64

Synthesis of RNA with Selective Isotopic Labels for NMR Structural Studies

by

Thomas J. Tolbert

B.S., Chemistry
Purdue University, 1991

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

DOCTOR OF PHILOSOPHY
IN BIOCHEMISTRY

at the

Massachusetts Institute of Technology

February 1998

Signature of Author.....
Department of Chemistry
November 19, 1997

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 JoAnne Stubbe..... .....
Chair

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

Professor Lawrence J. Stern.....

Synthesis of RNA with Selective Isotopic Labels for NMR Structural Studies

by

Thomas J. Tolbert

Submitted to the Department of Chemistry on November 19, 1997 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biochemistry

ABSTRACT

RNA is synthesized with specific isotopic labels (^{2}H, ^{13}C, and ^{15}N) to alleviate spectral crowding and broad linewidth problems encountered in large RNA NMR spectroscopy. A combined chemical and enzymatic synthesis is described in which glycerol (U-d8) is chemically converted into 3',4',5',5"-$^{2}H_4$-D,L-ribose, and then the 3',4',5',5"-$^{2}H_4$-D-ribose is enzymatically converted into the four NTPs of RNA. Enzymes of purine salvage and pyrimidine biosynthesis are overexpressed and purified in order to catalyze the nucleotide forming reactions. Several isotopically labeled forms of glucose are enzymatically converted into the four NTPs of RNA. The pentose phosphate pathway is utilized to convert glucose into ribose-5-phosphate and to accomplish specific hydrogen exchange of the H1' and H2' protons of ribose-5-phosphate. The enzymes previously used to convert ribose derived from glycerol into nucleotides are also used to convert ribose-5-phosphate derived from glucose into nucleotides. Methods for deuterating and heteronuclear labeling the base moieties of RNA are discussed. The isotopically labeled NTPs produced from glycerol and glucose are used in T7-transcription reactions to produce the 30 nucleotide HIV-2 TAR RNA. NMR experiments are conducted on the different isotopically labeled forms of TAR RNA to determine the usefulness of the different isotopic labeling patterns. Specific deuteration of RNA is demonstrated to have a dramatic effect on reducing spectral crowding and simplifying RNA NMR spectra. Specific deuteration is also shown to reduce relaxation rates of the remaining RNA protons. Spectral editing of heteronuclear RNA NMR spectra is demonstrated using specifically deuterated, heteronuclear labeled RNA derived from isotopically labeled glucose. Methods for improving both homonuclear and heteronuclear NMR spectra of large RNA molecules with labeling patterns produced using these techniques are discussed. All of the H1' and H2' chemical shifts for the 11 uridine nucleotides in a 65 nucleotide RNA are identified using specific deuteration patterns derived from these synthetic techniques. Improved transfer of magnetization is also demonstrated using specifically deuterated, ^{13}C labeled RNA. Application of the isotopic labeling patterns and synthetic strategies described in this thesis to NMR studies of RNA should prove very valuable.

Thesis Supervisor: James R. Williamson

Title: Associate Professor of Chemistry

ACKNOWLEDGMENTS

About six years ago Jamie Williamson told me of an idea he had to incorporate specific isotopic labels into RNA to simplify NMR structural studies. He jotted down a few chemical reactions, a little biology, waved his hands, and suddenly acetylene and formaldehyde had turned into RNA. The bulk of the scheme would be based upon previously reported synthetic procedures, and it was all so simple that it would probably not take more than a year. I joined the Williamson lab shortly afterward...

Things did not work out quite the way I thought they would. It took almost four years to make RNA, but I learned a great deal along the way. I must thank Jamie for support through the years, and also for creating a diverse research environment where one can learn about a wide range of topics. Jamie's lab is the only place that I know of where you can look at drosophila under a microscope and in the same day sit down at a 600 Mhz NMR spectrometer.

Jamie has brought together a group of scientists who have a wide range of expertise, ranging from molecular biology to NMR pulse programming. In my time here at MIT I have gone to just about every person in the Williamson lab and asked them for advice about the different stages of the project that this thesis is about. Without the advice and help of all of the members of the Williamson lab this thesis could not have been written. In particular I would like to thank Jody Puglisi for setting the tone of the lab in the first few years I was here. Thanks are due to graduate students Pat, Heidi, Robert, Helen, Martha, Pete, and Selen for good advice. I must thank John Battiste and Alex Brodsky for first introducing me to higher field NMR. Alex Brodsky deserves special thanks for getting Didinium, our 600 MHz Inova, in working condition, as well as helpful advice on the spectral crowding problems of TAR8. Christopher Cilley also deserves special thanks for keeping the computers running. Through the years there have been many postdocs in the lab who have brought a steady flow of new techniques and good advice. Dan, Jeff, Sharon, Feng, Phil, Kwaku, and Lincoln, thank you very much. I thank Lincoln Scott for providing a copy of Chemintosh which allowed chemical schemes to be put into the text of this thesis (since Chemdraw wasn't working). Special thanks are due to Kwaku Dayie for teaching me about heteronuclear NMR and for creating the CCH-TOCSY NMR experiment. Kwaku's good advice, NMR knowledge, and enthusiasm have been an incredible help.

I thank Professors JoAnne Stubbe, Kaj Frank Jensen, Charles L. Turnbough Jr., and Howard Zalkin for generously providing plasmids and overexpressing strains. I had never purified a protein before I came to MIT, and this project would not have gotten very far without their plasmids and overexpressing strains. I also thank Pete Wishnok for nucleotide triphosphate mass spectra, which were a great help in characterization.

I thank John Frost for giving me my first chance to work in a research laboratory as an undergraduate research assistant. Even though I broke a small fortune in glassware he still hired me to work during the summer. I will never forget the voice of my graduate student mentor, Jean-Luc Montchamp, calling my name in exasperation every time the sound of broken glass rang through the laboratory. I can't thank Jean-Luc enough for his friendship and advice over the years. Half of the synthetic problems discussed in this thesis would never have been solved without his help.

A great deal of thanks are due to my family. I thank Mom and Dad A., Pam, Debbie and the girls for visits to Texas which helped to keep me sane. The warm Texas sun and good cookies have helped me through many a difficult time. David, Elizabeth, and Catherine, thanks for being there to talk to over the years. Special thanks are due to my Mom and Dad for all of their love and support over the years, it has been a long road but I think I am finally out of school now. Finally, I want to thank my wife, Becky, for all of her love and help as well as for being the best editor I know. Her help in getting through graduate school and writing this thesis has been immeasurable.

Table of Contents

Table of Contents

Chapter 1

1. Introduction

Background

With the discovery of group I self-splicing introns the study of RNA has rapidly expanded (Cech, 1990). From the ever increasing amount of biochemical information on RNA, it is clear that it plays an important role in gene expression and cellular function. RNA forms a bridge between DNA and proteins, being intricately involved in protein synthesis, with the ribosome being made up of approximately 66% RNA (Voet & Voet, 1995), messenger RNA transporting genetic information from DNA to the ribosomes, and tRNA bringing amino acids to the ribosomes. Evidence suggests that RNA is directly involved in the peptidyl transferase activity of ribosomes. The most compelling of this evidence is the resistance of this peptidyl transferase activity of ribosomes to protein extraction procedures (Noller *et al.*, 1992). Intracelluar localization of mRNA establishes polarity in cells, and is very important in development. An example of this is in drosophila development where gradients of maternal *bicoid* mRNA in oocytes helps specify the development of the head, thorax, and abdomen (St. Johnston, 1995). RNA protein interactions are involved several cellular processes including activation of transcription, regulation of translation, and regulation of splicing. One example of RNA protein interactions is the Tat-TAR interaction of HIV-2, where the binding of the TAR RNA to the Tat protein is required for activation of transcription (Calnan *et al.*, 1991; Brodsky & Williamson, 1997). A second example of RNA protein interactions is the interaction of the iron responsive element (IRE) RNA with the IRE binding protein (IREbp), where the translation of the mammalian protein ferritin is suppressed by the binding of the IREbp to the IRE RNA at low iron concentration (Klausner *et al.*, 1993). A third example of RNA-protein interactions is the regulation of splicing that occurs when the HIV-1 Rev protein binds to the Rev response element (RRE) RNA (Olsen *et al.*, 1990; Zapp *et al.*, 1991;

Battiste *et al.*, 1996). RNA is also very important in the spliceosome which requires several small nuclear RNAs for assembly and function (Madhani & Guthrie, 1992). Group I and II self splicing introns catalyze reactions similar to the spliceosome without the help of proteins (McPheeters & Abelson, 1992). Finally, the ability of RNA to both catalyze chemical reactions and store genetic information has allowed *in vitro* selection methods to be designed which have produced RNAs with ligase activities (Wright & Joyce, 1997) and RNA aptamers which are able to bind small molecules (Dieckmann *et al.*, 1996).

The importance of RNA in the cell and the amount biochemical knowledge that has accumulated about functional RNAs over the past decade has been the impetus for several structural studies of RNA. Although the chemical structure of RNA differs from DNA only in the 2' hydroxyl on the ribose ring and uracil replacing thymine, the helical conformation differs in that RNA usually adopts A-form helical structure while DNA adopts B-form structure. RNA forms a wide variety of intramolecular structures where single stranded RNA folds back upon itself to form intramolecular base pairs. Many functional RNAs have large regions of single stranded RNA within their predicted secondary structures which are conserved phylogenetically (Gutell *et al.*, 1985). These supposedly single stranded regions are are most likely regions of "unknown structure" (Pace, 1993) rather than unstructured regions, and base triples, AA platforms (Doudna & Cate, 1997), 2'-hydroxyl hydrogen bonds, and other non-standard structure probably lurk within these undefined regions. It has been interest in these unknown structural regions and how RNA accomplishes its multitude of tasks that has fueled many RNA structural studies.

Nuclear Magnetic Resonance Structural Studies of RNA

NMR spectroscopy is a powerful tool for studying the structure of RNA oligonucleotides (Varani *et al.*, 1996), and it has been used to determine the structures of

RNA molecules of up to 40 nucleotides in size (Chang & Varani, 1997; Ramos *et al.*, 1997). Some examples of the wide variety of RNA and RNA complexes whose structures have been determined with NMR spectroscopy include tetraloops (Jucker & Pardi, 1995), pseudoknots (Puglisi *et al.*, 1990), hairpins (Puglisi *et al.*, 1997), 'kissing' hairpin interactions between two stem loops (Marino *et al.*, 1995), RNA-peptide (Puglisi *et al.*, 1995; Ye *et al.*, 1995; Battiste *et al.*, 1996; Peterson *et al.*, 1994) and protein complexes (Allain *et al.*, 1996), RNA aminoglycoside interactions (Fourmy *et al.*, 1996; Recht *et al.*, 1996), RNA amino acid complexes (Puglisi *et al.*, 1992; Brodsky & Williamson, 1997), RNA aptamers (Dieckmann *et al.*, 1996; Jiang *et al.*, 1996), and others (Greenbaum *et al.*, 1995; Nowakowski & Tinoco, 1996; Glemarec *et al.*, 1996). NMR structural studies of RNA complexes that have proven difficult to study by other methods have been successful in obtaining structures, including several cases that were highly sought after such as the HIV Tat-TAR (Puglisi *et al.*, 1992; Brodsky & Williamson, 1997; Aboul-ela *et al.*, 1995) and Rev-RRE (Peterson *et al.*, 1994; Battiste *et al.*, 1996) RNA interactions. Unfortunately extension of NMR structural studies to RNAs larger than 40 nucleotides has been difficult because of spectral crowding and broad linewidths in RNA NMR spectra. Since there are many biologically interesting RNAs that are significantly larger than 40 nucleotides, such as t-RNA (≈76-80 nucleotides), it is important to develop techniques that would allow the study of these larger RNAs using NMR spectroscopy.

Information which can be obtained from NMR spectroscopy (NMR Observables)

A general discussion of structure determination with NMR will be made here with emphasis on the use of the information that can be extracted from NMR spectra (NMR observables) in structure determinations. A much more detailed discussion of NMR spectroscopy and structure determination with NMR spectroscopy can be found in Derome (Derome, 1990), Wüthrich (Wüthrich, 1986), Cavanagh *et. al.* (Cavanagh *et al.*, 1996), and Carrington McLachlan (Carrington & McLachlan, 1967). Nuclear magnetic resonance

spectroscopy of biological macromolecules can provide several types of information which can be used to determine structures. The main types of information that are used in determining NMR structures of macromolecules are the nuclear Overhauser effect (NOE), scalar coupling (spin-spin coupling or J-coupling), and chemical shift.

Nuclear Overhauser Effect (NOE). The nuclear Overhauser effect (NOE) is a through space transfer of magnetization from one proton to another that can be used to estimate the distance between two protons. The intensity of an NOE has a $1/r^6$ dependence and can be observed for distances less than or equal to 5 Å. Nuclear Overhauser Effects are one of the most important types of NMR information because they can be used to link two parts of a macromolecule close in space even though they may be separated distantly in primary sequence. In addition, patterns of NOEs can be indicative of certain types of structural motifs, such as α-helical or β-sheets in proteins, and A-form or B-form helical structure in nucleic acids.

Scalar Coupling. Another important type of NMR information is scalar coupling between covalently bonded nuclei. Scalar coupling between nuclei can generally be observed through at least 3 bonds, and the magnitude of the scalar coupling constant (J) depends on several factors related to molecular orbital overlap including bond length, hybridization, and dihedral angle. Observation of coupling between two nuclei establishes a through bond interaction between the nuclei and is therefore useful in assigning NMR spectra. In addition, three bond coupling constants are used in structural studies to determine the dihedral angle between protons on adjacent carbons via the Karplus equation (Karplus, 1963) which relates the magnitude of the J-coupling to the dihedral angle. This yields information about the conformation of the carbon chain to which the protons belong.

Chemical Shift. Chemical shift is the frequency that a nucleus resonates at relative to an arbitrary standard (usually TSP or TMS). Chemical shift can be influenced by the hybridization of an atom and also the electronegativity of the substituents that are attached

to that atom or neighboring atoms. The chemical shift of a nucleus yields information about the type of chemical environment that the nucleus is in, and is mainly used to identify the type of chemical bonding that a nucleus is involved in. Changes in chemical shift can indicate structural changes, those changes due to ligand binding for example, and can be used to identify areas in macromolecules of particular interest.

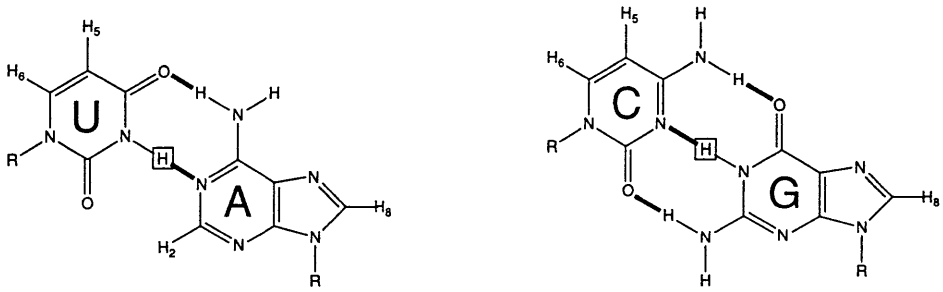
RNA NMR Observables

When NMR spectroscopy is conducted on a particular RNA, several of the NMR observables described above can be translated directly into structural information about the RNA. In particular, base pairing, sugar pucker, and glycosidic torsion information can be extracted directly from RNA NMR spectra, and these are summarized in Figure 1.1. More detailed information about the RNA structure can be obtained by assigning all of the interactions within RNA NMR spectra and using them as constraints in molecular modeling (Varani *et al.*, 1996).

Base pairing. NMR spectra of the exchangeable protons of RNA yield information about the base pairing of RNA oligonucleotides. The imino protons of guanine and uracil exchange with solvent rapidly and are unobservable in NMR spectra of the free nucleotides, but when guanine and uracil form base pairs in RNA the exchange rates of the imino protons are slowed and they often become observable with NMR spectroscopy, Figure 1.1A. Imino protons within base pairs resonate far downfield of other RNA protons at approximately 11-14 ppm, and so have little spectral crowding and are easily identified. Observation of imino protons within RNA NMR spectra obtained in 90% H_2O and 10% D_2O is very important in determining RNA structures because the observed imino protons indicate secondary structure.

Sugar pucker. Three bond scalar coupling constants between ribose protons can be used to determine the conformation of the ribose ring , or sugar pucker, via the Karplus

A. Secondary structure: observation of imino protons

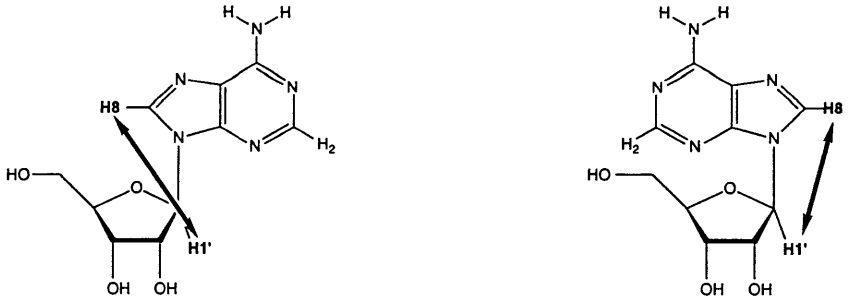


B. Sugar pucker: H1'-H2' scalar coupling

C3' *endo*
$J_{1'-2'} \approx 1$ hz

C2' *endo*
$J_{1'-2'} \approx 8$ hz

C. Glycosidic torsion: H1' to base NOEs



anti

syn

Figure 1.1 RNA NMR Observables

equation (Karplus, 1963). J-coupling between the H1' and H2' ribose protons are most often used to determine sugar pucker since H1'-H2' COSY crosspeaks occur in an uncrowded region of RNA NMR spectra. Nucleic acids usually adopt two sugar puckers, C3'-endo and C2'-endo, shown in Figure 1.1B. C3'-endo is the preferred conformation of sugar pucker in A-form RNA, and C2'-endo is the preferred conformation of sugar pucker in B-form DNA and free nucleotides. Unfortunately the coupling constant between the H1' and H2' for C3'-endo sugar pucker is very small, less than 3 Hz, and so is unobservable in large RNA NMR spectra. In contrast, the coupling constant between the H1' and H2' for C2'-endo is large enough to be observed, being greater than 8 Hz, and nucleotides with this conformation are generally observed in RNA NMR spectra. Observation of COSY crosspeaks for C2'-endo conformation within RNA NMR spectra usually indicates non A-form or unstructured regions of RNA.

Glycosidic torsion. The glycosidic conformation of a nucleotide can be determined by the magnitude of intra-nucleotide NOEs between non-exchangeable ribose and base protons, Figure 1.1C. Ribose H1' to base H6 or H8 NOEs are especially useful for the determination of the glycosidic torsion because the H1' to base NOEs occur in an uncrowded region of RNA NMR spectra and the H1' is close to the base protons. *Anti* conformation of the glycosidic torsion results in a H1' to base distance of approximately 3.8 Å, and is the energetically preferred conformation. *Syn* conformation of the glycosidic torsion results in a H1' to base distance of approximately 2.5 Å, and this conformation sometimes occurs in RNA structures. Observation of strong H1' to base NOEs can indicate *syn* glycosidic conformation and unusual RNA structure.

Sequential Assignment of RNA NMR spectra

NMR spectra of RNA contain a wealth of structural information, but unfortunately that information is encoded in a complicated manner. To translate the chemical shifts,

NOEs, and coupling constants observed in RNA NMR spectra into structures, a sequential assignment of RNA NMR spectra is necessary which involves correlating the chemical shifts of resonances to specific nuclei within an RNA molecule (Wüthrich, 1986). Once a specific assignment of the RNA NMR spectra has been made, the intensities of NOEs and magnitudes of coupling constants can be translated directly into distances and dihedral angles of specific parts of the RNA, and then used as constraints in molecular modeling to determine RNA structures (Varani *et al.*, 1996).

Obtaining a sequential assignment of RNA NMR spectra is not always a simple task, as can be demonstrated by considering a 30 nucleotide RNA. There are on average 8 non-exchangeable protons per nucleotide and thus approximately 240 non-exchangeable protons in a 30 nucleotide RNA. Most of the protons within the RNA will give rise to more than one NOE yielding several hundred NOEs in the entire molecule. Each and every NOE represents rough distance information about a specific proton-proton interaction, and if that NOE can be assigned to a specific part of the RNA and its intensity measured, then that distance information can be used in molecular modeling to determine the RNA's structure. If that NOE cannot be assigned and measured, then its presence simply complicates the assignment of other NOEs. When no sequential assignment of RNA NMR spectra can be made, structural studies are not possible.

NOESY Based Sequential Assignment of Non-exchangeable RNA Spectra. Patterns of NOEs in RNA NMR spectra which are due to the chemical structure of RNA and the conformations it adopts can be combined with the known primary structure of the RNA to obtain a sequential assignment of RNA NMR spectra. RNA in A-form helical geometry has a regular repeating structure which gives rise to patterns of inter-nucleotide and intra-nucleotide NOEs which can be used to identify the peaks in an NMR spectra. A list of some of the predicted inter-nucleotide and intra-nucleotide distances for which NOEs can be expected for an A-form RNA helix are shown below:

	Distances (Å)		
Intra-nucleotide		Inter-nucleotide (3' to 5')	
H1'⇔H6/H8	3.5	H1'⇔H6/H8	4.5
H2'⇔H6/H8	4.0	H2'⇔H6/H8	2.0
H3'⇔H6/H8	2.7-3.0	H3'⇔H6/H8	3.0
H5⇔H6	2.4	H6/H8⇔H6/H8	4.5

Table 1.1 Intra-nucleotide and inter-nucleotide distances for which NOEs can be expected between protons in A-form RNA (Arnott & Hukins, 1973; Puglisi, 1989). (Hx⇔Hy is the distance between Hx and Hy in Å.)

NOESY-based sequential assignment of RNA depends upon linking NOEs from protons on one nucleotide to the next nucleotide in the sequence, then repeating the process. Two NOEs (leaps through space) are required using this method to link one nucleotide to the next. The first leap through space is necessary to link the ribose protons of a nucleotide to the base protons of that same nucleotide, and the next leap through space is necessary to link the ribose protons to the base proton that is 3' to it in the RNA strand. Figure 1.2 shows a graphical representation of this where intra-nucleotide NOEs are shown as arrows numbered 1 and inter-nucleotide NOEs are shown as arrows numbered 2. By tracing the NOEs from H6/H8 protons to ribose protons on the same nucleotide, then from those ribose protons to the H6/H8 base proton that is 3' to that ribose, an NMR spectroscopist can follow the sequential assignment of the NMR spectra from one nucleotide to the next 3' nucleotide. This is termed a 'NOESY walk'.

The H1' to base (H2, H6, or H8 protons) region of RNA NOESY spectra is particularly important to the NOESY based sequential assignment of RNA, because the placement of the H1' proton in A-form RNA is such that NOEs can be expected for both inter-nucleotide and intra-nucleotide H1' to base proton interactions (table 1.1). This is advantageous because a single H1' proton can be used to link one nucleotide through space to the next, Figure 1.2. An additional advantage is that the H1' to base NOE region is

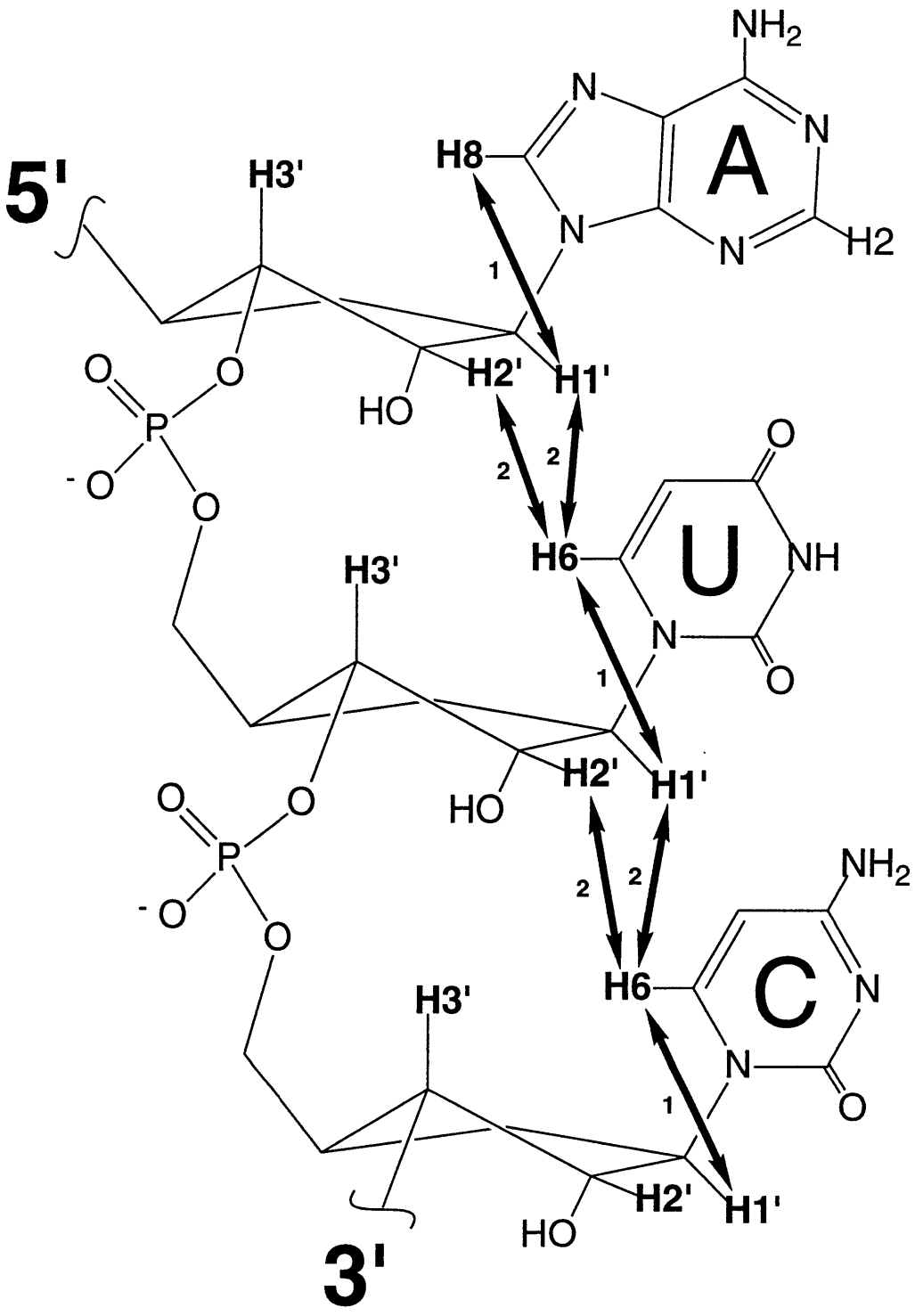


Figure 1.2 Pattern of ribose to base NOEs in A-form RNA that are used in NOESY based sequential assignment of NMR spectra. **1** = intra-nucleotide NOE, **2** = inter-nucleotide NOE.

relatively uncrowded, having only H1' to base NOEs and H5-H6 pyrimidine NOEs, making the assignment of this region much easier than other ribose to base NOE regions. H6 base protons from pyrimidines (Y) are distinguished from H8 base protons from purines (R) by the presence of H5-H6 crosspeaks for the H6 pyrimidine protons in COSY and NOESY spectra. Patterns of purines and pyrimidines in sequence such as 5'-RRYR-3' can then be identified knowing the primary sequence of the RNA as 5'-AGUG-3'.

Other ribose protons can also be used to help in NOESY based sequential assignment, most notably the H2' and H3' ribose protons. The H2' ribose proton is very close to the base proton that is 3' to it in A-form RNA, ≈2.0 Å, and thus gives a very strong NOE to the 3' base proton. This NOE is helpful in RNA sequential assignment because it can yield a second piece of evidence for the identity of the 3' base proton, Figure 1.2. The H3' ribose proton is much like the H1' ribose proton, often having both inter-nucleotide and intra-nucleotide NOEs (table 1.1). Unfortunately both the H2' to base and H3' to base NOEs fall in a very crowded region of RNA NMR spectra, and are thus difficult to assign.

While much of the RNA NMR structural work that has been done to date has relied, at least in part, upon NOESY based sequential assignment, this type of assignment has the drawback that A-form helical structure must be assumed during the assignment procedure. When A-form helical structure does not exist, the assignment procedure breaks down. Since some of the most interesting RNA structures are non A-form, this is a large drawback. Generally the A-form sections of RNAs are assigned leaving holes where non-A-form structure can then be assumed to exist. By assigning as many NOEs and coupling constants in these non A-form regions, the unusual structures can be determined. Much work in the field of heteronuclear NMR spectroscopy has focused on developing through-bond assignment procedures of RNA which rely less on assumptions and are less prone to

error than NOESY based sequential assignment (Dieckmann & Feigon, 1994; Pardi, 1995).

Problems with the application of NMR spectroscopy to larger RNAs

In solution NMR studies of large macromolecules both spectral crowding and increased resonance linewidth combine to prevent assignment of NMR spectra and ultimately limit the size of macromolecules that can be effectively studied (Wagner, 1993; Clore & Gronenborn, 1994). In the extreme case of very large macromolecules spectral crowding becomes severe because of the increased number of resonances, and most heteronuclear NMR experiments are inefficient due to rapid T_2 relaxation. Figure 1.3 shows a comparison of the NOESY spectra of a 30 nucleotide RNA with that of a 65 nucleotide RNA. The NMR techniques of 1994-95 were capable of determining the structure of the 30 nucleotide RNA, but the study of the 65 nucleotide RNA with NMR is still problematic in 1997 due to severe spectral crowding and broad linewidths.

Spectral Crowding. Spectral crowding is a large problem for RNA NMR spectroscopy because of poor distribution of the non-exchangeable RNA proton chemical shifts. While the base, imino, and H1' protons in RNA have good chemical shift dispersion, the remaining protons of the ribose ring, comprising over half of the non-exchangeable protons in RNA, resonate between 4-5 ppm. A 1-dimensional spectra of the nonexchangeable protons of RNA (Figure 1.4) demonstrates this, where the overlap of the H2', H3', H4', H5', and H5" ribose protons results in severe spectral overlap between 4-5 ppm. In a 2-dimensional NOESY spectra of RNA, this 1-dimensional overlap results in severe spectral crowding in NOESY regions where NOEs from the H2', H3', H4', H5', and H5" ribose protons fall, regions **A**, **B**, and **C** of the NOESY spectra of RNA shown in Figure 1.5. While the number of protons within an RNA increases linearly with the number of nucleotides, the spectral width that RNA protons resonate in does not change with the size of the RNA, and this causes proton resonances to pile up upon one

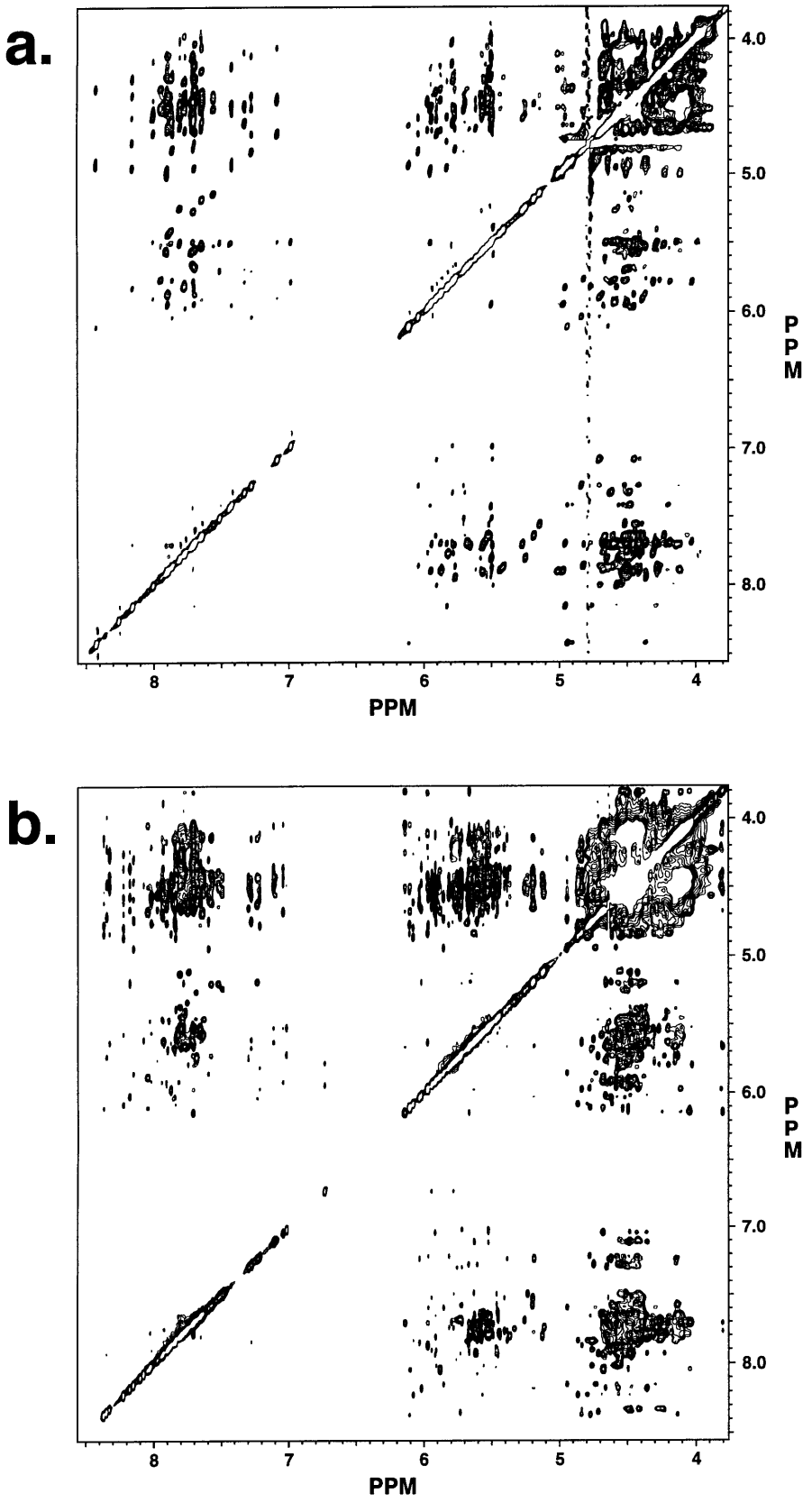


Figure 1.3 200 ms NOESY spectra: a. 30 nucleotide RNA b. 65 nucleotide RNA

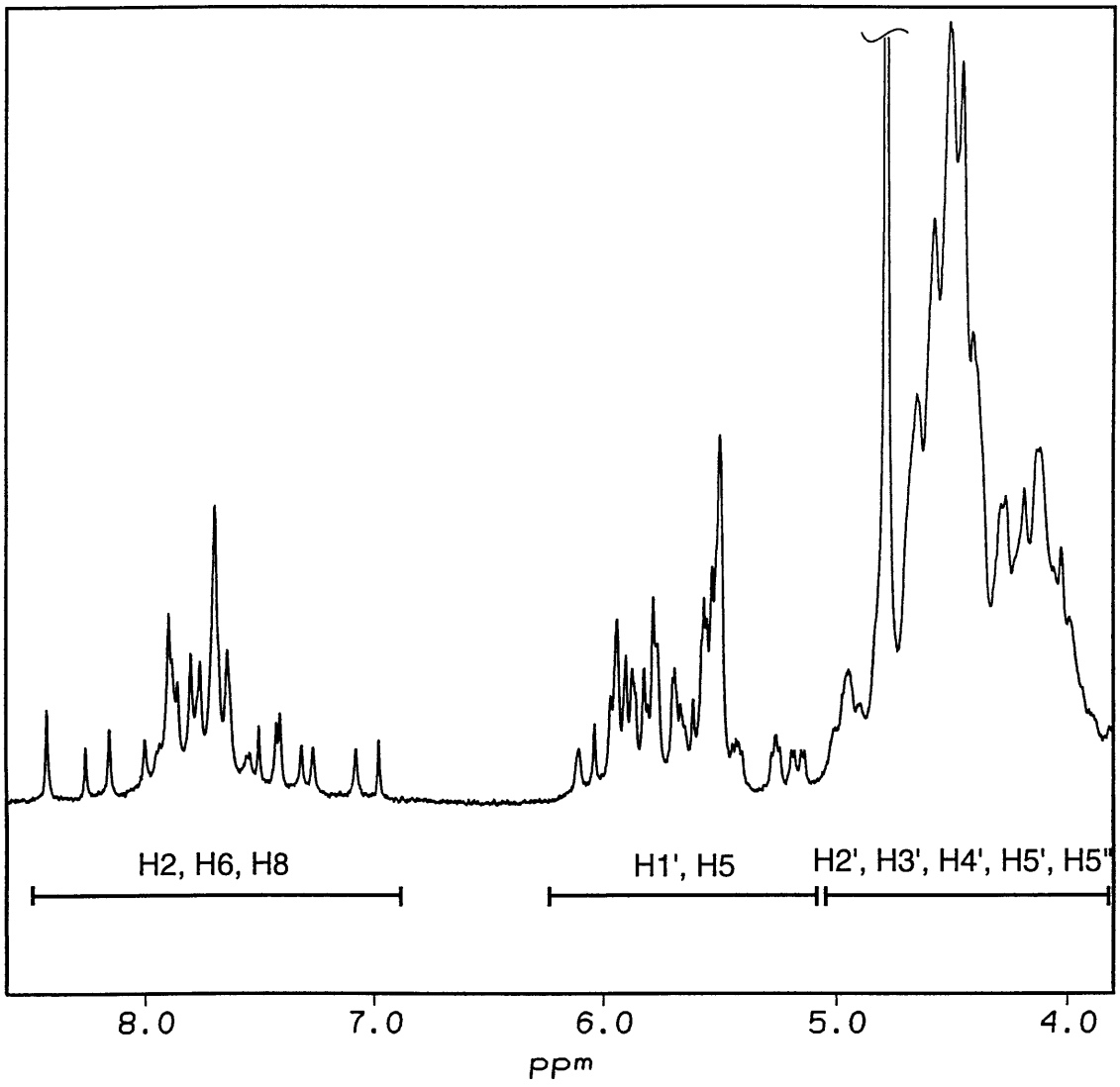


Figure 1.4 Spectral crowding in a 1D NMR spectra of the non-exchangeable RNA protons. While the base, imino, and H1' protons in RNA have good chemical shift dispersion, the remaining protons of the ribose ring, comprising over half of the non-exchangeable protons in RNA, resonate between 4-5 ppm.

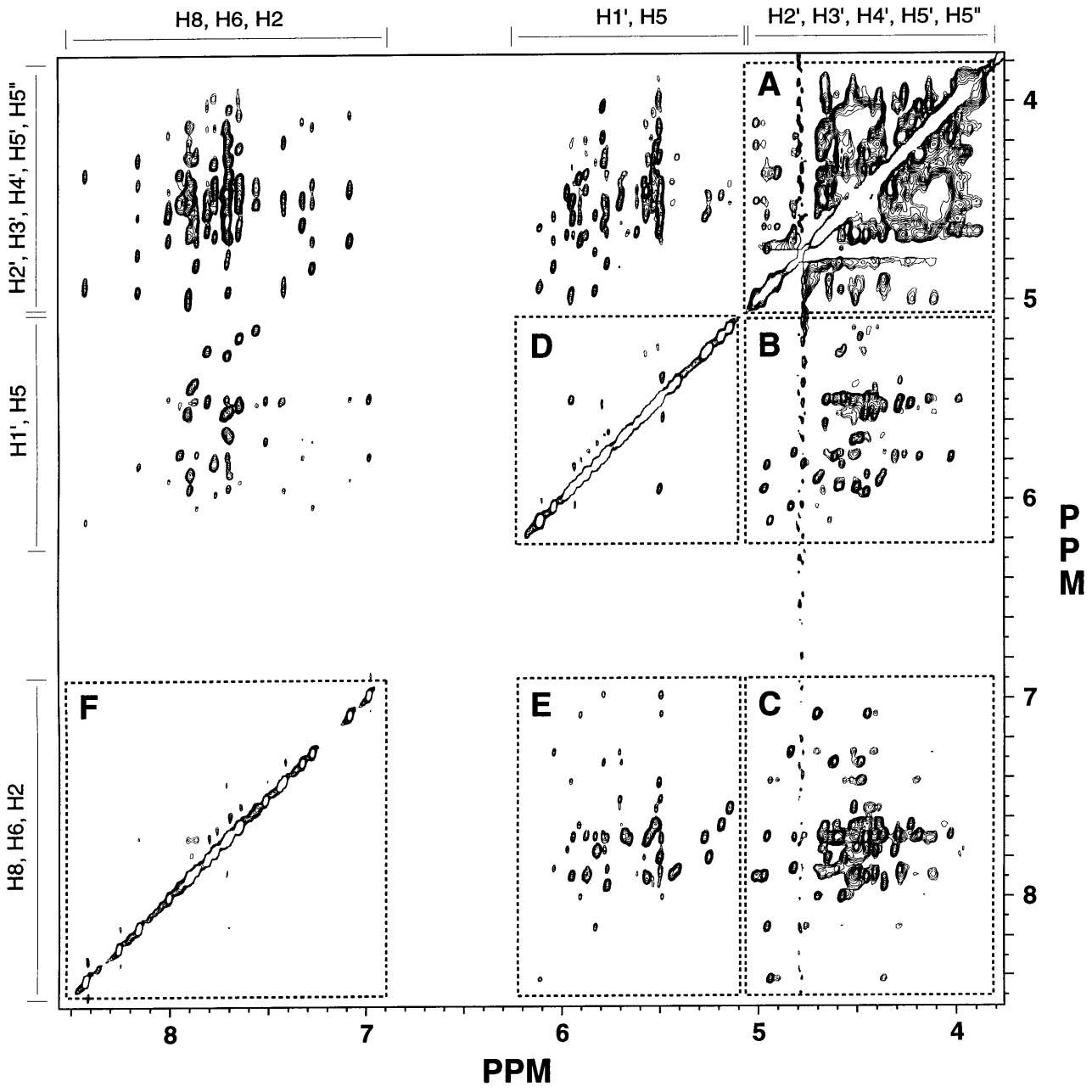


Figure 1.5 2-dimensional NOESY spectra of the non-exchangeable protons of RNA. The overlap of the ribose H2', H3', H4', H5', and H5" proton chemical shifts results in severe spectral crowding in NOESY regions A, B, and C.

another in larger RNA NMR spectra. Spectral overlap in RNA molecules larger than 40 nucleotides becomes so severe that assignment of the ribose 4-5 ppm regions of NMR spectra becomes impossible.

Line Broadening. As the macromolecules studied with NMR spectroscopy increase in size, the linewidths of resonances in NMR spectra become broader. This phenomenon is due to faster relaxation of magnetization in larger macromolecules as opposed to smaller molecules. A detailed discussion of the causes and effects of spin relaxation is outside the scope of this text, but a qualitative discussion of the effects that spin relaxation has on NMR spectra will be presented here. (More detailed discussions of relaxation can be found in Derome (Derome, 1990), Carrington and McLachlan (Carrington & McLachlan, 1967), or Cavanagh *et. al.* (Cavanagh *et al.*, 1996).) NMR experiments rely upon detecting non-equilibrium bulk magnetic moments which relax back to equilibrium with roughly exponential rates termed T_1 and T_2 relaxation rates (longitudinal and transverse) which describe the relaxation of z and x,y magnetization respectively. The lifetime of detectable magnetization during the acquisition phase of an NMR experiment (the duration of detectable signal in the Free Induction Decay or FID of an NMR experiment) determines the linewidth of the resulting NMR lines, and this lifetime depends mostly on the T_2 relaxation rate. Longer lived detectable magnetization produces narrower linewidths in NMR experiments. The rate of relaxation of a certain spin back to equilibrium magnetization is dependent upon fluctuations of local magnetic fields, which in turn are dependent on the thermal motions of a macromolecule. The overall rate of tumbling for a macromolecule is dependent upon the size and shape of the macromolecule, as well as the viscosity and temperature of the medium which it is in. These factors are often summarized as the overall correlation time of the macromolecule τ_c. In this manner, the broadness of NMR resonances can be related to relaxation times of magnetic spins and most importantly the size of the macromolecules. Figure 1.6 shows the fourier transform of the same free

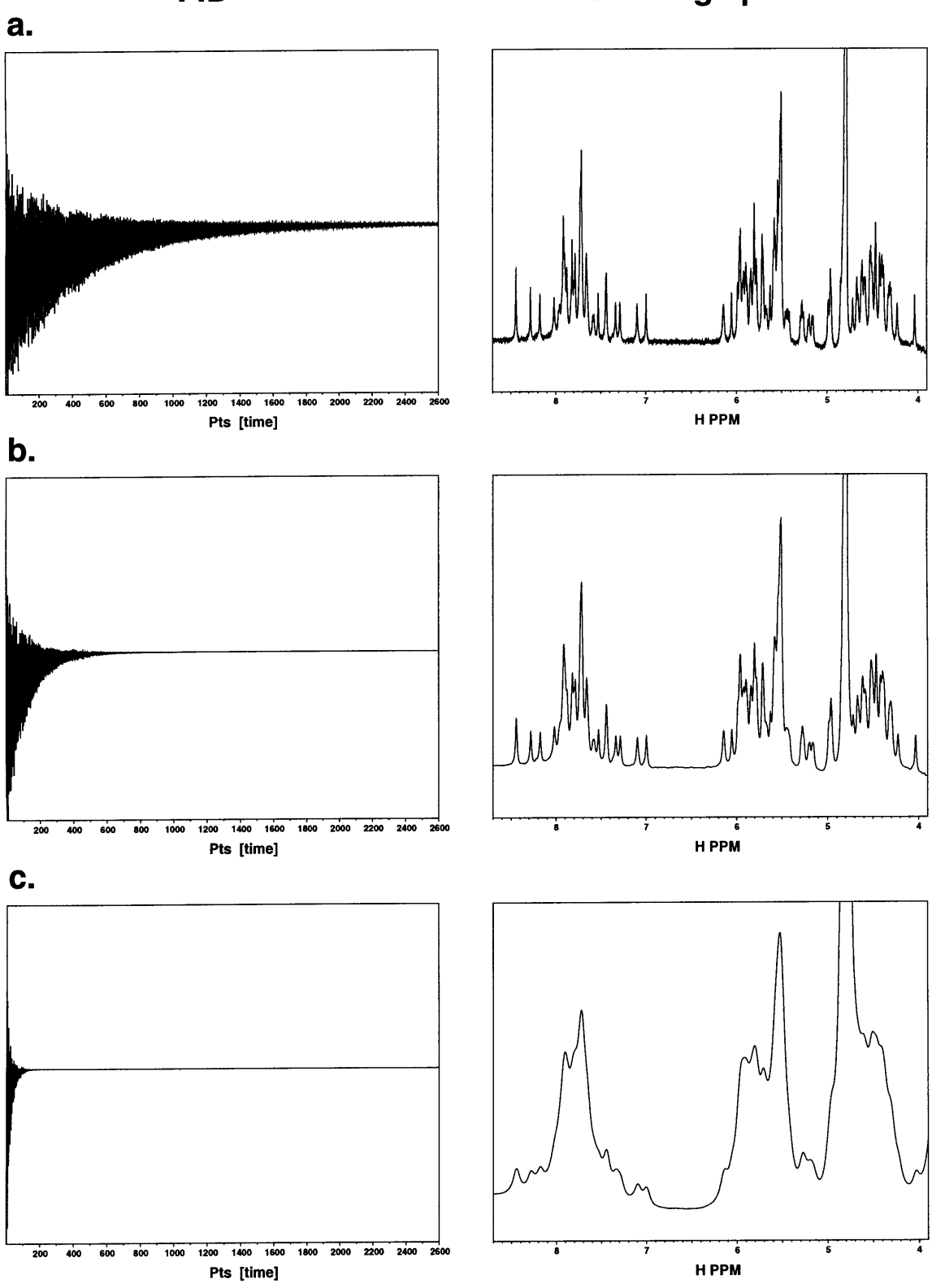


Figure 1.6 Effect of magnetization lifetime on linewidth.

induction decay (FID) with three different simulated T_2 relaxation times (The different T_2 relaxation times were simulated by multiplying the same FID by increasingly negative exponentials.). As lifetimes of the detectable magnetization become shorter in the FIDs, broader linewidths result in the NMR spectra. Peaks that were clearly distinct in the first spectra are blurred together in the last broadest NMR spectra. Similar line broadening effects are seen for larger macromolecules because of their increased correlation times relative to smaller molecules.

Shorter T_2 relaxation times also have an effect on the intensity of the signal that is observed. Relaxation back to equilibrium begins the instant non-equilibrium magnetization is established, so depending on the pulse sequence that is used, significant decay of detectable magnetization can occur by the time acquisition begins, which is the start of the FID. In the case of very large macromolecules with very short T_2's, detectable magnetization can decay so fast that there is no magnetization to detect during the acquisition phase of NMR experiments. This situation is often encountered in heteronuclear NMR spectroscopy of large macromolecules (Kay & Gardner, 1997).

Altering NMR spectra with isotopic labeling

Introduction of above normal levels of stable isotopes into macromolecules can greatly alter their NMR properties while only slightly altering their chemical properties (Kainosho, 1997). This can be utilized to change NMR spectroscopic properties of macromolecules so that their structures can be more easily be studied with NMR. Two types of isotopic labeling which have been used to increase the size of molecules that can be studied with NMR spectroscopy are deuteration and heteronuclear labeling, which can be incorporated into macromolecules either uniformly or specifically.

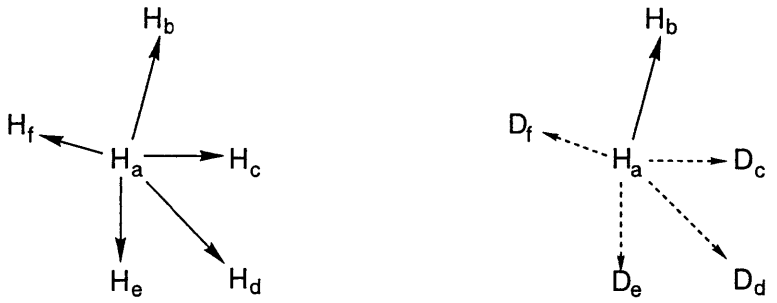


Figure 1.7 Deuteration of hydrogens within macromolecules removes those resonances from NMR spectra and also reduces dipolar relaxation of the remaining protons resulting in narrower linewidths. (solid arrows represent strong ^{1}H-^{1}H dipolar relaxation, and dashed arrows represent weaker ^{2}H-^{1}H dipolar relaxation)

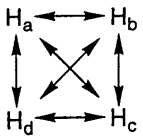
Deuterium Labeling. Deuterium has a natural abundance of 0.015%, while hydrogen has a natural abundance of 99.985%. Increasing the level of deuterium in macromolecules above the natural abundance can have several effects that are beneficial to NMR spectroscopy (Markley *et al.*, 1968; Crespi *et al.*, 1968; LeMaster & Richards, 1988; LeMaster, 1990). For example, replacing a hydrogen with deuterium will make that position silent in proton NMR spectroscopy. Even though deuterium has a spin, the energy of transition for deuterium is sufficiently different from that of hydrogen so that no deuterium resonances are observed in proton NMR spectra. Deuterium can be used to reduce the intensity of a proton or even remove certain resonances from NMR spectra. In addition, deuterium has a gyromagnetic ratio that is approximately 6.5 time smaller than hydrogen. Dipolar interactions which give rise to T_2 relaxation are greatly reduced when deuterium is substituted for hydrogen resulting in longer T_2 relaxation times and narrower linewidths for the remaining hydrogens. There are two types of deuterium labeling which have different effects on NMR spectra, one is termed random fractional deuteration and the other is specific deuteration (LeMaster, 1990).

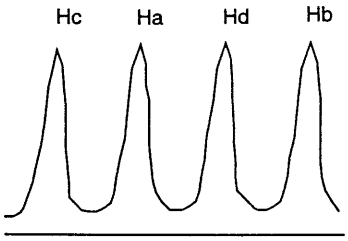
Random fractional deuteration is the term used to describe biological macromolecules which are labeled with deuterium by growing organisms on partially deuterated carbon source in a partial D_2O mixture. Scrambling of the protons of the carbon

source during metabolism places deuterium relatively randomly (depending on the carbon source and the metabolic pathway) on the proteins, nucleic acids, and other molecules produced by the organism during metabolism. Proteins overexpressed in bacteria grown on 75% deuterated medium will have a random fractional deuteration level of approximately 75%, which means that there is about a 75% probability that any nonexchangeable proton on the protein will be deuterium. The advantage of random fractional deuteration is that it greatly reduces dipole-dipole interactions which give rise to relaxation, resulting in narrower linewidths (LeMaster, 1990). The disadvantage of random fractional deuteration is that it reduces the intensity of all of the resonances within a macromolecule (Figure 1.8b). Random fractional deuteration of a biological macromolecule gives rise to an ensemble of deuterated molecules where each position of the macromolecule is deuterated in some of the molecules in the ensemble. The NMR spectra of the entire ensemble of molecules gives rise to all of the resonances of the unlabeled molecule, but with narrower linewidths and reduced intensity due to the level of deuteration. Random fractional deuteration has proven to be very useful in improving the spectra of the exchangeable protons in proteins and nucleic acids, because the nonexchangeable protons are deuterated to a high level reducing dipolar relaxation, but the exchangeable protons are present at 90% protonation due to exchange with solvent and therefore not having reduced intensities in NMR spectra.

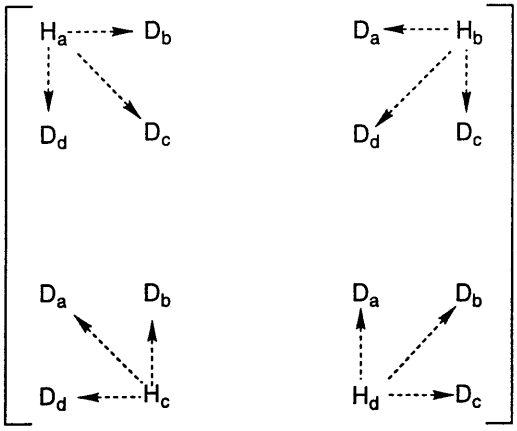
Specific deuteration of macromolecules refers to placing deuterium into specific positions of the macromolecule with either specific metabolic pathways or chemical synthesis, and has somewhat different effects than random fractional deuteration on the NMR spectra of macromolecules (LeMaster, 1990). Specific deuteration results in a homogeneous group of isotopically labeled macromolecules where certain protons have been deuterated and certain others have been left protonated. Deuteration of the specific protons leads to reduced dipolar interactions and narrower linewidths for the remaining protons as it does in random fractional deuteration, but specific deuteration does not result

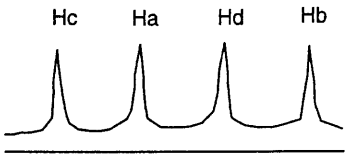
a. Unlabeled



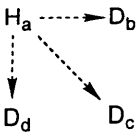


b. Random fractionally deuterated





c. Specifically deuterated



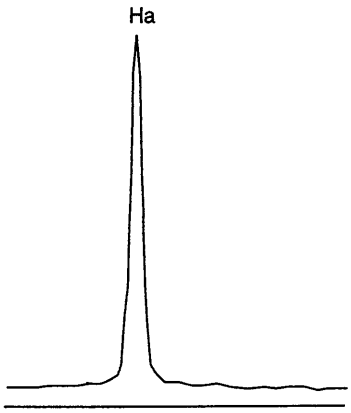


Figure 1.8 Effect of deuteration on NMR spectra. a) Unlabeled: all possible dipolar interactions are present, peaks are at full intensity. b) Random fractionally deuterated: NMR sample is made up of an ensemble of molecules with different deuteration patterns, linewidths are narrower due to reduced dipolar interactions and the intensity of all peaks are reduced because of deuteration. c) Specifically deuterated: linewidths of the remaining protons are narrower because of reduced dipolar interactions and the spectra is simplified by the specific removal of several resonances.

in a uniform reduction of intensity as random fractional deuteration does (Figure 1.8c). Instead, specific deuteration results in a close to 100% reduction in signal for the specifically deuterated protons, and no reduction of intensity for the protons which are not deuterated (Hsu & Armitage, 1992). Because of this, specific deuteration of macromolecules can be used to remove unwanted resonances from NMR spectra reducing spectral crowding and simplifying the interpretation of the remaining resonances.

Heteronuclear or ^{13}C and ^{15}N labeling. ^{13}C and ^{15}N are both spin 1/2 nuclei which have low natural abundance (1.11% ^{13}C, 0.37% ^{15}N). Incorporation of elevated levels ^{13}C and ^{15}N nuclei into macromolecules, termed here 'heteronuclear labeling', results in drastically different effects than deuteration (Wagner, 1993; Clore & Gronenborn, 1994; Pardi, 1995). Rather than reducing dipolar interactions like deuteration, introduction of heteronuclei into macromolecules actually introduces extra scalar and dipolar interactions which increase line broadening by adding large one bond ^{1}H-^{13}C and/or ^{1}H-^{15}N dipolar interactions into the NMR spectra. The advantage of heteronuclear labeling lies not in reducing linewidths or removing resonances from NMR spectra, but in being able to transfer magnetization from protons to heteronuclei and then back to protons through the large one bond scalar couplings between protons directly bound to heteronuclei. This allows multidimensional heteronuclear NMR and through bond assignment of macromolecules.

Multidimensional heteronuclear NMR reduces spectral crowding by spreading proton resonances out over a heteronuclear chemical shift dimension (Figure 1.9). HMQC (Heteronuclear Multiple Quantum Correlation spectroscopy) NMR spectra correlate ^{1}H chemical shifts with the chemical shifts of ^{13}C or ^{15}N nuclei which the ^{1}H is attached to. By combining HMQC NMR experiments with two dimensional proton-proton NMR experiments such as NOESY, COSY (Correlational SpectroscopY), or TOCSY (Total Correlational SpectroscopY), a third heteronuclear dimension can be added to the NMR

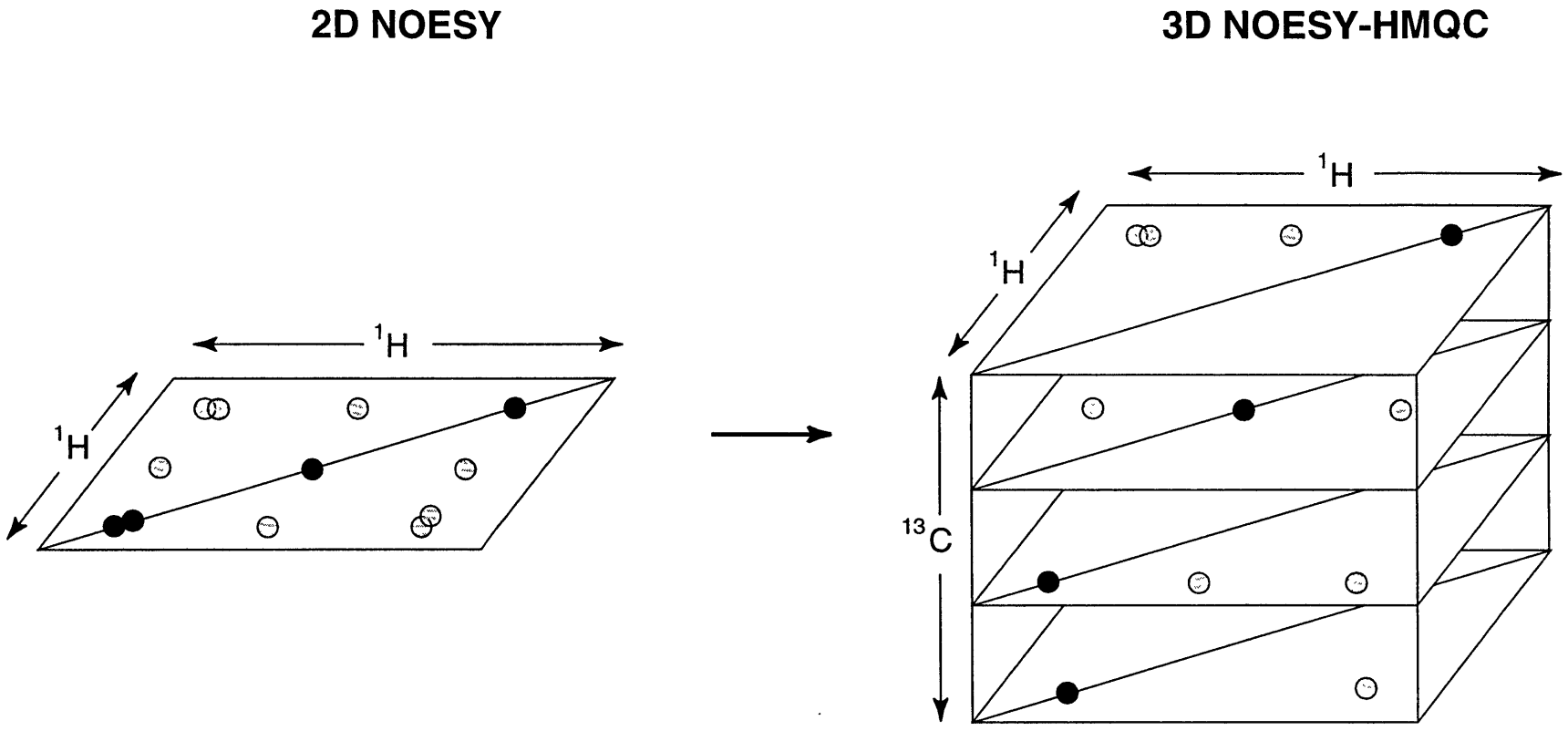


Figure 1.9 Heteronuclear labeling reduces spectral crowding by spreading out proton information over a heteronuclear dimension.

spectra. This additional chemical shift axis serves to reduce spectral crowding by increasing the overall spectral volume over which the same information is spread out over, and also separates proton information by the chemical shift of the carbon or nitrogen that it is directly attached to. The reduction in spectral overlap that results from multidimensional heteronuclear NMR has increased the size of proteins that can be studied with NMR from 10 to 25 kDa (Wagner, 1993; Clore & Gronenborn, 1994).

Heteronuclear labeling also enables through bond assignment of labeled macromolecules, which is less prone to error in interpretation than NOESY based assignment (Clore & Gronenborn, 1994). Heteronuclear TOCSY NMR experiments allow the transfer of magnetization between distantly separated protons by transferring magnetization through large one bond proton-heteronuclei and heteronuclei-heteronuclei scalar couplings. Observation of correlations between two protons in an HCCH-TOCSY NMR experiment (Figure 1.10) establishes a through bond relationship between the protons, indicating that they are in the same spin system (such as an amino acid side chain in a protein, or a ribose moiety from a nucleic acid).

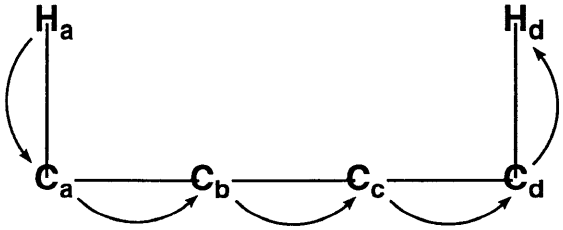


Figure 1.10 HCCH-TOCSY NMR experiments transfer magnetization from protons to heteronuclei and then back to protons through large one bond scalar couplings. Observation of correlations between H_a and H_d in an HCCH-TOCSY NMR experiment establishes a through bond relationship between the protons, indicating that they are in the same spin system.

Heteronuclear and deuterium labeling combined. Within the last five years heteronuclear labeling and deuteration have been combined to successfully study proteins

on the order of 40-60 kDa (Kay & Gardner, 1997; Grzesiek *et al.*, 1993; Markus *et al.*, 1994; Yamazaki *et al.*, 1994b; Farmer & Venters, 1995; Muhandiram *et al.*, 1995; Nietlispach *et al.*, 1996; Rosen *et al.*, 1996; Matsuo *et al.*, 1996; Shan *et al.*, 1996; Yamazaki *et al.*, 1997). This approach combines the benefits of deuteration with multidimensional heteronuclear NMR, and has started to be applied to RNA in the form of random fractional deuteration of RNA in combination with heteronuclear labeling (Batey *et al.*, 1996; Nikonowicz *et al.*, 1997). Deuteration greatly increases the T_2's of heteronuclei by replacing the strong dipolar coupling between ^{13}C and hydrogen with the weaker ^{13}C-deuterium interaction. The increase in the lifetime of carbon magnetization enables heteronuclear NMR experiments to work on larger protein molecules than would be possible without the advantages of deuteration.

Uniform vs. Specific Isotopic Labeling. Isotopic labeling for NMR purposes can be divided into roughly two types of labeling, uniform labeling and specific labeling, with each type having different advantages and disadvantages. For the purposes of this text, uniform isotopic labeling is defined as uniform incorporation of an isotope or isotopes into every position of a macromolecule with a roughly equal percentage of incorporation for each position, and specific isotopic labeling is defined as non-uniform incorporation of an isotope or isotopes into the positions of a macromolecule with unequal percentages of incorporation for the different possible positions. With these definitions random fractional deuteration will be considered uniform isotopic labeling, because in a 75% random fractionally deuterated protein, each hydrogen position has a roughly 75% probability of being deuterated, and incorporation of ^{13}C labeled G into otherwise unlabeled RNA will be considered specific labeling since the ^{13}C labeling will be approximately 100% for G carbons and 1% for the other carbons. These definitions make a distinction between the types of effects that are observed in NMR spectra where uniform isotopic labeling will affect the NMR spectra of every atom of the types that are labeled in the macromolecule, and specific labeling will only affect a subset of all of the atoms in the NMR spectra.

Chapter 1

Much of the NMR isotopic labeling that has been used over the last decade to extend the size of macromolecules that can be studied with NMR has been of the uniform isotopic labeling type. Uniform heteronuclear labeling of proteins and nucleic acids has greatly increased the size of macromolecules that can be studied with NMR (Clore & Gronenborn, 1994). One advantage of uniform isotopic labeling is that the benefits of isotopic labeling are experienced for the entire molecule in a single NMR sample, rather than for only certain parts of the molecule as in specific isotopic labeling. This allows all of the possible NMR information of a macromolecule to be obtained from a single NMR sample, or at most a few NMR samples, and because of this, less work need be spent preparing NMR samples when uniform isotopic labeling is used. Another advantage of uniform isotopic labeling is that it can be accomplished in a fairly straightforward manner for most biological macromolecules by harvest of the macromolecules or their precursors from bacteria grown on isotopically labeled minimal media (Kainosho, 1997). These two advantages make uniform isotopic labeling the most attractive option for the study of proteins or RNA with NMR until the size of macromolecules increase to the point that spectral crowding and relaxation problems become insurmountable by uniform isotopic labeling alone.

As the size of the macromolecules to be studied with NMR spectroscopy increases, specific isotopic labeling becomes a more attractive labeling option because it can be used to simplify NMR spectra by focusing on a subset of all of the resonances of a macromolecule. Specific isotopic labeling can be designed so that the subset of resonances that are observed in NMR spectra will reduce spectral overlap, simplify assignment, and contain the most structural information possible for the remaining resonances. For extremely large macromolecules where no specific assignment of resonances can be made due to spectral crowding, the use of specific isotopic labeling to focus on a subset of resonances can make assignment of NMR spectra possible where uniform labeling fails. In addition, specific deuterium labeling can be used to increase the relaxation times of the remaining protons

without the loss of signal that is inherent with random fractional deuteration. This can be valuable when large macromolecules with relaxation problems are studied. Unfortunately the synthetic methods required to accomplish specific isotopic labeling of macromolecules is usually more difficult and involved than uniform isotopic labeling, requiring either chemical synthesis or growth of bacteria on specifically labeled media (usually obtained by chemical synthesis). Even though specific isotopic labeling is chemically more difficult to accomplish, as NMR spectroscopist start to study larger macromolecules, specific labeling techniques have been utilized more frequently (Kay & Gardner, 1997).

Goal: Incorporate isotopic labels into RNA in order to facilitate NMR studies

The nature of the problems encountered with RNA NMR spectroscopy make RNA a good candidate for the application of specific isotopic labeling patterns. Spectral overlap problems encountered in RNA are severe for many reasons which specific isotopic labeling can alleviate. Unlike proteins, where there are 20 residues with widely varying side chains, RNA has only 4 residues, A, G, U, and C nucleotides. This means for an RNA and protein of the same number of residues, the RNA will most often have more repeated residues than the protein. In addition, each nucleotide of RNA contains a ribose moiety whose proton chemical shifts do not vary significantly with the type of base which is attached to the ribose. Because of this, the identification of the type of nucleotide which a ribose proton belongs to is not a simple matter since the ribose resonances for all nucleotide types overlap. This is unfortunate because ribose protons comprise over half of the non-exchangeable protons in RNA, and five out of the six non-exchangeable ribose protons overlap in proton chemical shift. This results in spectral crowding in RNA NMR spectra which is largely due to the overlap of different types of ribose protons and also overlap of ribose protons from different types of nucleotides. Isotopic labeling of RNA so that specific types of ribose protons and/or specific types of nucleotides were labeled differently would result in a great reduction of the spectral crowding encountered in RNA NMR

spectroscopy, and might make the study of larger RNA molecules with NMR possible. The work on this project was initiated to explore what utility specific isotopic labeling of RNA might have in the study of larger RNAs with NMR spectroscopy.

Good methods for incorporating a wide variety of specific deuterium and heteronuclear labeling patterns into RNA did not exist at the time this project was initiated, so a large part of the work described in this thesis deals with developing procedures which can be used to specifically isotopically label RNA. Previous procedures used to produce isotopically labeled nucleoside triphosphates and nucleosides, the precursors of RNA, were used as a starting point for the design of syntheses which would produce specific isotopic labeling patterns. In the end the goal was to develop RNA isotopic labeling patterns which would be of general use in the study of RNA structure with NMR spectroscopy, so much consideration was given to the cost, level of difficulty, and flexibility of the synthesis that was to be developed since if the cost of the synthesis was too high, the amount of work necessary too much, or the synthesis too inflexible, it would be unlikely that the synthesis would become generally used in the field of RNA NMR spectroscopy.

Chapter 2

2. Synthesis of d4-NTPs

As described in Chapter 1, structural studies of RNA using proton NMR spectroscopy are limited in the size of RNAs that can be effectively studied to about 20-35 nucleotides because of spectral crowding and linewidth problems. Isotopic labeling can be used to extend the size of macromolecules that can be studied with NMR by reducing spectral crowding and linewidth problems in larger macromolecules. It was the purpose of the research described in this thesis to explore the use of specific incorporation of ^{2}H, ^{13}C, and ^{15}N labels into RNA to extend NMR structural studies to larger RNA molecules. The first RNA isotopic labeling pattern to be synthesized in this thesis was the 3',4',5',5"-*d4*-RNA labeling pattern, which was designed to simplify homonuclear NMR spectra by specific removal of ribose protons which cause spectral crowding. By attempting to first synthesize a labeling pattern that only contained deuterium labels, it was possible to work out the problems of incorporating specific labels into RNA with relatively inexpensive deuterated starting materials as opposed to more expensive heteronuclear labeled starting materials.

The 3',4',5',5"-*d4*-RNA (D4-RNA) labeling pattern was designed to alleviate the spectral crowding and large linewidth problems that are encountered in large RNA NMR spectroscopy. The spectral crowding problems of RNA are quite evident in the one dimensional spectra of the non-exchangeable protons of RNA shown in Figure 1.4 in Chapter 1. While the base, imino, and H1' protons in RNA have good chemical shift dispersion, the remaining protons of the ribose ring, comprising over half of the non-exchangeable protons of RNA, resonate between 4-5 ppm. In a two dimensional NOESY spectra of RNA this overlap of the H2', H3', H4', H5', and H5" resonances results in severe spectral crowding in regions **A**, **B**, and **C** shown in Figure 1.5 of Chapter 1. NOEs falling in these regions can arise from interactions of any of the five types of ribose protons

that resonate between 4-5 ppm with other RNA protons. Selective deuteration of the ribose 3', 4', 5', and 5" protons would simplify RNA NMR spectra while still retaining much of the structural information available from RNA NMR spectroscopy. Glycosidic torsion, sugar pucker, and base pairing information can be obtained from NMR spectra using only the base protons, the ribose H1' and H2' protons, and the exchangeable protons as summarized in Chapter 1 Figure 1.2. Deuteration of the H3', H4', H5', and H5" protons would still permit this information to be obtained, but the spectral assignment would be greatly simplified due to the removal of over half of the non-exchangeable RNA protons from NMR spectra. In addition, the D4-RNA labeling pattern would reduce the linewidths of the remaining protons by substituting strong ^{1}H-^{1}H dipolar interactions with weaker ^{2}H-^{1}H interactions. The reduced dipolar relaxation of the remaining protons of the D4-RNA would result in longer relaxation times and narrower linewidths. For these reasons the D4-RNA labeling pattern would be a valuable tool for NMR structural studies of large RNAs, but unfortunately common methods for isotopically labeling RNA would not easily allow for selective deuteration of the 3', 4', 5' and 5" protons of RNA. New synthetic methods needed to be developed in order to allow for selective isotopic labeling of RNA in the D4-RNA labeling pattern, and also to make other selective isotopic labeling patterns possible.

Strategy

NMR structural studies of RNA usually require between 0.25-1.5 μmole of purified RNA to prepare a single RNA NMR sample. RNA can be synthesized in these NMR quantities by either chemical RNA synthesis (Wincott *et al.*, 1995) or T7 RNA polymerase catalyzed transcription (Wyatt *et al.*, 1991). These methods require either phosphoramidites or nucleoside triphosphates (NTPs) respectively to produce RNA, and so to isotopically label RNA it is first necessary to develop procedures for isotopically labeling these starting subunits. Although chemical RNA synthesis can incorporate

different types of isotopic labels at each nucleotide of an RNA sequence, it is not the most generally used method of incorporating isotopic labels into RNA. This is partially because chemical RNA synthesis has a low coupling yield, lower than chemical DNA synthesis, making it difficult to synthesize RNAs larger than 50 nucleotides in significant quantities, and partially because chemical RNA synthesis requires phosphoramidites as starting materials, which require several chemical steps to synthesize from isotopically labeled nucleosides. T7 RNA polymerase catalyzed transcription is by far the preferred method of incorporating isotopic labels into RNA, since it has the ability to produce NMR quantities of short RNA sequences as low as 12 nucleotides and also RNA sequences larger than 400 nucleotides. The nucleoside triphosphates required for T7 transcription reactions can be produced in a single enzymatic phosphorylation reaction from isotopically labeled nucleoside monophosphates derived from biological sources (Batey *et al.*, 1992), making the synthesis of the NTPs required for T7 transcription much easier than the synthesis of the phosphoramidites required for chemical RNA synthesis. T7 RNA polymerase catalyzed transcription was chosen as the method for synthesizing the RNA labeling patterns described in this thesis because the 3',4',5',5"-*d4*-RNA labeling pattern was to be applied to the study of RNAs larger than 50 nucleotides, and also to minimize the synthetic effort required to produce RNA.

Since T7 RNA polymerase catalyzed transcription (Wyatt *et al.*, 1991) was to be used to synthesize the 3',4',5',5"-*d4*-RNA labeling pattern, the goal was to prepare 3',4',5',5"-*d4*-NTPs (*d4*-NTPs). One method for producing isotopically labeled NTPs is to harvest total cellular RNA from bacteria that have been grown on isotopically labeled media, enzymatically digest the RNA to nucleoside monophosphates (NMPs), and finally enzymatically convert the NMPs to NTPs. Uniform ^{13}C, ^{15}N, and $^{13}C/^{15}N$ labeling of NTPs has been accomplished by this method by growing *E. coli* or *M. methylotrophus* on ^{13}C and/or ^{15}N labeled media (Nikonowicz *et al.*, 1992; Hoffman & Holland, 1995; Batey *et al.*, 1995). Unfortunately it is difficult to incorporate specific isotopic labels with this

method due to isotopic scrambling during biosynthesis. A second method that can be used for producing deuterated NTPs involves selective oxidation of one of the hydroxyl groups of ribose, followed by reduction with deuteride. This method has been used to deuterate both the H5' and H5" of thymidine, as well as the H3' of adenosine (Frank *et al.*, 1991; Robins *et al.*, 1990). Although schemes using selective oxidation/reduction of nucleosides to incorporate deuterium labels work well for incorporation of deuterium on a single carbon, they become cumbersome when applied to deuteration of multiple carbons, and also do not readily offer the possibility for ^{13}C and ^{15}N incorporation. A third method for incorporating isotopic labels into nucleotides is to isotopically label the ribose or base separately, followed by chemical coupling to form a nucleoside (Vorbrüggen *et al.*, 1981). This method allows a large flexibility in isotopic labeling since the ribose or the bases can be prepared and labeled separately, and then combined. A variety of isotopic labels have been incorporated into nucleosides using this general scheme including ribose deuterated by Raney nickel-D_2O exchange (Földesi *et al.*, 1992), H1 deuterated ribose (Toyama *et al.*, 1993), H2 deuterated ribose (Cook & Greenberg, 1994), and ribose ^{13}C labeled at the C1 or C2 (Lancelot *et al.*, 1993; Kline & Serianni, 1990). This method also presents problems, however, since to chemically couple ribose to a base, D-ribose must first be protected as 1-O-acetyl-2,3,5-tri-O-benzoyl-b-D-ribofuranoside or some equivalent, which is a multistep synthetic process. Furthermore, once nucleosides have been formed chemically, they must still be converted into NTPs, which requires additional chemical and enzymatic steps (SantaLucia *et al.*, 1995). Another method that has been used, and the one that we chose to adopt, is isotopic labeling of the ribose or base separately, followed by enzymatic coupling to form a nucleotide. Enzymatic coupling allows the direct conversion of ribose into nucleotides or nucleosides without the use of protecting groups, which greatly simplifies the synthetic effort required to produce isotopically labeled NTPs from ribose. This strategy has been used to produce 1'-^{14}C-AMP and 1'-^{3}H-AMP starting with isotopically labeled ribose and adenine using adenine phosphoribosyltransferase (Parkin *et*

al., 1984). Purine nucleoside phosphorylases have also been used to couple C8-^{13}C-purines to ribose-1-phosphate to form C8-^{13}C-purine nucleosides (SantaLucia *et al.*, 1995).

To synthesize 3',4',5',5"-*d*4-RNA we have developed a synthetic strategy that uses both chemical and enzymatic methods as outlined in Figure 2.1. By using T7 RNA polymerase catalyzed transcription to produce large RNAs we have avoided the chemical protection steps required to convert nucleosides into phosphoramidites. Greater flexibility of isotopic labeling is afforded by chemically synthesizing the isotopically labeled ribose separately and then coupling that ribose to bases. With this strategy the isotopically labeled ribose and bases can be derived by a variety of methods, and then coupled to form nucleotides. The use of enzymatic synthesis accomplishes coupling of ribose to the bases in a single coupled enzymatic reaction, avoiding the multistep, laborious process of coupling ribose to bases chemically. One drawback of using enzymatic synthesis to accomplish coupling of ribose to bases was that seven enzymes needed to be purified since they were not commercially available in forms that could be used. While this represented a major obstacle, it was reasoned that by expending a larger initial amount of effort in purifying the enzymes and working out the procedures for the enzymatic reactions, the production of subsequent isotopic labeling patterns would be simplified. This is not the case for chemical coupling of ribose to bases since each time it is repeated the same amount of work must be expended to produce nucleotides.

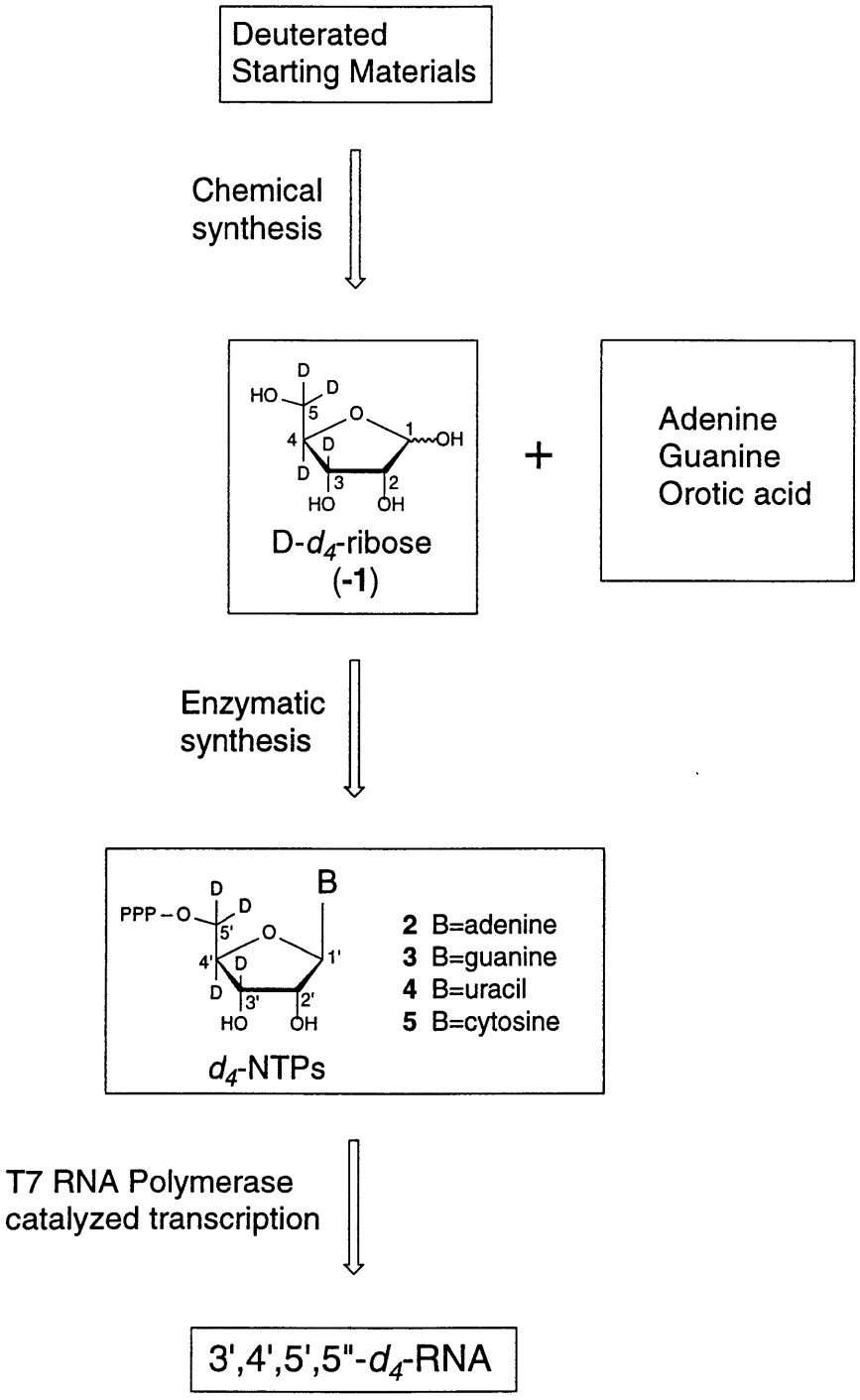


Figure 2.1 Synthetic strategy for the synthesis of 3',4',5',5"-d_4-RNA

Chapter 2

Chemical Synthesis.

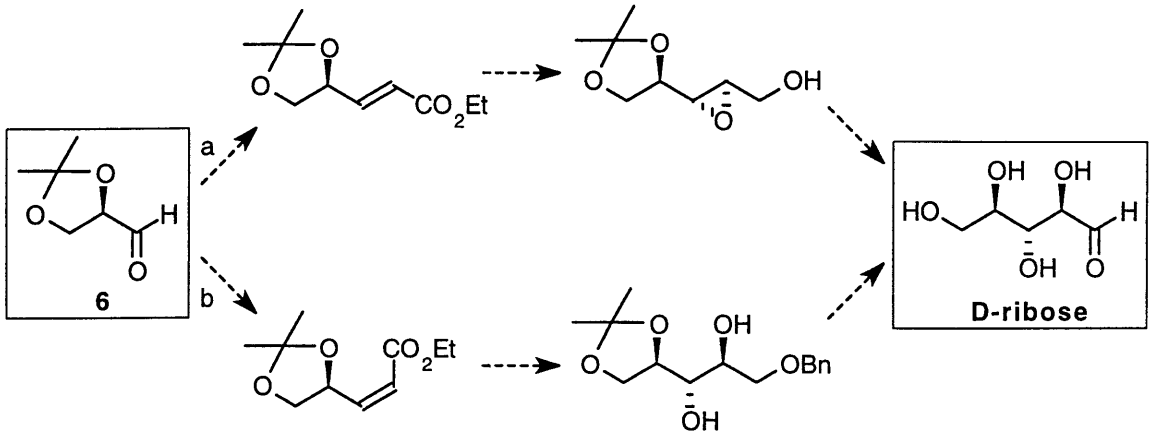


Figure 2.2 Schematic of previous syntheses of D-ribose from 2,3-O-isopropylidene-D-glyceraldehyde (**6**) using a) a route that utilizes Sharpless asymmetric epoxidation, or b) a route that utilizes osmium tetraoxide dihydroxylation.

For the chemical synthesis, our strategy was to build upon well-established carbohydrate syntheses to produce D-*d*4-ribose (**-1**) from commercially available, isotopically labeled starting materials. The ribose synthesis was based on acyclic carbohydrate syntheses, where 2,3-O-isopropylidene-D-glyceraldehyde (**6**) is extended by a Wittig reaction, and the resulting double bond is then dihydroxylated with either osmium tetroxide dihydroxylation (Cha *et al.*, 1984) or Sharpless epoxidation with subsequent ring opening, Figure 2.2 (Ko *et al.*, 1990; Katsuki *et al.*, 1982). The desired labeling pattern can be obtained if the protected D-glyceraldehyde is prepared perdeuterated, since the 3, 4, and 5 carbons of ribose will arise from D-glyceraldehyde. Unfortunately, common methods of synthesizing 2,3-O-isopropylidene-D-glyceraldehyde, such as the periodate oxidation of 1,2:5,6-di-O-isopropylidene-D-mannitol (Jackson, 1988), do not conveniently allow for deuteration, so a method for obtaining perdeuterated 2,3-O-isopropylidene-D-glyceraldehyde (**7**) was needed.

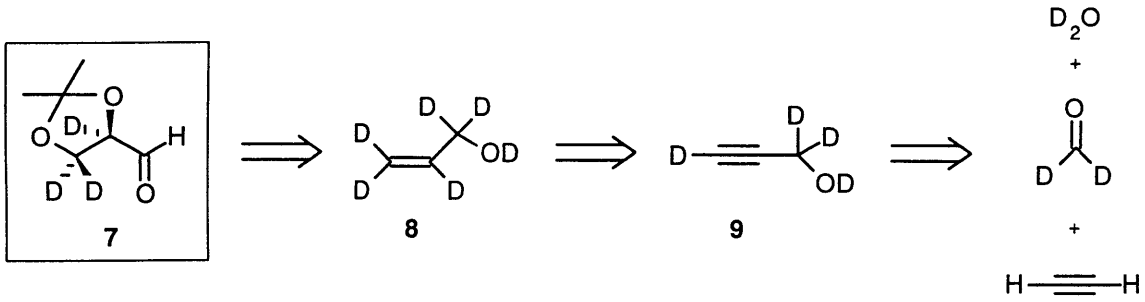


Figure 2.3 Retrosynthetic analysis of the synthesis of perdeuterated 2,3-O-isopropylidene-D-glyceraldehyde (**7**) using D_2O, deuterated formaldehyde, lithium aluminum deuteride, and acetylene as starting materials.

Initial attempts at preparing protected perdeuterated-D-glyceraldehyde (**7**) used acetylene, deuterated paraformaldehyde, lithium aluminum deuteride, and D_2O as starting materials. The retrosynthetic strategy is outlined in Figure 2.3. Acetylene and deuterated paraformaldehyde would be condensed to form 1,1-*d*2-propargyl alcohol, and then the ethynyl proton of the alcohol would be exchanged with D_2O under basic conditions to produce perdeuterated propargyl alcohol (**9**) (McMichael, 1967; Baldwin & Black, 1983). The perdeuterated propargyl alcohol would then be reduced with lithium aluminum deuteride, and the resulting aluminum-allyl alcohol adduct would be hydrolyzed with D_2O to form perdeuterated allyl alcohol (**8**). Perdeuterated allyl alcohol could then be stereospecifically converted into a protected form of glycerol by Sharpless asymmetric epoxidation and subsequent ring opening (Klunder *et al.*, 1986; Ko *et al.*, 1990), or asymmetric dihydroxylation of the protected allyl alcohol (Sharpless *et al.*, 1992). The protected glycerol would then be oxidized by Swern oxidation to form protected perdeuterated D-glyceraldehyde (**7**). The lithium aluminum hydride reduction of propargyl alcohol occurs with regioselectivity (Baldwin & Black, 1983), which allows the synthesis of virtually any deuteration pattern of D-glyceraldehyde by using this synthetic strategy with different combinations of deuterated and protonated starting materials. After much experimentation it was determined that this approach would require at least 7-10 synthetic

steps with some of the steps suffering from very low yield, making this strategy an impractical starting point for the synthesis of RNA.

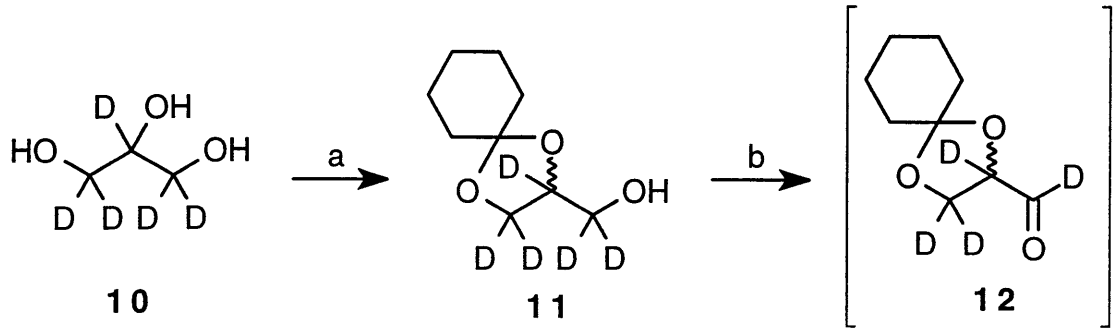


Figure **2.4**: Two step racemic synthesis of the cyclohexyl ketal of perdeuterated D,L-glyceraldehyde (**12**) from deuterated glycerol (**10**). a) cyclohexanone, $(MeO)_3CH$, H^+ b) oxalyl chloride, DMSO, Et_3N.

A more practical two step synthesis of protected perdeuterated D,L-glyceraldehyde (**12**) was developed that utilized commercially available deuterated glycerol (**10**) as a starting material, Figure 2.4. First deuterated glycerol (**10**) was protected as the cyclohexyl ketal (**11**), then Swern oxidation was used to oxidize the unprotected alcohol to form deuterated D,L-glyceraldehyde (**12**). The synthesis is racemic, but since the ultimate product of the chemical synthesis, D,L-ribose, can be subsequently resolved by the enzymes used to convert D-ribose into NTPs, the disadvantage of a racemic synthesis was offset by the high total yield of D-glyceraldehyde and ease of preparation. The cyclohexyl ketal of perdeuterated D,L-glyceraldehyde (**12**) was unstable to silica gel purification so the unpurified product of the Swern oxidation was submitted directly to Wittig reactions.

Synthesis of D,L-3',4',5',5"-*d4*-ribose (±**1**)

The synthesis of D,L-*d4*-ribose (±**1**) from glycerol (U-d8) (**10**) is shown in Figure 2.5. Glycerol (U-d8) (**10**) is converted into a protected form of D,L-glyceraldehyde (**12**) as described above, and then coupled to the Wittig reagent (carbethoxymethylene)triphenylphosphorane ($Ph_3P{=}CHCO_2Et$). The Wittig reagent

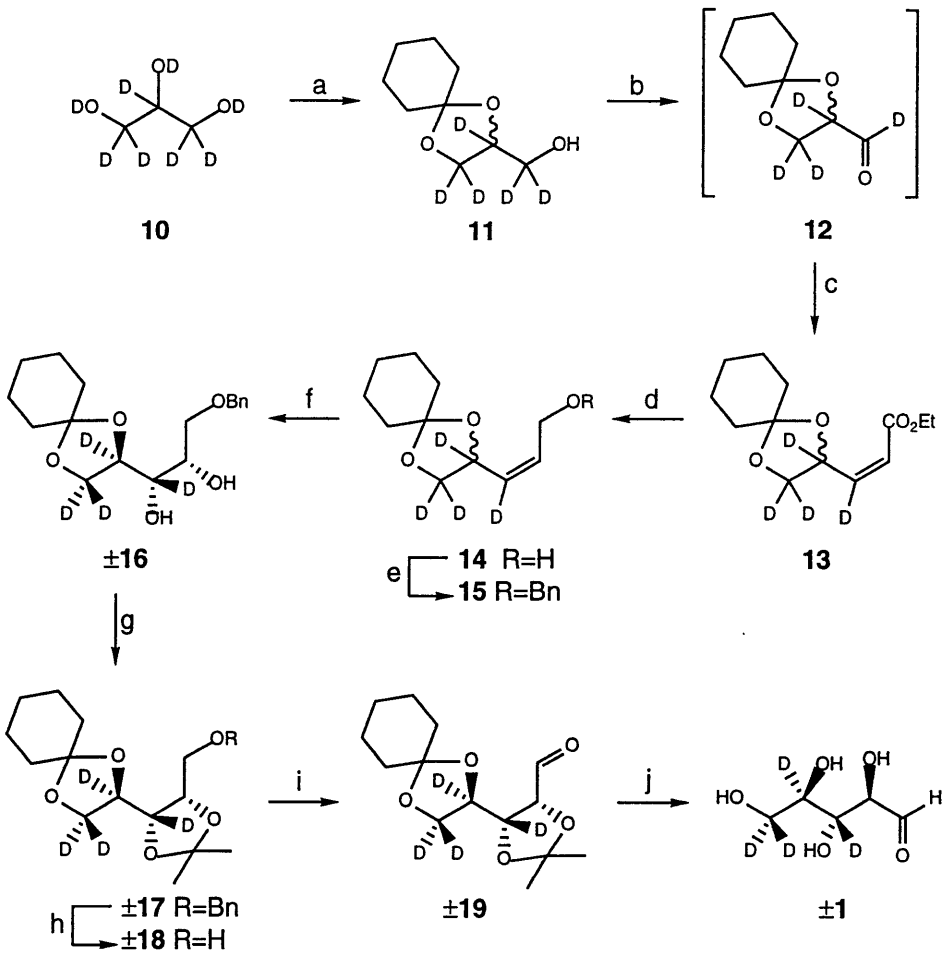


Figure 2.5 Synthesis of D,L-3',4',5',5"-$^2H_4$-ribose (±**1**).
(a) cyclohexanone, $(MeO)_3CH$, H^+ (b) oxalyl chloride, DMSO, Et_3N
(c) $Ph_3P{=}CHCO_2Et$, MeOH (d) DIBAL-H, CH_2Cl_2 (e) BnBr, NaH,
$(nBu)_4NI$, THF (f) OsO_4, N-methyl morpholine N-oxide, acetone:H_2O (8:1)
(g) 2-methoxy propene, H^+ (h) Pd on C, H_2 (i) oxalyl chloride, DMSO,
Et_3N (j) H^+, THF, H_2O

Chapter 2

$Ph_3P{=}CHCO_2Et$ was used in this synthesis because it can be derived from ethyl bromoacetate (Isler *et al.*, 1957), which is commercially available in several ^{13}C labeled forms offering the possibility of ^{13}C labeling of the C1 and/or the C2. In this way the C3, C4, and C5 carbons of ribose are derived from glycerol (U-d8), and the C1 and C2 carbons are derived from a Wittig reagent that can be ^{13}C labeled if desired. Protection of glycerol (U-d8) (**10**) is followed by Swern oxidation of ketal (**11**), resulting in D,L-1,2,3,3'-*d*4-glyceraldehyde ketal (**12**). The triethylamine from the Swern oxidation was carefully neutralized to prevent deuterium exchange by enolization that can occur under basic protic conditions (Ko *et al.*, 1990). The crude aldehyde was unstable to silica gel purification, and was directly submitted to a Wittig reaction with $Ph_3P{=}CHCO_2Et$ in methanol at 0° C, which preferentially forms the Z olefin (Katsuki *et al.*, 1982), to produce (**13**) in a 76% yield. Reduction of the ester with DIBAL-H, and protection of the resulting alcohol (**14**) afforded the benzyl ether (**15**). Facial diastereoselectivity in the osmium tetroxide cis dihydroxylation of the Z olefin (Cha *et al.*, 1984) resulted in the desired protected ribitol (±**16**) and protected lyxitol diastereomers in a 7:3 ratio. After chromatographic separation of the diastereomers, protection of the cis diol, and removal of the benzyl ether, Swern oxidation of (±**18**) gave the protected ribose (±**19**). Again, the triethylamine was carefully neutralized by an aqueous workup, to prevent epimerization under basic conditions (Ko *et al.*, 1990). The aldehyde (±**19**) was unstable to silica gel purification, and hydrolysis of the crude aldehyde (±**19**) afforded crude D,L-*d*4-ribose (±**1**). The yield of ribose was quantified by enzymatic assay of the crude racemic ribose mixture. The overall yield of the 10 step synthesis was 12% D-*d*4-ribose (-**1**), (24% ±**1**), as calculated from the starting material glycerol (U-d8) (**10**). The crude D,L-*d*4-ribose (±**1**) prepared in this manner was suitable for direct use in the subsequent enzymatic reactions.

Chapter 2

Enzymatic Synthesis

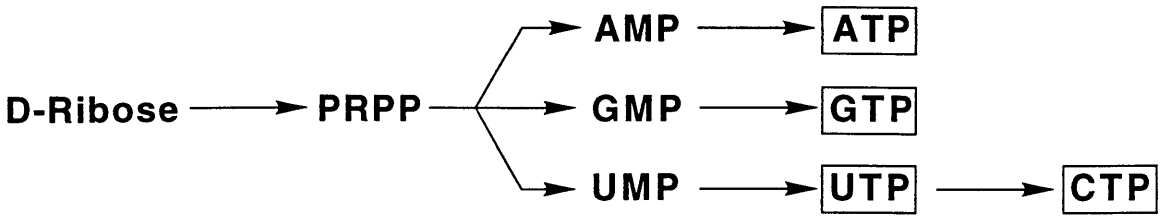


Figure 2.6 Strategy for the enzymatic synthesis.

The strategy used for the enzymatic conversion of D-ribose into the 4 NTPs of RNA synthesis is outlined in Figure 2.6. Phosphoribosyltransferases were utilized for coupling of 5-phospho-D-ribosyl α-1-pyrophosphate (PRPP) with nitrogenous bases to form nucleoside monophosphates. Reactions catalyzed by phosphoribosyltransferases can be easily linked to enzymatic PRPP synthesis and NTP formation, allowing conversion of ribose into a nucleoside triphosphate in a single coupled enzymatic reaction as has been shown previously by Schramm (Parkin *et al.*, 1984). Preparative scale enzymatic reactions for PRPP synthesis from ribose (Gross *et al.*, 1983), and NTP formation from NMPs (Simon *et al.*, 1989) have been described previously. Enzymatic procedures for converting ribose into AMP and UMP have been reported, offering enzymatic routes to ATP (Parkin *et al.*, 1984) and UTP (Gross *et al.*, 1983). We have developed enzymatic reactions that produce GTP from ribose and CTP from UTP, to complete the set of 4 NTPs required for RNA synthesis. By combining these methods we were able to form d_4-NTPs (**2-5**) from D-d_4-ribose (**-1**) in a minimum number of steps in high yield.

Seven of the enzymes necessary for the enzymatic conversions were commercially unavailable or only available in costly or low specific activity forms. *E. coli* overexpressing strains of ribokinase (RK), PRPP synthetase, orotate phosphoribosyltransferase (OPRT), OMP decarboxylase, and CTP synthetase were acquired by generous donation, and the enzymes adenine phosphoribosyltransferase (APRT) and xanthine-guanine phosphoribosyltransferase (XGPRT) were cloned from the

E. coli genome. Since partially purified enzymes were suitable for conducting the nucleotide forming reactions, a great deal of effort was avoided by not purifying all seven enzymes to homogeneity. The enzymes were purified using a general purification scheme which readily afforded partially purified enzymes. The enzyme purification scheme used requires at most one chromatographic step, and results in preparations that are free from undesirable contaminating activities, such as nonspecific ATPases. Each enzyme preparation from 1 L of bacterial culture yielded enough enzyme to catalyze several, ranging from 7 to 180, reactions (0.5 mmole scale) depending on the enzyme.

PRPP synthesis.

The PRPP used in NTP formation was generated *in situ* from D-ribose in a manner similar to that used by Gross *et. al.* (Gross *et al.*, 1983) D-*d*4-ribose (**-1**) was first phosphorylated by ribokinase to produce 3',4',5',5"-*d*4-ribose-5-phosphate (*d*4-R5P), which was pyrophosphorylated by PRPP synthetase to form 3',4',5',5"-*d*4-PRPP (*d*4-PRPP, **20**), as shown in Figure 2.7. In all of the enzymatic reactions, the ATP consumed was regenerated by a PEP/pyruvate kinase/myokinase system, where the phosphoenolpyruvate (PEP) used to drive the reaction was generated *in situ* from an excess of 3-phosphoglycerate (Simon *et al.*, 1989; Hirschbein *et al.*, 1982) (3PGA), also shown in Figure 2.7. Reactions containing PRPP synthetase were carried out in potassium phosphate buffer to prevent the inactivation of PRPP synthetase that occurs at low phosphate concentration (Gross *et al.*, 1983). One difficulty encountered in PRPP synthesis was the inhibition of PRPP synthetase by ADP in the presence of ribose-5-phosphate, resulting in extremely slow nucleotide formation in our reactions due to a buildup of ADP and ribose-5-phosphate (Switzer & Gibson, 1978). The buildup of ADP was prevented by reducing the amount of myokinase and increasing the amount of pyruvate kinase added to the reactions, to insure that the rate of ADP formation from AMP was much less than the rate of ATP formation from ADP and PEP.

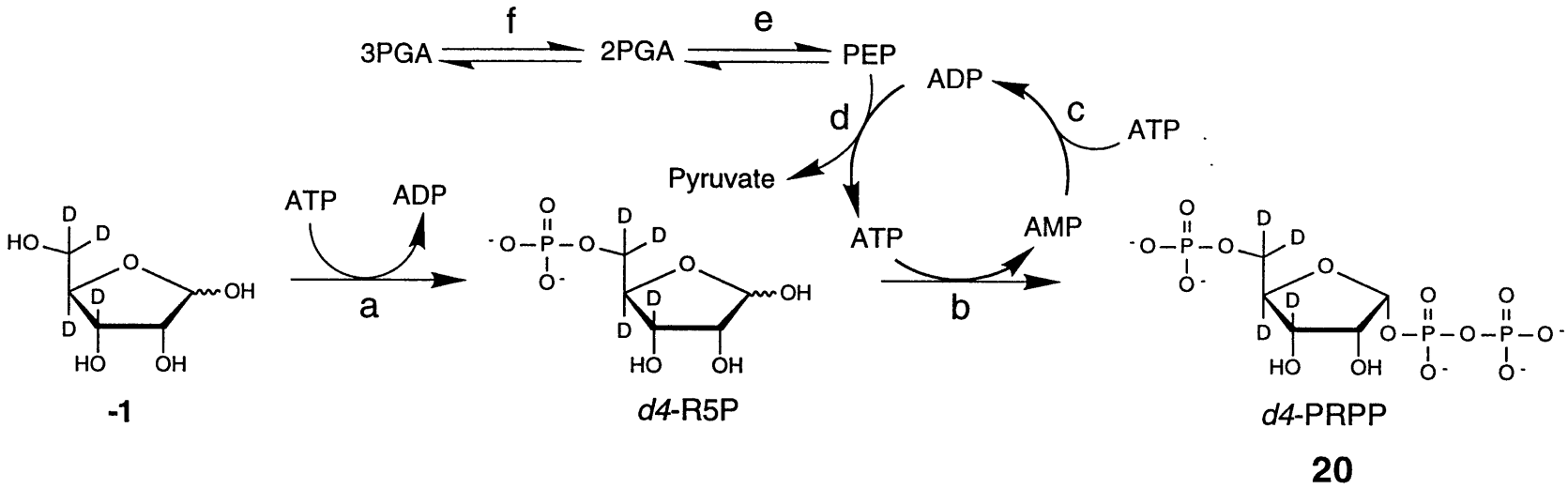


Figure 2.7 Enzymatic synthesis of labeled PRPP from labeled D-ribose. Also shown is ATP regeneration by PEP/PK/MK system driven by 3PGA. Enzymes are (a) ribokinase (b) PRPP synthetase (c) myokinase (d) pyruvate kinase (e) enolase (f) 3-phosphoglycerate mutase. PRPP synthesis was coupled directly to ATP, GTP and UTP formation shown in Figure 2.8 and 2.10.

ATP synthesis.

3',4',5',5"-*d4*-ATP (*d4*-ATP, **2**) was formed in a coupled enzymatic reaction from D-ribose in a manner similar to that described by Schramm (Parkin *et al.*, 1984). d4-PRPP (**20**) was first generated *in situ* as described above, and the purine salvage enzyme adenine phosphoribosyltransferase (APRT) was used to catalyze the condensation of *d4*-PRPP (**20**) and adenine to form *d4*-AMP, shown in Figure 2.8. *d4*-ATP (**2**) was then formed in the same reaction by the sequential action of myokinase and pyruvate kinase on *d4*-AMP. A catalytic amount of unlabeled ATP was added to the *d4*-ATP forming reaction to initiate the reaction. In all of the other reactions, *d4*-ATP was used as a cofactor to prevent dilution of the isotopic labels. The initial rate of ATP formation was very slow, but the reaction rate increases quickly as the concentration of ATP increases. *d4*-ATP (**2**) was produced from (**-1**) in 78% isolated yield by this method.

GTP synthesis.

The synthesis of 3',4',5',5"-*d4*-GTP (*d4*-GTP, **3**) was accomplished by utilizing the *E. coli* purine salvage enzyme xanthine-guanine phosphoribosyltransferase (XGPRT). First, *d4*-PRPP (**20**) was generated *in situ*, and then xanthine-guanine phosphoribosyltransferase was used to couple *d4*-PRPP (**20**) and guanine to form *d4*-GMP, as shown in Figure 2.8. Once *d4*-GMP had been formed, the sequential action of guanylate kinase and pyruvate kinase in the same reaction produced *d4*-GTP (**3**). Although guanine has a very low solubility in water, GMP formation occurs at an appreciable rate when solid guanine is added to the reaction in a slurry. Figure 2.9a shows HPLC traces during the course of a preparative GTP synthesis. Only a small guanine peak can be detected in the HPLC traces, but 2 equivalents (per ribose) of solid guanine were present in the reaction. During the course of the reaction, GMP formation can be observed at 9 and

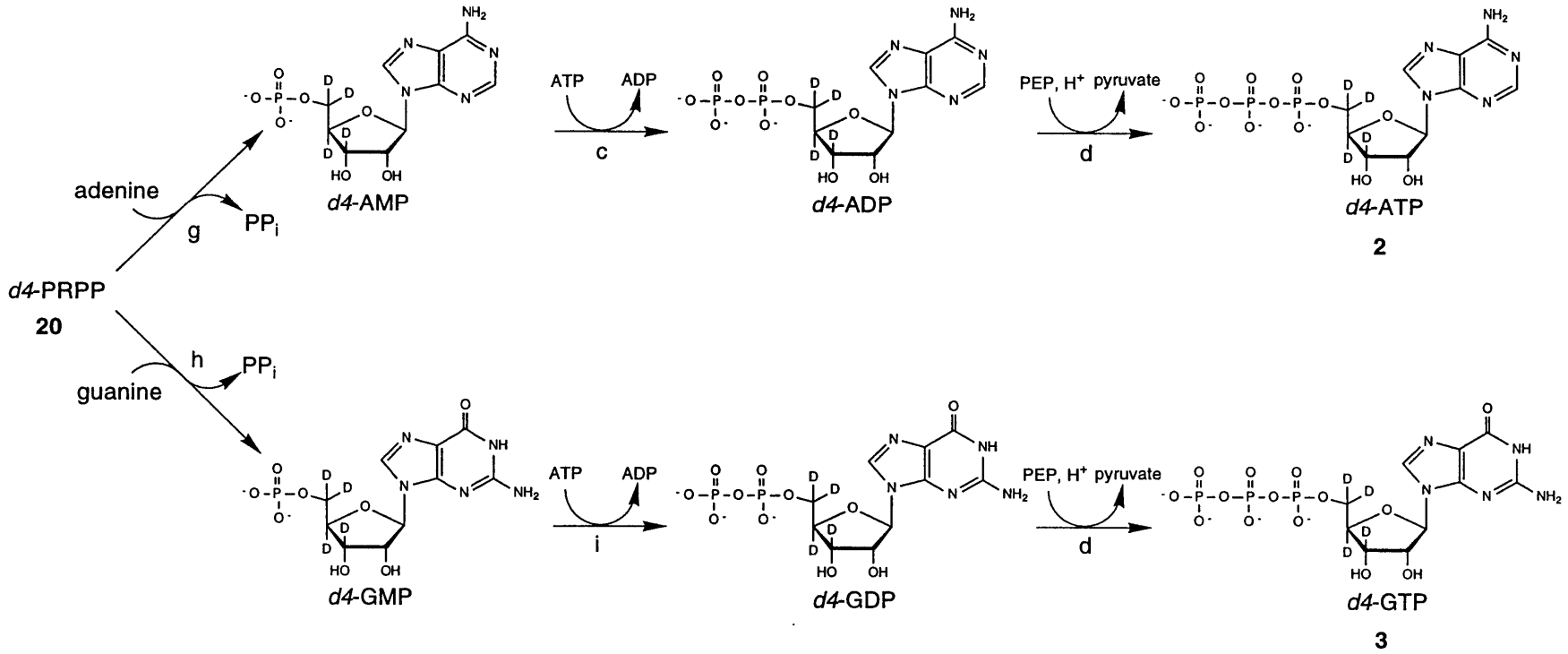


Figure 2.8 Enzymatic synthesis of purines ATP and GTP using purine salvage. Enzymes are (c) myokinase (d) pyruvate kinase (g) adenine phosphoribosyltransferase (h) xanthine-guanine phosphoribosyltransferase (i) guanylate kinase.

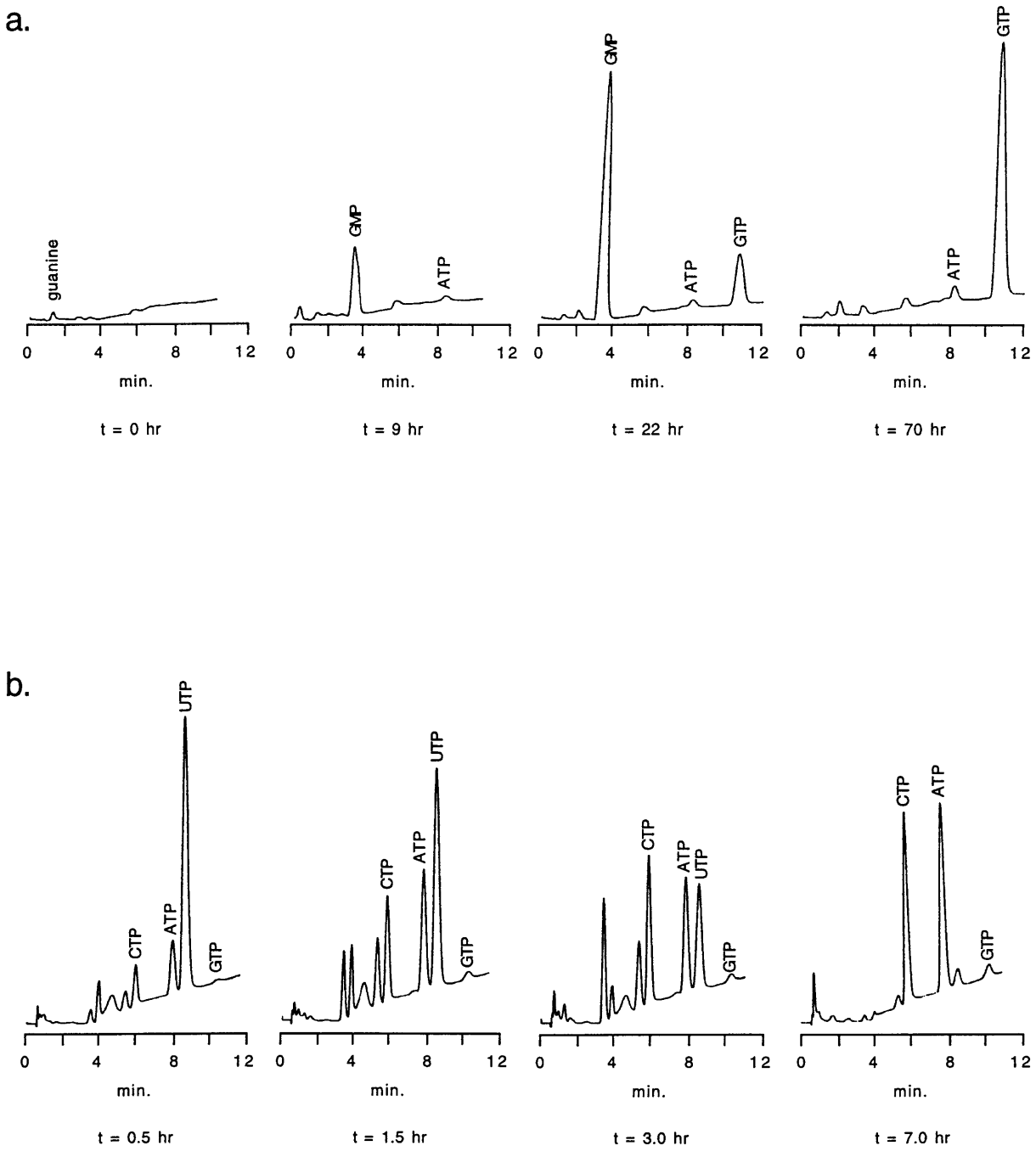


Figure 2.9 HPLC chromatograms of a) GTP forming reaction b) CTP forming reaction

22 hours, and the conversion of GMP to GTP is seen in the final HPLC trace. *d4*-GTP (**3**) was produced in 75% yield from D-*d4*-ribose (**-1**) by this method.

UTP synthesis.

3',4',5',5"-*d4*-UTP (*d4*-UTP, **4**) was synthesized from D-*d4*-ribose (**-1**) using two pyrimidine biosynthetic enzymes, orotate phosphoribosyltransferase and orotidine 5'-monophosphate decarboxylase (OMP decarboxylase), which catalyze the formation of UMP from PRPP and orotate. Gross *et. al.* (Gross *et al.*, 1983) have used these two enzymes to form UMP from PRPP and orotate, and we have extended this method to produce UTP from D-ribose and orotate. As in the *d4*-ATP (**2**) and *d4*-GTP (**3**) forming reactions, *d4*-PRPP (**20**) was first generated *in situ*, and then orotate phosphoribosyltransferase was used to catalyze the condensation of orotate and PRPP to form *d4*-OMP, which was then decarboxylated by OMP decarboxylase forming *d4*-UMP, shown in Figure 2.10. The sequential action of nucleoside monophosphate kinase and pyruvate kinase then produced *d4*-UTP (**4**) in 90% isolated yield from (**-1**).

CTP synthesis.

3',4',5',5"-*d4*-CTP (*d4*-CTP, **5**) was synthesized from *d4*-UTP (**4**) by ATP dependent amination of UTP with glutamine or ammonia using the *E. coli* pyrimidine biosynthetic enzyme CTP synthetase, as shown in Figure 2.10. CTP synthetase proved to be a difficult enzyme to work with since the activity of the enzyme is regulated by the concentration of substrates and products in solution (Long & Pardee, 1967; Anderson, 1983). The substrates ATP and UTP are activators for CTP synthetase, and the product CTP can act as an inhibitor. Using conditions similar to those used for synthesis of the other nucleotides, no CTP production was observed with CTP synthetase. After trying several different reaction conditions, it was found that by diluting UTP concentration from

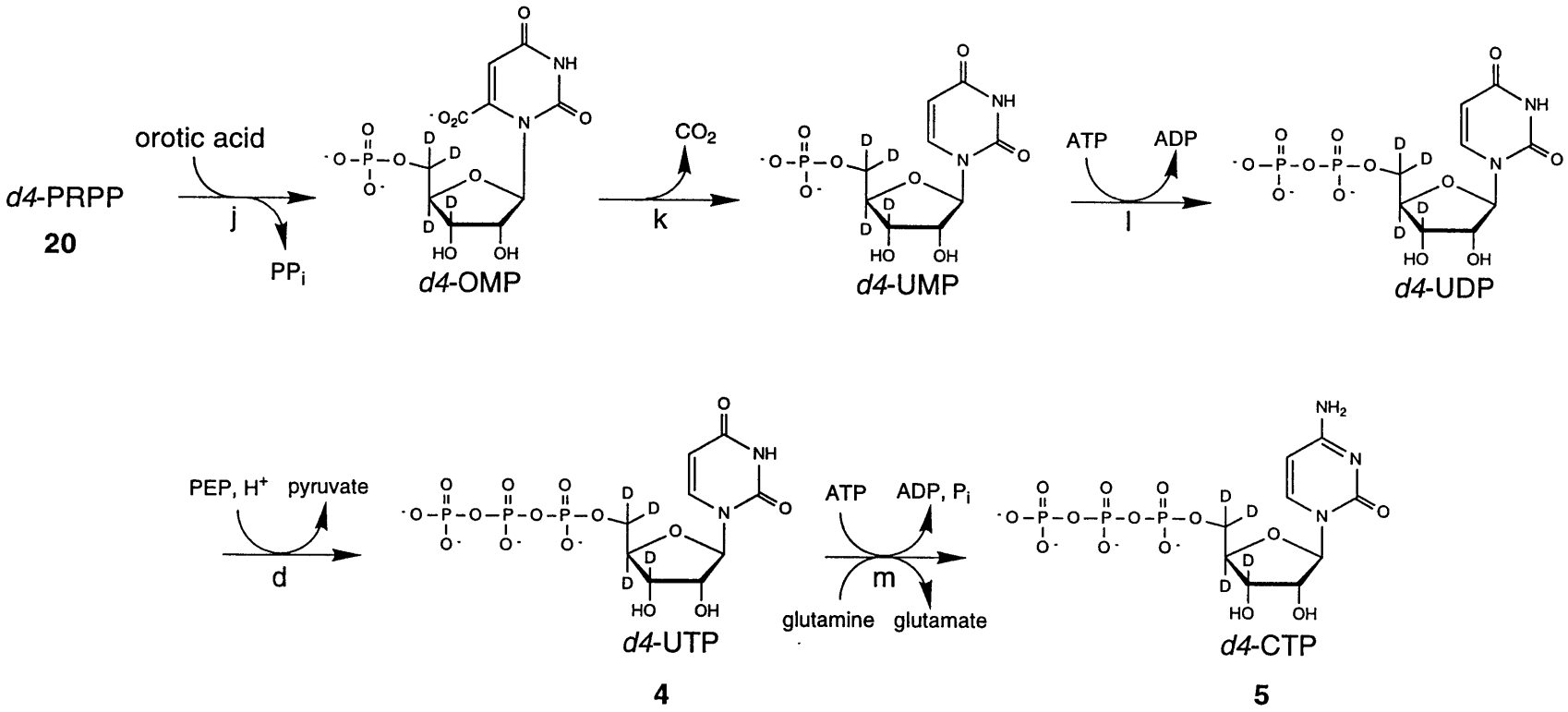


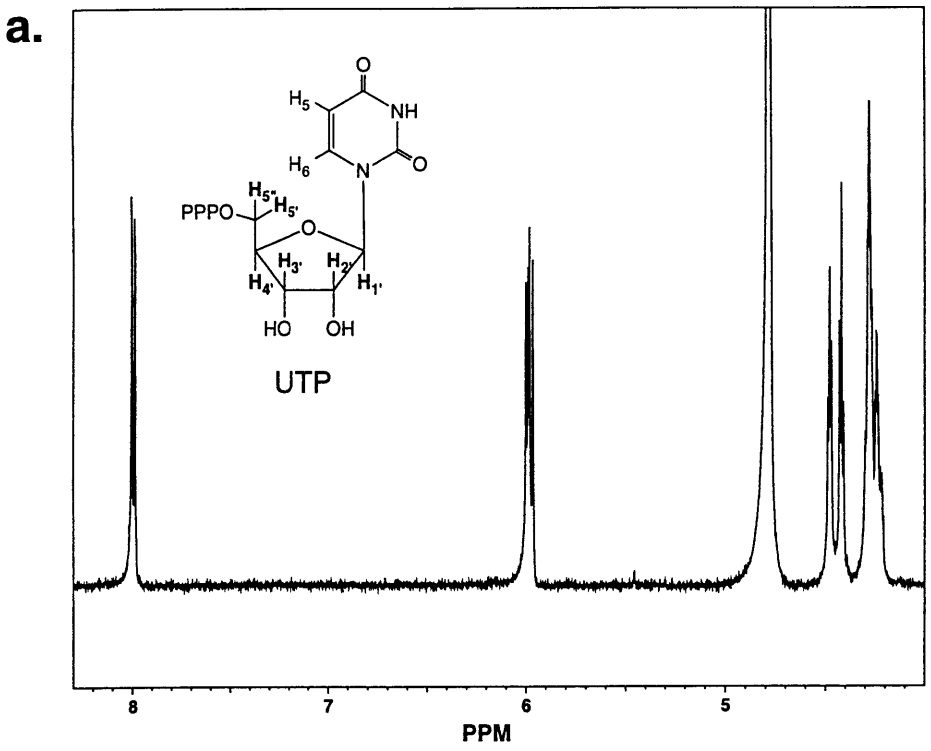
Figure 2.10 Enzymatic synthesis of UTP and CTP using pyrimidine biosynthesis. Enzymes are (d) pyruvate kinase (j) orotate phosphoribosyltransferase (k) OMP decarboxylase (l) nucleoside monophosphate kinase (m) CTP synthetase. In the preparative reactions, CTP synthesis is performed in a reaction separate from UTP formation.

approximately 5.5 mM to 1 mM and increasing the ATP concentration from 0.17 mM to 0.22 mM, it was possible to convert UTP into CTP. Under these conditions CTP concentration is low enough to prevent feedback inhibition, and the ATP concentration is high enough to keep the enzyme active at low UTP concentrations. Because of the problems encountered with CTP synthetase, CTP was produced from UTP rather than the large coupled enzymatic reactions where ATP, GTP, and UTP were produced directly from ribose. In addition, a small amount of GTP was added as a positive affector of glutamine dependent amine transfer (Long & Pardee, 1967). The conversion of UTP into CTP catalyzed by CTP synthetase was monitored by HPLC as shown in Figure 2.9b. The amount of ATP relative to the other nucleotides is much greater in the CTP forming reaction than in the GTP forming reaction, (Figure 2.9a) and additional ATP was added after 3.5 hours in order to increase the rate of the reaction. During the course of the conversion of UTP into CTP, a number of NDP species become visible, but these are eventually converted to NTPs in the final HPLC trace. After careful optimizing, this deceptively simple single enzyme reaction coupled with ATP regeneration produced *d*4-CTP (**5**) in 87% yield from *d*4-UTP (**4**).

Conclusion:

The chemical synthesis of D,L-*d4*-ribose (±**1**) described in this chapter uses perdeuterated glycerol as a starting material and produces (±**1**) in 24% overall yield. Three multi-enzyme reactions are discussed that mimic purine salvage and pyrimidine biosynthetic pathways to convert D-*d4*-ribose (-**1**) into *d4*-ATP (**2**), *d4*-GTP (**3**), and *d4*-UTP (**4**) in 78%, 75%, and 90% yields respectively. Another enzymatic reaction was used to convert *d4*-UTP (**4**) into *d4*-CTP (**5**) in 87% yield. These enzymatic reactions represent an efficient method for incorporating isotopically labeled ribose or bases into the four NTPs required for RNA transcription. The chemical and enzymatic synthesis that is described in this chapter was used to synthesize *d4*-NTPs (**2-5**) in quantities sufficient to produce at

least 2 RNA NMR samples. In Figure 2.11, the ^{1}H NMR spectra of unlabeled UTP is compared to *d4*-UTP. The aromatic and anomeric regions of both NMR spectra are similar, but the ribose region of the *d4*-UTP is much less crowded due to the removal of the H3', H4', H5' and H5" protons. Only the H2' ribose proton remains, and it is a doublet due to the removal of coupling to the H3' proton by deuteration. The conversion of the *d4*-NTPs (**2-5**) into D4-TAR RNA and the evaluation of the NMR spectroscopic properties of the D4-RNA labeling pattern is described in Chapter 4.



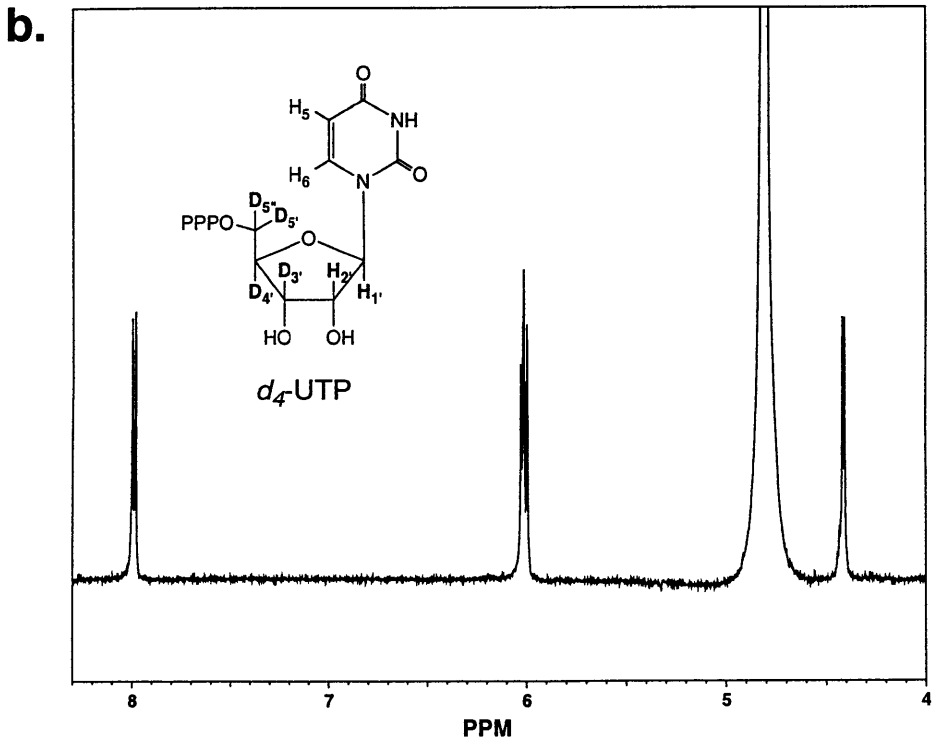


Figure 2.11 ^{1}H NMR spectra of a) UTP b) d_4-UTP

Chapter 3

3. Other Methods for Specific Isotopic Labeling of RNA

The successful synthesis of the D4-RNA labeling pattern paved the way for other specific isotopic labeling patterns of RNA to be synthesized. By purifying the enzymes and developing the enzymatic methods necessary for coupling ribose to bases to make the four NTPs of RNA, most of the work for converting ribose and bases into NTPs had already been accomplished. Isotopically labeled ribose or bases could be derived from a variety of sources and coupled enzymatically to form NTPs. In addition, any isotopically labeled form of UTP that was synthesized could be converted into CTP in a single enzymatic reaction. This would allow several isotopic labeling patterns to be synthesized which would have been difficult to produce before the enzymatic synthesis had been worked out, and there were several reasons to develop labeling patterns other than the D4-RNA labeling pattern.

While the D4-RNA labeling pattern was successful in reducing spectral crowding and altering relaxation properties of the remaining protons in the RNA (see Chapter 4) (Tolbert & Williamson, 1996) it would not solve all of the problems of studying large RNA molecules with NMR. There are certain areas of RNA NOESY spectra where the D4-RNA labeling pattern does not reduce spectral crowding. One area in particular, the H1', H5 to base NOE region, is very important in RNA sequential assignment and it was desirable to reduce spectral crowding in this area of an RNA NOESY spectra. In addition, even though the D4-RNA labeling pattern greatly reduces spectral crowding by removing all but the H2' proton in the 4-5 ppm region of RNA proton spectra, it was conceivable that RNA molecules would be studied that were so large that the sheer number of H2's in the molecule would cause severe spectral crowding in the 4-5 ppm region. For these reasons it was desirable to synthesize new specific deuterium labeling patterns of RNA which would simplify RNA NMR spectra even further than the D4-RNA labeling pattern.

It was also desirable to develop techniques for incorporating ^{13}C and ^{15}N labels into RNA. Shortly after the synthesis of the D4-RNA labeling pattern had begun, general methods for uniform incorporation of ^{13}C and/or ^{15}N labeling into RNA were published (Nikonowicz *et al.*, 1992; Batey *et al.*, 1992). These methods became widely used, and several heteronuclear multidimensional NMR techniques were developed that have had great impact on RNA structural studies (Pardi, 1995; Dieckmann & Feigon, 1994; Hall, 1995; Varani *et al.*, 1996). In the field of protein NMR spectroscopy, which has usually been one or two steps ahead of RNA NMR spectroscopy, researchers have combined ^{13}C and ^{15}N labeling of proteins with random fractional deuteration to help reduce the relaxation problems that were being encountered in the NMR spectroscopy of heteronuclear labeled proteins larger than 25 kDa. The combination of random fractional deuteration with heteronuclear labeling has made the study of proteins as large as 40-60 kDa possible with NMR spectroscopy (Grzesiek *et al.*, 1993; Markus *et al.*, 1994; Yamazaki *et al.*, 1994a; Farmer & Venters, 1995; Muhandiram *et al.*, 1995; Nietlispach *et al.*, 1996; Matsuo *et al.*, 1996; Shan *et al.*, 1996). Due to the success of this approach in proteins, methods for incorporating uniform ^{13}C and ^{15}N labeling in combination with random fractional deuteration of RNA have been developed, but the impact that these methods have on RNA NMR spectroscopy has yet to be fully assessed (Batey *et al.*, 1996; Nikonowicz *et al.*, 1997). More recently in protein NMR spectroscopy, there have been several methods reported which allow selective incorporation of protons into perdeuterated, $^{13}C/^{15}N$ labeled proteins, such as the methyl groups of aliphatic side chains (Rosen *et al.*, 1996) or the C^{α} protons (Yamazaki *et al.*, 1997). These methods allow NMR spectroscopist to benefit from gains in ^{13}C relaxation due to high levels of deuteration while still having protons within the molecule to transfer magnetization to and from. Taking the advances of RNA and protein heteronuclear NMR spectroscopy into consideration lead to the conclusion that incorporation of ^{13}C and ^{15}N labels in combination with specific deuterium labeling would be useful in reducing the problems encountered in large RNA NMR spectroscopy.

Chapter 3

Isotopic labeling of the ribose moieties of RNA with glucose

We wanted to develop a general approach to specific isotopic labeling of RNA where several isotopic labeling patterns, both ^{2}H and ^{13}C, could be incorporated into the ribose moieties of RNA with a minimum amount of effort from a starting material that was affordable and commercially available in several isotopically labeled forms. The chemical synthesis of the *d4*-ribose (**-1**) described in Chapter 2 could produce ribose with certain types of deuterium and ^{13}C labels, but the flexibility of the synthesis was limited, the amount of work necessary to produce the ribose was large, and the cost of the starting materials in ^{13}C labeled forms was too expensive to be practical. To overcome this we searched for a method for converting a commercially available starting material that is both inexpensive and available in several isotopically labeled forms into NTPs in a minimum amount of effort. The strategy we found was one where the pentose phosphate pathway is used to convert glucose into 5-phospho-D-ribosyl α-1-pyrophosphate (PRPP), and then the PRPP is converted to the four NTPs of RNA using phosphoribosyltransferases as described in Chapter 2, Figure 3.1 (Tolbert & Williamson, *in press*). This type of strategy has been used very successfully by Schramm to produce several isotopically labeled forms of AMP and NAD^+ including 1'-$^{2}H_1$-5'-$^{3}H_1$-AMP, 5'-$^{14}C_1$-AMP, and H_N1'-$^{3}H_1$-NAD^+ (Parkin *et al.*, 1984; Parkin & Schramm, 1987; Rising & Schramm, 1994), and it has also been used to incorporate ^{13}C labeled glucose into UTP (Gilles *et al.*, 1995).

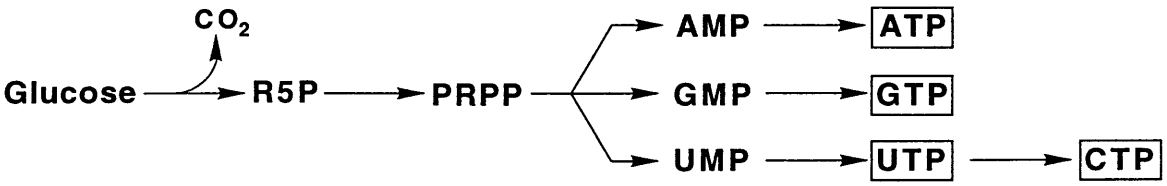


Figure 3.1 Synthetic strategy for incorporating glucose into NTPs. Glucose is converted into ribose-5-phosphate (R5P) by enzymes of the pentose phosphate pathway, and then procedures described in Chapter 2 are be used to convert R5P into ATP, GTP, UTP, and CTP.

Chapter 3

Glucose is an attractive starting material for incorporation of isotopic labels into the ribose moieties of RNA for NMR studies because it is commercially available in a variety of ^{13}C and/or deuterium labeled forms, including uniformly deuterated, uniformly ^{13}C labeled, and uniformly ^{13}C and ^{2}H labeled. In addition, it has been shown that glucose-6-phosphate isomerase, 6-phosphogluconic dehydrogenase, and ribose-5-phosphate isomerase can be used to exchange the protons on glucose destined to become the H1' and H2' protons on nucleotides (Lienhard & Rose, 1964; Parkin & Schramm, 1987; Rising & Schramm, 1994). By combining the wide range of commercially available isotopically labeled glucose with the ability of exchanging hydrogen or deuterium into the 1' or 2' position of nucleotides, many isotopic labeling patterns that will be useful for RNA NMR studies can be created. To demonstrate the flexibility and utility of these methods we synthesized several isotopic labeling patterns of nucleotides.

Enzymatic Synthesis.

The strategy for the enzymatic synthesis of isotopically labeled RNA from glucose utilizes enzymes of the pentose phosphate pathway to convert glucose into PRPP and also to exchange the H1' and H2' protons with solvent, Figure 3.2 (Parkin *et al.*, 1984; Parkin & Schramm, 1987; Rising & Schramm, 1994; Gilles *et al.*, 1995; Lienhard & Rose, 1964). Once PRPP has been formed from glucose it can be converted enzymatically into the four nucleoside triphosphates required for RNA by the methods described in Chapter 2. The enzymes of the pentose phosphate pathway required to convert glucose into PRPP are all commercially available. These enzymes are phosphoglycerate mutase, enolase, pyruvate kinase, glutamic dehydrogenase, hexokinase, glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, phosphoribose isomerase, PRPP synthetase, and glucose-6-phosphate isomerase. Of the remaining enzymes required to convert PRPP into the four NTPs, only adenine phosphoribosyltransferase (APRT), xanthine-guanine phosphoribosyltransferase (XGPRT), CTP synthetase, and uracil

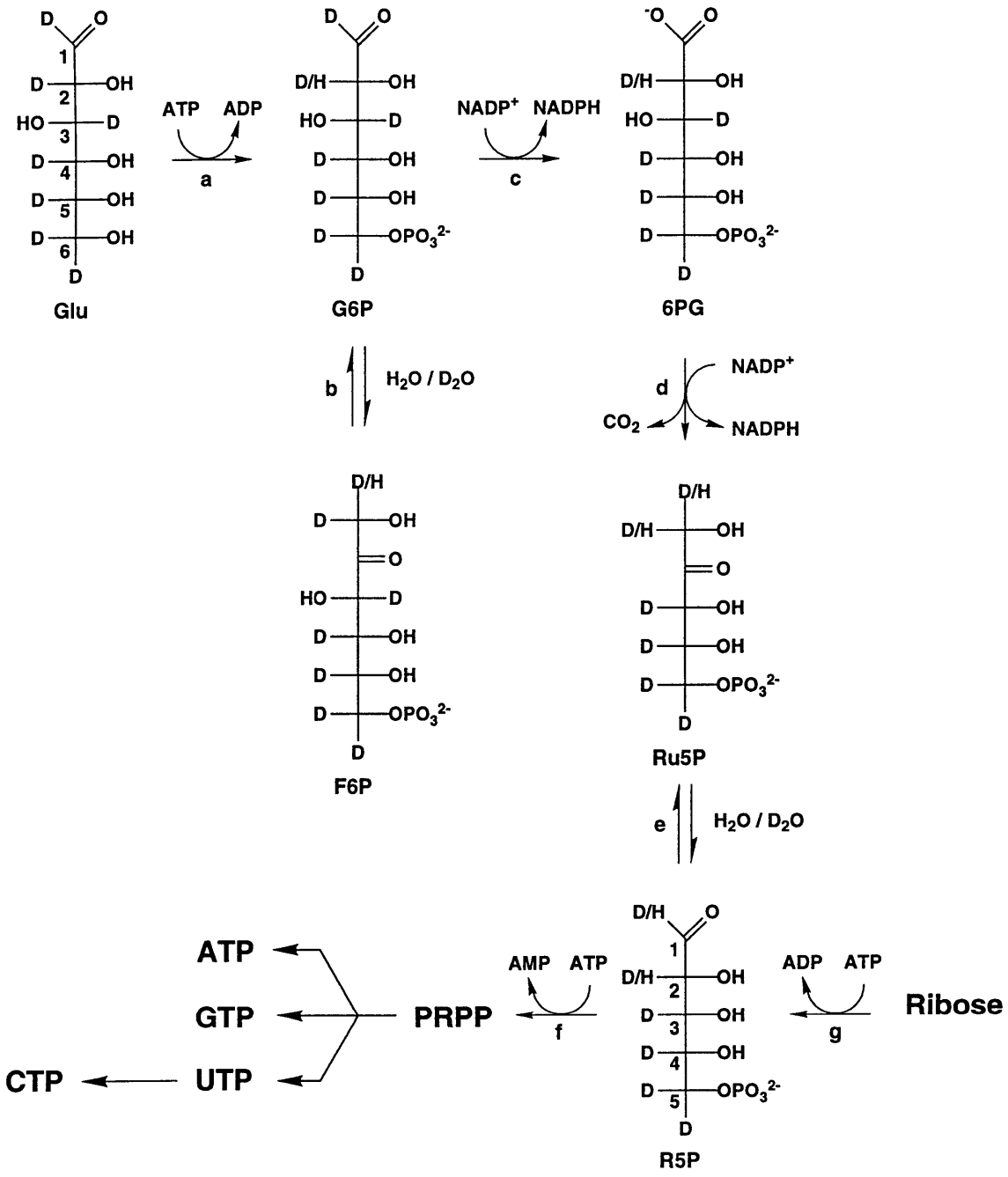


Figure 3.2 Enzymatic conversion of fully deuterated glucose into PRPP showing possible hydrogen exchange of what will become the H1' and H2' of NTPs and subsequent conversion of PRPP into the four NTPs of RNA. (a) hexokinase (b) glucose-6-phosphate isomerase (c) glucose-6-phosphate dehydrogenase (d) 6-phosphogluconate dehydrogenase (e) ribose-5-phosphate isomerase (f) PRPP synthetase (g) ribokinase

phosphoribosyltransferase (UPRT) are not commercially available in high activity forms. The purifications of the first three enzymes are described in Chapter 2 and the detailed experimental is in Chapter 6. The final enzyme, uracil phosphoribosyltransferase (Andersen *et al.*, 1992; Rasmussen *et al.*, 1986) was cloned and purified to allow uracil to be used to form the pyrimidine nucleotides rather than orotic acid as was described in Chapter 2. This was done first, because heteronuclear labeled forms of uracil are available in a wider variety and are less expensive than heteronuclear labeled forms of orotic acid, and secondly, it was desirable to conduct some of the glucose nucleotide forming reactions in D_2O without the deuteration of the H6 of the pyrimidines which occurs when orotic acid is converted into UMP in D_2O (described in greater detail in the base deuteration section of this chapter).

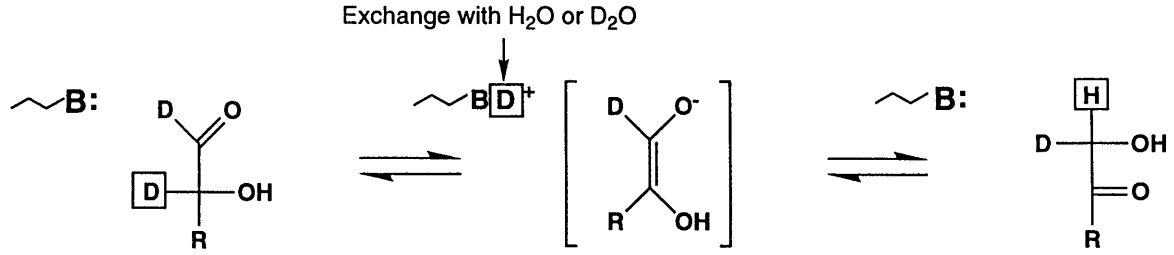
To reduce the number of steps required to produce labeled RNA from glucose, we attempted to produce all four nucleotides from glucose in one reaction, rather than synthesize each nucleotide separately as we did in Chapter 2. Since all four nucleotides are mixed together during transcription reactions, simultaneous synthesis of the nucleotides together would reduce the number of reactions and purifications necessary to produce uniformly labeled RNA. Unfortunately, CTP synthetase, which is required to convert UTP into CTP, is inhibited by some component or combination of components in the multienzyme reactions that form NTPs from glucose (Long & Pardee, 1967; Anderson, 1983), so it was not possible to produce CTP from glucose directly. Instead, it was necessary to convert glucose into ATP, GTP, and UTP in a single coupled enzymatic reaction, then half of that purified reaction mixture could be used to convert UTP into CTP in a separate reaction. In this way, all four NTPs can be produced from glucose in only two reactions.

Enzymatic conversion of glucose into nucleoside triphosphates requires 5 phosphate equivalents, supplied by ATP and PEP, and 2 oxidizing equivalents, supplied

by $NADP^+$. To prevent dilution of the isotopic label, and also to keep the cost of the reactions down, the cofactors ATP and $NADP^+$ were added in catalytic amounts and regenerated during the nucleotide forming reactions. The ATP that was consumed during the enzymatic reactions was regenerated by a PEP/pyruvate kinase/myokinase system where the PEP used to drive the reaction was generated *in situ* from an excess of 3-phosphoglycerate (Hirschbein *et al.*, 1982; Simon *et al.*, 1989). The $NADP^+$ required for oxidation of glucose 6-phosphate and 6-phosphogluconate was regenerated by reductive amination of α-ketoglutarate with NADPH and ammonia catalyzed by glutamic dehydrogenase (Rising & Schramm, 1994).

Enzyme catalyzed hydrogen exchange. The H1' proton of NTPs can be exchanged with solvent during enzymatic synthesis by using glucose-6-phosphate (G6P) isomerase, as shown in Figure 3.2 (Parkin & Schramm, 1987). G6P isomerase catalyzes the isomerization of a C1 aldose (G6P) into a C2 ketose fructose-6-phosphate (F6P) through an enediolate intermediate (Figure 3.3a). The mechanism of the isomerization involves a protein base abstracting a C2 proton and then replacing it on the C1 for the conversion of the aldose into the ketose. While the abstracted proton resides on the protein there is the possibility of exchange with solvent, and if the isomerization is carried out for a long enough time, complete exchange of the proton can be achieved. By exchanging the proton on the C2 of G6P, the proton that will become the H1' of NTPs is exchanged with solvent. This is due to the pentose phosphate pathway oxidatively removing the C1 of glucose, making what was the C2 of glucose into the C1 of ribose-5-phosphate (Figure 3.2). The H1' exchange can be accomplished by phosphorylation of glucose with hexokinase in the presence of G6P isomerase, allowing sufficient time to pass before beginning the conversion of glucose-6-phosphate into NTPs. In addition, elevating the temperature of the reaction to 34° C facilitates complete exchange with solvent. Approximately 60% H1' exchange with solvent was observed in the reaction that was used to produce nucleotides

a)



b)

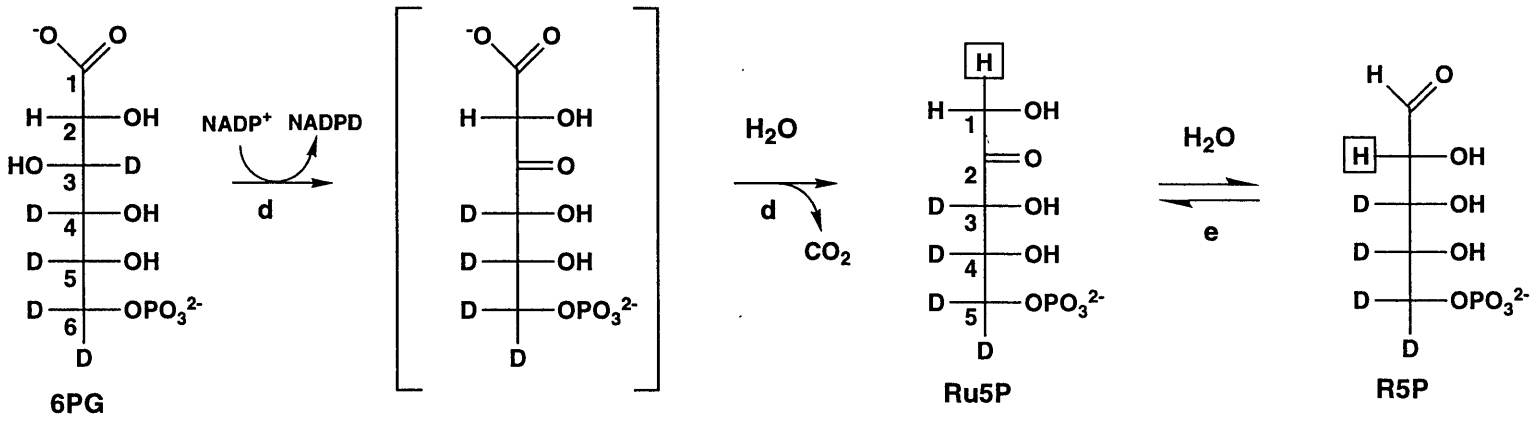


Figure 3.3 a) General mechanism for glucose 6-phosphate and ribose 5-phosphate isomerases showing exchange of sugar proton with solvent while it resides on a protein base b) Enzymatic conversion of 6-phosphogluconate into ribose-5-phosphate catalyzed by 6-phosphogluconate dehydrogenase and ribose-5-phosphate isomerase. The boxed hydrogen arises from solvent.

32-34 when the exchange reaction was carried out at room temperature for 2 days. In contrast, 100% exchange of the H1' was observed in the 3',4',5',5'-$^{2}H_4$-UTP (**4**) forming reaction when it was heated to 34° C for 20 hours while exchanging the H1' for approximately 2 days.

Exchange of the proton destined become the H2' proton of NTPs with solvent is unavoidable when using the pentose phosphate pathway to convert glucose into NTPs. This is due to 6-phosphogluconate dehydrogenase removing the only proton on the C3 of 6-phosphogluconate during the oxidation of the 3 hydroxyl group by $NADP^+$ (Figure 3.3b). During the decarboxylation of the resulting β-keto-carboxylic acid, a solvent proton is stereospecifically placed on the C1 of Ru5P (Lienhard & Rose, 1964). This former solvent proton is stereospecifically removed from the C1 of Ru5P by ribose-5-phosphate isomerase and replaced onto the C2 to produce R5P (reverse of the general reaction in Figure 3.3a). This results in the H2 of R5P being equilibrated with solvent as a side effect of the enzymatic conversion of 6-phosphogluconate to R5P. Because of this, the solvent that the oxidation and isomerization are conducted in must be the same hydrogen or deuterium composition that the H2' is to be labeled with. When ^{1}H labeling of the H2' is desired the reactions simply need be conducted in H_2O, but when deuterium labeling is desired, the chemicals and enzymes must be exchanged with D_2O and the reactions must be conducted in D_2O to achieve a high level of deuteration.

Different isotopic labeling patterns from glucose.

The use of glucose as a starting material for isotopic labeling of RNA allows several different isotopic labeling patterns to be created from glucose with only slight variations in the same general reaction conditions. By simply changing the type of isotopically labeled glucose that is used as a starting material, whether glucose-6-phosphate isomerase is added to the reaction, and whether the reaction is conducted in D_2O or H_2O, many different

labeling patterns can be derived, Figure 3.4. We have used this general strategy to produce three new isotopic labeling patterns, as well as a more efficient synthesis of the 3',4',5',5"-*d4*-labeling pattern which was produced from chemical synthesis in Chapter 2.

The synthesis of 1',2',3',4',5',5'-$^{2}H_6$[-ATP (**24**), -GTP (**25**), and -UTP(**26**)] (*d6*-NTPs) was achieved in a single reaction beginning with 1,2,3,4,5,6,6-$^{2}H_7$-D-glucose (**21**). No glucose-6-phosphate isomerase was added to this reaction since no exchange of the H1' of the NTPs was desired. The reaction was conducted in D_2O, and the enzymes and chemicals used in the reaction were exchanged into D_2O to insure a high level of deuteration at the H2' of NTPs. In this reaction 84% of the starting $^{2}H_7$-glucose was converted into *d6*-ATP (**24**), *d6*-GTP (**25**), and *d6*-UTP (**26**). *d6*-CTP (**27**) was prepared from *d6*-UTP (**26**) in a separate reaction catalyzed by CTP synthetase.

The synthesis of 3',4',5',5'-$^{2}H_4$-UTP (**4**) (*d4*-UTP) from 1,2,3,4,5,6,6-$^{2}H_7$-D-glucose (**21**) was carried out in H_2O so that the protons that would become the H1' and H2' of ribose were exchanged with solvent by the action of glucose-6-phosphate isomerase, 6-phosphogluconic dehydrogenase, and ribose-5-phosphate isomerase. First glucose was phosphorylated by hexokinase, and once glucose-6-phosphate had been formed, the reaction was heated to 34° C in the presence of glucose-6-phosphate isomerase to speed the exchange of the H2 of glucose-6-phosphate. Then the reaction was cooled and the remaining enzymes and chemicals required for UTP formation were added. This reaction converted 67% of the starting $^{2}H_7$-glucose into *d4*-UTP (**4**).

The synthesis of 1',2',3',4',5'-$^{13}C_5$[-ATP (**28**), -GTP (**29**), and -UTP (**30**)] (*$^{13}C_5$-ribose*-NTPs) was effected in a single reaction from 1,2,3,4,5,6-$^{13}C_6$-D-glucose (**22**). No glucose-6-phosphate isomerase was added to this reaction since no exchange of the H1' of the NTPs was desired. The reaction was conducted in H_2O since the desired labeling of the H2' of the NTPs was hydrogen. This reaction converted 88% of the starting $^{13}C_6$-glucose into *$^{13}C_5$-ribose*[-ATP (**28**), -GTP (**29**), and -UTP (**30**)]. *$^{13}C_5$-*

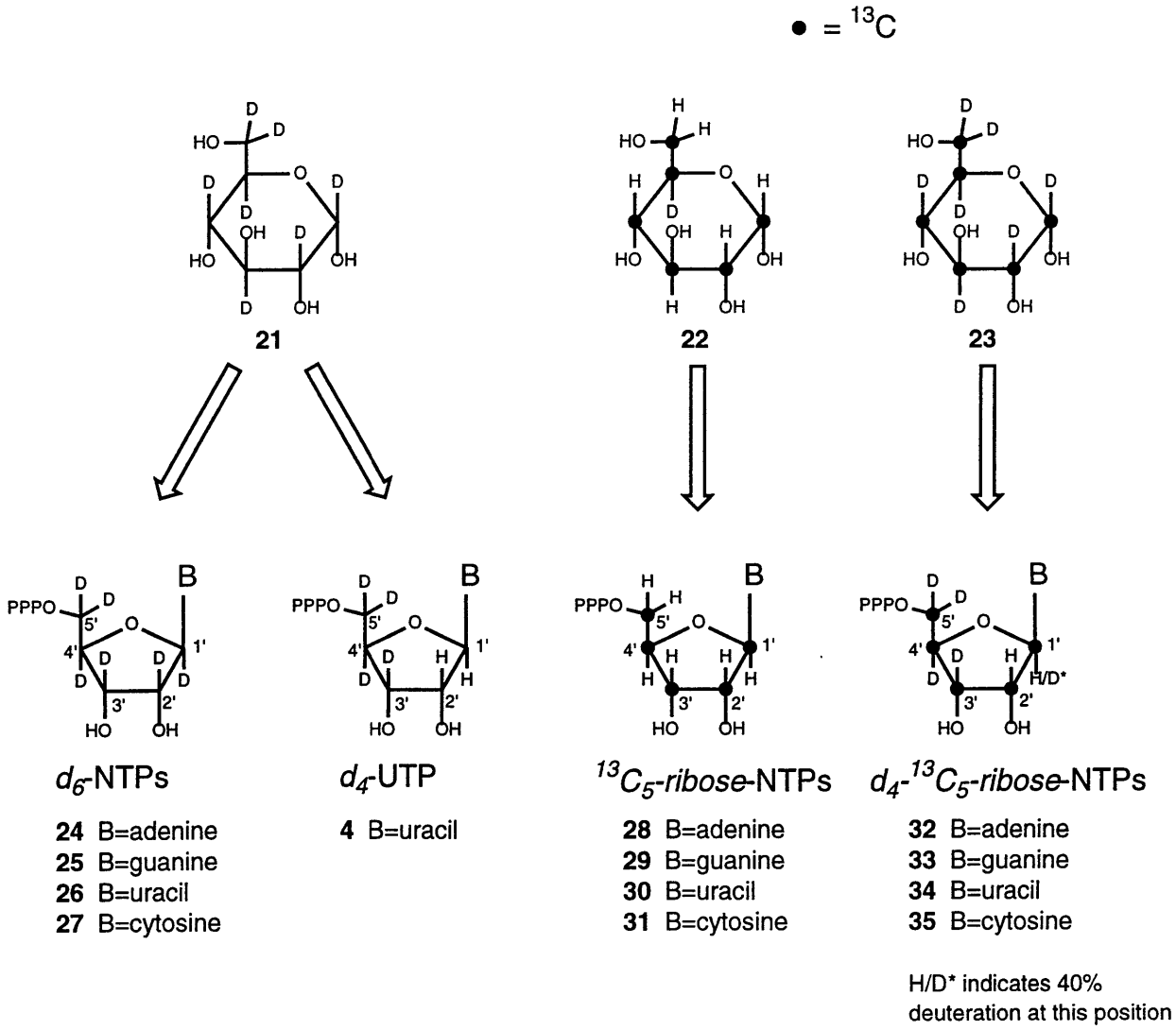


Figure 3.4 Different isotopically labeled NTPs synthesized from 1,2,3,4,5,6,6-d_7-glucose (**21**), 1,2,3,4,5,6-$^{13}C_6$-glucose (**22**), and 1,2,3,4,5,6,6-d_7-1,2,3,4,5,6-$^{13}C_6$-glucose (**23**) in this thesis.

ribose-CTP (**31**) was prepared from *$^{13}C_5$-ribose*-UTP (**30**) in a separate reaction catalyzed by CTP synthetase.

The synthesis of 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$[-ATP (**32**), -GTP (**33**), and -UTP (**34**)]* (*d4-$^{13}C_5$-ribose*-NTPs) was achieved from 1,2,3,4,5,6,6-$^{2}H_7$-1,2,3,4,5,6-$^{13}C_6$-D-glucose (**23**) in H_2O so that the protons that become the H1' and H2' of ribose are exchanged with solvent by the action of glucose-6-phosphate isomerase, 6-phosphogluconic dehydrogenase, and ribose-5-phosphate isomerase. This reaction was conducted in a manner similar to the reaction that produced *d4*-UTP (**4**), except that it was not heated above room temperature during the glucose-6-phosphate isomerase step, resulting in only approximately 60% exchange of the H1' position with H_2O. The glucose-6-phosphate exchange was carried out at room temperature for 2 days, then the remaining enzymes and chemicals required for ATP, GTP, and UTP formation were added to the reaction. In this reaction, 92% of the starting $^{2}H_7$-$^{13}C_6$-glucose was converted into *d4-$^{13}C_5$-ribose*[-ATP (**32**), -GTP (**33**), and -UTP (**34**)]. *d4-$^{13}C_5$-ribose*-CTP (**35**) was produced in a separate reaction from *d4-$^{13}C_5$-ribose*-UTP (**34**) with CTP synthetase.

* [The 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-NTPs prepared in this thesis are present as a 3:2 mixture of 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-NTPs and 1',3',4',5',5'-$^{2}H_5$-1',2',3',4',5'-$^{13}C_5$-NTPs due to partial deuteration of the H1'. The entire mixture is characterized in the experimental, but since the 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-NTPs give rise to the NMR spectra of interest, this mixture is referred to as *d_4-$^{13}C_5$*-NTPs.]

The strategy described here for labeling RNA with glucose is very efficient and flexible. This method requires only two reactions to convert glucose into all four NTPs required for RNA synthesis, with yields ranging from 67-92% conversion. The amount of labeled glucose required to make nucleotides is low when compared to other methods of isotopically labeling nucleotides with glucose. For instance, it has been reported that

harvesting nucleoside monophosphates from *E. coli* grown on isotopically labeled glucose medium results in 58 milligrams of NMPs per gram of glucose added (Batey *et al.*, 1995). In contrast, using these enzymatic methods over a gram of isotopically labeled NMPs can be prepared from one gram of glucose. Granted the isotopically labeled NMPs prepared by the enzymatic synthesis will only be labeled in the ribose moieties, but this difference in yield emphasizes the gains that can be obtained when isotopically labeled starting material is directly converted into molecules of interest rather than relying on cellular biosynthesis to make them as a byproduct of metabolism. By combining enzymatic conversion of labeled glucose into NTPs with enzyme catalyzed hydrogen or deuterium exchange, many RNA isotopic labeling patterns that will be useful in NMR structural studies can be produced with a minimum amount of effort. Several specifically deuterated forms of RNA can be produced by exchanging either hydrogen or deuterium into the H1' or H2' positions, and the same specifically deuterated forms of RNA can be made with ^{13}C labeling by simply starting with the ^{13}C labeled forms of glucose.

Incorporation of isotopic labels into the bases of RNA

Thus far the discussion of specific isotopic labeling of RNA has focused on the ribose moieties of nucleotides. Isotopic labeling need not be restricted to ribose moieties, however, and here the isotopic labeling of the base moieties of nucleotides is discussed. The enzymatic synthesis developed for incorporating isotopically labeled ribose moieties into RNA in Chapter 2 and earlier in this chapter can also be utilized to incorporate isotopically labeled bases into RNA. This can be done by simply adding the isotopically labeled bases to enzymatic nucleotide forming reactions since no protection of the bases is necessary unlike chemical methods of base coupling. In addition, the use of enzymatic synthesis to phosphorylate NMPs to form NTPs and also to convert UTP into CTP has advantages in labeling base moieties. Nucleoside monophosphates can be subjected to harsh chemical conditions which would dephosphorylate NTPs, and then the NMPs can

be enzymatically phosphorylated in a single reaction to NTPs. In this way the analysis of labeling reactions is not complicated by mixtures of different phosphorylated forms of the nucleotides. Also, the more chemically stable uracil can be labeled under conditions which would hydrolyze cytidine, and then that uracil can be converted enzymatically into cytidine. In this way both UTP and CTP can be labeled from a common precursor such as UMP or OMP.

Base Deuteration

NOEs from ribose to the non-exchangeable base protons of RNA are the basis of sequential NOESY "walk" assignments of RNA NMR spectra (see Chapter 1). Because an NOE requires two protons, one from which magnetization arises and one to which magnetization is transferred and detected on, NOESY spectra can be simplified by specifically removing one of the two protons within an interaction. Synthetic methods have been discussed in Chapter 2 and earlier in this chapter which deal with simplifying NMR spectra by deuterating ribose protons. Here synthetic methods are discussed for deuterating the other half of the ribose to base NOEs that are important for sequential assignment, the base protons.

H5 deuteration

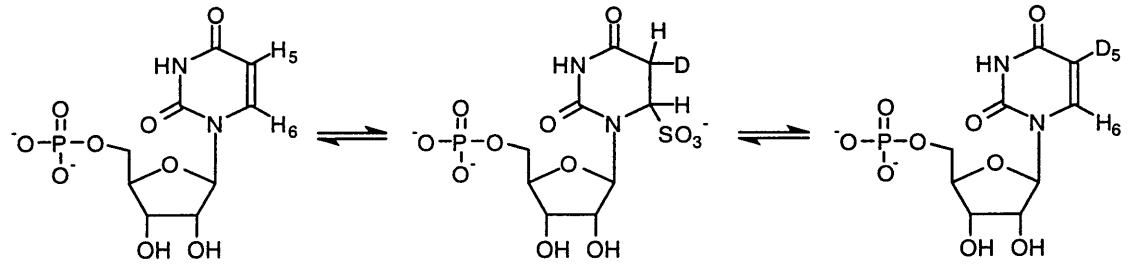


Figure 3.5 Treatment of UMP with Ammonium Sulfite in D_2O

Chapter 3

Treatment of UMP with ammonium bisulfite in D_2O at pH 8.5 and 55° C will result in the exchange of the H5 of uracil with D_2O by the mechanism shown in Figure 3.5 going through a sulfite adduct to the uracil (Wataya & Hayatsu, 1972). Adjusting the pH to 10 and allowing the reaction to cool to room temperature will cause conversion of all the sulfite adduct back to uracil. This reaction has been used to deuterate the H5 of uracil for simplification of NMR spectra of nucleic acids (Brush *et al.*, 1988; Puglisi *et al.*, 1990). A similar reaction, hydroxyl catalyzed H5 exchange (Rabi & Fox, 1973; Glemarec *et al.*, 1996), has also been used to exchange the H5 of uracil with D_2O. Cytidine undergoes H5 exchange in both of these reactions, but the conditions under which the hydrogen exchange occurs also causes deamination of the cytidine resulting in uridine. This makes it very difficult to achieve high levels of deuteration of the H5 of CMP without converting the CMP into UMP. To achieve high levels of deuteration of the H5 of cytidine nucleotides, UMP or uridine has first been exchanged with sulfite or hydroxyl, and then chemically converted into CMP or cytidine after the exchange reaction was completed (Glemarec *et al.*, 1996). We have taken a similar approach to deuteration of the H5 of pyrimidines except that we have used the enzyme CTP synthetase to enzymatically convert 5-d_1-UTP into 5-d_1-CTP.

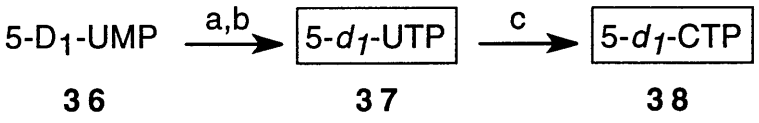


Figure 3.6 Enzymatic conversion of 5-d_1-UMP (**36**) into 5-d_1-UTP (**37**) and 5-d_1-CTP (**38**). a) nucleoside monophosphate kinase b) pyruvate kinase c) CTP synthetase

In order to produce both 5-d_1-UTP (**37**) and 5-d_1-CTP (**38**), UMP was first exchanged with sulfite, and then the sulfite was separated from the 5-d_1-UMP (**36**) with boronate chromatography at 4° C. Although boronate chromatography is conducted under

aqueous basic conditions which could possibly lead to loss of the deuterium label of 5-*d*1-UMP (**36**) before it is separated from the sulfite, no loss of the deuterium label was detected by NMR. In two separate reactions the 5-*d*1-UMP (**36**) was enzymatically phosphorylated into 5-*d*1-UTP (**37**), then half of the product from that reaction was used to produce 5-*d*1-CTP (**38**) with CTP synthetase (Figure 3.6). The nucleotides produced in this manner were used to produce isotopically labeled D5-pyrimidine RNA as described in Chapter 4.

H6 deuteration

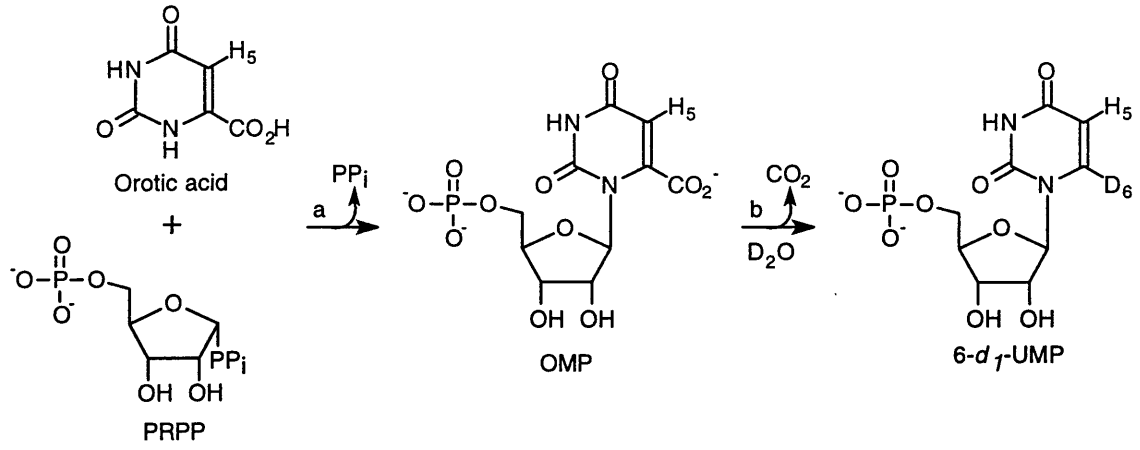


Figure 3.7 Exchange of H6 of pyrimidines. a) Orotate phosphoribosyltransferase b) OMP decarboxylase.

Exchange of the H6 of pyrimidines was accomplished enzymatically utilizing the pyrimidine biosynthesis enzymes that convert orotic acid and PRPP into UMP, Figure 3.7. Orotic acid is the biosynthetic precursor to uracil, and it is similar in structure to uracil except that it has a carboxyl group attached to the C6. In pyrimidine biosynthesis orotate phosphoribosyltransferase catalyzes the condensation of orotic acid and PRPP to form OMP, and then OMP decarboxylase catalyzes the decarboxylation of OMP to form UMP. The decarboxylation catalyzed by OMP decarboxylase occurs without any cofactors, and the proton that is placed onto the uracil C6 arises from solvent. Decarboxylation of OMP in D_2O results in deuteration of the H6 of UMP. From the H6 deuterated UMP, both 6-*d*1-

UTP and 6-*d1*-CTP can be produced enzymatically using phosphorylating enzymes and CTP synthetase. 1',2',3',4',5',5",6-*d7*-UTP (**39**) was produced from deuterated glucose and orotic acid in a reaction conducted in D_2O, and NMR spectra of an RNA transcribed with this nucleotide is discussed in Chapter 4.

H8 deuteration

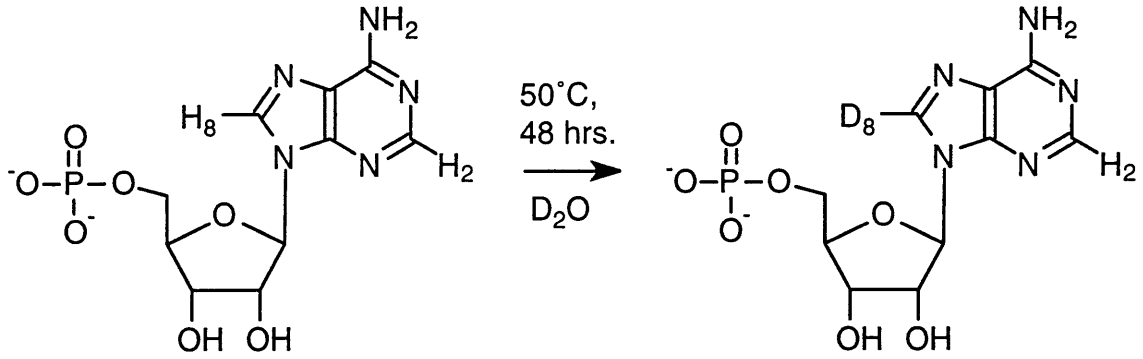


Figure 3.8 Exchange of the H8 of purines

The H8 of purines can be exchanged with D_2O by simply heating the purines to 50° C in D_2O (Benevides *et al.*, 1984; Brandes & Ehrenberg, 1986). The exchange of the H8 of AMP was accomplished by heating AMP to 50° C in D_2O for 48 hours (Figure 3.8). 8-*d1*-AMP was then enzymatically phosphorylated to produce 8-*d1*-ATP (**40**). While no RNA was produced with this nucleotide, deuteration of the H8 of purines will have similar effects on RNA NMR spectra as deuteration of the H6 of pyrimidines, except that no H5-H6 crosspeaks will be removed. GMP undergoes a similar H8 exchange with solvent when heated, so both purine H8s could be specifically deuterated. Removal of specific ribose to purine H8 NOEs could be used to simplify the assignments of very large RNAs.

Incorporation of heteronuclear labeled bases into RNA

Heteronuclear labeling of the bases of RNA can be accomplished by adding isotopically labeled bases to the enzymatic nucleotide forming reactions described in both

Chapter 2 and this chapter. The use of phosphoribosyltransferases to couple bases to PRPP offers a simple high yielding route to incorporate heteronuclear labeled adenine, guanine, orotic acid, or uracil into NTPs and then ultimately RNA. Several chemical methods for synthesizing heteronuclear labeled bases have been published, and some of the heteronuclear labeled forms are even commercially available (SantaLucia *et al.*, 1995). In addition, the bacterial CTP synthetase used to convert UTP into CTP in our enzymatic methods will accept ammonia as a substrate for the amination of UTP. This allows ^{15}N labeling of the 4 amino of CTP by adding ^{15}N-ammonium sulfate to the CTP forming reactions. Using enzymatic synthesis to couple bases to PRPP it will be possible to produce several heteronuclear labeling patterns of the base moieties of RNA in high yield and minimal effort. To demonstrate the incorporation of heteronuclear labeled bases into nucleotides, 1,2,3,4,5,6-$^{13}C_6$-1,2,3,4,5,6,6'-$^{2}H_7$-D-glucose was coupled to 6-amino,9-$^{15}N_2$-8-$^{13}C_1$-adenine, 7,9-$^{15}N_2$-8-$^{13}C_1$-guanine, and 1,3-$^{15}N_2$-2,4,5,6-$^{13}C_4$-uracil in an enzymatic reaction conducted in H_2O to produce the nucleotides **41**, **42**, and **43** (Figure 3.9).

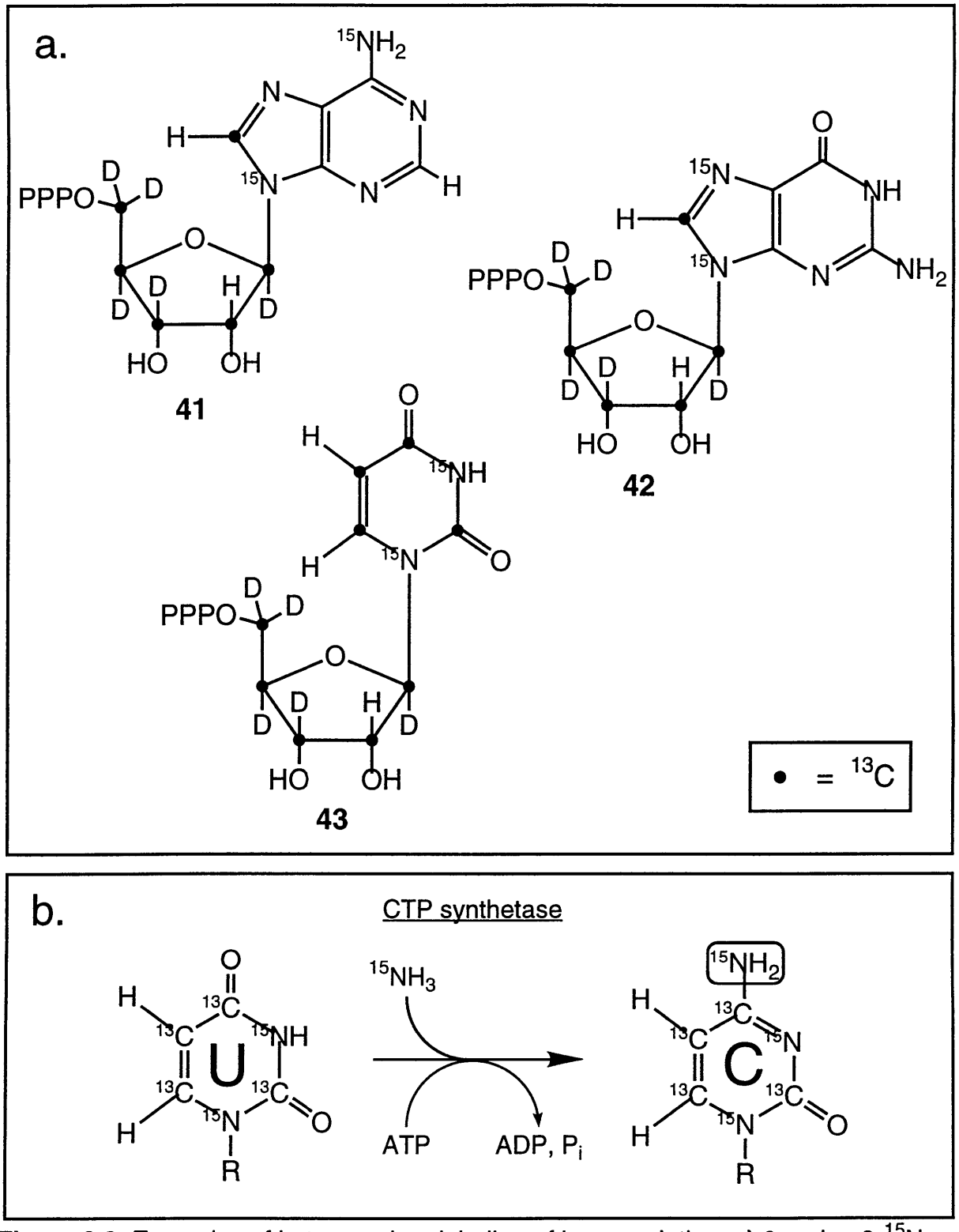


Figure 3.9 Examples of heteronuclear labeling of base moieties **a)** 6-amino,9-$^{15}N_2$-8,1',2',3',4',5'-$^{13}C_6$-1',3',4',5',5"-$^{2}H_5$-ATP (**41**), 7,9-$^{15}N_2$-8,1',2',3',4',5'-$^{13}C_6$-1',3',4',5',5"-$^{2}H_5$-GTP (**42**), and 1,3-$^{15}N_2$-2,4,5,6,1',2',3',4',5'-$^{13}C_9$-1',3',4',5',5"-$^{2}H_5$-UTP (**43**) **b)** The N4 of CTP can be ^{15}N labeled with $(^{15}NH_4)_2SO_4$ with the enzyme CTP synthetase.

Chapter 4

4. NMR of RNA with Specific Isotopic Labels

Chapters 2 and 3 describe the synthesis of nucleoside triphosphates with several different specific isotopic labeling patterns. In this chapter those specifically labeled NTPs are converted into RNA and NMR experiments are conducted on the resulting isotopically labeled RNA to demonstrate the utility of the different isotopic labeling patterns. Specific isotopic labeling can be used to increase the size of macromolecules which can be studied with NMR spectroscopy by focusing on a subset of the total resonances within a macromolecule. Proper design of specific isotopic labeling patterns will maximize the amount of structural information in NMR spectra while reducing spectral crowding, simplifying assignment, and producing narrower linewidths. Examples of these effects on the NMR spectra of RNA are demonstrated with the isotopic labeling patterns of RNA discussed in this chapter.

TAR RNA

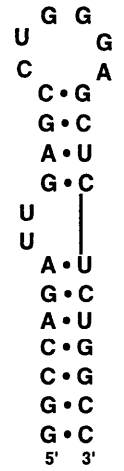


Figure 4.1 Secondary structure of the 30 nucleotide HIV-2 TAR RNA.

To test the different isotopic labeling patterns described in this chapter we used the 30 nucleotide HIV-2 TAR RNA, Figure 4.1. The TAR RNA has been well studied with

NMR spectroscopy in our laboratory, having several proton and heteronuclear NMR experiments conducted on it during the course of a high resolution structure determination of its complex with argininamide (Brodsky & Williamson, 1997). Sequential assignments of the NMR spectra of the RNA alone and in complex with argininamide have been made. The TAR RNA has well behaved NMR characteristics yielding very good NMR spectra at 25° C. In addition, in T7 RNA polymerase catalyzed transcription reactions the TAR RNA sequence yields fairly high amounts of RNA, requiring only 80-100 μmoles of each NTP to produce a 550 μl 1.0-1.7 mmolar TAR RNA NMR sample, meaning that large amounts of isotopically labeled NTPs do not have to be expended to produce TAR NMR samples.

The TAR RNA is a good RNA for evaluating how well a new isotopic labeling pattern relieves spectral crowding because the NMR spectra of the unlabeled molecule have already been assigned, so the reduction of spectral crowding or ease of assignment caused by a new labeling pattern can be immediately compared to the difficulties encountered with assigning the unlabeled RNA. Also, since the size of the TAR RNA molecule is not so large as to make heteronuclear NMR experiments difficult due to relaxation problems, it is a good RNA to develop new heteronuclear NMR experiments with. We have taken an iterative approach to the testing of RNA isotopic labeling patterns where new labeling patterns were tested in NMR experiments as TAR RNA, and then the results were used to guide the design of other isotopic labeling patterns.

Transcription of RNA. The different isotopically labeled forms of TAR RNA (5'GGC CAG AUU GAG CCU GGG AGC UCU CUG GCC3') were synthesized by *in vitro* transcription (Milligan & Uhlenbeck, 1989) with T7 RNA polymerase using unlabeled NTPs and the different isotopically labeled NTPs produced in Chapters 2 and 3 of this thesis. Transcription conditions were as described in Wyatt and Puglisi (Wyatt *et al.*, 1991), except that nucleotide concentrations were approximately 2 mM at the start of

transcription reactions and the $MgCl_2$ concentration was 35 mM. More detailed description of transcription reactions can be found in Chapter 6.

Specific Deuteration of RNA.

Specific deuteration of RNA protons can dramatically reduce the spectral crowding in RNA NMR spectra. Unlike random fractional deuteration, specific deuteration can be used to simplify NMR spectra as well as alter relaxation rates. In the different TAR RNA samples presented in this section, certain protons are removed from the NMR spectra to simplify the interpretation of the remaining spectral information. Structural information is retained in the spectra but the assignment is simplified due to the removal of several types of resonances. In addition, the removal of specific protons and replacing them with deuterium alters the dipolar interactions which give rise to relaxation of the remaining protons, lengthening the lifetime of their detectable magnetization and thus producing narrower linewidths.

NOESY spectra of the different deuterated forms of TAR RNA presented in this section are used to demonstrate the reduction in spectral crowding and simplification of assignment produced by specific deuteration. Figure 4.2 shows a 200 ms NOESY spectra of unlabeled TAR RNA. At 30 nucleotides the NOESY spectra of the unlabeled TAR RNA is fairly crowded, with assignment of the ribose regions of the spectra (regions **A**, **B**, and **C**) being very difficult because of spectral overlap. All of the NOESY spectra of TAR RNA within this section were collected with a 200 ms mixing time with the only differences being due to the different deuteration patterns. The different TAR deuteration patterns will be compared to Figure 4.2 to demonstrate the reduction of spectral crowding achieved by deuteration.

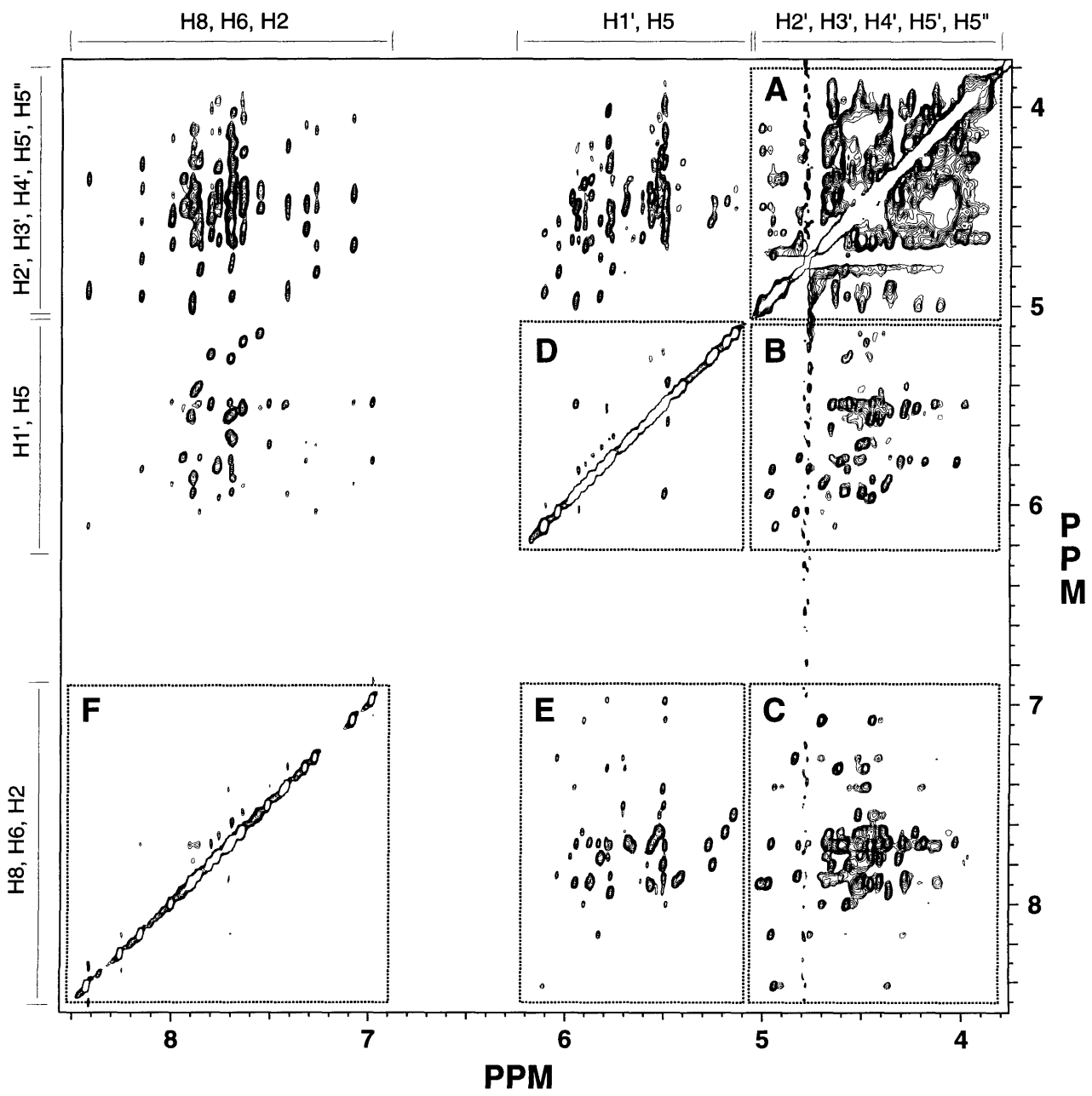


Figure 4.2 200 ms NOESY spectra of unlabeled TAR RNA

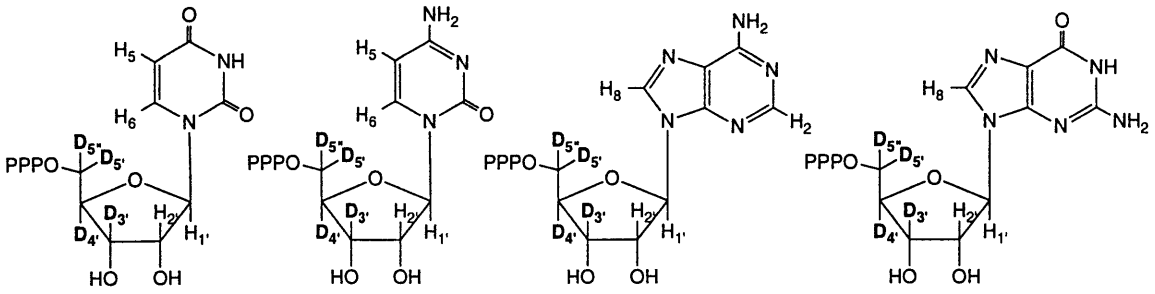
Chapter 4

3',4',5',5"-D4-TAR RNA.

3',4',5',5"-*d4*-TAR RNA (D4-TAR RNA) was produced by T7 RNA polymerase catalyzed transcription from 3',4',5',5"-*d4*-ATP (**2**), 3',4',5',5"-*d4*-GTP (**3**), 3',4',5',5"-*d4*-UTP (**4**), and 3',4',5',5"-*d4*-CTP (**5**), Figure 4.3a. The synthesis of **2-5** from perdeuterated glycerol is described in Chapter 2, and the detailed procedures for the synthesis are in Chapter 6. The D4-TAR RNA labeling pattern reduces spectral crowding in RNA NMR spectra by removing 4 of the 6 non-exchangeable ribose protons per nucleotide while still retaining much of the structural information that can be obtained from RNA NMR spectroscopy, Figure 4.3b. The remaining ribose protons, the H1' and H2', yield much of the possible structural information available from RNA NMR spectroscopy with J-coupling between the H1' and H2' yielding ribose sugar pucker, NOEs between the H1' and base protons determining glycosidic torsion, and NOEs between base protons and the H1' and H2' protons yielding sequential connectivity information.

The one dimensional (1D) ^1H NMR spectra of both unlabeled and D4-TAR RNA in D_2O are shown in Figure 4.4. A great reduction of spectral crowding is evident in the 4-5 ppm region of the D4-RNA NMR spectra where there are only 30 H2' ribose protons compared to the 150 ribose protons in the same region of the unlabeled RNA NMR spectra. In a 2-dimensional NOESY spectra, this reduction in the number of ribose protons in the 4-5 ppm region represents itself as a decrease in spectral crowding in the ribose NOE regions **A**, **B**, and **C**, Figure 4.5. The advantages of this deuteration pattern are dramatically illustrated in a comparison of the base to ribose region of a NOESY spectrum of TAR RNA acquired on unlabeled RNA and D4-RNA, shown in Figure 4.6a and 4.6b, respectively. The base to H2' NOEs are unambiguously identified in Figure 4.6b, and the spectral overlap in this region is greatly reduced. A similarly dramatic effect of the *d4*-ribonucleotides is observed in the ribose region of the NOESY spectrum, shown in Figure

a.



d_4-UTP **4** d_4-CTP **5** d_4-ATP **2** d_4-GTP **3**

b.

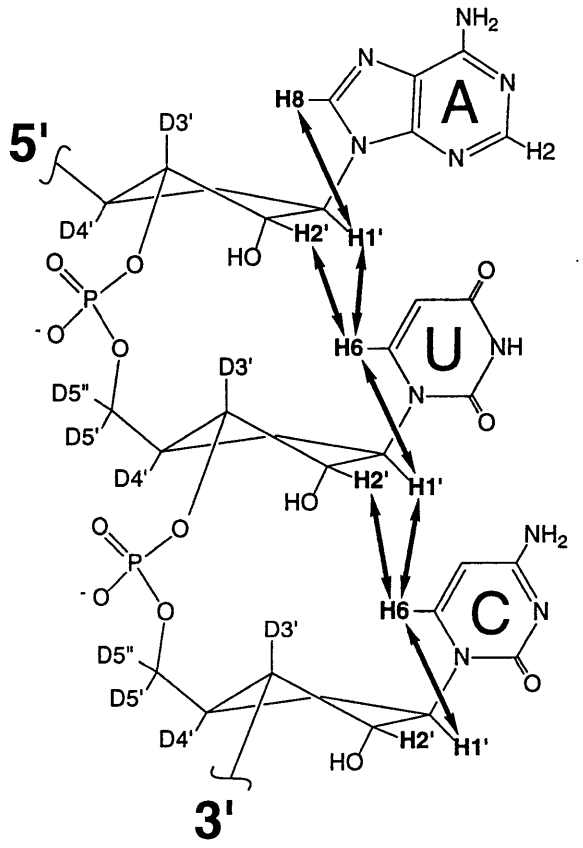


Figure 4.3 a) 3',4',5',5"-d_4-NTPs which were used to produce D4-TAR RNA b) schematic of D4-RNA in A-form helical structure. NMR spectra of D4-RNA are greatly simplified due to the deuteration of the H3', H4', H5', and H5" ribose protons, but base pairing, sugar pucker, glycosidic torsion, and sequential sequential connectivity information can still be obtained.

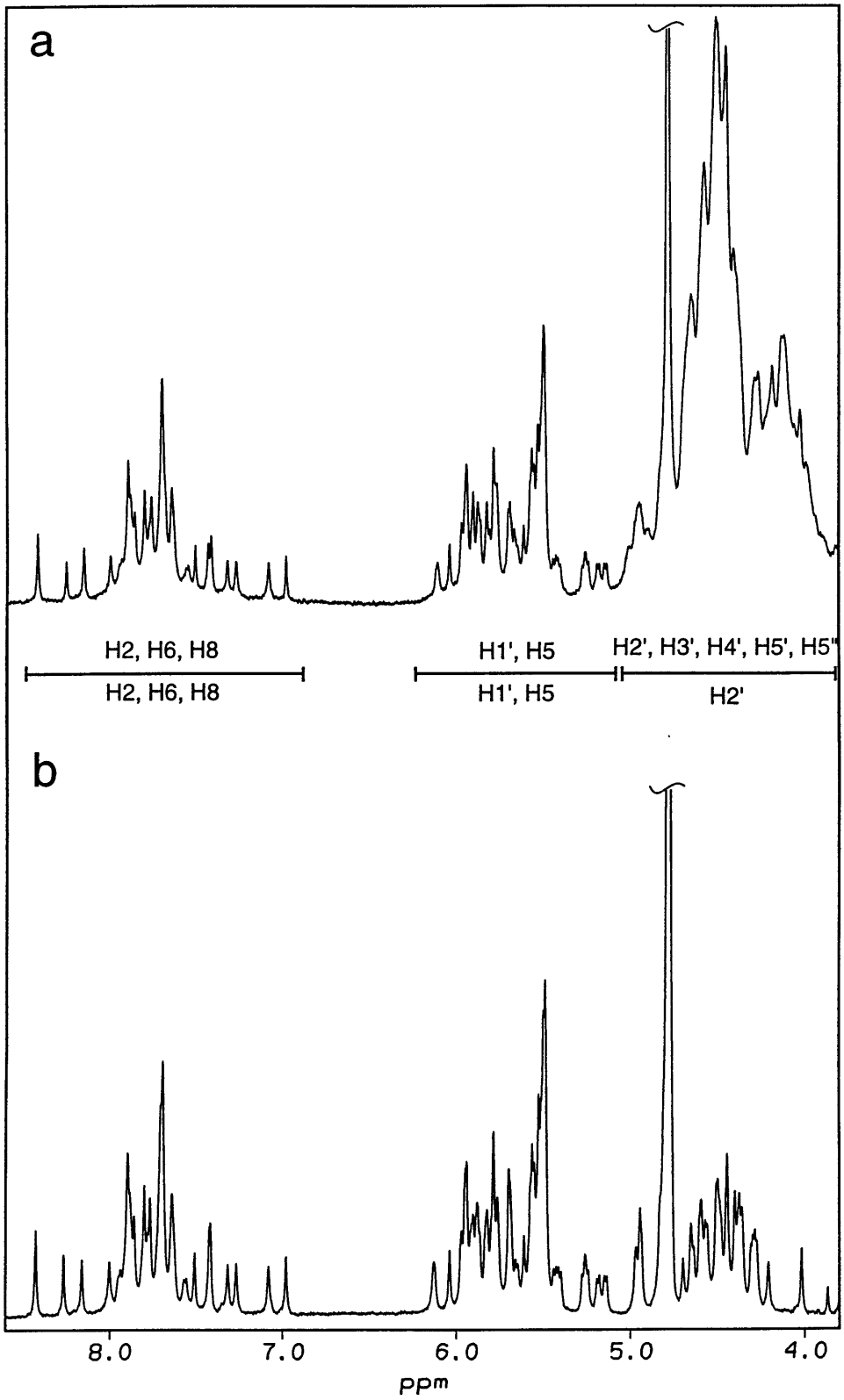


Figure 4.4 500 MHz 1H NMR spectra of HIV-2 TAR RNA
a) unlabeled b) 3',4',5',5"-D4-labeled.

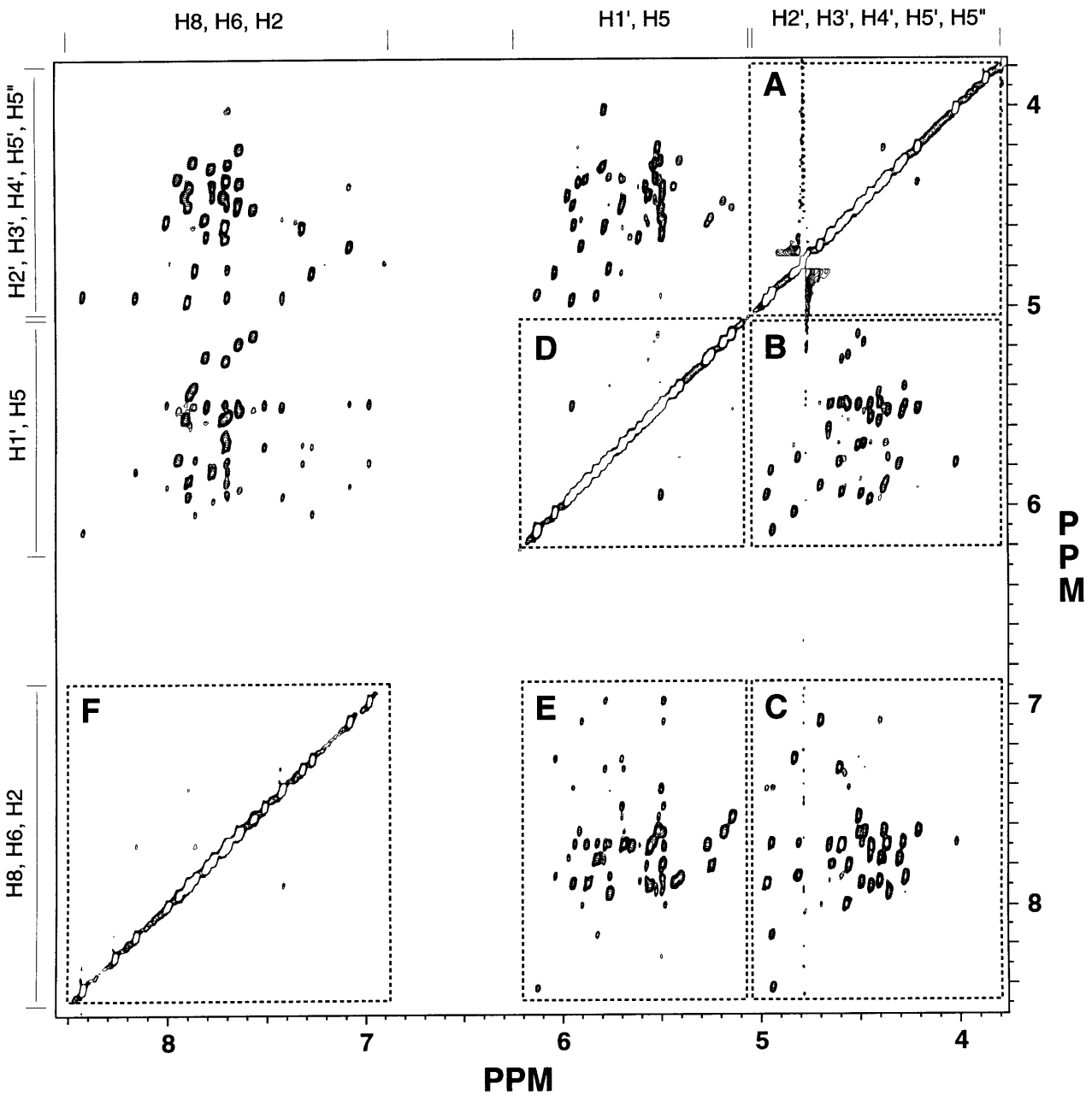


Figure 4.5 200 ms NOESY of D4-TAR RNA.

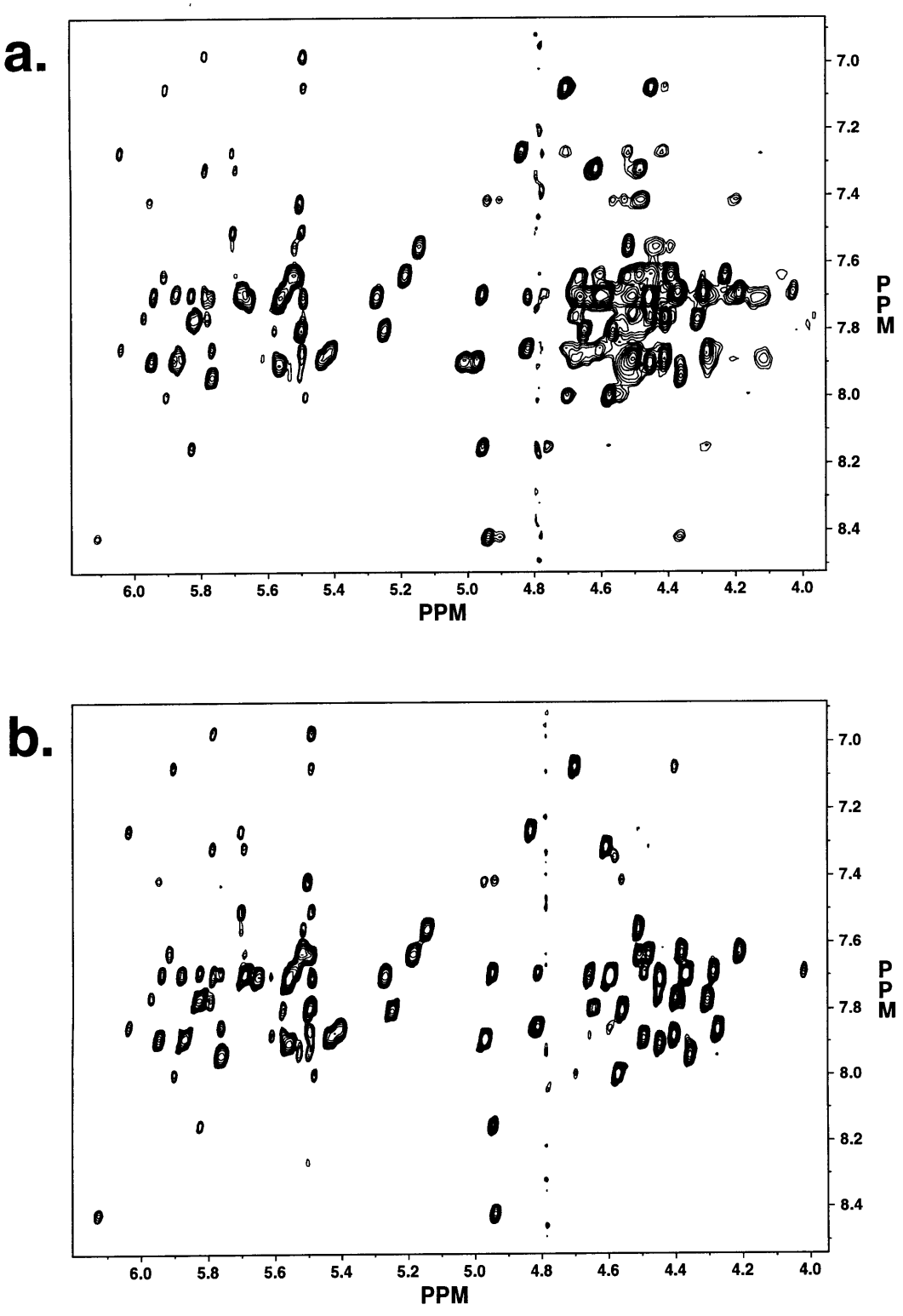


Figure 4.6 Base to ribose NOE region of a. unlabeled TAR b. D4-TAR

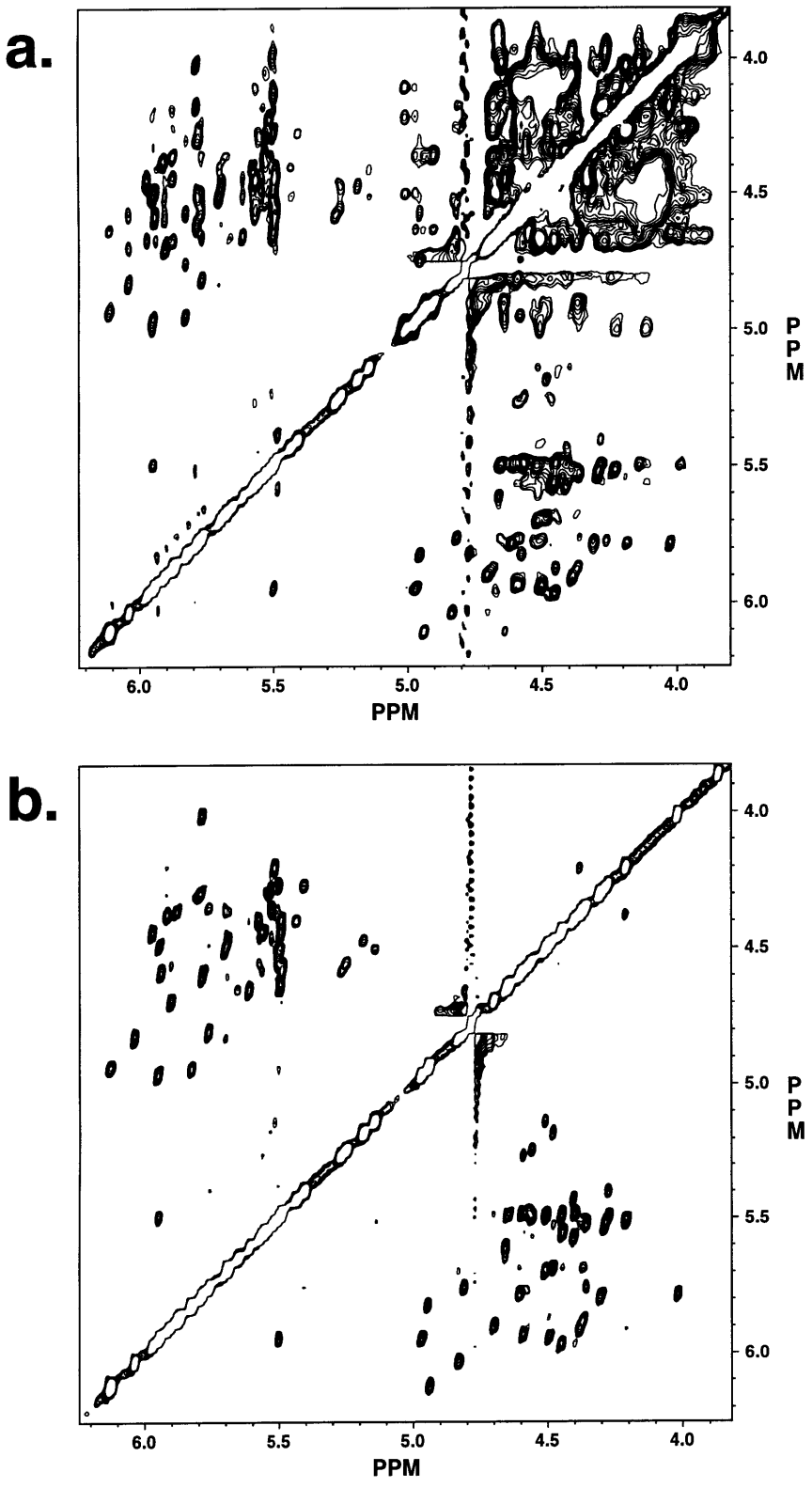


Figure 4.7 Ribose NOE region of a. Unlabeled TAR b. D4-TAR

4.7a and 4.7b. The H1' to H2' cross peaks can be unambiguously identified in Figure 4.7b. The diagonal for the region between 4-5 ppm is completely devoid of cross peaks, except for one weak peak that results from an unusual H2'-H2' NOE from the loop region of TAR. This peak was not previously identified in 3D- or 4D- ^{13}C resolved NOESY spectra using uniformly ^{13}C-labeled TAR RNA, even though all of the ribose protons were assigned (Brodsky & Williamson, 1997).

Relaxation times of D4-TAR vs. unlabeled TAR RNA

Although the cross peaks in the NOESY spectra of D4-RNA in Figures 4.6b and 4.7b appear marginally narrower than the unlabeled RNA in Figures 4.6a and 4.7a, relaxation rates were determined to provide a more quantitative measure of the effects of deuteration on dipolar relaxation. Nonselective T_1 and T_2 measurements were performed on unlabeled TAR and D4-TAR RNAs, and the relaxation rates were determined for a number of resolved peaks, summarized in Tables 1 and 2. The effective T_1s and T_2s for the D4-RNA were increased by approximately a factor of two compared to unlabeled RNA. These results are comparable to the effects of deuteration observed in proteins (Markus *et al.*, 1994) and DNA (Agback *et al.*, 1994). Although for the 30 nucleotide TAR RNA studied here, the deuteration does not qualitatively improve the NOESY spectrum by a relaxation effect, a 50% reduction of relaxation efficiency could significantly improve the spectra of RNAs that are much larger than 30 nucleotides.

Table 1		$\langle T_1 \rangle$[a]	
	TAR	D4-TAR	ratio[b]
H2'/(H3',H4',H5',H5")[e]	4.3±0.7	9.3±1.0	2.2±0.2
H1'/H5	5.6±0.7	10.3±1.3	1.8±0.1
H8/H2/H6	5.4±1.2	10.2±1.6	1.9±0.2

Table 2		$\langle T_2 \rangle$[c]	
	TAR	D4-TAR	ratio[d]
H2'/(H3',H4',H5',H5")[e]	0.04±.01	0.12±0.02	2.6±0.3
H1'/H5	0.10±0.02	0.17±0.03	1.8±0.3
H8/H2/H6	0.09±0.03	0.17±0.06	2.0±0.3

[a] Average T_1 relaxation time (s)
[b] Ratio of D4-TAR and TAR T_1 relaxation times
[c] Average T_2 relaxation time (s)
[d] Ratio of D4-TAR and TAR T_2 relaxation times
[e] The global average of relaxation times were calculated for these protons in the unlabeled sample due to spectral overlap in this region

Assignment of D4-RNA.

The assignment of the D4-TAR RNA NMR spectra is simplified compared to the assignment of unlabeled RNA NMR spectra because of the reduction in spectral crowding of the ribose regions of the NOESY spectra (regions **A**, **B**, and **C** of Figure 4.5). This allows easier identification of the H2' interactions in those regions because the only types of NOEs within those regions for the D4-RNA are H2' NOEs. The H2' ribose proton is useful in sequential assignment since it usually has a very strong NOE to the base proton that is 3' to it in A-form RNA (see list of distances between protons in A-form RNA Chapter 1), Figure 4.3b. Unfortunately the identification of H2' to base NOEs in unlabeled RNAs are complicated by the fact that within the spectral region that the H2'-base NOEs occur there is a great deal of spectral crowding because of the overlap of the H2',

H3', H4', H5' and H5" chemical shifts, region **C** in Figure 4.2. The D4-RNA labeling pattern simplifies the assignment of H2' to base NOEs because the H2' to base NOEs are the only type of NOE that occur within region **C** of NOESY spectra of D4-RNA, shown in Figure 4.5.

Sequential assignment information obtained from H2' to base NOEs can be linked to H1' to base NOE information by the intra-nucleotide H1'-H2' NOEs, thus building a more complete picture of the identity of the different proton resonances. Figure 4.8 demonstrates the linking of H1'-base NOEs to H2'-base NOEs through H1'-H2' crosspeaks for 4 out of the 6 uracil nucleotides within the TAR RNA. Figure 4.8a shows the connectivity for unlabeled TAR RNA, and Figure 4.8b demonstrates the same connectivity for the D4-TAR RNA. Notice that the spectral crowding in the H2' to base NOE regions and the H1' to H2' NOE regions are greatly reduced in the D4-TAR RNA, making the identification of the individual peaks and assignment a much easier task.

The NMR experiments presented here demonstrate the reduction of spectral crowding, increased relaxation times, and simplified assignment of NMR spectra that can be achieved with the D4-RNA isotopic labeling pattern. A great deal of structural information is retained within NMR spectra of D4-RNA even though the NMR spectra are largely simplified, with glycosidic torsion, sugar pucker and base pairing information still obtainable from D4-RNA NMR spectra. Unlike heteronuclear labeling, which spreads NMR structural information over an additional heteronuclear dimension with the disadvantage of broader linewidths, specific deuteration simplifies RNA NMR spectra by replacing certain protons with deuterium and results in narrower linewidths. This should allow specific deuteration to be effective on larger RNA molecules than heteronuclear labeling will be.

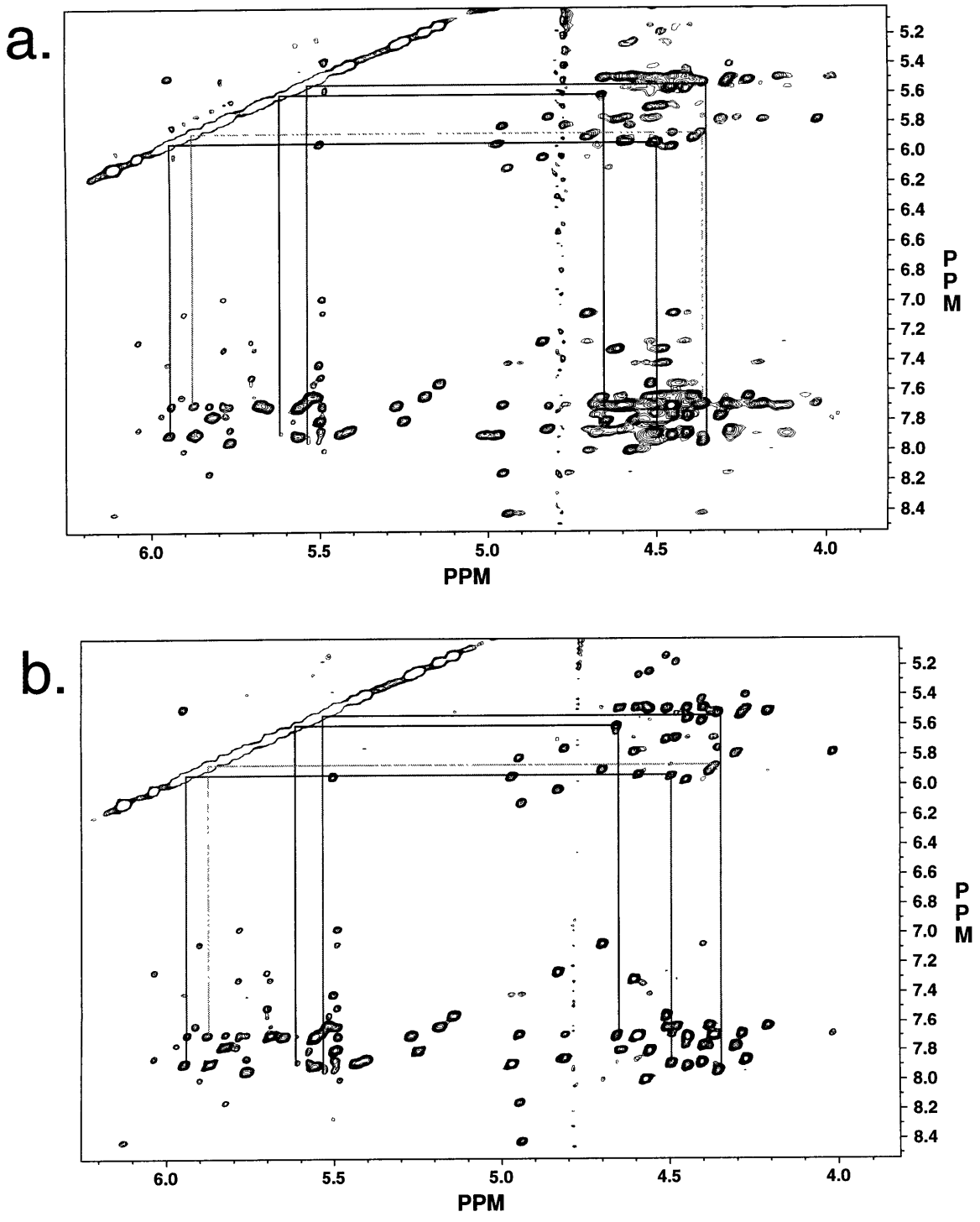


Figure 4.8 NOE connectivity of uridine nucleotides for a. unlabeled TAR RNA b. D4-TAR RNA. (From left to right: H1'-base NOEs, H1' crosspeaks on the diagonal, H1'-H2' NOEs, and H2'-base NOEs.)

Chapter 4

D6-TAR RNA.

As mentioned in Chapter 3, the D4-RNA isotopic labeling pattern would solve many of the problems of large RNA NMR spectroscopy, but it would not solve all of the problems. It could be foreseen that there might be spectral crowding problems in large RNA NMR spectroscopy that were so severe that the D4-RNA labeling pattern alone would not be sufficient to allow NMR structural studies, and it was with these concerns in mind that labeling patterns were produced that simplified RNA NMR spectra even further than D4-RNA pattern. By converting deuterated glucose into nucleotides in an enzymatic reaction conducted in D_2O, it was possible to produce nucleotides which had fully deuterated ribose moieties and fully protonated non-exchangeable base protons. Conversion of these nucleotides into RNA resulted in highly simplified NMR spectra.

1',2',3',4',5',5"-*d6*-TAR RNA (D6-TAR RNA) was produced by T7 RNA polymerase catalyzed transcription from 1',2',3',4',5',5"-*d6*-ATP (**24**), 1',2',3',4',5',5"-*d6*-GTP (**25**), 1',2',3',4',5',5"-*d6*-UTP (**26**), and 1',2',3',4',5',5"-*d6*-CTP (**27**), Figure 4.9a. The synthesis of **24-27** from deuterated glucose is described in Chapter 3, and the detailed procedures for the synthesis are in Chapter 6.

The D6-TAR RNA labeling pattern resulted in extremely simplified RNA NMR spectra as demonstrated by a NOESY spectra of D6-TAR RNA, Figure 4.10. Removal of all of the ribose protons from the RNA leaves the remaining non-exchangeable base protons relatively isolated (Figure 4.9c). This eliminates the majority of interactions which give rise to NOESY peaks in RNA NMR spectra, because most of the strong NOEs in a NOESY spectrum of RNA (Figure 4.2) arise from ribose-ribose and ribose-base NOEs. The only strong NOEs remaining in the D6-RNA NOESY are the intra-base pyrimidine H5-H6 crosspeaks (Figure 4.9b). While this type of labeling pattern does not yield very much structural information about the RNA, it does offer the possibility of spectral

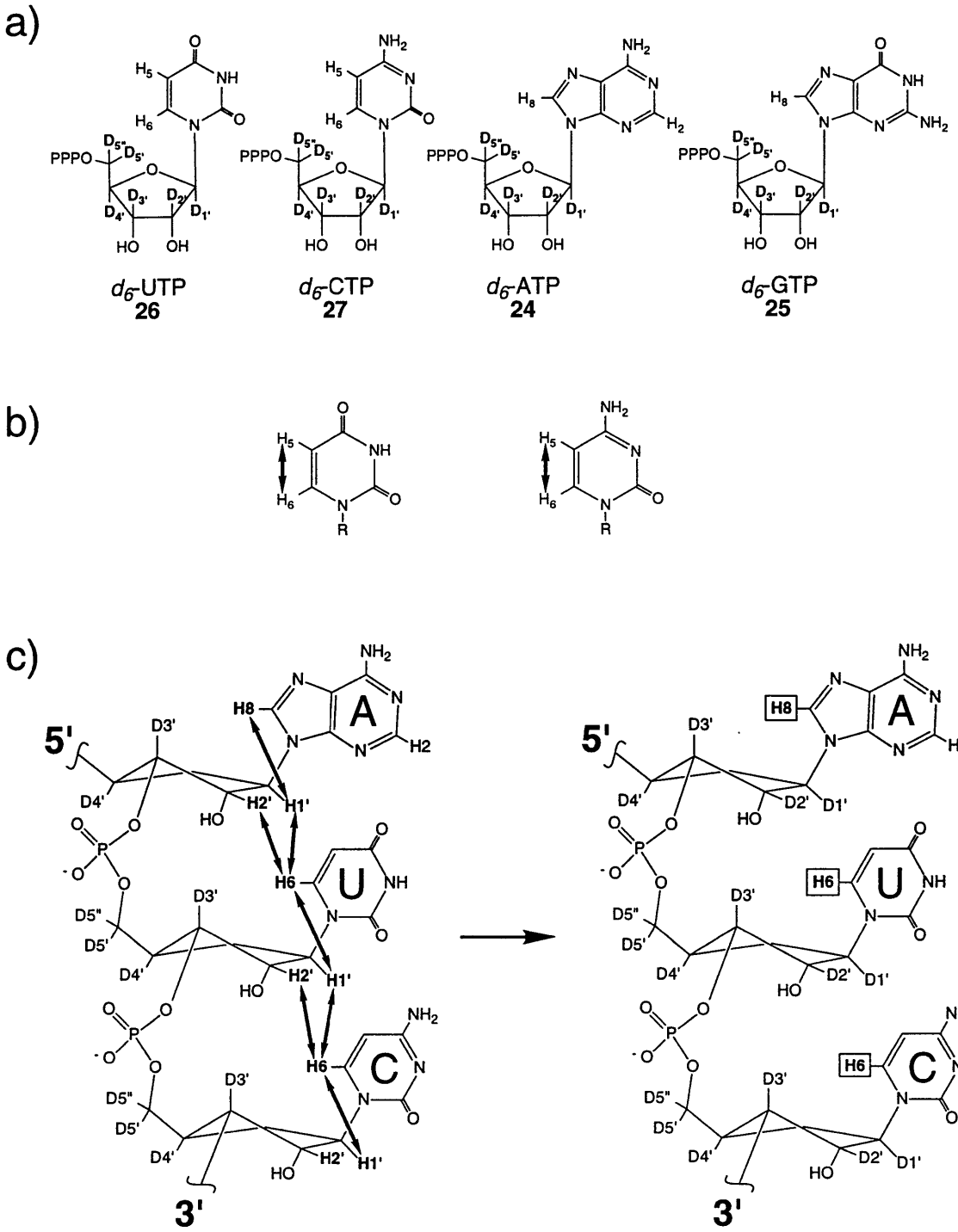


Figure 4.9 a) 1',2',3',4',5',5"-d_6-NTPs used to produce D6-TAR RNA b) the only strong NOEs remaining in D6-RNA are the intra-nucleotide pyrimidine H5-H6 crosspeaks. c) deuteration of all of the ribose protons in D6-RNA removes all ribose to base NOEs, leaving the non-exchangeable base protons relatively isolated.

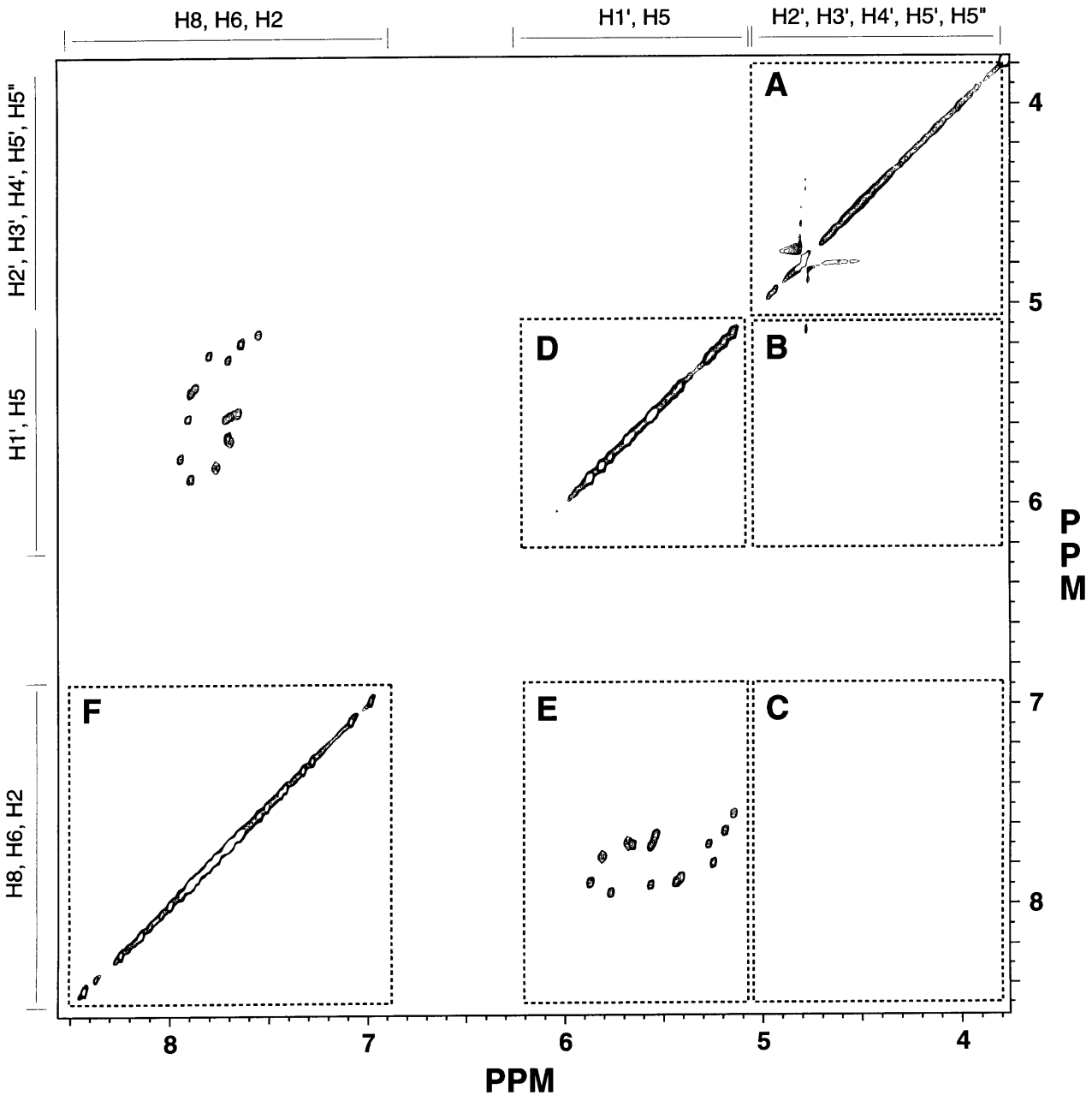


Figure 4.10 200 ms NOESY of D6-TAR RNA

simplification, either by using this labeling pattern in combination with another labeling pattern, or when studying an RNA-protein complex.

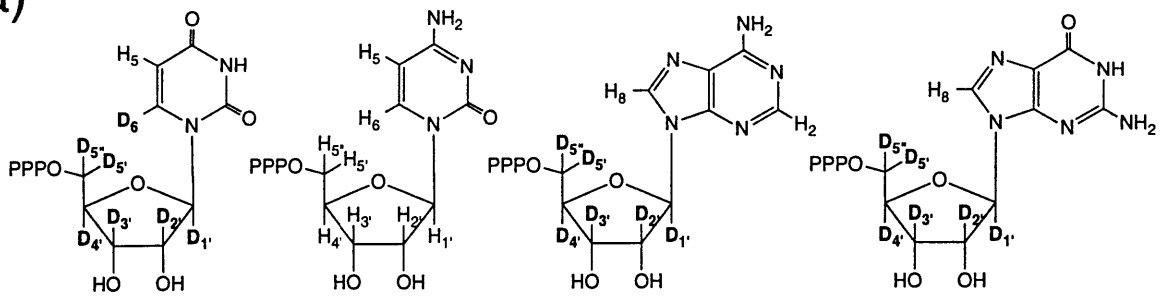
TAR-465 RNA

The TAR RNA discussed in this section, TAR 465, was produced for two reasons, it was synthesized to determine usefulness of mixing the D6-RNA labeling pattern with other labeling patterns, and also to test the usefulness of specifically deuterating the H6 proton of one of the pyrimidine nucleotides. The TAR 465 labeling pattern was produced by T7 RNA polymerase catalyzed transcription using 1',2',3',4',5',5"-*d6*-ATP (**24**), 1',2',3',4',5',5"-*d6*-GTP (**25**), 1',2',3',4',5',5",6-*d7*-UTP (**39**), and unlabeled CTP as starting materials, Figure 4.11a. The synthesis of **24**, **25**, and **39** from deuterated glucose, adenine, guanine, and orotic acid using D_2O is described in Chapter 3, and detailed reaction procedures are described in Chapter 6. The ribose portions of **24**, **25**, and **39** are derived from deuterated glucose and D_2O and are fully deuterated, and the H6 of uracil arises from D_2O during the decarboxylation of 1',2',3',4',5',5"-*d6*-orotidine monophosphate.

Mixing the D6-RNA labeling pattern with unlabeled nucleotides

The extreme spectral simplification caused by D6-RNA labeling pattern even though all of the base protons were still protonated suggested that mixing D6-RNA labeling with other types of RNA labeling might be profitable. A large amount of the important inter-nucleotide information that can be extracted from RNA NMR spectra consists of ribose to base interactions rather than ribose to ribose interactions, since there are relatively few inter-nucleotide ribose to ribose interactions. The TAR 465 RNA was synthesized with *d6*-ATP, *d6*-GTP, *d7*-UTP, and unlabeled CTP, to produce RNA in which all of the ribose to base interactions, except for uracil H6 NOEs, are present for the cytidine nucleotides, but none of those interactions are present for the *d6*-labeled or *d7*-labeled nucleotides,

a)



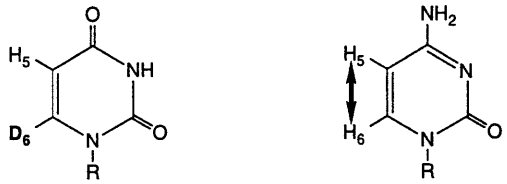
d_7-UTP
39

CTP

d_6-ATP
24

d_6-GTP
25

b)



c)

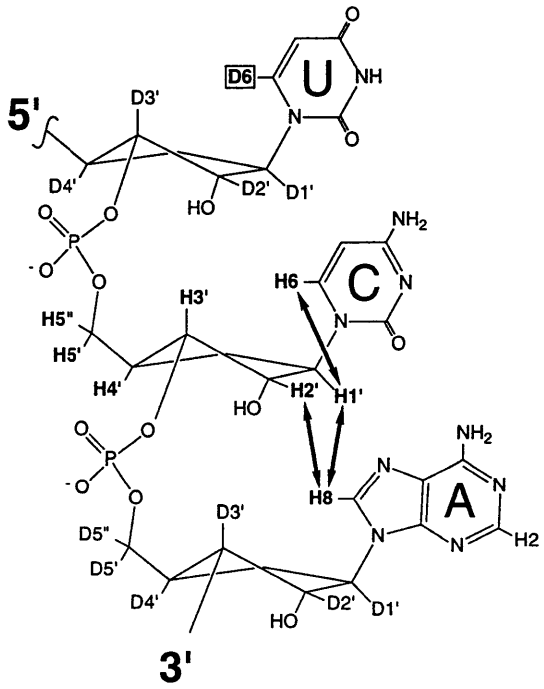


Figure 4.11 a) Nucleotides used to produce the TAR 465 labeling pattern b) Uracil's H5-H6 crosspeaks are removed from NOESY and COSY spectra because of deuteration of the H6, but cytidine H5-H6 crosspeaks remain. c) Inter-nucleotide and intra-nucleotide NOEs for C ribose protons are still present in TAR 465, but deuteration of A, G, and U ribose protons removes the corresponding A, G, and U NOEs. U-H6 NOEs are also removed by deuteration.

Figure 4.11c. The resulting spectral simplification is shown in Figure 4.12. All of the NOEs falling in regions **A**, **B**, and **C** of the NOESY spectra of TAR 465 arise from C ribose protons. The 30 nucleotide TAR RNA has 9 cytidines, so there are 21 less of each type of ribose proton to worry about in regions **A**, **B**, and **C** of Figure 4.12. In region **E** of the NOESY spectra, all of the H1' to base NOEs arise from cytidine H1's, and all of the H5-H6 NOEs arise from cytidines since the H6s of uracil are deuterated. While the mixing of unlabeled CTP with the D6-RNA labeling pattern did produce a great reduction in spectral crowding, the labeling pattern was a little disappointing since identification of individual peaks in Figure 4.12 is still difficult due to the spectral overlap of the different types of ribose protons in NOESY regions **A**, **B**, and **C.**

Deuteration of the H6 of uracil

Deuteration of the H6 of uracil in TAR 465 removed the strong uracil H5-H6 crosspeaks from NOESY and COSY spectra, and also removed all of the U-H6 to ribose NOEs from the TAR 465 NOESY spectra. Because only the U-H6 protons were deuterated, the assignment of pyrimidine H6 protons was greatly simplified in the TAR 465 RNA. An example of this is illustrated in Figure 4.13, in which the H5-H6 crosspeak region of a COSY spectra of unlabeled TAR RNA and TAR 465 are shown. There are 15 pyrimidines in TAR RNA and 15 H5-H6 crosspeaks can be observed in the COSY of unlabeled TAR RNA. Identification of which of the pyrimidine H5-H6 crosspeaks belong to U and which belong to C is not trivial in the unlabeled TAR RNA since the U and C crosspeaks are overlapping. In the TAR 465 COSY, only the cytidine H5-H6 crosspeaks are present since the H6 of uridines are deuterated. There are 9 cytidines in TAR, and 9 cytidine H5-H6 crosspeaks can be identified in the COSY of TAR 465. By comparing the COSY of TAR 465 with the COSY of unlabeled TAR RNA, the uracil H5-H6 crosspeaks can also be identified unambiguously.

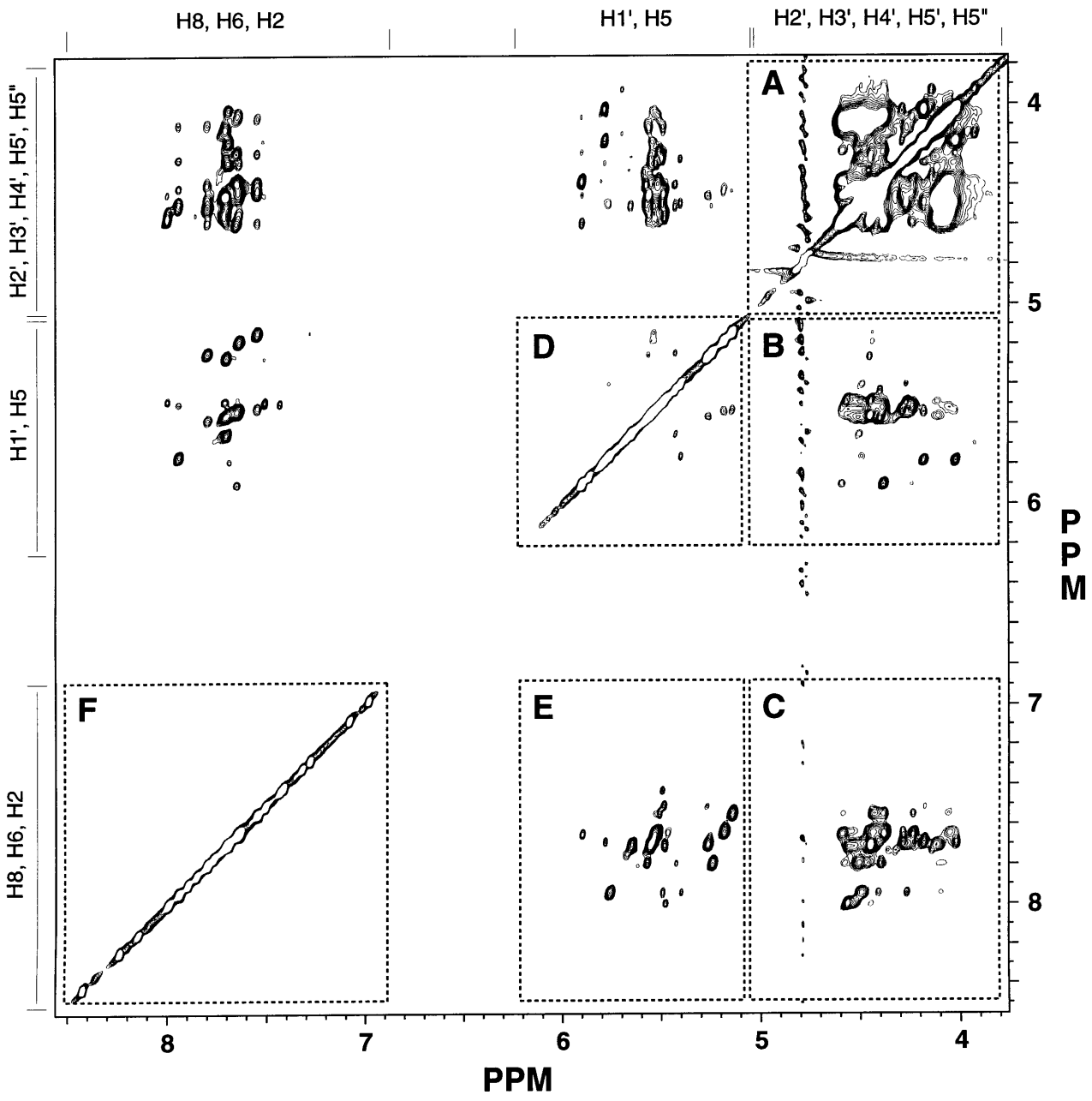


Figure 4.12 200 ms NOESY TAR 465

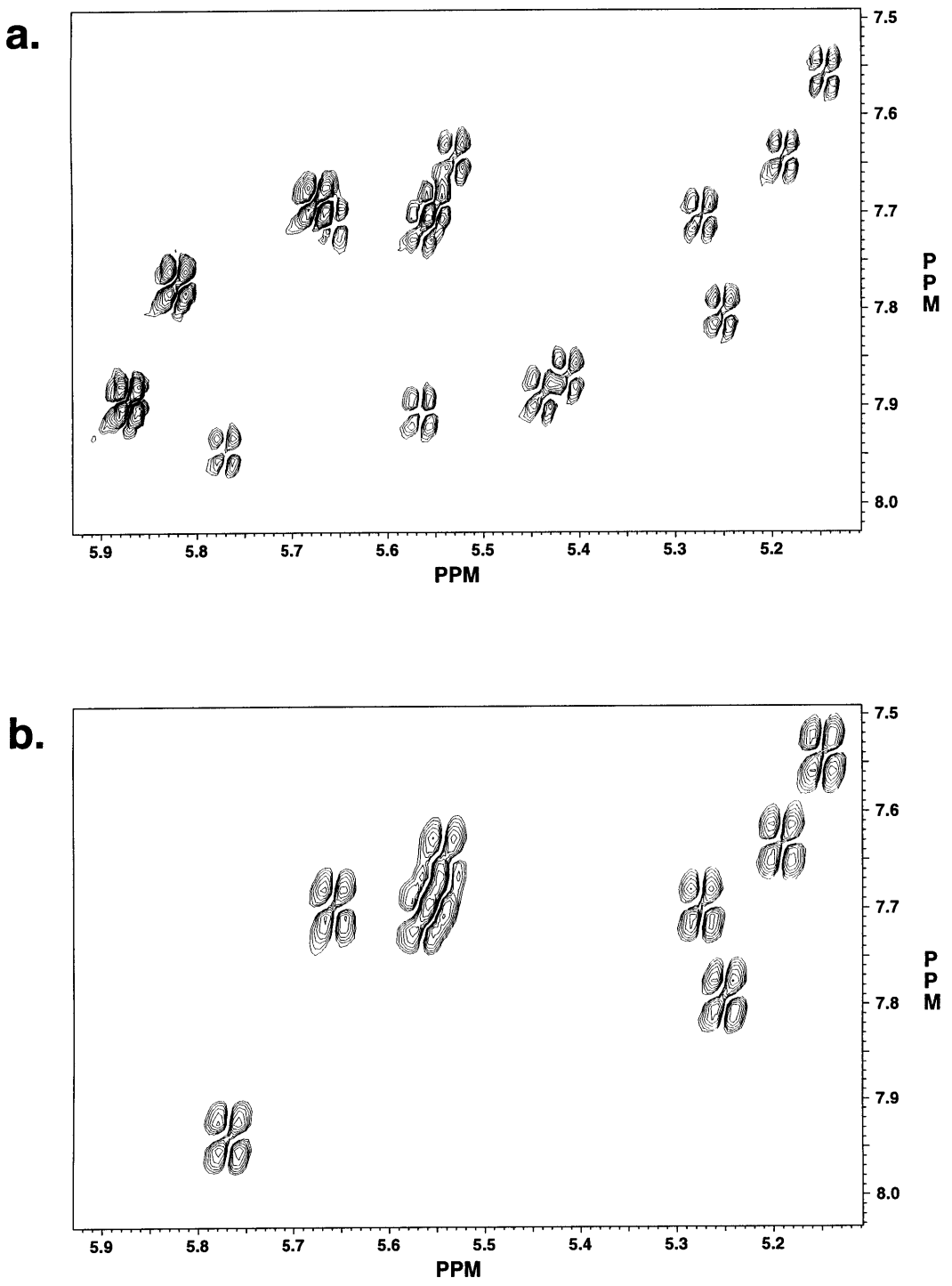


Figure 4.13 H5-H6 COSY crosspeak region of a) unlabeled TAR RNA with 15 pyrimidine H5-H6 crosspeaks b) TAR 465 RNA with only 9 H5-H6 cytidine crosspeaks due to the deuteration of the H6 of the six uridines in TAR 465.

Chapter 4

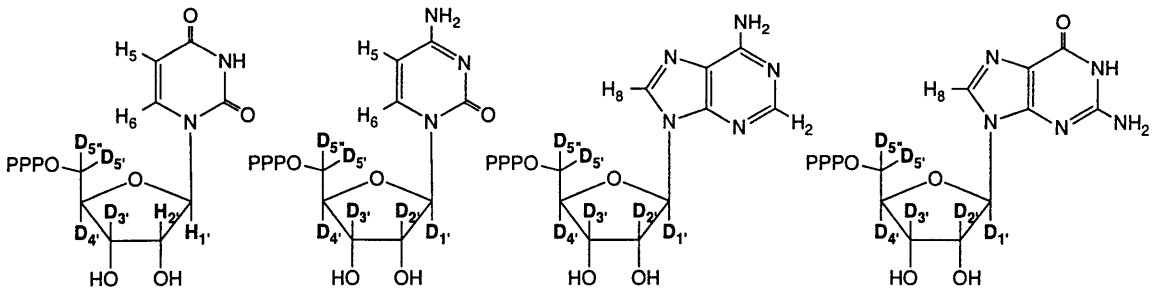
D4/D6-TAR RNA.

While the TAR 465 RNA type labeling pattern simplified NMR spectra greatly, there was still a lot of spectral crowding in the ribose regions of the spectra. NOESY regions **A**, **B**, and **C** of Figure 4.12 were still quite crowded and the identification of types of NOEs within these regions was complicated by the overlap of the H2', H3', H4', H5', and H5" protons. To solve some of the ribose overlap problems in NOESY spectra, the D4-RNA labeling pattern was mixed with the D6-RNA labeling pattern. RNA of this type, D4/D6-RNA, would benefit from the reduction in spectral crowding of ribose protons due to the D4-RNA labeling pattern for selected nucleotides, and the ribose protons of other nucleotides would be removed with D6-RNA labeling.

d4-U/*d6*-A,G,C-TAR RNA (D4/D6-RNA) was produced by T7 RNA polymerase catalyzed transcription using 1',2',3',4',5',5"-*d6*-ATP (**24**), 1',2',3',4',5',5"-*d6*-GTP (**25**), 1',2',3',4',5',5"-*d6*-CTP (**27**), and 3',4',5',5"-*d4*-UTP (**4**) as starting materials, Figure 4.14a. The synthesis of **24**, **25**, **27**, and **4** from deuterated glucose is described in Chapter 3, and the detailed procedures for the synthesis are in Chapter 6.

The only ribose protons in the D4/D6-TAR RNA are the 1' and 2' protons of uridine nucleotides, but the non-exchangeable base protons for all nucleotide types are still present. Because of this, all of the base to ribose interactions for uridine nucleotides in the D4-TAR RNA are still present in the D4/D6-TAR RNA, and all other base to ribose interactions have been removed, Figure 4.14b. A NOESY spectra of the D4/D6-TAR RNA, Figure 4.15, illustrates a dramatic decrease in spectral crowding as compared to the unlabeled RNA NOESY spectra, Figure 4.2, or even the D4-RNA, Figure 4.5. All of the NOEs in Figure 4.15 sections **B** and **C**, and several of the NOEs in section **E** arise from uridine 1' and 2' protons. There are 6 uridines in TAR RNA, and it is easy to locate the 6 strong intra-nucleotide H1'-H2' crosspeaks and the 6 strong inter-nucleotide H2'-

a.



d_4-UTP **4** d_6-CTP **27** d_6-ATP **24** d_6-GTP **25**

b.

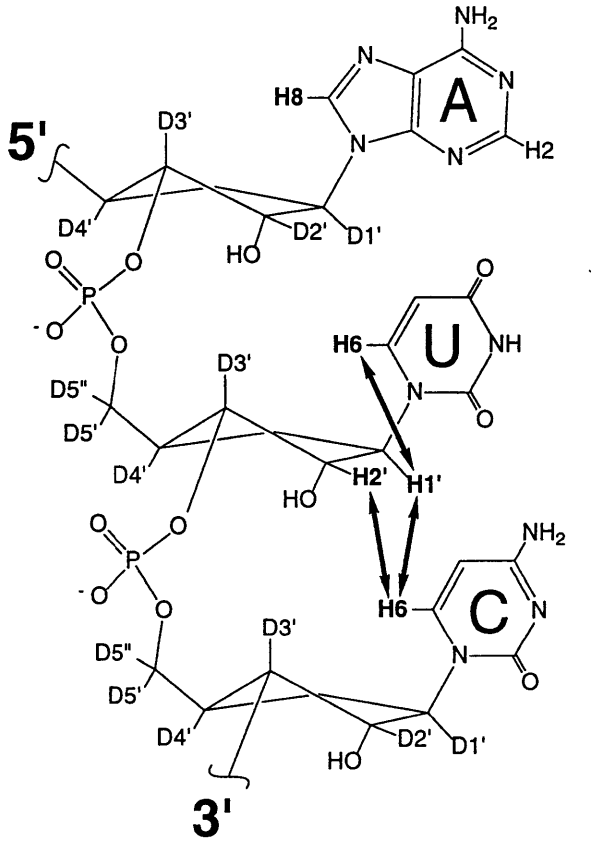


Figure 4.14 a) Nucleotides used to produce the D4/D6 TAR RNA labeling pattern b) Inter-nucleotide and intra-nucleotide NOEs for U-H1' and U-H2' ribose protons are still present in D4/D6-RNA, but deuteration of A, G, and C ribose protons removes all A, G, and C ribose NOEs.

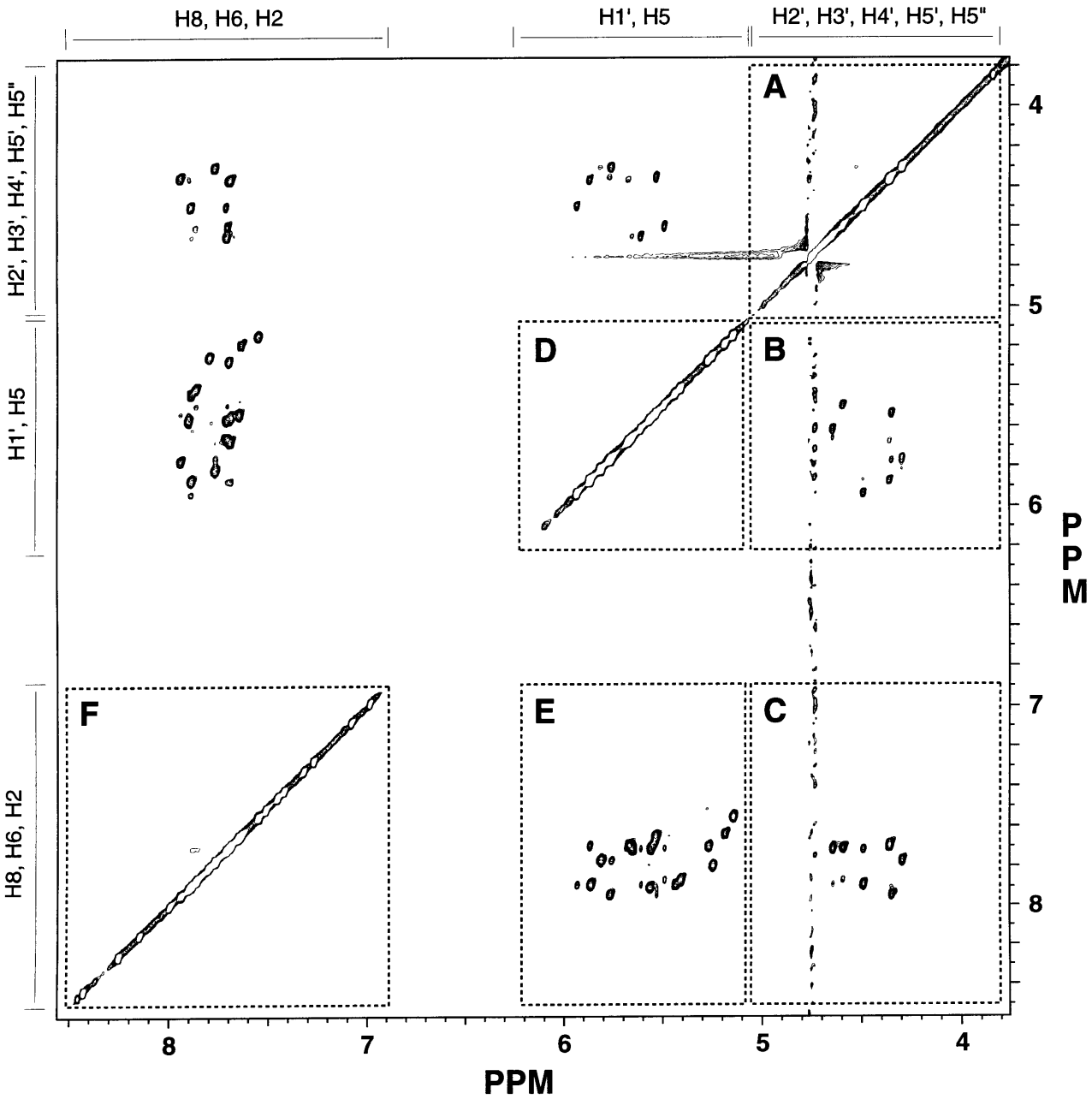


Figure 4.15 200 ms NOESY D4/D6 TAR RNA

base crosspeaks in Figure 4.15. While that information is also present in the D4-RNA spectra in Figure 4.5, the presence of the NOEs arising from A, G, and C ribose protons makes assignment of the uridine NOEs more difficult, although it is still a much simpler case than assigning uridine H1' and H2' NOEs in the unlabeled RNA, Figure 4.2. The selective spectral editing shown here might be useful in simplifying very complicated spectra of large RNAs.

Sequential assignment using the D4/D6-TAR RNA is similar to the sequential assignment of the D4-RNA labeling pattern except that only the U ribose protons are present in the spectra, Figure 4.16. H2' to base NOEs can be linked to H1' to base NOE information through the H1' and H2' crosspeaks as in the D4-TAR RNA, but the spectral crowding in regions **B**, **C**, and **E** is greatly reduced, making identification of the base protons that are 3' and 5' of the ribose protons much easier. Assignment of all of the uridine H1' and H2' protons of the D4/D6-TAR RNA is by inspection, without any overlap problems as illustrated by Figure 4.16b. Unfortunately, complete sequential assignment of the TAR RNA sequence cannot be made with the D4/D6-TAR RNA sample since only uridine sequential connectivity information is available. Other NMR samples containing *d4*-A, *d4*-G, and *d4*-C would have to be made to completely assign the RNA. While this may be tedious, it may offer a way of obtaining sequential assignments of very large RNAs where heteronuclear NMR is inefficient.

D5 pyrimidine-TAR RNA

An inspection of the base to H1', H5 region of a NOESY spectra of RNA, Figure 4.17, will reveal two types of NOEs. The first type of NOE is usually weaker than the other and arises from the interaction between H1' ribose protons and the nonexchangeable base protons (purine H8, adenine H2, pyrimidine H6); these peaks are shown as the normal peaks in Figure 4.17. These ribose to base NOEs are important because in A form

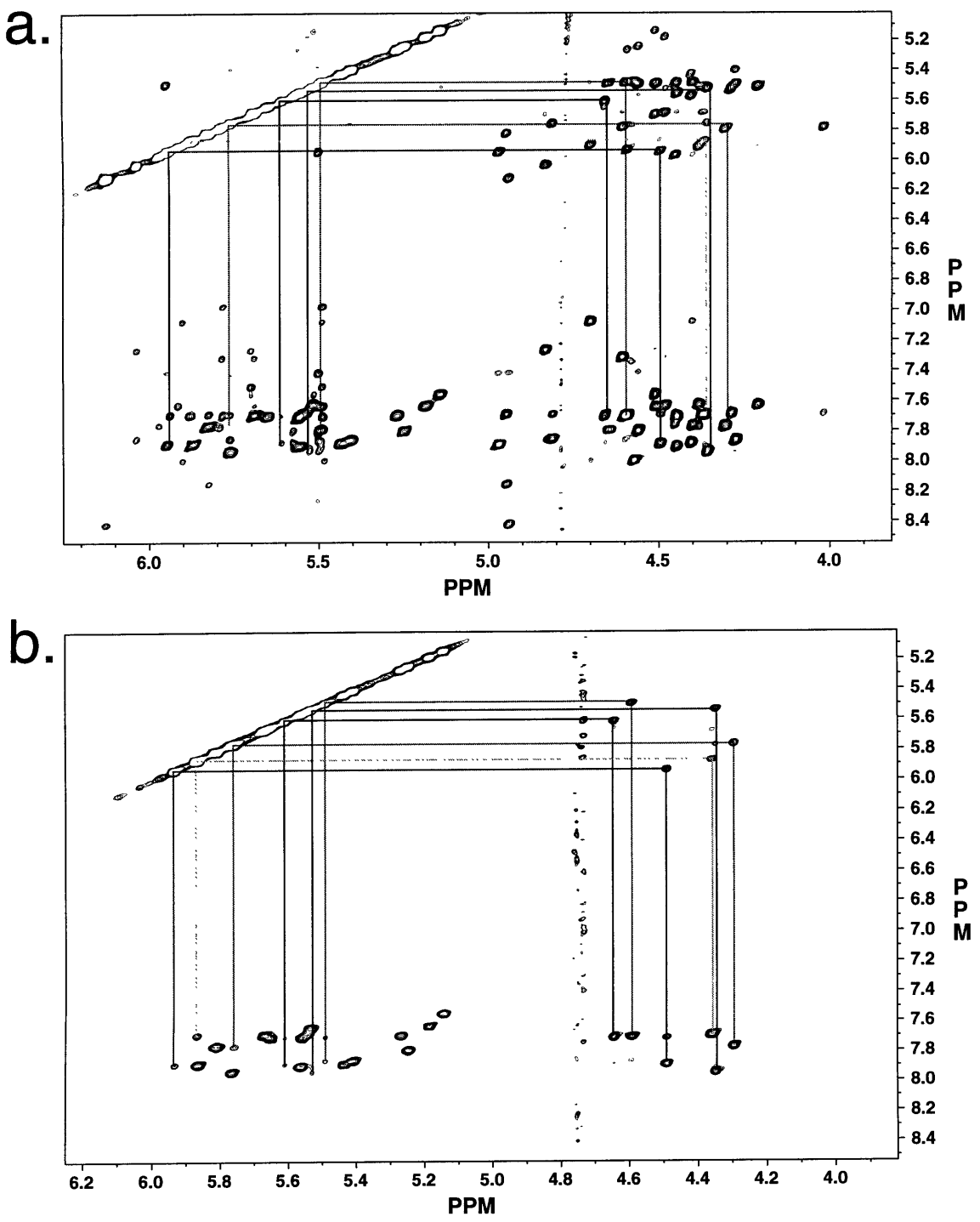


Figure 4.16 NOE connectivity of uridine nculeotides for a) D4-TAR RNA b) D4/D6-TAR RNA. (From left to right: H1'-base NOEs, H1' crosspeaks on the diagonal, H1'-H2' NOEs, and H2'-base NOEs.)

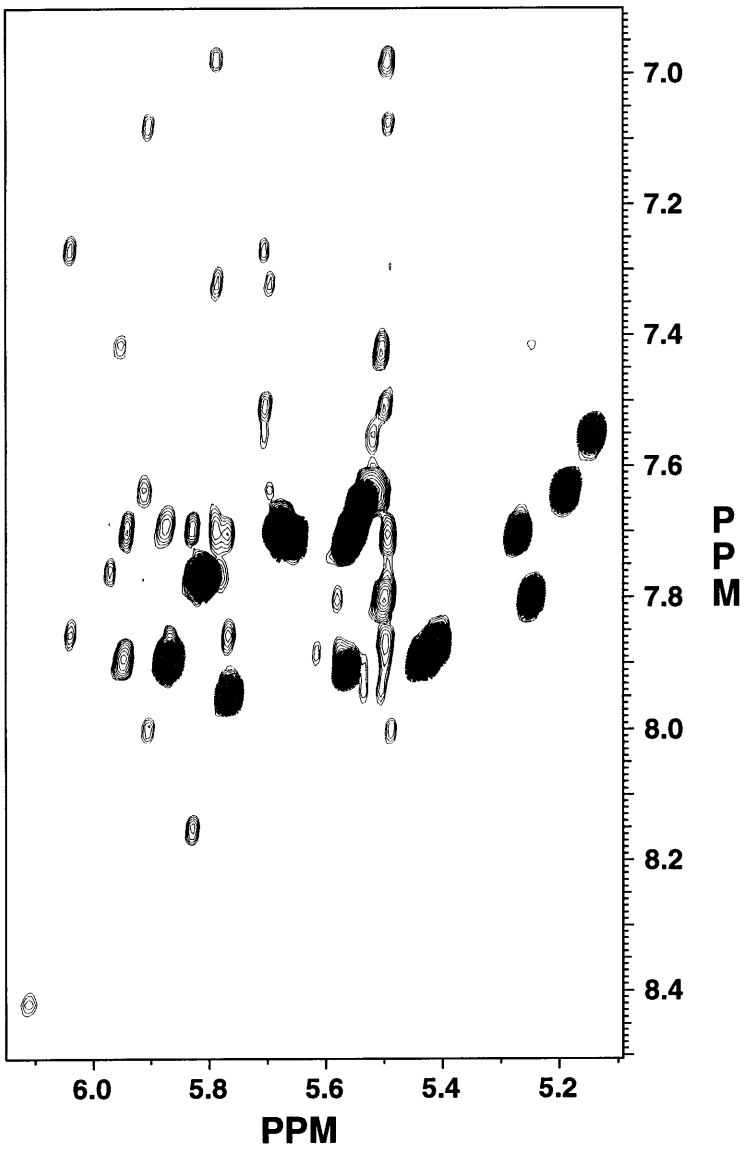


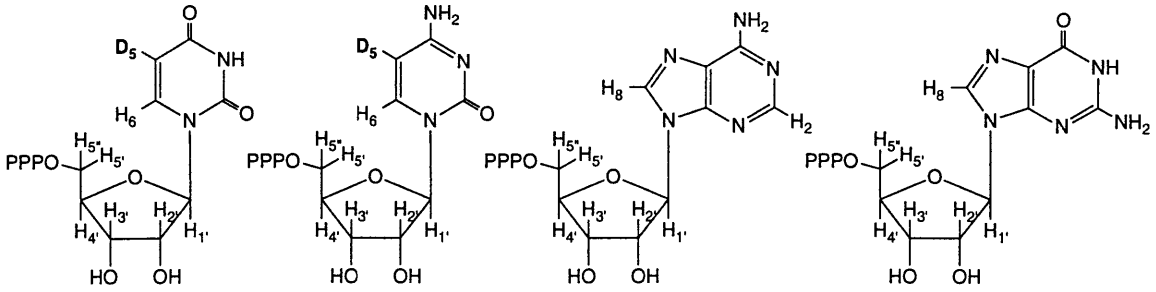
Figure 4.17 Base to H1', H5 region of TAR RNA 200 ms NOESY. There are two types of NOEs within this region, base-H1' NOEs (unshaded) and pyrimidine H5-H6 NOEs (shaded).

RNA they arise from both inter and intra-nucleotide interactions which can be used to walk from one nucleotide to another in NOESY based sequential assignment. The other type of NOE in this region is a strong NOE that arises from the intra-pyrimidine H5-H6 interaction; these peaks are shown as the shaded peaks in Figure 4.17. The distance between the H5 and H6 protons varies little and is relatively short range (2.4 Å), resulting in a strong NOE. These protons also yield crosspeaks in COSY spectra of RNA, which are used to help determine which base protons belong to pyrimidines in assignment of NMR spectra. Unfortunately, the strong H5-H6 NOEs fall in the H1'-base NOE region making NOESY-based sequential assignment difficult because the strong H5-H6 NOEs can eclipse some of the weaker H1' to base NOEs. In the TAR RNA, only a few of the H1' to base NOEs fall under H5-H6 crosspeaks, but this problem becomes much worse as larger RNA molecules are studied since the spectral width of the NOE region remains the same but the number of H5-H6 NOEs within the region increases linearly with the number of pyrimidines. In order to reduce this problem the H5 of pryrimidines can be deuterated chemically, leaving the H1'-H6 NOEs that are important for sequential assignment within the NMR spectra, but removing the strong H5-H6 NOEs which obscure sequential assignment, Figure 18b.

D5-pyrimidine-TAR RNA was produced by T7 RNA polymerase catalyzed transcription using unlabeled ATP, unlabeled GTP, 5-*d1*-UTP (**37**), and 5-*d1*-CTP (**38**) as starting materials, Figure 4.18a. The synthesis of **37** and **38** by ammonium bisulfite exchange of the H5 of UMP is described in Chapter 3, and the detailed procedures for the synthesis are in Chapter 6.

A NOESY spectra of the D5-pyrimidine RNA was acquired and compared to the NOESY spectra of unlabeled RNA. Most of the NOE regions of the D5-pyrimidine-TAR RNA are identical to the unlabeled RNA NMR spectra, but the H1' to base NOE region (region **E** Figure 4.2) there are no H5-H6 NOEs. A comparison of this region of NOESY spectra of D5-pyrimidine-TAR RNA and unlabeled TAR RNA is shown in Figure 4.19.

a.



b.

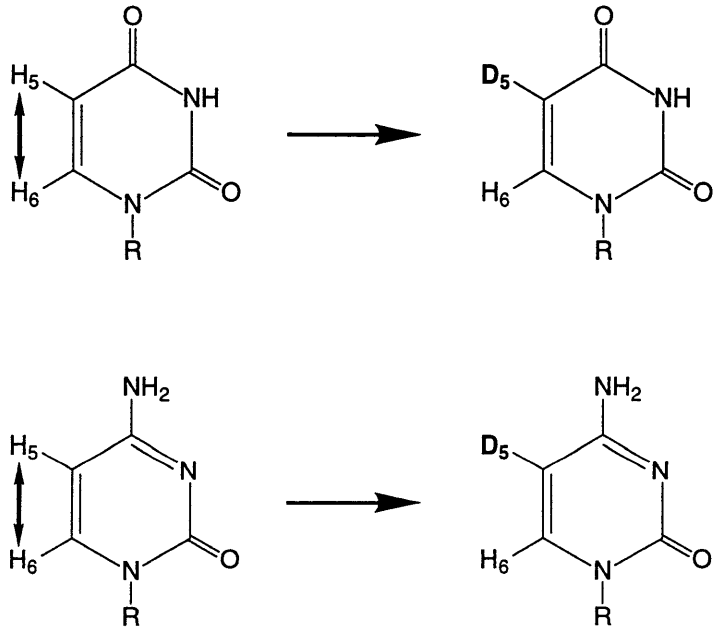


Figure 4.18 a) Nucleotides used to transcribed D5-pyrimidine TAR RNA b) deuteration of the H5 of pyrmidines will remove H5-H6 crosspeaks, but it will not remove ribose-H6 crosspeaks.

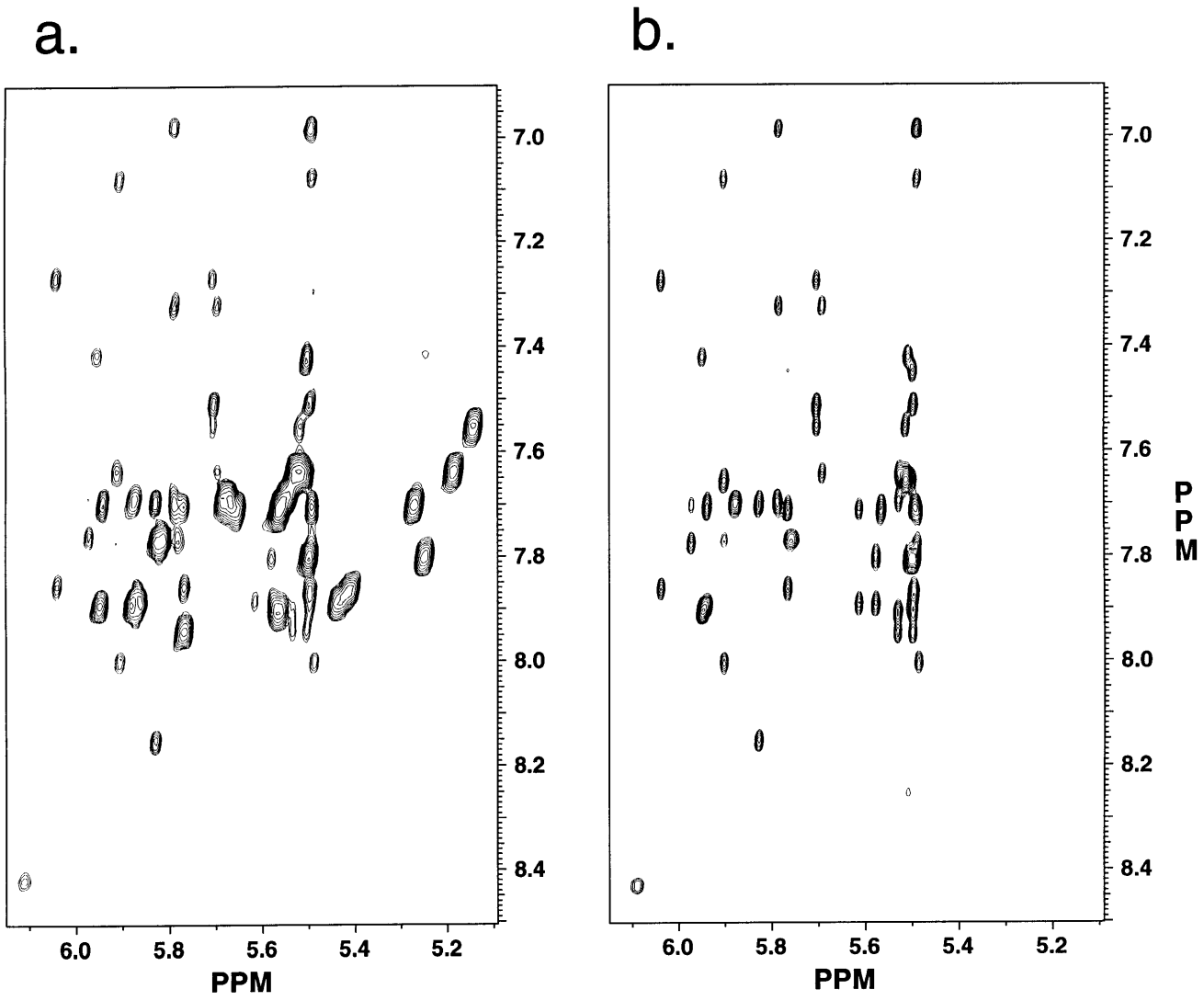


Figure 4.19 Base to H1', H5 NOE region of TAR RNA for a) unlabeled TAR RNA b) D5-pyrimidines TAR RNA. The H5-H6 crosspeaks of the D5-pyrmidine TAR RNA are removed, but all of the base-H1' NOEs are still present.

Notice that there are several H1' to base NOEs that are obscured by H5-H6 crosspeaks between 5.6 and 5.5 ppm in the unlabeled TAR RNA, which are uncovered in the D5-pyrimidine-TAR RNA. In a larger RNA there would be many more H5-H6 NOEs obscuring this region and covering up the H1' to base NOEs which help in sequential assignment. Deuteration of the H5 of pyrimidines may help in the sequential assignment of such larger RNAs.

Specific heteronuclear labeling of RNA.

Heteronuclear labeling has had a great impact on the way NMR structural studies are carried out on both RNA and proteins. In the past four years much of the NMR structural work on proteins and RNA has been conducted with heteronuclear labeling that use heteronuclear chemical shifts to spread out proton resonances and also utilize through bond assignment procedures (Kay & Gardner, 1997; Varani *et al.*, 1996). All of the isotopic labeling patterns that have been discussed so far in this chapter have been specific deuteration patterns. Using glucose as a starting material for preparing RNA as described in Chapter 3 allows all of the ribose labeling patterns and most of the base labeling patterns discussed so far to be synthesized with ^{13}C labels in addition to specific deuterium labels. By combining specific deuteration with ^{13}C labeling, RNA molecules can be created where fully protonated ^{13}C labeled carbons are adjacent to fully deuterated ^{13}C labeled carbons. Cases similar to this have been useful in obtaining sequential assignment and side chain assignment in protein NMR spectroscopy. Exchangeable amide protons of perdeuterated $^{13}C/^{15}N$ labeled proteins can be used to start NMR experiments and detect magnetization on (Kay & Gardner, 1997). With the carbons of the protein deuterated, magnetization has a much longer lifetime and longer transfers of magnetization on larger proteins can be achieved with deuteration than are otherwise possible . Here we start applying these ideas to RNA NMR spectroscopy, while also reducing overlap in heteronuclear spectra with specific deuteration.

Chapter 4

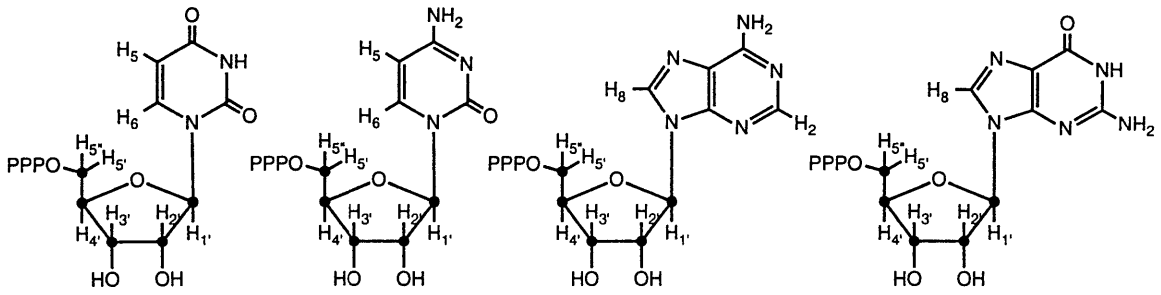
Simplification of heteronuclear spectra with specific deuteration.

1',2',3',4',5'-$^{13}C_5$-TAR RNA (*$^{13}C_5$-ribose*-TAR RNA) was produced by T7 RNA polymerase catalyzed transcription using 1',2',3',4',5'-$^{13}C_5$-ATP (**28**), 1',2',3',4',5'-$^{13}C_5$-GTP (**29**), 1',2',3',4',5'-$^{13}C_5$-UTP (**30**), and 1',2',3',4',5'-$^{13}C_5$-CTP (**31**) as starting materials, Figure 4.20a. The synthesis of **28-31** from ^{13}C-glucose is described in Chapter 3, and the detailed procedures for the synthesis are in Chapter 6.

1',2',3',4',5'-$^{13}C_5$-3',4',5',5"-*d4*-TAR RNA (*D4-$^{13}C_5$-ribose*-TAR RNA) was produced by T7 RNA polymerase catalyzed transcription using 1',2',3',4',5'-$^{13}C_5$-3',4',5',5"-*d4*-ATP (**32**), 1',2',3',4',5'-$^{13}C_5$-3',4',5',5"-*d4*-GTP (**33**), 1',2',3',4',5'-$^{13}C_5$-3',4',5',5"-*d4*-UTP (**34**), and 1',2',3',4',5'-$^{13}C_5$-3',4',5',5"-*d4*-CTP (**35**) as starting materials, Figure 4.20b. The synthesis of **32-35** from ^{2}H-^{13}C-glucose is described in Chapter 3, and the detailed procedures for the synthesis are in Chapter 6.

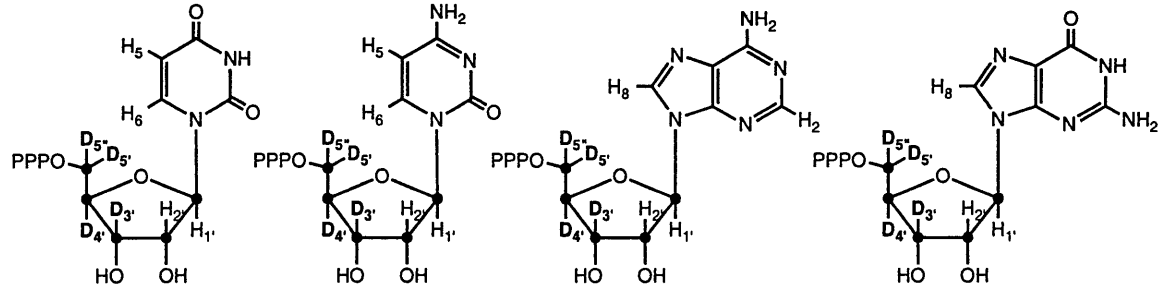
HSQC (Heteronuclear Single Quantum Coherence) spectra of *^{13}C-ribose*-RNA and *D4-^{13}C-ribose*-RNA were acquired, and are shown in Figure 4.21a and 4.21b. As can be seen in Figure 4.21a, the ^{13}C chemical shifts of ribose are fairly well dispersed by type of carbon, but the C2' and C3' carbon chemical shifts overlap. This is compounded by the overlap of the H2' and H3' proton chemical shifts, making the identification of some of the types of crosspeaks difficult. The crosspeaks within the boxes in Figure 4.21a are some of the H3'-C3' crosspeaks of TAR that overlap the H2'-C2' crosspeak region. Deuteration of the C3' of ribose in the *D4-^{13}C-ribose*-RNA, Figure 4.21b, makes the identification of exactly which peaks belong to C2' carbons much easier. The absence of the boxed crosspeaks in Figure 4.21b, that were present in Figure 4.21a, make it clear that they are H3'-C3' crosspeaks.

a.



$^{13}C_5$-UTP **30** $^{13}C_5$-CTP **31** $^{13}C_5$-ATP **28** $^{13}C_5$-GTP **29**

b.



d_4-$^{13}C_5$-UTP **34** d_4-$^{13}C_5$-CTP **35** d_4-$^{13}C_5$-ATP **32** d_4-$^{13}C_5$-GTP **33**

Figure 4.20 Nucleotides used to transcribe a) 1',2',3',4',5'-$^{13}C_5$-ribose TAR RNA b) 3',4',5',5"-d_4-1',2',3',4',5'-$^{13}C_5$-ribose TAR RNA (Black dots represent ^{13}C.)

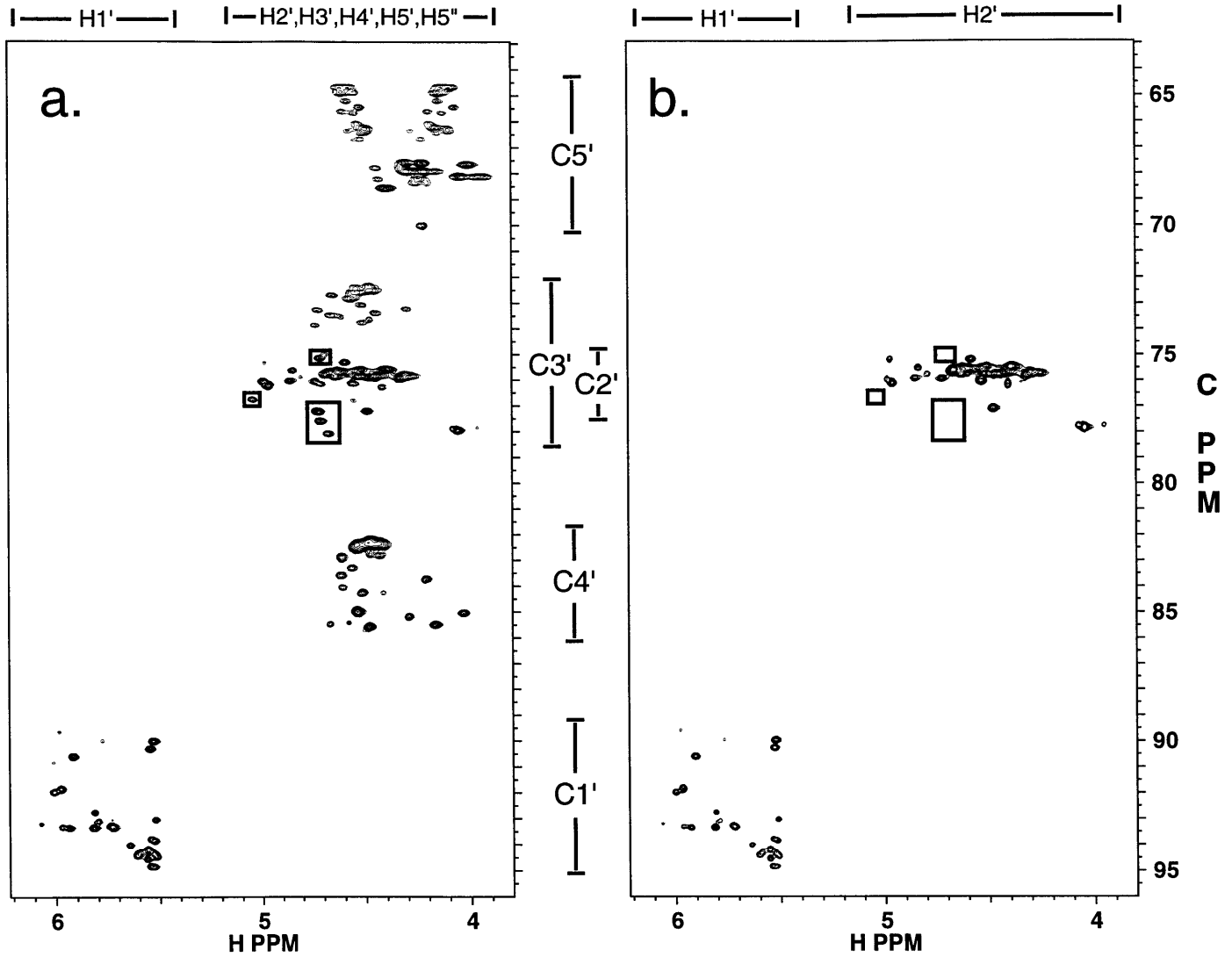


Figure 4.21 HSQC spectra of (a) $^{13}C_5$-*ribose*-TAR RNA and (b) d_4-$^{13}C_5$-*ribose*-TAR RNA. Chemical shift regions for different types of protons and carbons are indicated. The boxes inside the $^{13}C_5$-*ribose*-TAR RNA spectra outline some of the H3'-C3' crosspeaks that overlap the H2'-C2' crosspeak region, and the boxes inside the d_4-$^{13}C_5$-*ribose*-TAR RNA spectra show the absence of these H3'-C3' crosspeaks in the deuterated RNA spectra.

Chapter 4

^{13}C labeling of RNA from glucose offers the possibility of combining the advantages of multidimensional heteronuclear NMR with the spectral editing advantages of specific deuteration. Deuteration in this context will alter the relaxation properties of RNA, and it will be possible to create new types of through bond NMR experiments which utilize these labeling patterns (Dayie *et al.*, 1997). In addition, the *^{13}C-ribose*-RNA offers an inexpensive and easy way to incorporate ^{13}C labels into the ribose moieties of RNA.

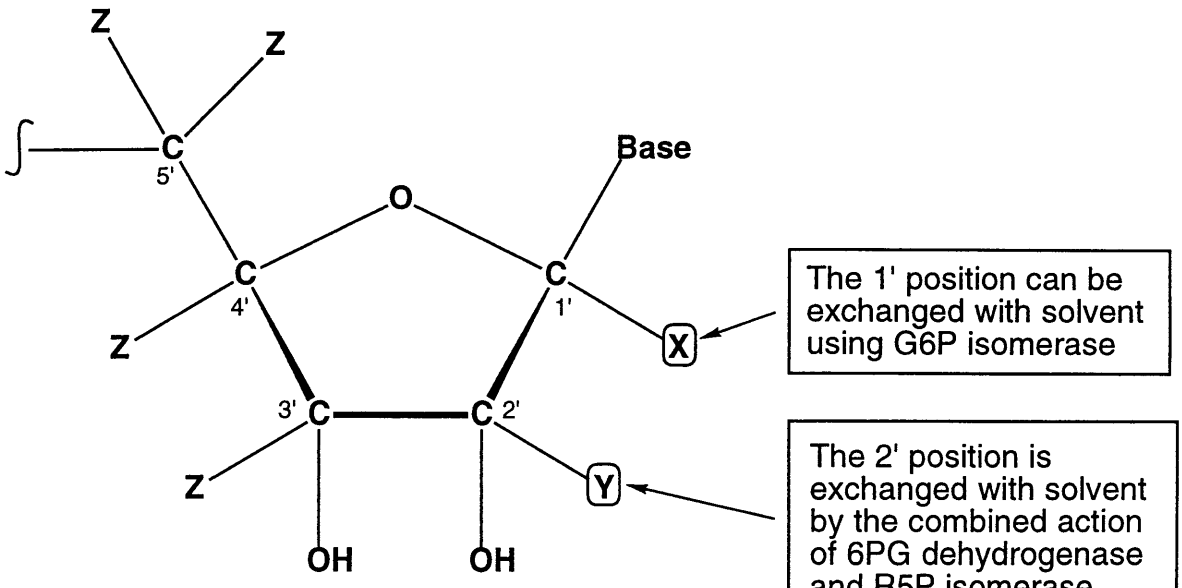
Chapter 5

5. Future Directions

As the field of solution NMR spectroscopy progresses the upper limit of the size of macromolecules which can be studied is constantly being pushed to new levels by advances in labeling and NMR experimental techniques (Wagner, 1997; Clore & Gronenborn, 1997; Kay & Gardner, 1997). The isotopic labeling strategies described in this thesis should increase the size of RNA molecules which can be effectively studied with NMR spectroscopy by reducing spectral crowding and linewidth problems which interfere with NMR structural studies. Using the techniques described in Chapters 2 and 3, the ribose moieties of RNA can be synthesized with a wide variety of specific deuterium and/or ^{13}C labeling patterns. As an example of the flexibility of these methods, Figure 5.1 summarizes how 16 different isotopically labeled ribose moieties can be produced using 4 types of commercially available labeled glucose. The bases of RNA can also be labeled with ^{2}H, ^{13}C, or ^{15}N as described in Chapter 3. Figure 5.2 summarizes how all of nonexchangeable base protons, except for the H2 of adenine, can be exchanged with deuterium, and examples of heteronuclear labeling of the base moietiesof RNA are shown in Figure 3.9 of Chapter 3. By combining the synthetic techniques described in this thesis with several commercially available, isotopically labeled starting materials, many isotopic labeling patterns can be synthesized which should increase the size of RNA molecules which can be studied with NMR spectroscopy. Methods for improving both homonuclear and heteronuclear NMR spectra of large RNA molecules with labeling patterns produced in this thesis are discussed in this chapter.

Improving Homonuclear RNA NMR Spectra with Specific Deuteration

One of the advantages of homonuclear NMR spectroscopy is that it has better relaxation characteristics than heteronuclear NMR spectroscopy. For two protons with the



a)

X (H1')	Y (H2')	Z (H3', H4', H5', H5")
D	D	D
H	D	D
H	H	D
D	H	D

b)

X (H1')	Y (H2')	Z (H3', H4', H5', H5")
H	H	H
D	H	H
D	D	H
H	D	H

Figure 5.1 Possible labeling of the ribose moieties of RNA which can be derived from glucose. **a)** Labeling patterns which can be derived from 1,2,3,4,5,6,6-D_7-glucose. **b)** Labeling patterns which can be derived from protonated glucose. ^{13}C labeling of the 1', 2', 3', 4', and 5' carbons can be incorporated into RNA by using 1,2,3,4,5,6-$^{13}C_6$-1,2,3,4,5,6,6-D_7-glucose, to produce ^{13}C labeled versions of the labeling patterns in the table in part **a**, and 1,2,3,4,5,6-$^{13}C_6$-glucose to produce ^{13}C labeled version of the labeling patterns in the table in part **b**.

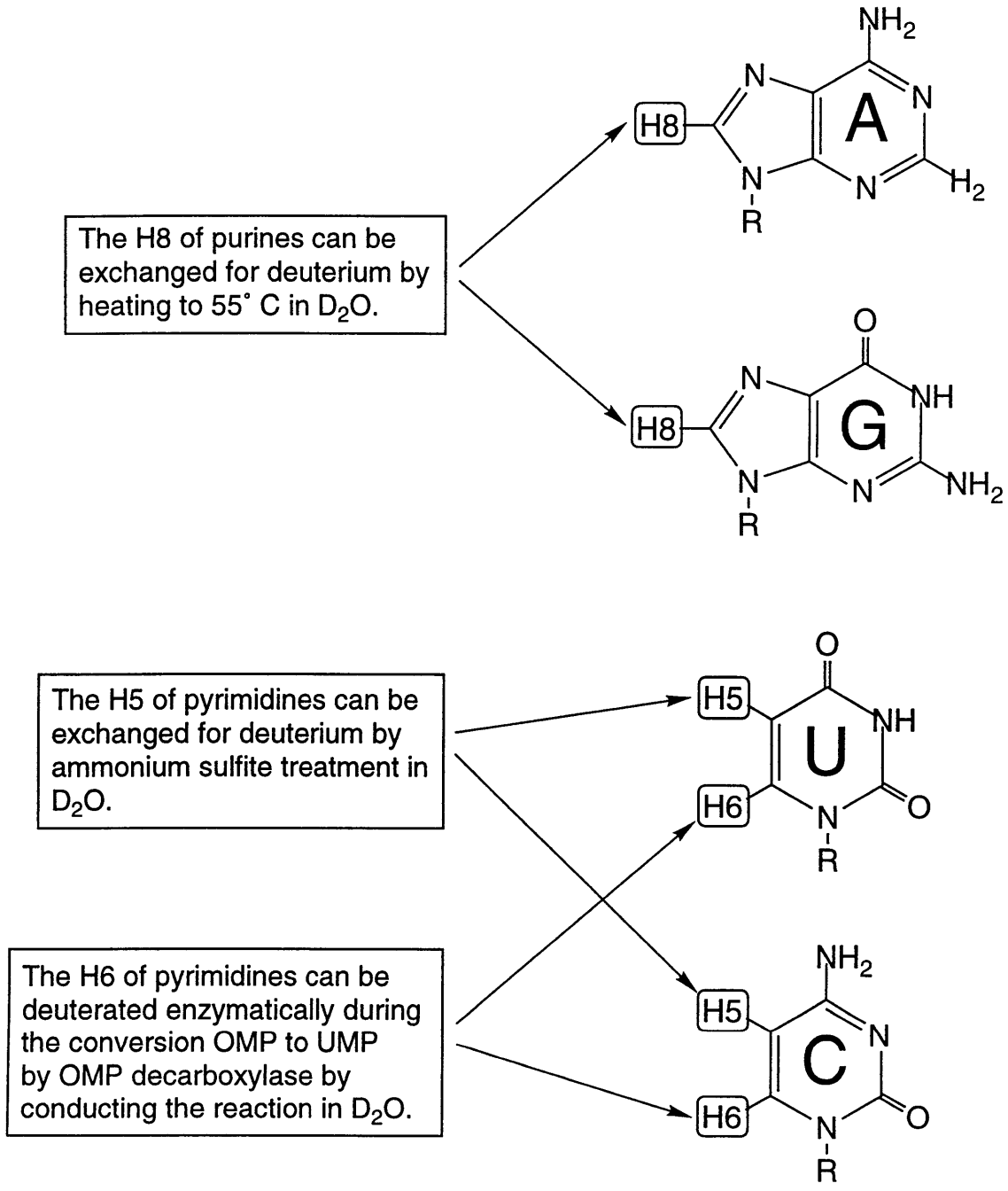


Figure 5.2 Summary of possible base proton deuteration discussed in this thesis.

same surroundings except for one being attached to ^{13}C and the other being attached to ^{12}C, the proton that is attached to ^{12}C will always have a slower relaxation rate because the ^{13}C attached proton has additional pathways for relaxation through interactions with the ^{13}C that it is bound to. This leads to the ironic situation of being able to acquire good quality proton NMR spectra of large, unlabeled macromolecules when heteronuclear NMR spectra of heteronuclear labeled versions of the same large macromolecules are of poor quality due to rapid relaxation. This situation is ironic because although the homonuclear NMR spectra may be of good quality in terms of signal to noise and linewidths, the proton NMR spectra of large macromolecules are usually hopelessly crowded due to the number of resonances in the spectra and the lack of a heteronuclear dimension into which to spread the information.

The specific deuteration patterns described in this thesis could be used to reduce spectral crowding in RNA NMR spectra so that even the homonuclear NMR spectra of very large RNA molecules would be simplified such that interpretation would be possible, Figure 5.3. Spectral crowding in the ribose regions of large RNA NMR spectra could be greatly simplified by the D4-RNA labeling pattern. Further reduction of spectral crowding could be obtained by mixing D4-RNA labeling with D6-RNA labeling as described in Chapter 4. Rather than making only a single NMR sample such as the [D4-U and D6-A,G,C]-TAR sample which was synthesized in Chapter 4, four NMR samples could be prepared with each sample having only one D4-labeled nucleotide so that there was one D4/D6-RNA for each type of nucleotide. With four D4/D6-RNAs of the type shown in Figure 5.4b [(D4-A, D6-G,U,C)-RNA, (D4-G, D6-A,U,C)-RNA, (D4-U, D6-A,G,C)-RNA, and (D4-C, D6-A,G,U)-RNA], all of the spectral information that was available in uniformly labeled D4-RNA, except for inter-nucleotide ribose to ribose information, would still be available from the combined spectra of the four D4/D6-RNAs. Spectral crowding could be reduced by as much as a factor of 4 when compared to uniformly D4-

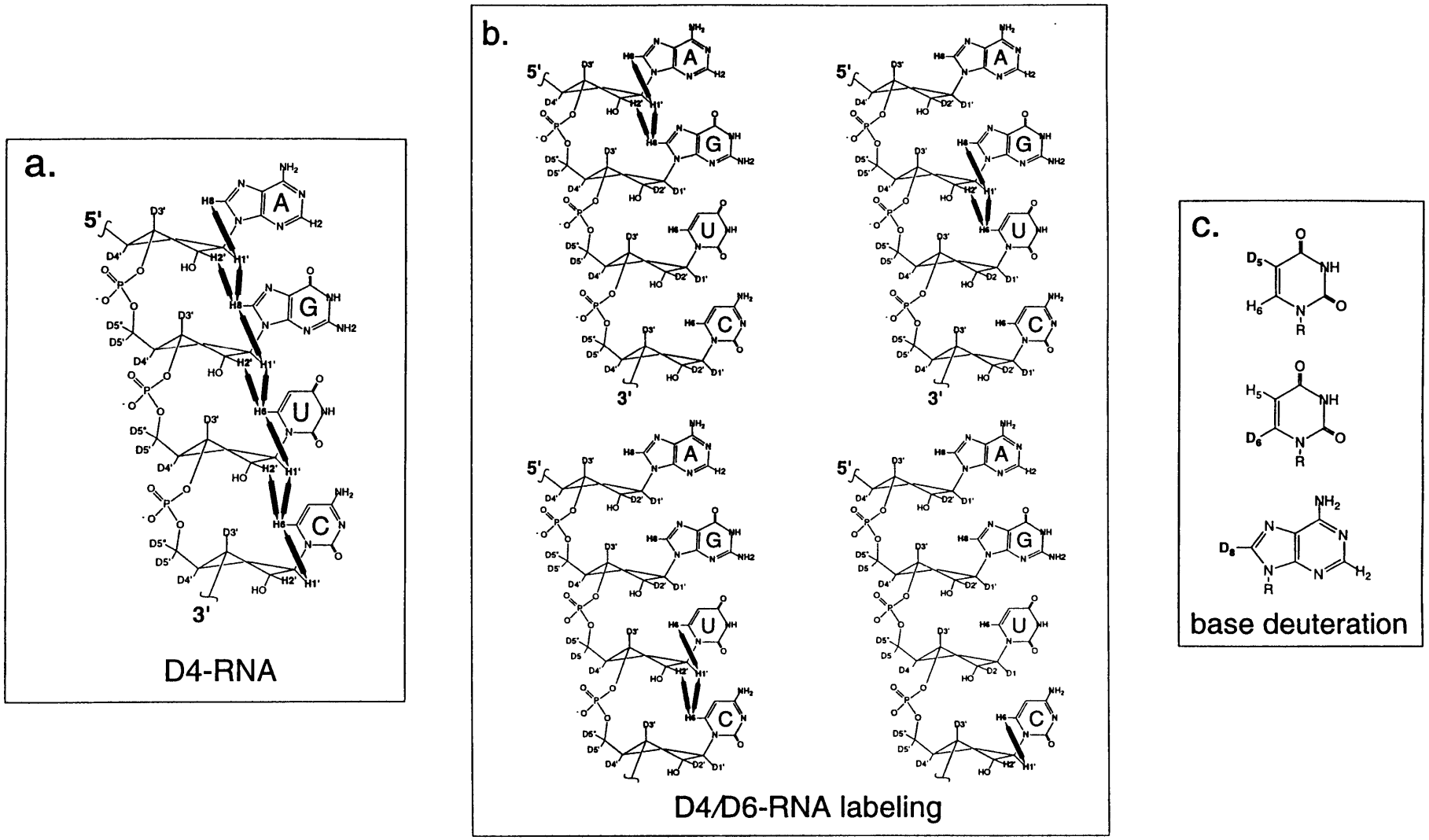


Figure 5.3 Reduction of spectral crowding with specific deuteration. a) Ribose regions of RNA NMR spectra can be simplified by the D4-RNA labeling pattern. b) Additional spectral simplification can be obtained by mixing nucleotide specific D4-RNA labeling with D6-RNA labeling. This should reduce spectral crowding by up to a factor of 4. c) Further spectral simplification can be obtained by using base proton deuteration (H5, H6, or H8). All of these labeling patterns can be synthesized in $^{13}C/^{15}N$ labeled forms.

labeled RNA. Assignment would also be easier because the identification of the type of ribose NOEs within NOESY spectra would be simplified, since there would be only one type of nucleotide with ribose protons within each NMR sample. Further spectral simplification could be obtained by deuterating base protons of the RNA. Deuteration of the H5 of pyrimidines could be used to remove H5-H6 NOESY crosspeaks, and deuteration of the H6 or H8 protons of bases could introduce another level of nucleotide specific spectral simplification. The specific deuteration patterns discussed above could be combined with heteronuclear labeling of the ribose moieties or bases to allow the use of multidimensional heteronuclear NMR spectroscopy to reduce spectral crowding even further if the relaxation properties of the RNA to be studied allowed heteronuclear NMR spectroscopy to be effective.

Example of reduction of spectral crowding in a large RNA with specific deuteration

As an example of the use of specific deuteration to reduce spectral crowding and simplify assignment of the NMR spectra of large RNAs, several deuterated forms of the 65 nucleotide RNA F22b have been synthesized. The ribosomal protein S15 is one of the first proteins to bind to the 16S ribosomal RNA in the process of ribosomal assembly. The interaction of the S15 ribosomal protein derived from *Bacillus stearothermophilus* with a minimized RNA fragment to which it binds, termed here F22b, has been biochemically characterized (Batey & Williamson, 1996a; Batey & Williamson, 1996b), and now attempts are being made to study the structure of the RNA and protein in complex with NMR spectroscopy. The S15 RNA binding site F22b is 65 nucleotides in size, and the *Bacillus stearothermophilus* S15 protein contains 88 amino acids. This makes the F22b-S15 RNA protein complex a very large and challenging system for NMR spectroscopic studies with a size of approximately 30 kDa.

Chapter 5

The interaction of the F22b RNA with the *Bacillus stearothermophilus* S15 protein exhibits several of the problems frequently encountered in large macromolecular NMR spectroscopy. The relaxation rates of the $^{13}C/^{15}N$ labeled complex are so fast that heteronuclear NMR experiments do not work to any useful extent, limiting NMR structural studies to homonuclear NMR spectroscopy for the moment. To help resolve some of the spectral crowding problems of the F22b-S15 NMR spectra on the RNA side of the complex, the F22b RNA was prepared with some of the specific deuterium labeling patterns described in this thesis. At 65 nucleotides the F22b RNA is much larger than the 36 nucleotide ATP-binding aptamer (Dieckmann *et al.*, 1996) or the 39 nucleotide bent helix hairpin interaction (Marino *et al.*, 1995), which are among the largest RNA NMR structures to have been determined so far. Although the imino spectra of some 76 nucleotide tRNAs have been assigned (Hall *et al.*, 1989; Choi & Redfield, 1992), no sequential assignment through the non-exchangeable ribose regions of an RNA of the size of the F22b RNA have been reported. The spectral crowding in the ribose regions of RNA NMR spectra has proven a formidable obstacle to the extension of NMR structural studies to RNAs larger than 40 nucleotides. Homonuclear NMR spectra of the F22b RNA are extremely crowded due to the large number of nucleotides in the RNA, preventing sequential assignment of the RNA in complex or alone.

The F22b RNA was synthesized with the D5-pyrimidine RNA labeling pattern, the D4-RNA labeling pattern, and the D4/D6-RNA labeling pattern that were described in Chapter 4. An unlabeled F22b RNA was also produced by T7 RNA polymerase catalyzed transcription for comparison using commercially obtained unlabeled nucleotides. D5-pyrimidine-F22b RNA was produced using unlabeled ATP, unlabeled GTP, 5-*d1*-UTP (**37**), and 5-*d1*-CTP (**38**) as starting materials (same nucleotides as in Figure 4.18a). *d4*--F22b RNA (D4-F22b) was produced using 3',4',5',5"-*d4*-ATP (**2**), 3',4',5',5"-*d4*-GTP (**3**), 3',4',5',5"-*d4*-UTP (**4**), and 3',4',5',5"-*d4*-CTP (**5**) as starting materials (same nucleotides as in Figure 4.3a). *d4-U/d6-A,G,C*-F22b RNA (D4/D6-F22b) was produced

using 1',2',3',4',5',5"-d_6-ATP (**24**), 1',2',3',4',5',5"-d_6-GTP (**25**), 1',2',3',4',5',5"-d_6-CTP (**27**), and 3',4',5',5"-d_4-UTP (**4**) as starting materials (same nucleotides as in Figure 4.14).

NOESY spectra of the unlabeled F22b (Figure 5.4) and the deuterated forms of F22b (Figures 5.5-5.7) demonstrate the dramatic reduction in spectral crowding afforded by deuteration. Figure 5.5 shows a comparison of the base to H1' NOE regions of unlabeled F22b and D5-pyrimidines F22b. Deuteration of the H5 of pyrimidines in Figure 5.5b results in the removal of the strong H5-H6 pyrimidine crosspeaks from the NOESY spectra. Several H1'-H2' and H2'-base crosspeaks are clearly resolved in the NOESY spectra of D4-F22b RNA, Figure 5.6, which were buried in spectral overlap in the unlabeled F22b NOESY Figure 5.4. Finally, the NOESY spectra of the D4/D6-F22b RNA, Figure 5.7, is extremely simplified compared to the unlabeled F22b RNA NOESY spectra, having only uridine H1' and H2' ribose protons.

Even without the S15 protein there are things which can be seen from the NMR spectra of the deuterated forms of F22b. For instance, the H2' proton that is shifted far downfield to 3.80 ppm can be deduced to be a U H2' from the presence of an H2' NOE at 3.80 ppm in the D4/D6-F22b RNA NMR spectra, Figure 5.7. The H1' to base NOE region of the F22b RNA is still quite crowded, even after the removal of the H5-H6 crosspeaks, Figure 5.5b, but this overlap of H1' to base NOEs should spread out after titration of the S15 protein. Examination of the D4/D6-F22b RNA NMR spectra allows identification of all eleven U-H1' and U-H2' chemical shifts. In addition, NOE contacts between the eleven U-H2' protons and non-exchangeable base protons are easily extracted from the D4/D6-F22b RNA NOESY spectra. Tables 5.1 and 5.2 summarize the chemical shift information described above which was extracted from the D4/D6-F22b RNA NOESY spectra.

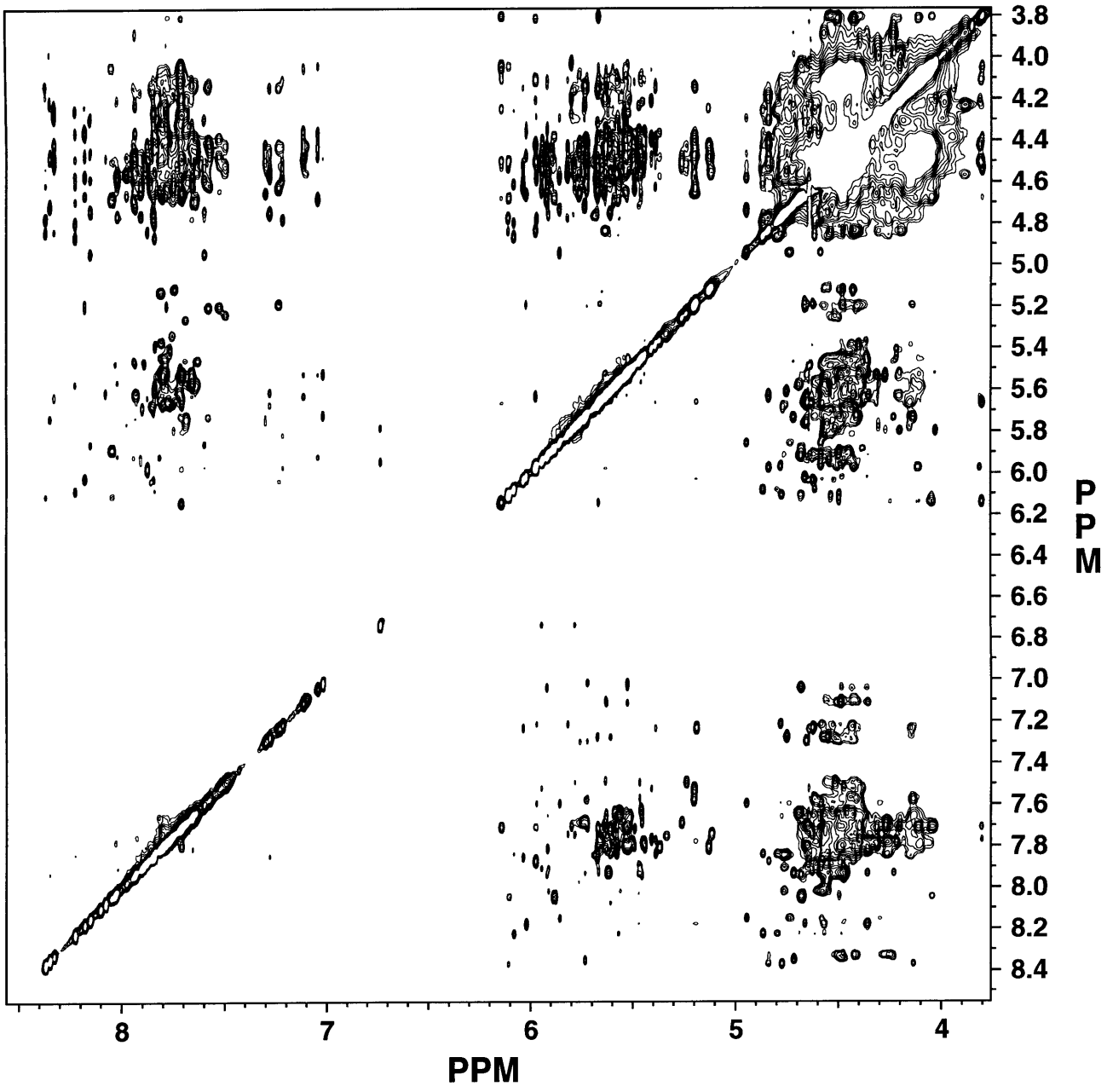


Figure 5.4 Unlabeled F22b 200 ms NOESY spectra

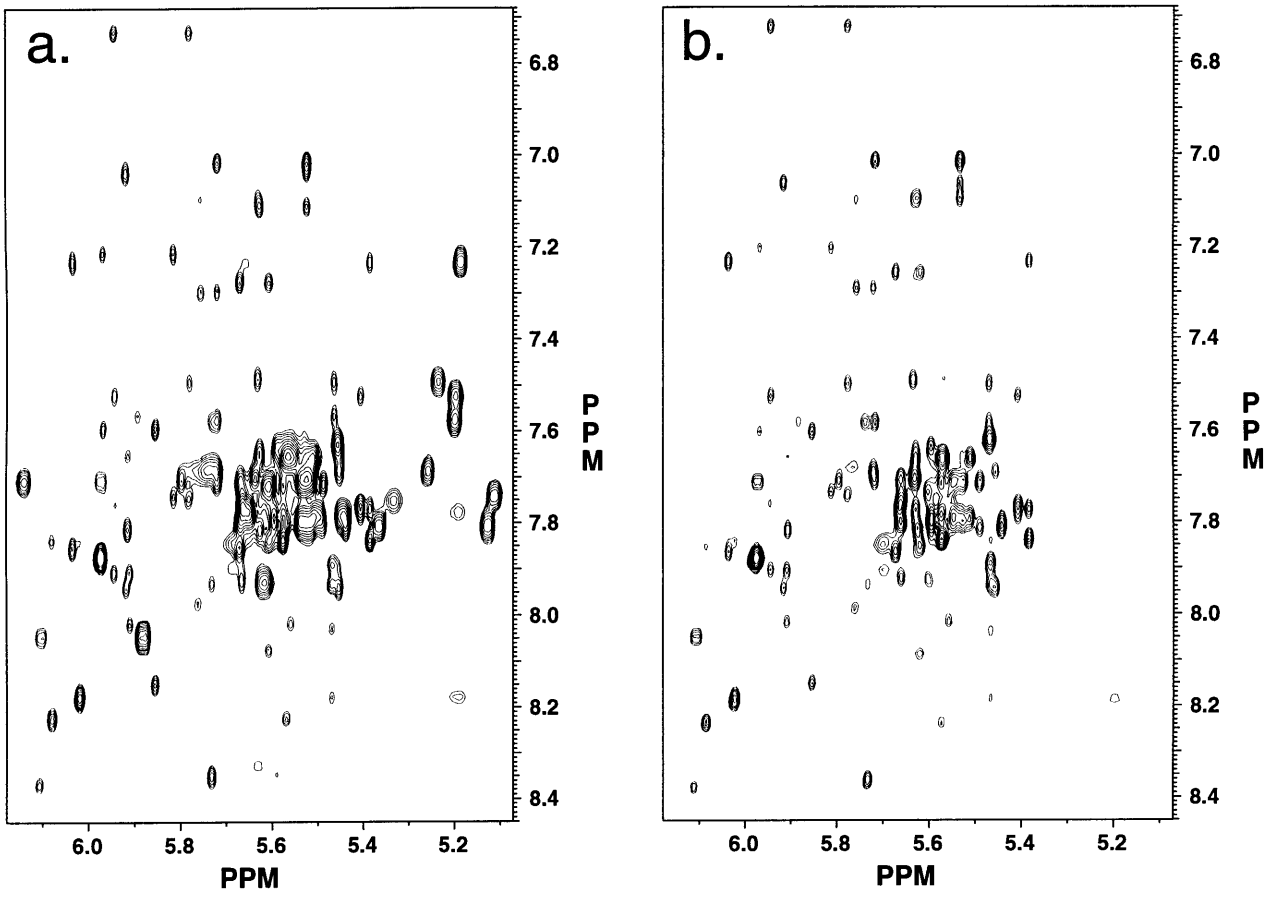


Figure 5.5 Base to H1', H5 NOE region of a 200 ms NOESY of a) unlabeled F22b RNA b) D5-pyrimidines F22b RNA. The H5-H6 crosspeaks of the D5-pyrimidine RNA are removed from the NOESY spectra, but all of the base to H1' NOEs are still present.

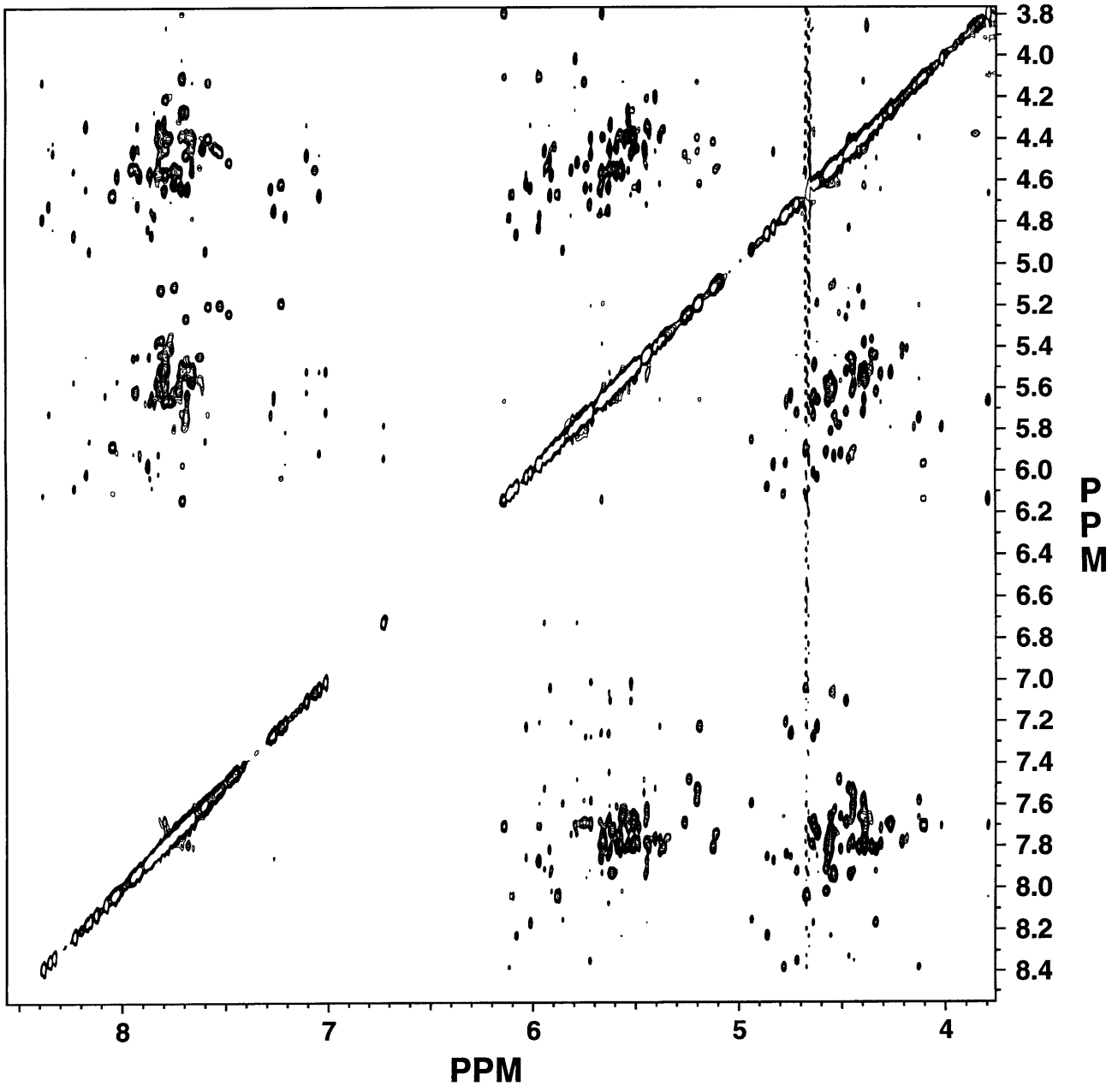


Figure 5.6 D4-F22b 200 ms NOESY spectra

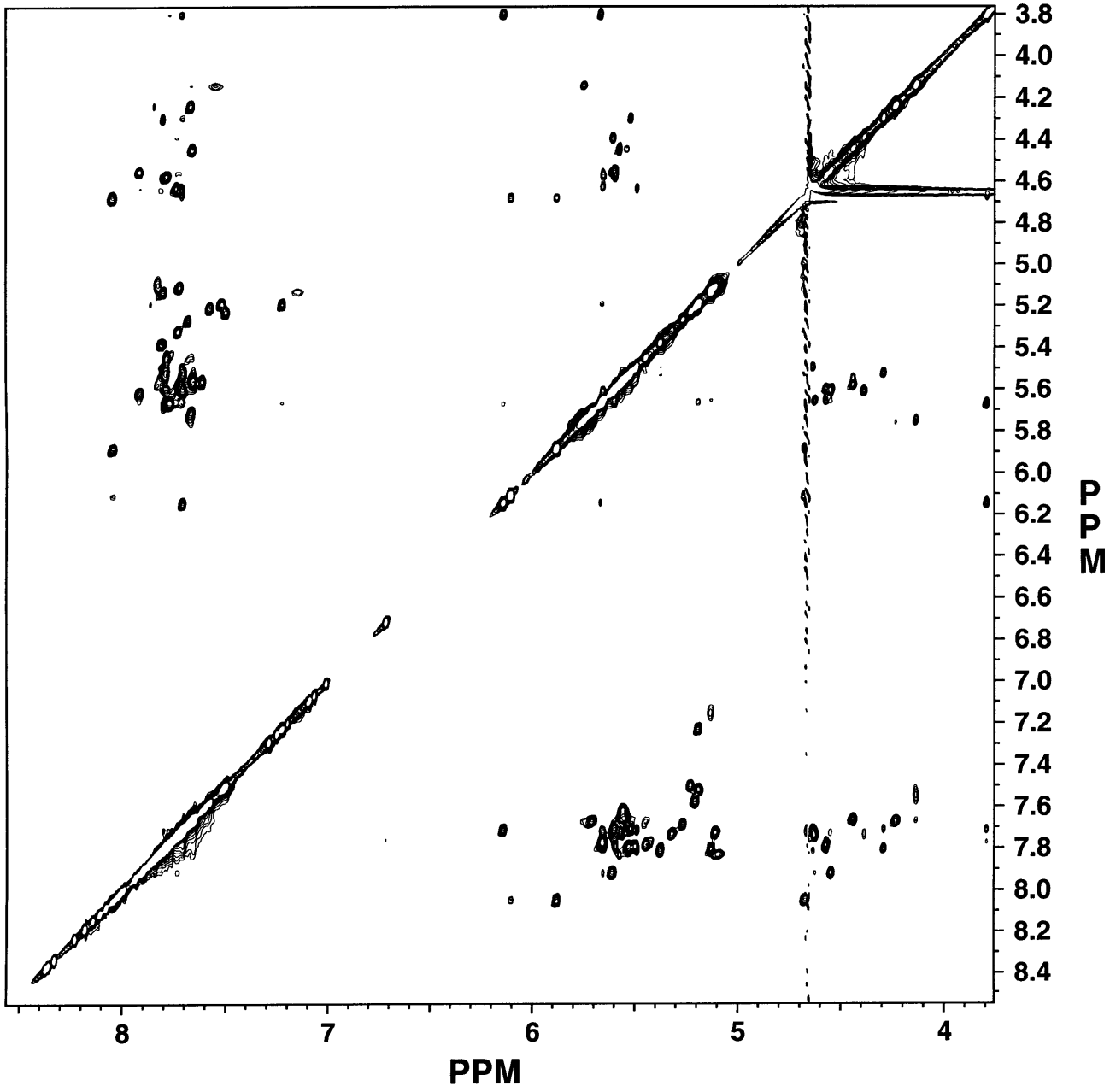


Figure 5.7 D4-U, D6-A,G,C F22b 200 ms NOESY spectra

Chapter 5

Table 5.1 Chemical Shifts of H1's and H2's of Uridine Nucleotides in F22b

U nucleotide	UH1' (ppm)	UH2' (ppm)
1	5.68	3.80
2	5.74	4.14
3	5.76	4.24
4	5.52	4.29
5	5.61	4.39
6	5.57	4.44
7	5.60	4.55
8	5.60	4.57
9	5.65	4.63
10	5.49	4.64
11	6.11	4.68

Table 5.2 Chemical Shifts of UH2' to Base Proton NOEs in F22b

U nucleotide	UH2' (ppm)	Base proton Strong NOE (ppm)	Base proton Weak NOE (ppm)
1	3.80	7.72	
2	4.14	7.55	7.67
3	4.24	7.68	
4	4.29	7.81	7.72
5	4.39	7.73	
6	4.44	7.67	
7	4.55	7.93	7.73
8	4.57	7.79	
9	4.63	7.75	
10	4.64	7.72	
11	4.68	8.06	

The information in Tables 5.1 and 5.2 establishes a starting point for the sequential assignment of F22b. With this information, all of the chemical shifts of the UH1's and UH2's within the F22b RNA have been identified. In addition, through space connectivity between UH2' protons and non-exchangeable base protons of yet unknown identity have been established. Extension of the D4/D6-RNA labeling strategy to nucleotides other than uridine would yield similar information for the other nucleotides and build a more complete picture of the sequential assignment of the F22b RNA.

Chapter 5

Specific deuteration and heteronuclear ($^{13}C/^{15}N$) labeling combined in RNA NMR spectroscopy

RNA heteronuclear NMR spectra can also be improved with the labeling patterns and strategies discussed in this thesis. Spectral overlap in heteronuclear NMR spectra can be removed by using specific deuteration in combination with heteronuclear labeling. An example of this has already been discussed in Chapter 4 where the chemical shift overlap of the H3'-C3' and H2'-C2' crosspeaks in an RNA HSQC spectra was removed with the *D4-$^{13}C_5$-ribose*-RNA labeling pattern. In addition to removing spectral overlap, the combination of specific deuteration with heteronuclear labeling can be used to extend the useful range of heteronuclear NMR experiments to larger RNA molecules.

Deuteration greatly increases the T_2's of heteronuclei by replacing the strong dipolar coupling between ^{13}C and hydrogen with the weaker ^{13}C-deuterium interaction. The increase in the lifetime of carbon magnetization increases the intensity of signal in NMR experiments which transfer magnetization through heteronuclei. In this way deuteration can be used to counteract the rapid relaxation of larger macromolecules, and thus increase the size of macromolecules where acceptable signal to noise ratios can be achieved in heteronuclear NMR experiments. The combination of deuteration with heteronuclear labeling in protein NMR spectroscopy has been very successful in extending the size of proteins where heteronuclear NMR spectroscopy can be utilized. $^{13}C/^{15}N$ labeled proteins which are also random fractionally deuterated to a high level have been applied successfully to the study proteins on the order of 40-60 kDa (Kay & Gardner, 1997). With such dramatic progress in the protein NMR field it is likely that some gains can be made by combining heteronuclear labeling with specific deuteration in RNA NMR spectroscopy.

One difference between protein NMR spectroscopy and RNA NMR spectroscopy is that protein NMR spectroscopy benefits from having an exchangeable amide proton central

to sequential assignment strategies. Protein NMR spectroscopists are able to deuterate proteins to high levels, but then conduct NMR experiments in H_2O to exchange the amide proton for ^1H. This allows protein NMR spectroscopist to benefit from narrow heteronuclear linewidths due to high deuteration levels, but also have good signal to noise ratios because the amide protons are present as 90% ^1H. Unfortunately RNA lacks an exchangeable proton which is strategically located to help with nonexchangeable proton sequential assignment in highly deuterated, heteronuclear labeled RNAs. Although the exchangeable imino protons are extremely important for the sequential assignment of RNA exchangeable proton NMR spectra, they are relatively isolated and do not help very much with RNA non-exchangeable sequential assignment. The synthetic techniques described in this thesis can be used to specifically place protons into highly deuterated, ^{13}C labeled ribose moieties in order to reap the relaxation benefits of high deuteration levels while still having protons present as ^1H to initiate and detect magnetization on during NMR experiments.

Example of specific deuteration in combination with heteronuclear labeling.

As a demonstration of the benefits of specific deuteration of RNA in combination with heteronuclear labeling, the C5' to H1' crosspeaks of an CCH-TOCSY NMR spectra (CCH-total correlation spectroscopy) (Dayie *et al.*, 1997) can be examined for the two ^{13}C labeled TAR RNAs produced in Chapter 4. The CCH-TOCSY NMR experiment was developed by Dr. Kwaku Dayie to utilize the benefits of the *D4-$^{13}C_5$-ribose*-RNA labeling pattern to improve RNA sequential assignment. The CCH-TOCSY NMR experiment takes advantage of the deuteration of the C3', C4', and C5' carbons in the *D4-$^{13}C_5$-ribose*-RNA labeling pattern by initiating magnetization on ^{13}C rather than ^1H. Starting magnetization on ^{13}C rather than ^1H has two effects; the length of the pulse sequence is reduced as compared to HCCH-TOCSY NMR experiments (which has signal to noise benefits in large RNA molecules), and the chemical shifts of fully deuterated carbons can be correlated to

the remaining ribose protons. The ability to observe magnetization from fully deuterated carbons in the CCH-TOCSY NMR experiment allows the observation of increased carbon magnetization lifetime due to deuteration.

The CCH-TOCSY experiment starts transverse magnetization on carbon, transfers magnetization through the ribose ring, and then transfers carbon magnetization to protons for detection. In a two dimensional CCH-TOCSY experiment, each ribose proton will have crosspeaks correlating it to every carbon within the ribose. For the C5' carbon, magnetization starts on the C5', and it is transferred through the C4', C3', C2', and C1', then C1' magnetization is transferred to the H1' and detected to produce a crosspeak correlating the C5' carbon to the H1' proton. This transfer of magnetization is schematized for the C5' in Figure 5.8. In Figure 5.9 the CCH-TOCSY spectra of *$^{13}C_5$-ribose*-labeled RNA and *D4-$^{13}C_5$-ribose*-RNA are shown. In the CCH-TOCSY of the non-deuterated *$^{13}C_5$-ribose*-RNA, the C5'-H1' crosspeaks are difficult to detect in large RNAs because of the rapid relaxation of heteronuclei. Deuteration of the C5', C4', and C3' in *D4-$^{13}C_5$-ribose*-RNA dramatically increases the intensity of C5'-H1' crosspeaks in CCH-TOCSY NMR spectra due to longer relaxation rates of the deuterated heteronuclei. The increased efficiency of magnetization transfer due to deuteration will allow heteronuclear NMR experiments like CCH-TOCSY to work on larger RNA molecules than otherwise possible.

Conclusion

The isotopic labeling patterns and strategies described in this thesis may aid in the assignment of large RNA NMR spectra. Already the specifically deuterated forms of the F22b RNA that were discussed in this chapter are simplifying the assignment of the RNA. The previously unidentified chemical shifts of the H1' and H2' protons of all 11 uridine nucleotides in F22b are now known. Our combination of specific deuteration with ^{13}C and ^{15}N labeling also shows promise in being useful in RNA NMR spectroscopy. There are several interesting possibilities, like the CCH-TOCSY experiments, to investigate that

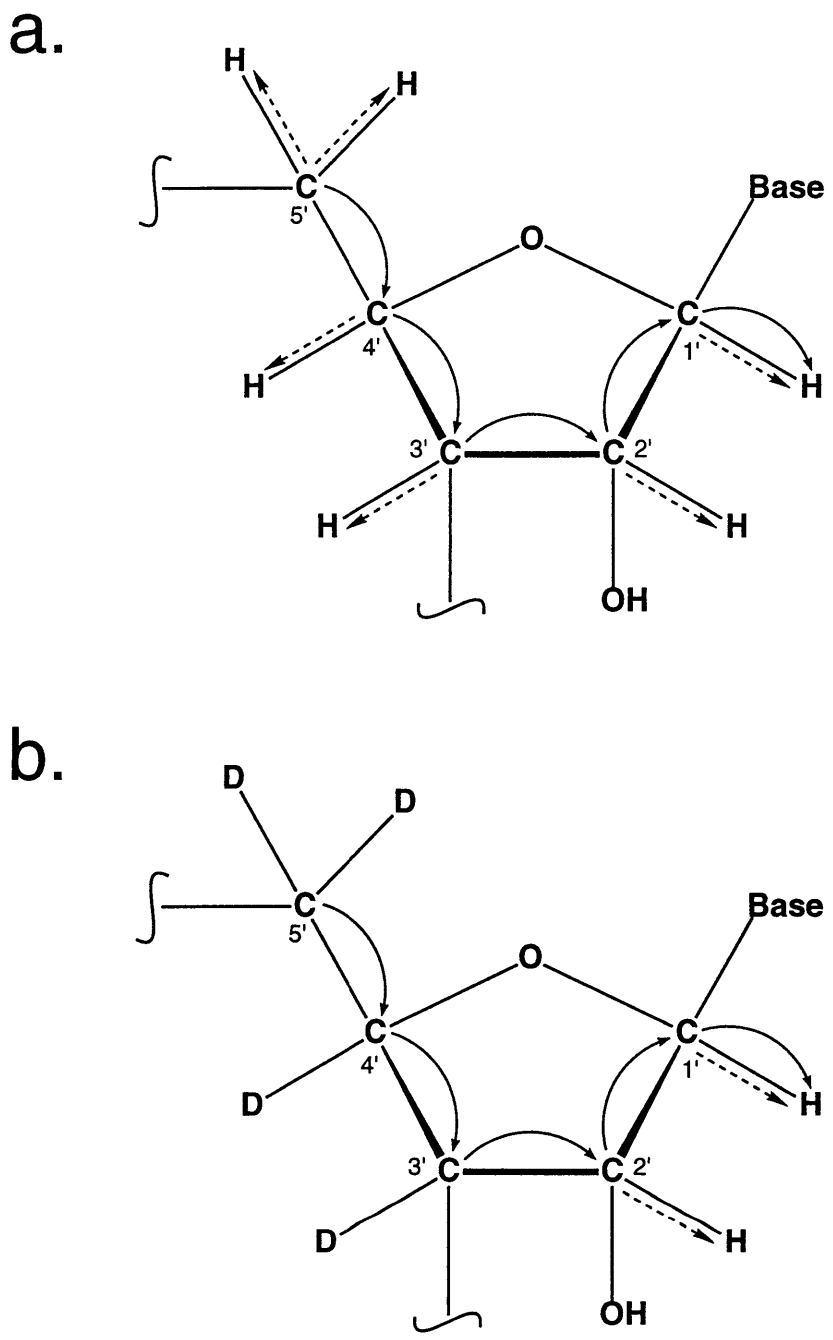


Figure 5.8 Transfer of magnetization from the C5' to the H1' using CCH-TOCSY for a) *$^{13}C_5$-ribose*-RNA b) 3',5',5',5"-d_4-$^{13}C_5$-ribose RNA . Curved solid arrows represent magnetization transfer in the CCH-TOCSY NMR experiment. Dashed arrows represent strong dipolar interactions between ^{13}C and ^{1}H which cause rapid relaxation of ^{13}C nuclei. Crosspeaks correlating the C5' with the H1' in figure 5.9 are much stronger for the *D4-$^{13}C_5$-ribose*-RNA because the relaxation rates of the C5', C4', and C3' are slower due to removal of the strong ^{13}C-^{1}H interactions.

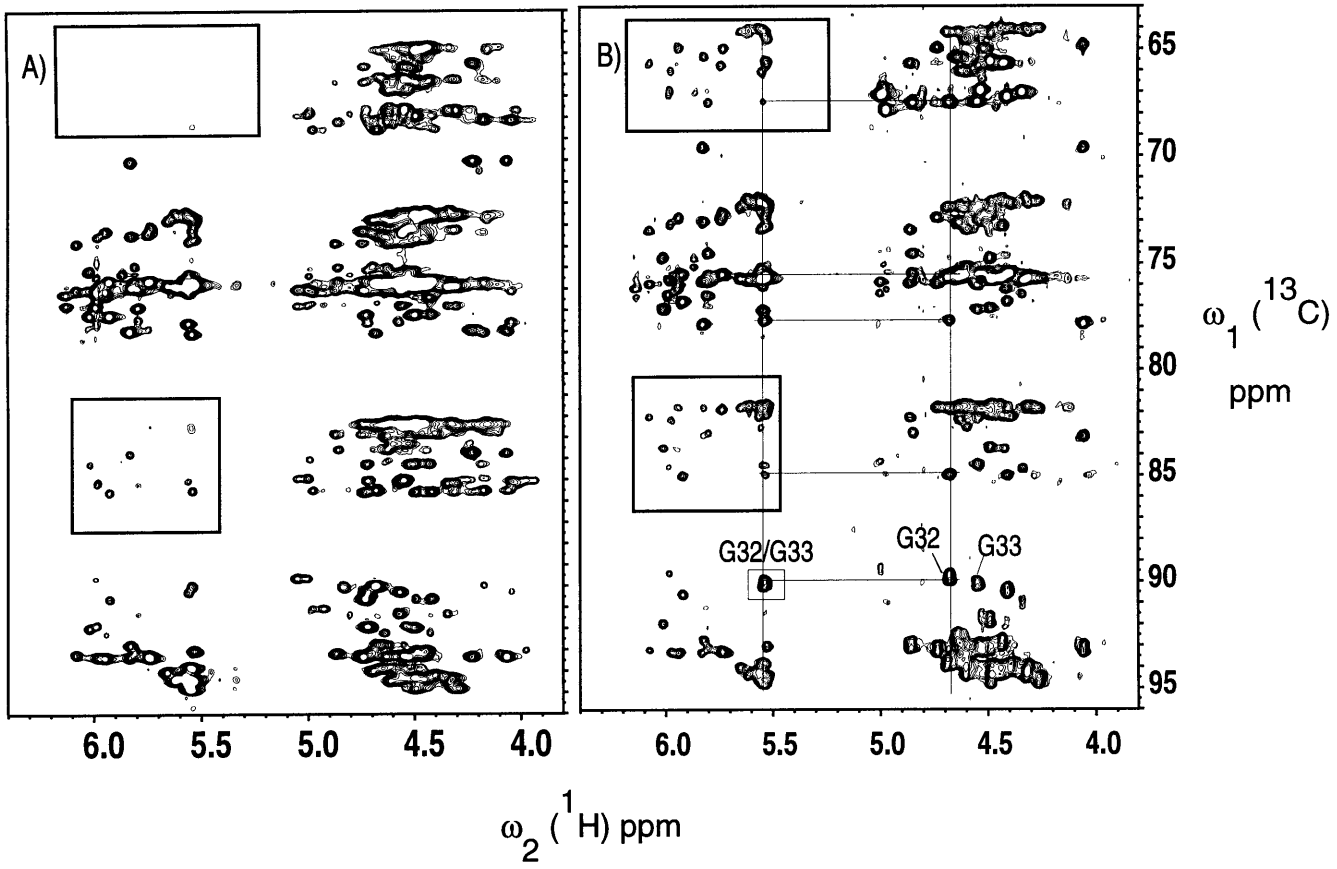


Figure 5.9 A 2D CCH-TOCSY spectra showing the efficiency of magnetization transfer in (A) *$^{13}C_5$-ribose*-TAR RNA and (B) *D4-$^{13}C_5$-ribose*-TAR RNA. The C5'-H1' crosspeaks are boxed in the upper left corner of the spectra. A parallel ladder of carbon resonances at the chemical shifts of the H1' and H2' are shown for G32.

may allow the through bond sequential assignment of large RNAs where rapid relaxation prevented through bond assignment before. Spectral crowding might be further reduced in RNA NMR spectra by combining the labeling strategies described in this thesis with "NMR window" (Yamakage *et al.*, 1993) or "segmental isotopic labeling" (Xu *et al.*, 1996) strategies. These strategies use chemical RNA synthesis or ligation of T7 transcripts to isotopically label specific sections of large RNAs in order to focus NMR studies on those sections. The use of glucose as a starting material for incorporating isotopic labels into RNA shows promise in reducing the cost of isotopically labeled RNA greatly. Fully $^{13}C/^{15}N$ labeled forms of UTP and CTP can be prepared from glucose and isotopically labeled uracil at a fraction of the cost required to produce the same nucleotides from bacteria grown on isotopically labeled media. This might make chemical synthesis of isotopically labeled RNA more feasible, since the starting materials (isotopically labeled nucleotides) for synthesizing phosphoramidites would be much less expensive, causing losses during the conversion of nucleotides into phosphoramidites to be less costly. The use of enzymatic synthesis to convert ribose and glucose into isotopically labeled nucleotides which would be useful in NMR spectroscopy was very successful, and it is possible that similar chemical and enzymatic techniques could be developed for the isotopic labeling of other biological molecules. Efficient methods for isotopically labeling DNA and oligosaccharides which would be of use in NMR spectroscopy could be developed using a combination of chemical and enzymatic synthesis. Application of the isotopic labeling patterns and synthetic strategies described in this thesis to NMR studies of RNA should prove very valuable.

Chapter 6

6. Experimental

General Procedures.

Materials and Methods. Reactions were conducted under argon atmosphere unless otherwise noted, and solvents were dried and distilled (Perrin & Armarego, 1988). Chemicals were purchased from Aldrich and Sigma. Glycerol (U-d8), 1,2,3,4,5,6-$^{13}C_6$-D-glucose (>99% ^{13}C), D_2O, and $CDCl_3$ were purchased from Cambridge Isotopes Laboratories. 1,2,3,4,5,6,6-$^{2}H_7$-D-glucose (>97% ^{2}H) and 1,2,3,4,5,6,6-$^{2}H_7$-1,2,3,4,5,6-$^{13}C_6$-D-glucose (>97% ^{2}H, >98% ^{13}C) were purchased from Martek Corporation. Nucleoside monophosphate kinase from beef liver (catalog number 107 948, 1 unit per mg protein) was purchased from Boehringer Mannheim. Phosphoglycerate mutase from rabbit muscle (catalog number P-9885, 470 units per mg protein), enolase from baker's yeast (catalog number E-6126, 73 units per mg protein), myokinase from chicken muscle (catalog number M-5520, 1500-3000 units per mg protein), pyruvate kinase type III from rabbit muscle (catalog number P-9136, 350-600 units per mg protein), guanylate kinase from porcine brain (catalog number G-9385, 20-60 units per mg protein), hexokinase type V from bakers yeast (catalog number H-5250, 40-60 units per mg protein), phosphoglucose isomerase type III from bakers yeast (catalog number P-5381, 500-800 units per mg protein), glucose-6-phosphate dehydrogenase type V from bakers yeast (catalog number G-7750, 230 units per mg protein), 6-phosphogluconic dehydrogenase from yeast (catalog number P-4553, 3-6 units per mg protein), phosphoriboisomerase from torula yeast (catalog number P-7434, 200-600 units per mg protein), and L-glutamic dehydrogenase type III from bovine liver (catalog number G-7882, 40 units per mg protein) were purchased from Sigma. One unit of enzyme corresponds to 1 μmole/min of activity under assay conditions. The pH of aqueous solutions was determined by colorpHast Indicator Strips pH 0-14 made by EM Science.

Flash chromatography was conducted on either a medium column (3.5 cm diameter, 10-15 cm height) or a small column (2 cm diameter, 6-12 cm height), depending on the amount of material to be purified, using ICN silica 32-63, 60A obtained from ICN Biomedicals. NMR spectra were recorded on a Varian XL-300, Varian VXR-500, or a Varian Inova-600 MHz spectrometer. Infrared spectra were recorded on a Perkin Elmer 1600 series FT-IR, and ultraviolet spectra were recorded on a Hitachi U-2000 UV/Vis spectrophotometer. Mass spectra were obtained with a Hewlett Packard 5890/5971 GC/mass spectrometer for low resolution, a Finnigan MAT 8200 for high resolution, or on a Hewlett Packard electrospray mass spectrometer for NTPs.

Preparation of sodium 3-phosphoglycerate from barium 3-phosphoglycerate. The sodium salt of 3-phosphoglycerate was prepared from the less expensive barium form of 3-phosphoglycerate by exchanging the barium for sodium with amberlite IR-120PLUS strongly acidic cation exchanger, sodium form (16-50 mesh). The cation exchanger (60 g) was first washed thoroughly with water, and then stirred with 5g of barium 3-phosphoglycerate and 75 mL of H_2O. After 2-1/2 hours the majority of the barium 3-phosphoglycerate, visible in the solution as an insoluble white precipitate, had dissolved. The mixture was filtered and the pH of the filtrate was adjusted to around 7 with NaOH. This resulted in approximately 80 mL of a solution that was 165 mM sodium 3-phosphoglycerate.

Boronate affinity purification of nucleotides. All of the nucleotides from the enzymatic reactions were purified by boronate affinity chromatography (Affigel 601 boronate affinity resin obtained from Biorad), which removes the majority of proteins and salts used in the enzymatic reactions and prepares the nucleotides for transcription. The nucleotide forming reactions were concentrated under vacuum and then dissolved in a minimum amount of 1M triethylammonium bicarbonate (TEABC) solution pH = 9.5-10 (which is prepared by bubbling carbon dioxide into a solution of 140 mL triethylamine and

860 mL of H_2O.). Once the residue was dissolved, the solution was allowed to rest at room temperature for 15-30 minutes while a white precipitate usually formed (probably precipitated protein), then the precipitate was removed from the solution by filtration or centrifugation. The filtrate was loaded onto an Affigel 601 boronate affinity column (containing 10 g of Affigel 601) that had been pre-equilibrated with 1M TEABC pH = 9.5-10. After loading, the column was washed with 3-5 column volumes of 1M TEABC pH = 9.5-10, then the nucleotides were eluted from the column with water that had been acidified by bubbling carbon dioxide through it (pH ≈ 4-5.5). All of the acidified H_2O elution was collected in a single flask, and concentrated under vacuum. Triethylammonium bicarbonate was removed from the residue by repeatedly dissolving the residue in ethanol and removing the ethanol under vacuum. This procedure does not separate nucleotides from one another, but does remove the majority of contaminants which interfere with transcription reactions. Nucleotides prepared in this manner a ready for transcription without any further purification steps. The boronate resin can be used repeatedly, and is stable when stored in H_2O at 4° C.

Chapter 2 Experimental

Chemical synthesis of D,L-3',4',5',5"-*d*4-ribose from perdeuterated glycerol (Figure 6.1)

D,L-2:3-O-cyclohexylidene-1,1',2,3,3'-*d*5-glycerol (11).

Cyclohexanone (3.49 g, 35.6 mmol) and trimethylorthoformate (3.77 g, 35.6 mmol) were dissolved in CH_2Cl_2 (150 mL). Glycerol (U-*d*8, 98%) (**10**) (3.56 g, 35.6 mmol) and a catalytic amount of p-toluenesulfonic acid (10 mg, 50 μmol) were added, and the heterogeneous mixture was stirred at room temperature. After stirring for 9 hours the solution was homogeneous and the reaction was complete as judged by TLC and the

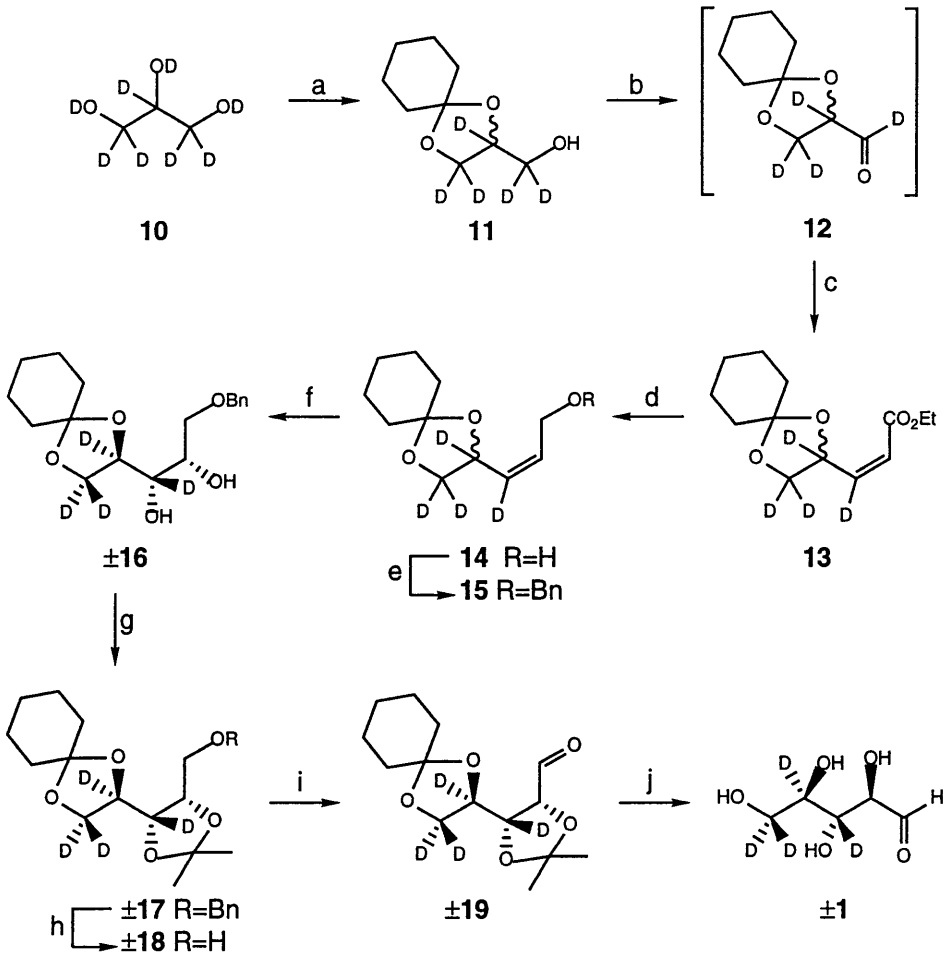


Figure 6.1 Synthesis of D,L-3',4',5',5"-$^2H_4$-ribose (±**1**).
(a) cyclohexanone, $(MeO)_3CH$, H^+ (b) oxalyl chloride, DMSO, Et_3N
(c) $Ph_3P{=}CHCO_2Et$, MeOH (d) DIBAL-H, CH_2Cl_2 (e) BnBr, NaH,
$(nBu)_4NI$, THF (f) OsO_4, N-methyl morpholine N-oxide, acetone:H_2O (8:1)
(g) 2-methoxy propene, H^+ (h) Pd on C, H_2 (i) oxalyl chloride, DMSO,
Et_3N (j) H^+, THF, H_2O

disappearance of the immiscible glycerol (U-d_8). The reaction mixture was concentrated and the product purified by flash chromatography (5 to 40% EtOAc in hexane) to yield 6.07 g (34.1 mmol, 95%) of **11** as colorless oil. ^{1}H NMR (300 MHz, $CDCl_3$) δ 2.86 (s br, 1H), 1.85-1.28 (m, 10 H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 109.7, 75.0 (t, J = 23 Hz), 64.6 (quintet, J = 23 Hz), 62.2 (quintet, J = 21 Hz), 36.2, 34.6, 24.9, 23.8, 23.6; IR (neat) 3442, 2944, 2860, 2215, 2106, 1449, 1367, 1333, 1288, 1257, 1231, 1169, 1105, 1063, 1011, 974, 941, 907; HRMS m/z 177.1409 (177.1408 calcd for $C_9H_{11}O_3{}^2H_5$, M^+)

D,L-2:3-O-cyclohexylidene-1,2,3,3'-*d*4-glyceraldehyde (12). Alcohol **11** was dissolved in CH_3OD and evaporated two times to exchange the hydroxyl proton, to remove protons and prevent ^{1}H exchange that can occur with the aldehyde under basic conditions. Oxalyl chloride (2.18 g, 17.2 mmol) was added to CH_2Cl_2 (200 mL) in a flask cooled to - 78° C, and DMSO (2.68 g, 34.4 mmol) was added slowly via syringe. The resulting mixture was stirred for five minutes before **11** (2.78 g, 15.6 mmol) was added. The mixture was stirred for 30 min. at -78° C, and then triethylamine (6.32 g, 62.5 mmol) was added slowly. After stirring for 15 minutes at -78° C, the dry ice/acetone bath was removed and the reaction was stirred at room temperature. After 30 min., the reaction was quenched with 75 mL of 0.55M HCl. The organic layer was separated and washed with saturated $NaHCO_3$ (75 mL) and saturated NaCl (75 mL). The organic layer was dried over $MgSO_4$ and then concentrated. The crude aldehyde **12** was carried on directly to the next reaction without purification.

Ethyl-(Z)-D,L-4:5-O-cyclohexylidene-3,4,5,5'-*d*4-2-pentenoate (13). The crude aldehyde **12** was dissolved in 100 mL of methanol and cooled to 4° C. $Ph_3P{=}CHCO_2Et$, (carbethoxymethylene)triphenylphosphorane obtained from Aldrich, was added, and the reaction was stirred for 24 hours at 4° C. After concentration *in vacuo*

the residue was purified by flash chromatography (5% Et_2O in petroleum ether) to yield 2.91 g (11.9 mmol, 76%) of a light yellow oil. ^1H NMR (300 MHz, $CDCl_3$) δ 5.83 (s, 1 H), 4.17 (q, 2 H, J = 7.1 Hz), 1.75-1.33 (m, 10 H), 1.29 (t, 3 H, J = 7.1 Hz); ^{13}C NMR (75 MHz, $CDCl_3$) δ 165.5, 149.0 (t, J = 23 Hz), 120.4, 110.1, 72.5 (t, J = 24 Hz), 68.2 (quintet, J = 23 Hz), 60.2, 36.1, 34.8, 25.0, 23.8, 23.7, 14.0; IR (neat) 2981, 2935, 2862, 2229, 2174, 2111, 1716, 1634, 1448, 1372, 1349, 1272, 1198, 1164, 1117, 1036; HRMS m/z 244.1610 (244.1608 calcd for $C_{13}H_{16}O_4{}^2H_4$, M^+)

(Z)-D,L-4:5-O-cyclohexylidene-3,4,5,5'-*d*4-2-penten-1-ol (14). Z ester **13** (4.21 g, 17.2 mmol) was added to 200 mL of CH_2Cl_2, and was cooled to O° C. Diisobutyl aluminum hydride (DIBAL-H) obtained from Aldrich (1 M in hexanes, 43.1 mmol) was slowly added with a syringe over a period of ten minutes. The mixture was stirred at 0-4° C for 30 minutes, then the reaction was quenched with 100 mL of 1 M NaOH. The sodium hydroxide solution was added dropwise to prevent explosive evolution of hydrogen. The quenched reaction was stirred for 8 hours at room temperature to hydrolyze the aluminum adducts, then the aqueous and organic layers were separated. The aqueous layer was extracted with 3x100 mL of EtOAc. The organic layers were combined and dried over $MgSO_4$, concentrated, and purified by flash chromatography (15 to 30% EtOAc in hexane), affording 3.24 g (16.0 mmol, 93%) of a light yellow oil (**14**). ^1H NMR (300 MHz, $CDCl_3$) δ 5.81 (t, 1 H, J = 6.5 Hz), 4.28 (dd, 1 H, J = 13.2, 6.9 Hz), 4.15 (dd, 1 H, J = 13.2, 6.1 Hz), 2.95 (s br, 1 H), 1.75-1.30 (m, 10 H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 132.9, 129.0 (t, J = 24 Hz), 109.9, 70.8 (t, J = 22 Hz), 68.2 (quintet, J = 24 Hz), 58.2, 36.2, 35.3, 24.9, 23.8, 23.7; IR (neat) 3416, 2934, 2861, 2226, 2175, 2108, 1448, 1366, 1338, 1285, 1234, 1166, 1116, 1081, 1035, 978, 928; HRMS m/z 202.1503 (202.1503 calcd for $C_{11}H_{14}O_3{}^2H_4$, M^+)

(Z)-D,L-1-benzyloxy-4:5-O-cyclohexylidene-3,4,5,5'-*d*4-2-pentene (15). Sodium hydride, 60% dispersion in mineral oil, (2.56 g, 64.1 mmol) was washed

with hexane to remove the mineral oil. THF (100 mL) was added to the sodium hydride. Allylic deuterated alcohol **14** (3.24 g, 16.0 mmol), benzyl bromide (3.01 g, 17.6 mmol), and tetrabutylammonium iodide (100 mg, 0.3 mmol) were added to this mixture. The reaction was stirred for 24 hours at room temperature, and was quenched by addition of saturated NaCl (75 mL). Water was added to dissolve precipitated salts, and the aqueous layer was separated. The aqueous layer was extracted with three 50 mL portions of EtOAc, then the combined organic layers were dried over $MgSO_4$ and concentrated. Purification by flash chromatography (5-15% Et_2O in petroleum ether) yielded 4.49 g of the yellow oil (**15**) (15.4 mmol, 96%). ^1H NMR (300 MHz, $CDCl_3$) δ 7.41-7.20 (m, 5 H), 5.80 (t, 1 H, J = 6.4 Hz), 4.53 (d, 1 H, J = 11.8 Hz), 4.48 (d, 1 H, J = 11.8 Hz), 4.11 (d, 2 H, J = 6.4 Hz), 1.75-1.33 (m, 10 H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 137.9, 130.5 (t, J = 24 Hz), 129.8, 128.1 (2 C), 127.5 (2 C), 127.4, 109.6, 71.9, 70.9 (t, J = 23 Hz), 68.1 (quintet, J = 22 Hz), 65.4, 36.1, 35.2, 24.9, 23.7, 23.6; IR (neat) 2934, 2860, 2226, 2174, 2107, 1451, 1365, 1166, 1086, 1028, 1011, 738, 698; HRMS m/z 292.1970 (292.1972 calcd for $C_{18}H_{20}O_3{}^2H_4$, M^+)

D,L-1-benzyloxy-4:5-O-cyclohexylidene-3,4,5,5'-*d*4-ribitol (±11). Benzyl deuterated allyl alcohol **15** (1.71 g, 5.85 mmol), N-methyl morpholine N-oxide (1.37 g, 11.7 mmol), and osmium tetroxide 2.5% in tBuOH (3.05 g, 0.3 mmol OsO_4) were combined in a solution of 1:8 D_2O/acetone (72 mL), and stirred for 12 hours at room temperature. The reaction was quenched by addition of 40 mL of 30% Na_2SO_3. After stirring for 1 hour, the solution was extracted with four 75 mL portions of EtOAc. The combined organic layers were dried over Na_2SO_4 and concentrated. Separation of diastereomers was accomplished by careful flash chromatography (15% EtOAc in hexane). The slower running diastereomer corresponds to the ribitol configuration (**±16**), 1.21 g (3.7 mmol, 63%) of the yellow oil **±16** was separated from the faster eluting diastereomer. ^1H NMR (300 MHz, $CDCl_3$) δ 7.41-7.22 (m, 5 H), 4.56 (d, 1 H, J = 11.9 Hz), 4.51 (d,

1 H, J = 11.9 Hz), 3.85 (t, 1 H, J = 4.9 Hz), 3.68 (d, 2 H, J = 4.9 Hz), 3.31 (s br, 1 H), 3.10 (s br, 1 H), 1.70-1.30 (m, 10 H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 137.5, 128.3 (2 C), 127.8, 127.7 (2 C), 109.7, 75.2 (t, J = 23 Hz), 73.5, 72.3 (t, J = 22 Hz), 71.5, 70.9, 65.2 (quintet, J = 23 Hz), 36.1, 34.6, 25.0, 23.9, 23.6; IR (neat) 3442, 2934, 2861, 2153, 1452, 1367, 1286, 1106, 1008, 909, 738, 699; HRMS m/z 326.2028 (326.2027 calcd for $C_{18}H_{22}O_5{}^2H_4$, M^+)

D,L-1-benzyloxy-4:5-O-cyclohexylidene-3,4,5,5'-*d*4-2:3-O-isopropylidene-ribitol (±17). Diol ±**16** (2.03 g, 6.2 mmol), 2-methoxy propene (0.54 g, 7.5 mmol), and CH_2Cl_2 (100 mL) were combined, and a single anhydrous crystal of p-toluene sulfonic acid was added. The reaction was stirred for 1 hour and then concentrated. Purification by flash chromatography (10-20% Et_2O in petroleum ether) yielded 2.11 g of the yellow oil ±**17** (5.8 mmol, 93%). ^1H NMR (300 MHz, $CDCl_3$) δ 7.46-7.22 (m, 5 H), 4.67 (d, 1 H, J = 12.1 Hz), 4.55 (d, 1 H, J = 12.1 Hz), 4.41 (dd, 1 H, J = 7.9, 3.1 Hz), 3.83 (dd, 1 H, J = 10.6, 3.1 Hz), 3.63 (dd, 1 H, J = 10.6, 7.9 Hz), 1.42 (s, 3 H), 1.34 (s, 3 H), 1.67-1.24 (m, 10 H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 137.9, 128.1 (2 C), 127.7 (2 C), 127.4, 109.9, 108.7, 77.2 (t, J = 22 Hz), 76.7, 73.3, 72.2 (t, J = 23 Hz), 68.5, 66.7 (quintet, J = 23 Hz), 36.3, 34.7, 27.7, 25.2, 24.9, 23.8, 23.7; IR (neat) 2934, 2861, 2117, 1452, 1380, 1370, 1271, 1243, 1219, 1182, 1109, 1070, 1052, 736, 698; HRMS m/z 366.2339 (366.2340 calcd for $C_{21}H_{26}O_5{}^2H_4$, M^+)

D,L-4:5-O-cyclohexylidene-3,4,5,5'-*d*4-2:3-O-isopropylidene-ribitol (±13). The benzyl ribitol ±**17** (2.34 g, 6.4 mmol) was dissolved in 75 mL of THF and hydrogenated over 5% palladium on carbon (0.4 g) at 1 atmosphere of hydrogen for 22 hours. The reaction was filtered through celite and the filtrate was concentrated. Purification by flash chromatography (20-40% Et_2O in petroleum ether) afforded 1.62 g (5.9 mmol, 92%) of alcohol (±**18**). ^1H NMR (300 MHz, $CDCl_3$) δ 4.36 (dd, 1 H, J = 7.3, 5.7 Hz), 3.90 (dd, 1 H, J = 11.9, 7.3 Hz), 3.83 (dd, 1 H, J = 11.9, 5.7 Hz), 2.99

(s, 1 H), 1.72-1.25 (m, 10 H), 1.40 (s, 3 H), 1.35 (s, 3 H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 110.4, 108.5, 77.6 (t, J = 24 Hz), 77.4, 72.2 (t, J = 23 Hz), 66.9 (quintet, J = 22 Hz), 60.6, 36.2, 34.6, 27.6, 25.0, 24.8, 23.9, 23.7; IR (neat) 3512, 2986, 2935, 2862, 2175, 2118, 1449, 1371, 1219, 1180, 1112, 1062, 861; HRMS m/z 276.1870 (276.1871 calcd for $C_{14}H_{20}O_5{}^2H_4$, M^+)

D,L-4:5-O-cyclohexylidene-3,4,5,5'-*d*4-2:3-O-isopropylidene-ribose (±19). Oxalyl chloride (53 mg, 0.42 mmol) was added to CH_2Cl_2 (200 mL), and cooled to - 78° C. DMSO (66 mg, 0.84 mmol) was added slowly via syringe, and the mixture was stirred for five minutes before the deuterated alcohol **±18** (104 mg, 0.38 mmol) was added to the flask as a solution in CH_2Cl_2 (30 mL). After stirring for 30 minutes at -78° C, triethylamine (154 mg, 1.52 mmol) was added slowly to the mixture. After stirring for 15 minutes a -78° C, the dry ice/acetone bath was removed and the solution was stirred at room temperature. Thirty minutes later, the reaction was quenched with an aqueous HCl solution (15 mL, 0.12 M). The organic and aqueous layers were immediately separated and the organic layer was washed with saturated aqueous $NaHCO_3$ solution (20 mL) and saturated aqueous NaCl solution (20 mL). The organic layer was dried over $MgSO_4$ and then concentrated. Purification by flash chromatography (20-40% Et_2O in petroleum ether) yielded 94 mg (0.34 mmol, 90%) of protected ribose (**±19**). ^1H NMR (300 MHz, $CDCl_3$) δ 9.75 (d, 1 H, J = 1.8 Hz), 4.63 (d, 1 H, J = 1.8 Hz), 1.73-1.28 (m, 10 H), 1.54 (s, 3 H), 1.38 (s, 3 H); ^{13}C NMR (75 MHz, $CDCl_3$) δ 197.2, 111.2, 110.7, 81.8, 78.4 (t, J = 23 Hz), 72.6 (t, J = 23 Hz), 66.5 (quintet, J = 23 Hz), 36.5, 34.5, 27.3, 25.4, 25.0, 23.9, 23.7; HRMS m/z 274.1715 (274.1714 calcd for $C_{14}H_{18}O_5{}^2H_4$, M^+)

D,L-3',4',5',5"-*d*4-ribose ±1. The protected deuterated ribose **±19** (22 mg, 78 μmol) was dissolved in 7 mL of H_2O : THF : AcOH (3:3:1). The flask was attached to a reflux condenser and heated to 60° C for 24 hours. The mixture was evaporated under vacuum affording crude D,L-3',4',5',5"-*d*4-ribose (**±1**). The amount of D-3',4',5',5"-

d_4-ribose (**-1**) produced was determined to be 31 μmoles using the enzymatic ribose assay described below. This corresponds to a 78% yield from the starting protected D-3',4',5',5"-d_4-ribose (**-19**), and a 39% yield from the racemic mixture of protected ribose (±**19**). NMR revealed the presence of the expected complex mixture of α and β anomers of the furanose and pyranose forms of ribose. HRFABMS (glycerol) m/z 137.0747 (137.0748 calcd for $C_5H_5O_4{}^2H_4$, $[M\text{-}OH]^+$)

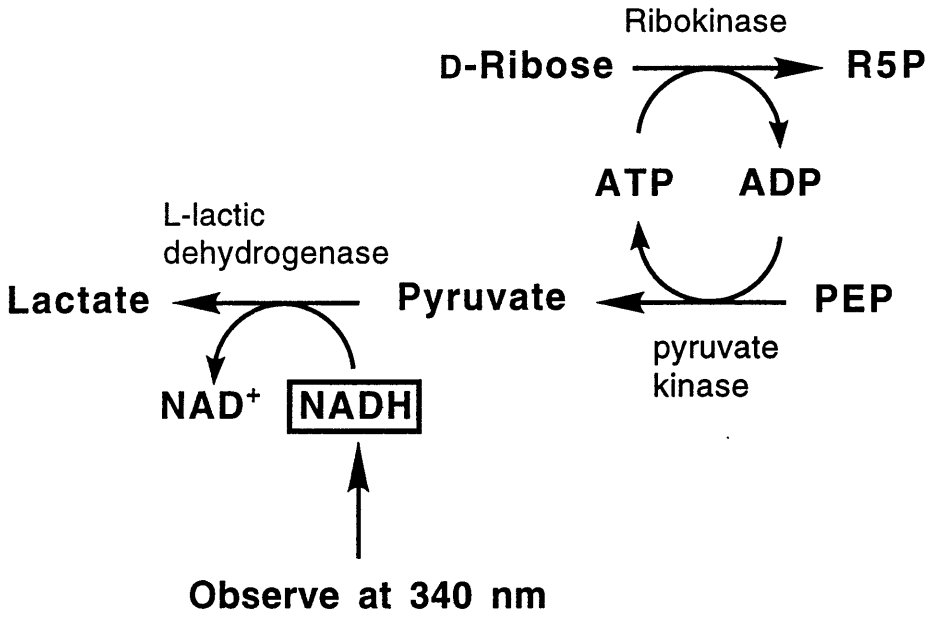


Figure 6.2 Enzymatic D-Ribose assay utilizing ribokinase, pyruvate kinase and L-lactic dehydrogenase to quantitate the amount of D-ribose produced by the chemical synthesis.

Enzymatic D-ribose Assay. The amount of D-d_4-ribose present in the crude racemic mixture produced by the chemical synthesis was quantitated by coupling D-ribose phosphorylation by ribokinase to the oxidation of NADH using the enzymes pyruvate kinase and lactate dehydrogenase, Figure 6.2. This assay is similar to the ribokinase assay of Gross *et. al.* (Gross *et al.*, 1983), except that the amount of D-ribose is limiting and the absolute amount of NADH consumed is measured. The ADP that is formed during the phosphorylation of ribose by ribokinase is subsequently rephosphorylated by PEP in a reaction catalyzed by pyruvate kinase. The conversion of ADP into ATP by pyruvate

kinase results in the formation of pyruvate, which is then reduced by lactate dehydrogenase using NADH as a cofactor to form lactic acid and NAD^+. The disappearance of NADH is monitored at 340 nm to determine the concentration of ribose in solution. In this assay there is a 1:1 relationship between NADH oxidation by lactate dehydrogenase and D-ribose phosphorylation by ribokinase, so the concentration of D-ribose can be calculated by determining the amount of NADH consumed. The assay solution (1 mL) contained 0.2 mM NADH, 1mM PEP, 10 mM $MgCl_2$, 3 mM ATP, 50 mM Tris-HCl buffer pH 7.8, 2 units of lactate dehydrogenase, 2 units of pyruvate kinase, and 1 unit of ribokinase. The assay solution was equilibrated after the addition of the enzymes, and then the assay was started by addition of a small aliquot of the unknown D-ribose solution. The absorbance of NADH at 340 nm was monitored until no further change was observed, and the total change in absorbance was used to calculate the amount of NADH consumed, and therefore the amount of D-ribose present in the aliquot, using an extinction coefficient of 6220 $cm^{-1}M^{-1}$.

Enzyme purification and cloning

Several of the enzymes that were required for the enzymatic nucleotide forming reactions are not commercially available, and those enzymes were obtained by purifying them from strains of *E. coli* which overexpress the desired enzymes. Many plasmids and overexpressing strains were generously donated to us by researchers who had previously constructed or obtained these strains. In three cases (APRT, XGPRT, and UPRT) where we were unable to obtain overexpressing strains, we constructed overexpressing strains from the reported gene sequences of the desired proteins. Overexpression of the desired enzymes allowed significant amounts of enzyme activity to be isolated with a minimum amount of work, and this allowed a single protein purification to produce enough enzyme

for multiple nucleotide forming reactions. Crude purifications (having only one column chromatography step) of the overexpressed enzymes were found to be sufficient in removing enzyme activities which would interfere with the nucleotide forming reactions, and a general procedure for these purifications is described in this section.

Cloning the gene for adenine phosphoribosyltransferase. The gene encoding *E. coli* adenine phosphoribosyltransferase (APRT) was cloned from the *E. coli* strain JM109 genome based on the reported gene sequence (Hershey & Taylor, 1986) using PCR with the oligonucleotides dCCG CGC GAA TTC ATG ACC GCG ACT GCA CAG CAG CTT and dCCG GCG CTG CAG TTA ATG GCC CGG GAA CGG GAC AAG as primers. The PCR product was digested with EcoR I and Pst I, and ligated into expression plasmid pKK223-3, that had been prepared by digestion with EcoR I and Pst I and dephosphorylation with calf intestinal alkaline phosphatase. Transformation of this construct into *E. coli* strain JM109 produced an IPTG inducible adenine phosphoribosyltransferase overproducing strain JM109/pTTA6. Overproduction of adenine phosphoribosyltransferase by strain JM109/pTTA6 was confirmed by comparing the activity of APRT in the crude lysate of IPTG induced strain JM109/pTTA6 with the APRT activity in untransformed JM109 crude lysate. The crude lysate of strain JM109/pTTA6 has over a hundred fold increase in adenine phosphoribosyltransferase activity as compared to untransformed JM109 crude lysate.

Cloning the gene for xanthine-guanine phosphoribosyltransferase. The gene encoding *E. coli* xanthine-guanine phosphoribosyltransferase (XGPRT) was cloned from the *E. coli* strain JM109 genome based on the reported gene sequence (Pratt & Subramani, 1983) using PCR with the oligonucleotides dCCG CGC GAA TTC ATG AGC GAA AAA TAC ATC GTC ACC and dCCG GCG CTG CAG TTA GCG ACC GGA GAT TGG CGG GAC as primers. The PCR product was digested with EcoR I and Pst I, and ligated into expression plasmid pKK223-3, that had been prepared by digestion with

EcoR I and Pst I and dephosphorylation with calf intestinal alkaline phosphatase. Transformation of this construct into *E. coli* strain JM109 produced an IPTG inducible xanthine-guanine phosphoribosyltransferase overproducing strain JM109/pTTG2. Overproduction of xanthine-guanine phosphoribosyltransferase by strain JM109/pTTG2 was confirmed by comparing the activity of XGPRT in the crude lysate of IPTG induced strain JM109/pTTG2 with the XGPRT activity in untransformed JM109 crude lysate. The crude lysate of strain JM109/pTTG2 has over a fifty fold increase in xanthine-guanine phosphoribosyltransferase activity as compared to untransformed JM109 crude lysate.

General Enzyme Purification scheme. All seven of the enzymes not obtained commercially were overexpressed and purified using a similar purification procedure. Cells were disrupted by sonication, and then cellular nucleic acids were precipitated with streptomycin sulfate. After the streptomycin sulfate precipitation, an ammonium sulfate precipitation of the desired proteins and subsequent dialysis of the resupended ammonium sulfate pellets was used as a purification step and also to concentrate and desalt the protein solution to prepare it for DEAE chromatography. The concentrated and desalted protein solution was then submitted to DEAE anion exchange chromatography as the final purification step. The general outline of this purification is given below, with specific details and exceptions for each enzyme following.

Cells were grown at 37° C in either LB media or A+B minimal media of Clark and Maaløe (Clark & Maaløe, 1967) containing 50 μg/mL of ampicillin and supplemented as indicated. All steps after cell growth were carried out in a 4° C cold room or on ice, unless otherwise specified. Cells were harvested by centrifugation at 6000 g for 15 min. The cell pellets were resuspended in the appropriate buffer and disrupted with thirty, 30 second sonication bursts using a Fisher Scientific 550 Sonic Dismembrator on a setting of 7, with a 2.5 minute interval between bursts. Cellular debris was removed by centrifugation at 31,000 g for 30 minutes. A 0.1 volume of a 20% streptomycin sulfate solution was then

added to the protein supernatant. After stirring for 15 minutes, the resulting precipitate was removed by centrifugation at 31000 g for 30 minutes. The supernatant was then fractionated by ammonium sulfate precipitation. The ammonium sulfate pellets were dissolved in the appropriate buffer and dialyzed against 2 L of buffer for 12 hours. The resulting solution was then loaded onto a 2.5 x 14 cm DEAE column (DEAE-650M Toyopearl resin from Supelco) that had been washed thoroughly with 3M KCl and then equilibrated with H_2O. After loading the protein solution, the column was washed with 100 mL of H_2O, and the enzymes were eluted with a salt gradient collecting fractions every 8 mL. The column fractions were assayed for protein, then pooled and concentrated by ammonium sulfate precipitation. Enzymes were stored at -20° C in 40-50% glycerol buffer solutions.

Ribokinase. The partial purification of ribokinase was accomplished by a simplified form of the method of Hope *et. al.* (Hope *et al.*, 1986) using *E coli.* strain MRi240 harboring the plasmid pJGK10. Hope *et. al.* report over 1000 fold increase in ribokinase activity when strain MRi240/pJGK10 is grown on a ribose media and induced with 3-β-indoleacrylic acid when compared to ribokinase activity of MRi240 which does not contain plasmid pJGK10. After a five step purification with a 6.9% total recovery of ribokinase activity, Hope *et. al.* report obtaining 780 units of ribokinase with a specific activity of 156 units/mg protein from 38.6 g of wet cells. Here we report a shortened version (with a higher % recovery) of the purification of Hope *et. al.* which results in partially purified ribokinase which is suitable for enzymatic nucleotide forming reactions.

Inducible strain MRi240/pJGK10 was grown on A+B minimal media (Clark & Maaløe, 1967) supplemented with glucose (4g/liter), ribose (4g/liter), and thiamine (7 mg/liter). One liter of media was inoculated with a 5 mL overnight of MRi240/pJGK10. The cells were grown to an OD of 0.6, then 1 mL of a solution of 3-β-indoleacrylic acid dissolved in ethanol (25 mg/mL) was added to the culture to induce ribokinase production.

Cells were harvested 14 hrs after the addition of 3-β-indoleacrylic acid, and the cell pellet was resuspended in 40 mL of 50 mM potassium phosphate buffer pH 7.5, 2 mM 2-mercaptoethanol (**buffer A**). Cell disruption, removal of cell debris, and streptomycin sulfate precipitation were carried out as described in the general procedure. The 0-80% saturation ammonium sulfate pellet was collected and dissolved in buffer A and dialysed against 2 L of buffer A (this step concentrates and desalts the protein solution to prepare it for DEAE chromatography). DEAE chromatography (2.5 x 14 cm DEAE column DEAE-650M Toyopearl resin from Supelco) was performed with a 500 mL linear gradient of 0 to 300 mM KCl in buffer A. Column fractions were assayed for ribokinase activity using the assay of Gross *et. al.* (Gross *et al.*, 1983), and the ribokinase activity eluted at approximately 160 mM KCl. Column fractions containing ribokinase activity were pooled and concentrated with ammonium sulfate precipitation. The final ammonium sulfate pellet was dissolved in a minimum amount of buffer A containing 40% glycerol. 700 units of ribokinase were obtained from 1 liter of bacterial culture.

5-Phosphoribosyl-α-1-pyrophosphate synthetase. The partial purification of 5-Phosphoribosyl-α-1-pyrophosphate synthetase (PRPP synthetase) was accomplished by a modified form of the procedure that Switzer *et. al.* (Switzer & Gibson, 1978) used to purify PRPP synthetase from *Salmonella typhimurium.* Switzer *et. al.* report obtaining 5400 units of PRPP synthetase with a specific activity of 103 units/mg protein from 500 g of *Salmonella typhimurium.* cells after a six step purification. Here we report a shortened version of the purification of Switzer *et. al.* conducted on cells from *E. coli* plasmid harboring strain HO561/pHO11 (Bjarne, 1985) which constitutively overproduces PRPP synthetase.

A 5 mL overnight of strain HO561/pHO11 (Bjarne, 1985), which constitutively expresses PRPP synthetase, was used to inoculate a one liter culture in LB media which was then grown for 12 hours. After cell harvest, the cell pellet was resuspended in 40 mL

of 50 mM potassium phosphate buffer pH 7.5, 2 mM 2-mercaptoethanol (**buffer A**). After sonication and removal of cellular debris, a 0.1 volume of 20% streptomycin sulfate solution was added to the supernatant and stirred at 0° C for 15 minutes. Before centrifuging, the streptomycin sulfate mixture (50 mL in a 125 mL glass Erlenmyer) was heat treated in a water bath at 55° C for 5 minutes, and then immediately poured into a 500 mL stainless steel beaker which had been cooled to 0° C in an ice bath. The solution was cooled at 0° C for 4-5 minutes, and then was poured into a centrifuge tube and the resulting precipitate was removed by centrifugation at 31000 g for 30 minutes. The supernatant was brought to 35% saturation of ammonium sulfate and stirred for 15 minutes. The ammonium sulfate pellet was collected and dissolved in 25 mL buffer A. The pH of this solution was adjusted to 4.6 by addition of 1 M acetic acid at 0° C, followed by immediate centrifugation for 10 minutes at 31000 g. The acid precipitate was collected and dissolved in buffer A. No DEAE chromatography was used in the purification of PRPP synthetase. The resulting solution was concentrated by ammonium sulfate precipitation and stored in buffer A containing 50% glycerol. PRPP synthetase activity was assayed using the method of Gross *et. al.* (Gross *et al.*, 1983). From 1 liter of bacterial culture, 47 units of PRPP synthetase were obtained.

Orotate phosphoribosyltransferase. The partial purification of orotate phosphoribosyltransferase (OPRT) was accomplished by a modified form of the procedure of Poulsen *et. al.* (Poulsen *et al.*, 1983). After a six step purification Poulsen *et. al.* report obtaining 2940 units of OPRT (specific activity of 397 units/mg protein) from 222 g of *E. coli* cells containing the plasmid pKK6 which causes the overexpression of OPRT. Here we report a shortened version of the purification of Poulsen *et. al.* conducted on *E. coli* strain JM109 harboring the plasmid pJTA43 (Jensen *et al.*, 1992) which causes similar overexpression of OPRT compared to the plasmid pKK6.

Strain JM109/pJTA43 (Jensen *et al.*, 1992), which constitutively expresses OPRT, was grown on LB media. One liter of medium was inoculated with a 25 mL overnight culture, and shaken for 14 hours at 37° C. The cells were harvested and resuspended in 40 mL of 100 mM Tris buffer pH 7.6, 2 mM EDTA, 2 mM 2-mercaptoethanol (**buffer B**). Sonication, removal of cellular debris, and streptomycin sulfate precipitation were carried out as described in the general procedure. The 0-75% saturation ammonium sulfate pellet was collected from the supernatant of the streptomycin sulfate precipitation, and then dialyzed against buffer B (this step concentrates and desalts the protein solution to prepare it for DEAE chromatography). The dialysate was subjected to DEAE chromatography (2.5 x 14 cm DEAE column DEAE-650M Toyopearl resin from Supelco), eluting with 500 mL linear gradient from 50-300 mM KCl in buffer B. Column fractions were assayed for OPRT activity according to the method of Schwartz and Neuhard (Schwartz & Neuhard, 1975). Fractions containing OPRT (eluting at approximately 170 mM KCl) were pooled and ammonium sulfate precipitated. The ammonium sulfate pellet was dissolved in a minimum amount of buffer B containing 40% glycerol. From one liter of culture, 31 units of orotate phosphoribosyltransferase were obtained.

Orotidine-5'-monophosphate decarboxylase. The partial purification of OMP decarboxylase was accomplished by a modified version of the procedure of Turnbough *et al.* (Turnbough *et al.*, 1987). After a five step purification Turnbough *et al.* report obtaining 294 units of OMP decarboxylase (specific activity 146 units/mg) from 2L of *E. coli* strain N100 containing the plasmid pDK26, which causes constitutive overexpression of OMP decarboxylase. Here we report a shortened version of the purification of Turnbough *et al.*, also conducted on strain N100/pDK26, which produces OMP decarboxylase which is suitable for use in enzymatic nucleotide forming reactions.

Strain N100/pDK26 (Turnbough *et al.*, 1987), which constitutively overproduces orotidine-5'-monophosphate decarboxylase (OMP decarboxylase), was grown on A+B

minimal media (Clark & Maaløe, 1967) supplemented with 8 g glucose, 0.5 g of casamino acids, 1 mL of 240 mM UMP, and thiamine (7 mg/liter). One liter of medium was inoculated with a 5 mL overnight culture, then shaken at 37° C for 12 hours. After cell harvest, the cells were resuspended in 40 mL of 64 mM Tris buffer (pH 7.8) with 5 mM 2-mercaptoethanol (**buffer C**). Sonication, removal of cellular debris, and streptomycin sulfate precipitation were carried out as described in the general procedure. The 45-75% ammonium sulfate pellet was collected from the supernatant of the streptomycin sulfate precipitation, and dissolved in buffer C for dialysis. The resulting dialysate was subjected to DEAE chromatography (2.5 x 14 cm DEAE column DEAE-650M Toyopearl resin from Supelco), eluting with a 500 mL linear gradient of 0-300 mM NaCl in buffer C. Column fractions containing OMP decarboxylase were detected by the method of Turnbough *et. al.*. (Turnbough *et al.*, 1987). Active fractions (eluting at approximately 140 mM NaCl) were pooled, ammonium sulfate precipitated, and stored in buffer C containing 40% glycerol. From one liter of culture, 400 units of OMP decarboxylase were obtained.

Adenine Phosphoribosyltransferase. Adenine phosphoribosyltransferase (APRT) was purified from an IPTG inducible, APRT overproducing strain which we constructed (JM109/pTTA6). The purification of APRT is based upon the general purification scheme that was used for the other enzymes in this section. Parkin *et. al.* (Parkin *et al.*, 1984) report a purification of APRT from *E. coli* strain K12 (no overexpression) where they obtain 2.5 units of APRT (specific activity 2.1 units/mg) from cells obtained from 12 liters of media. The overexpression of APRT in *E. coli* strain JM109/pTTA6 greatly increases the number of units of activity which can be obtained per liter of media.

IPTG inducible, adenine phosphoribosyltransferase (APRT) overproducing strain JM109/pTTA6 was grown on A+B minimal medium (Clark & Maaløe, 1967) supplemented with 22 mL glycerol, 2 g of glucose, and thiamine (7 mg/liter). The medium

was inoculated with a 5 mL overnight culture and grown for 12 hours, then induced with IPTG (.234g/liter) and grown for another 6 hours. Cell harvest and streptomycin sulfate precipitation were carried out as described in the general procedure using 64 mM Tris buffer (pH 7.8) with 5 mM 2-mercaptoethanol (**buffer C**). The 0-80% ammonium sulfate precipitate was collected from the supernatant of the streptomycin sulfate precipitation, and the ammonium sulfate pellet was dialyzed against buffer C (this step concentrates and desalts the protein solution to prepare it for DEAE chromatography). Then the protein solution was subjected to DEAE chromatography (2.5 x 14 cm DEAE column DEAE-650M Toyopearl resin from Supelco) with a 500 mL linear gradient of 0-300 mM KCl in buffer C. Column fractions containing APRT (eluting at approximately 150 mM KCl) were detected by the non radioactive APRT assay described below. The fractions containing APRT were pooled, ammonium sulfate precipitated, and finally dissolved in a minimum amount of buffer C containing 50% glycerol and stored at -20° C. From 1 liter of culture, 360 units of APRT were obtained.

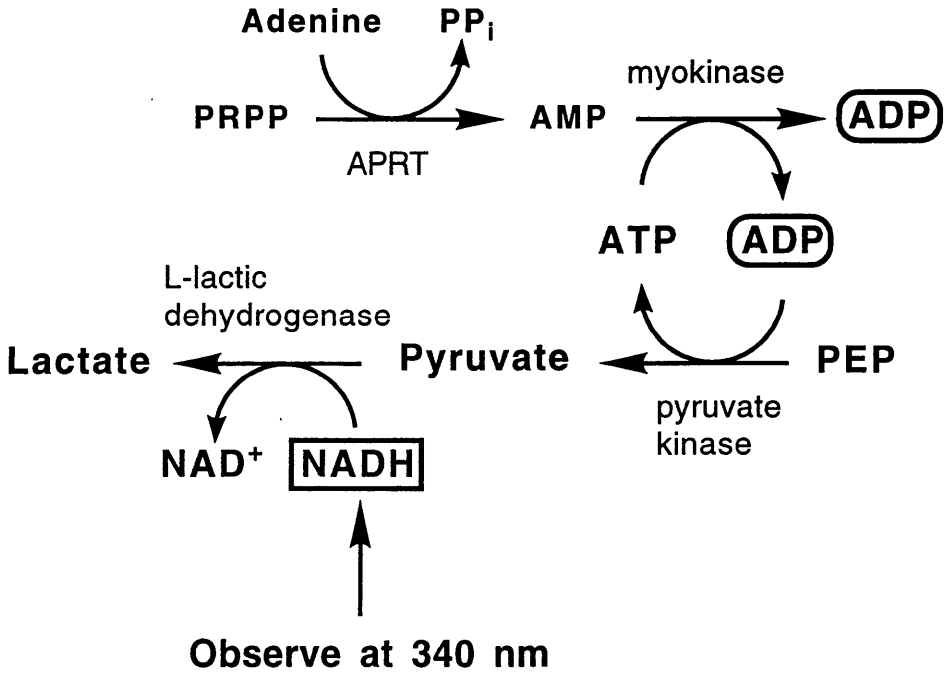


Figure 6.3 Adenine phosphoribosyltransferase (APRT) assay utilizing myokinase, pyruvate kinase, and L-lactic dehydrogenase to couple AMP formation by APRT to NADH oxidation. (Note: two equivalents of ADP are produced for each AMP formed by APRT.)

Adenine Phosphoribosyltransferase Assay. Adenine phosphoribosyltransferase (APRT) activity was measured by coupling the APRT reaction to oxidation of NADH with the enzymes myokinase, pyruvate kinase, and L-lactic dehydrogenase, Figure 6.3. This assay is similar to the PRPP synthetase assay of Gross *et al.* (Gross *et al.*, 1983). AMP that is formed by APRT is then phosphorylated by myokinase and ATP to form 2 equivalents of ADP. ADP is then phosphorylated by PEP and pyruvate kinase to form pyruvate, which is reduced by L-lactic dehydrogenase and NADH to produce lactate. Two equivalents of NADH are consumed for each AMP formed by APRT. Conversion of NADH to NAD^+ was monitored at 340 nm. The assay solution (1 mL) contained 0.2 mM NADH, 1 mM PEP, 10 mM $MgCl_2$, 1.5 mM PRPP, 1.5 mM adenine hydrochloride, 3 mM ATP, 50 mM Tris-HCl buffer pH 7.8, 2 units of lactate dehydrogenase, 2 units of pyruvate kinase, and 2 units of myokinase. The assay solution was equilibrated, a 20 µl aliquot of APRT solution was added to start the assay, and the absorbance change at 340 nm was monitored as a function of time. The activity was obtained using an extinction coefficient of 6220 $cm^{-1}M^{-1}$ for NADH.

Xanthine-Guanine Phosphoribosyltransferase. Xanthine-guanine phosphoribosyltransferase (XGPRT) was purified from an IPTG inducible, XGPRT overproducing strain which we constructed (JM109/pTTG2). The purification of XGPRT is based upon the general purification scheme that was used for the other enzymes in this section.

IPTG inducible, xanthine-guanine phosphoribosyltransferase (XGPRT) overproducing strain JM109/pTTG2 was grown on 1 liter of LB media inoculated with a 5 mL overnight culture for 12 hours, then induced with IPTG (.234g/liter) and grown for another 6 hours. Cells were harvested and streptomycin sulfate precipitated as in the general procedure using 64 mM Tris buffer (pH 7.8) with 5 mM 2-mercaptoethanol (**buffer C**). The 0-75% ammonium sulfate fraction was collected from the supernatant of

the streptomycin precipitation, then dialyzed against buffer C (this step concentrates and desalts the protein solution to prepare it for DEAE chromatography). The dialysate was subjected to DEAE chromatography (2.5 x 14 cm DEAE column DEAE-650M Toyopearl resin from Supelco) with a 500 mL gradient of 50-300 mM KCl in buffer C. Column fractions containing XGPRT were identified by SDS PAGE gel electrophoresis, pooled, concentrated by ammonium sulfate precipitation, and stored in buffer C containing 50% glycerol. Xanthine-guanine phosphoribosyltransferase activity was estimated by the method described below, and from 1 liter of culture, 28 units of XGPRT were obtained.

Xanthine-Guanine Phosphoribosyltransferase Assay. Xanthine-guanine phosphoribosyltransferase activity was estimated by observing GMP formation in a small (2 mL) GMP forming reaction by injecting aliquots onto an HPLC column as described in the preparative nucleotide synthesis section below. The assay conditions were as follows for a 2 mL reaction: 1 mM PRPP, 10 mM $MgCl_2$, 50 mM potassium phosphate buffer pH 7.5, 1 mg of solid guanine, and 1 unit of inorganic pyrophosphatase (obtained from Sigma). The reaction was stirred at room temperature and initiated by adding an aliquot of XGPRT solution. The time elapsed before the reaction was complete, determined by GMP formation observed by HPLC, was used to calculate the activity of the XGPRT solution. This method provides a very rough estimate of the activity of the XGPRT solution since the rate of the reaction is dependent upon the rate of the dissolution of guanine as well as the activity of the XGPRT solution. However, the assay is very similar to our GTP forming reactions, and thus suitable for empirical determination of the amount of XGPRT required for preparative reactions.

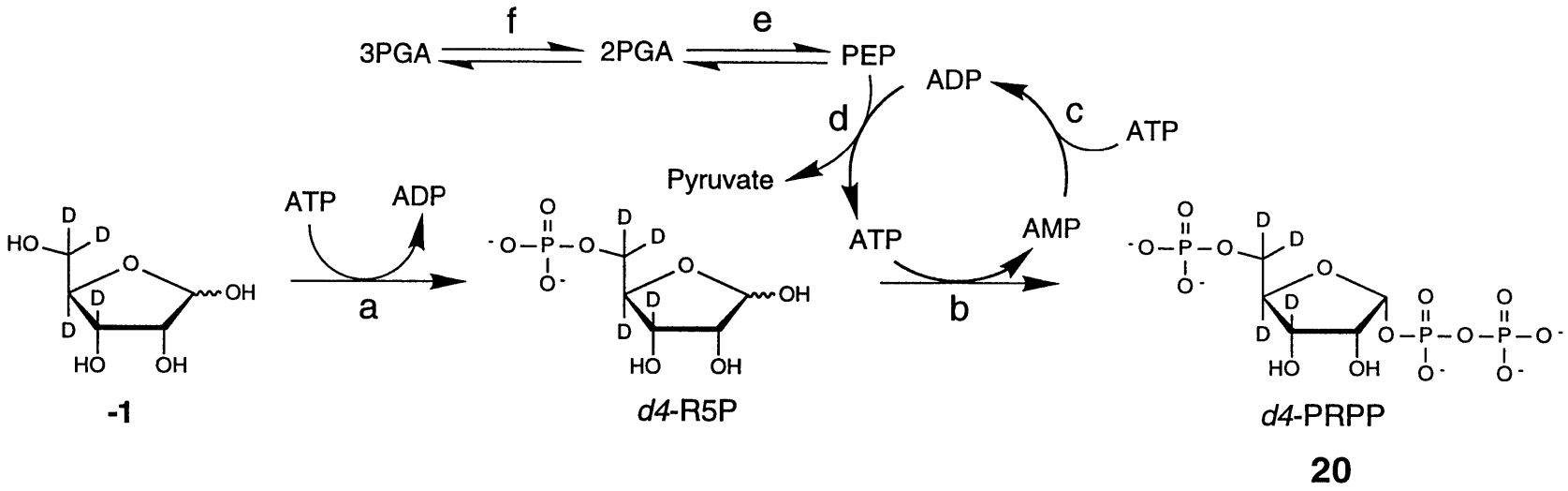
CTP Synthetase. CTP synthetase was purified by a modified form of the procedure of Anderson (Anderson, 1983). After a five step purification Anderson reports obtaining 96 units of CTP synthetase from 200g of *E. coli* B (no overexpression), specific activity 8.7 units/mg. Here we report a shortened version of the purification of Anderson

which is conducted on *E. coli* strain JM109 harboring the plasmid pMW5 (Weng et al., 1986), which causes a 50-fold increase in CTP synthetase activity.

A 1 liter flask of LB medium was inoculated with a 5 mL overnight culture of strain JM109/pMW5 (Weng et al., 1986), which constitutively overproduces CTP synthetase. The cells were grown for 12 hours, and then harvested. Sonication was performed as described in the general procedure in 64 mM Tris buffer (pH 7.8) with 5 mM 2-mercaptoethanol (**buffer C**) containing 20 mM glutamine. Removal of cellular debris and streptomycin sulfate precipitation were conducted as in the general procedure. From the resulting supernatant, the 0-65% ammonium sulfate pellet was extracted, and the ammonium sulfate pellet was dissolved in buffer C containing 4 mM glutamine and dialyzed against the same buffer (this step concentrates and desalts the protein solution to prepare it for DEAE chromatography). The dialysate was subjected to DEAE chromatography (2.5 x 14 cm DEAE column DEAE-650M Toyopearl resin from Supelco) with a 500 mL gradient of 50-300 mM potassium phosphate buffer, pH 7.5, containing 70 mM 2-mercaptoethanol and 4 mM glutamine. Column fractions were assayed by the method of Anderson (Anderson, 1983), and fractions containing CTP synthetase (eluting at approximately 170 mM potassium phosphate) were pooled, ammonium sulfate precipitated, and dissolved in a minimum amount of buffer B containing 50% glycerol. From 1 liter of culture, 38 units of CTP synthetase activity were obtained.

Enzymatic synthesis of NTPs from D-3',4',5',5"-d_4-ribose (Figures 6.4-6.6)

Previously reported procedures (Gross *et al.*, 1983, Parkin *et al.*, 1984, Simon *et al.*, 1989, Nikonowicz *et al.*, 1992; Batey *et al.*, 1992) were used as the starting point for the design of the enzymatic reactions presented here. The enzymatic reactions which convert ribose into NTPs require several enzymes some of which are commercially



20

Figure 6.4 Enzymatic synthesis of labeled PRPP from labeled D-ribose. Also shown is ATP regeneration by PEP/PK/MK system driven by 3PGA. Enzymes are (a) ribokinase (b) PRPP synthetase (c) myokinase (d) pyruvate kinase (e) enolase (f) 3-phosphoglycerate mutase. PRPP synthesis was coupled directly to ATP, GTP and UTP formation shown in Figure 2.8 and 2.10.

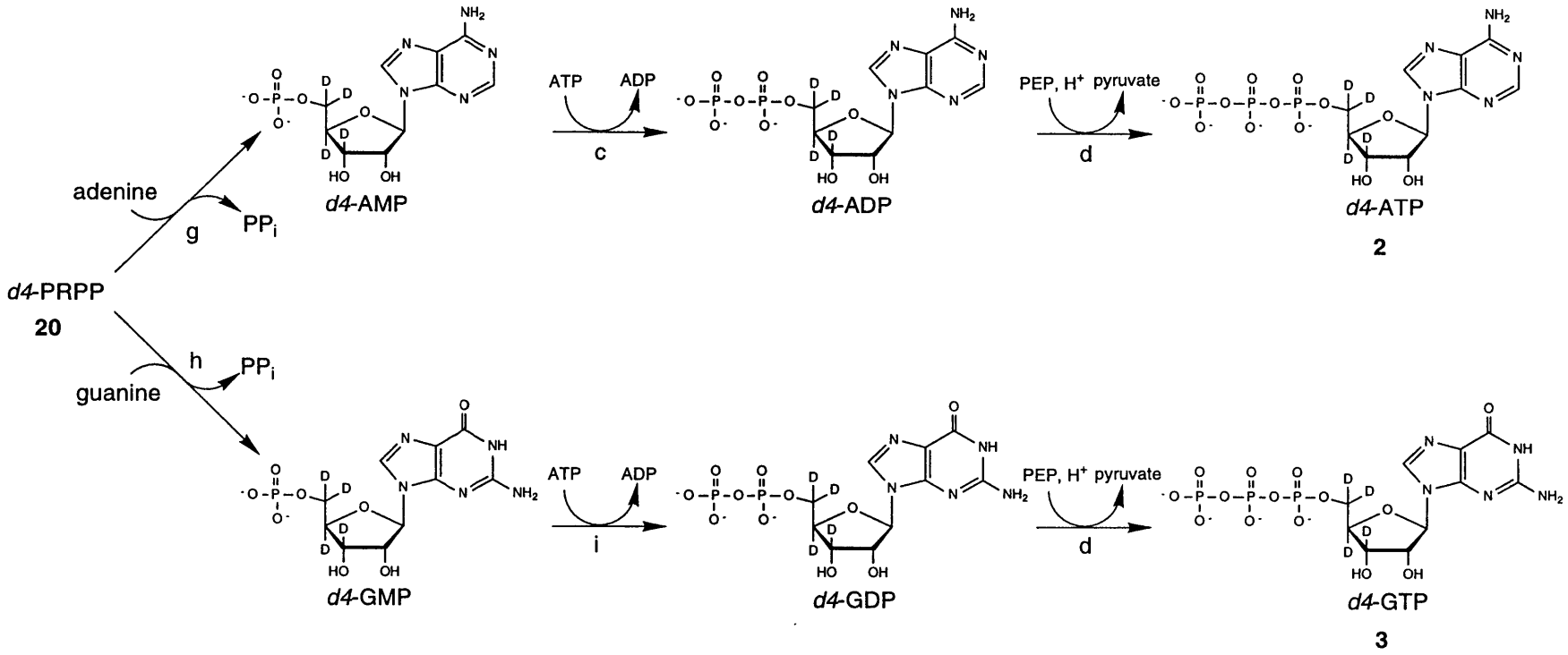


Figure 6.5 Enzymatic synthesis of purines ATP and GTP using purine salvage. Enzymes are (c) myokinase (d) pyruvate kinase (g) adenine phosphoribosyltransferase (h) xanthine-guanine phosphoribosyltransferase (i) guanylate kinase.

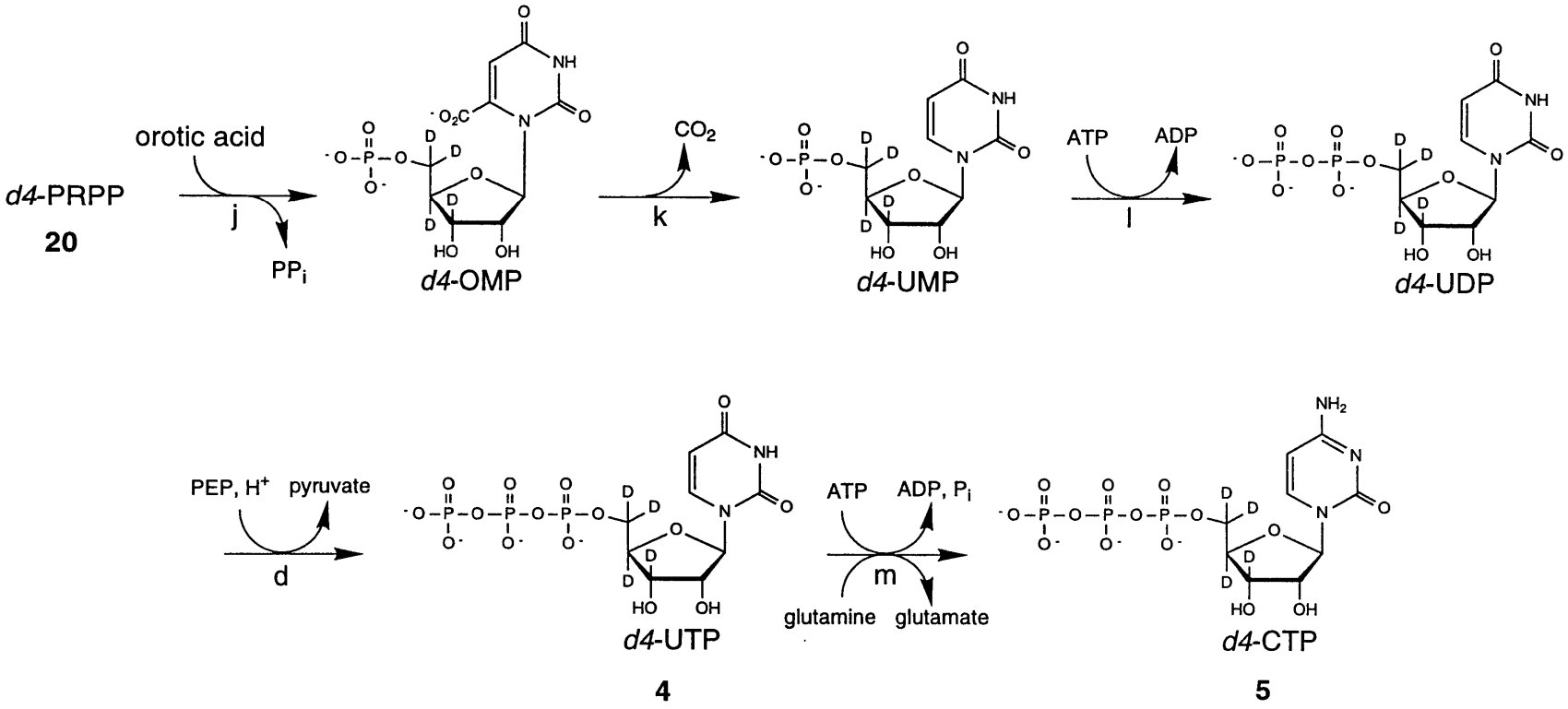


Figure 6.6 Enzymatic synthesis of UTP and CTP using pyrimidine biosynthesis. Enzymes are (d) pyruvate kinase (j) orotate phosphoribosyltransferase (k) OMP decarboxylase (l) nucleoside monophosphate kinase (m) CTP synthetase. In the preparative reactions, CTP synthesis is performed in a reaction separate from UTP formation.

available and others which had to be purified from overexpressing strains. Smaller amounts of the more expensive enzymes or those enzymes which had to be purified from overexpressing strains were used in the enzymatic reactions because those enzymes were difficult to obtain and valuable. Larger amounts of the inexpensive enzymes (such as the ATP regeneration enzymes enolase and pyruvate kinase) were added to the enzymatic reactions to insure that the rates of the reactions were not dependent upon those inexpensive enzyme activities. The rates of the nucleotide forming reactions were highly dependent upon the concentration of ATP in the reactions as well as the amount of the different enzymes added. Longer reaction times were required for the enzymatic GTP and UTP forming reactions compared to the ATP forming reaction because only small catalytic amounts of ATP were added to the GTP and UTP forming reactions. The enzymatic reactions were monitored by HPLC using a Vydac nucleotide column which allowed detection and rough quantitation of all of the NMPs, NDPs, and NTPs formed in the nucleotide reactions as well as the bases adenine and orotic acid (guanine was not detected by HPLC because of its low solubility in H_2O). The disappearance of starting material (adenine or orotic acid) and the appearance of the final product (NTPs) could be monitored in the enzymatic reactions using HPLC. In addition, ATP regeneration could also be monitored by HPLC by observing the phosphorylated state of the catalytic ATP in HPLC traces (that is, whether the majority of adenine nucleotides are present as ATP, ADP or AMP).

Preparative Nucleotide Synthesis. All enzymatic reactions were carried out under argon to reduce the rate of oxidation of enzymes. Reactions were monitored by HPLC on a 25x4.6 mm Vydac 303NT405 nucleotide column, using a linear gradient from 100% buffer A (0.045M ammonium formate brought to pH 4.6 with phosphoric acid) to 100% buffer B (0.5M NaH_2PO_4 brought to pH 2.7 with formic acid) in 10 min at a flow rate of 1 mL/min, with detection at 254 nm. Potassium phosphate buffer (50mM) was

used in reactions containing PRPP synthetase because the enzyme is inactivated in solutions with low phosphate concentration (Switzer & Gibson, 1978). The pH of the reactions was determined every 6-12 hours using pH paper, and adjusted periodically with 1M NaOH or HCl to maintain the pH between 7.0 and 7.9. NTPs were purified by using boronate affinity chromatography as described in the general procedures section of this chapter.

3', 4', 5', 5''-d_4-Adenosine 5'-triphosphate (2). Crude 3',4',5',5''-d_4-ribose (±**1**) containing 42.3 mg, 275 μmoles D-3',4',5',5''-*d*4-ribose, as determined by enzymatic ribose assay, was placed in a three neck flask containing adenine hydrochloride (87.5 mg, 510 μmol) and sodium 3-phosphoglycerate (2.55 mmol). This was dissolved in 50 mL of a solution containing 10 mM $MgCl_2$, 10 mM dithiothreitol, and 50 mM potassium phosphate buffer pH 7.5. Argon was bubbled through this solution for 5 minutes to remove oxygen. To this solution was added 125 units of phosphoglycerate mutase, 50 units of enolase, 200 units of pyruvate kinase, 25 units of myokinase, 5 units of ribokinase, 2 units of PRPP synthetase, and 2 units of adenine phosphoribosyltransferase. The reaction was initiated by adding a catalytic amount of ATP (0.5 mg, 0.9 μmol) to this solution. The reaction was stirred at 25° C and monitored by HPLC. The pH was maintained between 7.0 and 7.9 with 1M HCl. An additional 0.5 mmol of 3-phosphoglycerate was added every 5 hours. Formation of ATP was initially slow , but increased as the reaction proceeded due to the increase in ATP concentration. The reaction was stopped after 29 hours by concentration under vacuum and storage at -20° C. Purification of the 3',4',5',5''-d_4-ATP (**2**) was accomplished by boronate affinity chromatography. The 3',4',5',5''-d_4-ATP (**2**) was quantitated by absorption at 259 nm (ε_{259} = 15400 $cm^{-1}mol^{-1}$). This reaction yielded 213 μmoles of 3',4',5',5''-d_4-ATP (**2**) (78% yield). ^1H NMR (500 MHz, D_2O) δ 8.55 (s, 1H), 8.26 (s, 1H), 6.14 (d, 1H, J = 6.0 Hz), 4.80 (d, 1H, J = 6.0 Hz); MS m/z 510.1 (510.07 calcd for $C_{10}H_{11}{}^2H_4N_5O_{13}P_3$).

3',4',5',5''-d_4-Guanosine 5'-triphosphate (3). Crude 3',4',5',5''-d_4-ribose (±**1**) containing 42.3 mg, 275 μmoles D-3',4',5',5''-d_4-ribose, as determined by enzymatic ribose assay, was placed into a three neck flask containing guanine (77.1 mg, 510 μmol) and sodium 3-phosphoglycerate (2.55 mmol). This was mixed with 50 mL of a solution containing 10 mM $MgCl_2$, 10 mM dithiothreitol, and 50 mM potassium phosphate buffer pH 7.5. Argon was bubbled through this inhomogeneous solution for 5 minutes to remove oxygen. To this mixture was added 250 units of phosphoglycerate mutase, 100 units of enolase, 200 units of pyruvate kinase, 38 units of myokinase, 5 units of ribokinase, 2 units of PRPP synthetase, and (0.2 units) units of xanthine-guanine phosphoribosyltransferase. The reaction was initiated by adding 4.4 μmoles of 3',4',5',5''-d_4-ATP (**2**). The reaction was stirred at 25° C and monitored by HPLC. The pH was maintained between 7.0 and 7.9 with 1M HCl. Additional 3-phosphoglycerate was added to the reaction at a rate of 0.24 mmol every 5 hours. After four hours 2 units of guanylate kinase were added, and after 21 hours, an additional 4.4 μmoles of 3',4',5',5''-d_4-ATP (**2**) were added. The reaction was stopped after 72 hours by concentration under vacuum and storage at -20° C. Purification of the 3',4',5',5''-d_4-GTP (**3**) was accomplished by boronate affinity chromatography. The 3',4',5',5''-d_4-GTP (**3**) was quantitated by the absorbance at 253 nm (GTP ε_{253} = 13700 $cm^{-1}mol^{-1}$). This reaction yielded 206 μmoles of (**3**) (75% yield). ^1H NMR (500 MHz, D_2O) δ 8.14 (s, 1H), 5.93 (d, 1H, J = 6.0 Hz), 4.80 (d, 1H, J = 6.0 Hz); MS m/z 526.2 (526.07 calcd for $C_{10}H_{11}{}^2H_4N_5O_{14}P_3$).

3',4',5',5''-d_4-Uridine 5'-triphosphate (4). Crude 3',4',5',5''-d_4-ribose (±**1**) containing 84.7 mg, 550 μmoles D-3',4',5',5''-d_4-ribose, as determined by enzymatic ribose assay, was placed into a three neck flask containing potassium orotate (198 mg, 1.02 mmol) and sodium 3-phosphoglycerate (10.2 mmol). This was dissolved in 100 mL of a solution that contained 10 mM $MgCl_2$, 10 mM dithiothreitol, and 50 mM

potassium phosphate buffer pH 7.5. Argon was bubbled through this solution for 5 minutes to remove oxygen. To this solution was added 375 units of phosphoglycerate mutase, 150 units of enolase, 300 units of pyruvate kinase, 38 units of myokinase, 5 units of ribokinase, 2 units of PRPP synthetase, 1.5 units of orotate phosphoribosyltransferase, and 10 units of orotidine-5'-monophosphate decarboxylase. The reaction was initiated by adding 6.5 μmoles of 3',4',5',5"-*d*4-ATP (**2**), and then stirred at 25° C. The reaction was monitored by HPLC, and the pH was maintained between 7.0 and 7.9 with 1M HCl. An additional 0.35 mmol of 3-phosphoglycerate was added every 5 hours. After four hours 1 unit of nucleoside monophosphate kinase was added to catalyze the formation of UDP from UMP. After 21 hours, 10.9 μmoles of 3',4',5',5"-*d*4-ATP (**2**) was added to stimulate the rate of UTP formation. The reaction was stopped after 72 hours by concentration under vacuum and storage at -20° C. Purification of the 3',4',5',5"-*d*4-UTP (**4**) was accomplished by boronate affinity chromatography. The 3',4',5',5"-*d*4-UTP (**4**) was quantitated by the absorbance at 260 nm (ε_{260} = 10000 $cm^{-1}mol^{-1}$). This reaction yielded 495 μmoles of 3',4',5',5"-*d*4-UTP (**4**) (90% yield). ^1H NMR (500 MHz, D_2O) δ 7.96 (d, 1H, J = 8.1 Hz), 5.99 (d, 1H, J = 5.4 Hz), 5.97 (d, 1H, J = 8.4 Hz), 4.38 (d, 1H, J = 5.4 Hz); MS m/z 487.1 (487.05 calcd for $C_9H_{10}{}^2H_4N_2O_{15}P_3$).

3',4',5',5"-*d*4-Cytidine 5'-triphosphate (5). 3',4',5',5"-*d*4-UTP (**4**) (247 μmol), glutamine (369 mg, 2.52 mmol), and sodium 3-phosphoglycerate (505 μmol) were placed into a three neck flask. This was dissolved in 250 mL of a solution containing 10 mM $MgCl_2$ and 50 mM Tris-HCl buffer pH 7.8. Argon was bubbled through this solution for 5 minutes to remove oxygen. To this solution was added 250 units of phosphoglycerate mutase, 100 units of enolase, 200 units of pyruvate kinase, and 5.5 units of cytidine 5'-triphosphate synthetase. To initiate the reaction 26.9 μmoles of 3',4',5',5"-*d*4-ATP (**2**) and 2.5 μmoles of 3',4',5',5"-*d*4-GTP (**3**) were added. The reaction was stirred at 25° C and monitored by HPLC. After 3.5 hours, 26.9 μmoles of 3',4',5',5"-*d*4-

ATP, 2.5 μmoles of 3',4',5',5"-d_4-GTP, and 0.86 mmoles of 3-phosphoglycerate were added to stimulate the rate of the reaction. The reaction was stopped after 8 hours by concentration under vacuum and storage at -20° C. Purification of the 3',4',5',5"-d_4-CTP (**5**) was accomplished by DEAE chromatography. The 3',4',5',5"-d_4-CTP was eluted from the DEAE resin with a triethylammonium bicarbonate (TEABC) salt gradient (0-500 mM TEABC, the nucleotides eluted from 250-500 mM TEABC.). Boronate affinity chromatography was not possible due to the use of tris buffer, which binds to the boronate affinity resin. The 3',4',5',5"-d_4-CTP (**5**) was quantitated by absorbance at 259 nm and 280 nm, using the extinction coefficients ε_{259}ATP=15400, ε_{280}ATP=1911, ε_{259}CTP=7204, and ε_{280}CTP=6905. This reaction yielded 215 μmoles of 3',4',5',5"-d_4-CTP (**5**) (87% yield). ^{1}H NMR (500 MHz, D_2O) δ 8.07 (d, 1H, J = 7.7 Hz), 6.21 (d, 1H, J = 7.7 Hz), 5.99 (d, 1H, J = 4.4 Hz), 4.34 (d, 1H, J = 4.4 Hz); MS m/z 486.1 (486.06 calcd for $C_9H_{11}{}^2H_4N_3O_{14}P_3$).

Chapter 3 Experimental

Cloning and Purification of Uracil Phosphoribosyltransferase

Cloning the gene for uracil phosphoribosyltransferase. The gene encoding *E. coli* uracil phosphoribosyltransferase was cloned from the *E. coli* strain JM109 genome based on the reported gene sequence (Andersen *et al.*, 1992) using PCR with the oligonucleotides dCCG CGC GAA TTC TTG AAG ATC GTG GAA GTC AAA CAC and dCCG GCG AAG CTT TTT CGT ACC AAA GAT TTT GTC ACC as primers. The PCR product was digested with EcoR I and Hind III, and ligated into expression plasmid pKK223-3, that had been prepared by digestion with EcoR I and Hind III and dephosphorylation with calf intestinal alkaline phosphatase. Transformation of this construct into *E. coli* strain JM109 produced an IPTG inducible uracil

phosphoribosyltransferase overproducing strain JM109/pTTU2. Overproduction of uracil phosphoribosyltransferase by strain JM109/pTTU2 was confirmed by comparing the activity of UPRT in the crude lysate of IPTG induced strain JM109/pTTU2 with the UPRT activity in untransformed JM109 crude lysate. Due to the sensitivity of the spectrophotometric UPRT assay (described below), no UPRT activity could be detected in the untransformed JM109 crude lysate, but overproduction of UPRT in strain JM109/pTTU2 allowed detection of UPRT activity in its crude lysate (indicating at least a ten fold increase in UPRT activity).

Purification of Uracil Phosphoribosyltransferase Uracil phosphoribosyltransferase (UPRT) was purified from an IPTG inducible, UPRT overproducing strain which we constructed (JM109/pTTU2). The purification of UPRT is based upon the general purification scheme that was used for the other enzymes in this section. Rasmussen *et. al.* (Rasmussen *et al.*, 1986) report a purification of UPRT from a plasmid harboring strain of *E. coli* which overproduces UPRT 15-fold. From 72g of cells, Rasmussen *et. al.* report obtaining 73 units of UPRT with a specific activity of 6.6 units/mg.

IPTG inducible, uracil phosphoribosyltransferase (UPRT) overproducing strain JM109/pTTU2 was grown on 2 liters of LB media containing 50 μg/mL of ampicillin for 12 hours at 37° C after being inoculated with a 5 mL overnight culture, then induced with IPTG (0.23 g/liter) and grown for another 6 hours. All steps after cell growth were carried out in a 4° C cold room or on ice. Cells were harvested by centrifugation at 6000 g for 15 min. The cell pellets were resuspended in 65 mM Tris buffer (pH 7.8) with 5 mM 2-mercaptoethanol (buffer A) and disrupted with thirty, 30 second sonication bursts using a Fisher Scientific 550 Sonic Dismembrator on a setting of 7, with a 2.5 minute interval between bursts. Cellular debris was removed by centrifugation at 31,000 g for 30 minutes. A 0.1 volume of a 20% streptomycin sulfate solution was then added to the protein

supernatant. After stirring for 15 minutes, the resulting precipitate was removed by centrifugation at 31000 g for 30 minutes. The 0-75% ammonium sulfate fraction was collected from the supernatant of the streptomycin precipitation, then dialyzed against buffer A (this step concentrates and desalts the protein solution to prepare it for DEAE chromatography). The dialysate was subjected to DEAE chromatography (2.5 x 14 cm DEAE column DEAE-650M Toyopearl resin from Supelco) with a 500 mL gradient of 50-300 mM KCl in buffer A. Column fractions containing UPRT were identified by UPRT assay (peak fractions eluted at approximately 140 mM KCl), pooled, concentrated by ammonium sulfate precipitation, and stored at -20° C in buffer A containing 50% glycerol. Uracil phosphoribosyltransferase activity was determined by the method described below, and from 2 liters of culture, approximately 40 units of UPRT were obtained.

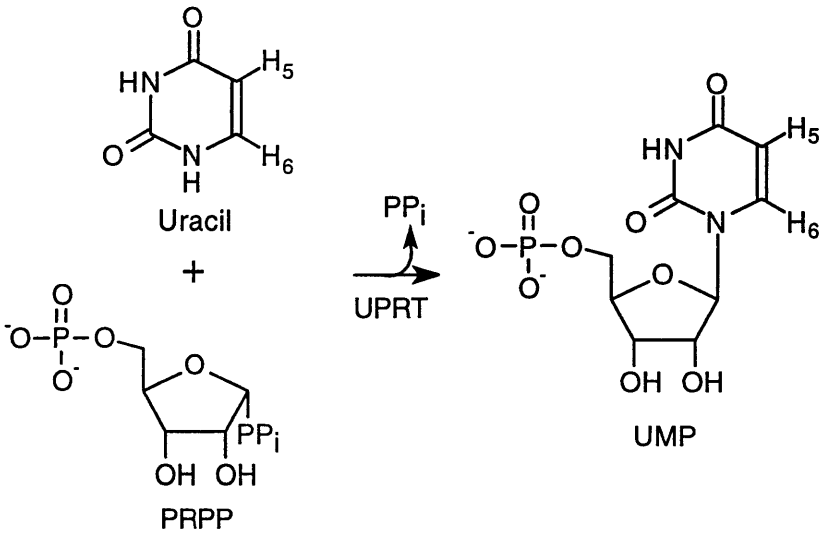


Figure 6.7 The spectrophotometric uracil phosphoribosyltransferase assay is based upon the change in extinction coefficients at 271 nm between uracil and UMP ($\Delta\varepsilon_{271}$=2763 $cm^{-1}M^{-1}$).

Uracil phosphoribosyltransferase assay. A spectrophotometric assay of uracil phosphoribosyltransferase (UPRT) activity was developed to detect and quantitate UPRT activity, Figure 6.7. This assay was based on the change of absorbance at 271 nm that occurs when uracil is converted into UMP. The assay solution (1 mL) contained 50

mM tris-HCl, 5 mM $MgCl_2$, 0.1 mM uracil, and 1.5 mM PRPP. The assay was started by addition of a small aliquot of UPRT solution, and the absorbance of the solution was monitored at 271 nm as a function of time. The activity of the UPRT solution was determined using a change in extinction coefficient at 271 nm of 2763 $cm^{-1}M^{-1}$ for the conversion of uracil into UMP.

Enzymatic synthesis of NTPs from Glucose (Figures 6.8 and 6.9)

The enzymatic conversions of glucose into the 4 NTPs of RNA described in this section were based upon previously reported procedures (Parkin *et al.*, 1984; Parkin & Schramm, 1987; Rising & Schramm, 1994) and also the enzymatic reactions which are described in Chapter 2 of this thesis. All of the additional pentose phosphate pathway enzymes required to convert glucose into PRPP are commercially available. Uracil phosphoribosyltransferase was cloned and purified in order to be able to use uracil as the starting base for the pyrimidine nucleotides instead of orotic acid. Instead of producing each nucleotide separately as described in Chapter 2, many of the procedures in this section describe the conversion of glucose into ATP, GTP, and UTP in a single enzymatic reaction. This was done to reduce the number of reactions and the amount of work necessary to produce isotopically labeled RNA from glucose. The rates of the glucose nucleotide forming reactions were highly dependent upon the concentration of cofactors (ATP and $NADP^+$) in the reactions as well as the amount of the different enzymes added. In reactions where ATP was to be synthesized in combination with other nucleotides, ATP was synthesized first in order to increase the activity of enzymes whose activities depended on ATP concentration, and then the other nucleotides were synthesized. Catalytic amounts of ATP and $NADP^+$ were used in the glucose enzymatic reactions, and this resulted in slow enzymatic reactions which could take up to a week or more to go to completion.

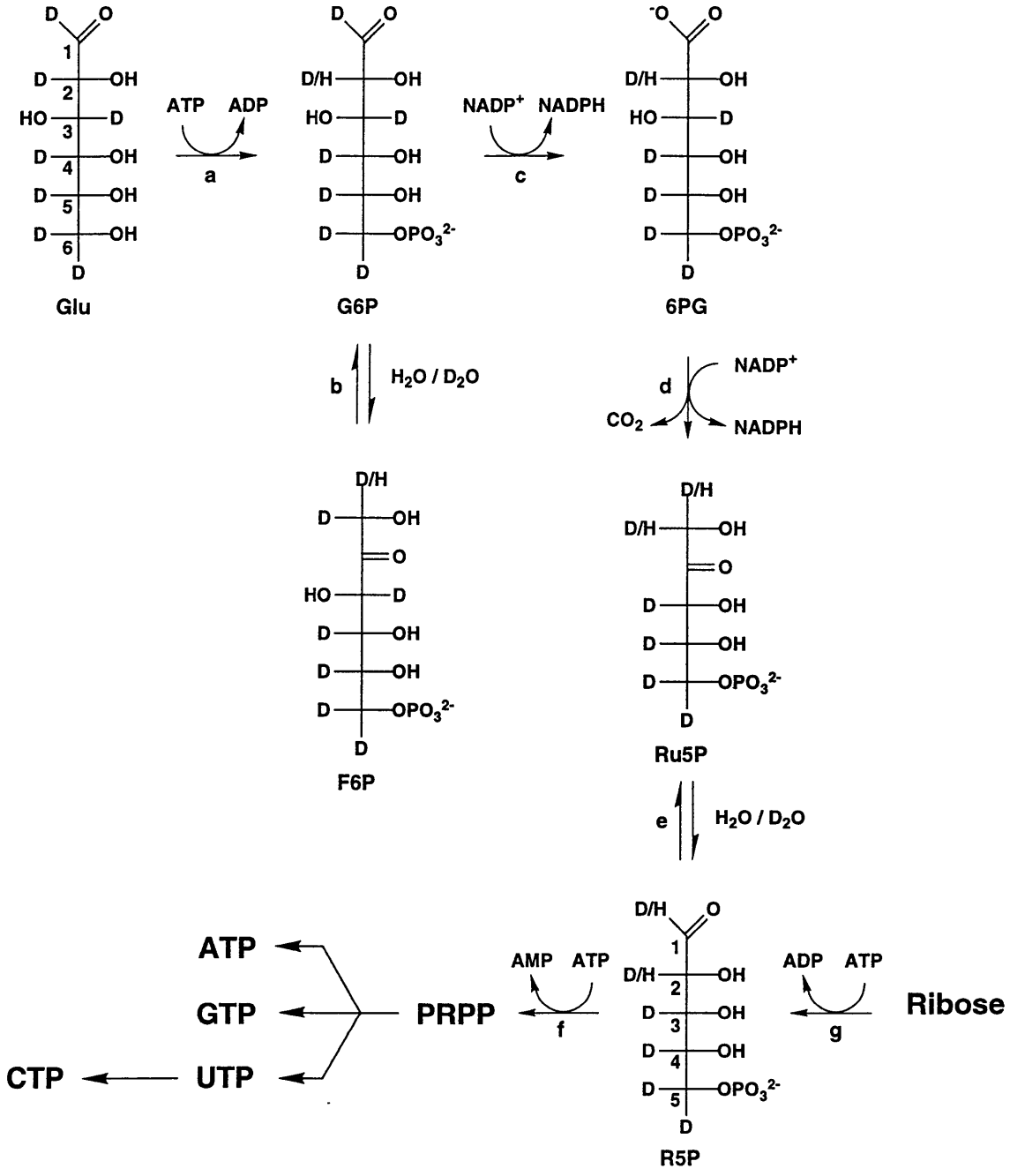


Figure 6.8 Enzymatic conversion of fully deuterated glucose into PRPP showing possible hydrogen exchange of what will become the H1' and H2' of NTPs and subsequent conversion of PRPP into the four NTPs of RNA. (a) hexokinase (b) glucose-6-phosphate isomerase (c) glucose-6-phosphate dehydrogenase (d) 6-phosphogluconate dehydrogenase (e) ribose-5-phosphate isomerase (f) PRPP synthetase (g) ribokinase

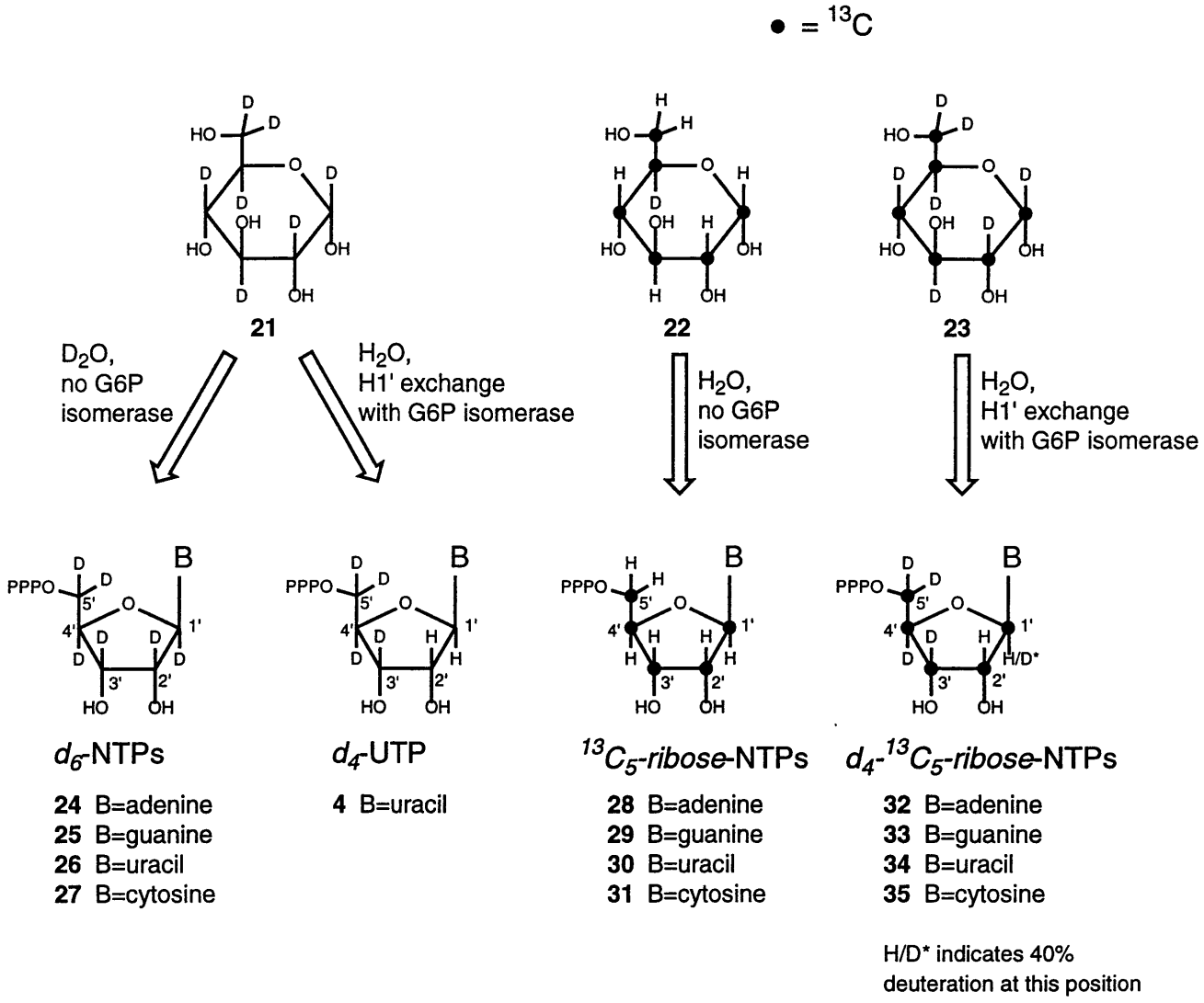


Figure 6.9 Different isotopically labeled NTPs synthesized from 1,2,3,4,5,6,6-d_7-glucose (**21**), 1,2,3,4,5,6-$^{13}C_6$-glucose (**22**), and 1,2,3,4,5,6,6-d_7-1,2,3,4,5,6-$^{13}C_6$-glucose (**23**) in this thesis. The solvent (which determines H2' labeling), and whether glucose-6-phosphate isomerase was used to exchange the H1' is indicated for each type of nucleotide forming reaction.

Preparative Nucleotide Synthesis. Reactions were monitored by HPLC on a 25x4.6 mm Vydac 303NT405 nucleotide column, using a linear gradient from 100% buffer A (0.045M ammonium formate brought to pH 4.6 with phosphoric acid) to 100% buffer B (0.5M NaH_2PO_4 brought to pH 2.7 with formic acid) in 10 minutes at a flow rate of 1 mL/min, with detection at 260 or 254 nm. Potassium phosphate buffer (50mM) was used in reactions containing PRPP synthetase because the enzyme is inactivated in solutions with low phosphate concentration (Switzer & Gibson, 1978). The pH of the reactions was determined every 6-12 hours using pH paper, and adjusted periodically with 1M NaOH or HCl (or NaOD or DCl for reactions run in D_2O) to maintain the pH between 7.0 and 7.9. In reactions that are to be conducted for an extended periods of time it is useful to add some ampicillin (≈50 μg/mL) to prevent bacterial growth in the reactions. Generally, in reactions where ATP was to be formed, a small catalytic amount of ATP was added to start the reaction, and the enzymes required for ATP formation were added first. Once a significant amount of ATP had formed in the reaction, the remaining enzymes for the other NTP formation were added. This was done to increase the rate of the reactions since the activities of many of the enzymes in the reactions are dependent on ATP concentration. In reactions where no ATP was to be formed, a greater initial amount of ATP is added to the reaction.

Preparation of 1',2',3',4',5',5'-$^2H_6$[-ATP (24), -GTP (25), and -UTP (26)]. Into a Millipore 10,000 MWC centricon were placed 188 units of phosphoglycerate mutase, 100 units of enolase, 75 units of pyruvate kinase, 50 units of myokinase, 75 units of L-glutamic dehydrogenase, 60 units of hexokinase, 15 units of glucose-6-phosphate dehydrogenase, 5 units of 6-phosphogluconic dehydrogenase, 100 units of phosphoriboisomerase, 2.7 units of PRPP synthetase, 1.3 unit of adenine phosphoribosyltransferase, 2 units of xanthine-guanine phosphoribosyltransferase, 2 units of uracil phosphoribosyltransferase, 1 unit of guanylate kinase, and 0.5 units of nucleoside

monophosphate kinase. The enzyme solution was centrifuged at 6000 rpm until 90-95% of its original volume had passed through the centricon dialysis membrane. Then 1.5 mL of D_2O were added to the centricon, and the centricon was again centrifuged at 6000 rpm until 90-95% of the liquid had passed through the filter. This was repeated twice more, and then the enzymes were considered to have been exchanged into D_2O and ready to be added to the reaction.

Into a 300 mL round bottom flask were placed sodium 3-phosphoglycerate (6.8 mmol), adenine hydrochloride (0.034 g, 0.2 mmol), α-ketoglutaric acid (0.580 g, 4.0 mmol), NH_4Cl (0.30 g, 5.6 mmol), $MgCl_2$ (0.6 mmol), dithiothreitol (0.21 g, 1.4 mmol), 3 mL of 1M potassium phosphate buffer pH = 7.5, and 60 mL of H_2O. The pH of the resulting solution was adjusted to 7.7 with 1M NaOH, and then the mixture was concentrated to dryness under vacuum. The residue was dissolved in 10 mL of D_2O and concentrated to dryness under vacuum, this was repeated three times total. The D_2O exchanged residue was dissolved in 60 mL of D_2O, and the pD of the solution was adjusted to 7.5 with DCl and NaOD. The D_2O exchanged enzyme solution was added to the mixture, and then *d6*-ATP (1.7 μmole) from a previous reaction, $NADP^+$ (0.005 g, 7 μmol), and 1',2',3',4',5',6',6-$^2H_7$-glucose (0.150 g, 0.8 mmol) were added to start the reaction. After 40 hours a significant amount of the adenine had been converted into *d6*-ATP (**24**), and so guanine (0.30 g, 0.2 mmol) and uracil (0.45 g, 0.4 mmol) were added to the reaction to begin GTP and UTP formation. Additional D_2O exchanged 3-PGA (1 mmol) was added to the reaction after 6 days, and the reaction was stopped by freezing it after 7 days. The mixture was purified by boronate chromatography, and nucleotides were quantitated by UV absorbance at 260 nm, assuming a 1:1:2 ratio of ATP:GTP:UTP. Using this quantification, 84% of the $^2H_7$-glucose was converted into nucleotides and recovered. Half of this product was used to make *d6*-CTP (**27**), and half was used directly in transcription reactions. Characterization was conducted on the mixture of products

resulting from this reaction, ^{1}H NMR and mass spectrometry analysis were consistent with a mixture of *d6*-ATP (**24**), *d6*-GTP (**25**), and *d6*-UTP (**26**) in a 1:1:2 ratio. ^{1}H NMR (600 MHz, D_2O) δ 8.55 (s, 1H), 8.29 (s, 1H), 8.13 (s, 1H), 7.97 (d, 1H, J = 8.4 Hz), 5.98 (d, 1H, J = 8.4 Hz); MS *m/z* 512.0, 528.0, 489.0 (512.08, 528.08, 489.06 calcd for $C_{10}H_9D_6O_{13}N_5P_3$, $C_{10}H_9D_6O_{14}N_5P_3$, $C_9H_8D_6O_{15}N_2P_3$ respectively).

Preparation of 1',2',3',4',5',5'-$^2H_6$-CTP (27) from a mixture of *d6*-ATP (24), *d6*-GTP (25), and *d6*-UTP (26). Into a 500 mL 3-neck flask were placed half of the purified nucleotide mixture from the previous reaction containing *d6*-ATP (**24**) (0.084 mmol), *d6*-GTP (**25**) (0.084 mmol), and *d6*-UTP (**26**) (0.168 mmol). To this were added NH_4Cl (0.535 g, 10.0 mmol), sodium 3-phosphoglycerate (0.5 mmol), and 250 mL of a solution containing 5 mM $MgCl_2$ and 1 mM dithiothreitol. To start the reaction, 62 units of phosphoglycerate mutase, 50 units of enolase, 50 units of pyruvate kinase, 5 units of myokinase, and 3 units of CTP synthetase were added to the reaction. This was stirred for 48 hours and then the reaction was purified by boronate chromatography. The yield for this reaction was 97% as determined by UV absorbance. Characterization was conducted on the mixture of products resulting from this reaction, ^{1}H NMR and mass spectrometry analysis were consistent with a mixture of *d6*-ATP (**24**), *d6*-GTP (**25**), *d6*-CTP (**27**) in a 1:1:2 ratio. ^{1}H NMR (600 MHz, D_2O) δ 8.53 (s, 1H), 8.27 (s, 1H), 8.13 (s, 1H), 7.99 (d, 1H, J = 7.2 Hz), 6.17 (d, 1H, J = 7.2 Hz); MS *m/z* 512.0, 528.0, 488.0 (512.08, 528.08, 488.07 calcd for $C_{10}H_9D_6O_{13}N_5P_3$, $C_{10}H_9D_6O_{14}N_5P_3$, $C_9H_9D_6O_{14}N_3P_3$ respectively).

Preparation of 3',4',5',5'-$^2H_4$-UTP (4). Into a three neck flask were placed sodium 3-phosphoglycerate (1.3 mmol), α-ketoglutaric acid (0.29 g, 2.0 mmol), and NH_4Cl (0.15 g, 2.8 mmol). This was dissolved in 40 mL of a solution containing 10 mM $MgCl_2$, 20 mM dithiothreitol, and 50 mM potassium phosphate buffer pH 7.5, and the pH of this solution was adjusted to 7.5 with 1M NaOH. After the solution had been

neutralized, *d6*-ATP (20 μmol) and 1,2,3,4,5,6,6-$^{2}H_7$-glucose (0.075 g, 0.4 mmol) were added to the mixture. The phosphorylation of glucose to G6P and G6P's isomerization was started by adding 188 units of phosphoglycerate mutase, 50 units of enolase, 75 units of pyruvate kinase, 38 units of myokinase, 60 units of hexokinase, and 75 units of glucose-6-phosphate isomerase. After 30 hours, the phosphorylation of glucose appeared to be complete by HPLC analysis of the ATP in the reaction, and the reaction was heated to 34° C for 20 hours to speed up the exchange of the H2 with solvent. The reaction was cooled to room temperature and 30 units of glutamic dehydrogenase, 6 units of glucose-6-phosphate dehydrogenase, 2.5 units of 6-phosphogluconic dehydrogenase, 100 units of phosphoriboisomerase, 1 unit of PRPP synthetase, 2 units of uracil phosphoribosyltransferase, 0.5 units of nucleoside monophosphate kinase, 125 units of phosphoglycerate mutase, 50 units of enolase, 50 units of pyruvate kinase, 38 units of myokinase, $NADP^+$ (0.009 g, 12 μmol), sodium 3PGA (2.7 mmol) were added to begin formation of *d4*-UTP (**4**). After 8 days the reaction was frozen to stop it, and purified by boronate chromatography. The yield of this reaction was 67% as determined by UV absorbance. ^{1}H NMR (600 MHz, D_2O) δ 7.95 (d, 1H, J = 8.4 Hz), 5.99 (d, 1H, J = 6 Hz), 5.98 (d, 1H, J = 8.4 Hz), 4.39 (d, 1H, J = 4.8 Hz); MS *m/z* 487.0 (487.05 calcd for $C_9H_{10}D_4O_{15}N_2P_3$).

Preparation of 1',2',3',4',5'-$^{13}C_5$[-ATP (28), -GTP (29), and -UTP (30)]. Into a three neck flask were placed sodium 3-phosphoglycerate (6.0 mmol), adenine hydrochloride (0.034 g, 0.2 mmol), α-ketoglutaric acid (0.58 g, 4.0 mmol), and NH_4Cl (0.30 g, 5.6 mmol). This was dissolved in 80 mL of a solution containing 10 mM $MgCl_2$, 20 mM dithiothreitol, and 50 mM potassium phosphate buffer pH 7.5, and the pH of this solution was adjusted to 7.5 with 1M NaOH. After the solution had been neutralized, ATP (8 μmol), $NADP^+$ (0.005 g, 7 μmol), and 1,2,3,4,5,6-$^{13}C_6$-D-glucose (0.150 g, 0.8 mmol) were added to the mixture. The reaction was started by adding 125

units of phosphoglycerate mutase, 50 units of enolase, 50 units of pyruvate kinase, 50 units of myokinase, 50 units of L-glutamic dehydrogenase, 60 units of hexokinase, 10 units of glucose-6-phosphate dehydrogenase, 2.5 units of 6-phosphogluconic dehydrogenase, 40 units of phosphoriboisomerase, 2 units of PRPP synthetase, 1 unit of adenine phosphoribosyltransferase, 2 units of xanthine-guanine phosphoribosyltransferase, 2 units of uracil phosphoribosyltransferase, 1 unit of guanylate kinase, and 0.5 units of nucleoside monophosphate kinase. At 24 hours a significant amount of ATP had formed, and guanine (0.030 g, 0.2 mmol), and uracil (0.045 g, 0.4 mmol) were added to the reaction to begin GTP and UTP formation. After 3 days an additional 2.0 mmol of sodium 3-phosphoglycerate was added to the reaction, and the reaction was stopped after five days. Boronate purification of the reaction and quantification of the purified products by UV absorbance indicated that 88% of the glucose had been converted into isotopically labeled nucleotides and recovered. Characterization was conducted on the mixture of products resulting from this reaction, ^{1}H NMR, ^{13}C HSQC, and mass spectrometry analysis were consistent with a mixture of 1',2',3',4',5'-$^{13}C_5$[-ATP (**28**), -GTP (**29**), and -UTP (**30**)] in a 1:1:2 ratio. ^{1}H NMR (600 MHz, D_2O) δ 8.55 (s, 1H), 8.28 (s, 1H), 8.13 (s, 1H), 7.98 (dd, 1H, J = 8.4, 2.4 Hz), 6.15 (d, 1H, J = 165.5 Hz), 6.00 (d, 1H, J = 170.4 Hz), 5.93 (d, 1H, J = 166.2 Hz), 5.99 (d, 1H, J = 8.4 Hz), 4.95-4.0 (m, 15 H); ^{13}C HSQC crosspeaks (δ proton, δ carbon): (6.18, 89.81), (6.03, 91.23), (5.95, 89.81), (4.85, 76.49), (4.83, 77.38), (4.60, 73.47), (4.45, 72.76), (4.42, 76.85), (4.43, 87.14), (4.38, 86.97), (4.32, 86.43), (4.30, 68.32), (4.28, 68.15), (4.26, 68.32); MS *m/z* 511.0, 527.0, 488.0 (511.06, 527.06, 488.04 calcd for ($^{12}C_5$)($^{13}C_5$)$H_{15}O_{13}N_5P_3$, ($^{12}C_5$)($^{13}C_5$)$H_{15}O_{14}N_5P_3$, ($^{12}C_4$)($^{13}C_5$)$H_{14}O_{15}N_2P_3$ respectively).

Preparation of 1',2',3',4',5'-$^{13}C_5$-CTP (31) from a mixture of 1',2',3',4',5'-$^{13}C_5$[-ATP (28), -GTP (29), and -UTP (30)]. 1',2',3',4',5'-$^{13}C_5$-CTP (**31**) was prepared from the purified mixture of NTPs **28-30** from the previous

reaction in a procedure similar to the one used to produce *d6*-CTP (**27**). The yield of this reaction was 80%. Characterization was conducted on the mixture of products resulting from this reaction, ^{1}H NMR, ^{13}C HSQC, and mass spectrometry analysis were consistent with a mixture of 1',2',3',4',5'-$^{13}C_5$[-ATP (**28**), -GTP (**29**), and -CTP (**31**)] in a 1:1:2 ratio. ^{1}H NMR (600 MHz, D_2O) δ 8.54 (s, 1H), 8.27 (s, 1H), 8.14 (s, 1H), 7.98 (d, 1H, J = 7.2 Hz), 6.16 (d, 1H, J = 7.2 Hz), 6.15 (d, 1H, J = 165.6 Hz), 6.01 (d, 1H, J = 169.2 Hz), 5.94 (d, 1H, J = 165.6 Hz), 4.94-4.00 (m, 15 H); ^{13}C HSQC crosspeaks (δ proton, δ carbon): (6.18, 89.81), (6.04, 92.11), (5.96, 89.81), (4.85, 76.49), (4.83, 77.38), (4.60, 73.47), (4.43, 87.14), (4.42, 72.23), (4.39, 86.97), (4.36, 77.20), (4.31, 85.90), (4.31, 68.32), (4.31, 67.79), (4.26, 68.32); MS *m/z* 511.0, 527.0, 487.0 (511.06, 527.06, 487.05 calcd for $(^{12}C_5)(^{13}C_5)H_{15}O_{13}N_5P_3$, $(^{12}C_5)(^{13}C_5)H_{15}O_{14}N_5P_3$, $(^{12}C_4)(^{13}C_5)H_{15}O_{14}N_3P_3$ respectively).

Preparation of 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$[-ATP (32), -GTP (33), -UTP (34)]. Into a three neck flask were placed sodium 3-phosphoglycerate (2.1 mmol), adenine hydrochloride (0.034 g, 0.2 mmol), α-ketoglutaric acid (0.76 g, 4.0 mmol), and NH_4Cl (0.30 g, 5.6 mmol). This was dissolved in 80 mL of a solution containing 10 mM $MgCl_2$, 20 mM dithiothreitol, and 50 mM potassium phosphate buffer pH 7.5, and the pH of this solution was adjusted to 7.5 with 1M NaOH. After the solution had been neutralized, *d6*-ATP (7 μmol) and 1,2,3,4,5,6,6-$^{2}H_7$-1,2,3,4,5,6-$^{13}C_6$-D-glucose (0.155 g, 0.8 mmol) were added to the mixture. The phosphorylation of the glucose to G6P and G6P's isomerization was started by adding 188 units of phosphoglycerate mutase, 50 units of enolase, 75 units of pyruvate kinase, 38 units of myokinase, 60 units of hexokinase, and 75 units of glucose-6-phosphate isomerase. After stirring for 48 hours, 30 units of glutamic dehydrogenase, 7.5 units of glucose-6-phosphate dehydrogenase, 2.5 units of 6-phosphogluconic dehydrogenase, 100 units of phosphoriboisomerase, 1 unit of PRPP synthetase, 1 unit of adenine

phosphoribosyltransferase, $NADP^+$ (0.007 g, 10 μmol), sodium 3PGA (2.7 mmol) were added to begin *d4-$^{13}C_5$*-ATP (**32**) formation. At 72 hours, guanine (0.030 g, 0.2 mmol), uracil (0.45 g, 0.4 mmol), 2 units of xanthine-guanine phosphoribosyltransferase, 2 units of uracil phosphoribosyltransferase, 1 unit of guanylate kinase, and 0.5 units of nucleoside monophosphate kinase were added to the reaction to begin GTP and UTP formation. After 5 days, additional 3- phosphoglycerate (1.7 mmol) were added to the reaction, and after 7 days the reaction was stopped and purified. Boronate purification of the reaction and quantification of the purified products by UV absorbance indicated that 92% of the glucose had been converted into isotopically labeled nucleotides and recovered. Characterization was conducted on the complex mixture of products resulting from this reaction. ^1H NMR, ^{13}C HSQC, and mass spectrometry analysis were consistent with a mixture of 3',4',5',5'-$^2H_4$-1',2',3',4',5'-$^{13}C_5$-ATP, 1',3',4',5',5'-$^2H_5$-1',2',3',4',5'-$^{13}C_5$-ATP, 3',4',5',5'-$^2H_4$-1',2',3',4',5'-$^{13}C_5$-GTP, 1',3',4',5',5'-$^2H_5$-1',2',3',4',5'-$^{13}C_5$-GTP, 3',4',5',5'-$^2H_4$-1',2',3',4',5'-$^{13}C_5$-UTP, and 1';3',4',5',5'-$^2H_5$-1',2',3',4',5'-$^{13}C_5$-UTP resulting from complete hydrogen exchange of the H2' in the nucleotides, and approximately 60% hydrogen exchange of the H1'. ^1H NMR (600 MHz, D_2O) δ 8.54 (s, 1H), 8.28 (s, 1H), 8.13 (s, 1H), 7.97 (d, 1H, J = 8.4 Hz), 6.14 (d, 0.6H, J = 173.4 Hz), 6.00 (d, 0.6H, J = 169.2 Hz), 5.93 (d, 0.6H, J = 167.4 Hz), 5.99 (d, 1H, J = 8.4 Hz), 4.81 (d, 1H, J = 150 Hz), 4.79 (d, 1H, J = 150 Hz), 4.39 (d, 1H, J = 151.2 Hz); ^{13}C HSQC crosspeaks (δ proton, δ carbon): (6.17, 89.81), (6.03, 91.23), (5.95, 89.81), (4.84, 76.49), (4.82, 77.20), (4.42, 76.67); MS *m/z* 515.1, 516.1, 531.1, 532.1, 492.0, 493.0 (515.09, 516.10, 531.08, 532.09, 492.06, 493.07 calcd for $(^{12}C_5)(^{13}C_5)H_{11}D_4O_{13}N_5P_3$, $(^{12}C_5)(^{13}C_5)H_{10}D_5O_{13}N_5P_3$, $(^{12}C_5)(^{13}C_5)H_{11}D_4O_{14}N_5P_3$, $(^{12}C_5)(^{13}C_5)H_{10}D_5O_{14}N_5P_3$, $(^{12}C_4)(^{13}C_5)H_{10}D_4O_{15}N_2P_3$, $(^{12}C_4)(^{13}C_5)H_9D_5O_{15}N_2P_3$ respectively).

Chapter 6

Preparation of 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-CTP (35) from a mixture of 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$[-ATP (32), -GTP (33), and -UTP (34)] 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-CTP (**35**) was prepared from the mixture of **32**, **33**, and **34** from the previous reaction in a procedure similar to the one used to produce *d6*-CTP (**27**). The yield of this reaction was 89%. Characterization was conducted on the complex mixture of products resulting from this reaction, ^{1}H NMR, ^{13}C HSQC, and mass spectrometry analysis were consistent with a mixture of 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-ATP, 1',3',4',5',5'-$^{2}H_5$-1',2',3',4',5'-$^{13}C_5$-ATP, 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-GTP, 1',3',4',5',5'-$^{2}H_5$-1',2',3',4',5'-$^{13}C_5$-GTP, 3',4',5',5'-$^{2}H_4$-1',2',3',4',5'-$^{13}C_5$-CTP, and 1',3',4',5',5'-$^{2}H_5$-1',2',3',4',5'-$^{13}C_5$-CTP. ^{1}H NMR (600 MHz, D_2O) δ 8.54 (s, 1H), 8.27 (s, 1H), 8.13 (s, 1H), 8.00 (d, 1H, J = 6.6 Hz), 6.15 (d, 0.6H, J = 166.2 Hz), 6.17 (d, 0.6H, J = 7.8 Hz), 5.93 (d, 0.6H, J = 165.6 Hz), 6.00 (d, 1H, J = 171.0 Hz), 4.81 (d, 2H, J = 150 Hz), 4.33 (d, 1H, J = 152.4 Hz); ^{13}C HSQC crosspeaks (δ proton, δ carbon): (6.17, 89.81), (6.03, 92.11), (5.96, 89.81), (4.84, 76.49), (4.82, 77.20), (4.36, 77.20); MS *m/z* 515.1, 516.1, 531.1, 532.1, 491.0, 492.0 (515.09, 516.10, 531.08, 532.09, 491.08, 492.08 calcd for $(^{12}C_5)(^{13}C_5)H_{11}D_4O_{13}N_5P_3$, $(^{12}C_5)(^{13}C_5)H_{10}D_5O_{13}N_5P_3$, $(^{12}C_5)(^{13}C_5)H_{11}D_4O_{14}N_5P_3$, $(^{12}C_5)(^{13}C_5)H_{10}D_5O_{14}N_5P_3$, $(^{12}C_4)(^{13}C_5)H_{11}D_4O_{14}N_4P_3$, $(^{12}C_4)(^{13}C_5)H_{10}D_5O_{14}N_3P_3$ respectively).

Chapter 6

Isotopic labeling of the base moieties of nucleotides

Deuteration of the H5 of pyrimidines with ammonium sulfite

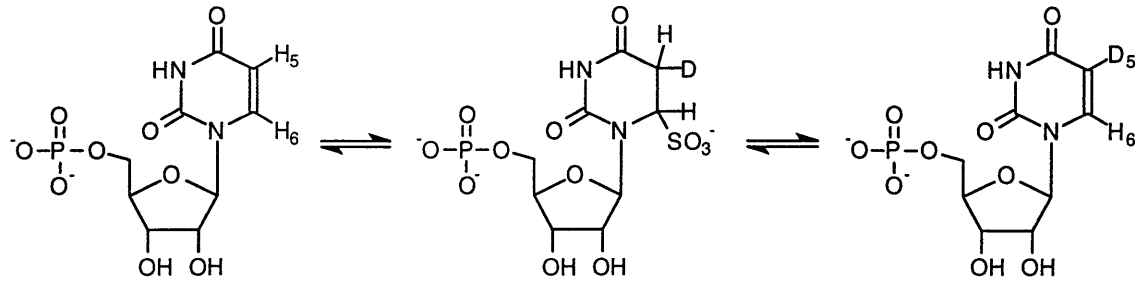


Figure 6.10 Treatment of UMP with Ammonium Sulfite in D_2O

Preparation of 5-$^2H_1$-UMP (36) from UMP. Into a 25 mL round bottom flask were placed ammonium sulfite (1.68 g), the sodium salt of UMP (0.30 g, 0.81 mmol), and 5 mL of H_2O. Once all of the solid had been dissolved in the water, the pH of the solution was adjusted to approximately 8.5 with 1M NaOH. The H_2O was removed under vacuum with a rotovap, and then the residue exchanged 4 times with 4 mL of D_2O. The exchanged residue was dissolved in 6 mL of D_2O and a reflux condenser was connected to the round bottom flask. The reaction was stirred and heated to 55° C for approximately 70 hours, at which time NMR analysis of the reaction indicated that all of the H5 protons of UMP had been exchanged for deuterium. The pH of the reaction was adjusted to between 9.5-10.5 with 20% NaOD in D_2O, and then stirred at room temperature for an hour. An NMR spectra of the reaction at this point showed that over 99% of the sulfite adduct had been converted back into UMP. The reaction was purified with boronate chromatography in a manner similar to the purification of the NTPs from the nucleotide forming reactions except that the column was conducted at 4° C to prevent exchange of the H5 with H_2O before the sulfite was separated from the UMP. No loss of the deuterium label was detected by NMR after the purification. This reaction yielded 0.66

mmol (81% yield) of 5-$^2H_1$-UMP (**36**) as determined by absorbance at 260 nm. ^1H NMR (300 MHz, D_2O) δ 7.84 (s, 1H), 5.93 (d, 1H, J = 6.0 Hz), 4.80-3.90 (m, 5H).

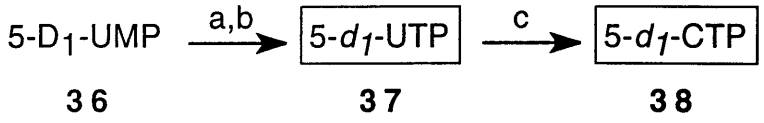


Figure 6.11 Enzymatic conversion of 5-*d*$_1$-UMP (**36**) into 5-*d*$_1$-UTP (**37**) and 5-*d*$_1$-CTP (**38**). a) nucleoside monophosphate kinase b) pyruvate kinase c) CTP synthetase.

Preparation of 5-$^2H_1$-UTP (37) from 5-$^2H_1$-UMP (36). 5-$^2H_1$-UMP (**36**) (0.66 mmol), sodium 3-phosphoglycerate (2.64 mmol), 10 mg of ampicillin, and ATP (0.033 g, 0.06 mmol) were placed into a 250 mL flask and dissolved in 60 mL of a solution containing 10 mM $MgCl_2$ and 10 mM dithiothreitol. The pH of the solution was adjusted to 7.2 with 1M NaOH, and then phosphoglycerate mutase (125 units), enolase (50 units), myokinase (25 units), pyruvate kinase (50 units), and nucleoside monophosphate kinase (0.5 units) were added to begin the reaction. The reaction was stirred at room temperature for 48 hours and then an additional 0.5 mmol of 3-phosphoglycerate was added to the reaction. The reaction was stopped after 80 hours and purified by boronate chromatography. This reaction produced 0.48 mmol (72% yield) of 5-$^2H_1$-UTP (**37**). Half of this reaction was used in the next reaction to produce 5-$^2H_1$-CTP (**38**) and half of the reaction was used directly in transcription reactions to produce D5-pyrimidine RNA.

Preparation of 5-$^2H_1$-CTP (38) from 5-$^2H_1$-UTP (37). 5-$^2H_1$-UTP (**37**) (0.24 mmol) and ATP (0.18 g, 0.33 mmol) were placed into a 500 mL round bottom flask and dissolved in 330 mL of a solution which contained 50 mM ammonium chloride, 5 mM $MgCl_2$, and 5 mM dithiothreitol. The pH of the solution was adjusted to 7.7 with 1M NaOH, and then phosphoglycerate mutase (125 units), enolase (50 units), pyruvate kinase (50 units), and CTP synthetase (2 units) were added to begin the reaction. The

reaction was stirred for 48 hours at room temperature and then concentrated. Approximately 0.23 mmol (96% yield) of 5-$^2H_1$-CTP (**38**) was produced by this reaction. The 5-$^2H_1$-CTP (**38**) was used directly in transcription reactions with 5-$^2H_1$-UTP (**37**) to produce D5-pyrimidine RNA.

Enzymatic deuteration of the H6 of Uracil

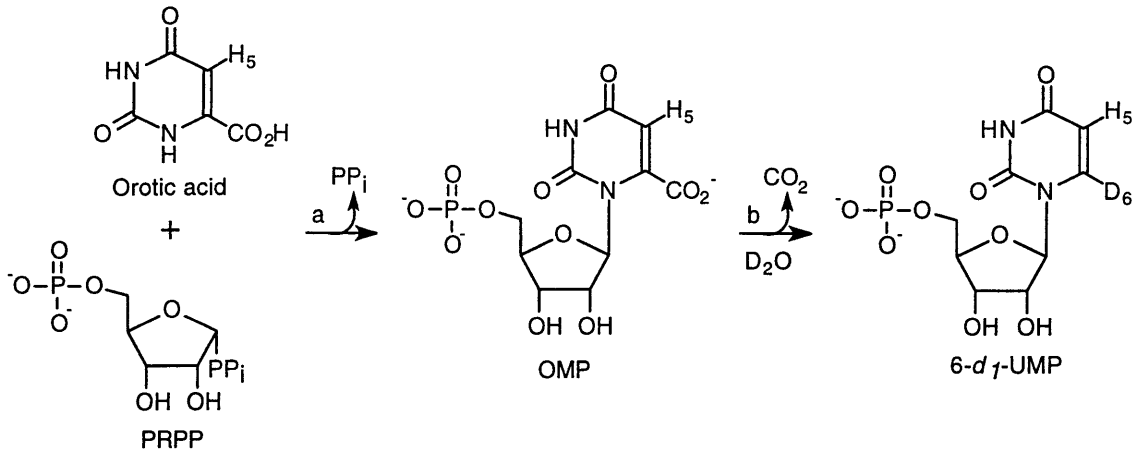


Figure 6.12 Exchange of H6 of pyrimidines. a) Orotate phosphoribosyltransferase b) OMP decarboxylase.

Preparation of 1',2',3',4',5',5'-$^2H_6$[-ATP (24), -GTP (25)] and 1',2',3',4',5',5',5-$^2H_7$-UTP (39). This reaction is very similar to the reaction used to form 1',2',3',4',5',5'-$^2H_6$-[ATP (**24**), GTP (**25**), and UTP (**26**)], but instead of forming UTP from uracil with uracil phosphoribosyltransferase, this reaction forms UTP from orotic acid with the enzymes orotate phosphoribosyltransferase and OMP decarboxylase. The reaction is conducted in D_2O to prevent the loss of the deuterium label at the 2' position, and this also has the effect of deuterating the H6 of UMP during the conversion of OMP into UMP by OMP decarboxylase (Figure 6.12). The synthesis of 1',2',3',4',5',5'-$^2H_6$-ATP (**24**) and 1',2',3',4',5',5'-$^2H_6$-GTP (**25**) has been described previously in the experimental of Chapter 3, but the synthesis of 1',2',3',4',5',5',5-$^2H_7$-UTP (**39**) is new and demonstrates how the H6 of pyrimidines can be deuterated using

OMP decarboxylase. The nucleotides produced in this reaction were used to produce the TAR 465 labeling pattern described in Chapter 4.

Into a Millipore 10,000 MWC centricon were placed 188 units of phosphoglycerate mutase, 100 units of enolase, 75 units of pyruvate kinase, 50 units of myokinase, 75 units of L-glutamic dehydrogenase, 60 units of hexokinase, 12 units of glucose-6-phosphate dehydrogenase, 5 units of 6-phosphogluconic dehydrogenase, 100 units of phosphoriboisomerase, 2.7 units of PRPP synthetase, 1.3 unit of adenine phosphoribosyltransferase, 2 units of xanthine-guanine phosphoribosyltransferase, 1.4 unit of orotate phosphoribosyltransferase, 2.9 units of OMP decarboxylase, 1 unit of guanylate kinase, and 0.5 units of nucleoside monophosphate kinase. The enzyme solution was centrifuged at 6000 rpm until 90-95% of its original volume had passed through the centricon dialysis membrane. Then 1.5 mL of D_2O were added to the centricon, and the centricon was again centrifuged at 6000 rpm until 90-95% of the liquid had passed through the filter. This was repeated twice more, and then the enzymes were considered to have been exchanged into D_2O and ready to be added to the reaction.

Into a 300 mL round bottom flask were placed sodium 3-phosphoglycerate (4.0 mmol), adenine hydrochloride (0.029 g, 0.17 mmol), α-ketoglutaric acid (0.37 g, 2.5 mmol), NH_4Cl (0.19 g, 3.5 mmol), $MgCl_2$ (0.5 mmol), dithiothreitol (0.18 g, 1.2 mmol), 2.5 mL of 1M potassium phosphate buffer pH = 7.5, and 50 mL of H_2O. The pH of the resulting solution was adjusted to 7.7 with 1M NaOH, and then the mixture was concentrated to dryness under vacuum. The residue was dissolved in 10 mL of D_2O and concentrated to dryness under vacuum, this was repeated three times total. The D_2O exchanged residue was dissolved in 50 mL of D_2O, and the pD of the solution was adjusted to 7.5 with DCl and NaOD. The D_2O exchanged enzyme solution was added to the mixture, and then *d6*-ATP (1.7 μmole) from a previous reaction, $NADP^+$ (0.003 g, 4.0 μmol), and 1',2',3',4',5',6',6-$^2H_7$-glucose (0.094 g, 0.5 mmol) were added to start the

reaction. After 40 hours a significant amount of the adenine had been converted into *d6*-ATP (**24**), and so guanine (0.025 g, 0.17 mmol) and orotic acid (0.032 g, 0.17 mmol) were added to the reaction to begin GTP and UTP formation. Additional D_2O exchanged 3-PGA (1 mmol) was added to the reaction after 3 days, and the reaction was stopped by freezing it after 4 days. The mixture was purified by boronate chromatography, and nucleotides were quantitated by UV absorbance at 260 nm, assuming a 1:1:1 ratio of ATP:GTP:UTP. Using this quantification, 95% of the $^2H_7$-glucose was converted into nucleotides and recovered. ^1H NMR (600 MHz, D_2O) δ 8.55 (s, 1H), 8.29 (s, 1H), 8.13 (s, 1H), 5.98 (s, 1H).

Deuteration of the H8 of Adenine

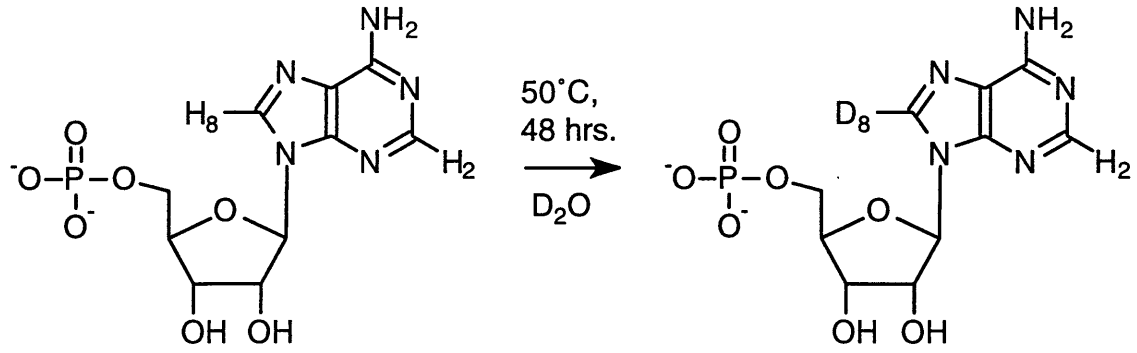


Figure 6.13 Exchange of the H8 of purines

Preparation of 8-$^2H_1$-ATP (40) from AMP. Adenosine-5'-monophosphate (0.5 g, 1.18 mmol) was placed into a 25 mL round bottom flask and exchanged 3 times with 3 mL of D_2O. The residue was dissolved in 4 mL of D_2O, a reflux condenser was attached to the flask, and the solution was heated to 70° C for 60 hours. A proton NMR spectra of the reaction at 60 hours revealed that the H8 proton had been exchanged with D_2O.

The D_2O was removed from the reaction under vacuum and the residue was dissolved in 100 mL of an H_2O solution which contained 10 mM $MgCl_2$, 10 mM

dithiothreitol, and 50 mM potassium phosphate buffer (pH ≈ 7.5). Sodium 3-phosphoglycerate (4.7 mmol) and a small catalytic amount of ATP (.001 g) were added to the mixture, and the pH of the solution was adjusted to 7.7 with 1M NaOH. Phosphoglycerate mutase (250 units), enolase (50 units), pyruvate kinase (100 units), and myokinase (125 units) were added to the mixture to start phosphorylation of the 8-$^{2}H_{1}$-AMP. After 30 hours the reaction was stopped by freezing. The 8-$^{2}H_{1}$-ATP (**40**) produced in this reaction was purified by boronate chromatography. 0.83 mmol of 8-$^{2}H_{1}$-ATP (**40**) (70% yield) were recovered from the boronate column. ^{1}H NMR (300 MHz, D_2O) δ 8.16 (s, 1H), 6.04 (d, 1H, J = 6.0 Hz), 4.80-3.90 (m, 5H).

Heteronuclear labeling of Bases ($^{13}C/^{15}N$)

Preparation of 6-amino,9-$^{15}N_2$-8,1',2',3',4',5'-$^{13}C_6$-1',3',4',5',5"-$^{2}H_5$-ATP (41), 7,9-$^{15}N_2$-8,1',2',3',4',5'-$^{13}C_6$-1',3',4',5',5"-$^{2}H_5$-GTP (42), and 1,3-$^{15}N_2$-2,4,5,6,1',2',3',4',5'-$^{13}C_9$-1',3',4',5',5"-$^{2}H_5$-UTP (43). Into a three neck flask were placed sodium 3-phosphoglycerate (8.0 mmol), 6-amino,9-$^{15}N_2$-8-$^{13}C_1$-adenine (0.035 g, 0.2 mmol), α-ketoglutaric acid (0.76 g, 4.0 mmol), and NH_4Cl (0.30 g, 5.6 mmol). This was dissolved in 80 mL of an H_2O solution containing 10 mM $MgCl_2$, 20 mM dithiothreitol, and 50 mM potassium phosphate buffer pH 7.5, and the pH of this solution was adjusted to 7.4 with 1M NaOH. After the solution had been neutralized, ATP (8 μmol), $NADP^+$ (0.007 g, 9 μmol), and 1,2,3,4,5,6-$^{13}C_6$-1,2,3,4,5,6,6'-$^{2}H_7$-D-glucose (0.155 g, 0.8 mmol) were added to the mixture. The reaction was started by adding 375 units of phosphoglycerate mutase, 200 units of enolase, 150 units of pyruvate kinase, 50 units of myokinase, 60 units of L-glutamic dehydrogenase, 120 units of hexokinase, 25 units of glucose-6-phosphate dehydrogenase, 5 units of 6-phosphogluconic dehydrogenase, 100 units of phosphoriboisomerase, 4 units of PRPP synthetase, 3 unit of adenine phosphoribosyltransferase, 4 units of xanthine-guanine phosphoribosyltransferase, 6 units

of uracil phosphoribosyltransferase, 1 unit of guanylate kinase, and 0.5 units of nucleoside monophosphate kinase. At 72 hours a significant amount of ATP had formed, and 7,9-$^{15}N_2$-8-$^{13}C_1$-guanine (0.043 g, 0.28 mmol), and 1,3-$^{15}N_2$-2,4,5,6-$^{13}C_4$-uracil (0.047 g, 0.4 mmol) were added to the reaction to begin GTP and UTP formation. After 7 days an additional 2.0 mmol of sodium 3-phosphoglycerate was added to the reaction, and the reaction was stopped after 11 days. Boronate purification of the reaction and quantification of the purified products by UV absorbance indicated that 80% of the glucose had been converted into isotopically labeled nucleotides and recovered. Characterization was conducted on the mixture of products resulting from this reaction. ^{1}H NMR (600 MHz, D_2O) δ 8.53 (dd, 1H, J = 216, 6 Hz), 8.26 (s, 1H), 8.11 (ddd, 1H, J = 216, 12, 12 Hz), 7.95 (dd, 1H, J = 186, 6 Hz), 5.97 (d, 1H, J = 180 Hz), 4.80 (d, 1H, J = 150 Hz), 4.78 (d, 1H, J = 156 Hz), 4.38 (d, 1H, J = 156).

Chapter 4 Experimental

Preparation of Isotopically Labeled RNA from Isotopically Labeled Nucleotides (Figure 6.14)

Isotopically labeled RNA was synthesized by *in vitro* transcription (Milligan & Uhlenbeck, 1989) with T7 RNA polymerase using slight variations of the procedure of Wyatt and Puglisi (Wyatt *et al.*, 1991). Small scale (20-25 µl) transcription trials containing trace amounts of [α-^{32}P] GTP (which allows detection of RNA transcripts with autoradiography) were used to find optimal transcription reaction conditions. RNA transcription reactions were optimized for the amount RNA produced per isotopically labeled nucleotide added to a transcription reaction because of the expense of isotopically labeled nucleotides. Transcription reactions were first optimized with commercially obtained unlabeled nucleotides to find optimal RNA transcription conditions, then small

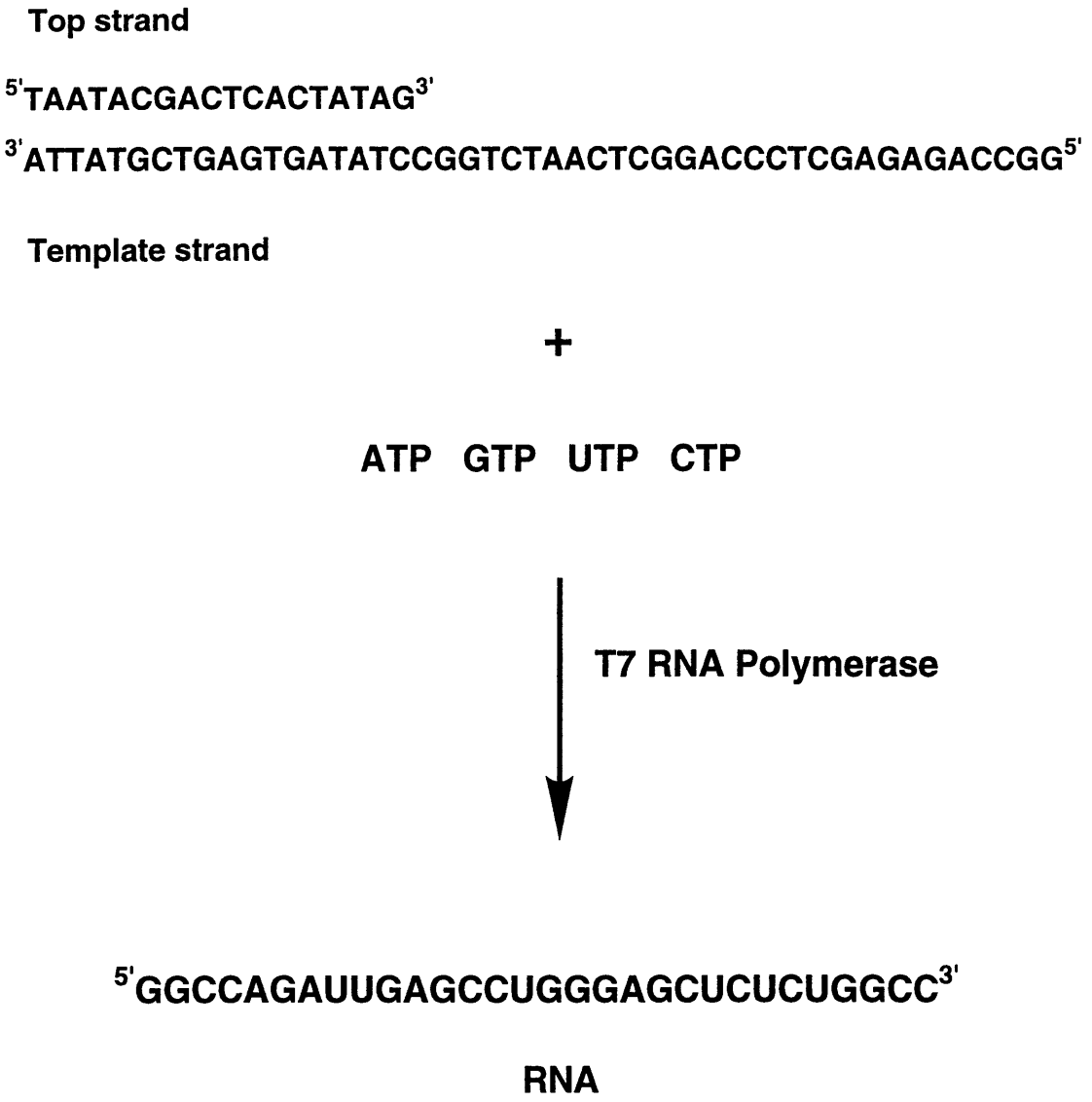


Figure 6.14 T7 RNA polymerase catalyzed transcription

scale transcription trials were conducted using the same transcription conditions for both unlabeled and isotopically labeled nucleotides to compare transcriptional efficiencies of the nucleotides derived from different sources. Once similar yields were obtained for both isotopically labeled nucleotides and commercially obtained nucleotides in RNA transcription trials, large scale transcription reactions were conducted as described below.

TAR RNA synthesis. TAR RNA (5'GGC CAG AUU GAG CCU GGG AGC UCU CUG GCC3') was synthesized by *in vitro* transcription (Milligan & Uhlenbeck, 1989) with T7 RNA polymerase using unlabeled NTPs from Sigma and the different isotopically labeled NTPs produced in this thesis. Transcription reactions were conducted in a solution which contained 40 mM Tris-HCl buffer (pH=8.1), 1 mM spermidine, 0.01% triton X-100, 12 mM dithiothreitol, 80 mg/mL polyethylene glycol, 35 mM $MgCl_2$, 450 nM top strand DNA, and 450 nM TAR template DNA. The procedures of Wyatt and Puglisi (Wyatt *et al.*, 1991) were followed for conducting transcription reactions. TAR RNA was purified by 20% PAGE, electro-eluted, and dialyzed against 50 mM NaCl, 0.1 mM EDTA, 10 mM sodium phosphate pH 6.4. For each labeling pattern produced in Chapter 4, the nucleotides used to produce it, the amount of each nucleotide used, and the concentration and volume of the final NMR sample produced from the transcription reaction is listed below.

Unlabeled TAR RNA. Unlabeled TAR RNA was produced in a 50 mL transcription reaction using unlabeled ATP (200 μmoles), unlabeled GTP (200 μmoles), unlabeled UTP (200 μmoles), and unlabeled CTP (200 μmoles). The final NMR sample produced from this transcription reaction was 0.83 mM, 500 μliters.

D4-TAR RNA. D4-TAR RNA was produced in a 45 mL transcription reaction using 3',4',5',5"-*d4*-ATP (**2**) (90 μmoles), 3',4',5',5"-*d4*-GTP (**3**) (90 μmoles),

3',4',5',5"-*d4*-UTP (**4**) (90 μmoles), and 3',4',5',5"-*d4*-CTP (**5**) (90 μmoles). The final NMR sample produced from this transcription reaction was 1.1 mM, 500 μliters.

D6-TAR RNA. D6-TAR RNA was produced in a 40.5 mL transcription reaction using 1',2',3',4',5',5"-*d6*-ATP (**24**) (81 μmoles), 1',2',3',4',5',5"-*d6*-GTP (**25**) (81 μmoles), 1',2',3',4',5',5"-*d6*-UTP (**26**) (81 μmoles), and 1',2',3',4',5',5"-*d6*-CTP (**27**) (81 μmoles). The final NMR sample produced from this transcription reaction was 1.6 mM, 500 μliters.

TAR-465 RNA. TAR-465 RNA was produced in a 40.5 mL transcription reaction using 1',2',3',4',5',5"-*d6*-ATP (**24**) (81 μmoles), 1',2',3',4',5',5"-*d6*-GTP (**25**) (81 μmoles), 1',2',3',4',5',5",6-*d7*-UTP (**39**) (81 μmoles), and unlabeled CTP (162 μmoles). The final NMR sample produced from this transcription reaction was 1.4 mM, 550 μliters.

D4/D6-TAR RNA. D4/D6-TAR RNA was produced in a 14 mL transcription reaction using 1',2',3',4',5',5"-*d6*-ATP (**24**) (22 μmoles), 1',2',3',4',5',5"-*d6*-GTP (**25**) (22 μmoles), 3',4',5',5"-*d4*-UTP (**4**) (31 μmoles), and 1',2',3',4',5',5"-*d6*-CTP (**27**) (43 μmoles). The final NMR sample produced from this transcription reaction was 0.46 mM, 300 μliters.

D5-pyrimidine TAR RNA. D5-pyrimidine TAR RNA was produced in a 40 mL transcription reaction using unlabeled ATP (87 μmoles), unlabeled GTP (80 μmoles), 5-*d1*-UTP (**37**) (79 μmoles), and 5-*d1*-CTP (**38**) (79 μmoles). The final NMR sample produced from this transcription reaction was 1.1 mM, 500 μliters.

***$^{13}C_5$-ribose*-TAR RNA.** *$^{13}C_5$-ribose*-TAR RNA was produced in a 40 mL transcription reaction using 1',2',3',4',5'-$^{13}C_5$-ATP (**28**) (80 μmoles), 1',2',3',4',5'-$^{13}C_5$-GTP (**29**) (80 μmoles), 1',2',3',4',5'-$^{13}C_5$-UTP (**30**) (80 μmoles), and

1',2',3',4',5'-$^{13}C_5$-CTP (**31**) (80 μmoles). The final NMR sample produced from this transcription reaction was 1.7 mM, 600 μliters.

***3',4',5',5''-D4-$^{13}C_5$-ribose*-TAR RNA.** *3',4',5',5''-D4-$^{13}C_5$-ribose*-TAR RNA was produced in a 40 mL transcription reaction using 3',4',5',5''-*d4*-1',2',3',4',5'-$^{13}C_5$-ATP (**32**) (80 μmoles), 3',4',5',5''-*d4*-1',2',3',4',5'-$^{13}C_5$-GTP (**33**) (80 μmoles), 3',4',5',5''-*d4*-1',2',3',4',5'-$^{13}C_5$-UTP (**34**) (80 μmoles), and 3',4',5',5''-*d4*-1',2',3',4',5'-$^{13}C_5$-CTP (**35**) (80 μmoles). The final NMR sample produced from this transcription reaction was 1.1 mM, 600 μliters.

TAR NMR experiments.

Unlabeled TAR and D4-TAR RNA: NOESY spectra of TAR and D4-TAR were recorded on a Varian VXR-500 spectrometer with a spectral width of 5500 Hz, acquiring 4096 complex points in t_2 and 644 in t_1, 64 or 32 scans per FID for TAR and D4-TAR respectively, a relaxation delay of 1.6 s, and a mixing time of 0.200 s. Sample conditions were 50 mM NaCl, 0.1 mM EDTA, 10mM sodium phosphate pH 6.4, 25°. RNA concentrations were 0.83 mM for TAR and 1.12 mM for D4-TAR. Twice the number of scans were acquired per FID for the unlabeled TAR sample to compensate for the difference in concentrations between the two RNA samples.

Nonselective T_1 relaxation rates for individual resonances were measured using inversion recovery, and nonselective T_2 relaxation rates were measured using a CPMG experiment. For both T_1 and T_2 measurements a relaxation delay of 75 seconds between each of 4 scans was used.

All other TAR NMR experiments. All other TAR NMR experiments were recorded on a Varian Inova-600 MHz spectrometer. RNA sample conditions were 50 mM

NaCl, 0.1 mM EDTA, 10mM sodium phosphate pH 6.4, and experiments were conducted at 25° C. NOESY spectra were acquired with a spectral width of 8000 Hz, acquiring 4096 complex points in t_2 and 1024 in t_1, with 32 scans per FID, a relaxation delay of 2.6 s, and a mixing time of 200 ms. HSQC-CT spectra were acquired with a constant time interval of 25 ms with spectral widths of 6500 and 5000 Hz for proton and carbon dimensions respectively. 1024 and 224 complex points were acquired for t_1 and t_2 respectively, with 8 scans per FID.

Chapter 5 Experimental

F22b RNA transcription and NMR sample preparation

F22b RNA synthesis. The 65 nucleotide F22b RNA (5'GGG UUG CGG UCG AAA GAC UUG AGG GCA GGA GAG GAC UUC GGU CUG GCC UGC ACC UGA CGC AAC CC3') was synthesized by *in vitro* transcription with T7 RNA polymerase using unlabeled NTPs from Sigma and the different isotopically labeled NTPs produced in this thesis. Transcription reactions were conducted in a solution which contained 40 mM Tris-HCl buffer (pH=8.1), 1 mM spermidine, 0.01% triton X-100, 12 mM dithiothreitol, 8 mM $MgCl_2$, and 63 μg/mL of SmaI digested pB16SF22b template DNA. The procedures of Wyatt and Puglisi were followed for conducting transcription reactions. F22b RNA was purified by 20% PAGE, electro-eluted, and dialyzed against 50 mM KCl, 0.1 mM EDTA, 10 mM potassium phosphate pH 6.5. For each F22b labeling pattern produced in Chapter 5, the nucleotides used to produce it, the amount of each nucleotide used, and the concentration and volume of the final NMR sample produced from the transcription reaction is listed below.

Unlabeled F22b RNA. Unlabeled F22b RNA was produced in a 40 mL transcription reaction using unlabeled ATP (80 μmoles), unlabeled GTP (80 μmoles),

unlabeled UTP (80 μmoles), and unlabeled CTP (80 μmoles). The final NMR sample produced from this transcription reaction was 0.89 mM, 500 μliters.

D5-pyrimidine F22b RNA. D5-pyrimidine F22b RNA was produced in a 40 mL transcription reaction using unlabeled ATP (87 μmoles), unlabeled GTP (80 μmoles), 5-*d1*-UTP (**37**) (80 μmoles), and 5-*d1*-CTP (**38**) (80 μmoles). The final NMR sample produced from this transcription reaction was 0.83 mM, 500 μliters.

D4-F22b RNA. D4-F22b RNA was produced in a 40 mL transcription reaction using 3',4',5',5"-*d4*-ATP (**2**) (86 μmoles), 3',4',5',5"-*d4*-GTP (**3**) (88 μmoles), 3',4',5',5"-*d4*-UTP (**4**) (88 μmoles), and 3',4',5',5"-*d4*-CTP (**5**) (80 μmoles). The final NMR sample produced from this transcription reaction was 0.82 mM, 500 μliters.

D4/D6-F22b RNA. D4/D6-F22b RNA was produced in a 33 mL transcription reaction using 1',2',3',4',5',5"-*d6*-ATP (**24**) (51 μmoles), 1',2',3',4',5',5"-*d6*-GTP (**25**) (51 μmoles), 3',4',5',5"-*d4*-UTP (**4**) (72 μmoles), and 1',2',3',4',5',5"-*d6*-CTP (**27**) (102 μmoles). The final NMR sample produced from this transcription reaction was 0.44 mM, 500 μliters.

NMR experiments

F22b NMR experiments. NMR experiments were recorded on a Varian Inova-600 MHz spectrometer. RNA sample conditions were 50 mM KCl, 0.1 mM EDTA, 10mM potassium phosphate pH 6.5, and experiments were conducted at 35° C. NOESY spectra were acquired with a spectral width of 8000 Hz, acquiring 4096 complex points in t_2 and 1024 in t_1, with 32 scans per FID, a relaxation delay of 2.6 s, and a mixing time of 200 ms.

2D CCH-TOCSY NMR experiments. The 2D CCH-TOCSY NMR experiments were conducted as described in Dayie *et. al.* (Dayie *et al.*, 1997) on both *$^{13}C_5$-ribose*-TAR RNA and *3',4',5',5''-D4-$^{13}C_5$-ribose*-TAR RNA.

Chapter 6

List of Improvements

While the procedures in this thesis work fairly well, there is always room for improvement...

1. Several times during the course of these studies the bacteria containing the plasmid for CTP synthetase overexpression and ampicillin resistance, plasmid pMW5, apparently lost the plasmid stopped producing CTP synthetase. A plasmid construct with a antibiotic resistance which would be more difficult to lose than ampicillin resistance would be a good improvement.

2. Several of the enzymes which were purified are only moderately overexpressed. Construction of plasmids which would highly overexpress PRPP synthetase, orotate phosphoribosyltransferase, and CTP synthetase would increase the yield of those enzymes in protein purifications and potentially reduce the amount of work necessary to produce isotopically labeled NTPs.

3. The rates of many of the enzymatic reactions discussed in this thesis are slow because of the small amount of catalytic ATP added to the reactions to prevent isotopic dilution of the final nucleotides. Developing methods to add larger initial amounts of ATP to the reactions and then being able to remove that ATP later would potentially speed up the rate of NTP formation. This might be done by using deoxy-ATP (which some of the enzymes will accept as a substrate) and then removing that dATP with a boronate column. It also might be done in UTP and GTP forming reactions by adding a large amount of ATP to catalyze UTP and GTP formation, and then adding an enzyme which will selectively dephosphorylate ATP to form AMP. This would allow the separation of unlabeled AMP from isotopically labeled UTP and GTP by anion exchange chromatography.

4. Developing methods to add larger initial amounts of $NADP^+$ to the glucose nucleotide forming reactions and then being able to remove that $NADP^+$ later would potentially speed up the rate of NTP formation.

5. Base protons on adenine, guanine and uracil could be exchanged in D_2O before the bases were coupled to PRPP. This could reduce the number of chemical steps required to produce nucleotides with deuterium labels in both the bases and the ribose moieties.

6. Developing some method that would allow CTP to be produced in the ribose and glucose nucleotide forming reactions would reduce the number of enzymatic reactions necessary to produced isotopically labeled RNA.

7. Warming the enzymatic nucleotide forming reactions to 37° C may speed the rate of product formation without any bad side effects.

8. For most of the enzymes purified in this thesis, LB media will probably work just as well or better than minimal media for purification of overexpressed enzymes.

9. Addition of antibiotics like ampicillin to all of the enzymatic nucleotide forming reactions may reduce bacterial contamination problems.

10. Ribulose phosphate 3-epimerase might be used to exchange the H3' with H_2O or D_2O in the reactions where nucleotides are enzymatically converted into NTPs.

References

Aboul-ela, F., Karn, J. & Varani, G. (1995) "The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein."*J. Mol. Biol.* 253, 313-332.

Agback, P., Maltseva, T. V., Yamakage, S.-I., Nilson, F. P. R., Földesi, A. & Chattopadhyaya, J. (1994) "The differences in the T_2 relaxation rates of the protons in the partially-deuteriated and fully protonated sugar residues in a large oligo-DNA ('NMR-window') gives complementary structural information."*Nucleic Acids Res.* 22, 1404-1412.

Allain, F. H., Gubser, C. C., Howe, P. W., Nagai, K., Neuhaus, D. & Varani, G. (1996) "Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation."*Nature* 380, 646-650.

Andersen, P. S., Smith, J. M. & Mygind, B. (1992) "Characterization of the *upp* gene encoding uracil phosphoribosyltransferase of *Escherichia coli* K12."*Eur. J. Biochem.* 204, 51-56.

Anderson, P. M. (1983) "CTP Synthetase from *Escherichia coli:* An Improved Purification Procedure and Characterization of Hysteretic and Enzyme Concentration Effects on Kinetic Properties."*Biochemistry* 22, 3285-3292.

Arnott, S. & Hukins, D. W. L. (1973) "Structures of Synthetic Polynucleotides in the A-RNA and A'-RNA Conformations: X-ray Diffraction Analyses of the Molecular Conformations of Polyadenylic acid-Polyuridylic acid and Polyinosinic acid-Polycytidylic acid."*J. Mol. Biol.* 81, 107-122.

Chapter 6

Baldwin, J. E. & Black, K. A. (1983) "Stereoselective Preparation of *trans* -2,3-Dideuterioprop-2-en-1-ol."*J. Org. Chem.* 48, 2778-2779.

Batey, R. T., Battiste, J. L. & Williamson, J. R. (1995) "Preparation of Isotopically Enriched RNAs for Heteronuclear NMR."*Methods Enzymol.* 261, 300-322.

Batey, R. T., Cloutier, N., Mao, H. & Williamson, J. R. (1996) "Improved large scale culture of *Methylophilus methylotrophus* for $^{13}C/^{15}N$ labeling and random fractional deuteration of ribonucleotides."*Nucleic Acids Res.* 24, 4836-4837.

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 Res.* 20, 4515-4523.

Batey, R. T. & Williamson, J. R. (1996a) "Interaction of the *Bacillus stearothermophillus* Ribosomal Protein S15 with 16S rRNA: I. Defining the Minimal RNA Site."*J. Mol. Biol.* 261, 536-549.

Batey, R. T. & Williamson, J. R. (1996b) "Interaction of the *Bacillus stearothermophilus* Ribosomal Protein S15 with 16S rRNA: II. Specificity Determinants of RNA-Protein Recognition."*J. Mol. Biol.* 261, 550-567.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996) "α Helix-RNA Major Groove Recognition in an HIV-1 Rev Peptide-RRE RNA Complex."*Science* 273, 1547-1551.

Benevides, J. M., Lemeur, D. & Thomas, G. J. (1984) "Molecular Conformations and 8-CH Exchange Rates of Purine Ribo- and Deoxyribonucleotides: Investigation by Raman Spectroscopy."*Biopolymers* 23, 1011-1024.

Bjarne, H.-J. (1985) "Cloning and characterization of the *prs* gene encoding phosphoribosylpyrophosphate synthetase of *Escherichia coli*."*Mol. Gen. Genet.* 201, 269-276.

Brandes, R. & Ehrenberg, A. (1986) "Kinetics of the proton-deuteron exchange at position H8 of adenine and guanine in DNA."*Nucleic Acids Res.* 14, 9491-9508.

Brodsky, A. S. & Williamson, J. R. (1997) "Solution Structure of the HIV-2 TAR-Argininamide Complex."*J. Mol. Biol.* 267, 624-639.

Brush, C. K., Stone, M. P. & Harris, T. M. (1988) "Selective Reversible Deuteration of Oligodeoxynucleotides: Simplification of Two-Dimensional Nuclear Overhauser Effect NMR Spectral Assignment of a Non-Self-Complementary Dodecamer Duplex."*Biochemistry* 27, 115-122.

Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D. & Frankel, A. D. (1991) "Arginine-Mediated RNA Recognition: The Arginine Fork."*Science* 252, 1167-1171.

Carrington, A. & McLachlan, A. D. (1967) "Introduction to Magnetic Resonance." in *Harper's Chemistry Series* (S. A. Rice, eds) pp. 266, Harper & Row, New York.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G. & Skelton, N. J. (1996) "Protein NMR Spectroscopy Principles and Practice." pp. 587, Academic Press, San Diego.

Cech, T. R. (1990) "Self-splicing and Enzymatic Activity of an Intervening Sequence RNA from *Tetrahymena* (Nobel Lecture)."*Angew. Chem. Int. Ed. Engl.* 29, 759-768.

Cha, J. K., Christ, W. J. & Kishi, Y. (1984) "On Stereochemistry of Osmium Tetraoxide oxidation of Allylic Alcohol Systems: Empirical Rule."*Tetrahedron* 40, 2247-2255.

Chang, K.-Y. & Varani, G. (1997) "Nucleic acids structure and recognition."*Nature Structural Biology* 4, 854-858.

Choi, B.-S. & Redfield, A. G. (1992) "NMR Study of Nitrogen-15-Labeled *Escherichia coli* Valine Transfer RNA."*Biochemistry* 31, 12799-12802.

Clark, D. J. & Maaløe, O. (1967) "DNA Replication and the Division Cycle in *Escherichia coli*."*J. Mol. Biol.* 23, 99-112.

Clore, G. M. & Gronenborn, A. M. (1994) "Multidimensional Heteronuclear Nuclear Magnetic Resonance of Proteins."*Methods Enzymol.* 239, 349-363.

Clore, G. M. & Gronenborn, A. M. (1997) "NMR structures of proteins and protein complexes beyond 20,000 M_r."*Nature Structural Biology* 4, 849-853.

Cook, G. P. & Greenberg, M. M. (1994) "A General Synthesis of C2'-Deuteriated Ribonucleosides."*J. Org. Chem.* 59, 4704-4706.

Crespi, H. L., Rosenberg, R. M. & Katz, J. J. (1968) "Proton Magnetic Resonance of Proteins Fully Deuterated except for ^{1}H-Leucine Side Chains."*Science* 161,

Dayie, K. T., Tolbert, T. J. & Williamson, J. R. (1997) "3D C(CC)H TOCSY Experiment for Assigning Protons and Carbons in Uniformly ^{13}C and Selectively ^{2}H Labeled RNA."*J. Magn. Reson., in press.*

Derome, A. E. (1990) "Modern NMR Techniques for Chemical Research." in *Organic Chemistry Series* (J. E. Baldwin, eds) pp. 280, Pergamon Press, Oxford.

Dieckmann, T. & Feigon, J. (1994) "Heteronuclear techniques in NMR studies of RNA and DNA."*Curr. Opin. Struct. Biol.* 4, 745-749.

Dieckmann, T., Suzuki, E., Nakamura, G. K. & Feigon, J. (1996) "Solution Structure of an ATP-binding RNA Aptamer Reveals a Novel Fold."*RNA* 2, 628-640.

Doudna, J. A. & Cate, J. H. (1997) "RNA structure: crystal clear?"*Curr. Opin. Struct. Biol.* 7, 310-316.

Farmer, B. T. & Venters, R. A. (1995) "Assignment of Side-Chain ^{13}C Resonances in Perdeuterated Proteins."*J. Am. Chem. Soc.* 117, 4187-4188.

Földesi, A., Nilson, F. P. R., Glemarec, C., Gioeli, C. & Chattopadhyaya, J. (1992) "Synthesis of 1'#, 2', 3', 4'#, 5', 5"-$^{2}H_{6}$-β-D-ribonucleosides and 1'#, 2', 3', 4'#, 5', 5"-$^{2}H_{7}$-β–D-2'-deoxyribonucleosides for Selective Suppression of Proton Resonances in Partially-deuterated Oligo-DNA, Oligo-RNA and in 2,5A core (^{1}H-*NMR window*)."*Tetrahedron* 48, 9033-9072.

Fourmy, D., Recht, M. I., Blanchard, S. C. & Puglisi, J. D. (1996) "Structure of the A Site of *Escherichia coli* 16S Ribosomal RNA Complexed with an Aminoglycoside Antibiotic."*Science* 274, 1367-1371.

Frank, B. L., Worth, L. J., Christner, D. F., Kozarich, J. W., Stubbe, J., Kappen, L. S. & Goldberg, I. H. (1991) "Isotope Effects on the Sequence-Specific Cleavage of DNA by Neocarzinostatin: Kinetic Partitioning between 4'- and 5'-Hydrogen Abstraction at Unique Thymidine Sites."*J. Am. Chem. Soc.* 113, 2271-2275.

Gilles, A.-M., Cristea, I., Palibroda, N., Hilden, I., Jensen, K. F., Sarfati, R. S., Namane, A., Ughetto-Monfrin, J. & Bârzu, O. (1995) "Chemienzymatic Synthesis of Uridine Nucleotides Labeled with [^{15}N] and [^{13}C]."*Anal. Biochem.* 232, 197-203.

Glemarec, C., Kufel, J., Földesi, A., Maltseva, T., Sandström, A., Kirsebom, L. A. & Chattopadhyaya, J. (1996) "The NMR structure of 31mer RNA domain of *Escherichia coli* RNase P RNA using its non-uniformly deuterium labelled counterpart [the 'NMR-window' concept]."*Nucleic Acids Res.* 24, 2022-2035.

Greenbaum, N. L., Radhakrishnan, I., Hirsh, D. & Patel, D. J. (1995) "Determination of the Folding Topology of the SL1 RNA from *Caenorhabditis elegans* by Multidimensional Heteronuclear NMR."*J. Mol. Biol.* 252, 314-327.

Gross, A., Abril, O., Lewis, J. M., Geresh, S. & Whitesides, G. M. (1983) "Practical Synthesis of 5-Phospho-D-ribosyl α-1-Pyrophosphate (PRPP): Enzymatic Routes from Ribose 5-Phosphate or Ribose."*J. Am. Chem. Soc.* 105, 7428-7435.

Grzesiek, S., Anglister, J., Ren, H. & Bax, A. (1993) "^{13}C Line Narrowing by ^{2}H Decoupling in $^{2}H/^{13}C/^{15}N$-Enriched Proteins. Application to Triple Resonance 4D J Connectivity of Sequential Amides."*J. Am. Chem. Soc.* 115, 4369-4370.

Gutell, R. R., Weiser, B., Woese, C. R. & Noller, H. F. (1985) "Comparative Anatomy of 16-S-like Ribosomal RNA."*Progress in Nucleic Acid Research* 32, 155-216.

Hall, K. B. (1995) "Uses of ^{13}C- and ^{15}N-Labeled RNA in NMR of RNA-Protein Complexes."*Methods Enzymol.* 261, 542-559.

Hall, K. B., Sampson, J. R., Uhlenbeck, O. C. & Redfield, A. G. (1989) "Structure of an Unmodified tRNA Molecule."*Biochemistry* 28, 5794-5801.

Hershey, H. V. & Taylor, M. W. (1986) "Nucleotide sequence and deduced amino acid sequence of *Escherichia coli* adenine phosphoribosyltransferase and comparison with other analogous enzymes."*Gene* 43, 287-293.

Hirschbein, B. L., Mazenod, F. P. & Whiteside, G. M. (1982) "Synthesis of Phosphoenolpyruvate and Its Use in Adenosine Triphosphate Cofactor Regeneration."*J. Org. Chem.* 47, 3765-3766.

Hoffman, D. W. & Holland, J. A. (1995) "Preparation of carbon-13 labeled ribonucleotides using acetate as an isotope source."*Nucleic Acids Res.* 23, 3361-3362.

Hope, J. N., Bell, A. W., Hermodson, M. A. & Groarke, J. M. (1986) "Ribokinase from *Escherichia coli* K12."*J. Biol. Chem.* 261, 7663-7668.

Hsu, V. L. & Armitage, I. M. (1992) "Solution Structure of Cyclosporin A and a Nonimmunosuppressive Analog Bound to Fully Deuterated Cyclophilin."*Biochemistry* 31, 12778-12784.

Isler, O., Montavon, M., Rüegg, R., Ryser, G. & Zeller, P. (1957) "Synthesen in der Carotinoid-Reihe Anwendun der Wittig-Reaktion zur synthese von Estern des Bixins und Crocetins."*Helvetica Chimica Acta* 139, 1242-1249.

Jackson, D. Y. (1988) "An Improved Preparation of (+)-2,3-O-isopropylidene-D-glyceraldehyde."*Synthetic Communications* 18, 337-341.

Jensen, K. F., Andersen, J. T. & Poulsen, P. (1992) "Overexpression and Rapid Purification of the *orfE/rph* Gene Product, RNase PH of *Escherichia coli*."*J. Biol. Chem.* 267, 17147-17152.

Jiang, F., Kumar, R. A., Jones, R. A. & Patel, D. J. (1996) "Structural Basis of RNA Folding and Recognition in an AMP-RNA Aptamer Complex."*Nature* 382, 183-186.

Jucker, F. M. & Pardi, A. (1995) "Solution Structure of the CUUG Hairpin Loop: A Novel RNA Tetraloop Motif."*Biochemistry* 34, 14416-14427.

Kainosho, M. (1997) "Isotope labelling of macromolecules for structural determinations."*Nature Structural Biology* 4, 858-861.

Karplus, M. (1963) "Vicinal Proton Coupling in Nuclear Magnetic Resonance."*J. Am. Chem. Soc.* 85, 2870-2871.

Katsuki, T., Lee, A. W. M., Ma, P., Martin, V. S., Masamune, S., Sharpless, K. B., Tuddenham, D. & Walker, F. J. (1982) "Synthesis of Saccharides and Related

Polyhydroxylated Natural Products. 1. Simple Alditols."*J. Org. Chem.* 47, 1373-1378.

Kay, L. E. & Gardner, K. H. (1997) "Solution NMR spectroscopy beyond 25 kDa."*Current Opinion in Structural Biology* 7, 722-731.

Klausner, R. D., Rouault, T. A. & Harford, J. B. (1993) "Regulating the fate of mRNA: the control of cellular iron metabolism."*Cell* 72, 19-28.

Kline, P. C. & Serianni, A. S. (1990) "^{13}C-Enriched Ribonucleosides: Synthesis and Application of ^{13}C-^{1}H and ^{13}C-^{13}C Spin-Coupling Constants To Assess Furanose and *N*-Glycoside Bond Conformations."*J. Am. Chem. Soc.* 112, 7373-7381.

Klunder, J. M., Ko, S. Y. & Sharpless, K. B. (1986) "Asymmetric Epoxidation of Allyl Alcohol: Efficient Routes to Homochiral β–Adrenergic Blocking Agents."*J. Org. Chem.* 51, 3710-3712.

Ko, Y. S., Lee, A. W. M., Masamune, S., Reed, L. A. I., Sharpless, K. B. & Walker, F. J. (1990) "Total Synthesis of the L-Hexoses."*Tetrahedron* 46, 245-264.

Lancelot, G., Chanteloup, L., Beau, J. & Thuong, N. T. (1993) "Selectively ^{13}C-Enriched DNA: ^{13}C and ^{1}H Assignments of the Lac Operator by Two-Dimensional Relayed HMQC Experiments."*J. Am. Chem. Soc.* 115, 1599-1600.

LeMaster, D. M. (1990) "Deuterium labelling in NMR structural analysis of larger proteins."*Quart. Rev. Biophys.* 23, 133-174.

LeMaster, D. M. & Richards, F. M. (1988) "NMR Sequential Assignment of *Escherichia coli* Thioredoxin Utilizing Random Fractional Deuteriation."*Biochemistry* 27, 142-150.

Lienhard, G. E. & Rose, I. A. (1964) "The Mechanism of Action of 6-Phosphogluconate Dehydrogenase."*Biochemistry* 3, 190-195.

Long, C. W. & Pardee, A. B. (1967) "Cytidine Triphosphate Synthetase of *Escherichia coli* B."*J. Biol. Chem.* 242, 4715-4721.

Madhani, H. D. & Guthrie, C. (1992) "A Novel Base-Pairing Interaction between U2 and U6 snRNAs Suggests a Mechanism for the Catalytic Activation of the Spliceosome."*Cell* 71, 803-817.

Marino, J. P., Gregorian, R. S., Csankovszki, G. & Crothers, D. M. (1995) "Bent Helix Formation Between RNA Hairpins with Complementary Loops."*Science* 268, 1448-1454.

Markley, J. L., Putter, I. & Jardetzky, O. (1968) "High-Resolution Nuclear Magnetic Resonance Spectra of Selectively Deuterated Staphylococcal Nuclease."*Science* 161, 1249-1251.

Markus, M. A., Dayie, K. T., Matsudaira, P. & Wagner, G. (1994) "Effect of Deuteration on the Amide Proton Relaxation Rates in Proteins. Heteronuclear NMR Experiments on Villin 14T."*J. Magn. Reson., B* 105, 192-195.

Matsuo, H., Li, H. & Wagner, G. (1996) "A Sensitive HN(CA)CO Experiment for Deuterated Proteins."*J. Magn. Reson., B* 110, 112-115.

McMichael, K. D. (1967) "Secondary Deuterium Isotope Effects on a Cyclic Allylic Rearrangement."*JACS* 89, 2943-2947.

McPheeters, D. S. & Abelson, J. (1992) "Mutational Analysis of the Yeast U2 snRNA Suggests a Structural Similarity to the Catalytic Core of Group I Introns."*Cell* 71, 819-831.

Milligan, J. F. & Uhlenbeck, O. C. (1989) *Methods Enzymol.* 180, 51-62.

Muhandiram, D. R., Yamazaki, T., Sykes, B. D. & Kay, L. E. (1995) "Measurement of ^{2}H T_1 and $T_{1\rho}$ Relaxation Times in Uniformly ^{13}C-Labeled and Fractionally ^{2}H-Labeled Proteins in Solution."*J. Am. Chem. Soc.* 117, 11536-11544.

Nietlispach, D., Clowes, R. T., Broadhurst, R. W., Ito, Y., Keeler, J., Kelly, M., Ashurst, J., Oschkinat, H., Domaille, P. J. & Laue, E. D. (1996) "An Approach to the Structure Determination of Larger Proteins Using Triple Resonance NMR Experiments in Conjuction with Random Fractional Deuteration."*J. Am. Chem. Soc.* 118, 407-415.

Nikonowicz, E. P., Michnicka, M., Kalurachchi, K. & DeJong, E. (1997) "Preparation and characterization of a uniformly $^{2}H/^{15}N$-labeled RNA oligonucleotide for NMR studies."*Nucleic Acids Res.* 25, 1390-1396.

Nikonowicz, E. P., Sirr, A., Legault, P., Jucker, F. M., Baer, L. M. & Pardi, A. (1992) "Preparation of ^{13}C and ^{15}N labelled RNAs for heteronuclear multi-dimensional NMR studies."*Nucleic Acids Res.* 20, 4507-4513.

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. J. (1996) "Conformation of an RNA Molecule That Models the P4/P6 Junction from Group I Introns."*Biochemistry* 35, 2577-2585.

Olsen, H. S., Cochrane, A. W., Dillon, P. J., Nalin, C. M. & Rosen, C. A. (1990) *Genes Dev.* 4, 1357.

Pace, N. R. "Biology Seminar, Massachusetts Institute of Technology.", 1993.

Pardi, A. (1995) "Multidimensional Heteronuclear NMR Experiments for Structure Determination of Isotopically Labeled RNA."*Methods Enzymol.* 261, 350-380.

Parkin, D. W., Leung, H. B. & Schramm, V. L. (1984) "Synthesis of Nucleotides with Specific Radiolabels in Ribose."*J. Biol. Chem.* 259, 9411-9417.

Parkin, D. W. & Schramm, V. L. (1987) "Catalytic and Allosteric Mechanism of AMP Nuclosidase from Primary, β-Secondary, and Multiple Heavy Atom Kinetic Isotope Effects."*Biochemistry* 26, 913-920.

Perrin, D. D. & Armarego, W. L. (1988) "Purification of Laboratory Chemicals." pp. 391 Pergamon Press Inc.

Peterson, R. D., Bartel, D. P., Szostak, J. W., Horvath, S. J. & Feigon, J. (1994) "1H NMR Studies of the High-Affinity Rev Binding Site of the Rev Responsive Element of HIV-1 mRNA: Base Pairing in the Core Binding Element."*Biochemistry* 33, 5357-5366.

Poulsen, P., Jensen, K. F., Valentin-Hansen, P., Carlsson, P. & Lundberg, L. G. (1983) "Nucleotide sequence of the *Escherichia coli pyrE* gene and of the DNA in front of the protein-coding region."*Eur. J. Biochem.* 135, 223-229.

Pratt, D. & Subramani, S. (1983) "Nucleotide sequence of the *Escherichia coli* xanthine-guanine phosphoribosyl transferase gene."*Nucleic Acids Res.* 11, 8817-8823.

Puglisi, E. V., Green, R., Noller, H. & Puglisi, J. D. (1997) "Structure of a conserved RNA component of the peptidyl transferase centre."*Nature Structural Biology* 4, 775-778.

Puglisi, J. D. (1989) "RNA Folding: Structure and Conformational Equilibria of RNA Pseudoknots.", Thesis, University of California at Berkeley.

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., Tan, R., Calnan, B. J., Frankel, A. D. & Williamson, J. R. (1992) "Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy."*Science* 257, 76-80.

Puglisi, J. D., Wyatt, J. R. & Tinoco, I. (1990) "Conformation of an RNA Pseudoknot."*J. Mol. Biol.* 214, 437-453.

Rabi, J. A. & Fox, J. J. (1973) "Nucleosides. LXXIX. Facile Base-Catalyzed Hydrogen Isotope Labeling at Position 6 of Pyrimidine Nucleosides."*J. Am. Chem. Soc.* 95, 1628-1632.

Ramos, A., Gubser, C. C. & Varani, G. (1997) "Recent solution structures of RNA and its complexes with drugs, peptides and proteins."*Curr. Opin. Struct. Biol.* 7, 317-323.

Rasmussen, U. B., Mygind, B. & Nygaard, P. (1986) "Purification and some properties of uracil phosphoribosyltransferase from *Escherichia coli* K12."*Bioch. et Biophys. Acta* 881, 268-275.

Recht, M. I., Fourmy, D., Blanchard, S. C., Dahlquist, K. D. & Puglisi, J. D. (1996) "RNA Sequence Determinants for Aminoglycoside Binding to an A-site rRNA Model Oligonucleotide."*J. Mol. Biol.* 262, 421-436.

Rising, K. A. & Schramm, V. L. (1994) "Enzymatic Synthesis of NAD^+ with the Specific Incorporation of Atomic Labels."*J. Am. Chem. Soc.* 116, 6531-6536.

Robins, M. J., Samano, V. & Johnson, M. D. (1990) "Periodinane Oxidation, Selective Primary Deprotection, and Remarkably Stereoselective Reduction of *tert*-

Butyldimethylsilyl-Protected Ribonucleosides. Synthesis of 9-(β-D-Xylofuranosyl)adenine or 3'-Deuterioadenosine from Adenosine."*J. Org. Chem.* 55, 410-412.

Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T. & Kay, L. E. (1996) "Selective Methyl Group Protonation of Perdeuterated Proteins."*J. Mol. Biol.* 263, 627-636.

SantaLucia, J., Shen, L. X., Cai, Z., Lewis, H. & Tinoco, I. (1995) "Synthesis and NMR of RNA with selective isotopic enrichment in the bases."*Nucleic Acids Res.* 23, 4913-4921.

Schwartz, M. & Neuhard, J. (1975) *J. Bacteriol.* 121, 814-822.

Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H. & Kay, L. E. (1996) "Assignment of ^{15}N, $^{13}C^{\beta}$, $^{13}C^{\beta}$, and HN Resonances in an ^{15}N, ^{13}C, ^{2}H Labeled 64 kDa Trp Repressor-Operator Complex Using Triple-Resonance NMR Spectroscopy and ^{2}H-Decoupling."*J. Am. Chem. Soc.* 118, 6570-6579.

Sharpless, K. B., Amberg, W., Bennani, Y. L., Crispino, G. A., Hartung, J., Jeong, K., Kwong, H., Morikawa, K., Wang, Z., Xu, D. & Zhang, X. (1992) "The Osmium-Catalyzed Asymmetric Dihydroxylation: A New Ligand Class and a Process Improvement."*J. Org. Chem.* 57, 2768-2771.

Simon, E. S., Grabowski, S. & Whitesides, G. M. (1989) "Convenient Syntheses of Cytidine 5'-Triphosphate, Guanosine 5'-Triphosphate, and Uridine 5'-Triphosphate and Their Use in the Preparation of UDP-glucose, UDP-glucuronic Acid, and GDP-mannose."*J. Org. Chem.* 55, 1834-1841.

St. Johnston, D. (1995) "The Intracellular Localization of Messenger RNAs."*Cell* 81, 161-170.

Switzer, R. L. & Gibson, K. J. (1978) "Phosphoribosylpyrophosphate Synthetase (Ribose-5-phosphate Pyrophosphokinase) from *Salmonella typhimurium*."*Methods in Enzymology* LI, 3-11.

Tolbert, T. J. & Williamson, J. R. (1996) "Preparation of Specifically Deuterated RNA for NMR Studies Using a Combination of Chemical and Enzymatic Synthesis."*J. Am. Chem. Soc.* 118, 7929-7940.

Tolbert, T. J. & Williamson, J. R. (1997) "Preparation of Specifically Deuterated and ^{13}C-Labeled RNA for NMR Studies Using Enzymatic Synthesis."*J. Am. Chem. Soc. in press.*

Toyama, A., Takino, Y., Takeuchi, H. & Harada, I. (1993) "Ultraviolet Resonance Raman Spectra of Ribosyl C(1')-Deuterated Purine Nucleosides: Evidence of Vibrational Coupling between Purine and Ribose Ribose Rings."*J. Am. Chem. Soc.* 115, 11092-11098.

Turnbough, C. L., Kerr, K. H., Funderburg, W. R., Donahue, J. P. & Powell, F. E. (1987) "Nucleotide Sequence and Characterization of the *pyrF* Operon of *Esherichia coli* K12."*J. Biol. Chem.* 262, 10239-10245.

Varani, G., Aboul-ela, F. & Allain, F. H. (1996) "NMR investigation of RNA structure."*Prog. NMR Spectrosc.* 29, 51-127.

Voet, D. & Voet, J. G. (1995) "Biochemistry." in eds) pp. 1360, John Wiley & Sons, Inc., New York.

Vorbrüggen, H., Krolikiewicz, K. & Bennua, B. (1981) "Nucleoside Synthesis with Trimethylsilyl Triflate and Perchlorate as Catalysts."*Chem. Ber.* 114, 1234-1255.

Wagner, G. (1993) "Prospects for NMR of large proteins."*J. Biomol. NMR* 3, 375-385.

Wagner, G. (1997) "An account of NMR in structural biology."*Nature Structural Biology* 4, 841-844.

Wataya, Y. & Hayatsu, H. (1972) "Effect of Amines on the Bisulfite-Catalyzed Hydrogen Isotope Exchange at the 5 Position of Uridine."*Biochemistry* 11, 3583-3588.

Weng, M., Makaroff, C. A. & Zalkin, H. (1986) "Nucleotide Sequence of *Escherichia coli pyrG* Encoding CTP Synthetase."*J. Biol. Chem.* 261, 5568-5574.

Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S. & Usman, N. (1995) "Synthesis, deprotection, analysis and purification of RNA ribozymes."*NAR* 23, 2677-2684.

Wright, M. C. & Joyce, G. F. (1997) "Continuous in Vitro Evolution of Catalytic Function."*Science* 276, 614-617.

Wüthrich, K. (1986) "NMR of Proteins and Nucleic Acids." in eds) pp. 292, John Wiley & Sons, New York.

Wyatt, J. R., Chastain, M. & Puglisi, J. D. (1991) "Synthesis and Purification of Large Amounts of RNA Oligonucleotides."*BioTechniques* 11, 764-769.

Xu, J., Lapham, J. & Crothers, D. M. (1996) "Determining RNA solution structure by segmental isotopic labeling and NMR: Application to *Caenorhabditis elegans* spliced leader RNA 1."*Proc. Natl. Acad. Sci. USA* 93, 44-48.

Yamakage, S.-I., Maltseva, T. V., Nilson, F. P., Földesi, A. & Chattopadhyaya, J. (1993) "Deuteriation of sugar protons simplify NMR assignments and structure determination of large oligonucleotide by the ^{1}H-NMR window approach."*Nucleic Acids Res.* 21, 5005-5011.

Chapter 6

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. (1994a) "A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of ^{15}N, ^{13}C, ^{2}H Labeled Proteins with High Sensitivity."*J. Am. Chem. Soc.* 116, 11655-11666.

Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H. & Kay, L. E. (1994b) "An HNCA Pulse Scheme for the Backbone Assignment of ^{15}N, ^{13}C, ^{2}H-Labeled Proteins: Application to a 37-kDa *Trp* Repressor-DNA Complex."*J. Am. Chem. Soc.* 116, 6464-6465.

Yamazaki, T., Tochio, H., Furui, J., Aimoto, S. & Kyogoku, Y. (1997) "Assignment of Backbone Resonances for Larger Proteins Using the ^{13}C-^{1}H Coherence of a $^{1}H_{\alpha}$-, ^{2}H, ^{13}C, and ^{15}N-Labeled Sample."*J. Am. Chem. Soc.* 119, 872-880.

Ye, X., Kumar, R. A. & Patel, D. J. (1995) "Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex."*Chemistry & Biology* 2, 827-840.

Zapp, M. L., Hope, T. J., Parslow, T. G. & Green, M. R. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 7734.