Nuclear Magnetic Resonance Studies of HIV TAR RNA Mutants

by

Heidi A. Erlacher

B.A. in Chemistry
Clark University, 1989

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

DOCTOR OF PHILOSOPHY
IN ORGANIC CHEMISTRY

at the

Massachusetts Institute of Technology

September, 1996

Signature of Author .....
Department of Chemistry
August 15, 1996

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

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

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

Professor John M. Essigmann..... Chair

Professor James R. Williamson..... Thesis Supervisor

Professor Peter T. Lansbury.....

Dedicated to

the memory of my Godparents:

Helen Rajtar Tait
and
Robert P. Rajtar, Sr.

and to the futures of my Godchildren:

Dayna Marie Nobile
and
Nicholas Edward Erlacher

Acknowledgments

Seven years- and many, many projects later- it appears to be over! Realizing that all of this thesis work was done in the past 18 months has made me wonder why I haven't lost my mind sooner (perhaps I have and haven't noticed!). It is a humbling thing to become aware of all the people one needs to thank after such a long time in one place. First and foremost, I need to thank the J-man, for everything. Jamie took me in as a 'used' graduate student, and gave me a second chance. The knowledge that he would support me in my career decisions, whatever they may be, was invaluable. And, mostly, I'd like to thank him for knowing when to say when, and telling me that it was time to leave the nest. I hope his excitement for science continues to be contagious to future members of his laboratory.

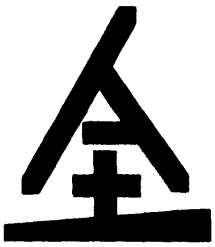
A heartfelt thanks also goes to all of the Williamson lab. The laboratory environment was stimulating and friendly. A few members merit special mention. I'd especially like to thank John Battiste for answering numerous questions about NMR. I think he could've graduated a year ago if it hadn't been for me bugging him all of the time. I'd also like to thank him for taking the time to help me with computer difficulties while in the throes of writing his own thesis. But mostly, for being a great person, and for not being too macho to have a 'girl' for a weight-lifting partner. Keep up the hard work, and be huge! Alex Brodsky and I collaborated on work with the UU-TAR-CGC mutant RNA. He was a tremendous help with NMR at the Bitter Lab, and data processing in general. Of course, I need to mention the computer Czar, Christopher Cilley. Management of a ton of data and huge data sets would have been nearly impossible if not for all his work with the computers. Thanks to him for all his help, advice, and 6 weeks on Newbury St. Thanks, too, to Pat Zarrinkar for keeping the lab well-stocked and running smoothly. In the early days, Jody Puglisi was a great source of inspiration and excitement about RNA. Chris Turner provided tremendous help with many experiments performed at the Bitter Lab. Robert Batey was always the first one I'd ask with questions about all things biological. Thanks, too, to Dan Treiber for always thinking that I've been working out, even when I haven't!

When my shopping partner, Sharon Parker, moved on to another post-Doc, I had no idea that her replacement (some guy from Indiana) would turn out to be such a wonderful person. Dr. Jeffrey Orr has been a tremendous source of support during the last few years. He is a great cook and an even better friend. I'd like to thank him for that - as well as the frequent use of his car - and for doing all the 'guy' things that had to be done.

Thanks must also go to Dr. Mark Turnbull and Dr. Dave Thurlow at Clark University. It was a great comfort to know that caring advice was only an hour (or an email) away. Mark's belief in me started me on this scientific journey. However, I must admit that much of what I have learned in the past seven years doesn't have much to do with science. Many people at MIT have enhanced my well-being while in graduate school. Linda Doerrer has been a great friend; desperation trips to Crossroads with her will not be forgotten (we're due after this one!). Annie Helgason and Jon Come get thanks (blame?) for convincing me to stick it out. Their friendships made the early days significantly more bearable, and have continued to brighten my life. Living in Worcester won't be too bad with Annie around. Hey, Girl, get over the soccer thing and stick with rugby! I need someone to commute with me to practice...

Monica Feest Moriarty has been a tremendous friend over the past 15 years. No matter where I go I always know she'll be there with me. Liz Harris has kept me well-stocked with contacts, glasses and sunglasses. Curt Burmeister taught me a lot about many things. Constant support has come from the people in Connecticut, from the Beach House Posse (ACP, VLR, LLE) to the members of my family. Thanks for understanding and trying to understand what it is like to be a graduate student.

A huge hug goes to my 'sistahs' on Beantown (1996 WOMEN'S RUGBY NATIONAL CHAMPIONS!) for giving me an invaluable outlet for my aggression. Rugby has really kept me going, and for the sake of sanity, I encourage everyone to try it at least once! Thanks to Joanie for getting us there, and for doing all the extras no other coach has ever done for her players. And to Walter, too, for loving Beantown as we all do - and reminding us of it. To Meaty, the best landlord ever, you're the greatest. For supporting me, and eating with me, thanks to the members of the Big Eaters Club -for no one truly enjoys pigging-out alone! A very special thank-you goes to fellow second-row Jan Rutkowski for so many things, not the least of which are teaching me how to play and really love the sport of rugby! I'm forever in her debt for the use of her home and her computer. She was a tremendous source of support when things got really ugly toward the end of this ordeal.



Nuclear Magnetic Resonance Studies of HIV TAR RNA Mutants

by

Heidi A. Erlacher

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

Abstract

Transcriptional regulation in the human immunodeficiency virus requires a specific interaction between the Tat protein and the transactivation responsive region (TAR) RNA, a stem-loop structure located at the 5' end of all viral mRNAs. The Tat binding site in the TAR element contains two helical stem regions separated by a three nucleotide bulge. A single arginine within the basic region of Tat mediates specific binding to TAR; a model system used for studying the interaction of Tat with TAR includes the binding of free arginine to RNA transcripts. The guanidinium group of arginine interacts with guanosine 26 (G26) and two phosphates. This binding is stabilized by formation of a U•A-U base-triple interaction between a bulge nucleotide (U23) and the A27-U38 base pair. In an effort to better understand the sequence requirements of base-triple formation, mutations were made at positions 23 and 38 of TAR. The conformations of the free RNAs and of these mutants with the amino acid analog argininamide were analyzed by nuclear magnetic resonance (NMR) spectroscopy. To facilitate enzymatic synthesis of mutant TARs, duplex RNAs were designed to contain a single uridine at position 23 of the bulge-containing strand. The effects of deoxythymidine and dihydrouridine at position 23 on base-triple formation were investigated. Similarly, a hairpin RNA containing an A-C mispair in place of the Watson-Crick A-U base pair binds argininamide and forms a solution conformation similar to that of the wild-type TAR-argininamide complex. The stability of the U•A-C triple is increased at low pH due to protonation of the adenine base with concomitant formation of a stable base-triple interaction.

Thesis Supervisor: James R. Williamson

Title: Associate Professor of Chemistry

Contents

1 Introduction

1.1 Ribonucleic Acids

Function

During the past 15 years there has been a tremendous growth in the field of ribonucleic acid (RNA) structural biology. This increased attention can largely be attributed to the discovery of a number of new biological functions carried out by RNA. Traditionally, involvement of RNA in cellular processes was thought to be quite passive. Genetic material, stored as DNA, is transcribed by a protein enzyme (polymerase) into a messenger RNA (mRNA) intermediate. This message is then translated into a protein at the ribosome, a complex of ribosomal RNA (rRNA) and protein. Translation is assisted by adapter molecules made of RNA called transfer RNA (tRNA). One end of a tRNA recognizes the language of the genetic code (written in the base sequence of the mRNA) while the other end carries an amino acid to add to the growing polypeptide chain.

RNA is now thought to be more actively involved in each step of the transmission and management of genetic information (Wyatt et al., 1991, P. Zarrinkar, pers. commun.). Many of these new functions were discovered after it had been realized that RNA could act as a biological catalyst, in the absence of all protein (Kruger et al., 1982, Cech, 1987). This landmark discovery of the ability of RNA to catalyze phosphodiester bond cleavage and formation propelled the scientific community to search for additional examples of non-traditional biological roles of this important biomolecule.

We are now much more aware of the variety and type of functions RNA possesses. For example, newly synthesized tRNAs must be modified prior to their use in translation.

The pre-tRNAs are processed by the enzyme RNase P, an RNA-protein complex (for a review, see Pace & Smith, 1990). The RNA component is now known to be responsible for the enzymatic activity of RNase P (Guerrier-Takada et al., 1983, Altman, 1989, Cech, 1993). Many mRNA transcripts must also be edited before they contain a coherent, translatable message. This processing involves the excision of large portions of the RNA that do not encode part of the protein sequence, and subsequent reattachment of the coding sections of the RNA. In the case of the RNA enzymes (ribozymes), this processing is carried out by the RNA itself (Kruger et al., 1982, Cech & Bass, 1986, Peebles et al., 1986, Veen et al., 1986). Other RNAs are processed by the spliceosome, a complex of small nuclear RNAs (snRNAs) and proteins (Moore et al., 1993, Black et al., 1985, Baserga & Steitz, 1993). There is current speculation that RNA is the active component of the spliceosome (Madhani & Guthrie, 1994). It has also been proposed that rRNA plays an active role in the ribosome. In addition to its structural role as a mere scaffold for ribosomal proteins, rRNA may be directly involved in the chemistry of the peptidyl transfer reaction (Moore, 1993, Noller, 1991, Noller et al., 1992).

Accompanying the many newly detected *in vivo* functions of RNA, *in vivo* selection techniques (Bartel & Szostak, 1994, Tuerk & Gold, 1990) have resulted in the discovery of a number of RNAs that posses novel activities. New ribozymes have been discovered (Bartel & Szostak, 1993, Lehman & Joyce, 1993), including those that can catalyze the cleavage of DNA substrates (Herschlag & Cech, 1990, Robertson & Joyce, 1990). RNAs have been found that are capable of binding a number of small ligands, including organic dyes and amino acids (Ellington & Szostak, 1990, Szostak & Ellington, 1993, Geiger et al., 1996).

In analogy with the structure-function relationship of proteins, the specific functions carried out by an RNA are directed by the conformation of the molecule. It is obvious from the variety of biological functions outlined above, that as a class of molecules RNA must be capable of forming a wide variety of three-dimensional structures. By studying the structure

of these molecules, the biophysical chemist hopes to gain insight into their active conformations.

Structure

The building blocks of nucleic acids are relatively simple. *Ribo*nucleic acids (RNA) and *deoxyribo*nucleic acids (DNA) have similar covalent structures. Each biopolymer is composed of monomer units of nucleotides. Canonical nucleotides consist of one of four different nucleotide bases attached to a furanose sugar (Figure 1.1). In both DNA and RNA, the sugars in a polynucleotide chain are connected by phoshodiester bonds between the 3' and 5' positions. One difference between the nucleic acids is replacement of uracil with the thymine base (5-methyluracil) in DNA. The most obvious distinction between the two biopolymers is the presence of the hydroxyl group at the 2' position in RNA (a hydrogen atom in DNA). This small chemical difference in the mononucleotides has a large effect on the global conformation of polynucleotide chains.

In order to relieve steric pressure, the 5-membered furanose rings pucker. Two major conformations are C3'-*endo* and C2'-*endo*. In C3'-*endo*, the 3' carbon is displaced approximately 0.4 Å above the plane of the other four ring atoms, in C2'-*endo* it is the C2' carbon that is displaced (Varani & Pardi, 1994). Mononucleotides are in fast exchange between the different conformations, with a strong preference for C2'-*endo* (Tinoco et al., 1987, Altona & Sundaralingam, 1972). In solution, single stranded nucleic acids have a tendency to stack in a helical fashion. As stacking occurs in the ribonucleotides, the sugar conformation shifts to predominantly C3'-*endo*. Oligodeoxynucleotides maintain their preference for the C2'-*endo* state (Altona, 1982).

Polynucleotide chains of both RNA and DNA form right-handed double helices of antiparallel strands (Saenger, 1984). The helix is stabilized by complementary Watson-Crick base pairing (hydrogen bonding) between purine (A, G) and pyrimidine (C, U [T])

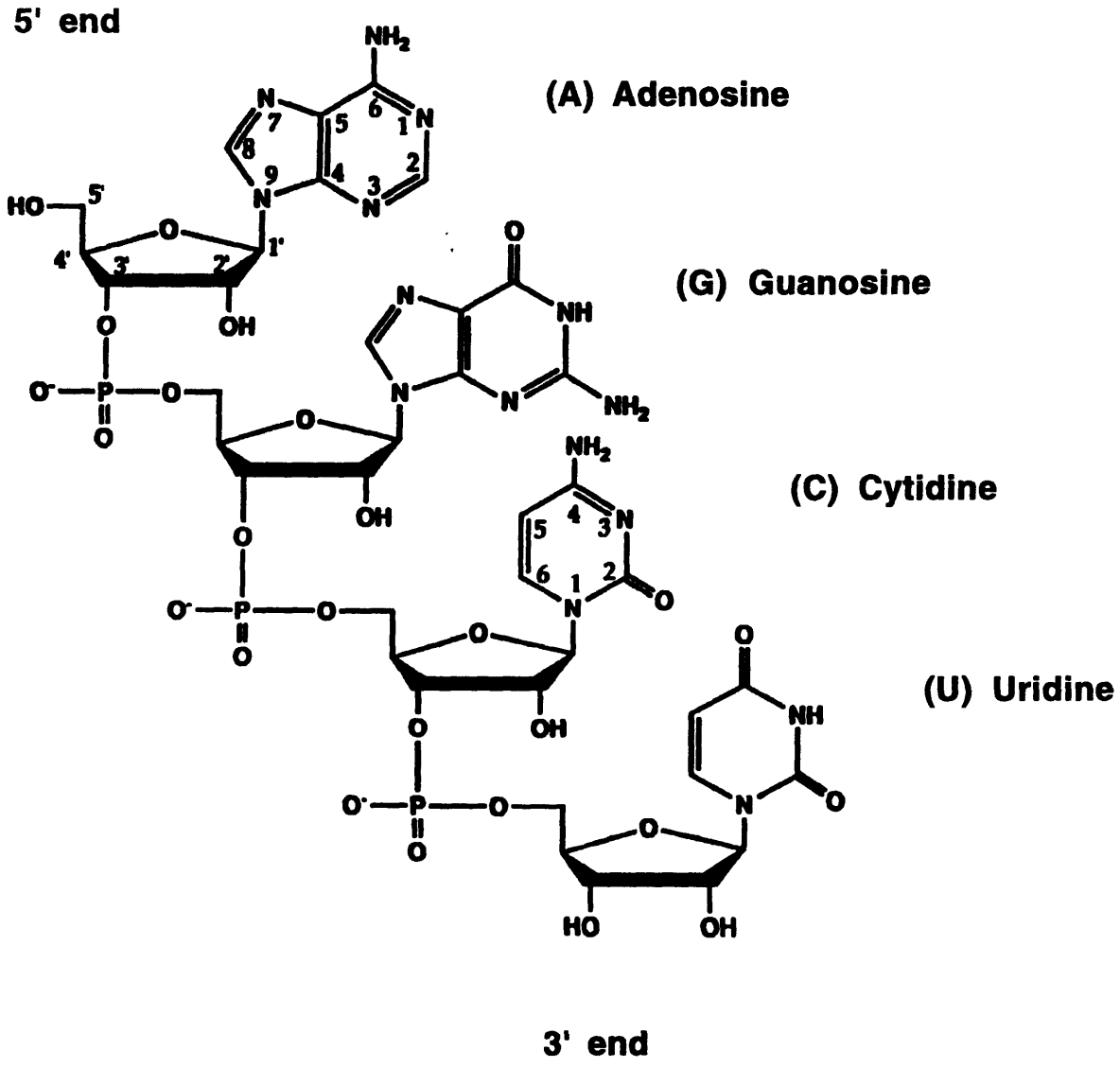


Figure 1.1) Nucleotide building blocks in RNA. Base and sugar atoms are numbered as shown. Purine (A,G) and pyrimidine (C,U) bases are attached to the ribose C1' by their N^9 or N^1 positions, respectively. Base H2, H6, and H8 are referred to as the aromatic protons. The ribose H1' is termed the anomeric proton.

nucleotides (Figure 1.2). Additional stabilization comes from stacking interactions between the planar, aromatic bases. Because of the different sugar puckers for the ribo- and deoxyribonucleotides, the helical geometry of RNA helices is very different from that of DNA helices. The C2'-*endo* sugar conformation in a DNA B-form helix results in a backbone phosphate-phosphate distance of approximately 7 Å (Figure 1.3). In RNA (A-form) helices the different sugar pucker shortens the interphosphate distance to 5.9 Å. Another major difference between the two helical forms is the displacement of the base pairs from the helical axis; whereas in the B-form helix, the center of the base pairs is close to the helical axis, in A-form RNA helices, the displacement is approximately 4 Å (Saenger, 1984). A result of the base pair displacement is a different helical groove size and shape. In A-form helices, the minor groove is wide (11 Å) and shallow (3 Å). However, the major groove is very narrow (3 Å) and deep (9 Å) in contrast to that of a B-form helix (12 Å wide and 9 Å deep) (Saenger, 1984). The different groove size and shape has important consequences for the biological functions of RNA. Most proteins interact with DNA helices by recognizing sequences in the major groove (Steitz, 1993), yet the A-form RNA major groove is too narrow to accommodate protein recognition. The minor groove of nucleic acids lacks suitable information markers (in the form of hydrogen-bond acceptors and donors) to distinguish between the G-C and A-U (or T) base pairs (Seeman et al., 1976). RNA circumvents this problem by assuming a wide variety of different shapes (see below) capable of recognition by proteins and other molecules.

Unlike the double-stranded DNA helix, the helical elements in RNA are predominantly intramolecular. A single polynucleotide chain folds back upon itself in order to form helical regions. The way in which an RNA strand folds upon itself is called the secondary structure. To date, a handful of RNA secondary structural motifs have been identified. These include double-helical regions (stems, duplexes), loops, stem-loop structures (hairpins), base mismatches, bulges, internal loops, and junctions between helices

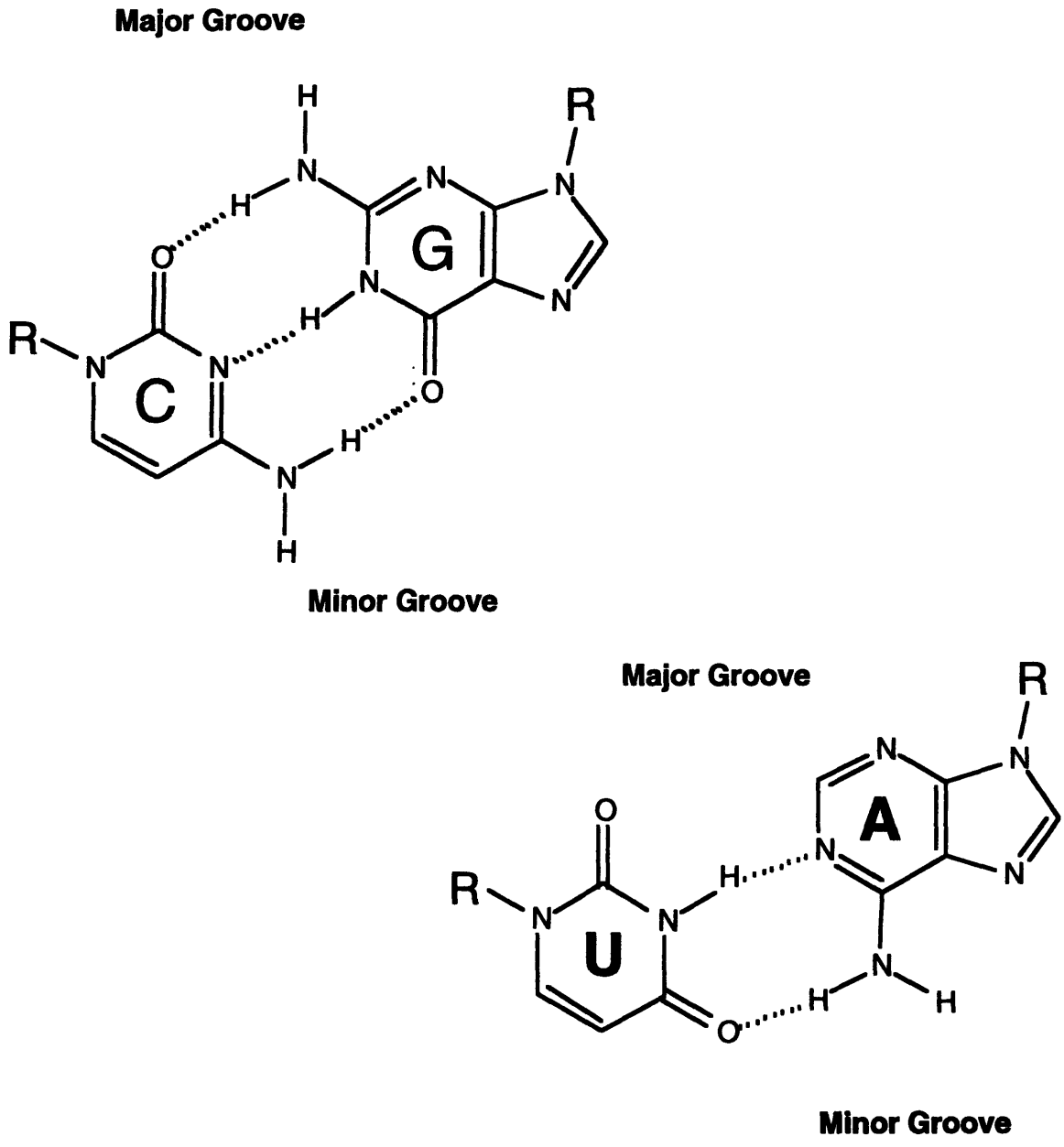


Figure 1.2) Canonical Watson-Crick base pairing in RNA. Complementary base pairing in RNA involves hydrogen bonding (dashed lines) between G and C bases, and between A and U bases. Ribose moieties are denoted by **R**. Both base pairs are similar in shape, with correspondingly similar ribose C1'-C1' distances (Saenger, 1984). The stacking of base-pairs in RNA causes the formation of right-handed nucleic acid helices.

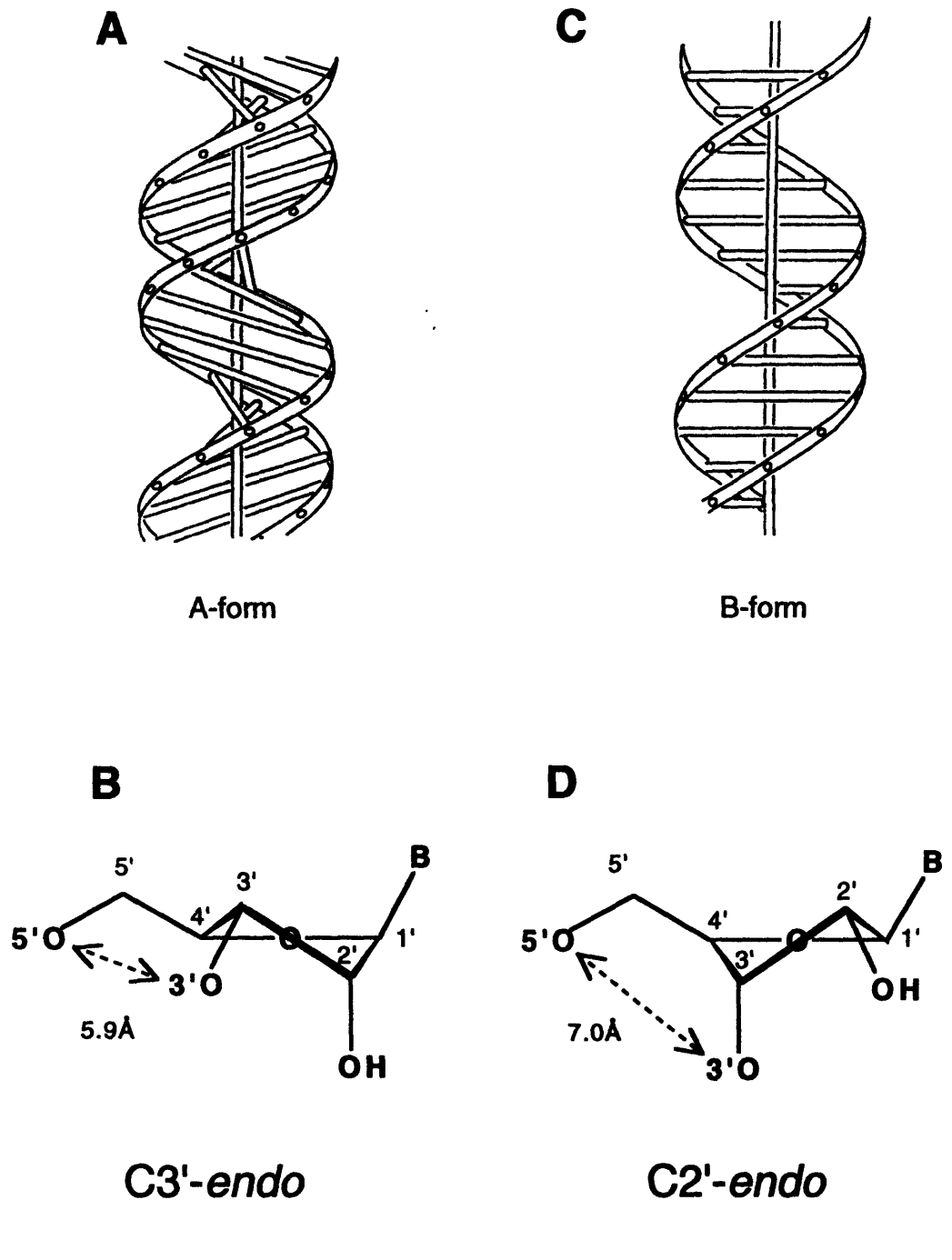


Figure 1.3) Nucleic acid helices. A) Schematic representation of an A-form RNA helix. **B)** The C3'-*endo* sugar pucker responsible for the geometry of an A-form helix. **C)** Schematic representation of a B-form helix as observed in DNA. **D)** The C2'-*endo* sugar pucker found in deoxynucleotides and unstacked ribonucleotides.

(for a review see Chastain & Tinoco, 1991, Tinoco et al., 1987). As shown in Figure 1.4, a single polynucleotide chain may contain several secondary structural motifs connected by single-stranded regions. These structural elements can further interact with each other to produce a compact, globular structure. Tertiary interactions result when secondary structural elements interact with single stranded regions or other structural motifs. These include the formation of base triples and triple helices, helix-helix interactions, loop-loop and loop-helix interactions. Frequently, divalent metal ions (e.g. magnesium) are required as counterions to the negatively charged nucleic acid backbone in order to stabilize compact RNA structures (Pan et al., 1993).

Although the field of RNA structural biology is still in its infancy, there have been a number of important structural studies recently performed (for reviews see Shen et al., 1995 and Moore, 1992). Advances in cloning and the ability to produce large quantities of RNA by chemical (Usman et al., 1987, Usman et al., 1992, Wincott et al., 1995) or enzymatic synthesis (Milligan et al., 1987, Wyatt et al., 1991) have enabled the structural biologist to produce large amounts of RNA of a defined sequence. Therefore, in addition to phylogenetic comparisons (Westhof & Michel, 1994) and chemical and enzymatic probing of secondary structure (Peattie, 1979, Ehresmann et al., 1987, Gaur & Krupp, 1993), high resolution studies using techniques that require large quantities of material, such as X-ray crystallography and nuclear magnetic resonance (NMR), are now feasible. A number of secondary structural elements have been studied using NMR and X-ray crystallography. Examples include RNA duplexes (Dock-Bregeon et al., 1989, Holbrook et al., 1991, Cruse et al., 1994, Leonard et al., 1994), bulges (Baeyens et al., 1995, Portmann et al., 1996), and hairpins (Puglisi et al., 1990b, Davis et al., 1993), including the ultra-stable tetraloops (Varani et al., 1991, Heus & Pardi, 1991a, Jucker & Pardi, 1995). An interesting study has just been published (Yang et al., 1996) in which the NMR structures of two hairpin RNAs that bind different amino-acids were determined. These

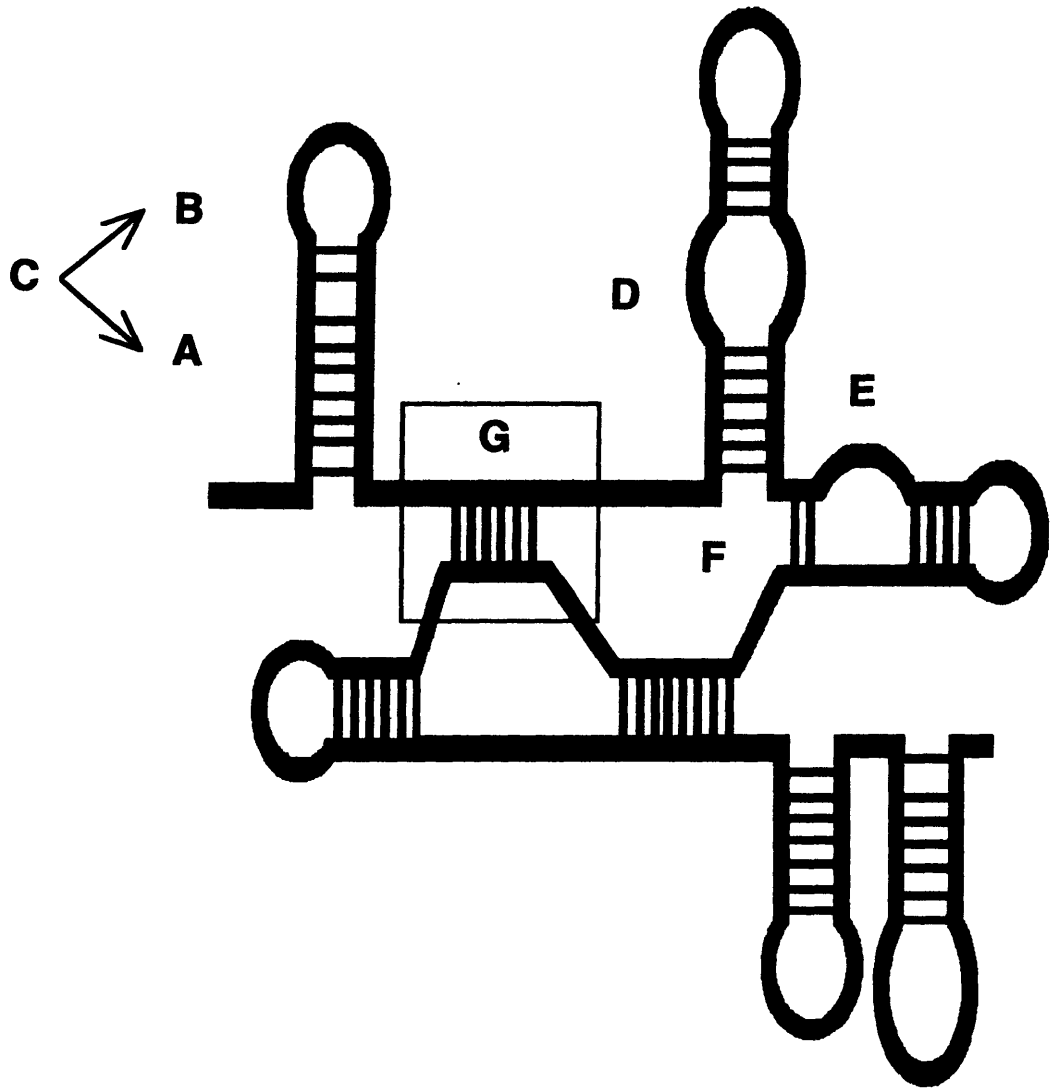


Figure 1.4) Complex secondary structure of an RNA. This schematic drawing of the core structure (see Davies *et al*, 1982 and Michel *et al*, 1982) of a group I intron illustrates a variety of known RNA secondary structural elements: **A)** helix; **B)** loop; **C)** hairpin; **D)** internal loop; **E)** bulge; and **F)** junction. Complex, tertiary interactions such as the long-range pairing shown in **G)** help this molecule to fold into a compact three-dimensional structure (Michel and Westhof, 1990). The sugar-phosphate backbone is represented by a thick line, base pairs are depicted by thinner horizontal or vertical lines.

RNAs, obtained using *in vitro* selection techniques (Famulok, 1994), differ at only three of 33 positions yet each is capable of selectively binding to either arginine or citrulline.

As a class of molecules, the most well-studied RNAs are the tRNAs. Indeed, until very recently, most of what was known about RNA tertiary interactions and RNA-protein interactions came from work with tRNA. Several crystal structures of free tRNAs, and those complexed with their aminoacyl-tRNA synthetases have been published (see, for example Kim et al., 1974, Woo et al., 1980, Westhof et al., 1985 Rould et al., 1989, Basavappa & Siegler, 1991, Biou et al., 1994). NMR studies that have provided information about RNA tertiary interactions include pseudoknot structures located at the 3' ends of plant viral RNAs (Puglisi et al., 1990a), a solution structure of an RNA tetraplex (Cheong & Moore, 1992), and a model of the hairpin-hairpin interaction in ColE1 transcripts RNA I and RNA II (Marino et al., 1995). Recent crystallographic studies have yielded the structure of a hammerhead ribozyme (Pley et al., 1994), as well as that of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin (Oubridge et al., 1994).

With the exception of tRNAs, most biologically important RNA molecules are too large and (perhaps) too flexible for high resolution studies using NMR or X-ray crystallography (Shen et al., 1995). Researchers therefore must use a 'reductionist approach' to designing RNA molecules for study by these techniques. Portions of the RNA of interest are examined for activity analogous to that of the parent molecule. Thus, functionally important structural subunits of the RNAs are structurally characterized. Recent examples of this approach include several NMR studies of fragments from ribosomal RNAs (Varani et al., 1989, White et al., 1992, Wimberly et al., 1993, Szewczak et al., 1993, Fountain et al., 1996), group I introns (Chastain & Tinoco, 1992, Allain & Varani, 1995, Nowakowski & Tinoco, 1996), and parts of the equine infectious anemia virus (EIAV) (Hoffman et al., 1993), the bovine immunodeficiency virus (BIV) (Puglisi et al., 1995), and

the human immunodeficiency virus (HIV) (Puglisi et al., 1992, Colvin et al., 1993, Jaeger & Tinoco, 1993, Battiste et al., 1994).

The library of known RNA structures is small, but it is quickly growing. From the studies of RNA fragments, some structural trends have become apparent. In addition to hydrogen bonding, base stacking interactions are important factors in determining RNA conformations (Shen et al., 1995). Non-Watson-Crick base pairings are common, and may play an important role in protein recognition. The structures adopted by hairpins, bulges and internal loops, as well as their interactions with other structural elements and/or proteins, are now thought to be significant in the many varied biological roles of RNA. As mentioned above, the structure of an A-form RNA helix limits recognition by proteins in the major groove. Structural elements such as mismatches, base triples, bulges, and internal loops, can distort the RNA helix facilitating protein recognition. Alternatively, proteins and other biomolecules can interact with local RNA structural motifs in a sequence-independent manner. With NMR and X-ray crystallography of small RNAs becoming more common, the focus of many research efforts is now being shifted toward determining how these structural building blocks are arranged to form global structures in large RNAs, and how these RNAs interact with proteins and other molecules to exert their biological functions.

1.2 HIV Biochemistry

Organization and Expression of Genes

Replication of the human immunodeficiency virus (HIV) (the virus responsible for the acquired immunodeficiency syndrome [AIDS]) is controlled by two important RNA-protein interactions. The virally encoded *trans*-activator (Tat) and the regulator of virion expression (Rev) regulatory proteins interact with stem-loop structures found in newly synthesized viral RNA. Because these interactions are crucial for viral replication, both the

proteins and their RNA ligands constitute possible targets for anti-HIV drug therapy. Thus, the structure of the RNAs and the nature of the RNA-protein interactions are of interest to both the structural biologist and the clinical scientist.

The AIDS virus is a member of a family of RNA viruses called retroviruses. There are three sub-families of retroviruses (Schüpbach, 1989): the oncoviruses, the lentiviruses, and the spumaviruses. HIV is classified as a lentivirus, a retrovirus associated with slowly progressing inflammatory and degenerative disorders (Schüpbach, 1989). The general structure of a retrovirus is comprised of a dense inner core composed of RNA and protein surrounded by an outer membrane containing glycoproteins. The nucleic acid in the ribonucleoprotein core exists as two copies of the single-stranded RNA genome. The RNAs are capped and polyadenylated, similar to the mRNAs of eucaryotic cells. In order for replication to occur the RNA must be converted via reverse-transcription into a double-stranded DNA intermediate that is integrated into the host cell genome as a provirus (Cullen, 1991).

At the simplest level, the retroviral genome contains three genes encoding proteins required for the virus life-cycle. These are *gag* (*g*roup specific *a*ntigen) that encodes the structural proteins, *pol* (RNA-dependent DNA *pol*ymerase) that encodes the enzyme responsible for the reverse transcription, and *env* which codes for the viral *env*elope glycoproteins (Schüpbach, 1989). These genes are always presented as gag-pol-env (Figure 1.5 A) and are flanked on both ends by regulatory sequences called R and U5 at the 5'-end, and R and U3 at the 3'-end. During reverse transcription these regulatory sequences are duplicated resulting in identical units of U3-R-U5 called LTRs (long terminal repeats) (Haseltine, 1991) at both ends of the genome (Figure 1.5 B). The 5' LTR promotes viral transcription, the 3' LTR is required for efficient polyadenylation of viral transcripts (Cullen, 1991).

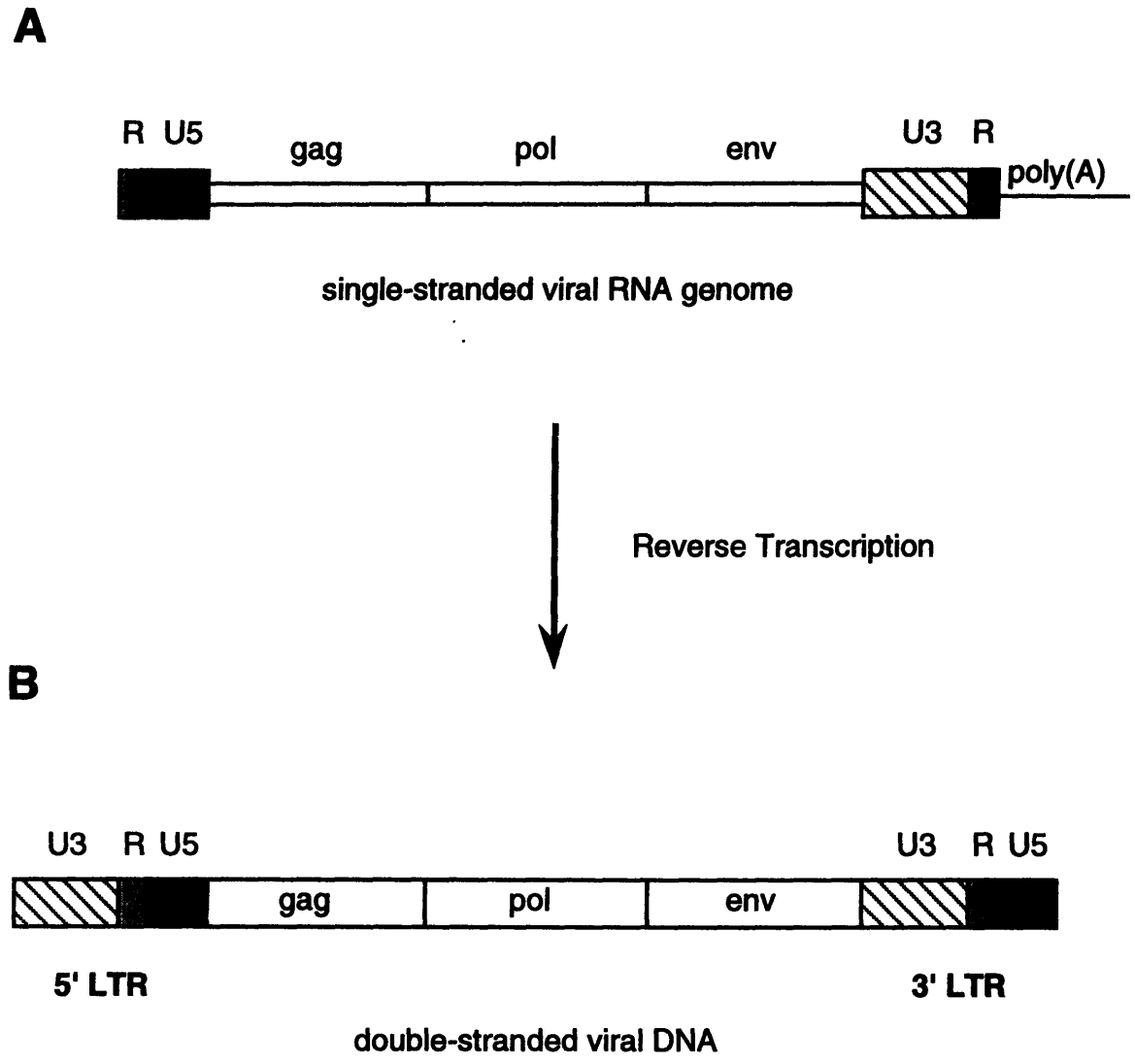


Figure 1.5) Genomic structure of a simple retrovirus. A) The single stranded RNA genome contains the *gag*, *pol* and *env* structural genes flanked by the U5, U3 and R regulatory sequences. **B)** Upon reverse transcription, the regulatory sequences are duplicated to produce identical long terminal repeats (LTR) at each end of the genome (after Schüpbach, 1989).

HIV is considered to be a complex retrovirus. Although the *gag*, *pol*, and *env* gene products are necessary for viral replication, they are not sufficient (Cullen, 1991). HIV encodes at least six other gene products that regulate the viral life cycle (Figure 1.6). In addition to *gag*, *pol*, and *env* genes, HIV contains the early regulatory genes *tat*, *rev*, and *nef*. The interaction of the Tat protein with the 5'-end of viral transcripts (the *trans*-activation response element, or TAR, region of the 5' LTR) serves to activate viral transcription. Similarly, Rev interacts with viral RNA (at the Rev responsive element, or RRE) to regulate the splicing of the primary transcript and transport of mRNA from the nucleus. The *nef* gene product, Nef protein, is essential for disease progression, but the mechanism of its action is poorly understood. Perhaps an understanding of the structure of the protein will give insight into its biological function. In fact, a crystal structure of the conserved core of this protein has recently been solved (Lee et al., 1996). HIV also employs the late regulatory genes *vif*, *vpu* and *vpr*, whose products affect infectivity, export and replication capacity of the newly made virus particles (Haseltine, 1991). A complicated splicing pattern is employed in order to produce many spliced mRNAs necessary to produce several gene products from a single primary transcript (Schwartz et al., 1990, reviewed in Vaishnav & Wong-Staal, 1991). The structural (capsid and envelope) proteins arise from un- (~9 kb) or singly-spliced (~4 kb) mRNA, while the regulatory gene products are produced from multiply-spliced (~2 kb) mRNA (Cullen, 1991).

Virus expression begins with the synthesis of a complete RNA copy of proviral DNA. The viral promoter and enhancer sequences are located in the LTR and span a region of only 250 nucleotides (Jones & Peterlin, 1994). The rate of transcription initiation is controlled by host proteins that affect the rate of cellular gene transcription (reviewed in Jones & Peterlin, 1994). The HIV LTR promoter is a weak promoter, allowing only a low basal level of HIV gene expression in infected cells. This is due, in part, to the poor

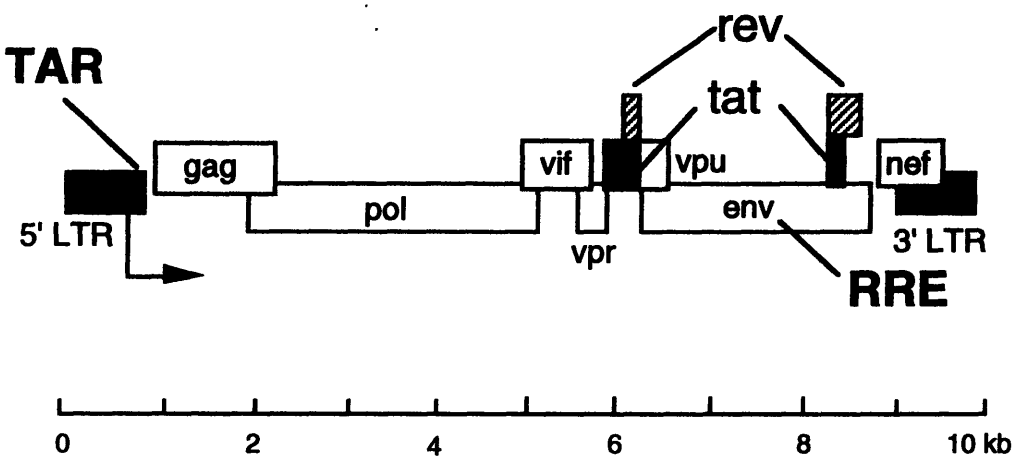


Figure 1.6) Complex genomic structure of the retrovirus HIV. The ca. 10 kb genome of HIV is organized as shown. In addition to the structural genes *gag*, *pol* and *env*, HIV contains the regulatory genes *tar*, *rev*, *nef*, *vif*, *vpr* and *vpu*. The *tat* gene product, the Tat protein, interacts with the mRNA in the TAR region of the 5' LTR. Rev binds to the RRE located in the *env* gene.

processivity of the RNA polymerase (Barry et al., 1991). These initial transcripts reach the cytoplasm for translation exclusively in the multiply-spliced 2 kb form (Kim et al., 1989, Cullen, 1991), resulting in the synthesis of the regulatory proteins Tat and Rev. The presence of a small amount of the regulatory protein Tat activates further transcription by allowing the synthesis of full length transcripts (Laspia et al., 1990, Marciniak et al., 1990a, Ritter et al., 1995), establishing a 'positive feedback loop' (Karn et al., 1994) that boosts Tat and Rev mRNA production 200-1000 fold (Ayra et al., 1985, Sodroski et al., 1985a Sodroski et al., 1985b). The increased transcription results in very high levels of HIV-specific RNA and protein synthesis in the infected cell (Cullen, 1991).

It is thought that the switch from production of the early regulatory proteins to that of the late, structural proteins requires a certain critical level of Rev (Pomerantz et al., 1990). When sufficient levels of Rev are present, unspliced and singly spliced mRNAs can be detected in the cytoplasm (Feinberg et al., 1986, Hammarskjold et al., 1989, Malim et al., 1989). HIV gene expression shifts away from the expression of the mRNAs for the regulatory proteins towards the production of mRNAs for the virion proteins (Sodroski et al., 1986). The synthesis of capsid and envelope glycoproteins can now be accomplished, and therefore new virus particles can be made.

HIV Tat

The HIV Tat protein is expressed from three different multiply spliced mRNAs (Schwartz et al., 1990). The predominant form of Tat found in HIV-infected cells is 86 amino acids long (16 kD) (Aldovini et al., 1986), and is encoded by two exons (Ayra et al., 1985). It has been shown (Sodroski et al., 1985a) that the N-terminal 72 amino acids (encoded by the first exon) are sufficient for full *trans*activation activity; substantial activity is seen with even shorter fragments corresponding to the first 60 amino acids (Kuppuswamy et al., 1989, Garcia et al., 1989). The HIV protein can be divided into

discrete segments (Figure 1.7) according to homology with Tat from other lentiviruses (Kuppuswamy et al., 1989, Dorn et al., 1990, Derse et al., 1991). These include the N-terminal, cysteine-rich, core, basic, glutamine-rich, and C-terminal regions. Based on mutagenesis and domain-swapping experiments (Rice & Carlotti, 1990a), the core, cys-rich, and amino terminal regions of Tat are thought to constitute an independent *trans*activation domain. Nuclear localization and the ability to bind to TAR are specified by the basic domain sequences (Ruben et al., 1989 Hauber et al., 1989). Circular dichroism analysis (Slice et al., 1992) of HIV Tat indicates the presence of 15-20% alpha-helical structure in aqueous solution, and the helical content is increased upon the addition of trifluoroethanol (TFE). A model of the structure of the full-length Tat protein based on NMR and molecular dynamics calculations has recently been published (Bayer et al., 1995). These authors find no evidence of stable secondary-structural elements in the intact protein, and no induction of alpha-helix upon addition of TFE.

In the absence of Tat, short transcripts (60-80 nucleotides) of the HIV mRNA predominate, terminating near the end of the viral LTR (Toohey & Jones, 1989, Ratnasabapathy et al., 1990 Kessler & Matthews, 1992). The presence of Tat allows the production of full-length mRNAs (Kao et al., 1987). Nuclear run-off experiments have shown that the majority of polymerases initiating transcription stall near the promoter in the absence of Tat (Laspia et al., 1989), and that upon addition of the protein, there is a dramatic increase in the number of polymerases found downstream of the promoter (Laspia et al., 1990, Feinberg et al., 1991, Marciniak & Sharp, 1991). Deletion analysis of the LTR indicated that the activity of Tat requires the *trans*activation response region (TAR), a regulatory element located immediately downstream of the initiation site for transcription (Rosen et al., 1985, Cullen, 1986, Muesing et al., 1987). The location of a regulatory element within a transcribed region allows for the possibility that Tat could be interacting with an RNA form of TAR rather than a DNA element. Indeed, TAR was found to only be

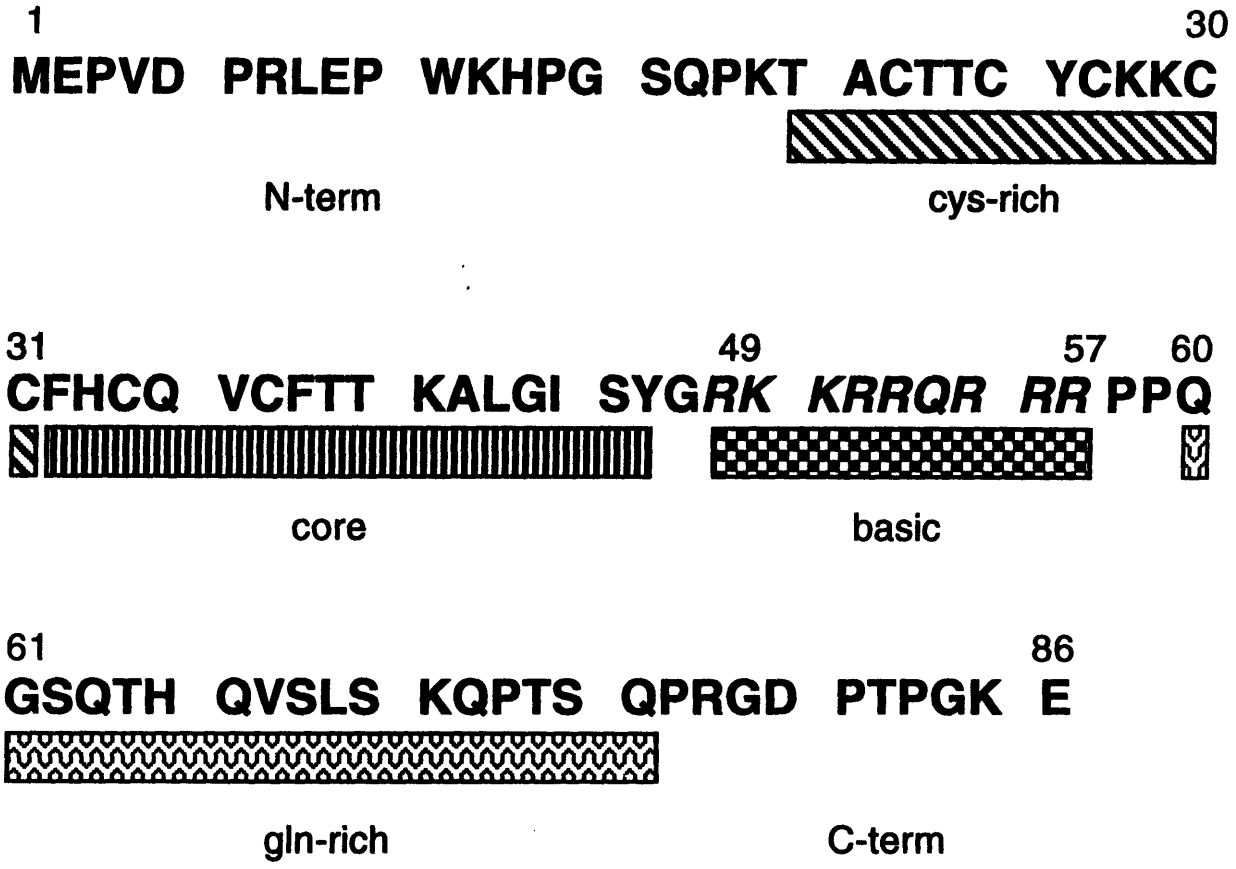


Figure 1.7) HIV-1 Tat protein sequence. Different regions of the HIV Tat protein are denoted by patterned boxes. Core, cysteine-rich and amino-terminal regions are thought to constitute an independent *trans*activation domain. The basic and core regions are highly conserved among all known Tat proteins, although the cysteine-rich region is absent in the equine infectious anemia virus (EIAV) Tat. The arginine-rich basic region (R49-R57) responsible for RNA binding is shown in italics.

functional when it is placed on the 3' side of the HIV promoter, and in the correct orientation and position (Feng & Holland, 1988, Selby et al., 1989, Berkhout et al., 1989a). When a TAR anti-sense sequence was inserted upstream of the normal TAR sequence, the formation of the TAR hairpin was prohibited, and the Tat response was inhibited (Berkhout et al., 1989a). That the TAR RNA forms a stable nuclease-resistant stem-loop structure also supported the hypothesis that Tat acts upon the RNA form of TAR (Muesing et al., 1987). *In vitro* binding assays confirmed that Tat binds to TAR RNA, and not to antisense TAR, cellular mRNAs, or to TAR DNA (Dingwall et al., 1989, Roy et al., 1990a). When mutations that inhibit Tat binding *in vitro* were introduced into the viral LTR, a reduction in *trans*activation was observed (Roy et al., 1990a, Dingwall et al., 1990).

Because of the high susceptibility of Tat to proteolytic cleavage (Frankel & Pabo, 1988), fragments corresponding to the C-terminal portions of the Tat protein (residues 49-72, and 49-86) were isolated by Weeks and coworkers (Weeks et al., 1990) while purifying recombinant full-length protein. These authors demonstrated that the proteolytic fragments were able to bind to TAR RNA with similar affinity and specificity as intact Tat. This and subsequent work (Cordingley et al., 1990, Roy et al., 1990a, Calnan et al., 1991b) indicated that the RNA binding capability of Tat was localized to only nine amino acids (R49-R57) located in the arginine rich basic region (Figure 1.7). This RNA-binding region is remarkably flexible. Frankel and coworkers (Calnan et al., 1991a, Calnan et al., 1991b) have shown that these nine amino acids can be scrambled, reversed, or even replaced with all arginines (or eight lysines and one arginine) and still retain TAR binding specificity similar to wild-type Tat peptide (RKKRRQRRR). A full-length Tat protein in which the basic region has been changed to -KKKRKKKKK- is able to *trans*activate at wild-type levels *in vivo* (Calnan et al., 1991a, Calnan et al., 1991b, Tao & Frankel, 1993). For electrostatic reasons, the number and position of charged residues in the basic region

are thought to be important for high affinity peptide binding and *trans*activation *in vivo* (Calnan et al., 1991a, Calnan et al., 1991b, Delling et al., 1991, Subramanian et al., 1991, Tao & Frankel, 1993, Long & Crothers, 1995).

The requirement for only one arginine (R52) in the basic peptides propelled Tao and Frankel to test free arginine and arginine analogs for TAR binding (Tao & Frankel, 1992). Arginine is capable of competing with Tat peptides for TAR RNA, whereas N-methylated arginine or other charged residues such as lysine are not. Interestingly, argininamide, an arginine analog in which the negatively charged carboxylate group has been replaced with a neutral amide functionality, was found to be slightly more effective at this competition than arginine (Tao & Frankel, 1992). Arginine binds to TAR with the same specificity as the Tat derived peptides, albeit with lower affinity. The binding constants (K_ds) for Tat protein and Tat peptides are in the nanomolar range (Weeks et al., 1990, Calnan et al., 1991b, Dingwall et al., 1990); that for arginine is in the millimolar range (Tao & Frankel, 1992, Tan & Frankel, 1992). Additionally, it was shown that the guanidine moiety alone is capable of TAR RNA recognition (Tan & Frankel, 1992). Circular dichroism studies have also shown that the TAR RNA exhibits a similar change in conformation upon binding of peptides, arginine, or guanidine (Tan & Frankel, 1992). The conformational changes upon arginine binding were also demonstrated by transient electric birefringence (Zacharias & Hagerman, 1995) and NMR (Puglisi et al., 1992, see below).

HIV TAR

The TAR region was initially mapped between nucleotides –17 and +80 of the HIV LTR (Rosen et al., 1985). Nucleotides 1-59 in TAR RNA form a hairpin structure (Figure 1.8 A). The most critical region for *trans*activation by Tat is composed of the stem-loop region of nucleotides +18 to +44 (Figure 1.8 A) (Feng & Holland, 1988, Hauber & Cullen, 1988, Garcia et al., 1989 Selby et al., 1989, Roy et al., 1990b). The binding site

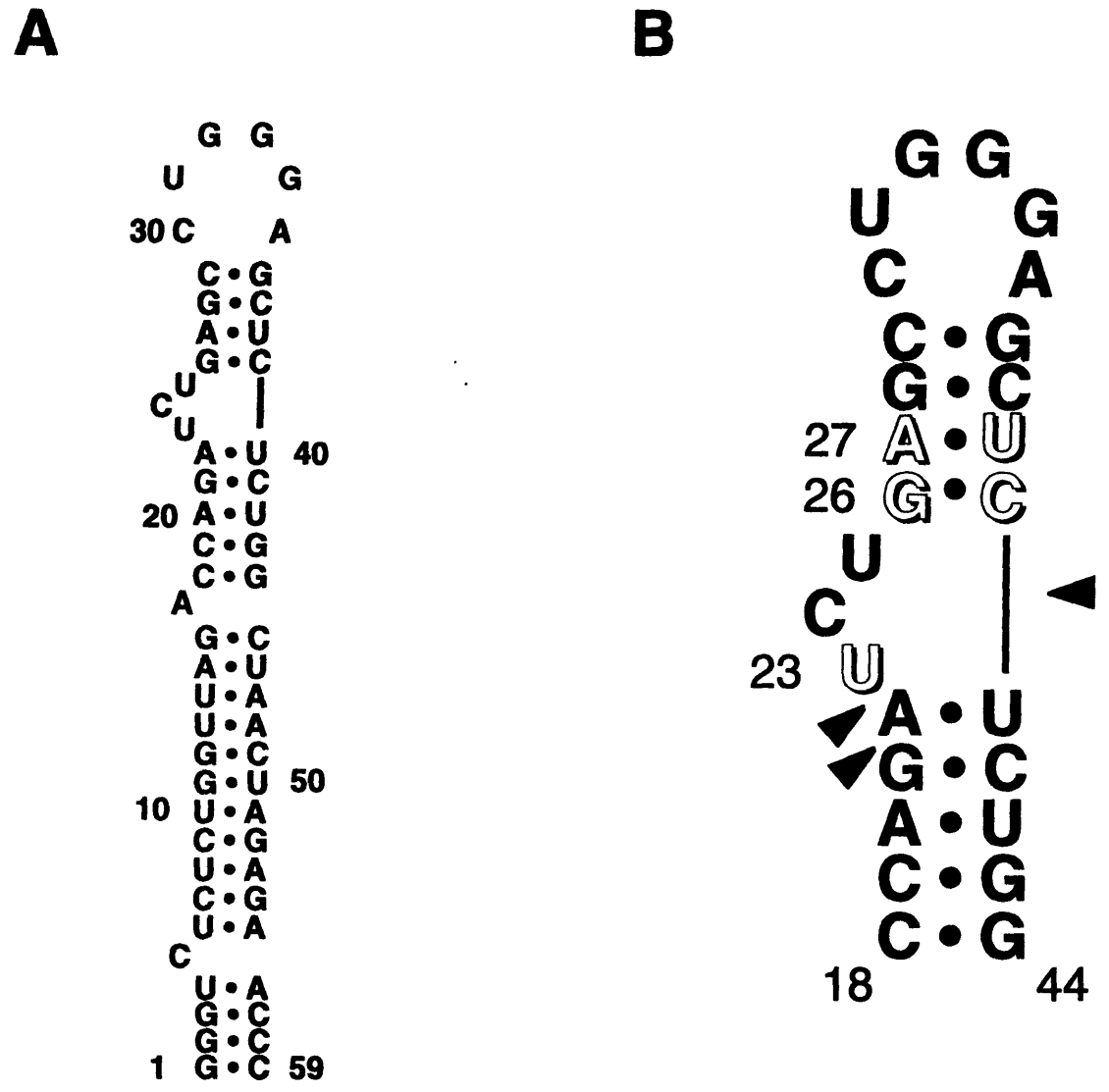


Figure 1.8) HIV TAR sequence. A) Hairpin structure of nucleotides 1-59 of the HIV TAR RNA region. B) Region of TAR critical for *trans*activation and Tat binding, C18- through G44. Base-pairs G26-C39 and A27-U38, as well as the bulge U23 (outline) are important for protein binding. Phosphates involved in protein binding (P22, P23, phosphates opposite from the bulge) are indicated by black arrow heads. The loop is involved in *trans*activation, but not Tat binding.

for Tat is located in the bulge region of TAR. Residues crucial for Tat binding include U23, and the two base pairs immediately above the bulge (G_{26}-C_{39}, and A_{27}-U_{38}) (Roy et al., 1990a, Dingwall et al., 1990, Berkhout & Jeang, 1991, Churcher et al., 1993, Delling et al., 1992) (Figure 1.8). The other bulge nucleotides do not appear to be critical for Tat binding, and may be replaced with other nucleotides, or a non-nucleotide spacer (Sumner-Smith et al., 1991). The bulge may serve to widen the major groove of the RNA helix, allowing for increased accessibility to the base pairs necessary for specific recognition (Weeks & Crothers, 1991). Mutational (Weeks & Crothers, 1991) and functional group (Hamy et al., 1993) analyses have shown that the Tat protein recognizes the RNA in the major groove, and that the N^7 positions of G_{26} and A_{27} are essential for this recognition. The N^3 position of U23 has also been shown to be important for *tat* binding (Churcher et al., 1993, Hamy et al., 1993). The phosphates between G_{21} and A_{22} (P_{22}) and A_{22} and U_{23} (P_{23}) have been shown via ethylation interference experiments to be important for interaction with *Tat, Tat* peptides and arginine (Calnan et al., 1991a, Tao & Frankel, 1992, Churcher et al., 1993). A recent study using methylphosphonate substitutions at specific positions of the backbone has indicated that phosphates P_{36}-P_{40} on the strand opposite from the bulge are also important for Tat binding (Pritchard et al., 1994). The hexanucleotide loop sequence appears to be important for *trans*activation (Feng & Holland, 1988, Berkhout & Jeang, 1989b), but not for Tat binding to the TAR element (Roy et al., 1990a, Weeks et al., 1990). It may be involved in an interaction with cellular proteins (Frankel, 1994, Marciniak et al., 1990b, Sheline et al., 1991, Wu et al., 1991).

As shown in Figure 1.9, the determinants in TAR required for *trans*activation, binding of Tat protein, Tat peptides or arginine are essentially the same (Frankel, 1994, Puglisi et al., 1993). As described above, the important functionalities of TAR are localized in a rather small region of the RNA. Also, the amino acid arginine alone is capable of mimicking the specific recognition of TAR by Tat. Taken together, this system

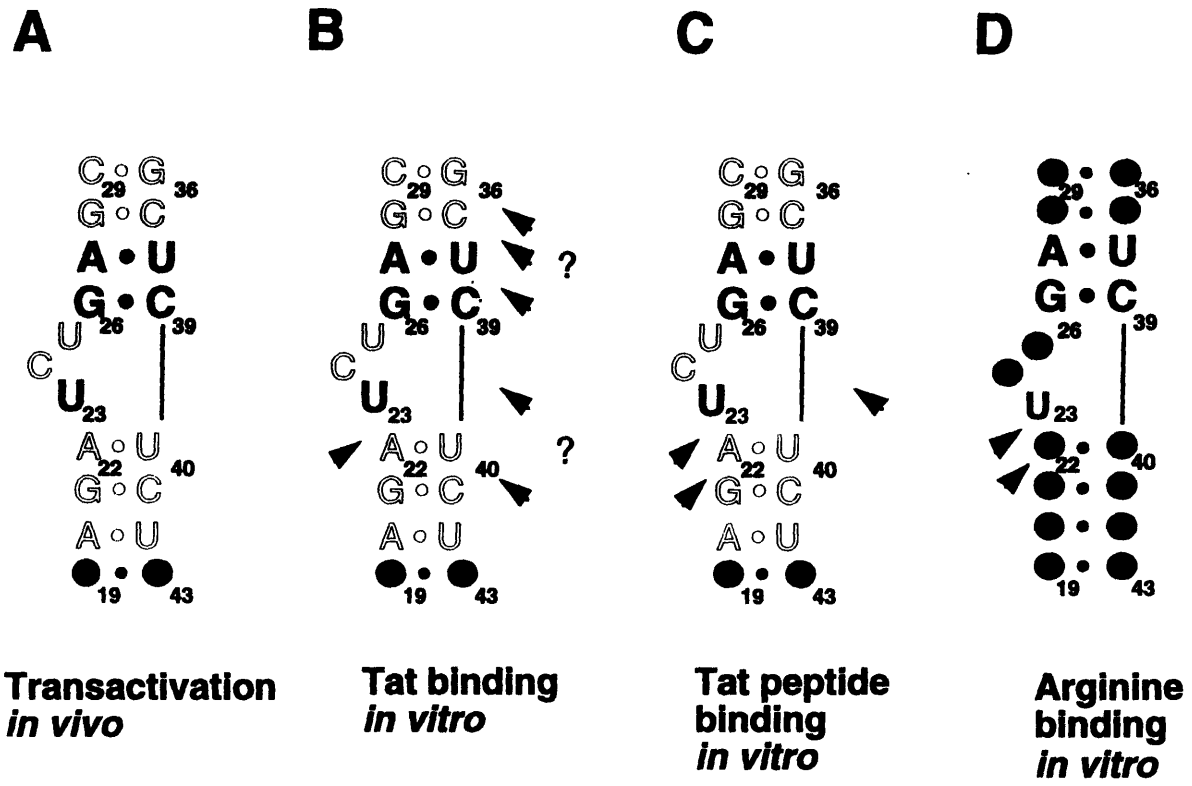


Figure 1.9) Summary of TAR sequence requirements. Nucleotides important for A) transactivation. B) Tat protein binding C) Tat peptide binding D) arginine binding. A black circle represents positions that have not been explicitly studied. Mutation of the outlined nucleotides had no effect on TAR function. Nucleotides critical for function are shown in black type. Phosphates thought to be involved in binding are indicated by an arrow. A question mark denotes phosphates that might be important. The loop is not necessary for Tat binding, therefore it is not shown.

lends itself perfectly to structural analysis of the Tat-TAR interaction using NMR spectroscopy.

Tat-TAR Structural Models

Structural studies of free TAR RNA and the complex with argininamide were performed in this laboratory by J.D. Puglisi (Puglisi et al., 1992). The results of his work are summarized below. The 31 nucleotide RNA corresponding to the minimal TAR is shown in Figure 1.10, and is hereafter referred to as TAR. Using NMR, he determined that TAR forms two A-form helical stem regions (the upper and lower stems) from G26 to A22 and from G26 to C29. The three bulge nucleotides (U23, C24 U25) are partially stacked between the two stems; however, this stacking does not cause a major distortion of the helical conformation opposite the bulge. The free conformation is consistent with a bending of the overall helical axis, as has been evidenced by circular dichroism (Tan & Frankel, 1992), gel electrophoresis (Riordan et al., 1992, Zacharias & Hagerman, 1995), and transient electric birefringence experiments (Zacharias & Hagerman, 1995). This helical kink has been shown to straighten upon complexation of the RNA with argininamide or Tat-derived peptides (Tan & Frankel, 1992, Zacharias & Hagerman, 1995, Long & Crothers, 1995).

In the NMR analysis, the addition of argininamide caused significant conformational changes in the spectra of TAR RNA. Chemical shift is a sensitive indicator of local environment of the nucleus being observed. Different chemical shifts can result from the molecule adopting a new conformation or from the binding of a ligand. Based on the changes in the NMR spectrum of the RNA, Puglisi *et al.* proposed that the bulge nucleotides become unstacked (evidenced by downfield shifting of their NMR resonances) with concomitant coaxial stacking of the two stems. U23 becomes positioned near A27 in the major groove of the upper-stem. Additionally, contacts were observed between

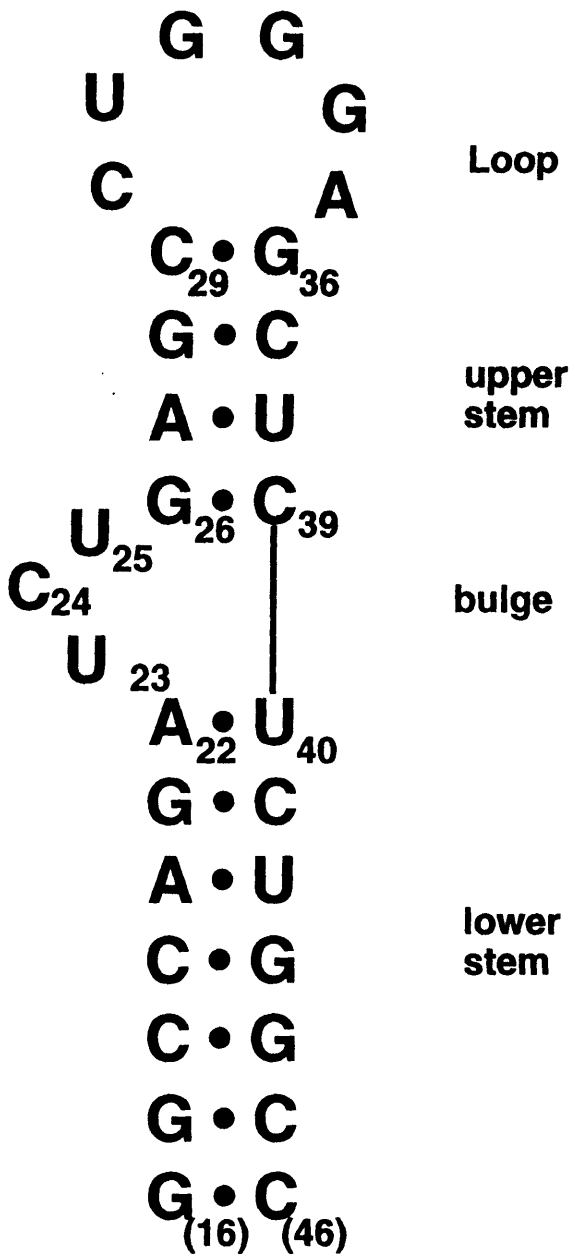


Figure 1.10) TAR RNA used for NMR. Analysis was performed on this 31 nucleotide RNA. The bases at the bottom of the lower stem were changed in order to increase transcription yields (compare with Figure 1.8). These changes do not effect binding of arginine.

argininamide and A22, U23, and A27. The binding constant for argininamide was estimated to be 2 mM. Lysine (50 mM) was found to have no effect on either the structure of the free RNA or the binding of argininamide. NMR analysis was also performed on mutant RNAs shown to have reduced binding affinity to Tat peptides. RNAs with the U23 to C23 mutation, or A27-U38 changed to U27-A38 were found to have a conformation similar to that of free wild-type TAR, even in the presence of 6 mM argininamide. No conformational changes were observed, and no contacts were seen between these mutants and argininamide.

Analysis was also performed on a complex between TAR and an 11 amino acid peptide (YKKKRKKKKKA). As discussed above, when placed in the context of the Tat protein, this sequence is capable of *trans*activation (Calnan et al., 1991a, Calnan, 1991b #151]). Similar results were obtained for the peptide-RNA complex and the argininamide-RNA complex. In neither case did the study focus on the structure of the loop. It was determined, however, that the loop does not undergo significant conformational change upon addition of argininamide or peptide. Other laboratories have focused their attention toward determining the loop structure (Jaeger & Tinoco, 1993, Colvin et al., 1993). These studies suggest that in the absence of any ligand, the loop structure is dynamic.

Based on the NMR data, a model of the conformation of the RNA bound to argininamide was developed (Figure 1.11). The bulge nucleotide at position 23 forms a Hoogsteen (Hoogsteen, 1959) base pair with A27. A base-triple interaction (Saenger, 1984) between U_{23} and the A_{27}-U_{38} base pair was proposed. The argininamide is positioned below U_{23}, near G_{26}, in the major groove and is thought to form hydrogen bonds to phosphates P_{22} and P_{23}. This model positions all of the nucleotides critical for Tat binding (U_{23}, G_{26}-C_{39}, A_{27}-U_{38}, P_{22} and P_{23}; see above) in close proximity in the presence of argininamide (they are distant in the absence of ligand). It also correlates with the chemical interference and mutational data discussed above.

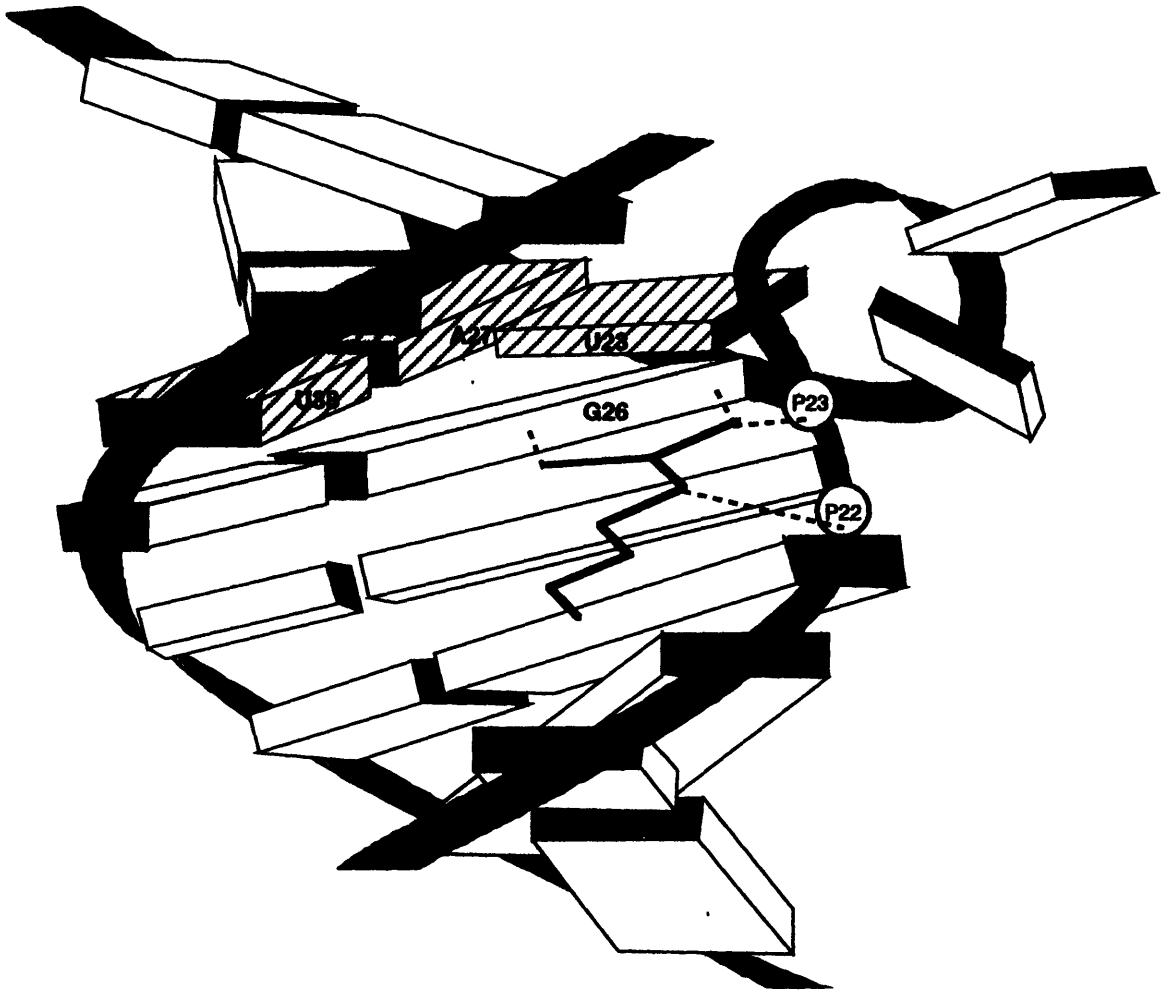


Figure 1.11) Schematic representation of TAR RNA bound to argininamide. Bases are depicted as rectangles, the backbone as a ribbon. Dashed lines represent hydrogen bonding of the argininamide to G26, P22 and P23. The bases involved in the triple interaction (U23•A27-U38) are hatched. Bulge nucleotides C24 and U25 are not well defined, but they are unstacked. The upper and lower stems are coaxially stacked.

In a subsequent study (Puglisi et al., 1993), Puglisi *et al.* synthesized an RNA in which the bases involved in the triple (U23, A27, U38) were mutated so as to form an isomorphic C^+•G-C triple (Figure 1.12). It was demonstrated that the two RNAs adopt similar conformations in the context of a complex with argininamide, and similar contacts to the amino acid were determined. As seen in DNA triplex work (Sklenar & Feigon, 1990), the triple mutant is stabilized at low pH by protonation of C23.

Other groups are using NMR to study the solution structure of TAR RNA. In a recently published article (Aboul-ela et al., 1995), Varani and coworkers have studied a complex between TAR and a 37 amino acid peptide containing the basic region of Tat. These authors observe similar conformational changes between the free and bound structures of the RNA, including the unstacking of the bulge residues upon peptide binding. They do, however, dispute the existence of the base triple interaction. Recently, single crystals have been obtained of TAR bound to a 14 amino acid peptide (McKenna et al., 1994). An X-ray structure of the TAR-peptide complex should provide insight into the conformation of the bound form of the RNA.

Using isotopically labeled RNA, detailed studies of a TAR molecule with a 2-base (UU) bulge have been performed in this laboratory in an effort to produce a more detailed model of the TAR-argininamide complex (A.S. Brodsky and J.R. Williamson, manuscript in preparation). In conjunction with this NMR and modeling project, it has been the purpose of this thesis work to investigate the conformational effects of mutations on the nucleotides involved in the base-triple using NMR techniques.

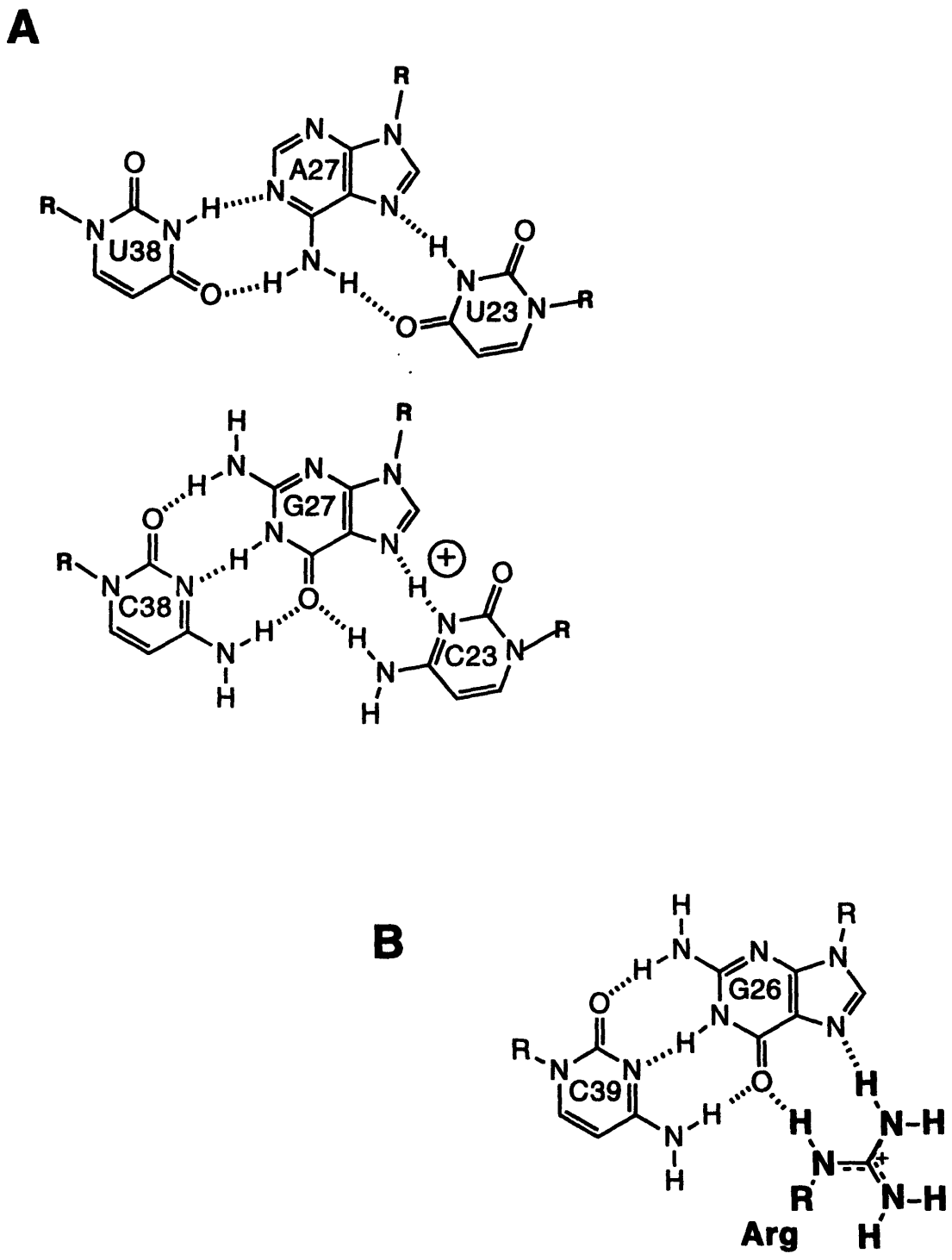


Figure 1.12) Comparison of TAR triples. A) The $C23^{+}$•G27-C38 mutant base-triple is isomorphic with the wild-type U23•A27-U38 base-triple. **B)** Proposed argininamide contact with G26. This interaction is the same in the mutant TAR as it is in the wild-type.

2 Materials and Methods

2.1 RNA Synthesis

Templates

RNAs were synthesized enzymatically using T7 runoff transcription (Milligan et al., 1987, Wyatt et al., 1991). This system employs the use of synthetic DNA templates, one of the promoter region plus the complement to the RNA to be synthesized, and a 17 nucleotide 'top-strand' to the promoter region (Figure 2.1). DNA oligos were synthesized on an Applied Biosystems PCR-Mate DNA Synthesizer using standard phosphoramidite chemistry. Typically, a 1 μMol synthesis was performed for each template of RNA for an NMR sample. Oligonucleotides were deprotected in 1 mL concentrated (30%) ammonium hydroxide in Teflon-coated screw-capped vials at 65 °C overnight. Samples were then lyophilized, resuspended in water, and diluted 2-fold with gel loading buffer (80% v/v formamide, 20 % v/v 0.5 M EDTA and trace xylene cyanol and bromophenol blue as tracking dyes). Purifications were performed on a 45x35x0.15 cm 20% polyacrylamide (19:1 crosslink) gels containing 8 M urea in TBE. One-half of each synthesis was loaded into an 11 cm well. Gels were typically run at 55 W until the tracking dyes were run off the bottom of the gel, allowing resolution between molecules differing in size by only one nucleotide. DNA bands were visualized by UV-shadowing, excised, and eluted in TBE using an Elutrap (Schleicher and Schuell) at 250 Volts. Aliquots of the eluent were removed every hour for two to three hours. DNAs were precipitated with 0.3 M sodium acetate (NaOAc, pH 5.2) and 2.5 volumes ice cold 100% ethanol at -20 °C for a minimum

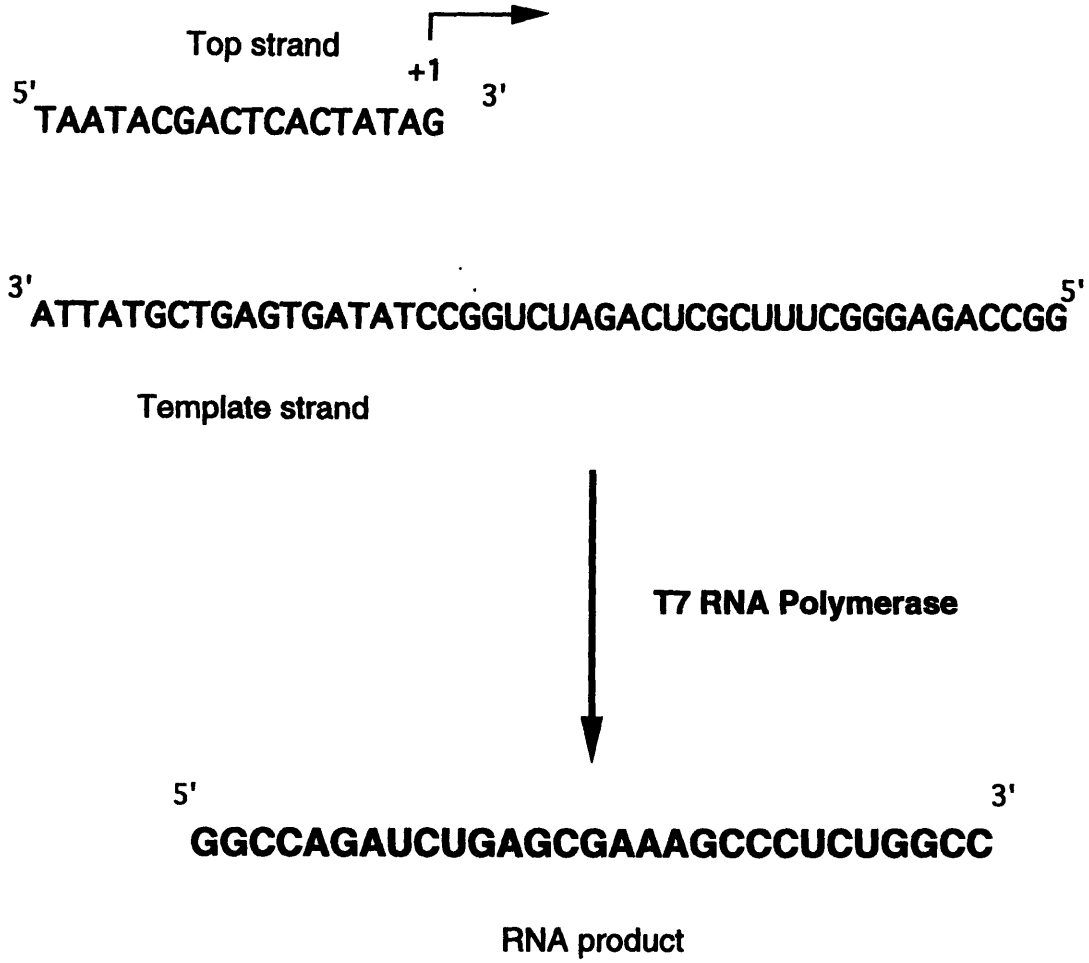


Figure 2.1) Synthesis of an RNA oligonucleotide. DNA templates provide a double-stranded promoter region for T7 polymerase. Synthesis of an RNA begins at the +1 position.

of 3 hours. DNA was resuspended in doubly-distilled, deionized, autoclaved water. Solution concentrations were calculated by UV absorbance on a Hitachi U-2000 spectrophotometer.

Transcriptions

Transcription conditions were essentially according to published procedures (Wyatt et al., 1991). Large quantities of T7 RNA polymerase were obtained by over-expression and purified as described (Grodberg & Dunn, 1988). For each template, reaction conditions were optimized for polymerase, nucleoside triphosphate usage (NTPs, purchased from Sigma), and the ratio of $MgCl_2$ concentration to that of NTPs + GMP. Optimization reactions were performed on a 20 µL scale, in the presence of trace amounts of [α-^{32}P] GTP (0.1 µL of 10 mCi/mL stock per reaction). Reactions were examined by electrophoresis on small, wide (25x35x0.05 cm) denaturing polyacrylamide gels. Optimum yields were estimated by incorporation of radioactivity as visualized by autoradiography. A wide range of conditions were explored for each template, however most of the RNAs studied in this thesis were synthesized under the standard T7 transcription conditions listed in Table 2.1. Reagents marked by a dot (•) were combined in a 10x stock solution and stored at -20 °C. Fresh DTT was added if the buffer was more than a month old. For large scale reactions, inorganic pyrophosphatase (Sigma) was added to remove the inorganic pyrophosphate by-product of the transcription reaction. A 100 µg vial of inorganic pyrophosphatase was suspended in 1 mL of T7 storage buffer (20 mM Na-phosphate pH 7.5, 1 mM DTT, 1 mM EDTA, 0.1M NaCl, 50% v/v glycerol). Reactions were incubated for 3.5 hours at 37 °C. Upon completion of the reaction, an extraction was performed with an equal volume of biotechnology grade phenol (Amaresco) to remove proteins. The aqueous layer was set aside, and the phenol was back-extracted with 1/10 volume of 40 mM Tris pH 8.1. Aqueous layers were then pooled and extracted

with an equal volume of 24:1 chloroform:isoamyl alcohol. After the extraction procedure 1/10 volume of 3 M NaOAc (pH 5.2) was added along with 2.5 volumes of ice cold 100% ethanol. Samples were then precipitated at -20 °C overnight.

Table 2.1
Transcription reaction conditions.

Reagent	Concentration
• Tris-HCl, pH 8.1	40 mM
• Spermidine	1 mM
• Triton X-100	0.01% (v/v)
• DTT	5-10 mM
$MgCl_2$	36 mM
Polyethylene glycol (8000 MW)	80 mg/mL
GMP	5 mM
DNA Template	300 nM
DNA Topstrand	300 nM
NTPs	4 mM each
Pyrophosphatase	1 μL/mL transcription

Transcription products were purified by denaturing gel electrophoresis as described for DNA oligos except that the gels were 0.3 cm thick instead of 0.15 cm, and they were pre-run for at least 3 hours prior to adding the sample. Pre-running allows for ample heating of the gel and produces a more homogeneous matrix. Typically, RNA from eight to fifteen mL of transcription was loaded into a 22 cm long well. Gels for 31 nucleotide (nt) samples were run at 55 W for 22 to 24 hours. Shorter RNAs (12 and 14 nts) were run for 11 to 14 hours. Excision and elution were as described for the DNA, except that elution was carried out at 4 °C and 100 to 150 V and checked every hour. Small RNAs (12-14 nts) were eluted at 100 V; eluent was removed every 30 to 45 minutes to minimize

loss of product through the membranes. The elution procedure for NMR samples of small RNAs took one to two days. Small RNAs were precipitated overnight at -20 °C.

RNA was resuspended in 0.5 to 1 mL of doubly-distilled, deionized sterile water. NMR samples were dialyzed in a microdialysis chamber (Bethesda Research) fitted with 1,000 molecular weight cutoff dialysis membrane (Spectrapor, regenerated cellulose). Samples were dialyzed in the following sequence at 4 °C against:

(1) 10 mM Na phosphate, pH 6.4, 5 mM EDTA, 100 mM NaCl for $\geq$ 12 hours.

(2) 10 mM Na phosphate, pH 6.4, 0.1 mM EDTA, 100 mM NaCl for $\geq$ 12 hours.

(3) ddH_2O for $\geq$ 24 hours.

Buffer was continuously circulated with a peristaltic pump, and the buffer in the reservoir was stirred. The samples were lyophilized and resuspended in the desired buffer (10 mM Na-phosphate, pH 5.5 or 6.4, 0.1 mM EDTA, 50 or 100 mM NaCl). For NMR experiments in H_2O, the sample was dissolved in 480 µL of water plus 60 µL of 10x buffer and 60 µL D_2O. Samples in D_2O were exchanged a minimum of 2 times with 500 µL D_2O (99.9%, Cambridge Isotope Labs), including once overnight. Samples were prepared in a glove box under argon gas with 99.996% D_2O (Isotec, Inc.). All samples included trimethyl-silyl-propionic acid (TSP, Sigma) as a chemical shift standard. New NMR tubes were washed with ethanolic KOH and repeatedly rinsed with ddH_2O to limit sample sticking. Tubes to be used in D_2O experiments were baked overnight in a 500 °C oven (laying flat), and cooled in a dessicator prior to use. All NMR pH titrations were performed in the NMR tube with 0.1 M NaOH or 0.1 M HCl, and were monitored using an NMR tube pH meter (Microelectrodes, Inc., MI-412). A 120 mM stock solution of argininamide (Sigma) was prepared in NMR buffer (10 mM Na-phosphate pH 6.4, 50 mM NaCl, 0.1 mM EDTA) for use in NMR titrations. NMR samples for duplex RNAs contained 100 mM NaCl instead of 50 mM.

Gel Filtration Chromatography

RNA to be used for NMR analysis was checked for dimerization by gel filtration chromatography. The high concentrations necessary for NMR (ca. 1 mM) may shift the equilibrium of some RNAs away from the hairpin structure toward dimers. Gel filtration chromatography, or size-exclusion chromatography, separates molecules based on their apparent molecular volume. RNAs were analyzed using a Beckman System Gold HPLC fitted with a Bio-Rad Biosil SEC 400 or Biosil SEC 125 column (with a guard column). Flow rate was typically 1 mL/minute, elution was monitored at 260 nm. Degassed, filtered NMR buffer (see above) was used. Samples known to dimerize were injected as size standards. A 1 μL injection of a 1 mM sample was followed with a 10 μL 'chase' of buffer into a 10 μL injection loop. Samples were then diluted 1/10 with NMR buffer (10 μL were injected) and 1/100 (50 μL were injected into 100 μL injection loop) to determine whether the molecularity of the RNA changes upon dilution (Figure 2.2).

2.2 Thermodynamic Stability of RNA

The transition between native and denatured RNA structures can be monitored using ultraviolet (UV) absorbance. The interaction between induced dipoles in the heterocyclic nucleotide bases causes the absorbance of a nucleic acid polymer to be considerably less than the sum of the absorbances of the nucleotides (Tinoco, 1960). The increase in light absorbance as a nucleic acid changes from an ordered (native) state to a disordered (denatured) state is termed hyperchromicity (Tinoco, 1960). The magnitude of the difference in absorbance (hyperchromicity) between the two states can be used as a measurement of base pairing and stacking (secondary structure) (Petersheim & Turner, 1983, Puglisi & Tinoco, 1989). Monitoring the change in absorbance of a nucleic acid

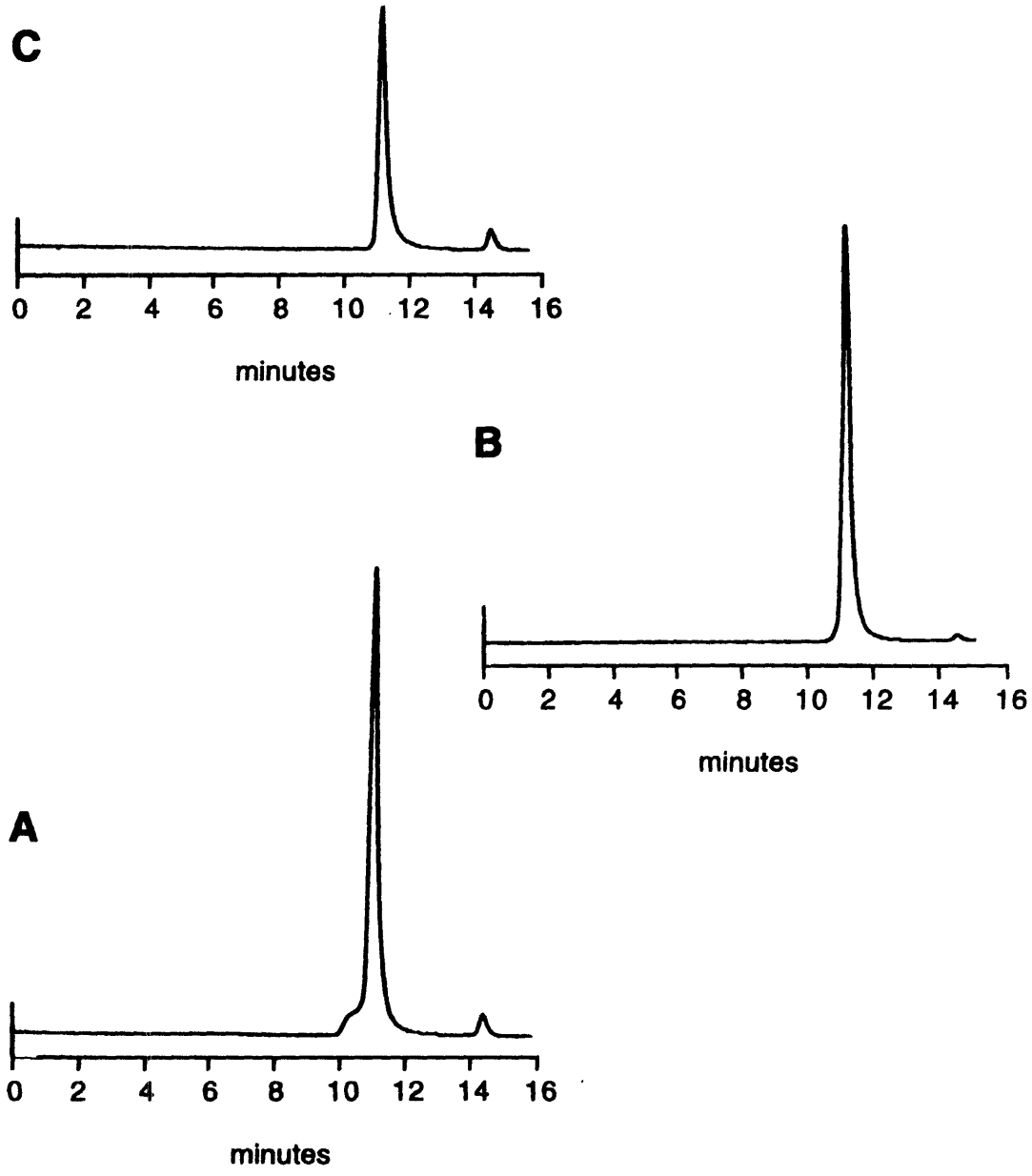


Figure 2.2) HPLC traces from size exclusion chromatography. Elution traces are shown for an experiment performed on a 31 nucleotide hairpin RNA with an elution time of 11 minutes. Trace **A** shows the sample at 1.8 mM. The molecule is predominantly a monomer, but the small shoulder eluting at 10.5 minutes indicates the presence of some dimer. Traces **B** and **C** show 1/10 and 1/100 dilutions, respectively. The dimer is completely absent in both of these samples. A 1 mL/minute flow rate was used in buffer containing 10 mM Na-phosphate pH 6.4, 50 mM NaCl, and 0.1 mM EDTA.

upon heat denaturation is referred to as 'melting'. The profile of absorbance as a function of temperature is called a 'melting curve'. From the melting curve, one can determine the fraction of one component (native, or denatured); the midpoint of the melting transition is defined as the melting temperature (T_m). Analysis of a melting curve can yield thermodynamic data for the melting transition.

Optical Spectroscopy

Melting curves in this study were performed on an AVIV 14DS spectrophotometer. Temperature in the sample compartment was controlled by the spectrophotometer which was connected to a circulating water or water:ethylene glycol bath (Neslab RTE110 cooler). Absorbance was monitored at 260 nm. Data were collected at a heating rate of one degree per minute, with 2 minutes of equilibration time per increment. The sampling rate was set to one second at each temperature point. A typical melt was performed over the range of 25-95 °C. Teflon stoppered-300 μL cuvettes with a path length of 1 cm were used with aluminum adapters to insure fast, even sample heating. The use of special cuvettes fitted with spacers allows one to achieve much shorter path lengths (l), and therefore more concentrated samples can be analyzed [$A = \varepsilon cl$, where A is absorbance, ε is the extinction coefficient of the sample, and c is the concentration]. These cuvettes are stopperless and silicone oil must be layered over the sample. A minimum amount of oil is used as excess oil may creep down the sides of the cuvette causing inaccurate absorbance readings. Pathlengths as small as 0.05 mm can be achieved with these spacers. Absorbances at all concentrations should be kept below 1.5 so as to minimize signal-to-noise errors.

Sample Preparation

RNA solutions used for melting curves were dialyzed in a microdialysis chamber (see above) at 4 °C for a minimum of 18 hours. Typical buffers contained 10 mM Na-phosphate, pH 6.4, 0.1 mM EDTA, and 25 to 100 mM NaCl. Buffers that have a very

low temperature coefficient must be selected. Typically, phosphate or cacodylate are used while common buffers such as Tris and HEPES must be avoided. Small pathlength cuvettes are especially prone to bubble formation, therefore care must be taken to degas the samples by bubbling nitrogen or argon gas through the sample for ca. 15 minutes, or by briefly subjecting the sample to vacuum for ca. 5 minutes. Prior to performing the melt, the samples should be heated to a temperature well above the expected T_m, and allowed to slowly cool to the starting temperature. Typically, after degassing, each sample was heated to 90 °C for 1 minute, then allowed to cool slowly (over 20 minutes) to the starting temperature. After heat-cooling, prior to beginning the melt, the sample was placed in the cuvette and equilibrated to the starting temperature in the AVIV. The change in absorbance over time was monitored. When this change was minimized to 0.005 absorbance units/minute the experiment was started.

Data Analysis

For the samples analyzed in this thesis, the transition from duplex to random coil (monomer) was assumed to be a two-state process, always at equilibrium. An all-or-none model was used, the RNA was assumed to be either single-stranded or in a duplex (i.e. no intermediate states were accounted for). This assumption is usually valid for duplexes of fewer than 12 base pairs (Puglisi & Tinoco, 1989). Although the duplexes studied for this work contain 14 base pairs, this assumption seems to be valid. No evidence of the molecules melting in stages was observed. Thus, duplex formation can be represented by the equation

$$S_A + S_B \leftrightarrows D \qquad (2\text{-}1)$$

where S_A and S_B represent the non-self complementary single strands A and B, and D is the duplex conformation. The transition between the single-stranded and double stranded

states is monitored by the change in absorbance (A) as a function of temperature. Other physical properties, such as circular dichroism or NMR chemical shift, can also be monitored (Puglisi & Tinoco, 1989). For the two-state model:

$$A = fA_D + (1 - f)A_S \tag{2-2}$$

where f is the fraction of bases paired, and A_S and A_D are the absorbances for single-stranded and double-stranded species, respectively. Equation 2-2 can be used to analyze a melting curve whereby the T_m is defined as the temperature where $f = 0.5$. An equilibrium constant (K) can then be obtained for the two-state transition as a function of f and c_T, the total strand concentration:

$$K = [D]/[S_A][S_B] = 2f/(1 - f)^2 c_T \tag{2-3}$$

In this study, the raw data were fit by the nonlinear least squares method (Petersheim & Turner, 1983, Puglisi & Tinoco, 1989) using Igor Pro software. Six parameters: ΔH°, ΔS°, and four parameters that specify the slopes (m_D and m_S) and intercepts (b_D and b_S) of the upper and lower baselines of the curve are obtained. In using this analysis, it is important that the baselines are well defined, otherwise erroneous values of ΔH° and ΔS° will be obtained. Sloping baselines in A vs. T plots result from a small change in the double stranded (lower baseline) or single stranded (upper baseline) states with temperature. The temperature dependence on absorbance of the two states is approximated by assuming a linear dependence of the extinction coefficients (ε) on temperature:

$$\varepsilon_D = m_D T + b_D \tag{2-4}$$

$$\varepsilon_S = m_S T + b_S \tag{2-5}$$

The melting curves are fit to Equation 2-6:

$$\varepsilon(T) = A(T)/lc_T = f\varepsilon_D + (1 - f) \qquad (2\text{-}6)$$

where $\varepsilon(T)$ and $A(T)$ are the extinction coefficient and absorbance of the solution, respectively, at temperature T. The pathlength is l, c_T is the total strand concentration, and ε_D and ε_S are the extinction coefficients defined in Equations 2-4 and 2-5. The fraction of strands in the double strand state (f) is related to the changes in enthalpy, ΔH°, and entropy, ΔS°, by:

$$K = \exp(\Delta H^\circ/RT + \Delta S^\circ/R) = 2f/(1 - f)^2 c_T \qquad (2\text{-}7)$$

An Igor Pro computer program fits the experimental curve treating ΔH°, ΔS°, m_D, m_S, b_D, and b_S as variable parameters. Using standard thermodynamic equations, one may obtain the standard enthalpies, entropies, and free energies per mole of the reaction (Equations 2-8 and 2-9):

$$\Delta G^\circ = -RT \ln K \qquad (2\text{-}8)$$

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \qquad (2\text{-}9)$$

Examples of raw and fit data are shown in Figure 2.3.

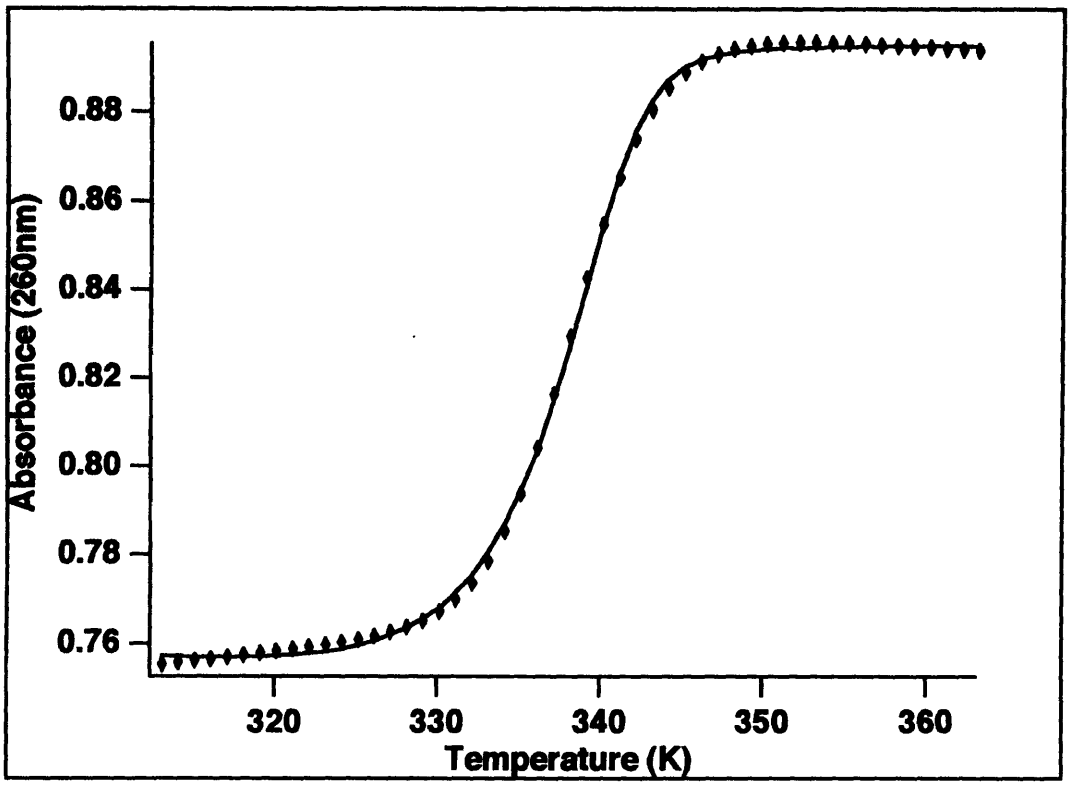


Figure 2.3) Example of a melting curve. Raw data from the melting curve of a 14 base pair RNA is shown as a plot of absorbance vs. temperature. Data (black diamonds) were fit (solid line) as described in the text. There is a reasonable correlation between the raw and fit data. Note the well-behaved upper- and lower- baselines. Low salt experimental conditions were used (10 mM Na-phosphate pH 6.4, 25 mM NaCl, 0.1 mM EDTA).

2.3 NMR Spectroscopy Applied to RNA

Due to the difficulty in obtaining suitable crystals, only recently have RNAs other than tRNAs been crystallized for X-ray diffraction studies (see Lietzke et al., 1995). Thus, alternative methods for the analysis of RNA structure and dynamics were necessarily developed. The most powerful method for studying the structure of molecules *in solution* is NMR. Detailed models based on NMR data require assignment of resonances for most of the atoms of the molecule. As discussed below, this process is greatly hampered by overlap of many of the sugar protons and by a poorly defined backbone conformation. Isotopic labeling techniques (Batey et al., 1992, Nickonowicz et al., 1992), and the corresponding development of complicated multi-dimensional NMR experiments to exploit this labeling, have alleviated some of the spectral overlap. The synthesis and assignment of labeled molecules and subsequent structural modeling (based on distance geometry and molecular dynamics calculations) are quite time consuming. To supplement previously published work on the detailed structure of TAR (Puglisi et al., 1992), a model of a 30 nucleotide TAR with a 2-base bulge (UU) using isotopically labeled RNA has been performed in this laboratory by Alex Brodsky (A.S. Brodsky and J.R. Williamson, manuscript in preparation). The aim of this thesis work was to investigate the structure and argininamide binding of a number of mutant TAR RNAs. As such, it was not feasible to synthesize isotopically labeled molecules for all of them. A number of general NMR experiments were performed, and global conformations of the mutant RNAs were determined. It should be pointed out, however, that a detailed modeling of their structures was not performed.

Information from NMR

Basic types of information obtained from NMR experiments include chemical shift, an indicator of the local environment of the nucleus of interest; scalar (spin-spin) couplings;

dipolar (through-space) interactions (NOEs); and exchange properties. These subjects will be discussed briefly here (for general reviews, see Derome, 1987, Ernst et al., 1987 Kessler et al., 1988, Wuthrich, 1986).

The **chemical shifts** (δ) of protons in an RNA molecule range from approximately 3.5 to 14.5 ppm. In general, there are two types of proton resonances: non-exchangeable protons bonded to atoms such as carbon, and exchangeable protons, bonded to nitrogen or oxygen. Non-exchangeable protons are observed in experiments in D_2O. As shown in Figure 2.4, protons attached to different types of carbon atoms resonate in different parts of an NMR spectrum. For example, protons attached to sp^3-hybridized ribose 2', 3', 4', 5' and 5" carbons (Figure 1.1) all resonate between 3.9 and 4.9 ppm, causing a great deal of spectral overlap, as alluded to above. Because C1' is attached to an oxygen (O4') and a nitrogen (from the aromatic base), the H1' protons resonate *down*field between 5 and 6.5 ppm. Aromatic purine H8 (and adenine H2) and pyrimidine H6 proton resonances are located between 6.9 and 8.5 ppm. The pyrimidine H5 chemical shifts are found between 5.0 and 6.3 ppm.

Exchangeable protons can be observed in 90% H_2O/10% D_2O solutions. They include imino protons bound to nitrogen atoms of the aromatic bases and amino protons on the exocyclic nitrogens. A comparison of Figure 2.5 with Figure 2.4 shows that the imino protons resonate far downfield (ca. 11-14.5 ppm) and do not overlap with other RNA resonances. These protons exchange rapidly with solvent (H_2O) and therefore are not usually observed in an NMR spectrum unless they are hydrogen bonded. Imino protons (attached to guanine N^1, and uracil N^3) within a base pair will therefore be observed and are considered to be an indication of secondary structure (Figure 1.2). Guanine imino protons in a G-C base pair give rise to characteristic NOEs (through space interactions) to C-amino protons, as do uracil iminos to adenine H2 protons. Interestingly, in the G-U 'mispair' there are two imino protons that give characteristically strong NOEs to each other,

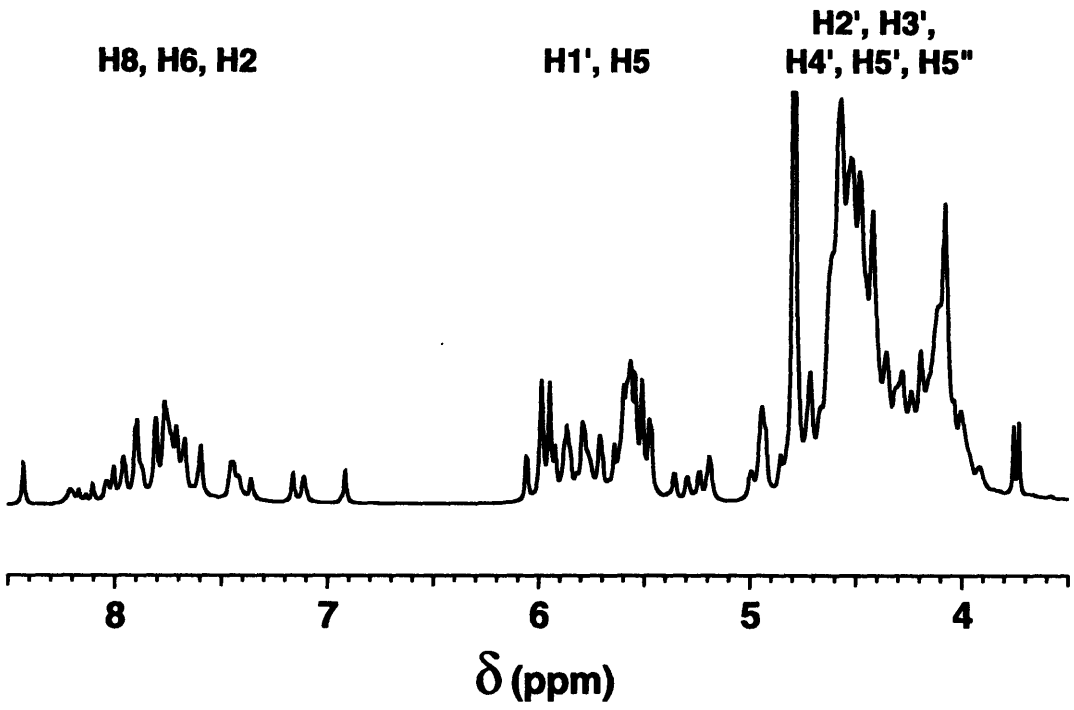


Figure 2.4) One-dimensional non-exchangeable spectrum of RNA. Three distinct regions of the spectrum are present. The aromatic H6, H8, and H2 protons resonate downfield at 6.9 to 8.5 ppm. The H1' protons are separated from the other sugar resonances, giving rise to peaks near pyrimidine H5s in the 5 to 6.5 ppm range. The remaining sugar protons all resonate in the narrow region from 3.9 to 4.9 ppm, causing a great deal of spectral overlap.

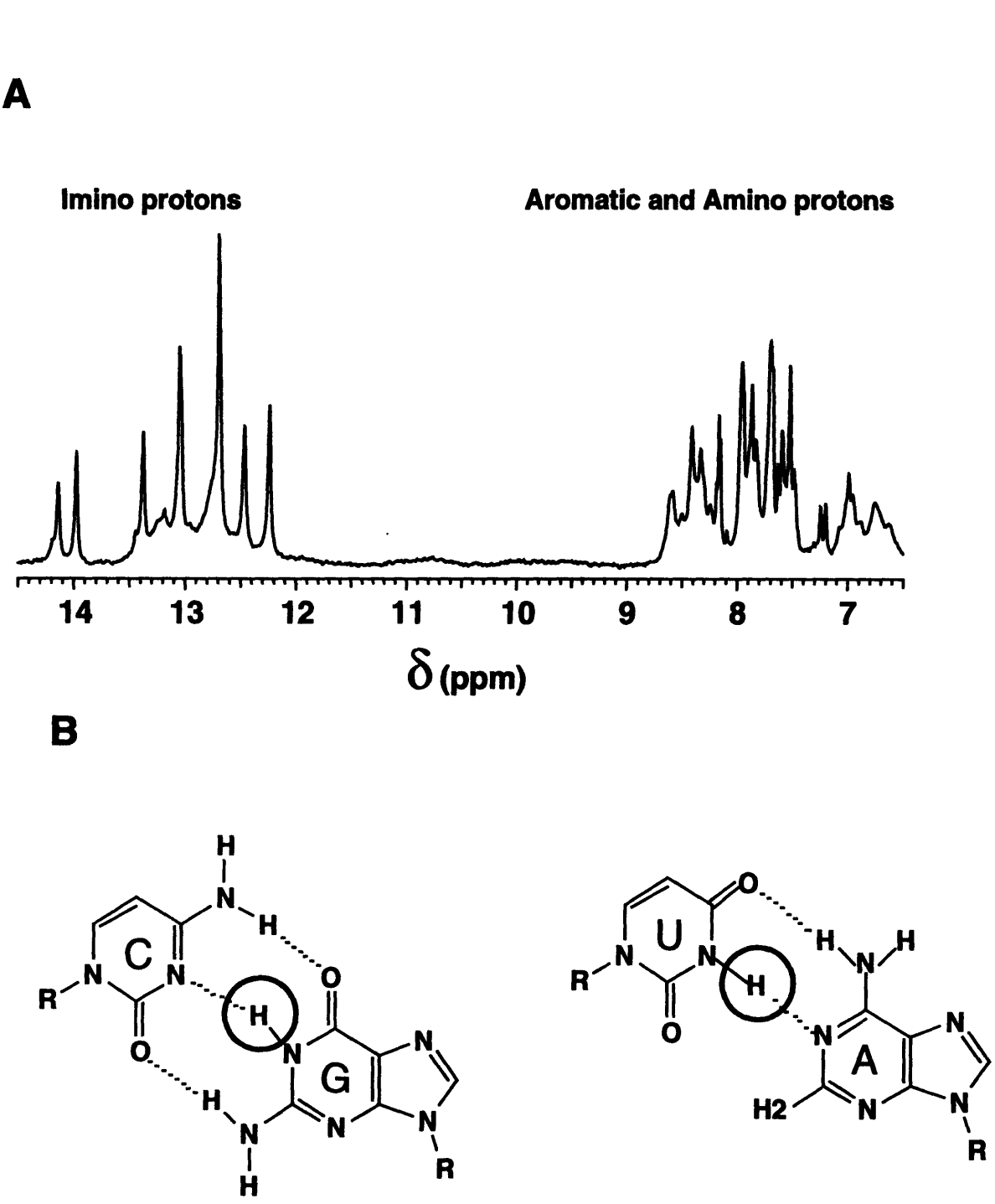


Figure 2.5 RNA imino protons. A) NMR spectrum of exchangeable protons in RNA. The imino protons resonate downfield in a well resolved portion of the spectrum. B) RNA base-pairs are shown with imino protons circled. G-imino protons are close in space to the amino protons on the base paired C. The U-imino protons are close to the A-amino and H2 protons. Imino protons from stacked base-pairs are also close to each other in space.

with one imino (usually the U) shifted upfield to ca. 10.5 ppm. NOEs between imino protons of stacked base pairs can also be observed and used to analyze secondary structure formation. Amino protons in nucleic acids resonate between 6 and 9 ppm, with the hydrogen bonded protons having more downfield chemical shifts.

Scalar or **J-coupling** is observed for nuclei that are covalently bonded. In 1-dimensional spectra, the effect of J-coupling is usually seen as splitting of NMR peaks. This splitting is often used to determine how many protons exist on a neighboring atom. For protons, two and three bond couplings (J) are usually on the order of 2-20 Hz. The magnitude of a coupling constant depends on factors such as the hybridization of the bonds, the electronegativity of the attached substituents, and the dihedral angle between the two coupled nuclei. The relationship between coupling constants and dihedral angle is important for nucleic acid structure determination and has been investigated in detail (Karplus, 1959, Haasnoot et al., 1980). Simply put, vicinal protons are strongly coupled when the dihedral angle between them is 0° or 180°. They are weakly coupled if it is 90°. For complicated molecules, the analysis of 1D spectra is not practical. Simple 2-dimensional DQF-COSY (Double Quantum Filtered COrrelation SpectroscopY) experiments (e.g. Muller et al., 1986, Piantini et al., 1982, Kessler et al., 1988) can be performed where observed cross peaks correspond to three bond couplings between vicinal protons (Figure 2.6). Coupling constant data can be obtained from these spectra, and consequently structural information as well. Unfortunately, many of the sugar resonances are severely overlapped (Figure 2.6, box C), making it difficult to obtain much information from standard homonuclear experiments. However, two other regions of a COSY experiment can provide useful information. Box A in Figure 2.6 shows a region of a COSY spectrum corresponding to cross peaks between pyrimidine H5 and H6 protons. Although the coupling constant data are particularly uninteresting (the conformation of the planar aromatic ring is locked), chemical shift information for all the pyrimidine H6 protons

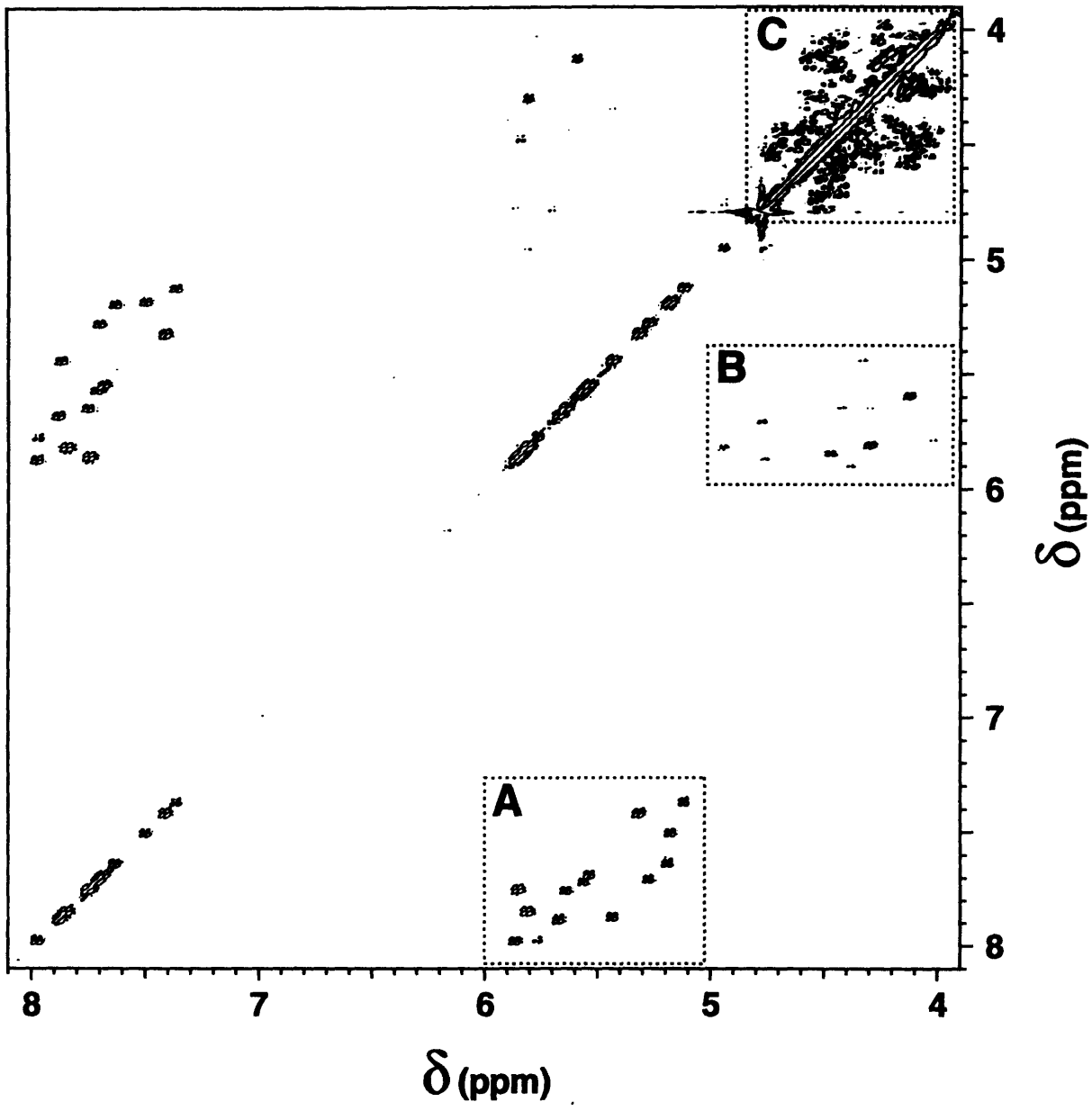


Figure 2.6) Example of a two-dimensional DQF-COSY. Three distinct regions of peaks corresponding to scalar couplings are seen in this spectrum. Peaks in region **A** result from pyrimidine H5-H6 correlations. Well resolved correlations between H1' and H2' protons are seen in box **B**. All other peaks corresponding to sugar couplings resonate in the crowded region marked **C**.

is useful in assigning through space connectivities (see below). Because the 1' protons resonate in a region of the spectrum separate from the other sugar protons, it is possible to observe H1'-H2' COSY cross peaks. However, in sugars with the C3'-*endo* conformation, the dihedral angle is approximately 90°, corresponding to an unobservably small $J_{1'-2'}$. Conversely, in C2'-*endo* sugars, the $J_{1'-2'}$ is large and detectable by NMR (Figure 2.7). Unfortunately, most sugars in structured RNA molecules are in the C3'-*endo* conformation (Section 1.1). Cross peaks corresponding to H1'-H2' resonances result from ribose moieties that have a significant C2'-*endo* population. In RNA these sugars may be in unstructured regions, at the ends of helices, near bulges, or a result of some other local distortion.

TOtal Correlation SpectroscopY (TOCSY) (Bax & Davis, 1985) experiments allow magnetization transfer through several bonds, not just the three bond couplings as seen with the COSY. Again, if the sugar is C3'-*endo*, there will be no observable coupling between the H1' and other sugar protons. If there is a sufficient $J_{1'-2'}$, magnetization will be transferred from H1' to H2' and H3'. Nucleotides with fast interconversion between C3'-*endo* and C2'-*endo* conformers will produce cross peaks between H1' and all other sugar protons (Varani & Tinoco, 1991).

Dipolar coupling involves a direct magnetic coupling between nuclei. This interaction causes the exchange of magnetic moment spin states through space. The change in magnetization of a spin upon perturbation of another spin is called the Nuclear Overhauser Effect (NOE). Because of spectral overlap, it is impractical to perform 1D NOE experiments on large molecules. For RNA, only the imino protons can be reasonably analyzed for 1D NOEs. Through space interactions between other nuclei can be detected in 2-dimensional Nuclear Overhauser Effect SpectroscopY (NOESY) experiments (Jeener et al., 1979). The intensity of an NOE has a $1/r^6$ dependence on the distance between two nuclei, and thus is only detected if they are less than 5 Å apart. Nevertheless, NOESY

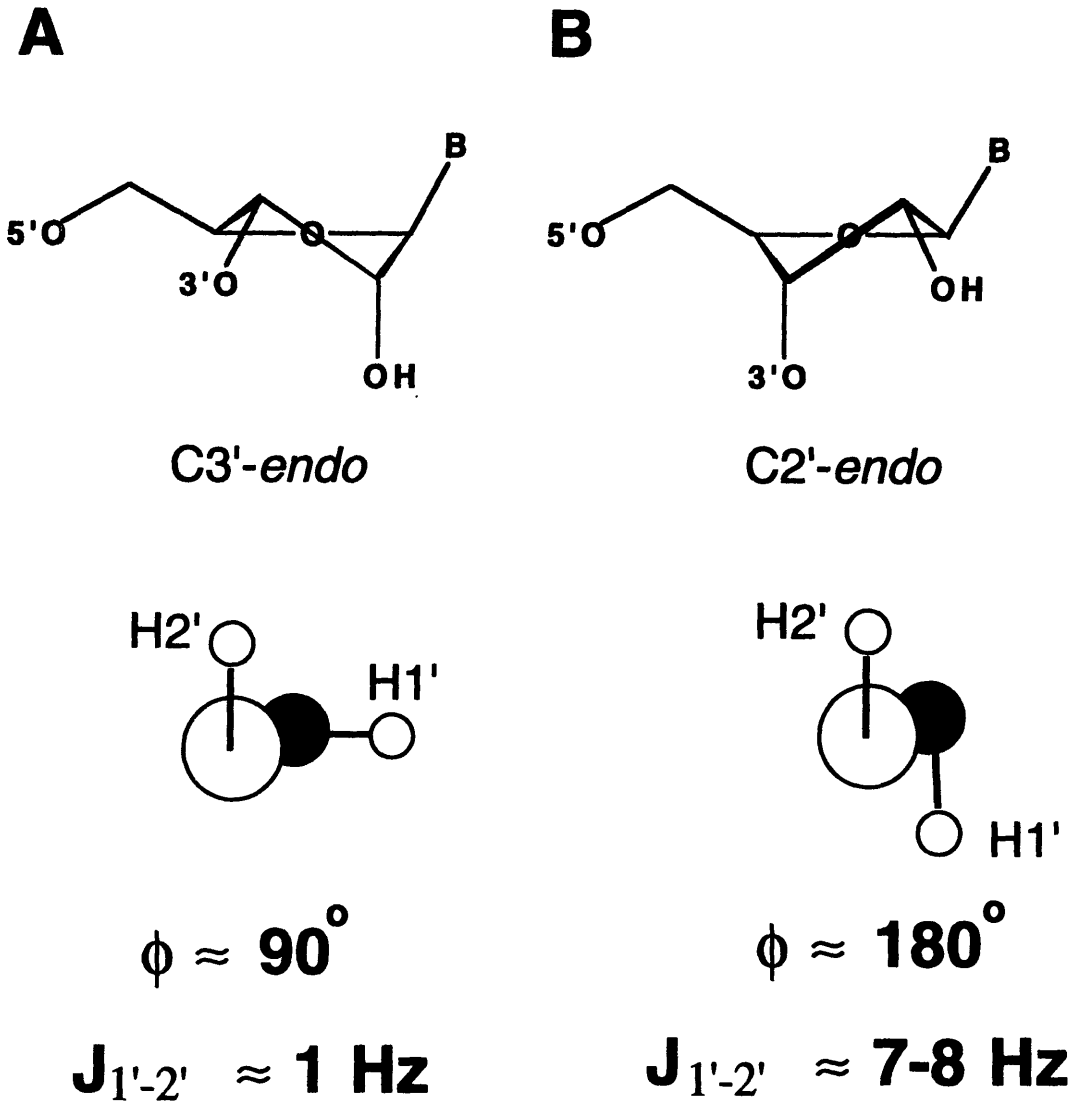


Figure 2.7) Dependence of coupling constant on sugar pucker. A) The C3'-*endo* sugar pucker. The dihedral angle (ϕ) between H1' and H2' is 90°, causing in a small coupling constant, $J_{1'-2'}$. **B)** The C2'-*endo* sugar with a 180° dihedral angle, resulting in a large coupling constant of 7 to 8 Hz.

experiments are extremely useful for assigning protons found in the context of an RNA helix. As mentioned above, NOEs are detected between imino and amino or adenine H2 (AH2) protons, and between pairs of imino protons in H_2O experiments. Figure 2.8 shows a typical NOESY spectrum. This spectrum was acquired with a well-behaved 29 nucleotide RNA molecule on a 750 MHz spectrometer. Separate regions of the spectrum correspond to interactions between different types of protons based on their chemical shifts. For example, region A of Figure 2.8 shows cross peaks between aromatic protons. The distance between aromatic protons from stacked base pairs within the context of an A-form helix is ca. 4.5 Å, barely within the experimental constraints for an NOE. These peaks are only observable in experiments performed with long mixing times (see below). Region B is perhaps the most useful region for assignment of RNA resonances. Here, NOEs are seen between the anomeric H1' protons and their own aromatic protons (distances of 3.5 Å). NOEs are also seen from pyrimidine H5-H6 interactions (2.4 Å), these peaks are easily assigned from their presence in COSY spectra. An anomeric proton in an A-form helix will also have an NOE to the aromatic proton on the base of the nucleotide to its 3' side (4.5 Å). Also in this region are NOEs from helical AH2s to the H1' proton of the following (3') residue and to the H1' on the sugar one nucleotide in the 5' direction of the opposite strand (ca. 4 Å). The location of AH2 resonances can be confirmed from the H_2O data. These NOEs are very useful for assignment. NOEs between aromatic protons and the remaining sugar protons are found in region C and are difficult to assign in proton only experiments. Through space interactions (inter- and intra-nucleotide) between H5 and H1' protons as well as internucleotide H5-H5 (3.7 Å) interactions between stacked pyrimidines give rise to resonances in region D. Anomeric to ribose correlations are observed in region E. In proton-only experiments spectral overlap renders the ribose-ribose region (F) useless.

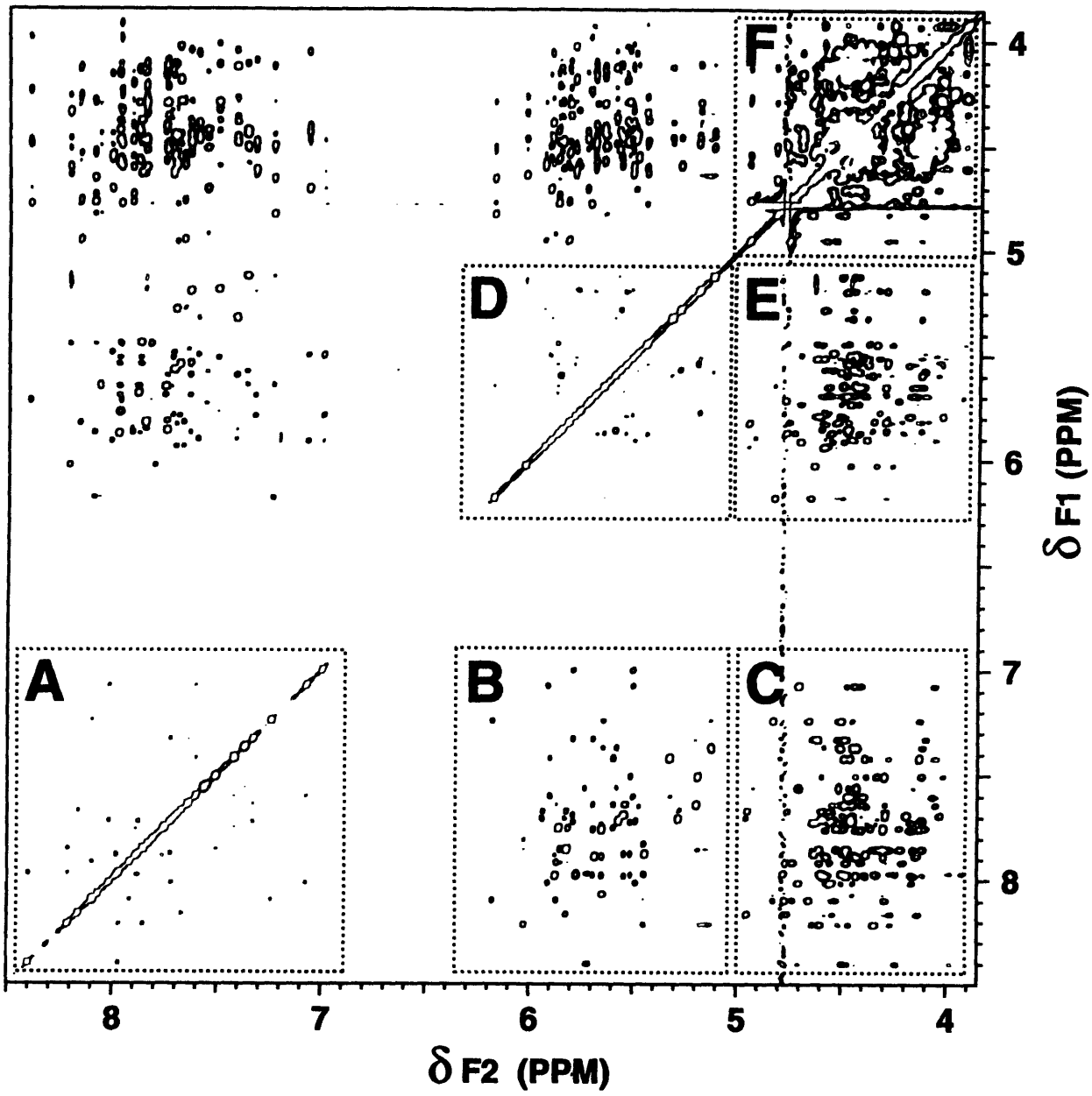


Figure 2.8) Example of a two-dimensional proton-proton NOESY. Crosspeaks in a NOESY spectrum result from through-space interactions between nuclei. For RNA, peaks in region **A** represent couplings between aromatic protons in stacked nucleotides. Information in box **B** corresponds to aromatic to H1' and H5 NOEs. This region is used for resonance assignment. Box **C** contains NOEs from aromatic protons to sugar protons (other than H1's). This region is useful for determining the sequential connectivity pathway in short mixing time experiments. Crosspeaks in **D** result from H1' and H5 correlations. Region **E** contains NOEs from H1's to other sugar protons. Ribose-ribose crosspeaks are overlapped in **F**.

Experiments are performed with different mixing times to allow observation of interactions between protons separated by different distances. For example, in short mixing time experiments (<50 ms), region C will contain resonances that arise mostly from aromatic H6 and H8 to the H2' proton of the nucleotide on its 5' side (2.0 Å). At short mixing times, region E is also simplified and will contain resonances mostly due to intranucleotide H1'-H2' NOEs (2.4 Å). An interesting NOESY comparison of A- and B-form helices has been performed (Haasnoot et al., 1984). A table of A-form interproton distances can be found in Wüthrich (Wuthrich, 1986).

Two types of **exchange** can be observed in NMR spectroscopy. These include chemical exchange (i.e. imino protons exchanging with solvent) and conformational exchange (i.e. flipping sugar puckers or dynamic stacking and unstacking of a nucleotide). Relative to the NMR time scale, exchange can be classified as slow, intermediate, or fast (Wuthrich, 1986). A proton in slow exchange between two states, A and B, will produce a spectrum with two different resonances at chemical shifts δ_A and δ_B (Figure 2.9). A particular exchange process is considered to be slow if the rate of exchange, k_{AB}, is much smaller than the difference between the chemical shifts ($\Delta\delta = |\delta_A - \delta_B|$), $k_{AB} << \Delta\delta$. Intermediate exchange occurs when $\Delta\delta \sim k_{AB}$. This results in broadening of a resonance and may cause its complete disappearance. Processes where the exchange rate is much greater than the chemical shift difference ($k_{AB} >> \Delta\delta$) result in one resonance with a chemical shift that is the weighted average of the relative populations of the two states. A simple indication of slow or intermediate conformational exchange of an RNA molecule is the presence of more H5-H6 cross peaks in a COSY spectrum than can be accounted for by the number of pyrimidines in a single conformer.

Assignment Procedure

The basic methodology for assignment of RNA molecules was essentially as published (Haasnoot et al., 1984, Ven & Hilbers, 1988, Varani & Tinoco, 1991).

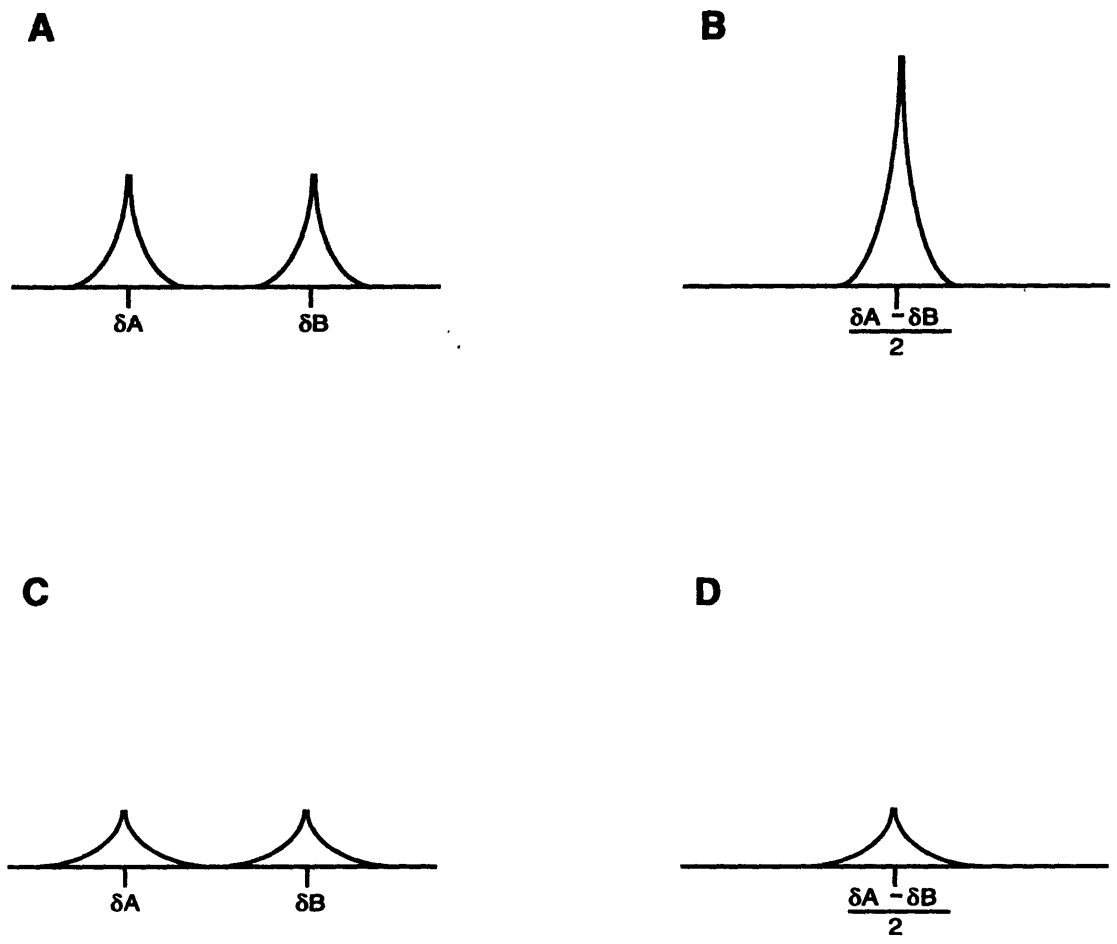


Figure 2.9) Types of exchange in NMR. A) An example of slow exchange ($k_{AB} << \Delta\delta$). This schematic representation of a spectrum shows two distinct resonances for state A and B. Peak intensities are proportional to the relative populations of each state. **B)** Fast exchange ($k_{AB} >> \Delta\delta$). One resonance is observed at a chemical shift that is the weighted average of the relative populations of the two states (here they are equal). **C)** Intermediate exchange ($k_{AB} \sim \Delta\delta$) resulting from an increase in the slow exchange depicted in **A**. **D)** As fast exchange (shown in **B**) decreases, the resonance broadens in intermediate exchange ($k_{AB} \sim \Delta\delta$).

Initial experiments are performed in 90% H_2O/10% D_2O. A 1D-exchangeable spectrum provides crude information of secondary structure. The presence of imino resonances indicates base pair formation (see above). The number of imino resonances should roughly correspond to the number of expected base pairs, allowing for the exception that base pairs at helix termini are dynamic and their imino protons usually exchange too rapidly to be observed. Position of base pairs within the helix can be determined based on the melting behavior of the imino proton spectrum. Peaks arising from weakly paired bases, non-Watson-Crick interactions, or base pairs closer to the end of the helix will melt out first. Saturation transfer NOE experiments can be performed in order to identify the types of base pairs and their stacking order. Selective irradiation of an imino proton in the context of a helix will affect imino protons in the base pairs above and below it, as well as the AH2 proton (if a uracil was irradiated) or the cytosine amino resonances (if a guanosine was irradiated). The spectral changes are observed by subtracting the data from a non-irradiated reference. This method is good for a first approximation, but it is often impossible to distinguish the type of base pair due to difficulty in locating the AH2 and amino protons.

More detailed assignment of imino, amino, and adenine H2 resonances can be undertaken by performing a NOESY experiment in water. This method is becoming increasingly more common because of advances in water suppression techniques (e.g. Piotto et al., 1992). Imino-imino resonances (an indication of stacking) are well resolved in a water NOESY. They resonate downfield, near the diagonal, well away from the water peak, thus assignment of the sequence of base pairs is possible. A water NOESY also allows one to find the adenine H2 protons which, as mentioned earlier, are helpful for the assignment procedure. Uridine iminos give one strong NOE to the adenine H2, while guanosine iminos give rise to two from the cytosine aminos. Other information can be obtained from a water NOESY that can be used to verify non-exchangeable assignments once they have been assigned.

The first step in assigning the non-exchangeable protons is a DQF-COSY. As pointed out above, the H5-H6 region can be examined for evidence of exchange. Identification of all the H6 protons is important in assigning the H1' and H2' protons of pyrimidines. The H1'-H2' region of a COSY spectrum also provides crude structural information. The presence of cross peaks allows identification of riboses near helix termini, or in other non-standard helical conformation (see above). TOCSY experiments will provide similar information. These experiments can be performed second, or after the NOESY data have been collected.

The next step in assignment of the non-exchangeable resonances involves acquiring NOESY spectra at several mixing times. At 500 MHz, 50, 100 and/or 200, and 400 ms mixing times are used, although at higher fields these times can be shortened. The 50 ms NOESY will only have cross peaks from nuclei close in space (ca. 2 - 2.5 Å), allowing identification of intranucleotide H1'-H2' resonances (Figure 2.8 box E). Also, aromatic protons on helical residues (n) will give NOEs to H2' protons on the residue to its 5' side (n-1) (Figure 2.8, box C). These data will help in interpreting the longer mixing time experiments. In region B of Figure 2.8, at a chemical shift corresponding to a particular aromatic proton (n), several NOEs are seen in a long mixing time experiment. Peaks corresponding to aromatic H6 or H8 to H1' (n) and to the H1' of residue n-1 are observed. Data from the short mixing time NOESY provide information on aromatic (n) to H2' (n-1) and n-1 H2' to H1'(n). Thus, it is possible to distinguish between the inter- and intra-residue aromatic to H1' peaks. The presence of a strong peak corresponding to a COSY H5-H6 peak allows one to determine that this resonance is a pyrimidine (Y). This procedure allows one to 'walk' through the sequence of helical nucleotides (Figure 2.10). One continues the sequential assignments until a unique pattern of purine (R) and pyrimidine (Y) nucleotides is obtained that can only arise from one part of the molecule. At

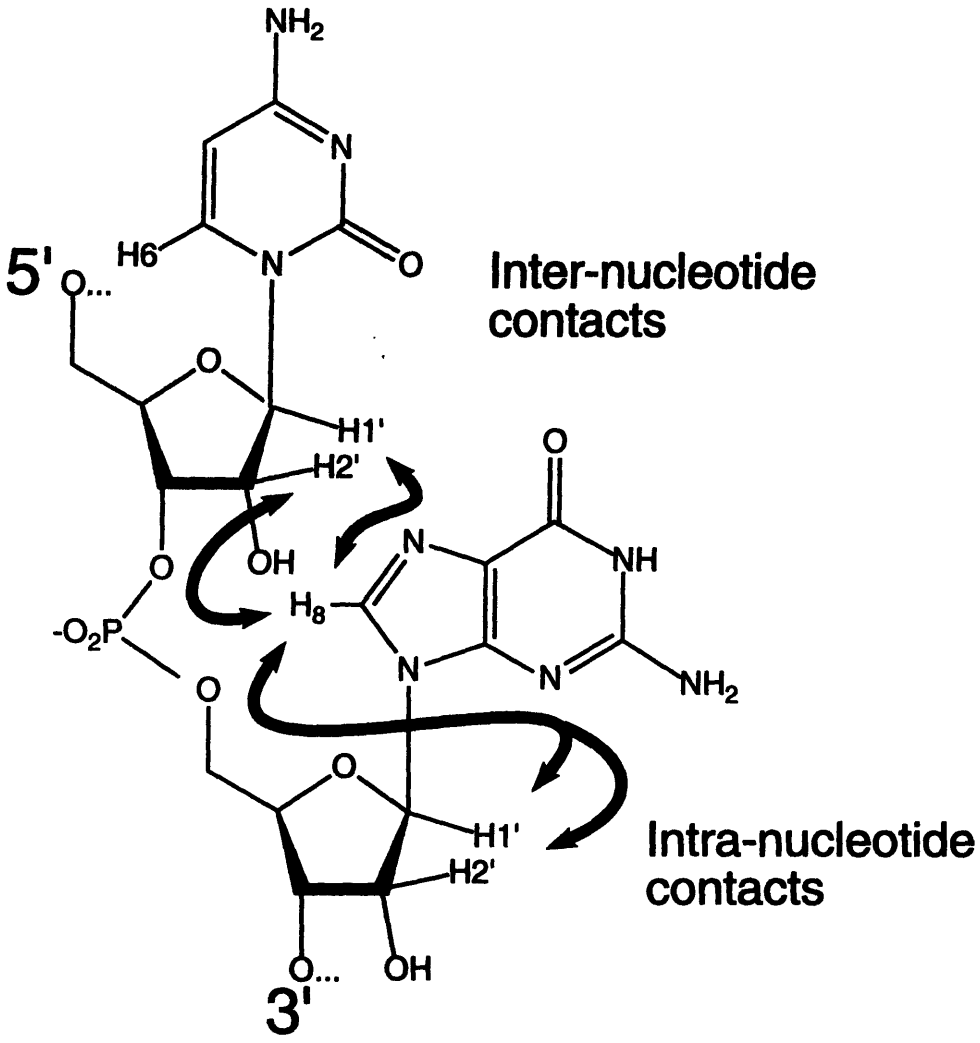


Figure 2.10) Sequential assignment pathway for A-form RNA helices. Short intranucleotide and internucleotide aromatic to sugar NOEs are shown by thick arrows. Aromatic H6 or H8 protons have short distances to their own H1' and H2' prorons, as well as to the H1' and H2' on the neighboring sugar in the 5' direction. Assignment of the connectivity pathway allows one to 'walk' through the sequence of an RNA molecule

long mixing times and with well behaved molecules, it is sometimes possible to observe NOEs between aromatic protons and the H5 of a neighboring (3') pyrimidine. COSY and TOCSY H1'-H2' data help to assign which resonances are at helix termini or near structural perturbations. Adenine H2 protons (assigned from H_2O spectra) give important NOEs to the n+1 H1', and to the opposite strand NOE (to the H1' of the residue in the preceding base pair). Information from Figure 2.8 regions A and D also help to confirm the connectivity pathway.

At this point, the H_2O NOESY is reexamined. Imino to aromatic and ribose protons resonances are assigned. For example, guanosine iminos should give NOEs to the H5 of its base paired cytosine, to the H1' of the nucleotide to its 3' side, and to the H1' on the ribose across the strand, one base pair to the 3' side (Heus & Pardi, 1991b). The arrangement of base pairs within the helix should now be unambiguous. Based on the above procedure, it is possible to assign the aromatic (H5, H6, H8), AH2, H1', H2' (some H3'), imino and amino protons of nucleotides in helical regions. Based on the assignments, structural features of the RNA can be determined.

Instrumentation

NMR experiments discussed in this thesis were performed on a variety of instruments. Much of the early work was done on a Varian VXR-500 MHz instrument at MIT. 500 and 600 MHz instruments, built at the Francis Bitter National Magnet Lab (FBNML) were used, as was a Varian 750 MHz magnet. The FBNML instruments were fitted with Nalorac triple resonance (^1H ^{13}C ^{15}N) probes. Sample volumes were 550 or 600 μL, with a minimum RNA concentration of 1 mM. Unless otherwise noted, all exchangeable spectra (90% H_2O, 10% D_2O) were taken at 15 °C, and non-exchangeable experiments (99.996% D_2O) were collected at 25 °C.

One-dimensional H_2O spectra were recorded using 1331 binomial solvent suppression (Hore, 1983) with a sweep width of 12000 Hz, and at least 8K complex data points and 64 scans. The excitation maximum was set to the imino region of the spectrum (3500 Hz on the 500 MHz instrument). Water NOESYs were taken at the FBNML using 3-9-19 WATERGATE (Piotto et al., 1992) solvent suppression. These experiments were done with a sweep width of 12000 Hz in both dimensions, 4K or 8K complex data points in t_2, 512 complex points in t_1, and 32 scans per slice. The recycle delay was set to at least 1.6 s, mixing times were either 150 or 200 ms. Water NOESY data were processed using a combination of gaussian and negative line broadening.

NOESY spectra of non-exchangeable proton resonances were obtained using standard pulse sequences with hypercomplex acquisition to obtain pure phase spectra (States et al., 1982). 4K data points were collected in t_2 with a sweep width of 5500 Hz. A total of 512 t_1 values were collected with 32 scans per t_1 with a recycle delay of 1.6 to 2.0 s. Mixing times were 400, 200, 100, and 50 ms. NOESY experiments were also processed using a combination of gaussian and negative line broadening. NOESYs on the 750 MHz instrument were collected with 4K complex data points in t_2, 512 complex points in t_1, and 16 scans per slice. Sweep widths of 7600 Hz were used. Experiments at the FBNML also involved SCUBA (Brown et al., 1988) experiments instead of NOESYs. The SCUBA experiment involves the addition of a water suppression pulse prior to the actual NOESY sequence. This technique allows recovery of proton resonances obliterated by saturation of the water. It was initially developed for use in observing protein C-alfa proton resonances normally found near the water peak, hence the name SCUBA: Stimulated Cross peaks Under Bleached Alphas (it allows the protons to breathe under water) (Brown et al., 1988).

Phase-sensitive DQF-COSY data were acquired using the standard pulse sequence (Piantini et al., 1982) using the same acquisition parameters as for the NOESYs.

Occasionally experiment time was reduced by collecting 16 t_1 scans. COSY spectra were processed using a shifted sine-bell apodization in both dimensions. TOCSY experiments were recorded with 125 ms MLEV-17 spin lock pulse (Bax & Davis, 1985). Gaussian and negative line broadening functions were used for processing. All two dimensional data were transferred to a Silicon Graphics work station and were analyzed using NMR Pipe version 1.100 (NIH + Laboratory of Chemical Physics, NIDDK).

3 Structural Analysis of Hairpin TAR RNAs

3.1 Peptide Binding Studies

The interaction between the HIV *Tat* protein with its target RNA, TAR, can be modeled *in vitro* using a simplified system. As described in Section 1.2, NMR analysis has been performed on a 31 nucleotide RNA (Figure 1.10) in the free form and in the bound conformation, in the context of complexes with the amino acid analog argininamide or with *Tat*-derived peptides. The model of the RNA bound to argininamide proposes a base-triple interaction between the bulge nucleotide U23 and the A27-U38 base pair in the upper stem. The argininamide binds to G26 via hydrogen bonds in the major groove of the RNA. The two base pairs discussed above (G26-C39 and A27-U38), along with the bulged U23, are the only RNA elements necessary for specific binding of the argininamide or peptides. The simplicity of the TAR-argininamide system allowed us to investigate the sequence requirements of base-triple formation using NMR and a number of TAR RNA mutants. The first class of RNAs, the 'hairpin TARs', were chosen based on the results of a peptide binding assay performed by our collaborator Alan Frankel (J. Tao & A.D. Frankel, unpublished results). For both wild-type TAR and the C•G-C TAR (Figure 1.12), mutations were made in the Watson-Crick base pair of the upper stem (positions 27 and 38). Six RNAs were tested for binding with the R52 peptide (YKKKRKKKKKA) at both pH 7.5 and 5.5. Results are reported in Table 3.1 as binding affinities relative to the wild-type U•A-U base triple.

Table 3.1
R52 binding to TAR mutants.

				Relative Binding Affinity	
				pH	
	N_{23}	N_{27}	N_{38}	7.5	5.5
1	**U**	**A**	**U**	1.0	1.0
2	U	A	C	0.5	1.0
3	U	A	Δ	0.5	0.5
4	C⊕	G	C	0.1	1.3
5	C⊕	G	U	0.15	2.0
6	C⊕	G	Δ	0.2	2.1

There is no difference in binding at the two different pHs for the wild-type RNA (row 1). The data in row 2 show the results of the binding study with RNA in which U38 has been changed to a C. At neutral pH, the A27-C38 base pair is probably in a wobble conformation (Boulard et al., 1992, Boulard et al., 1995). This apposition is expected to be stabilized by only one hydrogen bond (Figure 3.1). NMR (Boulard et al., 1995) and molecular dynamics (Cognet et al., 1995) studies of a DNA molecule containing an A-C apposition indicate that pH-dependent structural variations occur at the site of the mispair. At low pH, adenine can be protonated at the N1 position (Figure 3.1), allowing the formation of a more stable wobble base pair containing two hydrogen bonds. The pK_a for this proton is approximately 4 in free adenosine nucleotides (Saenger, 1984). However this value may increase one to two pK units when the protonated site is found within a base pair of a nucleic acid helix (Hunter et al., 1986, Brown et al., 1990, Wang et al., 1991, Boulard et al., 1992, Boulard et al., 1995). Therefore, the pH dependence of peptide binding by the TAR A-C mutants might fully be explained by structural changes in the RNA induced by protonation of the adenine base. While at both pHs the A and the C of the mispair are expected to be stacked into an RNA helix (Puglisi et al., 1990b), protonation of the mispair may introduce a slight stabilization of the upper stem. The mispair site is only one base pair removed from the trinucleotide bulge. At pH 5.5, the triple interaction in the TAR-UAC mutant may be stabilized by the Hoogsteen and U23-A27 interaction as well as a protonated A27-C38 wobble pair. The low pH peptide binding data indicate that this mutant is equivalent to the wild-type. A mutant in which the nucleotide at position 38 was deleted (denoted by a Δ) binds R52 at 50% of the wild-type level, regardless of the pH of the assay (row 3).

Interesting results were obtained with mutants in the C•G-C triple (Table 3.1, rows 4-6). As expected, all of the C•G-C mutants exhibit low binding to the peptide at pH 7.5 (10-20% of wild-type). However, upon protonation of the bulge C, a stable protonated

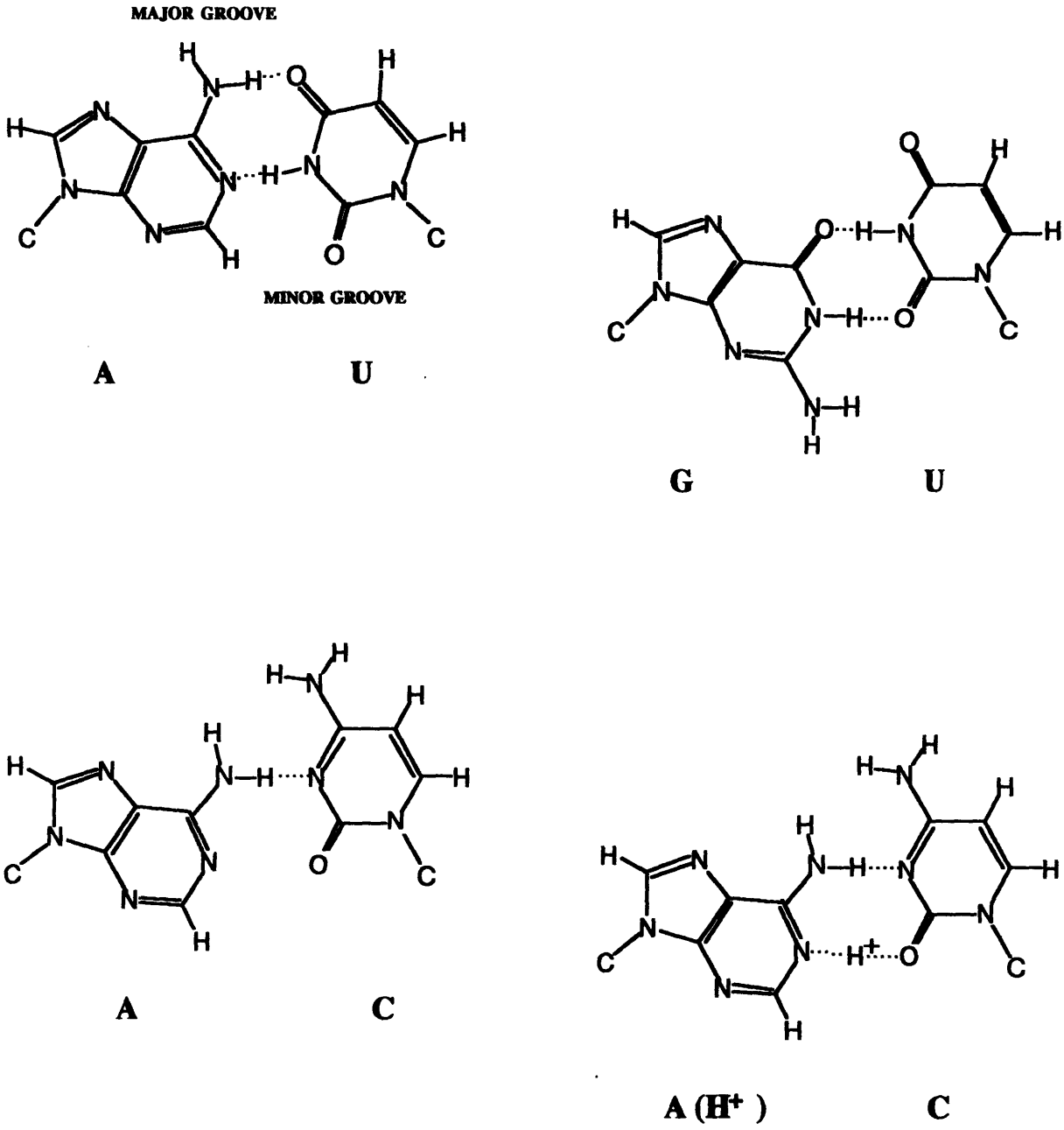


Figure 3.1) Comparison of hydrogen bonding patterns in A-U, G-U, A-C and A^+–C base pairs. At high pH, the A-C pair is stabilized by only one hydrogen bond (dashed line) in the wobble conformation. An additional hydrogen bond forms upon protonation of the adenine N1. The G-U mispair, commonly found in RNA structutres, is nearly as stable as an A-U pair. The C1'-C1' distance is approximately 10.3 Å in the A-C and G-U mispairs (0.7 Å shorter than in Watson-Crick base-pairs), causing only minor helical distortions (Puglisi et.al., 1990).

C•G Hoogsteen base pair can form in addition to the Watson-Crick G-C pair at positions 27 and 38. At low pH, all of the triples with the C•G Hoogsteen base pair bind to the peptide even better than the U•A-U RNA does. Data in row 4 correspond to the C•G-C mutant studied by NMR (Puglisi et al., 1993). This RNA was shown by NMR to adopt a structure equivalent to the wild-type RNA in an argininamide solution at low pH. The mutations in which the stem G-C Watson-Crick base pair is changed to a G-U mispair (row 5; Figure 3.1) or abolished by deletion of N38 (row 6) *increase* peptide binding two-fold.

Based on these results, the pH-dependent structures of one RNA with a mispair at the 27-38 position, and one with the deletion at nucleotide 38 were investigated. NMR analysis was performed at pH 6.4 and 5.5, in the presence and absence of argininamide, with the U•A-C and the C•G-Δ mutants (Figure 3.2). For reasons discussed below, both of these samples proved to be unsuitable for structural analysis.

3.2 Deletion Mutant TAR-CGΔ

Sample Preparation

The 47 nucleotide DNA template GGC CAG AGG CTC CCA GGC CCA GGT CTG GCC TAT AGT GAG TCG TAT TA was synthesized and purified as described in Materials and Methods. A 50 mL transcription was performed with NTP concentrations at 4 mM each. The ratio of Mg^{++} to (NTP + GMP) concentrations was 1.75:1. A total of 8 preparative gels were run for 19 hours at 60 W. 10 mg of product RNA was obtained after the final dialysis, yielding a 1.8 mM sample (600 μL). The NMR buffer was 10 mM Na-phosphate pH 6.4, 50 mM NaCl, 0.1 mM EDTA.

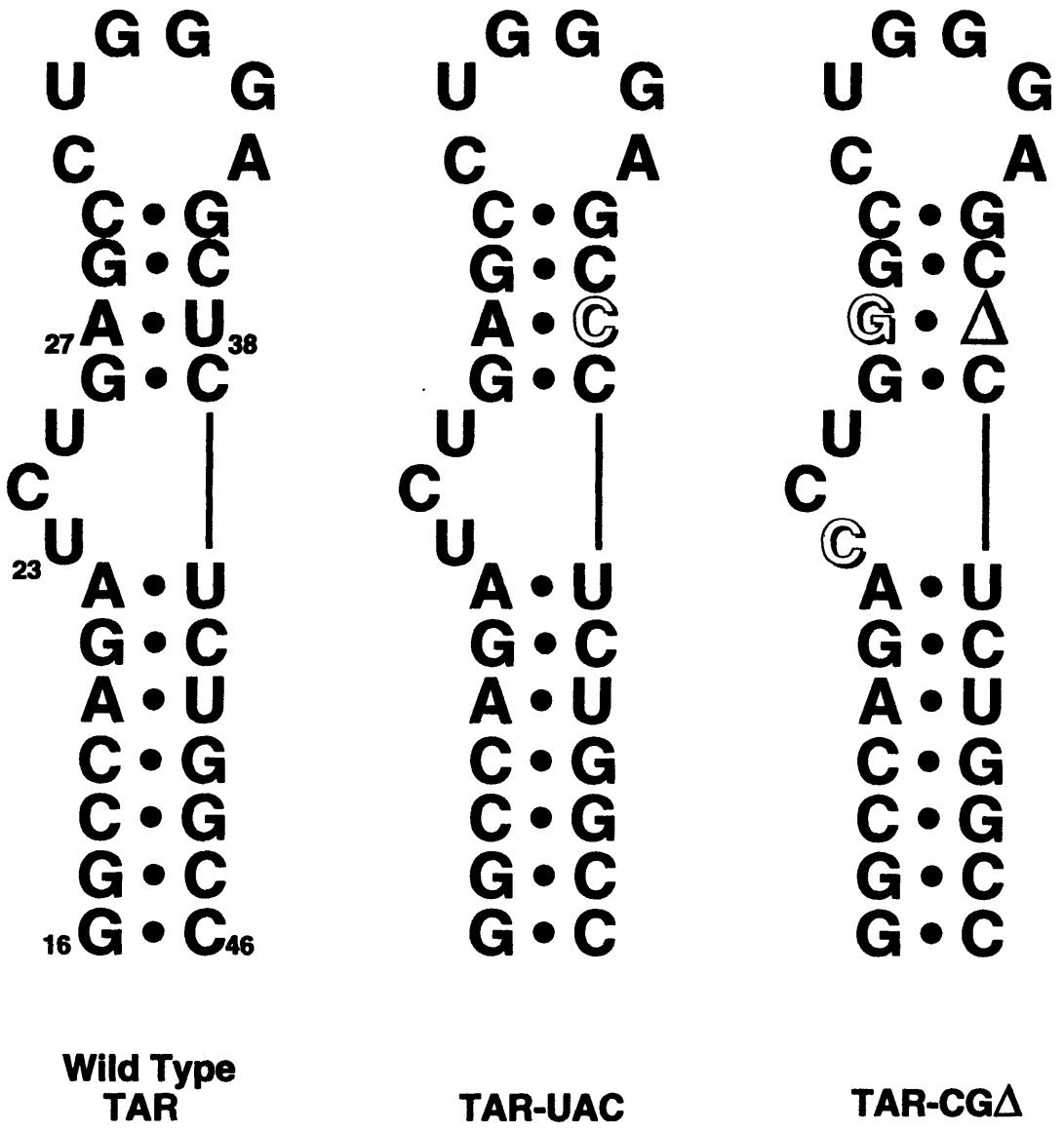


Figure 3.2) TAR mutants synthesized for NMR. The TAR-UAC mutant has an A_{27}-C_{38} mispair instead of the wild-type A-U pair. The bulge nucleotide U_{23} is unchanged. TAR-CGΔ is missing the base pair at the 27-38 position, but is expected to be able to form a C_{23}-G_{27} Hoogsteen pair at low pH. Mutations are shown in outline.

Characterization

Gel chromatography (see Materials and Methods) of TAR-C•GΔ at a concentration of 1.8 mM showed the presence of a negligible amount of dimer. Therefore, an NMR analysis of this molecule was undertaken. Initial characterization of the RNA secondary structure was performed in water. At least twelve imino resonances were observed (Figure 3.3). Only nine resonances are expected from the base pairing scheme of the wild-type TAR structure (Puglisi et al., 1992) (Figure 3.2). The abundance of broad peaks suggests that the molecule is undergoing some type of conformational exchange. The presence of an upfield resonance (ca. 10.4 ppm) giving a strong NOE to another imino proton (ca. 11.7 ppm) in saturation transfer experiments (data not shown) indicates the presence of a G-U mispair (see Chapter 2, also, Figure 3.1). A G-U pair is not expected to exist in the TAR-like structure of an RNA containing this sequence. TAR-C•GΔ may adopt the alternate structure shown in Figure 3.4. This RNA was exchanged into D_2O for analysis of the non-exchangeable proton resonances. Two-dimensional experiments in D_2O provided additional evidence for exchange. There are 15 pyrimidine nucleotides in TAR-C•GΔ. Therefore, 15 cross peaks in the H5-H6 region of a COSY spectrum are expected. Depending on experimental conditions, at least 17, and occasionally 19 resonances were observed. A comparison of the NOESY cross peak arrangement in the aromatic to H1' region also differed markedly from the wild-type spectrum.

TAR-C•GΔ was also tested for binding to argininamide. It was hoped that the presence of argininamide would cause a shift in equilibrium towards the wild-type conformation. Titration to 6 mM argininamide was monitored by changes in the imino spectrum in water. No significant spectral changes were observed. The presence of 6 mM argininamide only served to broaden the imino peaks. The sample was observed in the pH range of 7 to 5.5, and at a variety of temperatures. At no point did this sample adopt the wild-type conformation. While this molecule binds to *Tat*-derived peptide at low pH two

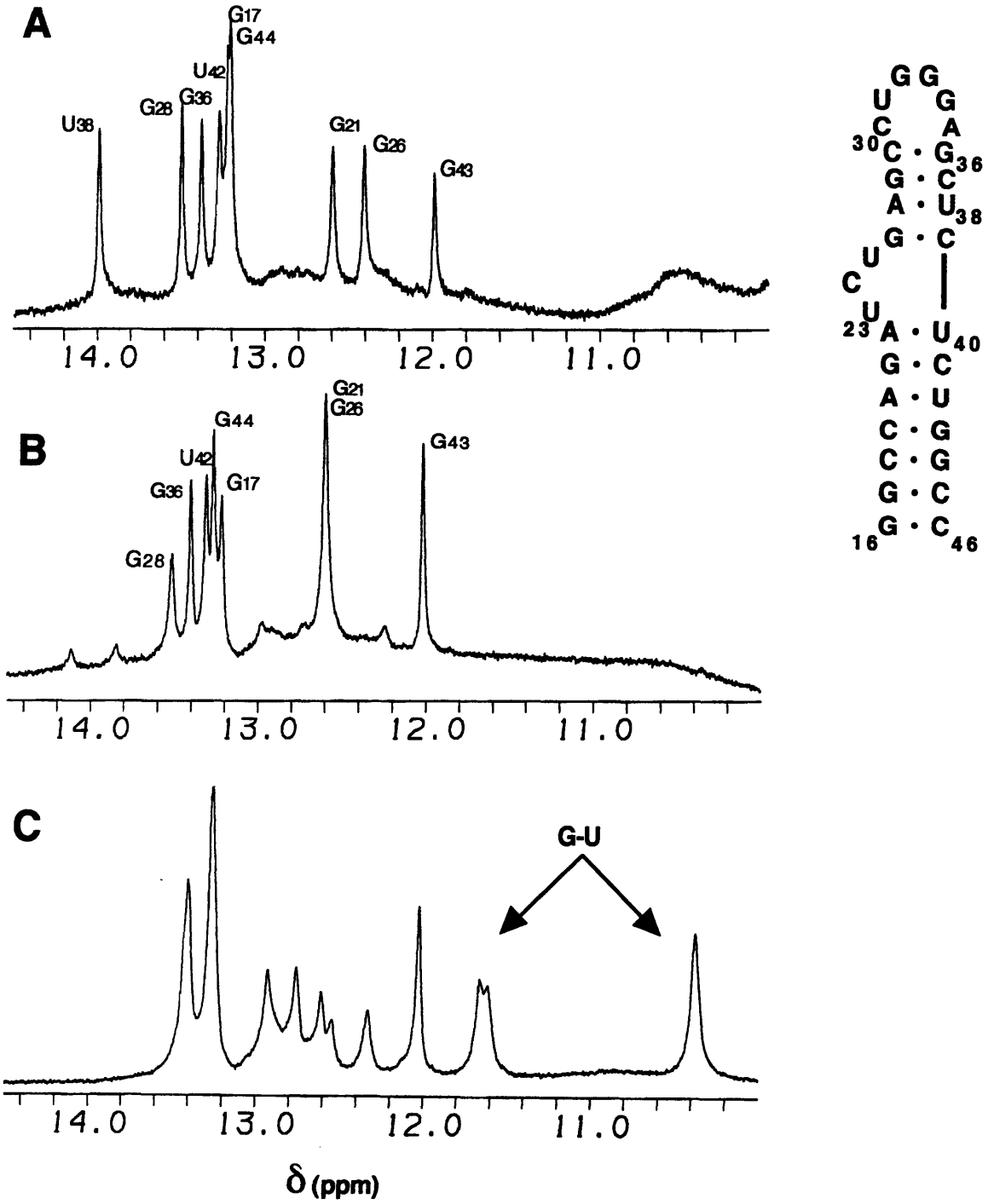


Figure 3.3) Comparison of imino spectra. The spectrum in **A** is the wild-type TAR RNA, peaks are assigned as labeled. In **B** the TAR-UAC mutant shows as similar arrangement of imino resonances,. with the exception of the U38 peak. G21 and G26 are overlapped while G28 is much weaker, suggesting a small conformational difference in base pairs near the bulge and A-C apposition relative to the wild-type. An additional set of weak resonances suggests a small population of an alternate structure. The spectrum of the TAR-CGΔ mutant (**C**) is markedly different from the other two. This RNA does not adopt a TAR-like structure. Peaks determined to be from the G-U pair are labeled.

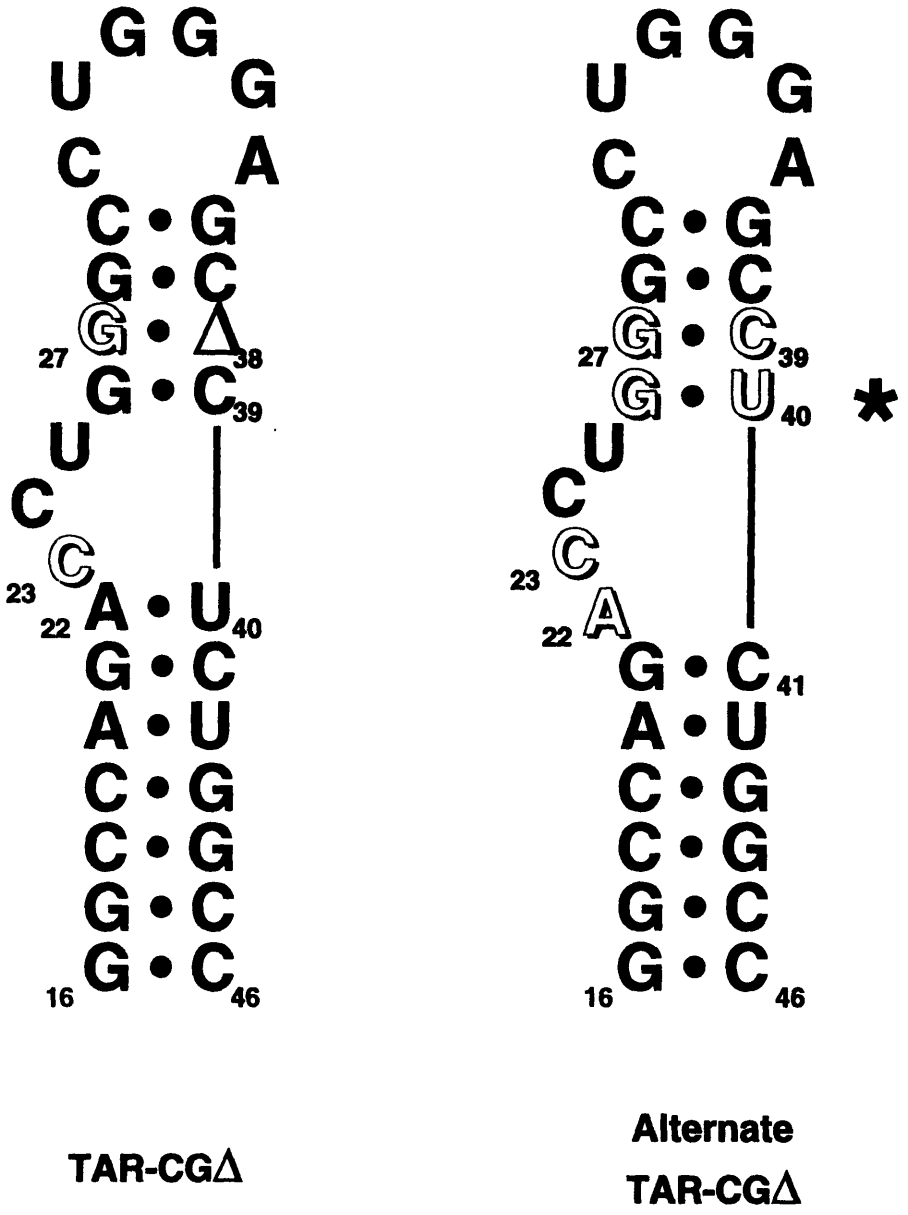


Figure 3.4) Alternate conformation of TAR-CGΔ. NMR experiments detected the presence of a G-U mispair in TAR-CGΔ. Accordingly, an alternate RNA conformation was proposed. The deletion at position 38 may cause a shift in one RNA strand, allowing C39 to pair with G27 and U40 to pair with G26. A22 is now a part of the four nucleotide bulge. Two stable stems result, including the formation of a G-U mispair marked by an asterisk (*).

fold better than the wild-type TAR (Table 3.1), it probably is doing so in a different conformation. There is evidence that short *Tat*-peptides have difficulty discriminating between wild-type TAR and RNAs containing different bulged nucleotides (Weeks & Crothers, 1991, Weeks & Crothers, 1992, Long & Crothers, 1995). A bulge-induced bend in the RNA helix may be important for Tat peptide binding. Although the TAR-CGΔ RNA may posses an interesting structure, the aim of this project is to investigate TAR-like structures. Hence, the sample was stored at -20 °C, perhaps to be investigated at another time.

3.3 Mispair Mutant TAR-UAC

Sample Preparation

A 31 nucleotide RNA was synthesized under standard conditions using T7 template-driven transcription as described in Materials and Methods. The complementary DNA sequence used was: GGC CAG AGG GCT CCC AGG CTC AGA TCT GGC CTA TAG TGA GTC GTA TTA. A total of 9 preparative gels were necessary to purify RNA from a 50 mL transcription. 6 mg of RNA were obtained, enough for a 1 mM sample of 600 μL. As with the deletion mutant, the molecularity of the sample was tested by gel filtration chromatography and found to be predominantly monomer. The sample was prepared in 90% H_2O/10% D_2O with 10 mM Na-phosphate pH 6.4, 50 mM NaCl, 0.1 mM EDTA.

Preliminary Analysis

The first step in characterization of the RNA was obtaining a spectrum of the imino resonances in order to determine the base pairing scheme. A comparison of the imino spectra for the wild-type TAR RNA, the deletion mutant (TAR-CGΔ) and the mispair

mutant (TAR-UAC) is shown in Figure 3.3. Unlike TAR-CGΔ, the TAR-UAC RNA spectrum is similar to that of the wild-type RNA. Seven resonances are observed, including one at 12.6 ppm corresponding to two protons. Therefore eight base pairs can be accounted for. As with the wild-type, no imino peaks are seen for the terminal G16-C46 base pair, or the A22-U40 pair found just below the bulge. The remaining base pairs are accounted for, with the exception of the A-C mispair which is not expected to show a resonance in this region (adenine and cytidine do not have imino protons). Preliminary D_2O experiments indicate that the mispair mutant adopts a TAR-like structure in solution. COSY spectra contained the requisite number (16) of H5-H6 peaks corresponding to pyrimidines. NOESY experiments suggest a structure similar to that of the wild-type TAR RNA. Preliminary assignments were made for the aromatic, H1' and H2' protons in the free RNA.

Imino proton spectra were recorded at 0.5 mM additions of argininamide to a final concentration of 6 mM. The change in RNA structure was monitored by the changing chemical shifts of the RNA imino protons. At 3.0 mM argininamide, the peak at 12.6 ppm (G21 and G26) separates into two resonances, indicating some argininamide-induced changes in RNA conformation. However, many of the other peaks broaden. It was suspected that this molecule is involved in some type of conformational exchange. The changes in the imino resonance chemical shifts ($\Delta\delta$) never reach the same magnitude as seen with the wild-type RNA. D_2O characterization was performed at 6.0 and 7.5 mM argininamide, under several different conditions. Experimental variables included temperature, pH and salt concentration. No experimental conditions allowed observation of the C38 H5-H6 cross peak in a COSY spectrum of an argininamide-containing sample. After some time, this molecule, too, was set aside for future analysis. At this point, it was decided to modify the RNA sequence in order to stabilize the upper stem of the molecule.

3.4 Tetraloop Mispair TAR (tTAR-UAC)

Rationale

The loop conformation of the wild-type TAR did not change significantly upon argininamide or *Tat* peptide binding (Puglisi et al., 1992); we were therefore confident that changing the loop sequence would not hinder the argininamide-binding properties of mutant TARs. The six nucleotide apical loop in TAR was replaced with the four nucleotide 'tetraloop' sequence GAAA. There are two reasons for choosing to use a tetraloop to replace the wild-type loop. The first is thermodynamic. A family of tetraloop sequences was determined to be unusually stable (Anato et al., 1991). These sequences include the UNCG, CUUG and GNRA tetraloops where N is any nucleotide, and R is a purine (A or G). It was hoped that the extra stability provided by the tetraloop would eliminate the exchange problems incurred by having an A-C mispair in the upper stem of the TAR molecule. The other reason is practical. High resolution structures have been determined for a number of tetraloop-containing RNAs (Heus & Pardi, 1991a, Varani et al., 1991, Hines et al., 1994, Jucker & Pardi, 1995, see also Shen et al., 1995). Frequently these structures give rise to unusual chemical shifts, allowing for a starting point in assigning NMR spectra. The chemical shift assignments of the GAAA tetraloop were kindly provided by Fiona Jucker (University of Colorado, Boulder). Because we are interested in an RNA containing a U•A-C triple, the GAAA tetraloop sequence was chosen as it lacked pyrimidine nucleotides. The GA_3 tetraloop also requires that the loop is closed by a C-G base pair, which is present in TAR RNA. A G-A pair is formed between the first and last loop nucleotides, leaving only two true loop nucleotides, and these are thought to stack on the 3' side of the loop (Heus & Pardi, 1991a), see Figure 3.5 . A similar rationale was recently employed by Varani and coworkers (Aboul-ela et al., 1995). These authors used a

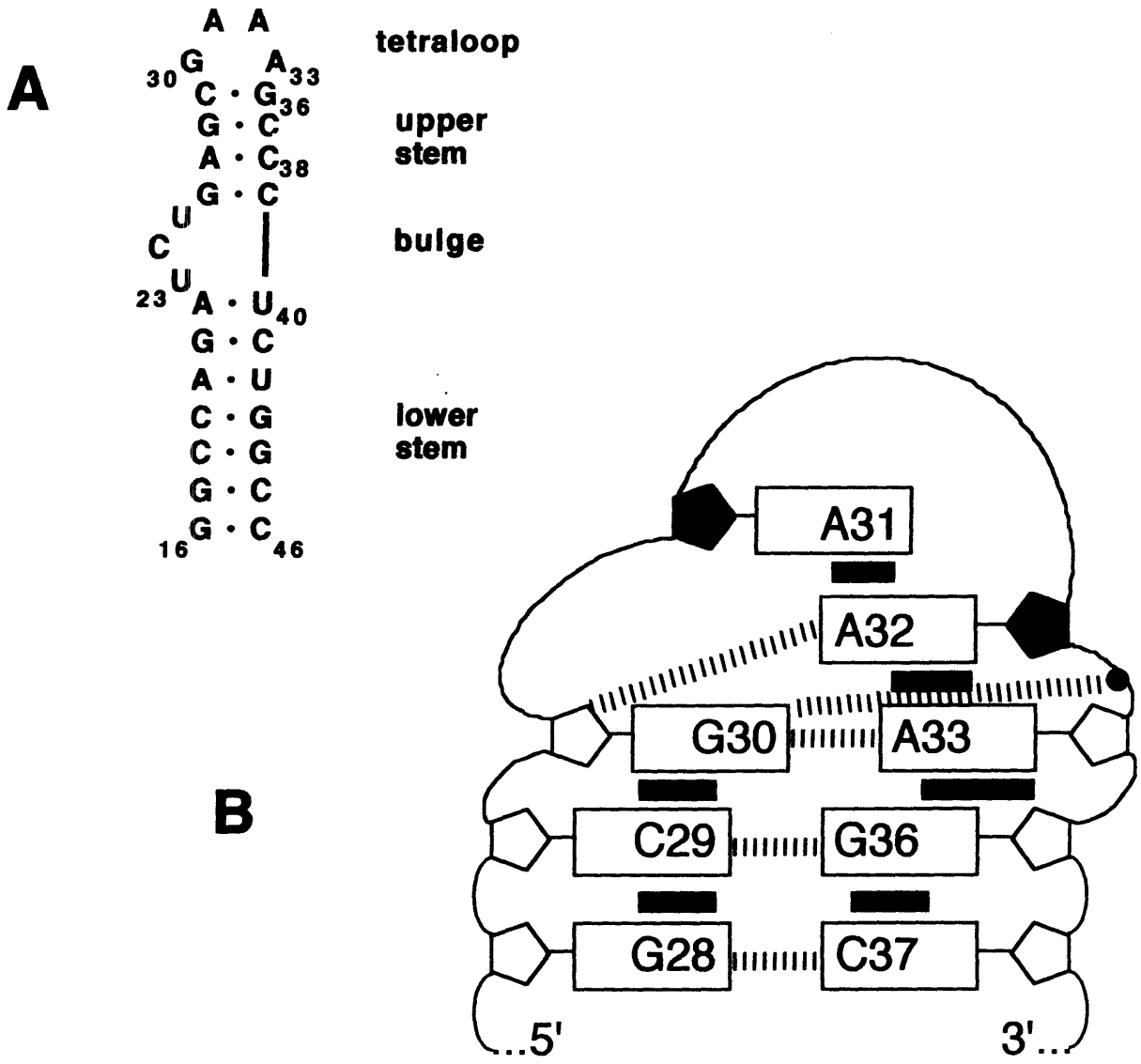


Figure 3.5) Structure of the GA_3 tetraloop. **A)** The secondary structure of the tTAR-UAC RNA. Numbering for all stem and bulge nucleotides is as in wild-type TAR, the tetraloop is numbered from 30 to 33. **B)** Schematic representation of the ultrastable GAAA tetraloop. For clarity, nucleotide numbering shown is as in **A**. Bases are represented by rectangles, C3'-*endo* sugars by pentagons (sugars with at least 30% C2'-*endo* conformation are shaded). Dashed lines indicate hydrogen bonds, thick black lines represent stacking interactions. Extensive stacking and hydrogen bonding in the loop account for the extra stability of tetraloops. A base pair between G30 and A33 continues the helix, leaving only A31 and A32 to bridge the stem. P32, shown as a black dot, forms a cross-strand hydrogen bond with the base of G30. An additional stabilizing hydrogen bond comes from the A32 (base) to G30 (sugar) hydrogen bond (Heus and Pardi, 1991). Adapted from Shen et. al., 1995.

UUCG tetraloop to stabilize a 27 nucleotide TAR hairpin for NMR analysis with a 37 residue *Tat* peptide.

Sample Preparation

A 20 mL transcription was sufficient for preparation of *two* 1.8 mM samples containing the 29-nt GAAA tetraloop TAR with an A-C opposition at positions 27 and 38 (tTAR-UAC). Several conditions were assayed, however standard conditions (4 mM ea. NTP, 1.75:1 $MgCl_2$:(NTP + GMP) produced optimum product yield. Production of this RNA was remarkably efficient, yielding 1.3 mg/mL transcription. Four 20% preparative gels were used, run 19 hours at 60 W. After dialysis, size exclusion HPLC analysis provided no evidence of RNA dimerization at a concentration of 2 mM. Initial NMR studies were performed in 10 mM Na-phosphate pH 6.4, 50 mM NaCl, 0.1 mM EDTA.

Assignment of Exchangeable Protons.

Exchangeable proton spectra of tTAR-UAC were collected at various temperatures. Nine imino resonances are observed and were assigned as indicated in Figure 3.6. Initial assignments were made from 1-dimensional NOE experiments, and a more detailed characterization was made via 2-dimensional WATERGATE NOESY data (see Materials and Methods). Spectra were collected at 15 °C, with a sweep width of 12000 Hz on a 500 MHz spectrometer. A mixing time of 200 ms was used. A total of 512 t_1 values were collected with 32 scans per t_1, and a relaxation delay of 1.8 s. 4K data points were collected in t_2.

Imino resonances give characteristic NOEs to their base pairing partners (see Materials and Methods). Seven guanosine iminos were determined to each give the characteristic NOE pattern to two cytosine amino resonances. Only one uridine was found; it gives a strong NOE to an adenine H2 at 7 ppm. The imino-imino NOEs show a stacking pattern of GGGUG. An additional pattern of GG was observed. These resonances were

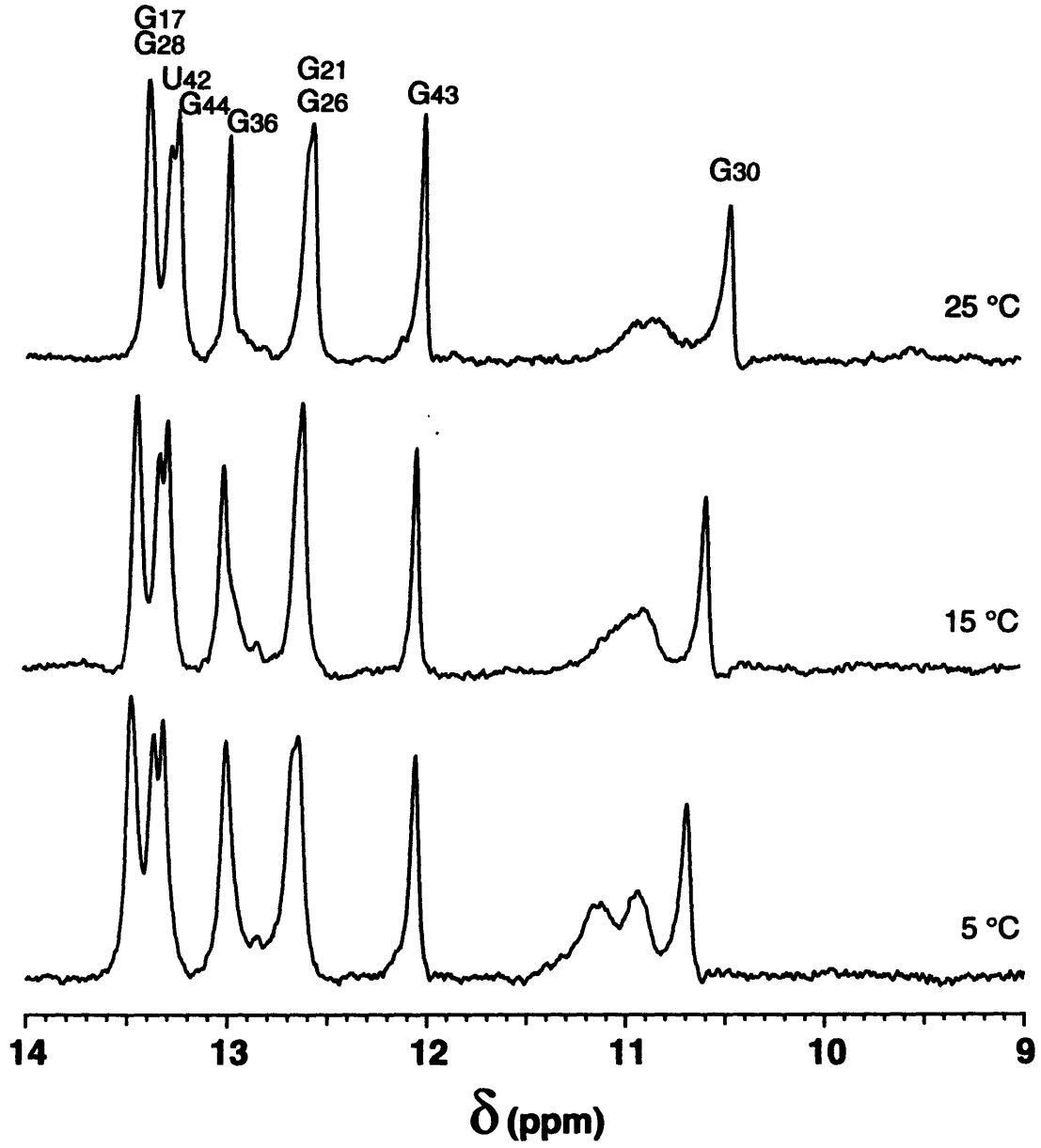
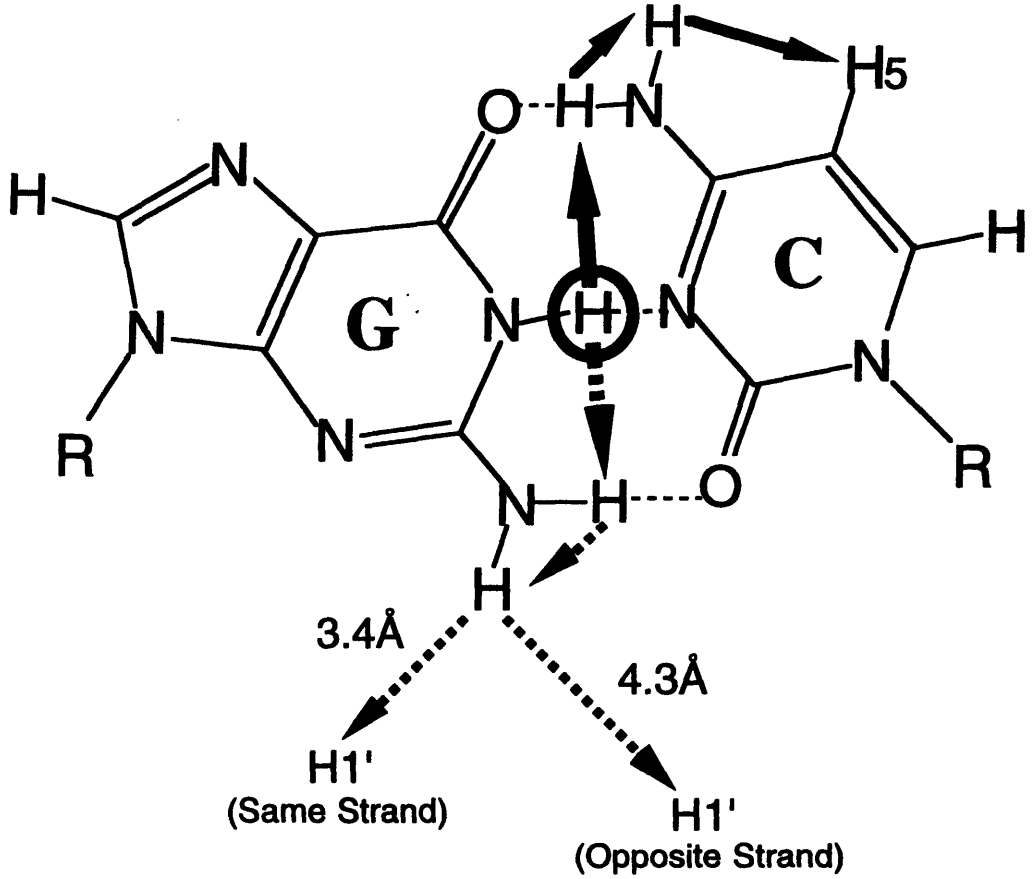


Figure 3.6) Imino resonances for tTAR-UAC. Eight base pairs were determined from the imino region of the RNA spectrum in water. Peaks are assigned as labeled. The imino resonances for free tTAR-UAC do not change much with temperature. The broad peak at ca. 11 ppm is due to non-hydrogen bonded imino protons. The G30 imino proton is protected from exchange with solvent by the unique conformation of the tetraloop (Heus and Pardi, 1991). The sample is 1.8 mM in 10 mM Na-phosphate pH 6.4, 50 mM NaCl, 0.1 mM EDTA. Chemical shifts are referenced to TSP.

easily assigned to the lower and upper stems, respectively. No imino resonances are found for G16 or U40. This is not surprising since base pairs at the ends of helices are often dynamic, increasing the exchange rate of the imino protons. Similarly, the U40 imino proton remained unassigned in wild-type TAR (Puglisi et al., 1992). An NOE was assigned to the G26 imino proton, although no stacking NOEs are observed to it, consistent with its placement between the bulge and the A-C mispair. There are no imino protons in an A-C apposition, so the lack of a resonance for this base pair is not at all surprising. Based on the imino-imino NOEs, the apical two G-C base pairs are determined to be stacked. The broad peak at ca. 11 ppm corresponds to non-hydrogen bonded bulge imino protons that exchange too rapidly to be observed. Accordingly, this resonance begins to sharpen as the temperature is lowered, slowing the rate of exchange (Figure 3.6). The resonance at 10.54 ppm gives no stacking NOEs. In agreement with the known GA_3 tetraloop assignments, this resonance is determined to belong to G30. Although the G-imino proton is not involved in the proposed G-A pair, Heus and Pardi suggest that this proton is in some way protected from exchange by the unique conformation of the tetraloop nucleotides (Heus & Pardi, 1991a).

Cytosine amino resonances are normally assigned from G-imino-C-amino cross peaks. In the absence of an observable imino resonance (i.e. for the terminal G16-C46 base pair), data from the non exchangeable NOESY experiments can help in assigning the C-amino protons. This is the case for the C46 amino protons. Upon assignment of the C46 H5 chemical shift from non exchangeable NOESY data, its amino resonances are determined. The guanine imino resonance pathway is useful for confirming non-exchangeable assignments. In addition to the common G-imino-C-amino-CH5 NOE pathway, helical G-imino protons will also give NOEs to H1' protons on nucleotides 3' to the G (same strand) and to H1' of the sugar on the opposite strand, one nucleotide 5' of the base paired C (Figure 3.7). Cross-strand NOEs are always important when assigning

MAJOR GROOVE



MINOR GROOVE

Figure 3.7) G-imino NOE pathway. Exchangeable and nonexchangeable NMR data are connected by the guanine imino resonance assignment pathway. The guanine imino proton gives NOEs to the hydrogen bonded amino protons in a G-C base pair. These protons are in chemical exchange with the non-hydrogen bonded aminos. Spin diffusion through the cytosine amino will generate NOEs to the H5 proton (solid lines). Similarly, in A-form helices, the guanine amino resonance will give one NOE to the H1' on the following (n+1) nucleotide, and one across the minor groove of the helix to the H1' of the nucleotide in the n-1 (3') direction (dashed lines). Adapted from Heus & Pardi, 1991b.

RNA helices. The G-imino-H1' assignment procedure proves to be very useful when ambiguities arise in assigning D_2O spectra. This was the case in assigning the lower stem of TAR molecules that contain two G-C base pairs in a row followed by two C-G base pairs. Most of the G-imino protons gave the expected amino, CH5, and H1' NOEs. In the region of the C-A mispair, cross peaks from the G28 imino (just above the A-C apposition) to C37 aminos, a weak NOE to C37 H5, C29 H1' and C38 H1' are observed. The G26-imino (just below the A-C) gives weak NOEs to A27 H1' and C39 H5, but the C39 aminos are not observed, nor is an NOE to the U40 H1' located at the base of the bulge. It appears that nucleotides in the A-C mispair are stacking into the helix, although there is some distortion from normal A-form geometry.

Because the U40 imino proton is not observed, a uridine imino to adenine H2 NOE is not obtained for the U40-A22 pair. An imino-AH2 NOE is not observed for A27 in the exchangeable spectra due to the absence of an imino proton in the A-C mispair. The U42 imino to A20 H2 NOE is observed as a very strong cross peak. The A20 H2 is in a well-resolved part of the D_2O NOESY and was a useful starting place for the assignment of non-exchangeable resonances. The exchangeable proton assignments are presented in Table 3.2 with the non-exchangeable data discussed below.

Assignment of Non-exchangeable Protons.

Non-exchangeable data presented here are from spectra taken at 25 °C. The most informative NOESYs were acquired with 300 ms and 40 ms mixing times on the 750 MHz Varian spectrometer at the FBNML. DQF-COSYs, 400 ms and 50 ms NOESYs were also acquired on a Varian 500 MHz instrument at MIT. Parameters were as described in Materials and Methods.

The non-exchangeable protons were assigned according to published procedures (Haasnoot et al., 1984, Varani & Tinoco, 1991) outlined in Materials in Methods.

Table 3.2
Assignments for tTAR-UAC (pH 6.4)
Chemical shifts are reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino	Amino
G16	8.16	NA	5.82	4.95		
G17	7.67	NA	5.93	4.59	13.45	
C18	7.71	5.28	5.57	4.45	NA	6.91/8.63
C19	7.72	5.57	5.49	4.55	NA	6.89/8.38
A20	8.02	7.00	5.91	4.70	NA	
G21	7.07	NA	5.50	4.47	12.63	
A22	7.60	7.64	5.90	4.39	NA	
U23	7.42	5.32	5.59	4.13		NA
C24	7.75	5.86	5.81	4.30	NA	
U25	7.85	5.82	5.85	4.47		NA
G26	7.92	NA	5.87	4.51	12.60	
A27	8.10	8.07	6.18	4.43	NA	
G28	7.24	NA	5.64	4.82	13.35	
C29	7.37	5.12	5.58	4.77	NA	6.78/8.27
G30	7.56	NA	5.73	4.47	10.54	
A31	8.40	7.83	5.72	4.78	NA	
A32	7.97	7.81	5.45	4.34	NA	
A33	8.21	8.10	6.02	4.64	NA	
G36	7.85	NA	3.64	4.29	13.00	
C37	7.50	5.19	5.51	4.59	NA	6.97/8.58
C38	7.98	5.77	5.67	4.42	NA	
C39	7.76	5.65			NA	
U40	7.89	5.68	5.65	4.30		NA
C41	7.98	5.87	5.54	4.38	NA	7.07/8.34
U42	7.87	5.44	5.51	4.63	13.32	NA
G43	7.72	NA	5.79	4.61	12.02	
G44	7.33	NA	5.69	4.49	13.27	
C45	7.64	5.20	5.53	4.29	NA	6.94/8.61
C46	7.69	5.54	5.79	4.03	NA	7.02/8.40

NA refers to not applicable protons, blank spaces indicate protons that were not assigned.

Chemical shift assignments are reported in Table 3.2. The aromatic-H1'/H5 region of the 300 ms NOESY (pH 6.4) used for assignments is shown in Figure 3.8. H5-H6 cross peaks were useful in assigning pyrimidine nucleotides. Their chemical shifts were obtained from COSY spectra (see below) and confirmed with H_2O NOESY data. Helpful NOEs also arise in this region of NOESY spectra acquired at long mixing times between pyrimidine H5 and aromatic protons on the preceding (5') nucleotide. The adenine H2 to (n+1) H1' and complementary strand H1' (n–1) are also located in this region. Stacking information can be derived from aromatic to aromatic and H5-H5 NOEs from other regions of NOESY spectra (Figure 2.8).

Starting points for the connectivity pathway of tTAR-UAC were the A20 H2 (assigned from exchangeable data described above) NOEs to G21 and G43 H1' protons. The known GA_3 tetraloop chemical shift assignments were also useful in beginning the assignments. The unique structure of the G-A pair (Heus & Pardi, 1991a) causes the H1' of the apical stem guanine to be located directly under the preceding adenine H8, causing a large ring current effect. Thus, in tTAR-UAC the G36 H1' was expected to be located far upfield of other H1' protons. A chemical shift of 3.64 ppm was assigned to G36 H1'.

General Features

The assignment procedure was straightforward. Two A-form helical regions are formed as expected from the secondary structure. Nucleotides G16 to A22 and C41 to C46 are stacked in normal A-form helical geometry as evidenced from strong H8/H6 to n-1 H2' NOEs. Despite the lack of a U40 imino resonance, A22 and U40 are found to stack in a helical geometry. Nucleotides in the upper stem are predominantly stacked as well, including the nucleotides in the A-C mispair. Extensive aromatic to aromatic, aromatic to H1' and H2', H5-H5 and H1'-H1' NOEs are observed in tTAR-UAC. In long mixing time experiments, aromatic and H1' to n-1 H5 NOEs were useful in clarifying assignments for portions of the sequence with clusters of pyrimidines (Figure 3.8). Adenine H2

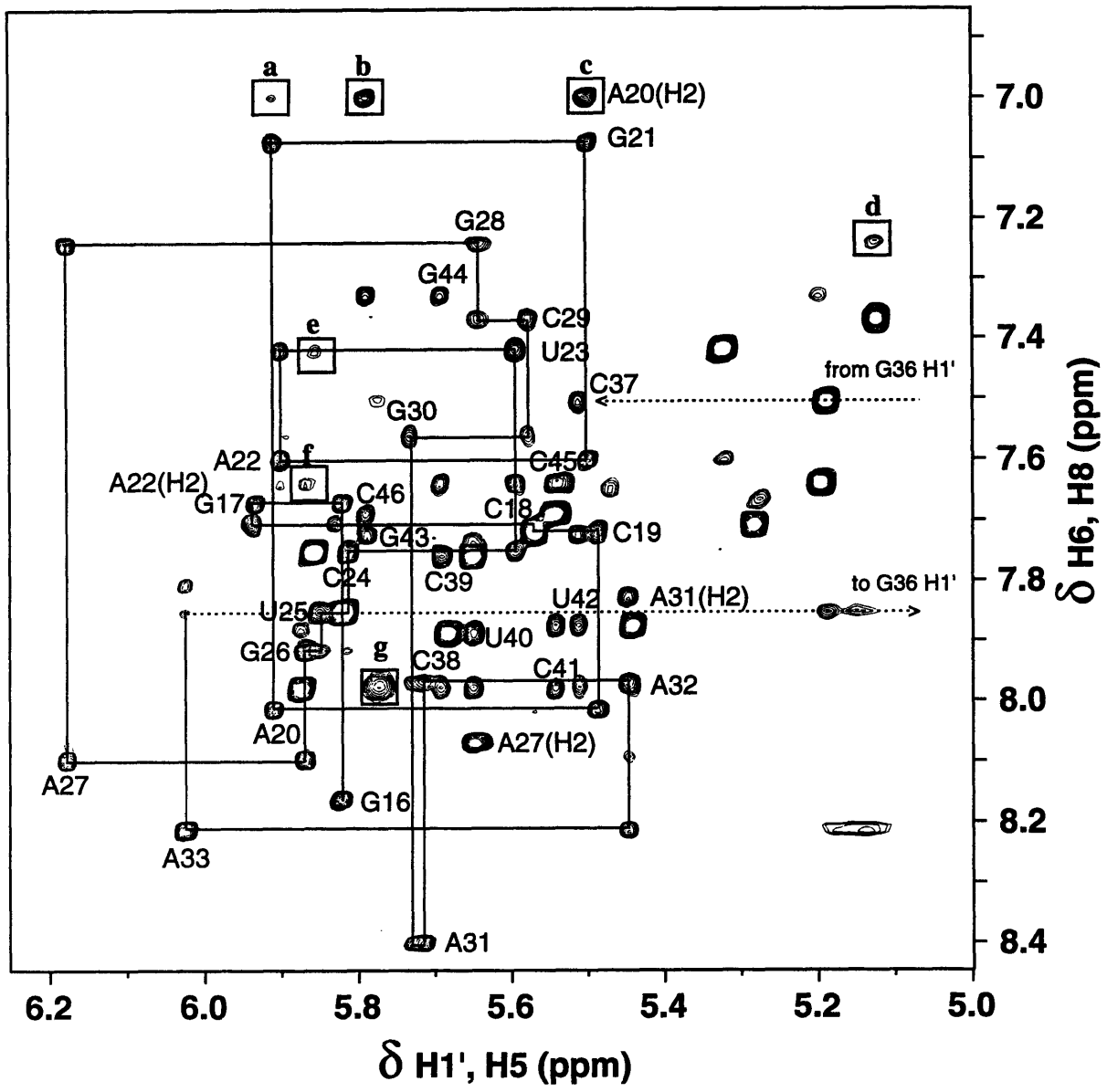


Figure 3.8) Aromatic to H1' region of the tTAR-UAC NOESY spectrum. The aromatic to H1' connectivity pathway for one strand of the molecule, residues 16-36 is shown by the solid line. A20 H2 was a starting point for assignment. The weak peak in box **a** is the A20 H2 NOE to its own H1'; boxes **b** and **c** are the NOEs to G43 H1' and G21 H1', respectively. The peak in box **d** represents an aromatic to H5 NOE from G28 to C29. A similar stacking NOE is shown in box **e** between U23 and C24. The bulge nucleotides U23, C24 and U25 are predominantly stacked between the upper and lower helices. Box **f** contains the weak NOE from A22 H2 to G26 H1', indicative of partial unstacking of the bulge, with concomitant coaxial stacking of the helices. The broad peak in box **g** is the C38 H5-H6 cross peak. This spectrum was recorded at 25 °C with a 1.8 mM RNA sample in 99.996% D2O, 10 mM Na-phosphate pH 6.4, 50 mM NaCl, 0.1 mM EDTA.

assignments were established for the remaining adenine nucleotides based on the expected interstrand and cross-strand NOE patterns. With the exception of C39, all aromatic, H2, H5, H1' and H2' protons were assigned. Assignments were confirmed with data from the water NOESY and DQF-COSY. Interesting features of the NOE patterns are discussed below.

Bulge Nucleotides

All three bulge nucleotides appear to be partially stacked between the two stems. This conformation is consistent with that of the wild-type TAR. Aromatic to H1' and H2', as well as aromatic-aromatic NOEs are observed between all three nucleotides. The A22 H2 also gives an NOE to U23 H1'. However, a weaker NOE is also seen between A22 H2 and G26 H1', indicating that the two stems are stacked at least part of the time, causing the extrusion of the bulge nucleotides. Table 3.3 lists NOEs between nucleotides in the bulge region. An aromatic-aromatic NOE is observed between C39 H6 and U40 H6. Upon comparison with the wild-type TAR conformation (Puglisi et al., 1992), it is expected that these nucleotides would be stacked in approximate A-form geometry. Recall that the NOE patterns for the exchangeable data (G26 imino) indicate some local distortion below the A-C apposition. The G26 imino to C39 H5 NOE is weak, and the imino-amino NOEs are not observed for this base pair. The G26 imino proton did not give the cross-strand to the U40 H1', an indication that this distance may be longer than the 4.3 Å predicted for normal A-form geometry (Figure 3.7). Another indication of A-form geometry would be the U40 to C39 H2' NOE. Unfortunately, the C39 H1' and H2' protons are not assignable. Difficulty in assigning C39 may be due, in part, to conformational flexibility. An NMR study (Puglisi et al., 1990b) of an RNA hairpin containing A-C and G-U mispairs showed this to be the case with the nucleotide located 3' to the C of an A-C apposition. It is likely that the proximity of the G26-C39 base pair to both the A-C apposition and the bulge serves to distort the local conformation of the helix, causing difficulties assigning these nucleotides.

Table 3.3
Internucleotide NOEs in the bulge region of tTAR-UAC (pH 6.4).

N	N+1		N	N+1
A22 H8	U23 H6		C38 H6	U40 H6
A22 H8	U23 H5			
A22 H1'	U23 H5			
A22 H1'	U23 H6			
A22 H2'	U23 H6			
A22 H2	U23 H1'			
U23 H6	C24 H6			
U23 H6	C24 H5			
U23 H5	C24 H5			
U23 H1'	C24 H6			
U23 H1'	C24 H5			
U23 H1'	C24 H1'			
U23 H2'	C24 H6			
C24 H6	U25 H6			
C24 H1'	U25 H6			
C24 H2'	U25 H6			
U25 H6	G26 H1'			
U25 H6	G26 H2'			

A-C Mispair

With the exception of C39, the upper stem, including the A-C mispair, was assigned. Aromatic-aromatic NOEs are seen between all bases on both strands of the helix, including C39 H6 to the H6s of U40 and C38. Inspection of the NOESY spectrum (Figure 3.8) shows one broad cross peak from A27 H2 (8.07 ppm) to G28 H1' (5.64 ppm). It is possible that this peak could represent two NOEs to protons with the same chemical shifts. C39 H1' may be located at 5.64 ppm, but the cross peak would be obscured due to overlap with its own H5-H6 resonance. C37 H6 gives NOEs to C38 H6, H5, H1', and H2' protons. H5-H5 and H1'-H1' NOEs are also seen for this pair. C38 seems to be incorporated into the helix in approximate A-form geometry. The connectivity pathway breaks down at the nucleotide just 3' to the C of the A-C mispair (C39). A27 gives all of the same-strand NOEs expected for a nucleotide in an A-form helix, although the presence of a cross-strand NOE from A27 H2 to C39 H1' would provide strong evidence to support this geometry. The deviation from classical A-form connectivity pathway on the 3' strand suggests some local distortion for nucleotides C38 through C41, just below the A-C mispair and across from the bulge.

DQF-COSY Spectra

Fourteen well resolved H5-H6 cross peaks are seen in the DQF-COSY spectrum of tTAR-UAC. The peak corresponding to C38 H5-H6 appears to be slightly weaker than the others. The COSY data were used to determine ribose moieties with partial C2'-*endo* sugar pucker. A total of 11 peaks are observed in the H1'-H2' region of the spectrum (see Figure 3.15). Three very strong peaks result from the predominantly C2'-*endo* character of the bulge nucleotides U23, C24, and U25. Two strong cross peaks were assigned to the 5'-terminal G (G16), and the tetraloop A31. Four medium resonances are observed. From these peaks, partial C2'-*endo* character was assigned to A22, located just below the bulge;

the 3' terminal nucleotide (C46); the tetraloop A32; and an unassignable peak at 4.77/5.87 ppm. It is possible that this corresponds to C39, although it was not possible to confirm this assignment with NOESY data. Two weak COSY peaks are also seen for U40 (just below the bulge), and for C38 of the A-C apposition. A summary of NMR data is provided by the schematic representation of the RNA conformation, including sugar pucker and NOE data, in Figure 3.9.

3.5 tTAR-UAC TAR + Argininamide Complex

Complex Formation and Exchangeable Assignments

The binding of argininamide was monitored by changes in the imino spectrum of the RNA. A 120 mM stock of argininamide in NMR buffer was added in small volumes. Spectra were taken at 25 °C at argininamide concentrations of 0.25, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, and 6.0 mM (Figure 3.10). The tetraloop G-imino resonance starts to disappear upon addition of as little as 0.25 mM argininamide. At 6.0 mM argininamide, the peak is completely broadened into the base line. The addition of argininamide appears to alter the conformation of the upper stem in such a way as to increase the exchange properties of the G30 imino proton. At 0.5 mM argininamide, the G26 imino begins a downfield shift, and overlaps with the G21 imino resonance until a concentration of 3.0 mM argininamide has been added. By 6.0 mM argininamide, the G26 is shifted downfield of G21. A similar shift is seen with the G28 imino proton. At 2.0 mM argininamide, G28 is shifted into the G17 resonance, and by 6.0 mM it has emerged upfield of G17. Imino resonances most affected by argininamide binding (G26 $\Delta\delta$ +0.18 ppm; G28 $\Delta\delta$ +0.19 ppm) are in base pairs flanking the A-C apposition, the mispair proposed to be involved in the formation of the U•A-C triple. Recall that in the model of the wild-type TAR-

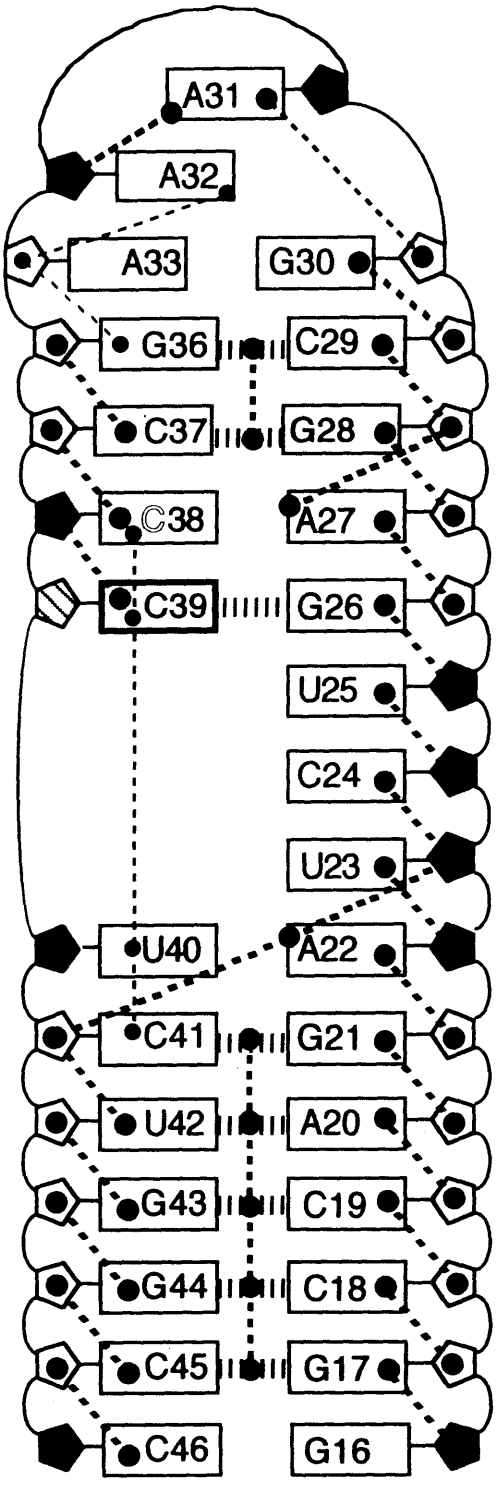


Figure 3.9) Schematic representation of tTAR-UAC NMR data.

Figure 3.9) NMR summary legend. Ribose rings are represented as pentagons, bases as rectangles. Horizontal dashed lines indicate hydrogen bonding. Thick, dashed lines represent NOEs. Shaded pentagons indicate sugars with C2'-*endo* character. The striped ribose was unassignable (C39). Bold rectangle indicates conformational flexibility of the C39 nucleotide. Hydrogen bonding is observed in the five central base-pairs in the lower stem, the apical two base-pairs of the upper stem, and in the G26-C39 base-pair just above the bulge. No hydrogen bonding was seen for the A22-U40 base-pair. The three bulge nucleotides are partially stacked between the two stems, they all have considerable C2'-*endo* composition, as evidenced by strong H1'-H2' COSY cross peaks. Extensive stacking is observed for the entire molecule. Normal A-form connectivities indicate overall A-form helical structure, except for nucleotides C38 through C41. Aromatic-aromatic NOEs (thin, vertical, dashed lines) are shown only for this region as an indication that the bases are close to each other in space; these stacking NOEs also occur elsewhere in the RNA molecule, in addition to the normal aromatic to H2' connectivities. Tetraloop nucleotides are also stacked. A weak NOE is present from G36 H8 to the A33 H1' (thin, diagonal, dashed line), but not to the H2' proton. Adenine H2 to neighboring H1' NOEs are seen for A22 to U23 and C41; A27 to G28 (but not to C39); A32 to A33; and A31 to A32.

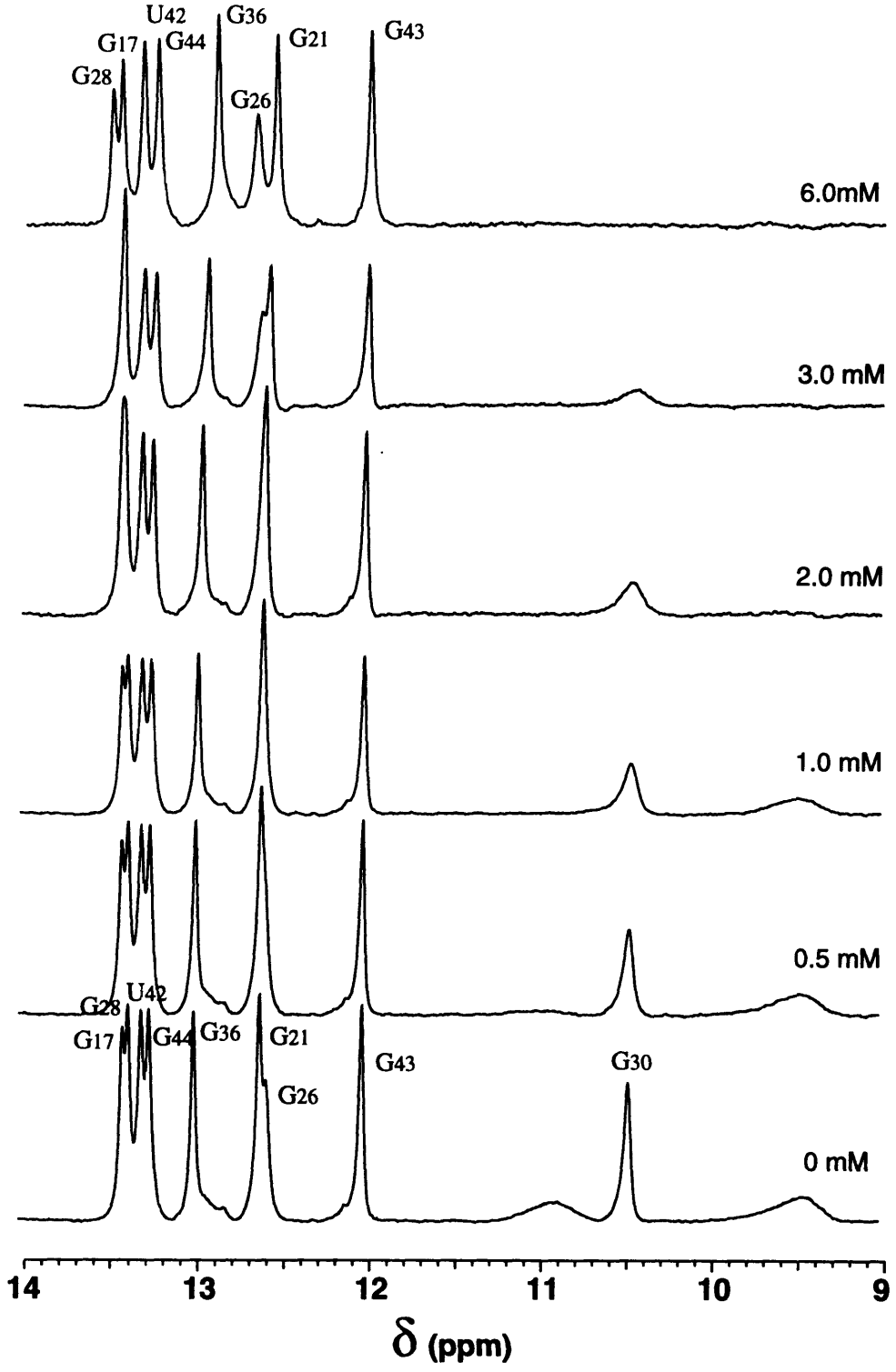


Figure 3.10) tTAR-UAC Argininamide titration at 25 °C. The effect of argininamide on the conformation of the RNA was monitored by observing changes in the imino spectrum. Addition of argininamide causes the G30 resonance to broaden into the base line. Significant chemical shift changes are observed for the imino protons G26 and G28. Both of these bases flank the A-C apposition, the nucleotides proposed to form a base-triple with U23 upon argininamide binding. The addition of argininamide causes changes in the appropriate portion of the molecule.

argininamide complex (Puglisi et al., 1992), G26 was proposed to make a direct contact with the argininamide. Changes in the 1-dimensional exchangeable proton spectrum upon argininamide binding suggest that the appropriate nucleotides are involved in formation of the mutant RNA-amino acid analog complex.

The effect of temperature on the imino resonances was also explored. Figure 3.11 shows imino spectra for tTAR-UAC + 6 mM argininamide (pH 6.4) at 25, 15, and 5 °C. Two broad resonances appear at 15 °C and sharpen at 5 °C. The resonance at 10.5 ppm is likely to be the G30 imino proton. Decreasing the temperature slows the rate of exchange of this proton with solvent. The other peak, resonating near 13.98 ppm is assigned to the U40 imino proton. This peak was absent in spectra of free tTAR-UAC (above), free wild-type TAR, and complexes of wild-type TAR with argininamide (Puglisi et al., 1992). However, an imino resonance was seen at this chemical shift in spectra of a complex of wild-type TAR + the *Tat* peptide YKKKRKKKKKA (J.D. Puglisi and J.R. Williamson, unpublished results). Changes in chemical shift were observed upon peptide binding for G36 ($\Delta\delta$ = 0.11 ppm) and for G21 ($\Delta\delta$ = 0.06 ppm) imino resonances. The U40 imino proton was also observed at low temperatures in an argininamide complex with the wild-type TAR containing a bulge of only two uridines (A.S. Brodsky & J.R. Williamson, manuscript in preparation).

As with the free RNA, a NOESY in H_2O was acquired (10 °C, 600 MHz) in order to assign the exchangeable protons. The basic procedure was as above. Imino-imino stacking NOEs are determined for the five central base pairs of the lower stem. The G16 imino is not observed due to fraying of the terminal base pair. The U40 imino resonance is quite broad and as such, no NOEs were seen to it. Cross peaks from the G-imino protons were used to assign C-aminos, and to confirm CH5s, as well as intra- and interstrand H1's (see above). Similar to the free RNA, the interesting imino resonances belong to G26 and G28. Both of these peaks shift downfield upon the addition of argininamide. Although the

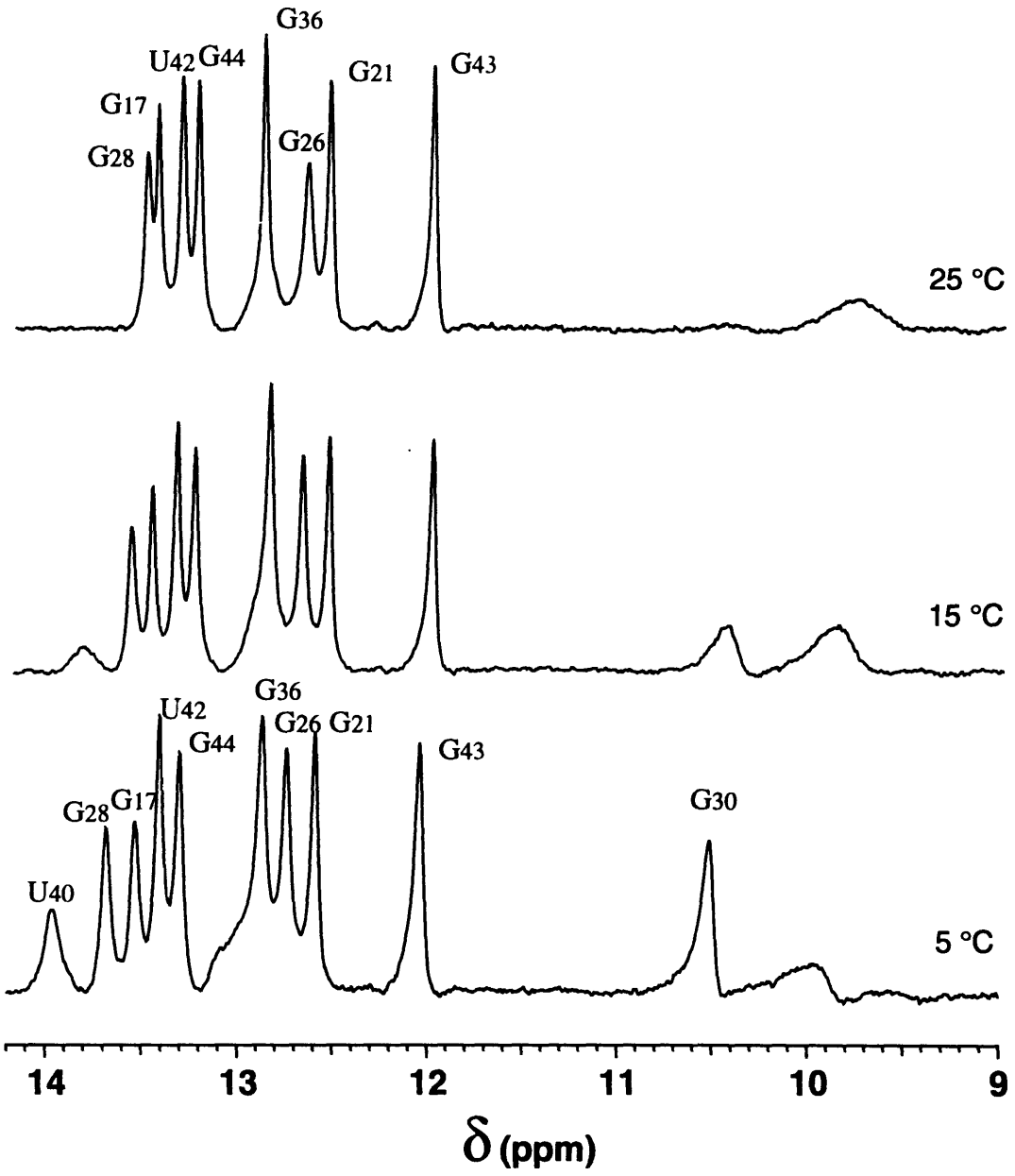


Figure 3.11) Temperature dependence of the tTAR-UAC + argininamide imino spectra. Decreasing the temperature of the argininamide-containing sample causes the appearance of two peaks. The peak at 10.6 ppm is the G30 imino proton. The broad peak at 13.95 ppm results from the argininamide-induced stabilization of the A22-U40 base-pair. This resonance was completely absent in spectra of the free RNA (Figure 3.6). The G26 imino resonance is increased by low temperature; this was not the case in the free sample.

cross peaks from the G28 imino proton are weak, with the exception of the C37 H5, all of the predicted NOEs are observed. As with the free RNA, the G26 imino proton gives an NOE to the A27 H1'. There are two differences in the NOE pattern in the argininamide complex. In contrast with the free RNA, there is no observable NOE to the C39 H5 (in neither case are there observed NOEs to the C39 aminos). The other, surprising difference is that an NOE is observed to the U40 H1' in the argininamide-containing sample. This cross strand NOE suggests that the residues spanning the bulge region maintain partial A-form character. The expected distance between the Guanine amino proton to the cross strand (n+1) H1' is 4.3 Å (Figure 3.7). An NOE would not be observed experimentally if this distance in the tTAR-UAC + argininamide complex was greater than 5 Å. For the wild-type RNA, significant chemical shift changes are only seen for the G21-C41 (0.1 ppm) and A27-U38 (+0.15 ppm) base pairs. In tTAR-UAC, there is no imino proton in the A27-C38 mispair. Large chemical shift differences between the free and bound RNA imino protons occur for the G21-C41 base pair ($\Delta\delta$ = +0.06), the G26-C39 base pair (-0.18 ppm), the G28-C37 base pair (-0.19 ppm) and the G36-C29 base pair (+0.11 ppm). Clearly, a lot of action is going on in the upper stem and bulge regions.

In direct correspondence with the free RNA, the U42 imino was the only U-imino proton assigned that gave an NOE to an AH2. The A20 H2 was assigned to 6.89 ppm. The difference in chemical shift ($\Delta\delta$) for this proton (free-bound) is 0.11 ppm (it was 0.04 ppm in the wild-type). No NOEs were seen to the broad U40 imino proton; thus, the A22 H2 was not assigned from the exchangeable data. The exchangeable proton assignments are presented with the non-exchangeable assignments in Table 3.4.

Assignment of Non-exchangeable Protons.

General Features

The sample was exchanged into D_2O for acquisition of the non-exchangeable

Table 3.4
Assignments for tTAR-UAC +6 mM Argininamide (pH 6.4)
Chemical shifts are reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino	Amino
G16	8.12	NA	5.78	4.92		
G17	7.75	NA	5.94	4.60	13.48	
C18	7.73	5.29	5.56	4.43	NA	6.93/8.66
C19	7.75	5.54	5.49	4.60	NA	6.90/8.44
A20	7.98	6.89	5.90	4.68	NA	
G21	7.09	NA	5.54	4.40	12.57	
A22	7.76	7.28 #	5.96	4.33	NA	
U23	7.49	5.50	5.83	4.33		NA
C24	7.88	6.05	5.99	4.41	NA	
U25	7.90	5.90	6.03	4.45		NA
G26	7.84	NA	5.93	4.93	12.78	
A27	8.12	8.11	6.18	4.85	NA	
G28	7.32	NA	5.69	4.46	13.54	
C29	7.38	5.08	5.58	4.46	NA	6.60/8.21
G30	7.50	NA	5.74	4.43	10.6 *	
A31	8.37		5.69	4.74	NA	
A32	7.95	7.75	5.43	4.32	NA	
A33	8.16	7.78	6.01	4.56	NA	
G36	7.88	NA	3.64	4.28	12.89	
C37	7.56	5.14	5.52	4.61	NA	6.92/8.58
C38	8.05	5.78	5.70	4.26	NA	
C39	7.64	5.62	5.44	4.37	NA	7.26/7.98
U40	7.99	5.56	5.50	4.41	13.95 *	NA
C41	7.94	5.72	5.51	4.26	NA	7.07/8.35
U42	7.86	5.33	5.47	4.57	13.35	NA
G43	7.69	NA	5.77	4.60	12.01	
G44	7.35	NA	5.70	4.46	13.26	
C45	7.68	5.17	5.53	4.30	NA	6.94/8.67
C46	7.69	5.55	5.77	4.02	NA	

NA refers to not applicable protons, blank spaces indicate protons that were not assigned, * denotes 5 °C data, # represents a tentative assignment.

spectra. 300 and 40 ms mixing time NOESYs were acquired on FBNML Varian 750 MHz spectrometer. The basic assignment procedure was as outlined above. Again, the A20 H2 and tetraloop G36 H1' protons served as starting points for assignment. Assignments were determined from the aromatic H6/H8 to H1'/H5 portion of the 300 ms NOESY shown in Figure 3.12. Chemical shift assignments (Table 3.4) for aromatic, H2/H5, H1' and H2' protons were obtained for the entire tTAR-UAC RNA/ argininamide complex. Aromatic to H2' (n-1) NOEs indicative of A-form helical geometry are observed for all upper and lower stem nucleotides. Not surprisingly, stacking NOEs are not detected between the tetraloop nucleotides A31 and G30. The A22 H2 is shifted 0.36 ppm upfield. A *very* weak cross strand NOE is observed from this proton to C41 H1', no NOE is seen to any other H1' proton. Another difference between the free and bound spectra is the intensity of the G28 cross peaks. The G28 intranucleotide H8-H1' resonance is shifted 0.08 ppm downfield, and gives a weak NOE to C29 H1' (not observable at the level shown in Figure 3.12). The U23 and C38 H5-H6 cross peaks are broad in the argininamide spectra.

Interestingly, a large chemical shift difference is observed for the H2 proton of the tetraloop nucleotide A33. This resonance is shifted 0.32 ppm upfield in the argininamide sample. Also, NOEs to the A31 H2 proton are not detected in this spectrum. Recall that the G30 imino resonance broadens upon addition of argininamide (Figure 3.10). Argininamide-induced conformational changes appear to occur for the tetraloop nucleotides. No argininamide NOEs were detected to any tetraloop nucleotides. Therefore, the binding of argininamide to nucleotides in the upper stem may indirectly affect the loop structure. Minor changes in the wild-type loop conformation were also observed in TAR upon binding of the *Tat* peptide YKKKRKKKKKA (J.D. Puglisi and J.R. Williamson, unpublished results).

Bulge Region

Dramatic changes in chemical shifts of the bulge nucleotides were observed with the

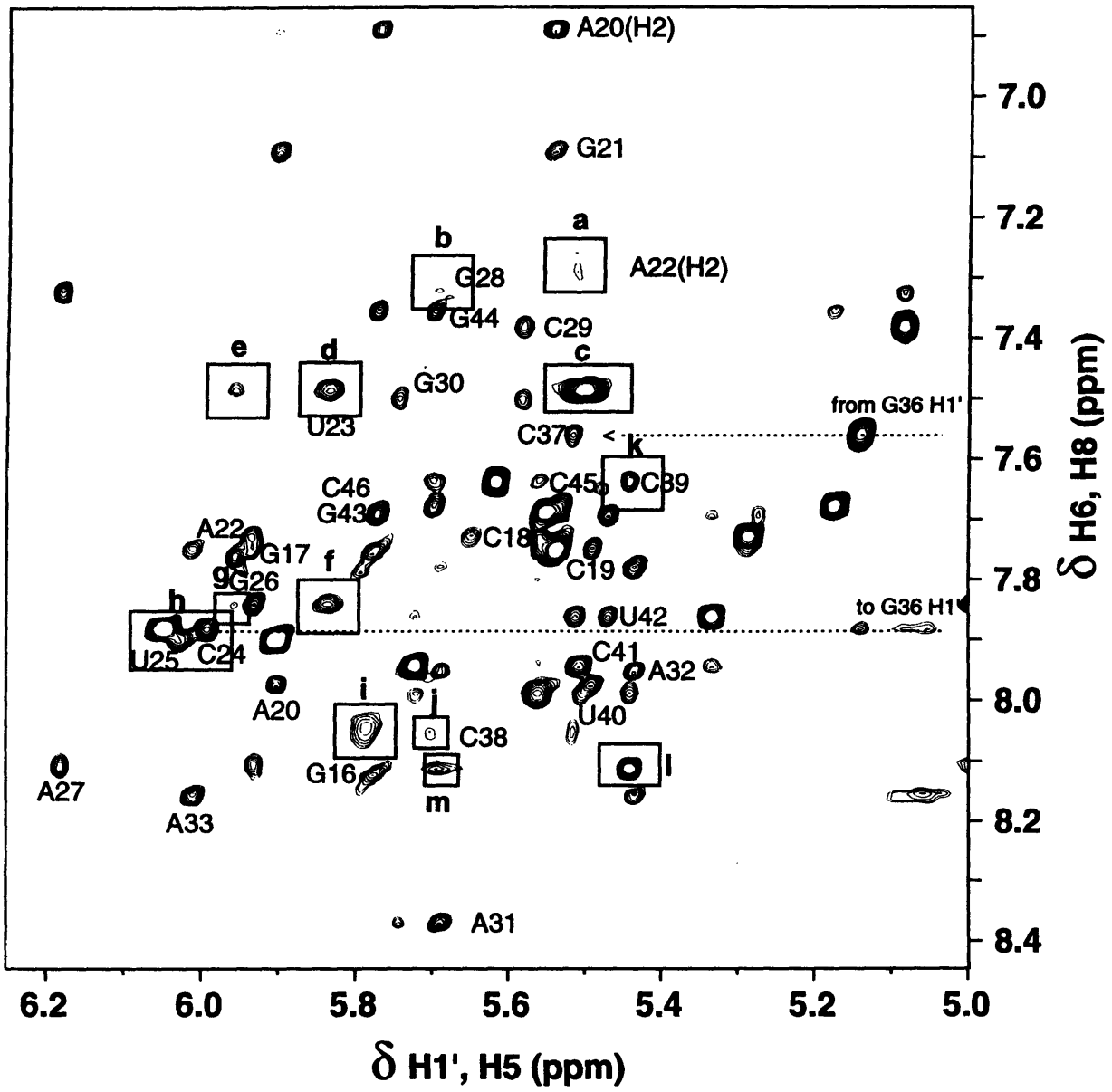


Figure 3.12) tTAR-UAC + argininamide NOESY. Aromatic to H1' connectivities were determined for the entire molecule with the exception of nucleotides C24 and U25 (box **h**). Box **a** encloses the potential crosspeak resulting from the NOE from A22 H2 to C41 H1'. The weak intranucleotide H6-H1' NOE for G28 is shown in **b**. Boxes **c**, **d**,and **e** contain NOEs from U23 H6 to U23 H5, G26 H1' and A22 H1' , respectively. The intranucleotide H8-H1' NOE for G26 is shown in **f**. A very weak NOE is seen between G26 H8 and A22 H1' (**g**). Boxes **i** and **j** contain intranucleotide NOEs from C38 H6 to the H5 and H1' protons, respectively. C39 is assignable in this spectrum (**k**). The A27 H2 NOEs to C39 H1' and G28 H1' (weak) are shown in **l** and **m**, respectively. Conditions were as in Figure 3.8.

wild-type RNA upon argininamide binding (Puglisi et al., 1992). All of the bulge nucleotides, including U23, become completely unstacked. No internucleotide NOEs were detected between U23 and A22. NOEs were seen between U23 H1' and G26 H8, G26 H3' and G26 H5'/H5". Based on this NOE pattern, Puglisi et al. (Puglisi et al., 1992) proposed that the U23 was positioned in the major groove of the upper stem, near A27. In their model, weak NOEs detected between A22 H2 and G26 H1, and between A22 H2' and G26 H8, suggested coaxial stacking of the two stem regions.

In the tTAR-UAC molecule, a lack of NOEs to the bulge nucleotides U24 and C25 suggest that they are no longer stacked between the upper and lower stems. H2' assignments for these nucleotides were made on the basis of COSY H1'-H2' data (see below). Dramatic chemical shift changes, indicative of a significant alteration in local environment, are observed for these nucleotides as a result of argininamide binding. In contrast to the wild-type TAR, however, the tTAR-UAC U23 appears to be partially stacked into the helix. The H5-H6 resonance is slightly broadened, indicating some conformational exchange. A strong NOE is seen between U23 H6 and G26 H1'. However, a weak cross peak is also observed to the A22 H1' (Figure 3.12). NOEs are also seen from U23 H6 to G26 *and* A22 H8s, with G26 giving the stronger NOE. No aromatic to H2' resonances are detected for G26 H8, either to U25, U23, or to A22. Similar to wild-type NMR data, a strong NOE is observed from U23 H1' to G26 H3', locating this nucleotide in the major groove of the upper stem. Identification of this NOE in homonuclear experiments was only possible because the G26 H3' proton resonates at 4.99 ppm, downfield of most other ribose protons (Figure 3.13). COSY and TOCSY spectra allowed confirmation of the G26 H3' assignment. The NOEs between U23 H1' and G26 H5'/H5" were not established, as the G26 H5' and H5" protons were not assigned in tTAR-UAC. A weak NOE is also observed between U23 H5 and A27 H1', an indication that U23 is close to the upper stem at least part of the time. Unlike the wild-type RNA,

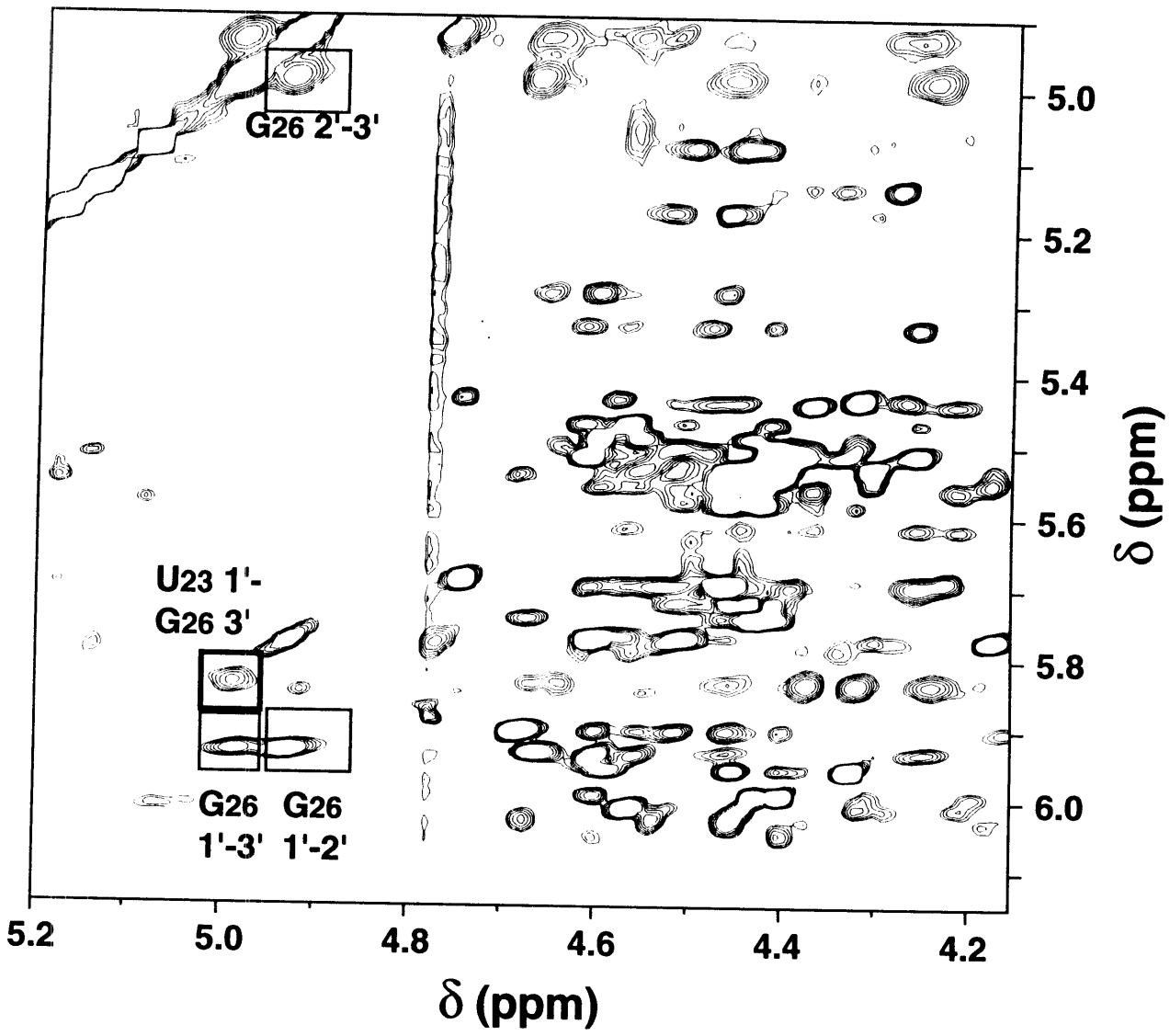


Figure 3.13) Downfield-shifted G26 protons. The identification of the NOE from G26 H3' to U23 H1' (bold box) was possible because G26 H2' and H3' protons (normal boxes) resonate downfield from most other ribose protons.

however, no NOE is observed between A22 H2 and G26 H1' (an NOE is barely seen to C41 H1'), suggesting a greater degree of distortion at the junction of the stacked upper and lower helices.

The strand opposite from the bulge has undergone argininamide-induced structural changes as well. An NOE is observed between U40 H6 and C39 H2'. This indication of A-type helical conformation was absent in the free tTAR-UAC RNA. C39 is better behaved in the argininamide complex. The H6, H5, H1' and H2' protons are assigned for this nucleotide. In addition to the U40 H6 to C39 H2' and H1' NOEs, cross peaks are observed for H6-H6 and H5-H5 interactions. An NOE from C39 H6 to U40 H5 is also observed. In correspondence with the wild-type RNA, the two stems are coaxially stacked. U23 may be stacked in-between the helices a small part of the time, but the presence of NOEs from U23 to G26 ribose protons (discussed above) positions the nucleotide near the upper stem a significant part of the time. NOEs determined for the bulge region are listed in Table 3.5 (compare with Table 3.3).

A-C Mispair

As mentioned above, the G28 resonances are broad in the argininamide spectra, suggesting conformational exchange in the region of the A C apposition. Nevertheless, most of the expected A-form NOEs are observed in this region. The exception is a lack of NOE between C29 H6 to G28 H2'. Observed NOEs include G28 H6 and H1' to C29 H5, and C28 H6 to C29 H6. A27 H2 gives the expected cross peak to C39 H1' and a weak NOE to G28 H1'.

The differences in chemical shifts between the free RNA and the argininamide complex are listed in Table 3.6, and are depicted graphically in Figure 3.14. The pattern of chemical shift differences between the free and argininamide samples of tTAR-UAC is very similar to that observed for the wild-type RNA (Puglisi et al., 1992). Base stacking and hydrogen bonding affect the chemical shift of a proton. Aromatic protons within the

Table 3.5
Internucleotide NOEs in the bulge region of tTAR-UAC + 6mM argininamide (pH 6.4).

N	N+1		N	N+1
A22 H8	U23 H6 (w)		C39 H6	U40 H6
A22 H1'	U23 H6 (w)		C39 H6	U40 H5
U23 H6	G26 H8		C39 H6	U40 H1'
U23 H6	G26 H1'		C39 H5	U40 H5
U23 H1'	G26 H3'		C39 H2'	U40 H6
U23 H1'	A27 H1'(w)		C39 H1'	U40 H5

(w) denotes a weak NOE

Table 3.6
Chemical shift differences for tTAR-UAC, Free data -Argininamide data (pH 6.4).

Δδ reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino
G16	+0.04	NA	+0.04	+0.03	
G17	-0.08	NA	-0.01	-0.01	-0.03
C18	-0.02	-0.01	+0.01	+0.02	NA
C19	-0.03	+0.03	0.0	-0.05	NA
A20	+0.04	+0.11	+0.01	+0.02	NA
G21	-0.02	NA	-0.04	+0.07	+0.06
A22	-0.16	+0.36	-0.06	+0.06	NA
U23	-0.07	-0.18	-0.24	-0.20	
C24	-0.13	-0.19	-0.18	-0.11	NA
U25	-0.05	-0.08	-0.18	+0.02	
G26	+0.08	NA	-0.06	-0.42	-0.18
A27	-0.02	-0.04	0.0	-0.42	NA
G28	-0.08	NA	-0.05	+0.36	-0.19
C29	-0.01	+0.04	0.0	+0.31	NA
G30	+0.06	NA	-0.01	+0.04	-0.06
A31	+0.03		+0.03	+0.04	NA
A32	+0.02	+0.06	+0.02	+0.02	NA
A33	+0.05	+0.32	+0.01	+0.08	NA
G36	-0.03	NA	0.0	+0.01	+0.11
C37	-0.06	+0.05	-0.01	-0.02	NA
C38	-0.07	-0.01	-0.03	+0.16	NA
C39	+0.12	+0.03			NA
U40	-0.10	+0.12	+0.15	-0.11	
C41	+0.04	+0.15	+0.03	+0.12	NA
U42	+0.01	+0.11	+0.04	+0.06	-0.03
G43	+0.03	NA	+0.02	+0.01	+0.01
G44	-0.02	NA	-0.01	+0.03	+0.01
C45	-0.04	+0.03	0.0	-0.01	NA
C46	0.0	-0.01	+0.02	+0.01	NA

NA refers to not applicable protons, blank spaces indicate protons that were not assigned for one or both data sets.

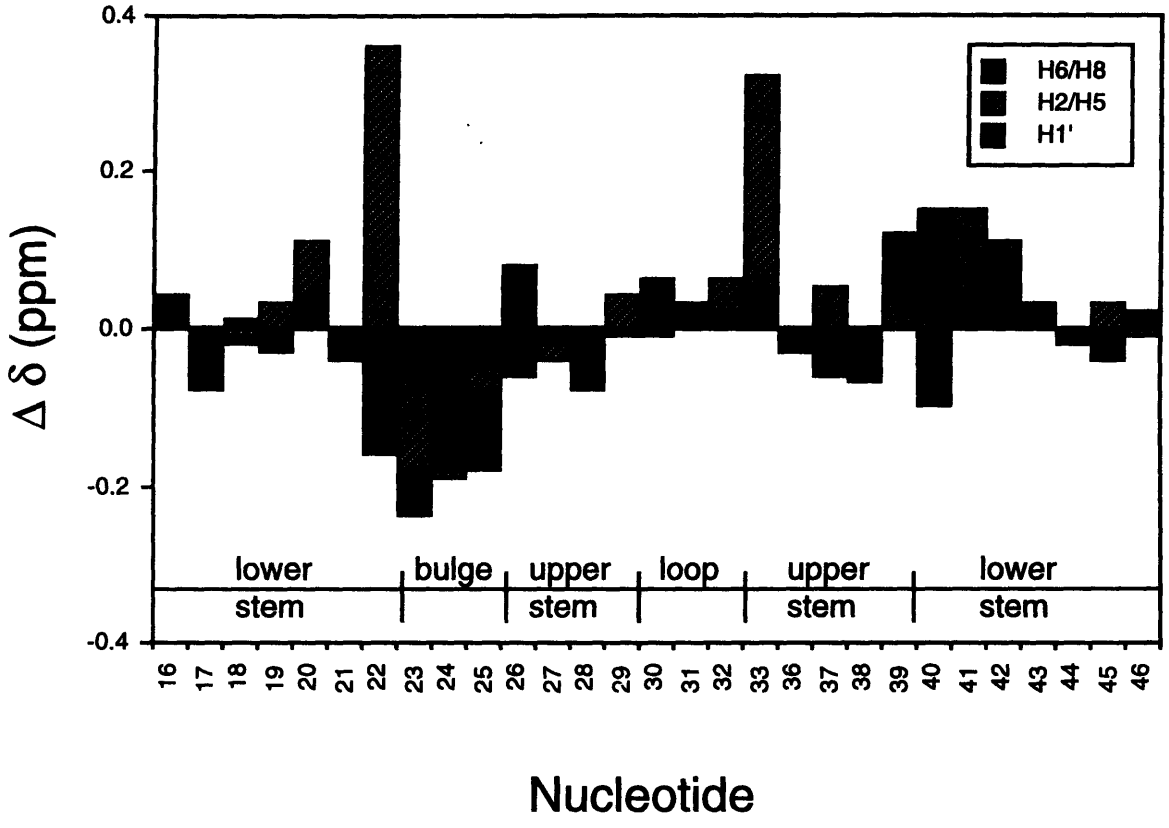


Figure 3.14) Plot of changes in chemical shift for tTAR-UAC resonances upon addition of 6 mM argininamide (pH 6.4). Nucleotide position, stem, bulge, and loop regions are marked. The H8/H6 proton chemical shift changes are indicated by solid bars, H2/H5 protons by hatched bars, and H1' protons by shaded bars. Large changes in chemical shift are observed for the bulge nucleotides as a result of argininamide binding.

context of a helix are shifted upfield relative to the single stranded state. The binding of argininamide causes a large upfield shift (+0.36 ppm) for the A22 H2 resonance, located just below the bulge. This trend is consistent with coaxial stacking of the two stems in TAR RNA, resulting in a more helical environment for the A22 base. A similar change in chemical shift (+0.32 ppm) is observed for A33 H2, the last nucleotide in the tetraloop. The geometry of the unusual tetraloop G-A base pair is altered by addition of argininamide to the RNA sample, although evidence for a specific interaction of the loop nucleotides with argininamide was not observed. Conversely, the H5 protons of the bulge nucleotides all experience large downfield changes in chemical shift; an indication that they are less stacked in the bound conformation of the RNA.

DQF COSY Spectra

Thirteen pyrimidines gave strong H5-H6 cross peaks in DQF-COSY spectra. In agreement with a broad appearance in the NOESY spectrum (Figure 3.12), the C38 H5-H6 resonance is very weak in the COSY data. Because the H5 and H6 protons are in a rigid conformation imposed by the 5-6 double bond, broadening must be a result of conformational exchange of the entire nucleotide.

A comparison of the H1'-H2' region from COSY spectra of the free and bound RNA is shown in Figure 3.15 . Surprisingly, no H1'-H2' cross peak is observed for C38, indicating minimal C2'-*endo* conformation for this nucleotide. C24 and U25 cross peaks are more intense in the argininamide spectrum, consistent with a greater degree of C2'-*endo* character resulting from unstacking. Conversely, the U23 resonance is significantly weaker in the argininamide spectrum, indicating a more structured environment in the bound conformation. A22 also appears to have slightly more C3'-endo conformation in the argininamide complex. A summary of NOE data is presented schematically in Figure 3.16.

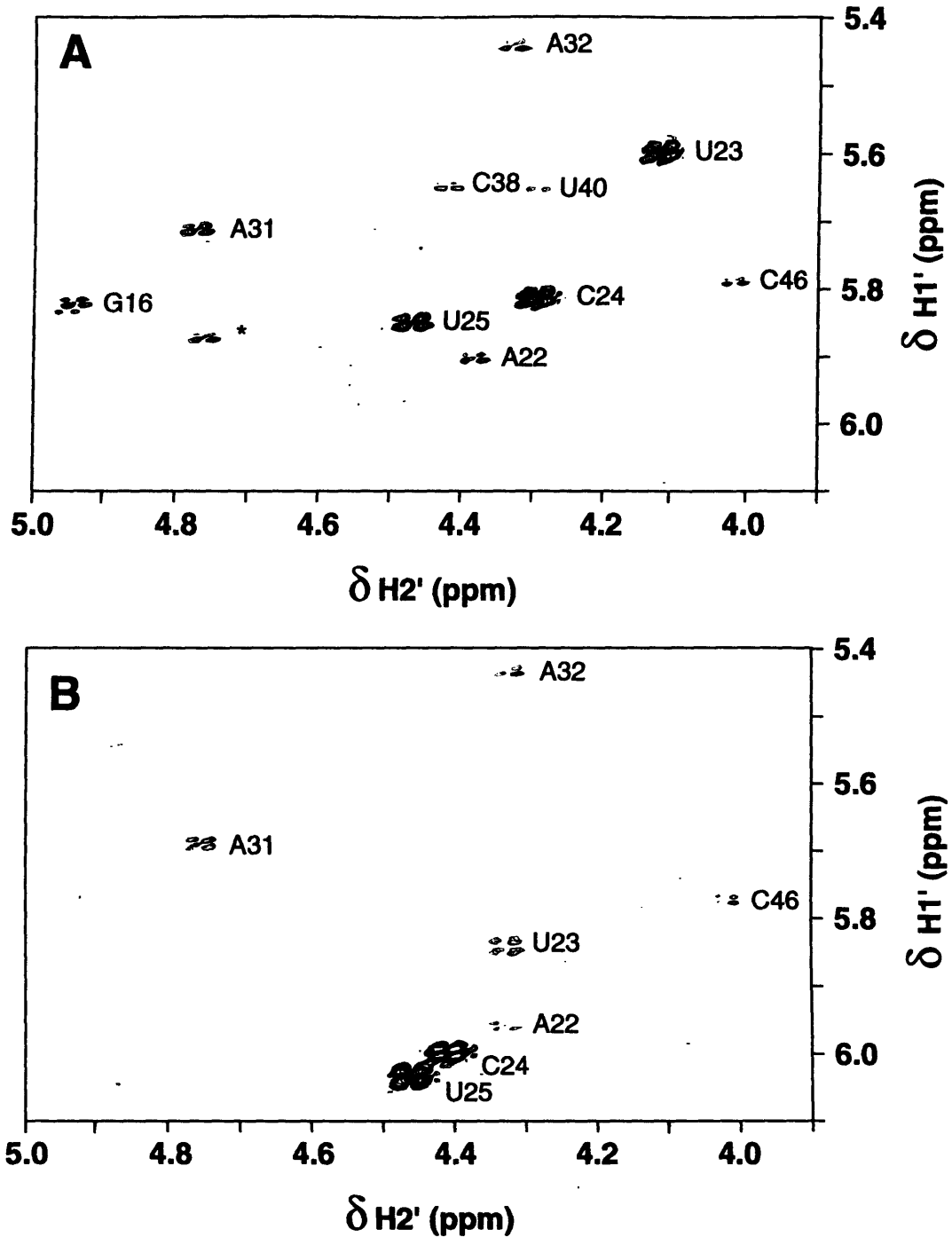


Figure 3.15) H1'-H2' regions of DQF-COSYs: A) tTAR-UAC free, B) tTAR-UAC +6mM argininamide. Comparison of the spectra demonstrates the stabilization of A-form helical conformation resulting from addition of argininamide. Fewer nucleotides exhibit C2'-*endo* composition in the argininamide complex. U23 has significantly less C2'-*endo* character, while the other bulge nucleotides (C24 and U25) have more. Large chemical shift changes are apparent for the bulge nucleotides as a result of argininamide binding. The cross peak labeled with an asterisk (*) may be C39.

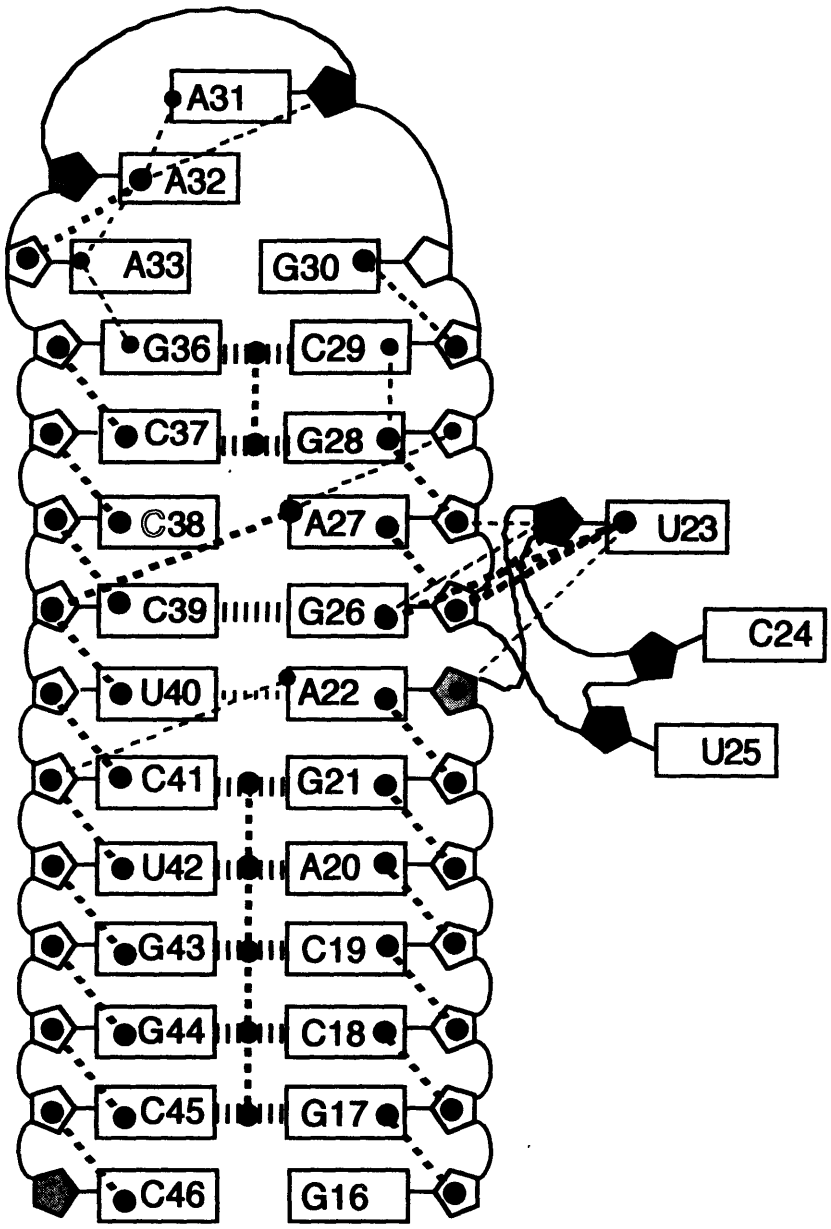


Figure 3.16) NMR summary for tTAR-UAC + 6 mM argininamide. Symbols are as in Figure 3.9. A-form helical character is seen for a majority of the molecule. Degree of C2'-endo character is indicated by relative intensities of shading. Bulge nucleotides are no longer stacked between the two stems as was seen in the free RNA. Based on NOE data between C39 and U40, the upper and lower stems are coaxially stacked. NOEs from U23 H1' to G26 H3' locate this nucleotide in the major groove of the upper stem, as seen in the wild-type TAR RNA-argininamide model. However, this nucleotide is stacked into the helix a small part of the time, indicated by the thin, diagonal dashed line between the U23 base and the A22 ribose. Characteristic A-form NOEs were not seen between G26 and A22, as would be expected if these nucleotides were stacked in a true A-form arrangement. The thin, horizontal dashed line between U40 and A22 represents hydrogen bonding detected by the presence of the U40 imino proton at low temperatures.

NOEs to Argininamide

In addition to causing large structural changes in the bulge region, argininamide was shown to give NOEs directly to important TAR RNA nucleotides. Puglisi et al. (Puglisi et al., 1992) demonstrated that the argininamide δ proton gave intermolecular NOEs to A22 H8, A22 H1', U23 H6, U23 H5, and A27 H8. The NOE to U23 H5 was the strongest. Similar NOEs are seen to argininamide in the tTAR-UAC mutant (Figure 3.17). In a long mixing time experiment, the argininamide δ proton gives NOEs to U23 H1', H5, and H6; to A22 H8 and to A27 H8, but not to A22 H1'. The γ proton demonstrated strong NOEs to U23 H5 and to A22 H8. The argininamide β proton also gives NOEs to U23 H5. Additional NOEs are observed from this proton to G28 H8, and C29 H5. In many ways, the mutant TAR resembles the wild-type RNA. Clearly, the argininamide is recognizing the RNA in the bulge region. However, the lack of an NOE from the argininamide δ proton to A22 H1' suggests local distortion in the RNA due to the presence of the A-C apposition in the upper stem. The presence of NOEs to the C29 H5 and G28 H8 protons also suggest the existence of a secondary argininamide binding site. These protons are located 10 to 12 Å away from the argininamide β proton in the detailed model of TAR containing a dinucleotide bulge (A.S. Brodsky and J.R. Williamson, manuscript in preparation).

A schematic representation of the wild-type model was presented in Figure 1.11. This model suggests that the functionally important moieties (U23, G26-C39, and A27-U38, P22 and P23) are brought together by the formation of the base-triple. The compact structure of the bound conformation of the RNA allows recognition of the critical nucleotides by a single arginine residue, be it as the free amino acid, in peptides or in the context of the *Tat* protein. The mutant tTAR-UAC studies illustrated above suggest that an RNA in which U38 has been mutated to a cytosine residue can mimic many facets of the wild-type structure. The conformation of nucleotides in the region of the A-C apposition

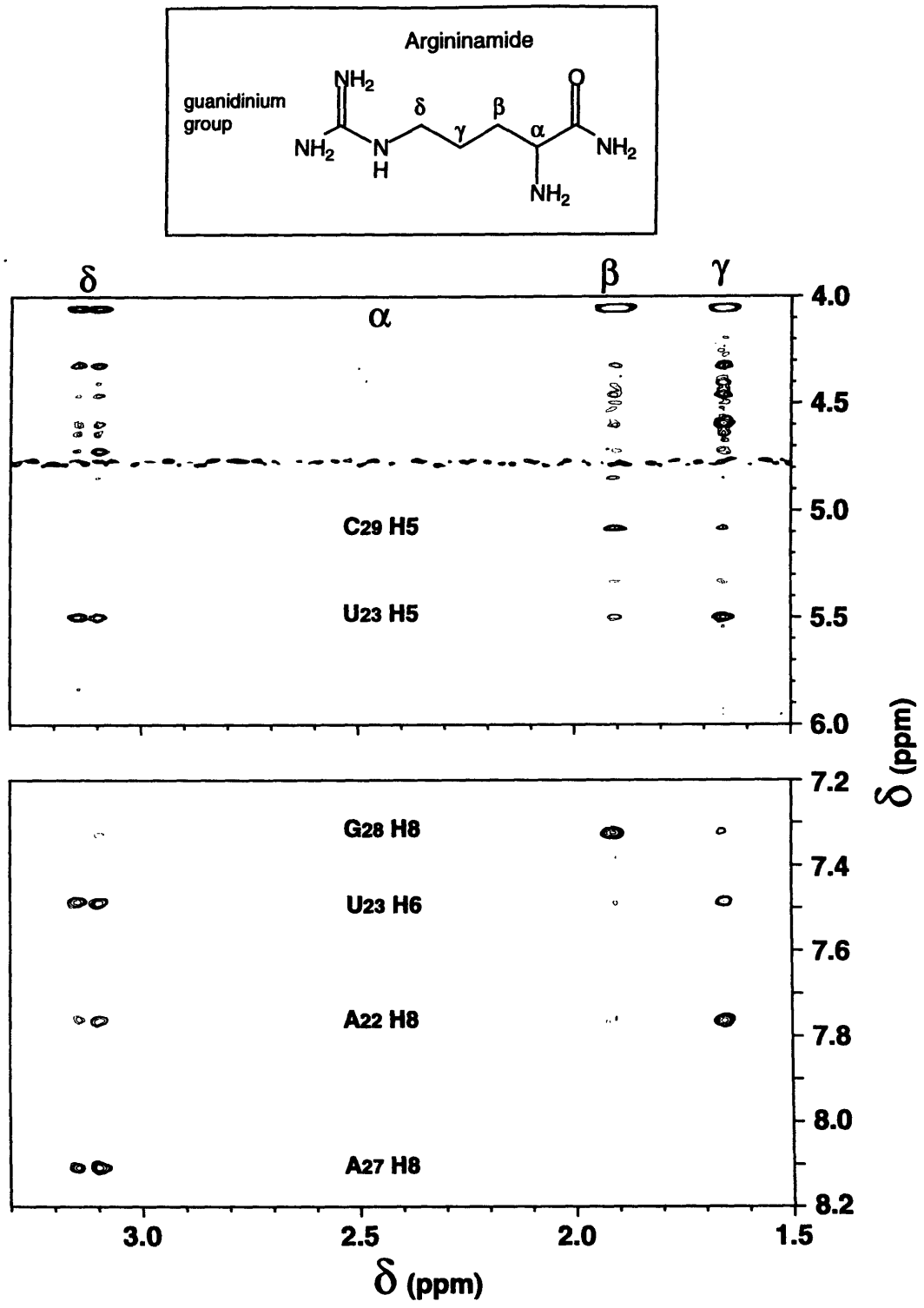


Figure 3.17) NOEs from tTAR-UAC to argininamide protons. Strong NOEs were observed from the argininamide δ proton to aromatic protons on U23, A22, and A27. Formation of the base-triple interaction brings these nucleotides close together in space. The argininamide β and γ protons also give NOEs to RNA protons, the strongest NOEs from the β proton are to C29 H5 and G28 H8, while the γ proton gives strong NOEs to both U23 H5 and A22 H8. NOESY data were recorded at 25 °C, pH 6.4 (750 MHz). The mixing time for this experiment was 300 ms.

appears to be dynamic, and at pH 6.4 the argininamide binding site may be distorted. The stabilization of A-C mispairs at low pH coupled with the improved *Tat* peptide binding to this RNA at pH 5.5, suggests that at low pH the conformation of this mutant RNA should be even more similar to the wild-type structure.

4 Low pH Characterization of tTAR-UAC

The effect of pH on the structure of an A-C mispair (Figure 3.1), and on the binding of *Tat* peptides to mutant TAR RNAs (Table 3.1), warranted an investigation of the pH dependence of the solution conformation of tTAR-UAC. Differences in the imino proton spectra of the RNAs in both the free and argininamide-containing samples were examined as a preliminary indication of global differences in RNA structure resulting from pH and temperature changes.

4.1 pH Titrations

Titrations were monitored on a Varian 500 MHz instrument. Small aliquots of 0.1 M NaOH or 0.1 M HCl were added to the sample, the pH was measured upon each addition. Measurement of the sample directly in NMR tube serves to minimize sample loss (and it ensures that the electrode has only been used with nuclease-free samples). NMR spectra were taken at 0.5 pH unit increments, temperature points were only taken at intermediate pH (Figures 3.6 and 3.11) and low pH (see below).

The line widths of imino resonances depend on the contributions of dipolar coupling and exchange (Chapter 2). If exchange with solvent is significant (as in terminal base pairs), the resonances should broaden with decreasing stability of the base pair, i.e. with increasing temperature. If this exchange is limited only by the opening of the base pair, the line width should depend on temperature, but not buffer concentration (Teitelbaum & Englander, 1975, Benight et al., 1988, Leroy et al., 1988) or pH (i.e. concentration of

base). However, if the exchange is base-catalyzed or the structure of the RNA is pH dependent, the exchange rate (and therefore line width) of imino resonances may depend on the pH as well as temperature.

The pH titration data for the free tTAR-UAC sample are shown in Figure 4.1. Raising the pH of free RNA sample to 7.8 does not dramatically affect the imino resonances of the G-C base pairs in the lower stem (G17, G44, G43, G21) or the apical G-C base pair (G36). pH does, however, affect peaks from imino protons in the base pairs flanking the A-C (G26, G28). Both peaks decrease significantly with increasing pH. The G26 resonance disappears entirely by pH 7.8. The G28 resonance decreases until pH 7.0, but appears to sharpen again at pH 7.8. The structure of this base pair depends on the conformation of the neighboring base pairs. An alteration in the structure of the A-C mispair at high pH may cause a different, potentially more stable, environment for this proton. The U42 resonance is broad at pH 7.8, however the neighboring base pairs do not seem to be affected by high pH. A more detailed analysis of the local environment of this base pair would be necessary to understand this behavior.

The G30 imino resonance is strongest at neutral pHs. As pointed out in Chapter 3, this nucleotide is involved in a unique G-A pair (Heus & Pardi, 1991a). The proposed structure of this pair is shown in Figure 4.2. At low pH, the stability of this base pair may be reduced by protonation of the N^1 of A33 , or any of the other surrounding adenines. The decrease in stability of this resonance coincides with the appearance of a broad peak at ca. 9.8 ppm. As discussed below, this is likely to be a resonance from an adenine amino group. The broad resonance at ca. 11 ppm is most likely from the non-hydrogen bonded bulge iminos. These resonances were also seen in the spectra acquired at pH 6.4 and low temperatures (Figure 3.6).

A similar pH titration was performed for the sample containing 6 mM argininamide. The spectra at various pH points are shown in Figure 4.3. As with the free RNA, the

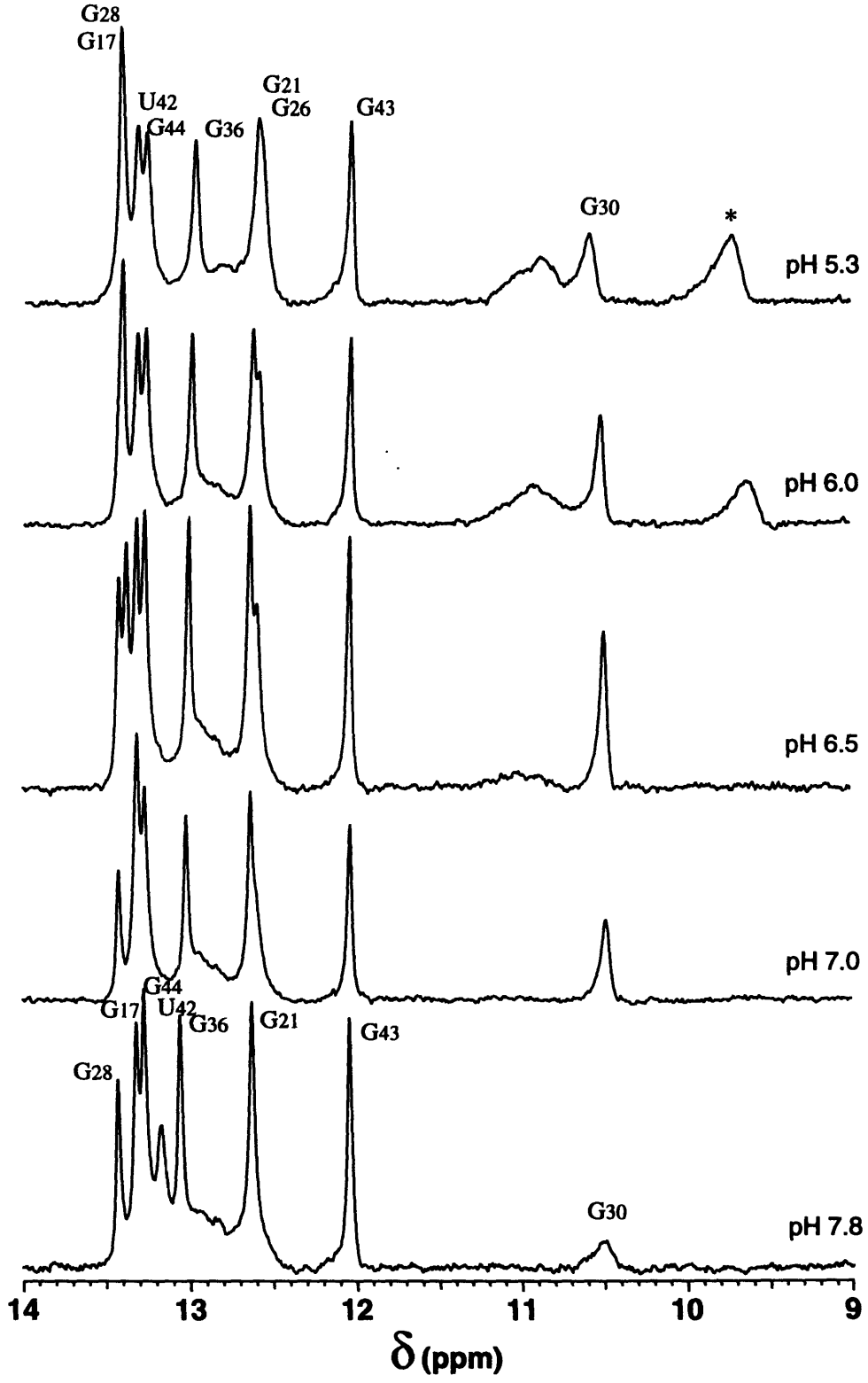


Figure 4.1) pH titration of tTAR-UAC at 25 °C (free). Increasing the pH causes imino resonances from G26, G28, and U42 to broaden. The tetraloop resonance (G30) is most stable at the intermediate pH (6.5). The broad peak at 11 ppm (low pHs) is due to non-hydrogen bonded bulge imino protons. The broad peak at ca. 9.8 ppm (*) in low pH spectra is from a shifted amino resonance.

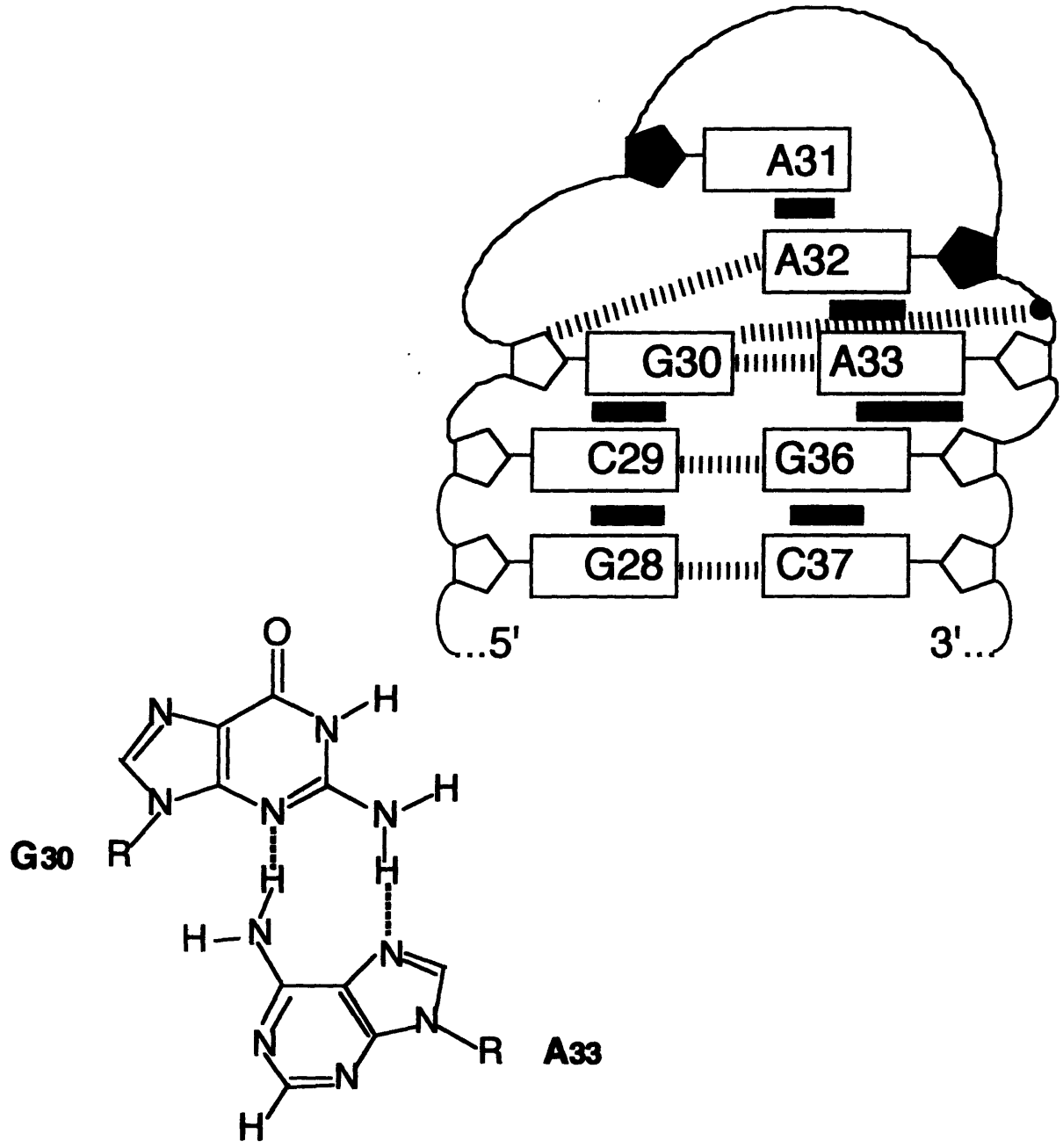


Figure 4.2) Unique G-A pair of the tetraloop GA_3. The G30-A33 pair in tTAR-UAC does not involve the G30 imino proton. The compact structure of the tetraloop protects this proton from exchange with solvent. The proposed hydrogen bonding of this pair involves the G30 and A33 *amino* protons. Temperature or pH effects on the amino groups of these nucleotides (or those of the neighboring adenines) may alter the slow exchange properties of the guanine imino group.

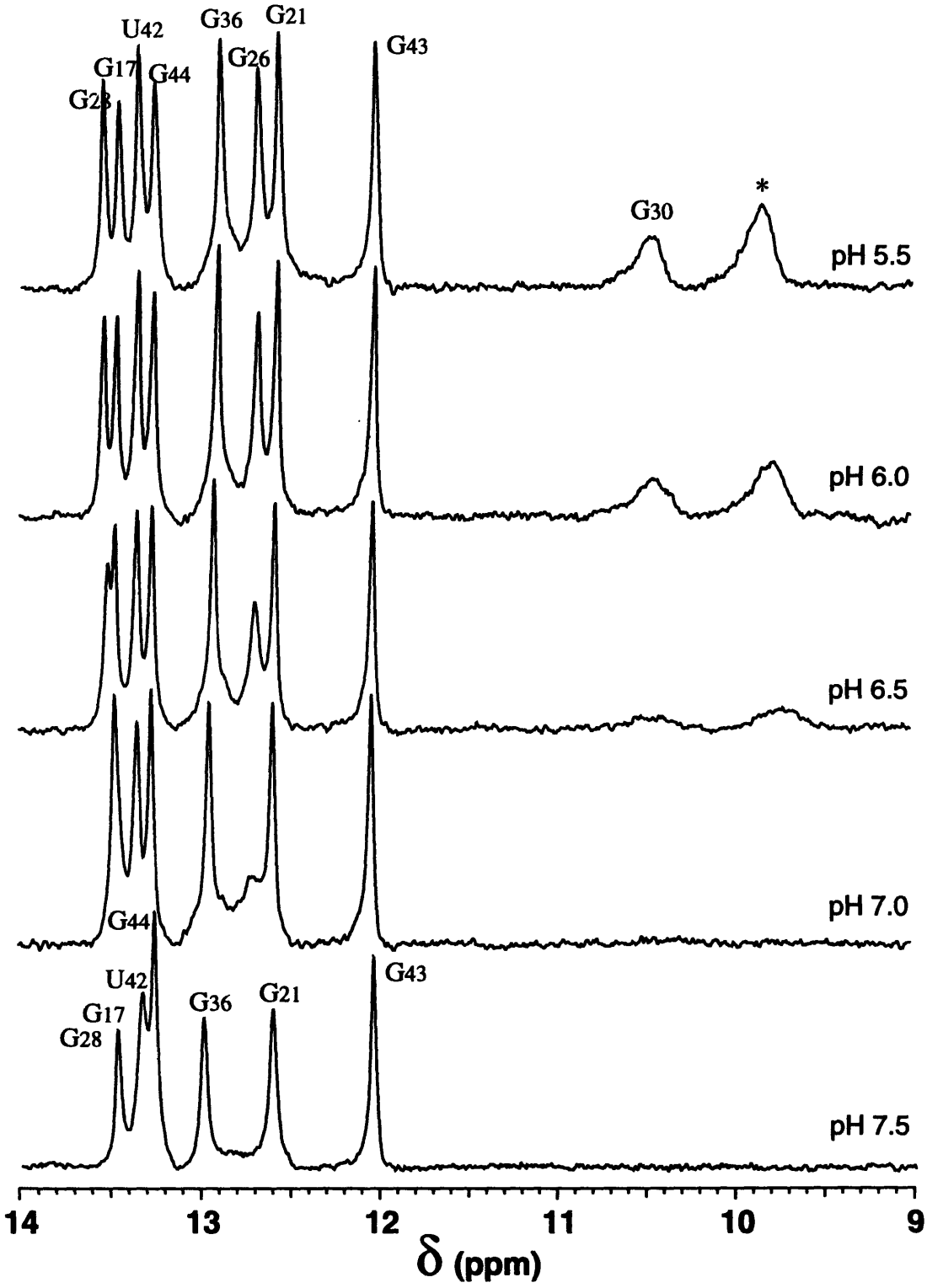


Figure 4.3) pH titration of tTAR-UAC + argininamide at 25 °C. Base-pairs in the upper stem are destabilized by increasing the pH of the sample. At high pH, the resonance from the G26 imino proton is completely absent. The broad peak at ca. 9.8 ppm (*) is due to a downfield shifted adenine amino proton.

exchange properties of the G-C pairs in the lower stem are not affected by the concentration of base. They appear to be well protected from exchange with solvent. The strength of the U42 imino resonance decreases with increasing pH, although not as dramatically as in the spectra of the free RNA. The stability of the G26 imino proton is also sensitive to the pH of the sample. At pH 6, this resonance begins to weaken, by pH 7.5, it is broadened into the base line. As with the free RNA, the G28 peak shifts downfield at increasing pH. It either disappears completely or is overlapped with the G17 resonance. Clearly, the pH of the sample affects the base pairs near the A-C apposition. This suggests that the exchange properties of these imino protons are not simply dependent on the stability of the base pairs, but also on the pH of the surrounding medium.

The tetraloop stability shows a pH dependence in the argininamide spectra as well as in the free sample. The G36 peak (the base pair just below the tetraloop) is slightly broadened at pH 7.5, due to changes in the tetraloop conformation. In the argininamide sample, the imino proton from the tetraloop guanine is broad even at a pH as low as 5.5. The disappearance of this resonance was observed upon addition of argininamide at pH. 6.4 (Figure 3.10). In agreement with the free sample, a broad peak at ca. 9.8 ppm is apparent in the argininamide sample at low pH. Based on the effect of temperature on the exchange kinetics of this proton (see below) it was determined that this peak corresponds to an amino proton.

The temperature dependence of the imino spectrum of the free RNA, pH 5.3, is shown in Figure 4.4. At this pH, the helical imino peaks appear to be relatively stable at 35 °C, and up to 45 °C (data not shown). Upon raising the temperature, the intensity of the G44 peak decreases slightly. The peak corresponding to the G21 and G26 resonances is the first to broaden . Thus, the base pairs flanking the A-C apposition are the first to weaken, or 'melt-out'. At 35 °C the tetraloop G36 has broadened, as have the non-hydrogen bonded iminos. Higher temperatures increase the exchange rate of the tetraloop

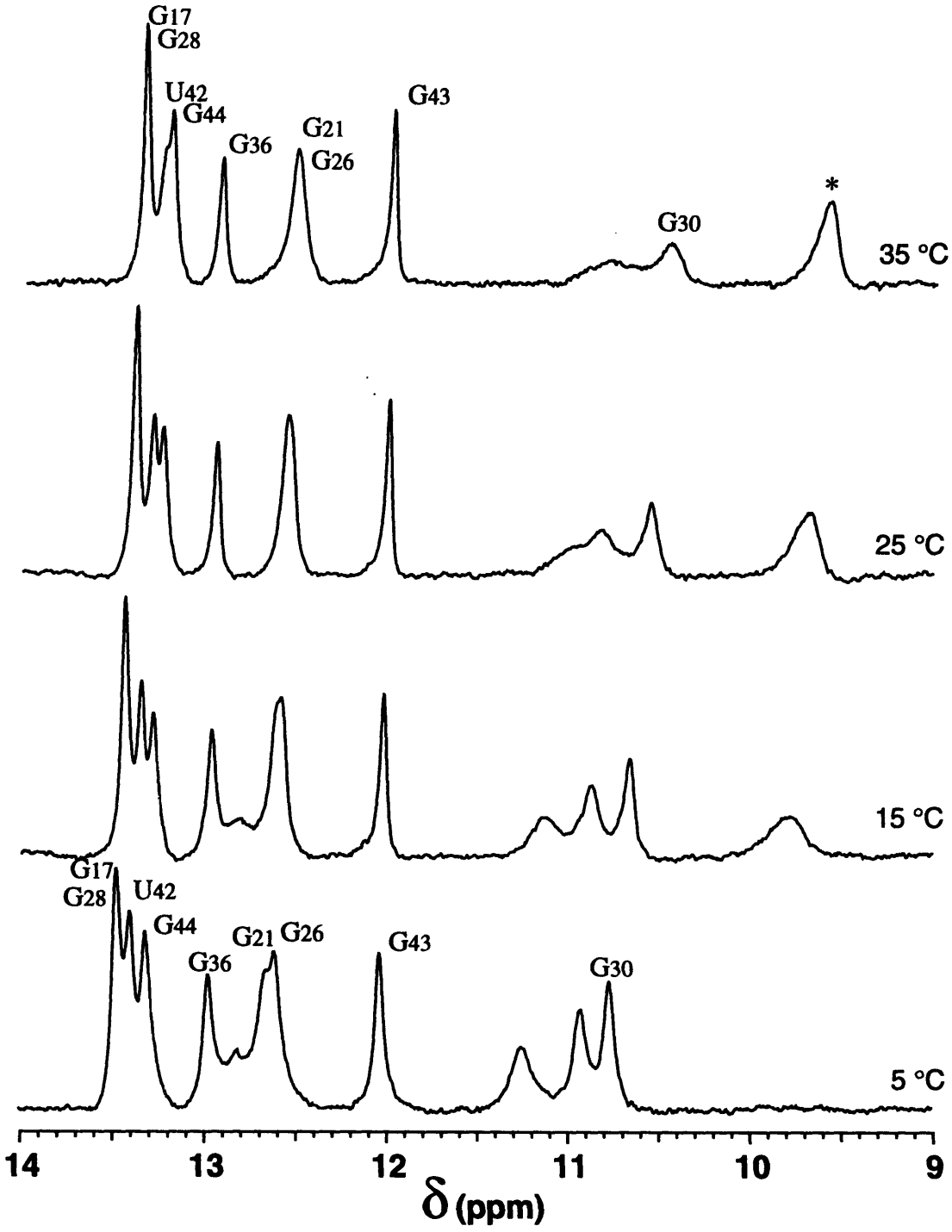


Figure 4.4) Temperature depencence of tTAR-UAC at pH 5.3. Most imino resonances are slightly broadened at low temperatures. All base pairs are stable at 35° C. The non-hydrogen bonded imino resonances (ca. 11-11.5 ppm) are the first peaks to broaden with increasing temperature. Unlike the other peaks, the amino resonance (*) sharpens at higher temperatures.

resonance and the other non-bonded iminos. The three medium peaks between 10.5 and 11.5 ppm correspond to these protons. No NOEs were seen to any of them, even at temperatures as low as 5 °C where the peaks are the sharpest.

Interestingly, in both the free and argininamide samples, the resonance at 9.8 ppm sharpens with increasing temperatures. At temperatures below 15 °C, this peak is broadened into the base line. This temperature dependent behavior suggests that this peak results from an amino proton. Normally amino resonances from adenine and guanine nucleotides are broad due to rotational exchange about the C-N bond (McConnell & Seawell, 1972), and are not easily observed. The rate of rotation increases at higher temperatures, causing the lines to sharpen into a single resonance for the two amino protons. Adenine amino resonances have been observed at low pH in exchangeable NMR studies of DNA (Kalnik et al., 1988, Boulard et al., 1992, Boulard et al., 1995) and RNA (Puglisi et al., 1990b) samples containing an A-C apposition. Protonation of a base results in large downfield shifts of amino protons due to the charged environment, Figure 3.1, (Raszka, 1974, de los Santos et al., 1989, Sklenar & Feigon, 1990, Boulard et al., 1995). For helical nucleotides, the N^1 position an adenine may be protonated at neutral pH, certainly it will have occurred at pH 5.5 (Section 3.1). The charged environment will effect neighboring nuclei, primarily the adenine N^6 amino proton. Therefore, the peak at 9.8 is likely due to the adenine amino proton in the A-C mispair. Exchangeable NOESY experiments at 10 °C and 15 °C were performed with 32 and 64 scans per t1 slice. A very weak NOE is detected from this proton to the G28 imino resonance at 13.48 ppm. The presence of the stacking NOE provided support for the assignment of the peak at 9.8 ppm to the A27 amino group. Attempts to increase the intensity of this NOE by changing experimental conditions were unsuccessful.

Temperature points for the argininamide sample are shown in Figure 4.5. A comparison with the free RNA (Figure 4.4) shows that at 35 °C the imino resonances

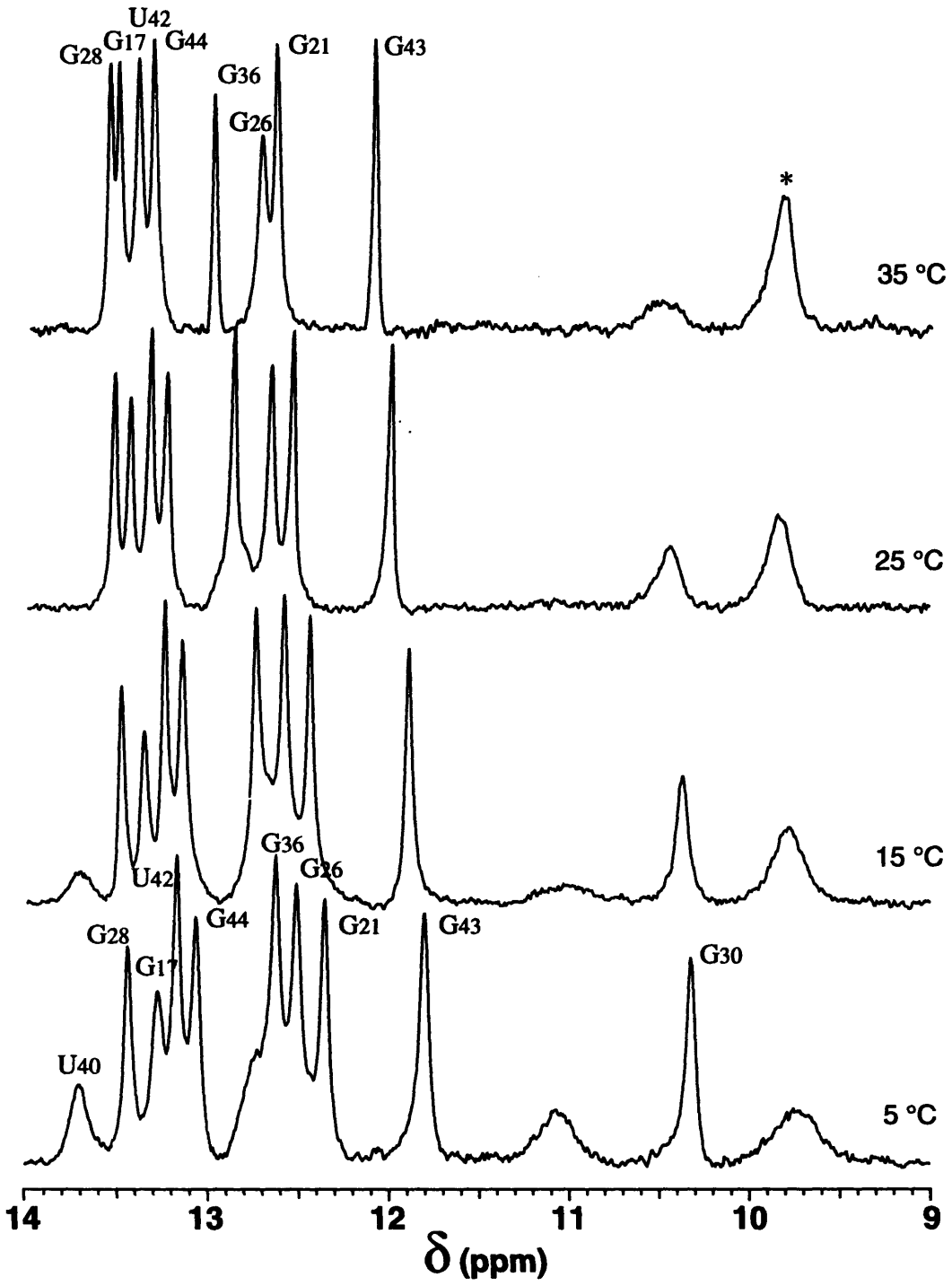


Figure 4.5) Temperature dependence of tTAR-UAC + argininamide complex at pH 5.5. The U40 imino resonance is apparent only at low temperatures. All of the other base pairs are stable at 35 °C. G26 is the first peak to broaden with increasing temperature. In the presence of argininamide the amino resonance (*) is visible at temperatures as low as 5 °C.

appear to be better protected from solvent exchange. This trend is especially true for the G26 imino proton; it is stronger at 35 °C in the pH 5.5 sample than it is at 25 °C at pH 6.4 (Figure 3.11). Even the tetraloop imino resonance is observable at 25 °C; this was not the case in the argininamide sample at higher pH. At 15 °C, all of the peaks are narrow, and well dispersed. At this temperature, the U40 resonance has started to appear. Consistent with the behavior of the TAR mutant at higher pH, the presence of argininamide in the sample stabilizes the hydrogen bonding of the A22-U40 base pair, located just below the bulge. Interestingly, only one broad peak is observed for the non-bonded imino protons (in addition to G30) in the 10.5-11.5 ppm range, even at 5 °C. The bulge nucleotides appear to be more susceptible to solvent exchange in the argininamide-containing sample. This behavior is consistent with the transition of the bulge nucleotides from a stacked environment to an unstructured one, as expected in the bound-conformation of the TAR molecule.

The amino resonance at ca. 9.9 ppm is also stronger in the argininamide spectra. NOESY experiments were performed at 15 °C (32 scans per t1, 512 increments) and 10 °C (256 increments at 64 scans per t1) on a 600 MHz NMR. Very weak NOEs are seen from this proton to imino protons of both base pairs flanking the mismatch site: G26 (12.73 ppm) and G28 (13.64 ppm). Recall that in the free RNA, an NOE is only observed to the G28 imino proton. At low pH, the bound conformation of the RNA further stabilizes stacking in the upper stem proximal to the A-C mispair.

4.2 Low pH Structural Studies – Free RNA

Exchangeable Data

The procedure for assignment of resonances at low pH was essentially as described for the samples at pH 6.4. Therefore, only the interesting characteristics of the spectral

analysis will be discussed. As noted above, the imino resonances that appeared at low pH do not give any NOEs in the exchangeable NOESY spectra. A similar set of imino-imino stacking NOEs is observed for the free RNA at low pH and at pH 6.4. The NOE from the G28 imino proton to the A27 H1' is very weak. At this pH, the U40 H1' and the C39 H1' overlap. It is therefore not possible to determine the identity of the NOE from G28 imino to 5.65 ppm. Similar to the spectra at higher pH, the C39 aminos are not determined. All other features of the exchangeable spectra were the same at low pH.

Non-exchangeable Data

There are few differences between the exchangeable spectra at the two pHs studied. Spectral assignments were made from a 400 ms mixing time SCUBA acquired at a field strength of 500 MHz. The data are presented in Figure 4.6. The main difference between this spectrum and that shown in Figure 3.8 (pH 6.4) is the peak shape. Although both data sets were processed exactly the same way, the high pH spectrum was acquired at 750 MHz, which results in much sharper resonances. The A22 H2 resonance gives strong NOEs to the cross-strand C41 H1', and to U23 H1'. A medium NOE is also seen to G26 H1', again indicating that the bulge nucleotides are not stacked between the upper and lower helices all of the time. However, the NOE to U23 H1' is the stronger one, indicating that this bulge nucleotide is predominantly stacked into the helix. In agreement with this observation is the presence of the expected stacking NOE from U23 H6 to C24 H5. The aromatic to H1' connectivities suggest that the bulge nucleotides (U23, C24, and U25) are stacked into the helix at both pH 6.4 and 5.3. The C39 H1' is again not observed, and is probably obscured its own H5-H6 cross peak. The A27 H2 NOE to G28 H1' is broad. This cross peak may result from the NOE to C39 H1' as well as that to G28 H1'. The only significant difference between the data at different pHs appears to lie in the NOEs from the C38 protons. Intranucleotide H6 to H1' and H5 NOEs are more intense at lower pH, and the H5-H6 cross peak is no more broad than that for any of the other pyrimidines. This

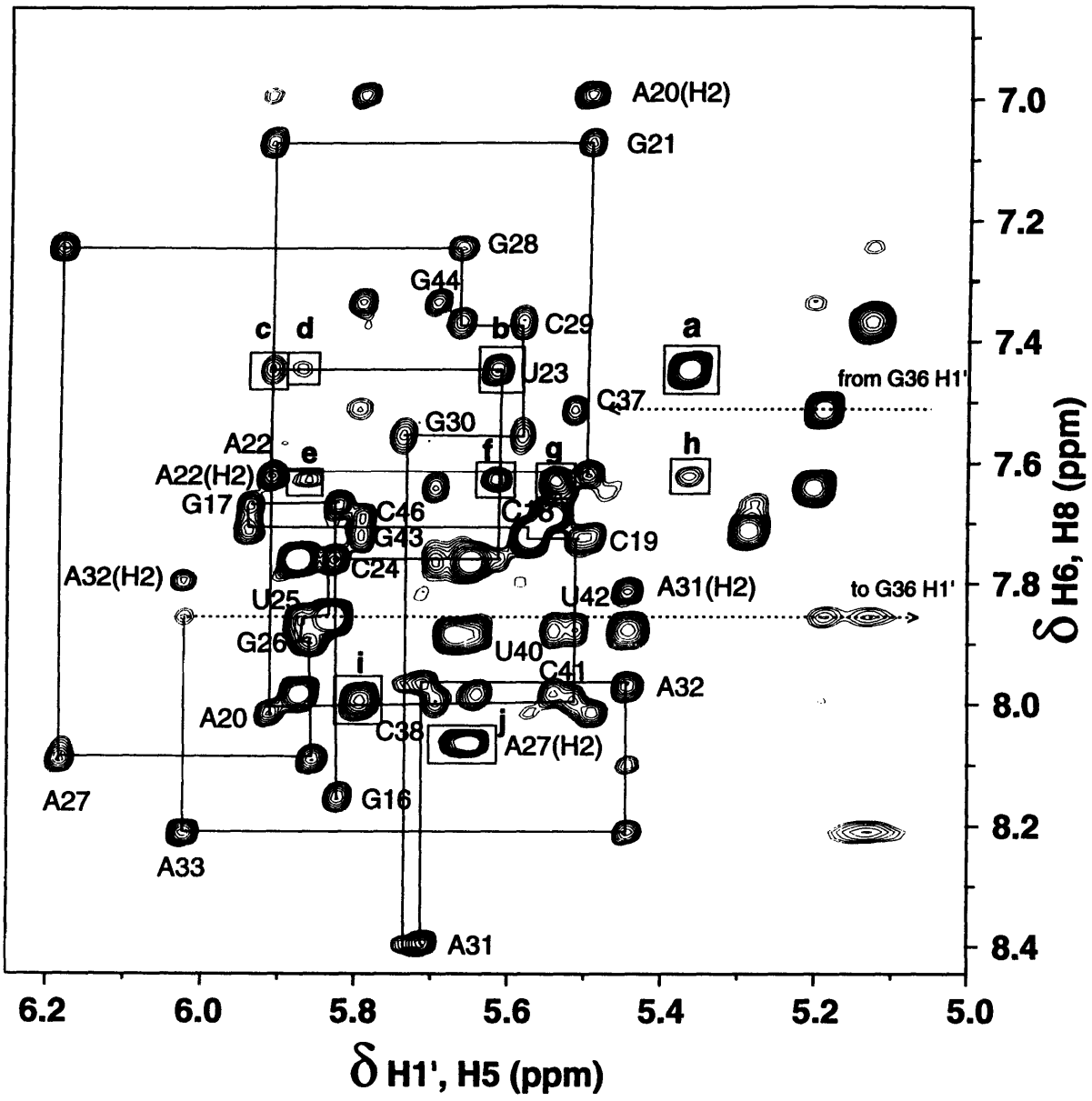


Figure 4.6) tTAR-UAC SCUBA pH 5.3. The aromatic to H1' connectivity pathway is shown for nucleotides 16 -33. The NOE pattern is indicative of a mixture of confomations, with the bulge nucleotides stacked between the upper and lower stems a significant part of the time. Boxed NOEs **a-d** are from U23 H6 to U23 H5, U23 H1', A22 H1', and G26 H1', respectively. **e**, **f**, and **g** are NOEs from G26 H1', U23 H1' and C41 H1' to A22 H2. The crosspeak in box **h** represents a stacking NOE between A22 H8 and U23 H5. The C38 H5-H6 crosspeak is sharper at this pH than it was at pH 6.4 (box **i**). The A27 H2 resonance is enclosed in box **j**. Data were acquired with a 400 ms mixing time at 25 °C in 99.996% D_2O, 10 mM Na-phosphate, 50 mM NaCl, 0.1 mM EDTA.

RNA appears to form a similar structure to that of the free wild-type TAR. Spectral assignments for tTAR-UAC at pH 5.3 are listed in Table 4.1.

4.3 Low pH Structural Studies - RNA/Argininamide Complex

Exchangeable Data

A NOESY experiment was performed at pH 5.5 and 10 °C, in 90% H_2O, 10% D_2O, with a 200 ms mixing time. As discussed with the 1-dimensional data, very weak stacking NOEs are seen from the C38 amino proton to the imino protons of the surrounding base pairs. Many of the other NOEs from the exchangeable protons are the same at pH 6.4 and pH 5.5. No NOEs are detected to the broad U40 imino proton. At low pH, the helical H1' NOEs to the G28 imino proton are much stronger than they are at pH 6.4, an indication that the conformation of the nucleotides in the upper stem is more similar to the A-form geometry than was observed at higher pH. The NOE to C37 H5 is weaker than either of the NOEs to C29 H1' or C38 H1'; it was not observed in the data collected at pH 6.4. The NOE from the G26 imino proton to C29 H1' was observed only in the free sample at pH 6.4. At low pH, however, this cross peak is observed in both the free and bound spectra. NOEs from the G26 imino proton are detected to the C39 amino protons at low pH, yet they were absent in the pH 6.4 data sets. The NOE to the U40 H1' proton is stronger at low pH than at pH 6.4. The stacking between the upper and lower helices is stabilized at lower pH.

Decreasing the pH has a small effect on the exchangeable spectra of the argininamide sample. The presence of the NOEs from the A27 amino proton to the G26 and G28 imino protons indicates that a protonated A-C pair is forming, and that the pair is part of a roughly A-form helical geometry. NOEs between the G26 imino resonance and the U40 H1' indicate that, as expected, the presence of argininamide causes the stacking of the

Table 4.1
Assignments for tTAR-UAC (pH 5.3)
Chemical shifts are reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino	Amino
G16	8.15	NA	5.82			
G17	7.67	NA	5.94		13.48	
C18	7.71	5.29	5.57		NA	6.95/8.65
C19	7.72	5.57	5.49		NA	6.93/8.38
A20	8.01	6.99	5.91		NA	
G21	7.07	NA	5.50		12.65	
A22	7.62	7.63	5.91		NA	
U23	7.45	5.37	5.62			NA
C24	7.76	5.87	5.83		NA	
U25	7.86	5.83	5.87			NA
G26	7.90	NA	5.86		12.61	
A27	8.09	8.06	6.18		NA	9.65 #
G28	7.24	NA	5.66		13.48	
C29	7.37	5.13	5.58		NA	6.87/8.26
G30	7.55	NA	5.74		10.75 *	
A31	8.39	7.81	5.71		NA	
A32	7.97	7.80	5.44		NA	
A33	8.21	8.10	6.02		NA	
G36	7.86	NA	3.61		12.99	
C37	7.51	5.19	5.51		NA	6.99/8.91
C38	7.99	5.79	5.69		NA	
C39	7.76	5.65			NA	
U40	7.88	5.67	5.65			NA
C41	7.98	5.87	5.54		NA	7.15/8.34
U42	7.88	5.44	5.51		13.39	NA
G43	7.72	NA	5.79		12.05	
G44	7.33	NA	5.70		13.32	
C45	7.64	5.20	5.53		NA	6.99/8.58
C46	7.69	5.54	5.79		NA	

NA refers to not applicable protons, blank spaces indicate protons that were not assigned, * denotes 5 °C data, # represents 35 °C data.

two stems in tTAR-UAC. The upper stem is most stable at low pH and in the presence of argininamide.

Non-exchangeable Data

A 400 ms mixing time SCUBA was acquired for the argininamide sample at 25 °C at 500 MHz. The aromatic to H1' region is shown in Figure 4.7. A comparison of these data with Figure 3.12 shows a difference in peak shape due to the difference in field strength (500 vs. 750 MHz). All expected aromatic to H1' and H2' connectivities are observed, including the region surrounding the A-C mispair. Important NOE data are summarized below for nucleotides in the vicinity of the bulge and A-C apposition. All assignments are listed in Table 4.2, and chemical shift differences between the free and bound RNA are listed in Table 4.3. A schematic representation of the chemical shift differences is also given in Figure 4.8. The $\Delta\delta$s calculated at pH 6.4 and 5.5 are extremely similar (compare with Table 3.6 and Figure 3.14).

Bulge Region

The most obvious difference between the argininamide spectra at the two pHs studied it that the A22 H2 NOEs are apparent at pH 5.5. A strong NOE to the C41 H1' proton is observed. Weak NOEs also occur to the inter-residue H1' and to G26 H1'. The NOE to G26 H1', although it is weak, is evidence for the displacement of the bulge nucleotides with concomitant stacking of the A22 and G26 nucleotides. An NOE may be present between G26 H8 and A22 H1'. These resonances are nearly overlapped in the spectrum, making identification of the cross peak difficult. This cross peak may be too broad to be observed in the low pH spectrum acquired at 500 MHz. The G26 H8 to A22 H1' NOE was present in the spectrum acquired at pH 6.4 and 750 MHz (Figure 3.12). Support for stacking of the two stems comes from the pair of nucleotides opposite from the bulge (C39 and U40). Similar to the pH 6.4 data, NOEs from U40 H6 to C39 H1', H2', and H6 are observed.

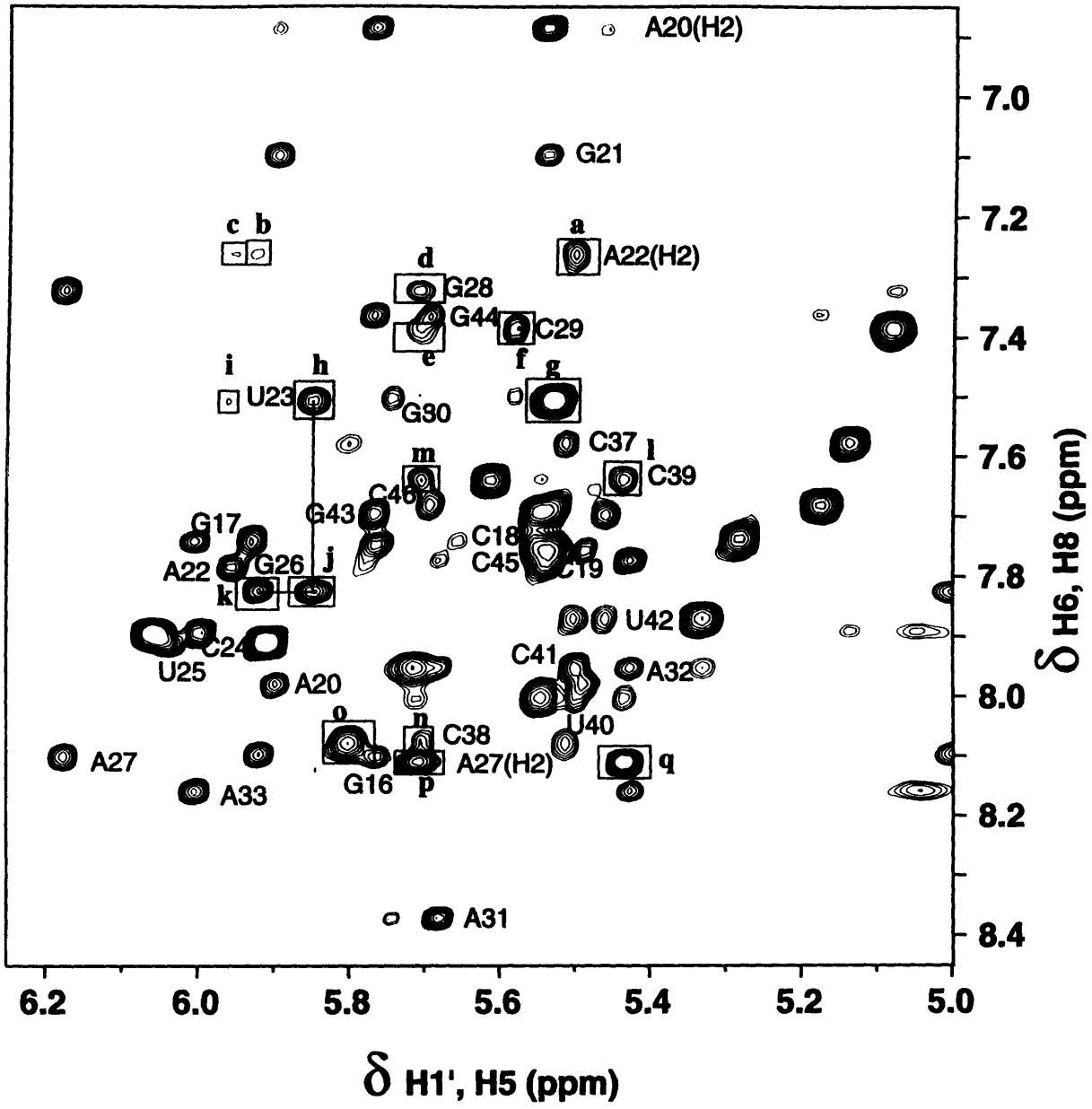


Figure 4.7) tTAR-UAC + argininamide SCUBA at pH 5.5. In contrast to the NOESY taken with the argininamide-containing sample at pH 6.4 (Figure 3.12), the A22 H2 proton is easily assigned. The NOE to C41 (box **a**) is much stronger at pH 5.5. Also observable are weak NOEs to G26 H1' (**b**) and to its own (A22) H1' (**c**). These NOEs are consistent with stacking of the upper and lower stems. NOEs **d** through **f** show the G28 through C29 aromatic to H1' connectivity. Similar to the A22 peaks, the G28 resonances are also stronger at low pH. The U23 NOEs are shown in boxes **g** through **i**. Box **g** contains the H5-H6 crosspeak, the H1' NOE is in **h**, and a very weak stacking NOE to A22 H1' is outlined in **i**. A strong NOE is detected from the U23 H1' to G26 H8 (**j**). Boxes **l**, **m**, and **n** contain the aromatic to H1' connectivities for C39 through C38. The C38 H5-H5 peak is in **o**. **p** and **q** are the A27 H2 resonances. Data were collected at 25 °C with a 400 ms mixing time.

Table 4.2
Assignments for tTAR-UAC + 6 mM Argininamide (pH 5.5)
Chemical shifts are reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino	Amino
G16	8.10	NA	5.77	4.91		
G17	7.74	NA	5.93	4.60	13.51	
C18	7.74	5.28	5.55	4.44	NA	6.95/8.68
C19	7.76	5.54	5.49	4.60	NA	6.90/8.45
A20	7.98	6.89	5.90	4.68	NA	
G21	7.10	NA	5.54	4.39	12.59	
A22	7.79	7.26	5.96	4.33	NA	
U23	7.50	5.53	5.85	4.34		NA
C24	7.90	6.06	6.00	4.40	NA	
U25	7.91	5.91	6.03	4.46		NA
G26	7.83	NA	5.92	4.95	12.74	
A27	8.10	8.11	6.18	4.84	NA	9.8 #
G28	7.32	NA	5.71	4.43	13.64	
C29	7.38	5.08	5.58	4.44	NA	6.65/8.20
G30	7.50	NA	5.74	4.82	10.35 *	
A31	8.37		5.68	4.75	NA	
A32	7.95	7.74	5.43	4.31	NA	
A33	8.16	7.77	6.00	4.55	NA	
G36	7.89	NA	3.63	4.27	12.89	
C37	7.58	5.14	5.51	4.61	NA	6.95/8.59
C38	8.08	5.80	5.70	4.24	NA	
C39	7.64	5.61	5.44	4.37	NA	7.35/8.54
U40	8.00	5.55	5.50	4.41	13.94 *	NA
C41	7.95	5.71	5.50	4.25	NA	7.10/8.37
U42	7.87	5.33	5.46	4.56	13.40	NA
G43	7.70	NA	5.77	4.60	12.04	
G44	7.36	NA	5.69	4.46	13.30	
C45	7.68	5.17	5.54	4.30	NA	6.97/8.67
C46	7.69	5.54	5.75	4.02	NA	

NA refers to not applicable protons, blank spaces indicate protons that were not assigned, * denotes 5 °C data, # represents 35 °C data.

Table 4.3
Chemical shift differences for tTAR-UAC, Free data–Argininamide data (pH 5.5).

Δδ reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino
G16	+0.05	NA	+0.05		
G17	-0.07	NA	+0.01		-0.03
C18	-0.03	+0.01	+0.02		NA
C19	-0.04	+0.03	0.0		NA
A20	+0.03	+0.10	+0.01		NA
G21	-0.03	NA	-0.04		+0.06
A22	-0.17	+0.37	-0.05		NA
U23	-0.05	-0.16	-0.23		
C24	-0.14	-0.19	-0.17		NA
U25	-0.05	-0.08	-0.16		
G26	+0.07	NA	-0.06		-0.13
A27	-0.01	-0.05	0.0		-0.15 # (amino)
G28	-0.08	NA	-0.05		-0.16
C29	-0.01	+0.05	0.0		NA
G30	+0.05	NA	-0.01		+0.40 *
A31	+0.02		+0.03		NA
A32	+0.02	+0.06	+0.01		NA
A33	+0.05	+0.33	+0.02		NA
G36	-0.03	NA	-0.02		+0.10
C37	-0.07	+0.05	0.0		NA
C38	-0.09	-0.01	-0.01		NA
C39	+0.12	+0.04			NA
U40	-0.12	+0.12	+0.15		
C41	+0.03	+0.16	+0.04		NA
U42	+0.01	+0.11	+0.05		-0.01
G43	+0.02	NA	+0.02		+0.01
G44	-0.03	NA	+0.01		+.02
C45	-0.04	+0.03	-0.01		NA
C46	0.0	0.0	+0.04		NA

NA refers to not applicable protons, blank spaces indicate protons that were not assigned for one or both data sets, * refers to 5 °C data, # denotes 35 °C data.

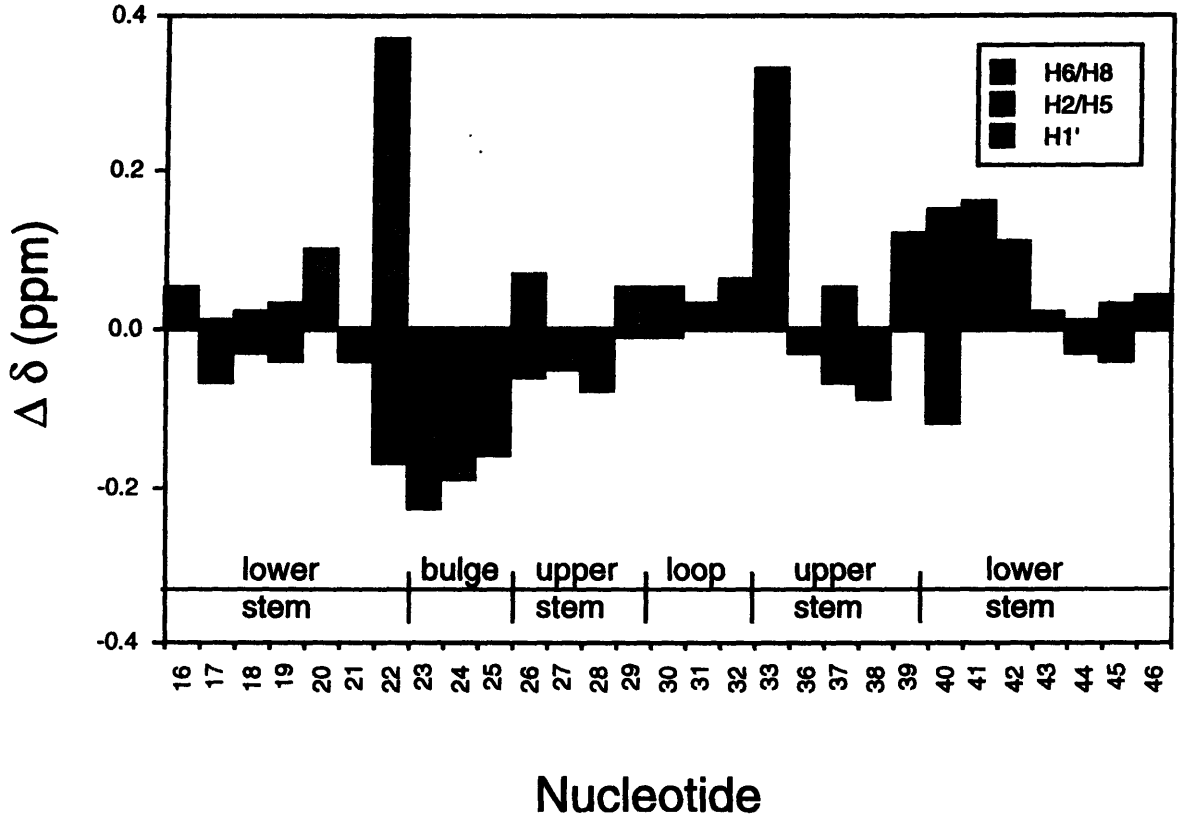


Figure 4.8) Plot of changes in chemical shift for tTAR-UAC resonances upon addition of 6 mM argininamide (pH 5.5). Nucleotide position, stem, bulge, and loop regions are marked. The H8/H6 proton chemical shift changes are indicated by solid bars, H2/H5 protons by hatched bars, and H1' protons by shaded bars. Similar chemical shift changes were observed at pH 6.4 (Figure 3.14).

Consistent with unstacking of the bulge, no NOEs were observed between U23, C24, and U25. The U23 H5-H6 cross peak is not significantly more broad than that observed for the other pyrimidines, indicating a lack of conformational exchange for this nucleotide; recall that this cross peak was broadened in the pH 6.4 spectrum. A very strong NOE is present between G26 H8 and U23 H1', indicating the proximity of these nucleotides. No cross peaks are identified that correspond to NOEs between G26 H8 and the 2' proton of any assigned ribose. Unfortunately, a weak NOE is also detected between U23 H6 and A22 H1'. Thus, U23 is stacked on A22 part of the time. However, similar to the data acquired at pH 6.4, an NOE is observed between U23 H1' and G26 H3', placing U23 in the major groove of the upper stem in the complex with argininamide.

A-C Region

All expected aromatic to H1' and H2' NOEs are observed for nucleotides in the upper stem. This includes connectivites on both strands- from G26 through G30 and G36 through C39. A27 H2 gives strong NOEs to both G28 H1' and C39 H1'. Despite the different peak shapes, the G28 and C38 resonances are much sharper in spectra taken at pH 5.5 than they are at pH 6.4. In general, the molecule appears more structured at low pH and in the presence of argininamide.

DQF-COSY Data

In DQF-COSY experiments of both the free and bound samples 14 strong, well resolved H5-H6 cross peaks are observed. This contrasts the data acquired at pH 6.4 where the C38 H5-H6 resonance was barely visible in the argininamide sample. As with the data obtained at higher pH, differences were observed between the H1'-H2' regions of the COSY spectra for the free and bound forms of tTAR-UAC. The spectra are presented in Figure 4.9 (compare with Figure 3.15). Similar changes in the cross peak intensities and chemical shifts occur between the free and argininamide samples at both pHs. At pH

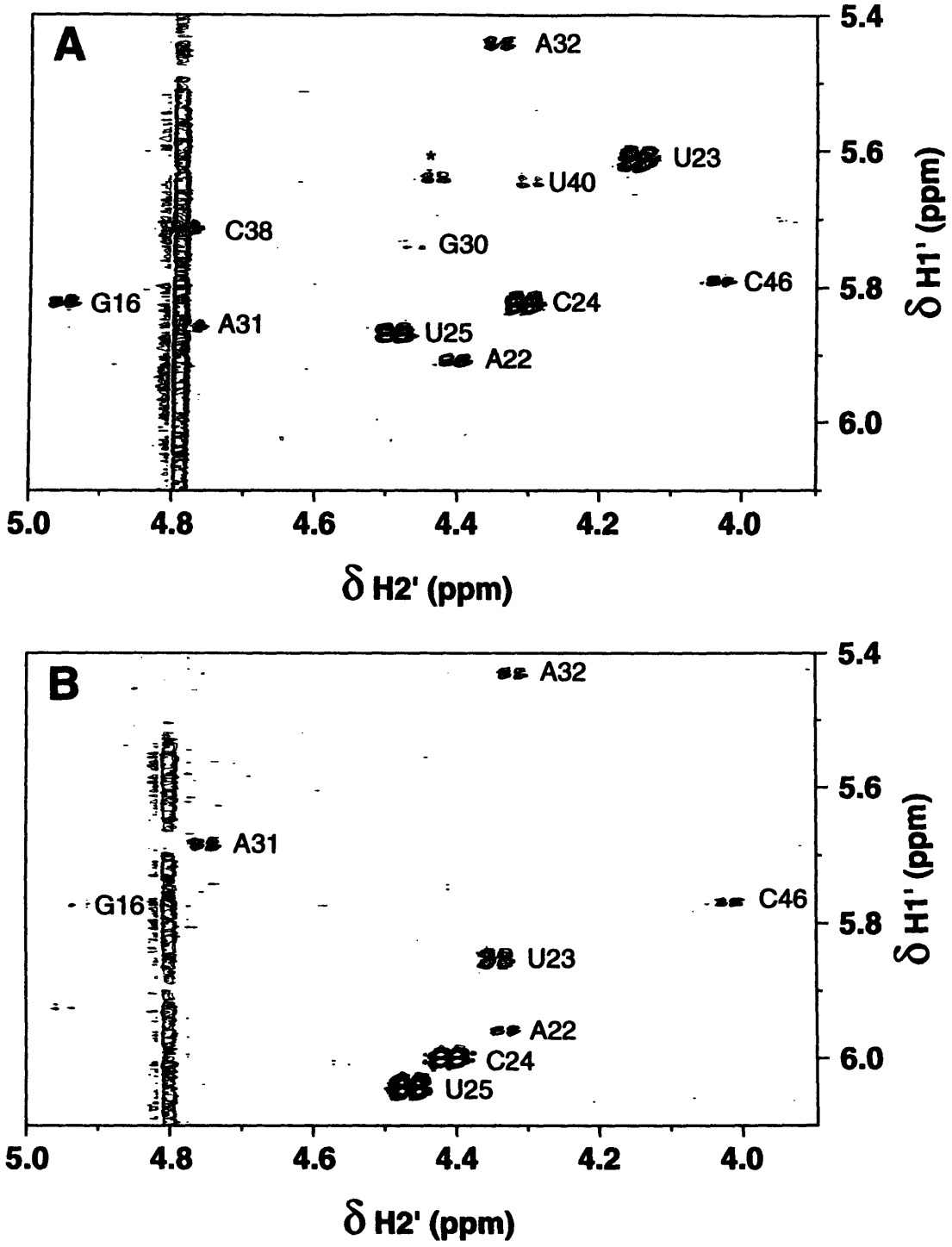


Figure 4.9) Comparison of DQF-COSYS at pH 5.5, free (A) and argininamide (B). Similar to the results obtained at higher pH, the additon of argininamide causes the bulge nucleotides (C24 and U25) to unstack, resulting in larger, down-field shifted H1'-H2' COSY peaks. The cosspeak for U23 is smaller in the argininamide spectrum (B).

5.5 in the argininamide sample, the C24 and U25 H1'-H2' cross peaks are more intense than in the free sample, while A22 and U23 are weaker. As discussed in Chapter 3, the data are consistent with unstacking of the bulge nucleotides and formation of the U23•A27-C38 base triple. The NMR data are summarized schematically in Figure 4.10.

NOEs to Argininamide

NOE data for the argininamide sample at pH 5.5 were collected at 500 MHz with mixing times of 400 and 200 ms. The pattern of NOEs in both experiments is similar to that observed at pH 6.4 (Figure 3.17). With the exception of the missing NOE from the argininamide δ proton to A22 H1', a similar set of NOEs was observed with this mutant (at both pHs) as in the wild-type RNA. The cross peaks observed from the β proton to G28 H8 and C29 H5 support the presence of an additional argininamide binding site in the tTAR-UAC mutant RNA. These nucleotides are located ca. 11 Å away from the argininamide β proton in the model proposed by Brodsky and Williamson (A.S. Brodsky and J.R. Williamson, manuscript in preparation). These NOEs are of equal intensity in shorter mixing time experiments, suggesting that the cross peaks result from direct interactions and are not simply due to spin diffusion through neighboring protons.

4.4 Conclusions

Detailed NMR experiments were not performed at pH 7.5, the high pH used for the peptide binding studies with TAR mutants (Chapter 3). At this pH the A-C mispair is likely to be unprotonated, and stabilized by only one hydrogen bond. The free energy of destabilization of an A-C pair (relative to an A-U pair) has been determined to be approximately 2 kcal/mole in an RNA hairpin (Puglisi et al., 1990b). Similar destabilization has been calculated for the equivalent DNA mismatch (Aboul-ela et al.,

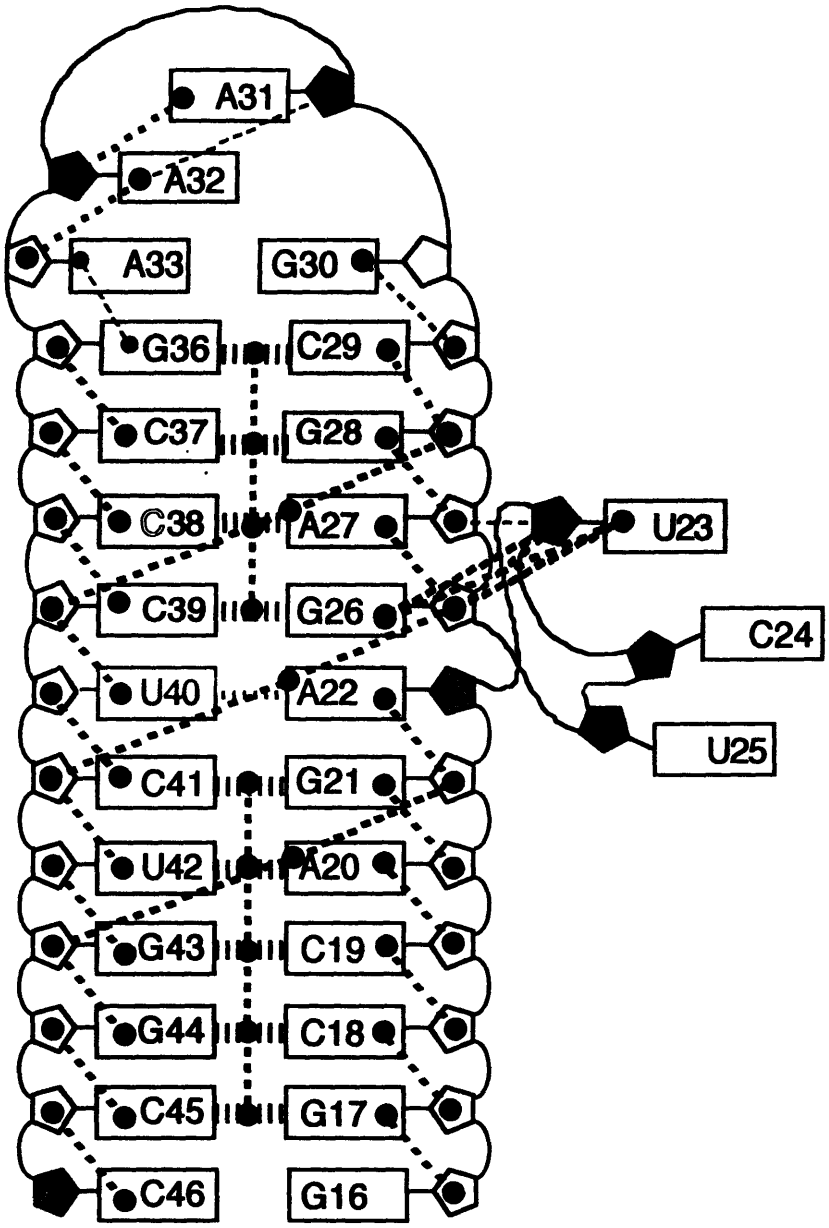


Figure 4.10) NMR summary for tTAR-UAC + 6 mM argininamide, pH 5.5. Symbols are as in Figures 3.9 and 3.16. A-form helical character is seen for a majority of the molecule. At this pH, NOEs are seen from A22 H2 to C41 H1' and G26 H1'. NOEs from U23 H1' to G26 H3' locate this nucleotide in the major groove of the upper stem, as seen at pH 6.4. NOEs were not seen between G26 H8 and A22 H1'. The thin, horizontal dashed line between U40 and A22 represents hydrogen bonding detected by the presence of the U40 imino proton at low temperatures. Similarly, the hydrogen bond between the A27 (protonated at the N1 position) and C38 is indicated. Stacking NOEs were seen from the A27 amino proton to the imino protons in neighboring base-pairs (vertical dashed lines).

1985). It is expected that the destabilization effects the conformation of base pairs surrounding the A-C mispair (Puglisi et al., 1990b). At pH 7.5 the imino resonance for the G26-C39 base pair (located below the A-C and above the bulge) is not detected (Figure 4.3). In the wild-type model, G26 is proposed to make a direct contact with the guanidinium group of the argininamide (Figure 1.12). The decreased binding of the peptide at high pH may be explained by local destabilization of the arginine binding site by the A-C apposition. High pH would further inhibit the formation of the argininamide binding pocket in this mutant RNA.

Similar conformations for RNA with U•A-U (Puglisi et al., 1992), C^+•G-C (Puglisi et al., 1993), and the U•A^+-C (this work) triples have been determined. Partial formation of the U•A^+-C triple interaction was observed at both pH 6.4 and 5.5, although in both cases there was evidence that U23 remains partially stacked on A22. Incorporation of an A-C mispair into a B-form DNA helix (Hunter et al., 1986, Kalnik et al., 1988) and an A-form RNA helix (Puglisi et al., 1990b) has been observed. The C1'-C1' distance for nucleotides in an A-C mispair is slightly shorter than in a standard Watson-Crick pair (Figure 3.1). The major structural difference for the mispair is that the purine is shifted toward the minor groove, and the pyrimidine toward the major groove. This minor distortion allows the formation of two hydrogen bonds between a protonated A and a C nucleotide. A minor perturbation of the triple formation may result from a slightly different major groove environment for U23. NOEs between the argininamide and the RNA were shown not to depend on the pH of the sample. At lower pH, however, the bound conformation of tTAR-UAC appears to be more stable and more similar to the wild-type model. The global conformation of the TAR mutant appears to be consistent with formation of a U•A^+-C triple at low pH. Consistent with peptide binding studies (Chapter 3) and the NMR results discussed above, recent modification interference experiments (Tao & Frankel, 1996) have provided evidence that the identity of the Watson-Crick nucleotide

at position 38 may be unimportant for triple formation. In this study, hydrazine modification of U38 in a TAR-like RNA (containing the wild-type U•A-U base-triple) did not interfere with arginine binding. The bulged U involved in the U23•A27 Hoogsteen interaction, however, was shown to be critical for arginine binding. Additionally, it has been shown (Hamy et al., 1993) that substitution of N^6-methyl-dA for A27 (a substitution that does not inhibit the Hoogsteen interaction) had only a moderate effect on *Tat* binding affinity. Chemical modification (Tao & Frankel, 1996, Hamy et al., 1993) and NMR (Puglisi et al., 1992, Puglisi et al., 1993, this thesis) data support the conclusion that the formation of a Hoogsteen interaction between positions 23 and 27, and the presence of a stable upper stem (including the required G26-C39 base pair) appear to be sufficient for argininamide binding to TAR RNA.

5 Duplex TAR RNAs

As outlined in Chapter 1, only a small set of nucleotides is important for specific recognition of TAT RNA by the TAR protein. These include a bulge uridine (U23), a G-C base pair located just above the bulge (G26-C39), and an A-U (A27-U38) pair capable of forming a base-triple interaction with U23 (Figure 1.12). In Chapters 3 and 4, evidence was presented that suggests that the identity of the Watson-Crick partner of A27 (N38) may be unimportant for base-triple formation. A stable environment for the Hoogsteen interaction between nucleotides 23 and 27, in addition to the G26-C39 base pair are necessary for *Tat* peptide or argininamide binding to TAR RNA. Thus, in addition to sequence requirements outlined in Chapter 1, the global structure of the RNA appears to be important for arginine binding.

A second approach toward investigating the requirements for formation of the base-triple interaction involves performing mutations at position 23 of TAR, the nucleotide involved in the Hoogsteen interaction. Previous work in our laboratory suggested the existence of the base-triple based on NMR constraints and molecular modeling studies (Puglisi et al., 1992, Puglisi et al., 1993). Unfortunately, the proposed hydrogen bonding for the Hoogsteen interaction was never directly observed via NMR (Figure 1.12). No U23 imino resonance was detected in the wild-type U•A-U triple. Similarly, the protonated C23 of the C^+•G-C mutant remained undetected. It was our hope to design an RNA molecule that would provide direct evidence for the presence of a base-triple.

As discussed in Chapter 2, imino protons in nucleic acids are only observable by NMR when their exchange with solvent protons is slowed by the hydrogen bonding of a

base-pairing interaction. Even within the context of the base-triple, the U23 imino proton exchanges too rapidly to be seen by NMR. A proton with a higher pK_a will exchange more slowly with solvent. We therefore reasoned that the incorporation of a modified uridine (one with higher pK_a for the N^3 proton) into TAR RNA at position 23 might allow detection of the hydrogen bonding involved in base-triple formation. The pK_a values of bases can be directly influenced by chemical modification (Saenger, 1984). Two bases with pK_as higher than uridine (9.3) are 5-methyluridine (9.7) and 5,6-dihydrouracil (>11) (Saenger, 1984). The difference between 5-methyluridine and uridine is too small to expect a significant difference in exchange properties. However, the pK_a of the dihydrouridine imino proton is sufficiently high so as to expect slow enough exchange with solvent to allow an observable imino resonance for this proton in the context of a base-triple interaction

Duplex TARs were synthesized via T7 template-driven transcriptions. To facilitate the incorporation of mutations specifically at position 23, TAR molecules were designed to contain only one U (at position 23) in the strand containing the bulge. Duplex RNAs have been used by a number of groups to investigate the nucleotide and functional group requirements of *Tat* binding to TAR RNA (Sumner-Smith et al., 1991, Hamy et al., 1993, Barnett et al., 1993, Pritchard et al., 1994, Wang & Rana, 1996). For the duplexes used in this work U25 was deleted - leaving a two nucleotide U, C bulge. Dinucleotide bulges have been shown to bind efficiently to *Tat* protein and peptides (Weeks & Crothers, 1991, Weeks & Crothers, 1992, Churcher et al., 1993) and argininamide (Tao & Frankel, 1996, A.S. Brodsky and J.R. Williamson, manuscript in preparation), as long as the nucleotide at position 23 remains a uridine.

The RNA sequences are shown in Figure 5.1. The core of nucleotides important for argininamide binding was unaltered. However, the sequences at the 5' ends of each strand were chosen so as to promote efficient transcription with T7 RNA polymerase and

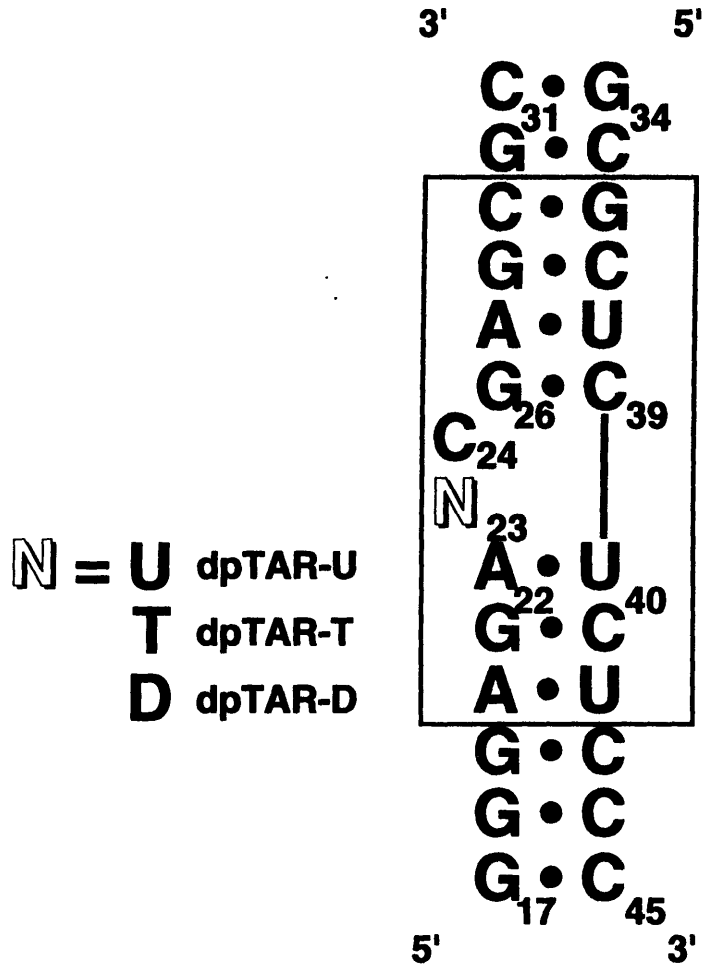


Figure 5.1) Sequences of duplex TAR mutants. Three TAR duplexes were synthesized using T7 template-driven transcription. The boxed nucleotides represent the 'core' of the TAR molecule. Mutations at position 23 are the only diferences from the wild-type sequence in this region. The 14 nucleotide strand containing the bulge is referred to as the 'A-strand', the 12 nucleotide complement is the 'B-strand'. Numbering is consistent with the wild-type hairpin TAR previously studied (Puglisi et.al., 1992).

minimize self-complementary duplex or hairpin formation. A control RNA (dpTAR-U) was synthesized in which the nucleotide at position 23 was a uridine. Two additional NMR samples were prepared. An RNA with U23 mutated to a deoxythymidine (dpTAR-T) was chosen as a control for the presence of a deoxynucleotide in this critical nucleotide. Because of the high pK_a of its imino proton (see above), dihydrouridine was also incorporated at position 23 (dpTAR-D).

5.1 dpTAR-U

Sample Preparation

A-strand

A 31 nucleotide DNA template with the sequence 5'- GCG CTC GAT CTC CCT ATA GTG AGT CGT ATT A -3' was synthesized as described in Materials and Methods. One half of the deprotected oligo from a 1 μmole synthesis was purified on a 20% polyacrylamide gel (25 x 35 x 0.15 cm) containing 8 M urea. The purification gel was run for 13 hours at 50 W. Electroelution was run at 150 V at room temperature, aliquots of eluent were collected every hour. Optimum transcription conditions were determined to be equivalent to the standard protocol (i.e. 4 mM ea. NTP, 1.75/1 $MgCl_2$/NTP). RNA from a 30 mL transcription was purified on 3 (25 x 35 x 0.3 cm) gels, run for 13 hours at 50 W. Electroelution of RNA was performed at 4 °C and 100 V, eluent was collected *every 45 minutes*. Frequent collection of product is necessary as small RNAs may pass through the electroelution membrane into the surrounding buffer. Elution typically took one to two days. Enough RNA for a 1 mM sample (600 μL) was obtained.

B-strand

An identical procedure was followed for the B-strand. The template sequence (29 nt) was 5'- GGG AGA GAG CGC TAT AGT GAG TCG TAT TA -3'. RNA synthesis

was performed under standard conditions, except that ATP was omitted from the reaction. Purification gels were run for 11.5 hours at 50 W. A 40 mL transcription produced enough RNA for 2 (1 mM) NMR samples. Elution was as above.

Complex Formation

Initial detection of dpTAR-U formation was performed using a crude gel-shift experiment. 3' ^{32}P end-labeled B-strand migrated as one band on a non-denaturing gel (0.5x TBE). An equimolar amount of A-strand and B-strand were annealed in NMR buffer (10 mM Na-phosphate pH 6.4, 100 mM NaCl, 0.1 mM EDTA). These samples were heated to 90 °C for one minute and allowed to cool slowly to room temperature over 20 minutes, and then were placed on ice. Heat-cooled duplexes produced one slow moving band on a non-denaturing gel. A gel-shift experiment was also performed with the duplex RNA and a *Tat*-derived peptide. A 36 amino acid peptide (ADP1-R52) with the sequence CFTTK ALGIS YG**KKK RKKKK** KPPQG SQTHQ VSLSK Q (the RNA binding site is in bold, see Chapter 1) was shown to bind to the RNA duplex but not to either of the monomer strands. Based on these crude results it was determined that the duplexes would be suitable for NMR analysis.

NMR Studies

RNAs were synthesized and dialyzed as outlined in Materials and Methods. Final NMR buffer contained 10 mM Na-phosphate pH 6.4, 100 mM NaCl, 0.1 mM EDTA. Titration of the two strands was performed in the NMR tube. The sample was heat-cooled after each addition of complementary strand, and allowed to equilibrate for 5 minutes prior to data acquisition. Formation of the duplex was monitored by changes in the imino spectrum at 25 °C. The titration of two strands of dpTAR-U, the wild-type model, is shown in Figure 5.2. The presence of weak imino peaks in the spectrum of the A-strand alone indicates that this monomer has some self-complementary structure. However, these resonances are weak and appear to disappear upon addition of the B-strand. It should be

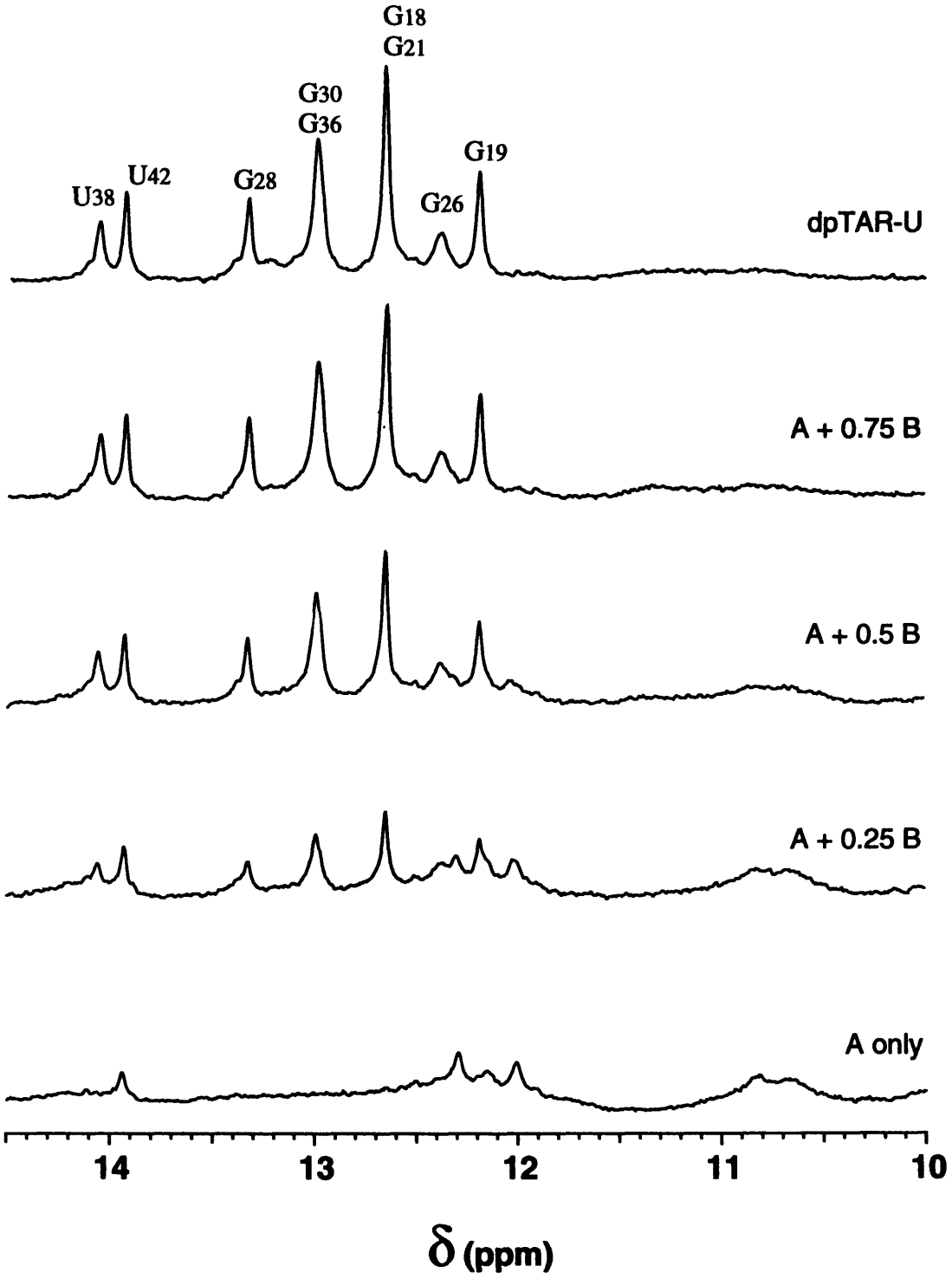


Figure 5.2) Titration of dpTAR-U RNA strands. A small amount of structure is apparent in the sample containing the A-strand alone. As the B-strand is added, imino resonances corresponding to duplex formation appear. Spectra were taken at 25 °C and 500 MHz. Sample buffer was 10 mM Na-phosphate pH 6.4, 100 mM NaCl, and 0.1 mM EDTA.

pointed out, however, that the G26 imino resonance is significantly weaker than the other peaks of the duplex RNA.

A complete set of two-dimensional experiments were performed with this molecule. A water NOESY was acquired at 15 °C, with a 200 ms mixing time. D_2O experiments included 400 and 50 ms SCUBAs, a DQF-COSY, and a TOCSY. Because this molecule was only used as a control for duplex formation, the specifics of the free RNA conformation will not be discussed here. The global structure is similar to the free form of other TAR molecules studied. The bulge nucleotides are stacked between the two stems; there was no evidence for hydrogen bonding between the A22 and U40 pair below the bulge. It was extremely difficult to assign this molecule. The B-strand contains a stretch of 9 pyrimidines. A majority of the H1' and H5 resonances are clustered in a very small part of the spectrum. In retrospect, the sequences of the non-critical parts of the RNA could have been more wisely chosen.

Argininamide titrations were performed as described for the TAR-UAC molecules. Two dimensional exchangeable and non-exchangeable experiments were also similarly carried out. The aromatic to H1' and H5 region of the 400 ms SCUBA of dpTAR-U + 6 mM argininamide is presented in Figure 5.3. Immediately obvious in this spectrum is the lack of NOEs to C24, suggesting that this nucleotide is not stacked. A very strong H1'-H2' cross peak is observed for this nucleotide in the COSY experiment. The A22 H2 resonances are well resolved in dpTAR-U. An intra-residue H1' NOE is seen, along with inter-residue NOEs to C41 H1' and to G26 H1'. There appears to be an NOE between U23 H1' and G26 H8, although it is difficult to distinguish G26 H8 from A22 H8 in this spectrum. The presence of the A22 H2 NOE to G26 H1' suggests that U23 is not stacked on A22, therefore this cross peak is likely to result from the NOE between U23 and G26. As with tTAR-UAC, an NOE is observed from the U23 H1' proton to G26 H3'. A weaker NOE was also detected from U23 H1' to G26 H2'. Argininamide protons give NOEs to

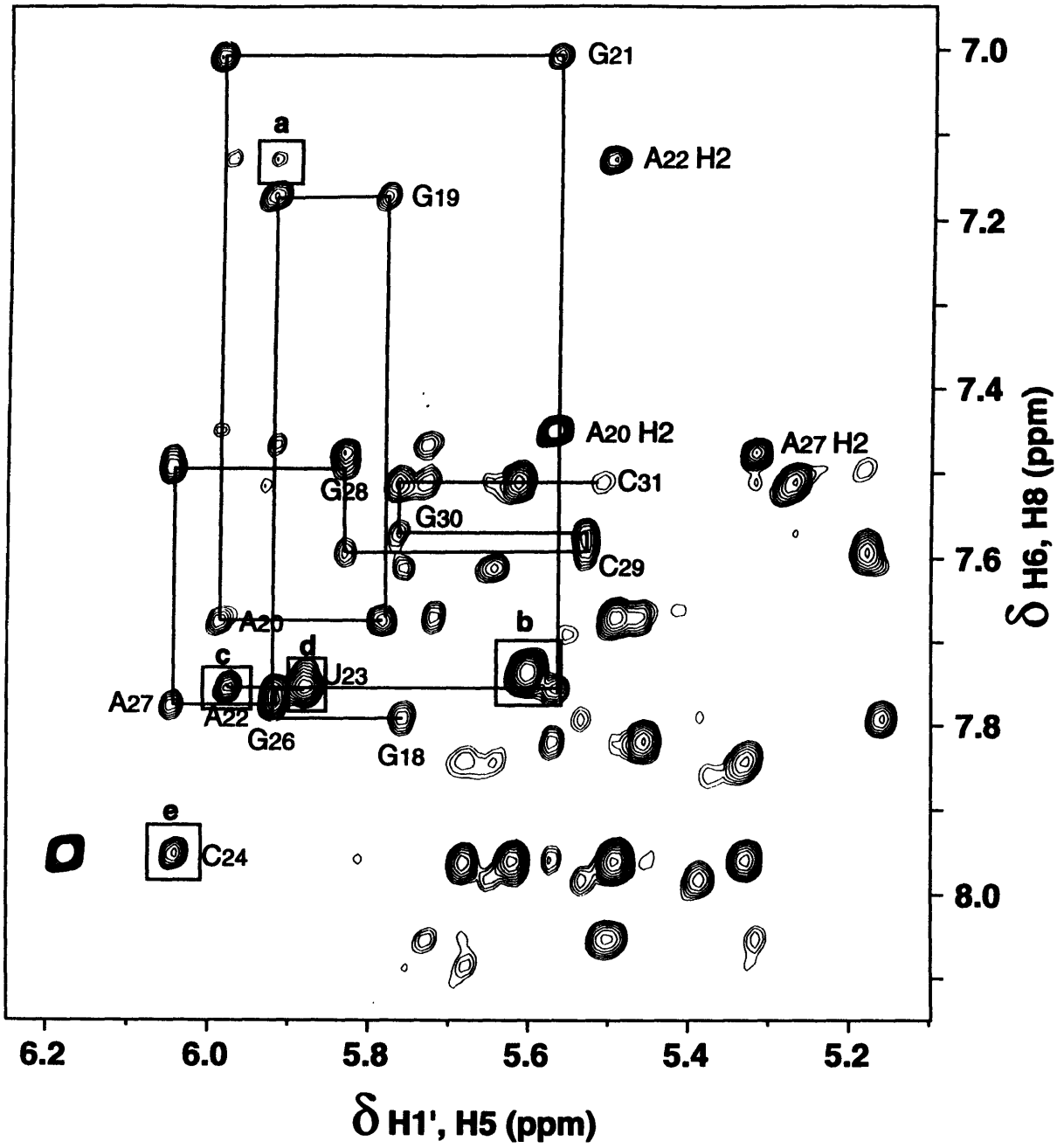


Figure 5.3) dpTAR-U + argininamide 400 ms SCUBA. The connectivity pathway for nucleotides G18 through C31 is indicated by the thick line. Box **a** shows the NOE between A22 H2 and G26 H1'. For reference, the U23 H5-H6 crosspeak is enclosed in box **b** and the H6 to H1' NOE is in box **d**. The broad intraresidue peak in **d** may also result from the U23 H1' to G26 H8 NOE. The A22 H8 is also in this region of the spectrum (box **c**). No NOEs were detected to C24 (box **e**). The data were acquired at 25 °C and 500 MHz in 10 mM Na-phosphate pH 6.4, 100 mM NaCl, 0.1 mM EDTA.

U23 H6, H5 H2', and H1'; to C29 H5; and to A22 H8 and H1'. It appears that the control duplex dpTAR-U forms a structure similar to that observed to hairpin TAR molecules.

5.2 dpTAR-T

Many modified nucleotides are commercially available only in the deoxyribose form. A duplex TAR molecule containing a deoxythymidine at position 23 was synthesized as a control for incorporation of deoxynucleotides into this critical position of TAR molecules. The presence of the 5-methyl group on the thymidine base, and the upfield-shifted 2' and 2" protons of the deoxyribose allow for detection of NOEs to argininamide or to other RNA protons in a well-resolved region of an NMR spectrum. Substitution of U23 by dU or 2'-O-methyl-U into TAR was shown to have no effect on *Tat* binding (Slim et al., 1991, Hamy et al., 1993). The RNAs containing these nucleotides were synthesized chemically. It is impractical for us to use chemical synthesis to produce the necessary quantities of RNA for NMR studies. There is only one uridine in the A-strand of dpTAR. The absence of UTP in a transcription reaction might induce T7 RNA polymerase to place dT at that one position.

The same templates are used for synthesis of dpTAR-T as were used for dpTAR-U. Trial transcription reactions (20 μL, see Materials and Methods) performed in the absence of UTP or dTTP produced no full-length RNA product. However, reactions containing dTTP, ATP, GTP and CTP at concentrations of 4mM each, under standard conditions, yielded an equivalent amount of the 14 nucleotide RNA as obtained with UTP transcriptions.

A 40 mL transcription, purified on 4 gels run for 13 hours at 55 W, produced enough of the A-strand of dpTAR-T for a 1.5 mM NMR sample. The B-strand was

prepared as described above. Electroelution and dialysis were as described above. NMR experiments were performed in 10 mM Na-phosphate pH 6.4, 100 mM NaCl, 0.1 mM EDTA.

NMR Analysis -Free dpTAR-T

Exchangeable Data

The titration of the A- and B-strands for dpTAR-T is shown in Figure 5.4. There are three imino resonances present in the spectrum containing only the A-strand. These disappear with addition of the B-strand. At a 1:1 complex, 7 peaks are apparent. Two are significantly larger than the others; they result from two protons each. Thus, nine base pairs are accounted for in the 1-dimensional exchangeable spectrum. A water NOESY was performed, imino proton assignments were made as shown in the figure. Protons from the two terminal base pairs and the A22-U40 pair located below the bulge are not observed. Imino-imino stacking NOEs are seen between the 4 central base pairs of the lower stem, and between U38, G28, and G36 imino protons. It is impossible to determine if G36 and G30 stack because of overlap of the two imino resonances, although it is likely that they do. A stacking NOE is not detected between G26 and U38.

The H_2O NOESY provided all of the expected NOEs from G-imino protons to H2, H5, H1' protons. Amino resonances are assigned for all cytidine residues except those in the terminal base pairs (C31 and C45), for C41 and C44 (obscured by spectral overlap). Surprisingly, a cross peak is observed at chemical shifts corresponding to G26 imino and U40 H1' protons. This NOE was not detected in RNAs containing a three nucleotide bulge in the absence of argininamide.

Non-exchangeable Data

The A27 and A20 H2 protons (assigned from the exchangeable data) served as starting points for interpretation of the non-exchangeable Data. Data from a 400 ms mixing

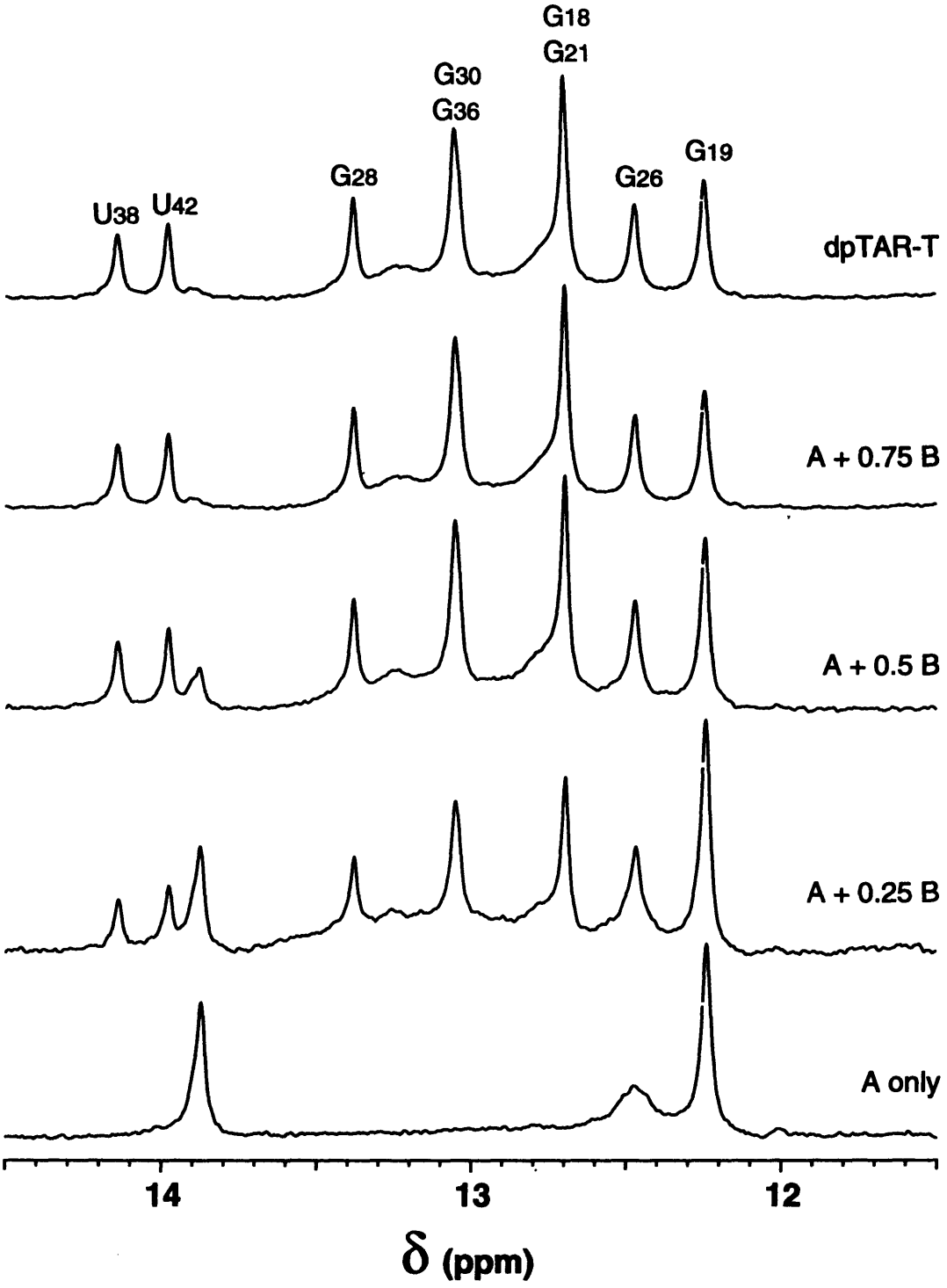


Figure 5.4) Titration of dpTAR-T RNA strands. The presence of imino resonances in the sample containing the A-strand alone indicates some self-complementary secondary structure. The disappearance of these peaks can be monitored as the duplex B-strand is titrated in. All expected base pairs are accounted for by the spectrum of the duplex RNA. Imino protons are not observed for the terminal base pairs and the A22-U40 pair below the bulge. Data were acquired at 25 °C and 500 MHz, in the standard NMR buffer containing 100 mM NaCl.

time NOESY (500 MHz, Varian) are presented in Figure 5.5. The H1' of the bulge nucleotide dT23 gives strong NOEs to its own H6 and to the H6 of C24; and C24 H1' gives an NOE to G26, consistent with stacking of the bulge nucleotides. However, there is also a weak NOE between dT23 H1' and G26 H8, suggesting partial unstacking of this nucleotide. Aromatic to H2' connectivities are assigned for dT23 to A22, C24 to dT23, G26 to C24 *and* G26 to dT23. Opposite from the bulge, standard NOEs are observed for stacking of C39 and U40. COSY experiments indicate H1'-H2' cross peaks suggestive of C2'-*endo* character for the 5'-terminal nucleotides on each strand (G17 and G34), A22, C24, and for dT23 1' to 2' & 2". A list of the assigned resonances is given in Table 5.1.

NMR Analysis -dpTAR-T + Argininamide

Exchangeable Data

The imino spectra recorded during the argininamide titration into the dpTAR-T sample are shown in Figure 5.6. The G21 imino resonance is the only one to exhibit a noticeable change in chemical shift. This resonance moves upfield 0.14 ppm upon addition of 6 mM argininamide. A broad peak at 14.39 ppm becomes apparent during the titration, and sharpens at low temperatures (data not shown). An additional resonance at 13.19 ppm is present throughout the titration. No NOEs are detected to either proton in 2-dimensional water NOESY experiments performed at 15 °C. The resonance at 14.39 ppm is likely to be the U40 imino proton. The peak at 13.19 ppm remains unassigned. In addition to the imino-imino NOEs observed for the free RNA, a stacking interaction between G26 and U38 is detected in the sample containing 6 mM argininamide. All imino to H5 and H1' NOEs are observed. Interestingly, the NOE from the G26 imino proton to U40 H1' is very strong; that to C39 H5 is considerably weaker than it was in the spectrum of the free RNA. The presence of argininamide appears to stabilize the stacking of the upper and lower stems in dpTAR-T.

Table 5.1
Assignments for dpTAR-T (pH 6.4)
Chemical shifts are reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino	Amino
G17	8.14	NA	5.79	4.89		
G18	7.46	NA	5.93	4.68	12.70	
G19	7.19	NA	5.79	4.62	12.24	
A20	7.68	7.50	5.99	4.75	NA	
G21	6.98	NA	5.51	4.49	12.70	
A22	7.67	7.59	5.98	4.45	NA	
T23	7.47	1.58 (CH_3)	6.17	2.26', 2.49"		NA
C24	7.86	5.93	5.91	4.48	NA	
G26	7.69	NA	5.70	4.80	12.47	
A27	7.86	7.49	6.04	4.78	NA	
G28	7.24	NA	5.70	4.48	13.38	
C29	7.58	5.19	5.54	4.55	NA	6.77/8.50
G30	7.59	NA	5.78	4.44	13.06	
C31	7.53	5.29	5.74		NA	
G34	8.15	NA	5.73	4.01		
C35	7.84	5.39	5.68	4.66	NA	6.79/8.63
G36	7.62	NA	5.78	4.56	13.03	
C37	7.71	5.25	5.54	4.42	NA	6.95/8.59
U38	7.90	5.44	5.60	4.66	14.17	NA
C39	7.72	5.66	5.53	4.55	NA	7.07/8.30
U40	7.96	5.40	5.52	4.49		NA
C41	7.97	5.79	5.49		NA	
U42	7.96	5.65	5.41		14.00	NA
C43	7.95	5.65	5.59	4.34	NA	6.98/8.32
C44	7.82	5.48	5.47	4.25	NA	
C45	7.68	5.47	5.74	4.44	NA	

NA refers to not applicable protons, blank spaces indicate protons that were not assigned.

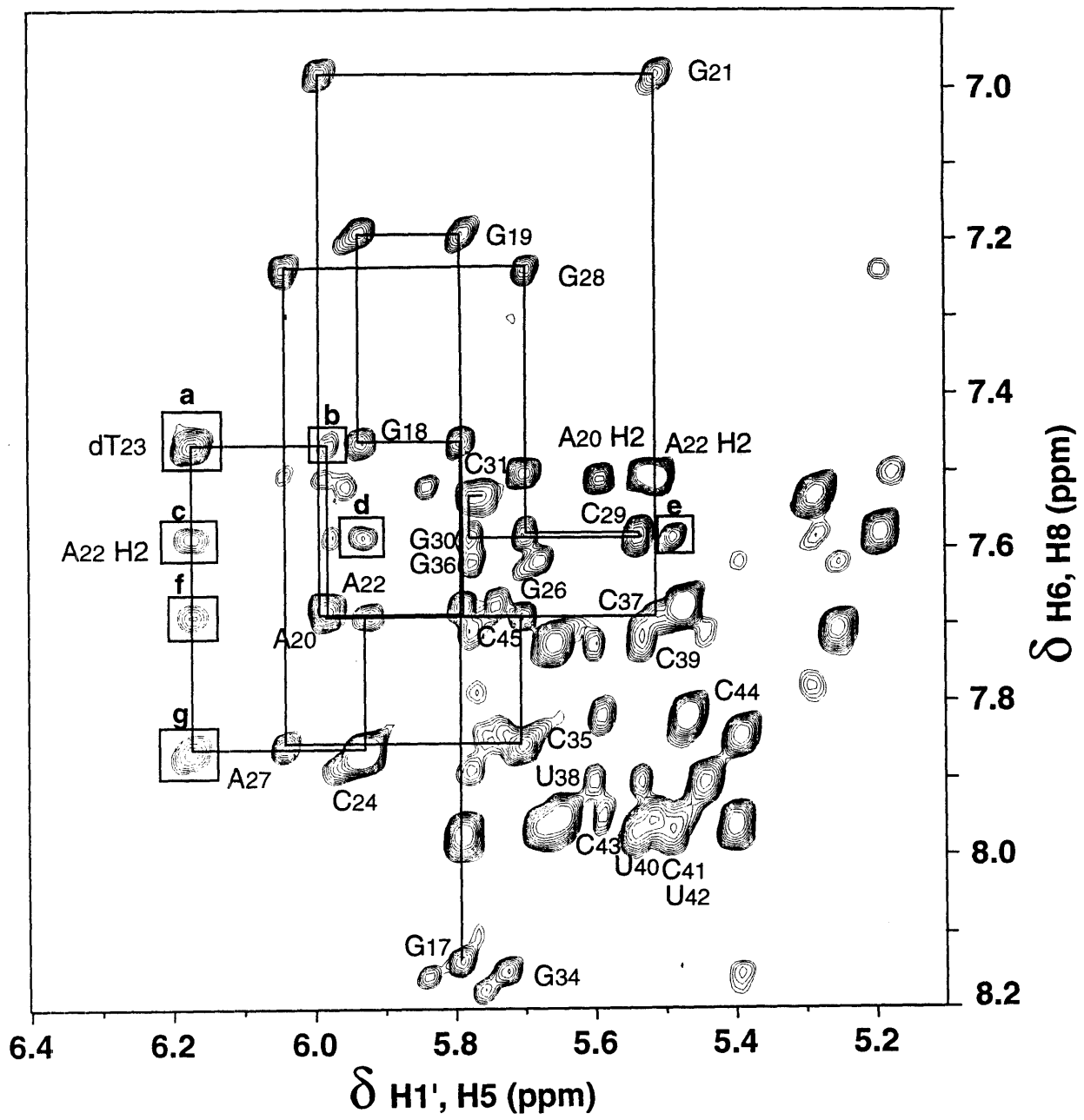


Figure 5.5) 400 ms NOESY of dpTAR-T. The aromatic to H1' region of the spectrum is shown. The connectivity pathway is traced for the A-strand of the duplex. The bulge nucleotides are only partially stacked. Box **a** shows the intraresidue H6-H1' NOE for dT23. In box **b** is the NOE from dT23 H6 to A22 H1', in box **g** is the NOE to C24 H6. There is also a strong NOE to G26 H8 (box **f**). Consistent with only partial stacking of dT23, A22 H2 gives NOEs to dT23 (**c**) and C24 (**d**) H1's in addition to the cross-strand NOE to C41 H1' (**e**). The spectrum was acquired at 25 °C and 500 MHz in standard NMR buffer containing 100 mM NaCl.

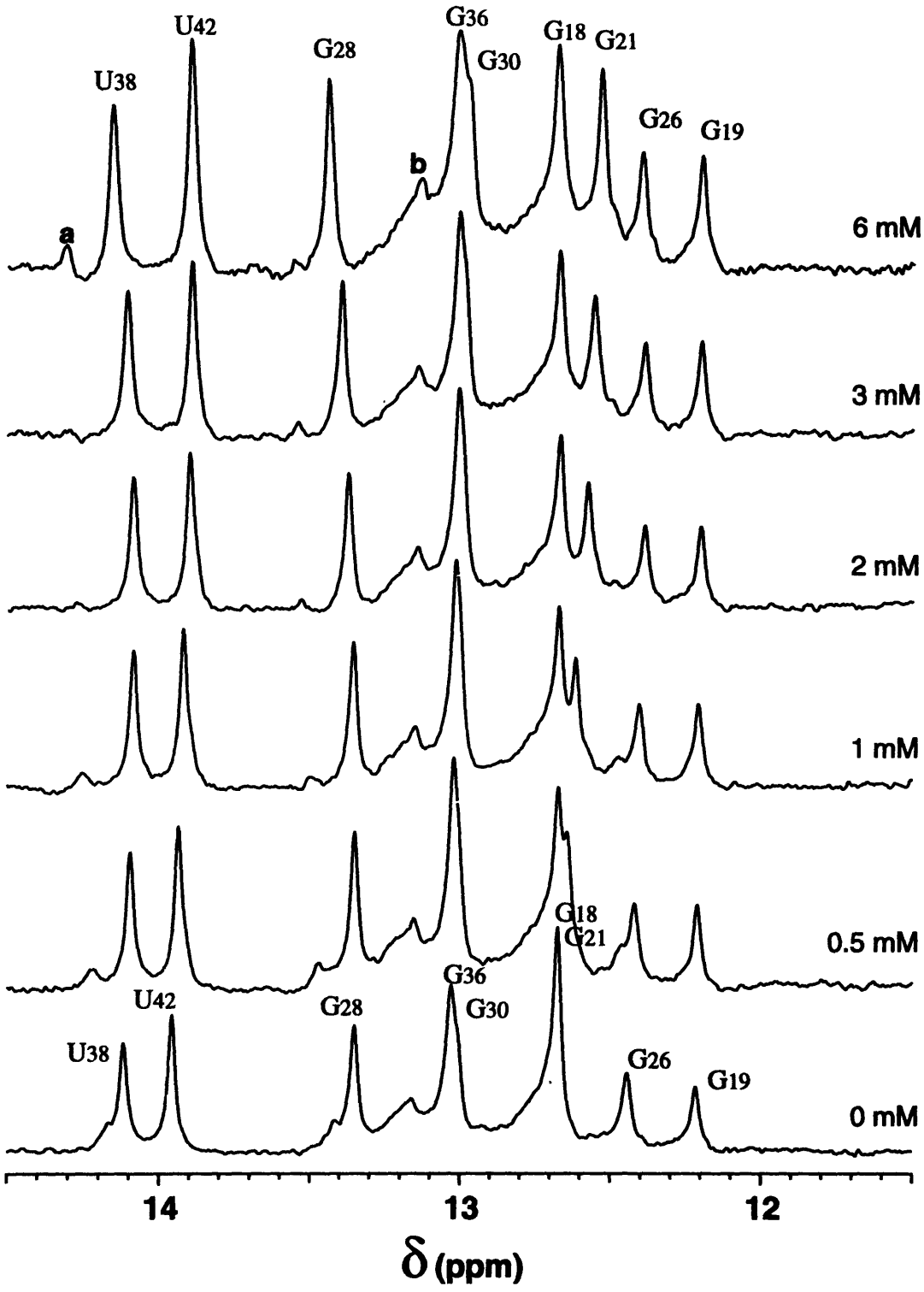


Figure 5.6) Argininamide titration of dpTAR-T at 25 °C. The argininamide titration was monitored by changes in the imino spectrum of the RNA. The G21 and G28 resonances experience the largest chemical shift differences upon argininamide binding. The weak peak labeled **a** is tentatively assigned to U40, while peak **b** remains unassigned.

Non-exchangeable Data

400 and 50 ms NOESYs were acquired at 25 °C and 500 MHz for assignment purposes. The region of the spectrum used for assignment is shown in Figure 5.7 Standard aromatic to 2' connectivities are traced for the entire B-strand, except where 2' resonances are unassignable (i.e. C45). H5-H5 and H6-H6 NOEs are also assigned between C39 and U40. An additional NOE between C39 H1' and U40 H5 is observed for this pair. As was the case with dpTAR-U, the assignment of the pyrimidine tract was very difficult. For the A-strand, connectivities are determined for G16 through A22 and G26 through C31. No NOEs are detected to C24. As expected for the complex of a TAR RNA with argininamide, there is evidence for stacking of the upper and lower stems. Cross peaks are assigned that correspond to NOEs from G26 H8 to A22 H2'. NOEs also exist between A22 H2 and G26 H1'. dT23 appears to be unstacked. An NOE is seen from dT23 H1' to either A22 H8 or G26 H8. These resonances are nearly overlapped in this spectrum. Data were acquired at shorter mixing times (150 ms, 500 MHz; 75 ms, 750 MHz), lower temperature (15 °C, 500 MHz, 400 ms) and higher temperature (35 °C, 500 MHz, 400 ms) in an effort to determine the identity of this NOE. It appears to be due to the dT23 H1' to G26 H8 interaction, although this is a tentative assignment. Evidence for the location of dT23 in the major groove of the upper stem comes from NOEs between dT23 H1' and G26 H1' and H3'. Cross peaks corresponding to these NOEs are assigned in NOESY spectra taken with 400 and 150 ms mixing times at 500 MHz and with a 75 ms mixing time at a field strength of 750 MHz . The NOE from dT23 to G26 H3' is still of medium intensity in a 150 ms experiment (500 MHz), indicating that these protons are close in space. An NOE of medium strength is seen in 400 ms experiments between A27 H8 and dT23 H1'.

The dT23 H2', H2" and methyl protons resonate far upfield of other H2' protons (2.57, 2.68, and 1.71 ppm, respectively). It was therefore possible to observe NOEs from

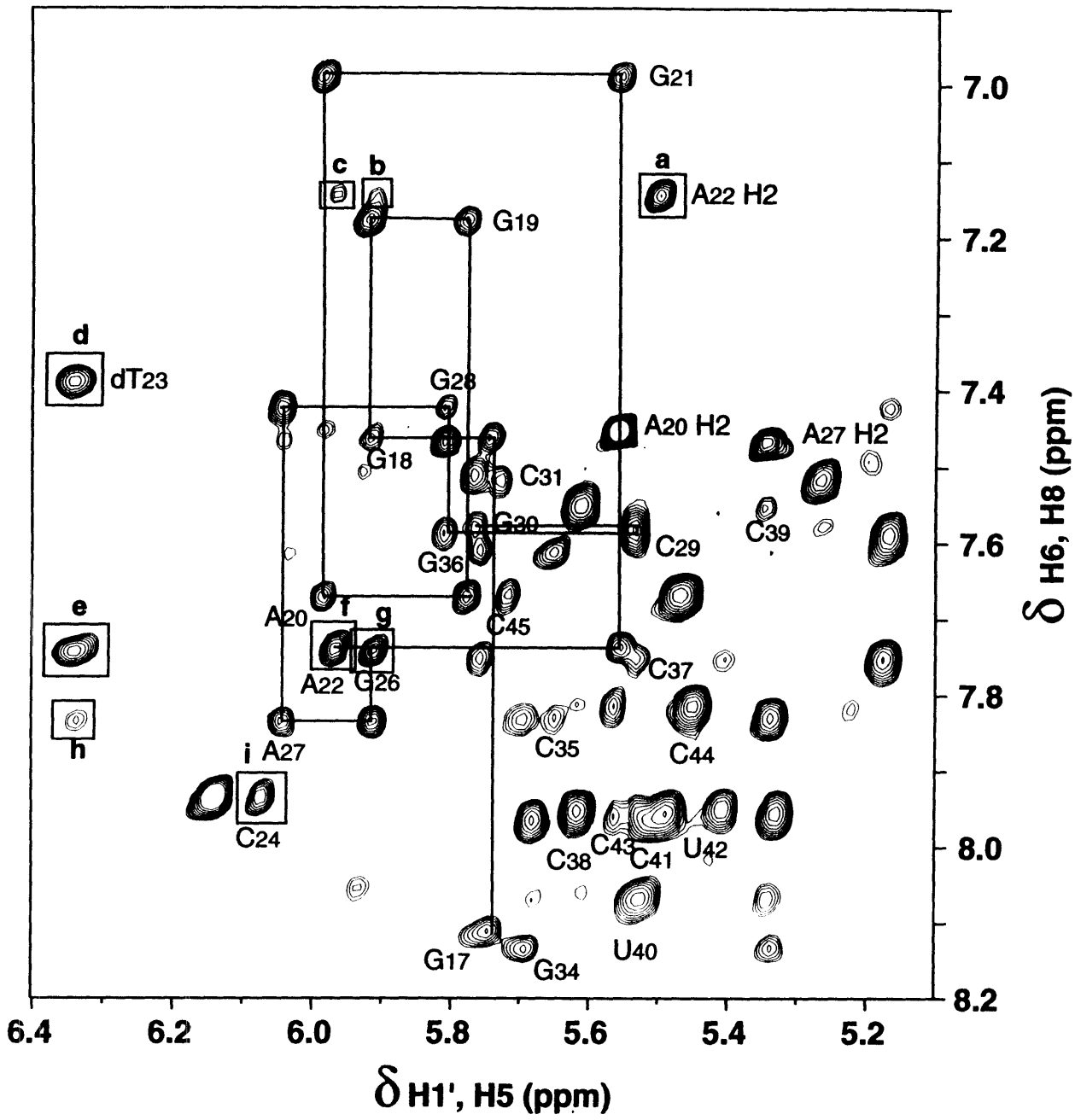


Figure 5.7) dpTAR-T + 6 mM argininamide 400 ms NOESY. The connectivity pathways for the A-strand (G17 through A22 and G26 through C31) are indicated by the solid line. The presence of argininamide has caused unstacking of C24 (**i**). dT23 is now completely unstacked. A22 H2 gives NOEs to C41 (**a**), G26 (**b**) and to its own H1' (**c**). No NOEs are seen to dT23 H6 except one to the intraresidue H1' (**d**). dT23 H1' gives an NOE at ca. 7.74 ppm. This NOE is either to A22 (box **f**) or to G26 (box **g**). Box **h** contains an NOE from dT23 H1' to A27 H8. Data were acquired at 25 °C and 500 MHz.

these protons to other nucleotides. In 400 ms experiments, NOEs are detected to a proton resonating at 7.74 ppm. As discussed above, the G26 H8 and A22 H8 resonances overlap and are located at this chemical shift. A cross peak observed at 6.34 and 7.74 ppm is tentatively assigned to the NOE between dT23 H1' and G26 H8. It is likely that the upfield NOEs also correspond to an interaction between the dT23 2', 2", and methyl protons with the G26 base. This molecule appears to adopt a conformation similar to that observed with the wild-type RNA.

Assignments for dpTAR-T (+ argininamide) are listed in Table 5.2. The chemical shift differences between the free and bound RNA are listed in Table 5.3. These data are also depicted graphically in Figure 5.8. The pattern of chemical shift changes for A22 H2 and the bulge nucleotides is very similar to the wild-type and tTAR-UAC data discussed above. In fact, the A22 H2 proton experiences an even larger upfield shift in dpTAR-T (0.45 ppm vs. 0.37 ppm for tTAR-UAC at pH 5.5). The bulge nucleotides in this RNA also experience the expected downfield shift.

A comparison of the H1'-H2' region of DQF-COSY spectra for the free and bound RNAs is shown in Figure 5.9. In the free RNA, the 5' termini (G17 and G34) contain some C2'-*endo* character. C24 and A22 also give strong COSY peaks in this region. Cross peaks corresponding dT23 H1' to H2' and H2" are also apparent. The argininamide spectrum has fewer peaks. The only strong peaks result from C24 and G34. The A22 H1'-H2' resonance is gone. This nucleotide is stacked at the junction of the two coaxially stems in the bound conformation of TAR, and it exhibits the C3'-*endo* conformation expected of a helical ribonucleotide. Surprisingly, the dT23 H1'-H2' and H2" peaks are significantly weaker in the argininamide sample. Conformational flexibility of the unstacked deoxyribose may account for the lack of detectable cross peaks for this nucleotide.

A schematic representation of the NMR data for the argininamide sample is presented in Figure 5.10. The global conformation is of an A-form RNA helix, consistent

Table 5.2
Assignments for dpTAR-T + 6mM Argininamide (pH 6.4)
Chemical shifts are reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino	Amino
G17	8.11	NA	5.75	4.86		
G18	7.46	NA	5.92	4.67	12.71	
G19	7.18	NA	5.78	4.60	12.23	
A20	7.67	7.45	5.98	4.73	NA	
G21	6.99	NA	5.56	4.47	12.56	
A22	7.74	7.14	5.97	4.29	NA	
T23	7.39	1.71 (CH_3)	6.34	2.57', 2.68"		NA
C24	7.93	6.14	6.08	4.53	NA	
G26	7.74	NA	5.91	4.72	12.43	
A27	7.83	7.47	6.05	4.82	NA	
G28	7.42	NA	5.81	4.53	13.50	
C29	7.59	5.17	5.53	4.53	NA	6.71/8.49
G30	7.58	NA	5.77	4.44	13.05	
C31	7.51	5.27	5.73		NA	
G34	8.13	NA	5.69	4.01		
C35	7.83	5.34	5.65	4.66	NA	6.82/8.63
G36	7.61	NA	5.76	4.54	13.01	
C37	7.75	5.18	5.53	4.33	NA	6.94/8.74
U38	7.95	5.41	5.63	4.78	14.22	NA
C39	7.55	5.61	5.35	4.49	NA	7.23/8.35
U40	8.07	5.53	5.53	4.42	14.39 *	NA
C41	7.96	5.68	5.50	4.19	NA	7.11/8.45
U42	7.95	5.33	5.48	4.45	13.95	NA
C43	7.96	5.61	5.56	4.32	NA	7.00/8.36
C44	7.82	5.45	5.46	4.25	NA	6.93/8.45
C45	7.66	5.47	5.72		NA	

NA refers to not applicable protons, blank spaces indicate protons that were not assigned, * denotes a tentative assignment at 5 °C.

Table 5.3
Chemical shift differences for dpTAR-T, Free data -Argininamide data (pH 6.4).

Δδ reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino
G17	+0.03	NA	+0.04	+0.03	
G18	0.0	NA	+0.01	+0.01	-0.01
G19	+0.01	NA	+0.01	+0.02	+0.01
A20	+0.01	+0.05	+0.01	+0.02	NA
G21	-0.01	NA	-0.05	+0.02	+0.14
A22	-0.07	+0.45	+0.01	+0.16	NA
dT23	+0.08	-0.13 (CH_3)	-0.17	-0.31 2', -0.19 2"	NA
C24	-0.07	-0.21	-0.17	-0.05	NA
G26	-0.05	NA	-0.21	+0.08	+0.04
A27	+0.03	+0.02	-0.01	-0.04	NA
G28	-0.18	NA	-0.11	-0.05	-0.11
C29	-0.01	+0.02	+0.11	+0.02	NA
G30	+0.01	NA	+0.01	0.0	+0.01
C31	+0.02	+0.02	+0.01		NA
G34	+0.02	NA	+0.04	0.0	
C35	+0.01	+0.05	+0.03	0.0	NA
G36	+0.01	NA	+0.02	+0.02	+0.02
C37	-0.04	+0.07	+0.01	+0.09	NA
U38	-0.05	+0.03	-0.03	-0.12	-0.05
C39	+0.17	+0.05	+0.18	+0.06	NA
U40	-0.11	-0.13	-0.01	+0.07	
C41	+0.01	+0.11	-0.01		NA
U42	+0.01	+0.32	-0.07		+0.05
C43	-0.01	+0.04	+0.03	+0.02	NA
C44	0.0	+0.02	+0.01	0.0	NA
C45	+0.02	0.0	+0.02		NA

NA refers to not applicable protons, blank spaces indicate protons that were not assigned in one or both data sets.

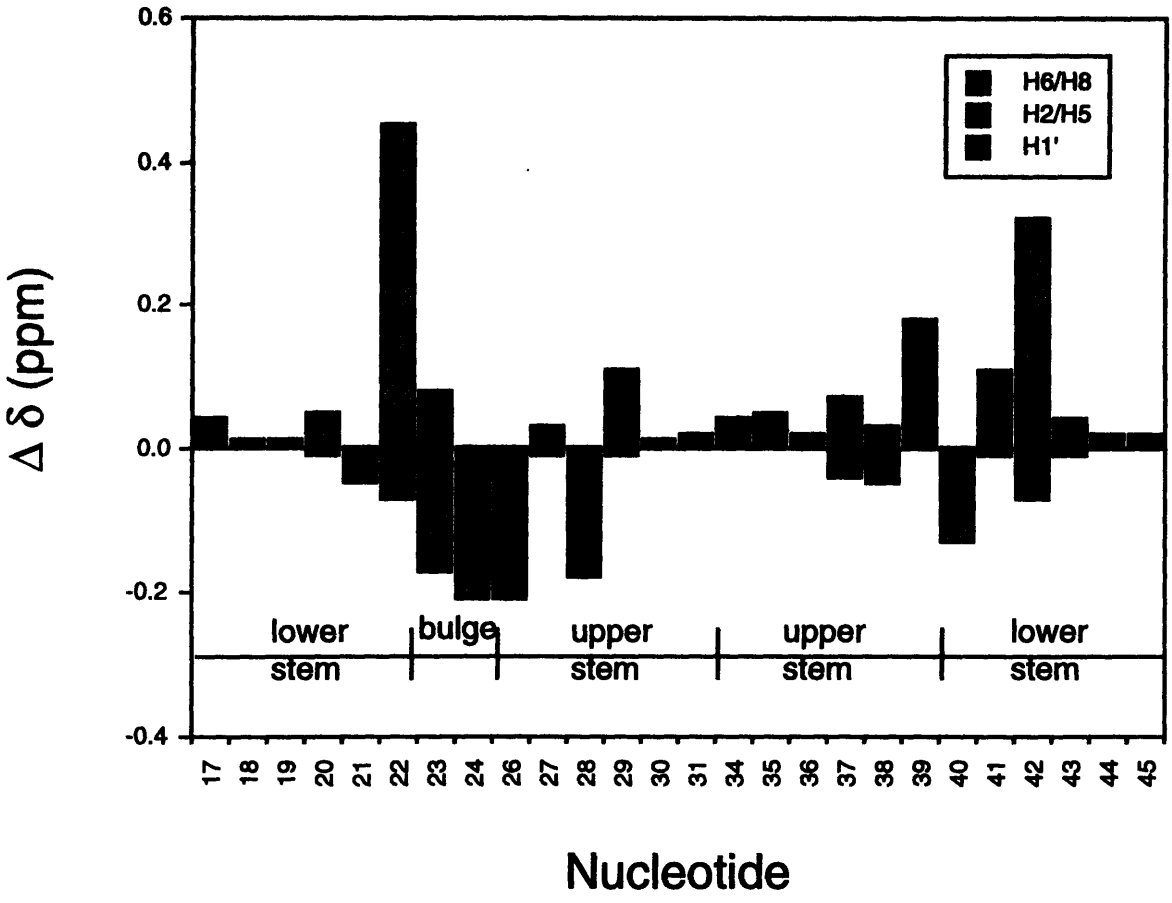


Figure 5.8) Plot of changes in chemical shift for dpTAR-T resonances upon addition of 6 mM argininamide (pH 6.4). Nucleotide position, stem, bulge, and loop regions are marked. The H8/H6 proton chemical shift changes are indicated by solid bars, H2/H5 protons by hatched bars, and H1' protons by shaded bars. Large changes in chemical shift are observed for the bulge nucleotides as a result of argininamide binding.

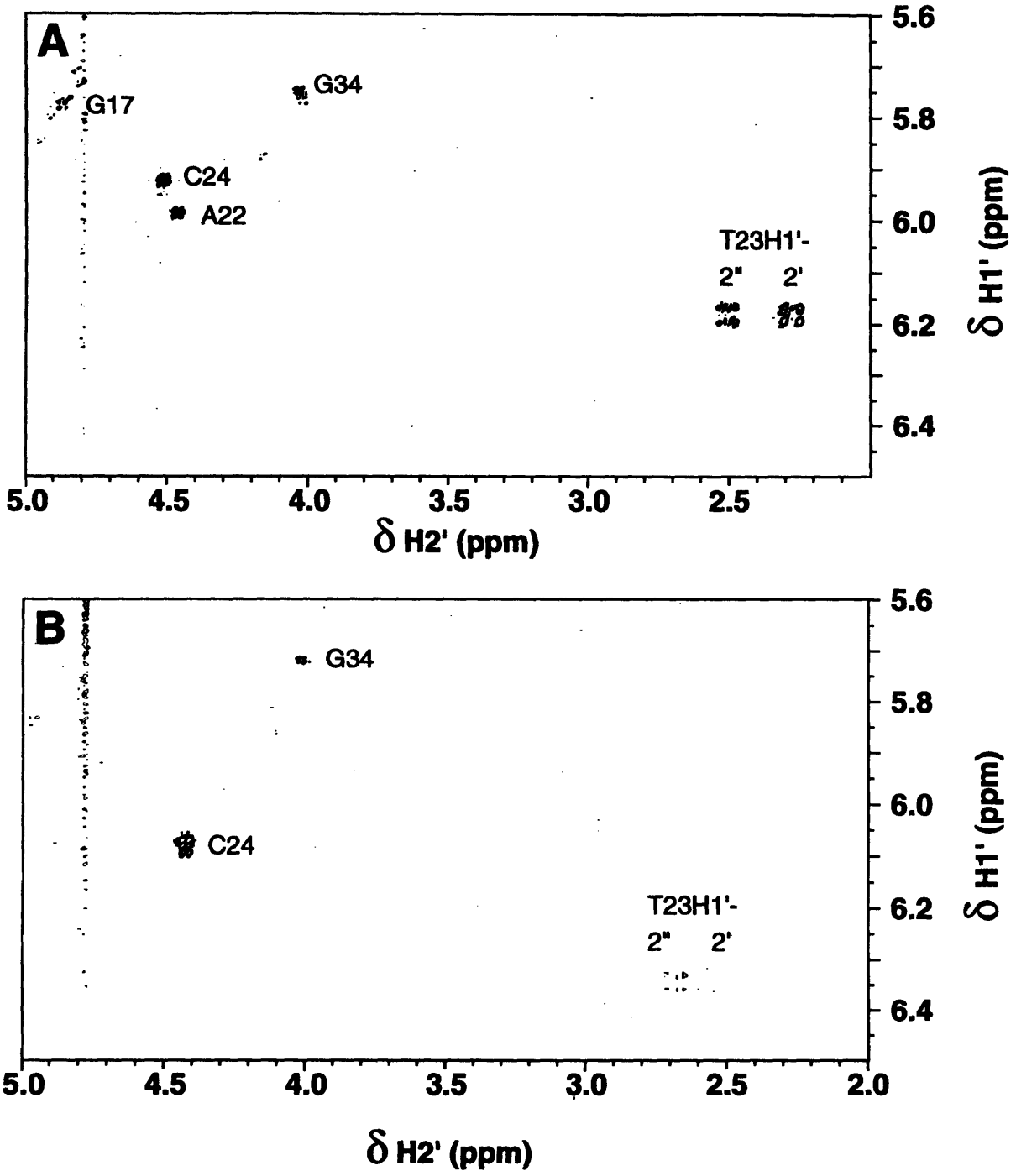


Figure 5.9) Comparison of dpTAR-T COSY spectra. The H1'-H2' region of the DQF-COSY spectrum of free dpTAR-T is shown in **A**. The 5' terminal nucleotide of each strand displays some conformational flexibility. C2'-*endo* character is observed for the nucleotide just below the bulge (A22) and the bulge nucleotide (C24). Deoxythymidine H1' to H2' and H2" crosspeaks are observed. The argininamide spectrum is shown in **B**. A22 no longer gives a crosspeak, indicating a more stacked environment for this nucleotide in the argininamide sample. Surprisingly, the presence of argininamide has weakened the dT23 crosspeaks.

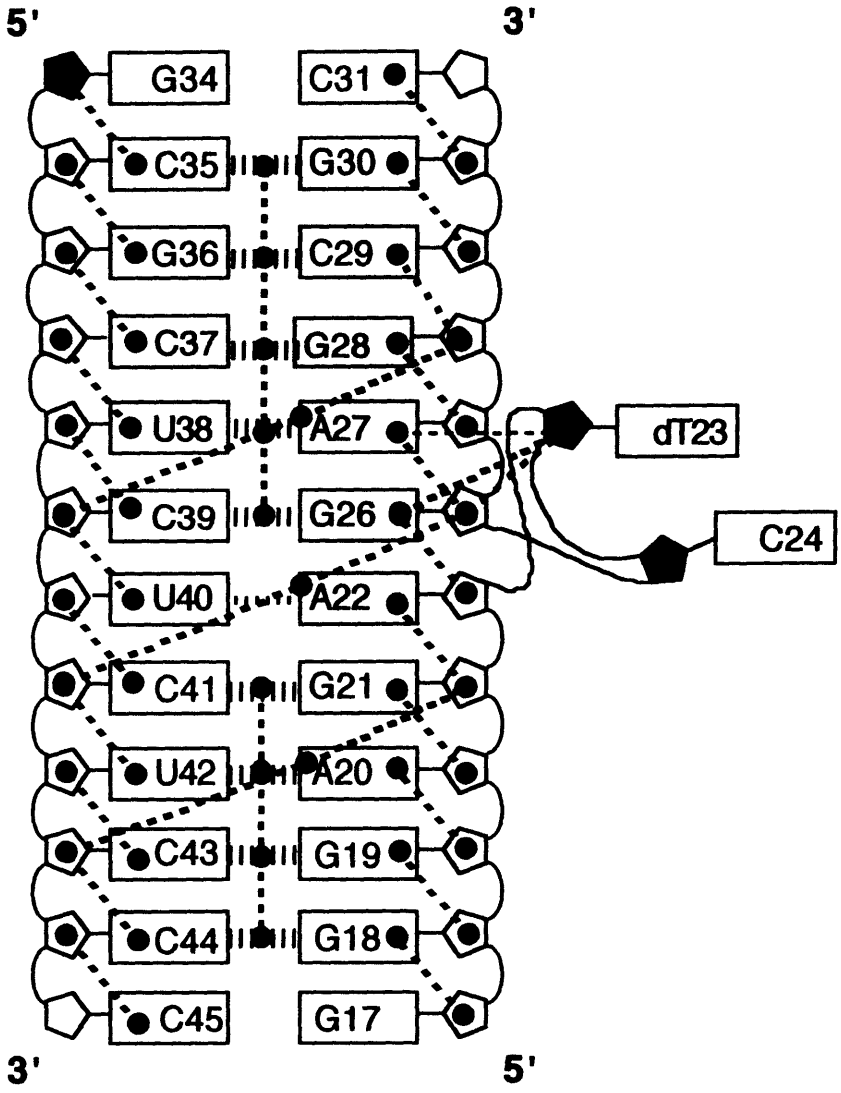


Figure 5.10) NMR summary for dpTAR-T + 6 mM argininamide. Symbols are as in Figures 3.9 and 3.16. A-form helical character is present in both strands of the duplex. NOEs are seen from A22 H2 to C41 H1' and G26 H1'. NOEs from dT23 H1' to G26 H1' and H3' and to A27 H8 locate this nucleotide in the major groove of the upper stem. The thin, horizontal dashed line between U40 and A22 represents hydrogen bonding detected by the presence of the U40 imino proton at low temperatures.

with coaxial stacking of the two stems. The bulge nucleotides are unstacked. No NOEs are detected to C24. A number of NOEs were detected that position dT23 in the major groove of the upper stem, close to A27. Consistent with the wild-type model, the NOE data suggest the formation of a base-triple interaction between dT23 and the A27-U38 base pair.

NOEs to Argininamide

Argininamide interacts with dpTAR-T in a similar manner as with the wild-type RNA. NOESY experiments performed with 400 ms mixing times provide a number of NOEs between the RNA and argininamide (Figure 5.11). At long mixing times, very strong NOEs are observed from the argininamide protons to the methyl group of dT23. The cross peaks corresponding to the methyl group and the argininamide γ proton are difficult to observe because they are obscured by the diagonal. Many of the NOEs from the other argininamide protons are still observed at short mixing times. At 150 ms (500 MHz), and 75 ms (750 MHz) the NOE from the δ proton is very strong, and those to the β and α protons are of medium intensity. In a 50 ms experiment performed at a field strength of 500 MHz, the NOE between the δ proton and the methyl group is still strong. Therefore, the distance between these protons is short, on the order of 2 Å.

NOEs are detected between argininamide protons and dT23 H6 (γ, δ) and a resonance at 7.74 ppm (β, γ, δ) corresponding to either G26 or A22 H8. Weaker cross peaks are observed for the NOEs between the γ and δ protons and A22 H1'. The δ proton also gives a strong NOE to A27 H8. The NOEs observed in different mixing time experiments are listed in Table 5.4.

The structural studies discussed above suggest that the duplex TARs form complexes with argininamide that are similar to that of the wild-type molecule. The chemical shift changes upon addition of the amino acid analog, and the pattern of NOEs

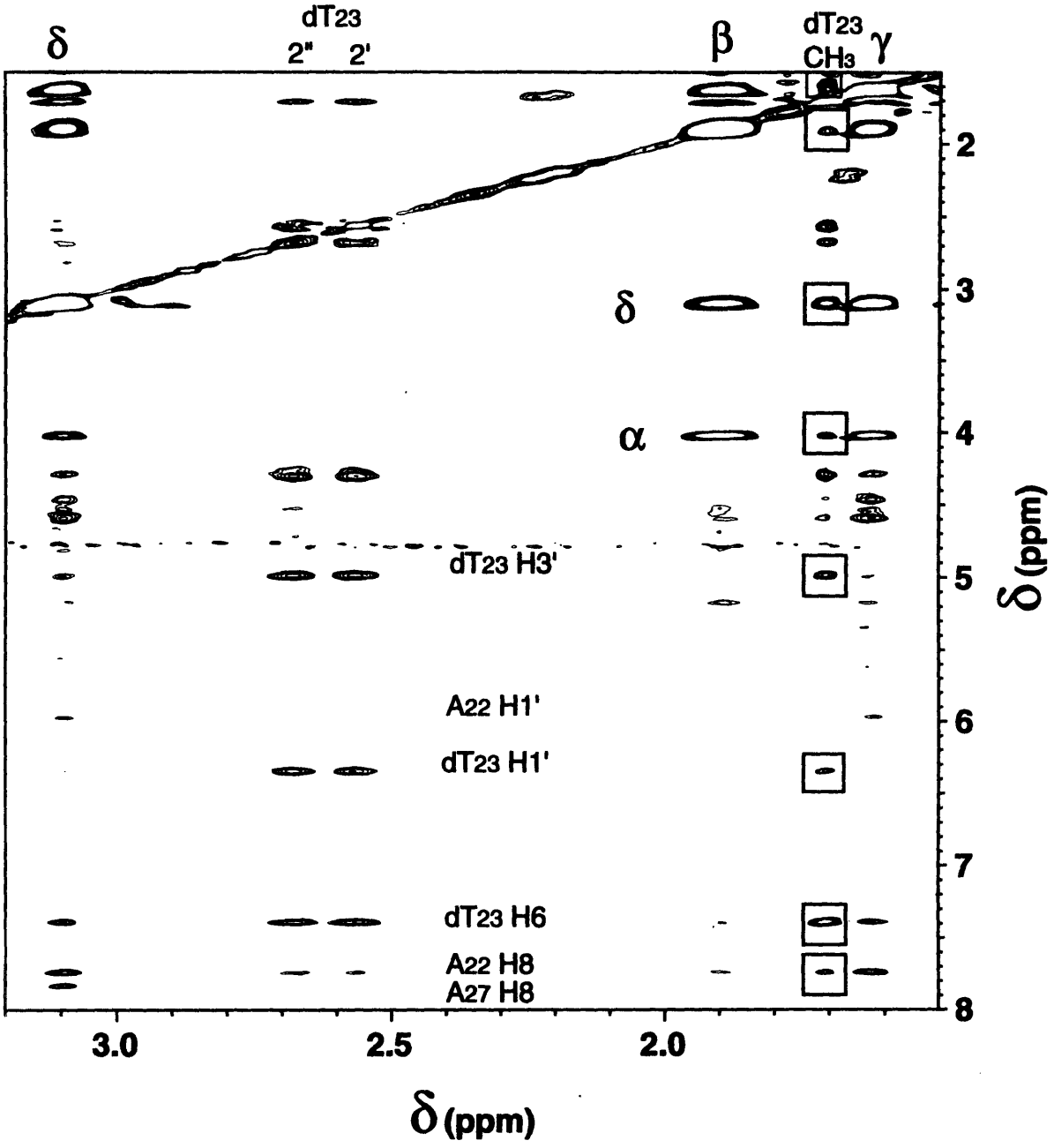


Figure 5.11) Argininamide NOEs to dpTAR-T. The methyl, 2' and 2" protons of deoxythymidine resonate far upfield of most other protons. The NOEs to the methyl group are boxed. Strong NOEs are observed to argininamide, even in short mixing time experiments. The expected TAR NOEs are seen to the argininamide δ proton. These data were collected from a 400 ms NOESY acquired at 25 °C. Buffer was 10 mM Na-phosphate pH 6.4, 100 mM NaCl, 0.1 mM EDTA.

Table 5.4

Argininamide NOEs to dpTAR-T at various mixing times.

	δ (3.10 ppm)				γ (1.63 ppm)				β (1.90 ppm)			
mix (ms) NOE	400	150	75	50	400	150	75	50	400	150	75	50
A22 H8 *	S	S	M	W	S	S	M		Mw			
A22 H1'	M		W		M							
dT23 H6	M	M	M		S	W	M		vW		M	
dT23 CH_3	vS	S	S	S	‡	‡	‡	‡	S	mS	M	M
A27 H8	S	M	M									
dT23 H1'	W											
dT23 H3'	M	vW	M		W				vW			

NOEs are characterized as very strong (vS), strong (S), mediun to strong (mS), medium (M), medium to weak (mW), weak (W), or very weak (vW). Blank spaces indicate the absence of an NOE. NOEs in the shaded box correspond to those reported by Puglisi et al., 1992, except U23 H5 was present instead of the dT methyl group. * may be A22 H8 or G26 H8. ‡ denotes cross peaks that are obscured by the diagonal. 75 ms experiments were acquired with a field strength of 750 MHz.

observed for the complex, suggest the formation of a similar argininamide binding site in the mutant RNAs. The presence of a deoxynucleotide at position 23 of the bulge does not appear to effect formation of the base-triple interaction with the A27-U38 base pair. It would be interesting to perform additional structural studies with TAR RNAs containing modified nucleotides at this position. A uridine analog in which a non-exchangeable proton has been substituted for the imino proton at N^3 of the aromatic ring would be a worthwhile pursuit. NOEs from this proton to the adenine base provide conclusive evidence for the existence of the base-triple interaction. That a deoxynucleotide can be used for these studies allows one to choose from a wide variety of commercially available deoxynucleotides. Additionally, the synthesis of a unique nucleotide designed specifically for this purpose would be facilitated by the presence of a deoxyribose sugar. It is expected that the use of an RNA with the modified nucleotide at only one position will allow for enzymatic synthesis of the large quantities of RNA necessary for structural studies.

5.3 dpTAR-D

In the absence of a uridine nucleotide with a truly non-exchangeable proton at the N^3 position, we chose to study one with a reduced exchange rate (see above). In order to observe the base-triple interaction in TAR, it was our goal to incorporate 5,6-dihydrouridine (D) into the bulge of a duplex TAR molecule. Dihydrouridine has been shown to form a base-triple with a G-C base pair in the crystal structure of *T. thermophilus* $tRNA^{Ser}$ complexed with its cognate aminoacyl-tRNA synthetase (Biou et al., 1994). Experiments were undertaken to study the solution conformation of TAR containing dihydrouridine in the base-triple. Incorporation of this nucleotide should provide detectable differences between the NMR spectra of the free RNA and that complexed with

argininamide. Preliminary exchangeable NMR analysis of free dihydrouridine-monophosphate (Sigma) indicated that the imino proton for this nucleotide is observable even in the absence of base pairing. One resonance of medium line width is observed at approximately 10 ppm (data not shown). The peak was shown to sharpen with decreasing pH.

Recently, the solution conformations of small, synthetic RNAs containing DHU have been analyzed (Narwot et al., 1995, Dalluge et al., 1996). Because of the difficulty in synthesizing the DHU phosphoramidites, the RNAs analyzed were limited to three nucleotides in length. Clearly, it was not desirable to produce the dihydrouridine-containing TAR synthetically. Previous studies with *E.coli* RNA polymerase suggested that DTP could be incorporated specifically into RNA, albeit at low yields (Roy-Burman et al., 1965, Roy-Burman et al., 1967). The efficiency of incorporation into large DNA templates was measured to be only 4% of that measured for UTP. Given our success with dTTP incorporation by T7 RNA polymerase (see above) and because we required incorporation of DHU into only one position, it was worthwhile to attempt an enzymatic synthesis the 14 nucleotide A-strand of the duplex TAR using T7 template-driven transcription.

Synthesis of DTP

Hydrogenation of UTP was performed essentially according to published procedures (Cohn & Doherty, 1956). The reaction was performed twice. The first time only 100 mg of UTP was reacted. Once the success of the reaction had been established, an additional 500 mg of UTP were reacted.

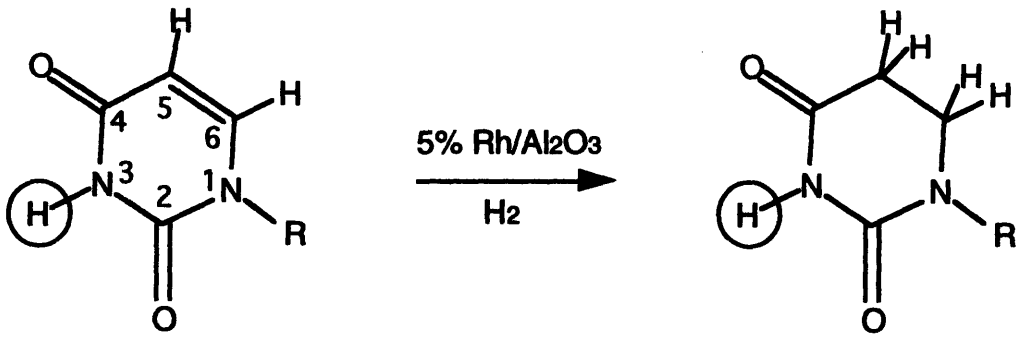
UTP (100 mg) was dissolved in 15 mL of cold, dilute aqueous HCl (pH 3.5). The reaction was stirred continuously and kept in an ice bath. Argon gas was bubbled through the solution for 10 minutes to purge the reaction of any oxygen. 35 mg of catalyst (5%

Rhodium on alumina, of unknown age and origin was kindly provided by S.J. Lippard) was then added followed by 5 additional minutes of purging with argon. The reaction was initiated by bubbling hydrogen gas through the solution. The hydrogen was passed through the solution continuously at atmospheric pressure. Reaction progress was monitored by the decrease in UTP absorbance at 260 nm. This was accomplished by centrifuging 20 μL of the reaction in an eppendorph tube to spin down the catalyst, and appropriately diluting the sample with water to obtain a reasonable absorbance reading. The reaction course is shown in Figure 5.12. The absorbance at 260 nm for the 100 mg reaction had decreased to 0.06 after 4.5 hours. The reaction was then judged to be complete. Argon gas was then bubbled through the reaction for 10 minutes to remove the hydrogen gas. The reaction mixture was then centrifuged at 4 °C at 14,000 rpm for 30 minutes to spin down the catalyst. The supernatant was filtered through a 0.2 μm filter (Schleicher and Schuell). The large scale reaction was treated similarly except that the reaction volume was 20 mL and it took approximately 11 hours to reach completion.

The reaction product was analyzed via NMR prior to performing trial transcriptions. Spectra in 90% H_2O / 10% D_2O indicated the presence of a single imino resonance at 9.9 ppm (data not shown). The sample was then exchanged into 99.996% D_2O. The non-exchangeable spectrum is compared with UTP (Sigma) and DMP (Sigma) in Figure 5.13. The absence of a downfield aromatic resonance, and the presence of upfield methylene protons confirmed that reduction of the 5-6 double bond had occurred. ^{31}P NMR confirmed that the product was in the triphosphate form (Figure 5.14).

Trial transcriptions were performed as described in Materials and Methods. The concentration of the DTP stock solution was not accurately determined. It was assumed that the reduction reaction had gone to completion and that no product was lost. The product was diluted with ddH_2O to a pseudo-concentration of 100 mM. The concentration of the stock solution is actually lower than this. Large scale reactions are based on relative

A



B

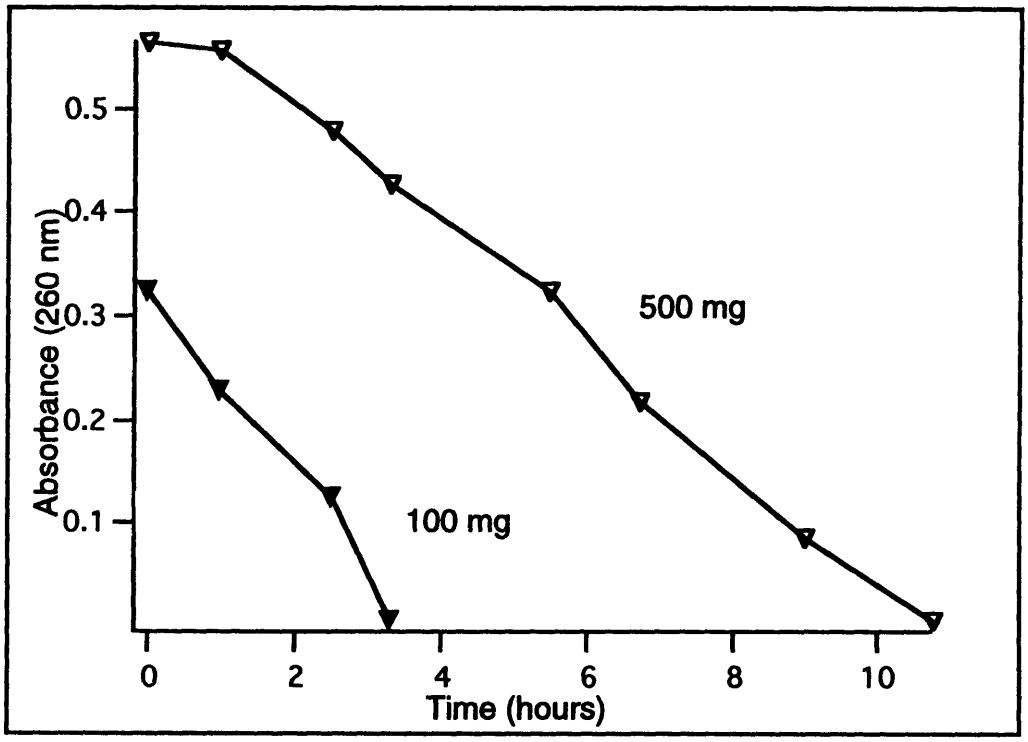


Figure 5.12) Reduction of UTP. Dihydrouridine 5'-triphosphate was synthesized as shown by equation A (R = ribose triphosphate). UTP was reacted with hydrogen gas in the presence of a rhodium catatlyst. The pKa of the imino proton (circled) increases as a result of the reduction of the 5-6 double bond. The reaction was monitored by measuring the decrease in absorbance of the starting material (UTP) at 260 nm (B). Filled triangles correspond to time points for the small scale reaction (100 mg), open triangles represent time points for the 500 mg reaction.

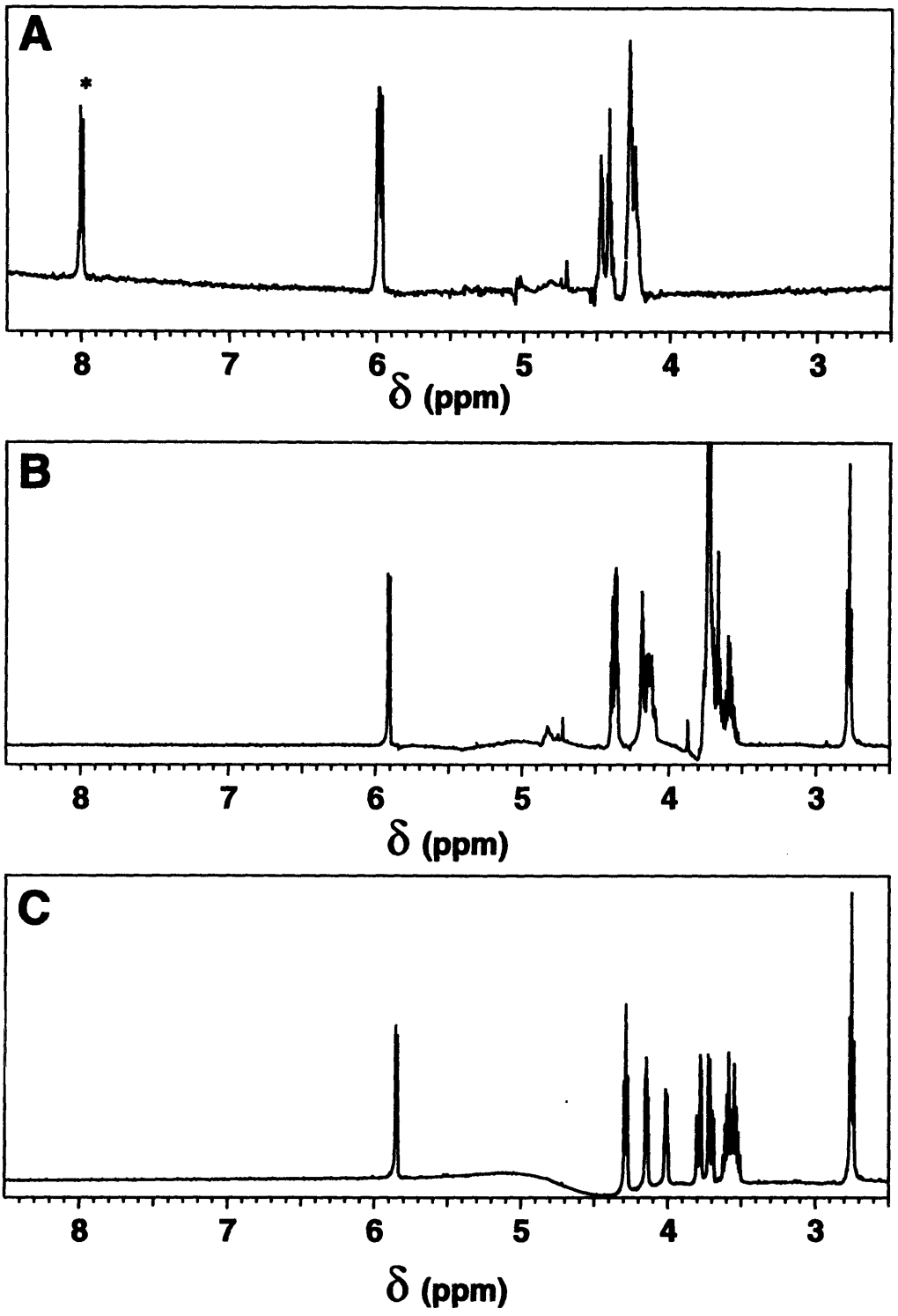


Figure 5.13) D_2O spectra of dihydrouridine species. **A)** Starting material (UTP). **B)** Reaction product. **C)** dihydrouracil. The absence of the aromatic peak (*) and the presence of upfield methylene protons in **(B)** suggest that the reduction reaction was successful. Spectra were recorded at 25 °C.

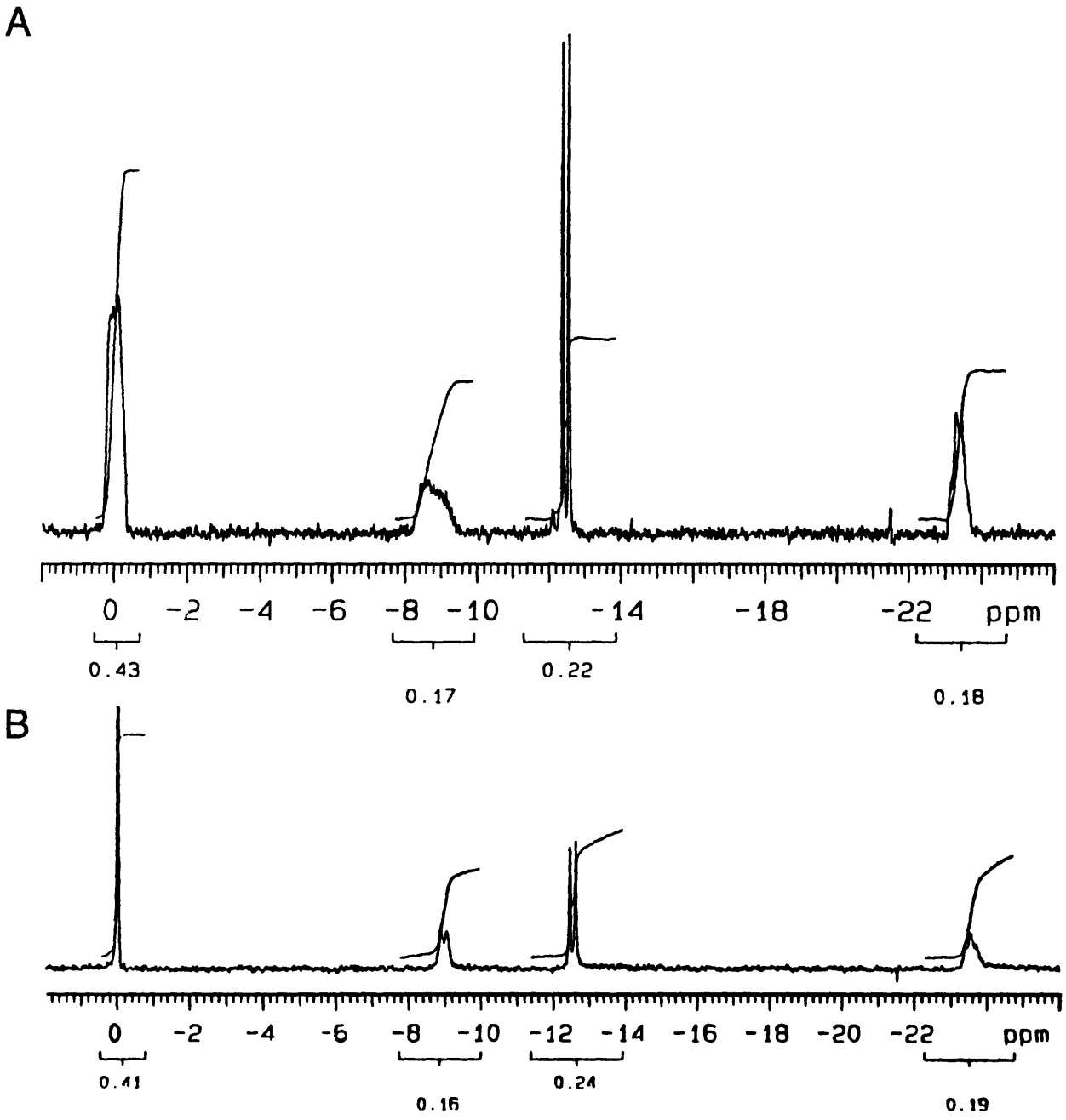


Figure 5.14) ^{31}P NMR spectra of UTP and reduction reaction product. Spectra were acquired in 10 mM Na-phosphate pH 5.5, 100 mM NaCl, 0.1 mM EDTA. Chemical shifts are referenced to the phosphate buffer (0 ppm). A comparison of the phosphorus spectrum of the reaction product (B) with UTP (A) indicates that the reaction product is still in the triphosphate form. Integration values are shown below each peak.

volumes used for optimization reactions. Thus, the actual concentration of DTP in the reaction is irrelevant as long as the same stock solution is used for both the trials and the preparative transcription. An example of an optimization experiment for the 14 nucleotide duplex TAR A-strand (see above) is shown in Figure 5.15. End-labeled A-strand from dpTAR-U was run on the gel as a control for product length (Figure 5.15, lane 1). No full-length RNA was observed in reactions without UTP or DTP (Figure 5.15, lane 2). Increasing concentrations of DTP were added for reactions run in lanes 3 through 9. The yields for reactions containing DTP at a concentration of approximately 5 mM were estimated to be equivalent to those performed with UTP at 4 mM. $MgCl_2$ was added at the same concentration used in standard transcriptions (i.e. 36 mM).

Synthesis of dpTAR-D A-strand

A 50 mL transcription produced enough RNA for an NMR sample (see above). The DTP concentration was roughly 5 mM, all other reaction conditions were as described above. Four preparative gels were sufficient for product purification. The B-strand was prepared as for the other dpTAR molecules. Interestingly, the solubility of the A-strand containing DTP is sensitive to salt concentration. During the dialysis procedure, the RNA remains in solution during the first step (5 mM EDTA, 100 mM NaCl, 10 mM Na-phosphate pH 5.5). During the second step (0.1 mM EDTA, 100 mM NaCl, 10 mM Na-phosphate pH 5.5) the RNA precipitates. Upon dialysis with ddH_2O, the RNA resolubilizes.

Formation of dpTAR-D

The titration of the A- and B-strands was monitored via NMR in H_2O. The titration is shown in Figure 5.16. As the B-strand is titrated into the sample, two resonances are seen for DHU. Upon completion of complex formation, one DHU imino resonance is observed at 10.15 ppm. The imino resonances were assigned using data from a NOESY in

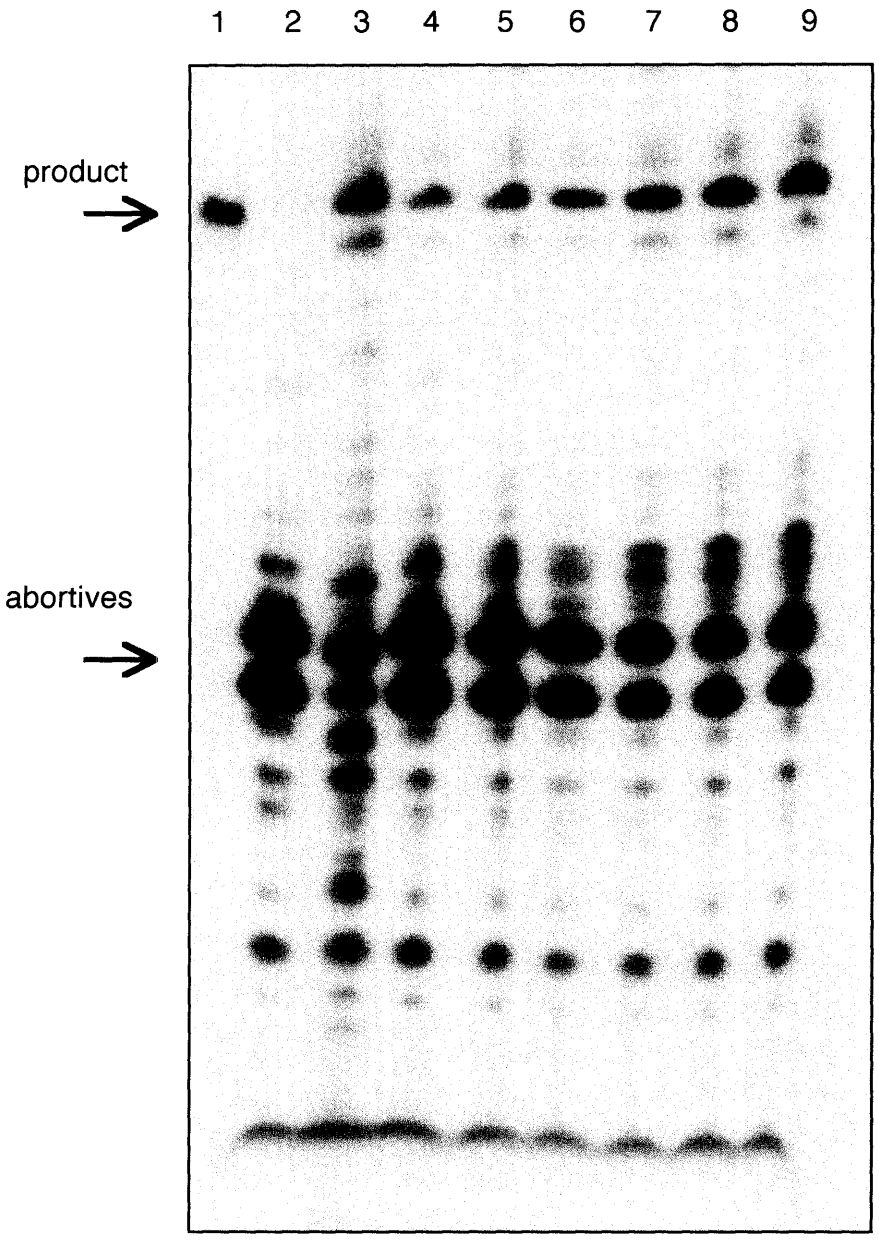


Figure 5.15) Dihydrouridine triphosphate trial transcriptions. One quarter of a 20 μL reaction was loaded into each lane. Lane **1** contains the 14 nucleotide A-strand from dpTAR-U (with a U at position 23). Lane **2** contains a control reaction in which UTP and DTP were omitted under otherwise standard conditions. Lane **3** is a standard transcription reaction (4 mM each ATP, GTP, CTP, UTP). Lanes 4 through 9 show reactions containing increasing amounts of DTP, along with 4 mM each of ATP, GTP and CTP. Concentrations of DTP for reactions in each lane were: lane **4**, 0.5 mM; lane **5**, 1 mM; lane **6**, 2 mM; lane **7**, 4 mM; lane **8**, 5 mM; lane **9**, 6.25 mM. Standard conditions are discussed in Materials and Methods.

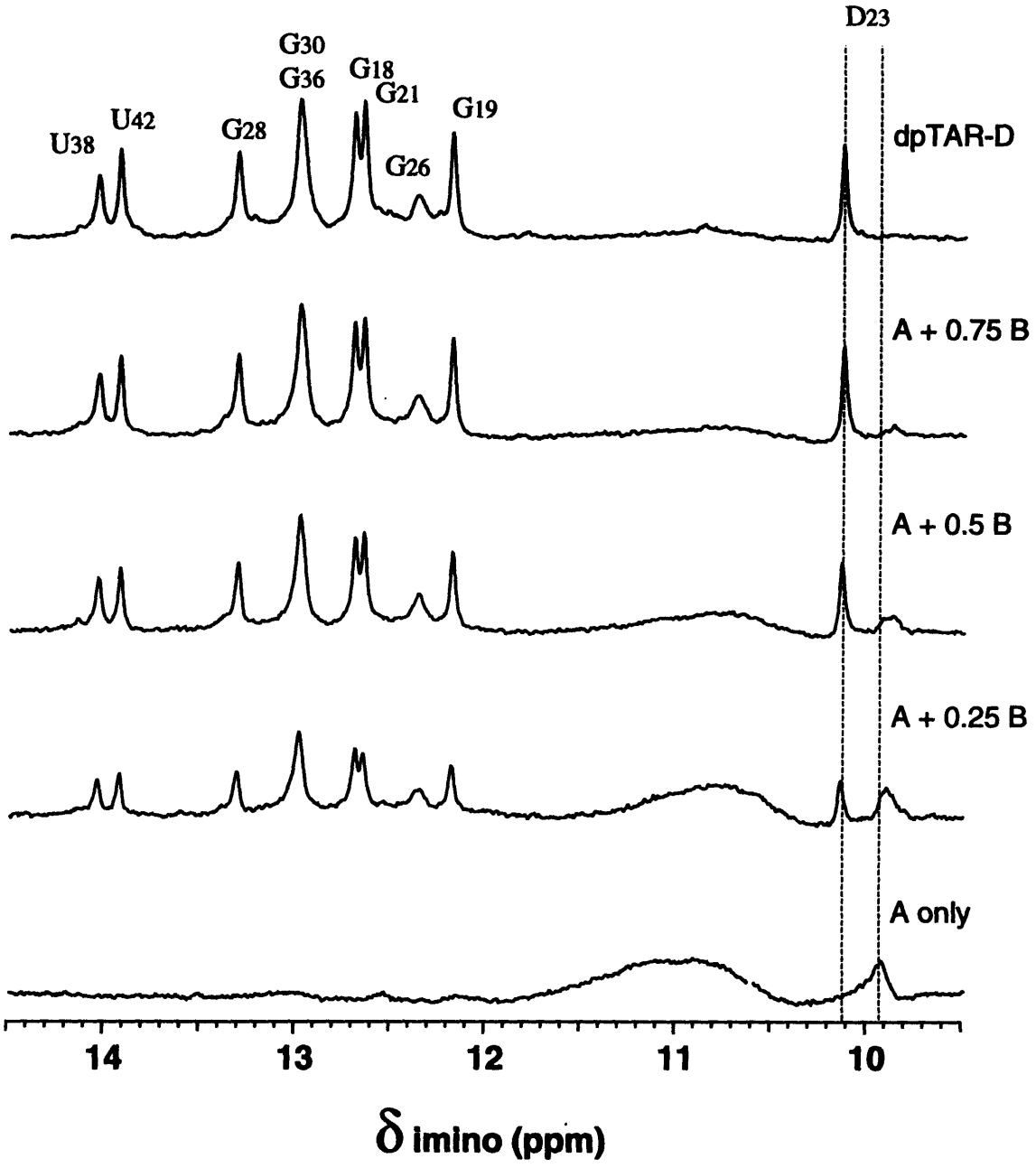


Figure 5.16) Formation of dpTAR-D. Titration of the two RNA strands produces imino resonances indicative of base pairing. The dihydrouridine resonance shifts downfield as the duplex is formed (dashed lines). The G26 imino resonance is significantly weaker than the others. Data were collected at 25 °C in 10 mM Na-phosphate pH 5.5, 100 mM NaCl, 0.1 mM EDTA on a 500 MHz magnet.

water (15 °C). Stacking NOEs are observed for the central four base pairs of the lower stem and the central three of the upper stem. No NOEs are seen to the broad G26 imino proton. The imino proton from the A22-U40 base pair, located below the bulge, is not observed. Nor are the protons from the two terminal base pairs. No NOEs are determined for the dihydrouridine imino resonance. The quality of spectra of the free RNA is poor, as the resonances are very broad and there is a considerable amount of overlap. Experiments in D_2O were largely unsuccessful. The free RNA was not further characterized.

Argininamide Titration

The titration of argininamide into the dpTAR-D sample is shown in Figure 5.17. Small chemical shift changes are seen for G21, G28, G30 and G36. The D23 imino proton shifts 0.5 ppm downfield at a final argininamide concentration of 6 mM. It was possible that this shift was indicative of hydrogen bonding of base-triple formation caused by the presence of argininamide. However, a NOESY in water at 15 °C did not provide any NOEs to this peak. Decreasing the temperature of the sample causes an additional downfield shift of the D23 resonance. Unfortunately, this resonance broadens considerably at low temperatures. An argininamide titration to a final concentration of 9 mM caused an increase in the downfield shift of this proton, as well as an increase in its line width. A number of different temperature, pH and salt conditions were explored. Despite the synthesis of three different samples, no suitable spectra were obtained for characterization of this molecule. The D23 imino proton shifts downfield upon the addition of argininamide to the sample. Thus, the 1-dimensional data suggest that the molecule undergoes an argininamide-induced change in conformation. However, the broadening of the D23 imino resonance also indicates the existence of argininamide-induced conformational exchange.

The structural effects of incorporation of dihydrouridine into an RNA duplex have not previously been studied. Association constant measurements indicate that the efficiency

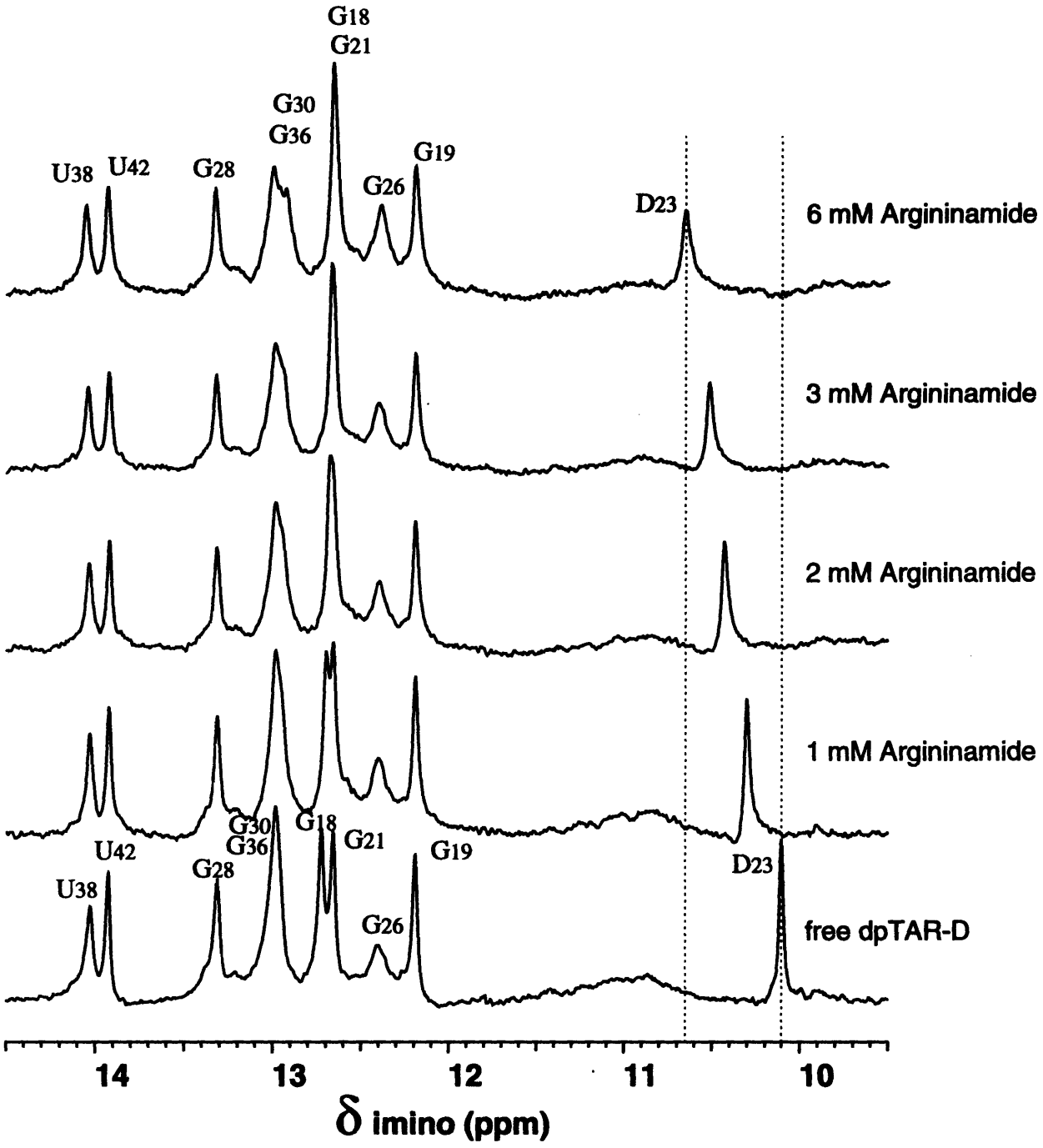


Figure 5.17) Argininamide titration of dpTAR-D at 25 °C. The addition of argininamide causes the dihydrouridine imino resonance to broaden and shift 0.55 ppm downfield. The G26 resonance sharpens, although it is still weaker than the other imino protons. G21 shifts into G18, and the G30, G36 resonance broadens.

of base pair formation of dihydrouracil with adenine is only 30% of that for uracil (Saenger, 1984). Dihydrouridine is also thought to promote conformational flexibility and dynamic motion in tRNAs (Dalluge et al., 1996). This flexibility and a resistance to stacking (Saenger, 1984, Dalluge et al., 1996) combine to prohibit the base-triple interaction and therefore the formation of the argininamide binding site in dpTAR-U.

The crystal structure of *T. Thermophilus* $tRNA^{Ser}$ indicates that this nucleotide is capable of forming of a base-triple interaction with a G-C base pair (Biou et al., 1994). The inability to detect the D23•A22-U38 triple in dpTAR-D may be related to experimental conditions. It is interesting that the A-strand possesses a salt-dependent solubility that differs from the 14-mer containing a U at position 23. Argininamide at a concentration of 6 mM may cause some type of salt-dependent destabilization of the dihydrouridine triple. It would be interesting to study the effect of salt on the solution structure of dpTAR-U in greater detail. The crude gel-shift experiments described in Section 5.1 indicated that dpTAR-D is capable of binding the peptide ADP1-R52. The presence of additional positively charged amino acids in a Tat-peptide might also alter the stability of the bound structure of the mutant RNA. The solution conformation of this RNA in the context of a complex with a Tat-peptide would be an interesting pursuit.

6 Duplex RNAs

The lack of success with the dpTAR-D sample prompted us to investigate the properties of RNAs containing dihydrouridine within the context of an RNA helix. NMR and thermodynamic studies have recently been performed on short (3 nt) unpaired oligonucleotides containing dihydrouridine (Dalluge et al., 1996). The results of this work suggest that dihydrouridine adopts a C2'-*endo* conformation and promotes destacking in neighboring nucleotides. A study was initiated to examine the structural and thermodynamic effects of a base paired dihydrouridine in a duplex RNA.

Using the procedures described above, three duplexes were prepared. The A-strands of the RNAs are identical to those used in dpTAR-U, dpTAR-T and dpTAR-D. The complementary strand is similar to the B-strand of the duplexes studied, except that two nucleotides have been added across from the bulge, allowing for the formation of 14 base pairs. The sequences of the duplexes are shown in Figure 6.1. The A-strands for all three RNAs were synthesized as described above. The sequence of the template for the B-strand was 5' -GGG AGA TCG AGC GCT ATA GTG AGT CGT ATT A -3'. Standard transcription conditions (see above) produced optimal product yield.

6.1 Thermodynamic Studies

Thermodynamic parameters for double strand formation of the three duplex RNAs were determined from melting curves as described in Chapter 2. RNAs were prepared as outlined in Materials and Methods. Small scale (1 mL) transcription reactions generated

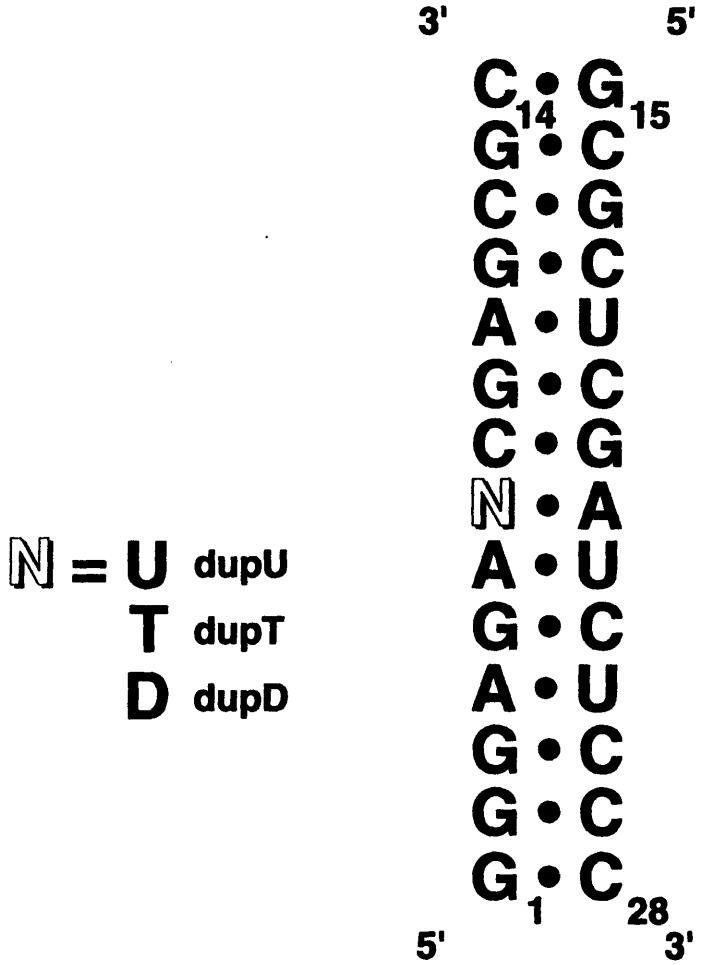


Figure 6.1) Sequences of duplex RNAs. Three 14 base pair RNAs were synthesized for thermodynamic analysis. The A-strand contains either a uridine, deoxythymidine or a dihydrouridine at position 7. The B-strand is similar to that used for duplex TAR structural analysis except that two base pairs are present in what formerly was the bulge region. Numbering is no longer consistent with that used for the TAR molecules, and is as indicated.

enough RNA for thermodynamic studies. In a separate tube for each duplex, equimolar amounts of the A- and B-strands were heated to 90 °C for 1 minute and allowed to cool slowly over 20 minutes in 10 mM Na-phosphate pH 5.5, 100 mM NaCl, and 0.1 mM EDTA. The annealed duplexes were then purified on a non-denaturing 20% polyacrylamide gel at 4 °C to remove any uncomplexed monomer. The RNAs were eluted, ethanol precipitated and dialyzed for at least 18 hours in a buffer containing 10 mM Na-phosphate pH 5.5, 100 or 25 mM NaCl, and 0.1 mM EDTA. A total of six melting curves was performed for each duplex at the two different salt concentrations (25 and 100 mM NaCl). The forward and reverse 'melts' were monitored for each experiment. The data are summarized in Table 6.1.

The melting temperature (T_m) obtained for each of the three duplexes is approximately 8.5 °C higher at 100 mM NaCl than at 25 mM NaCl. An increase of 17-19° is expected for each factor of 10 increase in NaCl concentration (Puglisi & Tinoco, 1989). The melting temperatures for dupU and dupT are similar at both salt concentrations. Incorporation if a single deoxyribonucleotide into the 28 nucleotide RNA does not appear to effect the stability of the duplex. The dihydrouridine molecule, however, melts at a temperature that is ca. 5 °C lower than the other duplexes under both sets of conditions. The decrease in T_m for dupD is reflected in a more unfavorable (more positive) ΔG° (37 °C). The free energy of destabilization of the dihydrouridine duplex ($\Delta\Delta G^\circ$ [37 °C]) is approximately 1 kcal/mole relative to the U and dT containing duplexes. This translates into an order of magnitude difference in the equilibrium constant for formation of the double stranded dihydrouridine molecule. As mentioned in Chapter 5, association constant of an A-D base pair is only 30% of that of an A-U base pair (Saenger, 1984). The decrease in association constant correlates with the low acidity of the imino proton in D (pK = 11 vs. 9.3 in U). An additional explanation for the decrease in T_m for dupD relative to dupU is the effect of base conformation on stacking of nucleotides in a helix. Saturation of the 5-6

Table 6.1
Thermodynamic parameters for duplex formation.

	Tm (°C)	Δ H° (kcal/mol)	Δ S° [cal/ (mol•K)]	Δ G° (37 °C) (kcal/mol)	ΔΔ H° [c] (kcal/mol)	ΔΔG° (37 °C) [c] (kcal/mol)
dupU[a]	65.4	– 87.2	– 259.6	– 7.4		
dup T[a]	65.5	– 82.3	– 243.1	– 7.0	+ 4.9	+ 0.4
dupD[a]	60.6	– 84.3	– 252.4	– 6.0	+ 2.9	+ 1.4
dupU[b]	74	– 86.1	– 247.5	– 9.3		
dupT[b]	73.7	– 87.1	– 251.1	– 9.3	- 1.0	0
dupD[b]	69.3	– 86.3	– 248.7	– 8.2	- 0.2	+ 1.1

[a] 10 mM Na-phosphate pH 5.5, 25 mM NaCl, 0.1 mM EDTA; [b] 10 mM Na-phosphate pH 5.5, 100 mM NaCl, 0.1 mM EDTA. [c] values are relative to dup U.

bond causes the protons on the now sp^3-hybridized carbons to interfere with each other. A result of this steric interaction is that these carbons pucker 0.3 to 0.7 Å out of the plane of the four remaining ring atoms (Saenger, 1984, Suck et al., 1971). The puckering of the dihydrouridine base is expected to destabilize stacking interactions within a nucleic acid helix (Saenger, 1984, Feldman, 1977). A lack of stacking is consistent with the structural studies performed with short RNA oligos (Dalluge et al., 1996). The presence of dihydrouridine in an unstructured region of tRNAs has led to the suggestion that this nucleotide is a natural 'helix breaker' (Saenger, 1984, Dalluge et al., 1996). The geometry of the base also affects the sugar pucker, shifting the conformation toward C2'-*endo* (Emerson & Sundaralingam, 1980, Dalluge et al., 1996). However, as evidenced by dupT, the presence of a nucleotide containing a C2'-*endo* sugar pucker does significantly not alter the stability of the 14 base pair duplex (Table 6.1). Thus, the destabilization of dupD results from the hydrogen bonding and stacking properties of the dihydrouridine base.

6.2 Structural Studies

Exchangeable Data

An NMR sample of dupD was prepared in order to investigate the structural effects of dihydrouridine within the context of a complementary helix. The A-strand was synthesized as described above (Chapter 5). A 40 mL transcription under standard conditions (Chapter 2) produced a sufficient amount of the 14 nucleotide B-strand for a 1 mM NMR sample (600 µL). Three purification gels were necessary. After dialysis (see above), the A- and B-strands were titrated together in the NMR tube in buffer containing 10 mM Na-phosphate pH 5.5, 100 mM NaCl, and 0.1mM EDTA. Immediately upon addition of 0.25 equivalents of the B-strand, the imino resonance for the dihydrouridine was shifted

into the region of the spectrum containing the other imino resonances. This contrasts with the dpTAR-D titration (Figure 5.16) in which the titration was monitored by the disappearance of the peak corresponding to the single stranded RNA with concomitant growth of a downfield peak corresponding to the duplex TAR molecule. For dpTAR-D, the D23 resonance never shifted into the other imino resonances. The imino spectrum of a 1:1 complex between the 14 nucleotide RNA strands consists of 12 equally strong peaks. The location of the dihydrouridine resonance was not obvious from the one dimensional data.

A WATERGATE NOESY was performed at 15 °C and 500 MHz with a 200 ms mixing time. The region of the spectrum containing the imino-imino NOEs is shown in Figure 6.2. It was surprising to observe stacking NOEs for the entire duplex, with the expected exception of the two terminal base pairs (G1 and G15). Based on the pattern of stacking NOEs, the dihydrouridine imino proton is assigned to a resonance at 12.77 ppm. It gives stacking NOEs to both of the neighboring base pairs, U23 and G21. The NOE between D7 and U23 is the strongest of all the crosspeaks. This is apparent at higher level plots of the spectrum. The identity of the dihydrouridine resonance was confirmed by the imino to amino/AH2 region of the spectrum shown in Figure 6.3. Four imino resonances are shown to have the characteristically strong NOEs to adenine H2 protons. These are assigned to the U23-A6, U25-A4, U19-A10 and D7-A22 base pairs. These crosspeaks appear to be of equal intensity. The remaining imino protons give two NOEs each to cytosine amino protons of their base pairing partners. The relative strength of the D7 / U23 crosspeak is apparent from this plot. The exchangeable data provide evidence for both the formation and stacking of the D7-A22 base pair.

Non-exchangeable Data

The dupD sample was exchanged in to 99.996% D_2O for characterization of the

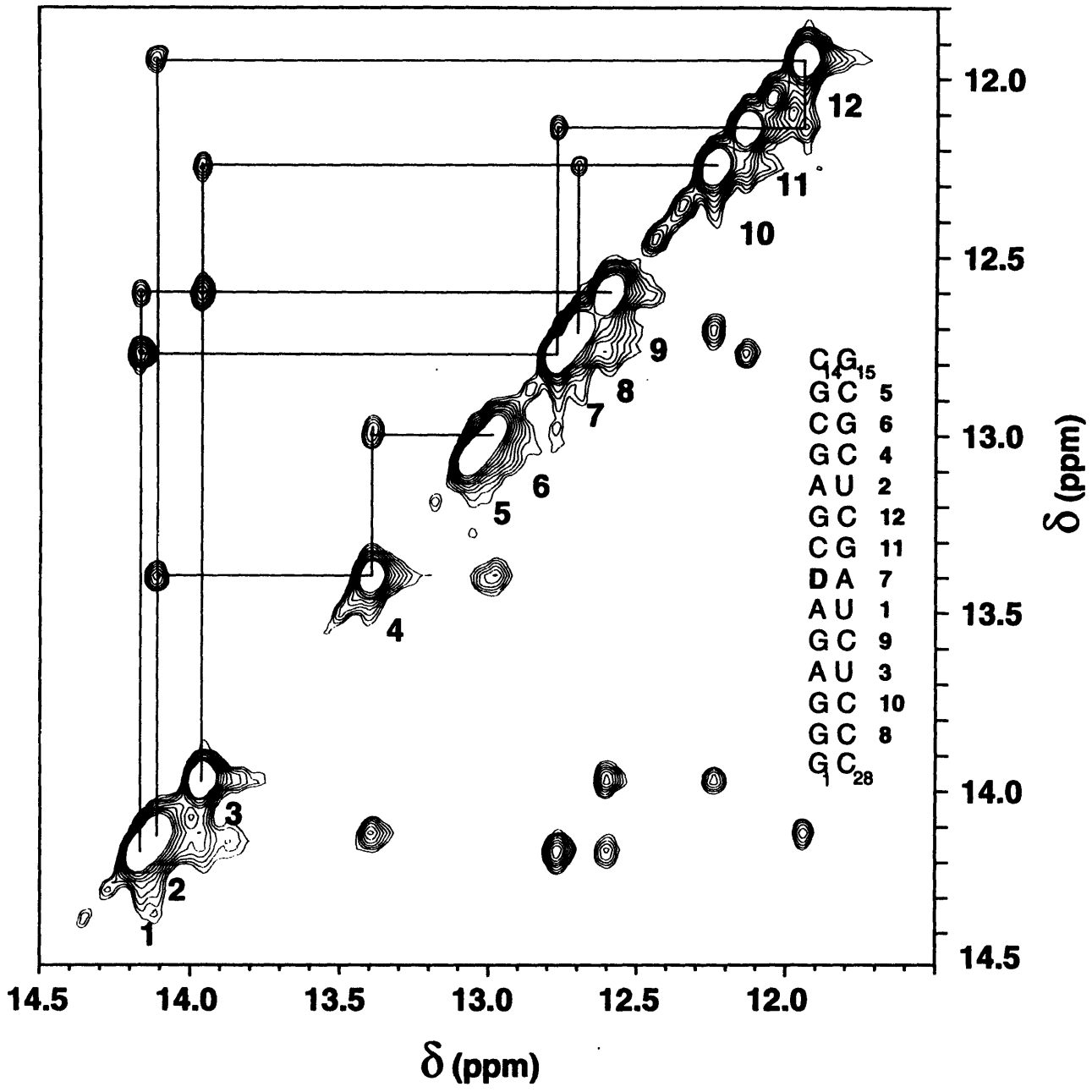


Figure 6.2) Imino-imino NOEs for dupD. Stacking NOEs were observed for all but the two terminal base pairs. Imino protons were assigned to base pairs as indicated by numbers on the RNA. The imino proton from the D-A base pair is peak 7, it gives NOEs to both of the neighboring base pairs (1 and 11). Data were collected at 15 °C in a 200 ms mixing time WATERGATE NOESY at 500 MHz.

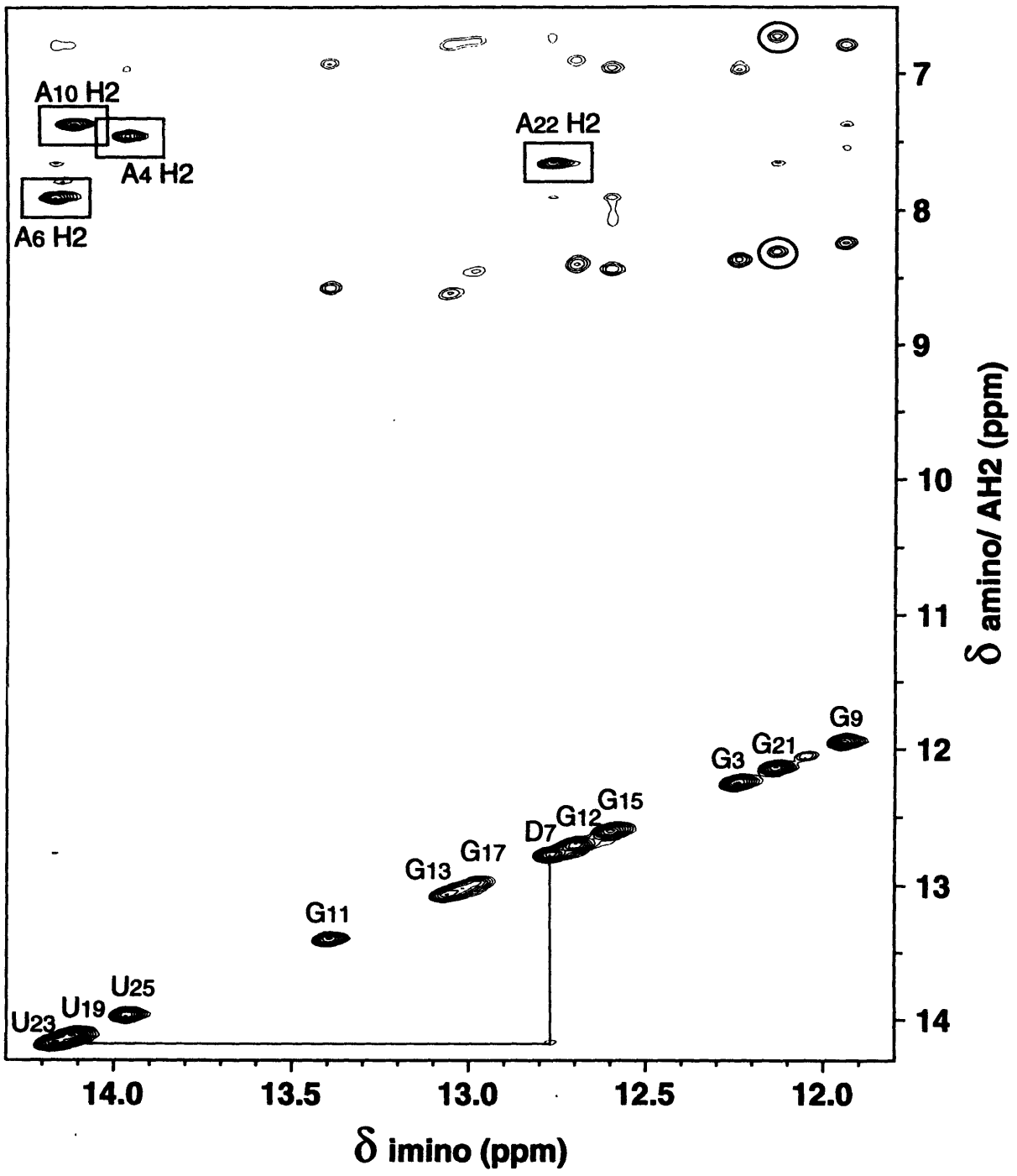


Figure 6.3) Imino to amino and A H2 NOEs for dupD. Characterisitc imino to adenine H2 NOEs were observed for the U23-A6, U19-A10, U25-A4 and D7-A22 base pairs (boxed cross peaks). For comparison, the G21 imino to C8 amino NOEs are circled. Solid lines connect the strong D7-U23 imino-imino NOE. Data were collected in a WATERGATE NOESY with 200 ms mixing time at 15 °C and 500 MHz.

non-exchangeable resonances. SCUBA experiments were performed at 25 °C and 500 MHz with 400 and 50 ms mixing times. The aromatic to H1' and H5 region of the spectrum is shown in Figure 6.4. The four adenine H2 resonances were used as starting points for the assignment of the non-exchangeable spectrum. As was the case with the duplex TAR molecules, the pyrimidine stretch of the B-strand was very difficult to assign. The aromatic, H2, H5, H1', and H2' protons were assigned for the A-strand. The chemical shifts are given in Table 6.2. Normal aromatic to H1' and H2' connectivities were traced for nucleotides G1 through A6 and C8 through C14. The non-aromatic D7 H6 and H5 resonate far upfield and are not located in this region of the spectrum. The A6 H2 proton gives an NOE to a peak at 5.30 ppm. C8 H6 also gives an NOE to a proton at this chemical shift. The peak at 5.30 ppm was identified as D7 H1'. All remaining adenine H2 protons give the expected NOEs.

An interesting feature of this spectrum is the chemical shift of U23 H5. U23 is located one nucleotide 3' to the base pairing partner of D7 (A22). The U23 H5-H6 crosspeak is located at 7.76/4.97 ppm. Pyrimidine H5 protons normally resonate between 5 and 6.3 ppm. The aromatic protons in a stacked helix resonate upfield relative to their position in a single stranded state. The upfield nature of the H5 chemical shift indicates a slight deviation from a normal stacking arrangement.

NOEs to the dihydrouridine base protons are depicted in Figure 6.5. At long mixing times, intraresidue NOEs are seen for both the H5 and H6 protons to the H1' and H2' protons. NOEs are also seen to C8 H5 and A6 H2'. A weak NOE is also detected form the D7 H6 to A6 H1'. The absence of H1'-H2' cross peaks in a DQF-COSY experiment indicated a lack of C2'-*endo* character for D7 or any neighboring nucleotides . This is a surprising result as it the studies with small, unpaired RNA oligonucleotides indicate that at least the dihydrouridine and the nucleotide in the 5' direction (on the same strand) would contain some C2'-*endo* character (Dalluge et al., 1996). It is possible that a

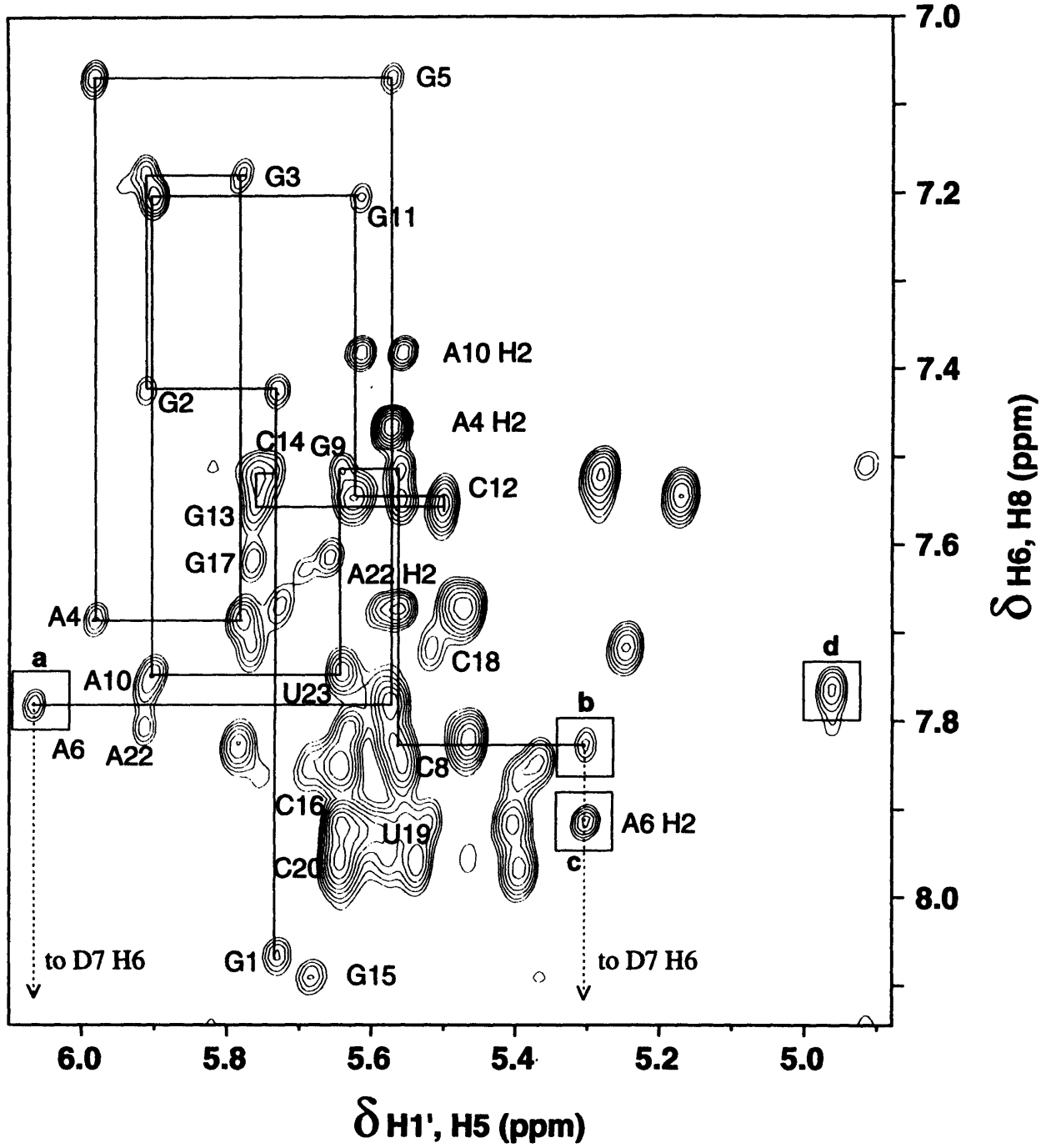


Figure 6.4) 400 ms SCUBA of dupD at 25 °C. The connectivity pattern for the A-strand is indicated by the solid line. The NOE in box **a** is the inter-residue A6 H8 to H1'. Box **b** contains the C8 H6 to D7 H1' cross peak. The A6 H2 to D7 H1' NOE is shown in box **c**. The upfield shifted U23 H5-H6 cross peak is indicated by box **d**.

Table 6.2
Chemical shift assignments for dupD (pH 5.5).
Values reported in ppm.

Residue	H6/H8	H2/H5	H1'	H2'	Imino	Amino
G1	8.06	NA	5.73	4.85		
G2	7.42	NA	5.91	4.65	12.70	
G3	7.17	NA	5.78	4.60	12.25	
A4	7.68	7.46	5.98	4.71	NA	
G5	7.07	NA	5.57	4.54	12.60	
A6	7.78	7.91	6.07	4.46	NA	
D7	2.99	2.28	5.30	4.26	12.77	NA
C8	7.83	5.78	5.56	4.51	NA	6.72/8.30
G9	7.51	NA	5.64	4.57	11.94	
A10	7.74	7.38	5.90	4.63	NA	
G11	7.20	NA	5.62	4.43	13.40	
C12	7.54	5.17	5.50	4.53	NA	6.76/8.45
G13	7.55	NA	5.76	4.44	13.06	
C14	7.52	5.29	5.74		NA	
G15	8.09	NA	5.69	4.74		
C16	7.84	5.37	5.65	4.65	NA	6.78/8.62
G17	7.62	NA	5.79	4.45	12.99	
C18	7.71	5.24	5.51	4.40	NA	6.93/8.58
U19	7.92	5.41	5.53	4.53	14.12	NA
C20	7.85	5.64	5.56	4.52	NA	6.79/8.24
G21	7.55	NA	5.63	4.63	12.14	
A22	7.81	7.67	5.92	4.36	NA	
U23	7.76	4.97	5.59		14.17	NA
C24	7.93	5.61			NA	6.96/8.43
U25	7.86	5.53			13.97	NA
C26					NA	6.97/8.37
C27					NA	6.90/8.39
C28					NA	

NA refers to not applicable protons, blank spaces indicate protons that were not assigned.

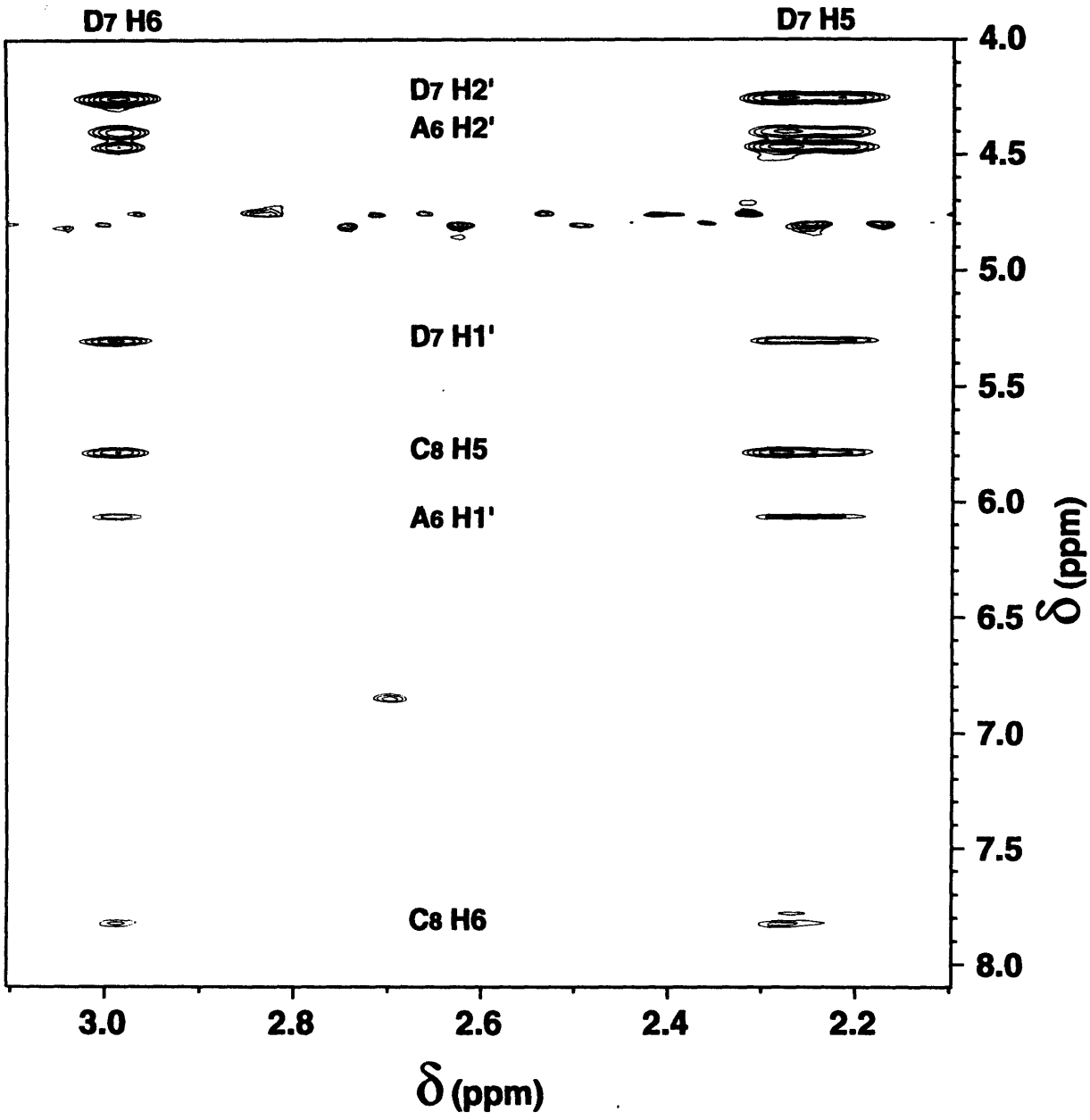


Figure 6.5) NOEs to D7 H6 and H5. The D7 H6 and H5 protons resonate upfield from the aromatic H6 and H5 protons. Intra-residue NOEs are seen to the H1' and H2' protons. Stacking NOEs are seen to C8 H5 and H6. An NOE is also observed to A6 H1'.

weak cross peak was not detected in the COSY experiment. TOCSY experiments are more sensitive to minor ribose C2'-*endo* populations. It would be beneficial to perform a TOCSY with this molecule. This work is currently in progress.

A schematic representation of the preliminary NMR data for dupD is given in Figure 6.6. An overall A-form helical geometry is observed. The exchangeable and non-exchangeable data indicate a slight deviation from A-form geometry in the region of the D-A pair. Additional experiments are necessary to determine the specific conformation of the duplex. The incorporation of a D-A base pair into an RNA helix does cause a slight (ca. 1 kcal/mole) destabilization relative to an U-A or T-A. This destabilization does not appear to be associated with a large conformational change. The strength of the base pairs (evidenced by the association constants with adenine) is apparently responsible for the decrease in stability, thus the lower T_m, of the dihydrouridine duplex.

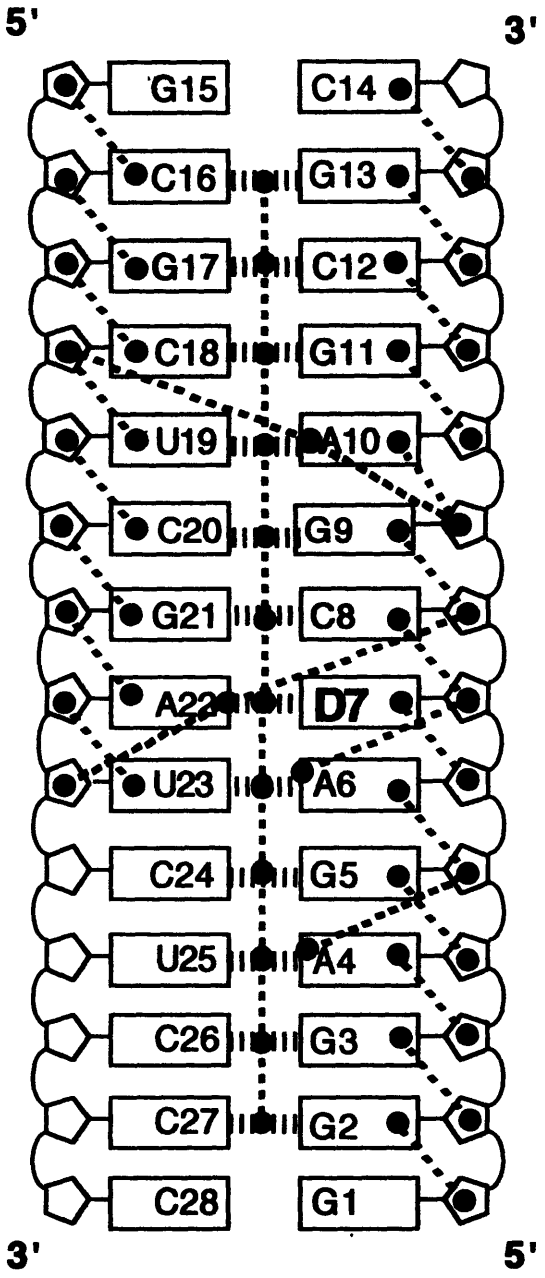


Figure 6.6) NMR summary for dupD. Symbols are as in Figures 3.9 and 3.16. A-form helical character is present in both strands of the duplex. The sugar protons were not assigned for the pyrimidine stretch of the B-strand. Twelve base pairs are observed. Imino-imino stacking NOEs are present in the entire duplex (including D7) with the exception of the terminal base pairs. NOEs are also seen from A22 H2 to C8 H1' and U23 H1', and for A6 H2 to D7 H1'.

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." *Journal of Molecular Biology* 253, 313-332.

Aboul-ela, F., Koh, D., Tinoco, I., Jr. & Martin, F. H. (1985) "Base-base mismatches. Thermodynamics of double helix formation for $dCA_3XA_3G + d\,CT_3YT_3G$ (X,Y = A, C, G, T)." *Nucleic Acids Research* 13, 4811-4824.

Aldovini, A., Debouck, C., Feinberg, M. B., Rosenberg, M., Ayra, S. K. & Wong-Staal, F. (1986) "Synthesis of the complete *trans*-activation gene product of human T-lymphotropic virus type III in *Eschericia coli*: Demonstration of immunogenicity in *vivo* and expression in *vitro*." *Proceedings of the National Academy of Sciences (USA)* 83, 6672-6676.

Allain, F. H.-T. & Varani, G. (1995) "Structure of the P1 helix from group I self-splicing introns." *Journal of Molecular Biology* 250, 333-353.

Altman, S. (1989) "Ribonuclease P: An enzyme with a catalytic RNA subunit." *Adv. Enzymol. Relat. Areas Mol. Biol.* 62, 1-36.

Altona, C. (1982) "Conformational analysis of nucleic acids. Determination of the backbone geometry of single-helical RNA and DNA in solution." *Recl. Trav. Chim. Pays-Bas* 101, 413-434.

Altona, C. & Sundaralingam, M. (1972) "Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation." *Journal of the American Chemical Society* 94, 8205-8212.

Anato, V. P., Lai, S. Y. & Tinoco, I., Jr. (1991) "A thermodynamic study of unusually stable RNA and DNA hairpins." *Nucleic Acids Research* 19, 5901-5905.

Ayra, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. (1985) " *Trans*-activator gene of human T-lymphotropic virus type III (HTLV-III)." *Science* 229, 69-73.

Baeyens, K. J., DeBondt, H. L. & Holbrook, S. R. (1995) "Structure of an RNA double helix including uracil-uracil base pairs in an internal loop." *Nature Structural Biology* B98, 56-62.

Barnett, R. W., Delling, U., Kuperman, R., Sonenberg, N. & Sumner-Smith, M. (1993) "Rotational symmetry in ribonucleotide strand requirements for binding of HIV-1 Tat protein to TAR RNA." *Nucleic Acids Research* 21, 151-154.

Barry, P. A., Pratt-Lowe, E., Unger, R. E. & Luciw, P. A. (1991) "Cellular factors regulate transactivation of human immunodeficiency virus type 1." *Journal of Virology* 65, 1392-1399.

Bartel, D. P. & Szostak, J. W. (1993) "Isolation of new ribozymes from a large pool of random sequences." *Science* 261, 1411-1418.

Bartel, D. P. & Szostak, J. W. (1994) "Study of RNA-protein recognition by *in vitro* selection." in *RNA-Protein Interactions* (K. Nagai and I. W. Mattaj, eds) pp. 248-268, IRL Press, New York, New York.

Basavappa, R. & Siegler, P. (1991) "The 3Å crystal structure of yeast initiator tRNA: Functional implications in initiator/elongator discrimination." *EMBO Journal* 10, 3105-3111.

Baserga, S. J. & Steitz, J. A. (1993) "The diverse world of small ribonucleoproteins." in *The RNA World* (R. F. Gesteland and J. F. Atkins, eds) pp. 359-381, Cold Spring Harbor Press, Cold Spring Harbor, NY.

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, 4515-4523.

Battiste, J. L., Tan, R., Frankel, A. D. & Williamson, J. R. (1994) "Binding of an HIV Rev peptide to Rev responsive element RNA induces formation of purine-purine base pairs." *Biochemistry* 33, 27412747.

Bax, A. D. & Davis, D. G. (1985) "MLEV-17 based two-dimensional homonuclear magnetization transfer spectroscopy." *Journal of Magnetic Resonance* 65, 355-360.

Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R. & Rosch, P. (1995) "Structural studies of HIV-1 Tat protein." *Journal of Molecular Biology* 247, 529-535.

Benight, A. S., Schurr, J. m., Flynn, P. F., Reid, B. R. & Wemmer, D. E. (1988) "Melting of a self-complementary DNA minicircle. Comparison of optical melting theory with exchange broadening of the nuclear magnetic resonance spectrum." *Journal of Molecular Biology* 377-399.

Berkhout, B. & Jeang, K.-T. (1989b) "*Trans*-activation of human immunodeficiency virus type 1 is sequence specific for both the single-stranded bulge and loop of the *trans*-acting responsive hairpin: A quantitative analysis." *Journal of Virology* 63, 5501-5504.

Berkhout, B. & Jeang, K.-T. (1991) "Detailed mutational analysis of TAR RNA: critical spacing between the bulge and loop recognition domains." *Nucleic Acids Research* 19, 6169-6176.

Berkhout, B., Silverman, R. H. & Jeang, K.-T. (1989a) "Tat *trans*-activates the human immunodeficiency virus through a nascent RNA target." *Cell* 59, 273-282.

Biou, V., Yaremchuk, A., Tukalo, M. & Cusak, S. (1994) "The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$." *Science* 263, 1404-1410.

Black, D. L., Chabot, B. & Steitz, J. A. (1985) "U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-messenger RNA splicing." *Cell* 42, 737-.

Boulard, Y., Cognet, J. A. H., Gabarro-Arpa, J., Bret, M. L., Carbonnaux, C. & Fazakerly, G. V. (1995) "Solution structure of an oncogenic DNA duplex, the K-*ras* gene and the sequence containing a central C.A or A.G mismatch as a function of pH: Nuclear magnetic resonance and molecular dynamics studies." *Journal of Molecular Biology* 246, 194-208.

Boulard, Y., Cognet, J. A. H., Gabarro-Arpa, J., Bret, M. L., Sowers, L. C. & Fazakerly, G. V. (1992) "The pH dependent configurations of the C.A mispair in DNA." *Nucleic Acids Research* 20, 1933-1941.

Brown, S. C., Weber, P. L. & Mueller, L. (1988) "Toward complete ^1H NMR spectra in proteins." *Journal of Magnetic Resonance* 77, 166-169.

Brown, T., Leonard, G. A., Booth, E. D. & Kneale, G. (1990) "Influence of pH on the conformation and stability of mismatch base-pairs in DNA." *Journal of Molecular Biology* 212, 437-440.

Calnan, B. J., Biancalana, S., Hudson, D. & Frankel, A. D. (1991b) "Analysis of arginine-rich peptides from the HIV-1 Tat protein reveals unusual features of RNA-protein recognition." *Genes and Development* 5, 201-209.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991a) "Arginine-mediated RNA recognition: the arginine fork." *Science* 252, 1167-1171.

Cech, T. R. (1987) "The chemistry of self-splicing RNA and RNA enzymes." *Science* 236, 1532-1539.

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* (R. F. Gesteland and J. F. Atkins, eds) pp. 239-269, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Cech, T. R. & Bass, B. (1986) "Biological catalysis by RNA." *Annual Review of Biochemistry* 55, 599-629.

Chastain, M. & Tinoco, I., Jr. (1991) "Structural Elements in RNA." *Progress in Nucleic Acid Research and Molecular Biology* 41, 131-177.

Chastain, M. & Tinoco, I., Jr. (1992) "A base triple structural domain in RNA." *Biochemistry* 31, 12733-12741.

Cheong, C. & Moore, P. B. (1992) "Solution structure of n unusually stable RNA tetraplex containing G- and U-quartet structures." *Biochemistry* 31, 8406-8414.

Churcher, M., Lamont, C., 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 tat protein requires amino acid residues flanking the basic domain and base pairs in the RNA stem." *Journal of Molecular Biology* 230, 90-110.

Cognet, J. A. H., Boulard, Y. & Fazakerly, G. V. (1995) "Helical parameters, fluctuations, alternative hydrogen bonding, and bending in oligonucleotides containing a mismatched base-pair by NOESY distance restrained and distance free molecular dynamics." *Journal of Molecular Biology* 246, 209-226.

Cohn, W. E. & Doherty, D. G. (1956) "The catalytic hydrogenation of pyrimidine nucleosides and nucleotides and the isolation of their ribose and respective ribose phosphates." *Journal of the American Chemical Society* 78, 2863-2866.

Colvin, R. A., White, S. W., Garcia-Blanco, M. A. & Hoffman, D. W. (1993) "Structural features of an RNA containing the CUGGGA loop of the human immunodeficiency virus type 1 trans-activation response element." *Biochemistry* 32, 1105-1112.

Cordingley, 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. (1990) "Sequence-specific interaction of Tat protein and Tat peptides with the *trans*activation-responsive sequence element of human immunodeficiency virus type 1 *in vitro*." *Proceedings of the National Academy of Sciences (USA)* 87, 8985-8989.

Cruse, W. B. T., Saludjian, P., Biala, E., Strazewski, P., Prange, T. & Kennard, O. (1994) "Structure of a mispaired RNA double helix at 1.6-Å resolution and the implications for the prediction of RNA secondary structue." *Proceedings of the National Academy of Sciences (USA)* 91, 4160-4164.

Cullen, B. R. (1986) "*Trans*-activation of human immunodeficiency virus occurs via a bimodal mechanism." *Cell* 46, 973-982.

Cullen, B. R. (1991) "Regulation of HIV-1 gene expression." *FASEB Journal* 5, 2361-2368.

Dalluge, J. J., Hashizume, T., Sopchik, A. E., McCloskey, J. A. & Davis, D. R. (1996) "Conformational flexibility in RNA: The role of dihydrouridine." *Nucleic Acids Research* 24, 1073-1079.

Davis, P. W., Thurmes, W. & Tinoco, I., Jr. (1993) "Structure of a small RNA hairpin." *Nucleic Acids Research* 21, 537-545.

de los Santos, C., Rosen, M. & Patel, D. J. (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." *Biochemistry* 28, 7282-7289.

Delling, U., Reid, L. S., Barnett, R. W., Ma, M. Y.-X., Climie, S., Sumner-Smith, M. & Sonenberg, N. (1992) "Conserved nucleotides in the TAR RNA stem of human immunodeficiency virus type 1 are critical for Tat binding and *trans*-activation: A model for TAR RNA tertiary structure." *Journal of Virology* 66, 3018-3025.

Delling, U., Sumner-Smith, M., Barnett, R., Reid, L., Rosen, C. & Sonenberg, N. (1991) "The number of positively charged amino acids in the basic domain of Tat is critical for *trans*-activation and complex formation with TAR RNA." *Proceedings of the National Academy of Sciences (USA)* 88, 6234-6238.

Derome, A. E. (1987) "Modern NMR Techniques of Chemistry Research." Pergamon Press, Oxford.

Derse, P., Carvalho, M., Carroll, R. & Peterlin, B. M. (1991) "A minimal lentivirus Tat." *Journal of Virology* 65, 7012-7015.

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 Journal* 9, 4145-4153.

Dingwall, C., Ernberg, I., Gait, M. J., Green, S. M., Heaphy, S., Karn, J., Lowe, A. D., Singh, M., Skinner, M. A. & Valero, R. (1989) "Human immunodeficiency virus 1 Tat protein binds trans-activation-responsive region (TAR) *in vitro*." *Proceedings of the National Academy of Sciences (USA)* 86, 6925-6929.

Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., de Bear, J. S., Gough, G. R., Gilham, P. T. & Moras, D. (1989) "Crystallographic structure of an RNA helix: $[U(UA)_6A]_2$." *Journal of Molecular Biology* 209, 459-474.

Dorn, P., DaSilva, L., Martarano, L. & Derse, D. (1990) "Equine infectious anemia virus Tat: Insights into the structure, function, and evolution of lentivirus trans-activator proteins." *Journal of Virology* 64, 1616-1624.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P. & Ehresmann, B. (1987) "Probing the structures of RNAs in solution." *Nucleic Acids Research* 15, 9109-9128.

Ellington, A. D. & Szostak, J. W. (1990) "*In vitro* selection of RNA molecules that bind specific ligands." *Nature* 346, 818-822.

Emerson, J. & Sundaralingam, M. (1980) "Structure of the potassium salt of the modified nucleotide dihydrouridine 3'-monophosphate hemihydrate: Correlation between the base pucker and sugar pucker and models for metal interactions with ribonucleic acid loops." *Acta Crystallogr. B* 36, 537-543.

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

Famulok, M. (1994) "Molecular recognition of amino acids by RNA aptamers: An L-citrulline binding RNA motif and its evolution into an L-arginine binder." *Journal of the American Chemical Society* 116, 1698-1706.

Feinberg, M. B., Baltimore, D. & Frankel, A. D. (1991) "The role of Tat in the human immunodeficiency virus life cycle indicates a primary effect on transcriptional elongation." *Proceedings of the National Academy of Sciences (USA)* 88, 4045-4049.

Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. (1986) "HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA." *Cell* 46, 808-817.

Feldman, M. Y. (1977) "Minor components in transfer RNA: The location-function relationships." *Prog. Biophys. Molec. Biol. 32*, 83-102.

Feng, S. & Holland, E. C. (1988) "HIV-1 tat *trans*-activation requires the loop sequence within TAR." *Nature* 334, 165-167.

Fountain, M. A., Serra, M. J., Krugh, T. R. & Turner, D. H. (1996) "Structural features of a six-nucleotide RNA hairpin loop found in ribosomal RNA." *Biochemistry* 35, 6539-6548.

Frankel, A. D. (1994) "Using peptides to study RNA-protein recognition." in *RNA-Protein Interactions* (K. Nagai and I. W. Mattaj, eds) pp. 221-247, IRL Press, New York, New York.

Frankel, A. D. & Pabo, C. O. (1988) "Tat protein from human immunodeficiency virus forms a metal-linked dimer." *Science* 240, 70-73.

Garcia, J. A., Harrich, D., Soultanakis, F., Wu, F., Mitsuyasu, R. & Gaynor, R. B. (1989) "Human immunodeficiency virus type I LTR TATA and TAR region sequences required for transcriptional regulation." *EMBO Journal* 8, 765-778.

Gaur, R. K. & Krupp, G. (1993) "Modification interference approach to detect ribose moieties important for the optimal activity of a ribozyme." *Nucleic Acids Research* 21, 21-26.

Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A. & Famulok, A. R. (1996) "RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity." *Nucleic Acids Research* 24, 1029-1036.

Grodberg, J. & Dunn, J. J. (1988) "*ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification." *Journal of Bacteriology* 170, 1245-1253.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. R. & Altman, S. (1983) "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." *Cell* 35, 849-857.

Haasnoot, C. A. G., Westerlink, H.P., van der Marel, G. A. & van Boom, J. H. (1984) "Discrimination between A-type and B-type conformations of double helical nucleic

acid fragments in solution by means of two-dimensional Nuclear Overhauser Experiments." *Journal of Biomolecular Structure and Dynamics* 2, 345-360.

Haasnoot, C. A. G., de Leeuw, F. A. A. M. & Altona, C. (1980) "The relationship between proton-proton NMR coupling constants and substituent electronegativities (I.) An empirical generalization of the Karplus equation." *Tetrahedron* 36, 2783-2792.

Hammarskjold, M.-I., Heimer, J., Hammarskjold, B., Sangwan, L., Albert, I. & Rekosh, D. (1989) "Regulation of the human immunodeficiency virus *env* expression by the *rev* gene product." *Journal of Virology* 63, 1959-1966.

Hamy, F., Asseline, U., Grasby, J., Iwai, S., Pritchard, C., Slim, G., Butler, P. J. C., 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." *Journal of Molecular Biology* 230, 111-123.

Haseltine, W. A. (1991) "Molecular biology of the human immunodeficiency virus type 1." *FASEB Journal* 5, 2349-2360.

Hauber, J. & Cullen, B. R. (1988) "Mutational analysis of the *trans*-activation response region of the human immunodeficiency virus type 1 long terminal repeat." *Journal of Virology* 62, 673-679.

Hauber, J., Malim, M.-H. & Cullen, B. R. (1989) "Mutational analysis of the conserved basic domain of the human immunodeficiency virus *tat* protein." *Journal of Virology* 63, 1181-1187.

Herschlag, D. & Cech, T. R. (1990) "DNA cleavage catalyzed by the ribozyme from *Tetrahymena*." *Nature* 344, 405-409.

Heus, H. A. & Pardi, A. (1991a) "Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops." *Science* 253, 191-194.

Heus, H. A. & Pardi, A. (1991b) "Novel ^{1}H nuclear magnetic resonance assignment procedure for RNA duplexes." *Journal of the American Chemical Society* 113, 4360-4361.

Hines, J. V., Landry, S. M., Varani, G. & Tinoco, I., Jr. (1994) "Carbon-proton scalar couplings in RNA: 3D heteronuclear and 2D isotope-edited NMR of a ^{13}C-labeled extrastable hairpin." *Journal of the American Chemical Society* 116, 5823-5831.

Hoffman, D. W., Colvin, R. A., Garcia-Blanco, M. A. & White, S. W. (1993) "Structural features of the trans-activation response RNA element of equine infectious anemia virus." *Biochemistry* 32, 1096-1104.

Holbrook, S. R., Cheong, C., Tinoco, I., Jr. & Kim, S.-H. (1991) "Crystal structure of an RNA double helix incorporating a track of non Watson-Crick base pairs." *Nature* 353, 579-581.

Hoogsteen, K. (1959) "The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine." *Acta Crystallogr.* 12, 822-823.

Hore, P. J. (1983) "Solvent suppression in Fourier transform nuclear magnetic resonance." *Journal of Magnetic Resonance* 55, 283-300.

Hunter, W. H., Brown, T., Anand, N. N. & Kennard, O. (1986) "Structure of an adenine-cytosine base pair and its implications for mismatch repair." *Nature* 230, 552-555.

Jaeger, J. A. & Tinoco, I., Jr. (1993) "An NMR study of the HIV-1 TAR element hairpin." *Biochemistry* 32, 12522-12530.

Jeener, J., Meier, B. H., Bachman, P. & Ernst, R. R. (1979) "Investigation of exchange processes by two-dimensional NMR spectroscopy." *Journal of Chemical Physics* 71, 4546-4553.

Jones, K. A. & Peterlin, B. M. (1994) "Control of RNA initiation and elongation at the HIV-1 promoter." *Annu. Rev. Biochem.* 63, 717-743.

Jucker, F. M. & Pardi, A. (1995) "Solution structure of the CUUG hairpin loop: A novel RNA tetraloop motif." *Biochemistry* 34, 14416-14427.

Kalnik, M. W., Kouchakdjian, M., Li, B. F. L., Swann, P. F. & Patel, D. J. (1988) "Base pair mismatches and carcinogen-modified bases in DNA: An NMR study of A-C and A-O^4meT pairing in dodecanucleotide duplexes." *Biochemistry* 27, 100-108.

Kao, S.-Y., Calman, A. F., Luciw, P. A. & Peterlin, B. M. (1987) "Anti-termination of transcription within the long terminal repeat of HIV-1 by *tat* gene product." *Nature* 330, 489-493.

Karn, J., Gait, M. J., Churcher, M. J., Mann, D. A., Mikaelian, I. & Pritchard, C. (1994) "Control of human immunodeficiency virus gene expression by the RNA-binding proteins Tat and Rev." in *RNA-protein Interactions* (K. Nagai and I. W. Mattaj, eds) pp. 192-220, IRL Press, New York, New York.

Karplus, M. (1959) "Contact electron-spin coupling of nuclear magnetic moments." *Journal of Chemical Physics* 30, 11-15.

Kessler, H., Gehrke, M. & Griesinger, C. (1988) "Two-dimensional NMR spectroscopy: background and overview of the experiments." *Angew. Chem. Int. Ed. Engl* 27, 490-536.

Kessler, M. & Matthews, M. B. (1992) "Premature termination and processing of human immunodeficiency virus type 1-promoted transcripts." *Journal of Virology* 66, 4488-4496.

Kim, S.-H., Suddath, F. L., Quigley, G. J., McPherson, A., Sussman, J. L., Wang, A. H. J., Seeman, N. C. & Rich, A. (1974) "Three-dimensional tertiary structure of yeast phenylalanine transfer RNA." *Science* 185, 435-440.

Kim, S.-Y., Byrn, R., Groopman, J. & Baltimore, D. (1989) "Temporal aspects of DNA and RNA synthesis during of human immunodeficiency virus infection: Evidence for differential gene expression." *Journal of Virology* 63, 3708-3713.

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. (1982) "Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena.*" *Cell* 31, 147-157.

Kuppuswamy, M., Subramanian, T., Srinivasan, A. & Chinnadurai, G. (1989) "Multiple functional domains of Tat, the *trans*-activator of HIV-1, defined by mutational analysis." *Nucleic Acids Research* 17, 3551-3561.

Laspia, M. F., Rice, A. P. & Matthews, M. B. (1989) "HIV-1 tat protein increases transcriptional initiation and stabilizes elongation." *Cell* 59, 283-292.

Laspia, M. F., Rice, A. P. & Matthews, M. B. (1990) "Synergy between HIV-1 Tat and adenovirus E1A is principally due to stabilization of transcriptional elongation." *Genes and Development* 4, 2397-2408.

Lee, C.-H., Saksela, K., Mirza, U. A., Chait, B. T. & Kuriyan, J. (1996) "Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain." *Cell* 85, 931-942.

Lehman, N. & Joyce, G. F. (1993) "Evolution *in vitro* of an RNA enzyme with altered metal dependence." *Nature* 361, 182-185.

Leonard, G. A., McAuley-Hecht, K. E., Ebel, S., Lough, D. M., Brown, T. & Hunter, W. N. (1994) "Crystal and molecular structure of r(CGCGAAUUAGCG): An RNA duplex containing two G(anti)-G(anti) base pairs." *Structure* 2, 483-494.

Leroy, J. L., Kochoyan, M., Huynh-Dinh, T. & Guéron, M. (1988) "Characterization of base-pair opening in deoxynucleotide duplexes using catalyzed exchange of the imino proton." *Journal of Molecular Biology* 200, 223-238.

Lietzke, S. E., Barnes, C. L. & Kundrot, C. E. (1995) "Crystallization and structure determination of RNA." *Current Opinion in Structural Biology* 5, 645-649.

Long, K. S. & Crothers, D. M. (1995) "Interaction of human immunodeficiency virus type 1 Tat-derived peptides with TAR RNA." *Biochemistry* 34, 8885-8895.

Madhani, H. D. & Guthrie, C. (1994) "Dynamic RNA-RNA interactions in the spliceosome." *Annual Review of Genetics* 28, 1-26.

Malim, M. H., Huber, J., Les, Y., Maizel, J. V. & Cullen, B. R. (1989) "The HIV-1 *rev trans*-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA." *Nature* 338, 254-257.

Marciniak, R. A., Calnan, B. J., Frankel, A. D. & Sharp, P. A. (1990a) "HIV *tat* protein *trans*-activates transcription in vitro." *Cell* 63, 791-802.

Marciniak, R. A., Garcia-Blanco, M. & Sharp, P. A. (1990b) "Identification and characterization of a HeLa nuclear protein that specifically binds to the *trans*-activation-response (TAR) element of human immunodeficiency virus." *Proceedings of the National Academy of Sciences (USA)* 87, 3624-3628.

Marciniak, R. A. & Sharp, P. A. (1991) "HIV-1 Tat protein promotes formation of more processive elongation complexes." *EMBO Journal* 10, 4189-4196.

Marino, J. P., Razmic S. Gregorian, J., Csankovszki, G. & Crothers, D. M. (1995) "Bent helix formation between RNA hairpins with complementary loops." *Science* 268, 1448-1454.

McConnell, B. & Seawell, P. C. (1972) "Proton exchange of nucleic acids. Amino protons of mononucleotides." *Biochemistry* 11, 4382-4392.

McKenna, M.-C., Anderson, D., Cascio, D. & Eisenberg, D. (1994) "Crystallization studies of the human immunodeficiency virus (HIV-1) Tat protein and its trans-activation response element (TAR) RNA." *Acta Crystallography* D50, 527-534.

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.

Moore, M. J., Query, C. C. & Sharp, P. A. (1993) "Splicing of precursors to mRNA by the spliceosome." in *The RNA World* (R. F. Gesteland and J. F. Atkins, eds) pp. 119-135, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Moore, P. B. (1992) "Recent RNA structures." *Current Opinion Structural Biology* 3, 340-344.

Moore, P. B. (1993) "Ribosomes and the RNA world." in *The RNA World* (R. F. Gesteland and J. F. Atkins, eds) pp. 119-135, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Muesing, M. A., Smith, D. H. & Capon, D. J. (1987) "Regulation of mRNA accumulation by a human immunodeficiency virus *trans*-activator protein." *Cell* 48, 691-701.

Muller, N., Ernst, R. R. & Wuthrich, K. (1986) "Multiple-quantum-filtered two-dimensional correlated spectroscopy of proteins." *Journal of the American Chemical Society* 108, 6482-6492.

Narwot, B., Malkiewicz, A., Smith, W. S., Sierzputowska-Gracz, H. & Agris, P. F. (1995) "RNA modified uridines VII: Chemical synthesis and initial analysis of tRNA D-loop oligomers with tandem modified uridines." *Nucleosides & Nucleotides* 14, 143-165.

Nickonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M. & Pardi, A. (1992) "Preparation of ^{13}C and ^{15}N labeled RNA for heteronuclear multidimensional NMR studies." *Nucleic Acids Research* 20, 4507-4513.

Noller, H. F. (1991) "Ribosomal RNA and translation." *Annu. Rev. Biochem.* 60, 191-227.

Noller, H. F., Hoffarth, V. & Zimniak, L. (1992) "Unusual resistance of peptidyl transferase to protein extraction procedures." *Science* 256, 1416-1419.

Nowakowski, J. & Tinoco, I., Jr. (1996) "Conformation of an RNA molecule that models the P4/P6 junction from group I introns." *Biochemistry* 35, 2577-2585.

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.

Pace, N. R. & Smith, D. (1990) "Ribonuclease P: Function and variation." *J. Biol. Chem.* 265, 3587-3590.

Pan, T., Long, D. M. & Uhlenbeck, O. C. (1993) "Divalent metal ions in RNA Folding and catalysis." in *The RNA World* (R. F. Gesteland and J. F. Atkins, eds) pp. 271-302, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Peattie, D. A. (1979) "Direct chemical method for sequencing RNA." *Proceedings of the National Academy of Sciences, USA* 76, 1760-1764.

Peebles, C. L., Perlman, P. S., Mecklenberg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. & Cheng, H.-L. (1986) "A self-splicing RNA excises an intron lariat." *Cell* 44, 213-223.

Petersheim, M. & Turner, D. H. (1983) "Base-stacking and base-pairing contributions to helix stability: Thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp." *Biochemistry* 22, 256-263.

Piantini, U., Sørensen, O. W. & Ernst, R. R. (1982) "Multiple quantum filters for elucidating NMR coupling networks." *Journal of the American Chemical Society* 104, 6800-6801.

Piotto, M., Saudek, V. & Sklenar, V. (1992) "Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions." *Journal of Biomolecular NMR* 2, 661-665.

Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994) "Three-dimensional structure of a hammerhead ribozyme." *Nature* 372, 68-74.

Pomerantz, R. J., Trono, D., Feinberg, M. B. & Baltimore, D. (1990) "Cells nonproductively infected with HIV-1 exhibit an aberrant pattern of viral RNA expression: A molecular model for latency." *Cell* 61, 1271-1276.

Portmann, S., Grimm, S., Workman, C., Usman, N. & Egli, M. (1996) "Crystal structures of an A-form duplex with single-adenosine bulges and a conformational basis for site-specific RNA self-cleavage." *Chemistry & Biology* 3, 173-184.

Pritchard, C. E., Grasby, J. A., Hamy, F., Zacharek, A. M., Singh, M., Karn, J. & Gait, M. J. (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 peptide-RNA 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." *Proceedings of the National Academy of Sciences (USA)* 90, 3680-3684.

Puglisi, J. D. & Tinoco, I., Jr. (1989) "Absorbance melting curves of RNA." *Methods in Enzymology* 180, 304-325.

Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992) "Conformation of TAR RNA-arginine complex by NMR spectroscopy." *Science* 257, 76-80.

Puglisi, J. D., Wyatt, J. R. & Tinoco, I., Jr. (1990a) "Conformation of an RNA pseudoknot." *Journal of Molecular Biology* 214, 437-453.

Puglisi, J. D., Wyatt, J. R. & Tinoco, I., Jr. (1990b) "Solution conformation of an RNA hairpin loop." *Biochemistry* 29, 4215-4226.

Raszka, M. (1974) "Mononucleotides in aqueous solution: Proton magnetic resonance studies of amino groups." *Biochemistry* 13, 4616-4622.

Ratnasabapathy, R., Sheldon, M., Johal, L. & Hernandez, N. (1990) "The HIV-1 long terminal repeat contains an unusual element that induces the synthesis of short RNAs from various mRNA and snRNA promoters." *Genes and Development* 4, 2061-2074.

Rice, A. P. & Carlotti, F. (1990a) "Mutational analysis of the conserved cysteine-rich region of the human immunodeficiency virus type 1 Tat protein." *Journal of Virology* 64, 1864-1868.

Riordan, F. A., Bhattacharyya, A., McAteer, S. & Lilley, D. M. J. (1992) "Kinking of RNA helices by bulged bases, and the structure of the human immunodeficiency virus *trans*-activator response element." *Journal of Molecular Biology* 226, 305-310.

Ritter, K., Churcher, M. J., Gait, M. J. & Karn, J. (1995) "The human immunodeficiency virus long terminal repeat includes a specialized initiator element which is required for Tat-responsive transcription." *Journal of Molecular Biology* 248, 562-580.

Robertson, D. L. & Joyce, G. F. (1990) "Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA." *Nature* 344, 467-468.

Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. (1985) "The location of *cis*-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat." *Cell* 41, 813-823.

Rould, M. A., Perona, J. S., Soll, D. & Steitz, T. A. (1989) "Structure of *E.coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Glu}$ and ATP at 2.8 Å resolution." *Science* 246, 1135-1142.

Roy, S., Delling, U., Chen, C.-H., Rosen, C. A. & Sonenberg, N. (1990a) "A bulge structure in HIV-1 TAR RNA is required for Tat binding and Tat-mediated *trans*-activation." *Genes and Development* 4, 1365-1373.

Roy, S., Parkin, N. T., Rosen, C., Itovitch, J. & Sonenberg, N. (1990b) "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." *Journal of Virology* 64, 1402-1406.

Roy-Burman, P., Roy-Burman, S. & Visser, D. W. (1965) "Incorporation of 5,6-dihydrouridine triphosphate into ribonucleic acid by DNA-dependent RNA polymerase." *Biochemical and Biophysical Research Communications* 20, 291-297.

Roy-Burman, P., Roy-Burman, S. & Visser, D. W. (1967) "Utilization of 5,6-dihydrouridine 5'-triphosphate in the reaction catalyzed by *Escherichia coli* RNA polymerase." *Biochimica Biopyhsica Acta* 142, 355-367.

Ruben, S., Perkins, A., Purcell, R., Joung, K., Sia, R., Burghoff, R., Haseltine, W. A. & Rosen, C. A. (1989) "Structural and functional characterization of the human immunodeficiency virus *tat* protein." *Journal of Virology* 63, 1-8.

Saenger, W. (1984) "Principles of Nucleic Acid Structure." Springer-Verlag, New York, New York.

Schwartz, S., Felber, B. K., Benko, D. M., Fenyo, E.-M. & Pavlakis, G. N. (1990) "Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1." *Journal of Virology* 64, 2519-2529.

Schüpbach, J. (1989) "Human Retrovirology: Facts and Concepts." Springer-Verlag, New York, New York.

Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) "Sequence-specific recognition of double helical nucleic acids by proteins." *Proceedings of the National Academy of Sciences (USA)* 73, 804-808.

Selby, M. J., Bain, E. S., Luciw, P. & Peterlin, B. M. (1989) "Structure, sequence and position of the stem-loop in TAR determine transcriptional elongation by Tat through the HIV-1 long terminal repeat." *Genes and Development* 3, 547-558.

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." *Genes and Development* 5, 2508-2520.

Shen, L. X., Cai, Z. & Tinoco, I., Jr. (1995) "RNA structure at high resolution." *FASEB, Journal* 9, 1023-1033.

Sklenar, V. & Feigon, J. (1990) "Formation of a stable triplex from a single DNA strand." *Nature* 345, 836-838.

Slice, L. W., Codner, E., Antelman, D., Holly, M., Wegrzynski, B., Wang, J., Toome, V., Hsu, M.-C. & Nalin, C. M. (1992) "Characterization of recombinant HIV-1 Tat and its interaction with TAR RNA." *Biochemistry* 31, 12602-12608.

Slim, G., Pritchard, C., Biala, E., Asseline, U. & Gait, M. J. (1991) "Synthesis of site-specifically modified oligoribonucleotides for studies of the recognition of TAR RNA by HIV-1 *tat* protein and studies of hammerhead ribozymes." *Nucleic Acids Research Symposium Series* 24, 55-58.

Sodroski, J., Goh, W. C., Rosen, C. A., Dayton, A., Terwilliger, E. & Haseltine, W. A. (1986) "A second post-transcriptional activator gene required for HTLV-III replication." *Nature* 321, 412-417.

Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. A. (1985a) "Location of the *trans*-acting region on the genome of human T-cell lymphotropic virus type III." *Science* 229, 74-77.

Sodroski, J., Rosen, C., Wong-Staal, F., Salahuddin, S. Z., Popovic, M., Ayra, S., Gallo, R. C. & Haseltine, W. A. (1985b) "*Trans*-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat." *Science* 227, 171-173.

States, D. J., Haberkorn, R. A. & Ruben, D. J. (1982) "A two-dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants." *Journal of Magnetic Resonance* 48, 286-292.

Steitz, T. A. (1993) "Similarities and differences between RNA and DNA recognition by proteins." in *The RNA World* (R. F. Gesteland and J. F. Atkins, eds) pp. 219-237, Cold Spring Harbor Press, Cold Spring Harbor, NY.

Subramanian, T., Govindarajan, R. & Chinnadurai, G. (1991) "Heterologous basic domain substitutions in the HIV-1 *tat* protein reveal an arginine-rich motif required for *trans*-activation." *EMBO Journal* 10, 2311-2318.

Suck, D., Saenger, W. & Zechmeister, K. (1971) "Molecular structure of the tRNA minor constituent dihydrouridine." *Acta Crystallogr.* B28, 596-605.

Sumner-Smith, M., Roy, S., Barnett, R., Reid, L. S., Kuperman, R., Delling, U. & Sonenberg, N. (1991) "Critical chemical features in *trans*-acting responsive RNA are required for interaction of human immunodeficiency virus type 1 Tat protein." *Journal of Virology* 65, 5196-5202.

Szewczak, A. A., Moore, P. B., Chan, Y.-L. & Wool, I. G. (1993) "The Solution Structure of the Sarcin/Ricin Loop from 28S Ribosomal RNA." *Proceedings of the National Academy of Sciences USA* 90, 9581-9583.

Szostak, J. W. & Ellington, A. D. (1993) "*In vitro* selection of functional RNA sequences." in *The RNA World* (R. F. Gesteland and J. F. Atkins, eds) pp. 511-533, Cold Spring Harbor Press, Cold Spring Harbor, New York.

Tan, R. & Frankel, A. D. (1992) "Circular dichroism studies suggest that TAR RNA changes conformation upon specific binding of arginine or guanidine." *Biochemistry* 31, 10288-10294.

Tao, J. & Frankel, A. D. (1992) "Specific binding of arginine to TAR RNA." *Proceedings of the National Academy of Sciences* 89, 2723-2726.

Tao, J. & Frankel, A. D. (1993) "Electrostatic interactions modulate the RNA-binding and transactivation specificities of the human immunodeficiency virus and simian immunodeficiency virus Tat proteins." *Proceedings of the National Academy of Sciences (USA)* 90, 1571-1575.

Tao, J. & Frankel, A. D. (1996) "Arginine-binding RNAs resembling TAR identified by in *vitro* selection." *Biochemistry* 35, 2229-2238.

Teitelbaum, H. & Englander, S. W. (1975) "Open states in native polynucleotides (I). Hydrogen-exchange study of adenine-containing double helices." *Journal of Molecular Biology* 92, 55-78.

Tinoco, I., Jr., Davis, P. W., Hardin, C. C., Puglisi, J. D., Walker, G. T. & Wyatt, J. (1987) "RNA Structure from A to Z." *Cold Spring Harbor Symposia on Quantitative Biology* LII, 135-146.

Tinoco, I., Jr. (1960) "Hyperchromism in polynucleotides." *Journal of the American Chemical Society* 82, 4785-4790.

Toohey, M. G. & Jones, K. A. (1989) "*In vitro* formation of short RNA polymerase II transcripts that terminate within the HIV-1 and HIV-2 promoter-proximal downstream regions." *Genes and Development* 3, 265-282.

Tuerk, C. & Gold, L. (1990) "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." *Science* 249, 505-510.

Usman, N., Egli, M. & Rich, A. (1992) "Large scale chemical synthesis, purification and crystallization of RNA-DNA chimeras." *Nucleic Acids Research* 20, 6695-6699.

Usman, N., Ogilvie, K. K., Jiang, M.-Y. & Cedergen, R. J. (1987) "Automated synthesis of long oligoribonucleotides using 2'-O-silylated ribonucleoside 3'-O-phosphoramidites on a controlled-pore glass support: Synthesis of a 43-nucleotide sequence similar to the 3'-half molecule of an *Eschericia coli* formylmethionine tRNA." *Journal of the American Chemical Society* 109, 7845-7854.

Vaishnav, Y. N. & Wong-Staal, F. (1991) "The biochemistry of AIDS." *Annual Reviews of Biochemistry* 60, 577-630.

Varani, G., Cheong, C. & Tinoco, I., Jr. (1991) "Structure of an unusually stable RNA hairpin." *Biochemistry* 30, 3280-3289.

Varani, G. & Tinoco, I., Jr. (1991) "RNA structure and NMR spectroscopy." *Quarterly Reviews of Biophysics* 24, 479-532.

Varani, G. & Pardi, A. (1994) "Structure of RNA." in *RNA-Protein Interactions* (K. Nagai and I. W. Mattaj, eds) pp. 1-24, IRL Press, New York, New York.

Varani, G., Wimberly, B. & Tinoco, I., Jr. (1989) "Conformation and dynamics of an RNA internal loop." *Biochemistry* 28, 7760-7772.

van der Veen, R. , Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F. & Grivell, L. A. (1986) "Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*." *Cell* 44, 225-234.

van de Ven, F. J. M. & Hilbers, C. W. (1988) "Nucleic Acids and Nuclear Magnetic Resonance." *European Journal of Biochemistry* 178, 1-38.

Wang, C., Gao, H., Gaffney, B. L. & Jones, R. A. (1991) "Nitrogen-15-labeled oligodeoxynucleotides. 3. Protonation of the adenine N^1 in the A-C and A-G mispairs of the duplexes $\{d[CG(^{15}N^1)AGAATTCCCG]\}_2$ and $\{d[CGGGAATTC(^{15}N^1)ACG]\}_2$." *Journal of the American Chemical Society* 113, 5486-5488.

Wang, Z. & Rana, T. M. (1996) "RNA conformation in the Tat-TAR complex determined by site-specific photo-cross-linking." *Biochemistry* 35, 6491-6499.

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.

Westhof, E., Dumas, P. & Moras, D. (1985) "Crystallographic refinement of yeast aspartic acid transfer RNA." *Journal of Molecular Biology* 184, 119-145.

Westhof, E. & Michel, F. (1994) "Prediction and experimental investigation of RNA secondary and tertiary foldings." in *RNA-Protein Interactions* (K. Nagai and I. W. Mattaj, eds) pp. 25-51, IRL Press, New York, New York.

White, S. A., Nilges, M., Huang, A., Brunger, A. T. & Moore, P. B. (1992) "NMR analysis of Helix I from the 5S RNA of *Escherichia coli*.." *Biochemistry* 31, 1610-1621.

Wimberly, B., Varani, G. & Tinoco, I., Jr. (1993) "The conformation of Loop E of eukaryotic 5S ribosomal RNA." *Biochemistry* 32, 1078-1087.

Wincott, F., Usman, N. & al, e. (1995) "Synthesis, deprotection, analysis and purification of RNA and ribozymes." *Nucleic Acids Research* 23, 2677-2684.

Woo, N. H., Roe, B. A. & Rich, A. (1980) "Three-dimensional structure of *E.coli* initiator $tRNA^{Met}$." *Nature* 286, 346-351.

Wu, F., Garcia, J., Sigman, D. & Gaynor, R. (1991) "Tat regulates the binding of the human immunodeficiency virus *trans*-activating region RNA loop-binding protein TRP-185." *Genes and Development* 5, 2128-2140.

Wuthrich, K. (1986) "NMR of Proteins and Nucleic Acids." Wiley, New York, New York.

Wyatt, J. R., Chastain, M. & Puglisi, J. D. (1991) "Synthesis and purification of large amounts of RNA oligonucleotides." *BioTechniques* 11, 764-769.

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.

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." *Proceedings of the National Academy of Sciences (USA)* 92, 6052-6056.