H8)	3.40	1.60	2.60
assign (resid 36 and name H8)(resid 35 and name H8)	3.40	1.60	2.60
assign (resid 36 and name H8)(resid 36 and name H1')	3.40	1.60	1.60
assign (resid 36 and name H8)(resid 36 and name H2')	3.40	1.60	1.60
assign (resid 36 and name H8)(resid 36 and name H3')	3.40	1.60	1.60
assign (resid 36 and name H8)(resid 36 and name H4')	3.40	1.60	1.60
assign (resid 36 and name H8)(resid 36 and name H5' or resid 36 and name H5")	3.90	2.10	1.10
assign (resid 36 and name H1)(resid 28 and name H1)	3.40	1.60	1.60
assign (resid 36 and name H1)(resid 29 and name H41)	3.90	2.10	2.10
assign (resid 36 and name H1)(resid 29 and name H42)	2.90	1.10	1.10
assign (resid 36 and name H1)(resid 29 and name H5)	3.90	2.10	2.10
assign (resid 36 and name H1)(resid 30 and name H1')	3.40	1.60	2.60
assign (resid 36 and name H1)(resid 37 and name H1')	3.40	1.60	2.60

assign (resid 37 and name H1')(resid 36 and name H2')	3.40	1.60	1.60
assign (resid 37 and name H2')(resid 37 and name H1')	2.40	0.60	0.60
assign (resid 37 and name H3')(resid 37 and name H1')	2.90	1.10	1.10
assign (resid 37 and name H4')(resid 37 and name H1')	2.90	1.10	1.10
assign (resid 37 and name H41)(resid 27 and name H61 or resid 27 and name H62)	3.90	2.10	2.10
assign (resid 37 and name H42)(resid 27 and name H61 or resid 27 and name H62)	3.90	2.10	2.10
assign (resid 37 and name H41)(resid 28 and name H1)	3.90	1.10	2.10
assign (resid 37 and name H41)(resid 36 and name H1)	3.90	2.10	2.10
assign (resid 37 and name H42)(resid 28 and name H1)	2.90	1.10	1.10
assign (resid 37 and name H42)(resid 36 and name H1)	3.90	2.10	2.10
assign (resid 37 and name H6)(resid 36 and name H1')	2.90	1.10	2.10
assign (resid 37 and name H6)(resid 36 and name H2')	2.40	0.60	0.60

194

```
assign (resid 37 and name H6 )(resid 36 and name H3' )                              3.40  1.60  1.60
assign (resid 37 and name H6 )(resid 36 and name H8 )                               3.40  1.60  2.60
assign (resid 37 and name H6 )(resid 37 and name H1' )                              3.40  1.60  1.60
assign (resid 37 and name H6 )(resid 37 and name H2' )                              3.40  1.60  1.60
assign (resid 37 and name H6 )(resid 37 and name H4' )                              3.40  1.60  1.60
assign (resid 37 and name H5 )(resid 28 and name H1 )                               3.90  2.10  2.10
assign (resid 37 and name H5 )(resid 36 and name H2' )                              3.65  1.85  2.35
assign (resid 37 and name H5 )(resid 36 and name H3' )                              3.65  1.85  1.35
assign (resid 37 and name H5 )(resid 36 and name H8 )                               3.65  1.85  2.35
assign (resid 37 and name H5 )(resid 36 and name H1 )                               3.65  1.85  2.35

assign (resid 38 and name H1' )(resid 27 and name H2 )                              3.90  2.10  2.10
assign (resid 38 and name H1' )(resid 37 and name H2' )                             3.40  1.60  1.60
assign (resid 38 and name H2' )(resid 38 and name H1' )                             2.40  0.60  0.60
assign (resid 38 and name H3' )(resid 38 and name H1' )                             3.40  1.60  1.60
assign (resid 38 and name H3' )(resid 38 and name H2' )                             2.40  0.60  0.60
assign (resid 38 and name H4' )(resid 38 and name H1' )                             2.90  1.10  1.10
assign (resid 38 and name H4' )(resid 38 and name H2' )                             3.40  1.60  1.60
assign (resid 38 and name H5' or resid 38 and name H5" )(resid 38 and name H3' )    3.40  1.60  0.60
assign (resid 38 and name H3 )(resid 26 and name H1 )                               3.40  1.60  1.60
assign (resid 38 and name H3 )(resid 28 and name H1 )                               3.40  1.60  1.60
assign (resid 38 and name H3 )(resid 27 and name H2 )                               2.40  0.60  0.60
assign (resid 38 and name H3 )(resid 27 and name H61 or resid 27 and name H62 )     2.90  1.10  1.10
assign (resid 38 and name H3 )(resid 37 and name H41 )                              3.90  2.10  2.10
assign (resid 38 and name H3 )(resid 37 and name H42 )                              3.90  2.10  2.10
assign (resid 38 and name H3 )(resid 38 and name H1' )                              3.40  1.60  1.60
assign (resid 38 and name H6 )(resid 37 and name H1' )                              2.90  1.10  2.10
assign (resid 38 and name H6 )(resid 37 and name H2' )                              2.40  0.60  0.60
assign (resid 38 and name H6 )(resid 37 and name H3' )                              2.90  1.10  1.10
assign (resid 38 and name H6 )(resid 37 and name H6 )                               3.40  1.60  2.60
assign (resid 38 and name H6 )(resid 37 and name H5 )                               3.40  1.60  2.60
assign (resid 38 and name H6 )(resid 38 and name H1' )                              2.90  1.10  1.10
assign (resid 38 and name H6 )(resid 38 and name H2' )                              3.40  1.60  1.60
assign (resid 38 and name H6 )(resid 38 and name H3' )                              2.40  0.60  0.60
assign (resid 38 and name H6 )(resid 38 and name H4' )                              3.40  1.60  1.60
assign (resid 38 and name H5 )(resid 37 and name H2' )                              3.90  2.10  2.10
assign (resid 38 and name H5 )(resid 37 and name H3' )                              3.40  1.60  1.60
assign (resid 38 and name H5 )(resid 37 and name H5 )                               3.65  1.85  2.35
assign (resid 38 and name H5 )(resid 37 and name H6 )                               3.65  1.85  2.35
assign (resid 38 and name H5 )(resid 38 and name H1' )                              3.65  1.85  2.35
assign (resid 38 and name H5 )(resid 38 and name H3 )                               3.40  1.60  2.60

assign (resid 39 and name H1' )(resid 27 and name H2 )                              2.40  0.60  0.60
assign (resid 39 and name H1' )(resid 38 and name H2' )                             3.40  1.60  1.60
assign (resid 39 and name H1' )(resid 38 and name H3 )                              3.40  1.60  2.60
assign (resid 39 and name H2' )(resid 26 and name H22 )                             2.40  0.60  3.60
assign (resid 39 and name H2' )(resid 26 and name H21 )                             2.40  0.60  3.60
assign (resid 39 and name H2' )(resid 39 and name H1' )                             2.40  0.60  0.60
assign (resid 39 and name H3' )(resid 38 and name H2' )                             3.40  1.60  1.60
assign (resid 39 and name H3' )(resid 39 and name H1' )                             3.40  1.60  1.60
assign (resid 39 and name H3' )(resid 39 and name H2' )                             2.40  0.60  0.60
assign (resid 39 and name H4' )(resid 38 and name H2' )                             3.40  1.60  1.60
assign (resid 39 and name H4' )(resid 39 and name H1' )                             2.90  1.10  1.10
assign (resid 39 and name H4' )(resid 39 and name H2' )                             3.40  1.60  1.60
```

```
assign (resid 39 and name H4' )(resid 39 and name H3' )                                    2.90  1.10  1.10
assign (resid 39 and name H5' or resid 39 and name H5" )(resid 38 and name H2' )           3.90  2.10  1.10
assign (resid 39 and name H5' or resid 39 and name H5" )(resid 39 and name H1' )           3.90  2.10  1.10
assign (resid 39 and name H5' or resid 39 and name H5" )(resid 39 and name H3' )           3.90  2.10  1.10
assign (resid 39 and name H5' or resid 39 and name H5" )(resid 39 and name H4' )           3.90  2.10  1.10
assign (resid 39 and name H41 )(resid 26 and name H1 )                                      3.90  2.10  2.10
assign (resid 39 and name H42 )(resid 26 and name H1 )                                      2.90  1.10  1.10
assign (resid 39 and name H42 )(resid 26 and name H22 )                                     3.90  2.10  2.10
assign (resid 39 and name H41 )(resid 27 and name H61 or resid 27 and name H62)            3.90  2.10  2.10
assign (resid 39 and name H42 )(resid 27 and name H61 or resid 27 and name H62)            3.90  2.10  2.10
assign (resid 39 and name H42 )(resid 38 and name H5 )                                      3.90  2.10  2.10
assign (resid 39 and name H41 )(resid 38 and name H5 )                                      3.90  2.10  2.10
assign (resid 39 and name H41 )(resid 38 and name H3 )                                      3.90  2.10  2.10
assign (resid 39 and name H42 )(resid 38 and name H3 )                                      3.90  2.10  2.10
assign (resid 39 and name H6  )(resid 38 and name H1' )                                     3.40  1.60  1.60
assign (resid 39 and name H6  )(resid 38 and name H2' )                                     2.40  0.60  0.60
assign (resid 39 and name H6  )(resid 38 and name H3' )                                     3.40  1.60  1.60
assign (resid 39 and name H6  )(resid 38 and name H6 )                                      3.40  1.60  2.60
assign (resid 39 and name H6  )(resid 39 and name H1' )                                     3.40  1.60  1.60
assign (resid 39 and name H6  )(resid 39 and name H2' )                                     3.40  1.60  1.60
assign (resid 39 and name H6  )(resid 39 and name H3' )                                     2.90  1.10  1.10
assign (resid 39 and name H6  )(resid 39 and name H4' )                                     3.40  1.60  1.60
assign (resid 39 and name H6  )(resid 39 and name H5' or resid 39 and name H5" )           3.90  2.10  0.10
assign (resid 39 and name H5  )(resid 27 and name H61 or resid 27 and name H62)            4.40  2.60  1.60
assign (resid 39 and name H5  )(resid 38 and name H2' )                                     3.65  1.85  2.35
assign (resid 39 and name H5  )(resid 38 and name H3 )                                      3.90  2.10  2.10
assign (resid 39 and name H5  )(resid 38 and name H6 )                                      3.65  1.85  2.35
assign (resid 39 and name H5  )(resid 38 and name H5 )                                      3.65  1.85  2.35
assign (resid 39 and name H5  )(resid 39 and name H1' )                                     3.65  1.85  2.35


! LOWER STEM
assign (resid 40 and name H1' )(resid 26 and name H22 )                                     3.90  2.10  2.10
assign (resid 40 and name H1' )(resid 26 and name H21 )                                     3.90  2.10  2.10
assign (resid 40 and name H1' )(resid 39 and name H2' )                                     3.40  1.60  1.60
assign (resid 40 and name H1' )(resid 39 and name H3' )                                     3.40  1.60  1.60
assign (resid 40 and name H2' )(resid 26 and name H22 )                                     2.40  0.60  3.60
assign (resid 40 and name H2' )(resid 26 and name H21 )                                     2.40  0.60  3.60
assign (resid 40 and name H2' )(resid 40 and name H1' )                                     2.40  0.60  0.60
assign (resid 40 and name H3' )(resid 39 and name H2' )                                     3.40  1.60  1.60
assign (resid 40 and name H3' )(resid 40 and name H1' )                                     2.90  1.10  1.10
assign (resid 40 and name H3' )(resid 40 and name H2' )                                     2.40  0.60  0.60
assign (resid 40 and name H4' )(resid 40 and name H1' )                                     2.90  1.10  1.10
assign (resid 40 and name H4' )(resid 40 and name H2' )                                     2.90  1.10  1.10
assign (resid 40 and name H5' or resid 40 and name H5" )(resid 39 and name H2' )           3.90  2.10  1.10
assign (resid 40 and name H5' or resid 40 and name H5" )(resid 40 and name H3' )           3.90  2.10  1.10
assign (resid 40 and name H3  )(resid 21 and name H1 )                                      3.40  1.60  1.60
assign (resid 40 and name H3  )(resid 22 and name H2 )                                      2.90  1.10  1.10
assign (resid 40 and name H3  )(resid 26 and name H1 )                                      3.40  1.60  1.60
assign (resid 40 and name H3  )(resid 39 and name H5 )                                      2.90  1.10  2.10
assign (resid 40 and name H3  )(resid 39 and name H42 )                                     3.90  2.10  2.10
assign (resid 40 and name H6  )(resid 39 and name H1' )                                     3.40  1.60  1.60
assign (resid 40 and name H6  )(resid 39 and name H2' )                                     2.40  0.60  0.60
assign (resid 40 and name H6  )(resid 39 and name H3' )                                     2.90  1.10  1.10
assign (resid 40 and name H6  )(resid 39 and name H6 )                                      3.40  1.60  2.60
```

```
assign (resid 40 and name H6  )(resid 40 and name H1' )                              3.40  1.60  1.60
assign (resid 40 and name H6  )(resid 40 and name H2' )                              3.40  1.60  1.60
assign (resid 40 and name H6  )(resid 40 and name H3' )                              2.90  1.10  1.10
assign (resid 40 and name H6  )(resid 40 and name H4' )                              3.40  1.60  1.60
assign (resid 40 and name H5  )(resid 39 and name H2' )                              3.40  1.60  2.60
assign (resid 40 and name H5  )(resid 39 and name H3' )                              3.15  1.35  1.85
assign (resid 40 and name H5  )(resid 39 and name H5  )                              3.65  1.85  2.35
assign (resid 40 and name H5  )(resid 39 and name H6  )                              3.65  1.85  2.35
assign (resid 40 and name H5  )(resid 40 and name H1' )                              3.65  1.85  2.35
assign (resid 40 and name H5  )(resid 40 and name H3' )                              3.65  1.85  2.35

assign (resid 41 and name H1' )(resid 22 and name H2  )                              2.90  1.10  1.10
assign (resid 41 and name H1' )(resid 40 and name H2' )                              3.40  1.60  1.60
assign (resid 41 and name H2' )(resid 41 and name H1' )                              2.40  0.60  0.60
assign (resid 41 and name H3' )(resid 40 and name H2' )                              3.40  1.60  1.60
assign (resid 41 and name H3' )(resid 41 and name H1' )                              2.90  1.10  1.10
assign (resid 41 and name H3' )(resid 41 and name H2' )                              2.40  0.60  0.60
assign (resid 41 and name H41 )(resid 21 and name H1  )                              3.90  2.10  2.10
assign (resid 41 and name H42 )(resid 21 and name H1  )                              2.90  1.10  1.10
assign (resid 41 and name H42 )(resid 21 and name H22 )                              3.90  2.10  2.10
assign (resid 41 and name H42 )(resid 40 and name H3  )                              3.90  2.10  2.10
assign (resid 41 and name H41 )(resid 40 and name H5  )                              2.90  1.10  1.10
assign (resid 41 and name H42 )(resid 40 and name H5  )                              3.90  2.10  2.10
assign (resid 41 and name H6  )(resid 40 and name H1' )                              3.40  1.60  1.60
assign (resid 41 and name H6  )(resid 40 and name H2' )                              2.40  0.60  0.60
assign (resid 41 and name H6  )(resid 40 and name H3' )                              2.90  1.10  1.10
assign (resid 41 and name H6  )(resid 40 and name H6  )                              3.40  1.60  2.60
assign (resid 41 and name H6  )(resid 41 and name H1' )                              3.40  1.60  1.60
assign (resid 41 and name H6  )(resid 41 and name H2' )                              3.40  1.60  1.60
assign (resid 41 and name H6  )(resid 41 and name H3' )                              2.90  1.10  1.10
assign (resid 41 and name H6  )(resid 41 and name H4' )                              3.40  1.60  1.60
assign (resid 41 and name H6  )(resid 41 and name H5' or resid 41 and name H5" )     3.90  2.10  1.10
assign (resid 41 and name H5  )(resid 21 and name H1  )                              3.65  1.85  2.35
assign (resid 41 and name H5  )(resid 40 and name H3  )                              3.65  1.85  2.35
assign (resid 41 and name H5  )(resid 40 and name H2' )                              3.65  1.85  2.35
assign (resid 41 and name H5  )(resid 40 and name H3' )                              3.65  1.85  2.35
assign (resid 41 and name H5  )(resid 40 and name H6  )                              3.65  1.85  2.35
assign (resid 41 and name H5  )(resid 40 and name H3  )                              3.65  1.85  2.35
assign (resid 41 and name H5  )(resid 40 and name H5  )                              3.65  1.85  2.35

assign (resid 42 and name H1' )(resid 21 and name H22 )                              3.90  2.10  2.10
assign (resid 42 and name H1' )(resid 21 and name H21 )                              3.90  2.10  2.10
assign (resid 42 and name H2' )(resid 42 and name H1' )                              2.40  0.60  0.60
assign (resid 42 and name H3' )(resid 41 and name H2' )                              3.40  1.60  1.60
assign (resid 42 and name H3' )(resid 42 and name H1' )                              3.40  1.60  1.60
assign (resid 42 and name H3' )(resid 42 and name H2' )                              2.40  0.60  0.60
assign (resid 42 and name H4' )(resid 42 and name H1' )                              2.90  1.10  1.10
assign (resid 42 and name H4' )(resid 42 and name H3' )                              2.40  0.60  0.60
assign (resid 42 and name H5' or resid 42 and name H5" )(resid 41 and name H2' )     3.90  2.10  1.10
assign (resid 42 and name H5' or resid 42 and name H5" )(resid 42 and name H3' )     3.40  1.60  1.60
assign (resid 42 and name H3  )(resid 19 and name H42 )                              3.90  2.10  2.10
assign (resid 42 and name H3  )(resid 20 and name H2  )                              2.90  1.10  1.10
assign (resid 42 and name H3  )(resid 21 and name H1  )                              3.40  1.60  1.60
assign (resid 42 and name H3  )(resid 41 and name H42 )                              3.90  2.10  2.10
```

197

```
assign (resid 42 and name H3  )(resid 42 and name H1' )          3.40  1.60  1.60
assign (resid 42 and name H6  )(resid 41 and name H1' )          3.40  1.60  1.60
assign (resid 42 and name H6  )(resid 41 and name H2' )          2.90  1.10  1.10
assign (resid 42 and name H6  )(resid 42 and name H1' )          3.40  1.60  1.60
assign (resid 42 and name H6  )(resid 42 and name H2' )          2.90  1.10  2.10
assign (resid 42 and name H6  )(resid 42 and name H3' )          2.90  1.10  1.10
assign (resid 42 and name H5  )(resid 21 and name H1  )          3.65  1.85  2.35
assign (resid 42 and name H5  )(resid 41 and name H2' )          3.65  1.85  2.35
assign (resid 42 and name H5  )(resid 41 and name H3' )          3.65  1.85  1.35
assign (resid 42 and name H5  )(resid 41 and name H41 )          3.90  2.10  2.10
assign (resid 42 and name H5  )(resid 41 and name H42 )          3.90  2.10  2.10
assign (resid 42 and name H5  )(resid 41 and name H5  )          3.65  1.85  2.35
assign (resid 42 and name H5  )(resid 42 and name H3' )          3.65  1.85  2.35
assign (resid 42 and name H5  )(resid 42 and name H3  )          3.90  2.10  2.10

assign (resid 43 and name H1' )(resid 20 and name H2  )          2.90  1.10  1.10
assign (resid 43 and name H1' )(resid 42 and name H2' )          3.40  1.60  1.60
assign (resid 43 and name H2' )(resid 43 and name H1' )          2.40  0.60  0.60
assign (resid 43 and name H3' )(resid 43 and name H1' )          2.90  1.10  1.10
assign (resid 43 and name H3' )(resid 43 and name H2' )          2.40  0.60  0.60
assign (resid 43 and name H4' )(resid 43 and name H1' )          2.90  1.10  1.10
assign (resid 43 and name H4' )(resid 43 and name H2' )          2.90  1.10  1.10
assign (resid 43 and name H4' )(resid 43 and name H3' )          2.40  0.60  0.60
assign (resid 43 and name H5' or resid 43 and name H5" )(resid 43 and name H3' )   3.90  2.10  1.10
assign (resid 43 and name H5' or resid 43 and name H5" )(resid 43 and name H4' )   2.90  1.10  0.10
assign (resid 43 and name H8  )(resid 42 and name H1' )          3.40  1.60  1.60
assign (resid 43 and name H8  )(resid 42 and name H2' )          2.40  0.60  0.60
assign (resid 43 and name H8  )(resid 42 and name H3' )          2.90  1.10  1.10
assign (resid 43 and name H8  )(resid 42 and name H6  )          3.40  1.60  2.60
assign (resid 43 and name H8  )(resid 43 and name H1' )          2.90  1.10  1.10
assign (resid 43 and name H8  )(resid 43 and name H2' )          3.40  1.60  1.60
assign (resid 43 and name H8  )(resid 43 and name H3' )          2.90  1.10  1.10
assign (resid 43 and name H8  )(resid 43 and name H4' )          3.40  1.60  1.60
assign (resid 43 and name H8  )(resid 43 and name H5' or resid 43 and name H5" )   3.90  2.10  1.10
assign (resid 43 and name H1  )(resid 18 and name H41 )          3.90  2.10  2.10
assign (resid 43 and name H1  )(resid 18 and name H42 )          2.90  1.10  1.10
assign (resid 43 and name H1  )(resid 19 and name H42 )          3.90  2.10  2.10
assign (resid 43 and name H1  )(resid 44 and name H1' )          3.90  2.10  2.10
assign (resid 43 and name H1  )(resid 19 and name H5  )          3.90  2.10  2.10
assign (resid 43 and name H1  )(resid 20 and name H2  )          3.90  2.10  2.10
assign (resid 43 and name H1  )(resid 42 and name H3  )          3.40  1.60  1.60
assign (resid 43 and name H22 )(resid 20 and name H1' )          3.90  2.10  2.10
assign (resid 43 and name H21 )(resid 20 and name H1' )          3.90  2.10  2.10
assign (resid 43 and name H22 )(resid 44 and name H1' )          3.90  2.10  2.10
assign (resid 43 and name H21 )(resid 44 and name H1' )          3.90  2.10  2.10

assign (resid 44 and name H1' )(resid 43 and name H2' )          2.90  1.10  1.10
assign (resid 44 and name H2' )(resid 44 and name H1' )          2.40  0.60  0.60
assign (resid 44 and name H3' )(resid 44 and name H1' )          2.90  1.10  1.10
assign (resid 44 and name H4' )(resid 44 and name H1' )          3.40  1.60  1.60
assign (resid 44 and name H4' )(resid 44 and name H3' )          2.40  0.60  0.60
assign (resid 44 and name H5' or resid 44 and name H5" )(resid 43 and name H2' )   3.90  2.10  1.10
assign (resid 44 and name H5' or resid 44 and name H5" )(resid 43 and name H3' )   3.90  2.10  1.10
assign (resid 44 and name H5' or resid 44 and name H5" )(resid 44 and name H1' )   3.90  2.10  1.10
```

198

```
assign (resid 44 and name H5' or resid 44 and name H5" )(resid 44 and name H3' )        3.40  1.60  0.60
assign (resid 44 and name H5' or resid 44 and name H5" )(resid 44 and name H4' )        2.90  1.10  0.10
assign (resid 44 and name H8  )(resid 43 and name H1' )                                  3.40  1.60  1.60
assign (resid 44 and name H8  )(resid 43 and name H2' )                                  2.40  0.60  0.60
assign (resid 44 and name H8  )(resid 43 and name H3' )                                  2.90  1.10  1.10
assign (resid 44 and name H8  )(resid 43 and name H8  )                                  3.40  1.60  2.60
assign (resid 44 and name H8  )(resid 44 and name H1' )                                  2.90  1.10  1.10
assign (resid 44 and name H8  )(resid 44 and name H2' )                                  3.40  1.60  1.60
assign (resid 44 and name H8  )(resid 44 and name H3' )                                  2.90  1.10  1.10
assign (resid 44 and name H8  )(resid 44 and name H4' )                                  3.40  1.60  1.60
assign (resid 44 and name H8  )(resid 44 and name H5' or resid 44 and name H5" )         3.90  2.10  1.10
assign (resid 44 and name H1  )(resid 17 and name H1  )                                  2.90  1.10  1.10
assign (resid 44 and name H1  )(resid 18 and name H41 )                                  3.40  1.60  2.60
assign (resid 44 and name H1  )(resid 18 and name H42 )                                  2.90  1.10  1.10
assign (resid 44 and name H1  )(resid 18 and name H5  )                                  3.90  2.10  2.10
assign (resid 44 and name H1  )(resid 19 and name H42 )                                  3.90  2.10  2.10
assign (resid 44 and name H1  )(resid 19 and name H5  )                                  3.90  2.10  2.10
assign (resid 44 and name H1  )(resid 43 and name H1  )                                  3.40  1.60  1.60

assign (resid 45 and name H1' )(resid 44 and name H2' )                                  3.40  1.60  1.60
assign (resid 45 and name H2' )(resid 45 and name H1' )                                  2.40  0.60  0.60
assign (resid 45 and name H3' )(resid 45 and name H1' )                                  2.90  1.10  1.10
assign (resid 45 and name H3' )(resid 45 and name H2' )                                  2.40  0.60  0.60
assign (resid 45 and name H4' )(resid 45 and name H1' )                                  2.90  1.10  1.10
assign (resid 45 and name H4' )(resid 45 and name H2' )                                  2.90  1.10  1.10
assign (resid 45 and name H4' )(resid 45 and name H3' )                                  2.40  0.60  0.60
assign (resid 45 and name H5' or resid 45 and name H5" )(resid 45 and name H3' )         3.90  2.10  1.10
assign (resid 45 and name H5' or resid 45 and name H5" )(resid 45 and name H4' )         3.40  1.60  0.60
assign (resid 45 and name H41 )(resid 17 and name H1  )                                  3.90  2.10  2.10
assign (resid 45 and name H42 )(resid 17 and name H1  )                                  2.90  1.10  1.10
assign (resid 45 and name H42 )(resid 18 and name H42 )                                  3.90  2.10  2.10
assign (resid 45 and name H42 )(resid 18 and name H5  )                                  3.90  2.10  2.10
assign (resid 45 and name H42 )(resid 44 and name H1  )                                  3.90  2.10  2.10
assign (resid 45 and name H6  )(resid 44 and name H1' )                                  3.40  1.60  1.60
assign (resid 45 and name H6  )(resid 44 and name H2' )                                  2.40  0.60  0.60
assign (resid 45 and name H6  )(resid 44 and name H3' )                                  2.90  1.10  1.10
assign (resid 45 and name H6  )(resid 44 and name H8  )                                  3.40  1.60  2.60
assign (resid 45 and name H6  )(resid 45 and name H1' )                                  2.90  1.10  1.10
assign (resid 45 and name H6  )(resid 45 and name H2' )                                  3.40  1.60  1.60
assign (resid 45 and name H6  )(resid 45 and name H3' )                                  2.90  1.10  1.10
assign (resid 45 and name H6  )(resid 45 and name H4' )                                  3.40  1.60  1.60
assign (resid 45 and name H6  )(resid 45 and name H5' or resid 45 and name H5" )         3.90  2.10  1.10
assign (resid 45 and name H5  )(resid 17 and name H1  )                                  3.90  2.10  2.10
assign (resid 45 and name H5  )(resid 18 and name H42 )                                  3.90  2.10  2.10
assign (resid 45 and name H5  )(resid 44 and name H2' )                                  3.65  1.85  2.35
assign (resid 45 and name H5  )(resid 44 and name H3' )                                  3.65  1.85  2.35
assign (resid 45 and name H5  )(resid 44 and name H8  )                                  3.65  1.85  2.35
assign (resid 45 and name H5  )(resid 44 and name H1  )                                  3.90  2.10  2.10
assign (resid 45 and name H5  )(resid 45 and name H3' )                                  3.65  1.85  2.35

assign (resid 46 and name H1' )(resid 45 and name H2' )                                  3.40  1.60  1.60
assign (resid 46 and name H2' )(resid 46 and name H1' )                                  2.40  0.60  0.60
assign (resid 46 and name H3' )(resid 45 and name H2' )                                  3.40  1.60  1.60
assign (resid 46 and name H3' )(resid 46 and name H1' )                                  2.90  1.10  1.10
```

```
assign (resid 46 and name H3' )(resid 46 and name H2' )                          2.40  0.60  0.60
assign (resid 46 and name H4' )(resid 45 and name H2' )                          3.40  1.60  1.60
assign (resid 46 and name H4' )(resid 46 and name H1' )                          2.90  1.10  1.10
assign (resid 46 and name H4' )(resid 46 and name H2' )                          2.90  1.10  1.10
assign (resid 46 and name H4' )(resid 46 and name H3' )                          2.40  0.60  0.60
assign (resid 46 and name H5' or resid 46 and name H5" )(resid 45 and name H2' ) 3.90  2.10  1.10
assign (resid 46 and name H5' or resid 46 and name H5" )(resid 45 and name H3' ) 3.90  2.10  1.10
assign (resid 46 and name H5' or resid 46 and name H5" )(resid 46 and name H3' ) 3.90  2.10  1.10
assign (resid 46 and name H5' or resid 46 and name H5" )(resid 46 and name H4' ) 3.40  1.60  0.60
assign (resid 46 and name H6  )(resid 45 and name H2' )                          2.40  0.60  0.60
assign (resid 46 and name H6  )(resid 45 and name H3' )                          2.90  1.10  1.10
assign (resid 46 and name H6  )(resid 46 and name H1' )                          3.40  1.60  1.60
assign (resid 46 and name H6  )(resid 46 and name H2' )                          3.40  1.60  1.60
assign (resid 46 and name H6  )(resid 46 and name H3' )                          2.90  1.10  1.10
assign (resid 46 and name H6  )(resid 46 and name H4' )                          2.90  1.10  1.10
assign (resid 46 and name H6  )(resid 46 and name H5' or resid 46 and name H5" ) 3.90  2.10  1.10
assign (resid 46 and name H5  )(resid 17 and name H1  )                          3.90  2.10  2.10
assign (resid 46 and name H5  )(resid 45 and name H2' )                          3.65  1.85  2.35
assign (resid 46 and name H5  )(resid 45 and name H3' )                          3.65  1.85  2.35
assign (resid 46 and name H5  )(resid 45 and name H5  )                          3.65  1.85  2.35
assign (resid 46 and name H42 )(resid 16 and name H1  )                          3.65  2.35  2.35

! INTERMOLECULAR ARGININAMIDE NOES
assign (resid 47 and name HC1 )(resid 38 and name H5  )                          3.40  1.60  2.60
assign (resid 47 and name HC2 )(resid 38 and name H5  )                          3.40  1.60  2.60
assign (resid 47 and name HB1 or resid 47 and name HB2 )(resid 27 and name H61 or resid 27 and name
H61)                                                                             3.90  2.10  2.10
assign (resid 47 and name HB1 or resid 47 and name HB2 )(resid 39 and name H41 ) 3.90  2.10  2.10
assign (resid 47 and name HB1 or resid 47 and name HB2 )(resid 39 and name H42 ) 3.40  1.60  2.60
assign (resid 47 and name HG1 or resid 47 and name HG2 )(resid 22 and name H8  ) 3.40  1.60  1.60
assign (resid 47 and name HG1 or resid 47 and name HG2 )(resid 23 and name H5  ) 2.90  1.10  1.10
assign (resid 47 and name HG1 or resid 47 and name HG2 )(resid 23 and name H6  ) 3.90  2.10  2.10
assign (resid 47 and name HD1 or resid 47 and name HD2 )(resid 22 and name H8  ) 3.40  1.60  1.60
assign (resid 47 and name HD1 or resid 47 and name HD2 )(resid 23 and name H5  ) 2.90  1.10  1.10
assign (resid 47 and name HD1 or resid 47 and name HD2 )(resid 39 and name H41 ) 3.40  1.60  2.60
assign (resid 47 and name HD1 or resid 47 and name HD2 )(resid 39 and name H42 ) 3.40  1.60  2.60
assign (resid 47 and name HE  )(resid 22 and name H2' )                          3.90  2.10  2.10
assign (resid 47 and name HE  )(resid 22 and name H3' )                          3.90  2.10  1.10
assign (resid 47 and name HE  )(resid 22 and name H8  )                          3.90  2.10  2.10
assign (resid 47 and name HE  )(resid 23 and name H6  )                          3.90  2.10  2.10
assign (resid 47 and name HE  )(resid 23 and name H5  )                          3.90  2.10  2.10
assign (resid 47 and name HH11 or resid 47 and name HH12 or resid 47 and name HH21 or resid 47 and
name HH22)(resid 22 and name H8)                                                 2.50  0.70  2.50
assign (resid 47 and name HH11 or resid 47 and name HH12 or resid 47 and name HH21 or resid 47 and
name HH22)(resid 26 and name H8)                                                 2.50  0.70  2.50
set echo = on end
set message = on end
```

Torsion Restraints:

These are the sugar pucker restraints used for the structure determination.

```
restraints
        dihedral
```

reset
!c2'endo puckers
    assign (resid 23 and name C1')(resid 23 and name C2')(resid 23 and name C3')
        (resid 23 and name C4') 1.0 -40.0 5.0 2
    assign (resid 23 and name C3')(resid 23 and name C4')(resid 23 and name O4')
        (resid 23 and name C1') 1.0 0.0 5.0 2
    assign (resid 25 and name C1')(resid 25 and name C2')(resid 25 and name C3')
        (resid 25 and name C4') 1.0 -40.0 5.0 2
    assign (resid 25 and name C3')(resid 25 and name C4')(resid 25 and name O4')
        (resid 25 and name C1') 1.0 0.0 5.0 2
    assign (resid 30 and name C1')(resid 30 and name C2')(resid 30 and name C3')
        (resid 30 and name C4') 1.0 -40.0 5.0 2
    assign (resid 30 and name C3')(resid 30 and name C4')(resid 30 and name O4')
        (resid 30 and name C1') 1.0 0.0 5.0 2
    assign (resid 31 and name C1')(resid 31 and name C2')(resid 31 and name C3')
        (resid 31 and name C4') 1.0 -40.0 5.0 2
    assign (resid 31 and name C3')(resid 31 and name C4')(resid 31 and name O4')
        (resid 31 and name C1') 1.0 0.0 5.0 2
    assign (resid 32 and name C1')(resid 32 and name C2')(resid 32 and name C3')
        (resid 32 and name C4') 1.0 -40.0 5.0 2
    assign (resid 33 and name C3')(resid 33 and name C4')(resid 33 and name O4')
        (resid 33 and name C1') 1.0 0.0 5.0 2
    assign (resid 33 and name C1')(resid 33 and name C2')(resid 33 and name C3')
        (resid 33 and name C4') 1.0 -40.0 5.0 2
    assign (resid 34 and name C1')(resid 34 and name C2')(resid 34 and name C3')
        (resid 34 and name C4') 1.0 -40.0 5.0 2
    assign (resid 34 and name C3')(resid 34 and name C4')(resid 34 and name O4')
        (resid 34 and name C1') 1.0 0.0 5.0 2
!A- helix deltas and nu2 for C3' endo pucker
    assign (resid 16 and name C5')(resid 16 and name C4')(resid 16 and name C3')
        (resid 16 and name O3')   1.0 82.0 5.0 2
    assign (resid 16 and name C1')(resid 16 and name C2')(resid 16 and name C3')
        (resid 16 and name C4')   1.0 38.0 5.0 2
    assign (resid 17 and name C5')(resid 17 and name C4')(resid 17 and name C3')
        (resid 17 and name O3')   1.0 82.0 5.0 2
    assign (resid 17 and name C1')(resid 17 and name C2')(resid 17 and name C3')
        (resid 17 and name C4')   1.0 38.0 5.0 2
    assign (resid 18 and name C5')(resid 18 and name C4')(resid 18 and name C3')
        (resid 18 and name O3')   1.0 82.0 5.0 2
    assign (resid 18 and name C1')(resid 18 and name C2')(resid 18 and name C3')
        (resid 18 and name C4')   1.0 38.0 5.0 2
    assign (resid 19 and name C5')(resid 19 and name C5')(resid 19 and name C3')
        (resid 19 and name O3')   1.0 82.0 5.0 2
    assign (resid 19 and name C1')(resid 19 and name C2')(resid 19 and name C3')
        (resid 19 and name C4')   1.0 38.0 5.0 2
    assign (resid 20 and name C5')(resid 20 and name C4')(resid 20 and name C3')
        (resid 20 and name O3')   1.0 82.0 5.0 2
    assign (resid 20 and name C1')(resid 20 and name C2')(resid 20 and name C3')
        (resid 20 and name C4')   1.0 38.0 5.0 2
    assign (resid 21 and name C5')(resid 21 and name C4')(resid 21 and name C3')
        (resid 21 and name O3')   1.0 82.0 5.0 2
    assign (resid 21 and name C1')(resid 21 and name C2')(resid 21 and name C3')
        (resid 21 and name C4')   1.0 38.0 5.0 2
    assign (resid 22 and name C5')(resid 22 and name C4')(resid 22 and name C3')

```
                    (resid 22 and name O3')    1.0 82.0  5.0 2
assign  (resid 22 and name C1')(resid 22 and name C2')(resid 22 and name C3')
        (resid 22 and name C4')    1.0 38.0  5.0 2
assign  (resid 26 and name C5')(resid 26 and name C4')(resid 26 and name C3')
        (resid 26 and name O3')    1.0 82.0  5.0 2
assign  (resid 26 and name C1')(resid 26 and name C2')(resid 26 and name C3')
        (resid 26 and name C4')    1.0 38.0  5.0 2
assign  (resid 27 and name C5')(resid 27 and name C4')(resid 27 and name C3')
        (resid 27 and name O3')    1.0 82.0  5.0 2
assign  (resid 27 and name C1')(resid 27 and name C2')(resid 27 and name C3')
        (resid 27 and name C4')    1.0 38.0  5.0 2
assign  (resid 28 and name C5')(resid 28 and name C4')(resid 28 and name C3')
        (resid 28 and name O3')    1.0 82.0  5.0 2
assign  (resid 28 and name C1')(resid 28 and name C2')(resid 28 and name C3')
        (resid 28 and name C4')    1.0 38.0  5.0 2
assign  (resid 29 and name C5')(resid 29 and name C4')(resid 29 and name C3')
        (resid 29 and name O3')    1.0 82.0  5.0 2
assign  (resid 29 and name C1')(resid 29 and name C2')(resid 29 and name C3')
        (resid 29 and name C4')    1.0 38.0  5.0 2
assign  (resid 36 and name C5')(resid 36 and name C4')(resid 36 and name C3')
        (resid 36 and name O3')    1.0 82.0  5.0 2
assign  (resid 36 and name C1')(resid 36 and name C2')(resid 36 and name C3')
        (resid 36 and name C4')    1.0 38.0  5.0 2
assign  (resid 37 and name C5')(resid 37 and name C4')(resid 37 and name C3')
        (resid 37 and name O3')    1.0 82.0  5.0 2
assign  (resid 37 and name C1')(resid 37 and name C2')(resid 37 and name C3')
        (resid 37 and name C4')    1.0 38.0  5.0 2
assign  (resid 38 and name C5')(resid 38 and name C4')(resid 38 and name C3')
        (resid 38 and name O3')    1.0 82.0  5.0 2
assign  (resid 39 and name C5')(resid 39 and name C4')(resid 39 and name C3')
        (resid 39 and name O3')    1.0 82.0  5.0 2
assign  (resid 39 and name C1')(resid 39 and name C2')(resid 39 and name C3')
        (resid 39 and name C4')    1.0 38.0  5.0 2
assign  (resid 40 and name C5')(resid 40 and name C4')(resid 40 and name C3')
        (resid 40 and name O3')    1.0 82.0  5.0 2
assign  (resid 40 and name C1')(resid 40 and name C2')(resid 40 and name C3')
        (resid 40 and name C4')    1.0 38.0  5.0 2
assign  (resid 41 and name C5')(resid 41 and name C4')(resid 41 and name C3')
        (resid 41 and name O3')    1.0 82.0  5.0 2
assign  (resid 41 and name C1')(resid 41 and name C2')(resid 41 and name C3')
        (resid 41 and name C4')    1.0 38.0  5.0 2
assign  (resid 42 and name C5')(resid 42 and name C4')(resid 42 and name C3')
        (resid 42 and name O3')    1.0 82.0  5.0 2
assign  (resid 42 and name C1')(resid 42 and name C2')(resid 42 and name C3')
        (resid 42 and name C4')    1.0 38.0  5.0 2
assign  (resid 43 and name C5')(resid 43 and name C4')(resid 43 and name C3')
        (resid 43 and name O3')    1.0 82.0  5.0 2
assign  (resid 43 and name C1')(resid 43 and name C2')(resid 43 and name C3')
        (resid 43 and name C4')    1.0 38.0  5.0 2
assign  (resid 44 and name C5')(resid 44 and name C4')(resid 44 and name C3')
        (resid 44 and name O3')    1.0 82.0  5.0 2
assign  (resid 44 and name C1')(resid 44 and name C2')(resid 44 and name C3')
        (resid 44 and name C4')    1.0 38.0  5.0 2
assign  (resid 45 and name C5')(resid 45 and name C4')(resid 45 and name C3')
```

                    (resid 45 and name O3')   1.0 82.0 5.0 2
        assign (resid 45 and name C1')(resid 45 and name C2')(resid 45 and name C3')
                    (resid 45 and name C4')   1.0 38.0 5.0 2
        assign (resid 46 and name C5')(resid 46 and name C4')(resid 46 and name C3')
                    (resid 46 and name O3')   1.0 82.0 5.0 2
        assign (resid 46 and name C1')(resid 46 and name C2')(resid 46 and name C3')
                    (resid 46 and name C4')   1.0 38.0 5.0 2
end

Hydrogen Bond Restraints:
These are all the distance restraints used to define the hydrogen bonding interactions for the structure determination.
set echo = off end
set message = off end
assign (resid 16 and name O6)(resid 46 and name H42)   1.91 0.20 0.20
assign (resid 16 and name O6)(resid 46 and name N4)   2.91 0.20 0.20
assign (resid 16 and name H1)(resid 46 and name N3)   1.95 0.20 0.20
assign (resid 16 and name N1)(resid 46 and name N3)   2.95 0.20 0.20
assign (resid 16 and name H22)(resid 46 and name O2)   1.86 0.20 0.20
assign (resid 16 and name N2)(resid 46 and name O2)   2.86 0.20 0.20
assign (resid 17 and name O6)(resid 45 and name H42)   1.91 0.20 0.20
assign (resid 17 and name O6)(resid 45 and name N4)   2.91 0.20 0.20
assign (resid 17 and name H1)(resid 45 and name N3)   1.95 0.20 0.20
assign (resid 17 and name N1)(resid 45 and name N3)   2.95 0.20 0.20
assign (resid 17 and name H22)(resid 45 and name O2)   1.86 0.20 0.20
assign (resid 17 and name N2)(resid 45 and name O2)   2.86 0.20 0.20
assign (resid 18 and name H42)(resid 44 and name O6)   1.91 0.20 0.20
assign (resid 18 and name N4)(resid 44 and name O6)   2.91 0.20 0.20
assign (resid 18 and name N3)(resid 44 and name H1)   1.95 0.20 0.20
assign (resid 18 and name N3)(resid 44 and name N1)   2.95 0.20 0.20
assign (resid 18 and name O2)(resid 44 and name H22)   1.86 0.20 0.20
assign (resid 18 and name O2)(resid 44 and name N2)   2.86 0.20 0.20
assign (resid 19 and name H42)(resid 43 and name O6)   1.91 0.20 0.20
assign (resid 19 and name N4)(resid 43 and name O6)   2.91 0.20 0.20
assign (resid 19 and name N3)(resid 43 and name H1)   1.95 0.20 0.20
assign (resid 19 and name N3)(resid 43 and name N1)   2.95 0.20 0.20
assign (resid 19 and name O2)(resid 43 and name H22)   1.86 0.20 0.20
assign (resid 19 and name O2)(resid 43 and name N2)   2.86 0.20 0.20
assign (resid 20 and name N1)(resid 42 and name H3)   1.70 0.20 0.20
assign (resid 20 and name N1)(resid 42 and name N3)   2.70 0.20 0.20
assign (resid 20 and name H62)(resid 42 and name O4)   1.95 0.20 0.20
assign (resid 20 and name N6)(resid 42 and name O4)   2.95 0.20 0.20
assign (resid 21 and name O6)(resid 41 and name H42)   1.91 0.20 0.20
assign (resid 21 and name O6)(resid 41 and name N4)   2.91 0.20 0.20
assign (resid 21 and name H1)(resid 41 and name N3)   1.95 0.20 0.20
assign (resid 21 and name N1)(resid 41 and name N3)   2.95 0.20 0.20
assign (resid 21 and name H22)(resid 41 and name O2)   1.86 0.20 0.20
assign (resid 21 and name N2)(resid 41 and name O2)   2.86 0.20 0.20
assign (resid 22 and name N1)(resid 40 and name H3)   1.70 0.20 0.20
assign (resid 22 and name N1)(resid 40 and name N3)   2.70 0.20 0.20
assign (resid 22 and name H62)(resid 40 and name O4)   1.95 0.20 0.20
assign (resid 22 and name N6)(resid 40 and name O4)   2.95 0.20 0.20
assign (resid 26 and name O6)(resid 39 and name H42)   1.91 0.20 0.20

assign (resid 26 and name O6)(resid 39 and name N4)	2.91	0.20 0.20
assign (resid 26 and name H1)(resid 39 and name N3)	1.95	0.20 0.20
assign (resid 26 and name N1)(resid 39 and name N3)	2.95	0.20 0.20
assign (resid 26 and name H22)(resid 39 and name O2)	1.86	0.20 0.20
assign (resid 26 and name N2)(resid 39 and name O2)	2.86	0.20 0.20
assign (resid 27 and name N1)(resid 38 and name H3)	1.70	0.20 0.20
assign (resid 27 and name N1)(resid 38 and name N3)	2.70	0.20 0.20
assign (resid 27 and name H62)(resid 38 and name O4)	1.95	0.20 0.20
assign (resid 27 and name N6)(resid 38 and name O4)	2.95	0.20 0.20
assign (resid 28 and name O6)(resid 37 and name H42)	1.91	0.20 0.20
assign (resid 28 and name O6)(resid 37 and name N4)	2.91	0.20 0.20
assign (resid 28 and name H1)(resid 37 and name N3)	1.95	0.20 0.20
assign (resid 28 and name N1)(resid 37 and name N3)	2.95	0.20 0.20
assign (resid 28 and name H22)(resid 37 and name O2)	1.86	0.20 0.20
assign (resid 28 and name N2)(resid 37 and name O2)	2.86	0.20 0.20
assign (resid 29 and name H42)(resid 36 and name O6)	1.91	0.20 0.20
assign (resid 29 and name N4)(resid 36 and name O6)	2.91	0.20 0.20
assign (resid 29 and name N3)(resid 36 and name H1)	1.95	0.20 0.20
assign (resid 29 and name N3)(resid 36 and name N1)	2.95	0.20 0.20
assign (resid 29 and name O2)(resid 36 and name H22)	1.86	0.20 0.20
assign (resid 29 and name O2)(resid 36 and name N2)	2.86	0.20 0.20

set echo = on end
set message = on end

Argininamide modeling restraints:

These are the hydrogen bonding and planarity restraints used to restrain the argininamide binding and base triple interactions.

!Argininamide - TAR H-bonding distances restraints

assign (resid 47 and name HH11)(resid 26 and name O6) 1.95 0.25 0.25
assign (resid 47 and name NH1)(resid 26 and name O6) 2.95 0.25 0.25
assign (resid 47 and name HH12)(resid 26 and name N7) 1.95 0.25 0.25
assign (resid 47 and name NH1)(resid 26 and name N7) 2.95 0.25 0.25

!Base triple H-bonding distance restraints

assign (resid 23 and name O4)(resid 27 and name H61) 1.95 0.25 0.25
assign (resid 23 and name O4)(resid 27 and name N6) 2.95 0.25 0.25
assign (resid 23 and name H3)(resid 27 and name N7) 1.95 0.25 0.25
assign (resid 23 and name N3)(resid 27 and name N7) 2.95 0.25 0.25

These are the planar restraints restraining the A27 base and the U23 base to be co-planar. These restraints are selected with the PLAN flag in the XPLOR scripts.

```
restraints
     planar
 group
     selection=(
         (not( name P or name O1P or name O2P or name O5' or
             name C5' or name H5' or name H5" or
             name C4' or name H4' or name C3' or name H3' or
             name C2' or name H2' or name O2' or name HO2' or
             name H1' or name C1' or name O4' or name O3') and
             residue 23 )
 or          (not( name P or name O1P or name O2P or name O5' or
             name C5' or name H5' or name H5" or
```

204

<pre>
                    name C4' or name H4' or name C3' or name H3' or
                    name C2' or name H2' or name O2' or name HO2' or
                    name H1' or name C1' or name O4' or name O3') and
                    residue 27)
)
weight=250.0
        end
end
</pre>

These are the planar restraints that constrain the G26 base and the argininamide guanidium group to be co-planar.

<pre>
restraints
        planar
  group
        selection=(
                (not(name P or name O1P or name O2P or name O5' or
                    name C5' or name H5' or name H5" or
                    name C4' or name H4' or name C3' or name H3' or
                    name C2' or name H2' or name O2' or name HO2' or
                    name H1' or name C1' or name O4' or name O3') and
                    residue 26)
or   (not(name N or name HN1 or name HN2 or name HN3 or
                    name CA or name HA or name CB or name HB1 or
                    name HB2 or name CG or name HG1 or name HG2 or
                    name CD or name HD1 or name HD2 or name NE or
                    name HE or name HC1 or name HC2 or name NC or
                    name O or name C) and residue 47)
)
weight=250.0
        end
end
</pre>

A-form torsion restraints:

These are the backbone torsion restraints which were used in various combinations as discussed in Chapter 3.

<pre>
restraints
        dihedral
        reset
!alfas
        assign (resid 16 and name O3')(resid 17 and name P)(resid 17 and name O5')
                (resid 17 and name C5') 1.0 -68.0 60.0 2
        assign (resid 17 and name O3')(resid 18 and name P(resid 18 and name O5')
                (resid 18 and name C5') 1.0 -68.0 60.0 2
        assign (resid 18 and name O3')(resid 19 and name P)(resid 19 and name O5')
                (resid 19 and name C5') 1.0 -68.0 60.0 2
        assign (resid 19 and name O3')(resid 20 and name P)(resid 20 and name O5')
                (resid 20 and name C5') 1.0 -68.0 60.0 2
        assign (resid 20 and name O3')(resid 21 and name P)(resid 21 and name O5')
                (resid 21 and name C5') 1.0 -68.0 60.0 2
!   assign (resid 26 and name O3')(resid 27 and name P)(resid 27 and name O5')
!   (resid 27 and name C5') 1.0 -68.0 60.0 2
        assign (resid 27 and name O3')(resid 28 and name P)(resid 28 and name O5')
                (resid 28 and name C5') 1.0 -68.0 60.0 2
</pre>

205

assign (resid 28 and name O3')(resid 29 and name P)(resid 29 and name O5')
     (resid 29 and name C5') 1.0 -68.0 60.0 2
assign (resid 36 and name O3')(resid 37 and name P)(resid 37 and name O5')
     (resid 37 and name C5') 1.0 -68.0 60.0 2
assign (resid 37 and name O3')(resid 38 and name P)(resid 38 and name O5')
     (resid 38 and name C5') 1.0 -68.0 60.0 2
assign (resid 38 and name O3')(resid 39 and name P)(resid 39 and name O5')
     (resid 39 and name C5') 1.0 -68.0 60.0 2
assign (resid 40 and name O3')(resid 41 and name P)(resid 41 and name O5')
     (resid 41 and name C5') 1.0 -68.0 60.0 2
assign (resid 41 and name O3')(resid 42 and name P)(resid 42 and name O5')
     (resid 42 and name C5') 1.0 -68.0 60.0 2
assign (resid 42 and name O3')(resid 43 and name P)(resid 43 and name O5')
     (resid 43 and name C5') 1.0 -68.0 60.0 2
assign (resid 43 and name O3')(resid 44 and name P)(resid 44 and name O5')
     (resid 44 and name C5') 1.0 -68.0 60.0 2
assign (resid 44 and name O3')(resid 45 and name P)(resid 45 and name O5')
     (resid 45 and name C5') 1.0 -68.0 60.0 2

!beta

assign (resid 17 and name P)(resid 17 and name O5')(resid 17 and name C5')
     (resid 17 and name C4') 1.0 178.0 60.0 2
assign (resid 18 and name P)(resid 18 and name O5')(resid 18 and name C5')
     (resid 18 and name C4') 1.0 178.0 60.0 2
assign (resid 19 and name P)(resid 19 and name O5')(resid 19 and name C5')
     (resid 19 and name C4') 1.0 178.0 60.0 2
assign (resid 20 and name P)(resid 20 and name O5')(resid 20 and name C5')
     (resid 20 and name C4') 1.0 178.0 60.0 2
assign (resid 21 and name P)(resid 21 and name O5')(resid 21 and name C5')
     (resid 21 and name C4') 1.0 178.0 60.0 2
assign (resid 28 and name P)(resid 28 and name O5')(resid 28 and name C5')
     (resid 28 and name C4') 1.0 178.0 60.0 2
assign (resid 29 and name P)(resid 29 and name O5')(resid 29 and name C5')
     (resid 29 and name C4') 1.0 178.0 60.0 2
assign (resid 37 and name P)(resid 37 and name O5')(resid 37 and name C5')
     (resid 37 and name C4') 1.0 178.0 60.0 2
assign (resid 38 and name P)(resid 38 and name O5')(resid 38 and name C5')
     (resid 38 and name C4') 1.0 178.0 60.0 2
assign (resid 39 and name P)(resid 39 and name O5')(resid 39 and name C5')
     (resid 39 and name C4') 1.0 178.0 60.0 2
assign (resid 41 and name P)(resid 41 and name O5')(resid 41 and name C5')
     (resid 41 and name C4') 1.0 178.0 60.0 2
assign (resid 42 and name P)(resid 42 and name O5')(resid 42 and name C5')
     (resid 42 and name C4') 1.0 178.0 60.0 2
assign (resid 43 and name P)(resid 43 and name O5')(resid 43 and name C5')
     (resid 43 and name C4') 1.0 178.0 60.0 2
assign (resid 44 and name P)(resid 44 and name O5')(resid 44 and name C5')
     (resid 44 and name C4') 1.0 178.0 60.0 2
assign (resid 45 and name P)(resid 45 and name O5')(resid 45 and name C5')
     (resid 45 and name C4') 1.0 178.0 60.0 2
assign (resid 46 and name P)(resid 46 and name O5')(resid 46 and name C5')
     (resid 46 and name C4') 1.0 178.0 60.0 2

!gammas

assign (resid 17 and name O5')(resid 17 and name C5')(resid 17 and name C4')
     (resid 17 and name C3') 1.0 60.0 30.0 2

assign (resid 18 and name O5')(resid 18 and name C5')(resid 18 and name C4')
    (resid 18 and name C3') 1.0 60.0 30.0 2
assign (resid 19 and name O5')(resid 19 and name C5')(resid 19 and name C4')
    (resid 19 and name C3') 1.0 60.0 30.0 2
assign (resid 20 and name O5')(resid 20 and name C5')(resid 20 and name C4')
    (resid 20 and name C3') 1.0 60.0 30.0 2
assign (resid 21 and name O5')(resid 21 and name C5')(resid 21 and name C4')
    (resid 21 and name C3') 1.0 60.0 30.0 2
assign (resid 28 and name O5')(resid 28 and name C5')(resid 28 and name C4')
    (resid 28 and name C3') 1.0 60.0 30.0 2
assign (resid 29 and name O5')(resid 29 and name C5')(resid 29 and name C4')
    (resid 29 and name C3') 1.0 60.0 30.0 2
assign (resid 37 and name O5')(resid 37 and name C5')(resid 37 and name C4')
    (resid 37 and name C3') 1.0 60.0 30.0 2
assign (resid 38 and name O5')(resid 38 and name C5')(resid 38 and name C4')
    (resid 38 and name C3') 1.0 60.0 30.0 2
assign (resid 39 and name O5')(resid 39 and name C5')(resid 39 and name C4')
    (resid 39 and name C3') 1.0 60.0 30.0 2
assign (resid 41 and name O5')(resid 41 and name C5')(resid 41 and name C4')
    (resid 41 and name C3') 1.0 60.0 30.0 2
assign (resid 42 and name O5')(resid 42 and name C5')(resid 42 and name C4')
    (resid 42 and name C3') 1.0 60.0 30.0 2
assign (resid 43 and name O5')(resid 43 and name C5')(resid 43 and name C4')
  (resid 43 and name C3') 1.0 60.0 30.0 2
assign (resid 44 and name O5')(resid 44 and name C5')(resid 44 and name C4')
  (resid 44 and name C3') 1.0 60.0 30.0 2
assign (resid 45 and name O5')(resid 45 and name C5')(resid 45 and name C4')
    (resid 45 and name C3') 1.0 60.0 30.0 2
!epsilon
assign (resid 16 and name C4')(resid 16 and name C3')(resid 16 and name O3')
    (resid 17 and name P) 1.0 -155.0 30.0 2
assign (resid 17 and name C4')(resid 17 and name C3')(resid 17 and name O3')
    (resid 18 and name P) 1.0 -155.0 30.0 2
assign (resid 18 and name C4')(resid 18 and name C3')(resid 18 and name O3')
    (resid 19 and name P) 1.0 -155.0 30.0 2
assign (resid 19 and name C4')(resid 19 and name C3')(resid 19 and name O3')
    (resid 20 and name P) 1.0 -155.0 30.0 2
assign (resid 20 and name C4')(resid 20 and name C3')(resid 20 and name O3')
    (resid 21 and name P) 1.0 -155.0 30.0 2
assign (resid 27 and name C4')(resid 27 and name C3')(resid 27 and name O3')
    (resid 28 and name P) 1.0 -155.0 30.0 2
assign (resid 28 and name C4')(resid 28 and name C3')(resid 28 and name O3')
    (resid 29 and name P) 1.0 -155.0 30.0 2
assign (resid 36 and name C4')(resid 36 and name C3')(resid 36 and name O3')
    (resid 37 and name P) 1.0 -155.0 30.0 2
assign (resid 37 and name C4')(resid 37 and name C3')(resid 37 and name O3')
    (resid 38 and name P) 1.0 -155.0 30.0 2
assign (resid 38 and name C4')(resid 38 and name C3')(resid 38 and name O3')
    (resid 39 and name P) 1.0 -155.0 30.0 2
assign (resid 40 and name C4')(resid 40 and name C3')(resid 40 and name O3')
    (resid 41 and name P) 1.0 -155.0 30.0 2
assign (resid 41 and name C4')(resid 41 and name C3')(resid 41 and name O3')
    (resid 42 and name P) 1.0 -155.0 30.0 2
assign (resid 42 and name C4')(resid 42 and name C3')(resid 42 and name O3')

```
                    (resid 43 and name  P ) 1.0 -155.0 30.0 2
          assign  (resid 43 and name C4')(resid 43 and name C3')(resid 43 and name O3')
                    (resid 44 and name  P ) 1.0 -155.0 30.0 2
          assign  (resid 44 and name C4')(resid 44 and name C3')(resid 44 and name O3')
                    (resid 45 and name  P ) 1.0 -155.0 30.0 2
          assign  (resid 45 and name C4')(resid 45 and name C3')(resid 45 and name O3')
                    (resid 46 and name  P ) 1.0 -155.0 30.0 2
!zetas

          assign  (resid 17 and name C3')(resid 17 and name O3')(resid 18 and name  P )
                    (resid 18 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 18 and name C3')(resid 18 and name O3')(resid 19 and name  P )
                    (resid 19 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 19 and name C3')(resid 19 and name O3')(resid 20 and name  P )
                    (resid 20 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 20 and name C3')(resid 20 and name O3')(resid 21 and name  P )
                    (resid 21 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 28 and name C3')(resid 28 and name O3')(resid 29 and name  P )
                    (resid 29 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 36 and name C3')(resid 36 and name O3')(resid 37 and name  P )
                    (resid 37 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 37 and name C3')(resid 37 and name O3')(resid 38 and name  P )
                    (resid 38 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 38 and name C3')(resid 38 and name O3')(resid 39 and name  P )
                    (resid 39 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 40 and name C3')(resid 40 and name O3')(resid 41 and name  P )
                    (resid 41 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 41 and name C3')(resid 41 and name O3')(resid 42 and name  P )
                    (resid 42 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 42 and name C3')(resid 42 and name O3')(resid 43 and name  P )
                    (resid 43 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 43 and name C3')(resid 43 and name O3')(resid 44 and name  P )
                    (resid 44 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 44 and name C3')(resid 44 and name O3')(resid 45 and name  P )
                    (resid 45 and name O5' ) 1.0 -68.0 60.0 2
          assign  (resid 45 and name C3')(resid 45 and name O3')(resid 46 and name  P )
                    (resid 46 and name O5' ) 1.0 -68.0 60.0 2
end
```

Appendix B.2 XPLOR scripts

XPLOR script 1: random.inp

This script generates random coordinates in a 20 Å box and then slowly increases the various geometric contributions to the energy function with the distance restraints kept high the whole time. The random seed of XPLOR was found to work not randomly so the program "gaseed" developed by Christopher Cilley which generates a random was used. To run the xplor scripts, a .csh file like the following was used:

```
remarks nmr/random.inp  adjusted dihe/cdih flags
remarks  The ultimate simulated annealing protocol for NMR structure determination!
remarks  The starting structure for this protocol can be completely arbitrary, such as random numbers.
remarks Note: the resulting structures need to be further processed by the dgsa.inp protocol.
remarks Author: Michael Nilges
remarks adapted by jrw 12/95 and asb 3/96


parameter                         {*Read the parameter file.*}
   @TOPPAR:gamma.par
end
parameter
   @TOPPAR:parallhdg.pro
end
structure @tar8_gamma.psf end          {*Read the structure file.*}
noe
  nres=3000         {*Estimate greater than the actual number of NOEs.*}
  class all
  @tar8final.tbl
end
@tar8dihed2.tbl               {*Read dihedral angle restraints.*}
noe                    {*Parameters for NOE effective energy term.*}
 ceiling=1000
 averaging  * R-6
 potential  * soft
 scale     * 1
 sqoffset   * 0.0
 sqconstant * 1.0
 sqexponent * 2
 soexponent * 1
 asymptote  * 2.0               {* Initial value - modified later. *}
 rswitch    * 1.0
end
restraints dihedral
      scale=100.
end
set message=off end
evaluate ($end_count=142)             {* Number of structures. *}
evaluate ($init_t = 1000 )     {* Initial simulated annealing temperature.*}

evaluate ($count = 0 )
{* copied from run that did well for comparison of new restraints *}
set seed=FOOBAR end
while ($count < $end_count ) loop main
```

```
    evaluate ($count=$count+1)
{* Generate a starting structure. *}
    vector do (x = (random()-0.5)*20) (all)
    vector do (y = (random()-0.5)*20) (all)
    vector do (z = (random()-0.5)*20) (all)
    vector do (fbeta=10) (all)     {*Friction coefficient for MD heatbath.*}
    vector do (mass=100) (all)          {*Uniform heavy masses to speed*}
parameter nbonds
        atom cutnb 100 tolerance 45 repel=1.2
        rexp=2 irexp=2 rcon=1.0 nbxmod 4
    end end
flags exclude * include bonds angle impr vdw noe end
    evaluate ($knoe = 1.0)
    evaluate ($kbon = 0.0001   )          {* Bonds.          *}
    evaluate ($kang = 0.0001   )          {* Angles.         *}
    evaluate ($kimp = 0.0  )          {* Impropers.      *}
    evaluate ($kvdw = 0.1)          {* Vdw.          *}
    constraints
        interaction (all) (all)
        weights bond $kbon angl $kang impr $kimp vdw $kvdw end
    end
    {* ================================================ High temperature dynamics. *}
    vector do (vx = maxwell($init_t)) (all)
    vector do (vy = maxwell($init_t)) (all)
    vector do (vz = maxwell($init_t)) (all)
    evaluate ($timestep = 0.04)
    evaluate ($nstep = 100)
    evaluate ($testframe = 0)
    evaluate ($totsteps = 0)
    while ($kbon < 0.01) loop stage1
        evaluate ($testframe = $testframe + 1)
            evaluate ($filename="testframe_"+encode($testframe)+".pdb")
            display total steps = $totsteps
            write coordinates output =$filename end
        evaluate ($kbon = min(0.25, $kbon * 1.25))
        evaluate ($kang = $kbon)
        evaluate ($kimp = 0)
        noe scale * $knoe end
        restraints dihed scale 0. end
        constraints
            interaction (all) (all)
            weights bond $kbon angl $kang impr $kimp vdw $kvdw end
        end
        dynamics  verlet
            nstep=$nstep   timestep=$timestep   iasvel=current
            tcoupling=true tbath=$init_t nprint=50  iprfrq=0
        end
            evaluate ($totsteps = $totsteps + $nstep)
    end loop stage1
    noe scale * 50 end
    parameter          {* Parameters for the repulsive energy term.  *}
      nbonds
      repel=0.9      {* Initial value for repel - modified later.  *}
        nbxmod=-3          {* Initial value for nbxmod - modified later. *}
```

```
     wmin=0.01
     cutnb=4.5 ctonnb=2.99 ctofnb=3.
     tolerance=0.5
   end
 end
 evaluate ($kbon=0.01)
 evaluate ($kvdw = 0.01)
 {* initialize dynamics *}
        evaluate ($testframe = $testframe + 1)
        evaluate ($filename="testframe_"+encode($testframe)+".pdb")
        display total steps = $totsteps
        write coordinates output = $filename end
   constraints
           interaction (all) (all)
           weights bond $kbon angl $kbon impr $kbon vdw $kvdw elec 0 end
 end
 dynamics verlet
        nstep=500   timestep=0.003   iasvel=maxwell
        firstt=$init_t
        tcoupling=true
        tbath=$init_t  nprint=50  iprfrq=0
 end
        evaluate ($totsteps = $totsteps + 500)
{* ================================= Write out the final structure(s).*}
 remarks ===========================================================
 remarks         overall,bonds,angles,improper,vdw,noe,cdih
 remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $cdih
{===>}                {*Name(s) of the family of initial structures.*}
  evaluate ($filename="t8finalinita_"+encode($count)+".pdb")
    display grand total steps = $totsteps
  write coordinates output =$filename end
restraints dihed scale 200. end
flags exclude * include bonds angle impr vdw noe cdih dihe end
        {* ramp up covalent terms *}
  evaluate ($kvdw = .01)
  while ($kbon < 1.0) loop rampcov
        evaluate ($testframe = $testframe + 1)
        evaluate ($filename="testframe_"+encode($testframe)+".pdb")
        remarks total steps = $totsteps
        write coordinates output = $filename end
        evaluate ($kbon = $kbon*1.5)
   constraints
           interaction (all) (all)
           weights bond $kbon angl $kbon impr $kbon vdw $kvdw  end
 end
 dynamics verlet
        nstep=500   timestep=0.003   iasvel=current
        tcoupling=true
        tbath=$init_t  nprint=50  iprfrq=0
 end
        evaluate ($totsteps = $totsteps + 500)
  end loop rampcov
{* ================================= Write out the final structure(s).*}
 remarks ===========================================================
```

```
  remarks          overall,bonds,angles,improper,vdw,dihe,noe,cdih
  remarks energies: $ener, $bond, $angl, $impr, $vdw, $dihe, $noe, $cdih
{===>}                    {*Name(s) of the family of covalent structures.*}
  evaluate ($filename="t8finalcova_"+encode($count)+".pdb")
    display grand total steps = $totsteps
  write coordinates output =$filename end
        {* ramp up vdw *}
  evaluate ($kbon=1.0)
  evaluate ($kvdw=0.01)
  while ($kvdw < 1.0) loop rampvdw
        evaluate ($testframe = $testframe + 1)
        evaluate ($filename="testframe_"+encode($testframe)+".pdb")
        display total steps = $totsteps
        write coordinates output = $filename end
        evaluate ($kvdw = $kvdw*1.5)
  constraints
          interaction (all) (all)
          weights bond $kbon angl $kbon impr $kbon vdw $kvdw elec 0 end
  end
  dynamics verlet
          nstep=500   timestep=0.003   iasvel=current
          tcoupling=true
          tbath=$init_t nprint=50 iprfrq=0
  end
          evaluate ($totsteps = $totsteps + 500)
  end loop rampvdw
  print threshold=0.5 noe
  evaluate ($rms_noe=$result)
  evaluate ($violations_noe=$violations)
  print threshold=2. cdih
  evaluate ($rms_cdih=$result)
  evaluate ($violations_cdih=$violations)
  print thres=0.05 bonds
  evaluate ($rms_bonds=$result)
  print thres=3. angles
  evaluate ($rms_angles=$result)
  print thres=5. impropers
  evaluate ($rms_impropers=$result)
  print thres=5. dihedrals
  evaluate ($rms_dihedrals=$result)
{* ============================== Write out the final structure(s).*}
  remarks ==========================================
  remarks          overall,bonds,angles,improper,vdw,dihe,noe,cdih
  remarks energies: $ener, $bond, $angl, $impr, $vdw, $dihe, $noe, $cdih
  {===>}                   {*Name(s) of the family of final structures.*}
  evaluate ($filename="randonly12_"+encode($count)+".pdb")
    display grand total steps = $totsteps
  write coordinates output =$filename end
end loop main
stop
```

XPLOR script 2: refine.inp

This script is run after the initial random.inp coordinates script. It cools the structures to 400 K through three 10 ps stops at 1000, 600, and 400 K. This protocol is modeled after the work of Jucker and Pardi (Jucker & Pardi, 1995).

```
remarks file refine.inp      -- Gentle simulated annealing refinement for NMR structure determination
remarks based on Pardi annealing protocol from Biochemistry on CUUG loop.
{====>}
evaluate ($init_t = 1000)                  {*SA temperature, in K.*}
{====>}
evaluate ($second_t = 600)
{====>}
evaluate ($third_t = 400)
{====>}
evaluate ($cool_steps = 10000 )            {*Total number of SA steps.*}
parameter                                  {*Read the parameter file.*}
{====>}
   @TOPPAR:gamma.par
end
parameter                                  {*Read the parameter file.*}
{====>}
   @TOPPAR:parallhdg.pro
end
{=====>} structure @tar8_gamma.psf end     {*Read the structure file. *}
noe
{====>}
  nres=3000       {*Estimate greater than the actual number of NOEs. *}
  class all
{====>}
   @tar8final.tbl                          {*Read NOE distance ranges.*}
   @tar8dist.tbl
end
@tar8dihed2.tbl                            {*Read dihedral angle restraints. *}
noe                    {*Parameters for NOE effective energy term.*}
  ceiling=1000
  averaging  * R-6
  potential  * square
  scale      * 100
  sqoffset   * 0.0
  sqconstant * 1.
  sqexponent * 2
  soexponent * 1
  asymptote  * 2.0                         {* Initial value - modified later. *}
  rswitch    * 1.0
end
parameter
    nbonds
    repel=0.9
    rexp=2 irexp=2 rcon=1.
    nbxmod=-2
    wmin=0.01
    cutnb=4.5 ctonnb=2.99 ctofnb=3.
    tolerance=0.5
```

213

```
        end
end
restraints dihedral        {* scale of 1000 = full value based on nucleotide tests *}
        scale=5.
end
set seed=FOOBAR end
evaluate ($end_count=41)        {*Loop through a family of 41 structures.*}
evaluate ($count = 0)
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)
{====>}                        {*Filename(s) for embedded coordinates.*}
  evaluate ($filename="final_"+encode($count)+".pdb")
  coor @@$filename
  flags exclude * include bond angl impr vdw noe cdih dihe end
{* in nucl. tests impr too high led to collapse of bases *}
  constraints
      interaction (all) (all)
      weights impr 0.5 end
  end
{*============================================ Intial minimization.*}
  minimize powell nstep= 100 nprint=25 end
{*============================================ Constant temperature SA.*}
  vector do ( fbeta=100. ) ( all )        {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$init_t
    tbath=$init_t  tcoup=true
    nprint=250 iprfrq=2500
  end
{*============================================ 700 step minimization.*}
  minimize powell nstep= 700 nprint=25 end
{*============================================ 600 K dynamics *}
  vector do ( fbeta=100. ) ( all )        {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$second_t
    tbath=$second_t tcoup=true
    nprint=250 iprfrq=2500
  end
{*============================================ 700 step minimization.*}
  minimize powell nstep= 700 nprint=25 end
{*============================================ 400 K dynamics *}
  vector do ( fbeta=100. ) ( all )        {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$third_t
    tbath=$third_t  tcoup=true
    nprint=250 iprfrq=2500
  end
{*============================================ Final minimization.*}
  minimize powell nstep= 1000 nprint=25 end
{*============================================ Write out the final structure(s).*}
  print threshold=0.2 noe
  evaluate ($rms_noe=$result)
```

```
evaluate ($violations_noe=$violations)
print threshold=5. cdih
evaluate ($rms_cdih=$result)
evaluate ($violations_cdih=$violations)
print thres=0.05 bonds
evaluate ($rms_bonds=$result)
print thres=5. angles
evaluate ($rms_angles=$result)
print thres=5. impropers
evaluate ($rms_impropers=$result)
print thres=5. dihedrals
evaluate ($rms_dihedrals=$result)
remarks ============================================================
remarks          overall,bonds,angles,improper,vdw,dihedrals,noe,cdih
remarks energies: $ener, $bond, $angl, $impr, $vdw, $dihe, $noe, $cdih
remarks ============================================================
remarks          bonds,angles,impropers,noe,cdih
remarks rms-d: $rms_bonds,$rms_angles,$rms_impropers,$rms_noe,$rms_cdih
remarks ============================================================
remarks          noe, cdih
remarks violations.: $violations_noe, $violations_cdih
remarks ============================================================
{====>}                  {*Name(s) of the family of final structures.*}
  evaluate ($filename="refine_"+encode($count)+".pdb")
  write coordinates output =$filename end
end loop main
stop
```

XPLOR script 3: repelno.inp

This script slowly cools the structures to 10K using the quartic non-bonding terms.

```
remarks file nmr/refinelow.inp       -- Gentle simulated annealing refinement
remarks                     for NMR structure determination
evaluate ($zero_t = 500)
evaluate ($init_t = 300)              {*SA temperature, in K.*}
evaluate ($second_t = 150)
evaluate ($third_t = 50)
evaluate ($fourth_t = 10)
evaluate ($cool_steps = 2000 )        {*Total number of SA steps.*}
parameter                     {*Read the parameter file.*}
@TOPPAR:gamma.par
end
parameter                     {*Read the parameter file.*}
@TOPPAR:parallhdg.pro
end
{====>} structure @tar8_gamma.psf end      {*Read the structure file.*}
noe
  nres=3000       {*Estimate greater than the actual number of NOEs.*}
  class all
  @tar8final.tbl                {*Read NOE distance ranges.*}
  @tar8dist.tbl
end
@tar8dihed2.tbl                 {*Read dihedral angle restraints.*}
```

215

```
noe                    {*Parameters for NOE effective energy term.*}
  ceiling=1000
  averaging  * R-6
  potential  * square
  scale      * 50
  sqoffset   * 0.0
  sqconstant * 1.
  sqexponent * 2
  soexponent * 1
  asymptote  * 2.0              {* Initial value - modified later. *}
  rswitch    * 1.0
end
{* lower repel radius 4/5/96 *}
parameter
    nbonds
    repel=0.9
    rexp=2 irexp=2 rcon=1.
    nbxmod=-5
    wmin=0.01
    cutnb=4.5 ctonnb=2.99 ctofnb=3.
    tolerance=0.5
    end
end
restraints dihedral        {* scale is 1000 = full value base on nt. tests *}
    scale=20.
end
vector do (mass=100) (all)
set seed=FOOBAR end
evaluate ($end_count= 27)        {*Loop through a family of 7 structures.*}
evaluate ($count = 0 )
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)
{====>}                         {*Filename(s) for embedded coordinates.*}
  evaluate ($filename="refine_"+encode($count)+".pdb")
  coor @@$filename
  flags exclude * include bond angl impr vdw noe cdih end
{* in nucl. tests impr too high led to collapse of bases *}
  constraints
      interaction (all) (all)
      weights impr 0.5 end
  end
vector do (fbeta=10) (all)     {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$zero_t
    tbath=$zero_t tcoup=true
    nprint=250 iprfrq=2500
  end
  {* ================================================= Constant temperature SA.*}
  vector do (fbeta=10) (all)     {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$init_t
    tbath=$init_t tcoup=true
```

```
    nprint=250 iprfrq=2500
  end
{*================================================ 100 K dynamics *}
  vector do (fbeta=10) (all)        {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$second_t
    tbath=$second_t tcoup=true
    nprint=250 iprfrq=2500
  end
{*================================================ 50 K dynamics *}
vector do (fbeta=10) (all)
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$third_t
    tbath=$third_t tcoup=true
    nprint=250 iprfrq=2500
  end
vector do (fbeta=10) (all)        {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$fourth_t
    tbath=$fourth_t tcoup=true
    nprint=250 iprfrq=2500
end
  {*================================================ Final minimization.*}
  minimize powell nstep= 1000 nprint=25 end
  {*================================== Write out the final structure(s).*}
  print threshold=0.2 noe
  evaluate ($rms_noe=$result)
  evaluate ($violations_noe=$violations)
  print threshold=5. cdih
  evaluate ($rms_cdih=$result)
  evaluate ($violations_cdih=$violations)
  print thres=0.05 bonds
  evaluate ($rms_bonds=$result)
  print thres=5. angles
  evaluate ($rms_angles=$result)
  print thres=5. impropers
  evaluate ($rms_impropers=$result)
  print thres=5. dihedrals
  evaluate ($rms_dihedrals=$result)
  remarks ================================================
  remarks         overall,bonds,angles,improper,vdw,dihedrals,noe,cdih
  remarks energies: $ener, $bond, $angl, $impr, $vdw, $dihe, $noe, $cdih
  remarks ================================================
  remarks         bonds,angles,impropers,noe,cdih
  remarks rms-d: $rms_bonds,$rms_angles,$rms_impropers,$rms_noe,$rms_cdih
  remarks ================================================
  remarks         noe, cdih
  remarks violations.: $violations_noe, $violations_cdih
  remarks ================================================
{====>}                  {*Name(s) of the family of final structures.*}
  evaluate ($filename="repelno1_"+encode($count)+".pdb")
```

```
write coordinates output =$filename end
end loop main
stop
```

XPLOR script 4: ljno.inp

This is the last XPLOR script which cools the structure to 10 K with Lennard-Jones force field turned on. Otherwise, it is the same protocol as repelno.inp.

```
remarks file ljno.inp
remarks no dihe force field to allow structures a chance to fix themselves
evaluate ($init_t = 300)              {*SA temperature, in K.*}
evaluate ($second_t = 150)
evaluate ($third_t = 50)
evaluate ($fourth_t = 10)
evaluate ($cool_steps = 2000 )        {*Total number of SA steps.*}
parameter                             {*Read the parameter file.*}
@TOPPAR:gamma.par
end
parameter                             {*Read the parameter file.*}
@TOPPAR:parallhdg.pro
end
{====>} structure @tar8_gamma.psf end     {*Read the structure file.*}
noe
{===>}
  nres=3000         {*Estimate greater than the actual number of NOEs.*}
  class all
{===>}
  @tar8final.tbl                      {*Read NOE distance ranges.*}
  @tar8dist.tbl
end
@tar8dihed2.tbl                       {*Read dihedral angle restraints.*}
noe                   {*Parameters for NOE effective energy term.*}
  ceiling=1000
  averaging  * R-6
  potential  * square
  scale      * 50
  sqoffset   * 0.0
  sqconstant * 1.
  sqexponent * 2
  soexponent * 1
  asymptote  * 2.0           {* Initial value - modified later. *}
  rswitch    * 1.0
end
parameter
nbonds
    tolerance=0.5
  cutnb=11.5 ctonnb=9.5 ctofnb=10.5 tolerance=0.5 rdie vswitch switch
    nbxmod=-5
  end
end
restraints dihedral            {* scale is 1000 = full value base on nt. tests *}
    scale=1000.
end
```

218

```
vector do (mass=100) (all)
set seed=FOOBAR end
evaluate ($end_count= 27)        {*Loop through a family of 7 structures.*}
evaluate ($count =0)
while ($count < $end_count ) loop main

  evaluate ($count=$count+1)
  {====>}                       {*Filename(s) for embedded coordinates.*}
  evaluate ($filename="repelno1_"+encode($count)+".pdb")
  coor @@$filename
  flags exclude * include bond angl impr vdw noe cdih end
{* in nucl. tests impr too high led to collapse of bases *}
  constraints
          interaction (all) (all)
          weights impr 0.5 end
  end
  {*=================================================== Constant temperature SA.*}
  vector do (fbeta=10) (all)      {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$init_t
    tbath=$init_t tcoup=true
    nprint=250 iprfrq=2500
  end
  {*=================================================== 100 K dynamics *}
  vector do (fbeta=10) (all)           {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$second_t
    tbath=$second_t tcoup=true
    nprint=250 iprfrq=2500
  end
  {*=================================================== 50 K dynamics *}
vector do (fbeta=10) (all)
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$third_t
    tbath=$third_t tcoup=true
    nprint=250 iprfrq=2500
    end
  {*=================================================== 10 K dynamics *}
vector do (fbeta=10) (all)            {* Coupling to heat bath. *}
  dynamics verlet
    nstep=$cool_steps timestep=0.001
    iasvel=maxwell firsttemperature=$fourth_t
    tbath=$fourth_t tcoup=true
    nprint=250 iprfrq=2500
end
  {*=================================================== Final minimization.*}
` minimize powell nstep= 1000 nprint=25 end
  {*=================================================== Write out the final structure(s).*}
  print threshold=0.1 noe
  evaluate ($rms_noe=$result)
  evaluate ($violations_noe=$violations)
```

```
print threshold=2. cdih
evaluate ($rms_cdih=$result)
evaluate ($violations_cdih=$violations)
remarks ===============================================
remarks          overall,bonds,angles,improper,vdw, noe,cdih
remarks energies: $ener, $bond, $angl, $impr, $vdw, $noe, $cdih
remarks ===============================================
remarks          bonds,angles,impropers,noe,cdih
remarks ===============================================
remarks              noe, cdih
remarks violations.: $violations_noe, $violations_cdih
remarks ===============================================
{====>}                    {*Name(s) of the family of final structures.*}
   evaluate ($filename="ljno_"+encode($count)+".pdb")
   write coordinates output =$filename end
end loop main
stop
```

Appendix B.3 PERL scripts

This appendix presents a suite of programs run by one master C shell script to analyze any set of RNA XPLOR .pdb files describing a RNA structure. A C shell script is used to run all the PERL programs mostly because PERL3 does not support multidimensional arrays straightforwardly. However, PERL5 does, and these programs probably could be upgraded, but despite their length, they work just fine.

Torsion.csh:

This is the master .csh script: torsion.csh that calls in some PERL scripts. The output includes a file called table which lists each nucleotide's average torsion with the standard deviation.   All that needs to be adjusted is the numer of total structures the "to" variable and the "nuc_min" and "nuc_max" variables which are the nucleotide numbering.

```
#!/bin/csh
set n=1
set to=20
set nuc_min=16
set nuc_max=46
set out=final
# this loop runs the xplor file torsion_core.inp that actually measures the torsions
# then the PERL script cleantor.pl deals with the date and spits out the all the torsions by structure into
# final.tor files
while ( $n <= $to )
      sed -e "s/FOOBAR/$out$n.pdb/" torsion_core.inp > torsion.inp2
      xplor < torsion.inp2 > tor_$n.log
      cat tor_$n.log | cleantor.pl > $out$n.tor
      mv tabtor tot_$n.tor
      @ n = $n + 1
end
#reset n to 1
set n=1
#this loop takes the xplor output and places sorts by
while ( $n <= $to )
      cat tor_$n.log | tornuc.pl > struct_$n.tor
      mv tabtor tot_$n.tor
```

```
        @ n = $n + 1
end
set i=$nuc_min
set imax=$nuc_max
# this loop takes the struct.tor files and makes files sorted by nucleotide
while ($i <= $imax )
        grep " $i " struct*.tor > nuc_$i.tor
        @ i = $i + 1
end
# reset i
set n=$nuc_min
# this loop takes the nucleotide sorted torsion files and calculates the average and standard deviation over all
# the structures by nucleotide position and makes a neat table in a file called table.
header.pl
sed -e "s/FOO/$to/" table.pl > table.pl2
chmod +x *.pl2
while ( $n <= $nuc_max )
echo this is file number $n
cat nuc_$n.tor | table.pl2
@ n =$n + 1
end
```

Header.pl

```perl
#!/usr/local/bin/perl
open(output ,"> table");
print output "residue  alfa      beta      gamma     delta     epsilon     zeta \n";
print output "         292       178       60        88        205         292 \n";
```

Table.pl

This PERL script takes input from the nuc.tor files and outputs a file called table which lists
the average torsion ± the standard deviation at each position.

```perl
#!/usr/local/bin/perl
open(output ,">> table");
#total structures are adjusted by the tor_master.csh script
$total_structures = FOO;
$alfa_total = 0;
$beta_total = 0;
$gamma_total = 0;
$epsilon_total = 0;
$zeta_total = 0;
$delta_total=0;
$i=0;  $j=0;
while ($i < $total_structures) {
        @alfa1[$i]=($j);
        $i = $i + 1;   }
$i=0;  $j=0;
while ($i < $total_structures) {
        @beta1[$i]=($j);
        $i = $i + 1;   }
$i=0;  $j=0;
while ($i < $total_structures) {
        @gamma1[$i]=($j);
```

221

```
        $i = $i + 1;   }
$i=0;  $j=0;
while ($i < $total_structures) {
        @epsilon1[$i]=($j);
        $i = $i + 1;   }
$i=0; $j=0;
while ($i < $total_structures) {
        @zeta1[$i]=($j);
        $i = $i + 1;   }
$i=0;  $j=0;
while ( $i < $total_structures)  {
        @delta1[$i]=($j);
        $i = $i + 1;    }
$count=0;
$i=0;
while(<>)  {
($file, $residue, $alfa, $beta, $gamma, $delta, $epsilon, $zeta) = split(/\s+/,$_);
if ( $alfa < 0 )  {
        $alfa1[$i] = $alfa + 360;   }
else  {
$alfa1[$i] = $alfa;   }
if ($beta < 0 )  {
        $beta1[$i] = $beta + 360;   }
else {
$beta1[$i] = $beta;    }
if ($gamma < 0 )  {
        $gamma1[$i] = $gamma + 360; }
else {
$gamma1[$i] = $gamma;    }
if ($epsilon < 0 )  {
        $epsilon1[$i] = $epsilon + 360;    }
else {
$epsilon1[$i] = $epsilon;    }
if ( $zeta < 0 )  {
        $zeta1[$i] = $zeta + 360;  }
else {
$zeta1[$i] = $zeta;    }
if ( $delta < 0 )  {
        $delta1[$i] = $delta + 360; }
else {
$delta1[$i]=$delta;  }
$i= $i+1;
$count= $count + 1;
#end of while loop   }
$i=0;
while ( $i < $total_structures) {
     $alfa_total = $alfa1[$i] + $alfa_total;
     $beta_total = $beta1[$i]  + $beta_total;
     $gamma_total = $gamma1[$i]  + $gamma_total;
     $epsilon_total = $epsilon1[$i]  + $epsilon_total;
     $zeta_total = $zeta1[$i]  + $zeta_total;
     $delta_total = $delta1[$i]  + $delta_total;
     $i =$i +1;        }
$alfa_avg= $alfa_total / $total_structures;
```

```perl
$beta_avg = $beta_total / $total_structures;
$gamma_avg = $gamma_total / $total_structures;
$epsilon_avg = $epsilon_total / $total_structures;
$zeta_avg = $zeta_total / $total_structures;
$delta_avg = $delta_total / $total_structures;
printf output "%4d %s ", $residue, $name;
$angle_count=1;
&standard_dev($alfa_avg, $angle_count, @alfa1);
$angle_count=2;
&standard_dev($beta_avg, $angle_count, @beta1);
$angle_count = 3;
&standard_dev($gamma_avg, $angle_count, @gamma1);
$angle_count=4;
&standard_dev($delta_avg, $angle_count, @delta1);
$angle_count=5;
&standard_dev($epsilon_avg, $angle_count, @epsilon1);
$angle_count=6;
&standard_dev($zeta_avg, $angle_count, @zeta1);
printf output "\n";
#this function calculates the standard deviation
sub standard_dev {
local($avg, $angle_name, @angle) = @_;
$sum=0;
$diff=0;
$diffa=0;
$stnd=0;
#print "this is the average angle: $avg \n";
$i=0;
#print "this is angle_res:$angle_name \n";
if ( $angle_name  == 1 ) {
      $name = alfa; }
if ($angle_name == 2) {
      $name = beta;  }
if ($angle_name == 3 ) {
      $name = gamma;  }
if ($angle_name == 4 ) {
      $name = delta;  }
if ($angle_name == 5 ) {
      $name = epsilon;  }
if ($angle_name == 6  ) {
      $name = zeta;  }
if ($angle[$i] < $avg) {
            $diff = $angle[$i] - $avg;        }
if ( $angle[$i] > $avg  ) {
            $diff=($avg - $angle[$i]);        }
      $diffa = $diff * $diff;
      $sum = $diffa + $sum;
      $i++;          }
      $stnd = sqrt($sum)/$total_structures;
# this section makes for pretty ouput of the table
if ( $stnd > 10 ) {
      if ($avg < 100 && $avg > 10 ) {
                  printf output "  %-4.2f ± %-4.2f ", $avg, $stnd;
                  }
```

223

```
            else { printf output " %-4.2f ± %-4.2f ", $avg, $stnd; }      }
else {
      if ( $avg == 0 ) {
            printf output "   %-4.2f ±  %-4.2f ", $avg, $stnd; }
      else {
          if ($avg < 100 && $avg > 10 ) {
                printf output "  %-4.2f ±  %-4.2f ", $avg, $stnd; }
          if ($avg < 10 ) {
                printf output "   %-4.2f ±  %-4.2f ", $avg, $stnd; }
          if ($avg > 100 )  {
                printf output " %-4.2f ±  %-4.2f ", $avg, $stnd; }              }     }
# end of stnd dev function
}
```

Energy.pl:

This script calculates the average value for all the energy terms used in structure calculations. Unfortunately, this was written before I figured out object oriented PERL programming, so it is a bit verbose. The energies line is greped from the .pdb files and then sent to the PERL script.

```
grep energies ljfinal*.pdb > foo
cat foo I energy.pl
```

```perl
#!/usr/local/bin/perl
open(output, "> total_energy");
$total_structures=20;
$i=0;  $j=0;
while ($i < $total_structures) {
      @bond_energy[$i]=($j);
      $i = $i + 1;  }
$i=0;  $j=0;
while ($i < $total_structures) {
      @cdih_array[$i]=($j);
      $i = $i + 1;   }
$i=0;  $j=0;
while ($i < $total_structures) {
      @bond_array[$i]=($j);
      $i = $i + 1;   }
$i=0;  $j=0;
while ($i < $total_structures) {
      @bond_improper[$i]=($j);
      $i = $i + 1;   }
$i=0;  $j=0;
while ($i < $total_structures) {
      @angle_array[$i]=($j);
      $i = $i + 1;   }
$i=0;  $j=0;
while ($i < $total_structures) {
      @vdw_array[$i]=($j);
      $i = $i + 1;   }
$i=0;  $j=0;
while ($i <$total_structures) {
      @diff_from_mean[$i]=($j);
      $i = $i + 1;   }
```

```perl
$energy=0;
$avg_energy=0;
$count=0;
$noe=0;
$cdih=0;
$avg_noe=0;
$avg_cdih=0;
$noe_high = 0;
$noe_low = 100;
$noe_high=0;
$noe_low=100;
$energy_low=0;
$energy_high=-500;
$j=0;
while(<>) {
        if( /\s*energies:\s*(\S*)\s*(\S*)\s*(\S*)\s*(\S*)\s*(\S*)\s*(\S*)\s*(\S*)\s*/) {
        #print "$2\n";
        $count++;
        $energy = $1 + $energy;
        if($1 > $energy_high) {
                $energy_high = $1;  }
        if($1 < $energy_low)  {
                $energy_low = $1;  }
        $noe = $6 + $noe;
        $cdih= $7 + $cdih;
        $bond = $2 + $bond;
        $angle = $3 + $angle;
        $improper = $4 + $improper;
        $vdw = $5 + $vdw;
        $cdih_array[$j] = $7;
        $bond_array[$j] = $2;
        $angle_array[$j] = $3;
        $improper_array[$j] = $4;
        $vdw_array[$j] = $5;
        $energy_array[$j] = $1;
        $j++;
        if($6 > $noe_high) {
                $noe_high=$6;  }
        if($6 < $noe_low)  {
                $noe_low=$6;  }
        if($7 > cdih_high)  {
                $cdih_high = $7;  }
        if($7 < $cdih_low)  {
                $cdih_low = $7;  }
        #end of if
        }
#end of while input loop
}
print "this is count $count\n";
$avg_energy = $energy / $count;
$avg_noe = $noe / $count;
$avg_cdih = $cdih / $count;
$avg_bond = $bond / $count;
$avg_angle = $angle / $count;
```

225

```
$avg_improper = $improper/$count;
$avg_vdw = $vdw/$count;
#calculate standard deviation
$test=0;
$i=0;
while ($i < $count) {
        $test = $avg_energy - $energy_array[$i];
        if( $test > 0 ) {
                $diff_from_mean[$i]= $test ;          }
        if( $test < 0 ) {
                $diff_from_mean[$i]= 0 - $test;       }
        $square = $diff_from_mean[$i] * $diff_from_mean[$i];
        $sum_stnd_energy = $sum_stnd_energy + $square;
     $i = $i + 1;        }
$i=0;
while ($i < $count) {
        $test = $avg_bond - $bond_array[$i];
        if( $test > 0 ) {
                $diff_from_mean[$i]= $test ;          }
        if( $test < 0 ) {
                $diff_from_mean[$i]= 0 - $test;       }
        $square = $diff_from_mean[$i] * $diff_from_mean[$i];
        $sum_stnd_bond = $sum_stnd_bond + $square;
     $i = $i + 1;   }
$i=0;
while ($i < $count) {
        $test = $avg_angle - $angle_array[$i];
        if( $test > 0 ) {
                $diff_from_mean[$i]= $test ;          }
        if( $test < 0 ) {
                $diff_from_mean[$i]= 0 - $test;       }
        $square = $diff_from_mean[$i] * $diff_from_mean[$i];
        $sum_stnd_angle = $sum_stnd_angle + $square;
     $i = $i + 1;     }
$i=0;
while ($i < $count) {
        $test = $avg_improper - $improper_array[$i];
        if( $test > 0 ) {
                $diff_from_mean[$i]= $test ;          }
        if( $test < 0 ) {
                $diff_from_mean[$i]= 0 - $test;
                #print "this is diff_from_mean $diff_from_mean[$i] \n";          }
        $square = $diff_from_mean[$i] * $diff_from_mean[$i];
        $sum_stnd_improper = $sum_stnd_improper + $square;
     $i = $i + 1;     }
$i=0;
while ($i < $count) {
        $test = $avg_vdw - $vdw_array[$i];
        if( $test > 0 ) {
                $diff_from_mean[$i]= $test ;   }
        if( $test < 0 ) {
                $diff_from_mean[$i]= 0 - $test;       }
        $square = $diff_from_mean[$i] * $diff_from_mean[$i];
        $sum_stnd_vdw = $sum_stnd_vdw + $square;
```

226

```
        $i = $i + 1;    }
$i=0;
while ($i < $count) {
        $test = $avg_cdih - $cdih_array[$i];
        if( $test > 0 ) {
                $diff_from_mean[$i]= $test ;        }
        if( $test < 0 ) {
                $diff_from_mean[$i]= 0 - $test;
                #print "this is r $r \n";            }
        $square = $diff_from_mean[$i] * $diff_from_mean[$i];
        $sum_stnd_cdih = $sum_stnd_cdih + $square;
        $i = $i + 1;      }
$stand_dev_cdih = sqrt($sum_stnd_cdih/)$total_structures;
$stand_dev_bond = sqrt($sum_stnd_bond)/ $total_structures;
$stand_dev_angle = sqrt($sum_stnd_angle)/ $total_structures;
$stand_dev_improper = sqrt($sum_stnd_improper)/ $total_structures;
$stand_dev_vdw = sqrt($sum_stnd_vdw)/ $total_structures;
$stand_dev_energy = sqrt($sum_stnd_energy)/ $total_structures;
print "this is the average total energy: $avg_energy \n";
print "This is the standard deviation of the total energy: $stand_dev_energy \n";
print "this is the average noe violation energy $avg_noe \n";
print "the total energy ranges from $energy_low to $energy_high \n";
print "the noe violation energy ranges from $noe_low to $noe_high \n";
print "this is the average cdih violation energy $avg_cdih \n";
print "the cdih violation energy ranges from $cdih_low to $cdih_high \n";
print "the standard deviation of the cdih violation energy is $stand_dev_cdih \n";
print " this is the average bond energy: $avg_bond \n";
print "the standard deviation of the bond violation energy is $stand_dev_bond \n";
print " This is the average angle energy: $avg_angle \n";
print " This is the standard deviation for the ANGLE energy $stand_dev_angle \n";
print " this is the average improper energy: $avg_improper \n";
print "This is the standard deviation for the improper energy: $stand_dev_improper \n";
print " This is the average non-bonding energy: $avg_vdw \n";
print "This is the standard deviation for the non-bonding energy: $stand_dev_vdw \n";
```

Appendix C: The Williamson WWW Home Page and NMR Database: HTML and cgi-bin Scripts

This Appendix presents some of the HTML and cgi-bin scripts that were part of the Williamson Lab home page at MIT. Only the main home page HTML is given as most of the information will be out of date and can be found on-line anyway. In addition to the home page, a crude NMR database was written to keep track of all the NMR experiments that are run in the lab. The database uses slightly different scripts and HTML depending on the dimension of the NMR experiment (2D, 3D, or 4D). These cgi-bin scripts are written in PERL3 and call in the standard cgi-lib.pl script which does a lot of the hard work of dealing with the data from the HTML form pages. A representative set of the form page, and PERL scripts is given for 4D data. A similiar set of files are used to deal with 2D and 3D NMR data.

Williamson Lab Home Page:

```
<HTML>
<HEAD>
<base href="http://volvox.mit.edu/home_us.html">
<TITLE>Williamson Lab: for internal use only</TITLE>
</HEAD>
<body bgcolor="000000" vlink="ff0000" link="88BBff" text="FFFFFF">
<table><tr><th>
<a href =links/figs/hiv2_tar.gif><img src="links/figs/hiv2_tar.gif" height=233 width=140
alt = Structure of HIV-2 TAR RNA binding Argininamide.></a></th>
<th><font size =21>
Williamson <a name="Group">Group<p></a></font></th>
<th><a href =figs/webpage1.gif> <img src="figs/webpage1.gif" height=233 width=194
alt = Structure of HIV RRE RNA and rev peptide Complex.></a></tr>
<tr><td><strong><center>Structure of HIV-2 TAR RNA binding Argininamide.</center></strong></td>
<td><strong><center>Structure of HIV RRE RNA - rev peptide Complex.
</center></strong></td></tr></table><blockquote>
The Williamson Lab is in the Dept. of Chemistry at MIT. We are interested in the fascinating molecule,
RNA. Our research focuses on understanding RNA's interactions with ligands, especially proteins, and on
how really large RNA molecules fold into complex structures. At this site, we have made available some
<a href="links/coords.html">pdb files</a> of the NMR models that have been developed in the lab.
</blockquote><font size=10><table border=7>
<tr><td align=center><strong><a href="links/members.html">Lab Personnel </a></strong> </td>
<td align=center><strong><a href="links/coords.html">NMR Models & Coordinates</a> </strong></td>
</tr><tr><td align=center><strong><a href="links/linkframe.html">Useful Links</a> </strong></td>
<td align=center><strong><a href="links/publications.html">Publications & Abstracts</a>
</strong></td></tr><tr>
<td align=center><strong><a href="http://www.wi.mit.edu/news/bioweek.html" target= "new">Biology
Week</a></strong> </td>
<td align=center><strong><a href="/usage/">Usage statistics for this server.</a>
</strong> </td></tr><tr>
<td align=center><strong><A HREF = figs/view.gif>A Lovely View of Boston from the
Lab</a></strong></td>
<td align=center><strong><A HREF = "http://www.document.com/cam/now.html" target ="new">
        Current WeatherCam View from Downtown Boston </A></strong></td></tr>
<tr><td align=center><strong><a href="links/new.html">What's New</a></strong></td>
<td align=center><strong><a href="labonly/lab.html">Williamson Lab Stuff</a></strong></td></tr>
```

<tr>
<td align=center>Williamson Lab Chair Museum</td>
<td align=center>Daily Funke Factoid</td></tr>
<tr><td align=center>The Tenure Bash! </td></tr>
</table><p>
<img src="/cgi-bin/Count.cgi?df=sample.dat" align=absmiddle> people have visited this page since May 1, 1997!
Any comments and/or questions about these pages or the Lab are welcome!
</BODY></HTML>

4d_nmr.html:

This is the form page for entering the NMR data to the NMR database.

<HTML>
<HEAD>
<base href="http://volvox.mit.edu/labonly/NMR/4d_nmr.html">
<TITLE>4D NMR FORM</TITLE></HEAD>
<body vlink="ff0000" link="88BBff" text="000000" bgcolor="#BFDFFF">
<center>
You are entering information for a 4-D experiment!
If this is not correct you can go back to the main NMR page by clicking here!
<A HREF="nmr_master.html"><IMG SRC="../../links/figs/back.gif" ALIGN="left" ALT="[Go Back]"></A> </center>
<br clear="all">
<FORM METHOD="POST" ACTION="http://volvox.mit.edu/cgi-bin/4D_magnetsort.pl">
<HR><center>
<h3>Enter all information carefully to ensure accuracy!! </h3>
<h3> DO NOT use any commas in any of the text boxes!! The program will not work properly! </h3>
</center><HR><table cellspacing=10><tr><td align=center>
Enter your name (choose a user name and use this for all entries!) </td>
<td align=left><INPUT TYPE="text" NAME="name" SIZE="20"> </td></tr>
<td align=center> Login id</td>
<td align=left><INPUT TYPE="text" NAME="login" SIZE="20"> </td></tr>
<td align=center>
Name of the sample e.g. tfr24-TAR8 </td>
<td align=left><INPUT TYPE="text" NAME="sample_name" SIZE="30"> </td></tr></table>
Please indicate the type of sample.

<INPUT type="radio" name="sample_label" value="peptide/protein only" > peptide/protein only
<INPUT type="radio" name="sample_label" value="RNA only" > RNA only
<INPUT type="radio" name="sample_label" value="peptide/protein-RNA complex" > peptide/protein-RNA complex
<INPUT type="radio" name="sample_label" value="RNA_ligand complex" > RNA-ligand complex
<INPUT type="radio" name="sample_label" value="nucleotide" > nucleotide
<INPUT type="radio" name="sample_label" value="other test sample" > other test samples
<INPUT type="radio" name="sample_label" value="other" > other
<P><table cellspacing=10>
<td align=center>
Please indicate labeling pattern of sample. </td>
<td align=left>
<SELECT name="label">
<OPTION>all molecules unlabeled
<OPTION>13C RNA

```
<OPTION>15N RNA
<OPTION>13C/15N RNA
<OPTION>13C/2H RNA
<OPTION>15N peptide or protein
<OPTION>15N/13C peptide or protein
<OPTION>2H random RNA
<OPTION>2H selective RNA
<OPTION>2H random peptide or protein
<OPTION>2H/13C/15N protein
<OPTION>2H/15N protein
<OPTION>2H protein
<OPTION>13C RNA in complex
<OPTION>15N RNA in complex
<OPTION>13C/15N RNA in complex
<OPTION>15N peptide or protein in complex
<OPTION>15N/13C peptide or protein in complex
<OPTION>2H random RNA in complex
<OPTION>2H selective RNA in complex
<OPTION>2H random peptide or protein in complex
<OPTION>2H/13C/15N protein in complex
<OPTION>2H/15N protein in complex
<OPTION>2H protein in complex        </select></P></td></tr>
<tr><td align=center><font size=2>
Please indicate the concentration of the sample e.g. 1 mM </font> </td>
<td align=left><INPUT TYPE="text" NAME="concentration"  SIZE="20"> </td></tr>
<td align=center> <font size=2>
Please indicate the approximate concentration of salt and what kind of salt e.g. 50mM NaCl </font> </td>
<td align=left><INPUT TYPE="text" NAME="salt"  SIZE="20"> </td></tr>
<tr><td align=center> <font size=2> Which Magnet did you use? </font> </td>
<td align =left>
   <SELECT name="magnet">
   <OPTION>None
   <OPTION>Didinium Varian 600
   <OPTION>Bitter Lab 500
   <OPTION>Bitter Lab 591
   <OPTION>Bitter Lab 750
   <OPTION>Wagner Varian 750
   <OPTION>Other        </SELECT></td></tr>
<tr><td align=center> <font size=2>When did you start on  the magnet? </font> </td>
<td align =left>
        <SELECT name="date">
        <OPTION>January
        <OPTION>February
        <OPTION>March
        <OPTION>April
        <OPTION>May
        <OPTION>June
        <OPTION>July
        <OPTION>August
        <OPTION>September
        <OPTION>October
        <OPTION>November
        <OPTION>December        </SELECT>
        <SELECT name="number">
        <OPTION>1
        <OPTION>2
        <OPTION>3
        <OPTION>4
        <OPTION>5
        <OPTION>6
```

```
<OPTION>7
<OPTION>8
<OPTION>9
<OPTION>10
<OPTION>11
<OPTION>12
<OPTION>13
<OPTION>14
<OPTION>15
<OPTION>16
<OPTION>17
<OPTION>18
<OPTION>19
<OPTION>20
<OPTION>21
<OPTION>22
<OPTION>23
<OPTION>24
<OPTION>25
<OPTION>26
<OPTION>27
<OPTION>28
<OPTION>29
<OPTION>30
```

<OPTION>31   </SELECT>

<SELECT name="year">

<OPTION>1997

<OPTION>1998

<OPTION>1999<OPTION>1997

<OPTION>1998

<OPTION>1999</SELECT></td></tr>

<tr><td align=center> <font size=2>When did you finish on the magnet? </td>

<td align=left><SELECT name="date_fin">

```
<OPTION>January
<OPTION>February
<OPTION>March
<OPTION>April
<OPTION>May
<OPTION>June
<OPTION>July
<OPTION>August
<OPTION>September
<OPTION>October
<OPTION>November
<OPTION>December        </SELECT>
```

<SELECT name="number_fin">

```
<OPTION>1
<OPTION>2
<OPTION>3
<OPTION>4
<OPTION>5
<OPTION>6
<OPTION>7
<OPTION>8
<OPTION>9
<OPTION>10
<OPTION>11
<OPTION>12
<OPTION>13
<OPTION>14
<OPTION>15
<OPTION>16
<OPTION>17
<OPTION>18
<OPTION>19
<OPTION>20
<OPTION>21
<OPTION>22
<OPTION>23
<OPTION>24
<OPTION>25
<OPTION>26
<OPTION>27
<OPTION>28
<OPTION>29
<OPTION>30
```

```
                <OPTION>31
                </SELECT><SELECT name="year_fin">
                <OPTION>1997
                <OPTION>1998
                <OPTION>1999  </SELECT></td></tr>
<tr><td align=center> <font size=2>Please enter the probe you used:</font></td>
<td align=left>
                <SELECT name="probe">
                <OPTION> None
                <OPTION> Didinium HCN Z-grad
                <OPTION> Didinium HCP Z-grad
                <OPTION> Bitter 500 HCN Z-grad
                <OPTION> Bitter 750 H only Z-grad
                <OPTION> Bitter 750 HX Z-grad
                <OPTION> Wagner 750 HCN Z-grad
                </SELECT></td></tr></table>
<hr><h3>Enter the experiment information requested you ran and whether they were considered
successful or not. (Note: default is successful ... ever the optimist!) </h3>
<table cellspacing=10><tr><td align=center> <font size=2>
Please Choose from the list of standard Experiments (If you choose other, enter the name in the textarea
below):    </font></td>
<td align=left>
                <SELECT name="stndexpt">
                <OPTION> 1H-13C HMQC-NOESY-HMQC
                <OPTION> OTHER
                </SELECT></td> </tr>
<tr><td align=center> <font size=2>
If you entered OTHER, please give a name for your experiment.  Otherwise you can
just leave this blank. (once you create a name, please use this name for all entries)
</td><td align=left><INPUT TYPE="text" NAME="nonstndexpt"  SIZE="25"> </td></tr>
<tr><td align=center> <font size=2>
Please enter the name of the Varian pulse sequence name (seqfil) or Bitter Lab macro to run this experiment.
</font></td><td align=left>
<INPUT TYPE="text" NAME="macro"  SIZE="30"></td></tr>
<tr><td align=center> <font size=2>Did your experiment work to your satisfaction? </Font></td>
<td align=left><INPUT type="radio" name="success" value="successful" checked> successful
<INPUT type="radio" name="success" value="not_successful"> not successful<br></td></tr>
</table><HR><center><h4>
Experimental Parameters: </h4>
Enter all pulse lengths in microseconds!<br>    </center>
<table cellspacing=10>
<tr><td align=center> <font size=2> Please enter the temperature at which you ran the experiment.
</font></td>
<td align=left><INPUT TYPE="text" NAME="temperature" SIZE="5"> </td></tr>
<tr><td align=center> <font size=2>Sweep Width for acquistion dimemsion (sw in Varian):
</font></td><td align=left><INPUT type="text" name="sw_t4" size="8"> </td></tr>
<tr><td align=center> <font size=2>Number of REAL points in acquisition dimension (np/2 in Varian):
</font></td><td align=left><INPUT type="text" name="pts_t4" size="8"> </td></tr>
<tr><td align=center> <font size=2>Enter the nucleus detected in the acquisition dimension.
</font></td><td align=left>
                <select name="hetnuc_t4">
                <option> none
                <option> proton
                <option> carbon
```

```
<option> nitrogen
<option> phosphorous
</select>          </td></tr>
<tr><td align=center> <font size=2>Sweep Width for first indirect dim. (sw1 in Varian):
</font></td><td align=left><INPUT type="text" name="sw_t3" size="8"></td></tr>
<tr><td align=center> <font size=2>Number of REAL points in first indirect dim. (ni in Varian):
</font></td><td align=left><INPUT type="text" name="pts_t3" size="8"></td></tr>
<tr><td align=center> <font size=2>Enter the nucleus acquired in the first indirect dimension.
</font></td><td align=left>
        <select name="hetnuc_t3">
        <option> none
        <option> proton
        <option> carbon
        <option> nitrogen
        <option> phosphorous
        </select>          </td></tr>
<tr><td align=center> <font size=2>Sweep Width for second indirect dim.(sw2 in Varian):
</font></td><td align=left><INPUT type="text" name="sw_t2" size="8"> </td></tr>
<tr><td align=center> <font size=2>Number of REAL points in second indirect dim. (ni2 in Varian):
</font></td><td align=left><INPUT type="text" name="pts_t2" size="8"> </td></tr>
<tr><td align=center> <font size=2>Enter the nucleus acquired in the second indirect dimension.
</font></td><td align=left>
        <select name="hetnuc_t3">
        <option> none
        <option> proton
        <option> carbon
        <option> nitrogen
        <option> phosphorous
        </select>          </td></tr>
<tr><td align=center> <font size=2>Sweep Width for third indirect dim.(sw2 in Varian):
</font></td><td align=left><INPUT type="text" name="sw_t1" size="8"> </td></tr>
<tr><td align=center> <font size=2>Number of REAL points in second indirect dim. (ni2 in Varian):
</font></td><td align=left><INPUT type="text" name="pts_t1" size="8"> </td></tr>
<tr><td align=center> <font size=2>Enter the nucleus acquired in the second indirect dimension.
</font></td><td align=left>
        <select name="hetnuc_t1">
        <option> none
        <option> proton
        <option> carbon
        <option> nitrogen
        <option> phosphorous
        </select>          </td></tr>
<tr><td align=center> <font size=2>Proton 90 pulse length:
</font></td><td align=left> <INPUT type="text" name="H_length" size="8">   </td></tr>
<tr><td align=center> <font size=2>Proton 90 pulse power level (typically tpwr in Varian):
</font></td><td align=left><INPUT type="text" name="H_power" size="8"> </td></tr>
<tr><td align=center> <font size=2>Enter the recycle delay in seconds (d1 in Varian / RD in Bitter Lab):
</font></td><td align=left><INPUT type="text" name="recycle" size="8"> </td></tr></table>
<hr><center> <h4> Heteronuclei information </h4></center>
<table cellspacing=10>
<tr><td align=center> <font size=2>
Did you use Carbon? <select name="hetnuc_carbon">
        <option> no
        <option> yes
```

```
            </select>  </font></td></tr>
<tr> <td align=center><font size=2>
Did you use Nitrogen? <select name="hetnuc_nitrogen">
            <option> no
            <option> yes
            </select>  </font></td></tr>
<tr> <td align=center><font size=2>
Did you use Phosphorus? <select name="hetnuc_phosphor">
            <option> no
            <option> yes
            </select>  </font></td></tr>
<tr><td align=center> <font size=4>
What transmitter offsets did you use? (List additional offsets in comments)</td></tr></table>
<center><table cellspacing=1><tr><td align=left>Offset in Hz according in "magnet units"</td>
<td align=center>Offset in approximate ppm </td></tr>
<tr><td align=center> <font size=2> Proton (tof in Varian):
</font></td><td align=left><INPUT type="text" name="H_xmit" size="8"> </td>
<td align=left><INPUT type="text" name="H_ppm" size="8"></td></tr>
<tr><td align=center> <font size=2>
Carbon:</font></td><td align=left><INPUT type="text" name="C_xmit" size="8"> </td>
<td align=left><INPUT type="text" name="C_ppm" size="8"></td></tr>
<tr><td align=center> <font size=2>
Nitrogen:  </font></td><td align=left><INPUT type="text" name="N_xmit" size="8"> </td>
<td align=left><INPUT type="text" name="N_ppm" size="8"></td></tr>
<tr><td align=center> <font size=2>
Phosphor:  </font></td><td align=left><INPUT type="text" name="P_xmit" size="8"> </td>
<td align=left><INPUT type="text" name="P_ppm" size="8"></td></tr>
<tr><td align=right> <font size=2>
Did you use deuterium decoupling? (Note: default is no.)</font></td>
<td align=left><INPUT type="radio" name="deut_dec" value="yes"> yes
<INPUT type="radio" name="deut_dec" value="no" checked> no </td></tr></table>
<center>If you used deuterium decoupling, please enter the parameters you used below.
<TEXTAREA NAME="deut_comments" ROWS="3" COLS="60"></TEXTAREA><p>
Enter all pulse lengths in microseconds!<br></center>
<table cellspacing=10><tr><td align=center> <font size=2>
Carbon 90 pulse length:</font></td>
<td align=left><INPUT type="text" name="carb_length" size="8">  </td></tr>
<tr><td align=center> <font size=2>
Carbon 90 pulse power level:</font></td><td align=left>
<INPUT type="text" name="carb_power" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Carbon decoupling pulse length:</font></td>
<td align=left><INPUT type="text" name="Cdec_length" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Carbon decoupling power level:</font></td>
<td align=left><INPUT type="text" name="Cdec_power" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Nitrogen 90 pulse length:</font></td>
<td align=left><INPUT type="text" name="nit_length" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Nitrogen 90 pulse power level:</font></td>
<td align=left><INPUT type="text" name="nit_power" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Nitrogen decoupling pulse length:</font></td>
```

```
<td align=left><INPUT type="text" name="Ndec_length" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Nitrogen decoupling power level:</font></td>
<td align=left><INPUT type="text" name="Ndec_power" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Phosphor 90 pulse length:</font></td>
<td align=left><INPUT type="text" name="phos_length" size="8">    <br>
<tr><td align=center> <font size=2>
Phosphor 90 pulse power level:</font></td>
<td align=left><INPUT type="text" name="phos_power" size="8"> <br>
<tr><td align=center> <font size=2>
Phosphor decoupling pulse length:</font></td>
<td align=left><INPUT type="text" name="Pdec_length" size="8"> </td></tr>
<tr><td align=center> <font size=2>
Phosphor decoupling power level:</font></td>
<td align=left><INPUT type="text" name="Pdec_power" size="8"> </td></tr></table>
<HR><h4> Comments Sections </h4>
<strong><table cellspacing=1><tr><td align=center font=2>
Check here if the comments below include an important general concern!!(Note:default is No)</td>
<td align=left>
<INPUT TYPE="checkbox" name="import_comments" value="yes"> Important general comments included
</td> </tr></table></strong>
<table cellspacing=10> <tr><td align=center font=3><strong>
DO NOT USE ANY COLONS IN THE COMMENTS BOXES!! </strong>
</td></tr></table><br>
Comments regarding any computer problems including hangups, tuning hangups, reboots, and networking.
<TEXTAREA NAME="computer_problems" ROWS="3" COLS="60"></TEXTAREA>
<br clear="all"><br>
Comments. Please indicate why you believe your experiment did not work:<BR>
<TEXTAREA NAME="comments" ROWS="3" COLS="60"></TEXTAREA>
<br clear="all"><br>
Comments. For the future, please enter all other important parameters required
to make this pulse sequence work eg. shaped pulse parameters, mix times, other
transmitter frequencies etc.
<TEXTAREA NAME="parameters" ROWS="6" COLS="60"></TEXTAREA>
<h3>Is all the information correct?  Once the submit button is clicked,
there is no going back!! </h3>
<h3> Did you use any commas??  If so, please go back and delete them!! </h3>
<P><INPUT TYPE=submit value="Submit the Form"> <INPUT TYPE=reset>
</center></FORM></body></html>
```

4D_magnetsort.pl:

This script takes the input from the form page, 4d_nmr.html, and first checks to ensure that all the critical fields have been entered before writing the data to temporary files. The script then sends the data back to the browser for the user to check a final time, before the data is officially submitted to the database.

```
#!/usr/local/bin/perl
#4d NMR perl script to deal with 4D data from the 4d_nmr.html form page
require ("/usr/people/brodsky/cgi-bin/cgi-lib.pl");
&ReadParse;
#handling the heteronuclear options
```

```perl
@hetnuc = split( "\0", $in{'hetnuc'} );
print "Content-type:text/html\n\n";
print "<html> \n";
print "<title>4D NMR response Page</title> \n";
#check to see if all input has been entered.
if ($in{'stndexpt'} eq "OTHER") {
        if( $in{'nonstndexpt'} eq "") {
                print "<h1> \n";
                print "You did not enter a name for your non-standard experiment ... Please go back and
resubmit the form!!";
                print "</h1>";
                exit(0);             }          }
if ($in{'magnet'} eq "None") {
        print "<h1> \n";
        print "You did not enter which magnet you used ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);             }
if ($in{'macro'} eq "") {
        print "<h1> \n";
        print "You did not enter a macro or pulse sequence ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);             }
if ($in{'name'} eq "") {
        print "<h1> \n";
        print "You did not enter your name  ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);             }
if ($in{'login'} eq "") {
        print "<h1> \n";
        print "You did not enter your e-mail address ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);             }
if ($in{'salt'} eq "") {
        print "<h1> \n";
        print "You did not enter the salt concentration ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);             }
if ($in{'concentration'} eq "") {
        print "<h1> \n";
        print "You did not enter the concentration of your sample ... Please go back and resubmit the
form!!";
        print "</h1>";
        exit(0);             }
if ($in{'sample_name'} eq "") {
        print "<h1> \n";
        print "You did not enter the type of sample you used  ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);             }
if ($in{'temperature'} == "\0") {
        print "<h1> \n";
        print "You did not enter a temperature ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);             }
if ($in{'probe'} eq "None") {
```

```perl
        print "<h1> \n";
        print "You did not enter any probe information ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);              }
if($in{'hetnuc_t3'} eq "none") {
        print "<h1> \n";
        print "You did not enter a nucleus in the acquisition dimension... Please go back and resubmit the
form!! \n";
        print "</h1> " ;
        exit(0);              }
if ($in{'sw_t4'} == "\0") {
        print "<h1> \n";
        print "You did not enter any sweep width in the acquisition dim. ... Please go back and resubmit
the form!!";
        print "</h1>";
        exit(0);              }
if ($in{'pts_t4'} == "\0") {
        print "<h1> \n";
        print "You did not enter any points in the acquisition dim. ... Please go back and resubmit the
form!!";
        print "</h1>";
        exit(0);              }
if($in{'hetnuc_t3'} eq "none") {
        print "<h1> \n";
        print "You did not enter a nucleus in the first indirect dimension... Please go back and resubmit the
form!! \n";
        print "</h1> " ;
        exit(0);              }
if ($in{'sw_t3'} == "\0") {
        print "<h1> \n";
        print "You did not enter any sweep width in the first indirect dim. ... Please go back and resubmit
the form!!";
        print "</h1>";
        exit(0);              }
if ($in{'pts_t3'} == "\0") {
        print "<h1> \n";
        print "You did not enter any points in the first indirect dim. ... Please go back and resubmit the
form!!";
        print "</h1>";
        exit(0);              }
if($in{'hetnuc_t2'} eq "none") {
        print "<h1> \n";
        print "You did not enter a nucleus in the second indirect dimension... Please go back and resubmit
the form!! \n";
        print "</h1> " ;
        exit(0);              }
if ($in{'sw_t2'} == "\0") {
        print "<h1> \n";
        print "You did not enter a sweep width in the second indirect dim. ... Please go back and resubmit
the form!!";
        print "</h1>";
        exit(0);              }
if ($in{'pts_t2'} == "\0") {
        print "<h1> \n";
```

```
        print "You did not enter any points in the second indirect dim. ... Please go back and resubmit the
form!!";
        print "</h1>";
        exit(0);              }
if($in{'hetnuc_t1'} eq "none") {
        print "<h1> \n";
        print "You did not enter a nucleus in the third indirect dimension... Please go back and resubmit
the form!! \n";
        print "</h1> " ;
        exit(0);              }
if ($in{'sw_t1'} == "\0")  {
        print "<h1> \n";
        print "You did not enter a sweep width in the third indirect dim. ... Please go back and resubmit
the form!!";
        print "</h1>";
        exit(0);              }
if ($in{'pts_t1'} == "\0")  {
        print "<h1> \n";
        print "You did not enter any points in the third indirect dim. ... Please go back and resubmit the
form!!";
        print "</h1>";
        exit(0);              }
if ($in{'H_length'} == "\0")  {
        print "<h1> \n";
        print "You did not enter a 90 proton pulse length ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);              }
if ($in{'H_power'} == "\0")  {
        print "<h1> \n";
        print "You did not enter a 90 proton power level ... Please go back and resubmit the form!!";
        print "</h1>";
        exit(0);              }
if( $in{'het_nuc'} eq "carbon" ) {
        if ($in{'C_xmit'} == "\0")  {
                print "<h1> \n";
                print "You used carbon and did not enter a carbon transmitter offset ... Please go back and
resubmit the form!!";
                print "</h1>";
                exit(0);              }
        if ($in{'carb_length'} == "\0")  {
                print "<h1> \n";
                print "You used carbon and did not enter a carbon 90 pulse length ... Please go back and
resubmit the form!!";
                print "</h1>";
                exit(0);              }
        if ($in{'carb_power'} == "\0")  {
                print "<h1> \n";
                print "You used carbon and did not enter a carbon 90 pulse power level. Please go back
and resubmit the form!!";
                print "</h1>";
                exit(0);              }          }
if( $in{'het_nuc'} eq "nitrogen" ) {
        if ($in{'N_xmit'} == "\0")  {
                print "<h1> \n";
```

```
                    print "You used nitrogen and did not enter a nitrogen transmitter offset ... Please go back
and resubmit the form!!";
                    print "</h1>";
                    exit(0);                    }
        if ($in{'nit_length'} == "\0")  {
                    print "<h1> \n";
                    print "You used nitrogen and did not enter a nitrogen 90 pulse length ... Please go back
and resubmit the form!! </h1>";
                    exit(0);                    }
        if ($in{'nit_power'} == "\0")  {
                    print "<h1> \n";
                    print "You used nitrogen and did not enter a nitrogen 90 pulse power level. Please go
back and resubmit the form!! </h1>";
                    exit(0);               }           }
if( $in{'het_nuc'} eq "phosphor" ) {
        if ($in{'C_xmit'} == "\0")  {
                    print "<h1> \n";
                    print "You used phosphor and did not enter a phosphor transmitter offset ... Please go
back and resubmit the form!! </h1>";
                    exit(0);              }
        if ($in{'phos_length'} == "\0")  {
                    print "<h1> \n";
                    print "You used phosphor and did not enter a phosphor 90 pulse length ... Please go back
and resubmit the form!! </h1>";
                    exit(0);               }
        if ($in{'phos_power'} == "\0")  {     print "<h1> \n";
                    print "You used phosphor and did not enter a phosphor 90 pulse power level. Please go
back and resubmit the form!!   </h1>";
                    exit(0);             }            }
print <<END_OF_MESSAGE;
<h1>Thank you $in{'name'} for filling out the 4D parameter form!</h1>
END_OF_MESSAGE
printf "4D EXPT <br> \n";
printf "USER: $in{'name'} <br> \n";
printf "Magnet: $in{'magnet'}  <br> \n";
printf "Start Date: $in{'date'} $in{'number'},$in{'year'}.  <br> \n";
printf "Finish Date: $in{'date_fin'} $in{'number_fin'},$in{'year_fin'}  <br> \n";
if($in{'stndexpt'} ne OTHER) {
          printf "4D Experiment: $in{'stndexpt'}  <br> \n";                      }
else {    printf "4D Experiment: $in{'nonstndexpt'}  <br> \n";                   }
printf "Macro/Pulse Sequence: $in{'macro'} <br> \n";
printf "Probe: $in{'probe'}  <br> \n";
printf "Sample: $in{'sample_name'}  <br> \n";
printf "Type of Sample: $in{'sample_label'}  <br> \n";
printf "Concentration of Sample: $in{'concentration'}  <br> \n";
printf "Labelling Pattern: $in{'label'}  <br> \n";
printf "Salt Concentration: $in{'salt'}  <br> \n";
printf "Temperature: $in{'temperature'}  <br> \n";
if($in{'success'} eq "successful") {
          printf "Success: YES  <br> \n";                    }
else    { printf "Success: NO  <br> \n";                     }
printf "Proton 90: $in{'H_length'} <br> \n";
printf "Proton 90 power: $in{'H_power'} <br> \n";
printf "Recycle Delay: $in{'recycle'} <br> \n";
```

239

```
printf "Acquisition detected $in{'hetnuc_t3'} with ";
printf "SW(t4): $in{'sw_t4'} \nPTS(t3):$in{'pts_t4'}    <br> \n";
printf "First Indirect dimension detected $in{'hetnuc_t3'} with ";
printf "SW(t3): $in{'sw_t3'} \nPTS(t3):$in{'pts_t3'}    <br> \n";
printf "Second indirect dimension detected $in{'hetnuc_t2'} with ";
printf "SW(t2): $in{'sw_t2'} \nPTS(t1):$in{'pts_t2'}    <br> \n";
printf "Third indirect dimension detected $in{'hetnuc_t1'} with ";
printf "SW(t1): $in{'sw_t1'} \nPTS(t1):$in{'pts_t1'}    <br> \n";
printf "PROTON_XMIT: $in{'H_xmit'} which is at $in{'H_ppm'} ppm   <br> \n";
if ( $in{'deut_dec'} eq "no" ) {       printf "Deuterium Decoupling: NO  <br> \n";                   }
else {    printf "Deuterium Decoupling: YES  <br> \n";
          printf "Deuterium Decoupling Parameters: \n $in{'deut_comments'} <br> \n";          }
if($in{'hetnuc_carbon'} eq "yes") {
              printf "Heteronucleus: CARBON <br> \n";
              printf "CARBON_XMIT: $in{'C_xmit'} which is at $in{'C_ppm'} ppm<br> \n";
              printf "CARBON 90: $in{'carb_length'}  <br>\nCARBON 90 POWER:
$in{'carb_power'} <br> \n";
              printf "CARBON DECOUPLING 90: $in{'Cdec_length'}  <br>\nCARBON DEC 90
POWER: $in{'Cdec_power'} <br> \n";                 }
      if($in{'hetnuc_nitrogen'} eq "yes") {
              printf "Heteronucleus: NITROGEN <br> \n";
              printf "NITROGEN_XMIT: $in{'N_xmit'} which is at $in{'N_ppm'} ppm<br> \n";
              printf "NITROGEN 90:$in{'nit_length'}  <br> \nNITROGEN 90
POWER:$in{'nit_power'}  <br> \n";
              printf "NITROGEN DECOUPLING 90:$in{'Ndec_length'}  <br> \nNITROGEN DEC
90 POWER:$in{'Ndec_power'}  <br> \n";              }
      if($in{'hetnuc_phosphor'} eq "yes") {
              printf "Heteronucleus: PHOSPHOR <br> \n";
              printf "PHOSPHOR_XMIT:$in{'P_xmit'} which is at $in{'P_ppm'} ppm<br> \n";
              printf "PHOSPHOR 90:$in{'phos_length'}  <br>\nPHOSPHOR 90
POWER:$in{'phos_power'} <br> \n";
              printf "PHOSPHOR DECOUPLING 90:$in{'Pdec_length'}  <br>\nPHOSPHOR DEC
90 POWER:$in{'Pdec_power'} <br> \n";             }
if ($in{'import_comments'} eq "yes" ) {
    printf "<strong> You have a critical GENERAL concern in the comments below </strong> <br> \n";
                      }
printf "COMPUTER COMMENTS: $in{'computer_problems'}<br> \n";
printf "SUCCESS COMMENTS: $in{'comments'} <br> \n";
printf "PARAMETER COMMENTS: $in{'parameters'}  <br> \n";
printf "<h3> Is this information correct? <br> \n";
printf "If this information is correct click the submit button, if not return to the form page using the
netscape back button and correct the wrong information! </h3> \n";
printf "<strong>Any problems or questions please contact Alex. Bugs/problems will not be fixed if not
reported! ThankYou! </strong> \n";
printf "<a href=mailto:brodsky@mit.edu>brodsky@mit.edu</a> \n";
printf "<form ACTION=http://volvox.mit.edu/cgi-bin/4D_submit.pl> \n";
print <<END_OF_MESSAGE;
<input type="radio" name="tmp_file" value="$in{'name'}_tmpfile" checked> <br>
END_OF_MESSAGE
printf '<P><INPUT TYPE=submit value="FINAL Submit to the Database">';
printf " \n </FORM> \n   </html> \n";
#to write to temporary data file to read if data is actually submitted to database
open(DATAFILE, "> /usr/ns-home/docs/labonly/NMR/$in{'name'}_tmpfile");
flock (DATAFILE, 2);
```

```
#printf DATAFILE "4D EXPT \n";
printf DATAFILE "USER: $in{'name'},";
printf DATAFILE "$in{'login'},";
printf DATAFILE "Magnet: $in{'magnet'}, ";
printf DATAFILE "Start Date: $in{'date'} $in{'number'} $in{'year'}, ";
printf DATAFILE "Finish Date: $in{'date_fin'} $in{'number_fin'} $in{'year_fin'},";
printf DATAFILE "Probe: $in{'probe'},";
if($in{'stndexpt'} ne OTHER) {
        printf DATAFILE "4D Experiment: $in{'stndexpt'},";              }
else {    printf DATAFILE "4D Experiment: $in{'nonstndexpt'},";              }
printf DATAFILE "Macro/Pulse Sequence: $in{'macro'},";
printf DATAFILE "Probe: $in{'probe'},";
printf DATAFILE "Sample: $in{'sample_name'},";
printf DATAFILE "Type of Sample: $in{'sample_label'},";
printf DATAFILE "Concentration of Sample: $in{'concentration'},";
printf DATAFILE "Labelling Pattern: $in{'label'},";
printf DATAFILE "Salt Concentration: $in{'salt'},";
printf DATAFILE "Temperature: $in{'temperature'},";
if($in{'success'} eq "successful") {   printf DATAFILE "Success: YES,";              }
else    { printf DATAFILE "Success: NO,";              }
printf DATAFILE "Proton 90: $in{'H_length'},";
printf DATAFILE "Proton 90 power: $in{'H_power'},";
printf DATAFILE "Recycle Delay:$in{'recycle'},";
printf DATAFILE "$in{'hetnuc_t4'},";
printf DATAFILE "SW(t3):$in{'sw_t4'},PTS(t3):$in{'pts_t4'},";
printf DATAFILE "$in{'hetnuc_t3'},";
printf DATAFILE "SW(t3):$in{'sw_t3'},PTS(t3):$in{'pts_t3'},";
printf DATAFILE "$in{'hetnuc_t2'},";
printf DATAFILE "SW(t2): $in{'sw_t2'},PTS(t2):$in{'pts_t2'},";
printf DATAFILE "$in{'hetnuc_t1'},";
printf DATAFILE "SW(t1): $in{'sw_t1'},PTS(t1):$in{'pts_t1'},";
printf DATAFILE "PROTON_XMIT:$in{'H_xmit'},$in{'H_ppm'},";
if ( $in{'deut_dec'} eq "no" ) {
        printf DATAFILE "Deuterium Decoupling: NO,";
        #printf DATAFILE "Deuterium Decoupling Parameters: None,";              }
else {    printf DATAFILE "Deuterium Decoupling: YES,";
        #printf DATAFILE "Deuterium Decoupling Parameters:  $in{'deut_comments'},";         }
        if($in{'hetnuc_carbon'} eq "yes") {
                printf DATAFILE "Carbon,";              }
        else {    printf DATAFILE "No,";              }
        if($in{'hetnuc_nitrogen'} eq "yes") {
                printf DATAFILE "Nitrogen,";              }
                else {    printf DATAFILE "No,";              }
        if($in{'hetnuc_phosphor'} eq "yes") {       printf DATAFILE "Phosphor,";              }
        else {    printf DATAFILE "No,";              }
                printf DATAFILE "CARBON_XMIT: $in{'C_xmit'},$in{'C_ppm'},";
                printf DATAFILE "CARBON 90: $in{'carb_length'},  CARBON 90 POWER:
$in{'carb_power'},";
                printf DATAFILE "CARBON DECOUPLING 90: $in{'Cdec_length'},  CARBON DEC
90 POWER: $in{'Cdec_power'},";
                printf DATAFILE "NITROGEN_XMIT: $in{'N_xmit'},$in{'N_ppm'},";
                printf DATAFILE "NITROGEN 90:$in{'nit_length'},  NITROGEN 90
POWER:$in{'nit_power'},";
```

```
                printf DATAFILE "NITROGEN DECOUPLING 90:$in{'Ndec_length'}, NITROGEN
DEC 90 POWER:$in{'Ndec_power'},";
                printf DATAFILE "PHOSPHOR_XMIT:$in{'P_xmit'},$in{'P_ppm'},";
                printf DATAFILE "PHOSPHOR 90:$in{'phos_length'}, PHOSPHOR 90
POWER:$in{'phos_power'},";
                printf DATAFILE "PHOSPHOR DECOUPLING 90:$in{'Pdec_length'}, PHOSPHOR
DEC 90 POWER:$in{'Pdec_power'},";
if ($in{'import_comments'} eq "yes" ) {    printf DATAFILE "yes,";              }
else {    printf DATAFILE "no,";              }
flock (DATAFILE, 8);
close DATAFILE;
open(DATAFILE, "> /usr/ns-home/docs/labonly/NMR/$in{'name'}_tmpfilecom");
flock (DATAFILE, 2);
printf DATAFILE "COMPUTER COMMENTS:$in{'computer_problems'}";
flock (DATAFILE, 8);
close DATAFILE;
open(DATAFILE, "> /usr/ns-home/docs/labonly/NMR/$in{'name'}_tmpfilecom1");
flock (DATAFILE, 2);
printf DATAFILE "SUCCESS COMMENTS:$in{'comments'}";
flock (DATAFILE, 8);
close DATAFILE;
open(DATAFILE, "> /usr/ns-home/docs/labonly/NMR/$in{'name'}_tmpfilecom2");
flock (DATAFILE, 2);
printf DATAFILE "PARAMETER COMMENTS: $in{'parameters'}";
flock (DATAFILE, 8);
close DATAFILE;
open(DATAFILE, "> /usr/ns-home/docs/labonly/NMR/$in{'name'}_tmpfilecom3");
flock (DATAFILE, 2);
printf DATAFILE "Deuterium Decoupling Parameters:$in{'deut_comments'}";
flock (DATAFILE, 8);
close DATAFILE;
#RECORD_ITEM
```

4D_submit.pl: This script reads the temporary file and writes the data to the different files for permanent record. The data is sorted by the dimension of the experiment, by the magnet it was run on, and to a master list of all experiments. Other sorting by type of experiment like NOESY or through-bond can be added. Also, the script send mail to the user and to Jamie for the user's records.

```
#!/usr/local/bin/perl
#4d NMR perl script to deal with 4D data from the 4D_response.html form page for FINAL
#submission to the database
require ("/usr/people/brodsky/cgi-bin/cgi-lib.pl");
&ReadParse;
print <<END_OF_MESSAGE;
<title>4D NMR Submission Response</title>
<h1> Thank you for submitting your data to the Williamson Lab NMR database </h1>
<h3>
<a href=http://volvox.mit.edu/labonly/NMR/nmr_master.html>Return to the main NMR page</a> </h3>
END_OF_MESSAGE
#read the temp file to get the information enterer by the user
open (input,"/usr/ns-home/docs/labonly/NMR/$in{'tmp_file'}");
while(<input>) {
($user, $login, $magnet, $start, $finish, $probe, $expt_name, $macro, $probe, $sample, $type, $conc,
$label, $salt, $temperature, $success, $proton_90, $proton_power, $recycle,$hetnuc_t4, $sw_t4,$pts_t4,
```

```
$hetnuc_t3,$sw_t3, $pts_t3, $hetnuc_t2,
$sw_t2, $pts_t2, $hetnuc_t1, $sw_t1, $pts_t1, $proton_xmit, $proton_ppm,$deut_no,
$het_carbon,$het_nit,$het_phos,$carbon_xmit,$carb_ppm,$carb_90,$carb_power,$carb_dec90,
$carb_decpower,
$nit_xmit,$nit_ppm,$nit_90,$nit_power,$nit_dec90,$nit_decpower,$phos_xmit,$phos_ppm,$phos_90,
$phos_power,$phos_dec90,$phos_decpower,$import_concern) = split(/,/, $_);
} #end of while input
close(input);
$/="";   #enable paragraph mode
$*=1;    #enable multi line patterns
open(com,"/usr/ns-home/docs/labonly/NMR/$in{'tmp_file'}com");
while(<com>) {
($computer_comment,$ccom)=split(/:/,$_);              }
close(com);
open(com1,"/usr/ns-home/docs/labonly/NMR/$in{'tmp_file'}com1");
while(<com1>) {
($success_comments,$scom)=split(/:/,$_);              }
close(com1);
open(com2,"/usr/ns-home/docs/labonly/NMR/$in{'tmp_file'}com2");
while(<com2>) {
($parameter_comments,$pcom)=split(/:/,$_);                 }
close(com2);
open(com3,"/usr/ns-home/docs/labonly/NMR/$in{'tmp_file'}com3");
while(<com3>) {
($deut_comments,$deutcom)=split(/:/,$_);                   }
close(com3);
#to establish the count of all experiments
open(input1,total_exptcount);
        while (<input1>) {
                ($count) = split(/\s+/,$_);
                $count= $count + 1;
                #print "<strong>You performed experiment number $count. </strong> ";          }
close(input1);
open(output, "> /usr/people/brodsky/cgi-bin/total_exptcount");
flock (output, 2);
printf output "$count";
flock (output, 8);
close(output);
#send mail to the submitter
$mailer=$login;
&write_mail($mailer);
#send mail to Jamie
$mailer=jrwill;
&write_mail($mailer);
#send mail to Kwaku
$mailer=kwaku;
&write_mail($mailer);
sub write_mail {
local($mailer) = @_;
#routine to send mail
#to send mail upon submission
open (MAIL, "| /usr/lib/sendmail -oi -n -t" );
print MAIL <<MAIL_MESSAGE;
To:$mailer
```

```
From:NMR database
MAIL_MESSAGE
printf MAIL "4D EXPT \n";
printf MAIL "Experiment number: $count \n";
printf MAIL "$user \n";
printf MAIL "$magnet \n";
printf MAIL "$start \n";
printf MAIL "$finish \n";
printf MAIL "$expt_name \n";
printf MAIL "$macro \n";
printf MAIL "$probe \n";
printf MAIL "$sample \n";
printf MAIL "$type \n";
printf MAIL "$conc \n";
printf MAIL "$label \n";
printf MAIL "$salt \n";
printf MAIL "$temperature \n";
printf MAIL "$success\n";
printf MAIL "$recycle \n";
printf MAIL "Acquisition nucleus:$hetnuc_t4 \n";
printf MAIL  "$sw_t4 \n";
printf MAIL  "$pts_t4 \n";
printf MAIL  "First indirect nucleus:$hetnuc_t3 \n";
printf MAIL "$sw_t3 \n";
printf MAIL "$pts_t3 \n";
printf MAIL  "Second indirect nucleus:$hetnuc_t2 \n";
printf MAIL "$sw_t2 \n";
printf MAIL "$pts_t2 \n";
printf MAIL  "Third indirect nucleus:$hetnuc_t1 \n";
printf MAIL "$sw_t1 \n";
printf MAIL "$pts_t1 \n";
printf MAIL "$proton_90 \n";
printf MAIL "$proton_power \n";
printf MAIL "$proton_xmit at $proton_ppm ppm \n";
if ( $deut_no eq "Deuterium Decoupling: NO" ) {
        printf MAIL "Deuterium Decoupling: NO  \n";                 }
else {
        printf MAIL "Deuterium Decoupling: YES  \n";
        printf MAIL "$deut_comments$deutcom \n";                    }
        if($het_carbon ne "No") {
                printf MAIL "Heteronucleus: CARBON \n";
                printf MAIL "$carb_xmit at $carb_ppm ppm\n";
                printf MAIL "$carb_90 \n$carb_power \n";
                printf MAIL "$carb_dec90\n$carb_decpower \n";               }
        if($het_nit ne "No") {
                printf MAIL "Heteronucleus: NITROGEN \n";
                printf MAIL "$nit_xmit at $nit_ppm ppm\n";
                printf MAIL "$nit_90\n$nit_power \n";
                printf MAIL "$nit_dec90\n$nit_decpower \n";                 }
        if($het_phos ne "No") {
                printf MAIL "Heteronucleus: PHOSPHOR \n";
                printf MAIL "$phos_xmit at $phos_ppm ppm\n";
                printf MAIL "$phos_90 \n$phos_power \n";
                printf MAIL "$phos_dec90\n$phos_decpower \n";               }
```

```perl
if ($import_concern eq "yes" ) {
        printf MAIL "GENERAL CONCERN LISTED IN COMMENTS \n";              }
printf MAIL "$computer_comment:$ccom\n";
printf MAIL "$success_comments:$scom \n";
printf MAIL "$parameter_comments:$pcom \n";
close MAIL;                 }
# determine the index and database file names to write to
# write to files for all 4D experiments tracking
$index="/disks/work0/brodsky/webpages/NMR/4D/all_index4D.html";
$data_file ="/disks/work0/brodsky/webpages/NMR/4D/all_database4D.html";
$data_link ="all_database4D.html";
#call the index subroutine
&write_index($index,$data_link);
#call subroutine to write to a database file
&write_database($data_file);
#write to files for all experiments
$index="/disks/work0/brodsky/webpages/NMR/all_index.html";
$data_file ="/disks/work0/brodsky/webpages/NMR/all_database.html";
$data_link ="all_database.html";
#call the index subroutine
&write_index($index,$data_link);
#call subroutine to write to a database file
&write_database($data_file);
#write to file if experiment was run on didinium
if( $magnet eq "Magnet: Didinium Varian 600" ) {
        $index="/disks/work0/brodsky/webpages/NMR/4D/didinium_index4D.html";
        $data_file ="/disks/work0/brodsky/webpages/NMR/4D/didinium_database4D.html";
        $data_link ="didinium_database4D.html";
        #call the index subroutine
        &write_index($index,$data_link);
        #call subroutine to write to a database file
        &write_database($data_file);            }
elsif ( $magnet eq "Magnet: Bitter Lab 591" ) {
        $index="/disks/work0/brodsky/webpages/NMR/4D/bitter591_index4D.html";
        $data_file ="/disks/work0/brodsky/webpages/NMR/4D/bitter591_database4D.html";
        $data_link ="bitter591_database4D.html";
        #call the index subroutine
        &write_index($index,$data_link);
        #call subroutine to write to a database file
        &write_database($data_file);    }
elsif ( $magnet eq "Magnet: Bitter Lab 750" ) {
        $index="/disks/work0/brodsky/webpages/NMR/4D/bitter750_index4D.html";
        $data_file ="/disks/work0/brodsky/webpages/NMR/4D/bitter750_database4D.html";
        $data_link ="bitter750_database4D.html";
        #call the index subroutine
        &write_index($index,$data_link);
        #call subroutine to write to a database file
        &write_database($data_file);            }
elsif ( $magnet eq "Magnet: Bitter Lab 500" ) {
        $index="/disks/work0/brodsky/webpages/NMR/4D/bitter500_index4D.html";
        $data_file ="/disks/work0/brodsky/webpages/NMR/4D/bitter500_database4D.html";
        $data_link ="bitter500_database4D.html";
        #call the index subroutine
        &write_index($index,$data_link);
```

```
            #call subroutine to write to a database file
            &write_database($data_file);            }
elsif ( $magnet eq "Magnet: Wagner Varian 750")  {
            $index="/disks/work0/brodsky/webpages/NMR/4D/wagner750_index4D.html";
            $data_file ="/disks/work0/brodsky/webpages/NMR/4D/wagner750_database4D.html";
            $data_link ="wagner750_database4D.html";
            #call the index subroutine
            &write_index($index,$data_link);
            #call subroutine to write to a database file
            &write_database($data_file);                    }
elsif ( $magnet eq "Magnet: Other" )  {
            $index="/disks/work0/brodsky/webpages/NMR/4D/other_index4D.html";
            $data_file ="/disks/work0/brodsky/webpages/NMR/4D/other_database4D.html";
            $data_link ="other_database4D.html";
            #call the index subroutine
            &write_index($index,$data_link);
            #call subroutine to write to a database file
            &write_database($data_file);            }
#subroutine to write an index file
sub write_index  {   local($index_file,$data_link) = @_;
open(DATAFILE20, ">> $index_file");
flock (DATAFILE20, 2);
#underscores to delimit records
printf DATAFILE20 "_____ <br>\n";
printf DATAFILE20 "<a href=$data_link#$count> \n";
printf DATAFILE20 "<h3>4D EXPT </h3> \n ";
printf DATAFILE20 "4D EXPT \n";
printf DATAFILE20 "Experiment number: $count \n";
printf DATAFILE20 "$user \n";
printf DATAFILE20 "$magnet \n";
printf DATAFILE20 "$finish  \n";
printf DATAFILE20 "$expt_name \n";
printf DATAFILE20 " $success \n";
if ($import_concern eq "yes" )  {
            printf DATAFILE20 "GENERAL CONCERN LISTED IN COMMENTS <br> \n";   }
printf DATAFILE20 "</a> \n";
printf DATAFILE <<END_OF_MESSAGE;
<strong><a href="#top_of_page">Return to the top of this page</a> </strong>
END_OF_MESSAGE
#RECORD_ITEM
flock (DATAFILE20, 8);
close DATAFILE20;                    }
#subroutine to write a database file that is passed to it.
sub write_database  {    local($database_file) = @_;
open(DATAFILE, ">> $database_file");
flock (DATAFILE, 2);
#underscores to delimit records
printf DATAFILE "_____ <br>\n";
printf DATAFILE "<a name = $count > \n";
printf DATAFILE "<h3>4D EXPT </h3>\n";
printf DATAFILE "</a> \n";
printf DATAFILE "Experiment number: $count <br>\n";
printf DATAFILE "$user <br>\n";
printf DATAFILE "$magnet  <br>\n";
```

```
printf DATAFILE "$start.  <br>\n";
printf DATAFILE "$finish  <br>\n";
printf DATAFILE "$expt_name  <br>\n";
printf DATAFILE "$macro <br>\n";
printf DATAFILE "$probe <br> \n";
printf DATAFILE "$sample <br> \n";
printf DATAFILE "$type <br> \n";
printf DATAFILE "$conc  <br>\n";
printf DATAFILE "$label <br> \n";
printf DATAFILE "$salt <br> \n";
printf DATAFILE "$temperature deg C<br> \n";
printf DATAFILE "$success  <br> \n";
printf DATAFILE "$recycle seconds<br> \n";
printf  DATAFILE "Acquisition detected $hetnuc_t4 with ";
printf  DATAFILE "<br>$sw_t4 \n$pts_t4    <br> \n";
printf  DATAFILE "First Indirect dimension detected $hetnuc_t3 with ";
printf  DATAFILE "<br>$sw_t3 \n$pts_t3   <br> \n";
printf  DATAFILE "Second indirect dimension detected $hetnuc_t2 with ";
printf  DATAFILE "<br>$sw_t2 \n$pts_t2    <br> \n";
printf  DATAFILE "Third indirect dimension detected $hetnuc_t1 with ";
printf  DATAFILE "<br>$sw_t1 \n$pts_t1    <br> \n";
printf DATAFILE "$proton_xmit at $proton_ppm ppm<br> \n";
if ( $deut_no eq "Deuterium Decoupling: NO" ) {
        printf DATAFILE "Deuterium Decoupling: NO  <br>\n";   }
else {
        printf DATAFILE "Deuterium Decoupling: YES <br> \n";
        printf DATAFILE "$deut_comments:<pre>$deutcom <br>\n";   }
        if($het_carbon ne "No") {
                printf DATAFILE "Heteronucleus: CARBON <br>\n";
                printf DATAFILE "$carb_xmit at $carb_ppm ppm<br> \n";
                printf DATAFILE "$carb_90 \n$carb_power <br>\n";
                printf DATAFILE "$carb_dec90\n$carb_decpower <br> \n";   }
        if($het_nit ne "No") {
                printf DATAFILE "Heteronucleus: NITROGEN <br> \n";
                printf DATAFILE "$nit_xmit at $nit_ppm ppm<br>\n";
                printf DATAFILE "$nit_90\n$nit_power  <br>\n";
                printf DATAFILE "$nit_dec90\n$nit_decpower <br> \n";    }
        if($het_phos ne "No") {
                printf DATAFILE "Heteronucleus: PHOSPHOR <br> \n";
                printf DATAFILE "$phos_xmit at $phos_ppm ppm<br>\n";
                printf DATAFILE "$phos_90 \n$phos_power  <br>\n";
                printf DATAFILE "$phos_dec90\n$phos_decpower <br> \n";                   }
if ($import_concern eq "yes" ) {
        printf DATAFILE "GENERAL CONCERN LISTED IN COMMENTS <br> \n";   }
printf DATAFILE "$computer_comment:$ccom <br>\n";
printf DATAFILE "$success_comments:$scom <br>\n";
printf DATAFILE "$parameter_comments:$pcom <br>\n";
printf DATAFILE <<END_OF_MESSAGE;
<strong><a href="#top_of_page">Return to the top of this page</a> <br></strong>
END_OF_MESSAGE
#RECORD_ITEM
flock (DATAFILE, 8);
close DATAFILE;       }
```

References

Aboul-ela, F., Karn, J. & Varani, G. (1995). The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J. Mol. Biol.* **253**, 313-332.

Addess, K. J., Basilion, J. P., Klausner, R. D., Rouault, T. A. & Pardi, A. (1997). Structure and Dynamics of the Iron Responsive Element RNA: Implications for Binding of the RNA by Iron Regulatory Binding Proteins. *Journal of Molecular Biology* **274**, 72-83.

Akke, M., Fiala, R., Patel, D. & Palmer, A. G. (1997). Base Dynamics in a UUCG Tetraloop RNA Hairpin Characterized by ^{15}N Spin Relaxation: Correlations with Structure and Stability. *RNA* **3**, 702-9.

Akke, M. & Palmer, A. G. (1996). Monitoring Macromolecular Motions on Microsecond to Millisecond Time Scales by $R_{1\rho}$-R_1 Constant Relaxation Time NMR Spectroscopy. *Journal of the American Chemical Society* **118**, 911-912.

Allain, F. H.-T., Gubser, C. C., Howe, P. W. A., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of Ribonucleoprotein Interaction Determined by RNA Folding during Complex Formation. *Nature* **380**, 646-650.

Allain, F. H.-T. & Varani, G. (1995). Structure of the P1 Helix from Group I Self-Splicing Introns. *J. Mol. Biol.* **250**, 333-353.

Allain, F. H.-T. & Varani, G. (1997). How Accurately and Precisely can RNA Structure be Determined by NMR? *Journal of Molecular Biology* **267**, 338-351.

Alonso, A., Derse, D. & Peterlin, B. M. (1992). Human Chromosome 12 is Required for Optimal Interactions between Tat and TAR of Human Immunodeficiency Virus Type 1 in Rodent Cells. *Journal of Virology* **66**, 4617-4621.

Baskerville, S., Zapp, M. & Ellington, A. (1995). High-Resolution Mapping of the Human T-cell Leukemia Virus type 1 Rex-Binding Element by in vitro Selection. *Journal of Virology* **69**, 7559-69.

Batey, R. T., Battiste, J. L. & Williamson, J. R. (1995). Preparation of Isotopically Enriched RNAs for Heteronuclear NMR. In *Methods in Enzymology* (James, T. L., ed.), Vol. 261, pp. 300-322. Academic Press, San Diego.

Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D. & Williamson, J. R. (1992). Preparation of Isotopically Labeled Ribonucleotides for Multidimensional NMR Spectroscopy of RNA. *Nucleic Acids Research* **20**(17), 4515-4523.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). α-helix-RNA Major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex. *Science* **273**, 1547-1551.

Battiste, J. L., Tan, R., Frankel, A. D. & Williamson, J. R. (1995). Assignment and Modeling of the Rev Response Element RNA Bound to a Rev Peptide using ^{13}C-Heteronuclear NMR. *J. Biomolecular NMR* **6**, 375-389.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schmeider, B. (1992). The Nucleic Acid Database: a Guide to the Use of a Relational Database of Nucleic Acid Crystal Structures. *Biophys. J.* **63**, 751-759.

Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. (1994). The 2.9 Å Crystal Structure of T. *thermophilus* Seryl-tRNA Synthetase Complexed with tRNA[SER]. *Science* **263**, 1404-1432.

Blackburn, E. H. (1993) Telomerase. In *The RNA World* (Gesteland, R. F. And Atkins, J. F., ed.), pp 557-576. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Brodsky, A. S. & Williamson, J. R. (1997). Structure of HIV-2 TAR Argininamide Complex. *J. Mol. Biol.* **30**, 624-639.

Brunger, A. T. (1992). *X-PLOR User Manual*, Yale University, New Haven.

Brunger, A. T. & Karplus, M. (1991). Molecular Dynamics Simlutations with Experimental Restraints. *Accounts of Chemical Restrainst* **24**, 54-61.

Buck, M., Schwalbe, H. & Dobson, C. M. (1995). Main-chain Dynamics of a Parially Folded Protein: 15N NMR Relaxation Measurements of Hen Egg White Lysozyme Denatured in Trifluoroethanol. *Journal of Molecular Biology* **257**, 669-683.

Cai, Z. & Tinoco, I. (1996). Solution Structure of Loop A from the Hairpin Ribozyme from Tobacco Ringspot Virus Satellite. *Biochemistry* **35**(19), 6026-6036.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991). Arginine-Mediated RNA Recognition: the Arginine Fork. *Science* **252**, 1167-1171.

Cate, J. H. & Doudna, J. A. (1996). Metal-Binding Sites in the Major Groove of a Large Ribozyme Domain. *Structure* **4**, 1221-1229.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kondrot, C. E., Cech, T. R. & Doudna, J. A. (1996). Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science* **273**, 1678-1685.

Cate, J. H., Hanna, R. L. & Doudna, J. A. (1997). A Magnesium Core at the Heart of a Ribozyme Domain. *Nature Structural Biology* **4**, 553-558.

Cech, T. R. (1993) Structure and Mechanism of the Large Catalytic RNAs: Group I and Group II Introns and Ribonuclease P. In *The RNA World* (Gesteland, R. F. And Atkins, J. F., ed.), pp 239-270. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Chang, Y. N. & Jeang, K. T. (1992). The Basic RNA-Binding Domain of HIV-2 Tat Contributes to Preferential trans-Activation of a TAR2-Containing-LTR. *Nucleic Acids Research* **20**, 5465-72.

Churcher, M. J., Lamont, C., Hamy, F., Dingwall, C., Green, S. M., Lowe, A. D., Butler, P. J. G., Gait, M. J. & Karn, J. (1993). High Affinity Binding of TAR RNA by the Human Immunodeficiency Virus Type-1 *tat* Protein Requires a Base-Pairs in the RNA Stem and Amino Acid Residues Flanking the Basic Region. *J. Mol. Biol.* **230**, 90-110.

Colvin, R. A., White, S. W., Garcia-Bianco, M. A. & Hoffman, D. W. (1993). Structural features of an RNA containing the CUGGGA loop of the human immunodeficiency virus type 1 trans-activating response element. *Biochemistry* **32**, 1105-1112.

Cordingly, M. G., LaFemina, R. L., Callahan, P. L., Condra, J. H., Sardana, V. V., Graham, D. J., Nguyen, T. M., LeGrow, K., Gotlib, L., Schlabach, A. J. & Colonno, R. J. (1990). Sequence-Specific Interaction of tat Protein and tat Peptides with the Transactivations-Responsive Sequence Element of Human Immnuodeficiency Virus type 1 *in vitro*. *Proc. Natl. Acad. Sci.* **87**, 8985-9090.

Dayie, K. T., Brodsky, A. S. & Williamson, J. R. (1997). RNA Base Dynamics by Solution ^{15}N, ^{13}C NMR Spectroscopy. Spectral Density Mapping Using Interference Effects. *in preparation..*

Dayie, K. T. & Wagner, G. (1994). Relaxation Measurements for ^{15}N -^{1}H Groups with Pulse-Field Gradients and Preservation of Coherence Pathways. *Journal of Magnetic Resonance A* **111**, 121-6.

Dayie, K. T., Wagner, G. & Lefevre, J.-F. (1996). Theory and Practice of Nuclear Spin Relaxation in Proteins. *Annual Review in Physical Chemistry* **47**, 243-82.

Delaglio, F., Grzesiek, S., Vuister, G., Zu, G., Pfeiffer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277-293.

Derse, D., Carvalho, M., Carroll, R. & Peterlin, B. M. (1991). A Minimal Lentivirus Tat. *Journal of Virology* **65**, 7012-7015.

Dieckmann, T. & Feigon, J. (1994). Heteronuclear techniques in NMR studies of RNA and DNA. *Current Opinion in Structural Biology* **4**, 745-749.

Dieckmann, T., Suzuki, E., Nakamura, G. D. & Feigon, J. (1996). Solution Structure of an ATP-Binding RNA Aptamer Reveals a Novel Fold. *RNA* **2**, 628-640.

Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skiner, M. A. & Valerio, R. (1989). Human Immunodeficiency Virus 1 Tat Protein Binds trans-Activation-Responsive Region (TAR) RNA *in vitro*. *Proc. Natl. Acad. Sci.* **86**, 6925-6929.

Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M. & Skinner, M. A. (1990). HIV-1 Tat Protein Stimulates Transcription by Binding to a U-rich Bulge in the Stem of the TAR RNA Structure. *EMBO J.* **9**, 4145-4153.

Dongen, M. v., Heus, H. A., Wymenga, S. S., Marel, G. A. v. d., Boom, J. H. v. & Hilbers, C. W. (1996). Unambiguous Structure Characterization of a DNA-RNA Triple Helix by ^{15}N- and ^{13}C-Filtered NOESY Spectroscopy. *Biochemistry* **35**, 1733-39.

Fan, P., Suri, A. K., Fiala, R., Live, D. & Patel, D. J. (1996a). Molecular Recognition in the FMN-RNA Aptamer Complex. *Journal of Molecular Biology* **258**, 480-500.

Fan, P., Suri, A. K., Fiala, R., Live, D. & Patel, D. J. (1996b). Molecular Recognition in the FMN-RNA Aptamer Complex. *J. Molecular Biology* **258**, 480-500.

Farmer, B. T., Mueller, L., Nikonowicz, E. P. & Pardi, A. (1994). Unambiguous Through-Bond Sugar-to-Base Correlations for Purines in $^{13}C,^{15}N$-labeled Nucleic Acids: The $H_sC_sN_b$, $H_sC_s(N)_bC_b$, and $H_bN_bC_b$ experiments. *Journal of Biomolecular NMR* **4**, 129-133.

Farmer, B. T., Muller, L., Nikonowicz, E. P. & Pardi, A. (1993). Unambiguous Resonance Assignments in $^{13}C,^{15}N$-Labeled Nucleic Acids by 3D Triple-Resonance NMR. *Journal of the American Chemical Society* **115**, 11040-11041.

Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1997). Characterization of the Backbone Dynamics of Folded and Denatured States of an SH3 Domain. *Biochemistry* **36**, 2390-402.

Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A. & Kay, L. E. (1995). Spectral Density Function Mapping Using ^{15}N Relaxation Data Exclusively. *Journal of Biomolecular NMR* **6**, 153-162.

Fiala, R., Jiang, F. & Patel, D. J. (1996). Direct Correlation of Exchangeable and Nonexchageable Protons on Purine Bases in $^{13}C,^{15}N$-Labeled RNA Using a HCCNH-TOCSY Experiment. *Journal of the American Chemical Society* **118**, 689-690.

Foldesi, A., Nilsson, F. P. R., Glemarec, C., Gioeli, C. & Chattopadhyaya, J. (1992). Synthesis of 1'#,2',3',4'#,5',5"-$^2H_6$-β-D-ribonucleosides and 1'#, 2',2",4'#5',5"-$^2H_7$-β-D-2'-deoxyribonucleosides for Selective Suppression of Proton Resonances in Partially-deuterated Oligo-DNA, Oligo-RNA and in 2,5A core (^1H-NMR window)*Tetrahedron* **48**, 9033.

Foldesi, A., Yamakage, S.-I., Nilsson, F. P. R., Maltseva, T. V. & Chattopadhyaya, J. (1996). The Use of Non-Uniform Deuterium Labelling ['NMR-window'] to Study the NMR Structure of a 21mer RNA Hairpin. *Nucleic Acids Research* **24**(7), 1187-1194.

Fourmy, D., Recht, M. I., Blanchard, S. C. & Puglisi, J. D. (1996). Structure of the A site of Escherichia coli 16S Ribosomal RNA Complexed with an Aminoglycoside Antibiotic. *Science* **274**, 1367-1371.

Gaffney, B. L., Goswami, B. & Jones, R. A. (1993). Nitrogen-15-Labeled Oligonucleotides. 7. Use of ^{15}N NMR to Probe H-Bonding in an O^6MeG-C Base Pair. *Journal of the Americal Chemical Society* **115**, 12607-12608.

Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. (1991). A Common Sense Approach to Peak Picking in Two-, Three- and Four-Dimensional Spectra using Automatic Computer Analysis of Contour Programs. *Journal of Magnetic Resonance* **95**, 214-220.

Gavis, E. R. & Lehmann, R. (1992). Localization of nanos RNA Controls Embryonic Polarity. *Cell* **71**, 301-313.

Gelbin, A., Schneider, B., Clowney, L., Hsieh, S.-H., Olsen, W. K. & Berman, H. M. (1996). Geometric Parameters in Nucleic acids: Sugar and Phophate Constituents. *Journal of the American Chemical Society* **118**, 519-529.

Gesteland, R. F. & Atkins, J. F., Eds. (1993). The RNA World: Cold Spring Harbor Laboratory Press.

Giessner-Prettre, C. & Pullman, B. (1987). Quantum Mechanical Calculations of NMR Chemical Shifts in Nucleic Acids. *Quarterly Review of Biophysics* **20**, 113-172.

Giver, L., Bartel, D., Zapp, M., Pawul, A., Green, M. & Ellington, A. D. (1993). Selective Optimization of the Rev-binding Element of HIV-1. *Nucleic Acids Research* **21**, 5509-5516.

Glemarec, C., Kukel, J., Foldesi, A., Maltseva, T., Sandstrom, A., Kirsebom, L. A. & Chattopadhyaya, J. (1996). The NMR Structure of 31mer RNA Domain of Escherichia coli RNase P RNA using its non-Uniformly Deuterium Labelled Counterpart [the 'NMR-Window' Concept]. *Nucleic Acids Research* **24**(11), 2022-2035.

Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. (1995). Diversity of Oligonucleotide Functions. *Annual Review of Biochemistry* **64**, 763-97.

Gorenstein, D. G. & Luxon, B. A. (1979). High-Resolution Phosphorus Nuclear Magnetic Resonance Spectra of Yeast Phenylalanine Transfer Ribonucleic Acid. Melting Curves and Relaxation Effects. *Biochemistry* **18**, 3796-3804.

Green, H. & Freeman, R. (1991). Band-Selective Radiofrequency Pulses. *J. Magnetic Resonance* **93**, 93-141.

Gueron, M. & Shulman, R. G. (1975). ^{31}P Magnetic Resonance of tRNA. *Proc. Nat. Acad. Sci. USA* **72**, 3482-3485.

Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P. J. G., Karn, J. & Gait, M. J. (1993). Hydrogen-bonding Contacts in the Major Groove are Required for Human Immunodeficiency Virus type-1 tat Protein Recognition of TAR RNA. *J. Mol. Biol.* **230**, 111-123.

Hamy, F., Felder, E. R., Heizmann, G., Lazdins, J., Aboul-ela, F., Varani, G., Karn, J. & Klimkait, T. (1997). An Inhibitor of the Tat/TAR RNA Interaction that Effectively Suppresses HIV-1 Replication. *PNAS* **94**, 3548-3553.

Hart, C. E., Saltarelli, M. J., Galphin, J. C. & Schochetman, G. (1995). A Human Chromosome 12-Associated 83-Kilodalton Cellular Protein Specifically Binds to the Loop Region of Human Immnuodeficiency Virus Type 1 Trans-Activation Response Element RNA. *Journal of Virology* **69**, 6593-6599.

Hauber, J., Malim, M. H. & Cullen, B. R. (1989). Mutational Analysis of the Conserved Basic Domain of Human Immunodeficiency Virus Tat Protein. *Journal of Virology* **63**, 1181-1187.

Heus, H. & Pardi, A. (1991). Structural Features that Give Rise to Unusual Stability of RNA Hairpins Containing GNRA Loops. *Science* **253**, 191-194.

Heus, H. A., Wijmenga, S. S., Ven, F. J. M. v. d. & Hilbers, C. W. (1994). Sequential Backbone Assignment in ^{13}C-labeled RNA via Through-Bond Coherence Transfer Using Three-dimensional Triple Resonance Spectroscopy(^{1}H, ^{13}C, ^{31}P) and Two-dimensional Hetero-TOCSY. *J. American Chemical Society* **116**, 4983-4984.

Hines, J. V., Landry, S. M., Varani, G. & Jr., I. T. (1994). Carbon-Proton Scalar Couplings in RNA: 3D Heteronuclear and 2D Isotope-Edited NMR of a ^{13}C-Labeled Extra-Stable Hairpin. *Journal of the American Chemical Society* **116**, 5823-5831.

Hines, J. V., Varani, G., Landry, S. M. & I. Tinoco, J. (1993). The stereospecific assignment of H5' and H5" in RNA using the sign of two-bond carbon-proton scalar coupling. *Journal of the American Chemical Society* **115**, 11002-11003.

Hoogstraten, C. G. & Markley, J. L. (1996). Effects of Experimentally Achievable Improvements in the Quality of NMR Distance Constraints on the Accuracy of Calculated Protein Structures. *Journal of Molecular Biology* **258**, 334-48.

Ishima, R. & Nagayama, K. (1995). Protein Backbone Dynamics Revealed by Quasi Spectral Density Function Analysis of Amide N-15 Nuclei. *Biochemistry* **34**, 3162-3171.

Jaeger, J. & Tinoco, J. I. (1993). An NMR Study of the HIV-1 TAR Element Hairpin. *Biochemistry* **32**, 12522-12530.

Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. (1994). High-Resolution Molecular Discriminaion by RNA. *Science* **263**, 1425-1429.

Jensen, K. B., Green, L., MacDougal-Waugh, S. & Tuerk, C. (1994). Characterization of an in Vitro-selected RNA Ligand to the HIV-1 Rev Protein. *Journal of Molecular Biology* **235**, 237-247.

Jiang, F., Kumar, R. A., Jones, R. A. & Patel, D. J. (1996). Structural Basis of RNA Folding and Recognition in an AMP-RNA Aptamer Complex. *Nature* **382**, 183-186.

Jiang, L., Suri, A. K., Fiala, R. & Patel, D. J. (1997). Saccahride-RNA Recognition in an Aminoglycoside Antibiotic-RNA Aptamer. *Chemistry and Biology* **4**, 35-50.

Johnson, B. A. & Blevins, R. A. (1994). NMRView: A Computer Program for the Visualization and Analysis of NMR Data. *Journal of Biomolecular NMR* **4**, 603-614.

Jones, K. A. & Peterlin, B. M. (1994). Control of RNA Initiation and Elongation at the HIV-1 Promoter. *Annual Reviews in Biochemistry* **63**, 717-743.

Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M. & Pardi, A. (1996). A Network of Heterogeneous Hydrogen Bonds in GNRA Tetraloops. *J. Mol. Biol.* **264**, 968-980.

Jucker, F. M. & Pardi, A. (1995). Solution Structure of the CUUG Hairpin Loop: a Novel RNA Tetraloop Motif. *Biochemistry* **34**, 14416-14427.

Karplus, M. (1969). Proton Spin Coupling by Pi Electrons. *J. Chem. Phys.* **33**, 1842-50.

Kay, L. E., Ikura, M. & Bax, A. (1990). Proton-Proton Correlation via Carbon-Carbon Couplings: A Three-Dimensional NMR Approach for the Assignment of Aliphatic Resonances in Proteins Labeled with Carbon-13. *Journal of the American Chemical Society* **112**, 888-889.

Kay, L. E., Muhandirm, D. R., Farrow, N. A., Aubin, Y. & Forman-Kay, J. D. (1996). Correlation Between Dynamics and High Affinity Binding in an SH2 Domain Interaction. *Biochemistry* **35**, 361-8.

Keen, N. J., Gait, M. J. & Karn, J. (1996). Human Immunodeficiency Virus Type-1 Tat is an Integral Component of the Activated Transcription-Elongation Complex. *Proceedings of the National Academy Sciences USA* **93**, 2505-2510.

King, G. C., Harper, J. W. & Xi, Z. (1995). Isotope Labeling for ^{13}C Relaxation Measurements on RNA. *Methods in Enzymology* **261**, 436-450.

Krishnan, V. V. & Rance, M. (1995). Influence of Chemical Exchange among Homonuclear Spins in Heteronuclear Coherence-Transfer Experiments in Liquids. *Journal of Magnetic Resonance, Series A* **116**, 97-106.

Kundrot, C. (1996). Rapid Identification of Ordered and Disordered Domains in NMR Structures. *Journal American Chemical Society* **118**, 8725-8726.

Kuppuswamy, M., Subramanian, T., Srinivasan, A. & Chinnadurai, G. (1989). Multiple Functional Domains of Tat, the trans-Activation of HIV-1, Defined by Mutational Analysis. *Nucleic Acids Research* **17**, 3551-3561.

Lefevre, J. F., Dayie, K. T., Peng, J. W. & Wagner, G. (1996). Internal Mobility in the Parially Folded DNA Binding and Dimerization Domains of GAL4: NMR Analysis of the N-H Spectral Density Functions. *Biochemistry* **35**, 2674-86.

Legault, P., Jucker, F. M. & Pardi, A. (1995). Improved measurement of 13C, 31P J coupling constants in isotopically labeled RNA. *FEBS Letters* **362**, 156-160.

Legault, P. & Pardi, A. (1997). Unusual Dynamics and pK_a Shift at the Active Site of a Lead-Dependent Ribozyme. *Journal of the American Chemical Society* **119**, 6621-628.

Lipari, G. & Szabo, A. (1982a). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society* **104**, 4546-4559.

Lipari, G. & Szabo, A. (1982b). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Analysis of Experimental Results. *Journal of the American Chemical Society* **104**, 4559-4570.

Lippens, G., Dhalluin, C. & Wieruszeki, J.-M. (1995). Use of a Water Flip-Back Pulse in the Homonuclear NOESY Experiment. *Journal of Biomolecular NMR* **5**, 327-331.

Liu, Y., Zhao, D., Altman, R. & Jardetzky, O. (1992). A Systematic Comparison of Three Structure Determination Methods from NMR Data: Dependence upon Quality and Quantity of Data. *Journal Biomolecular NMR* **2**, 373-388.

Long, K. S. & Crothers, D. M. (1995). Interaction of Human Immunodeficiency Virus Type 1 Tat-Derived Peptides with TAR RNA. *Biochemistry* **34**, 8885-8895.

Macaya, R., Wang, E., Schultze, P., Sklenar, V. & Feigon, J. (1992). Proton Nuclear Magnetic Resonance Assignments and Structural Characterization of an Intramolecular DNA triplex. *Journal of Molecular Biology* **225**, 755-773.

Marino, J. P., Diener, J. L., Moore, P. B. & Griesinger, C. (1997). Multiple-Quantum Coherence Dramatically Enhances the Sensitivity of CH and CH2 Correlations in Uniformly ^{13}C-Labeled RNA. *Journal of the American Chemical Society* **119**, 7361-7366.

Marino, J. P., Schwalbe, H., Anklin, C., Bermel, W., Crothers, D. M. & Griesinger, C. (1994). A three-dimensional triple-resonance ^{1}H, ^{13}C, ^{31}P experiment: sequential through-bond correlation of ribose protons and intervening phosphorous along the RNA oligonucleotide backbone. *J. American Chemical Society* **116**, 6472-6473.

Marino, J. P., Schwalbe, H., Anklin, C., Bermel, W., Crothers, D. M. & Griesinger, C. (1995). Sequential Correlation of Anomeric Ribose Protons and Intervening Phosphorus in RNA Oligonucleotides by a ^{1}H,^{13}C,^{31}P Triple Resonance Experiment: HCP-CCH-TOCSY. *Journal of Biomolecular NMR* **5**, 87-92.

Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA-Polymerase and Synthetic DNA Templates. *Nucleic Acids Research* **15**, 8783-8798.

Mohebbi, A. & Shaka, A. J. (1991). *Journal of Chemical Physics Letters* **178**, 374-378.

Moore, P. B. (1993) Ribosomes and the RNA World. In *The RNA World* (Gesteland, R. F. And Atkins, J. F., ed.), pp 119-136. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Moore, M. J., Query, C. C., & Sharp, P. A. (1993) Splicing of Precursors to mRNAs by the Spliceosome. In *The RNA World* (Gesteland, R. F. And Atkins, J. F., ed.), pp 303-358. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Mori, S., Abeygunawardana, C., Johnson, M. O. N. & Zijl, P. C. M. v. (1995). Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme that Avoids Water Saturation. *Journal of Magnetic Resonance, Series B* **108**, 94-98.

Mueller, L., Legault, P. & Pardi, A. (1995). Improved RNA Structure Determination by Detection of NOE Contacts to Eschange-Broadened Amino Protons. *Journal of the American Chemical Society* **117**(45), 11043-11048.

Mujeeb, A., Bishop, K., Peterlin, B. M., Turck, C., Parslow, T. G. & James, T. L. (1994). NMR Structure of a Biologically Active Peptide Containing the RNA-Binding Domain of Human Immunodeficiency Virus Type 1 Tat. *Proceeding National Academy Sciences USA* **91**, 8248-52.

Newstein, M., Lee, I. S., Venturini, D. S. & Shank, P. R. (1993). A Chimeric Human immunodeficiency Virus Type 1 TAR Region which Mediates High Level Trans-Activation in both Rodent and Human Cells. *Virology* **197**, 825-828.

Nikonowicz, E. P. & Pardi, A. (1992a). Simple Procedure for Resonance Assignment of the Sugar Protons in ^{13}C-Labeled RNAs. *Journal of the American Chemical Society* **114**, 9202-9303.

Nikonowicz, E. P. & Pardi, A. (1992b). Three-dimensional Heteronuclear NMR Studies of RNA. *Nature* **355**, 184-186.

Nikonowicz, E. P. & Pardi, A. (1993). An Efficient Procedure for Assignment of the Proton, Carbon and Nitrogen Resonances in $^{13}C/^{15}N$ Labeled Nucleic Acids. *Journal of Molecular Biology* **232**(4), 1141-1156.

Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M. & Pardi, A. (1992). Preparation of 13C and 15N labelled RNAs for Heteronuclear Multi-dimensional NMR Studies. *Nucleic Acids Research* **20**(17), 4507-4513.

Nilges, M. (1996). Structure Calculation from NMR Data. *Current Opinion in Structural Biology* **6**, 617-623.

O'Brien, T., Hardin, S., Greenleaf, A. & Lis, J. T. (1994). Phosphorylation of RNA Polymerase II C-Terminal Domain and Transcriptional Elongation. *Nature* **370**, 75-77.

Otting, G., Liepinsh, E. & Wuthrich, K. (1991). Protein Hydration in Aqueous Solution. *Science* **254**, 974-980.

Otting, G. & Wuthrich, K. (1989a). Extended Heteronuclear Editing of 2D ^{1}H NMR Spectra of Isotope-Labeled Proteins, Using the X(ω1, ω2) Double Half Filter. *Journal of Magnetic Resonance* **85**, 586-594.

Otting, G. & Wuthrich, K. (1989b). Studies of Protein Hydration in Aqueous Solution by Direct NMR Observation of Individual Protein-Bound Water Molecules. *Journal of the American Chemical Society* **111**, 1871-1875.

Otting, G. & Wuthrich, K. (1990). Heteronuclear Filters in Two-Dimensional [1H,1H]-NMR Spectroscopy: Combined use with Isotope Labelling for Studies of Macromolecular Conformation and Intermolecular Interactions. *Quarterly Reviews of Biophysics* **23**(1), 39-96.

Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.

Palmer, A. G., Williams, J. & McDermott, A. (1996). Nuclear Magnetic Resonance Studies of Biopolymer Dynamics. *Journal of Physical Chemistry* **100**, 13923-13310.
Pardi, A. (1995). Multidimensional Heteronuclear NMR Experiments for Structure Determination of Isotopically Labeled RNA. In *Methods in Enzymology* (James, T. L., ed.), Vol. 261, pp. 350-380. Academic Press, San Diego.

Pardi, A. & Nikonowicz, E. P. (1992). Simple Procedure for Resonance Assignment of the Sugar Protons in 13C-Labeled RNAs. *Journal of the American Chemical Society* **114**, 9202-9203.

Peng, J. W. & Wagner, G. (1992). Mapping of Spectral Density Functions Using Heteronuclear NMR Relaxation Measurements. *Journal of Magnetic Resonance* **98**, 308-332.

Peng, J. W. & Wagner, G. (1992b). Mapping of the Spectral Densities of N-H Bond Motions in Eglin c Using Heteronuclear Relaxation Experiments. *Biochemistry* **31**, 8571-8586.

Peng, J. W. & Wagner, G. (1994). Investigation of Protein Motions via Relaxation Measurements. *Methods in Enymology* **239**, 563-596.

Peng, J. W. & Wagner, G. (1995). Frequency Spectrum of NH Bonds in Eglin c from Spectral Density Mapping at Multiple Fields. *Biochemistry* **34**, 16733-16752.

Piotto, M., Saudek, V. & Skelnar, V. (1992). Gradient Tailored Excitation for Single Quantum NMR Spectroscopy of Aqueous Solution. *J. Biomolecular NMR* **2**, 661-.

Pley, H., Flaherty, K. M. & McKay, D. B. (1994a). Three-Dimensional Structure of a Hammerhead Ribozyme. *Nature* **372**, 68-74.

Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994b). Model of an RNA Tertiary Interaction from the Structure of an Intermolecular Complex Between a GAAA Tetraloop and an RNA helix. *Nature* **372**, 111-113.

Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharek, A. M., Singh, M., Karn, J. & Gait, M. (1994). Methylphosphonate Mapping of Phosphate Contacts Critical for RNA Recognition by the Human Immunodeficiency Virus Tat and Rev Proteins. *Nucleic Acids Research* **22**, 2592-2600.

Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution Structure of a Bovine Immunodeficiency Virus Tat TAR RNA-Peptide Complex. *Science* **270**, 1200-1203.

Puglisi, J. D., Chen, L., Frankel, A. D. & Williamson, J. R. (1993). Role of RNA structure in arginine recognition of TAR RNA. *Proc. Natl. Sci. USA* **90**, 3680-3684.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992). Conformations of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science* **257**, 76-80.

Quant, S., Wechselberger, R. W., Wolter, M. A., Worner, K.-H., Schell, P., Engels, J. W., Griesinger, C. & Schwalbe, H. (1994). Chemical Synthesis of ^{13}C-labelled Monomers for the Solid-Phase and Template Controlled Enzymatic Synthesis of DNA and RNA Oligomers. *Tetrahedron Letters* **35**(36), 6649-6652.

Reif, B., Hennig, M. & Griesinger, C. (1997). Direct Measurement of Angles Between Bond Vectors in High-Resolution NMR. *Science* **276**, 1230-1233.

Roberts, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C. & Klug, A. (1974). Structure of Yeast Phenylalanine tRNA at 3 Å Resolution. *Nature* **250**, 546-551.

Rould, M. A., Perona, J. J., Soll, D. & Steitz, T. A. (1989). Structure of *E. Coli* Glutaminyl-tRNA Synthetase Complexed with tRNA[Gln] and ATP at 2.8 Å Resolution: Implications for tRNA Discrimination. *Science* **246**, 1135-1142.

Roy, S., Delling, U., Chen, C.-H., Rosen, C. A. & Sonenberg, N. (1990b). A Bulge Structure in HIV-1 TAR RNA is Required for Tat Binding and Tat-mediated *trans*-Activation. *Genes Development* **4**, 1365-1373.

Roy, S., Parkin, N. T., Rosen, C. A., Itovitch, J. & Sonenberg, N. (1990a). Structural Requirements for trans-activation of Human Immunodeficiency Virus Type 1 Long Terminal Repeat-Directed Gene Expression by Tat: Importance of Base-Pairing, Loop Sequence and Bulges in the Tat-Responsive Sequence. *J. Virology* **64**, 1402-1406.

Ruben, S., Perkins, A., Purcell, R., Joung, K., Sia, R., Burghoff, R., Haseltine, W. A. & Rosen, C. A. (1989). Structural and Functional Characterization of Human Immnudeficiency Virus Tat Protein. *Journal of Virology* **63**, 1-8.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991). Class II Aminoacyl Transfer RNA Synthetases: Crystal Structure of Yeast Aspartyl-tRNA Synthetase Complexed with tRNA(Asp). *Science* **252**, 1682-9.

Saenger, W. (1983). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Salemink, P. J. M., Swarthof, T. & Hilbers, C. W. (1979). Studies of Yeast Phenylalanine-Accepting Transfer Ribonucleic Acid Backbone Structure in Solution by Phosphorous-31 Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **18**, 3477-3485.

SantaLucia, J., Shen, L. X., Cai, Z., Lewis, H. & Tinoco, I. (1995). Synthesis and NMR of RNA with Selective Isotopic Enrichment in the Bases. *Nucleic Acids Research* **23**(23), 4913-4921.

Santos, C. d. l., Rosen, M. & Patel, D. (1989). NMR Studies of DNA (R+)n.(Y-)n.(Y+)n Triple Helices in Solution: Imino and Amino Proton Markers of T.A.T and C.G.C+ Base-Triple Formation. *Journal of Molecular Biology* **28**, 7282-7289.

Schumacher, T. N. M., Mayr, L. M., Jr., D. L. M., Milhollen, M. A., Burgess, M. W. & Kim, P. S. (1996). Identification of D-Peptide Ligands Through Mirror-Image Phage Display. *Science* **271**, 1854-1857.

Schwalbe, H., Marino, J. P., Glaser, S. J. & Griesinger, C. (1995). Measurement of H,H-Coupling Constants Associated with v_1, v_2, and v_3 in Uniformly ^{13}C Babeled RNA by HCC-TOCSY-CCH-E. COSY. *J. A. C. S.* **117**, 7251-7252.

Schwalbe, H., Marino, J. P., King, G. C., Wechselberger, R., Bermel, W. & Griesinger, C. (1994). Determination of a complete set of coupling constants in ^{13}C-labeled oligonucleotides. *Journal of Biomolecular NMR* **4**(5), 631-644.

Scott, W. G., Finch, J. T. & Klug, A. (1995). The Crystal Structure of an All-RNA Hammerhead Ribozyme: A Proposed Mechanism for RNA Catalytic Cleavage. *Cell* **81**, 995.

Selby, M. J. & Peterlin, B. M. (1990). Trans-activation by HIV-1 Tat via a Heterologous RNA Binding Protein. *Cell* **62**, 769-776.

Shaka, A. J., Keeler, J. & Freeman, R. (1983). Evaluation of a New Broadband Decoupling Sequence: WALTZ-16. *Journal of Magnetic Resonance* **53**, 313-340.

Sheline, C. T., Milocco, L. H. & Jones, K. A. (1991). Two Distinct Nuclear Transcription Factors Recognize Loop and Bulge Residues of the HIV-1 TAR RNA Hairpin. *EMBO Journal* **14**, 5995-6009.

Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1996). Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **274**, 1531-4.

Simorre, J.-P., Zimmermann, G., Mueller, L. & Pardi, A. (1996a). Triple-Resonance Experiments for Assignment of Adenine Base Resonances in $^{13}C/^{15}N$-Labeled RNA. *Journal of the American Chemical Society* **5316-5317**.

Simorre, J.-P., Zimmermann, G. R., Pardi, A., II, B. T. F. & Mueller, L. (1995). Triple resonance HNCCCH experiments for correlating exchangeable and nonexchangeable cytidine and uridine base protons in RNA. *Journal of Biomolecular NMR* **6**, 427-432.

Simorre, J. P., Zimmermann, G. R., Mueller, L. & Pardi, A. (1996b). Correlation of the Guanosine Exchangeable and Nonexchangeable Base Protons in $^{13}C-/^{15}N$-Labeled RNA with an HNC-TOCSY-CH Experiment. *Journal of Biomolecular NMR* **7**, 153-156.

Skélnar, V., Brooks, B. R., Zon, G. & Bax, A. (1987). Absorption Mode two-Dimensional NOE Spectroscopy of Exchangeable Protons in Oligonucleotides. *FEBS Letters* **216**(2), 249-252.

Skélnar, V., Miyashiro, H., Zon, G. & Bax, A. (1986). Assignment of the ^{31}P and ^{1}H Resonances in Oligonucleotides by Two-Dimensional NMR Spectroscopy. *FEBS Letters* **208**, 94-98.

Skélnar, V., Peterson, R. D., Rejante, M. R. & Feigon, J. (1993a). Two- and Three-Dimensional HCN Experiments for Correlating Base and Sugar Resonances in $^{15}N,^{13}C$-Labeled RNA Oligonucleotides. *Journal of Biomolecular NMR* **3**, 721-727.

Skélnar, V., Peterson, R. D., Rejante, M. R., Wang, E. & Feigon, J. (1993b). Two-Dimensional Triple-Resonance HCNCH Experiment for Direct Correlation of Ribose H1' and Base H8, H6 Protons in $^{13}C,^{15}N$-Labeled RNA Oligonucleotides. *Journal of the American Chemical Society* **115**, 12181-12182.

Soll, Dieter (1993). Transfer RNA: an RNA for All Seasons. In *The RNA World* (Gesteland, R. F. And Atkins, J. F., ed.), pp 157-184. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Southgate, C., Zapp, M. L. & Green, M. R. (1990). Activation of Transcription by HIV-1 Tat Protein Tethered to Nascent RNA Through Another Protein. *Nature* **345**, 640-642.
Southgate, C. D. & Green, M. R. (1991). The HIV-1 Tat Protein Activates Transcription from an Upstream DNA-Binding Site: Implications for Tat Function. *Genes Development* **5**, 2496-2507.

Staley, J. P. & Kim, P. S. (1994). Formation of a Native-like Subdomain in a Partially Folded Intermediate of Bovine Pancreatic Trypsin Inhibitor. *Protein Science* **3**, 1822-1832.

Sumner-Smith, M., Roy, S., Barnett, R., Reid, S. L., Kuperman, R., Delling, U. & Sonenberg, N. (1991). Critical features in *trans*-acting-responsive RNA are required for

interaction of human immunodeficiency virus type 1 Tat protein. *J. Virology* **65**, 5196-5202.

Szyperski, T., Luginbuhl, P., Otting, G., Guntert, P. & Wuthrich, K. (1993). Protein Dynamics Studied by Rotating Frame ^{15}N Spin Relaxation Times. *Journal of Biomolecular NMR* **3**, 151-164.

Tan, R., Brodsky, A, Williamson, J. R., & Frankel, A. D. (1997) *Seminars in Virology*, **in press**.

Tao, J. & Frankel, A. D. (1992). Specific Binding of Arginine to TAR RNA. *Proc. Natl. Acad. Sci. USA* **89**, 2723-2726.

Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A. & Clore, G. M. (1997). Defining Long Range Order in NMR Structure Determination from the Dependence of Heteronuclear Relaxations Times on Rotational Diffusion Anistropy. *Nature Structural Biology* **4**, 443-449.

Tolbert, T. J. & Williamson, J. R. (1996). Preparation of Specifically Deuterated RNA for NMR Studies Using a Combination of Chemical and Enzymatic Synthesis. *Journal of the American Chemical Society* **118**, 7929-7940.

Tolbert, T. J. & Williamson, J. R. (1997). Preparation of Specifically Deuterated and ^{13}C-Labeled RNA for NMR Studies Using Enzymatic Synthesis. *Journal of the American Chemical Society* **in press**.

Ullman, K. S., Powers, M. A. & Forbes, D. J. (1997). Nuclear Export Receptors: from Importin to Exportin. *Cell* **90**, 967-970.

Valegard, K., Murray, J. B., Stockley, P. G., Stonehouse, N. J. & Liljas, L. (1994). Crystal Structure of a Bacteriophage-RNA Coat Protein-Operator Complex. *Nature* **371**, 623-626.

Varani, G., Aboul-ela, F., Allain, F. & Gubser, C. C. (1995). Novel Three-Dimensional ^{1}H-^{13}C-^{31}P Triple Resonance Experiments for Sequential Backbone Correlations in Nucleic Acids. *Journal of Biomolecular NMR* **5**, 315-320.

Varani, G., Aboul-ela, F. & Allain, F. H.-T. (1996). NMR Investigation of RNA Structure. *Progress in Nuclear Magnetic Resonance Spectroscopy* **29**, 51-127.

Varani, G. & Ignacio Tinoco, J. (1991). RNA Structure and NMR Spectroscopy. *Quarterly Reviews of Biophysics* **24**(4), 479-532.

Vlieg, J. d. & Gunsteren, W. F. v. (1991). Combined Procedures of Distance Geometry and Molecular Dynamics for Determining Protein Structure from Nuclear Magnetic Resonance Data. In *Methods in Enzymology*, Vol. 202, pp. 368-300.

Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, R. & Bax, A. (1993). Increased Resolution and Improved Spectral Quality in Four-Dimensional ^{13}C/^{13}C-Separated HMQC-NOESY-HMQC Spectra Using Pulsed Field Gradients. *Journal of Magnetic Resonance, Series B* **101**, 210-213.

Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A. & Crothers, D. M. (1990). Fragments of the HIV-1 Tat Protein Specifically Bind TAR RNA. *Science* **249**, 1281-1285.

Weeks, K. M. & Crothers, D. M. (1991). RNA Recognition by Tat-derived Peptides: Interaction in the Major Groove. *Cell* **66**, 577-588.

Weeks, K. M. & Crothers, D. M. (1992). RNA binding assays for tat-derived peptides: implications for specificity. *Biochemistry* **31**, 10281-10287.

Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). An All Atom Forcefield for Simulations of Proteins and Nucleic Acid. *J. Computational Chemistry* **7**, 230-252.

Wijmenga, S. S., Heus, H. A., Leeuw, H. A. E., Hoppe, H., Graaf, M. v. d. & Hilbers, C. W. (1995). Sequential Backbone Assignment of Uniformly ^{13}C-Labeled RNAs by a Two-Dimensional P(CC)H-TOCSY Triple Resonance NMR Experiment. *Journal of Biomolecular NMR* **5**, 82-86.

Williamson, J. R. & Boxer, S. G. (1989). Multinclear NMR Studies of DNA Hairpins. 1. Structure and Dynamics of d(CGCGTTGTTCGCG). *Biochemistry* **28**, 2819-31.

Wishart, D. S. & Sykes, B. D. (1994). Chemical Shifts as a Tool for Structure Determination. *Methods in Enzymology* **239**, 363-392.

Wishart, D. S., Sykes, B. D. & Richards, F. M. (1991). Relationship between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure. *Journal of Molecular Biology* **222**, 311-333.

Wu, F., Garcia, J., Sigman, D. & Gaynor, R. (1991). Tat Regulates Binding of the Human Immunodeficiency Virus trans-Activating Region RNA Loop-Binding Protein TRP-185. *Genes Development* **5**, 2128-2140.

Wu-Baer, F., Lane, W. S. & Gaynor, R. B. (1995). The Cellular Factor TRP-185 Regulates RNA Polymerase II Binding to HIV-1 TAR RNA. *EMBO J.* **14**, 5995-6009.

Wu-Baer, F., Lane, W. S. & Gaynor, R. B. (1996). Identification of a Group of Cellular Cofactors that Stimulate the Binding of RNA Polymerase II and TRP-185 to Human Immunodeficiency Virus 1 TAR RNA. *Journal of Biological Chemistry* **271**, 4201-4208.

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

Xu, J., Lapham, J. & Crothers, D. M. (1996). Determining RNA Solution Structure by Segmental Isotopic Labeling and NMR: Application to *Caenorhabditis elegans* Spliced Leader RNA 1. *Proceedings of the National Academy of Sciences USA* **93**, 44-48.

Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. (1996). Structural Basis of Ligand Discrimination by Two Related RNA Aptamers Resolved by NMR spectroscopy. *Science* **272**, 1343-1347.

Ye, X., Gorin, A., Ellington, A. D. & Patel, D. J. (1996). Deep Penetration of an α-helix into a Widened RNA Major Groove in the HIV-1 Rev Peptide-RNA Aptamer Complex. *Nature Structural Biology* **3**, 1026-1033.

Ye, X., Kumar, R. A. & Patel, D. J. (1995). Molecular Recognition in the Bovine Immunodeficiency Virus Tat Peptide TAR RNA Complex. *Chemistry & Biology* **2**, 827-840.

Zacharias, M. & Hagerman, P. J. (1995). The Bend in RNA Created by the trans-Activation Response Element bulge of Human Immunodeficiency Virus is Straightened by Arginine and by Tat-derived Peptide. *Proc. Natl Acad. Aci. USA* **92**, 6052-6056.

Zimmerman, G., Jenison, R. D., Wick, C. L., Simorre, J.-P. & Pardi, A. (1997). Interlocking Structural Motifs Meidate Molecular Discrimination by a Theophylline-Binding RNA. *Nature Structural Biology* **4**, 644-649.