Structural and Thermodynamic Studies of RNA

Graeme Leslie Conn

A thesis presented for the degree of Doctor of Philosophy

University of Edinburgh

1996

Abstract

The nucleic acids, RNA in particular, are conformationally complex biomolecules. The structures that they can form and the subtle sequence dependent modifications of conformation are of pivotal importance to their function *in vivo*. The advent of solid phase chemical synthesis methodologies for DNA and, more recently RNA, has fuelled the rapid growth in our knowledge of nucleic acid structure.

The work described here is concerned with the development of rapid and reliable synthesis and purification protocols for the automated chemical synthesis of RNA. The use of RNA phosphoramidite monomers incorporating two different 2'-hydroxyl protecting groups is described, along with ion-exchange and reversed phase HPLC protocols for the purification of oligoribonucleotides. The quality of the material obtained by these methods has allowed detailed biophysical investigations by UV melting, high field NMR and X-ray crystallography.

Two different RNA systems were studied:

1. An RNA duplex $r(CGCAAAUUUGCG)_2$: the first study of RNA hydration by NMR, detailed thermodynamic analysis and the growth of single crystals that diffract to 1.8Å resolution are described.
2. DNA.RNA Hybrids: the thermodynamic stability and solution conformation of DNA.RNA hybrids with purine (R) and pyrimidine (Y) strands rich were compared with those of the equivalent homoduplexes. The stability was found to be in the order: rR.rY>rR.dY>dR.dY>dR.rY. Computer modelling from NMR constraints showed that the hybrids adopt conformations intermediate to those of DNA and RNA, and different from each other. The growth of single crystals for X-ray crystallography is also described.

Contents

Table of Contents

Abbreviations

^{13}C	carbon [NMR]
^{1}H	proton [NMR]
2′-OH	2′-hydroxyl
^{31}P	phosphorus [NMR]
5′-OH	5′-hydroxyl
A or Ade	adenine
Å	Angstrom unit ($=10^{-10}$m)
A_{260}	absorbance at 260nm
ATP	adenosine triphosphate
AUFS	absorbance units full scale
Bz	benzoyl [base protecting group]
C	cytosine
CD	circular dichroism
CE	capillary electrophoresis
CEP	cyanoethyl-N,N-diisopropyl phosphoramidite
COSY	correlated spectroscopy
CPG	controlled pore glass
d	DNA strand
DCM	dichloromethane
DMF	dimethylformamidinyl [base protecting group]
DMSO	dimethylsulphoxide
DMT	dimethoxytrityl
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DOT	dioctaoxytrityl [5′-hydroxyl protecting group]
ε	molar extinction coefficient
EDTA	ethylenediaminetetraacetic acid
EtOH	ethanol

FAB MS	fast atom bombardment mass spectrometry
FPLC	fast protein liquid chromatography
G	guanine
Hepes	4-(2-hydroxyethyl)-1-piperazine ethanesulphonic acid
HPLC	high performance liquid chromatography
Hz	Hertz
iBu	isobutyryl [base protecting group]
IE	ion exchange [HPLC]
K	degrees Kelvin
M	molar
MeOH	methanol
mer	nucleotides in length
mg	milligram
µg	microgram
MHz	Megahertz
ml	millilitre
µl	microlitre
mM	millimolar
µM	micromolar
mol	mole
MPD	2-methyl-2,4-pentane diol
N-PAC	N-phenoxyacetyl [base protecting group]
NAP	nucleic acid purification [disposable sephadex column]
nm	nanometre
NMP	N-methylpyrrolidinone
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	two-dimensional Nuclear Overhauser enhancement spectroscopy
ns	nanosecond

OD_{260}	optical density [absorbance] at 260nm
p	phosphate
P_s	pseudorotation angle
PAGE	polyacrylamide gel electrophoresis
PC	personal computer
PEG	polyethylene glycol
ppm	parts per million
r	RNA strand
RNA	ribonucleic acid
RNase	ribonuclease
RP	reversed phase [HPLC]
T	thymidine
TBAF	tetrabutylammonium fluoride
tBDMS	tertiarybutyldimethylsilyl
TCA	trichloroacetic acid
TEA	triethylamine
TEA.3HF	triethylammonium tris-hydrofluoride
THF	tetrahydrofuran
tlc	thin layer chromatography
Tm	melting temperature
Tris	tris(hydroxymethyl)aminomethane
U	uridine
UV	ultraviolet
UV melting	ultra-violet thermal melting [experiment]
°C	degrees Centigrade

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Tom Brown, for his continuous support and enthusiasm. This project would not have been possible but for the generous financial support and provision of facilities by Oswel Research Products Limited. I am also most grateful to the Caledonian Trust for the Universities of Scotland for their Scholarship.

I would especially like to thank Dr. Andrew Lane for many enlightening discussions throughout the course of my project and during the preparation of this thesis. I am also indebted to the people who have read this thesis and offered many helpful comments: Sarah Allinson, Jeevan Bijapur, Dr. V. Ramesh (all of the University of Southampton) and Dr. William Hunter (University of Dundee). And to those lucky to escape that pleasure: John McClean, Paddy Finn, Helen Braven, Richard Payne, Lynda Lucas and assorted 'young 'uns'.

Oligonucleotide NMR spectroscopy, CD and gel mobility studies were conducted by Drs. Andrew Lane, Jeff Gyi and Maria 'Sasi' Conte at the National Institute for Medical Research, London.

Modified DNA phosphoramidites thio dC and dU, and propyne dU were synthesized by the research group of Professor Dick Walker (University of Birmingham) and Mr. Jeevan Bijapur respectively. Oligonucleotide mass spectrometry was conducted by Dr. John Langley of the University of Southampton. X-ray crystallographic data was recorded at Daresbury Laboratories and the ESRF, Grenoble by Drs. William Hunter and Gordon Leonard (of the ESRF, Grenoble).

Chapter 1

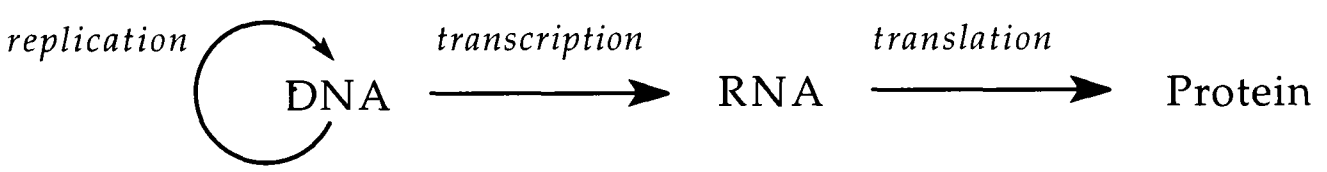
Introduction

1. Introduction

The aim of this chapter is to provide a general overview of nucleic acid structure, and how this can be related to its wide and varied function. The different methods that can be used to study nucleic acid structure and stability will be described briefly here. Before any biophysical study can begin, a source of material must be found. The methods chosen for the synthesis and purification of RNA, and the growth of crystals for diffraction studies are described in Chapters 2 and 3 respectively. The results of the specific thermodynamic and high resolution structural studies on RNA undertaken during the course of this work can be found in Chapters 4 and 5.

1.1 Background

By the 1970s, the roles of nucleic acids as informational molecules and of RNA as a structural component of essential ribonucleoprotein particles were well established [1]. According to the 'Central Dogma' of molecular biology:



where the arrows indicate the direction of transfer of genetic information. DNA acts as a template for both its own replication and the synthesis of proteins through RNA intermediates. Three major classes of RNA were recognised: messenger RNA (mRNA), transfer RNA (tRNA) and ribosomal RNA (rRNA) [2]. mRNA is the copy of the DNA 'transcribed' by RNA polymerase enzymes. This mRNA is used as a template for the

synthesis of proteins on ribosomes with tRNAs acting as 'adapter molecules' specific for each amino acid. rRNA is a key component of the ribosome: it maintains the correct functional ribosome structure and is involved in the orientation of the mRNA.

In little more than a decade, this simple picture has changed dramatically. First with the discovery that RNA could act as a template for the synthesis of DNA in retroviruses, and more significantly, that not all transcribed mRNA is used in protein synthesis. It has become clear that maturation of pre-mRNA by the excision of 'intron' sequences and re-ligation of 'exons' is a universal and important phenomenon in the regulation of gene expression. This has generated a great deal of excitement as it was found that the RNA molecules themselves were able to catalyse the splicing reaction in simple organisms. Since this initial discovery a variety of 'catalytic RNAs' have been found. For the first time, a biological molecule had been shown to have informational, regulatory and catalytic properties. This offers a possible solution to the 'chicken and egg' problem of the origin of polynucleotides and polypeptides through a pre-biotic 'RNA world' [3].

The advances in chemical synthesis of DNA and RNA have been a major force in many recent advances in molecular biology [1]. However, there is considerable lack of detailed structural information on RNA. The growth of RNA crystals that diffract X-rays to high resolution is still a considerable challenge and NMR studies are severely limited in the size of RNA that can be studied. As a result, our current knowledge of RNA structure relies heavily on other low resolution experimental and theoretical techniques.

1.2 Chemical Composition

Ribonucleic acid (RNA) is a linear polymer of repeating 'nucleotide' monomer subunits, linked by phosphodiester bonds. Naturally occurring RNAs can be anything from 75 to many thousands of nucleotides in length [4]. Each nucleotide is the phosphate ester of the related nucleoside, which has two main components: a cyclic furanose type sugar and one of four heterocyclic bases. In RNA, the bases are the purines adenine and guanine, and the pyrimidines cytosine and uracil. The structures of these and one complete nucleotide are shown in Figure 1.1. Each base is linked via a nitrogen to the C1′ of the ribose, to give the four nucleosides: adenosine, guanosine, cytidine and uridine respectively. This bond lies on the same face of the ribose ring as the C5′ hydroxy methyl group and is therefore known as a β-glycosidic linkage.

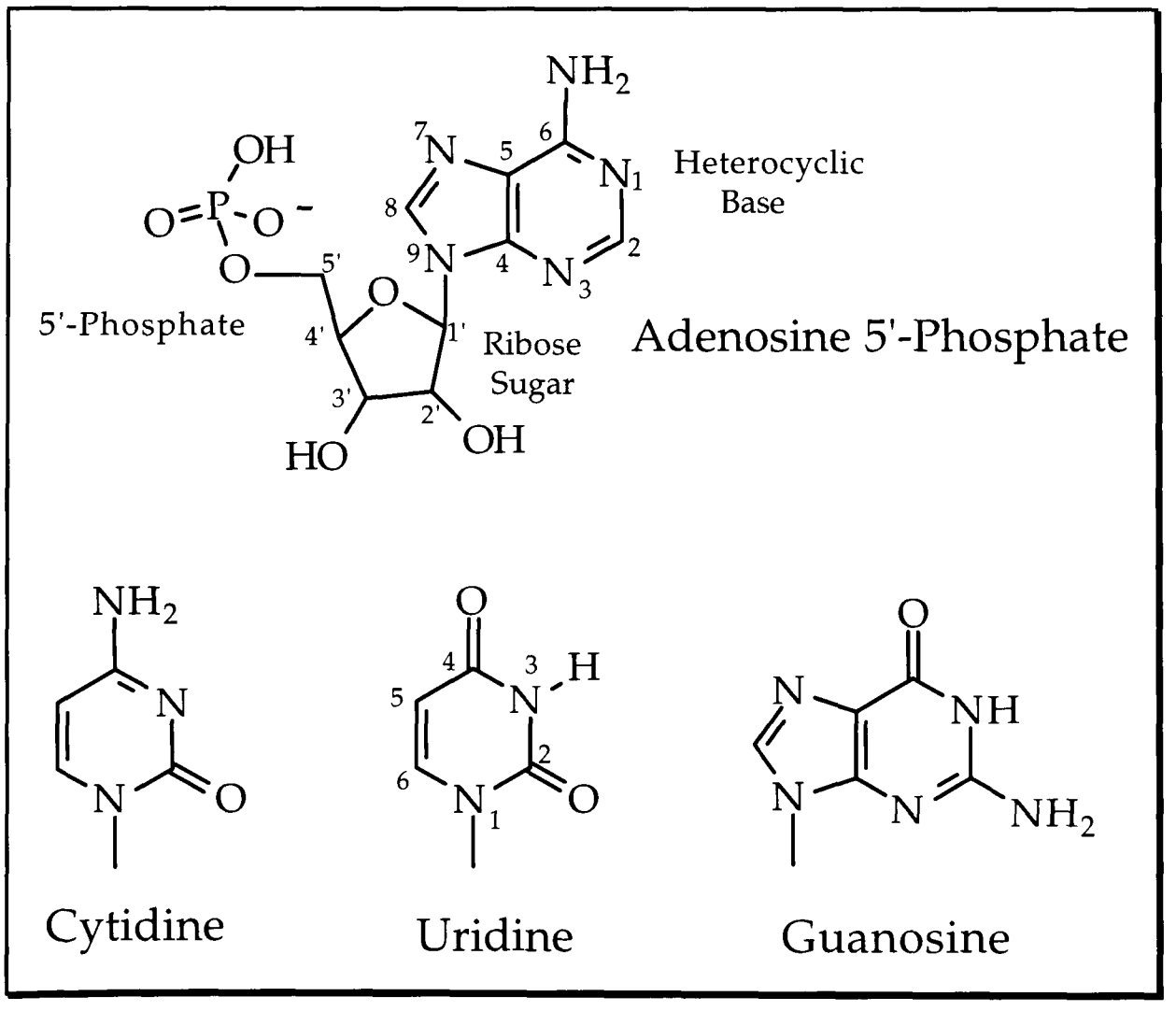


Figure 1.1 RNA Nucleotides

The chemical composition of DNA is identical to that of RNA, except that the sugar lacks a 2′-OH (hence the name *deoxy*ribonucleic acid) and uracil is replaced by thymine.

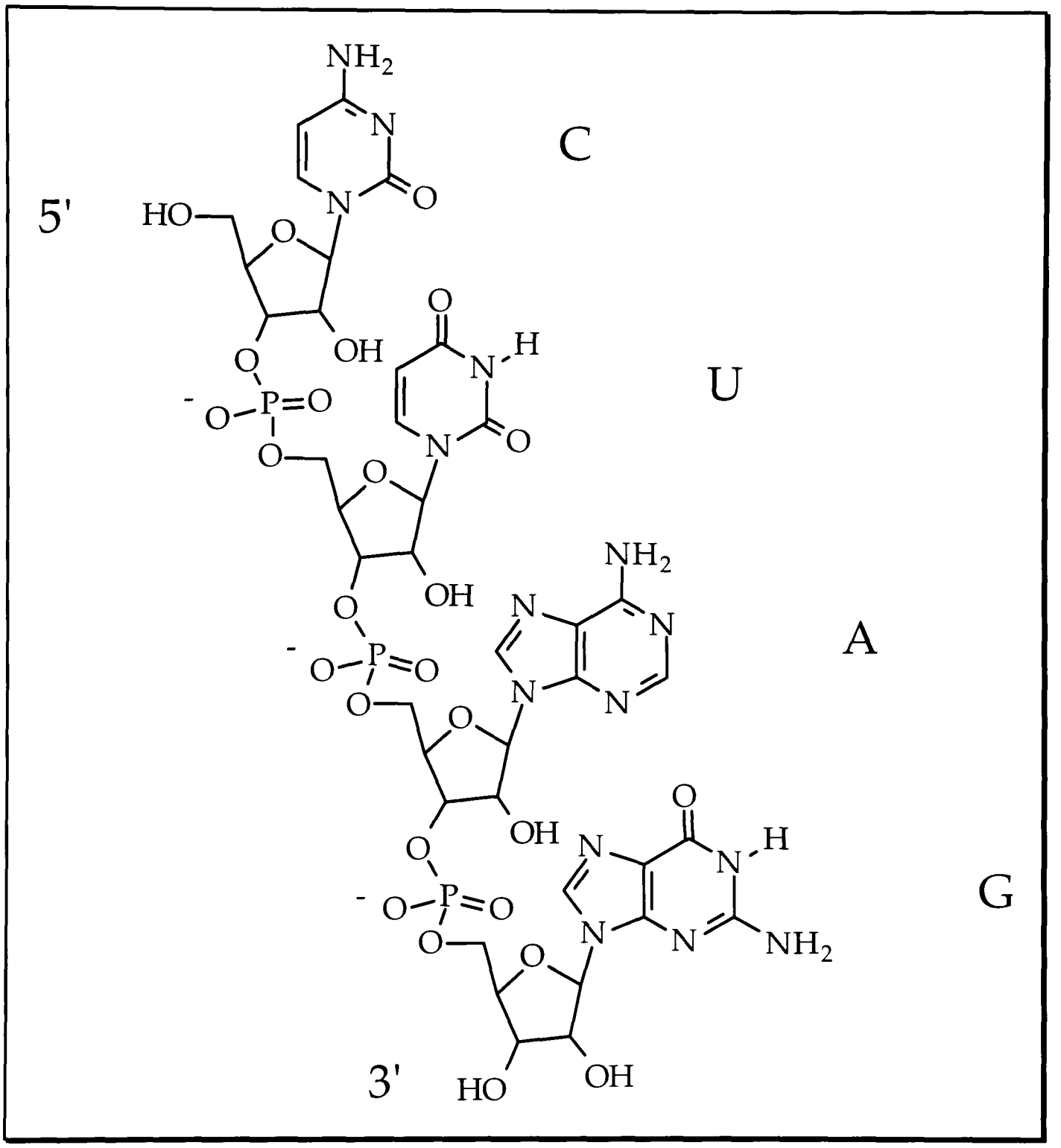


Figure 1.2 An RNA oligonucleotide

In a strand of RNA, nucleotides are attached, through the phosphate, from the 3′-OH of one to the 5′-OH of the next (3′,5′-phosphodiester bonds), as shown in Figure 1.2. The 5′-3′ linkage is actually thermodynamically less stable than the alternative 5′-2′ linkage, but this is

compensated by many other stabilising interactions, such as the stacking of the bases (π-π interactions) upon duplex formation. RNA oligonucleotides are polyanionic and are therefore soluble in water and precipitated by alcohol. At neutral pH, only the phosphates carry a charge, with all the bases existing in their amino-keto tautomeric forms.

1.3 Defining Terms in Nucleic Acid Structure

The conformation of a nucleotide within a strand can be uniquely defined in terms of a number of torsion angles. The torsion angle is defined, for a sequence of atoms A-B-C-D, as the angle between the projected bonds A-B and C-D when looking along the bond B-C. It is defined as 0° if A-B and C-D are eclipsed and the sign is positive if the far bond is rotated clockwise with respect to the near bond. To fully describe a nucleotide in such terms requires:

- six backbone torsion angles: α, β, γ, δ, ε and ζ.
- five torsion angles in the furanose ring: θ_0-θ_4.
- one angle for the glycosyl bond: χ.

The bonds these angles relate to are shown in Figure 1.3a; two of these angles, δ and θ_1, describe the same bond. In a chain of nucleotides many of these angles are interdependent, so the shapes of the nucleotides may be more simply discussed in terms of four parameters: the sugar pucker, the *syn/ anti* conformation of the glycosyl bond, the orientation of the C4′-C5′ bond and the phosphate moiety.

1.3.1 Sugar Pucker

The five-membered furanose rings are twisted in order to minimise non-bonded interactions. This 'puckering' has probably the most dramatic effect on nucleic acid structure, and in particular, the striking difference between DNA and RNA. The type of sugar pucker is identified by the major displacement of C2′ or C3′ from the median plane of C1′-O4′-C4′. The pucker is termed *endo* if it is to the same side as C5′ and the base. For single nucleotides two conformers, C3′-*endo* and C2′-*endo* (Figure 1.3b),

dominate and rapidly interconvert. These sugar puckers are located in the north (N) and south (S) of the pseudorotation cycle of the furanose ring, a notation commonly used by spectroscopists. Values of pseudorotation angle (P_S) may then be given for each possible sugar conformation (a more detailed discussion can be found in [4] and references therein).

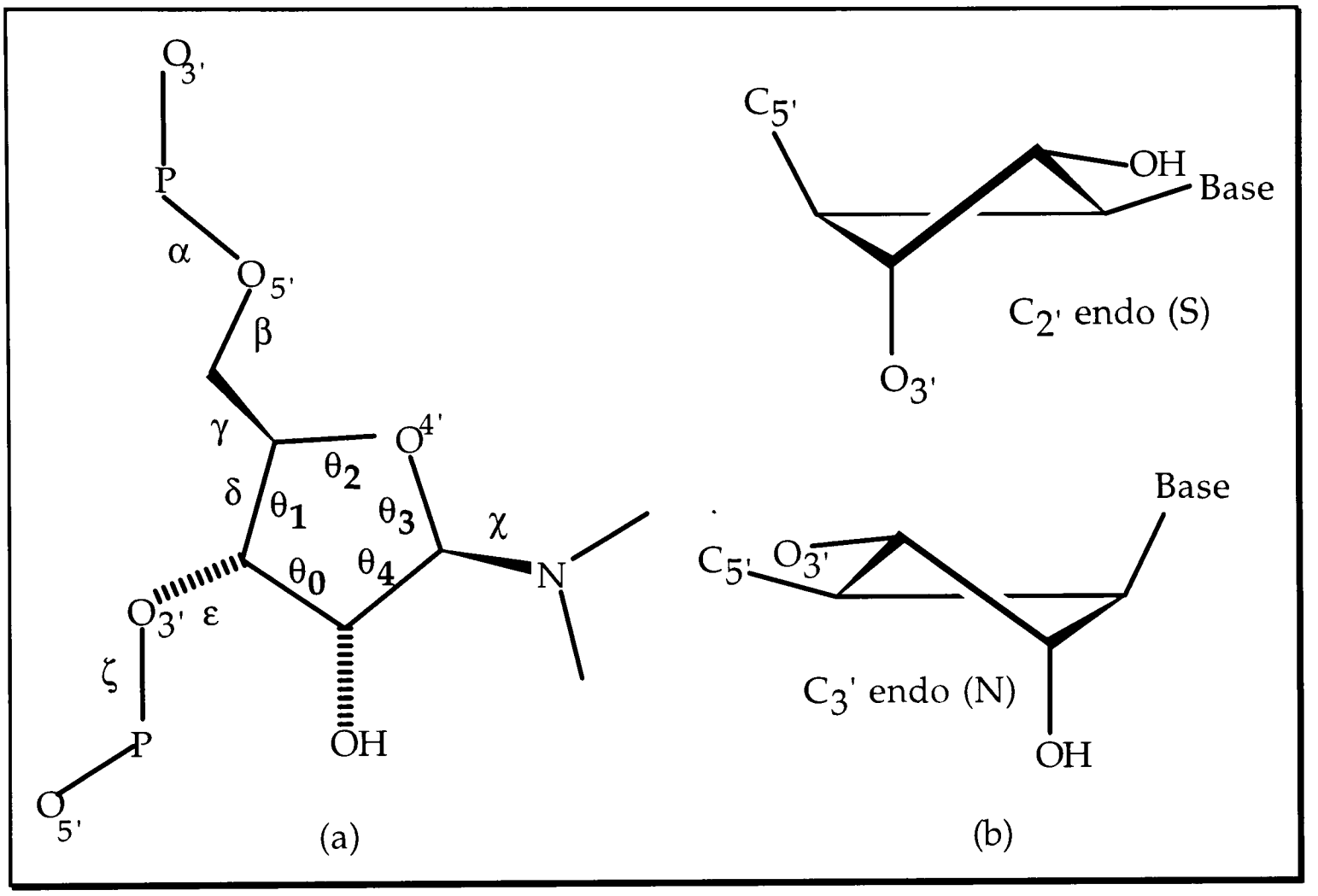


Figure 1.3 (a) Torsion angles used to describe nucleotide conformation and (b) the two predominant sugar puckers.

In helical RNA the predominant pucker is C3′-*endo* (P_S = 0-36°), while in DNA it is C2′-*endo* (P_S = 144-180°). This difference is mainly due to the steric constraints of the 2′-hydroxyl group and the possibility of H-bonding to O4′ of the next residue [5,6], which both favour the hydroxyl in an axial position.

1.3.2 *Syn/ Anti* Conformation

Two main orientations of the base, with respect to the sugar, about the C1′-N glycosylic linkage are possible. These essentially refer to the cases where the base lies above the sugar (*syn*) or away from it (*anti*), and they are defined by the torsion angle χ. In the *syn* conformation, the O2 of the pyrimidines or the N3 of the purines is above the ring, while for *anti* the smaller H6 or H8 respectively are in this position. As a result, it is the *anti* conformation that is found in nearly all DNA and RNA nucleotides and oligomers. The only exceptions to this rule are modified nucleosides with bulky groups on the base (such as 8-bromoguanosine [7]) and guanine in oligomers of the type d(CpGpCpG), or Z-DNA [8]. These special cases show a *syn* conformation.

1.3.3 C4′-C5′ Orientation

The orientation about the C4′-C5′ bond determines the position of the 5′-phosphate with respect to the sugar and is defined by the torsion angle γ. Three main conformations, with the substituents staggered, are possible but almost all nucleotides prefer the synclinal (+sc) or *gauche*, *gauche* (Figure 1.4). Again, the only exception is the unusual Z-DNA structure which forces an antiperiplanar (ap) or *gauche*, *trans* conformation (Figure 1.4) for *syn* guanine deoxynucleotides.

1.3.4 Phosphate Moiety

The phosphate diesters are tetrahedral at the phosphorus and show an antiperiplanar geometry for C5-O5′ and C3′-O3′ bonds. There is very little conformational variation in these linkages in helical regions, so that they are often considered as fixed. This can greatly simplify initial calculations

in otherwise complex structural investigations. Different P-O conformations are found in non-helical regions and also in Z-DNA.

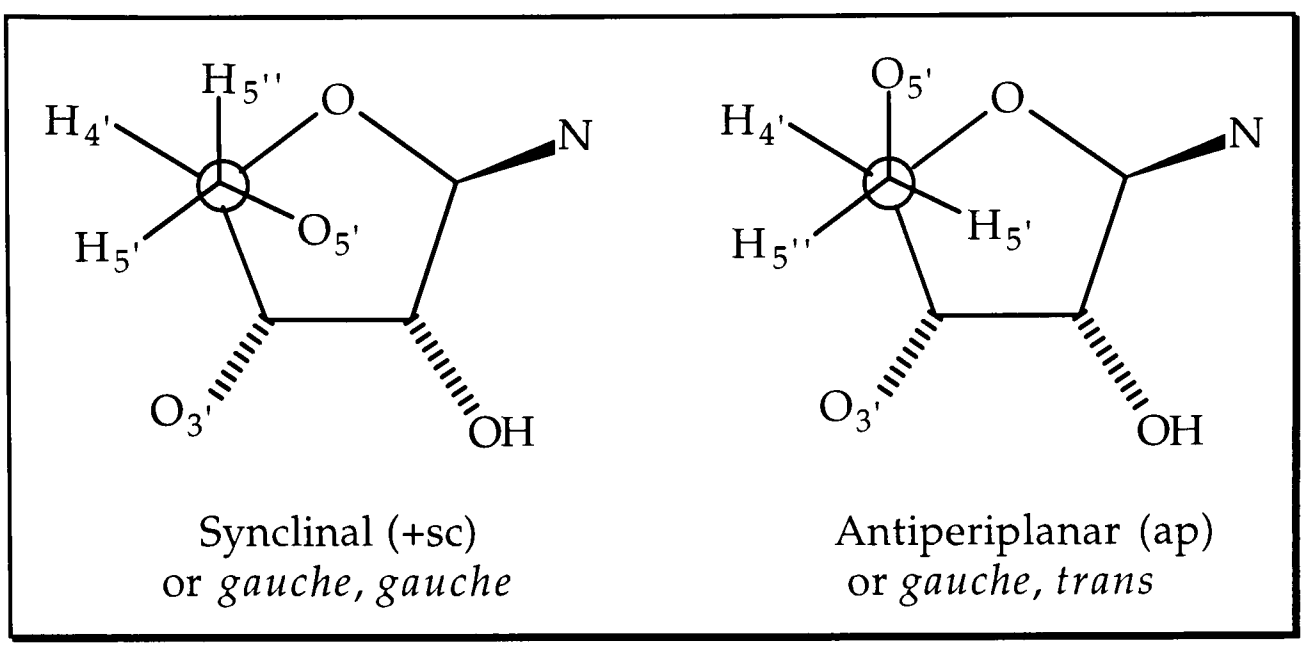


Figure 1.4 Orientation about the C4'-C5' bond

Clearly, nucleic acids are conformationally complex molecules, even at the level of the individual building blocks, the nucleotides. Nothing has yet been said about the many different types of helical conformation nor the interesting secondary and tertiary structures that nucleic acids, RNA in particular, can form. The primary, secondary and tertiary structures are discussed in detail in the next section.

1.4 Nucleic Acid Structure: Primary, Secondary and Tertiary.

1.4.1 Primary Structure

In both DNA and RNA each nucleotide unit is joined to its neighbours through phosphodiester bonds, one at the 5′- and one at the 3′-hydroxyl group. No other linkages are found in regular nucleic acid primary structures. The uniqueness of any given primary structure therefore comes only from the order of its bases. This order is commonly referred to as the DNA or RNA sequence. For example, the primary structure, or sequence, of one RNA oligonucleotide used in this study is shown below.

5′-r(GAAGAGAAGC)-3′

The sequence is commonly written in the fully condensed form and from 5′ to 3′, the direction in which enzymatic synthesis occurs. The prefixes 'r' and 'd' are used to denote ribo- and deoxyribonucleotides respectively. Various methods may be used to determine the content and sequence of nucleic acids. All of these rely on some method of forming labelled fragments of all possible lengths at specific base positions and subsequent analysis by PAGE. Short RNA fragments can be analysed by specific enzymatic degradation (section 1.5.2). For DNA, fragments are generated by treatment with specific chemical reagents followed by alkaline hydrolysis. Longer RNAs and DNAs are generally sequenced using a technique known as dideoxy sequencing. New strands are generated enzymatically (reverse transcriptase for RNA, and DNA polymerase for DNA) from a template annealed to a primer for the enzyme using normal nucleotide triphosphates and radiolabelled ATP, doped with a small amount of one 2′,3′-dideoxynucleotide. The dideoxy acts as a 'chain terminator' as there is no new 3′-OH for chain extension. Comparison of the fragment 'ladder' generated with each dideoxy nucleotide gives the

sequence. This method is the basis of automated sequencing, but fluorescent dyes and laser detection are used in place of the radiolabel.

1.4.2 Secondary Structure

Secondary structure is the pattern of base pairing that allows the formation of antiparallel double strands. In both DNA and RNA there is mutual recognition between A and T or U, and G and C through a pattern of hydrogen bonding to form the canonical Watson-Crick base pairs (Figure 1.5).

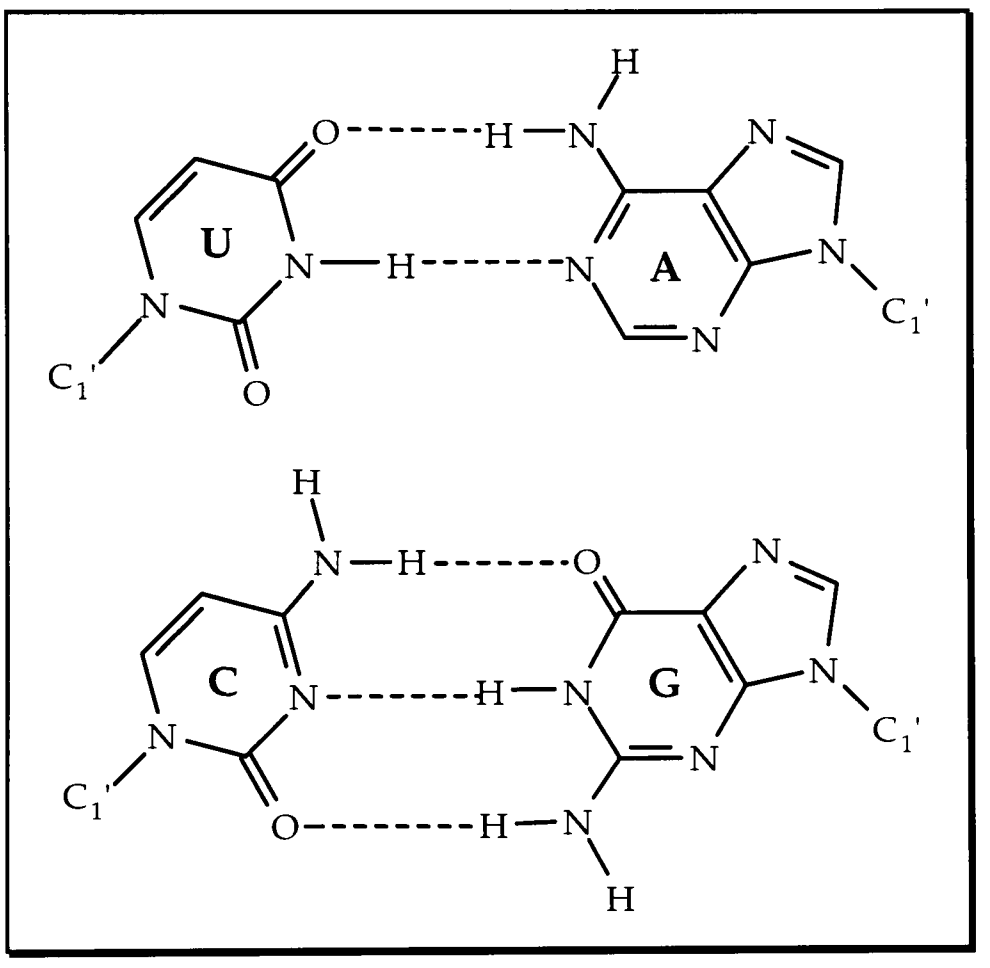


Figure 1.5 Watson-Crick base pairs: A-U and G-C

The amino groups of the bases are good hydrogen bond donors, while the carbonyl oxygens and ring nitrogens are better hydrogen bond acceptors than the oxygens of either the sugar or phosphate. The hydrogen bonds are largely electrostatic in nature and have an average strength of 8-

12kJmol^{-1}. As shown in Figure 1.5, G-C base pairs have three such bonds, while A-U (and A-T) have only two. The C1′-C1′ distances are the same for each base pair (around 10.6Å), which means that they are essentially isomorphous and so can be built into regular helices.

Early fibre diffraction studies showed that there were two different global duplex conformations for DNA, termed A-form and B-form [1]. Subsequently, RNA helices were found to be exclusively A-form. With the advent of solid phase methodology for producing synthetic DNA for crystallographic studies, an entirely new DNA structure, Z-DNA, was discovered [9]. However, as more high resolution structures have become available it has become clear that DNA and RNA duplexes do not conform to the standard structures of these three families. Under different conditions, different conformations may be observed: C-, D-, and T-DNA are all subtle variations on the standard B-form structure [4]. Furthermore, duplex conformation can be sequence dependent, as demonstrated by one of the very first oligonucleotide crystal structures [10]. Although a duplex may be global B-form, individual base pairs may show considerable variability in their orientation with respect to both the helix axis and each other. These subtle variations allow the minimisation of non-bonded interactions and maximisation of base stacking. These differences and the requirements for each of the three families of duplex conformation are discussed below.

1.4.2.1 A-form RNA

RNA forms right handed antiparallel double helices which are always A-form. Sugar pucker is C3′-*endo* and the base is in an *anti* conformation. A-form RNA helices have 11 base pairs per helix turn of about 30Å, giving a rise per base pair of 2.73-2.81Å. The bases are displaced 4.4Å from the helix axis and tilted by 16-19°, resulting in a hollow core down the centre of the

double helix approximately 3Å in diameter. The bases are not co-planar, but are twisted about the hydrogen bonds ('propeller twist') by 7-12°. (Illustration of base and base pair movements is given in Figure1.6). The backbone has skewed phosphate ester bonds and antiperiplanar conformation for adjacent C-O ester bonds. The overall effect is to give the double helix a very deep major groove and shallow minor groove.

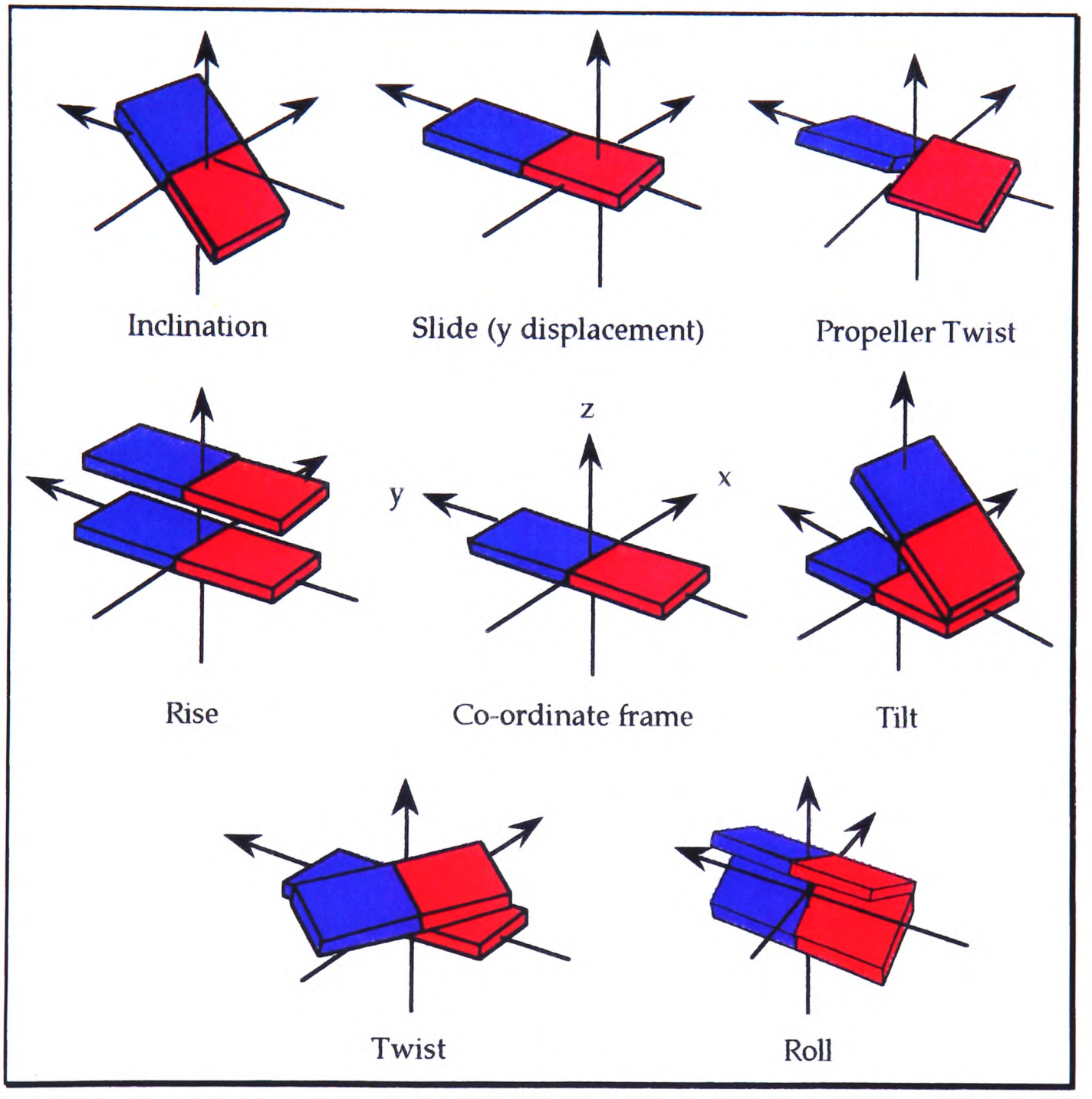


Figure 1.6 Movements of bases in sequence dependent structures

The only other conformation found for RNA duplexes is the A'-form, which occurs under high salt conditions. In A'-form RNA, there are 12

base pairs per rise of helix of 36Å, giving an axial rise of 3.0Å per residue. This is reflected in a lower base pair tilt angle of 10°. The A- and A′-forms of RNA are otherwise very similar. The global helix parameters for A-form, A′-form and the other helix conformations are summarised in Table 1.1.

1.4.2.2 A-form DNA

The A-form DNA structure is very similar to that of A-form RNA with 11 base pairs per turn of helix. It is generally found under conditions of low humidity and high salt. There is an average turn per base pair of 2.56Å giving a pitch height of 28Å. A-form DNA also has the characteristic deep major groove and shallow minor groove, and has an intra-strand phosphate separation of 5.4Å. Many A-form DNA structures have been observed by X-ray crystallography, but some of these have been found to be B-form in solution [11].

Parameter	A-RNA	A′-RNA	A-DNA	B-DNA	Z-DNA
Residues per turn	11	12	11	10	12
Twist per bp, Ω°	32.6	30.0	32.7	36.1	-9, -51
Displacement, Da Å	4.4	4.4	4.5	-0.2 to -0.8	-2 to -3
Rise per bp, Å	2.8	3.0	2.56	3.4	3.7
Base tilt, τ°	16-19	10	20	-6	-7
Sugar pucker	$C_{3'}$-*endo*	$C_{3'}$-*endo*	$C_{3'}$-*endo*	$C_{2'}$-*endo*	$C_{3'}$-*endo*
Base conformation	*anti*	*anti*	*anti*	*anti*	*syn* (G)
Propeller, Π°	-7 to -12	-7 to -12	-8.3	-11.1	-1.3
Slide Dy, Å	-	-	-1.6	0.4	5.4/ -1.1
Roll, P°	-	-	6.3	0.6	-5.8/ 5.8

Table 1.1 Average helix parameters for the major helix conformations (adapted from [1])

1.4.2.3 B-form DNA

B-form DNA helices are found under conditions of high humidity and low salt. They are, again, right handed antiparallel helices with the sugar pucker C2'-*endo* and the bases in an *anti* conformation. Unlike A-form helices, the base pairs sit on the helix axis and are very nearly perpendicular to it. This leads to major and minor grooves of similar depth. There are 10 base pairs per turn of helix of about 34Å, giving a rise per base pair of 3.4Å. The separation between adjacent phosphates is considerably larger for B-form than A-form at 6.7Å as compared to 5.4Å.

At the level of the base pairs, the main difference between the A- and B-forms comes in the degree of slide and roll (Figure 1.6). A-DNA shows a high roll with negative slide, while B-DNA shows no roll and a small positive slide. This has the effect of making a greater hydrophobic surface area of the bases exposed in an A-DNA per base pair. Therefore, with its lower energy of solvation, B-form is more stable at high humidity, and it may be this effect that determines whether a given sequence forms an A- or B-type structure. The grooves of B-DNA both contain large amounts of water, which interacts directly with both the functional groups of the bases and the phosphate backbone. In A-T regions, the narrower minor groove is seen to have an ordered 'spine of hydration' [12].

The other, less significant, families of DNA helix (C-, D- and T-DNA) are all subsets of the B-form, generally found under quite specific conditions. For example, D-DNA is found only for alternating regions of A-T bases and has an overwound structure with only eight base pairs per turn of helix.

1.4.2.4 Z-form DNA

The most striking difference between Z-DNA and the other DNA helices is the fact that it is a *left handed* helix! This DNA conformation is typified by sequences such as $d(CGCGCG)_2$, but has also been found for the duplex $d(CGATCG)_2$ [8]. The two backbone strands run downward on the left of the minor groove and up on the right, the opposite of the A- and B-forms. In an idealised left handed helix, such chain reversal would require all the bases to be in the *syn* conformation. This conformation is indeed observed for G (χ is approximately 60°). However, for C the clash of the O2 with the sugar prohibits the *syn* conformation, and as a result it is found in the usual *anti* conformation. The name Z-DNA is derived from this alternating base conformation which leads to localised chain reversal and a zig-zag effect in the phosphate backbone path.

The two dinucleotide steps also show a large difference in the amount of twist and slide per base pair: CpG has values of -9° and 5.4Å, and GpC - 50.6° and -1.1Å respectively. The overall effect is to create a minor groove so deep it contains the helical axis. The 'major groove' is a convex surface on which the cytidine C5 and guanosine N7 and C8 are exposed.

1.4.2.5 Other Secondary Structure Motifs in RNA

In contrast to DNA, which almost invariably exists in one of the helical forms discussed above, much of the RNA in a cell is single stranded. In order to form helical regions by base pairing, the strand must fold back upon itself in order to obtain the correct anti-parallel orientation. This leads to complex looped structures and, with the other motifs (Figure 1.7), gives large RNA molecules a high degree of flexibility. In this respect, RNA is much more similar to protein than it is to its close relative DNA. The single stranded regions may be stabilised by additional hydrogen bonding interactions. A recent NMR study of an RNA hexaloop showed a

number of unusual interactions including 2'-hydroxyl groups forming hydrogen bonds [13].

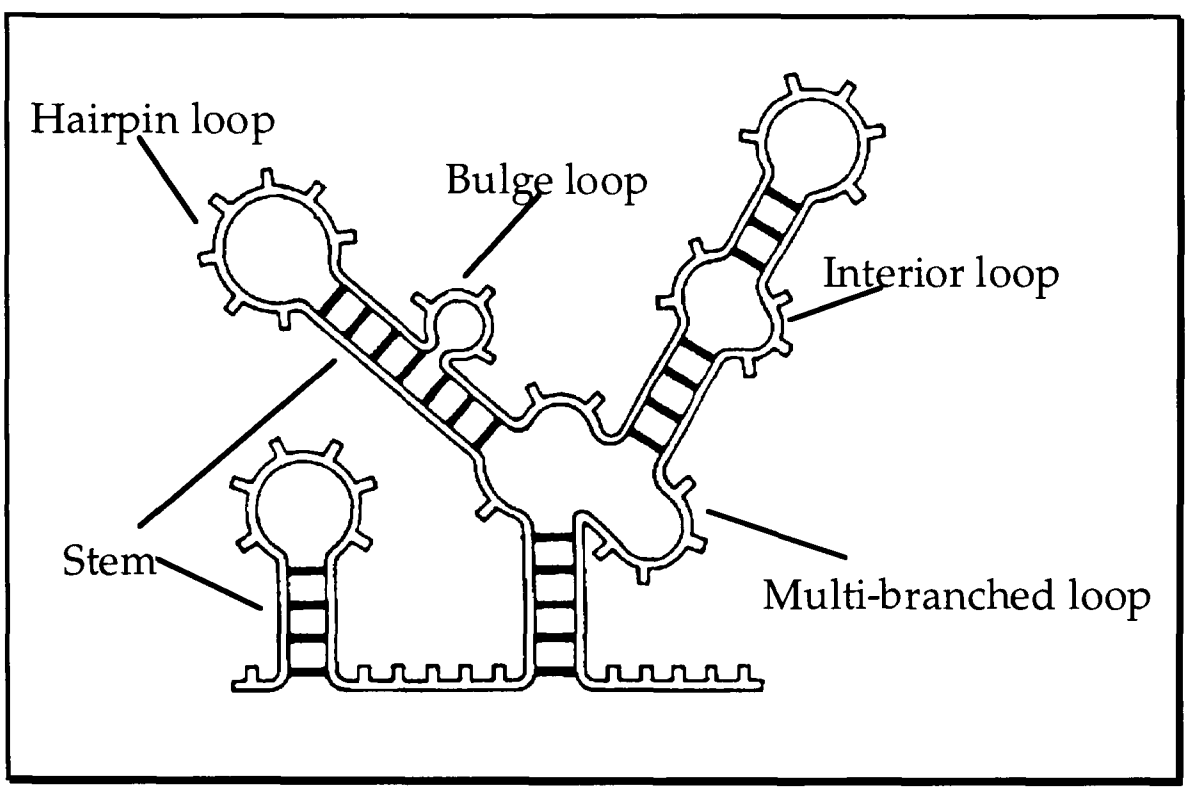


Figure 1.7 Secondary structure motifs in RNA

The secondary structure of an RNA can be determined using various experimental techniques, such as enzymatic or chemical probing. Phylogenetic analysis is now becoming increasingly powerful in predicting RNA secondary structures theoretically. A fuller review of these techniques is given in section 1.5.2.

1.4.3 Tertiary Structure

Interactions between separated regions of RNA secondary structure are commonly known as tertiary structure. RNA function *in vivo* is highly dependent on a wide variety of very specific three dimensional structures. Probably the simplest examples of a tertiary structural unit are pseudoknots. These involve base pairing between one strand of an internal loop with a distant single stranded region or loop (Figure 1.8). RNA tertiary structures are often stabilised by metal ions. Draper has

shown that specific magnesium and ammonium ion binding sites exist that can stabilise RNA tertiary structure [14,15]. However, tertiary structure stability is often not high (only several $kJmol^{-1}$), resulting in a large degree of conformational flexibility. This flexibility is certain to be crucial to the function of RNA *in vivo*.

The tRNAs also have a number of different tertiary interactions which, in this case, lead to very stable structures indeed [4]. The interactions observed include non-standard base pairs, base triples and hydrogen bonds to the phosphates and 2'-hydroxyl groups.

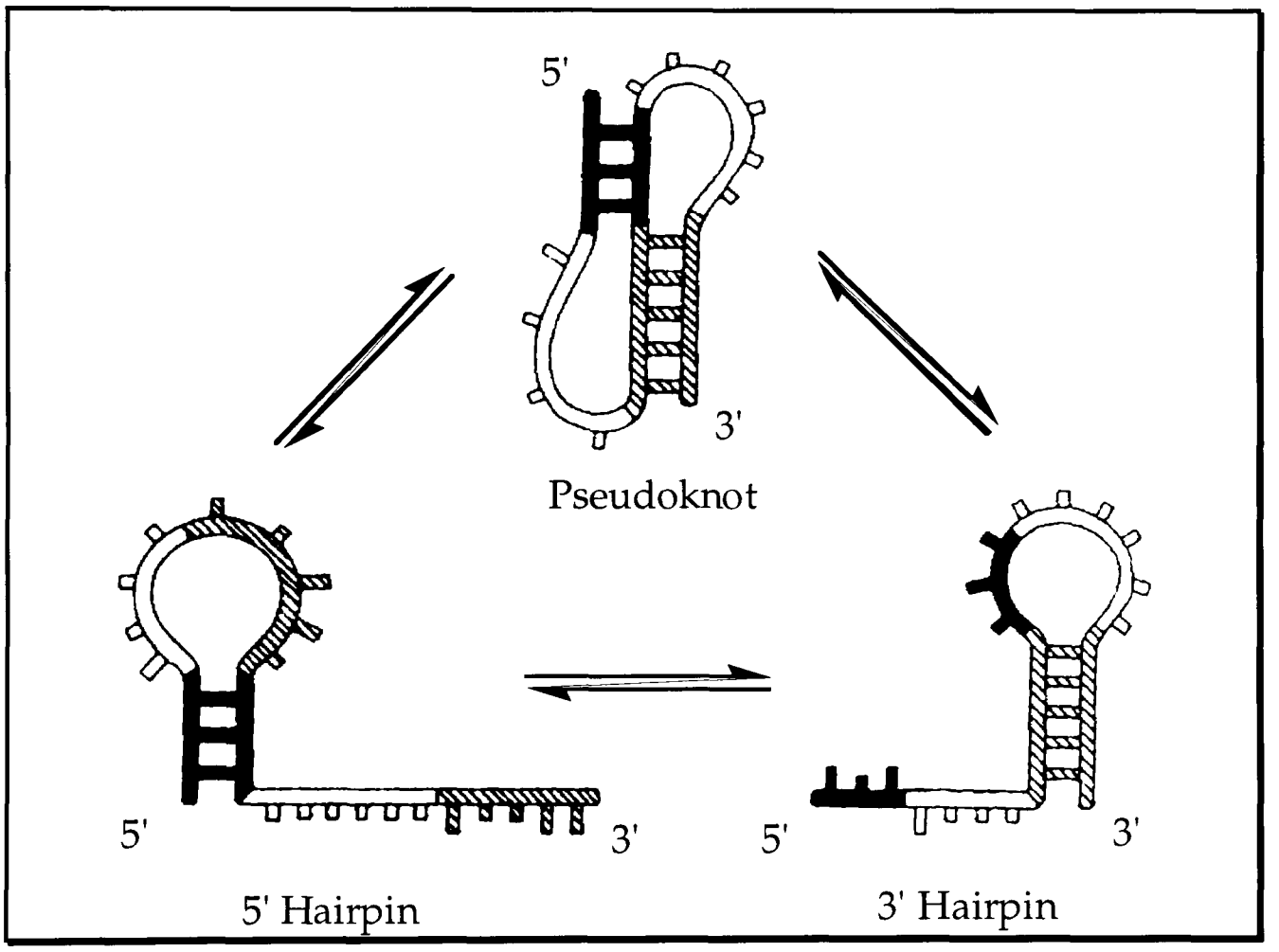


Figure 1.8 Equilibrium between pseudoknot and hairpin loop.

1.5 Structural and Thermodynamic Studies of RNA

Cellular RNAs perform a diverse range of functions, from ribozymes to targets for processing enzymes and regulatory proteins. Such varied functions can only be possible through a variety of structural conformations and stabilities. The starting point in the understanding of biomolecule function must therefore lie in the elucidation of the three dimensional structure. A combination of techniques may be used for this purpose, both experimental and theoretical. X-ray diffraction studies have long provided detailed structural information on biomolecules. However, further evidence that the results of crystallographic work relate directly to the biological function of the molecule *in vivo* is essential. Reliance on crystal structures can lead to the mistaken notion of rigid molecules with little flexibility. Often it is the conformational flexibility of a biomolecule that facilitates its function *in vivo*. Despite this, crystallography remains the most powerful tool in the study of large macromolecular systems.

Nuclear magnetic resonance (NMR) is now undoubtedly the most useful technique for the study of biomolecular structure in solution. Although its scope, in terms of the size of biomolecules that can be studied, is more limited, NMR can provide detailed structural information under physiological conditions. Most significantly, the structure produced is not static: changes in conformation, for example through ion concentration or temperature variation, can be monitored in real-time.

1.5.1 Modelling and Phylogeny

The difficulties encountered in obtaining detailed experimental structural information on large functional RNAs has led to an increase in the use of theoretical methods. With such an approach, the first step towards predicting three dimensional structure is the prediction of secondary

structure. Secondary and tertiary structures in RNA are not as inter-dependent as they are in proteins. Secondary structural units are quite stable in the absence of tertiary interactions, which are generally of low stability. Two main techniques can be employed in RNA structure prediction: phylogenetic analysis and computer modelling using thermodynamic constraints.

Phylogenetic analysis involves searching for complementary changes in RNAs of the same class or having the same function. The aim is to find structural features conserved during evolution [16]. The assumption is made that for RNAs that perform the same function a similar arrangement of secondary structural units (helices, loops, bulges, etc) will be found regardless of the primary structure. Alternatively, a search can be made for co-variations in single RNA sequences which indicate where important interactions for structure formation are conserved. These techniques are now being extended to include tertiary interactions, for example in the group I intron [17], and can form the basis of computer modelling studies.

Computer modelling techniques use empirical or semi-empirical thermodynamic parameters in constructing structures with minimised free energy. A number of algorithms that fold RNA into its most stable secondary structure have been developed [18,19]. However, difficulties may be encountered with non-canonical structures, tertiary interactions and determining exact contributions of loops or bulges. It should also be remembered that RNAs are likely to adopt a number of conformations. Despite this, tertiary structure models have been proposed for a number of different RNAs [20]. One major problem is that a manual approach is generally used in model assembly introducing significant bias in the final model. Some recent studies have reported the use of automated modelling techniques to overcome this problem [20,21]. Although

modelling techniques can arrive at different solutions, for large functional RNAs this does not necessarily mean they are incorrect. Often apparently conflicting results may be a consequence of the inherent flexibility in all or part of the RNA.

While the two theoretical techniques are quite limited on their own, together they can be very useful in RNA structure prediction. When combined with experimental techniques, such as chemical and enzyme probing, crosslinking, and fibre diffraction, they become powerful tools to the structural biologist. Furthermore, the results generated can serve as an invaluable aid in directing high resolution studies on smaller manageable fragments.

1.5.2 Chemical and Enzymatic Probing

Chemical and enzymatic probing techniques have been widely used in the study of nucleic acid structure, RNA in particular. They have the distinct advantage in that unlike biophysical techniques that require large amounts of highly purified material, structure probing experiments are extremely sensitive [22].

The general strategy involves two steps: first, the challenging of the RNA with a variety of enzymatic and chemical reagents and second, detection of the RNA fragments generated. Most nuclease enzymes cleave the backbone of single stranded RNA sequence specifically, with the exceptions of S1 that cleaves without specificity and RNase V1 that cleaves double stranded or structured regions [23]. Similarly, chemical reagents can be used to target specific functional groups, cause modifications or chain cleavage with a range of specificity. For some chemical reagents, additional chemical treatment is required for chain cleavage [22]. Whichever type of reagent is used, the reactions must be

designed such that cleavages or modifications are introduced at a statistical and low level (less than one per molecule). Two techniques may then be used for the detection of the fragments produced. For shorter RNA sequences (up to 200 nucleotides), prior end labelling with ^{32}P allows the analysis of fragments by PAGE. Determination of the size of the fragments may be achieved by comparison of each sequencing reaction with an alkaline hydrolysis ladder. A better approach for longer RNAs is primer extension. In this case, the RNA strand is reverse transcribed into DNA, the enzyme stopping at the site of cleavage or modification. The fragments may then be analysed by PAGE once more.

Reactions may be performed under native, semi-denaturing and fully denaturing conditions, allowing for the correct assignment of cleavages or lack of them. Enzymatic and chemical probing are often used in tandem and can be regarded as complementary. Enzymes have the advantage of being very specific but as large molecules are themselves sensitive to steric hindrance and may alter the structure of the RNA, inducing artefact cuts. In contrast, chemical reagents suffer no such steric problems and are therefore extremely useful in the identification of tertiary structures.

A full structure probing study can be a considerable piece of work. However, the sensitivity and low requirement for material make them extremely useful for examining interactions in functionally intact RNAs such as tertiary folding in ribosomal RNA or RNA-protein interactions. As with the theoretical approaches, these studies also act as invaluable guides for directed biophysical studies on smaller, more manageable fragments.

1.5.3 UV Spectroscopy

Nucleic acids exhibit a strong absorbance in the UV region of the electromagnetic spectrum, with a maximum around 260nm and a molar extinction coefficient of the order of 10^5-10^6 Mcm^{-1}. This absorption, due almost entirely to the heterocyclic bases, allows for a rapid and sensitive measure of the amount of nucleic acid present in solution, down to microgram quantities. The absorbance at 260nm, sometimes called the optical density (OD_{260}), can be used to calculate the nucleic acid concentration in solution from the Beer-Lambert Law:

$$A = (I/I_0) = \varepsilon cl$$

where, I and I_0 are the intensities of transmitted and incident light respectively, ε the molar extinction coefficient, c the concentration and l the path length. Single strand oligonucleotide extinction coefficients can be estimated from nearest neighbour interactions [24], or determined experimentally by enzymatic hydrolysis to constituent nucleotides (section 3.2).

1.5.3.1 UV Thermal Melting (UV Melting)

Nucleic acids generally exist in ordered double helical or folded structures in solution. UV absorbance can be conveniently used to monitor the transition between this native ordered state and a disordered random coil as a function of temperature. As the stacking of bases is disrupted, the absorbance increases, approaching the sum of the absorbances of the constituent nucleotides. The lowering of absorbance in helical or folded structures is known as *hypochromicity* and is a measure of the base pairing and base stacking in the secondary structure. A typical UV absorbance 'melting curve' is shown in Figure 1.9. The midpoint of the

transition is defined as the point at which half the molecules have melted and is called the *melting temperature* (Tm) of the duplex or strand under the solution conditions used. This makes the assumption of a two-state (all-or-none) model where single strands are in equilibrium with only one base paired native structure [25]. This approximation is most appropriate for duplexes up to around 12 base pairs in length [26]. The absorbances of the two states, native and random coil, are both intrinsically temperature dependent, giving rise to the variable upper and lower baselines as shown in Figure 1.9.

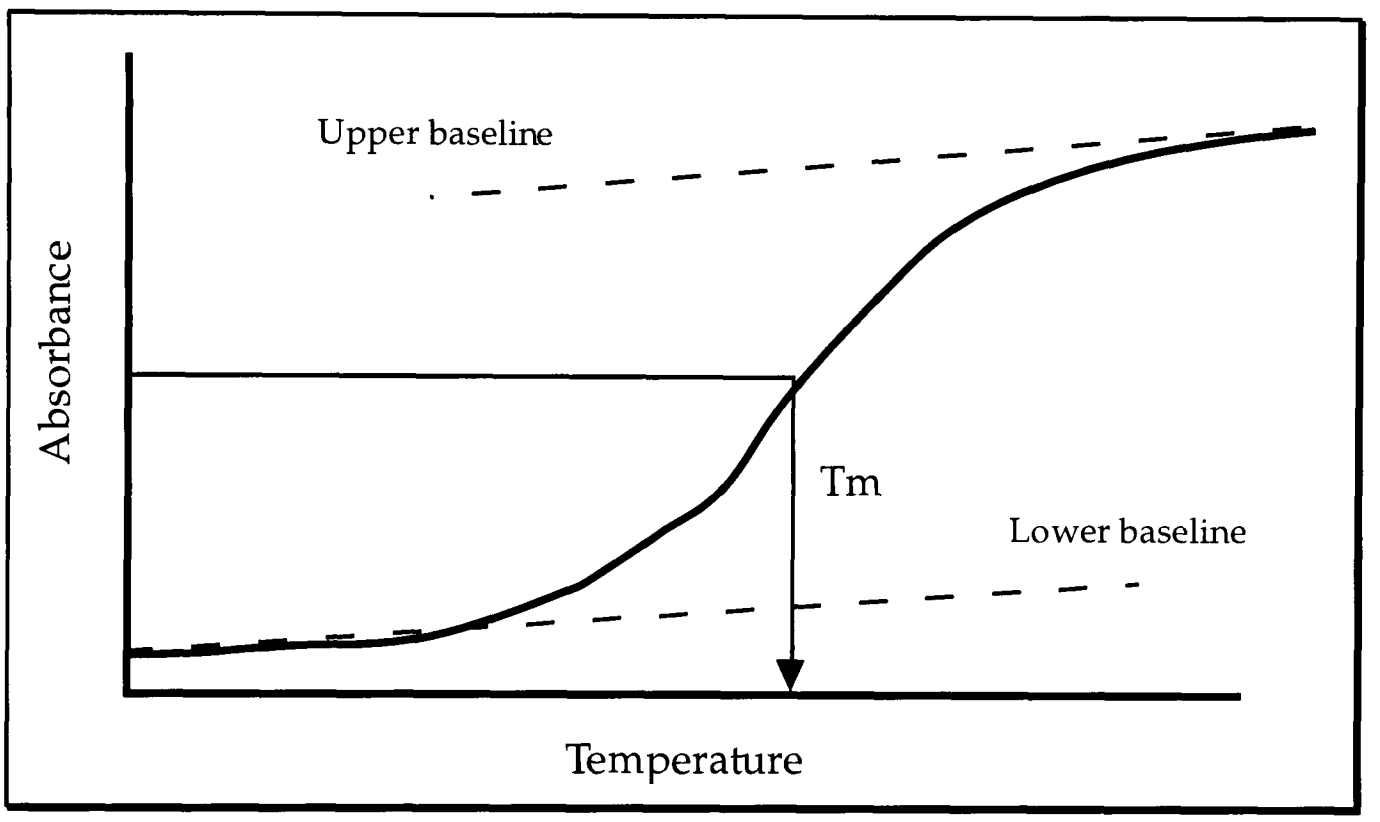


Figure 1.9 A typical UV melting curve, showing upper and lower baselines.

A convincing explanation of the origin of hypochromicity was first developed by Tinoco and Rhodes in the 1950s [25]. It is based on the interactions between the different electronic states of neighbouring chromophores, which for nucleic acids are the heterocyclic bases. The precise quantum mechanical description is not simple, but a classical analogy may be used. When a chromophore is irradiated with light of the

correct wavelength, electronic excitation can occur. The transition dipoles associated with these electronic transitions lie in the planes of the aromatic rings of the nucleic acid bases and their magnitude determines the intensity of absorption. In base stacked structures, the dipole is affected by nearby chromophores. Even if these chromophores do not absorb, they are affected by the incident oscillating electromagnetic field. As the molecules are polarisable, a dipole is induced and the dipole moment can be in or out of phase with the exciting light. In a parallel stack, such as a nucleic acid helix, all dipoles are arranged in a mutually repelling fashion (Figure 1.10). As a result, the transition dipole is reduced leading to a lower absorption. The strength of hypochroism varies with the inverse cube of the distance between interacting chromophores. The effect is therefore only observed when chromophores are packed close together, and is dominated by nearest neighbours. Variation of the angle between the planes of the bases affects the hypochromicity dramatically. The hypochromicity of a native structure is given by:

$$\text{Hypochromicity} = \frac{(A_S - A_D)}{A_S}$$

where A_S and A_D are the absorbance of random coil and native structure respectively. Hypochromicity is an approximately linear function of the number of stacked bases and, as such, is an effective monitor of the fraction of stacked bases as a native structure is denatured. Monitoring at different wavelengths can also provide information about base content as the wavelength of maximum hypochromicity differs for G-C (280nm) and A-T/U (260nm) base pairs.

UV melting experiments can yield both qualitative information and quantitative data in terms of standard thermodynamic parameters. Typically, experiments are carried out in phosphate buffer, which has a low temperature coefficient, with various concentrations of salt, metal

ions or ligands of interest. Heating rates of between 0.25 - 1.0° per minute are generally used, with the faster rates being mainly applicable to the study of short oligomer duplexes. A correct melting curve is an equilibrium measurement and therefore should be independent of heating rate.

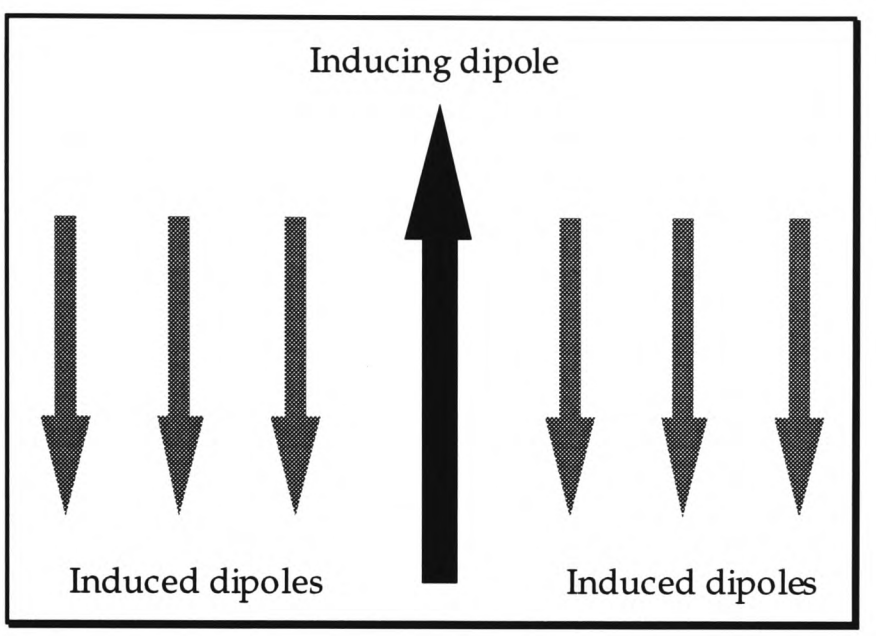


Figure 1.10 Origin of hypochromicity

With the assumption of a two-state system, equilibrium constants can be expressed in terms of fractions of fully base paired and random coil species [24]. This then allows the calculation of standard thermodynamic parameters ΔH°, ΔS° and ΔG° via the van't Hoff relation:

$$\frac{d\ln K}{d(1/T)} = \frac{-\Delta H^\circ}{R}$$

and standard thermodynamic relationships:

$$\Delta G^\circ = -RT\ln K \qquad \Delta S^\circ = \frac{\Delta H^\circ - \Delta G^\circ}{T}$$

There are a number of ways in which to extract this information from UV melting curves. All but one of these require the sloping upper and lower

baselines to be removed before accurate data can be obtained. This can be fairly subjective and as a result is a source of significant error. In some cases, with very low or very high Tm, accurate estimation of the baseline may be difficult. Probably the most accurate and efficient method for obtaining thermodynamic parameters is from the concentration dependence of Tm, which involves no baseline assumptions. This method is the most widely reported in the literature, and is thought to give the most reliable results in comparisons between different laboratories. It is the method used in this study. All that must be measured are melting temperatures at various oligonucleotide concentrations, ideally over a range of between 70-100 fold concentration. Thermodynamic parameters are then determined from plots of reciprocal Tm versus the natural logarithm of strand concentration, Ct for self-complementary sequences or Ct/4 for non-self complementary sequences [24]. The Tm and shape of melting curve for unimolecular (hairpin to random coil) transitions are independent of the total strand concentration. For these transitions, an alternative method must be used to extract the data direct from the melting curves.

UV melting experiments have also found application in the study of large complex RNAs with extensive secondary, and often tertiary structures, among them tRNA, rRNA, viroids and viral RNA. The melting of these RNAs is not two state: the native structure can be in equilibrium with a number of partially unfolded states as observed for tRNA [27]. The extraction of quantitative data from complex melting curves is not simple. Draper has developed algorithms for the deconvolution of first derivatives into sequential transitions and applied them to the unfolding of ribosomal RNA fragments [28]. The process involves the iterative calculation of three variables associated with each transition: the hypochromicity, enthalpy and Tm. The number of transitions is specified by the user, according to a model for the RNA unfolding which is based

on calorimetric measurements. This has allowed the study of the RNA tertiary structure and the various effects of magnesium and ammonium ions [14,15], antibiotics and protein [29,30].

1.5.3.2 Circular Dichroism

Circular dichroism (CD) measurements on nucleic acids may be used to provide information on global helix conformation and base stacking. The technique depends upon the differential absorption of the left and right circularly polarised components of plane polarised light by a chromophore that is either intrinsically chiral or that is in an asymmetric environment. The results are generally expressed as the difference in the left- and right-handed molecular extinction co-efficients as a function of wavelength. Nucleic acid bases are planar and therefore optically inactive. However, in nucleic acid strands the glycosylic linkage to C1′ of the furanose sugar imparts a chiral perturbation to their UV absorption [1].

The information available from CD studies is limited in comparison to high resolution techniques but CD has the advantage of rapid data collection and analysis and requires very little material. CD has proved particularly useful in the study of nucleic acid homopolymers [31], transitions between helical forms under different solution conditions [32], and DNA interactions with ligands and proteins. In these cases, changes in CD are quite dramatic. For large RNAs, changes in CD are likely to be much smaller in general, limiting its usefulness.

1.5.4 Nuclear Magnetic Resonance (NMR) Spectroscopy

Since the initial observation of RNA imino proton resonances in the early 1970s, NMR has become extensively used in the study of RNA structure and dynamics. Advances in spectrometer hardware, the

development of superconducting magnets of very high field combined with increased computing power have led to nothing short of a revolution in the field. Today, Fourier transform NMR is undoubtedly an invaluable alternative and complement to X-ray diffraction.

NMR is based on the fact that certain nuclei absorb radiofrequency radiation at a characteristic frequency when oriented by a strong magnetic field. The precise frequency depends not only upon the particular nucleus but also its environment. Furthermore, the magnetisation can be transferred between nuclei through bonds and through space, providing information on relative orientations and internuclear distances. Thus, high resolution RNA structures may be generated from NMR data provided resonances can be assigned to each nuclei and structural constraints measured. Three major problems make this difficult:

- the protons of the bases are not scalar coupled to their corresponding ribose, so there is no direct link between them. Furthermore, protons within a single ribose may not constitute a single scalar spin system either [33].
- there is no easy way to distinguish pyrimidine proton resonances.
- resonances are generally predictable but spectral overlap is often worst in the regions of most value.

Assignment of imino proton resonances is still useful in the study of RNA structure as it is effectively equivalent to determining the base pairing, for example in demonstrating the formation of a hairpin structure [34]. However, since the late 1980s, emphasis has shifted towards full conformational characterisation and two main methods for the sequential assignment of NMR data for RNA have developed [35]. Homonuclear proton based methods for assigning nucleic acid spectra which start with the base resonances and work in towards the backbone are now well established. These 'bases-in' strategies tend to rely heavily

upon NOE data. More recently, attention has focused on heteronuclear methods, that work from the backbone out to the bases. These methods depend more upon scalar couplings.

Bases-in strategies generally begin with the assignment of H1′ protons. H1′ NOEs to aromatic protons H6/H8 are found in one of the better resolved regions of an RNA NOESY spectrum. In helical regions, H1′ are always within NOE distance of the H6/H8 of the same nucleotide and the H6/H8 of the next nucleotide in the sequence. It is therefore possible to trace these resonances along the chain in a 'sequential walk', assigning each resonance in turn. The sequential walk can be assigned to the sequence by correlation with the assigned imino proton resonances. Following the H1′ assignment, the other sugar resonances can be addressed in the same manner. Spin coupling information from TOCSY and COSY experiments is generally only useful for DNA, which adopts a C2′-*endo* sugar pucker. In RNA, the C3′-*endo* pucker means that the H1′-H2′ coupling is small so that little useful information is available. Unfortunately, bases-in strategies rarely yield a full set of assignments and the information available is heavily conformation dependent. Problems may be encountered in non-helical regions, which are often the interesting and functional parts of the molecule, as H1′-H6/H8 walks are not available.

The use of ^{31}P-^{1}H spectroscopy is a step towards an alternative 'backbone-out' strategy. ^{31}P chemical shifts are sensitive to backbone conformation [36], allowing the identification of unusual conformations. More qualitative information, such as chemical shift dispersion, is also available which can be useful in comparing spectra of different duplexes. Scalar couplings can be observed from phosphorus to H3′ on the 5′-side, and to H5′ and/ or H5″ on the 3′-side. The first backbone out sequential assignment was based on these couplings but was limited to small RNA fragments [35]. Since then, the development of ^{31}P/^{1}H hetero-TOCSY

experiments has improved the usefulness considerably. However, this solution is not complete: the couplings are still dependent on conformation and their dispersion is low because RNA always adopts A-form helices.

A true backbone-out strategy requires the use of isotopic labelling. Spectral resolution is improved by the dispersing of the proton-proton cross peaks using the associated heteroatom shifts. Assignments are conformation independent because the one and two bond coupling constants exploited are largely independent of torsion angle [35]. ^{13}C and/ or ^{15}N labelled samples can be prepared using nucleotides obtained by digestion of RNA from bacteria grown on labelled medium. A ^{13}C-^{15}N nucleotide is a *single* spin system. Thus the complete sequential assignment of all ribose protons and their correlation with associated base protons is possible. However, the solution is not ideal as heteronuclear labelling can open up additional relaxation pathways. It may be that these new methods will only serve to improve the assignment of fragments of similar size to those studied by the bases-in methods, rather than extend the range of sizes.

At present, spectroscopists are generally limited to studying nucleic acids with up to around 30 nucleotides. This is likely to improve significantly in the future with the development of new heteronuclear approaches and improvements in magnetic field strength. One example is the use of ^{19}F spectra to study the conformational flexibility of RNA fragments containing fluoro uridine residues [37].

The analysis of NMR data gives information on sugar puckers, glycosidic torsion angles and various distance constraints. NOE distances are generally ranked as strong, medium or weak. This information can then be used to generate models of biomolecular conformation using

restrained molecular dynamics and various computer algorithms that work in distance and torsion angle space. Generally, data is supplied in terms of the NOE based distances, coupling constant derived torsion angles with the known base pairing and models are constructed with the aid of standard bond lengths and angles. The models are generated to be consistent with the experimental data and then refined. The refinement invariably involves the minimisation of the structure's nominal energy and a reduction in the deviation from the experimental data. This ensures that bond lengths and angles are sensible and prevents close contacts. It does, of course, also mean that the models are to a certain degree invented by the modelling process. It is possible to 'fix' areas of structure that are less well defined by the experimental data so that anomalous conformations are avoided. For example, the backbone in RNA helical regions can be fixed in the limits of an A-form helix, from which it is unlikely to deviate.

1.5.5 X-Ray Diffraction

During the second half of this century, X-ray diffraction has revolutionised our understanding of nucleic acid structure and function. The information from early fibre diffraction studies was crude by today's standards, but provided much useful structural information. After all, the first detailed model of DNA structure was based on X-ray fibre diffraction data. With the combination of modern X-ray sources and computing power, X-ray diffraction studies can now provide unambiguous and complete three dimensional structures of large biomolecules. For high resolution X-ray analysis, diffraction quality single crystals are required. Although DNA oligonucleotides do crystallize relatively easily, crystals with good diffraction properties are less common. The situation is likely to be the same or possibly worse for complex RNAs because of their greater conformational flexibility.

1.5.5.1 What is a Crystal?

A crystal of any material consists of a regular repeating array of atoms or molecules. The basic building block is the unit cell, a parallelepiped fully described by the dimensions a, b and c, and the angles between them α (b and c), β (a and c) and γ (a and b). A number of different unit cells may be used to describe a given lattice, but one is chosen such that the full symmetry of the lattice is displayed and by the convention that the angles should be the nearest possible to 90°. A cell which contains only a single copy of the repeating structure (a single lattice point) is known as a 'primitive' cell. Sometimes it is more convenient to use a cell with two or four copies of the repeating structure. There are seven different crystal systems, or unit cell shapes, defined by the minimum symmetry of the unit cell: triclinic, monoclinic, orthorhombic, tetragonal, trigonal/ rhombohedral, hexagonal and cubic. In three dimensions, there are 14 distinct crystal lattices (Bravais lattices) that can be built on these crystal systems, using the seven space lattices and adding face centred and body centred to a number of these [38]. The addition of symmetry elements to the Bravais lattices gives the 230 possible crystal space groups [39]. As nucleic acids only contain one optical isomer of the chiral ribose, only 65 of these space groups are available. In most structures some atoms in each unit cell are related to others in the same unit cell by a symmetry operation, such as a rotation axis. This smallest structural motif from which the complete structure can be obtained through the space group symmetry operations is known as the *asymmetric unit*. Therefore the problem is reduced to locating the atoms in the asymmetric unit rather than the whole unit cell.

1.5.5.2 Diffraction of X-rays

X-rays are suitable for diffraction studies as they lie in the region of the electromagnetic spectrum of similar wavelength to interatomic distances

(about 1.5Å or 0.15nm). Two different types of X-ray source are commonly used for today's high resolution diffraction studies [40]:

- *(rotating) anode generators*: X-rays are generated by accelerating a beam of electrons into a metal anode. The type of metal used determines the wavelength of radiation produced, for example copper produces X-rays with a characteristic wavelength of 0.152 nm. The X-ray beam is monochromatised using either a thin metal foil (which absorbs most of the unwanted radiation), or by using intense low order diffraction from a graphite crystal.
- *synchrotron radiation*: X-rays are produced when electrons, accelerated to high velocity, are bent by a magnet. Synchrotron sources are much more intense than rotating anode generators, which allows for more rapid collection of data. Furthermore, the X-ray spectrum produced is continuous over a range of wavelengths (generally 0.05 to 0.3nm), making multi-wavelength experiments possible.

The scattering or diffraction of X-rays by the electrons of atoms in a crystal is an interference phenomenon. The intensities of diffraction maxima show variations in different directions and also vary significantly with the angle of scattering. The diffraction of X-rays can be considered in a manner analogous to diffraction by a grating but extended to three dimensions, or alternatively in terms of reflections from planes of atoms within the lattice. The latter approach was used by William Bragg in his pioneering work in the determination of the structure of sodium chloride [36]. Waves scattered from adjacent planes will be just in phase only for certain angles of scattering. At these angles, the difference in the path lengths travelled by these waves will be an integral multiple of the wavelength of the radiation ($n\lambda$). This is expressed in terms of the perpendicular spacing of the lattice planes (d) and the complement of the angle of incidence (θ) of the X-ray beam in the Bragg equation:

$$n\lambda = 2d \sin\theta$$

Thus, maxima only occur when the relation of wavelength, interplanar spacing and angle of incidence is appropriate. The results of X-ray diffraction by a crystal are recorded, as a pattern of spots with various intensities, either photographically or electronically. Single crystal diffraction data contain all the information needed to reconstruct the three dimensional structure of the unit cell and the molecules, except the phases. The problem of determining the phase angles is still the object of much investigation and a considerable problem in solving certain structures [36]. A number of ways round the phase problem are available:

- *Multiple isomorphous replacement* involves the incorporation of a 'heavy atom' into the structure which alters the diffraction pattern observed. A comparison of native and derivative data can then be used to derive the phases. Clearly, the heavy atom derivative must crystallize in the same form as the original molecule for this method to be useful. The heavy atom may be incorporated through soaking or crystallizing with the heavy atom in solution, or directly into the molecule itself (for example with a 5-bromo uridine residue).
- *Molecular replacement* allows the determination of phase angle without the preparation of any additional material. In this case, a related or model structure is used in searching for solutions to the phase problem.

1.5.5.3 Steps Involved in Crystal Structure Determination

The first step in any crystal structure determination is obtaining diffraction quality crystals. Producing such crystals of nucleic acids alone or complexed with proteins is not simple and is often the rate limiting step in many crystal structure analyses. Generally a wide range of buffers, pH, precipitants and counter ions are examined before suitable crystals are

obtained. With crystals of sufficient size (0.1-0.3 mm on edge) the unit cell dimensions and space group may be determined from the initial diffraction tests. From this information, a decision must be made as to whether the ultimate resolution possible is sufficient to provide detailed information about the structure. If this is the case a number of diffraction patterns are recorded over a range of angles of orientation of the crystal to comprise a data set for the structural determination. Using the observed structure factor amplitudes (|F|) from the diffraction pattern and calculated phase angles, an approximate electron density map at all points in the unit cell may be generated by Fourier synthesis. This trial structure is then refined to give the best possible fit of observed and calculated structure factor amplitudes. A measure of the correctness of the model is then given by the crystallographic residual, R:

$$R = \Sigma \, | \, (|Fobs| - |Fcalc|) / \Sigma \, (|Fobs|)$$

Generally, the lower the value of R the better the structure. For DNA, R values of 0.12 for well resolved to 0.25 for poorly resolved structures are obtained. However, it should be remembered that this represents the closeness of fit of the model to the data, which is not necessarily the same as the accuracy of the model.

When structural determinations by X-ray crystallography are successful, precise information on atomic positions, bond angles and interatomic distances is available. A vast amount of structural information is now available for DNA. While less crystal structures are currently available for RNA, they are appearing in the literature at an increasing rate.

The relevance of the results of these studies in a biological context is still open to some debate: it is known that crystal packing forces can stabilise or induce important alterations in structure. However, it is unlikely that the

true state of nucleic acid structure in the semi-solid cellular environment is modelled perfectly by either solid or solution state structural studies. Thus, in order to obtain the most reliable information on nucleic acid structure it is desirable to use a combination of as many of the techniques available as is possible.

1.6 Structure-Function Relationships

Nucleic acid structure is undoubtedly intimately associated with its biological function. Whether this is at the local nucleotide level within helical structures, or through the formation of intricate folded tertiary structures, it is the three dimensional shape that facilitates the many different roles nucleic acids play.

1.6.1 Protein-Nucleic Acid Interactions

The recognition and interactions of nucleic acids and proteins are crucial at many levels of cell function. DNA-protein recognition is inherently simpler to rationalise than the analogous situation for RNA, because of the predominance of helical structures. Two types of recognition can be envisaged: a 'direct readout' of base sequence or an 'indirect readout' through helix conformation [41]. In the regular helical structure formed by DNA, the main variations are in the availability of the hydrogen bond donors/ acceptors presented by the base pairs in the grooves. Sequence recognition through specific hydrogen bonding patterns is most likely to occur in the major groove where greater discrimination is possible [42]. This 'direct readout' mechanism has proved to be correct for many DNA-binding proteins. Alternative 'indirect readout' mechanisms have been found, in which subtle variations in backbone conformation or easily deformable regions of helix are recognised. Protein induced deformations of DNA structure are common. One highly specific interaction is that of the double stranded sequence d(GAATTC) by Eco RI endonuclease. Interactions are observed between the protein and the DNA major groove and backbone, with binding causing abrupt distortions from canonical B-form [43]. Protein induced conformational changes are also observed in d(TATAAAAG) complexed with a binding protein [44,45]. In this case, the binding induces sharp kinks at either end of the duplex with an

unwinding of the helix and a widening of the minor groove. Protein-DNA contacts are seen down the minor groove and sugar-phosphate backbone, and at each end of the duplex through the intercalation of phenylalanine into the helix.

RNAs differ from helical DNA in a number of ways that affect the scope of protein recognition. RNA helices are A-form, with wide shallow minor grooves and deep narrow major grooves. This may limit RNA protein interactions to the minor groove, except where distortions are found [46]. These may be quite common as much RNA is single stranded in the form of bulges, internal and hairpin loops. These motifs can form specific but irregular structures that present mismatched and unstacked bases [131,47]. For many RNAs it is also necessary to consider the shape of the molecule formed through various tertiary interactions. Finally, the deformability seen for some DNA sequences is likely to play a major role in RNA recognition. RNA molecules are inherently more flexible and many of the structural units have low energies of formation to allow facile conformation changes. In all, these differences offer RNA binding proteins a much greater range of binding strategies than for DNA, with an 'indirect readout' mechanism being widespread.

1.6.2 Catalytic RNA

Catalysis of biological reactions was long considered to be the exclusive domain of proteins. However, since the discovery of RNA self-splicing activity in *Tetrahymena thermophilia* by Cech and his colleagues in 1981 [48,49], a wide variety of RNA molecules with catalytic properties have been identified. The term 'ribozyme' was first used by Cech to describe these RNAs although most are not true catalysts. The *Tetrahymena* pre-mRNA, for example, excises an intron and ligates the two exon sequences with single product turnover. However, the RNA does exhibit many

enzymatic properties: the lowering of activation energy through the formation of a precise three dimensional structure, an active site for phosphoester transfer reactions and a site for the binding of a co-factor. The situation is further complicated by the fact that cellular RNAs are generally associated with proteins. However, where proteins are necessary, it is thought that the RNA still forms the catalytic centre [50]. Often, some activity is retained in the absence of proteins with magnesium ions or spermine as a substitute [51].

Four general classes of catalytic RNA have been identified:

- ***Group I Intron*** RNAs, as found in *Tetrahymena*, undergo a self-splicing reaction in which the intron sequence is excised and the exons ligated. The reaction involves two consecutive transesterification reactions, the first being initiated by the binding of a guanosine cofactor. The self-catalysed intramolecular cyclisation of the excised intervening sequence is also common, following ligation of the exons. It is thought that the majority of group I introns require specific proteins for activity, although some are active *in vitro* with elevated Mg^{2+} concentrations [50]. Group I introns consist of four conserved sequence elements giving rise to a characteristic secondary structure, but they have little overall sequence conservation.

- ***Group II Introns*** are self-splicing RNAs found in eubacteria and eubacterial derived organellar genomes [52]. The splicing reaction occurs via a different mechanism from group I introns, proceeding though a branched lariat intermediate structure. The lariat is produced by the attack of a 2′-OH of an internal bulged A residue on the phosphate of the 5′-splice site. Thus, there is no requirement for an additional nucleotide cofactor. The mechanism is very similar to that of eukaryotic pre-mRNA splicing, except that it is an inherent property of the RNA alone, suggesting the possibility of a common origin for

the two processes. Few group II introns are active *in vitro*, without associated proteins, and those which are often require quite extreme conditions.

- ***RNase P***, an endonuclease involved in the maturation of 5′-ends of tRNAs, was the first truly catalytic RNA discovered [53]. The enzyme consists of an RNA with associated protein and requires a divalent ion, magnesium or manganese, for activity. The catalytic properties are associated solely with the RNA component [54], which is active *in vitro* in the absence of protein. The protein alone is completely inactive. Considerable primary structure variations are found between different organisms but the RNA secondary structure is well conserved.

- ***Self Cleavage Domains*** occur in a wide range of plant pathogenic RNAs, RNA transcripts from newt satellite II DNA and from *Neurospora* mitochondrial DNA plasmid. A number of distinct self cleavage domains have been identified [55], the most well characterised being the hammerhead motif. The intramolecular cleavage of the RNA proceeds by a transesterification reaction where the 2′-OH attacks the adjacent phosphodiester generating a 2′,3′-cyclic phosphate and free 5′-OH. The reaction is dependent on magnesium or other divalent cation, but has no requirement for proteins. The hammerhead motif consists of a 'conserved core' of nucleotides which may be separated into different strands, one of which contains the catalytic properties. Thus, true RNA enzymes can be generated that exhibit multiple turnover of substrate [56].

In all of these RNA cleavage reactions, the catalytic activity is only seen directed towards the cleavage of phosphodiester bonds. However, the scope of RNA as a catalyst is not limited to these reactions. The active site

of the *tetrahymena* group I intron can be engineered to catalyse the hydrolysis of an aminoacyl ester bond [57]. *In vivo*, RNAs are known to make direct contributions to substrate recognition and possibly catalysis in the ribosome [58] and in the spliceosome [59].

The specificity of ribozyme RNAs makes them potential therapeutic agents through the control of gene expression. A mechanism similar to that of antisense oligonucleotides (section 5.1.2) can be envisaged, but with the added advantage that the RNA itself is able to eliminate the target essentially irreversibly.

1.7 Project Overview

High resolution structures of RNA from NMR and crystallographic studies are still relatively sparse in the literature. Given that until very recently the preparation of sufficient quantities of pure RNA for such studies was a major hurdle, previous workers in our laboratory have had some considerable success in this area [60,61,62]. At the commencement of this project, however, reliable HPLC protocols were not available for the purification of multi-milligram quantities of RNA.

State of the art RNA synthesis and HPLC purification facilities are available in our laboratory. The first step in this project was the implementation of chemical synthesis and HPLC purification protocols for the preparation of RNA. Of course, this does not mean that these protocols were finalised before any of the biophysical studies could be conducted. Although the description of the work here may lead to an impression of a linear approach, it should be noted that the synthesis, deprotection and HPLC methods were modified continuously throughout the course of the project. This was done both in response to the challenge of increasing length and complexity of target RNA and advances published in the literature. This work is the subject of the next chapter.

Our aim was to be able to conduct thermodynamic and high resolution structural studies on two RNA systems (Chapters 4 and 5). The first involves RNA duplexes with various central A-U tracts in an examination of sequence dependent stability, hydration and structure. In DNA, these factors have been found to be important for recognition and function *in vivo*. Our second area of interest was the thermodynamic and conformational variability found in DNA.RNA hybrids. Such duplexes occur extensively in biology and are the basis of antisense technology.

1.8 References

1. Blackburn, G.M. and Gait, M.J. (1996). *Nucleic Acids in Chemistry and Biology*, 2nd Edition, Oxford University Press Inc., New York.

2. Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987). *The Molecular Biology of the Gene*, 4th Edition, The Benjamin/ Cummings Publishing Company Ltd, California.

3. Gestland, R.F. and Atkins, I.F. (Ed.) (1993). *The RNA World*, Cold Spring Harbour Laboratory Press, New York.

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

5. Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Moras, D., deBear, J.S. Gough, G.R., Gilham, P.T., and Johnson, J.E. (1988). *Nature*, **335**, 375-378.

6. Westhof, E., Dumas, P. and Moras, D. (1985). *J. Mol. Biol.*, **184**, 119-145.

7. Tavale, S.S. and Sobell, H.M. (1970). *J. Mol. Biol.*, **48**, 109-123.

8. Kennard, O. and Hunter, W.N. (1991). *Angew. Chem. Int. Ed. Engl.*, **30**, 1254-1277.

9. Wang, A.H.-J., Quigley, G.J., Kolpack, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979), *Nature*, **282**, 680-686.

10. Viswamitra, M.A., Kennard, O., Jones, P.G., Sheldrick, G.M., Salisbury, S.A., Favello, L. and Shakked, Z. (1978). *Nature*, **273**, 687-688.

11. Gottesfeld, J.M., Blanco, J., Tennant, L.L. (1987). *Nature*, **329**, 460-462.

12. Drew, H.R. and Dickerson, R.E. (1981). *J. Mol. Biol.*, **151**, 535-556.

13. Huang, S., Wang, Y.-X. and Draper, D.E. (1996). *J. Mol. Biol.*, **258**, 308-321.

14. Laing, L.G., Gluick, T.C. and Draper, D.E. (1994). *J. Mol. Biol.*, **237**, 557-587.

15. Lu, M. and Draper, D.E. (1994). *J. Mol. Biol.*, **244**, 572-585.

16. Olsen, G.J. and Woese, C.R. (1993). *FASEB Journal*, **7**, 113-123.

17. Michel, F. and Westhof, E. (1990). *J. Mol. Biol.*, **216**, 585-610.

18. Abrahams, J.P., van den Berg, M., van Batenberg, E. and Pleij, C. (1990). *Nucleic Acids Res.*, **18**, 3035-3044.

19. Gutlyaev, A.P., van Batenberg, F.H.D. and Pleij, C.W.A. (1995). *J. Mol. Biol.*, **250**, 37-51.

20. Ogata, H., Akiyama, Y. and Kanehisa, M. (1995). *Nucleic Acids Res.*, **23**, 419-426.

21. Hubbard, J.M. and Hearst, J.E. (1991). *J. Mol. Biol.*, **221**, 889-907.

22. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J.-P. and Ehresmann, B. (1987). *Nucleic Acids Res.*, **15**, 9109-9128.

23. Knapp, G. (1989). *Meth. Enzymol.*, **180**, 192-212.

24. Puglisi, J.D. and Tinoco, I., Jr. (1989). *Meth. Enzymol.*, **180**, 304-325.

25. Cantor, C.R. and Schimmel, P.R. (1980). *Biophysical Chemistry, Part III*, W.H. Freeman and Company, San Fransico.

26. Breslauer, K.J., Sturtevant, J.M. and Tinoco, I., Jr. (1975). *J. Mol. Biol.*, **99**, 549-565.

27. Cole, P.E., Yang, S.K. and Crothers, D.M. (1972). *Biochemistry*, **11**, 4358-4374.

28. Laing, L.G. and Draper, D.E. (1994). *J. Mol. Biol.*, **237**, 560-576.

29. Ryan, P.C., Lu, M. and Draper, D.E. (1991). *J. Mol. Biol.*, **221**, 1257-1268.

30. Xing, Y. and Draper, D.E. (1995). *J.Mol. Biol.*, **249**, 319-331.

31 Vorlickova, M., Kypr, J., Jovin, T.M. and Planck, M. (1990). *Biopolymers*, **29**, 385-392.

32. Ivanov, V.I., Minchenkova, L.E., Minyat, E.E., Frank-Kamenetskii, M.D. and Schyolkina, A.K. (1974). *J. Mol. Biol.* **87**, 817-833.

33. de Leeuw, F.A.A.M. and Altona, C. (1982). *J. Chem. Soc., Perkin Trans.* 2, **3**, 375-384.

34. Ramesh, V. (1993). *Nucleic Acids Res.*, **21**, 5485-5488.

35. Moore, P.B. (1995). *Acc. Chem. Res.*, **28**, 251-256.

36. Gorenstein, D.G. (1992). *Meth. Enzymol.*, **211**, 254-286.

37. Collier, A.K., Arnold, J.R.P. and Fisher, J. (1996). *Mag. Reson. Chem.*, **34**, 191-196.

38. Glusker, J.P. and Trueblood, K.N. (1985). *Crystal Structure Analysis*, Oxford University Press, New York.

39. Hahn, T. (ed) (1987). *International Tables for X-ray Crystallography*. D. Reidel Publishing Co., Dordrecht, Netherlands.

40. Sawyer, L. and Turner, M.A. (1992). In *Crystallization of Nucleic Acids and Proteins*, Ducruix, A. and Giegé, R. (ed), IRL Press at Oxford University Press, New York.

41. Draper, D.E. (1995). *Annu. Rev. Biochem.*, **64**, 593-620.

42. Seeman, N.C., Rosenberg, J.M. and Rich, A. (1976). *Proc. Natl. Acad. Sci. USA*, **73**, 804-808.

43. McClarin, J.A., Frederick, C.A., Wang, B.-C., Greene, P., Boyer, H.W., Grable, J. and Rosenberg, J.M. (1986). *Science*, **234**, 1526-1541.

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

45. Juo, Z.S., Chui, T.K., Leiberman, P.M., Baikalov, I., Berk, A.J. and Dickerson, R.E. (1996). *J. Mol. Biol.*, **261**, 239-254.

46. Weeks, K.M. and Crothers, D.M. (1993). *Science*, **261**, 1574-1577.

47. Varani, G., Cheong, C. and Tinoco, I. (1991). *Biochemistry*, **30**, 3280-3289.

48. Cech, T.R., Zaug, A.J. and Grabowski, P.J. (1981). *Cell*, **27**, 487-496.

49. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982). *Cell*, **31**, 147-157.

50. Cech, T.R. (1990). *Annu. Rev. Biochem.*, **59**, 543-568.

51. Uhlenbeck, O.C., (1991). *Curr. Op. Str. Biol.*, **1**, 459-463.

52. Michel, F. and Ferat, J.-L. (1995). *Annu. Rev. Biochem.*, **64**, 435-461.

53. Guerrier-Takade, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1993). *Cell*, **35**, 849-857.

54. Altman, S. (1990). *Angew. Chem. Int. Ed. Eng.*, **29**, 749-758.

55. Symons, R.H. (1992). *Annu. Rev. Biochem.*, **61**, 641-671.

56. Uhlenbeck, O.C. (1987). *Nature*, **328**, 598-600.

57. Piccirilli, J.A., McConnell, T.S., Zaug, S., Noller, H.F and Cech, T.R (1992). *Science*, **256**, 1420-1424.

58. Noller, H.F. (1991). *Annu. Rev. Biochem.*, **60**, 191-227.

59. Guthrie, C. (1991). *Science*, **253**, 157-163.

60. Lane, A.N., Ebel, S. and Brown, T. (1993). *Eur. J. Biochem.*, **215**, 297-306.

61. Ebel, S., Brown, T. and Lane, A.N. (1994). *Eur. J. Biochem.*, **220**, 703-715.

62. Leonard, G.A., McAuley-Hecht, K.E., Ebel, S., Lough, D.M., Brown, T. and Hunter, W.N. (1994). *Structure*, **2**, 483-494.

Chapter 2

Preparation of Synthetic Oligoribonucleotides

2. Preparation of Synthetic Oligoribonucleotides

2.1 Introduction

In order to conduct the physical studies on RNA by UV melting, and in particular NMR and X-ray crystallography, sufficient quantities of pure material were required. The synthesis and purification protocols developed to meet this demand are described in this chapter. The methods chosen were primarily driven by the equipment available and the current expertise in the laboratory.

2.1.1 Methods of Synthesis

The method of RNA synthesis for biochemical and biophysical studies depends largely on the aim of the experiment and the demands that it places on the material required. Traditionally, short RNA sequences were either isolated from cellular extracts or synthesized enzymatically, using a bacteriophage RNA polymerase, such as T7, and nucleoside triphosphates [1,2]. More recently, solid phase chemical synthesis of RNA has become possible. The development of protocols for chemical RNA synthesis has lagged behind that of DNA primarily due to the problem of protecting the extra 2'-hydroxyl. Each method has its advantages and disadvantages, and a crucial part of any investigation into the structure and function of RNA is deciding which is most suitable.

2.1.1.1 Enzymatic RNA Synthesis

RNA oligonucleotides can be synthesized enzymatically using nucleoside triphosphates (NTPs) and a DNA template [1]. Highly active RNA

polymerases have been isolated from bacteriophages such as T7 and SP6, and both are now commercially available. The DNA template can be prepared synthetically or from linearised plasmids, into which the sequence of interest has been cloned. The quality of RNA obtained is the same for each method but the latter, although more labour intensive, places no limit on the length of the template. Whichever method is used, the template must include a 'promoter region' to which the enzyme binds. This region must be double stranded but the sequence following it is single stranded and can be varied (Figure 2.1). Once initiation has taken place, the RNA is synthesized from 5′ to 3′.

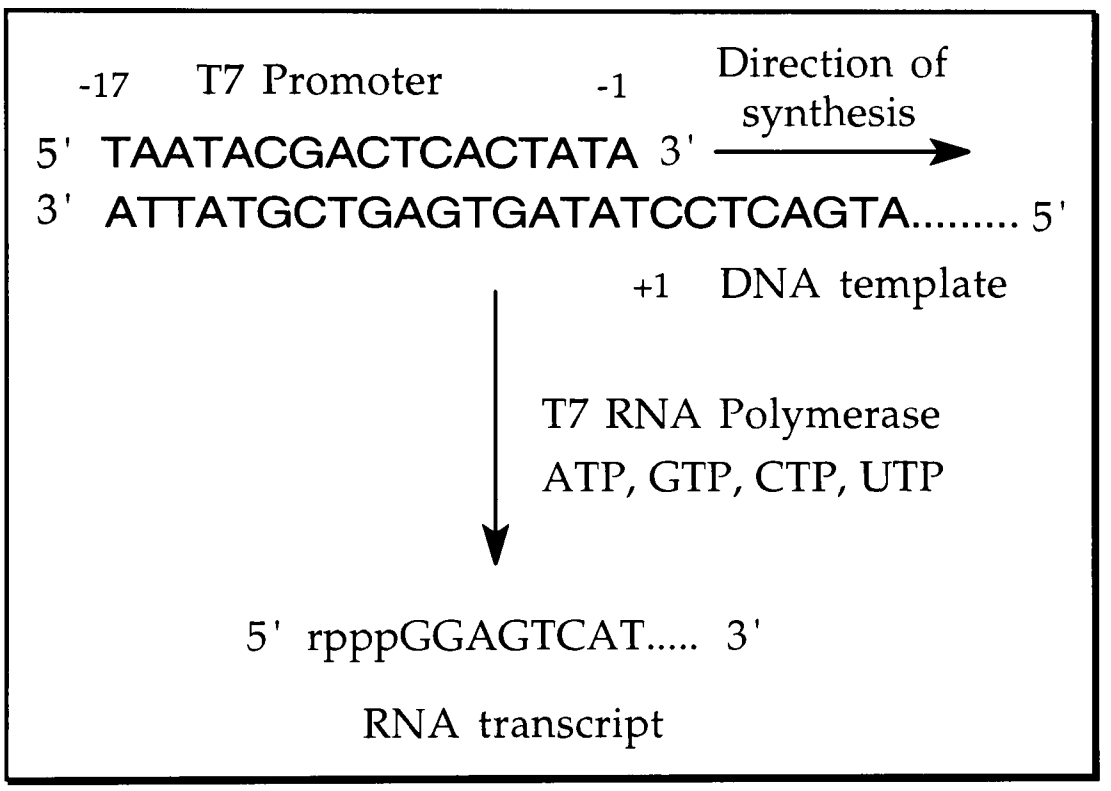


Figure 2.1 Enzymatic RNA synthesis using T7 RNA polymerase

Enzymatic RNA synthesis has several features that make it an attractive method:

- the promoter region is determined by the enzyme and does not vary.
- long RNA sequences can be synthesized without loss of yield.
- 1 mole of template gives thousands of moles of RNA.
- some modified nucleotides can be incorporated.

However, there are several major disadvantages which make this method less suitable for preparing RNA for some biophysical studies:

- conditions must be optimised for each reaction.
- initiation is the limiting step and aborted reactions give around 50% of the material as 2-6 nucleotide fragments.
- dinucleotides can be used to initiate polymerisation but their effect varies with the enzyme/ promoter system used.
- sequence heterogeneity can occur through mis-incorporation by the enzyme or heterogeneity in the template.
- some sequences are not synthesized as well as others, for example regions with more than eight A or U residues [3].
- generally more than one product of the approximate size expected is obtained, so that sequencing is often required to determine which is the desired product.
- modified nucleotides may not be incorporated, or may be incorporated but not initiate transcription (for example inosine [4]).
- modified nucleotides are incorporated non-specifically.

Clearly, for the type of investigations proposed here, enzymatic synthesis is not a viable proposition. It would not provide a simple route to the large amounts of highly pure material required for NMR and X-ray crystallography. A way around some of the major problems was devised by Price *et al.*, in their synthesis of RNA for crystallographic studies [5]. They transcribed the desired sequence flanked by two ribozymes, so that it was automatically excised in the presence of magnesium ions. This means that the same optimised conditions can be used for different sequences. Purification is also simplified as the fragments produced are of considerably different sizes. It does not, however, overcome the most significant problem, the site specific introduction of modifications.

2.1.1.2 Chemical Synthesis

The chemical synthesis of RNA involves the condensation of suitably protected activated nucleotides to a growing chain. Several groups have reported successful solution phase syntheses [6,7] of short (2-5 mer) oligoribonucleotides and their subsequent ligation with T4 RNA ligase. Not surprisingly, however, the yields were very low and the process time consuming and laborious. Clearly, this method is not a viable route to the large quantities of material required for biophysical studies, but it did show that chemically synthesized RNA is biologically active [8].

Solid phase synthesis was pioneered by Merrifield for the synthesis of peptides and was subsequently successfully applied to nucleic acid synthesis [9]. The major advantage of solid phase over solution reaction is that it allows the use of large excesses of reagents which are removed simply by washing the resin. This helps to drive reactions to completion and alleviates the need for any stepwise purification.

The most popular method of solid phase oligonucleotide synthesis involves the use of β-cyanoethyl phosphoramidite monomers and is the subject of a thorough review by Beaucage and Iyer [10]. The first nucleotide is attached to a non-swelling solid support, such as controlled pore glass (CPG) or a polystyrene derivative, via a linker, at the 3′ position. This, and subsequent nucleotides added in solution, must have orthogonal 5′-OH and base protecting groups so that selective coupling at the 5′-position of the growing chain can be achieved. These are usually an acid labile trityl type group on the 5′-OH and a base labile acyl on the exocyclic amines of A, G and C. The additional problem in the synthesis of RNA is finding a protecting group for the 2′-OH, which must be:

- easily and specifically introduced (2′ only).

- unable to migrate under the conditions used in the preparation of the phosphoramidite monomers.
- completely stable to the conditions of chain extension, cleavage from solid support and deprotection of phosphate and base protecting groups.
- completely and easily removable under conditions in which the RNA is stable.

A wide variety of different protecting groups have been suggested, but very few meet the criteria above fully. In 1967, Reese first suggested the use of acid labile acetal protection for the 2′-hydroxyl in oligoribonucleotide synthesis with the methoxytetrahydropyranyl group (Mthp, Figure 2.2) [11]. The major advantage of this type of protection is that it can be specifically introduced at the 2′-position following temporary blocking of the 3′- and 5′-OH using a Markiewicz reagent. Furthermore, once introduced, these groups are totally non-migratory. Unfortunately, this group proved to be too labile to withstand the repeated exposure to the relatively harsh acidic conditions required for 5′-hydroxyl deprotection [12,13]. To overcome this problem Reese and co-workers began designing new protecting groups where the ring oxygen was replaced with a tertiary amino function. They reasoned that by carefully varying the pK_a of this function they could obtain a protecting group protonated at the nitrogen and thus sufficiently stable in the pH range 0 to 2.5, but susceptible to hydrolysis under milder conditions [14]. This work led to the development of the 1-(2-chloro-4-tolyl)-4-methoxypiperidin-4-yl (Ctmp) and 1-(2-fluorophenyl)-4-methoxypiperidin-4-yl (Fpmp) protecting groups (Figure 2.2). Protonation at the nitrogen under the conditions of 5′-OH deprotection gives them the desired stability but at higher pH, they are readily hydrolysed. Due to its greater ease of synthesis, it is the Fpmp protecting group that has become commercially available and the most widely used of this type.

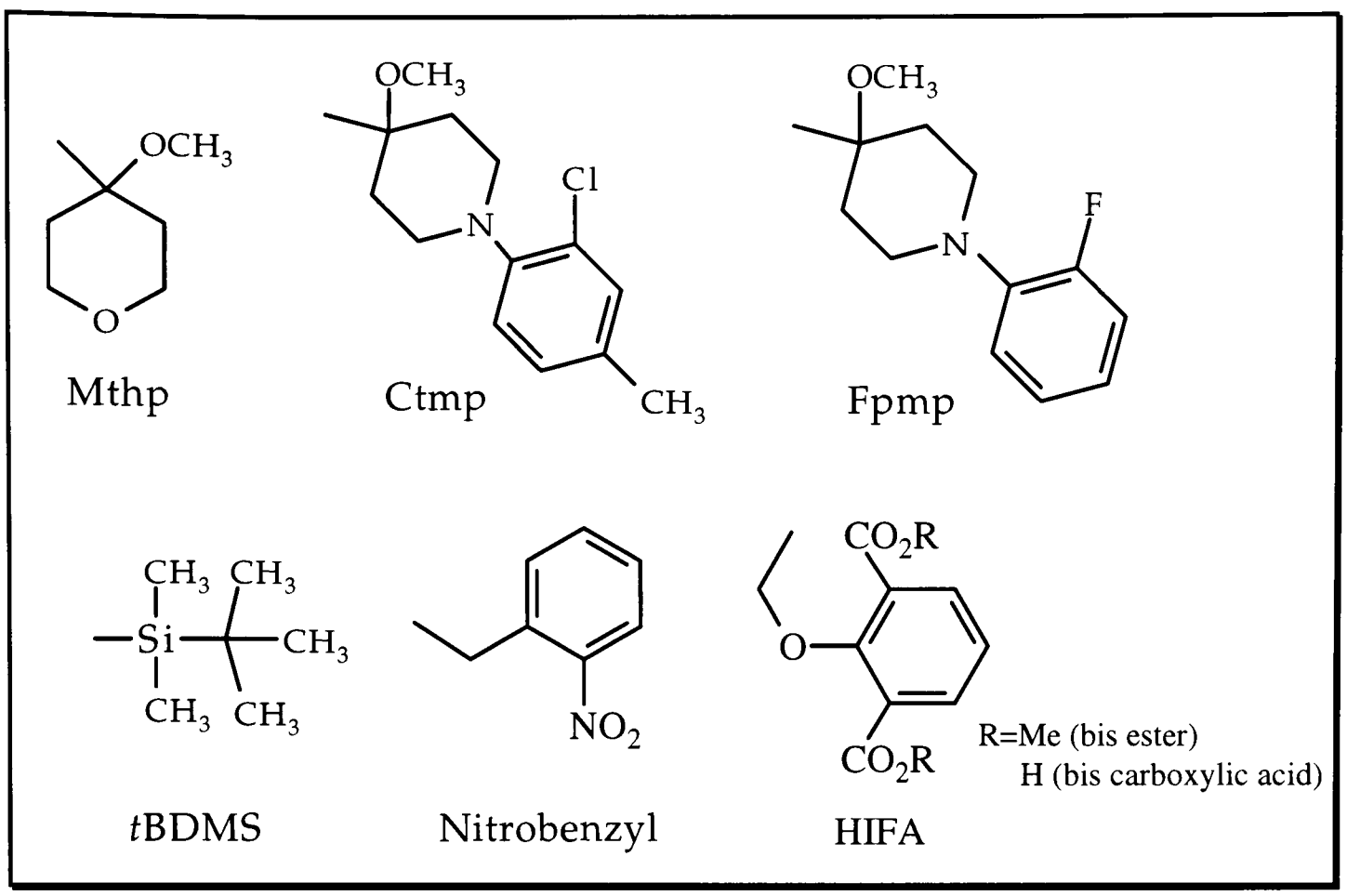


Figure 2.2 Protecting groups for the 2'-hydroxyl

An alternative strategy involves the use of alkylsilyl ether protecting groups for the 2′-OH, first introduced by Ogilvie in the 1970s [15]. Phosphoramidite monomers containing protecting groups of this type have been shown to be suitable for solid phase synthesis [16] and have subsequently been applied to the preparation of biologically active RNA [17]. The most commonly used group is *tertiary*-butyldimethylsilyl (tBDMS, Figure 2.2) which is compatible with the protocols used for the synthesis of DNA oligonucleotides, requiring only an extended coupling time. The major problems associated with this method are in the preparation of the phosphoramidite monomers. The tBDMS group is introduced non-specifically to 2′- and 3′-hydroxyl groups and the desired isomer obtained by silica column chromatography. Furthermore, the 2′-O-tBDMS isomer may be susceptible to migration under the conditions used for phosphitylation of the 3′-hydroxyl [18]. However, the isomeric integrity of 2′-O-tBDMS cyanoethyl phosphoramidite monomers has been

demonstrated by NMR [16]. Despite these difficulties in the preparation of tBDMS phosphoramidites, this method has now become by far the most popular and widely available commercially.

A number of other 2'-hydroxyl protecting groups have been investigated, such as the photolabile nitrobenzyl group [19,20] (Figure 2.2). While this group proved useful in the synthesis of short oligonucleotides, its complete removal was found to be difficult in longer syntheses [21]. A recent development in the design of acid labile groups has suggested that this type of protection may still prove to be the most suitable for solid phase RNA synthesis [22]. The new group, 2-hydroxyisophthalate formaldehyde acetal (HIFA, Figure 2.2) is stable under the conditions of detritylation as the bis ester. During the removal of base and phosphate protecting groups in alkali, the bis ester is converted into the bis carboxylic acid, allowing the subsequent 2'-deprotection with dilute acid. The results presented showed no significant improvement over the currently available Fpmp method and no oligonucleotides longer than dimers were synthesized. However, it is likely that, with some refinement, this type of protecting group could rival the tBDMS group as the most suitable for the protection of the 2'-hydroxyl in solid phase RNA synthesis.

2.1.2 Purification Techniques

Whichever method is used to synthesise RNA, some form of purification will be required to obtain a homogeneous sample. The two main tools available are described in detail below. Both have a number advantages and disadvantages in certain circumstances, which will determine the method most suitable for a particular application.

2.1.2.1 Polyacrylamide Gel Electrophoresis

Polyacrylamide gel electrophoresis (PAGE) is a standard tool for the molecular biologist and has been widely used in the purification of biological macromolecules, including short nucleic acids [23]. The technique involves the migration of nucleic acids through a buffered gel matrix under the influence of an applied electrical field. The separation is therefore on the basis of charge, which, at neutral pH, is directly proportional to the number of phosphate groups. The migration can be affected by the conformation of the molecule, so a denaturant such as urea is added. (This effect can be used to study the hydrodynamic properties of nucleic acid duplexes on non-denaturing gels, as discussed in Chapter 5.) The effective range of separation of a gel is determined by the length of the chains and the degree of crosslinking. A 20% acrylamide gel is generally used for strands of 6-100 base pairs but can be as low as 3.5% for 1000-2000 base pairs. Within these ranges, strands that differ in length by as little as 0.2% (or 1 base pair in 500) can be resolved, making PAGE a potentially powerful purification technique. It is most useful for the purification of small amounts of enzymatically or chemically synthesized RNA, particularly that which has been radiolabelled. However, it does have a number of significant disadvantages in the preparation of large amounts of material for biophysical studies:

- the loading capacity is low (in the μg range), making the purification of large amounts time consuming and laborious.
- resolution can be impaired by overloading the gel, resulting in contamination of the full length product with deletion sequences.
- the desired band on the gel must be located under UV light and cut from the gel in a short time to minimise exposure.
- for chemically synthesized RNA, incompletely deprotected oligonucleotides run as a smear behind the main band.

- the RNA must be recovered from the gel either by crushing and soaking in buffer, or by electroelution. Either way, large amounts of material are likely to be lost, particularly for shorter sequences.
- acrylamide contaminants from the gel can be difficult to remove and can cause problems in NMR and hinder crystallization.

2.1.2.2 High Performance Liquid Chromatography

A more recent development in the purification of oligonucleotides is the use of high performance liquid chromatography (HPLC). The development of HPLC methods has been hindered mainly by the lack of suitable stationary phase materials. Traditionally, only semi-rigid or non-rigid gels with a low mechanical strength have been available. The low flow rates and large particle size associated with these materials led to poor resolution and long elution times [24]. More suitable stationary phase materials have now been commercially developed and allow the application of a number of chromatographic techniques to the purification of nucleic acids. It is unlikely that chemical synthesis of DNA and RNA would have been as successful without these developments.

The basic principle of HPLC is that there should be a separation of the components of a solution by adsorption and/ or partition between the column material (stationary phase) and a solvent (mobile phase). The two most successful techniques applied to nucleic acids are reversed phase and ion-exchange HPLC. Both are discussed in greater detail in section 2.3. Size exclusion chromatography, often used in protein purification, has found only limited use in nucleic acid purification and is generally only used to remove unwanted salts and small molecules. HPLC techniques have a number of major advantages over PAGE:

- HPLC has a much higher loading capacity than PAGE and different column sizes may be used according to the scale of the work.

- HPLC is very fast, typical run times being measured in minutes rather than hours.
- the absorbance maximum for most nucleic acids is between 260 and 280nm allowing accurate detection of product elution through an in-line UV detector and flow cell. Data from this may be captured using a PC with appropriate software and subjected to more detailed analysis.
- buffer systems have been developed that allow the simple removal of residual salts by evaporation, gel filtration or sublimation. Furthermore, the majority of these do not pose significant problems for NMR or crystallization.
- HPLC protocols are easily optimised to improve the resolution of deletion sequences and incompletely deprotected species such as 2′-O-silylated RNA.

Clearly HPLC is the most powerful technique for the purification of short to medium length nucleic acids. Material of much higher purity can be obtained in a fraction of the time than with PAGE. As with any technique, however, there are certain cases where HPLC is less useful:

- long nucleic acids can exhibit extensive and stable secondary structures, which can produce anomalous elution profiles. This can be overcome by denaturing DNA with high pH buffers (for example, containing 1.0M NaOH). For RNA the use of high pH is not possible but chaotrophic agents, such as urea, may be added to the loading buffer and the HPLC column heated to prevent refolding.
- in the purification of radiolabelled RNA, where it is important to keep the material contained, PAGE is considerably safer.

For the preparation of RNA for structural and thermodynamic studies, it is clear that chemical synthesis coupled with a reliable HPLC purification protocol is the method of choice. It was an absolute requirement for the

work described here that modifications to RNA sequences could be made site specifically. Furthermore, rapid purification protocols were required to prepare the relatively large amounts of pure RNA required for NMR and crystal growth.

2.2 Chemical Synthesis

The automated solid phase synthesis of DNA is now well established. For the studies described here, DNA oligonucleotides were produced using standard synthesis cycles and deprotection protocols [9]. Two different precursors for RNA synthesis were commercially available at the commencement of this project, and the use of both of these is described here. Both are designed to be compatible with existing protocols for DNA synthesis, which with the use of the Applied Biosystems 394 eight port synthesizers, available in our laboratory, allows the synthesis of chimeric oligonucleotides. The two different strategies for RNA synthesis were applied to the synthesis of a number of oligoribonucleotides, allowing a critical comparison of the two methods (section 2.6.2).

2.2.1 Synthesis Precursors: Protected RNA Phosphoramidites

The two different types of RNA precursor used differ mainly in the 2′-hydroxyl protecting group. The 5′-hydroxyl and phosphorous protecting groups are the same: dimethoxytrityl (DMT) and cyanoethyl-N,N-diisopropyl phosphoramidite (CEP) respectively (Figure 2.3). Protecting groups are also required for the amine groups of the bases, except uridine which has none. There are a number of possible strategies, and the one used generally depends on the other protecting groups. The standard protecting group strategy is benzoyl (Bz) for A and C and isobutyryl (iBu) for G (Figure 2.4). These are used for the Fpmp protected phosphoramidite monomers, as it is completely stable during the long basic deprotection step. Applied Biosystems recently launched a new range of 'fast base deprotecting groups' for use with tBDMS protecting groups. These are N-phenoxyacetyl A (N-PAC A), dimethyl formamidine G (DMF G) and isobutyryl C (Figure 2.4). These phosphoramidite monomers were used for all RNA syntheses using 2′-O-tBDMS protection.

RNA phosphoramidites using only N-PAC base protection are also commercially available with tBDMS 2′-hydroxyl protection. However, these were found to give significantly poorer results both in terms of stability in storage and coupling yields. Synthesis of oligoribonucleotides containing modified nucleosides were all made with the tBDMS protection, a reflection of the wider commercial availability of precursors for this approach.

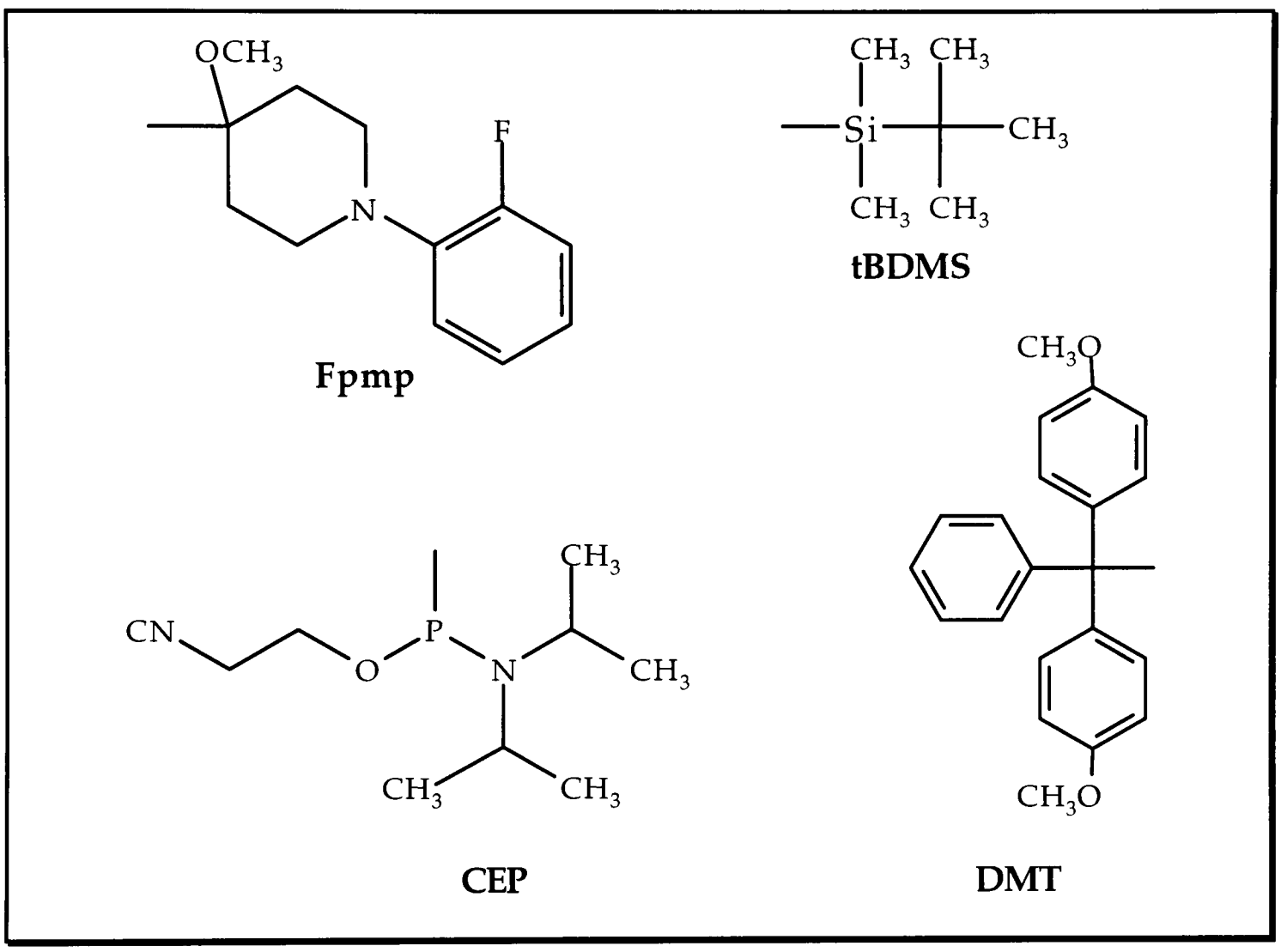
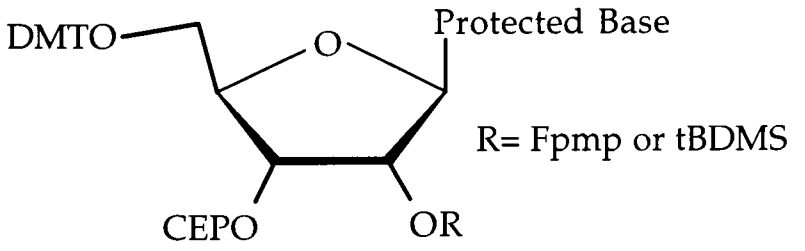


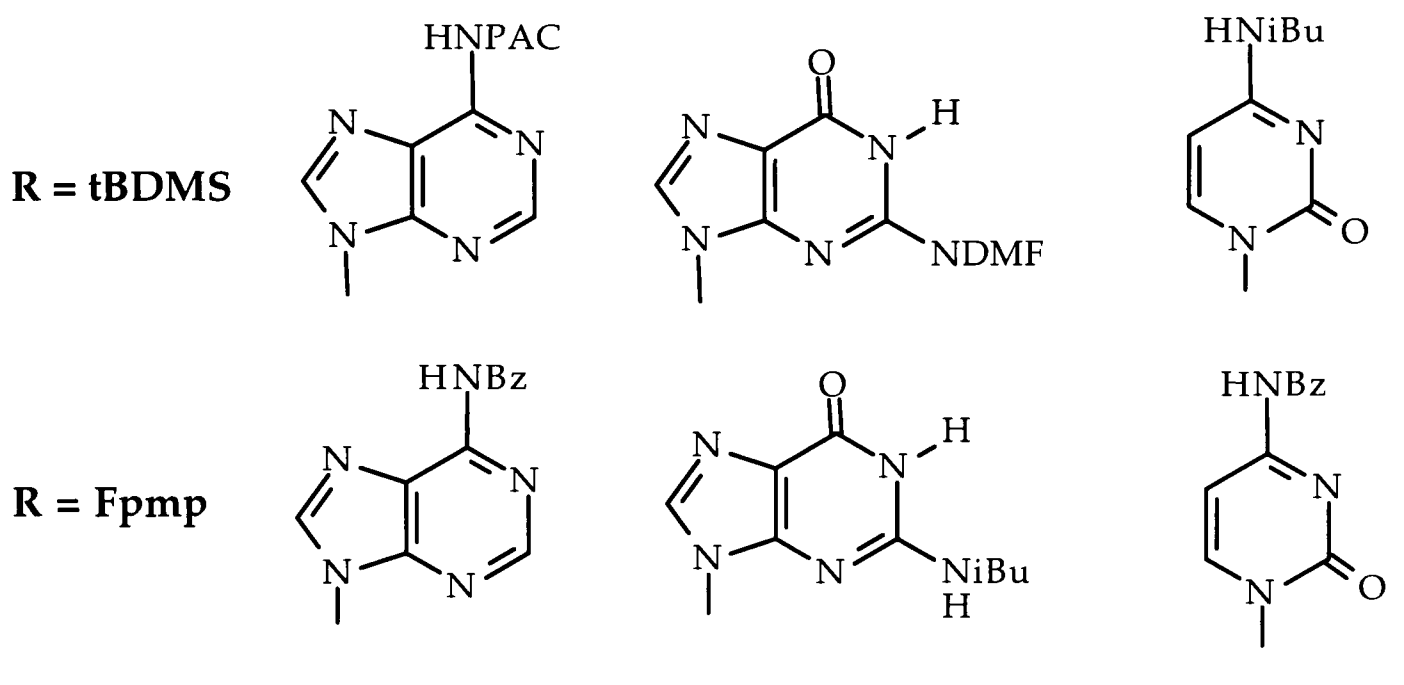
Figure 2.3 Sugar and phosphorous protecting groups.

2.2.2 The Solid Support

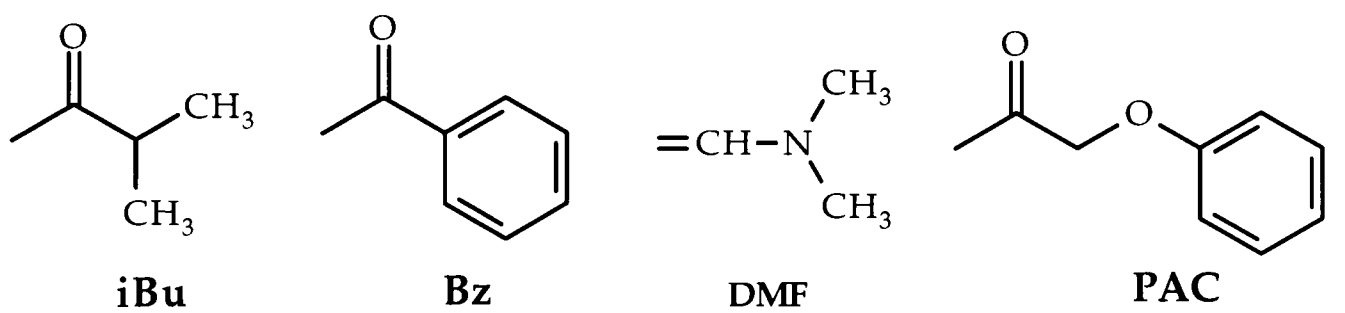
Traditionally, controlled pore glass (CPG) solid supports have been used for DNA synthesis and these have been commercially available with derivatised RNA nucleotides for some time. Different pore sizes (500Å or 1000Å) can be used depending on the scale and length of the synthesis.



(a) A fully protected RNA phosphoramidite monomer



(b) Protected bases for the tBDMS and Fpmp methodologies



(c) Structures of the base protecting groups

Figure 2.4 The base protecting group strategies

More recently, derivatised polystyrene supports used in RNA synthesis [25] have become commercially available. This type of support is reported to have improved liquid flow-through properties [26,27]. The 2′-hydroxyl

is protected with an acetyl group giving improved oligonucleotide cleavage from the support [28] and making it compatible with both types of monomer.

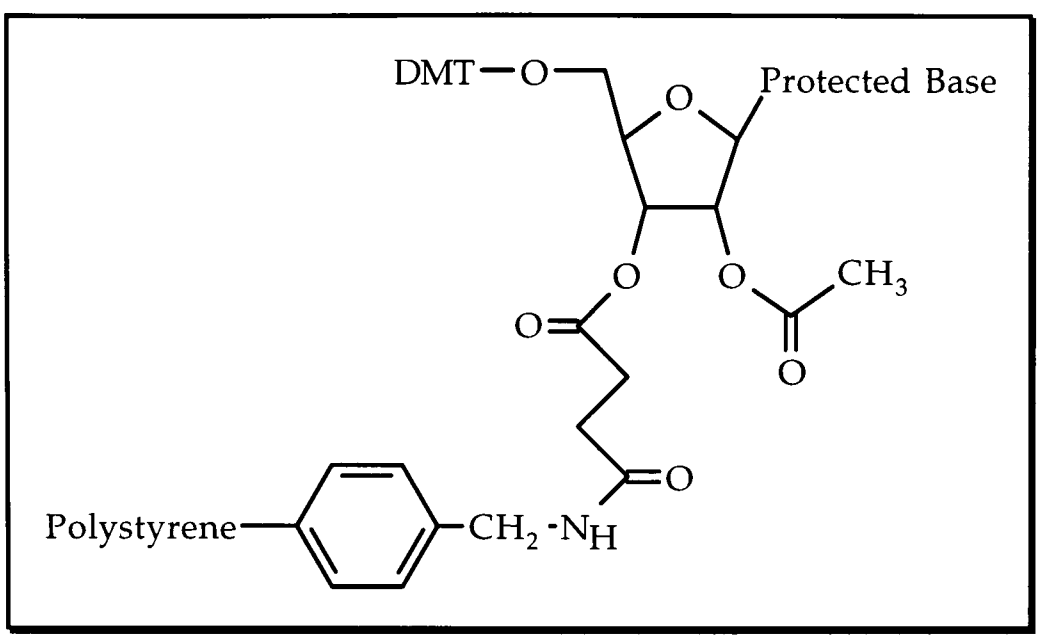


Figure 2.5 Derivatised Polystyrene Support

Polystyrene supports were used for the majority of the oligoribonucleotide syntheses in this work. Their use required a number of minor modifications to the standard Applied Biosystems 394 synthesizer RNA cycle, partly to overcome the effect of the larger body size of the column casing and volume of resin:

- acetonitrile wash at the start of the cycle was increased to 30s.
- monomer delivery time was increased to a total of 6s.
- iodine delivery was increased to 12s, followed by a 2s reverse flush and a further delivery of iodine of 2s.
- an extra delivery of capping reagent was included after the iodine oxidation to help dry the resin.

2.2.3 The Synthesis Cycle

The synthesis cycle (Figure 2.6) consists of four main steps, which are repeated in sequence for each nucleotide addition:

1. Deprotection of the 5′-hydroxyl
2. Coupling of the free nucleotide
3. Capping of unreacted chains
4. Oxidation of the phosphorous group

The first step involves the removal of the terminal 5′-hydroxyl protecting group on the growing oligonucleotide chain. This is achieved using 2% trichloroacetic acid in dichloromethane (TCA/ DCM). The reaction is forced to completion by repeated washing of the solid support with the TCA/ DCM solution. The protecting group is released as the trityl cation, a bright orange colour. The release can be monitored manually using the absorbance of the solution at 498nm, or automatically using conductivity measurements. Comparison of the measurements for each cycle gives an average stepwise yield. Both the Fpmp and tBDMS monomers consistently coupled at better than 98% per cycle, generally at about 99%. It was found that the automated monitor occasionally gave very low or very high readings (85-90% or 100%). However, there was rarely any difference in the quality of the synthesis as observed by HPLC (section 2.3). The problem was attributed to either too high or too low a removal of the DMT cation in an initial TCA/ DCM wash *before* the monitoring period. This was confirmed simply by observing the intensity of colour of the trityl release.

The next nucleotide in the sequence is then added in solution, as the phosphoramidite monomer, with tetrazole. The tetrazole acts as an activator by displacing the more stable diisopropylamine group, facilitating coupling with the free 5′-OH on the growing oligonucleotide

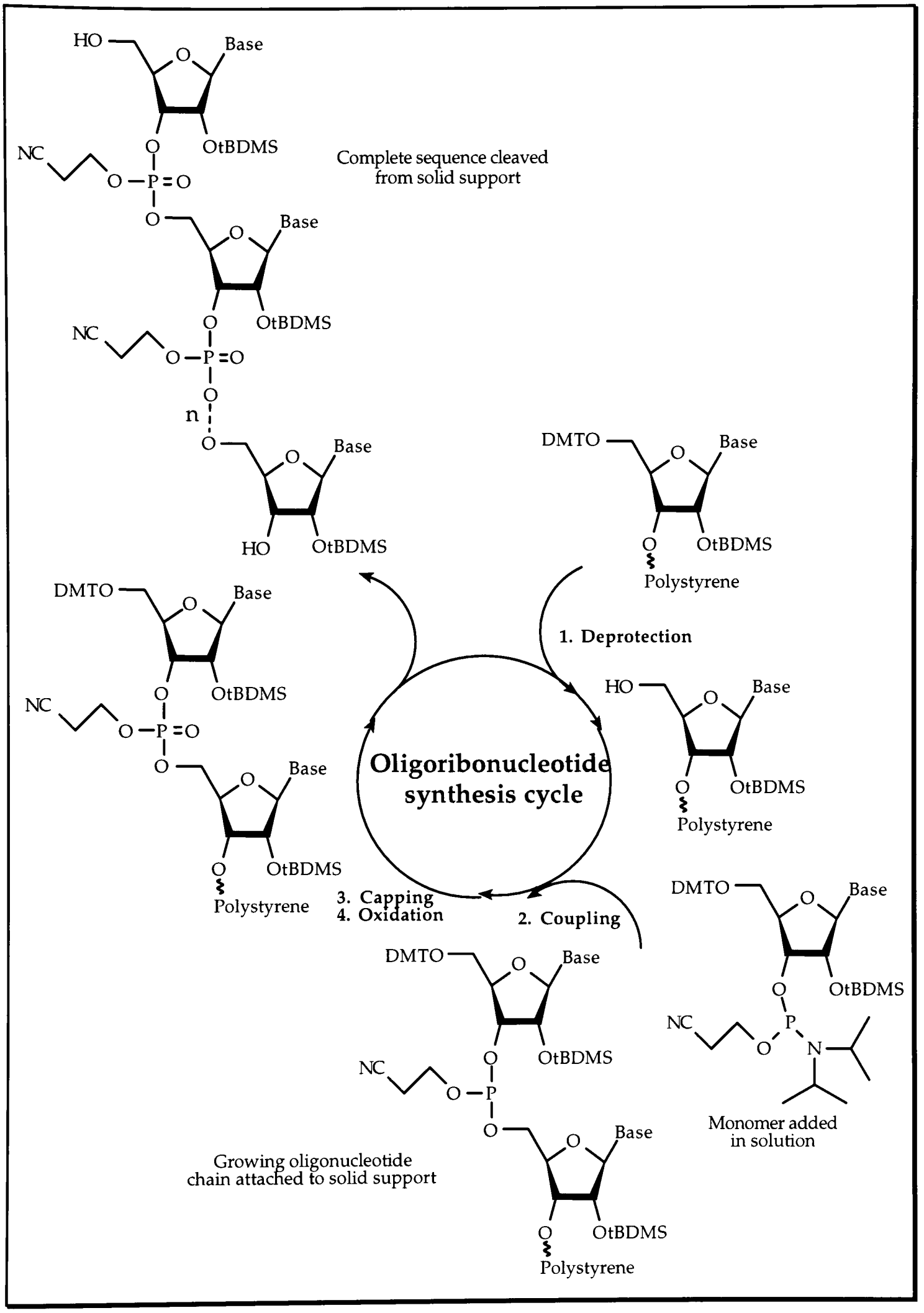


Figure 2.6 The RNA Solid Phase Synthesis Cycle (shown with 2′-O-tBDMS type phosphoramidites).

chain. A large excess of phosphoramidite and tetrazole ensure that the reaction is driven as near to completion as possible. The coupling time of RNA phosphoramidites needs to be considerably longer than for DNA, mainly due to the steric bulk of the 2'-OH protecting group. The RNA synthesis cycle was modified to give a coupling time of 600s, with a 1s acetonitrile delivery followed by a 1s reverse flush half way through. This was designed to help the mixing of the reactants in contact with the solid support.

As mentioned, coupling efficiencies for RNA phosphoramidites were generally very high (98% or better). Even so, at each step a small number of 'deletion sequences' are generated where the coupling has failed. These sequences must be 'capped' to prevent propagation of these chains. Otherwise, the major impurity would be only one nucleotide less than the full length product, which, even for a moderate length of synthesis, would be almost impossible to remove completely.

Any free 5'-OH groups are acetylated using a mixture of acetic anhydride and N-methyl imidazole, the reactive intermediate being generated *in situ*. The reaction time is brief to minimise possible side reactions at the base or cyanoethyl protecting groups.

The final step is the oxidation of the phosphite to give the stable phosphate. This is done using a mixture of iodine, pyridine and water in THF. Other reagents can be used, for example in the preparation of phosphorothioate oligonucleotides [29]. The iodine-pyridine complex forms an adduct with the phosphite which then reacts with water. This step follows the capping reaction to prevent any hydrolysis of the anhydride, which would reduce its efficiency. The brief wash with capping reagent which *follows* the oxidation step is designed to completely dry the resin.

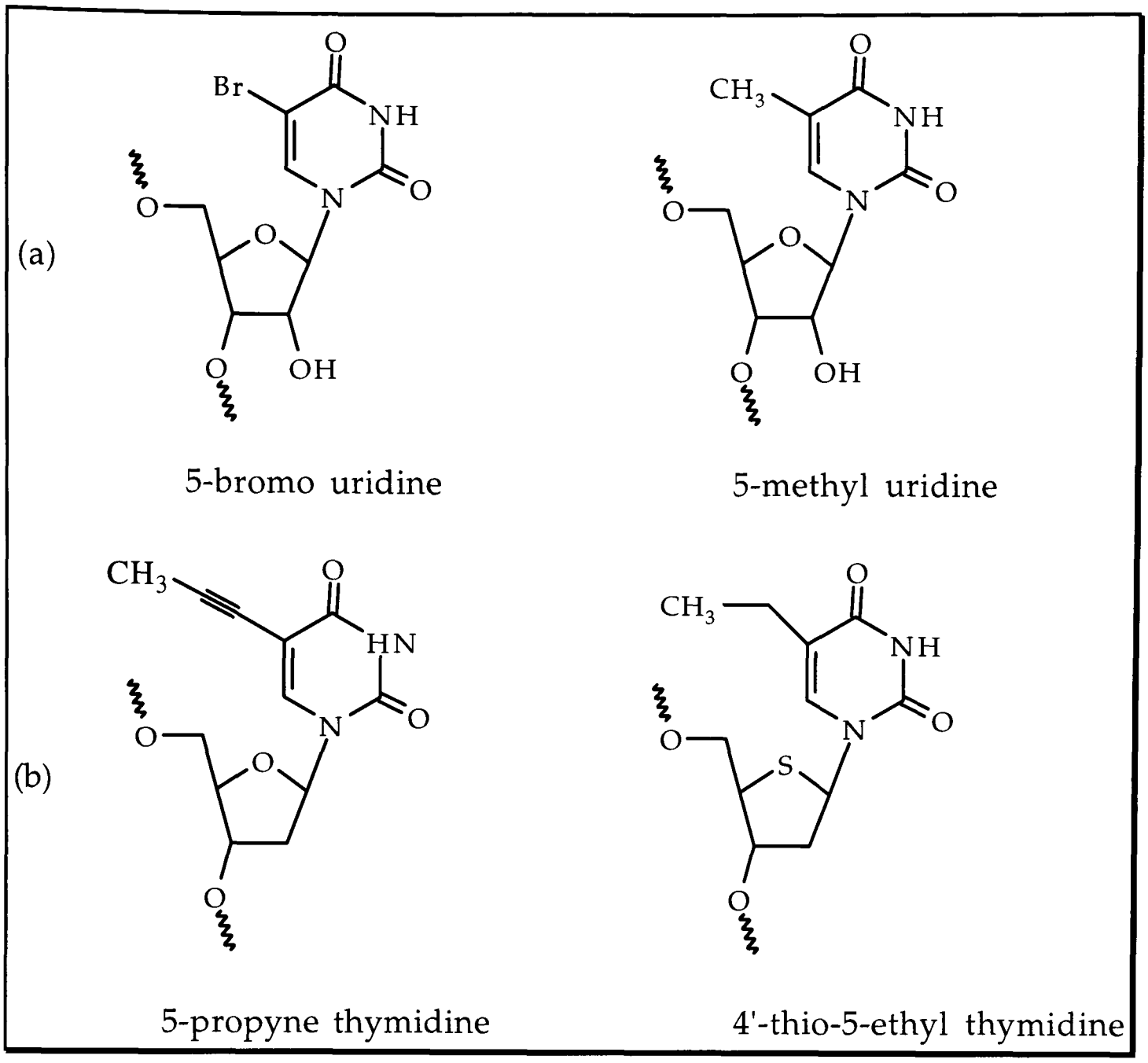


Figure 2.7 Modified bases incorporated into (a) RNA and (b) DNA by solid phase synthesis

The major advantages of solid phase chemical synthesis are immediately obvious: it allows the synthesis of any moderate length RNA sequence at will and the site specific incorporation of modified nucleotides. These can be modifications to the base (for example alkylation, halogenation, etc), sugar (*deoxy*nucleotides) or labelling molecules (such as biotin, fluorescene, or dinitrophenol). A number of modified bases have been used in this work to study their effects on nucleic acid structure and stability (Figure 2.7).

2.2.4 Deprotection Procedures

Once the synthesis cycle is complete, the RNA must be cleaved from the solid support and the protecting groups removed. First, the base and phosphorus protecting groups are removed. Where 2′-O-Fpmp protection has been used the 5′-terminal trityl can be left intact on the full length product. This allows an initial 'trityl-on' reversed phase HPLC purification (section 2.3.1). The second deprotection step is the removal of the 2′-protecting group.

2.2.4.1 Cleavage From the Solid Support and Base Deprotection

Cleavage from the solid support and removal of both the cyanoethyl groups and the base protecting groups can be accomplished using concentrated base (Figure 2.8). On some synthesizers, such as those in our laboratory, oligonucleotides are cleaved from the resin automatically and collected in 3ml screw cap vials ready for deprotection. For DNA, concentrated aqueous ammonia (cNH_4OH) is used, with the sample heated to 55°C (for five hours or more) or 70°C (for two hours).

These conditions can be used, with moderately extended incubation times, for oligonucleotides protected with Fpmp as it is completely stable under such conditions. cNH_4OH is, however, an extremely harsh basic reagent and is not suitable for deprotecting RNA with tBDMS 2′-protecting groups. The tBDMS groups can be lost, leading to phosphoryl migration and chain cleavage. Addition of 25% (v/v) ethanol gives a much less harsh reagent which allows complete removal of base protecting groups and is still compatible with automated cleavage from the resin. However, a small number of tBDMS protecting groups are still lost during the extended incubation times required for the removal of the standard Bz and iBu base protecting groups. Wu *et al.* proposed the use of phenoxyacetyl for the protection of the amine groups of exocyclic bases

and showed they could be completely removed using anhydrous ammoniacal methanol [30]. Their results clearly showed the superiority of this reagent compared to incubations of around 20 hours in EtOH/ cNH_4OH. An alternative solution to the problem is to minimise the exposure time to EtOH/ cNH_4OH by using more labile base protecting groups. 'Fast deprotecting' strategies have been developed [31] and became commercially available during the course of this work.

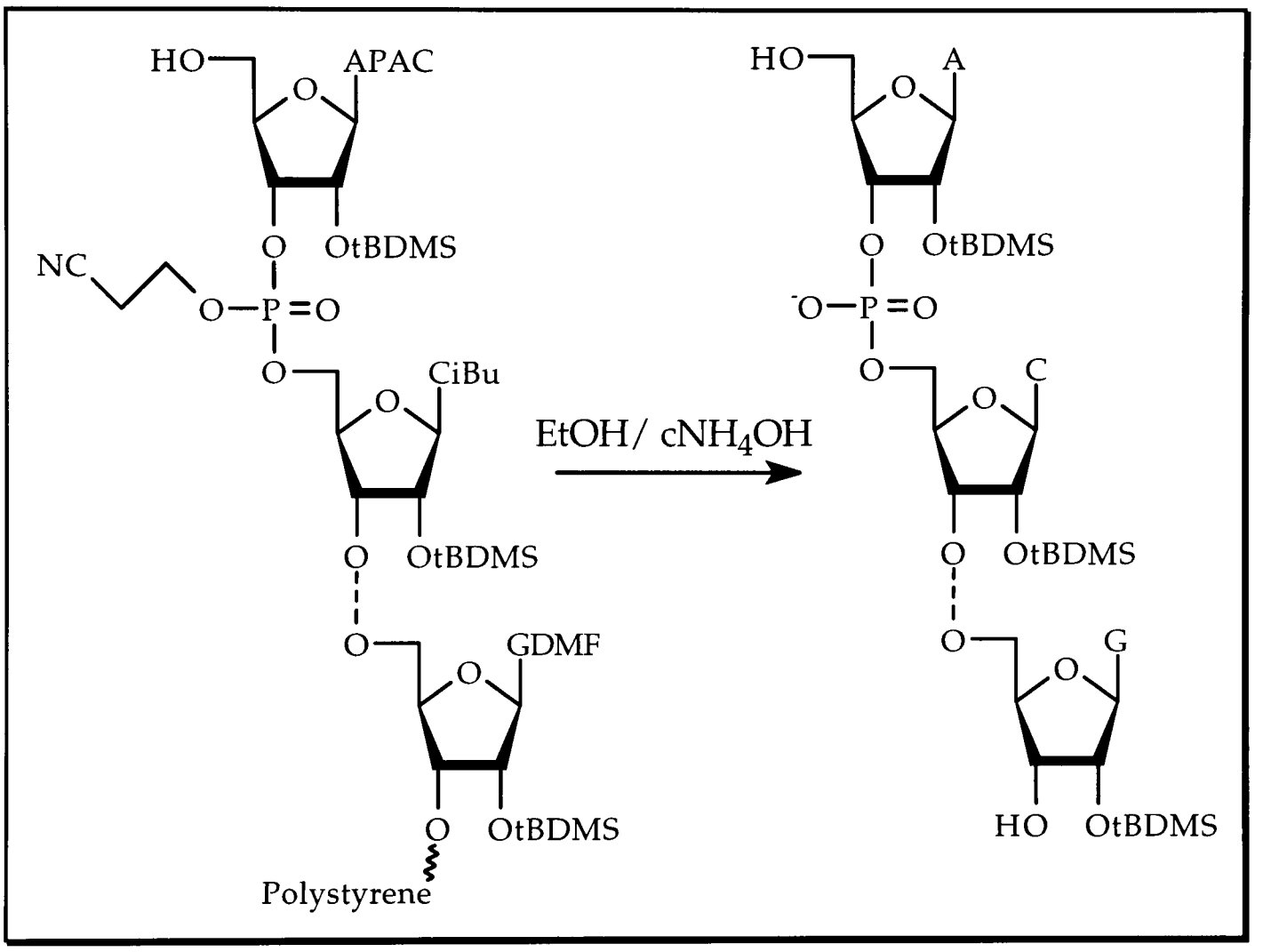


Figure 2.8 Cleavage from the solid support and removal of the phosphorus and base protecting groups

EtOH/ cNH_4OH was used in all cases for deprotecting oligoribonucleotides in this project. Of the two methods, it was the one compatible with automated cleavage from the resin. For Fpmp protected RNAs, incubation times of 8-15 hours at 55°C are recommended, though it was

found that shorter times (5 hours) gave equally satisfactory results. Where 2′-tBDMS groups were used, base deprotection was accomplished with a four hour incubation at 55°C. The final step in the RNA synthesis is then the removal of the 2′-protecting group.

2.2.4.2 Silyl Deprotection

The 2′-O-tBDMS protecting groups are removed using a fluoride reagent (Figure 2.9). Until recently, 1.0M tetrabutylammonium fluoride (TBAF) in THF was the reagent used. However, problems with its use have been encountered [32] and often complete removal of all tBDMS groups is not possible. This has been attributed to the variable, and often high, water content of the TBAF solution [33]. In 1992, Gasparutto *et al.* reported the use of TEA.3HF as a desilylating agent [34], and subsequent studies have shown it to be greatly superior to TBAF. In the early stages of the development of the protocols described here problems were also encountered in the reversed phase HPLC purification of RNAs deprotected with TBAF. Tetrabutyl ammonium salts first had to be removed using a cation exchange resin.

Neat TEA.3HF was used to deprotect a number of RNA oligonucleotides but with variable results. With most moderate length sequences desilylation was achieved in 12-16 hours at room temperature. In these cases the RNA was collected as a precipitate on the addition of water and cooling to -20°C for 2 hours. Precipitation occurred preferentially for longer sequences simplifying the subsequent HPLC purification considerably. Generally, this was observed for sequences like rCGCAAAUUUGCG that contained an approximately equal purine and pyrimidine content. The first problem encountered for other sequences was dissolving the RNA. It has been suggested that solubility in TEA.3HF

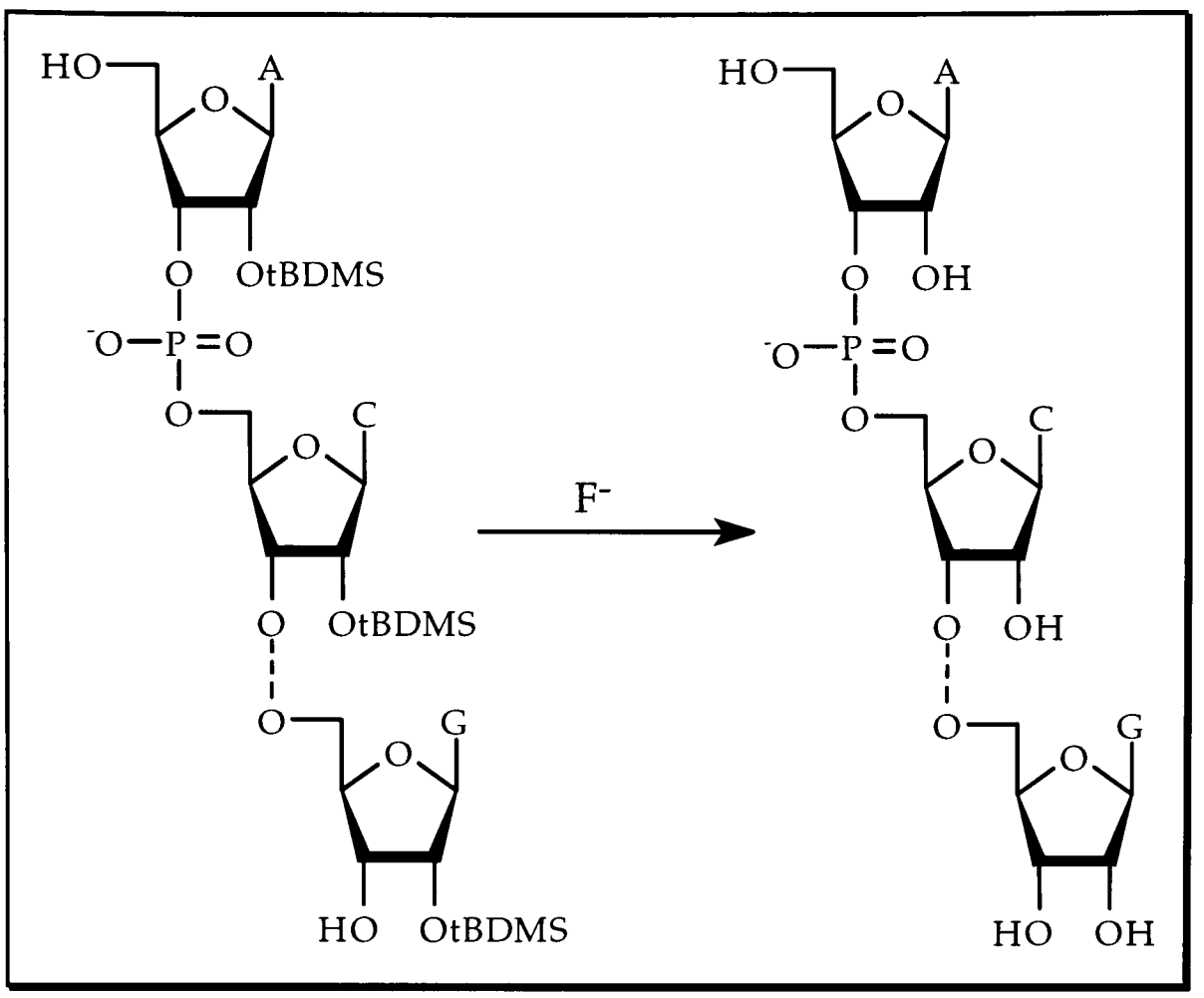


Figure 2.9 Removal of 2'-O-tBDMS protecting groups

is length and sequence dependent [35], and problems were indeed often encountered when trying to deprotect sequences longer than about 30 nucleotides in length. Surprisingly, the method also failed to fully deprotect a 10 mer rGAAGAGAAGC, suggesting that it is high purine content that causes the solubility problem. Conversely, the complementary strand rGCUUCUCUUC dissolved readily in TEA.3HF but would not precipitate!

The problem of initial solubility in the reagent was overcome by first dissolving the dried RNA in DMSO. A volume was used such that DMSO composed 10% of the deprotection mixture when the TEA.3HF was added. A higher percentage of DMSO (up to 50%) may be used for very long or insoluble RNAs with an extended incubation time [personal communication, A.K. Collier, University of Leeds]. Using this method, all

oligonucleotides prepared in this study could be efficiently deprotected. However, problems arose in the removal of the deprotection mixture. Two approaches were used: size exclusion chromatography with sephadex and direct application to the anion exchange HPLC column. The latter required that the pH of the deprotection mixture be brought carefully to around neutral by the addition of sodium hydroxide. Neither method was entirely successful. Although a major product peak was always seen by HPLC, the final yield of pure material was often less than expected based on absorbance readings of the crude material and analytical injections. Often, material was observed behind the main product peak in the preparative HPLC. If the organic solvents used in the deprotection were not removed properly, it is possible that they could cause aggregation of the RNA and consequent late elution from the HPLC column. It was anticipated that the TEA.3HF/ DMSO mixture would have no affinity for the anion exchange column and would be easily washed away. However, contamination was observed in an NMR sample prepared in this way. Problems with aggregation of the sequence $r(CGCAAAUUUGCG)_2$ were also observed by NMR, after using this method of purification.

During the course of this work, an alternative TEA.3HF deprotection mixture, allowing considerably reduced deprotection times, was suggested by Wincott *et al.* [25]. This method uses a 3:1.5:2 mixture of N-methylpyrrolidinone, triethylamine and TEA.3HF (NMP/TEA/TEA.3HF) at 65°C. Deprotection times are reduced to around 30 minutes for short sequences with up to 90 minutes required for longer sequences. The volume of deprotection mixture is relatively small, allowing the use of a butanol precipitation to recover the RNA. This method was successfully used in the deprotection of the sequences prepared in this study, with significantly improved yields.

2.2.4.3 *Fpmp Deprotection*

Removal of the 2′-hydroxyl protecting group and the terminal DMT group is achieved using an incubation in dilute aqueous acid. Originally, 0.01M HCl, pH2.0, was used for this purpose and provided satisfactory results with a number of sequences [14]. However, it was later shown that at this pH, concomitant phosphoryl migration and chain cleavage can occur [36]. One of the main requirements of the 2′-hydroxyl protecting group was that it should be easily removed under conditions to which the RNA was stable. Although products of chain cleavage are simple to remove by HPLC, it would be virtually impossible to remove strands with 2′-5′ linkages. Given that some sequences, but not others, can be successfully deprotected at pH 2.0, it is clear that the degree of phosphoryl migration is dependent on the length and sequence of the RNA under study. It has been shown that even sequences more susceptible to acid catalysed hydrolysis, such as polyuridylic acid, can be successfully deprotected at higher pH [37,38].

Fpmp RNA oligonucleotides prepared during the course of this work were deprotected using a sodium acetate buffer, pH 3.0-3.2 (Figure 2.10). A number of precautions were required to ensure the success of the deprotection and to minimise loss of material. The deprotection step follows an initial trityl-on HPLC purification (section 2.3.1.1). The HPLC buffer had to be removed completely to ensure the correct pH for deprotection was obtained in the sample. The purified RNA was lyophilised twice to remove the majority of the ammonium acetate by sublimation. As the Fpmp groups greatly lower the solubility of the RNA in water, a 2% ammonia solution was used to redissolve the oligonucleotides until after their removal. The desalting process was completed by gel filtration on sephadex and the RNA dried in a 3ml screw cap vial. The acidic deblock buffer was then added and the turbid sample vortexed vigorously for 24-36 hours at room temperature. After this time

a small volume of neutralising buffer was added. Following desalting on sephadex, the RNA was purified further by HPLC.

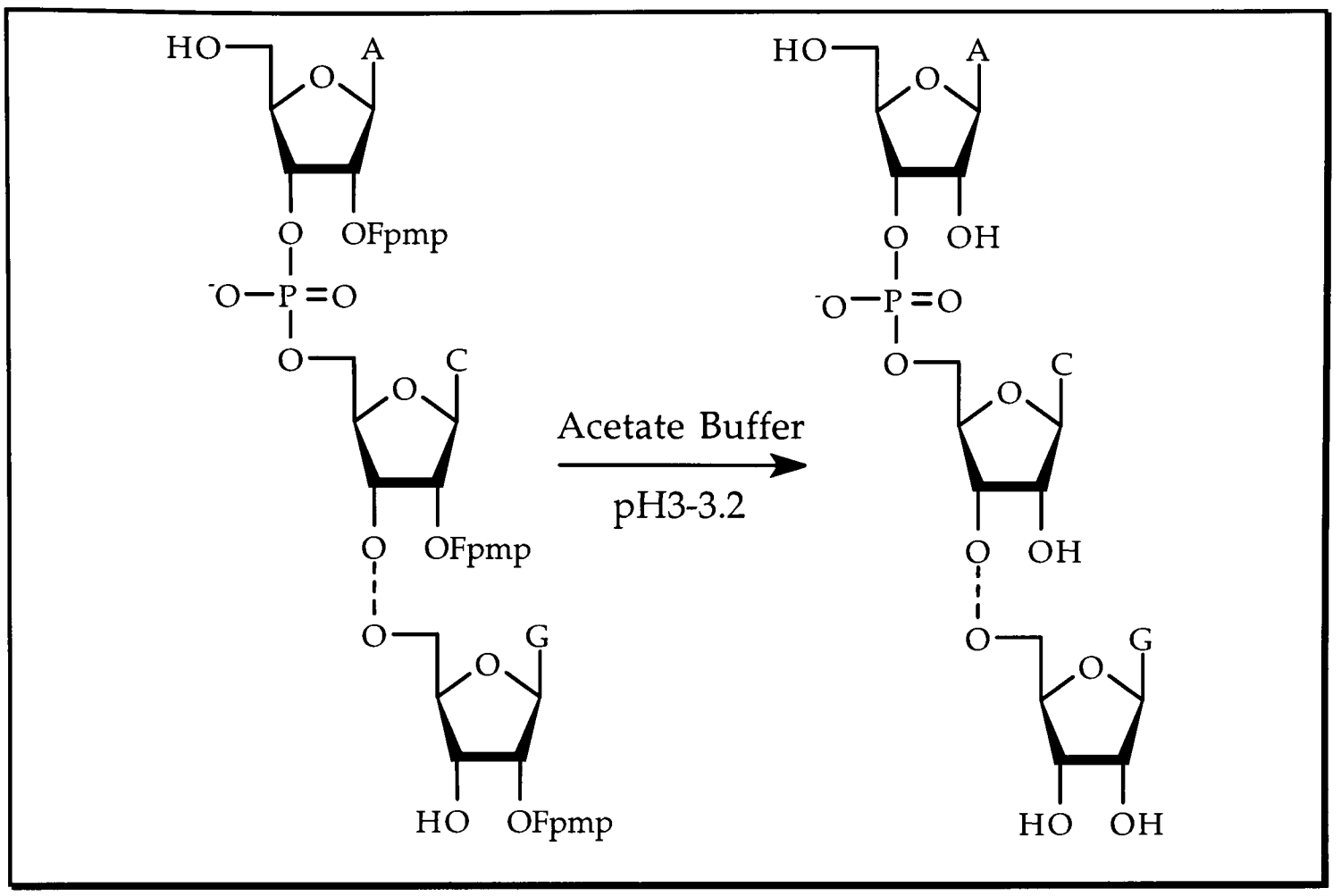


Figure 2.10 Removal of 2'-O-Fpmp protecting groups

This method proved to be satisfactory for short to moderate length oligoribonucleotides. However, the yields of pure material tended to be quite variable and generally somewhat lower than for the tBDMS method. This can be explained by the low solubility of the Fpmp RNA in the deblock buffer. Close examination of the deprotection mixture revealed that while much of the material formed a cloudy suspension, part of it was present in small particles. The amount of material present in these particles varied from sample to sample. Fpmp RNA in such particles would receive only a very small exposure to the deblock buffer, even on quite long incubation times. This can explain why the yields obtained were often quite low and variable, but little evidence was seen by HPLC for partially deprotected oligonucleotides (see section 2.3.2.2). To

help reduce this problem samples were sonicated briefly to break up large particles prior to the incubation period.

For longer and structured RNAs, it was found to be impossible to ensure complete removal of the Fpmp groups. Anion exchange HPLC elution profiles following the deblock showed a sharp peak followed by a broad shoulder. PAGE analysis of each of these fractions collected separately is shown in Figure 2.11. Although pure material could be obtained, the yields were often very low (1-10 OD_{260} from a 1μmole synthesis).

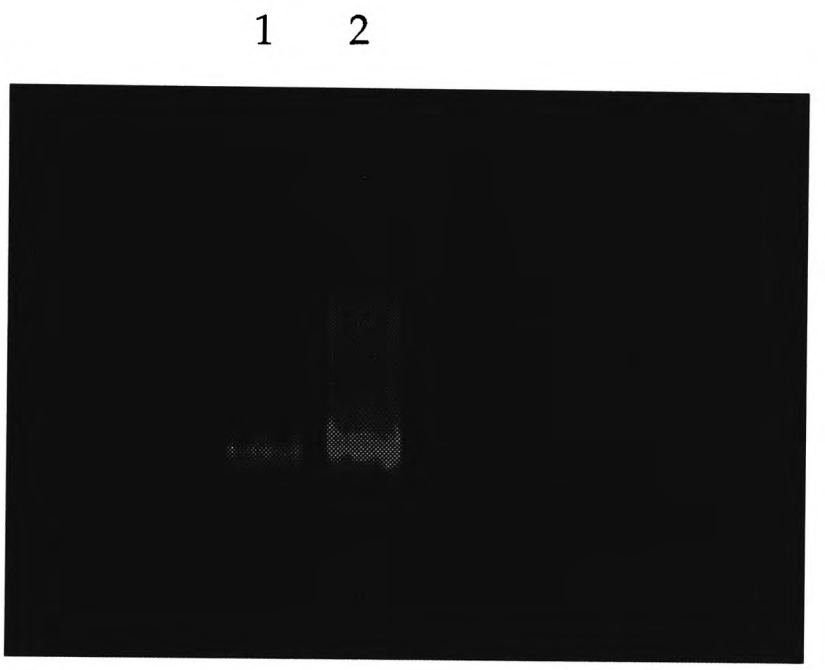


Figure 2.11 PAGE analysis of a 29 mer RNA ribozyme: sharp peak on HPLC (lane 1) and shoulder on HPLC (lane 2)

2.3 HPLC Purification

HPLC is the method of choice for the purification of synthetic RNA for biophysical studies. HPLC methods have been designed to take advantage of the different properties of RNA molecules and the effects of the protecting groups used in the chemical synthesis:

- RNAs are polyanions and the charge increases proportionally with chain length so that separation according to length can be achieved by an ionic interaction with the column material.
- RNAs contain lipophilic nucleobases whose composition and sequence will determine their chromatographic behaviour on reversed phase columns. Small impurities of base modifications and oligonucleotides containing incompletely deprotected base and 2′-OH groups may be resolved through an increased hydrophobic interaction.
- at neutral pH the phosphate groups exist as salts which allows the use of ion-pair reversed phase chromatography, where the interaction with the stationary phase is through the alkylammonium salts rather than the nucleobases.
- the highly lipophilic terminal 5′-trityl protecting group may be left intact, greatly increasing the interaction of the full length oligonucleotide with the reversed phase column [39].

The use of anion exchange and reversed phase HPLC for purification of oligonucleotides has been reported [35,40]. A number of these techniques have also been investigated in this study. The methods chosen depend entirely on which 2′-hydroxyl protecting group has been used during the chemical synthesis. The protocols developed for purifying both 2′-O-Fpmp and 2′-O-tBDMS RNA are described in the next section.

2.3.1 Fpmp Oligoribonucleotides

RNA oligonucleotides synthesized using 2′-O-Fpmp phosphoramidites are generally purified by reversed phase HPLC. Reversed phase HPLC refers to a system that has a non-polar stationary phase and a polar mobile phase. For oligonucleotide purification these are most commonly silica based, derivatised with octyl or octadecylsilane to generate the hydrophobic stationary phase. This system, with its reversal of polar nature, is distinct from other methods of hydrophobic interaction chromatography where the stationary phase is intrinsically hydrophobic.

Reversed phase HPLC is generally carried out at neutral pH, where the phosphate groups exist as salts. The separation of the components of an oligonucleotide solution is then based on the competition between the stationary phase, the hydrophobic bases and an organic solvent, generally acetonitrile. Oligonucleotides are adsorbed to the surface of the stationary phase until a certain critical concentration of organic solvent is reached. The strength of the hydrophobic interaction generally increases with chain length but is highly dependent on base composition and sequence, such that the separation is not strictly on the basis of size [41].

2.3.1.1 Trityl-on Reversed Phase HPLC

For RNA synthesized with Fpmp phosphoramidites, the terminal 5′-O-trityl is left intact ('trityl-on') in the full length product. The first purification is performed with all the hydrophobic 2′-O-Fpmp groups and the terminal DMT group still present. It is these groups that dominate the hydrophobic interaction with the stationary phase. Ammonium acetate buffer was used at all times for reversed phase HPLC purification. This is important for the trityl-on purification as this means that there is no additional hydrophobic interactions with the stationary phase.

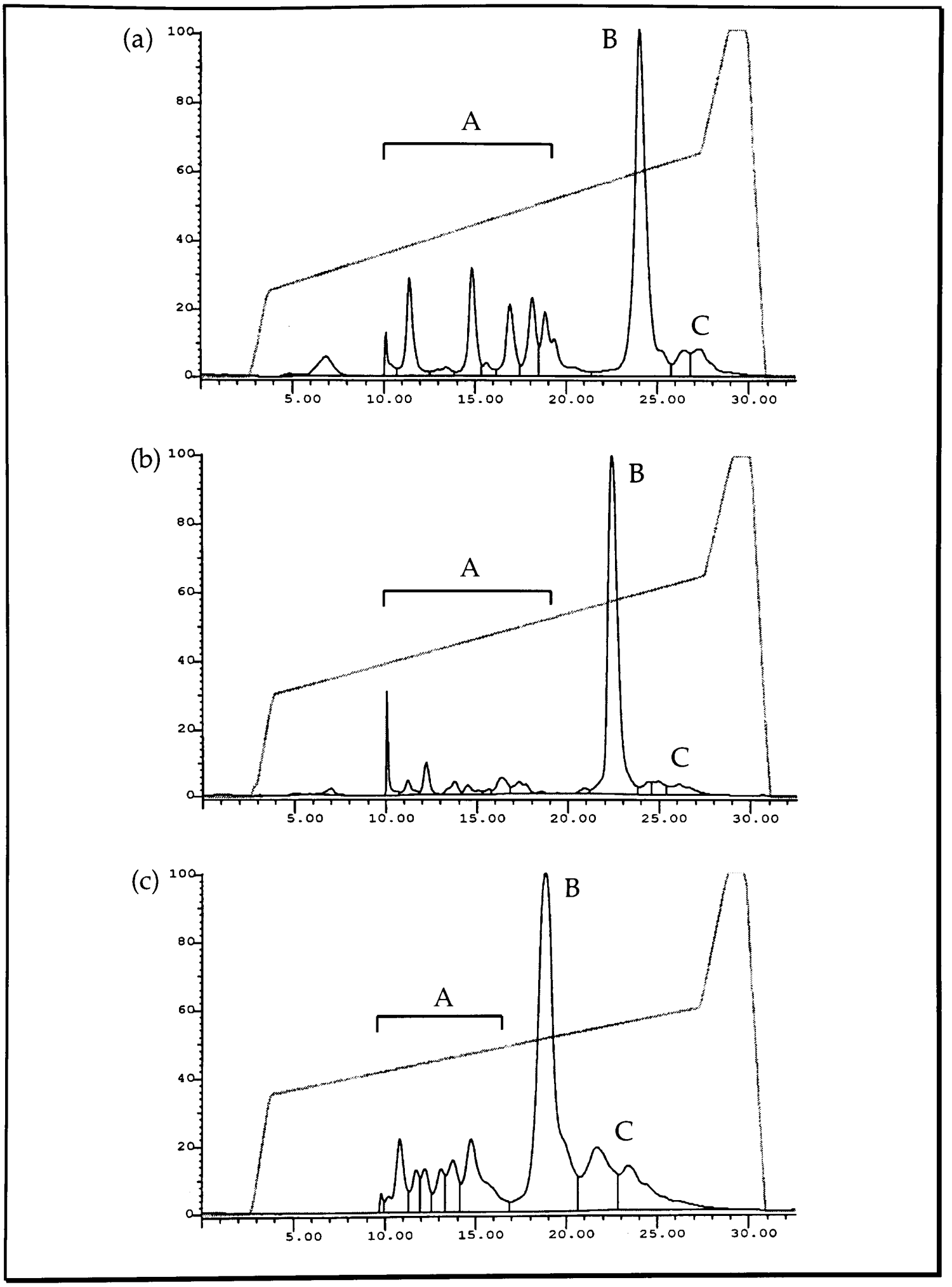


Figure 2.12 Trityl-on RP HPLC purification, (a) R02, rGAAGAGAAGC, (b) R06, rCGCAAAUUUGCG, (c) R09, 23-mer hairpin RNA

The strength of interaction of a fully Fpmp protected RNA with terminal DMT is such that the acetonitrile content of the mobile phase needs to be brought to between 50 and 70% on a linear gradient for the RNA to be eluted. Examples of trityl-on HPLC traces are shown in Figure 2.12. The full length product is the major peak (B) in each case. The effects of improvements made in the synthesis cycle are clearly evident from a comparison of the amount of deletion sequences (region A) in Figure 2.12 (a) and (b).

The usefulness of the normal DMT group in this type of purification is, however, limited. As the number of Fpmp groups in the molecule increases, the effect of the DMT becomes less. The actual length where resolution of the deletion sequences is lost was found to be sequence dependent making the technique somewhat unpredictable for longer sequences (more than 25-30 nucleotides). This problem and improvements to this step through a modification of the trityl group are discussed in section 2.4. A second problem with this method is the appearance of material after the main product peak (region C). The amount of material in this region was found to increase with the length of the sequence being synthesized. This suggests that the Fpmp group may not be completely stable to the repeated acid treatments required to remove the trityl groups during the synthesis cycle. Evidence that this could be the case came from a number of observations. Further treatment with base deprotection conditions did not reduce the amount of material in this 'shoulder' region. When collected and subjected to standard deprotection procedures, it was found to contain a number of overlapping peaks eluting after the expected elution time for the full length product on anion exchange HPLC. Furthermore, the anion exchange HPLC analysis of a 5'-fluorescene labelled RNA at 500nm indicated that there were deletion sequences containing the label. This could arise either through the generation of branched species during synthesis, through the

loss of an Fpmp group, or by acid catalysed hydrolysis during deblocking of the protecting groups. As previous studies have shown that RNA is almost completely stable under the conditions used for acidic deblocking (pH3.25) we have concluded that the problem must be due to loss of Fpmp during synthesis. A modest reduction in the delivery of TCA during the synthesis cycle and using the automatic monitor (which involves an extended TCA treatment) only at the final addition, considerably reduced the amount of material in this shoulder region but did not remove it completely.

Following the trityl-on purification the DMT and Fpmp groups are removed. The advantage of using ammonium acetate as the HPLC buffer is that is can be removed easily by sublimation on lyophilisation of the sample. This is important as the deprotection step is highly dependent on the pH of the solution (see section 2.2.4.3). The fully deprotected RNA may then be purified further by a second reversed phase HPLC.

2.3.1.2 Trityl-off Reversed Phase HPLC

A second RP HPLC step was used to remove any undeprotected RNA, free Fpmp groups and traces of shorter sequences that may be generated during the acidic deprotection. The buffer system used was the same as for the trityl-on purification, but with a much shallower gradient. With the removal of the hydrophobic Fpmp and trityl groups, a much lower concentration of organic solvent is required to elute the RNA from the column. Examples of trityl-off RP HPLC purifications are shown in Figure 2.13. Peak A corresponds to the deprotected full length RNA.

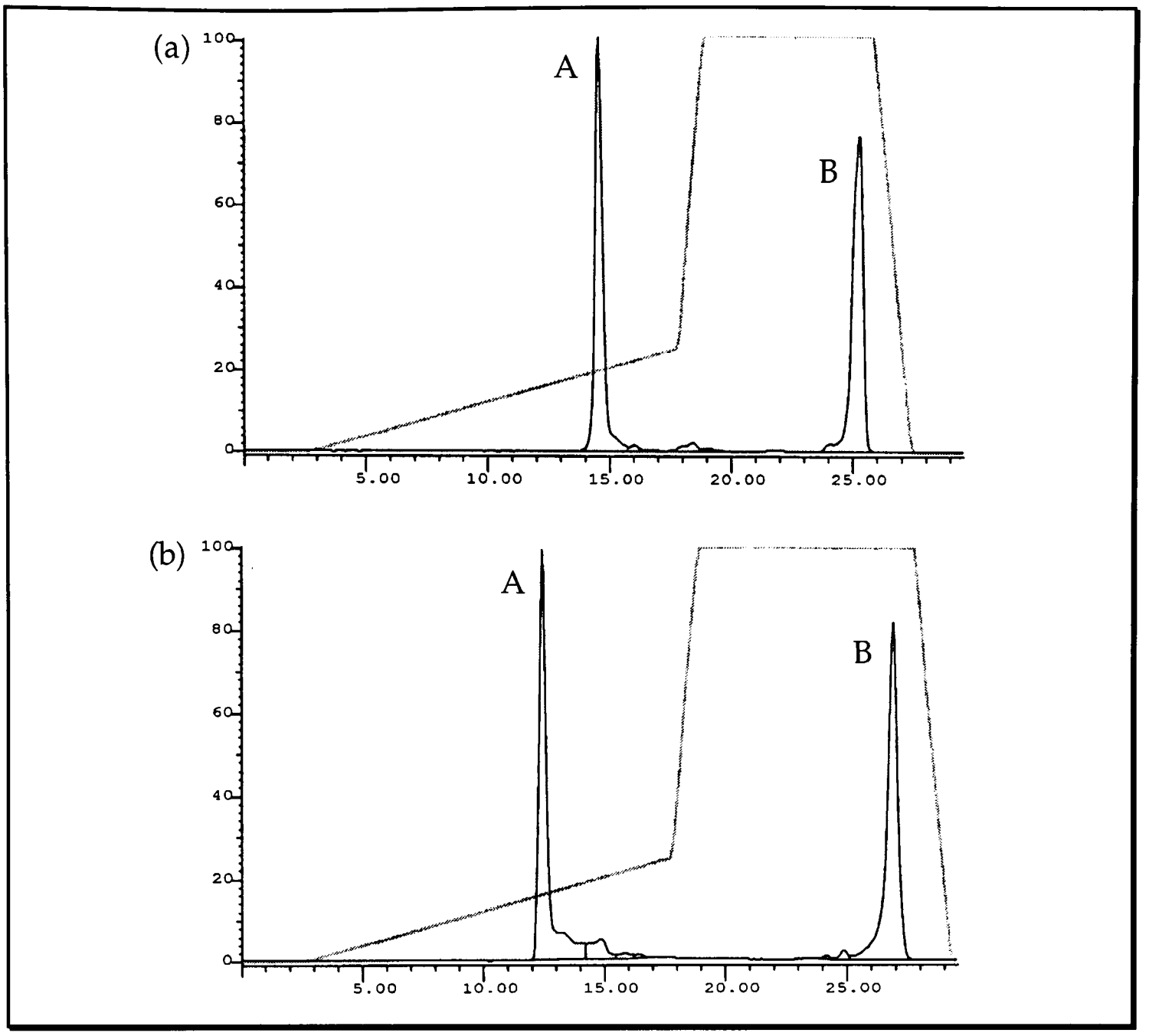


Figure 2.13 Trityl-off RP HPLC purification; (a) R02, rGAAGAGAAGC and (b) R06, rCGCAAAUUUGCG

The later eluting peak (B) was initially thought to be the free Fpmp groups. For one early synthesis it was almost completely removed by desalting the sample on sephadex and had a UV spectrum with a maximum around 240nm. However, the traces shown in Figure 2.13 were preparative injections following desalting on sephadex. That the peak should still be present suggests it is in fact due to fully protected RNA. The absence of significant quantities of partially protected species can be explained by the formation of small particles of RNA during the deprotection (see section 2.2.4.3). Figure 2.14 shows an analytical injection of an Fpmp RNA oligonucleotide during the acidic deblock step. Peaks

corresponding to both the fully protected RNA (B) and the free Fpmp groups (C) are clearly seen. The misassignment of the later eluting peak is undoubtedly significant in the lower yields often encountered using the Fpmp method.

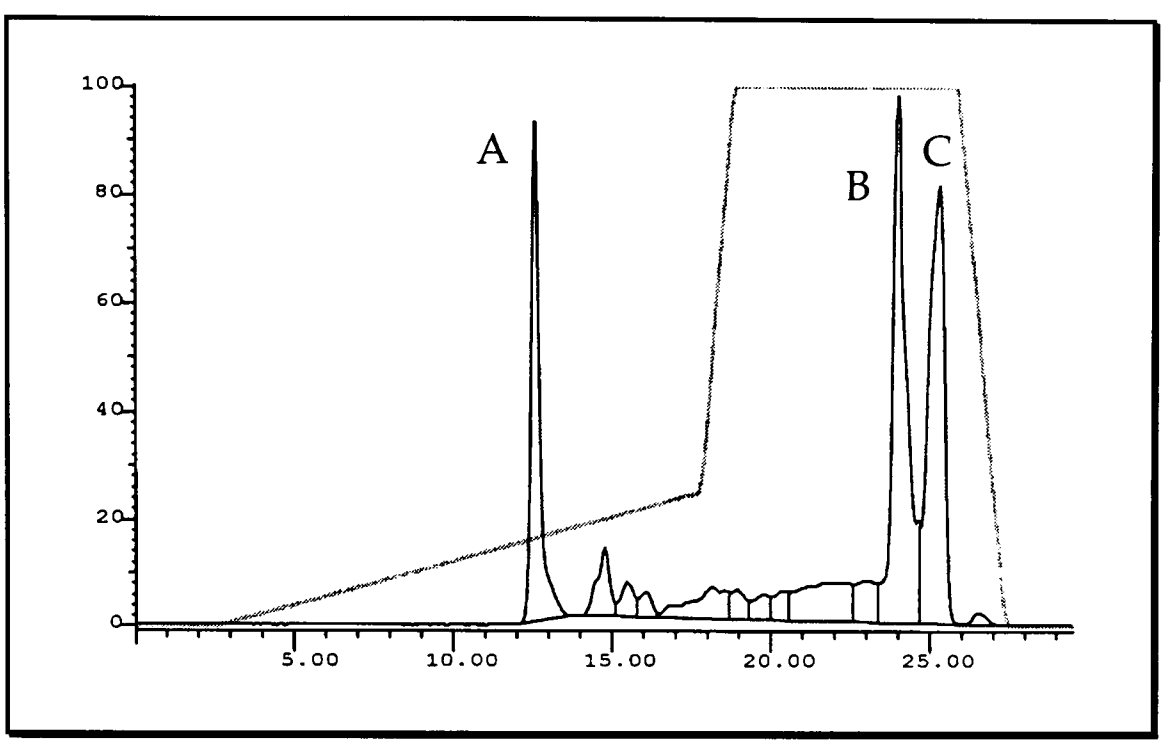


Figure 2.14 Reversed phase HPLC trace showing incomplete deprotection of an Fpmp RNA.

An alternative approach to the two-step RP HPLC is to fully deprotect the RNA prior to purification and use the protocols below for RNA synthesized using tBDMS phosphoramidite monomers.

2.3.2 tBDMS Oligoribonucleotides

RNA oligonucleotides synthesized using tBDMS phosphoramidite monomers were fully base and 2'-OH deprotected before HPLC purification. The use of a trityl-on RP HPLC purification has been reported [25] but was not investigated here. RNA oligonucleotides were purified using a combination of strong anion-exchange and RP HPLC.

For anion exchange HPLC a stationary phase with positively charged groups is required, most commonly quaternary alkylamine derivatised silica, bound to a support matrix. The oligonucleotides are bound through an electrostatic interaction at the phosphate groups and are displaced from the column by the negative ions of an increasing salt gradient. As the number of phosphate groups is directly proportional to the chain length the separation is based on the length of the sequence. The only exceptions are where interactions other than electrostatics alone are in effect, such as folding of the RNA into compact structures that shield some of the phosphate groups. However, in general, the combination of anion exchange and reversed phase HPLC make possible the removal of both deletion sequences and any undeprotected or modified oligonucleotides.

2.3.2.1 Ion Exchange HPLC

During the course of this work a new anion exchange column became available from Dionex. The anion exchange medium had properties suitable for use in an HPLC system and was stable to heating up to 90°C, over a wide range of pH and to chaotrophic agents up to a concentration required to denature most oligonucleotides. The stationary phase consists of methylacrylate beads (0.1μm), derivatised with alkylammonium sites, electrostatically bound to highly crosslinked resin beads (25μm), thus imparting solvent resistance and mechanical stability.

A gradient of ammonium chloride in tris (pH 7) buffer was used for all anion exchange HPLC purifications with the column heated to 55°C. The slope and the length of the gradient was chosen depending on the length of the oligonucleotide. For the sequences used in this work a 20 minute gradient was generally sufficient to give highly pure material (Figure 2.15). [Similar gradients of 40 or 60 minutes were used to purify RNA oligomers up to 50 nucleotides and DNA.RNA chimeric oligomers up to

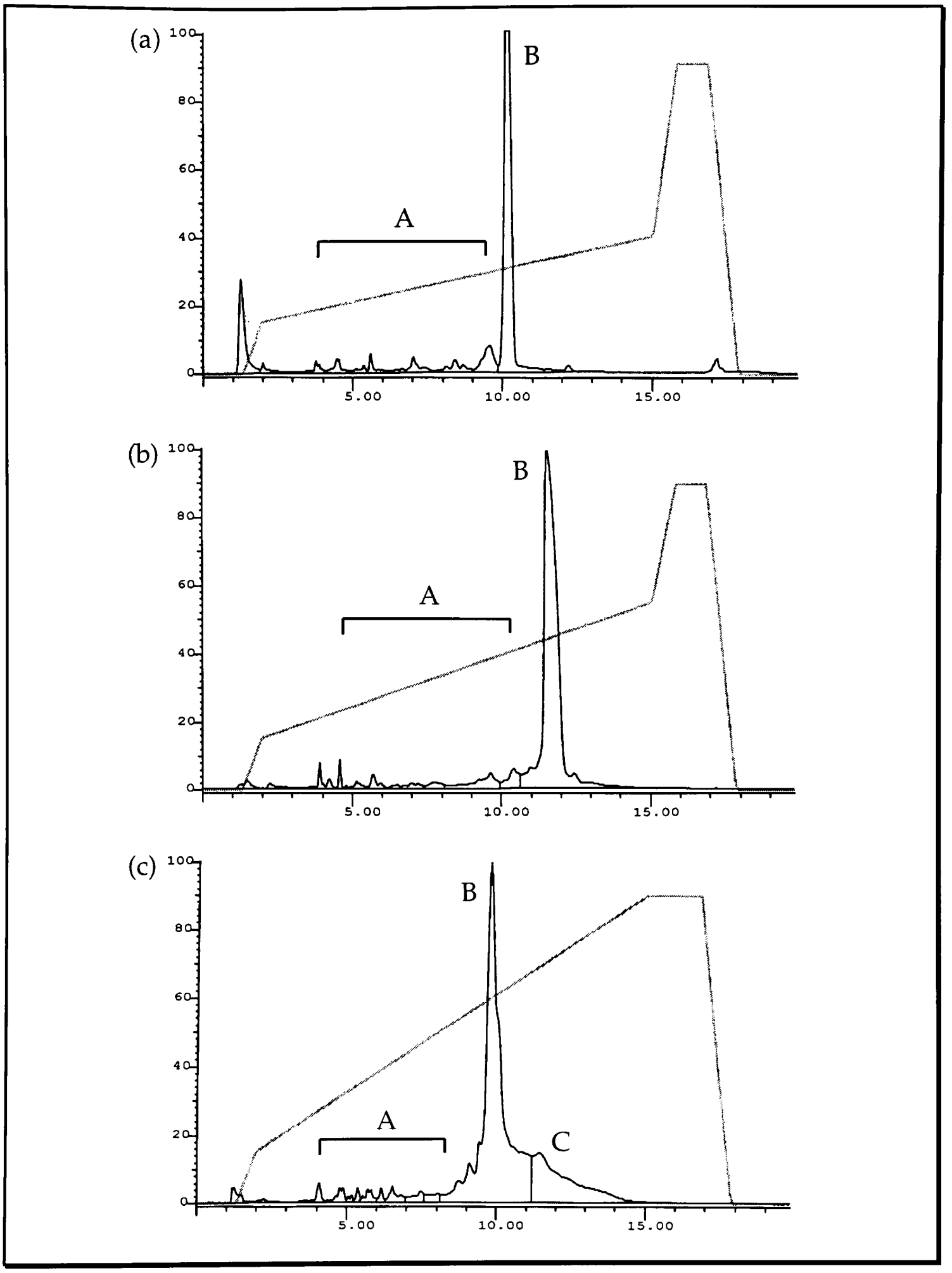


Figure 2.15 Anion exchange (Nucleopac PA-100) HPLC traces: (a) R02, rGAAGAGAAGC analytical injection, (b) R06, rCGCAAAUUUGCG and (c) R09, 23mer hairpin RNA

65 nucleotides in length for other projects.] The deletion sequences are clearly resolved (region A) from the full length product (B). The broadening of the product peak on going from an analytical (Figure 2.15a) to preparative injection (Figure 2.15b,c) results from overloading of the column (greater than 5% of the total column loading capacity). The desired level of purification could still be achieved under these conditions, known as high performance displacement chromatography (HPDC) [42].

The 23 mer hairpin RNA was loaded from buffer containing 8M urea and the sample heated to 90°C for 2 minutes prior to injection. Without this treatment, a second peak was observed at the end of the trailing region (C in Figure 2.15c) which was attributed to the formation of a relatively stable dimer. It is likely that the full length oligonucleotide is eluting in a folded conformation and that the trailing is due to some unfolding of the hairpin structure. A 21 mer hairpin (R04, Appendix B), with the same stem loop structure but without the 5'-GG overhang, had a melting temperature in excess of 90°C even in the presence of 10% methanol. Clearly, completely denaturing such stable secondary structures for HPLC purification is not a simple or necessarily desirable task. Indeed, the use of different metal ions under non-denaturing conditions to elute RNA in folded conformations from anion exchange columns has been examined [43]. The purity of material generated by the methods described here was found to be suitable for high resolution studies by NMR [personal communication, Dr. V. Ramesh, University of Southampton] and for growing single crystals (Chapter 3).

2.3.2.2 Reversed Phase HPLC

Following the removal of the deletion sequences, a reversed phase HPLC step is used to remove any remaining silylated material (Figure 2.16). Due

to the hydrophobicity of the tBDMS group, silylated material is expected to be well resolved behind the fully deprotected RNA. The method is essentially as described for Fpmp RNA, but as the oligonucleotides are fully deprotected a much lower concentration of acetonitrile is used (0 to 20%).

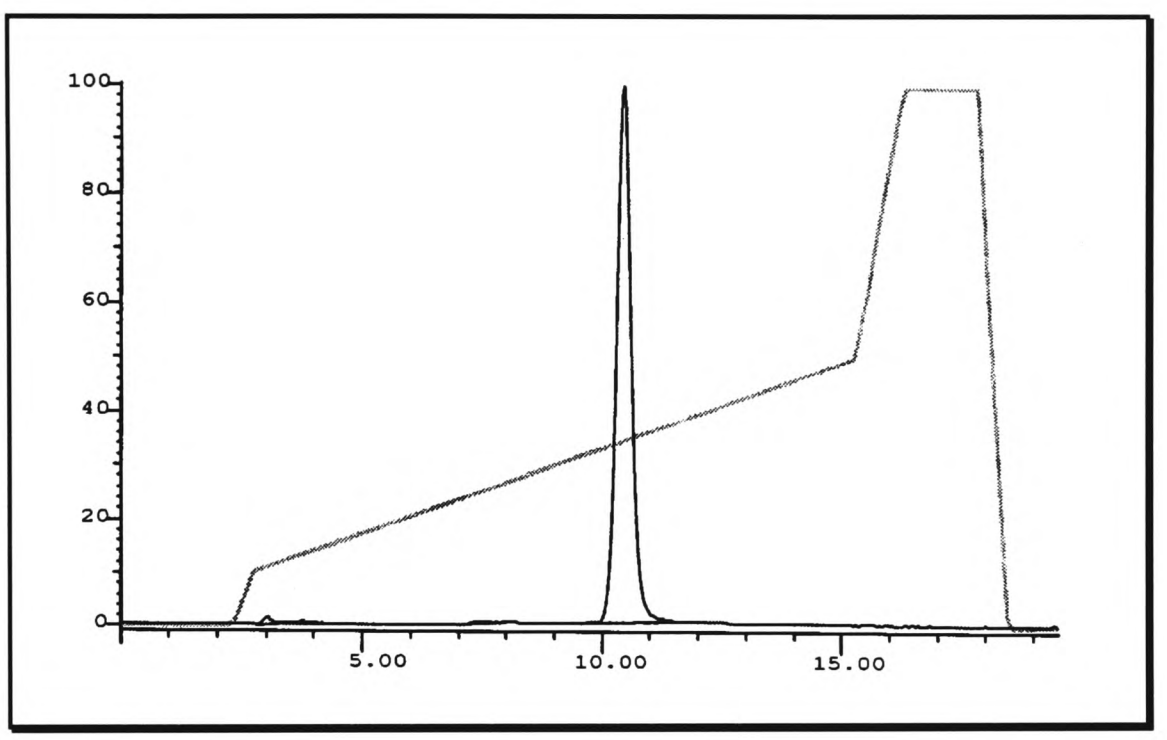


Figure 2.16 Reversed phase HPLC purification of R02, rGAAGAGAAGC

In general, the efficiency of the desilylation with NMP/ TEA.3HF/ TEA is such that the anion exchange HPLC gives complete purification in a single step (as observed by an analytical injection on a reversed phase column). For most biochemical applications this level of purity is more than sufficient. However, the reversed phase step was included in the preparation of material for these studies for two major reasons:

- the RNA was required to be of maximum possible purity for crystal growth and NMR work (even small traces of tBDMS groups could cause problems for NMR due to the very sharp peaks they give).
- the anion exchange step generates a very high concentration of salt in the sample which would not be completely removed by desalting on sephadex. By including the reversed phase step, the salt is completely

removed as it has no affinity for the reversed phase column. The lower concentrations of ammonium acetate buffer are simply removed by sublimation followed by desalting on sephadex.

2.4 Improvement of Fpmp RNA Purification by a Modified Trityl Group

The initial reversed phase HPLC purification of 2′-O-Fpmp protected RNAs constitutes a significant advantage for this method over others. Although we have found that for biophysical studies such as those described here, additional purification is desirable, a single trityl-on purification is more than sufficient for most biochemical applications. The significant problem is, however, that with increasing chain length the effect of the terminal 4,4′-dimethoxytrityl group becomes less due to the increasing number of hydrophobic Fpmp groups. Overlap of the trityl-off deletion sequences and the trityl-on full length product becomes a problem from around 25-30 nucleotides in length. The problem also seems to be sequence specific, which means that development of general protocols is not possible even for these intermediate length oligonucleotides.

The syntheses of monoalkoxytrityl groups, with C_{10} [10] and C_{16} [44] chains, have been reported and used to improve the HPLC purification of DNA [45]. It is likely that these have not been employed further for DNA as the synthesis chemistry has been developed to such efficiency that they are not necessary. The situation for RNA is still somewhat different. A modified 5′-OH protecting group that is fully compatible with the current methodologies but that gives a far superior means of purification is still desirable.

2.4.1 Synthesis of Phosphoramidite Monomers With a Dioctaoxytrityl Group

The modified trityl group was synthesized from 4,4'-dihydroxybenzo phenone (Figure 2.17). The alkylation step was initially accomplished using the method of Görtz and Seliger [44].

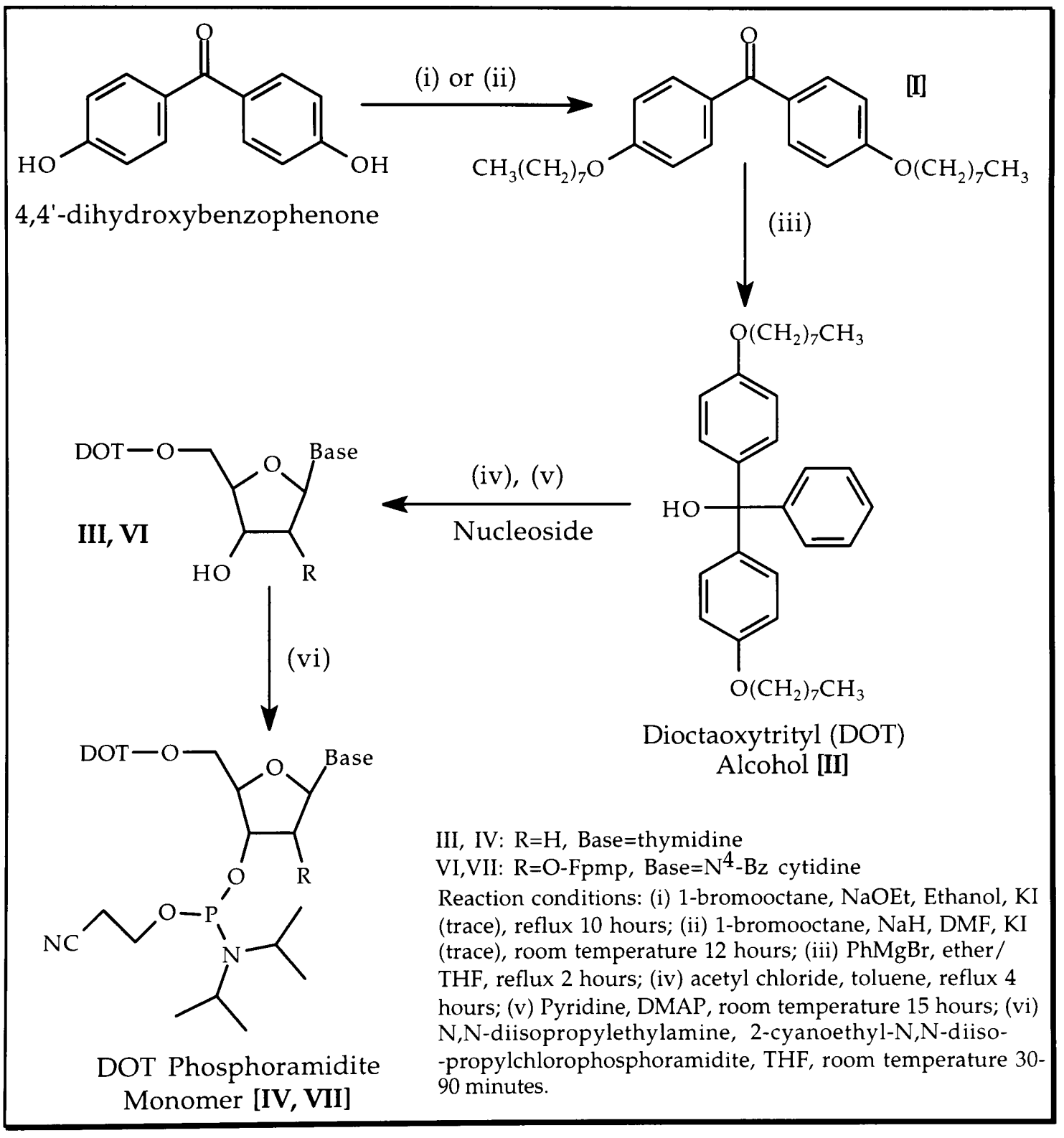


Figure 2.17 Synthesis of phosphoramidite monomers with a dioctaoxytrityl group

However, it was found that the dialkylated product precipitated from the reaction mixture so that the reaction generally had to be stopped at around 70-80% complete, giving an overall yield of 60%. A second method using NaH in DMF was found to be more successful, giving an 83% yield with no optimisation of reaction conditions. The trityl alcohol was synthesized by a Grignard reaction with phenyl magnesium bromide. The trityl chloride was then generated *in situ* and coupled to the 5'-OH of both thymidine and 2'-O-Fpmp-N^4-benzylcytosine. Phosphoramidite monomers were generated with 2-cyano-N,N-diisopropylchloro phosphoramidite and dried thoroughly *in vacuo*.

2.4.2 Oligonucleotide Synthesis and Purification Using the DOT Group

Initial tests on the effect of a DOT group on reversed phase HPLC retention time were carried out using polythymidine DNA oligonucleotides (T12, T35 and T55). The DOT thymidine phosphoramidite monomer was found to couple very efficiently during automated solid phase synthesis, with coupling yields of 97-99.5% per cycle. HPLC analysis showed a marked increase in retention time for all polythymidine oligonucleotides containing DOT as opposed to DMT. Figure 2.18 shows the results of co-injections of DMT (peak A) and DOT (peak B) T35 and T55 oligonucleotides on reversed phase HPLC. (The peaks were identified from individual injections of each oligonucleotide.) The same samples were also analysed by capillary electrophoresis, the DOT group causing an increase in the retention time, as shown for T35 (Figure 2.19).

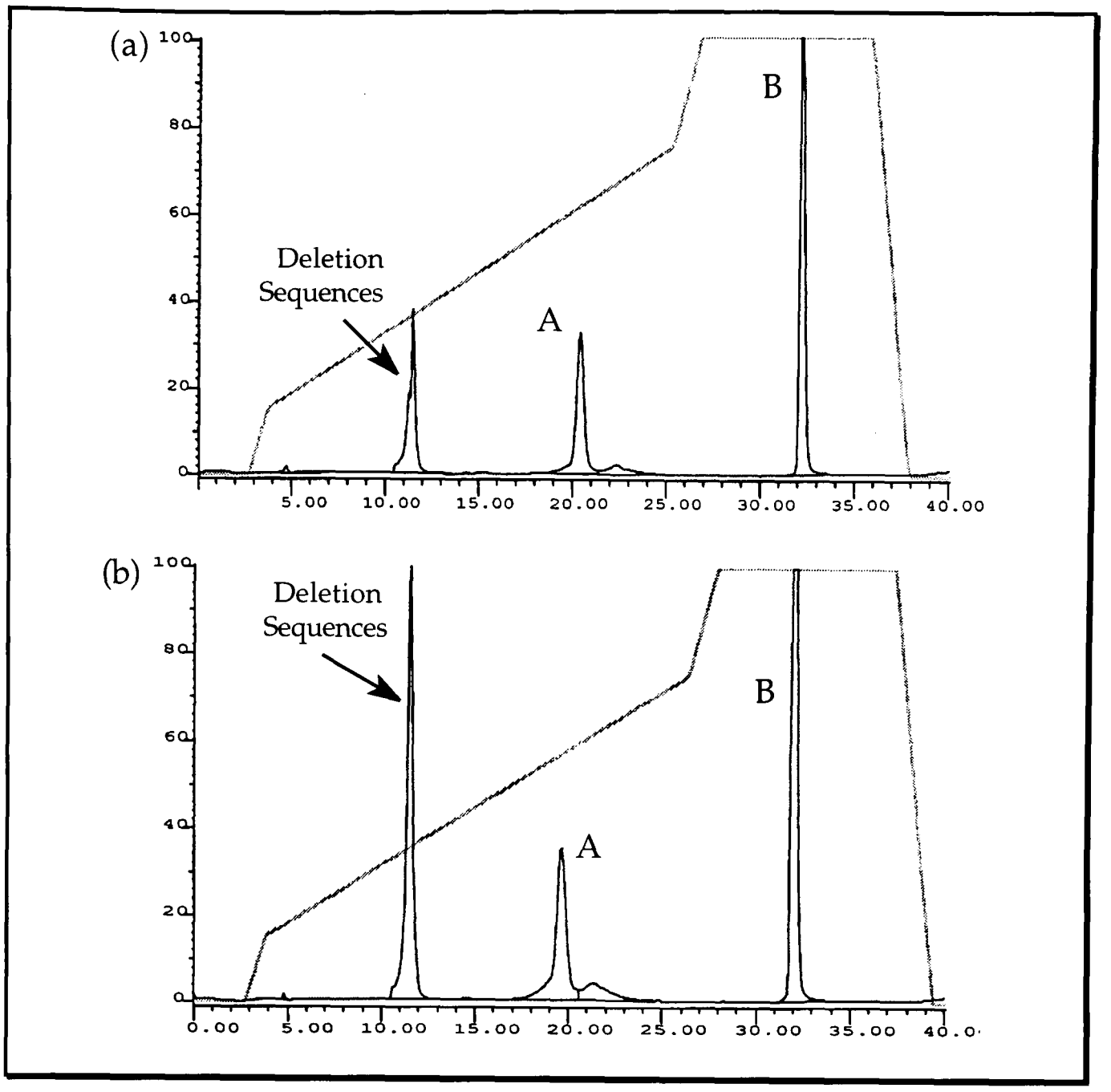


Figure 2.18 Reversed phase HPLC traces: (a) co-injection of DMT-T35 and DOT-T35 and (b) co-injection of DMT-T55 and DOT-T55

The DOT thymidine group was used to aid the purification of mixed sequence DNA oligonucleotides of 30, 65 and 101 nucleotides in length. DOT thymidine phosphoramidite was used for each addition of thymidine in the synthesis. HPLC analysis gave profiles almost identical to the polythymidine oligonucleotides even for the longest sequence. It was expected that as the length of the oligonucleotide increased, the relative effect of the DOT group in purification would diminish. Within the range studied this did not happen, indicating that the DOT group would make possible the purification of very long DNA oligonucleotides by reversed phase HPLC.

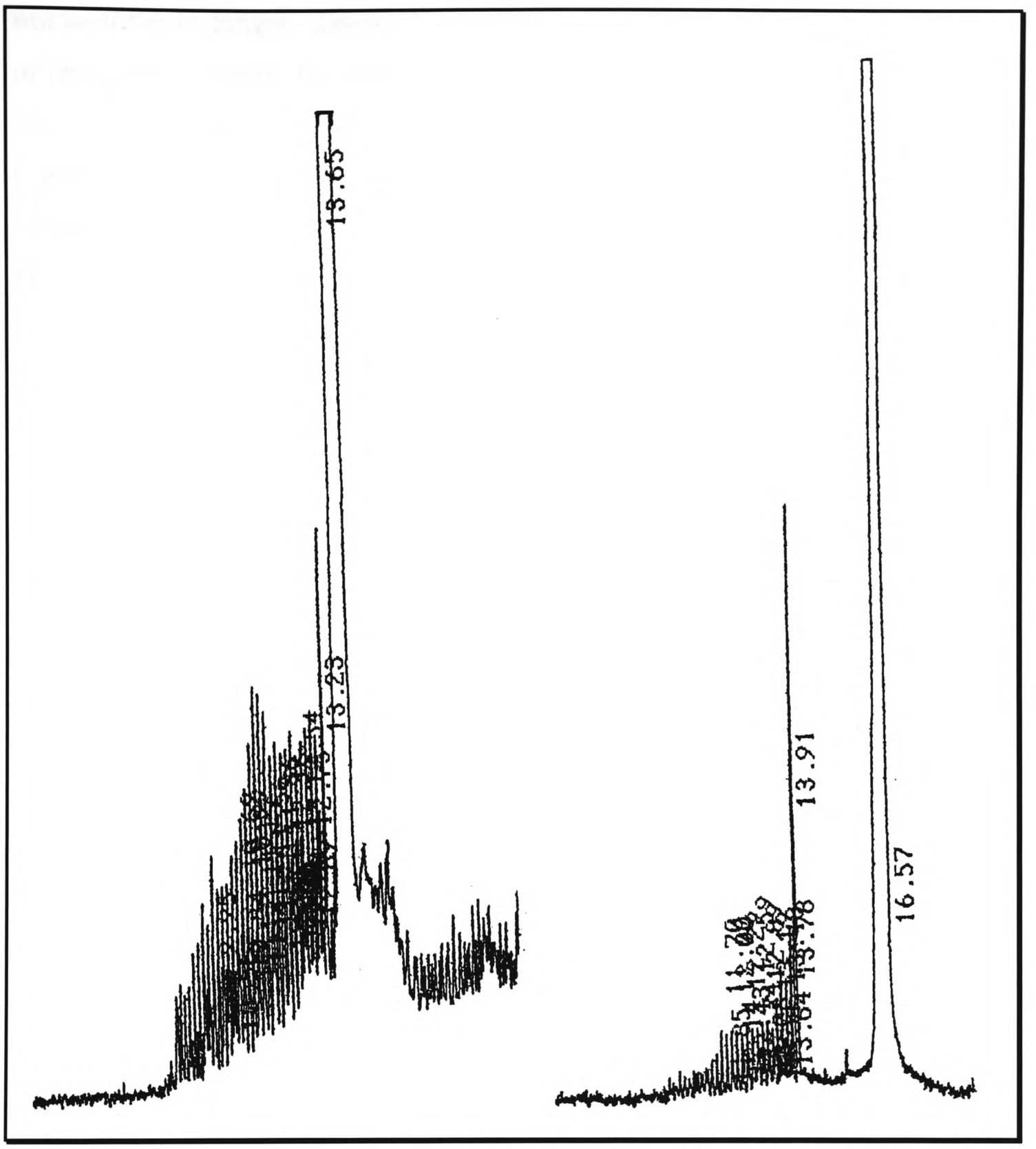


Figure 2.19 CE traces of DMT-T35 (left) and DOT-T35 (right)

2.4.3 2′-O-Fpmp RNA Synthesis and Purification using the DOT Group

A number of RNA/DNA oligonucleotides of various lengths were synthesized and purified using the DOT group incorporated through a

terminal thymidine. The longest, a hammerhead ribozyme was 29 nucleotides in length. The DOT group gave considerably better resolution of the product (peak B) from the deletion sequences (peak A) (Figure 2.20). The traces shown were obtained with buffer B containing 50% acetonitrile, which had proved useful for shorter Fpmp RNA oligonucleotides purified with the terminal DMT group intact. As the DOT RNA did not elute on the linear part of the gradient the concentration of acetonitrile was increased to 70% for subsequent purifications.

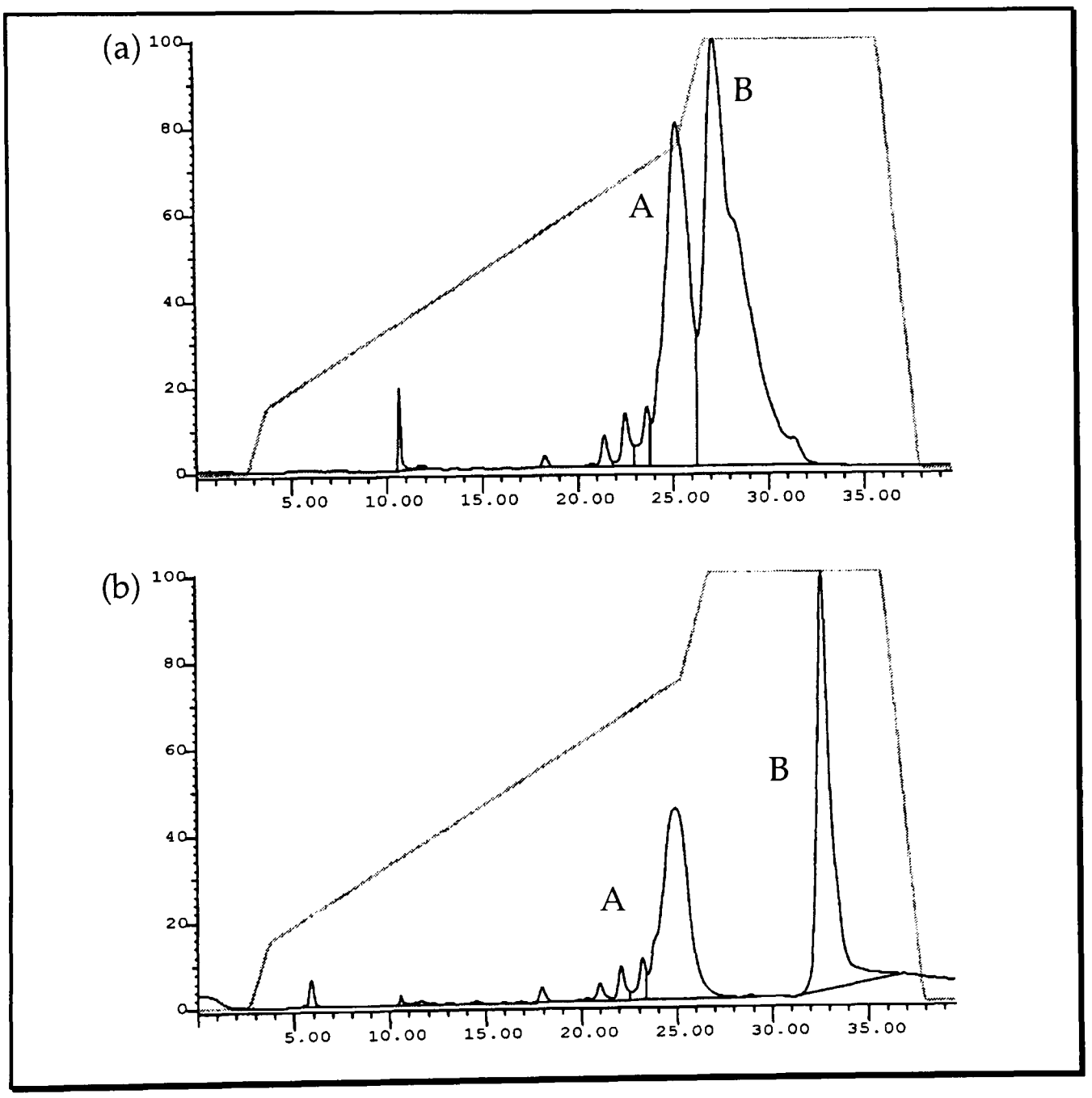


Figure 2.20 Reversed phase HPLC purification profiles: (a) 29 mer ribozyme DMT and (b) 29 mer ribozyme DOT.

The DOT group was incorporated into an RNA phosphoramidite to allow the synthesis of full RNA oligonucleotides. A DOT cytosine phosphoramidite was synthesized from 2′-O-Fpmp-N^4-benzoyl cytosine (a gift of Cruachem Ltd). A comparison of the reversed phase HPLC elution profiles of rC_{12} DMT (peak A) and rC_{12} DOT (peak B) is shown in Figure 2.21.

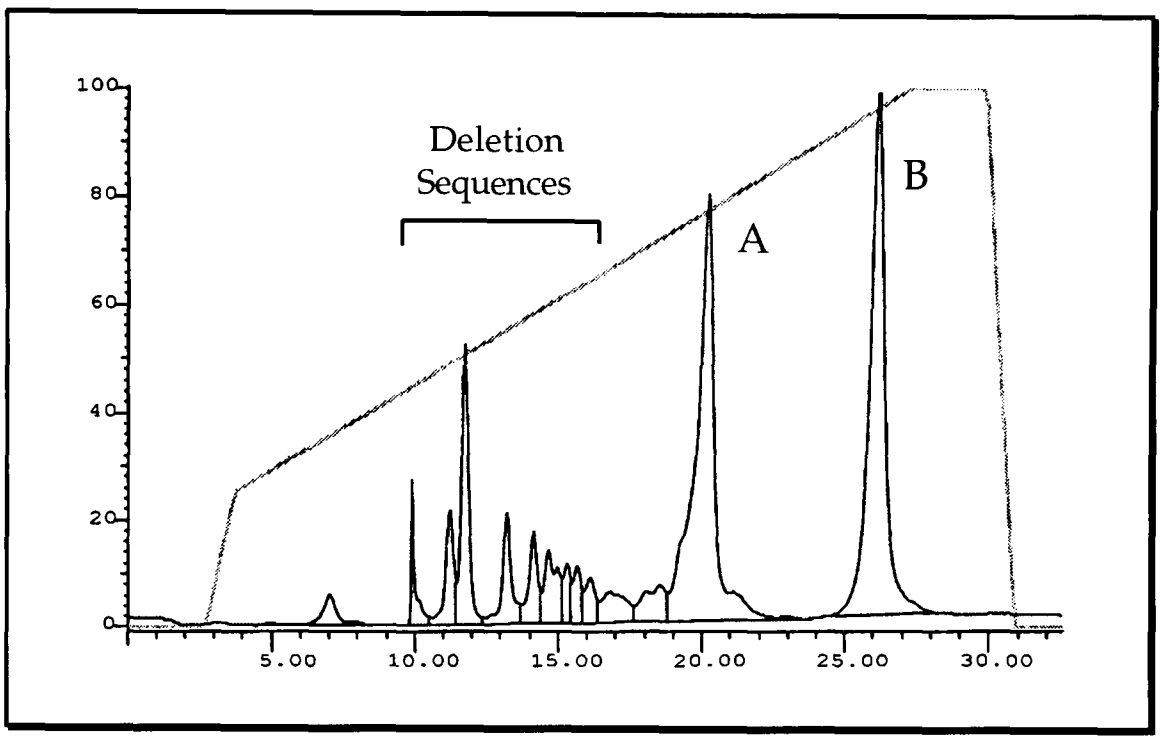


Figure 2.21 Reversed phase HPLC analysis of a co-injection of rC12 DMT (peak A) and rC12 DOT (peak B)

To test the usefulness of the DOT group in the purification of very long RNA oligonucleotides using Fpmp phosphoramidites, a 48 mer RNA was synthesized with terminal DMT and DOT groups. This RNA corresponds to part of the HIV type-1 tat protein binding region (TAR RNA), which is important in the regulation of transcription in eukaryotic cells [46]. The HPLC elution profiles show the greater hydrophobic effect of the DOT group even in such a long Fpmp RNA (Figure 2.22). The product (peak B) is clearly better resolved from the deletion sequences (peak A) using the DOT group. However, little improvement is seen in the resolution of the higher molecular weight impurity (peak C). When compared to the profiles obtained for the 29 mer hammerhead ribozyme sequence (Figure

2.20), these results also clearly demonstrate the sequence dependent nature of the purification. Although the 48 mer is considerably longer, the resolution obtained with the DMT group is much better. In all cases, however, the DOT group has greatly improved the product resolution from deletion sequences in trityl-on reversed phase HPLC.

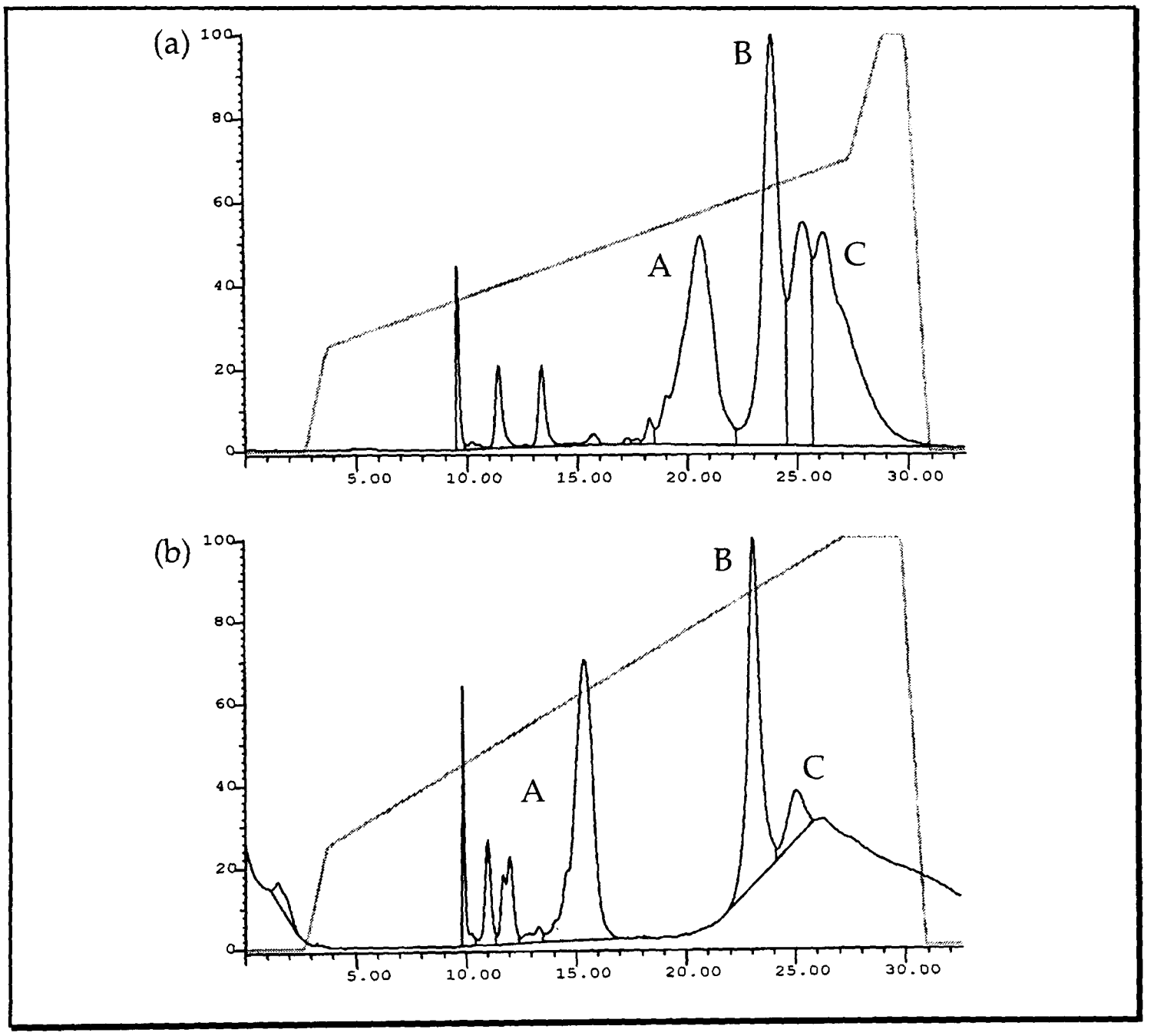


Figure 2.22 Reversed phase HPLC purification of (a) 48 mer RNA (DMT) and (b) 48 mer RNA (DOT)

2.5 Verification of Oligoribonucleotide Purity and Integrity

2.5.1 Analytical Anion Exchange HPLC

The Dionex anion exchange HPLC system (section 2.3.2.1) was found to be very useful in determining the success of the purification protocol. A small aliquot of the purified RNA was redissolved in HPLC buffer A and the elution profile monitored. The gradients and buffers used were the same as for the preparative injections. For a successful purification a single peak at the appropriate elution time was observed (Figure 2.23).

2.5.2 Capillary Electrophoresis

Oligonucleotide purity was also analysed using an Applied Biosystems 270A capillary electrophoresis (CE) system (Figure 2.24). The profile for the self-complementary sequence r(CGCAAAUUUGCG)$_2$ was reproducible, using a number of different samples. The conditions for CE analysis are mildly denaturing, so that the profile shows a mixture of single stranded and double stranded RNA. The relative size of the single strand peak to the double strand peak was found to vary with the concentration of the stock used (more duplex for a higher concentration) and could be changed by heating the sample tube immediately before loading onto the capillary (Figure 2.24).

2.5.3 Base Composition Analysis of 5-Bromouridine RNA

Brominated derivatives of DNA and RNA are often required to aid crystallographic structure solution. It is important that the 'heavy atom' be specifically introduced, and stable under the conditions used to remove the protecting groups after synthesis. A base composition analysis by the

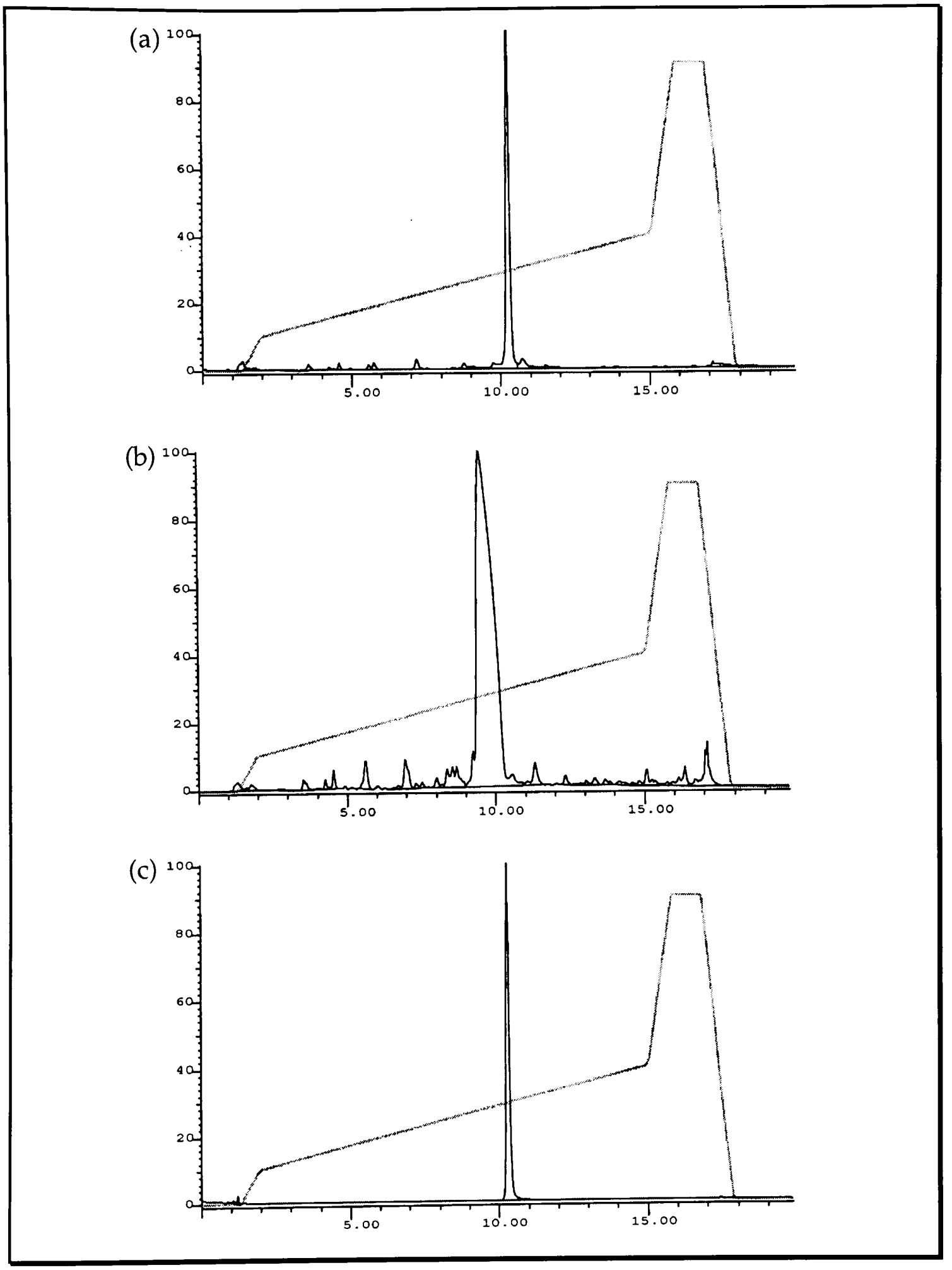


Figure 2.23 Anion exchange HPLC of R03 rGCUUCUCUUC: (a) before purification, (b) preparative injection, and (c) after purification.

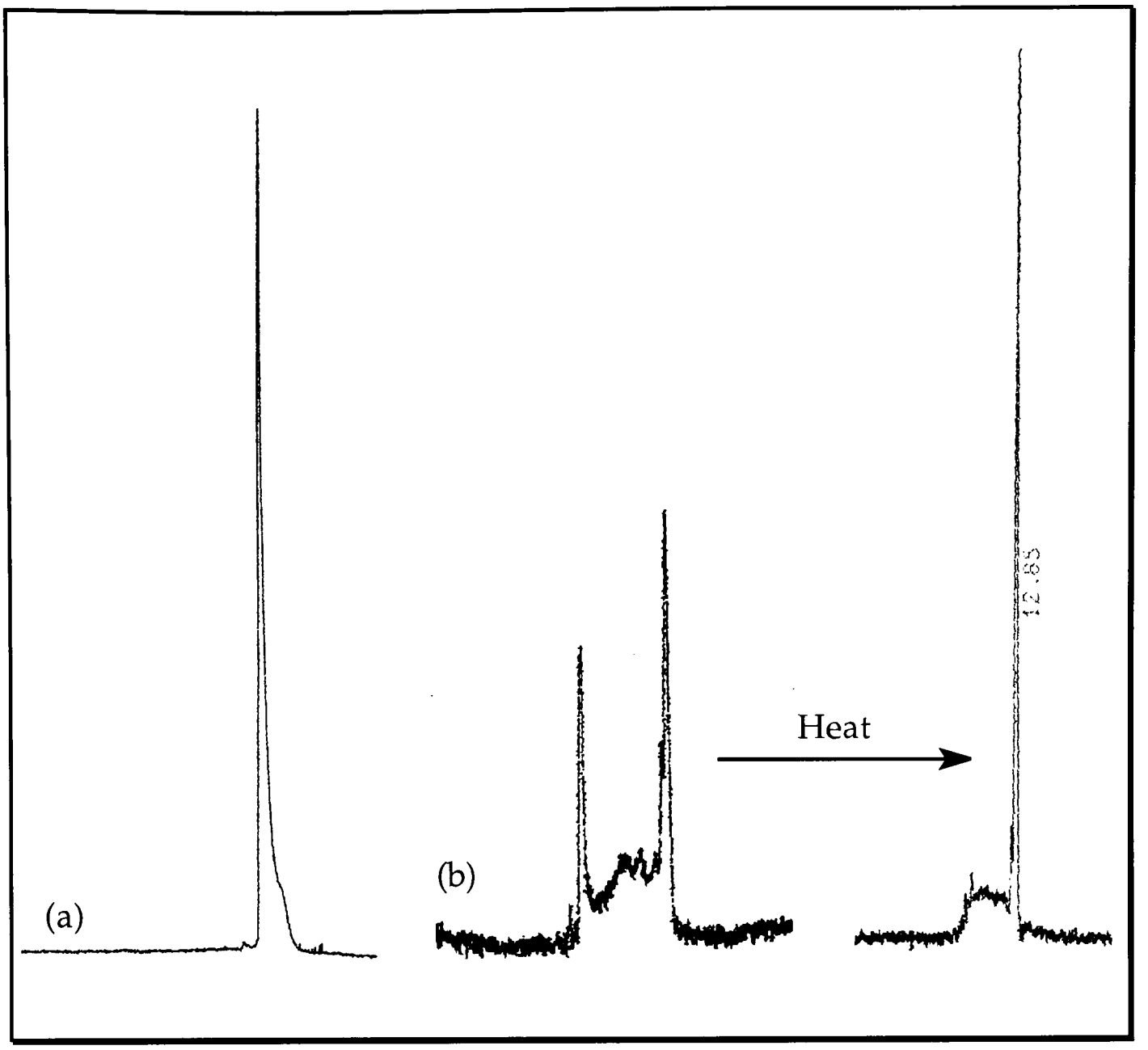


Figure 2.24 CE analysis of purified (a) R03 rGCUUCUCUUC and (b) R06 rCGCAAAUUUGCG

method of Eadie *et al.* [47] was performed on the duplex r(CGCAAAUU^{5Br}UGCG)$_2$. The RNA was prepared using tBDMS phosphoramidite monomers. The product was deprotected by treatment with cNH$_4$OH/ EtOH at 55°C for 4 hours followed by NMP/ TEA.3HF/ TEA for 60 minutes at 65°C, and then purified as described above (section 2.3). The 5-bromouridine was shown to be present by HPLC analysis of the enzyme digest, but an accurate base composition was not possible due to incomplete digestion. We have encountered this problem in our laboratory in the past for other duplexes.

To ensure that no modification of the bromouridine was occurring under the deprotection conditions, several tests were carried out with a 5-bromouridine standard. Treatment with cNH_4OH/ EtOH at 55°C for 5 hours, or room temperature for 24 hours caused no change, as observed by reversed phase HPLC or thin layer chromatography. An extended incubation of 24 hours at 55°C did cause some degradation, emphasising the need for labile base protecting groups. No degradation was observed after a 24 hour incubation of the nucleoside in NMP/ TEA.3HF/ TEA at 65°C.

2.5.4 Mass Spectrometry of 5-Bromouridine RNA

Confirmation of the integrity of the sequence rCGCAAAUU^{5Br}UGCG was obtained by electrospray mass spectrometry. A series of peaks was observed from mass 4220.2 upwards, corresponding to the RNA molecular ion peak with an increasing number of associated sodium ions. The expected average mass for the sequence is 4221.5. No peaks were observed at the mass calculated for the non-brominated RNA.

This result, taken with the enzyme digest and 5-bromonucleoside standard clearly indicate that brominated RNA is stable to the conditions used to prepare the material for these studies. This allowed the preparation of crystals of brominated RNA for X-ray diffraction experiments.

2.6 Conclusions

2.6.1 Preparation of RNA for Biophysical Studies

Automated chemical synthesis coupled with HPLC purification provides a rapid and reliable source of RNA for biophysical studies. The advantages of chemical synthesis are clear: any sequence may be synthesized at will and nucleotide modifications incorporated site specifically. Recent improvements in synthesis materials and deprotection protocols have meant that the rapid preparation of RNA in multi-milligram quantities is possible. The generation of highly pure material by our anion exchange and reversed phase HPLC protocols has allowed a number of detailed studies by high field NMR and X-ray crystallography. These studies are the subject of the next three chapters

2.6.2 Comparison of the Fpmp and tBDMS Methodologies for Oligoribonucleotide Synthesis

The relative merits of the Fpmp and tBDMS groups have been discussed in terms of the preparation of suitable building blocks for oligoribonucleotide assembly (see section 2.1.1.2). In this, the Fpmp group appears a more attractive proposition. However, in the synthesis and purification of RNA oligonucleotides for biophysical studies, the tBDMS group is considerably more suitable.

While phosphoramidite monomers with each type of protection are commercially available, there are many suppliers of tBDMS monomers compared with only one for Fpmp. This has often led to considerable delays in supply of Fpmp precursors. More significantly, the wider use of tBDMS monomers has led to a much wider availability of modified

nucleotide containing phosphoramidites using this protection strategy. This was of particular importance in this study. Also, considerably more effort has been invested by a number of research groups in the development of the tBDMS synthesis protocols and deprotection chemistry. These improvements have made the synthesis, deprotection and purification of tBDMS RNAs considerably more efficient and straightforward than Fpmp RNAs.

Probably the most important factor, however, is the quality of the final product as this will determine the success or otherwise of the planned studies. Both strategies proved suitable for the synthesis of short to moderate length RNAs (up to 25 mers). With the Fpmp protection, considerable problems were encountered with longer sequences, both in the integrity of the product and the removal of the Fpmp groups. Modification of the trityl group improved the initial purification step, but cannot remove the deprotection problem. The only significant problem with the tBDMS group is guaranteeing the stereochemical integrity of the monomer units. We have collected some evidence that commercial tBDMS monomers may be contaminated with very small amounts of the incorrect isomer and that this may be batch dependant. Despite this one reservation, it is clear that tBDMS is currently the most suitable protecting group for the 2'-hydroxyl in chemical RNA synthesis.

2.7 References

1. Milligan, J.F. and Uhlenbeck, O.C. (1989). *Meth. Enzymol.*, **180**, 51-62.

2. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987). *Nucleic Acids Res.*, **15**, 8783-8798.

3. Groebe, D.R and Uhlenbeck, O.C. (1988). *Nucleic Acids Res.*, **16**, 11725-11735.

4. Axelrod, V.D. and Kramer, F.R. (1985). *Biochemistry*, **24**, 5716-5723.

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

6. Middleton, T., Herlihy, W. C., Schimmel, P. R. and Munro, H. N. (1985). *Anal. Biochem.*, **144**, 110-117.

7. Ohtsuka, E., Tanaka, S., Tanaka, T., Miyake, T., Markham, A.F., Nakagawa, E., Wakabayashi, T., Taniyama, Y., Nishikawa, S., Fukumoto, R., Uemura, H., Doi, T., Tonkunga, T. and Ikehara, M. (1981). *Proc. Natl. Acad. Sci. USA*, **78**, 5493-5497.

8. Yu, W. (1984). *Acc. Chem. Res.*, **17**, 393-397.

9. Letsinger, R.L. and Mahadevan, V. (1965). *J. Am. Chem. Soc.*, **87**, 3526-3527.

10. Beaucage , S.L. and Iyer, R.P. (1992). *Tetrahedron*, **48**, 2223-2331.

11. Reese, C.B., Saffhil, R. and Sulston, J.E. (1967). *J. Am. Chem. Soc.*, **89**, 3366-3368.

12. Reese, C.B. and Skone, P.A. (1985). *Nucleic Acids Res.*, **13**, 5215-5331.

13. Christodoulou, C., Agrawal, S. and Gait, M.J. (1986). *Tetrahedron Lett.*, **27**, 1521-1522.

14. Rao, M.V., Reese, C.B., Schehlmann, V. and Yu, P.S. (1993). *J. Chem. Soc. Perkin Trans. I*, 43-55.

15. Ogilvie, K.K., Beucage, S.L., Schifman, A.L., Theriault, N.Y. and Sadana, K.L. (1978). *Can. J. Chem.*, **56**, 2768-2780.

16. Usman, N., Ogilvie, K.K., Jiang, M.-Y., and Cedegren, R.J. (1987). *J. Am. Chem. Soc.*, **109**, 7845-7854.

17. Scaringe, S.A., Francklyn, C. and Usman, N. (1990). *Nucleic Acids Res.*, **18**, 5433-5441.

18. Jones, S.S. and Reese, C.B. (1979). *J. Chem. Soc. Perkin Trans. I*, 2762-2764

19. Ohtsuka, E., Fujiyama, K. and Ikehara, M. (1981). *Nucleic Acids Res.*, **9**, 3503-3522.

20. Tanaka, T., Tamatsukuri, S. and Ikehara, M. (1986). *Nucleic Acids Res.*, **14**, 6265-6279.

21. Hayes, J.A., Brunden, M.J., Gilham, P.T. and Gough, G.R. (1985). *Tetrahedron Lett.*, **26**, 2407-2410.

22. Rastogi, H. and Usher, D.A. (1995). *Nucleic Acids Res.*, **23**, 4872-4877.

23. Ogden, R.C. and Adams, D.A. (1987). *Meth. Enzymol.*, **152**, 61-87.

24. Johns, D. (1989). In *HPLC of Macromolecules - A Practical Approach*, R.W.A. Oliver Ed., IRL Press at Oxford University Press, Oxford.

25. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995). *Nucleic Acids Res.*, **23**, 2677-2684.

26. McCollum, C. and Andrus, A. (1991). *Tetrahedron Lett.*, **32**, 4069-4072.

27. Vinayak, R. (1993). *Methods: A Companion to Methods in Enzymology*, **5**, 7-18.

28. Applied Biosystems (1995). *User Bulletin 91: RNA synthesis.*

29 Gait, M.J., Pritchard, C. and Slim, G. (1991). In *Oligonucleotides and Analogues*, Eckstein, F. (Ed.), Oxford University Press.

30. Wu, T., Ogilvie, K.K. and Pon, R.T. (1989). *Nucleic Acids Res.*, **17**, 3501-3517.

31. Vinayak, R., Anderson, P., McCollum, C., and Hampel, A. (1992). *Nucleic Acids Res.*, **20**, 1265-1269.

32. Westman, E. and Stromberg, R. (1994). *Nucleic Acids Res.*, **22**, 2430-2431.

33. Hogrefe, R.I., McCaffrey, A.P., Borozdina, L.U., McCampbell, E.S. and Vaghefi, M.M. (1993). *Nucleic Acids Res.*, **21**, 4739-4741.

34. Gasparutto, D., Livache, T., Bazin, H., Duplaa, A., Guy, A., Khorlin, A., Molko, D., Roget, A. and Téoule, R. (1992). *Nucleic Acids Res.*, **20**, 5159-5166.

35. Murray, J.B., Collier, A.K. and Arnold, J.R.P. (1994). *Anal. Biochem.*, **218**, 177-184.

36. Rozners, E., Westman, E. and Strömberg, R, (1994). *Nucleic Acids Res.*, **22**, 94-99.

37. Morgan, M.A., Kazakov, S.A. and Hecht, S.M. (1995). *Nucleic Acids Res.*, **23**, 3949-3953.

38. Capaldi, D.C. and Reese, C.B. (1994). *Nucleic Acids Res.*, **22**, 2209-2216.

39. Brown T. and Brown, D.J.S. (1992). *Meth. Enzymol.*, **211**, 20-35.

40. Pingoud, A., Fliess, A. and Pingoud, V. (1989). In *HPLC of Macromolecules - A Practical Appraoch*, Oliver, R.W.A. (Ed.), IRL Press at Oxford University Press, Oxford.

41. Haupt, W. and Pingoud, A. (1983). *J. Chromatography*, **260**, 419-427.

42. Dionex Corporation (1990). *Installation Instructions and Troubleshooting Guide for Nucleopac PA-100,* Document No. 034411-01.

43. Anderson, A.C., Scaringe, S.A., Earp, B.E. and Frederick, C.A. (1996), *RNA*, **2**, 110-117.

44. Görtz, H.-H and Seliger, H. (1981). *Angew. Chem. Int. Ed. Eng.*, **20**, 8, 681-683.

45. Görtz, H.-H and Seliger, H. (1981). *Angew. Chem. Int. Ed. Eng.*, **20**, 8, 683-684.

46. Churcher, M.J., Lowe, A.D., Gait, M.J. and Karn, J. (1995). *Biochemistry*, **92**, 2408-2412.

47. Eadie, J.S., McBride, L.J., Efcavitch, J.W., Hoff, L.B. and Cathcart, R. (1987). *Anal. Biochem.*, **165**, 442-447.

Chapter 3

Crystallization of Oligonucleotides

3. Crystallization of Oligonucleotides

3.1 Introduction

X-ray crystallography is the most powerful tool available for detailed structural studies on nucleic acids. A vast array of crystal structures are available of DNA duplexes and DNA complexed with proteins and a variety of small molecule ligands [1,2]. Such studies have provided invaluable information on the nature of interactions between DNA and proteins or ligands, as well as the effects of mispairings and chemical modifications on DNA structure. Crystallographic information on RNA structure, with the notable exception of the tRNAs, is presently considerably more sparse. The main driving force behind the rapid build up of DNA crystal structure data was the success of solid phase DNA synthesis. Only recently has it become possible to produce RNA of similar purity and in the quantities required, as the chemical synthesis of RNA has lagged behind that of DNA. As a result, the difficulties in obtaining RNA crystals that diffract to high resolution and of obtaining heavy atom derivatives have severely limited the number of RNA structures solved by X-ray crystallography.

The majority of RNA crystal structures solved to date have been of short duplexes [3,4,5], although some of these contain interesting mismatched base pairs [6,7,8]. Very recently, a major advance in RNA crystallography was made with the structure of part of a Group I intron at 2.8Å resolution [9]. A number of RNA-protein co-crystal X-ray structures have also been solved. These include three aminoacyl tRNA synthetases with their corresponding tRNAs [10], the MS-2 bacteriophage coat protein with an RNA hairpin [11] and the U1A protein with a 21 nucleotide hairpin [12].

The latter structure nicely illustrates a number of RNA-protein interactions and shows how both become restricted on binding.

Advances in RNA chemical synthesis have meant that the production of material for crystallization is now a considerably less formidable task. This is reflected in another recent milestone publication: the first crystal structure of a full RNA hammerhead ribozyme [13,14]. It has become clear that in growing single crystals, the RNA sequence itself is at least as important a factor as the conditions under which the crystals are grown. An earlier study demonstrated that a single base pair change in the sequence of an RNA hammerhead ribozyme with an all DNA substrate led to different crystal forms under some but not all crystallizing conditions [15]. By selecting the best of these crystals for analysis, the first crystal structure of a hammerhead ribozyme was determined [16]. Similarly, in the co-crystallization of the U1A protein [12], the crystal quality was greatly improved by altering the final base pair of the hairpin stem and the introduction of an overhanging U nucleotide. Even with the use of a number of different RNA sequences, a range of conditions and relative concentrations of the components of the solution may need to be examined. These conditions include the crystallizing agent, temperature, buffer, pH and additives such as polyamines and metal ions. The most commonly used crystallizing agents for nucleic acids include 2-methyl-2,4-pentane diol (MPD), isopropanol, polyethylene glycols (PEGs; 400-8000MW) and ammonium sulphate. Condition screens of these agents with various buffers and additives have been published, and have proven successful in the crystallization of RNA [13,17].

3.2 Mixing of Complementary Strands

Following the extensive purification procedures described in Chapter 2, the oligonucleotide strands must, if necessary, be mixed accurately and then placed under conditions suitable for crystallization. Self-complementary, or palindromic, duplexes are often used for crystallographic studies as this removes the need for accurate mixing of two different strands. The dodecamer DNA and RNA duplexes discussed later (Chapter 4) are of this type. Similarly, for RNA hairpins no mixing is required as the helix stem formation is intramolecular. However, for the DNA.RNA hybrids system studied here (Chapter 5) it was a requirement that each strand be rich in either purine or pyrimidine nucleotides. As a result, the duplexes could not be palindromic and an accurate way of mixing the strands had to be found.

A number of methods have been examined previously in our laboratory with the aim of mixing non-complementary strands for crystallographic studies. These include analysing mixtures of various volumes of stocks by non-denaturing capillary electrophoresis and non-denaturing FPLC purification of annealed strands [18]. Unfortunately, the success of these methods was never validated by the growth of single crystals. As stated previously, the absorbance at 260nm of an oligonucleotide solution can provide an accurate measure of the concentration if the molar extinction coefficient is known (section 1.5.3). Accurate molar extinction coefficients are known for all common nucleotides and nucleosides. Thus, the concentration of an oligonucleotide solution may be determined by measuring the absorbance of a sample before and after complete enzymatic digestion.

Nuclease P1 digests were used to determine extinction coefficients for both single stranded and duplex oligonucleotides. The increase in

absorbance upon addition of the enzyme (hyperchromicity) was monitored and, from known values for the nucleotide monophosphates [19], observed extinction coefficients could be calculated. Complete digestion generally took 10-20 minutes for single strands and 30-60 minutes for duplexes. The only exception was for RNA duplex R06, $r(CGCAAAUUUGCG)_2$, for which complete digestion could not be obtained. For particularly unstable duplexes, observed molar extinction coefficients were considerably higher than expected (corresponding to a low hyperchromicity). This was a particular problem with palindromic sequences, as hairpin rather than duplex formation became prevalent. The addition of counterions, Na^+ and Mg^{2+}, to the digest buffer was found to sufficiently stabilise the duplex for values to be calculated. The time taken for complete digestion under these conditions was around two hours. The values obtained (Appendix B) compare favourably with estimated values which are expected to be accurate to within 10% [20].

3.3 Growth of Crystals for X-ray Crystallographic Studies

One of the main objectives of this work was to obtain single crystals, suitable for X-ray diffraction studies, of a number of different RNAs. The two different methods used to obtain crystals are described in detail below. For the first, sitting drop vapour diffusion, conditions were used that have previously proven successful in our laboratory for similar sequences [8,21]. However, this method did not provide any crystals for a number of sequences, so the use of a conditions screen developed by Scott *et al.* [13] was examined.

Crystals large enough for diffraction tests or later, data collection, were sealed in glass capillaries (Figure 3.1). Crystals were drawn from the droplets by capillary action and the mother liquor carefully removed using a smaller diameter glass capillary (shown on the far right in Figure 3.1). Finally, the ends of the capillary were sealed with wax to prevent loss of solvent from the crystal. Data collection was primarily performed using synchrotron radiation sources at Daresbury Laboratories and the ESRF, Grenoble.

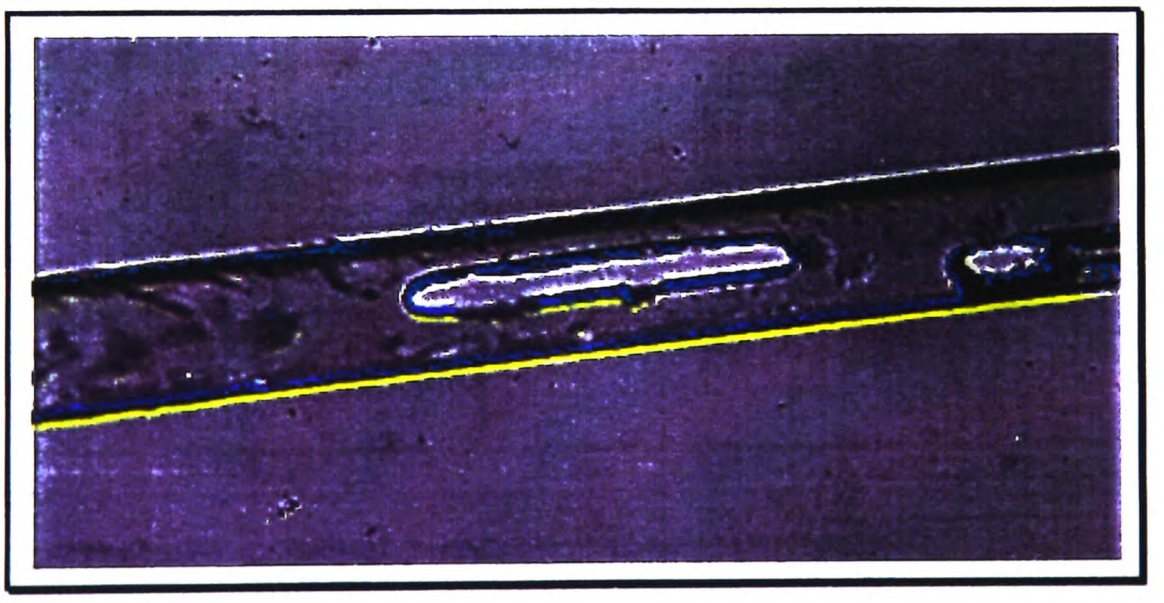


Figure 3.1 Mounting an oligonucleotide crystal for X-ray diffraction.

3.3.1 Sitting Drop Vapour Diffusion Method

Oligonucleotides were dissolved in 50mM sodium cacodylate buffer (pH 6.5) to a concentration of approximately $0.5OD_{260}/\mu l$ (about 3mM). Various concentrations of $MgCl_2$, 50% aqueous MDP and spermine were screened, over the ranges given in Table 3.1.

Solution	Stock concentration (mM)	Drop concentration (mM)
Cacodylate	50	12.5
$MgCl_2$	250	6 - 60
Spermine.4HCl	10	0.5 - 4.5
MPD	50% aqueous	2.5 - 15%

Table 3.1 Range of concentrations used in sitting drop crystallization trials

Sitting drops of 20-25µl, with a final concentration of RNA of 0.6-0.9mM, were sealed in glass plates with a central reservoir of 50% aqueous MPD for the initial trials (Figure 3.2a). The saturation point in the oligonucleotide drops is reached as the central reservoir absorbs water from the sealed atmosphere, slowly increasing the concentration of the precipitating agent. As this can otherwise be quite a slow process, the concentration of the central reservoir can be increased once ideal conditions have been found.

3.3.1.1 RNA Duplex r(CGCAAAUUUGCG)$_2$

Initial trials with the target duplex, R06 r(CGCAAAUUUGCG)$_2$, at best gave very small needles or micro-crystals but generally only amorphous precipitate over a wide range of spermine, $MgCl_2$ and MPD concentrations. To obtain quality single crystals more than just careful variation of precipitant and ion concentration was required. For an RNA

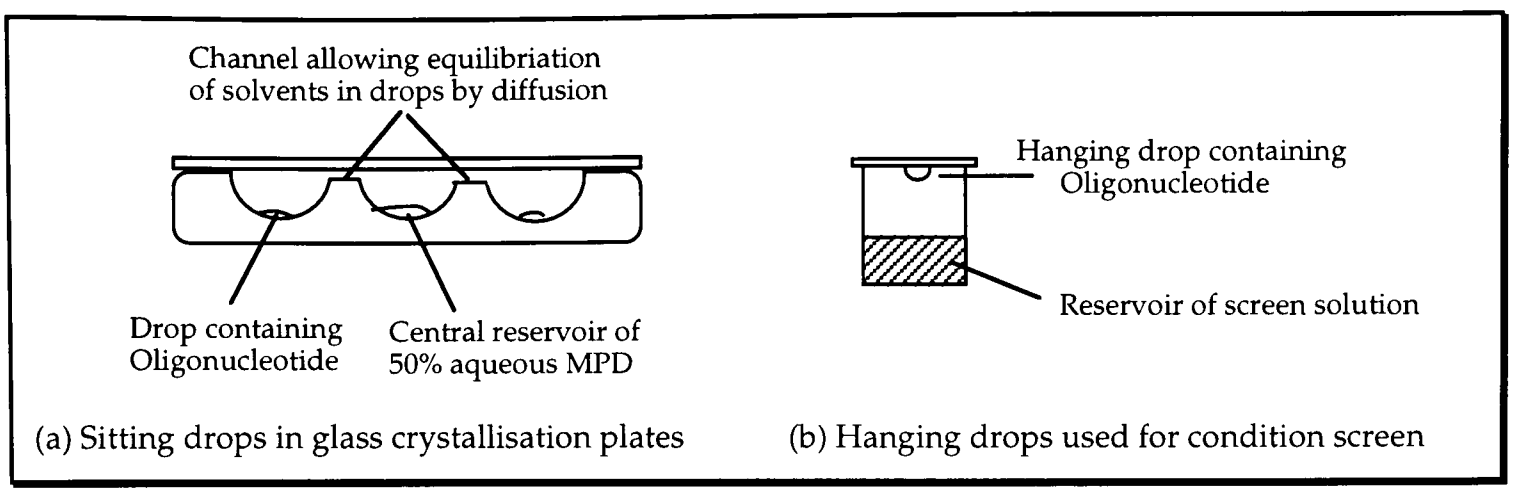


Figure 3.2 Methods of Oligonucleotide Crystallization

duplex a most considerable change to crystal packing is likely to arise from the addition of an overhanging base to either the 3′ or 5′-end (giving eight possible sequences). However, this should not significantly disrupt the area of interest in the main body of the duplex. A single 1.0μmole synthesis of each 13-mer sequence of this type provided sufficient material for a range of precipitant concentrations for each, giving the results shown in Table 3.2.

Construct	Duplex (5′-3′)	Crystals Obtained
R06	$(CGCAAAUUUGCG)_2$	Very small needles/ Microcrystals
3′-A	$(CGCAAAUUUGCGA)_2$	None (amorphous precipitate)
3′-G	$(CGCAAAUUUGCGG)_2$	Large hexagonal prisms
3′-C	$(CGCAAAUUUGCGC)_2$	Microcrystals
3′-U	$(CGCAAAUUUGCGU)_2$	Thin plates
5′-A	$(ACGCAAAUUUGCG)_2$	None (amorphous precipitate)
5′-G	$(GCGCAAAUUUGCG)_2$	Very small prisms
5′-C	$(CCGCAAAUUUGCG)_2$	Microcrystals
5′-U	$(UCGCAAAUUUGCG)_2$	Fibrous crystals

Table 3.2 Crystals obtained for the sequence r$(CGCAAAUUUGCG)_2$ and eight terminal overhang constructs.

The sequence r(CGCAAAUUUGCGG)$_2$ gave the most promising crystals and crystallizing conditions were varied carefully to optimization. The best crystals were obtained within 7-14 days from drops containing 25mM $MgCl_2$, 1.5mM spermine and 2.5% MPD with a central reservoir of 100% MPD (Figure 3.3a,b). The most important factor influencing the crystal growth for this sequence was found to be the MPD concentration. At higher concentrations (around 5-7%) considerably poorer crystals were obtained (Figure 3.3c). Crystals of the bromouridine derivative r(CGCAAAUU$^{5\text{-Br}}$UGCGG)$_2$ were obtained under almost identical conditions with no further optimization (Figure 3.3d).

A crystal of dimensions 0.10 x 0.10 x 0.15 mm was sealed in a capillary and used for data collection on station PX9.5 at the synchrotron radiation source at Daresbury Laboratory. Data to 2Å were recorded and processed. The unit cell is trigonal, R32, with a = b = 39.79, c = 215.79Å and $\alpha = \beta = 90$, $\gamma = 120°$. The data is 98% complete with redundancy of 5, R-merge 4.5%. The asymmetric unit contains a single duplex, and molecular replacement indicates infinite helices along the c-axis. These crystals were shown to diffract to 1.65Å in a resolution test. However, the largest native and the brominated RNA crystals, obtained under fully optimized conditions, have not yet been used for diffraction.

3.3.1.2 DNA.RNA Chimera

Two DNA.RNA chimeric duplexes were prepared, based on the well characterised Drew-Dickerson dodecamer sequence [1]. Both were self-complementary duplexes containing six RNA nucleotides covalently attached to six DNA nucleotides: R14 [r(CGCGAA)d(TTCGCG)]$_2$ and R15 [d(CGCGAA)r(UUCGCG)]$_2$. Each base pair in the duplex consists of one ribo- and one deoxyribonucleotide.

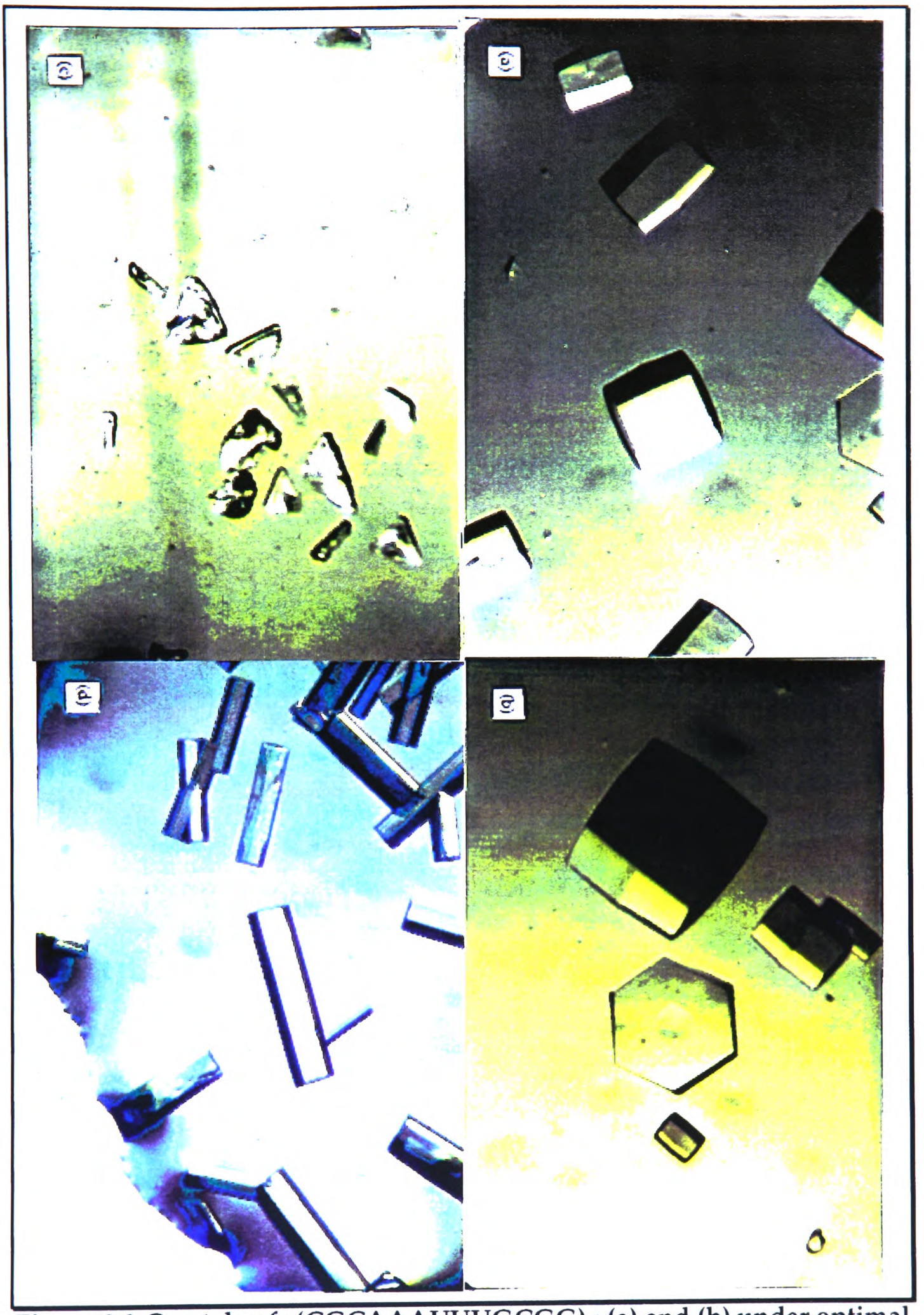


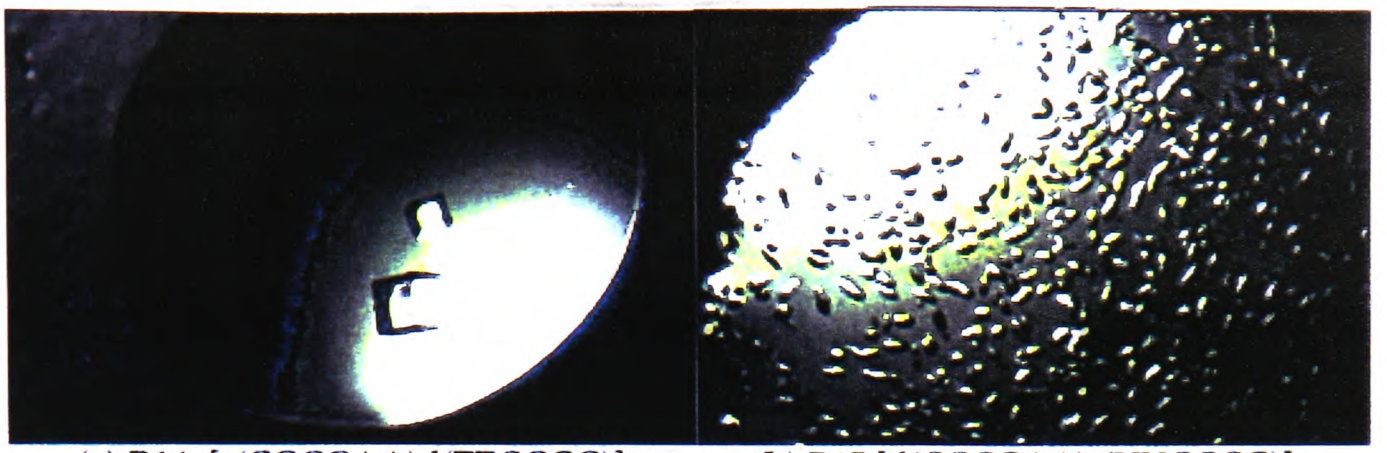
Figure 3.3 Crystals of r(CGCAAAUUUGCGG)$_2$; (a) and (b) under optimal conditions, (c) high MPD concentrations, (d) 5-bromouridine derivative under optimal conditions.

Crystals of both duplexes were obtained in the initial screens of spermine, MPD and $MgCl_2$ concentration. However, the actual conditions, crystal forms and rate of growth were very different for the two sequences. Crystals of the chimeric duplex R14 formed best in solutions containing around 12.5mM $MgCl_2$ and between 2 and 4mM spermine. The crystals grew in 4 to 8 weeks with a central reservoir of 50% aqueous MPD, as large polyhedra with sharp edges (Figure 3.4a). The crystals appeared to be of good optical order, as indicated by a sharp extinction when viewed with polarised light. In contrast for chimera R15, under the same conditions very small needles were obtained within minutes of mixing the solutions. The largest needles were obtained in solutions containing around 37.5mM $MgCl_2$ and 0.5 to 1.5mM spermine (Figure 3.4b). The majority of these crystals grew in 1-3 days, but could not be grown to a size suitable for diffraction tests. Over longer periods larger sheets of crystal grew under the same conditions (Figure 3.4c). These and a number of the R14 crystals were tested at the Daresbury Laboratories.

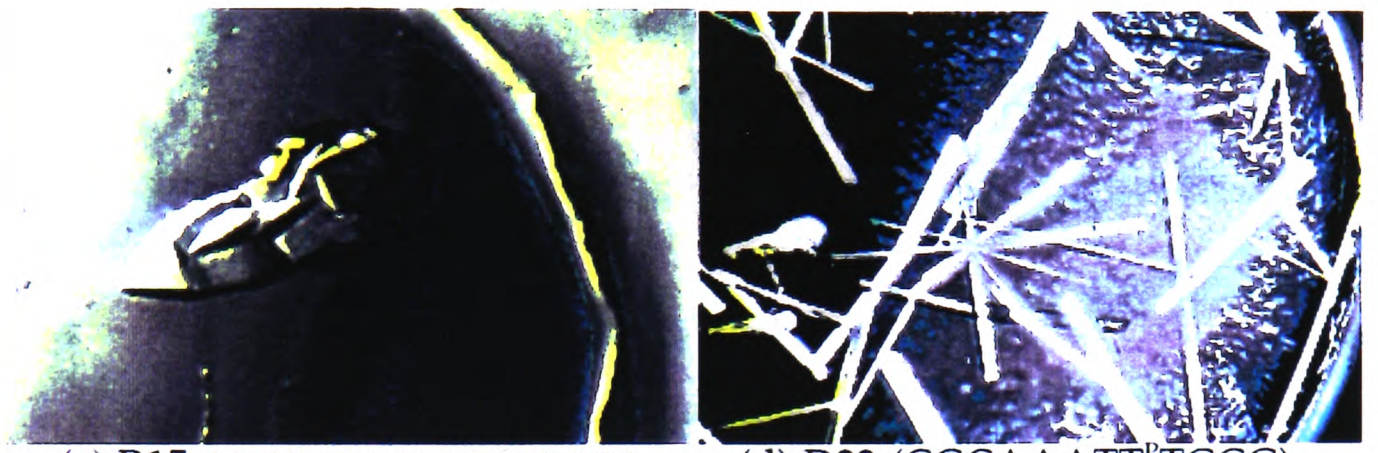
Crystals of the R14 chimera were found to be hexagonal, with a=b=50 and c=85Å, diffracting to a resolution of around 2.5Å. Despite the apparent good optical order, the diffraction patterns obtained were quite poor. This can be attributed to the rotational disorder of the helices and the high degree of solvation of nucleic acid crystals [1]. The larger crystals of the R15 chimera showed diffraction patterns similar to those in fibre diffraction experiments. The initial diffraction data suggested a unit cell with one very long edge. No further data was collected on crystals of either duplex.

3.3.1.3 DNA Duplexes

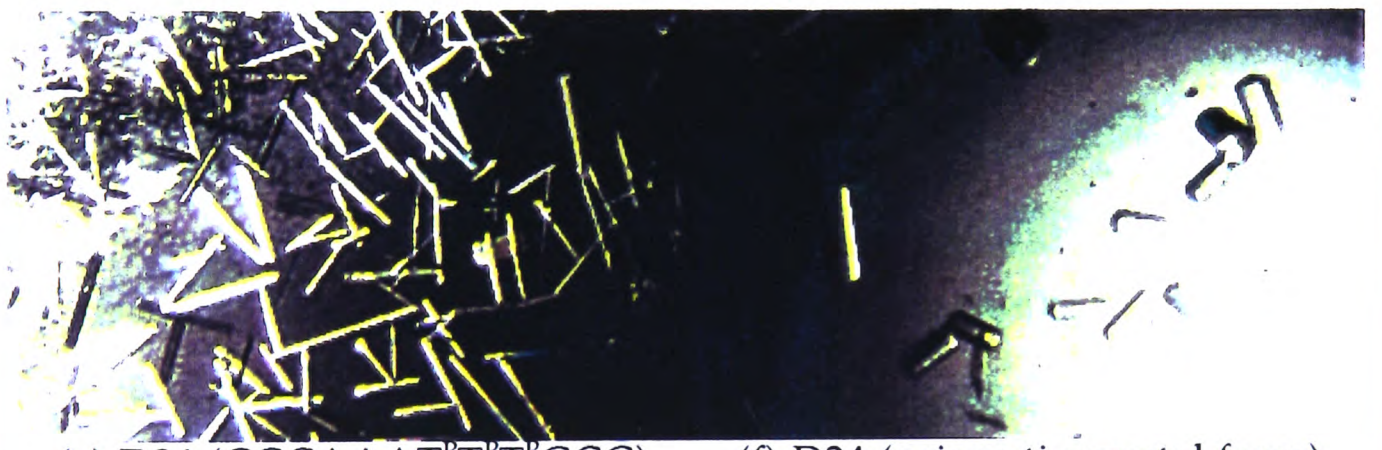
The effects of a number of chemical modifications on the structure and stability of nucleic acids was examined during this work. Modified DNA



(a) R14, $[r(CGCGAA)d(TTCGCG)]_2$ (b) R15 $[d(CGCGAA)r(UUCGCG)]_2$



(c) R15 $[d(CGCGAA)r(UUCGCG)]_2$ (d) D22 $(CGCAAATT^{P}TGCG)_2$



(e) D22 $(CGCAAAT^{P}T^{P}T^{P}GCG)_2$ (f) D24 (prismatic crystal form)

(g) 4′-thio-5-ethyl dU DNA (h) as (g), conditions optimized

Figure 3.4 Crystals of RNA, DNA and chimeric duplexes.

nucleotides were introduced to two well characterised systems: the B-DNA Drew-Dickerson type dodecamer and the Z-DNA $d(CGCGCG)_2$.

(a) 5-Propyne Deoxyuridine

The presence of propyne groups dramatically increases the stability of nucleic acid duplexes (see Section 5.2.1), but the reasons behind this are largely unknown at present. A total of four DNA duplexes with the sequence $d(CGCAAATTTGCG)_2$ were prepared for crystallization trials. These contained a single modification at the 3'- and central thymidines (D21 and D22), a double modification of these two thymidines (D23) and modification of all three positions (D24).

Crystals were obtained of each sequence over a range of spermine and $MgCl_2$ concentrations. Crystals generally formed as large needles (Figure 3.4d,e), as expected from previous work with the native duplex [22]. In some cases however, chunkier prismatic crystals were obtained (Figure 3.4f). All crystals exhibited a sharp extinction of polarised light, and a number of each sequence grew large enough for diffraction tests. The diffraction patterns obtained indicated that the crystals were quite fibrous and not suitable for a full structural determination. Given the wealth of information that has been derived from studies of similar sequences with crystals of the same form, this seems an unusual finding. It is likely that the crystals deteriorated, as considerable delays between growing the crystals and obtaining beam time were common. Ageing of the crystals, such as microscopic cracking, could cause them to become fibrous with time. It is possible, of course, that the modifications made to the duplex do result in poor quality crystals under these conditions.

(b) 4'-Thio-5-Ethyl Deoxyuridine

The modified nucleotide was used in the preparation of the Drew-Dickerson dodecamer $d(CGCGAAT^*T^*CGCG)_2$, where T* is 4'-thio-5-ethyl

deoxyuridine. Large needle crystals were obtained as expected, the largest being approximately 2mm on the longest edge (Figure 3.4g,h). As with the propyne modified duplexes, only fibre diffraction patterns were obtained indicating severe ageing of the crystals.

(c) 4′-Thio Cytidine

The 4′-thio cytidine (C*) residue was incorporated into the sequences $d(CGCGC^*G)_2$ and $d(C^*GC^*GC^*G)_2$. The native duplex is Z-form DNA and is known to crystallize under conditions similar to those used previously but using isopropanol [23] in the place of MPD. In this case the alcohol diffuses into the oligonucleotide droplet to reach the saturation point of the solution initiating crystallization.

Initial trials, with a central reservoir of 50% isopropanol, gave microcrystallites for $d(CGCGC^*G)_2$ but only amorphous precipitate for $d(C^*GC^*GC^*G)_2$. With lower concentrations of isopropanol reservoir no precipitation was observed. The sparse matrix conditions screen (section 3.3.2) was also used and the two sequences again behaved differently. On dissolving in 10mM sodium cacodylate buffer solution with 1mM spermine, the fully modified sequence precipitated from solution. Clearly, the introduction of the modification considerably alters the structure of the oligonucleotide, with the multiple modifications having the most dramatic effect. No crystals were obtained at 8°C or room temperature for either sequence.

3.3.2 Sparse Matrix Screen

For crystallizing RNA and DNA.RNA hybrid sequences that had either proven difficult by the methods described above, or for which no guide was available in the literature, a screen of conditions was carried out. The screen chosen here follows the crystallizing conditions of Scott *et al.*, used

for their work with the hammerhead ribozyme [13]. The screen consists of 48 different solutions containing one of four buffers at a number of pHs, with various crystallizing agents and additives (the full details can be found in Chapter 7). The two types of sequence for which the screen proved most successful are described in detail below. A summary of the results with all the sequences tested is shown in Table 3.3.

3.3.2.1 DNA.RNA Hybrids

Crystals of the DNA.RNA hybrid R02/D03 r(GAAGAGAAGC). d(GCTTCTCTTC) were obtained within 1-3 days under a variety of crystallizing conditions (Table 3.3). A number of these (conditions 1, 12 and 22) were prismatic and large enough for diffraction tests from the initial screen (Figure 3.5a,b). Crystals obtained using condition 14 appeared as thick needles in the initial screen. When the crystal growth was slowed, either by adding 2µl of water to the hanging drop or by increasing the ratio of RNA solution to matrix solution, these crystals grew considerably larger as long prisms. Crystals from condition 22 were used to collect data to a resolution of 1.7Å on BL19 of the European Synchrotron Radiation Facility, Grenoble, at 0.905Å wavelength. The unit cell is trigonal with dimensions a=b=39.79, c=30.07Å and $\alpha=\beta=90$, $\gamma=120^\circ$. The space group of the crystals is P3 with one duplex in the asymmetric unit. Molecular replacement indicates that there are infinite helices parallel to the c-axis.

Crystals were also obtained for the same hybrid duplex with modifications in the DNA strand under different conditions (Table 3.3). Both the hybrids with deoxyuridine and 5-propyne deoxyuridine formed long thick needles (Figure 3.5c,d). When the best conditions were repeated with a range of oligonucleotide concentrations, significantly better crystals were obtained at higher concentrations for both sequences. To date, these crystals have not been used for data collection.

Code	Sequence	Temp.	Crystals
R16/D16	5′-rGCGCA_4GCGC 3′-dCGCGT_4CGCG	RT	3, 11, 12, 13, 17, 18, 20, 33, 38, 40, 41, 43, 45
R17/D17	5′-dGCGCA_4GCGC 3′-rCCGCU_4CGCG	RT	33
R02/R03	5′-rGAAGAGAAGC 3′-rCUUCUCUUCG	RT	7, 23, 33, 40, 41, 45, 47
D02/D03	5′-dGAAGAGAAGC 3′-dCTTCTCTTCG	RT	4, 12, 13, 14, 23, 25, 26, 28, 33, 40, 41, 43, 44
D02/D03	5′-dGAAGAGAAGC 3′-dCTTCTCTTCG	8°C	4, 12, 13, 25, 26, 33, 38, 40, 41, 43, 44
D02/R03	5′-dGAAGAGAAGC 3′-rCUUCUCUUCG	RT	12, 13, 14, 23, 33, 40, 43, 44
D02/R03	5′-dGAAGAGAAGC 3′-rCUUCUCUUCG	8°C	4, 12, 13, 14, 23, 33, 40, 43
R02/D03	5′-rGAAGAGAAGC 3′-dGCTTCTCTTC	RT	1, 4, 5, 7, 9, 12, 13, 14, 17, 22, 23, 24, 25, 26, 28, 33, 36, 38, 40, 42, 43, 44, 45, 47, 48
R02/D19	5′-rGAAGAGAAGC 3′-dCU^{P}U^{P}CU^{P}CU^{P}U^{P}CG (U^P=5-propyne dU)	RT	1, 2, 3, 4, 6, 8, 9, 10, 11, 16, 18, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 39, 41, 42, 44
R02/D29	5′-rGAAGAGAAGC 3′-dCUUCUCUUCG	RT	26, 33, 40, 41, 43, 45, 47
R04	21 mer hairpin	20°C	Amorphous only (22 conditions)
R04 (5′A)	21 mer hairpin (5′-A overhang)	20°C	1, 3, 5, 7, 13, 14, 15, 17, 18, 20, 22, 23, 26, 29, 31

Table 3.3 Summary of crystallization results using the conditions screen of Scott *et al.* [13] (underlined conditions gave best crystals for optimization, double underlined conditions gave crystals suitable for diffraction tests).

R04 (5'G)	21 mer hairpin (5'-G overhang)	20°C	1, 3, 5, 7, 20, 28, 29, 30, 42, 44
R09	21 mer hairpin (5'-GG overhang)	20°C	3, 5, 7, 9, 13, 15, 20, 29, 31
R06	$(CGCAAAUUUGCG)_2$	20°C	45
R10	$(CGCAUAUAUGCG)_2$	20°C	46
R11	$(CGCUUUAAAGCG)_2$	20°C	42, 46
R14	$[r(CGCGAA)d(TTCGCG)]_2$	20°C	44
R15	$[d(CGCGAA)r(UUCGCG)]_2$	20°C	13, 17, 33, 40, 41, 47
R15 (3'A)	$[d(CGCGAA)r(UUCGCGA)]_2$	20°C	None
R15 (3'G)	$[d(CGCGAA)r(UUCGCGG)]_2$	20°C	None

Table 3.3 (Continued).

The conditions screen proved less successful with the opposite hybrid duplex D02/R03 d(GAAGAGAAGC).r(GCUUCUCUUC) and the RNA and DNA homoduplexes of this sequence. Crystals of each duplex were obtained but the conditions need to be optimized to obtain a size suitable for diffraction tests.

3.3.2.2 RNA Hairpins

Crystals of a number of hairpin RNAs based on the 21 mer sequence R04 rAGCCCGCCUAAUGAGCGGGCU were grown. This hairpin forms the terminator loop in the mRNA transcript of the *trp* leader segment in *E.*

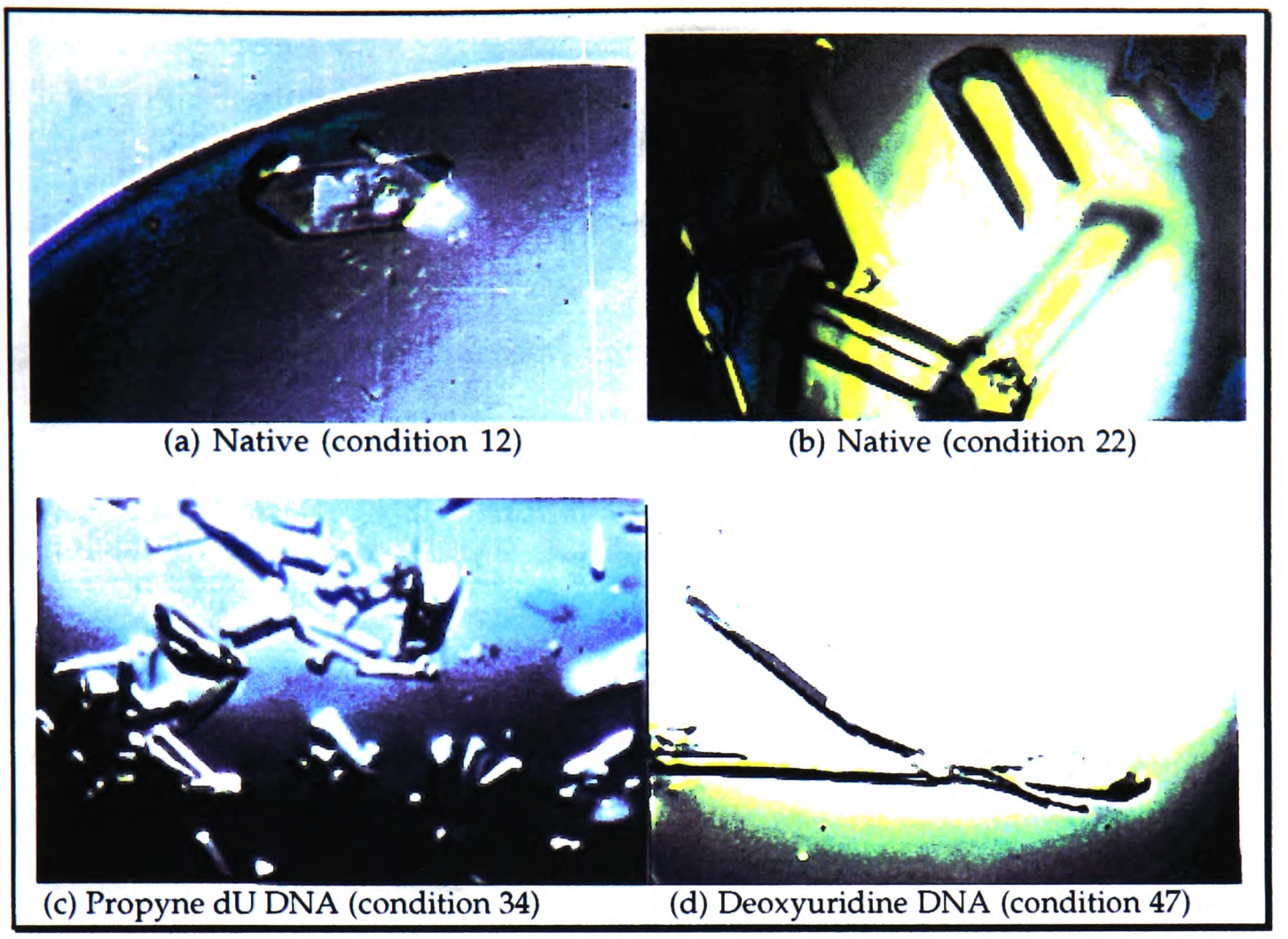


Figure 3.5 Crystals of DNA.RNA hybrid r(GAAGAGAAGC). d(GCTTCTCTTC): (a) & (b) native, (c) propyne dU DNA and (d) dU DNA.

Coli [24]. The effect of an overhanging nucleotide is clearly seen from the results in Table 3.3. The condition screens for R04, R04 (5'G) and R09 were conducted at similar RNA concentrations, under identical crystallizing conditions. For the sequence with no overhang only amorphous precipitate was obtained. For the two sequences with overhangs, crystals formed under a number of conditions. The best crystals of R04 (5'G) from conditions 20, 28, 29, 30 and 42 were small but chunky prisms. Similar crystals were obtained for R09 with conditions 9, 29, 31. The effect the overhanging nucleotide exerts on the crystallization is probably to some extent one of crystal packing, as different interactions become possible between adjacent molecules. However, comparing these sequences, it seems highly likely that the differences observed are due in part to the

added stability the overhanging base gives to the stem structure. The hairpin RNA was originally synthesized as part of an NMR study. It was found that the addition of a 5'-GG overhang greatly reduced the degree of 'fraying' at the end of the helical region [personal communication, Dr V. Ramesh, University of Southampton]. Reducing such fraying, and thereby reducing the conformational flexibility of the strand, is certain to have a dramatic effect on the likelihood of the sequence crystallizing.

The best crystals of all the hairpin RNAs were obtained for the 5'A overhang under conditions 14, 22 and 31 (prisms or thick needles). Crystals of this sequence formed under more conditions than for any of the other RNA hairpins (Table 3.3). While the nature of the overhanging base may play a part in this, the most significant difference lies in the concentration of the RNA. The screen for the 5'A overhang was conducted at a considerably *lower* concentration of RNA (0.2 OD_{260} compared to 0.5 OD_{260} per drop). For this type of sequence a lower concentration of RNA appears to have significantly improved the size and quality of crystals formed. Higher concentrations tended to lead to smaller crystals and considerably more amorphous or microcrystalline precipitate. To date, no diffraction data has been obtained on these crystals. However, a small number of the crystals obtained were of a suitable size for initial diffraction tests and conditions are known which may yield large crystals with some optimization.

3.3.2.3 Other RNA Sequences

The sparse matrix screen was used to obtain crystals of a number of other RNA and DNA.RNA duplexes: R16/D16, R17/D17, R06, R10, R11, R14 and R15 (Table 3.3). For these sequences, very few crystals formed and where they did, they were generally of poor quality.

The chimera R14 was expected to form crystals given the relative success with this sequence previously (see section 3.3.1.2). However, only very small crystals were obtained and under only one crystallizing condition. Increasing the concentration of spermine by adding 2µl of a 6mM stock to each droplet failed to increase the amount of precipitation, crystalline or amorphous. With the opposite chimera, the only promising crystals obtained were at the highest concentration (0.55 OD_{260} per drop) of oligonucleotide for R15. Unfortunately, to date, no data has been collected on these crystals. The screens at lower concentration (0.38 OD_{260} per drop) of R15 (3'A) and R15 (3'G) yielded no crystals at all. Similarly, no suitable crystals were obtained for any of the three RNA duplexes R06, R10 or R11.

When taken together, the results of the crystallization trials using the sparse matrix screen suggest that the concentration of the oligonucleotide in the droplets is crucial to the likelihood of success. For some sequences such as the RNA hairpins, a lower concentration favoured crystal growth. For others better crystals were obtained at higher concentrations. Where few or no crystals were obtained, it is likely that a suitable RNA concentration was not found. Fortuitously, a concentration was chosen that gave high quality crystals of the DNA.RNA hybrid R02/D03 with and without modifications almost immediately. It would appear, however, that the best approach would be to screen oligonucleotide concentrations in addition to crystallization conditions for any new sequence.

3.4 Conclusions

The success of the RNA synthesis and purification protocols described in Chapter 2 have been validated by the preparation of single crystals for X-ray diffraction studies. However, this work does also demonstrate that apparent crystal quality does not necessarily correlate with suitability for diffraction studies. For each crystal that diffracts well and to high resolution, many do not. Two solutions to this problem are presented here and in the recent literature. Firstly, the conditions of crystallization may be altered. While, at present, there are very few if any 'rules' to guide the choice of conditions, the use of sparse screens with many different conditions has proved successful. It was found, however, that the quality and quantity of crystals obtained in the initial screens depended quite heavily upon the concentration of the oligonucleotides in the drops. Secondly, and possibly most importantly, the RNA sequence may be subtly varied. Changes can be made that affect crystal packing and morphology, leading to crystals with improved diffracting power, but with minimal disruption to the area of interest.

Two high quality, high resolution data sets for two new oligonucleotides have been obtained: one of the RNA duplex r(CGCAAAUUUGCGG)$_2$ and the other of a DNA.RNA hybrid r(GAAGAGAAGC).d(GCTTCTCTTC). In the first case, where a guide to suitable conditions was available, a variation in the target sequence with overhanging nucleotides provided high quality single crystals. A sparse matrix conditions screen was used to obtain crystals of the second duplex, and a number of other RNA and DNA.RNA molecules. Once solved, both of these structures will make invaluable additions to the structural and thermodynamic data collected during the course of this work.

3.5 References

1. Kennard, O. and Hunter, W.N. (1991). *Angew. Chem. Int. Ed. Engl.*, **30**, 1254-1277.

2. Wahl, M.C. and Sundaralingam, M. (1995). *Curr. Op. Str. Biol.*, **5**, 282-295.

3. Dock-Bregeon, A.C., Chevrier, B., Podjarny, A., Moras, D., deBear, J.S., Gough, G.R., Gilham, P.T. and Johnson, J.E. (1988). *Nature*, **335**, 375-378.

4. Schindelin, H., Zhang, M., Bald, R., Fürste, J.-P., Erdmann, V.A. and Heinemann, U. (1995). *J. Mol. Biol.*, **249**, 595-603.

5. Portmann, S., Usman, N. and Egli, M. (1995). *Biochemistry*, **34**, 7569-2684.

6. Holbrook, S.R., Cheong, C., Tinoco, I. Jr. and Kim, S.-H. (1991). *Nature*, **353**, 579-581.

7. Cruse, W.B.T., Saludjian, P., Biala, E., Strazewski, P., Prangé, T. and Kennard, O. (1994). *Proc. Natl. Acad. Sci. USA*, **91**, 4160-4164.

8. Leonard, G.A., McAuley-Hecht, K.E., Ebel, S., Lough, D.M., Brown, T. and Hunter, W.N. (1994). *Structure*, **2**, 483-494.

9. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. and Doudna, J.A. (1996). *Science*, **273**, 1678-1685.

10. Rould, M.A., Perona, J.J. and Steitz, T.A. (1991). *Nature*, **352**, 213-218.

11. Valegard, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. and Liljas, L. (1994). *Nature*, **371**, 623-626.

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

13. Scott, W.G., Finch, J.T., Grenfell, R., Fogg, J., Smith, T., Gait, M.J. and Klug, A. (1995). *J. Mol. Biol.*, **250**, 327-332.

14. Scott, W.G., Finch, J.T. and Klug, A. (1995). *Cell*, **81**, 991-1002.

15. Pley, H.W., Lindes, D.S., DeLuca-Flaherty, C. and McKay, D.B. (1993). *J. Biol. Chem.*, **268**, 19656-19658.

16. Pley, H.W., Flaherty, K.M. and McKay, D.B. (1994). *Nature*, **372**, 68-74.

17. Doudna, J.A., Grosshans, C., Gooding, A. and Kundrot, C.E. (1993). *Proc. Natl. Acad. Sci. USA*, **90**, 7829-7833.

18. Barlow, T. (1996). *Ph.D. Thesis*, University of Edinburgh, Scotland.

19. Fasman, G.D. (Ed.) (1975). *Handbook of Biochemistry and Molecular Biology. Nucleic Acids Vol. I*, 3rd Edition, CRC Press.

20. Puglisi, J.D. and Tinoco, I. Jr. (1989).*Meth. Enzymol.*, **180**, 304-315.

21. Brown, T. Hunter, W.N., Kneale, G. and Kennard, O. (1986). *Proc. Natl. Acad. Sci. USA*, **83**, 2402-2406.

22. Coll, M., Fredrick, C.A., Wang, A.H.-J. and Rich, A. (1987). *Proc. Natl. Acad. Sci. USA*, **84**, 8385-8389.

23. Wang, A.H.-J., Quigley, G.J., Kolpack, R.J., Crawford, J.L., van Boom, J.H., van der Marel, G.A. and Rich, A. (1979). *Nature*, **282**, 680-686.

24. Ramesh. V. (1993). *Nucleic Acids Res.* **21**, 5485-5488.

Chapter 4

Sequence Dependent Structures in RNA

4. Sequence Dependent Structure in RNA

4.1 Introduction

The three dimensional structure of DNA is known to depend subtly on its sequence [1]. This property, combined with the inherent deformability of some sequences, is crucial in biological function such as the regulation of gene expression [2]. One area of special interest has been the sequence dependent structures that are formed in A-T tracts in DNA. These regions exhibit a number of unusual structural properties and have been implicated in a variety of biological roles:

- in $d(A_nT_n)$ tracts a highly ordered 'spine of hydration' is found in the minor groove (where n=2, or more) [3]. High propeller twists maximise base stacking and allow the formation of three centre hydrogen bonds down the major groove [4,5]. These factors make A-T tracts quite stiff and thermally extra stable. No ordered spine of hydration is observed for the opposite orientation of the base pairs, $d(T_nA_n)$ [6].
- base pairs in $d(A_nT_n)$ tracts are considerably less susceptible to chain cleavage by DNase I [7] and hydroxyl radicals [8]. The narrowing of the minor groove has been implicated in this decrease in activity, but equally the low flexibility in these regions may be significant.
- the sequence $d(GAATTC)_2$ is the specific recognition site for the endonuclease *Eco* RI. The *Eco* RI-DNA complex crystal structure shows contacts with the bases in the major groove only and a severely distorted DNA conformation [9].
- A-T tracts are specific sites for a variety of drugs that bind in the minor groove such as distamycin A [10,11], netropsin [12], Hoechst 33258 [13] and bisamidinium compounds [14].

RNAs exhibit a far wider range of secondary structural motifs than DNA and often have complex tertiary structures. These motifs undoubtedly offer a much richer variety of structures for recognition and function *in vivo*. Despite this, double helical regions are present in natural RNAs and can offer sites for protein recognition and/ or ligand binding. One example is ribosomal protein S8, which binds to an irregular but relatively simple double helix [15].

The question addressed here is whether RNA A-U tracts exhibit similar groove hydration, increased thermodynamic stability or conformational anomalies that might facilitate recognition or ligand binding. The problem of groove hydration can be examined by NMR and is discussed in more detail later in this chapter. First, however, the thermodynamic stability of the RNA duplex r(CGCAAAUUUGCG)$_2$ was compared to that of the well characterised DNA duplex of the same sequence. As discussed in the next section, these were found to be of almost identical thermal stability. We then wondered if this was solely due to the extra stability of the DNA A-T tract or whether the RNA was unusually unstable. To answer this question a series of RNA and DNA dodecamer duplexes with different arrangements of the A-U and A-T base pairs, but the same overall composition, was prepared:

RNA	r(CGCAAAUUUGCG)$_2$	R06 rA_3U_3
Duplexes	r(CGCAAATTTGCG)$_2$	R18 rA_3T_3
	r(CGCAUAUAUGCG)$_2$	R10 $r(AU)_3$
	r(CGCUUUAAAGCG)$_2$	R11 rU_3A_3
	r(GACUGAUCAGUC)$_2$	R12 Rmix
DNA	d(CGCAAATTTGCG)$_2$	D06 dA_3T_3
Duplexes	d(CGCATATATGCG)$_2$	D10 $d(AT)_3$
	d(CGCTTTAAAGCG)$_2$	D11 dT_3A_3
	d(GACTGATCAGTC)$_2$	D12 Dmix

4.2 UV Thermal Melting

4.2.1 Duplex Stability

Duplex melting temperatures (Tm) were measured in 10mM phosphate buffer, pH7.0, containing 1.0M NaCl, with a heating rate of 1° per minute. The high salt concentration was used to stabilise the double stranded helix over the alternative hairpin structure which can be formed with self-complementary sequences. Monophasic melting curves were obtained, and the formation of duplexes demonstrated by the dependence of the melting temperature on the concentration of the strands. A small degree of hysteresis was noted for both DNA and RNA duplex melting curves. In general, Tm values for melting were 2°C higher than for annealing of the strands with the same sample. However, the Tm values determined for either melting or annealing were found to be *independent of heating rate,* indicating that correct equilibrium measurements were being made [16]. For consistency, Tm values measured for melts were used in the calculation of thermodynamic parameters for all duplexes.

The thermodynamic parameters for each duplex were determined from the concentration-dependence of the melting temperatures, over a 50-fold or greater duplex concentration range, using the relationship:

$$\frac{1}{Tm} = \frac{R}{\Delta H^\circ} \ln Ct + \frac{\Delta S^\circ}{\Delta H^\circ}$$

where Ct is the total strand concentration, R the gas constant, ΔH° the enthalpy and ΔS° the entropy of melting (see section 1.5.3). The standard state was taken to be at 1μM strands and 25°C. The thermodynamic

parameters determined from van't Hoff plots of 1/Tm versus lnCt (Figure 4.1) are given in Table 4.1. The full Tm data can be found in Appendix C.

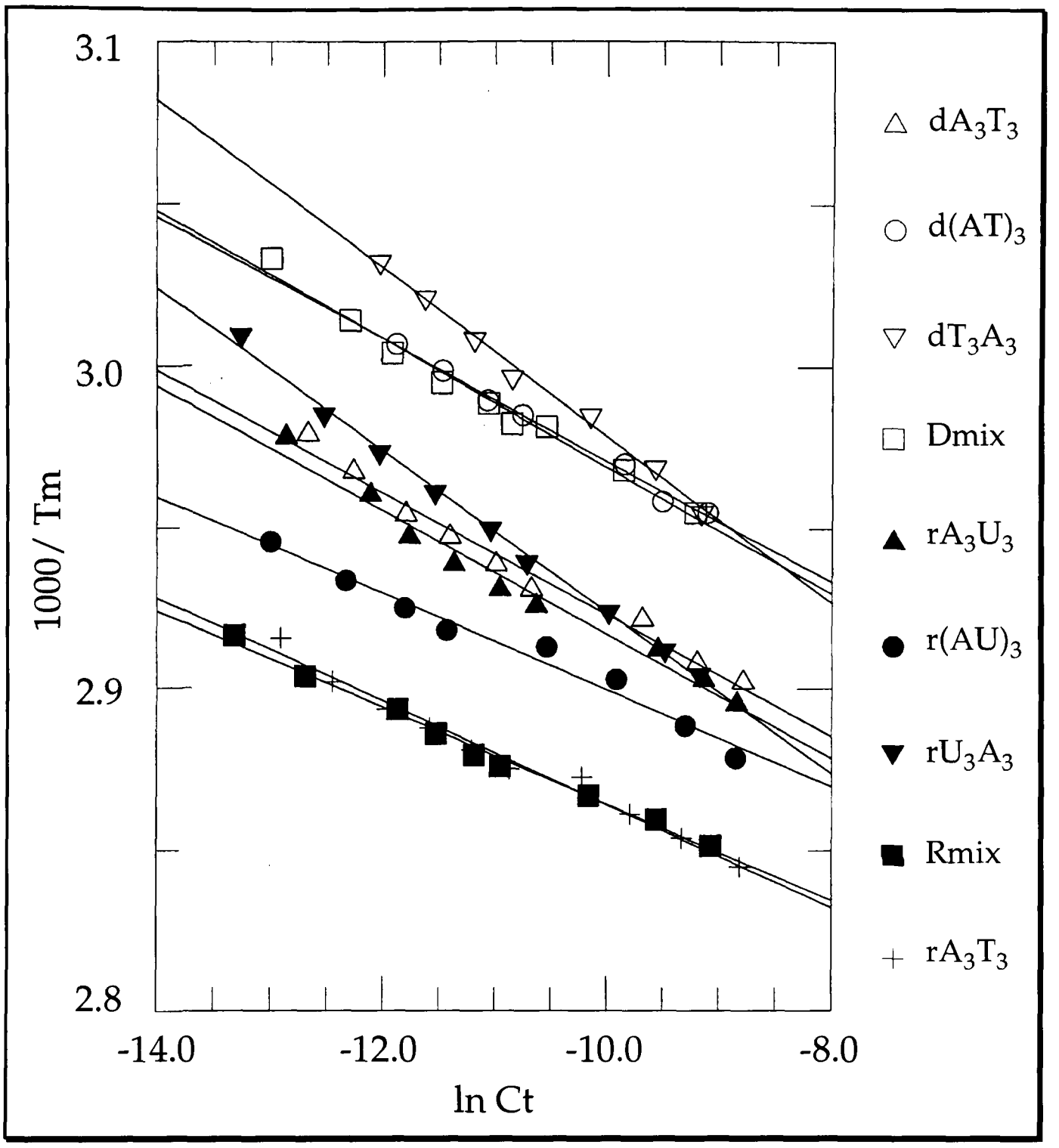


Figure 4.1 van't Hoff plots (1000/Tm versus lnCt) for the DNA and RNA dodecamers.

Errors in the values of lnCt are relatively small (1% or less) using the observed molar extinction coefficients (about 5% error) and Tm values are reproducible to within +/- 0.5°. While this reproducibility does not necessarily give a fair reflection of the absolute accuracy of the

measurements (as, for example, the temperature of the cell was measured rather than the solution directly), any other errors involved are the same in all Tm values. Thus the results for the different duplexes in this study can be compared relatively accurately, but care must be taken in comparisons with measurements made elsewhere. The errors in the calculated thermodynamic parameters ΔH°, ΔS° and ΔG° were determined from linear fits to the data points in the van't Hoff plots. Thus differences in duplex stability are only considered outwith these errors, which correspond to around 1-3 $kJmol^{-1}$ in ΔG°.

DNA Duplex	$-\Delta H^\circ$ ($kJmol^{-1}$)	$-\Delta S^\circ$ ($Jmol^{-1}K^{-1}$)	$-\Delta G^\circ_{298K}$ ($kJmol^{-1}$)	Tm, K (1μM)
Dmix	**422 ±20** (341)	**1170 ±59** (938)	**73.4 ±2.3** (61.5)	**328** (324)
A_3T_3	**440 ±23** (407)	**1204 ±67** (1117)	**81.6 ±3.0** (74.1)	**334** (334)
$(AT)_3$	**443 ±16** (374)	**1233 ±49** (1034)	**74.6 ±1.8** (65.9)	**329** (325)
T_3A_3	**321 ±11** (396)	**873 ±34** (1089)	**60.9 ±1.3** (71.4)	**325** (328)

RNA Duplex	$-\Delta H^\circ$ ($kJmol^{-1}$)	$-\Delta S^\circ$ ($Jmol^{-1}K^{-1}$)	$-\Delta G^\circ_{298K}$ ($kJmol^{-1}$)	Tm, K (1μM)
Rmix	**558 ±19** (483)	**1510 ±54** (1323)	**108 ±2.9** (88.7)	**343** (336)
A_3U_3	**433 ±25** (408)	**1179 ±73** (1111)	**81.5 ±3.3** (76.9)	**334** (333)
$(AU)_3$	**556 ±33** (413)	**1528 ±98** (1121)	**100 ±4.5** (78.6)	**338** (334)
U_3A_3	**332 ±7** (394)	**889 ±20** (1069)	**67.5 ±1.0** (75.4)	**332** (333)
A_3T_3	**522 ±24** (-)	**1410 ±70** (-)	**101 ±3.0** (-)	**342** (-)

Table 4.1 Thermodynamic parameters for DNA and RNA dodecamers (values in parentheses were calculated from nearest neighbour parameters as described in the text; no parameters are available for rT).

Despite the identical base pair composition, for both the DNA and RNA dodecamers, there is a clear ordering of the experimental thermodynamic stabilities in terms of ΔG° and Tm:

$$\text{RNA} \quad r[Rmix] > r[A_3T_3] \sim r[(AU)_3] > r[A_3U_3] > r[U_3A_3]$$
$$\text{DNA} \quad d[A_3T_3] > d[(AT)_3] \sim d[Dmix] > d[T_3A_3]$$

In general, RNA duplexes are expected to be more stable than their DNA counterparts [17]. For all the dodecamer duplexes studied here this is true, except for A_3U_3. In the RNA series, the mixed sequence has higher Tm and more negative ΔG° than all the other duplexes. Alternating A-U base pairs are slightly destabilising in ΔG° (+8 $kJmol^{-1}$) and Tm (6°). In contrast for DNA, this arrangement has no effect on duplex stability. For RNA and DNA the arrangements U_3A_3 and T_3A_3 are highly destabilising, giving the lowest Tm and largest $\Delta\Delta G^\circ$ compared to the randomized sequences (+40.5 and +12.5 $kJmol^{-1}$ respectively).

The extra stability of the DNA duplex A_3T_3 can be attributed to the stable spine of hydration, bifurcated hydrogen bonding and improved base stacking possible in A_nT_n tracts. In contrast, for RNA the A_3U_3 duplex is considerably destabilised relative to the randomised sequence (+26.7 $kJmol^{-1}$ in ΔG°). Both of these effects making the overall stability of the duplexes the same. In both cases, changes are seen in the enthalpy component, with an increase for DNA (-18.4 $kJmol^{-1}$) and a substantial decrease for RNA (+125 $kJmol^{-1}$). This is consistent with large differences in base stacking in these duplexes [18] and suggests that the rigid structures involving large propeller twists and bifurcated hydrogen bonds present in DNA are not formed in RNA. These differences are also manifest in the groove hydration, as no stable spine of hydration is observed in the RNA (section 4.3).

The addition of methyl groups to the RNA duplex, rA_3T_3, increases the stability dramatically, such that it is almost at the same level as the mixed sequence RNA. The addition of the six methyl groups increases the Tm (at 1μM) by 8°C and decreases ΔG° by -18.4 $kJmol^{-1}$. This corresponds to a stabilisation of -3 $kJmol^{-1}$ per methyl group. It is thought that methyl groups increase the base stacking proficiency in the duplex and that this makes a major contribution to the increase in stability [19]. The large increase in the enthalpy component of ΔG° (-89 $kJmol^{-1}$) is indeed consistent with much improved base stacking. Clearly, in the context of this destabilised RNA duplex the addition of methyl groups has a substantial effect on the base stacking interactions available. This could arise from the change in polarisability that the methyl group imparts on the base [20], allowing for more favourable van der Waals interactions with the neighbouring bases. Alternatively, there could be a change in conformation induced by the presence of the methyl groups which allows better stacking interactions.

DNA duplexes with the opposite ordering of the central bases have structures quite distinct from A_nT_n tracts in solution [21] and in the solid state [22]. Widening of the minor groove around a central T_2A_2 segment was observed in both cases, and this has been shown to disrupt the ordered spine of hydration [6]. Furthermore, TpA steps are very flexible in this context compared to the A_nT_n orientation [23]. The thermodynamic analysis of the DNA T_3A_3 duplex shows a very low enthalpy component, consistent with poor base stacking. This is most likely to be as a direct result of the changes in conformation and hydration in this orientation. The RNA duplex U_3A_3 also shows a very low enthalpy of helix formation. In fact, it is considerably less favourable compared to the randomised sequence than for DNA. No stable spine of hydration is observed in RNA therefore the low stability can be attributed to a considerable reduction in base stacking interactions and possibly less favourable backbone

conformations. A complete understanding of the reasons behind the low stability of this duplex will require a full structural analysis.

Duplex stability can be estimated using nearest neighbour parameters for both RNA and DNA duplexes [24,25]. The thermodynamic parameters calculated for the RNA and DNA dodecamers are also shown in Table 4.1. For most of the duplexes, reasonable agreement of Tm was observed, but the calculated ΔG° and ΔH° were less similar to the experimentally determined values. However, for the RNA series the order of stability was predicted correctly. For DNA, the A_3T_3 duplex was predicted to be the most stable, as observed. The duplex with alternating A-T base pairs was predicted to be slightly more stable than the randomised duplex than was observed experimentally. The most significant difference in the observed and predicted values was for the DNA duplex T_3A_3 (10.4 $kJmol^{-1}$ in ΔG°). This duplex was clearly the least stable but was predicted to be almost as stable as the alternative A_3T_3 orientation. This suggests that the nearest neighbour model is not sufficient for this duplex.

While many of the trends in the data were reproduced, the quantitative agreement was not good. Most disappointing was that the predictions for the mixed sequence duplexes were so poor. In fact, closest predictions were obtained for the A_3T_3 and A_3U_3 duplexes (particularly the Tm values). This is somewhat surprising, especially for the DNA duplex. The stabilising structures observed require longer A-T tracts than one would expect to be accounted for in the nearest neighbour model. In almost all cases, the values predicted were lower than observed experimentally. The only exception to this was for the U_3A_3 and T_3A_3 duplexes, which were predicted to be more stable than observed. Clearly, for these duplexes the nearest neighbour model is not appropriate for predicting their thermal stability. This is most likely to be as a result of co-operative changes in conformation, and hence base stacking interactions, throughout the

duplex, caused by the central T_3A_3 and U_3A_3 regions. Such long ranging effects cannot, implicitly, be incorporated into the nearest neighbour data.

4.2.2 Effect of Salt Concentration on Duplex Tm

Ions or small molecules that bind preferentially to either the duplex or single strand oligonucleotides will alter the Tm. Salt binds preferentially to the state with the greatest charge density, which for nucleic acids is the double stranded helix. The equilibrium between double and single strands can then be written:

$$M_n{\cdot}D \rightleftharpoons M_{n'}{\cdot}S_1 + M_{n''}{\cdot}S_2 + \Delta nM$$

where M is the ion (or small molecule), D is the double helix and S_1 and S_2 the single strands. Upon melting of the duplex, $\Delta n = n - n' - n''$ molecules of M are released. From a consideration of the equilibrium constant for helix formation, the effect of ligand concentration can be determined using the van't Hoff relationship [26], giving the simple result:

$$\frac{d\mathrm{Tm}}{d[\ln(\mathrm{M})]} = -\Delta n \frac{\mathrm{RTm}^2}{\Delta \mathrm{H}^0}$$

ΔH^0 is the enthalpy of helix formation at Tm and is large and negative. Thus, raising the concentration of salt or small molecule will increase the Tm, if that ion or molecule binds in greater numbers to the double helix than to the single strands. If the small molecule binds preferentially to the single strands then a decrease in Tm would be anticipated. Experimentally, up to 1M the Tm is found to increase with increasing salt concentration [16], and a plot of Tm against the logarithm of salt

concentration is a straight line [26]. At higher concentrations, the effect diminishes and Tm may begin to decrease. Similar arguments can be used for the differential binding, for example of magnesium ions, to double and triple helices.

This simple theory is sufficient to explain the qualitative effects of ionic strength, but fails in quantitative terms. A more rigorous treatment requires the application of polyelectrolyte theory [27,28]. Two separate phenomena are then considered: the electrostatic shielding by the counterion atmosphere and the direct binding of counter ions onto the nucleic acid. The final result predicts similar behaviour as the simple model above, but with a much reduced sensitivity to salt concentration. This makes sense intuitively: the affinity of the salt for the duplex will decrease as ions bind since the net charge is decreased by electrostatic shielding.

Dodecamer duplex melting temperatures were measured at NaCl concentrations between 0.1 and 1.0M, with a fixed duplex concentration. As expected, all the DNA and RNA dodecamers showed a strong dependence of Tm on the concentration of sodium chloride (Table 4.2). Plots of Tm against log [Na^+] gave straight lines for the RNA duplexes and DNA A_3T_3 and mixed sequences (Figure 4.2). For the DNA duplexes $(AT)_3$ and T_3A_3 the transitions were not two-state below 0.5M salt. At low salt concentrations, a significant lower temperature transition was observed, corresponding to the melting of a single stranded hairpin structure. For T_3A_3 at 0.1M salt the low temperature transition was dominant (with a Tm of approximately 310K).

For the RNA dodecamers, $d\mathrm{Tm}/d\log\,[Na^+]$ was in the range 9.0 to 14.0 (mean 11.9), whereas the slopes for mixed sequence DNA and A_3T_3 were

NaCl conc. (M)	A_3U_3	$(AU)_3$	U_3A_3	Rmix
0.10	329.1	330.8	nd	337.6
0.25	333.9	336.0	330.9	341.4
0.50	337.2	340.6	335.4	344.6
1.0	340.3	343.3	339.0	346.8
dTm/$d$log[Na^+]	11.0	14.0	13.5	9.0

NaCl conc. (M)	A_3T_3	$(AT)_3$	T_3A_3	Dmix
0.10	330.3	nd	~310	325.2
0.25	334.5	~323	nd	329.4
0.50	337.2	331.1	328.1	332.2
1.0	339.3	333.0	331.2	333.8
dTm/$d$log[Na^+]	9.0	~6	~10	8.6

Table 4.2 Tm (in K) over a range of salt concentrations for RNA (top) and DNA duplexes (bottom) [nd: not a duplex].

smaller (8.6 and 9.0 respectively). The slight difference in salt sensitivity between RNA and DNA was more obvious when the effect of magnesium was considered (section 4.2.3). For the RNA dodecamers, the least sensitive to salt concentration was the most stable mixed sequence. The lower sensitivity of this RNA could be due to a greater relative affinity of the salt for the single stranded state. However, it is also consistent with considerable differences in the global conformations of these duplexes, as discussed earlier.

4.2.3 Effect of Metal Ions on Duplex Tm

Bulk electrolytes, such as NaCl or KCl, stabilise nucleic acid duplexes through the neutralising of the surface charge due to the phosphate

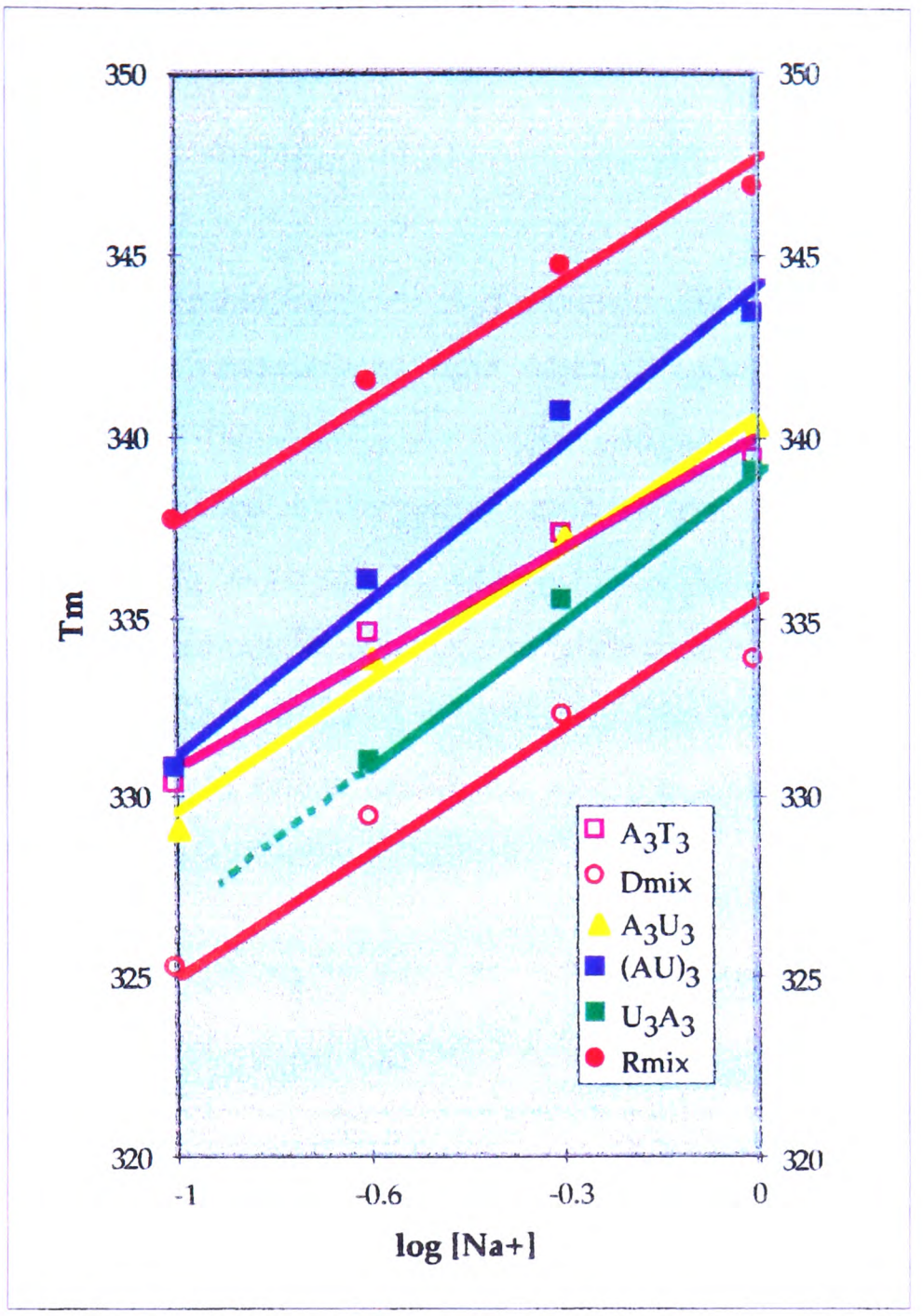


Figure 4.2 Tm versus log [Na^+]: the dependence of Tm on salt concentration for RNA and DNA dodecamer duplexes.

groups. For simple duplexes or hairpins, a similar mode of action can be envisaged for magnesium [29]. In contrast for more complex folded RNAs with areas of high charge density, magnesium and other divalent ions tend to bind more specifically and with much higher affinity. Magnesium ions have been shown to be crucial in RNA catalysis and to rRNA folding, through binding to a specific site by direct inner sphere coordination [29]. Magnesium ions are relatively compact with a high

charge density, making it possible for them to bridge phosphate groups. The preference for oxygen ligands means that there is extensive bonding via the waters of hydration along the back bone, usually in an octahedral environment [30].

Duplex melting temperatures were measured for each of the dodecamer duplexes in buffers containing 1mM Mg^{2+}, 10mM Mg^{2+} and 1mM Ca^{2+}, with both 0.1 and 1M NaCl. At high concentrations of metal ions, deprotonation can lead to aggregation and precipitation of insoluble hydrous oxides [31]. Hence, the effect of calcium ions could only be examined at 1mM concentration. In the high salt buffer no increase in Tm was observed for DNA or RNA duplexes with the added divalent ions. This is consistent with there being no *specific* coordination sites within the duplexes for magnesium or calcium.

RNA Duplex	Tm (K)	Tm (K) (+1mM Mg^{2+})	Tm (K) (+10mM Mg^{2+})	Tm (K) (+1mM Ca^{2+})
A_3U_3	329.1	330.1	337.1	329.0
$(AU)_3$	330.8	332.0	341.0	332.2
U_3A_3	nd	nd	337.6	nd
Rmix	337.6	338.2	345.5	338.2

DNA Duplex	Tm (K)	Tm (K) (+1mM Mg^{2+})	Tm (K) (+10mM Mg^{2+})	Tm (K) (+1mM Ca^{2+})
A_3T_3	330.3	330.9	334.5	330.4
$(AT)_3$	nd	nd	328.0	nd
T_3A_3	~310	~310	323.7	~310
Dmix	325.2	325.3	328.1	325.2

Table 4.3 Duplex Tm in 0.1M NaCl with divalent ions: RNA (top) and DNA (bottom).

To examine the effects of divalent ions, low concentrations of NaCl are often used to avoid competition with binding [16]. The Tm data in lower salt concentration does show considerable differences between RNA and DNA (Table 4.3) not apparent previously. In the low salt buffer, many of the duplexes exist as either predominantly hairpin structures or a mixture of hairpin and duplex. Tm values are given where one major transition was observed in the melting curve, corresponding to a clear single maximum in the derivative. The approximate values given in some instances are for the low temperature hairpin to random coil and 'nd' denotes a mixture of states from which no single Tm could be obtained.

All the RNAs existed predominantly as duplexes in 0.1M NaCl, except for U_3A_3. However, the addition of 10mM Mg^{2+} to the buffer sufficiently stabilised the duplex relative to the hairpin structure that a single maximum corresponding to the duplex to single strand transition was observed. As noted previously, the DNA duplexes $(AT)_3$ and T_3A_3 were considerably less stable in the lower salt buffer. The duplex $(AT)_3$ exists as a mixture of hairpin and duplex under conditions of low salt and low divalent ion concentration, but predominantly duplex at high divalent ion concentration. For T_3A_3, the main transition observed corresponded to the hairpin to random coil transition, until 10mM Mg^{2+} was present in the buffer. No difference in stability was observed for any duplex between 1mM Mg^{2+} and 1mM Ca^{2+} buffer. Again, this is consistent with purely electrostatic shielding rather than direct binding. The results are not conclusive however, as the Tm differences on adding divalent ion at 1mM are very small.

Where accurate Tm differences are available, it is clear that the RNA duplexes are stabilised to a greater extent than the equivalent DNA duplexes. In 1mM Mg^{2+} and Ca^{2+}, RNA duplex Tm values increase by about 1°. No change is observed for the DNA. In 10mM Mg^{2+}, the increase

in Tm for RNA was 8-11°, but only 3-4° for DNA (Table 4.4). Comparison with the salt range Tm data (Table 4.2) shows that 10mM Mg^{2+} ions are as effective as 500 mM Na^+ in stabilising the RNA duplexes. In contrast for DNA, the stabilisation is equivalent to only 100-150 mM Na^+. However, 10mM Mg^{2+} greatly stabilises the DNA duplexes $(AT)_3$ and T_3A_3 relative to the hairpin structures (Table 4.3), whereas the additional 150 mM Na^+ does not (Table 4.2). The major differences between RNA and DNA in terms of counterion binding must result from the presence of the 2'-OH and the overall conformational differences between the A- and B-forms. The results suggest that coordination of the metal ions to the 2'-OH, either directly or via waters of hydration, could play a role in stabilising the duplexes. The difference in counterion sensitivity probably also reflects the conformational difference between the two duplex forms where the linear charge density is quite different.

RNA Duplex	ΔTm (°)	DNA Duplex	ΔTm (°)
A_3U_3	8.0	A_3T_3	4.2
$(AU)_3$	10.2	$(AT)_3$	-
U_3A_3	-	T_3A_3	-
Rmix	7.9	Dmix	2.9

Table 4.4 ΔTm for RNA and DNA duplexes in 0.1M NaCl with and without 10mM Mg^{2+}

4.2.4 Drug Binding

A considerable number of minor groove binding drugs, such as distamycin A, netropsin, Hoechst 33258 or berenil, show a marked preference for A-T tracts in DNA duplexes [32]. While many of these drugs are too toxic to be of any pharmacological use, for example in chemotherapy, the synthesis of analogues with more favourable toxicity

but retaining DNA-binding specificity has raised much interest [33]. The binding of these drugs is very sensitive to the shape of the minor groove, which is very narrow in A-T tracts. The drugs are, superficially, very similar crescent shaped molecules (Figure 4.3) and it is this shape that is important in groove binding. For molecules such as distamycin A, two modes of binding are possible: either one drug molecule, or two side by side with the charged groups at either end. Both modes of binding have been studied by NMR, but generally only the 1:1 binding mode has been observed by X-ray crystallography [34]. Drugs with charged groups at each end only bind in a 1:1 mode. While the site preference of each drug is similar, the range of direct (hydrogen bonding) interactions observed is diverse. Analysis of structures involving this family of drugs has shown that the sequence recognition is largely due to the geometric characteristics of the minor groove rather than being driven by hydrogen bonding [35]. In essence, each of these molecules has the pre-defined shape to 'fit' into the binding site. Once there, they take up what hydrogen bonding sites are available. Indeed, the specificity for the shape of the minor groove is such that binding of netropsin and distamycin A can reverse B- to A- and B- to Z- transitions induced by solution conditions [36].

The preferential binding of drug molecules to the duplex state would be expected to give an increase in Tm (as discussed earlier for the counterion effects). An increase in the entropy component is expected, due to the disruption of the ordered spine of hydration found in A_nT_n tracts. Enthalpically, favourable binding will include hydrophobic and hydrogen bonding interactions in the groove and electrostatic interaction between positively charged drug molecules and the phosphates of the backbone [37].

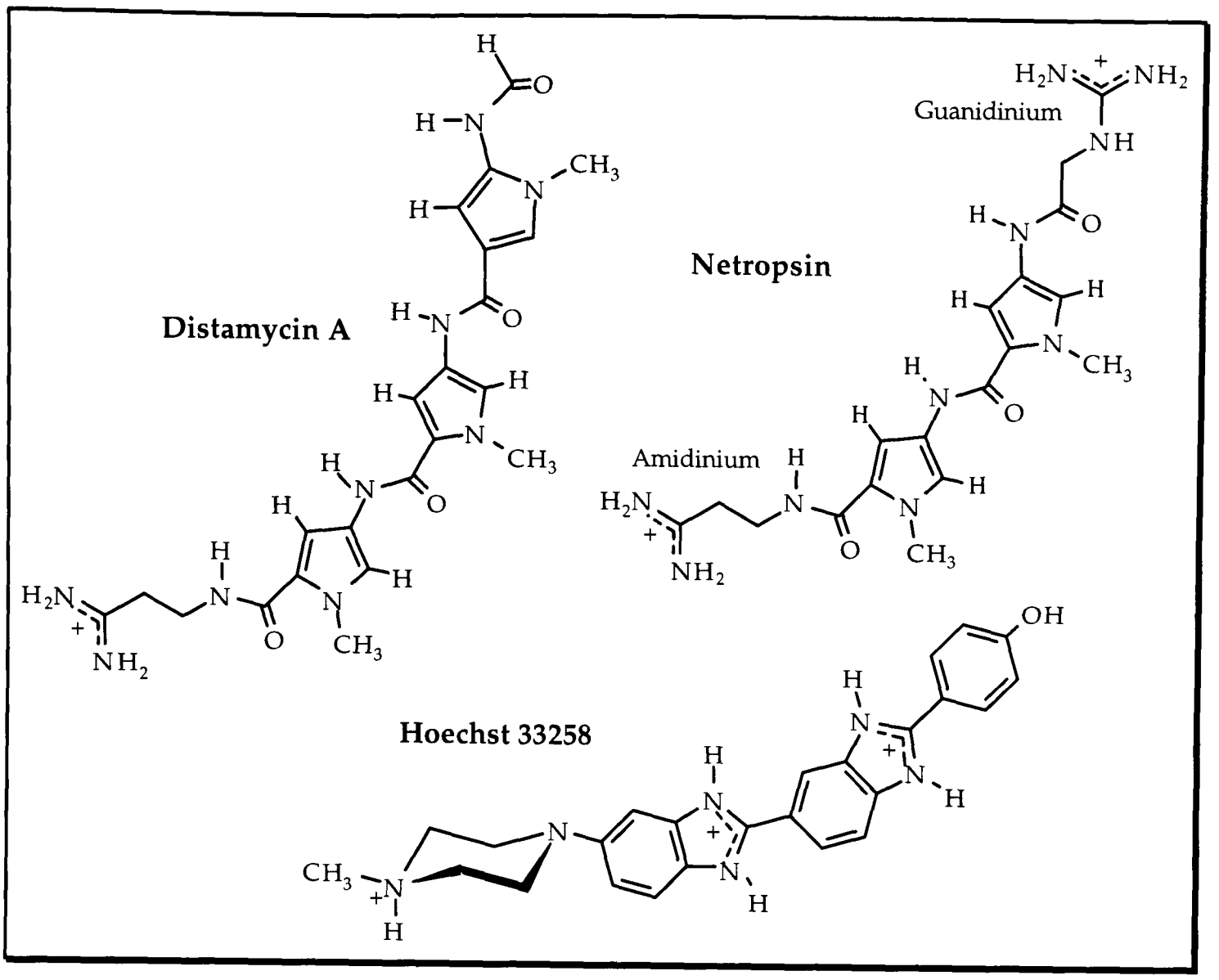


Figure 4.3 Structures of some DNA minor groove binding drugs.

The effects of groove binding drugs distamycin A and netropsin, the intercalator daunomycin, and spermine on the Tm of the RNA duplex A_3U_3 were examined in 1.0M NaCl. Under these conditions, none of the ligands caused an increase in Tm. In contrast, the Tm of the DNA duplex A_3T_3 with distamycin A increased dramatically (ΔTm=12.4°). As weak electrostatic interactions could be masked by the high salt concentration, the melting temperature of each DNA and RNA duplex was obtained in 0.1M NaCl with 10μM distamycin A. A clear stabilisation was observed for all DNA duplexes (Table 4.5). The preference for the A-T tract is demonstrated by the difference in ΔTm between A_3T_3 and Dmix (15.3 and 7.6° respectively). For the DNA T_3A_3 and $(AT)_3$ the addition of the drug greatly stabilised duplex relative to hairpin structures. No stabilisation of

any RNA was observed, indicating that these drugs have little or no affinity for RNA duplexes. This, again, reflects the considerable overall conformational differences between RNA and DNA helices.

Duplex	Tm (K) No Drug	Tm (K) + Distamycin A	ΔTm
A_3T_3	330.1	345.4	15.3
$(AT)_3$	nd	339.4	-
T_3A_3	nd	338.7	-
Dmix	324.9	332.5	7.6
A_3U_3	328.6	328.4	0
$(AU)_3$	331.1	331.1	0
U_3A_3	nd	nd	0
Rmix	337.0	336.7	0
A_3T_3	336.2	336.0	0

Table 4.5 Tm for RNA and DNA duplexes with distamycin A in 0.1M NaCl (nd=not duplex)

4.3 Hydration of the Duplex r(CGCAAAUUUGCG)$_2$ in Solution

Nucleic acids are strongly hydrated in both the crystal state and in solution. For DNA, the degree of hydration determines the overall helix conformation: B-form at high water activity and A-form at low water activity. Hydration patterns in many crystal structures of DNA and RNA have been reported [38]. The relatively recent success in NMR detection of water molecules on the surface and in the interior of proteins in solution [39,40] has led to similar studies in DNA. The ordered spine of hydration first seen in crystal structures of dA_nT_n tracts have been observed in solution by NMR [41,42]. Water residence times have been noted to be similar to that in the interior of proteins, at about 1ns or longer. To date, no NMR information is available on RNA hydration in solution.

We have undertaken the study of hydration in the RNA duplex r(CGCAAAUUUGCG)$_2$ by NMR. The results can be directly compared to the well characterised DNA analogue. The acquisition of spectra and assignment of resonances has been described in detail previously [43]. Water molecules can be considered to be bound if effective correlation times significantly longer than 0.5 ns are observed. Both grooves of the RNA duplex were seen to be in contact with water, as found for the DNA duplex. Water molecules with relatively long residence times were found close to H1′ despite the wide, shallow minor groove. At the adenine H2 and H8/H6 only small effective correlation times were observed, consistent with only transiently bound water.

In the DNA duplex, the ordered spine of hydration found in A_nT_n tracts is thought to be responsible for the relatively slow exchange of NMR visible water molecules from around the adenine H2 [6,41]. The observation of slow exchange in the narrow minor groove of these regions but not the wider major groove supports the hypothesis that groove width may be

the determining factor. In a wide groove, waters of hydration will have greater freedom and accordingly short residence times. In the narrow minor groove, with less freedom, longer residence times are expected and are observed. However, RNA has a wide, shallow minor groove and so the groove width hypothesis predicts that there should be no slowly exchanging water molecules. Indeed, no slow exchanging water around the adenine H2 was observed. However, other factors may be important. For example, the minor groove of RNA is lined with 2′-OH groups and is therefore more hydrophilic than the minor groove of DNA. Experimentally, kinetically bound waters of hydration were observed in the RNA minor groove that were not seen in the DNA. Our results show that it is possible for water protons to be sufficiently close to H1′ for a direct NOE to be observed and hydrogen bonding to the nearby 2′-OH is most likely to be responsible for the relatively long residence time. In contrast for DNA, there is no nearby 2′-OH to stabilise the water molecule and under similar conditions, no NOE between water and H1′ was observed [44]. This observation is also consistent with an X-ray study of an RNA duplex where water molecules interacting directly with the 2′-OH were observed [45].

The fully exposed groups in the grooves interact with water molecules only transiently, by virtue of the high concentration of water. The results here suggest that slowly exchanging water molecules may interact predominantly by hydrogen bonding interactions. In DNA A_nT_n tracts, the high propeller twists allow bridging interactions of waters between base pairs, while in RNA, interactions at the 2′-OH may play a crucial role. Slow exchange kinetics may therefore, in general, be determined by chemical bonding rather than conformational factors.

4.4 Molecular Structure of r(CGCAAAUUUGCG)$_2$

The RNA dodecamer r(CGCAAAUUUGCG)$_2$ was found to be considerably destabilised relative to a randomised duplex of the same length and composition (section 4.2). In order to determine the conformational reasons behind this, we have undertaken a full structural study of the duplex in the crystal and solution states.

Crystals of the duplex with a 3'-G overhang have been used to collect X-ray diffraction data to a resolution of 2.0Å (Chapter 3). The same crystallising conditions also provided crystals of the duplex with a 5-bromo uridine residue that will be used to aid structure solution. The crystal structure will provide useful comparisons with both the analogous DNA duplex and our solution state hydration and structural data.

NMR assignments have been used to examine the hydration (section 4.3) and helix conformation of the RNA A_3U_3 duplex. A number of helical parameters were found to deviate quite substantially from both the standard and energy minimised A-form helix (Table 4.6). The A_3U_3 dodecamer is characterised by a low twist, a high axial rise and a smaller inclination of the base pairs to the helical axis. This indicates a somewhat underwound duplex, with approximately 12 base pairs per turn of helix.

Parameter	Standard A-form	Energy minimised A-form	A_3U_3 dodecamer
Twist (°)	32.6	32.6	30.5
Axial rise (Å)	2.8	2.8	3.2
Inclination (°)	-16.0	-11.6	-3.7

Table 4.6 Comparison of some helical parameters for RNA A_3U_3 and standard and energy minimised A-forms

The width of the minor groove was also estimated from the C3′ to C5′ distance across the minor groove. For the standard RNA helix structure this gives a width of 12Å. In the A_3U_3 dodecamer, the width varies between 8 and 10Å, with the narrowest regions in the central A-U tract.

The high propeller twists found in DNA A_nT_n tracts, which improve base stacking interactions, were not seen in the RNA duplex. The poor base stacking in the A_3U_3 tract is consistent with the low thermal stability and thus quite different from other RNA structures [45] and even the GC base pairs of the same molecule. This indicates that the sequential stacking of A-U base pairs is enthalpically poor, leading to duplexes of low thermal stability. The situation is better for alternating A-U base pairs but considerably worse for the opposite orientation, U_3A_3 (section 4.2).

Egli *et al.* have suggested that the increased enthalpic stabilisation of RNA compared to DNA arises from an increased number of bound water molecules, despite the unfavourable entropic contribution this entails [46]. However, this largely ignores the difference in conformation between DNA and RNA. In general, the degree of base stacking is much greater in RNA than DNA and, as this is a strongly exothermic process, RNA would be expected to be enthalpically more stable. This is what is generally observed experimentally. However, in the case of the RNA duplex A_3U_3 the base stacking is less favourable than is typical for RNA, leading to a lower enthalpy and decreased net thermodynamic stability. This is despite the observation of hydration water in the minor groove.

4.5 Conclusions

The sequence dependent stability and structure in RNA duplexes, containing various orientations of base pairs in A-U tracts, has been confirmed by UV melting and NMR. Major differences were also observed between the RNA duplexes and their DNA equivalents.

A clear pattern of thermodynamic stability in the RNA duplexes was observed: Rmix > rA_3T_3 ~ $r(AU)_3$ > rA_3U_3 > rU_3A_3. The stabilities of the RNA duplex A_3U_3 and DNA duplex A_3T_3 were found to be the same. This can be attributed to a combination of relative destabilisation in the RNA and stabilisation in the DNA. Marked differences were observed in the sensitivity of the RNA and DNA duplexes to salt and magnesium ion concentration. While no evidence was found for specific binding in the RNA, these duplexes were always stabilised to a greater degree than their DNA counterparts. The differences can be attributed to conformational differences between DNA and RNA and also the presence of the 2′-OH.

The RNA duplex $r(CGCAAAUUUGCG)_2$ was studied in greater detail by NMR. The structure was found to deviate considerably from that expected for A-form RNA with a substantial decrease in base stacking in an underwound central region. This offered a suitable explanation for the low thermal stability observed. The hydration pattern in this duplex was also studied by NMR and compared to the DNA analogue. The ordered spine of hydration which stabilises the DNA duplex was not observed. However, kinetically bound water molecules were observed near H1′, where stabilisation via the 2′-OH is possible. These waters of hydration have not been observed in DNA. These results indicate that slow exchange kinetics in nucleic acid hydration may be driven by chemical bonding rather than conformation.

4.6 References

1. Hunter, C.A. (1996). *BioEssays*, **18**, 157-162.

2. Juo, Z.S, Chiu, T.K., Leiberman, P.M., Baikalov, I., Berk, A.J. and Dickerson, R.E. (1996). *J. Mol. Biol.*, **261**, 239-254.

3. Drew, H.R. and Dickerson, R.E. (1981). *J. Mol. Biol.*, **151**, 535-556.

4. Nelson, H.C.M., Finch, J.T., Luisi, B.F. and Klug, A. (1987). *Nature*, **330**, 221-226.

5. Edwards, K.J., Brown, D.G., Spink, N., Skelly, J.V. and Neidle, S. (1992). *J. Mol. Biol.*, **226**, 1161-1173.

6. Liepinsh, E, Leupin, W. and Otting, G. (1994). *Nucleic Acids Res.*, **22**, 2249-2254.

7. Sutton, D., Conn, G.L., Brown, T. and Lane, A.N. (1996). *Biochemical Journal*, in press.

8. Fox, K.R. (1992). *Nuleic Acids Res.*, **20**, 6487-6493.

9. Frederick, C.A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobsen, L., Wang, B.C., Greene, P., Boyer, H.W. and Rosenberg, J.M. (1984). *Nature*, **309**, 327-330.

10. Pelton, J.G. and Wemmer, D.E. (1989). *Proc. Natl. Acad. Sci., USA*, **86**, 5723-5727.

11. Coll, M. Frederick, C.A., Wang, A.H.-J. and Rich, A. (1987). *Proc. Natl. Acad. Sci., USA*, **84**, 8385-8389.

12. Tabernero, L., Verdaguer, N., Coll, M., Fita, I., van der Marel, G.A., van Boom, J.H., Rich, A. and Aymami, J. (1993). *Biochemistry*, **32**, 8403-8410.

13. Spink, N., Brown, D.G., Skelly, J.V. and Neidle, S. (1994). *Nucleic Acids Res.*, **22**, 1607-1612.

14. Searle, M.S. (1993). Prog. NMR Spectr., **25**, 403-480.

15. Allmang, C., Mougel, M., Westhoff, E., Ehresmann, B. and Ehresmann, C. (1994). *Nucleic Acids Res.*, **22**, 3708-3714.

16. Puglisi, J.D. and Tinoco, I. Jr., (1989). *Meth. Enzymol.*, **180**, 304-325.

17. Lesnik, E.A. and Freier, S.M. (1995). *Biochemistry*, **34**, 10807-10815.

18. Searle, M.S. and Williams, D.H. (1993). *Nucleic Acids Res.*, **21**, 2051-2056.

19. Wang, S. and Kool, E.T. (1995). *Biochemistry*, **34**, 4125-4132.

20. Sowers, L.C., Shaw, B.S. and Sedwick W.D. (1987). *Biochem. Biophys. Res. Comm.*, **148**, 790-794.

21. Chuprina, V.P., Lipanov, A.A., Fedoroff, O.Y., Kim, S.-G., Kintanar, A., and Reid, B.R. (1991). *Proc. Natl. Acad. Sci. USA.*, **88**, 9087-9091.

22. Quintana, J.R., Grzeskowiak, K., Yanagi, K. and Dickerson, R.E. (1992). *J. Mol. Biol.*, **225**, 379-395.

23. Lane, A.N. and Lefèvre, J.-F. (1994). *Meth. Enzymol.*, **239**, 596-619.

24. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986). *Proc. Natl. Acad. Sci. USA*, **83**, 9373-9377.

25. SantaLucia, J., Allawi, H.T. and Seneviratne, P.A. (1996). *Biochemistry*, **35**, 3555-3562.

26. Cantor, C.R. and Schimmel, P.R. (1980). *Biophysical Chemistry, Part III.* W.H. Freeman and Company, San Francisco.

27. Record, M.T. and Andersen, C.F. (1978). *Q. Rev. Biophys.*, **11**, 103-178.

28. Manning, G.S. (1987). *Q. Rev. Biophys.*, **11**, 179-246.

29. Laing, L.G., Gluick, T.C. and Draper, D.E. (1994). *J. Mol. Biol.*, **237**, 577-587.

30. Black, C.B., Huang, H.-W. and Cowan, J.A. (1994). *Coordination Chemistry Reviews*, **135/136**, 165-202.

31. Pan, T., Long, D.M. and Uhlenbeck, O.C. (1993). In: *The RNA World*, Gestland, R.F. and Atkins, J.F (Ed.), Cold Spring Harbour Laboratory Press, New York.

32. Kennard, O. and Hunter, W.N. (1991). *Angew. Chem. Int. Ed. Eng.*, **30**, 1254-1277.

33. Animati, F., Arcamone, F.M., Conte, M.R., Felicetti, P. Galeone, A., Lombardi, P., Mayol, L., Paloma, L.G. and Rossi, C. (1995). *J. Med. Chem.*, **38**, 1140-1149.

34. Wahl, M.C. and Sundaralingham, M. (1995). *Curr. Op. Str. Biol.*, **5**, 282-295.

35. Neidle, S., Nunn, C.M., Spink, N., Jenkins, T.C., Laughton, C.A. and Wood., A.A. (1995). *Meeting Abstract - NACON III*, Sheffield, UK.

36. Burckhardt, G. Walter, A. and Zimmer, C. (1996). *J. Biomol. Str. Dyn.*, **13**, 671-676.

37. Eckstein, F. and Lilley, D. (Eds) (1987). *Nucleic Acids in Molecular Biology I*, Springer-Verlag.

38. Berman, H.M. (1994). *Curr. Op. Str. Biol.*, **4**, 345-350.

39. Otting, G., Leipinsh, E. and Wüthrich, K. (1991). *Science*, **254**, 974-980.

40. Gerothanassis, I.P., Birdsall, B., Bauer, C.J., Frenkiel, T.A. and Feeney, J. (1992). *J. Mol. Biol.*, **226**, 549-554.

41. Leipinsh, E., Otting, G. and Wüthrich, K. (1992). *Nucleic Acids Res.*, **20**, 6549-6553.

42. Kubinec, M.G. and Wemmer, D.E. (1992). *J. Amer. Chem Soc.*, **114**, 8739-8740.

43. Conte, M.R., Conn, G.L., Brown, T. and Lane, A.N. (1996). *Nucleic Acids Res.*, **24**, 3693-3699.

44. Lane, A.N., Jenkins, T.C. and Frenkiel, T.A., unpublished data.

45. Leonard, G.A., McAuley-Hecht, K.E., Ebel, S., Lough, D.M., Brown, T. and Hunter, W.N. (1994). *Structure*, **2**, 483-494.

46. Egli, M., Portmann, S. and Usman, N. (1996) . *Biochemistry*, **35**, 8489-8494.

Chapter 5

DNA.RNA Hybrids

5. DNA.RNA Hybrids

5.1 Introduction

Hybrid duplexes comprising a DNA and an RNA strand occur in a number of important biological processes such as DNA replication [1], transcription [2] and in the retroviral life cycle [3]. Such duplexes are known to have different structures to pure DNA and RNA duplexes, and these differences are crucial to their recognition and function *in vivo*. DNA.RNA hybrids are also substrates for the ubiquitous enzyme RNase H, which catalyses the hydrolysis of the RNA strand only when present in a hybrid duplex [4]. RNase H shows little sequence specificity [5], and it has been suggested that the enzyme recognises a difference in conformation between hybrid and pure RNA duplexes [6,7]. DNA.RNA hybrids are also exploited in antisense technology. An understanding of the stability and nature of interaction of DNA with RNA is crucial to the development of antisense oligonucleotides as therapeutics.

5.1.1 Biological Occurrence of DNA.RNA Hybrids

During DNA replication, both strands of the unwound duplex are copied simultaneously at the replication fork. The fact that DNA polymerases are only able to catalyse the condensation of a nucleoside triphosphate with a free 3′-OH, therefore presents a problem: how can both daughter strands be synthesised simultaneously in the 5′-3′ direction from two chains of opposing direction? The answer lies in the fact that one strand is not copied continuously but in smaller fragments which are later ligated. The synthesis of the strand in the 5′-3′ direction ('leading strand') generally precedes that of the strand appearing to grow 3′-5′ ('lagging strand'). This

difference in growth rate produces a region of single stranded DNA at the replication fork which is converted to a double strand using a short primer known as an Okazaki fragment. Replication thus proceeds in a 5′-3′ direction until it reaches the terminus of a previously synthesised Okazaki fragment. The primers for Okazaki fragments are made by a special class of enzymes called primases, which recognise specific DNA sequences in the single stranded region [1]. These primers are now known to be short sequences of RNA, such that DNA.RNA hybrids are created during replication of the lagging strand. Studies of model Okazaki fragments by X-ray crystallography [8] and NMR [9] have shown them to be conformationally distinct from pure DNA and RNA duplexes. These differences in helix conformation are likely to play a vital role in the recognition of these regions by the enzymes involved in DNA replication.

DNA.RNA hybrids are also found during the transcription of DNA into RNA. This type of hybrid is distinct from Okazaki fragments in that they consist of one strand of pure DNA and another of pure RNA. The RNA polymerase enzyme contains a subunit, σ, that allows the specific and efficient initiation of polymerisation [2]. The RNA polymerase unwinds the helix as it moves along the DNA, creating a single stranded DNA template. RNA synthesis then occurs from one DNA strand only, using normal complementary base pairing to create a DNA.RNA hybrid duplex within the enzyme complex. After the enzyme has passed a given region of template, the RNA dissociates from the DNA allowing the DNA double helix to reform. The RNA is also able to begin folding into its specific secondary and possibly tertiary structures.

A similar type of DNA.RNA hybrid is formed in the retroviral life cycle, where the RNA genome of the virus is reverse transcribed into DNA. This is done to enable the virus to integrate its genetic material into a DNA chromosome of the host cell, in order to direct the synthesis of

more viral units. Since the initial discovery of reverse transcription in the retroviral life cycle, it has become recognised as a much more widespread phenomenon in eukaryotic cells and viruses [3].

5.1.2 Antisense Therapeutics

In recent years, a growing understanding of disease at a molecular level has allowed the development of a rational approach to drug design. The combination of increasing identification of specific gene products involved in disease and the failure of conventional therapeutic agents to combat these diseases has led to the search for alternative forms of therapy. Endogenous antisense mechanisms are known to exist in the control of gene expression, and it is possible they will be quite widely found [10]. Synthetic oligonucleotides may provide a potential new class of therapeutics based on this action, in the form of antisense agents [11]. Through complementary base pairing, the antisense oligonucleotides may interact in a rational manner with the mRNA of a disease related protein. Two mechanisms have been proposed for the action of antisense agents [12] (Figure 5.1):

- the antisense oligonucleotide exerts a steric interference to ribosome binding and translation or to excision of non-coding introns and recoupling of coding exon sequences. Antisense oligonucleotides can also enter the nucleus where they can inhibit splicing or block transport of the mRNA out of the nucleus (Figure 5.1b).
- mRNA is cleaved by RNase H, which specifically recognises DNA.RNA hybrid duplexes. The enzyme activity is promoted by the binding of the antisense DNA oligonucleotide and permanently eliminates the mRNA.

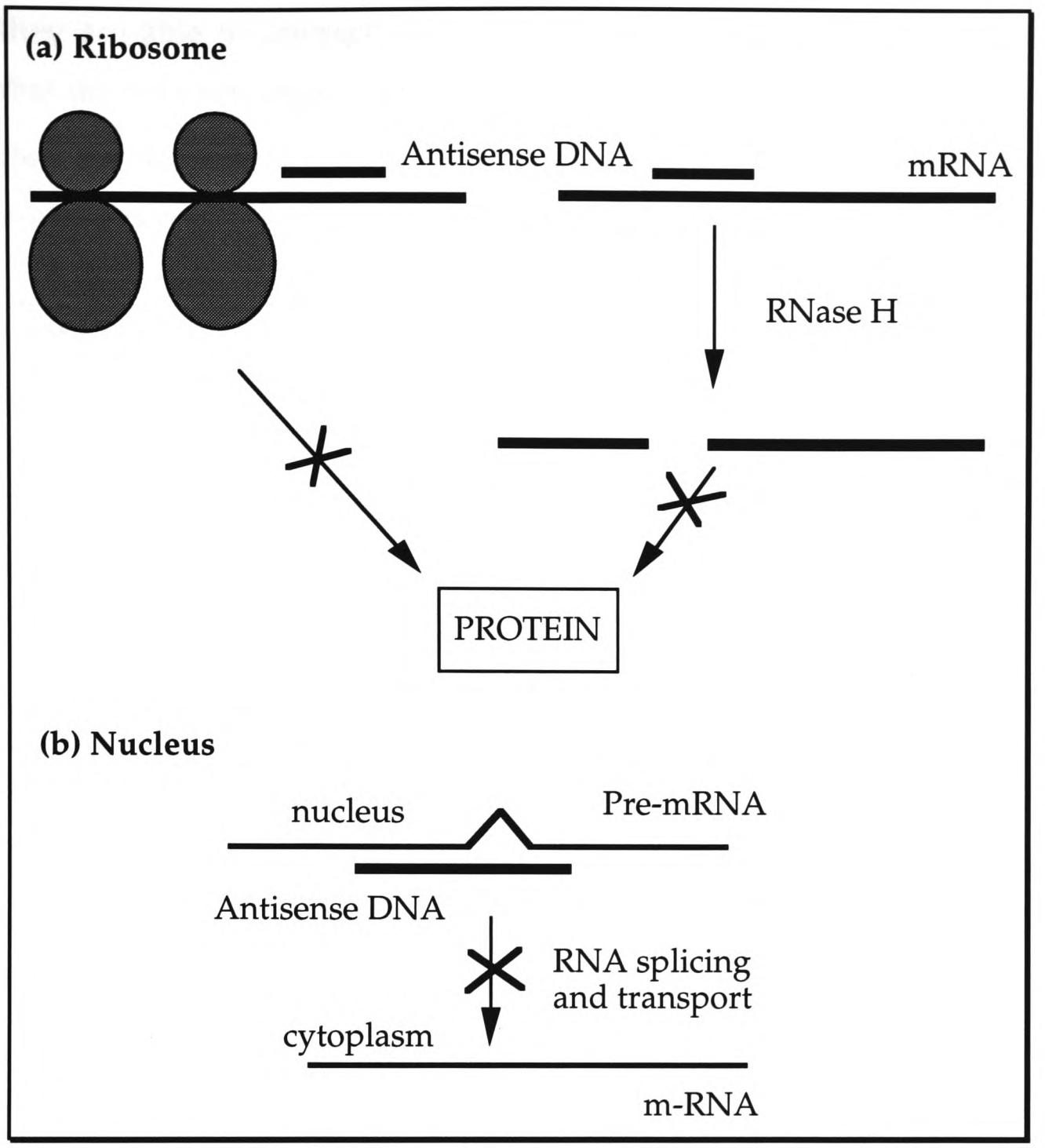


Figure 5.1 Proposed mechanisms of antisense activity

The activity of antisense oligonucleotides has been demonstrated extensively in model systems [11], but their use as therapeutic agents requires that a number of criteria are met:

- they can be synthesised easily and in bulk.
- they are stable *in vivo*.
- they are able to enter the cell and are retained at a sufficiently high concentration to be active.

- they are able to interact with their cellular targets at concentrations that do not exert significant toxicity to the cell.
- they do not interact in a non-sequence specific manner with other macromolecules within the cell.

The development of solid phase methods for the synthesis of short DNA oligonucleotides has greatly simplified the problem of producing antisense agents. However, the scale at which oligonucleotides would be required for therapeutic use still raises questions of the viability of this approach, not least in the cost of producing the oligonucleotides.

The exact mechanism of oligonucleotide entry into cells is currently still under investigation. As oligonucleotides are polyanions, passive diffusion across cell membranes is not possible. Uptake must therefore be an active process, which will be dependent on the length, overall charge, lipophilicity and concentration of the oligonucleotide. Certain modifications may improve cellular uptake, for example a methylphosphonate backbone [13] which removes the charge on the phosphate, or terminal lipophilic groups such as cholesterol [14]. Alternative approaches involve the use of liposome complexes or virus transport systems, both of which significantly improve oligonucleotide uptake [12]. The question of stability to cellular nucleases has also been addressed in some detail, with modifications made to the phosphate backbone, sugar and base [15]. These significantly increase the cellular lifetime of the oligonucleotides relative to natural DNA or RNA which are degraded rapidly.

The specificity of antisense action depends upon both the uniqueness of the target and the strength of the interaction. It can be calculated that the shortest continuous sequence required for a unique target in the human genome is 12-15 bases [16]. More stable hybrids will be formed with longer

sequences but this brings an increased risk of hybridisation to non-targeted mRNAs, with a number of mismatched base pairs. It has been demonstrated that oligonucleotides with mismatched base pairs can still hybridise and induce mRNA degradation [17]. Furthermore, non-sequence specific pharmacological effects and interactions with cellular proteins are possible with the high concentrations of oligonucleotides required. Information on the pharmacology of antisense agents *in vivo* is sparse. However, the longer the oligonucleotide and the longer the exposure, the higher the toxicity is likely to be. Accumulation of breakdown products in non-target organs has been observed and may ultimately govern the overall toxicity, and hence, application of antisense oligonucleotides [12].

A final consideration is choice of target in the mRNA. Areas of mRNA with little secondary structure are often targeted, such as around the initiation codon or 5′ cap. It is often found, however, that even within these regions shifts of only a few base pairs can greatly affect the efficiency of gene inhibition. Furthermore, different targets in the same mRNA can have very different activities [18,19]. Although secondary structure formation in the mRNA may be partly responsible, a full explanation of this phenomenon is not available.

Currently the most promising antisense agents are the phosphorothioate DNAs, which are in phase I clinical trials [15]. Oligonucleotides of this type have been shown to have improved resistance to cellular nucleases, while retaining affinity for the target and RNase H activity [20]. However, the stability of the phosphorothioate DNA and target RNA duplex is considerably lower. As the target RNA is likely to be present at low concentration, optimal stability of the hybrid duplex is desirable. The introduction of other modifications, such as 5-propyne deoxyuridine or deoxycytidine, can help increase the stability of antisense agents [21,22]. As

discussed in the next section, it has also been shown that the stability of DNA.RNA hybrids is sequence dependent. Therefore the elucidation of a relationship between sequence dependent conformation and stability is important not only for extending our fundamental concepts but also in the development of a successful antisense drug technology.

5.2 Structure and Stability of DNA.RNA Hybrids

The structure of DNA.RNA hybrids has been studied extensively in the solid state by X-ray crystallography and in solution by CD and NMR. DNA duplexes can adopt either A- or B-form double helices, and the transition from B- to A-form is induced by lowering the relative humidity [23,24]. In contrast, RNA has never been found to adopt the B-form. It might be expected, therefore, that the conformation of any DNA.RNA hybrid would be A-type to accommodate the RNA strand [25]. Indeed, an early X-ray structure of a DNA.RNA hybrid supported this theory: the duplex $[r(GCG)d(TATACGC)]_2$ was found to have a conformation similar to that of A-RNA, with all sugars in the C3'-*endo* pucker [26]. However, examination of the same sequence in solution did not show the complete conversion to canonical A-form [27]. Clearly, the conformational situation within the DNA and RNA strands is a little more complex.

More recent studies have shown the structures of DNA.RNA hybrids to be intermediate between the A-form of RNA and B-form of DNA. Crystal structures of hybrid duplexes containing only one ribonucleotide have been found to have an A-like conformation [28]. The structures deviated from the regular A-form but maintained a number of key characteristics, such as high propeller twist and inclination of the bases with respect to the helical axis. The conversion of B-DNA to A-form hybrid by a single ribonucleotide was further demonstrated in a separate study [29]. In the crystal structure of an Okazaki fragment, no conformational transition was observed between a three base pair DNA.RNA hybrid portion and a seven base pair DNA duplex portion [28]. In solution, however, drastic changes in the values of a number of backbone torsion angles were seen at the hybrid-DNA duplex junction [30]. High resolution NMR studies of hybrids containing complete strands of DNA and RNA have also demonstrated the A-like structure of hybrid duplexes [6,7,31,32]. These

studies have shown that the sugar conformations in the RNA and DNA strands are different in solution. The riboses are found in the C3'-*endo* pucker (N-type), typical of A-form helices, whereas the deoxyriboses are near to C2'-*endo* (S-type). This has been interpreted in two different ways: either that the sugar pucker is the unusual O4'-endo [9], or that there is an equilibrium between the two favoured puckers, C3'- and C2'-*endo*, with the majority in an S-type conformation [30,33]. The assumption of a two state equilibrium is also favoured by Lane *et al.* [6], and is used in the interpretation of the NMR data of this study.

A number of studies have addressed the question of thermodynamic stability of hybrid duplexes. The order of stability observed for sequences of mixed composition was RNA > DNA > DNA.RNA hybrids [34,35]. In the same study, CD and imino proton resonances suggested that the hybrid duplexes were most similar in structure to the A-form RNA duplex [34]. However, both were found to have a much more favourable entropy but less favourable enthalpy of duplex formation than for either of the pure duplexes. These two terms can be associated mainly with backbone flexibility and degree of base stacking respectively [36]. A-form helices generally have a higher propensity for base stacking with concomitant increase in enthalpy of helix formation [37]. From the lower stability observed, this is clearly not the case for the hybrid duplexes. It is possible that the high entropy term is due to an increased flexibility in the RNA and DNA strands caused by the opposing desire for A- and B- forms respectively. Such an increase in flexibility might be expected to reduce the degree of base stacking, giving duplexes of low stability.

Initial studies of DNA.RNA hybrids with polymer strands indicated that the stability ordering might be composition dependent [38,39]. This has now been confirmed with short oligonucleotides of defined sequence [35,40,41]. In these cases, it was reported that the thermodynamic stability

of a hybrid duplex with a purine rich DNA strand and a pyrimidine rich RNA strand (dR.rY) is much less than the corresponding duplex with purine RNA and pyrimidine DNA (rR.dY). In each case the relative order of stability was found to be RNA > rR.dY > DNA > dR.rY. The origin of the differences has been rationalised in terms of the relative contributions of the 5-methyl group in the DNA strand and the 2′-OH in the RNA strand [42]. Contributions to the stability of the hybrid duplex rR.dY were found to come from both 2′-OH and methyl effects. However, the influence of different conformations was not addressed. There is evidence from electrophoretic mobility and CD studies that the global conformation of the two hybrids dR.rY and rR.dY is different. One such study concluded that the two hybrids belonged to different structural classes, intermediate between canonical A- and B-forms [41]. It was noted from difference CD spectra that the DNA strand of rR.dY undergoes only minimal change from the free single strand on hybrid duplex formation. A separate study showing the same general trends concluded that although the structures were intermediate between A- and B-form, the hybrid rR.dY was more like A-RNA than dR.rY [35]. The higher thermodynamic stability of rR.dY was attributed to this similarity. It has recently been suggested that there may be a continuum of structures between the A- and B-forms modulated by the relative purine-pyrimidine content of the two strands [40].

The conformational contribution to the thermodynamic stability of purine-pyrimidine rich DNA.RNA hybrids has not been addressed in detail. For this study, four decamer duplexes composed of one purine and one pyrimidine rich strand were prepared:

rR.rY	r(GAAGAGAAGC).r(GCUUCUCUUC)
rR.dY	r(GAAGAGAAGC).d(GCTTCTCTTC)
dR.rY	d(GAAGAGAAGC).r(GCUUCUCUUC)

dR.dY d(GAAGAGAAGC).d(GCTTCTCTTC)

Their properties were examined using a combination of UV melting, electrophoresis, CD and ^{1}H and ^{31}P NMR, allowing a full comparison between the hybrids and homoduplexes.

The effects of various base modifications on duplex stability were also examined in this model system. First, the contribution of the methyl groups to duplex stability was determined by adding them to the RNA strand (5-methyl uridine, M) and removing them from the DNA strand (deoxyuridine, U) of the relevant hybrid and homoduplexes. A modification relevant to antisense technology was the replacement of thymidine with 5-propynyl deoxyuridine (P). The modified duplexes prepared were thus:

rR.rY(M)	r(GAAGAGAAGC).r(GCTTCTCTTC)
dR.rY(M)	d(GAAGAGAAGC).r(GCTTCTCTTC)
rR.dY(U)	r(GAAGAGAAGC).d(GCUUCUCUUC)
dR.dY(U)	d(GAAGAGAAGC).d(GCUUCUCUUC)
rR.dY(P)	r(GAAGAGAAGC).d(GCU^{P}U^{P}CU^{P}CU^{P}U^{P}C)
dR.dY(P)	d(GAAGAGAAGC).d(GCU^{P}U^{P}CU^{P}CU^{P}U^{P}C)

5.2.1 Thermodynamic Stability

Duplex melting temperatures (Tm) were measured in 10 mM phosphate buffer, pH 7.0, containing 0.1M NaCl, with a heating rate of 1° per minute. Monophasic melting curves were obtained, and the formation of duplexes demonstrated by the dependence of the melting temperature on the concentration of the strands. The composition of the strands does raise the possibility of triplex formation. However, no evidence for this was found by NMR: under similar buffer conditions but at higher nucleotide

concentration, conditions which would be expected to favour their formation, no triplexes were observed. Examples of melting curves and first derivatives are shown in Figure 5.2. The percentage hyperchromicities on melting increased in the order (approximate mean values in parenthesis): dR.rY (17%) < dR.dY (19%) < rR.dY (23%) < rR.rY (26%).

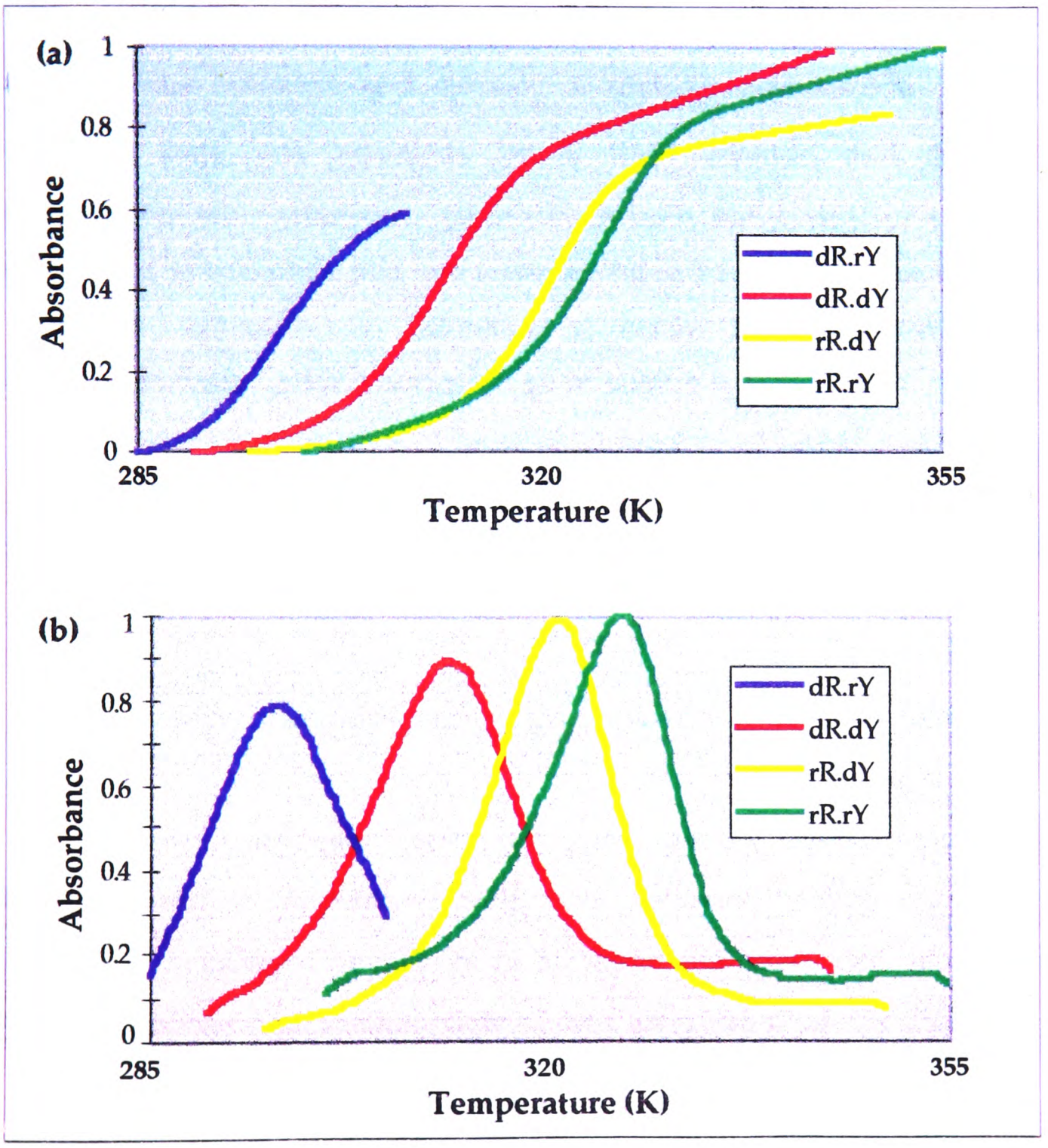


Figure 5.2 (a) UV melting curves, and (b) first derivatives for native dR.dY, dR.rY, rR.dY and rR.rY.

The thermodynamic parameters for each duplex were determined from the concentration-dependence of the melting temperatures using the relationship:

$$\frac{1}{Tm} = \frac{R}{\Delta H^\circ} \ln \frac{Ct}{4} + \frac{\Delta S^\circ}{\Delta H^\circ}$$

where Ct is the total strand concentration, R is the gas constant, and ΔH° and ΔS° the enthalpy and entropy of melting (see section 1.5.3). The standard state was taken to be at 1μM strands and 25°C. The thermodynamic parameters, determined from van't Hoff plots of 1/Tm versus ln(Ct/4) (Figure 5.3), are given in Table 5.1. The full Tm data can be found in Appendix D. Estimates of errors in the thermodynamic parameters were derived as described in section 4.2.

As expected, of the unmodified duplexes, the RNA was found to be the most stable. The order of stability, based on either ΔG° or Tm, was found to be:

$$rR.rY > rR.dY > dR.dY > dR.rY$$

This is in agreement with previous studies with homopurine-homopyrimidine duplexes [35,40]. It is noteworthy that the order of stability also follows the ΔH° term for these duplexes. This is consistent with the poorest base stacking being in the least stable duplex dR.rY [36].

Nearest neighbour parameters are available that allow the prediction of Tm and thermodynamic parameters for DNA, RNA and DNA.RNA hybrid duplexes. However, the parameters are based on measurements made in 1M salt and therefore cannot be used in direct comparison with

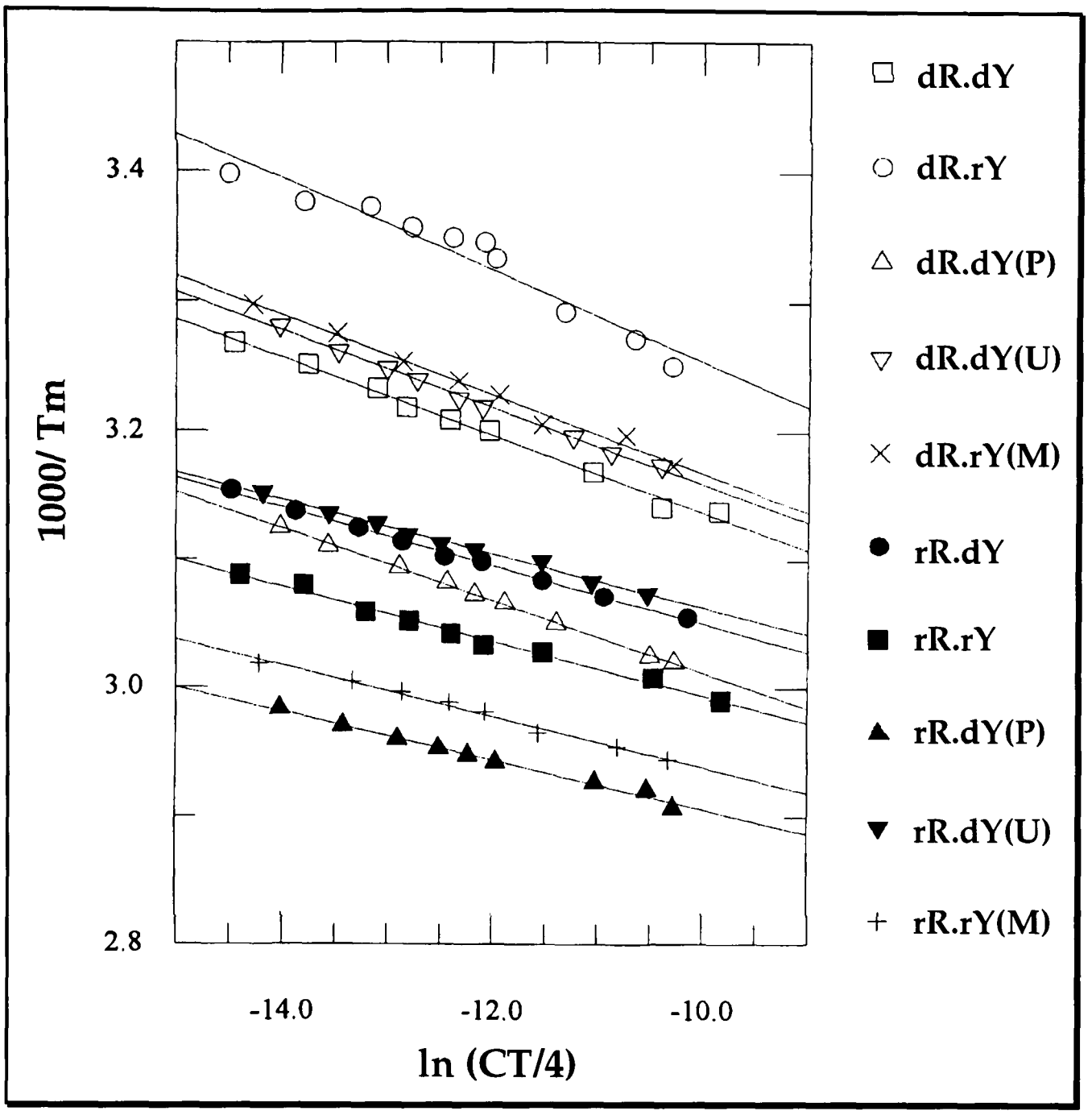


Figure 5.3 van't Hoff plots of 1000/Tm versus ln(Ct/4) for the DNA, RNA and DNA.RNA hybrid duplexes.

the data here. A useful comparison can be made with the data of Lesnik and Freier [40] for duplexes of similar composition. Melting temperatures were obtained under identical solution conditions but ΔG° obtained differently (by direct fit of the data to a two-state model). The values of $\Delta G^\circ_{37^\circ C}$ agree very well once the number of base pairs has been taken into account (Table 5.2).

Duplex	ΔH° ($kJmol^{-1}$)	ΔS° ($Jmol^{-1}K^{-1}$)	$\Delta G^\circ_{25^\circ C}$ ($kJmol^{-1}$)	Tm (K) 1μM strand	Tm (K) 1mM strand
dR.dY	-280 ±9.5	-797 ±27	-42.4 ±1.5	303	323
dR.rY	-238 ±21	-692 ±62	-31.8 ±2.5	291	313
rR.dY	-367 ±8.1	-1036 ±23	-58.3 ±1.3	316	332
rR.rY	-392 ±17	-1091 ±47	-66.9 ±3.0	322	338
rR.rY(M)	-420 ±17	-1150 ±47	-77.3 ±3.0	329	344
dR.rY(M)	-275 ±13	-789 ±38	-39.9 ±1.7	300	320
rR.dY(U)	-397 ±13	-1133 ±37	-59.4 ±2.0	315	330
dR.dY(U)	-283 ±5.8	-811 ±17	-41.3 ±0.8	302	322
rR.dY(P)	-441 ±16	-1199 ±44	-83.7 ±2.9	333	348
dR.dY(P)	-296 ±4.2	-808 ±12	-54.9 ±0.7	317	338

Table 5.1 Thermodynamic parameters for the hybrid and homoduplexes

Duplex	$-\Delta G^\circ_{37^\circ C}$ ($kJmol^{-1}$)			
	rR.rY	rR.dY	dR.dY	dR.rY
9mer	51.1 (**5.7**)	41.9 (**4.6**)	30.6 (**3.4**)	24.4 (**2.7**)
9mer	49.8 (**5.5**)	42.3 (**4.7**)	27.2 (**3.0**)	21.3 (**2.4**)
10mer	53.8 (**5.4**)	45.8 (**4.6**)	32.9 (**3.3**)	23.5 (**2.4**)

Table 5.2 Comparison of ΔG(37°) values in this study (10mer) with those of Lesnik and Freier (9 mers) [40]; values per base pair in parenthesis.

5.2.1.1 5-Methyl Uridine (Ribothymidine) RNA

The addition of methyl groups at the C5 position increases the Tm and thermal stability of both the pure RNA and the hybrid dR.rY (Table 5.3). Comparison of the free energies of the native and modified duplexes indicates a stabilisation of 1.6 and 2.1$kJmol^{-1}$ per methyl group for the hybrid and RNA duplexes respectively. This is considerably lower than

the stabilisation observed per methyl group for an unrelated dodecamer RNA duplex r(CGCAAATTTGCG)$_2$. One difference is that in the case of the dodecamer, the methyl groups are positioned consecutively in the centre of the duplex. In the 10mer duplexes the methyls are more evenly distributed.

Duplex	$-\Delta\Delta H^\circ$ ($kJmol^{-1}$)	$-\Delta\Delta S^\circ$ ($Jmol^{-1}K^{-1}$)	$-\Delta\Delta G^\circ_{25^\circ C}$ ($kJmol^{-1}$)	ΔTm, 1μM (K)
dR.rY(M)	37	97	8.1	9
rR.rY(M)	28	59	10.4	7

Table 5.3 Changes in thermodynamic parameters for 5-methyl uridine duplexes.

For both duplexes, a substantial increase in enthalpy is observed, consistent with an increase in base stacking interactions. While this component is larger in the case of the hybrid duplex, it is accompanied by a less favourable change in the entropic term. For the native duplexes a considerable amount of flexibility was observed in the DNA strand by NMR (section 5.2.3.3). The increase in magnitude of the entropic term may reflect an increase in backbone restriction due to the stronger base stacking. However, NMR analysis of the hybrid dR.rY(M) showed few conformational differences from the unmodified hybrid structure. The unfavourable entropy could therefore reflect changes in the ordered hydration in the grooves of the helix.

In a recent study, Wang and Kool examined the separate effects of C-5 methyl and 2'-OH groups on duplex stability [42]. For duplexes of similar composition, methyl groups were always found to be stabilising, with $\Delta\Delta G^\circ$ similar but always slightly lower than observed here. Furthermore, a comparison of duplexes with consecutive and alternating methyl groups

with the unmethylated case showed only a small preference for adjacent methyl groups. Therefore, hydrophobic methyl-methyl interactions cannot fully account for the stabilisation effect. The evidence presented suggests that the methyl group increases the base stacking efficiency (through an increase in polarisability [43] of the base), which is consistent with the increased enthalpy term observed here. The slight differential preference for adjacent positions can also be explained in this way. The increase in polarisability will allow more favourable van der Waals interactions with neighbouring bases; when the neighbour is also methylated the induced dipole-induced dipole interactions become even larger [42].

5.2.1.2 Deoxyuridine DNA

While the addition of methyl groups to RNA strands has been found to be stabilising, their removal from the DNA strands of pure DNA duplexes and rR.dY type hybrids has been shown to have a detrimental effect on helix stability [42]. However, for the duplexes with deoxyuridine in this study, dR.dY(U) and rR.dY(U), no significant difference in stability compared to the methylated duplexes was observed. This was true in terms of both the measured Tm values and the calculated ΔG° (Table 5.4).

For the pure DNA, dR.dY(U), no significant change was observed in either ΔH° or ΔS° compared to the unmodified duplex. As a result, no difference is observed in the overall free energy. In contrast for the hybrid, large changes in the ΔH° (-30 $kJmol^{-1}$) and ΔS° (-97 $Jmol^{-1}K^{-1}$) terms were observed. In this case, the favourable enthalpic increase is compensated almost exactly by an unfavourable entropic increase with the overall result again being no net difference in free energy. These changes in the hybrid duplex thermodynamic parameters are likely to reflect changes in helix conformation.

Duplex	$-\Delta\Delta H^\circ$ ($kJmol^{-1}$)	$-\Delta\Delta S^\circ$ ($Jmol^{-1}K^{-1}$)	$-\Delta\Delta G^\circ_{25^\circ C}$ ($kJmol^{-1}$)	ΔTm, 1μM (K)
rR.dY(U)	30	97	1.1	1
dR.dY(U)	3	4	-1.1	-1

Table 5.4 Changes in thermodynamic parameters for duplexes containing deoxyuridine.

The results presented here seem to contradict the findings of Wang and Kool, namely that methyl groups are always stabilising [42]. It may simply be that the general rules presented in that study do not always hold. While this may be true, the sequences used were very similar to those of this study. A plausible explanation is apparent when the solution conditions are considered. Of all the native and modified duplexes, the deoxyuridine containing duplexes were least sensitive to salt concentration (section 5.2.2). Thus, in going from 0.1M NaCl, as used in this study, to higher ionic strength, the *lowest increase* in Tm will be observed for the deoxyuridine duplexes. In contrast, for the native duplexes, with a higher sensitivity to salt concentration, a much more significant increase in Tm is observed with increasing salt concentration. In this study, the order of stability in Tm has always mirrored that of free energy for the hybrid duplexes. As discussed previously (section 4.2.3), 10mM Mg^{2+} stabilises the duplex as effectively as 500 mM NaCl for RNA and around 150 mM NaCl for DNA. It is reasonable to assume that the figure will be intermediate for the DNA.RNA hybrid case. The Tm data used by Wang and Kool was collected in 0.1M NaCl with an additional 10 mM Mg^{2+}. It is highly likely that the higher ionic strength has contributed significantly to the increased stability of the native over the deoxyuridine duplexes in that study. As demonstrated here, little difference in stability is observed at low ionic strength. It is clear therefore, that a careful

consideration of the solution conditions is necessary when examining the effects of modifications on helix stability.

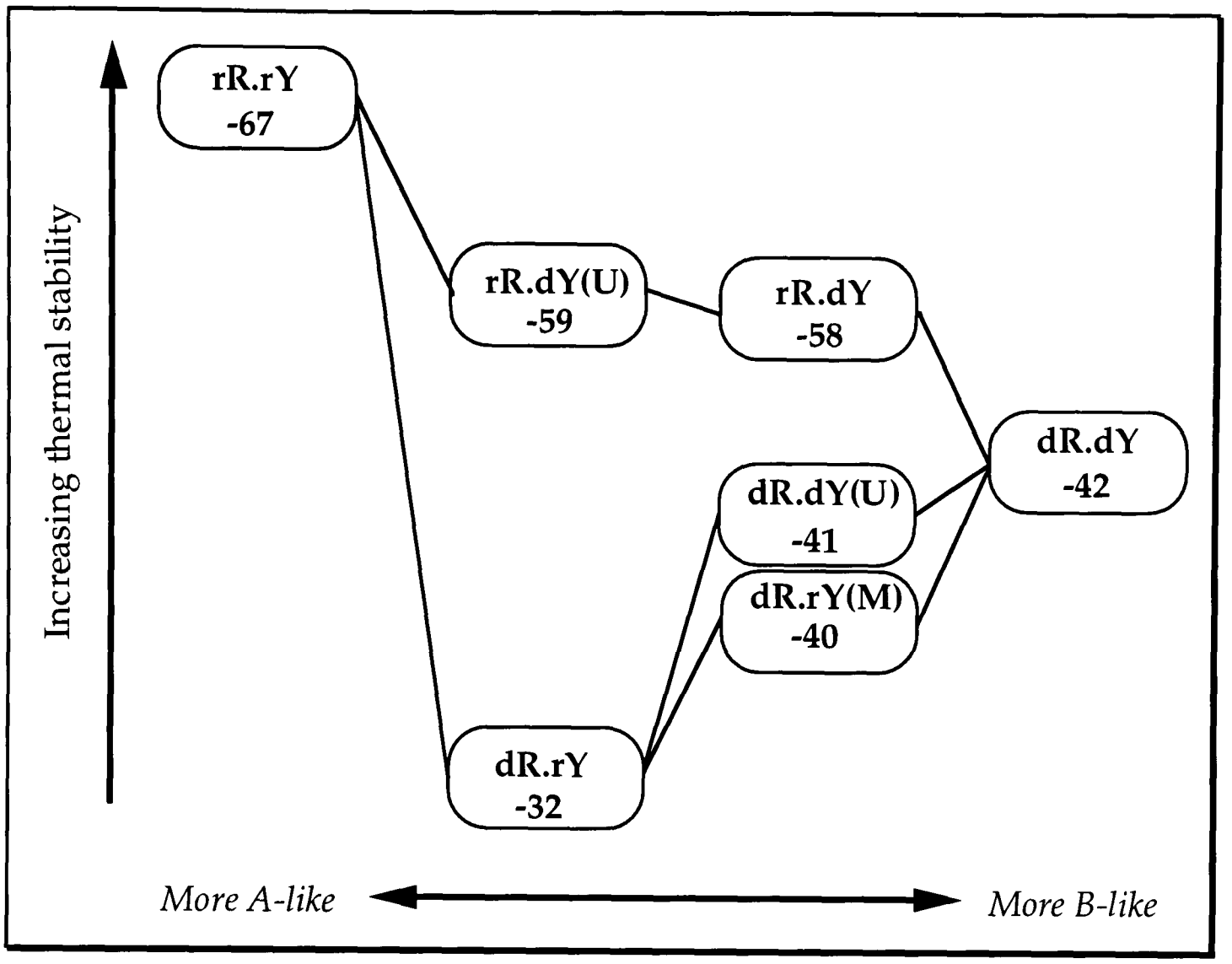


Figure 5.4 Thermodynamics of converting rR.rY into dR.dY.

The free energies involved in the conversion of the pure RNA duplex into the pure DNA duplex can be charted using the values determined for the deoxyuridine and 5-methyl uridine duplexes (Figure 5.4). The largest changes all occur in the removal or addition of hydroxyl groups. Removal of hydroxyl groups is always destabilising, except for the conversion of the unstable hybrid to the deoxyuridine DNA. The most dramatic change ($\Delta\Delta G^\circ$ 35 $kJmol^{-1}$) occurs for the removal of hydroxyls from the purine strand of the pure RNA. Under the solutions conditions used here methyl groups have little effect on helix stability, except for the stabilising effect on addition to the pyrimidine strand of dR.rY. At higher ionic

strength, the hybrid duplexes rR.dY(U) and dR.dY(U) would be expected to be of lower stability relative to the other duplexes.

5.2.1.3 5-Propyne Deoxyuridine DNA

Pyrimidine bases substituted at the C5 position with propyne groups are known to stabilise double helix formation over the natural bases [21]. This is thought to be mainly through an increase in base stacking interactions involving the planar propyne group [44]. A favourable entropy change could also arise from the displacement of ordered waters of hydration from the grooves of the helix by the hydrophobic propyne group.

To date, no high resolution study of duplexes with and without C5-propyne modification is available to support either hypothesis. Despite this, the potential application of the groups in antisense technology has been readily accepted. The propyne modification has two attributes that make it favourable for antisense technology: the interaction of DNA and RNA is greatly stabilised and the increased hydrophobicity may facilitate non-receptor mediated entry into cells [21]. Gene-selective antisense agents as short as seven nucleotides in length containing C5-propyne groups have been reported [22].

The extra stability of DNA.RNA hybrids with purine rich RNA and pyrimidine rich DNA has been conclusively demonstrated, both here and in the recent literature. For antisense application, it is important that the benefits of modifications in the DNA strand, such as the introduction of C5-propyne groups, are still apparent in this class of hybrid. Thus, the effect of replacing the C5-methyl groups of each thymidine with C5-propyne groups in the dGCTTCTCTTC hybrid strand was examined.

The stabilising effect of the propyne group is clear for both duplexes (Table 5.5), though it is much greater in the hybrid ($\Delta\Delta G^\circ$= -25.4 kJmol^{-1}) than the pure DNA duplex ($\Delta\Delta G^\circ$= -12.5 kJmol^{-1}). The average increase in free energy per propyne group is -5.1 kJmol^{-1} for the hybrid but only -2.5kJmol^{-1} for the DNA duplex. In both cases, the increased stability arises from an increase in the enthalpy term. For the hybrid, this increase is extremely large (-74 kJmol^{-1}), but is accompanied by a moderately unfavourable increase in entropy. In contrast for the DNA duplex, the increase in enthalpy is moderate (-16 kJmol^{-1}) but no significant change is observed in the entropy term.

Duplex	$-\Delta\Delta H^\circ$ (kJmol^{-1})	$-\Delta\Delta S^\circ$ (Jmol^{-1}K^{-1})	$-\Delta\Delta G^\circ_{25^\circ C}$ (kJmol^{-1})	ΔTm, 1μM (K)
rR.dY(P)	74	163	25.4	17
dR.dY(P)	16	11	12.5	14

Table 5.5 Changes in thermodynamic parameters for duplexes containing 5-propynyl deoxyuridine.

For both duplexes, the results are consistent with a considerable increase in base stacking interactions involving the propyne groups. The differences compared to the native duplexes are most probably a consequence of the different conformations these duplexes adopt (section 5.2.3). For the hybrid duplex, the increase in entropy is consistent with a significant loss of conformational freedom in the helix with the improved base stacking. By analogy for the propyne DNA duplex, less conformational rearrangement would appear to be required to accommodate the additional base stacking interactions. An alternative explanation is possible: a favourable increase in entropy, through the displacement of ordered water molecules, combined with an unfavourable loss of entropy, through a restriction of conformational

freedom, could also result in no net change in entropy. Conclusive arguments cannot be made from the thermodynamic data alone. The determination of the origin of the stabilising effect of the propyne group will require high resolution studies by NMR and/ or X-ray crystallography. Such studies are currently underway in our laboratory. Whatever the origin of the effect, it is clear that addition of propyne groups to pyrimidine rich DNA greatly stabilises its interaction with RNA. Thus, the use of 5-propyne DNA has tremendous potential in the application of oligonucleotides as antisense agents.

5.2.2 Effect Of Salt Concentration

Duplex melting temperatures were measured in solutions of different ionic strength. As expected, the Tm was strongly dependent on the concentration of NaCl, increasing with increasing salt (section 4.2.2). Tm values for each duplex at three salt concentrations are given in Table 5.6.

Duplex	Tm (K) 0.1M NaCl	Tm (K) 0.25M NaCl	Tm (K) 1.0M NaCl	ΔTm (1.0M-0.1M)
dR.dY	310.7	316.0	320.7	10.0
dR.rY	297.9	302.4	308.6	10.7
rR.dY	321.1	325.2	328.7	7.6
rR.rY	327.6	332.1	336.9	9.3
rR.rY(T)	338.0	340.1	342.4	4.4
dR.rY(T)	313.1	315.6	317.2	4.1
rR.dY(U)	323.5	325.5	326.9	3.4
dR.dY(U)	312.6	315.9	317.1	4.5
rR.dY(P)	343.6	345.5	351.1	7.5
dR.dY(P)	327.4	330.4	331.9	4.5

Table 5.6 Duplex melting temperatures at various salt concentrations.

For the native duplexes plots of Tm against log [Na^+] gave slopes of between 8 and 10. The less stable hybrid, dR.rY, was the most sensitive to salt concentration and the stable hybrid, rR.dY, least sensitive. The slopes for all the modified duplexes were considerably lower between 3 and 7. The Tm values at the highest salt concentration for most of these duplexes deviated slightly from this linear relationship, and is in part responsible for this difference. It has been shown that duplex Tm increase with increasing salt concentration up to around 1M and can then decrease [45]. Thus, to thoroughly examine the effect of salt concentration on the stability of these duplexes, data will need to be collected at much lower concentrations (10-100 mM NaCl). Differences in salt sensitivity between the native and modified hybrids might arise from large changes in helix conformation. It is also possible that changes in base stacking and/ or ordered waters of hydration are responsible for these different properties.

5.2.3 Solution Conformation

To assess the conformational contribution to the thermodynamic stability of the DNA.RNA hybrids their solution structures were examined using a number of techniques. Comparisons could then be made with the DNA and RNA homoduplexes to help explain the different thermal stabilities determined for these duplexes.

5.2.3.1 Gel Mobility

The mobility of each of the four native duplexes was measured by non-denaturing PAGE (Figure 5.5). The extra intensity of the dR.rY band is due to the higher loading concentration required for this unstable duplex. All the duplexes have identical compositions (other than the substitution of U for T in the RNA pyrimidine strand) and identical charges. Therefore, any differences in electrophoretic mobility should arise primarily from

differences in hydrodynamic volume, or shape. This could be due to gross differences in conformation (such as A- and B-form helices), large differences in hydration or counterion condensation (dependent upon overall conformation) or differential interaction with the gel matrix.

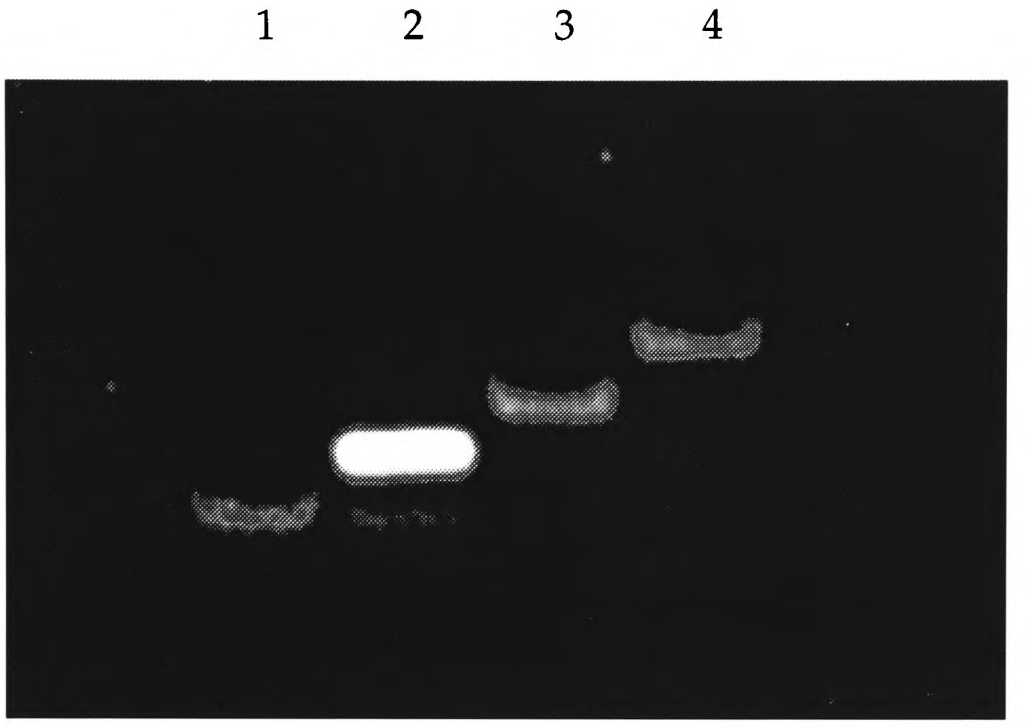


Figure 5.5 Non-denaturing PAGE of the native duplexes (visualised with ethidium bromide): lanes 1 to 4 are dR.dY, dR.rY, rR.dY and rR.rY.

The order of mobility was found to be (μ_{rel} values in parenthesis):

$$dR.dY\ (1.00) > dR.rY\ (0.92) > rR.dY\ (0.84) > rR.rY\ (0.77)$$

in agreement with the order reported by Ratmeyer *et al.* [35]. The ordering of the mobilities clearly indicates that the native hybrid duplexes are different from both the homoduplexes and each other. The intrinsic frictional resistance of RNA and DNA duplexes of the same number of base pairs appears to be about the same [personal communication, Dr. Andrew Lane]. This suggests that the differences are due primarily to differences in the ionic atmosphere associated with each duplex and not the overall size and shape. The pattern observed suggests that the more

stable hybrid, rR.dY, may be more similar to the RNA duplex while the less stable hybrid, dR.rY, is more like the DNA duplex.

The relative mobilities of the DNA.RNA hybrids containing strands with modified bases have also been measured (Table 5.7). The addition of methyl groups to the RNA strand in dR.rY(M) causes a moderate decrease in mobility (0.05 compared to 0.08 on average between the native duplexes). For the unmodified hybrids, the differences in mobility can be explained by the large differences in conformation and hence ionic atmosphere. However, the NMR data suggest that there are no overall conformational changes in going from dR.rY to dR.rY(M). The difference in mobility may therefore result from another factor, such as differences in the hydration of the duplexes.

Duplex	dR.rY	dR.rY(M)	rR.dY	rR.dY(U)	rR.dY(P)
μ_{rel}	0.92	0.87	0.84	0.87	0.79

Table 5.7 Comparison of the relative electrophoretic mobilities of the native and base modified hybrid duplexes.

The other type of modification, replacement of the C5-methyl with C5-H or C5-propyne also has a moderate effect on mobility (again the changes are smaller than between the native hybrids). Interestingly, the removal of the methyl groups, to give rR.dY(U), increases the mobility such that it is closer to the other hybrid with the RNA pyrimidine strand. Conversely, the replacement of the methyl groups with bulkier, more hydrophobic propyne groups decreases the mobility relative to the native. The mobilities are thus ordered in terms of both C5-substituent bulk and hydrophobicity (lowest to highest):

rR.dY(U) (0.87) > rR.dY (0.84) > rR.dY(P) (0.79)

The preliminary NMR data for these modified hybrids suggests that conformational rearrangement is more significant and widespread than for dR.rY(M). Thus, the differences in mobility of rR.dY(U) and rR.dY(P) are probably due in part to gross conformational changes. Given the changes in relative hydrophobicity, it is likely that changes in the ordered hydration also has a significant effect on the mobility.

5.2.3.2 Circular Dichroism

The absorption spectrum for each duplex was measured and found to be similar, with λ_{max} between 257 and 258nm. The near UV CD spectra of all four duplexes were recorded at a concentration of 40 μM duplex in 0.5M KCl at 15°C (Figure 5.6). From the UV melting data, it is clear that duplex formation is highly favoured for all duplexes under these conditions.

The pure DNA and RNA duplexes show spectra expected for A- and B-form helices respectively [23]. The RNA duplex showed a typical non-conservative spectrum with an intense positive peak at 264 nm and a weak negative peak at 232 nm. The spectrum of the DNA duplex is typically conservative and integrates to approximately zero. It shows a positive peak at 281 nm and a negative peak at 249 nm. The intensity in the spectrum of the RNA duplex is approximately 2.5 fold higher than that in the DNA spectrum. The spectra of the two hybrids are intermediate between these extremes, though both are non-conservative with features more A-like than B-like. Most significantly, the two hybrids are clearly very different from each other. The less stable hybrid duplex, dR.rY, shows a spectrum with a very intense positive peak at 268 nm similar to the RNA duplex but also a negative peak at 242 nm of similar intensity to that of the DNA duplex. Conversely, the stable hybrid, rR.dY, has a much less intense positive peak at 278 nm similar to the DNA duplex but only a very weak negative peak at 240 nm reminiscent of the

RNA duplex. The order of the spectral integration values is dR.dY < dR.rY < rR.dY < rR.rY, identical to the pattern observed in the hydrodynamic study (section 5.2.3.1).

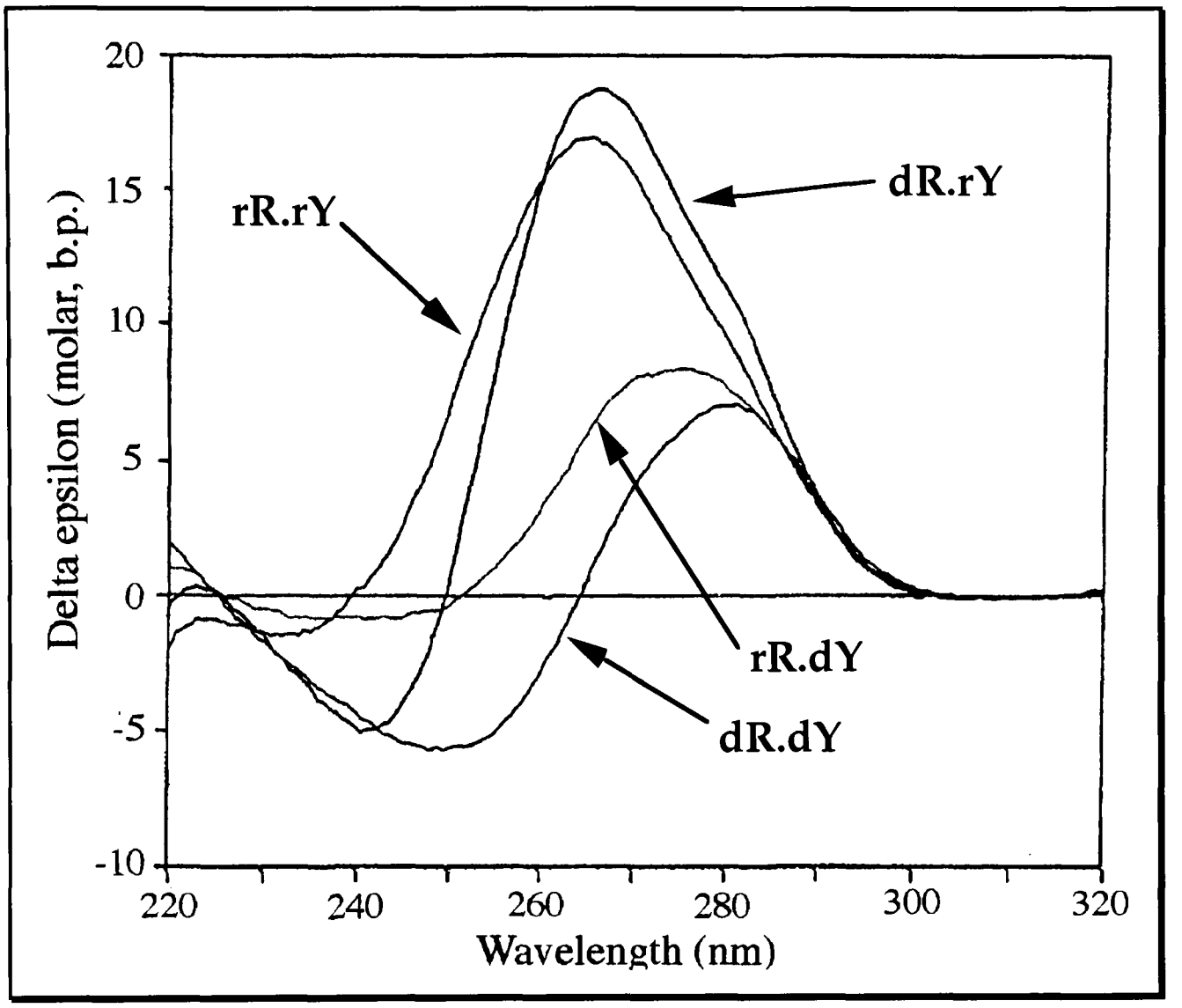


Figure 5.6 CD spectra of the four native duplexes.

These results are comparable to those reported previously for dodecamer duplexes consisting of alternating GA.CT(U) base pairs [35]. In the duplexes studied here, the purine strands have identical composition whereas the pyrimidine strands differ only in the presence or absence of the 5-methyl group in U and T of the ribo- and deoxyribonucleotides respectively. It is possible that different transition moments for U and T could produce significantly different spectra. However, in polymeric RNA of mixed sequence [23], the spectral properties are very similar to those of the RNA duplex in this study. It is therefore reasonable to conclude from

the CD spectra of the four duplexes that the DNA.RNA hybrids adopt conformations intermediate between those of the pure RNA or DNA duplexes with considerably different degrees of base stacking. Further, they are conformationally different from each other, reflected in their markedly different thermodynamic stabilities. Clearly, there must be considerable differences between the hybrid duplexes at the nucleotide level.

The CD spectra of each of the modified DNA. RNA hybrids has also been recorded (Figure 5.7). The addition of methyl groups to the rY strand to give dR.rY(M) has shifted both the positive (270 nm) and negative peak (244 nm) slightly to higher wavelength, towards the positions observed for these peaks in the DNA spectrum. The overall spectrum is, however, still more A- than B-like. Both peaks are more intense, though the positive peak is much broader while the negative peak is narrower than for the native duplex. These changes are consistent with a change in the base stacking interactions, as indicated by the different thermal stabilities.

The modifications at the thymidine C5 position have quite different effects on the CD spectra of the hybrids. Removal of the methyl groups, dR.dY(U), has little effect on the position of the positive peak (278 nm) although it is broader and more intense. The major difference occurs at lower wavelength, where a shoulder on the positive peak (240-250 nm) is observed rather than the very weak negative peak. The spectrum of the propyne modified duplex, rR.dY(P), has a yet more intense, broad positive peak which is shifted to lower wavelength (269 nm). Again the major difference lies at lower wavelength, where a new positive peak is observed (238 nm) compared to the weak negative peak of the native duplex. Both modified duplexes are clearly different from the native. In the case of the propyne modification, the changes probably reflect mainly

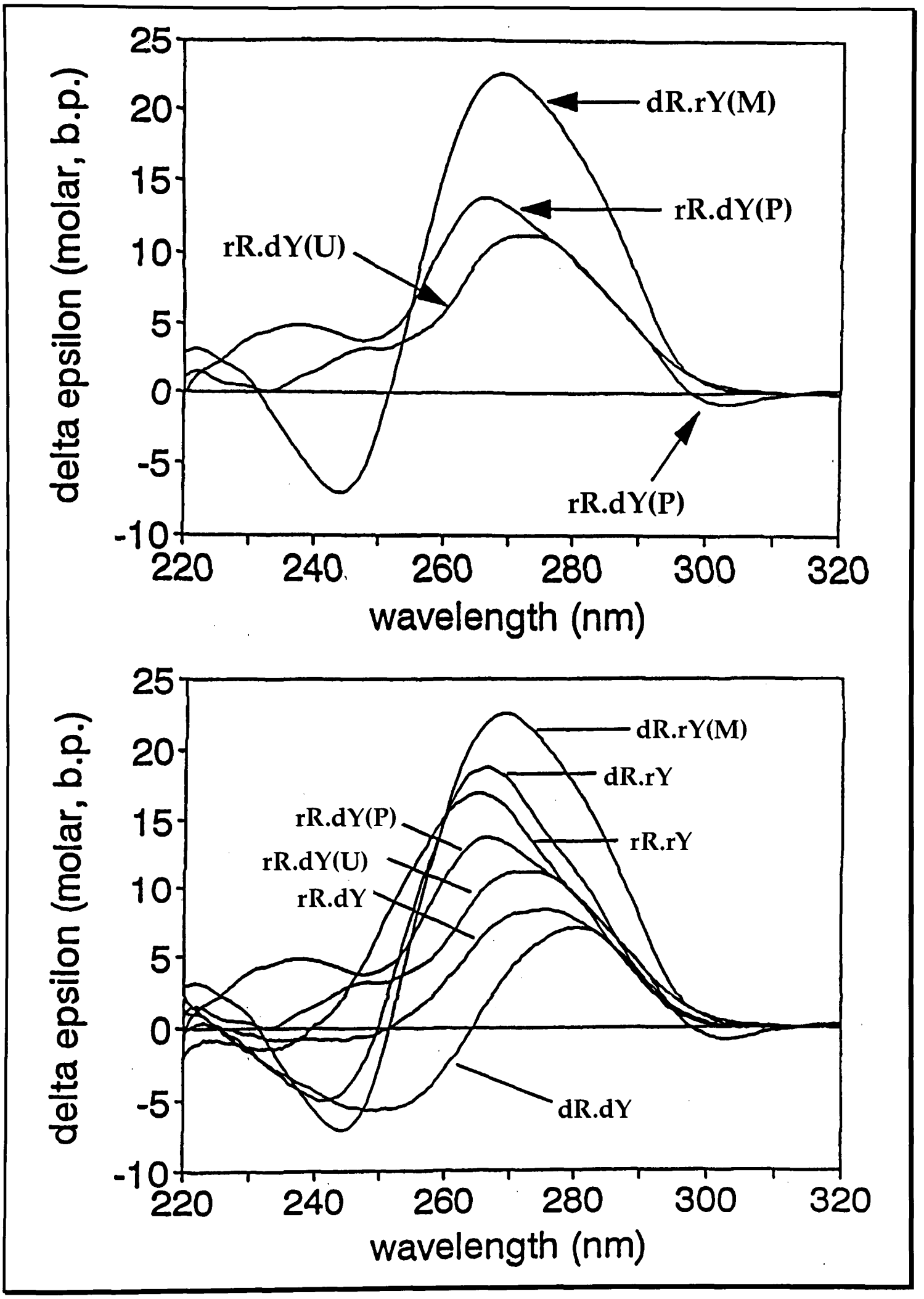


Figure 5.7 CD spectra of modified duplexes (top) and comparison with native duplexes (bottom).

the considerable increase in base stacking interactions (although some differences from purely electronic effects cannot be ruled out). In both cases, the differences are also consistent with some conformational rearrangement as indicated by the gel mobility.

5.2.3.3 Nuclear Magnetic Resonance

^{1}H and ^{31}P NMR was used to probe the conformational differences between the hybrid duplexes in greater detail. The preparation of NMR samples and recording of spectra has been described previously [46]. Assignments were made from NOESY, DQF-COSY and TOCSY spectra. An example of a sequential walk used in the assignment of sugar protons is shown in Figure 5.8. Full assignment of non-exchangeable protons was possible for the DNA duplex, hybrid DNA strands and the RNA purine strand, with the exception of the unresolved H5′/H5″. For the pyrimidine RNA strand, spectral overlap precluded the unambiguous assignment of all H3′ and H4′ resonances. For the two hybrids and the RNA duplex intense cross-peaks were present between Ade H2(i) and H1′(i+1) on both strands. This is consistent with substantial propeller twisting of the A-U base pairs in these duplexes.

The duplexes were compared in terms of chemical shift differences, backbone conformation, sugar pucker and glycosidic torsion angle. The distance and angle constraints generated were then used to build preliminary models of each duplex, showing the differences in global conformation.

Differences in chemical shifts for the assigned resonances were calculated as hybrid deoxyribose proton shifts minus DNA duplex shifts, and hybrid ribose shifts minus RNA shifts. The largest differences were observed in the DNA strands for both hybrids. This suggests that the greatest

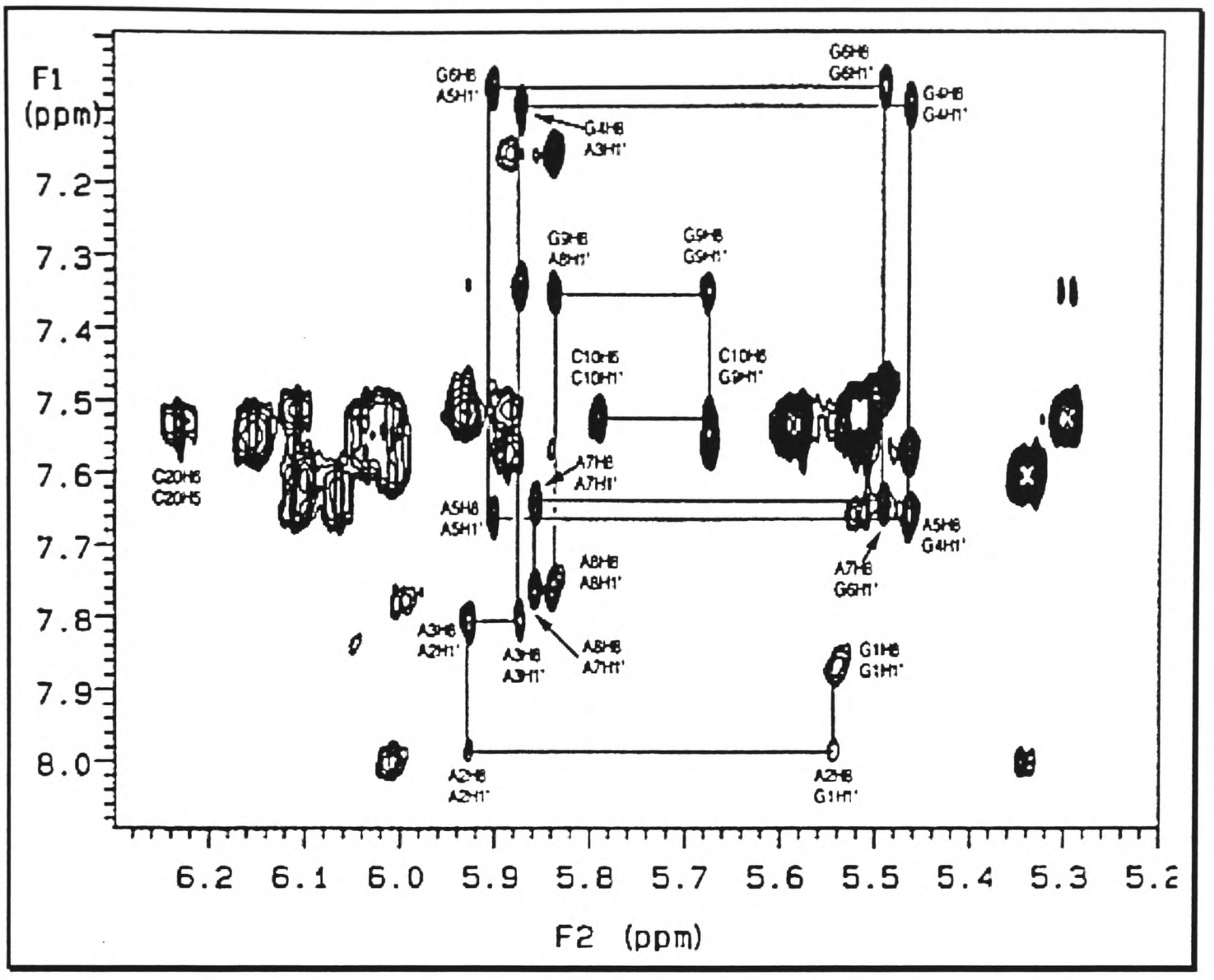


Figure 5.8 Sequential walk H1'-H6/8

conformational rearrangement occurs in the DNA strands, consistent with the hybrid conformations being most similar to that of the RNA duplex. The signs and absolute magnitudes of the shift differences were different for the two hybrids, indicating that they are conformationally distinct. Further evidence for this came from ^{31}P spectra where the chemical shift dispersions were in the order:

dR.dY (0.45 ppm) < dR.rY (0.60 ppm) < rR.rY (0.70 ppm) < rR.dY (0.72 ppm)

For comparison, a chemical shift dispersion of 0.5 to 0.6 would be expected for mixed sequence DNA duplexes [47], and the RNA duplex $r(CGCAAAUUUGCG)_2$ had a shift dispersion of 0.78 ppm under identical

conditions. This is consistent with the conformation of the hybrid rR.dY being closer to the A-conformation of the RNA duplex. The hybrid dR.rY clearly has a more intermediate conformation at the level of the backbone torsion angles.

Sugar conformations were determined from sums of coupling constants derived from cross-sections of DQF-COSY and NOESY spectra, and where possible 1D spectra. The riboses in all three RNA containing duplexes were found to have sugars with the expected C3′-*endo* pucker (N domain) with P_s -32 to 32°. Only for the terminal residues was some conformational freedom observed.

The distance H1′-H4′ ($r_{1'4'}$) could also be estimated from NOESY spectra. This distance is at a minimum for P_s=90° (O4′-*endo*) and increases sharply in the ranges $100<P_s<62$ and $90>P_s>18$ [48]. It can therefore be used to distinguish mixtures of N and S conformations [49] from a single unique conformation with P_s~90° [9]. In the pure DNA duplex, DQF-COSY cross-peaks indicate a predominance of sugars with C2′-*endo* pucker (S domain) for all but the terminal nucleotides. A more detailed analysis of the deoxyribose purine sugar conformation was possible. A two-state equilibrium model was used as the sums of coupling constants and $r_{1'4'}$ distances cannot be rationalised as a single unique conformation. Conformational mixtures were systematically generated as a function of P_s and f_s (the fraction with pseudorotation angle P_s) and compared with the experimental data using the program pfit [49]. The mean values of P_s and f_s produced for the DNA strands in the different duplexes are given in Table 5.8. The value of f_s for the DNA duplex is relatively high, and P_s is in the S domain, typical of B-form DNA. Fewer data were available for the pyrimidine strand, however, a similar structure is suggested but with a slightly lower pseudorotation phase angle.

Duplex	P_s ± s.d.	f_s ± s.d.
<u>dR</u>.dY	141 ±13	0.89 ±0.03
dR.<u>dY</u>	125 ±12	0.83 ±0.07
rR.<u>dY</u>	136 ±20	0.67 ±0.07
<u>dR</u>.rY	134 ±13	0.79 ±0.06

Table 5.8 Mean Ps and fs values for the DNA strands (underlined)

As for the pure DNA duplex, the data for the DNA strands in the hybrids was found to be inconsistent with a single unique conformation. It is clear from Table 5.8 that the hybrid DNA strands have different conformations from their counterparts in the DNA duplex. The mean values of P_s and f_s are considerably lower than the equivalent strands in the pure DNA duplex. This indicates an increase in the degree of flexibility in the hybrid DNA strands, in agreement with previous detailed studies on DNA.RNA hybrids [6,33]. The results also show that the level of flexibility is greater on average in the pyrimidine DNA strand of hybrid rR.dY. Flexibility within the DNA strands was detected at the nucleotide level: how widespread this is, and the contributions to the different thermodynamic stabilities will require a full three-dimensional structure determination. This analysis is now in its final stages.

The final conformational parameter examined was the glycosidic torsion angle. Intranucleotide NOE intensities were typical of an *anti* conformation about the glycosidic bond. For the DNA strands, these were consistent with χ in the range -90 to -120° in the pure duplex, and -100 to -130° in the hybrid duplexes. For the ribonucleotides a χ of around -160° was observed.

In most cases, the glycosidic torsion angle could be analysed in greater detail using the program NUCFIT [50]. A mixture of N and S sugar

conformations for DNA strands were incorporated into the calculations. In general, the torsion angles were found to be highest for the deoxynucleotides in the DNA duplex followed by the deoxynucleotides in the hybrids. The lowest glycosidic torsion angles were found for the ribonucleotides. The values for the RNA strands were similar whether they were in a hybrid or pure duplex. In contrast, large changes were observed for the DNA strands, demonstrating again that the conformational rearrangements occur in the DNA strands of the hybrids.

Computer generated models of the four native duplexes, based on the NMR data, clearly show the global differences in conformation (Figure 5.9). The most striking differences are in the width and depth of the major groove: deep and narrow in RNA but relatively wide and shallow in DNA. The two hybrids are clearly intermediate between these two extremes.

A detailed analysis of the solution conformation of the modified hybrids has also been undertaken. For the methylated RNA hybrid, dR.rY(M), this work is now near completion. The conformational analysis for this duplex indicates that there are no major rearrangements compared to the native. Chemical shift differences have been calculated as for the native duplexes (above). On methylation of the RNA in this hybrid, very little change is observed in the chemical shifts (Figure 5.10). Only minor changes are observed, at either H1′ or base positions. These results suggest that the difference in duplex stability for dR.rY(M) result purely from increased base stacking and changes in the helix hydration.

Conformational analysis of the hybrids rR.dY(U) and rR.dY(P) is at an earlier stage. However, the chemical shift differences have been calculated for these duplexes (Figure 5.10). A small number of moderate differences are observed for the deoxyuridine hybrid, mainly in the sugar protons H2′

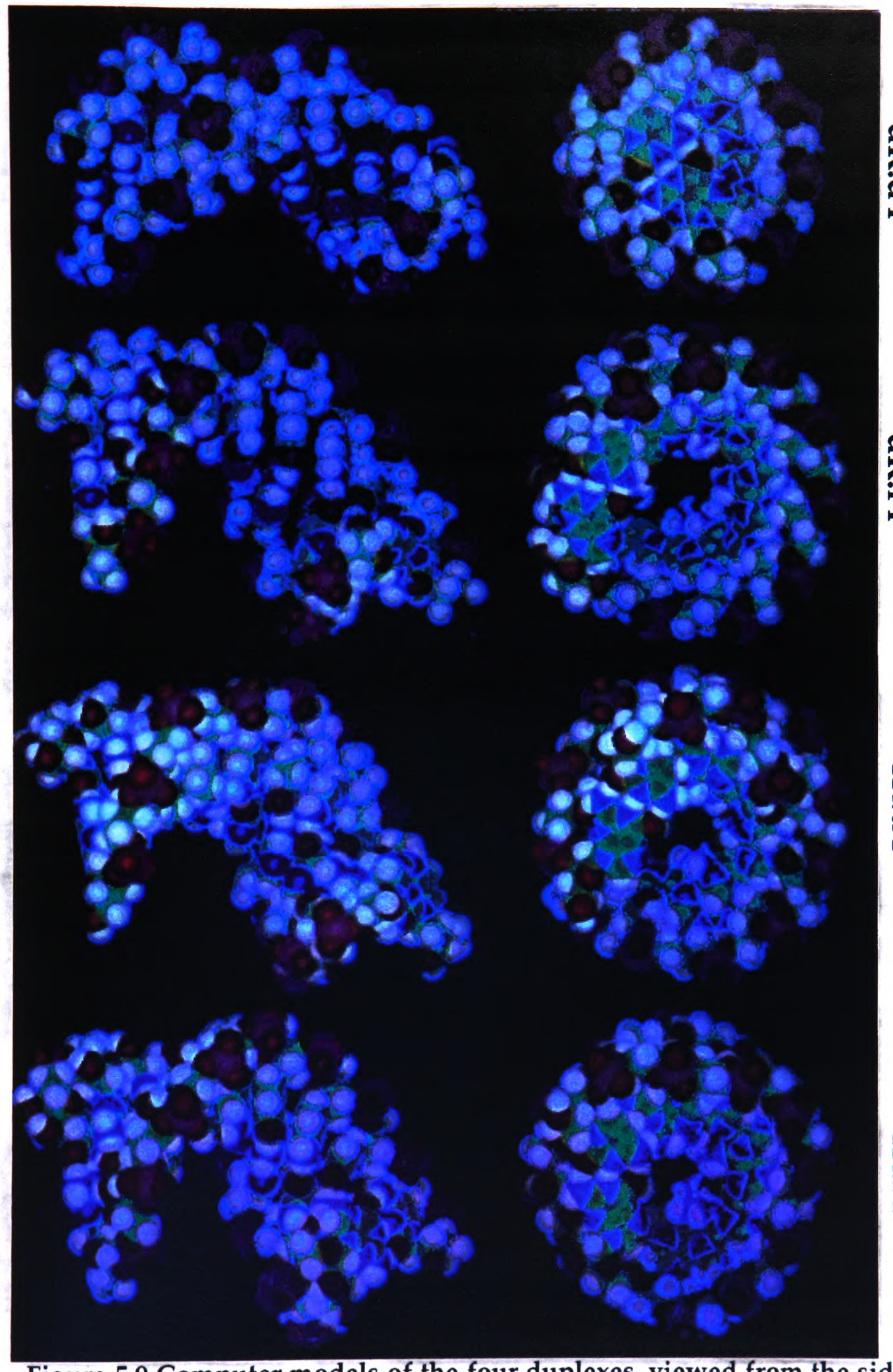


Figure 5.9 Computer models of the four duplexes, viewed from the side and from above.

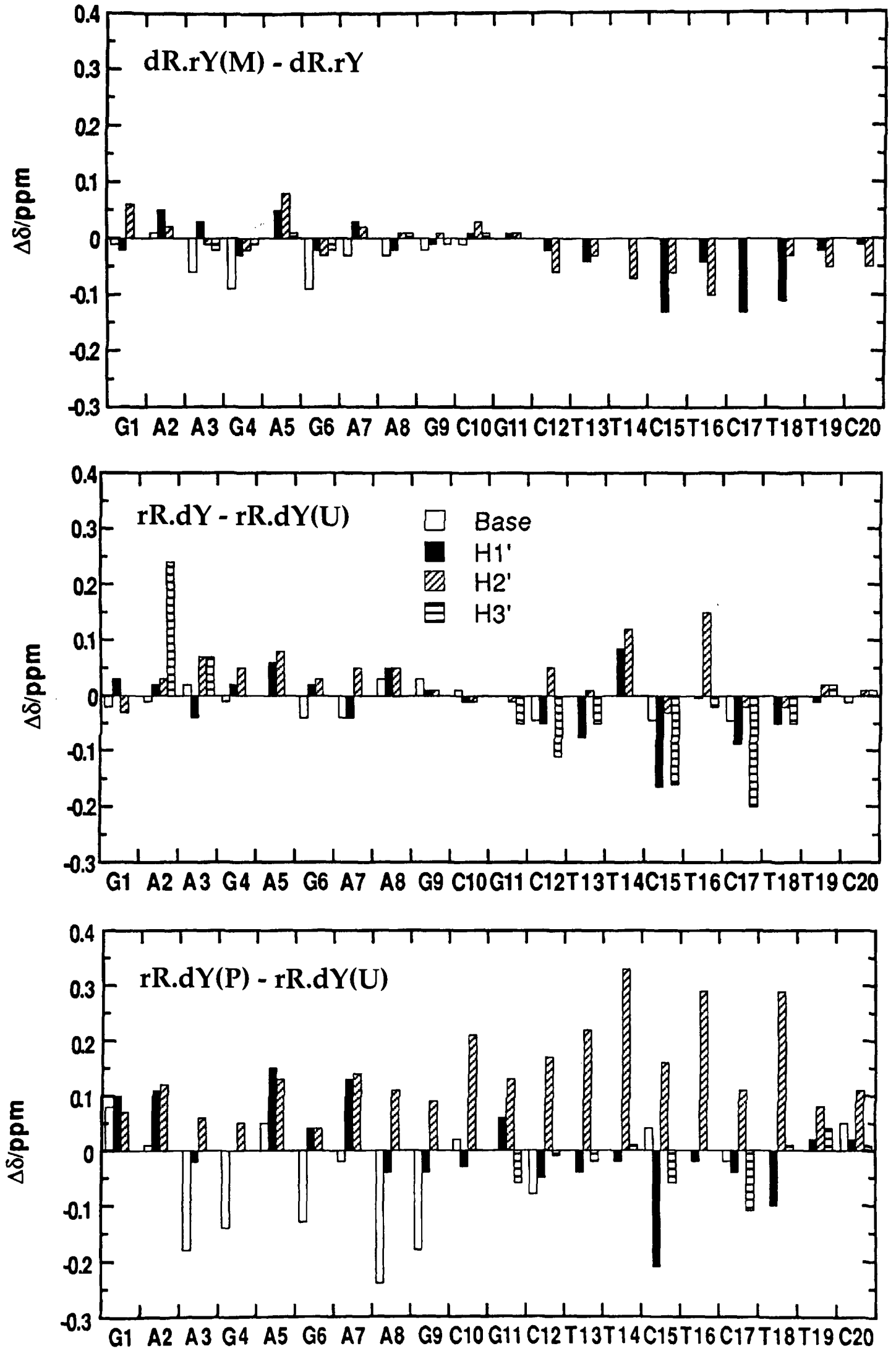


Figure 5.10 Chemical shift differences for the base modified hybrids.

and H3′ (positive and negative differences). In contrast, quite major differences are observed for the propyne substituted hybrid. These occur both in the DNA strand (H2′, positive) and the RNA strand (some H1′ and H2′ positive, base protons negative). These results support the observation from the gel mobility and CD spectra that conformational rearrangements are occurring. These rearrangements appear to be far more dramatic where the helix must accommodate the addition of the propyne group compared to the methyl.

5.2.3.4 Qualitative Enzyme Assays

DNase I is a non-specific nuclease that acts from the minor groove, and cleaves phosphodiester bonds on one strand at a time. Although it is relatively non-specific in its action on DNA, it does not cleave RNA. DNA.RNA hybrids could in principle be substrates as the conformation of the deoxynucleotides is similar to that in DNA [6]. The DNA.RNA hybrids of this study have been shown to be conformationally distinct, raising the possibility of different enzyme activity on each. The enzyme activity may be monitored simply by observing the increase in absorbance at 260nm on mixing with the oligonucleotide (Figure 5.11). These enzyme assays were conducted with a duplex concentration of around 2.5 μM. The rapid increase in absorbance with the pure DNA indicates that this duplex is rapidly cleaved. The overall increase in absorbance of 0.056 represents a hyperchromicity of around 13%. As digestion to mononucleotides would be expected to give a hyperchromicity of around 30%, this result is consistent with significant digestion of the duplex (fragments around four base pairs or less are not substrates). In contrast, for both native DNA.RNA hybrids and the methylated hybrid dR.rY(M), a very slow rate of increase in absorbance is observed. This indicates only very minimal activity with these substrates. It should be noted that it is not actually the rate of bond cleavage that is being measured but the dissociation of

cleaved products. However, the lack of activity of DNase I against these substrates has been confirmed using a spectrophotometric assay of the rate of bond cleavage rate [51].

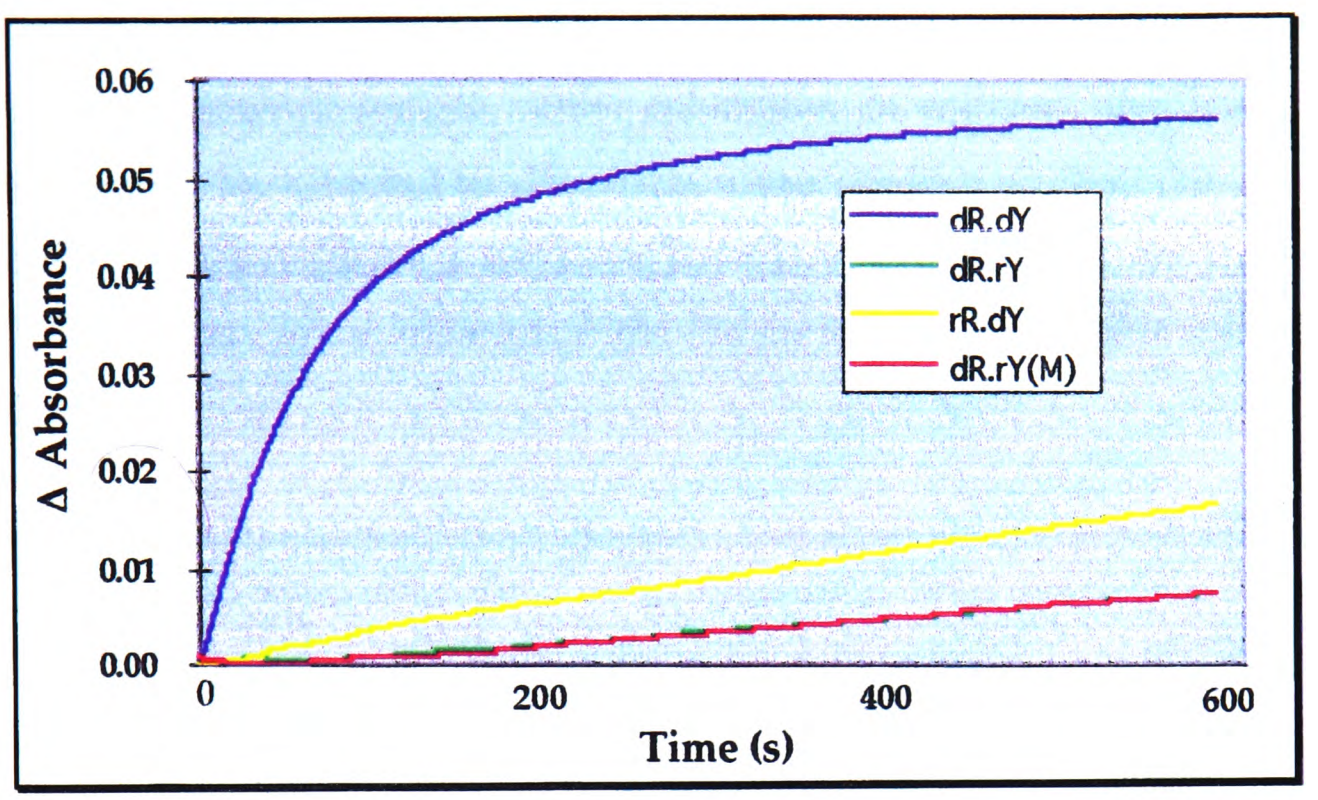


Figure 5.11 DNase I activity monitored using absorbance at 260nm.

The differences in solution conformation of the two different DNA.RNA may also affect the rate of activity of RNase H. From the models based on our NMR data, it is apparent that the backbone conformation and minor groove widths of the two classes of hybrid differ considerably. It is possible that with such extreme cases of conformational variability that the activity of the enzyme RNase H could be sensitive to the changes. As antisense action is often stated to rely on RNase H mediated cleavage of mRNA, knowledge of differential rates on the two classes of hybrid will be important to the application of DNA oligonucleotides as antisense agents. Qualitative assays were performed in a similar manner to those with DNase I. The absorbance curves obtained were somewhat unusual but clearly indicate that the enzyme is active on each hybrid (Figure 5.12). The rate of increase in absorbance appears to be significantly higher for the more stable hybrid rR.dY. However, no firm conclusions can be made

regarding the relative rates of *bond cleavage* as the hyperchromic effect relies on the dissociation of the cleaved strands rather than the bond cleavage event. The rate of dissociation of products is likely to be quite different for the two hybrids, given their different stabilities. It is possible that cleaved fragments of the stable hybrid rR.dY do not dissociate rapidly and might therefore remain viable substrates. In contrast, the less stable hybrid would be expected to dissociate more readily, so that the cleaved product no longer forms a substrate for the enzyme. Thus, the rate and overall increase in absorbance might be greater for the stable hybrid (as observed).

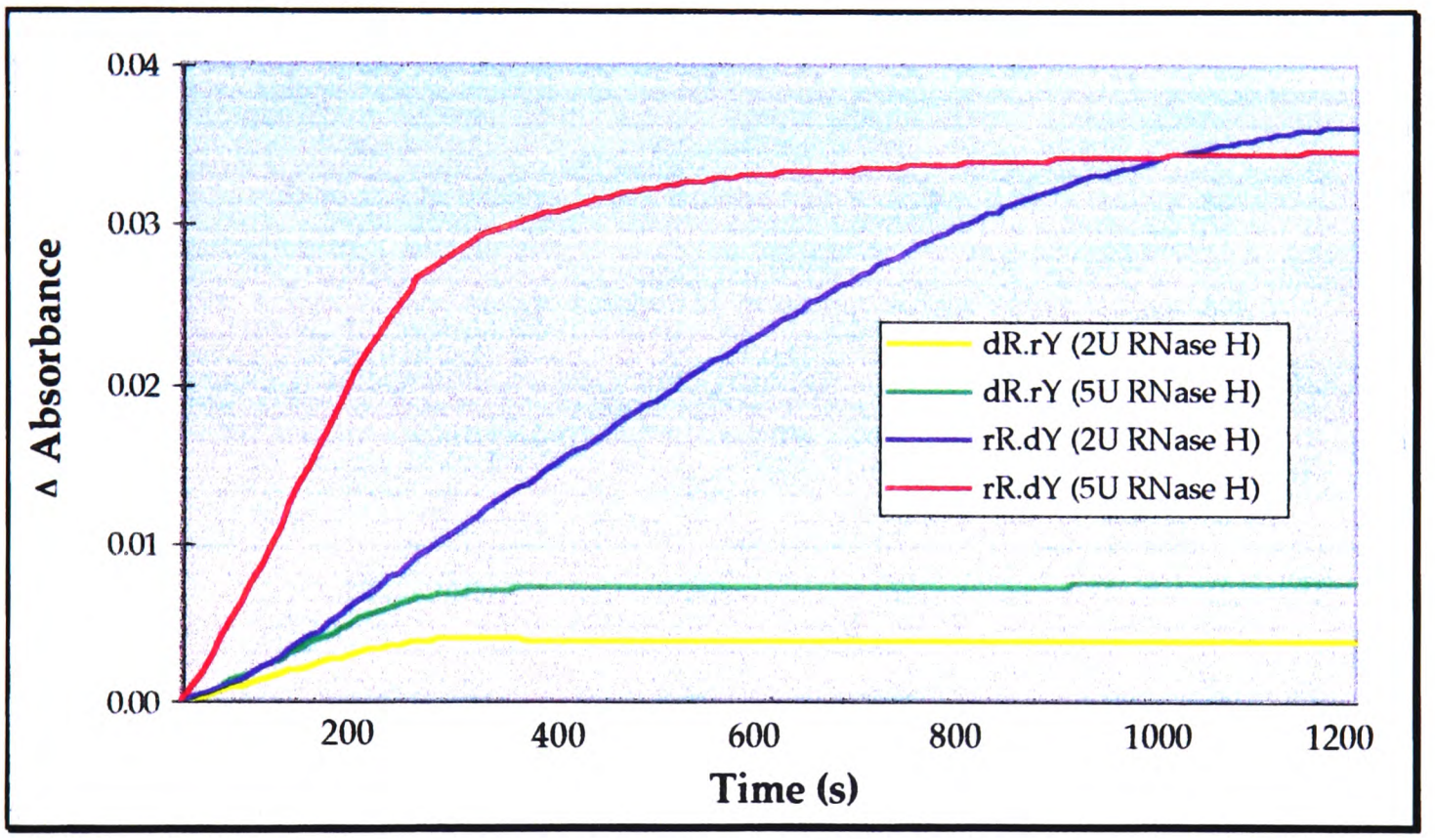


Figure 5.12 RNase H activity monitored using absorbance at 260 nm, at two concentrations of enzyme and 1.25 μM duplex (U = units of enzyme).

To properly assess the effect of conformational differences on RNase H activity it is necessary to obtain quantitative information on its activity against the two classes of hybrid. Unfortunately, it is not possible to use the spectrophotometric technique used for DNase I [51], as the indicator dye Cresol Red is not stable to the buffer conditions required. However,

we are currently investigating the use of alternative indicators, such as *p*-nitrophenol, for this purpose.

5.2.4 Structural Studies by X-ray Crystallography

X-ray crystallographic studies of the DNA.RNA hybrids are now well underway. Data has been collected from crystals of the native hybrid rR.dY and the structure is currently being refined. Large crystals of the two modified duplexes rR.dY(U) and rR.dY(P) have been obtained and will be used for diffraction in the near future.

5.3 Conclusions

Thermodynamic stability of DNA.RNA hybrids was found to be highly dependent on the base composition of the strands. The order of stability observed was rR.rY > rR.dY > dR.dY > dR.rY. The global solution conformations of the hybrids, while more A- than B-like were found to be considerably different by CD, electrophoresis and NMR. The equivalent pure RNA and DNA duplexes were found to be the expected A- and B-form respectively. Electrophoretic mobility, apparent correlation times and ^{31}P NMR shift dispersions were all consistent with a ranking of conformations from dR.dY (most B-like) through dR.rY and rR.dY to rR.rY (most A-like). Preliminary models constructed from the NMR data clearly showed the differences in global conformation.

Within the strands, the RNA sugars adopted the expected C3′-*endo* pucker. For the DNA sugars, a two-state equilibrium between C3′- and C2′-*endo* was found to fit the experimental data best. The fraction of DNA sugars in the C2′-endo pucker was greatest for the pure duplex and considerably lower for the hybrids. Glycosidic torsion angles were always found to be *anti*.

Comparisons of the individual hybrid strands with their equivalent in the pure duplexes, showed that the major conformational rearrangements always occur in the deoxyribose strands. Furthermore, the conformational changes are not the same in the two hybrids. The DNA strands were also found to have a much higher degree of flexibility in the hybrids than the pure DNA duplex. The differences at a local level translate into differences in base stacking, as observed by CD, sequential NOE intensities and backbone conformation. The difference in base stacking may, in part, explain the observed thermodynamic stability differences.

The addition of methyl groups to the pyrimidine RNA strand and propyne groups to the pyrimidine DNA strands was found to increase helix stability. In both cases the results suggest that the extra stability is due to an increase in base stacking interactions. The contribution of entropic factors was also suggested but is considerably less clear. Removal of methyl groups from thymidines in the pyrimidine DNA strand was found to have little effect on stability under the solution conditions used. Significant changes in gel mobility and CD spectra as a result of these modifications were observed. This is also consistent with considerable differences in base stacking on adding methyl and propyne groups. The NMR chemical shift differences suggest that major conformational changes only occur with the addition of propyne groups, with wide ranging perturbations observed in both grooves and on both strands. The methylation of the RNA strand of the hybrid dR.rY has little effect on the conformation. The stabilising effect must therefore arise primarily through increased stacking and possibly hydration effects. The combination of solution and crystal state studies will place us in the best possible position to explain the origins of the effects that methyl and propyne groups exert on helix stability. As well as increasing our fundamental understanding of nucleic acid conformation we may be able to begin to understand the complex role that the ordered hydration plays in nucleic acid conformation and stability.

The extra stability gained by the placement of pyrimidine nucleotides in the DNA strand combined with the effect of groups such as C5-propyne is also of considerable relevance in antisense technology. With increasing knowledge of the structure and stability of these hybrids, oligonucleotide therapeutics can be designed to maximise their specificity and pharmacological application.

5.4 References

1. Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987). *The Molecular Biology of the Gene*, 4th Edition, Chapter 10, The Benjamin/ Cummings Publishing Company Ltd, California.

2. Hansen, U.M. and McClure, W.R. (1980). *J. Biol. Chem.*, **255**, 20, 9564-9570.

3. Varmus, H. (1988). *Science*, **240**, 1427-1435.

4. Stein, H. and Hausen, P. (1969). *Science*, **166**, 393-395.

5. Oda, Y., Iwai, S., Ohtsuka, E., Ishikawa, M., Ikehara, M. and Nakamura, H. (1993). *Nucleic Acids Res.*, **21**, 4690-4695.

6. Lane, A.N., Ebel, S. and Brown, T. (1993). *Eur. J. Biochem.*, **215**, 297-306.

7. Fedoroff, O.Y., Salazar, M. and Reid, B.R. (1993). *J. Mol. Biol.*, **233**, 509-523.

8. Egli, M., Usman, N. and Rich, A. (1993). *Biochemistry*, **32**, 3221-3237.

9. Salazar, M., Champoux, J.J. and Reid, B.R. (1993). *Biochemistry*, **32**, 739-744.

10. Hawkins, J.W. and Nellen, W. (1994). *Antisense Res. Dev.*, **4**, 127-129.

11. Stein, C.A. and Cheng, Y.-C. (1993). *Science*, **261**, 1004-1011.

12. Bennett, M.R. (1995). *J. Drug Dev. Clin. Pract*, **7**, 225-235.

13. Miller, P.S. (1996). *Prog. Nucleic Acid Res. Mol. Biol.*, **52**, 261-291.

14. MacKellar, C., Graham, D., Will, D.W., Burgess, S. and Brown, T. (1992). *Nucleic Acids Res.*, **20**, 3411-3417.

15. De Mesmaeker, A., Haner, R., Martin, P. and Moser, H.E. (1995). *Acc. Chem. Res.*, **28**, 366-374.

16. Lewin, B. (1990). *Genes*, Oxford University Press, New York.

17. Woolf, T.M., Melton, D.A., and Jennings, C.G.B. (1992). *Proc. Natl Acad. Sci., USA*, **89**, 7305-7309.

18. Chiang, M.Y., Chan, H., Zounes, M.A., Freier, S.M., Lima, W.F. and Bennett, C.F. (1991). *J. Biol. Chem.*, **266**, 18162-18171.

19. Abe, J., Zhou, W., Taguchi, J., Takuwa, N., Miki, K., Ozaka, H., Kurokawa, K., Kamuda, M. and Takuwa, Y. (1994). *Biochem. Biophys. Res. Comm.*, **198**, 16-24.

20. Stein, C.A., Subasinghe, C., Shinozuka, K. and Cohen, J.S. (1988). *Nucleic Acids Res.*, **16**, 3209-3221.

21. Froehler, B.C., Wadwani, S., Terhorst, T.J. and Gerrard, S.R. (1992). *Tetrahedron Lett.*, **33**, 5307-5310.

22. Wagner, R.W., Matteucci, M.D., Grant, D., Huang, T. and Froehler, B.C. (1996). *Nature Biotechnology*, **14**, 840-844.

23. Ivanov, V.I., Minchenkova, L.E., Minyat, E.E., Frank-Kamenetskii, M.D. and Schyolkina, A.K. (1974). *J. Mol. Biol.* **87**, 817-833.

24. Leslie, A.G.W., Arnott, S., Chandrasekaran, R. and Ratliff, R.L (1980). *J. Mol. Biol.*, **143**, 49-72.

25. Hall, K.B. (1993). *Curr. Opin. Struct. Biol.*, **3**, 336-339.

26. Wang, A.H.-J., Fujii, S., van Boom, J.H., van der Marel, G.A., van Boeckel, S.A.A., and Rich, A. (1982). *Nature*, **299**, 601-604.

27. Mellema, J.-R., Haasnoot, C.A.G., van der Marel, G.A., Wille, G., van Boeckel, S.A.A., van Boom, J.H. and Altona, C. (1983). *Nucleic Acids Res.*, **11**, 5717-5738.

28. Egli, M., Usman, N. and Rich, A. (1993). *Biochemisry*, **32,** 3221-3237.

29. Ban, C., Ramakrishnan, B. and Sundaralingam, M. (1994). *J. Mol. Biol.*, **236**, 275-285.

30. Nishizaki, T., Iwai, S., Ohkubo, T., Kojima, C., Nakamura, H., Kyogoku, Y. and Ohtsuka, E. (1996). *Biochemistry*, **35**, 4016-4025.

31. Chou, S.-H., Flynn, P. and Reid, B. (1989). *Biochemistry*, **28**, 2435-2443.

32. González, C., Stec, W., Reynolds, M.A. and James, T.L. (1995). *Biochemistry*, **34**, 4969-4982.

33. González, C., Stec, W., Kobylanska, A., Hogrefe, R.I., Reynolds, M.A. and James, T.L. (1994). *Biochemistry*, **33**, 11062-11072.

34. Hall, K.B. and McLaughlin, L.W. (1991). *Biochemistry*, **30**, 10606-10613.

35. Ratmeyer, L., Vinayak, R., Zhong, Y.Y., Zon, G. and Wilson, W.D. (1994). *Biochemistry*, **33**, 5298-5304.

36. Searle, M.S. and Williams, D.H. (1993). *Nucleic Acids Res.*, **21**, 2051-2056.

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

38. Chamberlin, M.J. and Patterson, D. (1965). *J. Mol. Biol.* **12**, 410-428.

39. Riley, M., Maling, B. and Chamberlin, M.J. (1966). *J. Mol. Biol.* **20**, 359-389.

40. Lesnik, E.A. and Freier, S.M. (1995). *Biochemistry*, **34**, 10807-10815.

41. Hung, S.-H., Yu, Qin, Gray, D.M. and Ratliff, R.L. (1994). *Nucleic Acids Res.*, **22**, 4326-4334.

42. Wang, S. and Kool, E.T. (1995). *Biochemistry*, **34**, 4125-4132.

43. Sowers, L.C., Shaw, B.S. and Sedwick, W.D. (1987). *Biochem. Biophys. Res. Comm.*, **148**, 790-794.

44. Gutierrez, A.J., Terhorst, T.J., Matteucci, M.D. and Froehler, B.C. (1994). *J. Am. Chem. Soc.*, **116**, 5540-5544.

45. Puglisi, J.D. and Tinoco, I. Jr. (1989). *Meth. Enzymol.*, **180**, 304-325.

46. Gyi, J., Conn, G.L., Lane, A.N. and Brown, T. (1996). *Biochemistry*, **35**, 38, 12538-12548.

47. Gorenstein, D.G. (1992). *Meth. Enzymol.*, **211**, 254-286.

48. van de Ven, F.J.M. and Hilbers, C.W. (1988). *Eur. J. Biochem.*, **178**, 1-38.

49. Conte, M.R., Bauer, C.J. and Lane, A.N. (1996). *J. Biomol. NMR*, **7**, 190-206.

50. Lane, A.N. (1990). *Biochem. Biophys. Acta*, **1049**, 189-204.

51. Sutton, D., Conn, G.L., Brown, T. and Lane, A.N. (1996). *Biochemical Journal*, in press.

Chapter 6

Conclusions

6. Conclusions

6.1 General Discussion

The overall aim of this project was to develop RNA synthesis and purification protocols to provide a source of highly pure material for biophysical studies. Once implemented these protocols allowed detailed structural and thermodynamic studies on RNA and RNA.DNA duplexes by UV melting, X-ray crystallography and NMR.

6.1.1 Oligoribonucleotide Synthesis

Until recently, the preparation of large amounts of pure RNA for biophysical studies was a significant problem. Enzymatic methods were, and to some degree still are, commonly used for NMR and crystallographic studies. The main problem with such methods is that they do not allow the specific introduction of modified nucleotides or labels. As a result, there has been considerable interest in developing chemical synthesis methods based on those of DNA. The main point of debate has been the choice of protecting group for the 2'-hydroxyl. Two groups, Fpmp and tBDMS, have proven most successful and were available commercially at the commencement of this project. Both methods were used in RNA synthesis for this work and were found to be suitable for the preparation of short to moderate length oligoribonucleotides. However, problems such as the build up of side reaction products and difficulty in removing the protecting group completely were encountered with the Fpmp method for longer sequences. Removal of the tBDMS group was improved considerably using TEA.3HF rather than the traditional TBAF reagent. The best reagent

for deprotection was found to be a mixture of NMP/TEA/TEA.3HF (3:1.5:2) as proposed by Wincott *et al.*. This allowed the complete removal of the tBDMS groups within 30-90 minutes. The ease of handling and deprotection, combined with the significantly higher quality of product for longer sequences, make the 2'-O-tBDMS protecting group strategy the most suitable for chemical RNA synthesis presently available.

PAGE purification of RNA was abandoned in favour of HPLC methods early in this work. A combination of reversed phase and anion exchange HPLC as an orthogonal purification system was examined. Anion exchange HPLC was used to remove short deletion sequences generated by the capping step during chemical synthesis. Separation is achieved on the basis of charge, which correlates directly to the number of phosphate groups and hence the length of the RNA. With reversed phase HPLC, resolution is obtained through differences in hydrophobicity, and allows the removal of any incompletely deprotected species. The removal of tBDMS groups with TEA.3HF was found to be clean enough to allow single-step anion exchange purification of RNA samples. However, the reverse phase step was still included to remove the high concentration of buffer salts generated. The success of the synthesis and purification protocols developed was validated in a number of studies by high field NMR and by the growth of single crystals for diffraction.

6.1.2 Crystallization of Oligonucleotides

A wide variety of sequences were crystallized during the course of this work, among them: RNA duplexes, RNA hairpins, DNA.RNA chimeric and hybrid duplexes and DNA duplexes with modified nucleotides. It was found that variation of sequence was as useful as variation of crystallization conditions in the growth of high quality single crystals. By adding an overhanging 3'-G nucleotide to the target duplex

$r(CGCAAAUUUGCG)_2$ large single crystals were obtained that diffracted to a resolution of better than 2Å. The same conditions were used to grow crystals of a brominated derivative for multiple isomorphous replacement.

The use of a crystallization conditions screen was also examined. A critical factor in the success of the screen was found to be the concentration of oligonucleotide. For most sequences, an increase in concentration improved the number and quality of crystals obtained. However, in the case of a 21 mer hairpin, a lower concentration of RNA was found to be most suitable. The conditions screen provided many small crystals of a number of sequences which may prove useful, with optimisation of conditions, in diffraction experiments. Diffraction quality crystals of the DNA.RNA hybrid r(GAAGAGAAGC).d(GCTTCTCTTC) with thymidine, deoxyuridine and 5-propyne deoxyuridine in the DNA strand were produced almost immediately. Data was collected to 1.7Å resolution on the native hybrid duplex and is currently being used for a structural determination.

6.1.3 Sequence Dependent Structure in RNA

Runs of A-T base pairs are known to be responsible for unusual structural properties in DNA and are important in a number of biological functions. In the orientation A_nT_n, the duplexes are known to be thermodynamically stabilised due to the presence of an ordered spine of hydration and additional hydrogen bonds. A-T tracts are also sites for the specific binding of many minor groove binding drugs. The thermodynamic stability of a number of RNA duplexes with central A-U regions was compared with mixed sequence RNA and the equivalent DNA duplexes.

For RNA duplexes, the order of stability was found to be: Rmix > rA_3T_3 ~ $r(AU)_3$ > rA_3U_3 > rU_3A_3. In comparison, for DNA duplexes of the same sequence the order was: dA_3T_3 > $d(AT)_3$ ~ dMix > dT_3A_3. The A_3T_3 motif is stabilising in DNA, but in RNA A_3U_3 it is destabilising. The large positive change in the enthalpy component of ΔG° for this RNA is consistent with unusually poor base stacking. The introduction of methyl groups to the RNA, giving $r(A_3T_3)$, greatly increased the stability of the duplex (3.4 $kJmol^{-1}$ per base pair). In the opposite orientation, the destabilisation was even greater for DNA and RNA, attributed in both cases to very poor base stacking interactions. The effect of counter ion concentration on duplex stability was examined for each duplex. The slope dTm/ $d$log[Na^+] was generally highest for the less stable RNA duplexes compared to the randomised sequence. The DNA duplexes exhibited a lower sensitivity to salt and Mg^{2+} ion concentration than the RNA. These differences are consistent with the differences in the global conformations of the helical forms. Changes in counterion condensation and hydration are certain to play a major role in the varying duplex stability observed in the different sequence dependent structures. The general extra sensitivity of the RNA duplexes also suggests that the 2'-OH may be directly involved in ion coordination.

The binding of a number of ligands was examined by UV melting. In 0.1M NaCl, the increase in Tm of DNA A_3T_3 on distamycin A binding was twice the increase observed with the randomised sequence. No increase in Tm was observed for any RNA duplex with distamycin A, or a number of other groove binders and intercalators. This suggests that these molecules have little or no affinity for the RNA which is conformationally very different from that of DNA.

The hydration of the RNA duplex $(CGCAAAUUUGCG)_2$ was examined by NMR. Both grooves of the RNA were found to be clearly in contact

with water. Water protons were observed sufficiently close to H1' for direct NOE to be observed, which are not present for the equivalent DNA duplex. These relatively long residence times are most probably due to hydrogen bonding. Ordered spines of hydration, as observed in the DNA sequence, were not found in the RNA. The results suggested that slow exchange kinetics may, in general, be determined by hydrogen bonding rather than conformational factors.

Detailed structural calculations from NMR data to explain the lower thermodynamic stability are now underway. The initial data suggest that this duplex deviates substantially from the canonical A-form helix. The minor groove appears considerably narrowed in the central A-U in an underwound helix. This is consistent with poorer base stacking as observed in the UV melting analysis. Structural studies have also been conducted on this RNA by X-ray crystallography. Thus, detailed comparisons of the structure and hydration in the solution and crystal states will soon be possible.

6.1.4 DNA RNA Hybrids

DNA.RNA hybrids are important intermediates in DNA replication, transcription and in the retroviral life cycle. They also form the basis of the antisense therapeutic strategy. It is known that the stability of DNA.RNA hybrids can vary with their composition. For antisense technology, it is desirable to maximise the hybrid stability as the concentration of target RNA in cells is low.

Two DNA.RNA hybrids with purine and pyrimidine rich strands were compared to the equivalent RNA and DNA parent duplexes. The hybrid with purine rich RNA (rR.dY) was found to be considerably more stable than the opposite hybrid with pyrimidines in the RNA strand (dR.rY).

This gave the overall order of stability as: rR.rY > rR.dY > dR.dY > dR.rY. The stability of hybrid rR.dY was unaffected by the removal of the methyl groups in the DNA strand but stabilised dramatically by their replacement with 5-propyne. A large increase in the enthalpic contribution to ΔG° observed is consistent with much improved base stacking. This is certain to involve direct interaction of the unsaturated propyne in the π–π stacking of the bases. The effects of conformational restriction and hydration are also indicated by changes in the entropy of duplex formation. However, these considerations are considerably more complex, and will require high resolution analysis to interpret conclusively.

Analysis by CD and PAGE showed the native hybrids to be conformationally distinct from both the parent duplexes and each other. Differences in base stacking as observed in CD spectra may account in part for the differences in thermodynamic stability. Analysis of NMR data showed the two classes of hybrid to be overall more A- than B-like, but different from each other both globally and at a nucleotide level. The major differences were found to occur in the DNA strands. Spin coupling data indicated a higher degree of flexibility in the DNA strands of the hybrids, with a greater effect for the deoxypyrimidine strand. ^{1}H chemical shift changes, in comparison with the homoduplexes, were also much greater for the DNA strands and different for each hybrid. In the propyne modified hybrid, wide ranging changes in the shifts were observed in both strands. In contrast, addition of methyl groups to the pyrimidine DNA strand has considerably less effect, while methyl groups added to the RNA cause no conformational effects at all. Detailed analysis of the conformational flexibility at the nucleotide level and the contribution that this makes to duplex stability requires a full three dimensional structure determination. Such an analysis is currently underway in our laboratory.

6.2 Future Work

At present, very few RNA crystal structures are known. The work described here and in the recent literature suggests that many more will soon be available. We have collected high resolution diffraction data on pure RNA and DNA.RNA hybrid duplexes. Refinement of these structures is currently in progress. Conditions for the crystallization of a number of other RNAs and DNA.RNA duplexes are available from this work. Some of these require optimisation but should produce crystals large enough for X-ray diffraction studies. These will provide a considerable addition to the growing bank of RNA structural information. Furthermore, it will be possible for us to address in detail questions such as how the methyl and propyne groups affect duplex hydration, stability and conformation.

In the two areas of main interest in this project, a number of further experiments are possible. In the application of DNA oligonucleotides as antisense agents, modifications need to be made to afford cellular stability. The effects of modifications to the backbone, such as methyl phosphonate or phosphorothioate DNA, on the different classes of DNA.RNA hybrid are being examined at present. It has become clear that the backbone conformation and/ or groove widths of the two hybrids are significantly different. A full examination of the conformational dependence of RNase H activity is required. If differential activity is observed, as suggested here, this will provide an additional facet to the design of optimal antisense agents.

The relative effects of conformation, counterion condensation and hydration in the different stabilities of RNA and DNA duplexes deserve to be examined carefully. We have examined the solution phase hydration and conformation of the RNA duplex r(CGCAAAUUUGCG)$_2$.

The results of these studies have already suggested reasons behind the low thermal stability observed. The opposite orientation of the bases resulted in even lower stability, while alternating A-U base pairs had little effect. The structural reasons behind these considerable differences deserve a fuller understanding.

Chapter 7

Experimental

7. Experimental

7.1 Preparation of Synthetic Oligonucleotides

All oligonucleotides were synthesised on an Applied Biosystems 394 solid phase DNA/ RNA synthesiser, on a 1.0μmol scale. The 5′-O-(4,4′-dimethoxytrityl)-2′-O-(t-butyldimethylsilyl) nucleoside-3′-O-(2-cyanoethyl-N,N-diisopropyl) phosphoramidites, 1.0μmol RNA polystyrene supports, DNA phosphoramidites, DNA solid supports and ancillary reagents were purchased from Applied Biosystems. 5′-O-(4,4′-dimethoxytrityl)-2′-O-[1-(2-fluorophenyl)-4-methoxypiperidin-4-yl) nucleoside-3′-O-(2-cyanoethyl-N,N-diisopropyl) phosphoramidites were purchased from Cruachem Ltd. 5-Bromouridine and 5-methyluridine phosphoramidites were purchased from Glen Research and Chemgenes respectively. Chemicals used for deprotection of the oligonucleotides were purchased from Aldrich, unless otherwise stated.

7.1.1 Solid Phase Chemical Synthesis of DNA

DNA oligonucleotides were synthesised using a standard assembly cycle for the 394 synthesiser. Following base deprotection in concentrated ammonia at 55°C for 5 hours, the oligonucleotide was evaporated to dryness. Purification was achieved using anion exchange and/ or reversed phase HPLC (section 7.1.6).

7.1.2 Solid Phase Chemical Synthesis of RNA

Phosphoramidites were dissolved in anhydrous acetonitrile, under argon, to give a concentration of 0.1M. The standard RNA assembly cycle,

according to Applied Biosystems, was used with the following changes [the complete cycle is given in Appendix A]:

- the coupling time was extended to 600s, with a 1s acetonitrile delivery followed by 1s reversed flush half way through.
- the acetonitrile wash at the start of the cycle was extended to 30s.
- the iodine delivery time was increased to 15s, with a 2s reversed flush and further delivery of iodine of 2s.
- an extra 'capping' step was added after oxidation.

For Fpmp protected RNA oligonucleotides, the 5'-O-DMT group was left intact. In the case of the silyl RNA oligonucleotides the terminal trityl was removed as normal during the final synthesis cycle.

The stepwise and overall yield for the syntheses were monitored by an automatic trityl assay (based on conductivity). These 'trityl monitors' were found to vary considerably (in extremes up to 10%) with, in general, no obvious difference in synthesis quality as monitored by HPLC. These figures were therefore regarded as only a guide to the performance of the synthesiser and phosphoramidites.

7.1.3 Cleavage From the Solid Support and Base Deprotection

The cleavage of the oligonucleotide chain and the removal of the base protecting groups was achieved using one of two alternative methods:

a) Cleavage from the solid support was completed on the synthesiser with the standard Applied Biosystems RNA end-procedure using concentrated NH_3/ ethanol (3:1). Base protecting groups were removed by heating the solution to 55°C for four hours (Applied Biosystems 2'-O-tBDMS phosphoramidites) or 12 hours (2'-O-Fpmp phosphoramidites).

b) The solid support, with the bound fully protected oligonucleotide, was transferred to a 3ml screw cap vial and freshly prepared saturated ammoniacal methanol (1.5-2ml) added. The resulting mixture was heated at 55°C for 8-12 hours to complete both cleavage from the support and base deprotection. The solid support was removed by filtration, washed twice with 1.0ml EtOH/MeCN/water (3:1:1) and these washings added to the solution of RNA in a fresh 3ml vial.

In both cases, the samples were dried in a GeneVac evaporator. Silyl protected RNA oligonucleotides were then ready for 2'-OH deprotection (section 7.1.4). In the case of Fpmp protected RNA oligonucleotides, a trityl-on reversed phase HPLC purification step (section 7.1.6.1) was possible before deprotection (section 7.1.5).

7.1.4 Silyl Deprotection Procedures

The removal of the *tertiary*-butyldimethylsilyl (tBDMS) group can be accomplished in several ways, all involving the attack of a fluoride ion on the silicon.

7.1.4.1 TBAF, 1.0M in THF

1.0M TBAF in THF (1.0ml) was added to the dried oligonucleotide and left for at least 24 hours at room temperature. After this time the reaction was quenched by the addition of 1.0M ammonium acetate solution (1ml) and the mixture evaporated to dryness. Following desalting on sephadex, dowex AG50W-AX cation exchange resin (Biorad) was used to remove tBDMS salts. [Dowex columns were prepared by thoroughly washing a NAP-10 sephadex column casing with sterile water and adding dowex resin (approximately 2g). The resin was supplied as the Na^+ form ready for use. Columns were regenerated by washing with 0.1M HCl, sterile water,

0.1M NaOH (Na^+ form) or 0.1M NH_4OH solution (NH_4^+ form), and finally sterile water once again. At each step the pH of the eluant was monitored with pH paper.] The oligonucleotide was eluted from the dowex resin with sterile water (10ml), and dried ready for HPLC purification.

7.1.4.2 TEA.3HF/ 10%DMSO

The dried RNA was first dissolved in DMSO (0.1ml) and TEA.3HF (0.9ml) added. Deprotection was completed in 12-16 hours at room temperature. Sterile water (1ml) was added to the deprotection solution causing it to become cloudy. Precipitation of the RNA was encouraged by cooling the solution to -20°C for approximately two hours. The samples were centrifuged for 10 minutes at 13000rpm in a benchtop centrifuge internally cooled to 4°C. The supernatant was decanted, the pellet washed quickly with a small volume of sterile water and then redissolved in 1ml of water or 50mM tris (pH7.4) solution (anion exchange HPLC buffer). The supernatant was checked for RNA oligonucleotide content by absorbance measurements at 260nm. If a significant amount of RNA was present, the deprotection mixture was removed either by gel filtration on NAP 10 sephadex (pre-equilibrated with 15ml buffer A) or by direct application to the anion exchange column. For direct application to the column, the solution first had to be carefully brought to around neutral pH using 5N NaOH.

7.1.4.3 TEA.3HF/ NMP/ TEA [1]

A mixture of *N*-methylpyrrolidinone (3ml), TEA.3HF (2ml) and TEA (1.5ml) was prepared and sealed under argon. The TEA.3HF/ NMP/ TEA solution (0.25ml) was added to the dried oligonucleotide in a 3ml screw cap vial and vortexed to completely dissolve the RNA (often some heating was required to aid dissolution). The resulting mixture was heated to 65°C for 30-90 minutes and allowed to cool to room temperature. The deprotected RNA was transferred to a microfuge tube

and precipitated from solution by adding 3M NaOAc (25µl) followed by *n*-BuOH (1ml). The resulting mixture was vortexed briefly and then cooled to -70°C for at least 1 hour. The RNA was recovered as a pellet by centrifuging for 30 minutes at 13 000rpm in a benchtop centrifuge internally cooled to 4°C. The supernatant was decanted and the pellet washed with 70% ethanol solution then dried, ready for HPLC purification.

7.1.5 Fpmp Deprotection

Fpmp protected RNA purified by an initial trityl-on reversed phase HPLC (section 7.1.6.1) was lyophilised and then desalted by gel filtration (NAP 10 sephadex column). The resulting RNA solution was then dried to a white powder in a 3ml screw cap vial and acid deblock buffer (Cruachem Ltd) added according to Table 7.1. This cloudy mixture was sonicated briefly and then vortexed for 20-36 hours at room temperature. This procedure removes both the 2'-O-Fpmp groups and the terminal trityl group.

RNA Length (bp)	Vol. Acid Buffer (µl)	Vol. Neutralising Buffer (µl)
2-12	500	100
13-25	1000	200
26+	1500	300

Table 7.1 Volumes of Buffers used during Fpmp deprotection according to sequence length.

7.1.6 HPLC Purification

Two Gilson HPLC systems were available for purification of DNA and RNA oligonucleotides. Each consisted of two Gilson 306 pumps, an 805

manomeric unit and an 811C dynamic mixer controlled by Gilson 712 HPLC software on a PC. Elution of oligonucleotides was monitored using variable wavelength UV detectors and recorded on both a chart recorder and the PC. Buffers were prepared in 1 litre bottles, autoclaved and filtered prior to use. Analytical injections of small aliquots of oligonucleotide solution were performed on both systems with the detector set at 270nm, 0.1 AUFS. Preparative injections were monitored at 290-295nm, 1.0 AUFS and oligonucleotides collected manually.

7.1.6.1 Trityl-on Reversed Phase HPLC (Fpmp method)

Trityl-on reversed phase HPLC (RP HPLC) was used exclusively for Fpmp protected RNA oligonucleotides. The terminal 5'-O-trityl group, left intact in the final synthesis cycle, gives the full length RNA oligonucleotide a greatly increased hydrophobic interaction with the reversed phase media. This allows almost complete removal of the capped deletion sequences. The terminal trityl group is then removed during the normal deprotection of the 2'-O-Fpmp groups (section 7.1.5).

RP HPLC was performed on a Gilson system fitted with an Applied Biosystems Brownlee™ Aquapore octyl reversed phase cartridge column (250x10mm). The base deprotected oligonucleotide solution was evaporated to near dryness and buffer A1 (0.75ml) added. The resulting solution was then immediately injected onto the reversed phase HPLC column (as prolonged exposure to slightly acidic or even neutral conditions can cause significant detritylation). The buffer system is shown below and a typical gradient in Table 7.2.

Buffer A1 0.1M ammonium acetate, pH 7.4
Buffer B1 70% acetonitrile, 30% buffer A1

Time (minutes)	% Buffer B1	Flow (ml/min)
0	0	5
2	0	5
3	25	5
27	65	5
28	100	5
30	100	5
31	0	5
32.9	0	5
33	0	0

Table 7.2 A typical gradient used for trityl-on reversed phase HPLC (Fpmp method)

7.1.6.2 Trityl-off Reversed Phase HPLC (Fpmp method)

Reversed phase HPLC was used as a second purification step following 2'-O-Fpmp deprotection. This was done using the Gilson system as described and with the same buffer system as for trityl-on purification (section 7.1.6.1). The gradient used is shown in Table 7.3.

Time (minutes)	% Buffer B1	Flow (ml/min)
0	0	5
3.0	0	5
19.0	25	5
20.0	100	5
26.0	100	5
27.0	0	5
29.9	0	5
30.0	0	0

Table 7.3 Gradient for trityl-off reversed phase HPLC purification (Fpmp method)

7.1.6.3 Anion Exchange HPLC (tBDMS method)

Ion exchange HPLC was performed on a Gilson system fitted with a Dionex Nucleopac PA-100 (9 x 250mm) strong anion exchange column. A column oven was used to heat the column, generally to 55° but to a maximum of 90°C. Dionex anion exchange HPLC was used to purify

crude oligonucleotides following 2'-O-tBDMS deprotection. The formation of stable secondary structures on the column was overcome, where necessary, by the addition of urea to the sample (to 8M) and heating to 90°C for 2 minutes immediately before injecting. The duration of the run and the exact gradient depended very much on the sequence and length of oligonucleotide. For less than 25 nucleotides, a 20 minute gradient (Table 7.4) was generally sufficient. A 40 or 60 minute gradient with a similar profile was used in all other cases. Samples were injected in 0.75ml buffer A2. The buffer system used was:

Buffer A2 0.1M NH_4Cl, 50mM Tris pH 7.4
Buffer B2 1.0M NH_4Cl, 50mM Tris pH 7.4

Time (minutes)	% Buffer B2	Flow (ml/min)
0	0	5
1.5	0	5
2.0	20	5
15.0	50	5
15.5	100	5
17.0	100	5
18.0	0	5
19.9	0	5
20.0	0	0

Table 7.4 A typical gradient used for Dionex anion exchange HPLC

7.1.6.4 Trityl-off Reversed Phase HPLC (tBDMS method)

Reversed phase HPLC was used as a second purification step following Dionex anion exchange HPLC. This was essentially the same as the other RP HPLC protocols but using a lower concentration of acetonitrile, with the gradient shown in Table 7.5:

Buffer A3 0.1M ammonium acetate, pH 7.4 (=A1)

Buffer B3 20% acetonitrile, 80% buffer A3

Time (minutes)	% Buffer B2	Flow (ml/min)
0	0	5
1.5	0	5
2.0	20	5
15.0	50	5
15.5	100	5
17.0	100	5
18.0	0	5
19.9	0	5
20.0	0	0

Table 7.5 A typical gradient for RP HPLC (tBDMS method)

7.2 Analysis of Oligonucleotides

7.2.1 Anion Exchange HPLC

Dionex anion exchange HPLC (section 7.1.6.3) was used to examine the degree of purity of the full length product. Approximately 0.2 OD_{260} oligonucleotide were re-injected using the same profile as for purification, with the absorbance detector set at 270nm, 0.1 AUFS. The percentage area under the main peak was determined automatically by the Gilson HPLC software.

7.2.2 Capillary Electrophoresis

Capillary electrophoresis was performed on an Applied Biosystems 270A capillary electrophoresis system linked to a Spectra Physics ChromJet Integrator. Micro-gel_{100} capillaries and buffer (75mM Tris-phosphate, 10% methanol, pH 7.6) were purchased from Applied Biosystems. Oligonucleotides were loaded onto the capillary automatically from dilute salt-free aqueous solution. The actual loading time was varied (2-10s) depending on the concentration of the oligonucleotide solution (generally 2-5µl of stock solution in 0.5ml water).

7.2.3 Denaturing Polyacrylamide Gel Electrophoresis (PAGE)

Analytical PAGE was performed using a Hoefer Mighty Small II mini gel kit. Oligonucleotide samples were loaded onto 20% polyacrylamide gels containing 5% bisacrylamide, 7M urea and 1x TBE (89mM Tris, 89mM boric acid, 2mM EDTA). Loading mixtures contained either formamide (90%) or glycerol (50%) and either one or both of the marker dyes

bromophenol blue and xanthene cyanol. Gels were run at 10-15mA for approximately 2 hours.

7.2.4 Base Composition Analysis

The procedure used for base composition analysis was based on the method of Eadie *et al.* [2]. Enzymes phosphodiesterase I (type VII: from Crotalus atrox venom) and calf intestine alkaline phosphatase were purchased from Sigma and Boehringer Mannheim respectively. The reaction mixture consisted of 0.5-1.0 OD_{260} oligonucleotide mixed with phosphodiesterase I (0.02U), alkaline phosphatase (5U), Tris pH 7.5 (2µl, 1M) and $MgCl_2$ (2µl, 0.5M) in a total volume of 78.2µl (ie, 32mM Tris and 16mM $MgCl_2$). The mixture was incubated at 37°C for at least 12 hours.

The protein was then removed: sodium acetate, pH 5.0 (10µl, 2.5M) and 95% ethanol (234µl) were added successively and the mixture vortexed briefly and chilled to -70°C for 30 minutes. After centrifugation at 13000rpm for 15 minutes at 4°C the supernatant containing the isolated nucleosides, was decanted, diluted to 1ml with 95% ethanol, chilled to -70°C for a further 30 minutes and respun. The resulting supernatant was decanted, taken to dryness in a GeneVac evaporator and then redissolved in water (1ml) for HPLC analysis.

HPLC analysis of the RNA digest was performed on the reversed phase HPLC (as described in section 7.1.6) using the buffer system:

A4	50mM Potassium phosphate
B4	50% A4 + 50% Methanol

A flow rate of 1ml/min was used, with a linear gradient from 0 to 40% buffer B4. Elution times on this gradient were confirmed using nucleoside standards purchased from Sigma. Peak areas were generated automatically by the Gilson HPLC software.

7.2.5 Calculation of Oligonucleotide Extinction Coefficients

Extinction coefficients for single strand and duplex oligonucleotides were determined by Nuclease P1 digestion. As nuclease P1 acts considerably more slowly on duplex than single strand, each duplex digest sample was held at 70°C (the enzymes temperature optimum) for at least 10 minutes. However, in all cases initial and final absorbance measurements were made at 25°C.

Nuclease P1 (2µl, 1mg/ml solution) was added to the oligonucleotide (generally around 0.5 OD units) in 50mM sodium acetate pH 5.6 in a 1ml UV cell. For very unstable duplexes, NaCl and $MgCl_2$ were also added to the buffer to concentrations of 0.1M and 2-10mM respectively. The increase in absorbance on digestion was monitored and extinction coefficients calculated from known nucleotide values [3].

7.3 UV Thermal Melting

UV melting experiments were performed on a Perkin Elmer Lambda 2 UV/ Vis spectrometer, with a PTP-1 temperature programmer. A heating rate of 1° per minute was used (lower heating rates did not affect the Tm measured to within +/- 0.5°). Absorbance measurements were made at 260nm with data captured on a PC at an interval of 10s. Melting curves were smoothed and first derivatives obtained using the Perkin Elmer PECSS2 software. Conversion from time to temperature of the Tm values was done manually. All experiments were performed in 10mM sodium phosphate buffer, pH 7.0, containing 1mM EDTA and 0.1 or 1.0M NaCl unless otherwise stated. Melts were repeated to obtain reproducible results (to within 0.5°).

7.3.1 Duplex Concentration Dependence Studies

Thermodynamic parameters for duplex stability were determined from duplex concentration dependent melting temperature analysis. Using 1mm and 10mm path length cells (Hellma) a range of 50-100 fold duplex concentration was attainable.

For the 10mm pathlength cell (1 or 2.5 ml volume) the highest concentration sample (approximately 1.7-2.0 absorbance at 260nm) was prepared by mixing the appropriate amounts of stocks: 20mM sodium phosphate buffer (2x) and 4M NaCl, sterile water and the two single stranded oligonucleotides. The amounts of single strand used were determined from the absorbance of the oligonucleotide stock and the observed molar extinction coefficient. Lower concentrations were obtained by serial dilution of the sample with 1 x buffer with the appropriate salt concentration.

For the 1mm path length cell, aliquots of each single strand were lyophilised from water and redissolved in 1x buffer (phosphate/ NaCl) to give a sufficiently high concentration. A 100µl stock of the highest concentration of duplex (approximately 1.0 absorbance at 260nm) was prepared for each sample. Before use, the cell was washed three times with filtered 1x buffer (50µl) and twice with the oligonucleotide sample (10µl). The cell was then filled with the oligonucleotide sample (10µl) and any air bubbles carefully removed.

7.3.2 Salt Concentration Ranges

The effect of salt concentration on duplex melting temperature was measured using a 10mm pathlength cell (1ml volume) at a single duplex concentration. Samples were prepared as described above, but with a fresh sample for each salt concentration.

7.3.3 Metal Ion and Drug Effects

The effects of various ions and nucleic acid binding drugs on duplex stability were determined by UV melting. Samples were prepared in 10mm path length cells (1ml volume) as described above. Single melting temperatures, with no extra metal ion or drug, were determined and checked against the previously determined value. Small volumes of concentrated stock solutions of the metal ions (in water) and drugs (in methanol) were then added. To avoid dilution effects the volume added was always less than 5% of the total sample volume. The melts were then repeated under otherwise identical conditions.

7.4 Qualitative Enzyme Activity Assays

7.4.1 DNase I Assays

DNase I was purchased from Boehringer Mannheim. UV absorbance was measured using a Perkin Elmer Lambda 15 UV/Vis spectrometer, with samples in 1ml UV cuvettes held at 25°C. Single strand oligonucleotides were added to a solution containing 0.2M KCl, 1mM $CaCl_2$, 5mM $MgSO_4$ and $NaHCO_3$ at pH8.2, to give duplex concentrations of approximately 3μM. After 5 minutes, 10μl of DNase I solution (0.62mg/ml stock) was added and the absorbance at 260nm monitored for 10 minutes.

7.4.2 RNase H Assays

The activity of the enzyme RNase H (Amersham) was monitored essentially as described for DNase I (section 7.4.1). Assays were performed at 37°C in a solution containing 10mM Mg^{2+}, 25mM NaCl, 0.1mM EDTA 0.1M dithiothreitol, 20mM Tris pH8.0 and 1.5μM oligonucleotide duplex. The reaction was initiated by the addition of RNase H (0.5U/μl).

7.5 Oligonucleotide Crystallization

7.5.1 Sitting Drop Method

Glass plates, with three sets of three wells connected by small grooves, were siliconised by soaking in trimethylsilyl chloride 5% (v/v) in DCM for 20 minutes. The plates were washed with acetone and de-ionised sterile water, and oven dried overnight before use. Lyophilised oligonucleotides were dissolved in sodium cacodylate buffer to a concentration of 0.5 OD_{260}/µl and 5µl of this solution used in each 20µl drop. This gives a final concentration for RNA/DNA dodecamer duplexes of approximately 0.75mM duplex. The crystallization trials were performed at various concentrations of magnesium ion and precipitant. The exact conditions required depend greatly on the sequence, but the approximate initial ranges used are summarised in Table 7.6.

Solution	Stock concentration (mM)	Drop concentration (mM)
Cacodylate	50	12.5
$MgCl_2$	250	6 - 60
Spermine.4HCl	10	0.5 - 4.5
50% aq. MPD	N/A	2.5 - 15%

Table 7.6 Range of concentrations used in sitting drop crystallization trials

7.5.2 Sparse Matrix Screen by a Hanging Drop Method

Chemicals used were of the highest grade available (generally Molecular Biology grade, RNase free), and 5-10x stock solutions were prepared with sterile water. A 48 solution sparse matrix [4] was prepared by mixing the appropriate volumes of these stocks, solids and water to a final volume of 10ml. (This gave the concentrations of buffer, precipitants and salts indicated in Table 7.7). The crystallization conditions were screened in 4µl

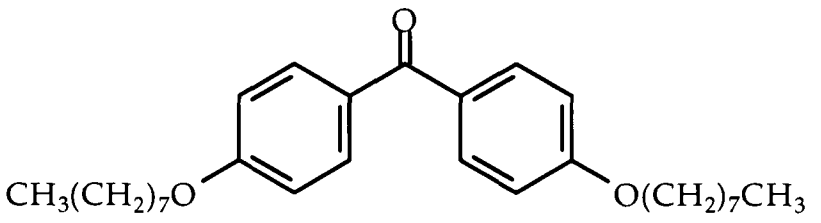
pH, 50mM buffer	Precipitating Agent	Salts and additives
1. 6.0, Cacodylate	5% MPD	40mM $MgCl_2$
2. 5.6, Mes	5% PEG 8000	200mM KCl + 10mM $MgCl_2$
3. 6.0, Cacodylate	1.0M Li_2SO_4	10mM $MgCl_2$
4. 6.0, Mes	15% Isopropanol	20mM $MgCl_2$
5. 7.0, Hepes	1.6M $(NH_4)_2SO_4$	10mM $MgCl_2$ + 0.2mM $Co(NH_3)_6Cl_3$
6. 5.6, Mes	10% PEG 400	10mM $MgSO_4$ + 200mM KCl
7. 6.0, Cacodylate	1.7M $(NH_4)_2SO_4$	15mM $Mg(OAc)_2$
8. 7.5, Tris	5% Isopropanol	10mM $MgCl_2$
9. 6.5, Cacodylate	10% PEG 4000	200mM KCl + 10mM $MgCl_2$
10. 7.0, Hepes	15% MPD	5mM $MgSO_4$
11. 6.0, Mes	5% PEG 4000	5mM $MgSO_4$
12. 5.6, Mes	20% MPD	100mM $Mg(OAc)_2$
13. 6.5, Cacodylate	15% PEG 400	80mM $Mg(OAc)_2$ + 0.2mM $Co(NH_3)_6Cl_3$
14. 7.5, Tris	10% PEG 4000	200mM KCl + 50mM $MgCl_2$
15. 7.5, Tris	1.6M $(NH_4)_2SO_4$	10mM $MnCl_2$
16. 7.0, Hepes	5% PEG 8000	20mM $MgCl_2$
17. 6.5, Cacodylate	10% Isopropanol	150mM $Mg(OAc)_2$ + 0.2mM $Co(NH_3)_6Cl_3$
18. 6.5, Cacodylate	1.3M Li_2SO_4	10mM $Mg(OAc)_2$ + 0.2mM $Co(NH_3)_6Cl_3$
19. 7.5, Tris	10% MPD	10mM $MgCl_2$
20. 5.6, Mes	1.1M Li_2SO_4	10mM $MgCl_2$
21. 7.0, Hepes	5% PEG 400	100mM KCl + 10mM $MgCl_2$
22. 7.0, Hepes	5% PEG 4000	200mM NH_4OAc + 150mM $Mg(OAc)_2$
23. 6.5, Cacodylate	10% PEG 8000	100mM $Mg(OAc)_2$ + 200mM KCl
24. 6.0, Mes	10% PEG 400	100mM KCl + 10mM $MgCl_2$
25. 7.0, Hepes	25% Dioxane	5mM $MgCl_2$
26. 6.0, Cacodylate	15% Isopropanol	5mM $MgCl_2$ + 2mM $CoCl_2$
27. 7.5, Tris	10% Dioxane	5mM $MgCl_2$
28. 5.6, Mes	20% PEG 8000	10mM $MgCl_2$ + 100mM $(NH_4)_2SO_4$
29. 6.0, Cacodylate	10% PEG 4000	200mM KCl + 10mM $CaCl_2$
30. 7.0, Hepes	10% PEG 400	200mM KCl + 10mM $CdSO_4$
31. 6.5, Cacodylate	10% PEG 4000	200mM NH_4OAc+ 10mM $CaCl_2$
32. 7.5, Tris	5% PEG 8000	100mM KCl + 5mM $CdSO_4$
33. 6.0, Cacodylate	30% MPD	40mM $Mg(OAc)_2$
34. 7.0, Hepes	10% PEG 400	100mM KCl + 10mM $CaCl_2$
35. 7.0, Hepes	20% Hexane diol	200mM KCl + 10mM $MgCl_2$
36. 7.5, Tris	20% Hexane diol	5mM $CdSO_4$
37. 6.5, Cacodylate	10% Hexane diol	0.1mM $Co(NH_3)_6Cl_3$ + 5mM $MgCl_2$
38. 5.6, Mes	10% PEG 2000	200mM KCl + 10mM $MgCl_2$ + 0.2mM $Co(NH_3)_6Cl_3$
39. 8.0, Tris	5% PEG 2000	200mM KCl + 10mM $MgCl_2$
40. 6.0, Cacodylate	30% PEG 4000	200mM NH_4Cl+ 10mM $CaCl_2$
41. 6.5, Cacodylate	30% PEG 4000	80mM $Mg(OAc)_2$
42. 7.0, Hepes	30% Hexane diol	200mM NH_4Cl+ 10mM $MgCl_2$
43. 6.5, Cacodylate	30% PEG 8000	200mM NH_4OAc + 10mM $Mg(OAc)_2$
44. 8.5, Tris	30% PEG 400	100mM KCl + 10mM $MgCl_2$
45. 6.0, Cacodylate	1.8M Li_2SO_4	10mM $MgSO_4$
46. 7.0, Hepes	1.6M Li_2SO_4	50mM $MgSO_4$
47. 6.5, Cacodylate	2.0M $(NH_4)_2SO_4$	10mM $MgSO_4$
48. 8.5, Tris	1.8M $(NH_4)_2SO_4$	25mM $MgSO_4$

Table 7.7 Crystallization sparse matrix conditions

drops suspended on siliconised cover slips over a 0.75ml reservoir of the matrix solution in 24-well Linbro tissue culture plates. For the initial screening procedure, the drop was composed of 2µl of matrix solution mixed with 2µl of oligonucleotide (0.5-2mM) in a solution containing sodium cacodylate buffer (10mM, pH6.5) and spermine.4HCl (1mM). The plates were checked daily for crystal growth.

7.6 Synthesis of a Modified Trityl Group

4,4′-Dioctaoxybenzophenone (Method 1) [I]



4,4′-Dihydroxybenzophenone (17.14g, 0.08mol), 1-bromooctane (30.9g, 27.6ml, 0.16mol) and KI (trace) were added to a solution of sodium ethoxide (0.16mol) in ethanol. After refluxing for 10 hours the ethanol was removed *in vacuo* and the residue taken up in DCM/ 1.0M NaOH. The aqueous phase was extracted a number of times with DCM and the combined organic phases washed with water. After drying over Na_2SO_4, the organic solvent was removed *in vacuo*, and the product purified by wet flash column chromatography (DCM/ MeOH 0-10%), yielding a white solid (21.3g, 60%).

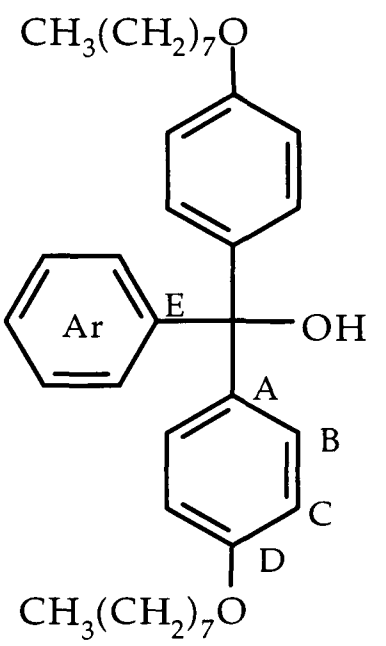
4,4′-Dioctaoxybenzophenone (Method 2) [I]

NaH (60% dispersion in mineral oil; 0.41g, 10.3mmol) was stirred in hexane (50ml) for 2 hours, after which time, the hexane was carefully removed with a pipette and DMF (20ml) added. The solution was cooled to 0°C on ice and 4,4′-dihydroxybenzophenone (1.0g, 4.7mmol) added. The resulting cloudy orange mixture was stirred for 2 hours and then allowed to warm to room temperature. 1-Bromooctane (3.6g, 3.2ml, 18.8mmol) and KI (trace) were then added and the mixture stirred for 12 hours. After this time, the reaction mixture was taken up in water/ DCM (1:1, 50ml) and the aqueous layer extracted with DCM (2 x 25ml). The combined organic fractions were then washed successively with water (25ml) and saturated brine (25ml) and dried over Na_2SO_4. The organic solvent was removed *in vacuo*, and the product purified by wet flash column chromatography (DCM/ MeOH 0-10%), yielding a white solid (1.7g, 83%).

^{1}H NMR (200MHz, $CDCl_3$); 0.88 (t, 6H, J=6.5Hz, 2 x $C\underline{H}_3$), 1.22-1.49 (m, 20H, 10 x $C\underline{H}_2$), 1.73-1.86 (quintet, 4H, J=6.7Hz, 2 x $OCH_2C\underline{H}_2$), 4.00 (t, 4H, J=6.5Hz, 2 x $OC\underline{H}_2$), 6.92 (d, 4H, J=8.8Hz, 4 x H_B), 7.75 (d, 4H, J=8.8Hz, 4 x H_C). ^{13}C NMR (50MHz, $CDCl_3$); 13.9 (2 x $\underline{C}H_3$), 22.5 ($\underline{C}H_2CH_3$), 25.8 ($\underline{C}H_2CH_2CH_3$), 29.0 ($\underline{C}H_2CH_2$), 29.2 ($\underline{C}H_2CH_2$), 31.6 ($OCH_2\underline{C}H_2$), 68.0 ($O\underline{C}H_2$), 113.7 (Ar C_B), 130.3 (Ar C_A), 132.0 (Ar C_C), 162.3 (Ar C_D), 194.3 (C=O). FAB MS m/z [$(M+H)^+$ $C_{29}H_{42}O_3$ requires 439.3211, found 439.32252.

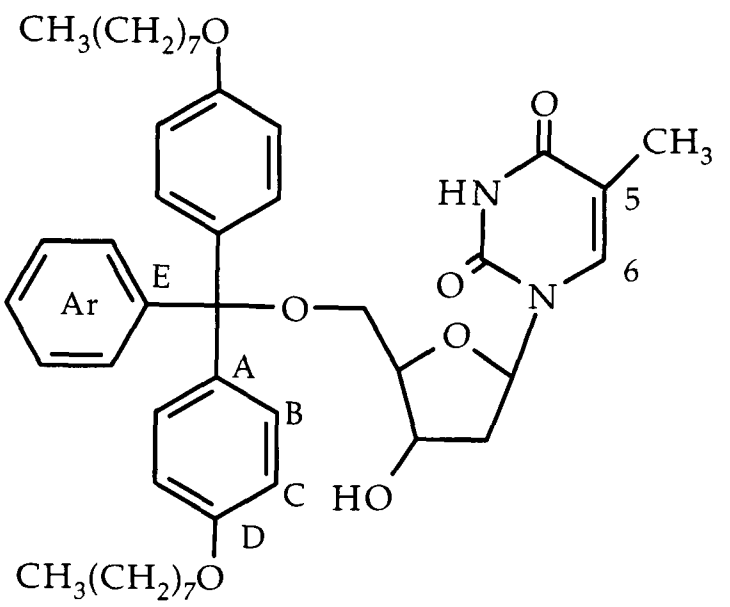
Dioctaoxytrityl alcohol [II]

A solution of 4,4′-dialkylated benzophenone [I] (4.4g, 0.01mol) in anhydrous THF (20ml) was dropped into a stirred solution of phenyl magnesium bromide (0.02mol) in ether. After spontaneous reflux had died down, the solution was heated to reflux for a further 2 hours. The reaction mixture was then allowed to cool to room temperature and poured onto ice. The resulting mixture was extracted twice with DCM and the combined organic fractions washed with water, $NaHCO_3$ solution (1M) and saturated KCl solution and finally dried over $MgSO_4$. Evaporation of the solvent gave an orange oil (3.9g, ~75%). ^{1}H NMR (200MHz, $CDCl_3$); 0.85 (t, 6H, J=6.5Hz, 2 x $C\underline{H}_3$), 1.18-1.42 (m, 20H, 10 x $C\underline{H}_2$), 1.70 (quintet, 4H, J=6.7Hz, 2 x $OCH_2C\underline{H}_2$), 3.85 (t, 4H, J=6.5Hz, 2 x $OC\underline{H}_2$), 6.74 (d, 4H, J=8.8Hz, 4 x H_B), 7.08 (d, 4H, J=8.8Hz, 4 x H_C), 7.20 (m, 5H, Ar-H). ^{13}C NMR (50MHz, $CDCl_3$); 14.3 (2 x $CH_2\underline{C}H_3$), 22.8 ($\underline{C}H_2CH_3$), 26.2 ($\underline{C}H_2CH_2CH_3$), 29.4-29.5 (3 x $\underline{C}H_2CH_2$), 32.0 ($OCH_2\underline{C}H_2$), 68.1 ($O\underline{C}H_2$), 81.6 ($Ph_3\underline{C}OH$), 113.8 (Ar C_B), 127.2-129.3 (5 x Ar C), 130.2 (Ar C_E), 139.4 (Ar C_A), 147.6 (Ar C_C), 158.4 (Ar C_D). FAB MS m/z [$(M+H)^+$ $C_{35}H_{48}O_3$ requires 516.3605, found 516.36100].



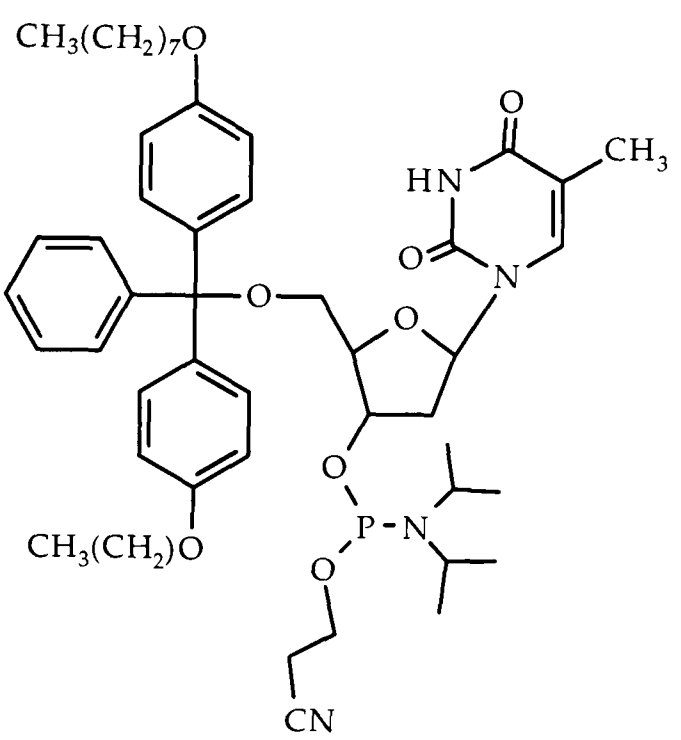
5′-O-Dioctaoxytrityl thymidine [III]

Dioctaoxytrityl alcohol [II] (3.9g, 7.6mmol) was coevaporated with anhydrous toluene (3x10ml) and the residue taken up in toluene (15ml) and acetyl chloride (10ml) added. The resulting mixture was refluxed for 4 hours after which time the solvent and volatile by-product were removed *in vacuo*.



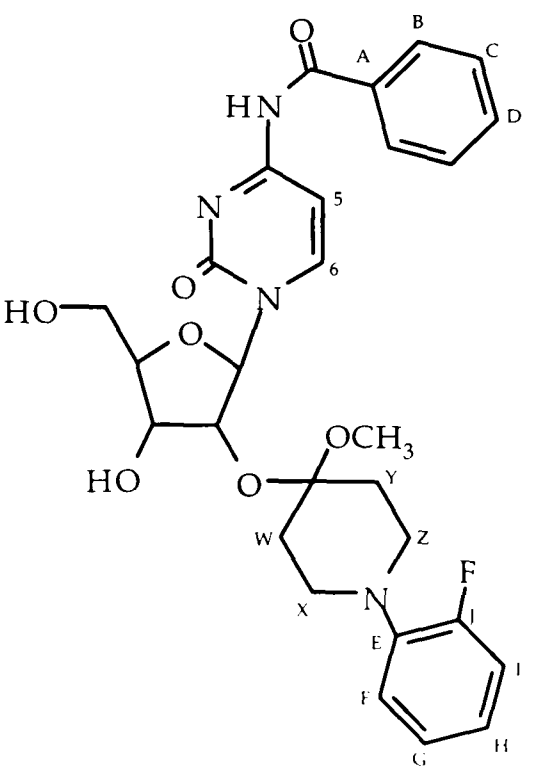
The residue was taken up in anhydrous pyridine (5ml) and thymidine (1.7g, 7mmol) and DMAP (10mg) added. The resulting mixture was stirred for 15 hours. After this time, the reaction was quenched by the addition of 5%NH_4HCO_3 (20ml)/ ethyl acetate (20ml) and the aqueous component extracted twice with ethyl acetate (20ml). Following washing with saturated KCl and drying over $MgSO_4$ the organic solvent was removed and the product purified by flash column chromatography (DCM/MeOH 0-10%), giving a pale yellow solid (3.78g, 68%). ^{1}H NMR (250MHz, $CDCl_3$); 0.87 (t, 6H, J=6.7Hz, 2 x C$\underline{H}_3$), 1.25-1.39 (m, 20H, 10 x C$\underline{H}_2$),1.44 (d, 3H, J=1.0Hz, CH_3), 1.70-1.81 (quintet, 4H, J=6.9Hz, 2 x OCH$_2$C$\underline{H}_2$), 2.30-2.43 (m, 2H, 2 x H-2′), 3.32-3.48 (m, 2H, H-5′$_{a,b}$), 3.91 (t, 4H, J=6.5Hz, 2 x OC$\underline{H}_2$), 4.08 (d, 1H, J=2.6Hz, H-4′), 4.55 (m, 1H, H-3′), 6.45 (m, 1H, H-1′), 6.80 (d, 4H, J=8.9Hz, 4 x H_B), 7.25 (d, 4H, J=8.7Hz, 4 x H_C), 7.25 (m, 5H, Ar-H), 7.61 (d, 1H, J=1.2Hz, H-6′). ^{13}C NMR (62.9MHz, $CDCl_3$); 11.6 (Thy-$\underline{C}H_3$), 13.9 (2 x $CH_2\underline{C}H_3$), 22.5 ($\underline{C}H_2CH_3$), 25.8 ($\underline{C}H_2CH_2CH_3$), 29.0-29.2 (3 x $\underline{C}H_2CH_2$), 31.6 (OCH$_2\underline{C}H_2$), 40.8 (C-2′), 63.5 (C-5′), 67.8 (O$\underline{C}H_2$), 72.4 (C-4′), 84.6 (Ph$_3\underline{C}$OH), 86.2 (C-3′), 86.8 (C-1′), 111.2 (Thy C-5), 113.6 (Ar C_B), 126.9-127.9 (Ar C), 129.9 (Ar C_C), 134.9 (Ar C_E), 135.6 (Thy C-6), 144.3 (Ar C_A), 150.5 (Thy C-4), 158.1 (Ar C_D), 163.9 (Thy C-2). FAB MS m/z [(M+H)$^+$ $C_{45}H_{60}O_7N_2$ requires 741.44788, found 741.44561].

5′-O-Dioctaoxytrityl -3′-O-cyanoethyl-N,N-diisopropyl-thymidine phosphoramidite [IV]



5′-O-DOTr protected thymidine [III] (1.95g, 2.6mmol) was coevaporated with anhydrous THF (3 x 15ml) and then taken up in THF (30ml). N,N-diisopropylethylamine (1.98g, 1.84ml) and 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (1.2 equivalents) were added and the mixture stirred for 30 minutes. Ethyl acetate (50ml) was added and the mixture washed with saturated KCl solution. Following evaporation of the organic solvent the product was purified by flash column chromatography (Hex/EtOAc 0-50%), yielding a white foam (1.18g, 48%). ^{31}P NMR (81.02MHz, $CDCl_3$), 149.1, 149.5 (2 x s, isomeric forms). Automated DNA synthesis average stepwise coupling 99.5% in synthesis of T35 test oligonucleotide.

2′-O-Fpmp-benzoyl cytosine [V]

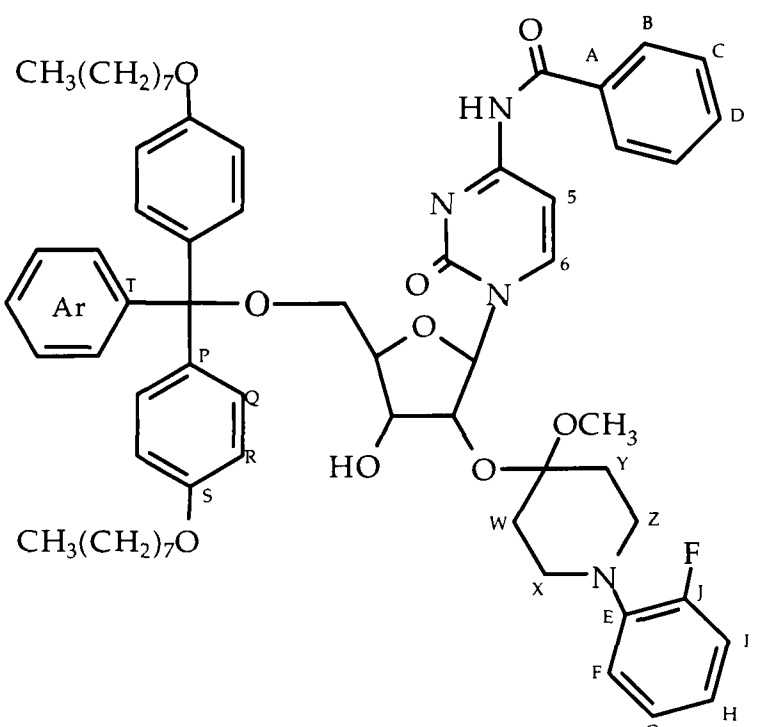


2′-O-Fpmp-benzoyl cytosine [V] was the kind gift of Cruachem Ltd, Glasgow, and was used without further purification. ^{1}H NMR (300MHz, DMSO), 2.69-2.93 (m, 4H, $CH_{2W,Y}$), 2.94 (s, 3H, $OC\underline{H}_3$), 2.94-3.20 (m, 4H, $CH_{2X,Z}$), 3.59-3.72 (m, 2H, H-5′$_{a,b}$), 3.95-3.99 (m, 1H, H-4′), 4.04-4.10 (m, 1H, H-3′), 4.45 (′t′, 1H, J=6Hz, H-2′), 5.10-5.50 (broad ′s′, ~2H, 3′-OH, 5′-OH), 6.16 (d, 1H, J=7.7Hz, H-1′),6.89-7.12 (m, 4H, Ar Fpmp), 7.29 (d, 1H, J=8.6Hz, Cyt H-5), 7.42-7.62 (m, 3H, $\underline{m}$-,$\underline{p}$-benzoyl),

8.00 (d, 2H, J=7.3Hz, o-benzoyl), 8.36 (d, 1H, J=8.6Hz, Cyt H-6). ^{13}C NMR (75.5MHz, DMSO), 32.3 ($CH_{2X/Z}$), 33.8 ($CH_{2X/Z}$), 47.1 (OCH_3), 47.7 ($CH_{2Y/W}$), 51.4 ($CH_{2Y/W}$), 61.5 (C-5′), 70.7 (C-3′), 73.4 (C-2′), 86.4 (C-4′), 87.0 (C-1′), 115.8 (Ar C_I), 116.0 (Ar C_F), 119.6 (Ar C_H), 122.2 (Cyt C-5), 124.8 (Ar C_G), 128.4 (Ar C_C), 128.6 (Ar C_B), 132.5 (Ar C_D), 134.1 (Ar C_E), 139.9 (Ar C_J), 140.0 (Ar C_A), 145.1 (Cyt C-6), 155.2 (Cyt C-2), 164.2 (Cyt C-4), 168.3 (PhCO).

5′-O- Dioctaoxytrityl-2′-O-Fpmp-benzoyl cytosine [VI]

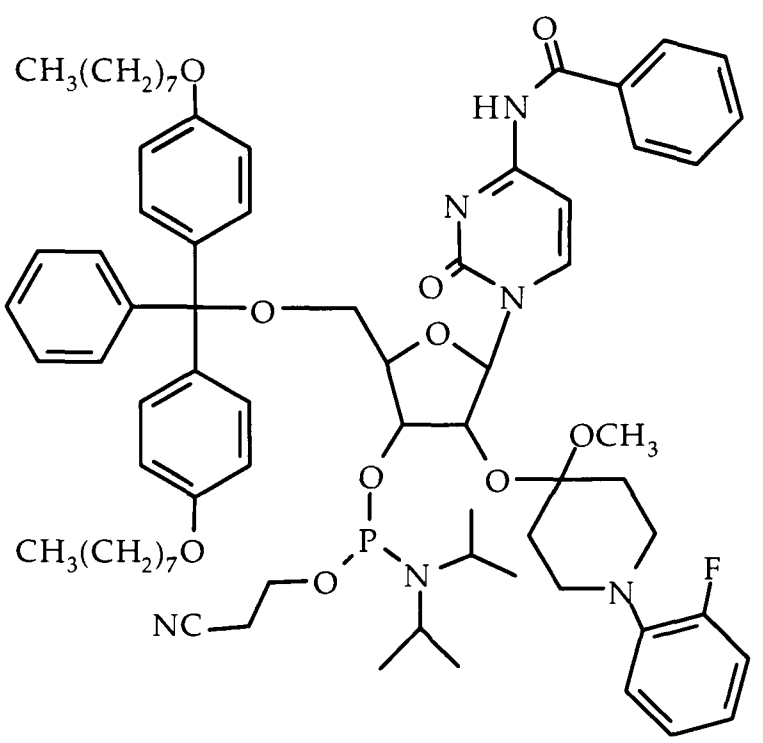
The same procedure as for the synthesis of [III] was employed with 2′-O-Fpmp-benzoyl cytosine [V] (3.32g, 6mmol). Volumes and amounts of other solvents and reagents were adjusted to scale. The product [VI] was obtained as a pale yellow solid (4.6g, 73%). ^{1}H NMR (300MHz, DMSO), 0.85 (t, 6H, J=6Hz, 2 x CH_2CH_3), 1.18-1.31 (m, 20H, 10 x CH_2), 1.31-1.43 (m, 4H, 2 x OCH_2CH_2), 1.66 (t, 4H, J=6.4Hz, OCH_2CH_2), 2.78-2.96 (m, 4H, $CH_{2W,Y}$), 3.08 (s, 3H, OCH_3), 3.10-3.42 (m, 4H, $CH_{2X,Z}$), 3.89-3.96 (m, 2H, H-5′$_{a,b}$), 4.11 (m, 1H, H-4′), 4.22 (m, 1H, H-3′), 4.55 (t, 1H, J=4.3Hz, H-2′), 5.35 (d, 1H, 3′-OH), 6.16 (d, 1H, J=4.7Hz, H-1′), 6.87 (d, 4H, J=8.6Hz, 4 x H_Q), 6.90-7.18 (m, 4H, Ar H Fpmp), 7.26 (d, 4H, J=8.6Hz, 4 x H_R), 7.28-7.65 (m, 9H, *m*-, *p*-benzoyl, Cyt H-5, 5 x Ar H), 8.01 (d, 2H, J=7.7Hz, *o*-benzoyl), 8.24 (d, 1H, J=8.6Hz, Cyt H-6). ^{13}C NMR (75.5MHz, DMSO), 14.0 (2 x CH_2CH_3), 22.2 (CH_2CH_3), 25.6 ($CH_2CH_2CH_3$), 28.7 (3 x $CH_2CH_2CH_2$), 31.3 (OCH_2CH_2), 47.5 (OCH_3), 63.5 (C-5′), 67.4 (OCH_2CH_2), 70.2 (C-3′), 73.4 (C-2′), 84.0 (C-4′), 86.2, (Ph_3CO), 87.4 (C-1′), 113.8 (C_Q), 115.8 (C_I),116.0 (C_F), 119.6 (C_H), 122.2 (Cyt C-5), 127.0-128.0 (5 x Ar C), 128.5 (Ar C_C), 128.6 (Ar C_B), 129.8 (Ar C_R), 132.5 (Ar C_D), 134.9 (Ar C_E/ Ar C_T), 139.4 (Ar C_P), 140.0 (Ar C_J/ Ar



C_A), 144.4 (Cyt C-6), 165.5 (Cyt C-2), 157.8 (Ar C_S), 164.4 (Cyt C-4) 168.3 (Ph<u>C</u>O). Nominal MS m/z $(M+H)^+$ $C_{63}H_{77}O_9N_4F$ requires 1053.6, found 1053.3 (awaiting FAB data).

5′-O- Dioctaoxytrityl-2′-O-Fpmp-benzoyl cytosine-3′-O- cyanoethyl-N,N-diisopropyl-phosphoramidite [VII]

The same procedure as for the synthesis of [IV] was employed with 5′-O- Dioctaoxytrityl-2′-O-Fpmp-benzoyl cytosine [VI] (3.32g, 6mmol), but using 1.4 equivalents of 2-cyanoethyl-N, N -diisopropylchlorophosphor-amidite and quenching the reaction with DCM (20ml). Volumes and amounts of other solvents and reagents were adjusted to scale. The product [VII] was obtained as a pale yellow foam (1.53g, 65%). ^{31}P NMR (121.5MHz, DMSO), 149.9, 151.6 (2 x s, isomeric forms). Automated RNA synthesis average stepwise coupling 97.0-99.0% in synthesis of poly-C oligoribonucleotides.



7.7 References

1. Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., and Usman, N. (1995). *Nucleic Acids Res.*, **23**, 2677-2684.

2. Eadie, J.S., McBride, L.J., Efcavitch, J.W., Hoff, L.B. and Cathcart, R. (1987). *Anal. Biochem.*, **165**, 442-447.

3. Fasman, G.D. (Editor) (1975). *Handbook of Biochemistry and Molecular Biology. Nucleic Acids Vol. I*, 3rd edition, CRC Press.

4. Scott, W.G., Finch, J.T., Grunfell, R., Fogg, J., Smith, T., Gait, M.J. and Klug, A. (1995). *J. Mol. Biol.*, **250**, 327-332.

Appendices

Appendix A: RNA Synthesis Cycle (for ABI 394).

STEP	FUNCTION NAME	NUM	TIME	ACTIVE	SAFE
1)	Begin	106			Yes
2)	18 to Waste	64	3.0		Yes
3)	18 to Column	42	30.0	AGCT5678	Yes
4)	Reverse Flush	2	10.0	AGCT5678	Yes
5)	Block Flush	1	4.0	AGCT5678	Yes
6)	Phos Prep	101	3.0		Yes
7)	Column 1 On	140			Yes
8)	Block Vent	111	2.0	AGCT5678	Yes
9)	Tet to Waste	58	1.7		Yes
10)	B+Tet to Column	33	2.5	AGCT5678	Yes
11)	Tet to Column	34	1.0	AGCT5678	Yes
12)	B+Tet to Column	33	3.5	AGCT5678	Yes
13)	Push to Column	43			Yes
14)	Column 1 Off	141			Yes
15)	Column 2 On	142			Yes
16)	18 to Waste	64	4.0		Yes
17)	Block Flush	1	3.0	AGCT5678	Yes
18)	Block Vent	111	2.0	AGCT5678	Yes
19)	Tet to Waste	58	1.7		Yes
20)	B+Tet to Column	33	2.5	AGCT5678	Yes
21)	Tet to Column	34	1.0	AGCT5678	Yes
22)	B+Tet to Column	33	3.5	AGCT5678	Yes
23)	Push to Column	43			Yes
24)	Column 2 Off	143			Yes
25)	Column 3 On	144			Yes
26)	18 to Waste	64	4.0		Yes
27)	Block Flush	1	3.0	AGCT5678	Yes
28)	Block Vent	111	2.0	AGCT5678	Yes
29)	Tet to Waste	58	1.7		Yes
30)	B+Tet to Column	33	2.5	AGCT5678	Yes
31)	Tet to Column	34	1.0	AGCT5678	Yes
32)	B+Tet to Column	33	3.5	AGCT5678	Yes
33)	Push to Column	43			Yes
34)	Column 3 Off	145			Yes
35)	Column 4 On	146			Yes
36)	18 to Waste	64	4.0		Yes
37)	Block Flush	1	3.0	AGCT5678	Yes
38)	Block Vent	111	2.0	AGCT5678	Yes
39)	Tet to Waste	58	1.7		Yes
40)	B+Tet to Column	33	2.5	AGCT5678	Yes
41)	Tet to Column	34	1.0	AGCT5678	Yes
42)	B+Tet to Column	33	3.5	AGCT5678	Yes
43)	Push to Column	43			Yes
44)	Column 4 Off	147			Yes
45)	Wait	103	300.0	AGCT5678	Yes
46)	18 to Waste	64	4.0		Yes
47)	18 to Column	42	1.0	AGCT5678	Yes
48)	Wait	103	5.0	AGCT5678	Yes
49)	Reverse Flush	2	1.0	AGCT5678	Yes
50)	Wait	103	300.0	AGCT5678	Yes
51)	Cap Prep	102	3.0		Yes
52)	18 to Waste	64	4.0		Yes
53)	Reverse Flush	2	7.0	AGCT5678	Yes
54)	Block Flush	1	3.0	AGCT5678	Yes
55)	Cap to Column	39	10.0	AGCT5678	Yes
56)	Wait	103	5.0	AGCT5678	Yes
57)	18 to Waste	64	4.0		Yes
58)	Reverse Flush	2	7.0	AGCT5678	Yes
59)	Block Flush	1	3.0	AGCT5678	Yes
60)	15 to Column	41	12.0	AGCT5678	Yes
61)	Reverse Flush	2	2.0	AGCT5678	Yes

62)	15 to Column	41	2.0	AGCT5678	Yes
63)	18 to Waste	64	4.0		Yes
64)	Block Flush	1	3.0	AGCT5678	Yes
65)	Wait	103	15.0	AGCT5678	Yes
66)	18 to Column	42	10.0	AGCT5678	Yes
67)	Flush to Waste	4	6.0	AGCT5678	Yes
68)	18 to Column	42	10.0	AGCT5678	Yes
69)	Reverse Flush	2	5.0	AGCT5678	Yes
70)	18 to Column	42	10.0	AGCT5678	Yes
71)	Reverse Flush	2	3.0	AGCT5678	Yes
72)	18 to Column	42	10.0	AGCT5678	Yes
73)	Reverse Flush	2	7.0	AGCT5678	Yes
74)	Block Flush	1	3.0	AGCT5678	Yes
75)	Cap Prep	102	3.0		Yes
76)	18 to Waste	64	4.0		Yes
77)	Reverse Flush	2	7.0	AGCT5678	Yes
78)	Block Flush	1	3.0	AGCT5678	Yes
79)	Cap to Column	39	10.0	AGCT5678	Yes
80)	Wait	103	5.0	AGCT5678	Yes
81)	18 to Waste	64	4.0		Yes
82)	Reverse Flush	2	7.0	AGCT5678	Yes
83)	Block Flush	1	3.0	AGCT5678	Yes
84)	Start Detrityl	105			Yes
85)	18 to Waste	64	4.0		Yes
86)	18 to Column	42	10.0	AGCT5678	Yes
87)	Reverse Flush	2	7.0	AGCT5678	Yes
88)	Block Flush	1	3.0	AGCT5678	Yes
89)	If Monitoring	167			Yes
90)	19 to Column	44	35.0	AGCT5678	Yes
91)	14 to Column	40	3.0	AGCT5678	Yes
92)	Monitor trityls	135			Yes
93)	14 to Column	40	60.0	AGCT5678	Yes
94)	Monitor noise	136			Yes
95)	14 to Column	40	10.0	AGCT5678	Yes
96)	Stop monitor	137			Yes
97)	18 to Column	42	10.0	AGCT5678	Yes
98)	Reverse Flush	2	8.0	AGCT5678	Yes
99)	If not Montring	168			Yes
100)	14 to Column	40	6.0	CT5678	No
101)	Trityl Flush	3	5.0	CT5678	No
102)	14 to Column	40	6.0	AGCT5678	No
103)	Wait	103	5.0	AGCT5678	No
104)	Trityl Flush	3	5.0	AGCT5678	No
105)	14 to Column	40	6.0	AGCT5678	No
106)	Wait	103	5.0	AGCT5678	No
107)	Trityl Flush	3	5.0	AGCT5678	No
108)	14 to Column	40	6.0	AGCT5678	No
109)	Wait	103	5.0	AGCT5678	No
110)	Trityl Flush	3	5.0	AGCT5678	No
111)	18 to Column	42	10.0	AGCT5678	No
112)	Trityl Flush	3	8.0	AGCT5678	No
113)	End Monitoring	169			Yes
114)	18 to Column	42	8.0	AGCT5678	Yes
115)	Reverse Flush	2	7.0	AGCT5678	Yes
116)	Block Flush	1	4.0	AGCT5678	Yes
117)	End	107			Yes

Appendix B: Oligonucleotide Sequences and Extinction Coefficients

(a) RNA oligonucleotides

Code	Oligonucleotide Sequence	Est. ε_{ss} (10^3Mcm)	Obs. ε_{ss} (10^3Mcm)
R02	GAAGAGAAGC	94.3	103.0
R03	CGUUCUCUUC	68.8	81.7
R04	AGCCCGCCUAAUGAGCGGGCU	179.5	nd
R09	GGAGCCCGCCUAAUGAGCGGGCU	213.0	nd
R13	GCTTCTCTTC	61.1	74.4
R14	r(CGCGAA)d(TTCGCG)	94.3	nd
R15	d(CGCGAA)r(UUCGCG)	97.2	nd
R16	GCGCAAAAGCGC	99.5	107.3
R17	GCGCUUUUGCGC	89.4	97.9

nd: not determined

(b) DNA Oligos

Code	Oligo Sequence	Est. ε_{ss} (10^3Mcm)	Obs. ε_{ss} (10^3Mcm)
D02	GAAGAGAAGC	94.3	89.3
D03	GCTTCTCTTC	61.1	77.0
D16	GCGCTTTTGCGC	83.6	91.2
D17	GCGCAAAAGCGC	99.5	107.7
D19	$GCT^PT^PCT^PCT^PT^PC$, P = 5-propyne	nd	51.2
D29	GCUUCUCUUC	68.8	81.9

nd: not determined (no extinction coefficient available for 5-propyne dU nucleotide)

(c) Duplexes

Code	Duplex Sequence	Obs. ε_{ds} (10^3Mcm)
D02.D03 [dR.dY]	d(GAAGAGAAGC.GCTTCTCTTC)	146
D02.R03 [dR.rY]	d(GAAGAGAAGC).r(GCUUCUCUC)	149
R02.D03 [rR.dY]	r(GAAGAGAAGC).d(GCTTCTCTTC)	154
R02.R03 [rR.rY]	r(GAAGAGAAGC.GCUUCUCUUC)	152
R02.D19 [rR.dY(P)]	r(GAAGAGAAGC).d(GCT^{P}T^{P}CT^{P}CT^{P}T^{P}C)	143
D02.D19 dR.dY(P)	d(GAAGAGAAGC. GCT^{P}T^{P}CT^{P}CT^{P}T^{P}C)	135
R02.R13 [rR.rY(M)]	r(GAAGAGAAGC.GCTTCTCTTC)	150
D02.R13 [dR.rY(M)]	d(GAAGAGAAGC).r(GCTTCTCTTC)	143
R02.D29 [rR.dY(U)]	r(GAAGAGAAGC).d(GCUUCUCUUC)	158
D02.D29 [dR.dY(U)]	d(GAAGAGAAGC.GCUUCUCUUC)	151
R06	$(CGCAAAUUUGCG)_2$	nd
R10	$(CGCAUAUAUGCG)_2$	176
R11	$(GCGUUUAAAGCG)_2$	178
R12	$(GACUGAUCAGUC)_2$	176
R18	$(CGCAAATTTGCG)_2$	179
D06	$(CGCAAATTTGCG)_2$	178
D10	$(CGCATATATGCG)_2$	182
D11	$(CGCTTTAAAGCG)_2$	183
D12	$(GACTGATCAGTC)_2$	181

nd: not determined (see section 3.2)

Appendix C: Tm Data for A-U/T Dodecamers (Chapter 4)

Set	RNA [R06, (A_3U_3)]		DNA [D06, (A_3T_3)]	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	145	345.4	143	344.6
2	108	344.5	92.5	343.9
3	72.0	343.4	60.6	342.3
4	24.2	341.8	22.1	341.2
5	17.5	341.2	16.0	340.3
6	11.6	340.3	10.4	339.3
7	7.80	339.3	7.08	338.5
8	5.50	337.8	4.37	337.0
9	2.60	335.8	2.95	335.7

Set	RNA [R10, $(AU)_3$]		DNA [D10, $(AT)_3$]	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	144	347.4	112	338.4
2	91.8	346.2	74.9	338.0
3	49.6	344.5	53.2	336.7
4	26.6	343.3	21.5	335.0
5	10.9	342.7	15.7	334.5
6	7.48	341.9	10.5	333.5
7	4.43	340.9	6.94	332.6
8	2.27	339.5		
9				

Set	RNA [R11, (U_3A_3)]		DNA [D11, (T_3A_3)]	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	104	344.3	106	338.4
2	76.4	343.4	69.8	336.8
3	46.1	342.0	39.2	335.0
4	22.2	340.2	19.5	333.7
5	16.1	339.0	13.9	332.4
6	9.80	337.7	8.94	331.0
7	5.96	336.3	6.00	329.8
8	3.65	335.0		
9	1.74	332.3		

Set	RNA [R12, mixed]		DNA [D12, mixed]	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	115	350.7	100	338.4
2	70.9	349.7	52.6	336.9
3	38.8	348.8	26.5	335.4
4	17.5	347.7	19.5	335.3
5	13.9	347.3	15.8	334.6
6	9.90	346.5	10.4	333.9
7	7.00	345.6	6.68	332.9
8	3.07	344.4	4.58	331.8
9	1.65	342.9	2.29	329.7

Set	RNA [R18, (A_3T_3)]	
	Ct (μM)	Tm (K)
1	149	351.5
2	88.9	350.4
3	55.9	349.5
4	36.4	348.1
5	19.0	347.8
6	13.6	347.1
7	9.33	346.3
8	6.37	345.6
9	3.92	344.6
10	2.48	343.0

Appendix D: Tm Data for DNA.RNA Hybrids (Chapter 5)

(a) Native duplex concentration range Tm.

No.	dR.dY		dR.rY	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	212	318.6	135	307.6
2	123	318.3	94.9	305.6
3	64.0	315.5	48.6	303.7
4	23.9	312.4	25.2	300.0
5	16.4	311.6	22.7	298.9
6	10.8	310.7	16.8	298.6
7	8.22	309.3	11.4	297.9
8	4.25	307.6	7.56	296.5
9	2.12	306.0	4.09	296.2
10			2.01	294.3

No.	rR.dY		rR.rY	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	158	327.3	217	334.4
2	71.4	325.5	114	332.5
3	39.6	324.2	39.6	330.2
4	22.3	322.7	22.6	329.7
5	15.6	322.3	16.7	328.7
6	10.4	321.1	11.1	327.6
7	6.88	320.0	7.42	326.9
8	3.77	318.7	4.09	324.7
9	2.06	317.0	2.26	323.8

(b) 5-Methyl uridine duplex concentration range Tm.

No.	rR.rY(M)		dR.rY(M)	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	132	339.7	136	315.0
2	81.5	338.5	87.4	312.7
3	38.1	337.3	39.2	311.9
4	23.0	335.4	26.1	309.7
5	16.4	334.6	17.7	308.7
6	10.4	333.7	10.4	307.3
7	6.53	332.8	5.60	305.3
8	2.68	331.3	2.52	303.3

(c) 5-Propyne deoxyuridine duplex concentration range Tm.

No.	rR.dY(P)		dR.dY(P)	
	Ct (μM)	Tm (K)	Ct (μM)	Tm (K)
1	138	344.0	139	331.1
2	108	342.4	110	330.6
3	65.7	341.7	45.2	327.8
4	25.4	339.9	27.6	326.1
5	19.5	339.3	20.8	325.4
6	14.8	338.7	16.0	324.4
7	10.0	337.9	10.1	323.2
8	5.96	336.7	5.16	321.5
9	3.27	335.2	3.26	320.0

(d) Deoxyuridine duplex concentration range Tm.

No.	rR.dY(U)		dR.dY(U)	
	Ct (mM)	Tm (K)	Ct (mM)	Tm (K)
1	109	325.3	122	315.0
2	63.3	324.3	75.5	314.0
3	39.2	322.6	52.8	312.8
4	20.8	321.7	22.3	310.6
5	15.0	321.2	17.7	310.0
6	11.0	320.5	11.9	308.6
7	8.18	319.6	9.00	307.8
8	5.18	318.8	5.64	306.8
9	2.76	317.2	3.24	304.8

Appendix E: Publications

Biochemistry **1996,** *35,* 12538−12548

Comparison of the Thermodynamic Stabilities and Solution Conformations of DNA·RNA Hybrids Containing Purine-Rich and Pyrimidine-Rich Strands with DNA and RNA Duplexes†

Jeffrey I. Gyi,‡ Graeme L. Conn,§ Andrew N. Lane,*,‡ and Tom Brown§

Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K., and Department of Chemistry, University of Southampton, Southampton SO17 1BJ, U.K.

Received April 19, 1996; Revised Manuscript Received July 12, 1996®

ABSTRACT: The conformations and thermodynamic stabilities of duplexes containing purine-rich (GAAGAGAAGC) and pyrimidine-rich (GCTTCTCTTC or GCUUCUCUUC) DNA and RNA strands have been measured by UV melting, electrophoresis, circular dichroism, and NMR spectroscopy. The free energies of stabilization ($-\Delta G$) were in the order rR·rY > rR·dY > dR·dY > dR·rY. The two DNA·RNA hybrid duplexes showed conformational properties intermediate between those of DNA·DNA and RNA·RNA duplexes and also different from one another. Differences between ^{1}H chemical shifts of the DNA strands in the two hybrid duplexes and those of the DNA duplex were larger than analogous shift differences for the RNA protons, and the differences were larger for the purine than the pyrimidines in the DNA strands. Detailed analysis of the nucleotide conformations using both NOE and scalar coupling data showed that the sugar conformations of the ribonucleotides are all near C3′-endo. The deoxyribonucleotides were in the "S" domain, i.e., near C2′-endo in the DNA duplex, and C1′-exo to C2′-endo in the two hybrids. However, the deoxyriboses in the two hybrids appear more flexible than in the DNA duplex, with the fraction in the "N" (C3′-endo) state increasing in the order dR·dY < dR·rY < rR·dY. Globally, the pure DNA duplex was B-form and the pure RNA duplex A form. The two DNA·RNA hybrids were neither A nor B, but closer globally to the A than the B form. The less stable dR·rY duplex has a significantly different conformation from rR·dY both at the local nucleotide level and globally.

Hybrid duplexes comprising a DNA and an RNA strand occur in several important biological processes. They are intermediates in transcription (Hanson & McClure, 1980), in DNA replication (Adams et al., 1986), and in the synthesis of retroviral cDNA by reverse transcription (Varmus, 1988). They are substrates for the ubiquitous RNAseH which catalyzes the hydrolysis of RNA only when it is present in an RNA·DNA hybrid duplex and not when it is part of an RNA·RNA duplex (Stein & Hausen, 1969). RNAse H shows little if any sequence specificity (Oda et al., 1993), and it has been recently suggested that the enzyme recognizes a difference in conformation between RNA·DNA hybrid and RNA·RNA duplexes (Fedoroff et al., 1993; Lane et al., 1993).

DNA·RNA hybrids are also exploited in antisense technology (Stein & Cheng, 1993). However, because the target RNA is likely to be present at low concentration, optimal thermodynamic stability of the proposed hybrids is desirable. It has been reported, however, that the thermodynamic stability of hybrid duplexes in which the DNA strand consists of purines and the RNA strand consists of pyrimidines (dR·rY) is much less than the corresponding duplex containing a purine RNA strand and a pyrimidine DNA strand (rR·dY) (Ratmeyer et al., 1994; Hung et al., 1994; Wang & Kool, 1995; Lesnik & Freier, 1995). These studies demonstrated that the thermodynamic stabilities of the four kinds of duplex are dependent on the base composition. Further, the measured free energies of stabilization agree reasonably well with nearest neighbor calculations (Breslauer et al., 1986, Freier et al., 1986, Sugimoto et al., 1995). The origin of the differences was recently rationalized in terms of the chemical differences arising from the 5-Me of T in the DNA and the 2′-OH of the riboses in RNA (Wang & Kool, 1995). The influence of different conformations was not addressed in this study. It has been shown that the solution conformation of DNA·RNA hybrids of mixed sequence is neither the A-form of typical of duplex RNA nor the B form typical of duplex DNA, though the overall conformation is closer to the A form than to the B form (Fedoroff et al., 1993; Lane et al., 1993; Gonzalez et al., 1995). Further, both electrophoretic properties and circular dichroism indicated that the rR·dY and dR·rY hybrids have different global conformations; rR·dY is more like RNA (i.e., A form) whereas dR·rY is intermediate between the DNA and RNA duplexes (Ratmeyer et al., 1994; Hung et al., 1994). Recent work has suggested that there may be a continuum of structures between the A and B forms depending on the composition of the strands (Lesnik & Freier, 1995).

To assess the conformational contribution to the thermodynamic stability of DNA·RNA hybrids of different base composition, it is necessary to determine their structures in solution, and compare them with their "parent" homodu-

† This work was supported by the Medical Research Council of the U.K. and a Royal Society of Edinburgh Caledonian Research Fellowship to G.C.

* Address correspondence to this author at Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K. Telephone: 0181 959 3666, ext. 2052. Fax: 0181 906 4477. E-mail: a-lane@nimr.mrc.ac.uk.

‡ National Institute for Medical Research.

§ University of Southampton.

® Abstract published in *Advance ACS Abstracts,* September 1, 1996.

S0006-2960(96)00948-8 CCC: $12.00

plexes. To this end, we have made four decamer duplexes containing a purine-rich and a pyrimidine-rich strand, and we have compared their properties using a combination of thermodynamics, electrophoresis, circular dichroism, and ^1H and ^{31}P NMR spectroscopy. In this article we report the differences in both local and global conformational features and dynamic aspects of the deoxynucleotide strands.

MATERIALS AND METHODS

Materials. d(GAAGAGAAGC) and d(GCTTCTCTTC) were synthesized and purified using standard methods as previously described (Brown & Brown, 1991, 1992). The analogous RNA strands r(GAAGAGAAGC) and r(GCUUCUCUUC) were synthesized on an ABI 349A automated synthesizer on a 1 μmol scale using 2′-*O*-*tert*-butylsilyl protected ribonucleotide phosphoramidites and derivatized polystyrene supports (Applied Biosystems). The 2′-OH protecting group was removed with triethylammonium fluoride in NMP for 30−60 min at 65 °C (Wincott et al., 1995). Oligonucleotides were cleaved from the solid support and the bases deprotected by treatment with concentrated ammonia/ethanol (3:1) at 55 °C for 4 h. The crude products were purified by reverse-phase HPLC (Brownlee Aquapore, 250 by 10 mm) using a gradient of 0−20% acetonitrile in 0.1 M ammonium acetate. Pooled fractions were desalted on Sephadex G10 and lyophilized. For brevity, we refer to these strands as dR10 and dY10 for the predominantly purine and pyrimidine DNA strands, respectively, and analogously rR10 and rY10 for the ribonucleotide strands.

Duplexes were formed by mixing equimolar concentrations of the appropriate strands as determined from their absorption coefficients. The duplexes were annealed, cooled on ice, and dialyzed at 4 °C against 10 mM sodium phosphate and 100 mM KCl, pH 7.0, using a 3.5 kDa cutoff membrane (Spectrapor), lyophilized, and redissolved in 0.6 mL of D_2O or 90% H_2O/10% D_2O for NMR analysis.

Methods. Melting temperatures, T_m, were determined as a function of concentration (range 70−100-fold) for each of the four possible duplexes on a Perkin-Elmer Lambda 2 UV/vis spectrophotometer using the PECSS2 software as previously described (Ebel et al., 1994). Unless otherwise stated, the melting curves were recorded in 10 mM sodium phosphate and 0.1 M NaCl, pH 7.0. Oligonucleotide concentrations were calculated from the absorbance and the measured absorption coefficients. Thermodynamic parameters were derived according to

$$1/T_m = (R/\Delta H)\ln(C_t/4) + \Delta S/\Delta H \quad (1)$$

where C_t is the concentration of stands, R is the gas constant, ΔH is the enthalpy, and ΔS is the entropy of melting. The standard state was taken to be at 1 μM strands and 25 °C.

UV spectra were recorded on a Hitachi U_3210 spectrophotometer equipped with thermostated cuvette holders. Absorption coefficients for the duplexes were determined by recording a spectrum under standard conditions in 0.05 M sodium acetate and 10 mM magnesium acetate, pH 5, followed by digestion with nuclease P1 (Pharmacia, Sweden). After the reaction was complete as monitored by the increase in the absorbance at 260 nm, the spectrum was recorded again. The concentration was determined from the known absorption coefficients of the mononucleotides (Fasman, 1975) and the composition. This was then used to calculate the absorption coefficient of the undigested sample. CD[1] spectra were recorded in 0.5 M KCl at 15 °C on a Jasco spectropolarimeter as previously described (Lane et al., 1993).

Comparative electrophoresis was carried out on minigels using 20% acrylamide and Tris borate−EDTA buffer. The gels were typically run at 150 V for 60−90 min at 4 °C, stained with ethidium bromide, and visualized using its fluorescence.

^1H NMR spectra were recorded at 11.75 and 14.1 T on Varian UnityPlus and Unity spectrometers, respectively. Spectra in H_2O were recorded at 10 °C using the Watergate method (Piotto et al., 1992) and at 30 and 35 °C in D_2O. Phase-sensitive two-dimensional NMR spectra were recorded using the method of States et al. (1982). Acquisition times were typically 0.4−0.5 s in t_2 and 0.05−0.06 s in t_1. TOCSY spectra were recorded using MLEV-17 for isotropic mixing (Bax & Davis, 1985) with mixing times of ca. 50 ms and a spin-lock field strength of 9 kHz. Data tables were zero-filled once in t_2 and twice in t_1 before Fourier transformation. Spectra were referenced to internal 2,2-dimethylsilapentane-5-sulfonate. Effective rotational correlation times were determined using time-dependent one-dimensional NOEs as previously described (Lane et al., 1986; Birchall & Lane, 1990).

^{31}P NMR spectra were recorded at 25 °C at 9.4 T on a Bruker AM400 spectrometer. The spectra were referenced to external methylene diphosphonate at 0 ppm. The acquisition time was 1.3 s, and a relaxation delay of 2 s was used. The spectra were processed using a Lorentz−Gauss function.

Relaxation rate constants R_1 and R_2 were measured as previously described (Forster & Lane, 1990). The heteronuclear NOE was measured using the gated decoupler experiment with a recycle time of $5T_1$. The rotational correlation time and CSA were estimated from the relaxation data by searching through values and finding the minimum of the target function:

$$\text{error} = |(R_1(\text{obs}) - R_1(\text{calc}))/R_1(\text{obs})| + (R_2(\text{obs}) - R_2(\text{calc}))/R_2\text{obs}| + (\text{NOE(obs)} - \text{NOE(calc)})/\text{NOEobs}| \quad (2)$$

where

$$R_1(\text{calc}) = 133.33\kappa^2\omega_p{}^2J(\omega_p)10^{-6} + \sum 9.1/r^6[J(\Delta\omega) + 3J(\omega_p) + 6J(\sum\omega)] \quad (3)$$

$$R_2(\text{calc}) = 22.22\kappa^2\omega_p{}^2[4J(0) + 3J(\omega_p)]10^{-6} + \sum 4.5/r^6[4J(0) + J(\Delta\omega) + 3J(\omega_p) + 3J(\omega_p) + 6J(\sum\omega)] \quad (4)$$

$$\text{NOE} = 1 + 2.5[6J(\sum\omega) - J(\Delta\omega)]/R_1 \quad (5)$$

κ is the effective CSA, which is the product of the chemical shift anisotropy and the asymmetry parameter (Lane, 1994).

As the observed NOE was small, the values of R_1 and R_2 are dominated by the CSA term, so an initial estimate of the correlation time can be estimated from the ratio of R_2/R_1 (Williamson & Boxer, 1989).

[1] Abbreviations: DQF-COSY, double quantum filtered correlation spectroscopy; NOESY, nuclear Overhauser spectroscopy; TOCSY, total correlation spectroscopy; CSA, chemical shift anisotropy; CD, circular dichroism.

$$\tau = (1/2\omega_p)[6R_2/R_1 - 7]^{0.5} \quad (6)$$

An initial estimate of κ was obtained from the difference function $R_2 - 0.5R_1$ using the initial value of τ

$$R_2 - 0.5R_1 = 88.88\kappa^2\omega_p{}^2\tau 10^{-6} + \sum 4.5/r^6[4J(0) + 6J(\sum\omega)] \quad (7)$$

An initial value for r was taken as 2.1 Å (Forster & Lane, 1990), which is based on field-dependent measurements of the ^{31}P relaxation times and calculations on B and A form structures. The search was then conducted for values of κ and τ in the neighborhood of these initial values, for different values of r. As the dipolar term is small compared with the CSA term, the result is not very sensitive to r. The search gives a set of solutions that are compatible with the experimental results within experimental error and therefore provides an estimate of the variance of each fitted parameter.

Conformational Analysis. Sums of coupling constants were determined from the NOESY recorded with a long mixing time (250 ms) and DQF-COSY spectra. H1′−H4′ distances were estimated from cross-sections through H1′ of NOESY spectra. The ratio, R, of the cross-peak areas in cross-section of H1′−H4′ to H1′−H2″ is equal to the ratio of the cross-peak volume as H1′ is the common diagonal peak. This NOE is not significantly distorted by spin diffusion, and because $r_{1'2''}$ is essentially independent of sugar conformation ($r_{1'2''} = 2.35 \pm 0.05$ Å), $r_{1'4'}$ can be calculated from R as

$$r_{1'4'} = 2.35R^{-1/6} \quad (8)$$

Because H2″ relaxes faster than H4′, R is overestimated. This is compensated for by making measurements from NOESY spectra recorded at different mixing times and extrapolating back to zero mixing time.

Distances and coupling constants were simultaneously evaluated using a grid-search method assuming a two-state equilibrium between N and S sugar puckers as previously described (Conte et al., 1996). In these searches, the pseudorotation phase angle was held constant at 9°, and a systematic search was conducted over a range of the pseudorotation phase angle for the S state (P_s) and the fraction of the S state, f_s. The search was repeated for different values of the pucker amplitude ϕ_m between 30° and 42°, assuming that the amplitude is the same for both the N and S states. The combination of the coupling constants and these distances leads a better determination of the sugar conformations than either set of data individually (Conte et al., 1996).

Glycosidic torsion angles were determined by the combined grid-search and nonlinear regression to NOE build-up curves with the program NUCFIT (Lane, 1990), which accounts for spin diffusion, multiple sugar conformations, and the effects of incomplete longitudinal relaxation (Lane & Fulcher, 1995).

Distances between protons in sequential nucleotides, H2′-(i)−H8/H6(i+1) in the RNA strands and H2″(i)−H8/H6-(i+1) in the DNA strands, were calculated from NOESY cross-peak volumes normalized to those of the cytosine or uridine H6−H5 cross-peaks. These distances are diagnostic of A- and B-like helices. Volumes were calculated using Felix 95.0 (Biosym) either by direct volume integration

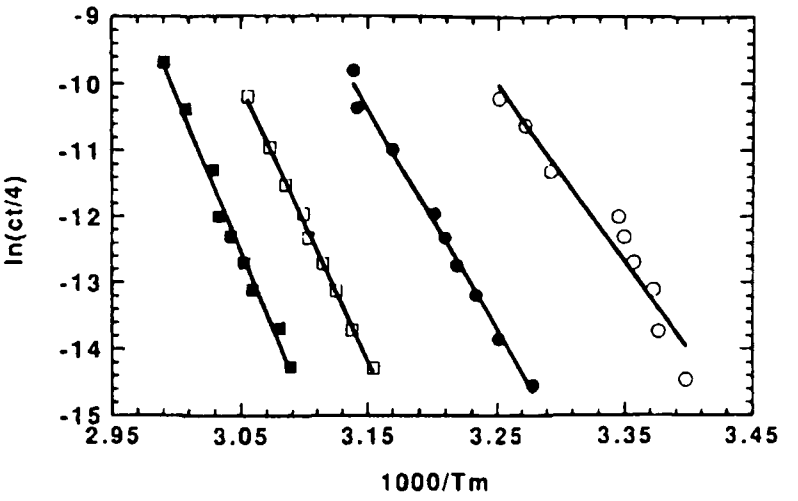


FIGURE 1: Dependence of melting temperatures on concentration. Melting temperatures for the four duplexes were measured as described in Materials and Methods. (●) dR10·dY10, (□) rR10·dY10, (○) dR10·rY10, (■) rR10·rY10.

or by fitting Gaussian functions to cross-sections in both dimensions through cross-peaks for overlapping peaks. For isolated peaks, these two methods agreed within 20%. Normalized volumes from spectra recorded at different mixing times (typically 50, 100, and 250 ms) were extrapolated linearly back to zero mixing time, and the distance was calculated assuming that r(H6−H5) = 2.45 Å. Simulations showed that, for the correlation times measured for these decamers, the linear extrapolation from 100 ms provides a reasonably precise estimate of the normalized volume at zero mixing time.

RESULTS

Thermodynamic Stability. The thermodynamic stability of all four duplexes has been determined from the concentration dependence of the melting temperatures determined optically as described in Materials and Methods. Monophasic melting curves were obtained, and the dependence of the T_ms on concentration of strands (Figure 1) demonstrated the involvement of at least two strands in the observed transitions. The sequences of the oligonucleotides raise the possibility of triplex formation. However, at higher concentrations of nucleotides and comparable buffer conditions, the NMR spectra recorded at temperatures between 10 and 40 °C showed no evidence of triple helices (see below), which have characteristic spectral properties (Feigon et al., 1995). It is therefore unlikely that triple helices would form at the lower concentrations used in the optical studies. Furthermore, the kinds of intermolecular triple helices that might form would be expected to be very unstable in the absence of magnesium (Pilch et al., 1990), especially at neutral pH where the cytosines in the Hoogsteen strands of the parallel RY·Y triplex (with overhanging G on the third strand) would be unprotonated. The electrophoresis experiments (see below) showed no evidence of mixtures of duplexes and triple helices. The UV-melting curves showed no evidence of low-temperature transitions. Additional control experiments carried out with excess of one or the other strand at different pH values, in the absence or presence of magnesium and monitored at both 260 and 280 nm, failed to detect any triplex formation. We conclude that the transitions measured optically correspond primarily to melting of duplexes into single strands, and they have been analyzed accordingly. The thermodynamic parameters are collected in Table 1. The data have been obtained also at

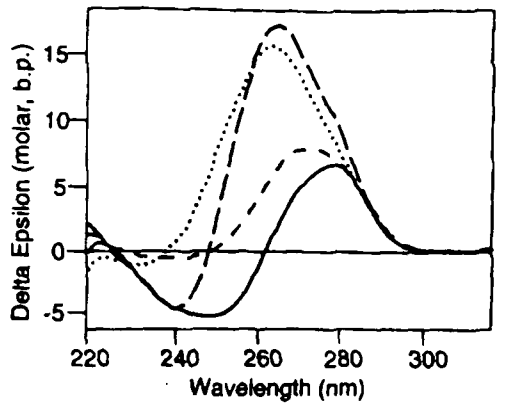


FIGURE 2: CD spectra of the four duplexes. Spectra were recorded at 15 °C as described in Materials and Methods. (—) dR·dY; (—·—) dR·rY; (—) rR·dY; (···) rR·rY.

Table 1: Thermodynamic Stability of the Four Decamer Duplexes[a]

duplex	$\Delta G(298)$ (kJ mol^{-1})	ΔH (kJ mol^{-1})	ΔS (kJ mol^{-1}/K)	T_m (1 μM) (K)	T_m (1 mM) (K)
dR10·dY10	−42.9	−290	−0.829	305	325
dR10·rY10	−31.8	−238	−0.693	291	313
rR10·dY10	−55.8	−343	−0.963	315	333
rR10·rY10	−66.9	−394	−1.098	322	338

[a] Thermodynamic parameters were determined from the concentration dependence of the T_m measured from UV melting curves as described in Materials and Methods. The T_m at 1 mM was calculated from the T_m measured at 1 μM and ΔH using eq 1 in the text.

different ionic strengths, from 0.1 to 1 M in salt at a fixed concentration of nucleotide. The slope of the T_m versus log [Na^+] was between 9 and 10, which is similar to the value reported by Ratmeyer et al. (1994) when the number of base pairs is taken into account. As expected, and consistent with other work (Hall & McLaughlin, 1991; Ratmeyer et al., 1994; Lesnik & Freier, 1995; Sugimoto et al., 1995; Wang & Kool, 1995), the RNA duplex is the most stable under all conditions. The order of stability based either on $\Delta G(298)$ or T_m is rR10·rY10 > rR10·dY10 > dR10·dY10 > dR10·rY10. Interestingly, the stabilities also follow the ΔH term, with the least stable dR10·rY10 duplex showing a low enthalpy, consistent with the poorest base-stacking.

Optical Spectroscopy. The absorption spectra of the four duplexes were similar, with λ_{max} between 257 and 258 nm, and the hyperchromicity for decomposing the duplexes into the component mononucleotides varied between 0.4 for the RNA duplex and the rR10·dY10 hybrid to 0.49 for the DNA duplex and dR10·rY10 hybrid.

The near-UV CD spectra of all four duplexes were recorded at a concentration of 40 μM duplex in 0.5 M KCl at 15 °C (Figure 2), where the oligonucleotides are >95% duplex (Table 1). The RNA duplex showed a typical nonconservative CD spectrum, with an intense positive peak at 264 nm and a weak negative peak at 232 nm. The DNA showed a typical conservative CD spectrum with a positive peak at 281 nm and a negative peak at 249 nm. The intensity in the spectrum of the RNA duplex was approximately 2.5-fold higher than that in the DNA spectrum (Figure 2). These spectral properties are typical of A- and B-like helices, respectively (Ivanov et al., 1974). The spectra of the two hybrids were intermediate between these two extremes, though in both cases the spectra were nonconservative, with features more A-like than B-like.

These results are comparable to those reported by Ratmeyer et al. (1994) for dodecamers consisting of alternating GA·CT base pairs. In all four duplexes studied here, the purine strands have identical base compositions, whereas in the pyrimidine strands there is the difference between U and T in the deoxy- or ribonucleotides. It is possible that different transition moments for T and U could produce significantly different spectra. However, in polymeric RNA of mixed sequence, the spectral properties are very similar to those of the RNA duplex reported here (Ivanov et al., 1974).

Table 2: Hydrodynamic Properties of the Four Decamer Duplexes[a]

duplex	μ_{rel}	τ_{cyt} (ns)	τ_p (ns)	CSA (ppm)
dR10·dY10	1.0	3.2 ± 0.3	3.6 ± 0.3	147 ± 7
dR10·rY10	0.92	nd	nd	nd
rR10·dY10	0.84	nd	2.95 ± 0.24	141 ± 7
rR10·rY10	0.77	2.9 ± 0.2	3.2 ± 0.2	154 ± 5

[a] μ_{rel} is the relative electrophoretic mobility on 20% acrylamide gels at 4 °C. τ_{cyt} and τ_p are effective correlation times at 30 °C determined by cross-relaxation between the H6 and H5 of cytosine and ^{31}P NMR relaxation, respectively. CSA is the effective chemical shift anisotropy as defined in the text, averaged over all phosphodiesters in each duplex. nd, not determined.

Electrophoresis. All duplexes have identical compositions and sequences (apart from the U for T substitutions in the RNA strands) and should have identical charges. Differences in electrophoretic or hydrodynamic properties therefore should arise primarily from differences in hydrodynamic volume (or shape). This could be due to gross differences in shape (such as B and A form), large differences in hydration or counterion condensation (which would depend on overall conformation), or differential interaction with the gel matrix. We have measured the electrophoretic mobility as summarized in Table 2. The order of mobility is dR10·dY10 > dR10·rY10 > rR10·dY10 > rR10·rY10, which agrees with the order reported by Ratmeyer et al. (1994). The clear ordering of mobilities suggests that the global shapes of the two hybrids are not identical, but that rR10·dY10 is more akin to the RNA duplex, and dR10·rY10 is more similar to the DNA duplex.

NMR Spectroscopy. The thermodynamic, spectroscopic, and hydrodynamic data confirm that there are differences between the two DNA·RNA hybrids. To probe the conformational differences in greater detail, we have applied ^1H NMR spectroscopy. The protons of the four duplexes have been assigned using NOESY, DQF-COSY, and TOCSY.

d(GAAGAGAAGC)·d(GCTTCTCTTC). Figure 3 shows the low-field region of the ^1H NMR spectra of the four decamers. In each case, 9−10 resonances can be observed, falling into two clusters corresponding to the imino protons of GC and AT(U) base pairs. The imino protons were assigned using NOESY in H_2O (not shown) which also provided the assignments of the Ade C2H. The N4H amino protons of the cytosines were assigned from NOEs between the GN1H of the base-paired G residue, and the H5 of the same cytosine residue. The hydrogen-bonded and non-hydrogen-bonded amino protons could also be discriminated using these NOEs. In none of the spectra was there any evidence for disproportionation into triple helices at any temperatures, nor in the presence of higher salt concentrations.

Despite the unfavorable sequences, the purine-rich strand gave good resolution in the base and H1′ regions, making the sequential assignment straightforward (Figure 4A). The pyrimidine resonances were much less well resolved, espe-

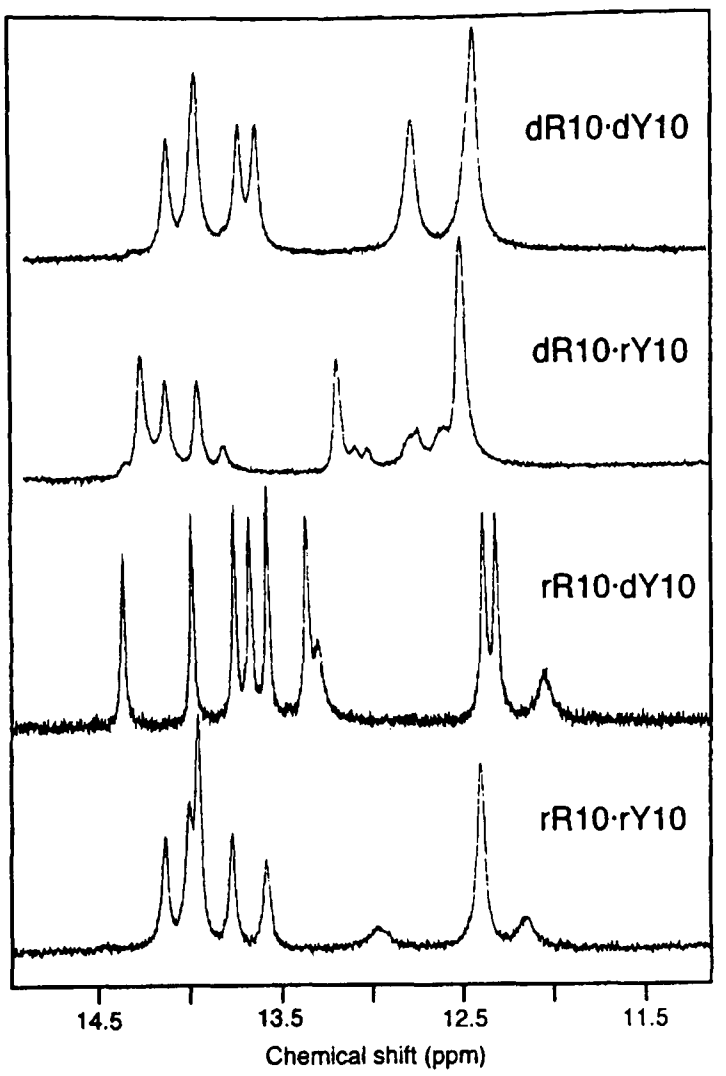


FIGURE 3: NMR spectra of the four duplexes in ^1H_2O. Spectra were recorded at 600 MHz, 10 °C, as described in Materials and Methods. Free induction decays were apodized using a 2 Hz line-broadening exponential. Low field region showing imino proton resonances.

cially those of thymidine H6, which can be attributed to the small ring current fields of the pyrimidines compared with the purines. Nevertheless, the sequential base−H1′ pathway could be traced. The remaining sugar protons were assigned using the NOEs from the assigned H1′ to H2′ and H2″, which were assigned stereospecifically because the NOE to H2″ is always larger than to H2′. The base and H2′/H2″ resonances were checked using the sequential NOEs between the base protons and the H2′, H2″ resonances. H3′ resonances were assigned using the sequential walk H8/H6(i)−H3′(i)−H8/6(i+1) in the same manner as for the base/H1′ resonances. The H3′ were confirmed using NOEs to the H2′ and from the cross-peaks H2′−H3′ in the DQF-COSY spectrum (see below). The H4′ were initially assigned using the NOEs from H1′ and confirmed using the H3′−H4′ cross-peaks in the DQF-COSY spectrum. With the exception of the unresolved H5′/H5″, all of the nonexchangeable protons were assigned.

r(GAAGAGAAGC)·d(GCTTCTCTTC). Figure 4B shows portions of a NOESY spectrum of the hybrid duplex r(GAAGAGAAGC)·d(GCTTCTCTTC). The H8−H1′ region of the RNA strand is well dispersed, making the sequential assignments via this pathway straightforward. The H2′ were assigned from the intense H1′−H2′ intraresidue cross-peaks. These assignments were confirmed from the sequential pathway H8(i)−H1′(i)−H2′(i)−H8(i+1) in NOESY spectra recorded with mixing times of 50 and 100 ms. H3′ resonances were assigned using both sequential and intraresidue NOEs to the assigned H8 and in some instances from DQF-COSY cross-peaks to the assigned H2′ resonances as previously described (Lane et al., 1993). The H4′ resonances were then found from correlations from the H1′ resonances in NOESY spectra and by noting the different cross-peak fine structure in the F_2 dimension for cross-peaks to H3′ and H4′, owing to their very different coupling patterns.

Although the H6−H1′ region of the NOESY spectra was relatively crowded, the combination the three sequential pathways as described for the DNA duplex (see above), the COSY connectivities for H1′ to both H2′ and H2″ as well as H2′−H3′ and H3′−H4′ made it possible to assign all of the nonexchangeable protons of the deoxypyrimidine nucleotides (except the H5′/H5″).

Also notable in Figure 4B is the presence of intense cross-peaks between the Ade C2H(i) and the H1′(i+1), both sequentially and across the duplex, compared with absence of the intraresidue AH2−H1′ (for which the distance is ca. 4.8 Å). This is consistent with substantial propeller twisting of the AU base pairs.

d(GAAGAGAAGC)·r(GCUUCUCUUC). Figure 4C shows a NOESY spectrum of the hybrid duplex d(GAAGAGAAGC)·r(GCUUCUCUUC). As for the DNA duplex, assignment of the purine strand was straightforward, and all but the H5′/H5″ resonances of the deoxyribose strand could be assigned using methods described for the DNA duplex above. The assignment of the ribose pyrimidine strand was more difficult because of the poorer dispersion, spectral overlap, and lack of correlation between H1′ and H2′ in the DQF-COSY experiment for all but the terminal residues (see below). However, by analyzing NOESY spectra recorded at different temperatures and mixing times, unambiguous assignments for all of the base and H1′ resonances were obtained. The H2′ resonances were assigned using the H1′−H2′ NOEs, which were checked against the base protons with the sequential pathways H6(i)−H1′(i)−H2′(i)−H6(i+1) (Figure 4C) as described above. Some H3′ resonances were assigned using H2′−H3′ correlations in the COSY experiment and H3′−H6 NOEs. Some H4′ were assigned from the H1′−H4′ NOE. However, the extensive overlap precluded complete assignment of the H3′ and H4′ resonances.

Notably strong cross-peaks were present for between Ade C2H(i) and H1′(i+1) on both strands. This is consistent with substantial propeller twisting of the AU base pairs, as observed with the other DNA·RNA hybrid (see above).

r(GAAGAGAAGC)·r(GCUUCUCUUC). Figure 4D shows a NOESY spectrum of the RNA duplex at 30 °C. The base and H1′ resonances were assigned using the NOE sequential pathway H8/H6(i)−H1′(i)−H8/H6(i+1), and H2′ from the H1′−H2′ intraresidue NOEs. The H2′ assignments were confirmed from the sequential H2′(i)−H8/H6(i+1) NOEs, which all appear intense. H4′ resonances were assigned from the H1′−H4′ NOE. The H3′ resonances were identified by their intra- and interresidue NOEs with the base protons, after eliminating the H2′ resonances. In some instances, H3′ and H4′ assignments could be confirmed from the DQF-COSY spectrum. However, some H4′ resonances could not be assigned, and some H3′ resonances in the pyrimidine strand remain ambiguous. Strong cross-peaks between AH2(i) and H1′(i+1) were present as observed in the two hybrid duplexes (see above).

The compositions and sequences of individual strands in the duplexes are identical. Hence, changes in chemical shifts between pure duplexes and the two hybrids should largely reflect differences in conformation. Figure 5 shows the

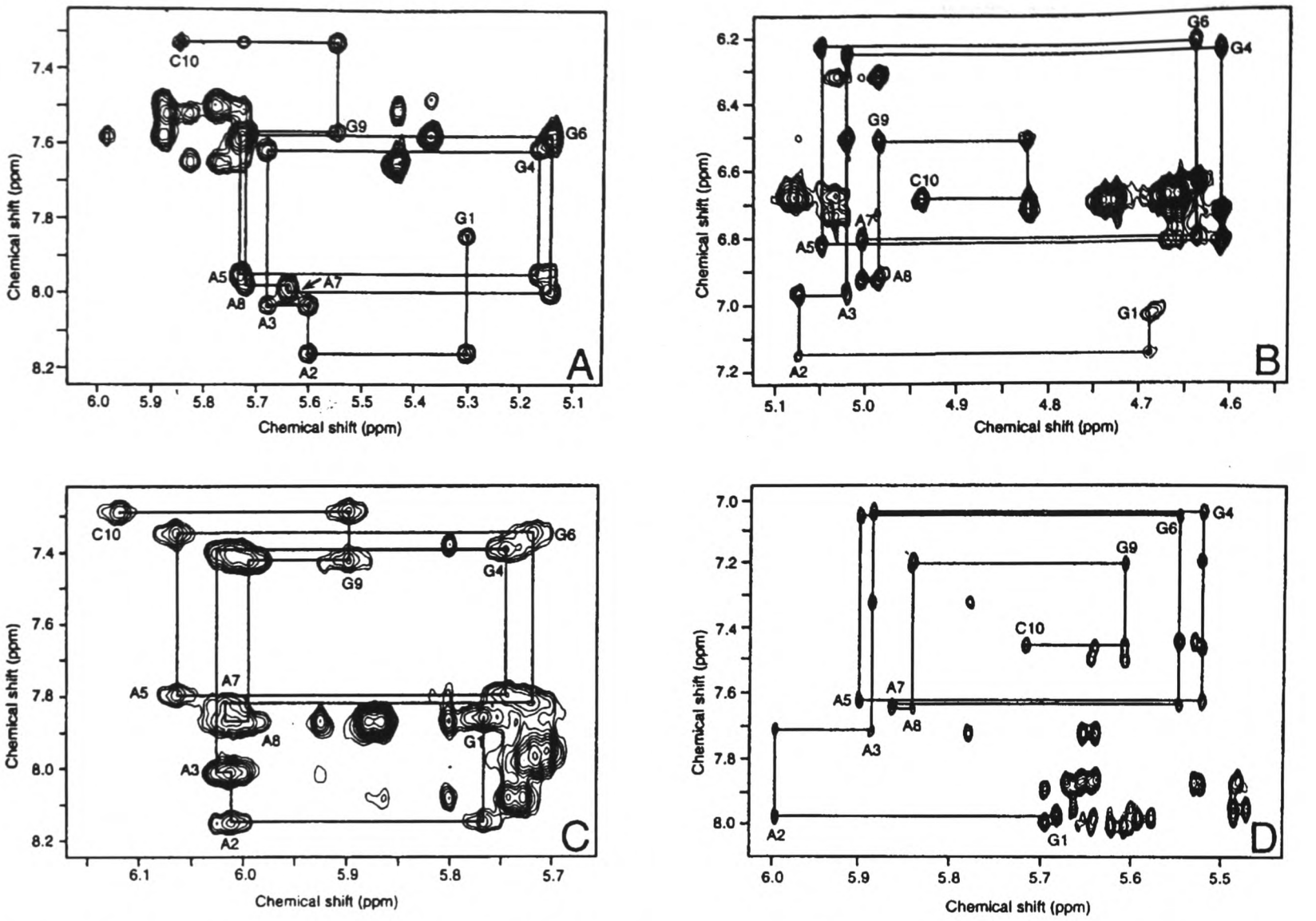


FIGURE 4: NOESY spectra of the duplexes in ^2H_2O. Spectra were recorded at 600 MHz, 30 °C, as described in Materials and Methods. Base proton−H1′ and C/U H6−H5 connectivities are shown. Spectra were zero-filled to 8192 × 2048 complex points, and apodized using a Gaussian function in both dimensions. (A) dR10·dY10. (B) rR10·dY10. (C) dR10·rY10. (D) rR10·rY10.

differences in proton chemical shifts between the hybrids and those of the DNA and RNA duplexes. The chemical shift differences were calculated as hybrid deoxyribose proton shifts minus DNA shifts and hybrid ribose proton shifts minus RNA shifts. We consider chemical shift changes of less than 0.05 ppm to be insignificant, as shown by the horizontal line at ±0.05 ppm. In the DNA strand of dR10·rY10, there are large positive shift differences for the H1′ protons and negative shift differences for the H8 and Ade C2H. In contrast, the deoxypyrimidines in rR10·dY10 in general show smaller shift differences, with the largest effects of H2′ (positive) and H2″ and H5 (negative). In the ribose purine strand, the shift differences are smaller on average than for the analogous comparison with the deoxypurines, and the signs of the changes are reversed. The smaller change in shifts for the ribose nucleotides is consistent with the conformation in the hybrids being similar to that in the pure RNA duplex. In contrast, the larger changes observed for the deoxynucleotides suggest that there is a substantial difference in conformation in the DNA strands between the pure DNA duplex and the DNA·RNA hybrids. Further, as the signs and absolute magnitudes of the shift differences differ for the two hybrids, it is plausible that the conformations are also somewhat different, which is supported by the electrophoretic and optical spectroscopy data (see above).

^{31}P NMR. ^{31}P chemical shifts have been reported to depend on backbone conformation (Gorenstein, 1992; Chou et al., 1992; Lane et al., 1992; Legault, & Pardi, 1994). We therefore recorded ^{31}P NMR spectra at 9.4 T as described in Materials and Methods. The chemical shift dispersions of the duplexes were in the order dR10·dY10 (0.45 ppm) < dR10·rY10 (0.60 ppm) < rR10·rY10(0.70 ppm) < rR10·dY10 (0.72 ppm) (spectra not shown). For comparison, the shift dispersion of an unrelated RNA duplex, r(CGCAAAU-UUGCG)$_2$ was 0.78 ppm under identical conditions (M. R. Conte, G. L. Conn, T. Brown, and A. N. Lane, unpublished data), and mixed-sequence DNA duplexes typically show a dispersion of shifts in the range 0.5−0.6 ppm (Gorenstein, 1992). Further, the center of the ^{31}P NMR spectra were essentially the same (approximately −17.3 ppm from methylene diphosphonate), suggesting similar backbone torsion angles on average. It appears that the range of phosphate shifts found in RNA is somewhat larger than in DNA duplexes. We also observed the relatively wide shift dispersion of 0.87 ppm in an octameric DNA·RNA hybrid of mixed sequence, which was shown to be globally closer to the A form than to the B form (Lane et al., 1993). The present results are consistent with the conformation of the rR10·dY10 hybrid duplex being closer to the A conformation adopted by RNA duplexes than the dR10·rY10 hybrid, which may have a more intermediate conformation at the level of

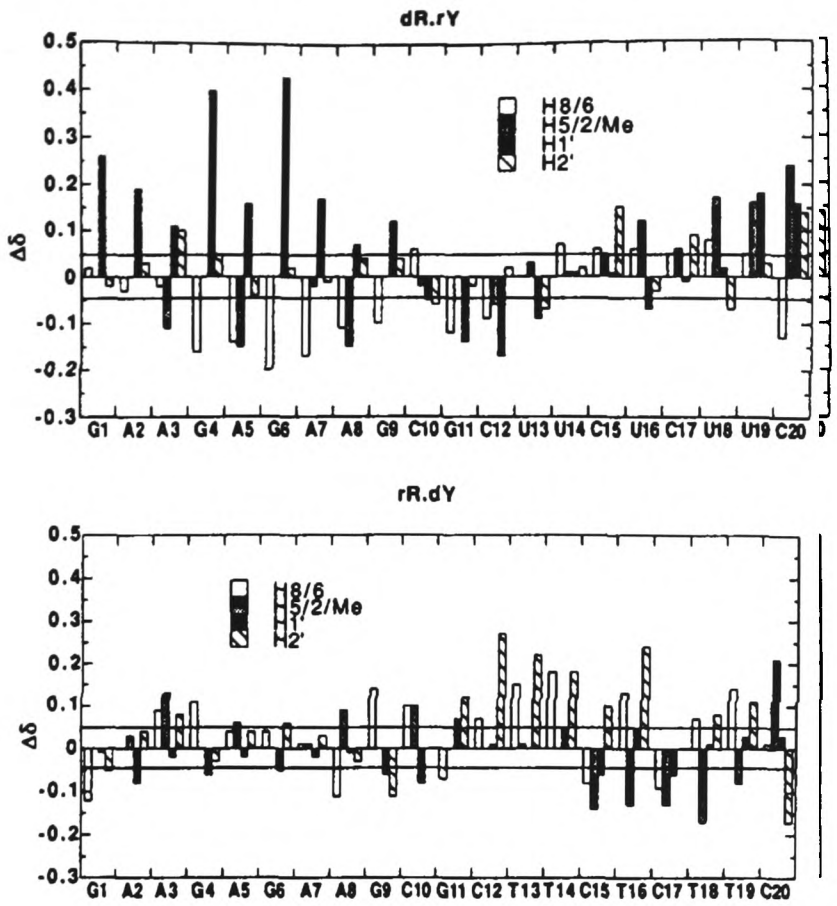


FIGURE 5: Chemical shift differences between the hybrid duplexes and the pure DNA and RNA duplexes. Chemical shift differences for selected protons were calculated from assignments made at 30 °C as described in the text. (Upper) Shift differences for the deoxyribose strands: *dR10*·rY10−*dR10*·dY10 and rR10·*dY10*−dR10·*dY10*. (Lower) Shift differences for the ribose strands: dR10·*rY10*−rR10·*rY10* and *rR10*·dY10−*rR10*·rY10.

the backbone torsion angles. This is consistent with the other spectroscopic data (see above).

Rotational Correlation Times. Rotational correlation times for the DNA and RNA duplexes were determined in two ways. First, measurements of the NOE build-up curves for the cytosine H6−H5 and uridine H6−H5 vectors were measured using 6−7 time points, and the effective correlation time was calculated from the cross-relaxation rate constant (Lane et al., 1986). The mean correlation time determined from the derived cross-relaxation rate constants was 3.2 ± 0.3 ns at 30 °C for the DNA duplex, which is typical of B-DNA in D_2O solution under these conditions (Birchall & Lane, 1990). Similarly, we obtained an effective rotational correlation time of 2.9 ± 0.3 ns at 30 °C for the RNA duplex.

Second, the spin−lattice and spin−spin relaxation rate constants of the phosphates were measured at 9.4 T and analyzed as described in Materials and Methods. Although the spectra were too crowded to assign by heteronuclear correlation experiments, the relaxation rate constants reflect mobility of the backbone. The relaxation data are summarized in Table 2. The correlation time for the DNA duplex was 3.6 ± 0.3 ns by ^{31}P relaxation, i.e., slightly larger than that obtained by $^1H-^1H$ cross-relaxation. The RNA duplex had an average correlation time of 3.2 ± 0.2 ns, again slightly larger than obtained from $^1H-^1H$ cross-relaxation (Table 2). This difference has been noted previously for DNA and may reflect different sensitivity of the two methods to internal motions (Lane, 1994). Although the correlation times for DNA and RNA are only slightly different compared with the experimental error, the consistent difference in two kinds of experiment is consistent with the RNA being more compact than the DNA as would be expected for A- and B-type duplexes, respectively. The correlation time determined by ^{31}P NMR for the rR10·dY10 hybrid duplex was 2.95 ± 0.24 ns, which is essentially identical to that determined for the RNA duplex and is consistent with a similar overall shape.

The effective chemical shift anisotropy was essentially identical for each duplex at ca. 150 ppm, which is comparable to values measured on other DNA duplexes in DNA in general (Forster & Lane, 1990), and demonstrates that ^{31}P line widths are dominated by CSA at magnetic field strengths of 9.4 T or higher.

Sugar Conformations of the Riboses. With the exception of terminal residues, there were no observable correlations between H1′ and H2′ of the riboses in DQF-COSY spectra recorded with acquisition times up to 0.09 s in t_1. Strong cross-peaks were observed, however, for H3′−H4′. The ribose H1′ that were resolved in the 1D spectra appeared as singlets, with line widths of <3 Hz. This is also evident in the NOESY spectra (Figure 4B−D) where the ribose H1′ resonances are narrow along F_2, compared with the H1′ resonance of the deoxyriboses, which appear as triplets. This indicates $^3J_{1'2'}$ of <2−3 Hz, which places the ribose conformations in the N domain (−32 < P < 32). This indicates that the riboses in all three RNA-containing duplexes have sugar conformations in the "N" domain. Even for terminal residues, the H1′−H2′ cross-peaks in the DQF-COSY spectra were very weak, also indicating that these residues, while having possibly greater conformational freedom than internal residues, are also largely in the "N" domain.

Sugar Conformations of the Deoxyriboses. Sums of coupling constants derived from cross-sections of DQF-COSY and NOESY, and where possible 1D spectra, are given in Table 3. In addition, the distance H1′−H4′ was estimated from NOESY spectra as described in Materials and Methods. This distance is at a minimum for $P_s = 90°$ (O4′-endo) and increases sharply in the ranges $100 < P_s < 162$ and $90 > P_s > 18$ (van de Ven & Hilbers, 1988). It is therefore a useful adjunct to coupling constants for differentiating between unique conformations with $P \approx 90°$ (Salazar et al., 1993) and mixtures of N and S conformations (Conte et al., 1996).

In the pure DNA duplex, the intensities of the DQF-COSY H1′−H2′ cross-peaks are much greater than the H1′−H2″ cross-peaks, and similarly for H2′−H3′ versus H2″−H3′. The cross-peaks also show considerable antiphase fine structure, viz., ++−− for H2′−H1′ and +−+− for H2″−H1′ (Figure 6). Indeed, this structure is evident for essentially all of the cross-peaks (C20 is an exception where H2′ and H2″ are nearly isochronous) indicating significantly different values of $^3J_{1'2'}$ and $^3J_{1'2''}$. Further, $\Sigma_{1'}$ (= $^3J_{1'2'}$ + $^3J_{1'2''}$) is >14.5 Hz for all residues except the terminal nucleotides, showing that the nucleotides are predominantly in the "S" domain (van Wijk et al., 1992). This is supported by the relatively small values of $\Sigma_{3'}$ and the value of $r_{1'4'}$ (>3 Å). Because extensive data for the sugars of the purine strand are available, a detailed examination of the sugar conformation can be made.

The sums of the coupling constants and $r_{1'4'}$ cannot be rationalized as a single, unique conformation. We have therefore used a two-state N/S equilibrium model to analyze these data. Conformational mixtures as a function of f_s and

Table 3: Sugar Conformations of Deoxyriboses[a]

nucleotide	$\Sigma_{1'}$ (Hz)	$\Sigma_{2'}$ (Hz)	$\Sigma_{2''}$ (Hz)	$\Sigma_{3'}$ (Hz)	$r_{1'4'}$ (Å)	P_s (deg)	f_s
dR10·dY10							
G1	15.1	30.4	21.2	nd	3.0	140	0.92
	$^3J_{1'2'} = 9.5$, $^3J_{1'2''} = 5.6$						
A2	15.3	30	21	nd	3.05	167	0.92
	$^3J_{1'2'} = 9.8$, $^3J_{1'2''} = 5.5$						
A3	15.0	nd	nd	nd	3.0	148	0.91
	$^3J_{1'2'} = 9.5$, $^3J_{1'2''} = 5.5$						
G4	14.6	30	21	nd	2.9	126	0.88
A5	nd	30	21	nd	3.1	153	0.88
G6	14.4	nd	nd	12.9	2.9	126	0.85
A7	15.0	nd	nd	14.4	3.0	135	0.91
	$^3J_{1'2'} = 8.8$, $^3J_{1'2''} = 6.2$						
A8	15.3	nd	nd	nd	3.0	S	0.93
G9	14.4	30.3	23.2	15.9	3.1	135	0.84
	$^3J_{1'2'} = 8.0$, $^3J_{1'2''} = 6.4$						
C10	14	29.2	nd	nd	3.0	135	0.78
mean	14.9	30.1	21.5	14.4	3.0	141	0.89
sd	±0.4	±0.2	±0.9	±1.2	±0.07	±13	±0.03
G11	14.2	29.8	nd	16.2	3.2	144	0.7
C12	nd	nd	nd	16	2.85	≈120	≈0.7
T13	15.0	nd	nd	16.2	nd	S	0.9
	$^3J_{1'2'} = 9.2$, $^3J_{1'2''} = 5.8$						
T14	15.2	31	nd	nd	2.85	126	0.92
C15	14.1	nd	nd	15.5	nd	≈120	0.7
T16	nd	30.5	nd	12	nd	≈130	≈0.9
C17	14.9	nd	nd	16	3.0	135	0.86
T18	14.7	nd	nd	nd	2.9	S	0.83
T19	14.6	nd	24	nd	3.0	108	0.82
C20	13.8	nd	nd	16.2	3.0	S	0.68
mean	14.7	30.4		15.3	2.97	125	0.83
sd	±0.4	±0.5		±1.5	±0.12	±12	±0.07
rR10·dY10							
G11	13.2	30	25	nd	nd	103	0.67
C12	13.3	28	nd	15	3.0	125	0.7
T13	13.1	29	26	nd	2.8	140	0.7
T14	13.3	28	23	12	2.8	150	0.66
C15	12.7	28	nd	14	2.9	148	0.58
T16	13.2	28	nd	12	2.8	162	0.66
C17	12.3	27	26	14	2.8	152	0.55
T18	13.0	29	nd	nd	2.8	110	0.67
T19	14.2	30	27	12	2.9	138	0.82
C20	13.9	29	25	15	2.9	128	0.75
mean	13.1	28.6	25.3	13.2	2.85	136	0.67
sd	±0.5	±1	±1.2	±1.2	±0.07	±20	±0.07
dR10·rY10							
G1	14.2	28	23	14	2.85	130	0.78
A2	nd	nd	nd	14	nd	S	nd
A3	14.5	nd	24	16.7	2.8	108	0.87
G4	14.5	nd	nd	12	2.95	135	0.87
A5	14.0	31	23	nd	3.0	135	0.78
G6	14.5	30	23	15	3.0	130	0.86
A7	14.1	nd	nd	11	3.1	158	0.78
A8	13.8	31	25	nd	2.95	140	0.75
G9	13.5	30	25	nd	3.0	135	0.69
C10	14.0	30	26	nd	2.9	131	0.78
mean	14.1	30	23.8	13.8	2.95	134	0.79
sd	±0.4	±1.1	±0.9	±1.9	±0.08	±13	±0.06

[a] $\Sigma_{3'}$ includes $^3J_{H3'-P}$ (4−6 Hz). Values of P_s and f_s calculated as described in the text. The standard deviations were estimated as ±0.5 Hz for $\Sigma_{1'}$, ±1 Hz for $\Sigma_{2'}$, and ±2 Hz for $\Sigma_{2'}$ and $\Sigma_{3'}$. The lower limit to $r_{1'4'}$ was set to 0.1 Å below the extrapolated value at zero mixing time, and the upper limit to this value +0.5 Å. Errors on P_s and f_s are estimated at ±20° and ±0.1, respectively. nd, not determined.

P_s were systematically generated, and the calculated coupling constants and distances were compared with the experimental data using the program pfit (Conte et al., 1996). The resulting values for f_s and P_s are given in Table 3. Not surprisingly, the values of f_s are relatively high (>0.85), and P_s is in the "S" range (ca. 130−160°), which is typical of B-DNA. Fewer data were obtained for the pyrimidine strand, though the combination of the values of $\Sigma_{1'}$ and $r_{1'4'}$ and the observation that $^3J_{1'2'} > {}^3J_{1'2''}$ (see above) suggest that the pyrimidines have on average similar sugar conformations as the purines. There is no significant difference in $\Sigma_{1'}$ between the two strands, on average, though there may be a tendency for the pyrimidines to adopt a lower pseudorotation phase angle (Table 3).

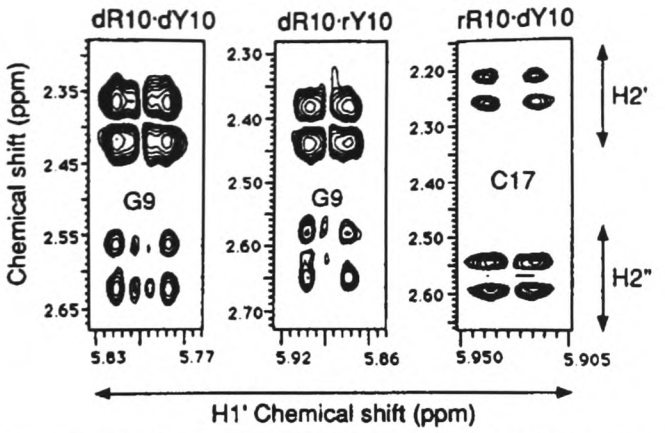


FIGURE 6: Comparison of DQF-COSY cross-peaks in the DNA and DNA·RNA hybrid duplexes for the H1′ to H2′/H2″ regions. Spectra were recorded at 500 MHz and 30 °C as described in Materials and Methods. Spectra were transformed from matrices of 8192 × 2048 complex points. with a shifted Gaussain function to provide apodization and mild resolution enhancement. The digital resolution was 1.22 Hz/point along F_2 and 4.88 Hz/point along F_1. Left, G9 of dR10·dY10; middle. G9 of dR10·rY10; right, C17 of rR10·dY10

In contrast to the DNA duplex, the intensities of the H1′−H2′ and H1′−H2″ cross-peaks are more similar in the hybrid duplex rR10·dY10, and there is little fine structure (Figure 6). Also, the H2″−H3′ cross-peaks are relatively more intense (not shown). Further, as Table 3 shows, the values of $\Sigma_{1'}$ are significantly smaller, in the range 12.3−14.2 Hz (mean = 13.3 ± 0.5 Hz). This indicates a lower conformational purity and/or rather smaller values of P_s. In addition, there is much less fine structure in the cross-peaks, which indicates that ${}^3J_{1'2'} \approx {}^3J_{1'2''}$. This is consistent with a unique sugar conformation only if $P_s \approx 200-220°$. However, neither the values of $\Sigma_{2'}$ and $\Sigma_{2''}$ nor $r_{1'4'}$ are consistent with such a conformation. Further, the $r_{1'4'}$ distances are significantly smaller (ca. 2.8−2.9 Å on average) than in the DNA duplex. This requires a smaller value of P_s for than in the DNA duplex (van Wijk et al., 1992). We note, for a single conformation, that a value of $P_s \approx 108°$ is required to account of the low value of $r_{1'4'}$, but this then predicts that $\Sigma_{1'} \approx 16$ Hz, which is much larger than is observed (Table 3). We conclude that a unique conformation is not consistent with the data. The conformations given in Table 3 were calculated assuming an N/S equilibrium, as described for the DNA duplex.

The hybrid duplex dR10·rY10 appears more similar to the DNA duplex, with the cross-peak intensities H1′−H2′ ≫ H1′−H2″ (Figure 6) and H2′−H3′ > H2″−H3′. However, the fine structures are significantly different. Thus, whereas the H1′−H2′ cross-peak of G9 in the hybrid shows a simple +− fine structure, the fine structure is ++−− in the DNA duplex. Similarly, in the hybrid, the H1′−H2″ fine structure for G9 is +−, whereas in the DNA duplex, it is +−+−. The values of $\Sigma_{1'}$ (Table 3) are in the range 13.4−14.5 Hz (mean 14 ± 0.35 Hz), which is smaller than for the purine strand in the pure DNA duplex but larger than for the pyrimidines in rR10·dY10. Further, values of $r_{1'4'}$ of 2.9−3 Å on average, and $\Sigma_{3'} < 16$ Hz (including the ${}^3J_{H3'P}$ of ca. 4−6 Hz) rule out major conformations with $P_s < 108°$ (van Wijk et al., 1992). The data in Table 3 were analyzed using the N/S two-state equilibrium, from which typical values for P_s of ca. 135°, and f_s values in the range 0.7−0.9 were obtained.

The data given in Table 3 show that the sugar conformations in the DNA duplex are typical of B-DNA, with the possibility that P_s in the pyrimidine strand on average is smaller than P_s in the purine strand. It is clear that the DNA strands in the DNA·RNA hybrids have different conformations from their counterparts in the pure DNA duplex, with possibly a lower value of P_s, and a significantly smaller value of f_s, which indicates a higher degree of flexibility. This is in agreement with our previous findings (Lane et al., 1993) and a recent detailed study of a mixed-sequence DNA·RNA hybrid (Gonzalez et al., 1994, 1995). It is also evident that the two DNA·RNA hybrids are not equivalent, such that the deoxyriboses in rR10·dY10 are more flexible on average than those in dR10·rY10.

Table 4: Glycosidic Torsion Angles[a]

base	*dR10*·dY10 (−χ/deg)	*dR10*·rY10 (−χ/deg)	*rR10*·dY10 (−χ/deg)	base	rR10·*dY10* (−χ/deg)	dR10·*dY10* (−χ/deg)
G1	125	nd	170	G11	145	anti[b]
A2	115	125	155	C12	107	120
A3	115	125	170	T13	125	120
G4	125	130	130	T14	130	120
A5	nd	125	150	C15	108	115
G6	125	125	150	T16	120	anti
A7	110	125	170	C17	130	115
A8	110	125	155	T18	115	120
G9	nd	128	150	T19	150	anti
C10	130	135	170	C20	115	130

[a] Torsion angles were determined by nonlinear regression to intranucleotide NOE time courses as described in Materials and Methods. Estimated errors are ±5° for deoxynucleotide and ±10° for ribonucleotides. Data refer to the italicized strand. [b] Based on H1′−H8 intranucleotide NOE intensity.

Glycosidic Torsion Angles. Intranucleotide NOE intensities were typical of the *anti* conformation about the glycosidic bond. For the deoxynucleotides, the H8/H6−H2′ NOEs were much larger than either H8/H6−H1′ or H8/H6−H3′, and consistent with χ in the range −90° to −120° in the DNA duplex, and −100° to −130° in the hybrid duplexes. For the ribonucleotides, the H8/H6−H1′ and H2′ NOEs were very weak compared with the H8/H6−H3′ NOE, as expected for a C3′-endo pucker and $\chi \approx -160°$. Where sufficient cross-peaks were resolved, the glycosidic torsion angles were analyzed in greater detail with NUCFIT (Lane, 1990) using the time courses of the NOEs from H8/H6 to H1′, H2′(H2″), and H3′. The deoxynucleotides were analyzed as a conformational blend of S and N states, with sugar conformations and the fraction of the S state sampled at values around those determined from the coupling constants (see above). The ribonucleotides were assumed to consist of a single conformation with $P \approx 9°$. The precise value of P within a conformational range (i.e., S or N) has little influence on the determination of the glycosidic torsion angle (Lane, 1990). The values of χ determined in this way are given in Table 4. Because there is a high ratio of data to parameters in these calculations, the glycosidic torsion angles are well determined. In general it can be seen that the torsion angles are highest for the deoxynucleotides in the DNA duplex, followed by the deoxynucleotides in the two hybrids, and are lowest in the ribonucleotides. Indeed, the glycosidic torsion angles of the ribonucleotides are similar whether they are in a pure RNA duplex or DNA·RNA hybrid, whereas the angles are significantly different between the pure DNA duplex and the two hybrid duplexes. This also suggests that much of the conformational differences between the hybrid duplexes arises from the nucleotide conformations in the DNA strands. This is consistent with the observed changes

Table 5: Relative H2′(*i*)−H8/H6(*i*+1) Distances in the RNA Strands[a]

rR10·rY10		rR10·*rY10*	
G1−A2	2.8	G11−C12	2.9
A2−A3	3.1	C12−U13	2.7
A3−G4	2.8	U13−U14	2.5
G4−A5	2.7	U14−C15	(2.5)[b]
A5−G6	3.0	C15−U16	2.6
G6−A7	2.7	U16−C17	(2.6)[b]
A7−A8	2.8	C17−U18	2.6
A8−G9	2.9	U18−U19	2.4
G9−C10	2.7	U19−C20	2.7
mean ± sd	2.8 ± 0.1		2.6 ± 0.1
rR10·dY10		dR10·*rY10*	
G1−A2	3.0	G11−C12	2.3
A2−A3	3.0	C12−U13	nd
A3−G4	2.7	U13−U14	2.3
G4−A5	2.8	U14−C15	nd
A5−G6	2.8	C15−U16	nd
G6−A7	2.7	U16−C17	nd
A7−A8	2.7	C17−U18	nd
A8−G9	2.8	U18−C19	2.2
G9−C10	2.6	U19−C20	2.2
mean ± sd	2.8 ± 0.1		

[a] Cross-peak volumes were measured using Felix 95.0 and normalized to the average of the nonterminal Cyt H6−H5 or UH6−H5 cross-peak volumes. The distances were calculated as $r = 2.45V_n^{-1/6}$, where V_n is the normalized volume, and extrapolating back to zero mixing time. Data refer to the italicized strand. nd, not determined. [b] Overlapping peaks.

in chemical shifts for the DNA protons between the hybrid duplexes and the DNA duplex (see above).

Global Conformations. The CD spectra indicate that the DNA duplex is B form and the RNA duplex is A form. This was confirmed by the NMR results, which showed nucleotide conformations typical for the B and A forms, respectively. In addition, the internucleotide NOE intensities were also consistent with B and A. The CD spectra of the two hybrids indicate the presence of conformations at the level of base-stacking interactions of conformations intermediate between the A and B structures. The electrophoretic mobilities, apparent correlation times, and ^{31}P NMR shift dispersions are consistent with a ranking of conformations from dR10·dY10 (most B-like) through dR10·rY10, rR10·dY10 to rR10·rY10 (most A-like).

The H2′(*i*)−H8/6(*i*+1) distance is short (<2.5 Å) in the A conformation and long (>3.5 Å) in the B-conformation, and therefore the corresponding NOE intensities provide a sensitive indicator of the structure of the ribose strands. We have measured the intensities of the H2′(*i*)−H8/H6(*i*+1) cross-peaks in the RNA strands as a function of mixing time. The distances have been calculated assuming a distance of 2.45 Å for the Cyt H6−H5 vector (Table 5). In the RNA·RNA duplex, the normalized cross-peak volumes for the pyrimidines are near unity for all mixing times (50−400 ms), which implies that the distance is about 2.5 Å. In contrast, the normalized volumes for the purines approach unity at long mixing times (250−400 ms) but decrease to ca. 0.2−0.3 at the shortest mixing time. This indicates that the H2′(*i*)−H8(*i*+1) distance in the purine strand is longer than that of the reference vector (2.45 Å) and is closer to 3 Å. We note that this is not predicted by the standard A structure ($r \approx 1.6$ Å) or even one that has been energy minimized ($r \approx$ 2.3 Å). This indicates some structural asymmetry between the two strands of this particular RNA duplex.

A similar pattern was observed in the two hybrids. The normalized H2′(*i*)−H8/H6(*i*+1) cross-peak intensity in rR10·dY10 decreased with decreasing mixing time (250 to 50 ms), reaching a limiting value of ca. 0.3 at short mixing times. This is equivalent to a distance of around 3−3.2 Å. In contrast, the normalized intensities for the pyrimidine strand in dR10·rY10 increased slightly with decreasing mixing time, reaching a limiting value of 1−1.5 at short mixing time, and consistent with a distance of 2.3−2.4 Å. Hence, the sequential H2′(*i*)−H8(*i*+1) and H2′(*i*)−6(*i*+1) are preserved in the RNA strands of the two hybrids. Overall, these distances indicate that the RNA strands are globally closer to the A conformation than to the B conformation, in the hybrids as well as in the RNA duplex.

In contrast, the H2′(*i*)−H8/H6(*i*+1) intensities in the DNA strands were weak compared with the H2″−H8/H6 intensities. In the pure DNA duplex, the observed intensity largely reflects spin diffusion via H2″. The observed H2″−H8/H6 distances were ca. 3 Å (not shown), which is longer than expected for standard B-DNA but is typical of DNA in solution.

These data, and the analysis of the conformations of the deoxy strands, clearly show significant conformational differences between the two DNA·RNA hybrids, and of course compared with the pure DNA and RNA duplexes. Further, the conformations of the RNA strands in the hybrids are similar to their conformations in the RNA duplex, whereas there are differences between the conformations of the DNA strands in the hybrids and the DNA duplex.

DISCUSSION

We have confirmed the extensive thermodynamic behavior of this class of DNA·RNA hybrids and also the global conformational features (Ratmeyer et al., 1994; Hung et al., 1994; Lesnik & Freier, 1995; Sugimoto et al., 1995; Wang & Kool, 1995). The present data show that the DNA duplex, as expected, is in the B family of conformations, and the RNA duplex is in the A conformation. Clearly the DNA·RNA hybrids, while overall more A-like than B-like, are in an intermediate state. Moreover, this intermediate state is clearly different for the two hybrids, both globally and at the nucleotide level in the DNA strands. According to the spin−spin coupling data, this is manifested in particular in the degree of flexibility of the DNA strands, which is greater in the hybrids than in the DNA duplex, and on average the deoxypyrimidine nucleotides are more flexible than the deoxypurines. Also, the changes in ^{1}H chemical shifts are much larger for the DNA strands than for the RNA strands, but the pattern of shifts for the two nucleotides is quite different. This indicates that the major conformational rearrangements between the pure duplexes and the hybrids occur in the DNA strands, and that the conformational changes are not the same in the two hybrid duplexes. The differences at the local level translate into differences in base stacking, as observed in the CD spectra and sequential NOE intensities, backbone conformation as seen in the ^{31}P NMR spectra, and overall in the differences in hydrodynamic and electrophoretic behavior. The differences in base-stacking may account, at least in part, for the observed differences in thermodynamic stabilities. It is clear that there is more

flexibility in the DNA·RNA hybrids than in either of the two pure duplexes, and presumably this in part affects the differences in stability through the configurational entropy changes between the strand and duplex states. We have detected experimentally flexibility at the nucleotide level. How widespread this is and the contributions to the different thermodynamic stabilities must await full three-dimensional structure determinations. These calculations are in progress. However, because of the effects of conformational averaging, rather more detailed structures calculations have to be carried out for these duplexes, such as the use of time averaged distance restraints as recently applied to a DNA·RNA hybrid (Gonzalez et al., 1995), or based on local conformational averaging, where appropriate ensembles of conformations are calculated that account for the experimental NOEs and coupling constants.

Based on our results and information in the literature regarding the solution conformation of DNA·RNA hybrids (Fedoroff et al., 1993; Lane et al., 1993; Gonzalez et al., 1994, 1995), it is likely that the backbone conformation and/or minor groove widths of the two classes of hybrid also differ significantly. If the activity of RNAseH is sensitive to minor groove width and backbone conformation, then we would expect that the enzyme would cleave rR10·dY10 at a different rate to dR10·rY10. It has been suggested that RNAse H activity increases the *in vivo* efficiency of antisense technology (by destroying the target mRNA and releasing the DNA) (Walder & Walder, 1988). This could provide an additional facet to the design of optimum antisense DNA molecules. It would therefore be of considerable practical interest to determine the relative rates of cleavage of mixed-sequence DNA·RNA hybrids versus the all-pyrimidine and all purine RNA strands.

ACKNOWLEDGMENT

We thank Dr. S. Martin for assistance with the circular dichroism. We are grateful to Dr. M. R. Conte for critical appraisal of the manuscript.

SUPPORTING INFORMATION AVAILABLE

Four tables and three figures of NMR assignments, UV melting curves, additional NOESY, and DQF-COSY spectra (14 pages). Ordering information is given on any current masthead page.

REFERENCES

Adams, R. L. P., Knowler, J. T., & Leader, D. P. (1986) *The Biochemistry of the Nucleic Acids*, 10th ed., Chapter 6, Chapman and Hall, London.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.
Birchall, A. J., & Lane, A. N. (1990) *Eur. Biophys. J. 19*, 73–78.
Breslauer, K. J., Frank, R., Blocker, H., & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3746–3750.
Brown, T., & Brown, D. J. S. (1991) in *Oligonucleotides and Analogues, a Practical Approach* (Eckstein, F., Ed.) pp 1–23, IRL Press, Oxford.
Brown, T., & Brown, D. J. S. (1992) *Methods Enzymol. 211*, 20–35.
Chou, S.-H., Cheng, J.-W., & Reid, B. R. (1992) *J. Mol. Biol. 228*, 138–155.
Conte, M. R., Bauer, C. J., & Lane, A. N. (1996) *J. Biomol. NMR 7*, 190–206.
Ebel, S., Brown, T., & Lane, A. N. (1994) *Eur. J. Biochem. 220*, 703–715.
Fasman, G. D., Ed. (1975) *Handbook of Biochemistry and Molecular Biology. Nucleic Acids*, Vol. I, 3rd ed., p 589, CRC Press, Boca Raton, FL.
Fedoroff, O. Y., Salazar, M., & Reid, B. R. (1993) *J. Mol. Biol. 233*, 509–523.
Feigon, J., Koshlap, K. M., & Smith, F. W. (1995) *Methods Enzymol. 261*, 225–255.
Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T., & Turner, D. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 9373–9377.
Forster, M. J., & Lane, A. N. (1990) *Eur. Biophys. J. 18*, 347–355.
Gonzalez, C., Stec, W., Kobylanska, A., Hogrefe, R. I., Reynolds, M., & James, T. L. (1994) *Biochemistry 33*, 11062–11072.
Gonzalez, C., Stec, W., Reynolds, M., & James, T. L. (1995) *Biochemistry 34*, 4969–4982.
Gorenstein, D. G. (1992) *Methods Enzymol. 211*, 254–286.
Hall, K. B., & McLaughlin, L. W. (1991) *Biochemistry 30*, 10606–10613.
Hansen, U. M., & McClure, W. R. (1980) *J. Biol. Chem. 255*, 9564–9570.
Hung, S.-H., Yu, Q., Gray, D. M., & Ratliff, R. L. (1994) *Nucleic Acids Res. 22*, 4326–4334.
Ivanov, V. I., Minchenkova, L. E., Minyat, E. E., Frank-Kamenetskii, M. D., & Schyolkina, A. K. (1974) *J. Mol. Biol. 87*, 817–833.
Lane, A. N. (1990) *Biochim. Biophys. Acta 1049*, 189–204.
Lane, A. N. (1994) *Prog. NMR Spectrosc. 25*, 481–505.
Lane, A. N., & Fulcher, T. (1995) *J. Magn. Reson. B 106*, 34–42.
Lane, A. N., Lefèvre, J.-F., & Jardetzky, O. (1986) *J. Magn. Reson. 66*, 201–208.
Lane, A. N., Jenkins, T. C., Brown, D., & Brown, T. (1991) *Biochem. J. 279*, 269–281.
Lane, A. N., Martin, S. R., Ebel, S., & Brown, T. (1992) *Biochemistry 31*, 12087–12095.
Lane, A. N., Ebel, S., & Brown, T. (1993) *Eur. J. Biochem. 215*, 297–306.
Legault, P., & Pardi, A. (1994) *J. Magn. Reson. 103B*, 82–86.
Lesnik, E. A., & Freier, S. M. (1995) *Biochemistry 34*, 10807–10815.
Oda, Y., Iwai, S., Ohtsuka, E., Ishikawa, M., Ikehara, M., & Nakamura, H. (1993) *Nucleic Acids Res. 21*, 4690–4695.
Pilch, D. S., Levenson, C., & Shafer, R. H. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1942–1946.
Piotto, M., Saudek, V., & Sklenar, V. (1992) *J. Biomol. Struct. Dyn. 2*, 661–665.
Ratmeyer, L., Vinayak, R., Zhong, Y. Y., Zon, G., & Wilson, W. D. (1994) *Biochemistry 33*, 5298–5304.
Salazar, M., Champoux, J. J., & Reid, B. R. (1993a) *Biochemistry 32*, 739–744.
Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S., & Reid, B. R. (1993b) *Biochemistry 32*, 4207–4215.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Stein, C. A., & Cheng, Y.-C. (1993) *Science 261*, 1004–1012.
Stein, H., & Hausen, P. (1969) *Science 166*, 393–395.
Sugimoto, N., Nakano, S.-i., Katoh, M., Matsamura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., & Sasaki, M. (1995) *Biochemistry 34*, 11211–11216.
van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1–38.
Varmus, H. (1988) *Science 240*, 1427–1435.
Walder, R. Y., & Walder, J. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5011–5015.
Wang, S. H., & Kool, E. T. (1995) *Biochemistry 34*, 4125–4132.
van Wijk, J., Huckriede, B. D., Ippel, J. H., & Altona, C. (1992) *Methods Enzymol. 211*, 286–306.
Williamson, J. R., & Boxer, S. G. (1989) *Biochemistry 28*, 2819–2831.
Wincott, F., DiRenzo, A., Shaffer, C., Grimm, S., Tracz, D., Workman, C., Sweedler, D., Gonzalez, C., Scaringe, S., & Usman, N. (1995) *Nucleic Acids Res. 23*, 2677–2684.

BI960948Z

Nucleic Acids Research, 1996, Vol. 24, No. 19 3693–3699

Hydration of the RNA duplex r(CGCAAAUUUGCG)$_2$ determined by NMR

Maria R. Conte, Graeme L. Conn[1], Tom Brown[1] and Andrew N. Lane*

Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK and [1]Department of Chemistry, University of Southampton, Southampton SO17 1BJ, UK

Received July 4, 1996; Revised and Accepted August 16, 1996

ABSTRACT

The so-called spine of hydration in the minor groove of AnTn tracts in DNA is thought to stabilise the structure, and kinetically bound water detected in the minor groove of such DNA species by NMR has been attributed to a narrow minor groove [Liepinsh, E., Leupin, W. and Otting, G. (1994) *Nucleic Acids Res.* 22, 2249–2254]. We report here an NMR study of hydration of an RNA dodecamer which has a wide, shallow minor groove. Complete assignments of exchangeable protons, and a large number of non-exchangeable protons in r(CGCAAAUUUGCG)$_2$ have been obtained. In addition, ribose C2′-OH resonances have been detected, which are probably involved in hydrogen bonds. Hydration at different sites in the dodecamer has been measured using ROESY and NOESY experiments at 11.75 and 14.1 T. Base protons in both the major and minor grooves are in contact with water, with effective correlation times for the interaction of ~0.5 ns, indicating weak hydration, in contrast to the hydration of adenine C2H in the homologous DNA sequence. NOEs to H1′ in the minor groove are consistent with hydration water present that is not observed in the analogous DNA sequence. Hydration kinetics in nucleic acids may be determined by chemical factors such as hydrogen-bonding more than by simple conformational factors such as groove width.

INTRODUCTION

Nucleic acids are strongly hydrated in both the crystal state and in solution and the degree of hydration determines the overall conformation. Hydration in DNA has been measured by a variety of methods including NMR (1–4), thermodynamics (5), and crystallography (6,7). Thus at low water activity, DNA adopts the A conformation, whereas at high water activity, it is found in the B conformation. In contrast, RNA is always in the A conformation. Chemically, the important difference between DNA and RNA is the 2′-OH on the sugar in RNA, though the methyl group of thymine affects primarily the thermodynamic stability of duplexes of either RNA or DNA (8).

Although there are numerous X-ray structures of DNA in both the B and A forms, there are no high resolution solution structures of DNA in the A form, and only a few solution structures of RNA. Many of the crystal structures have been reported on the hydration of DNA and RNA, and the results have been summarised in detail (9). In contrast, there have been only a few reports of hydration of nucleic acids using NMR (1–4,10) and none on RNA in solution.

The concentration of water is very high (~55 M), and therefore statistically can be expected to be close to all exposed sites on the surface of a solute particle. In the absence of any interaction between water and solute, the rate constant for dissociation of a water molecule from a solute surface depends on the diffusion constant, and can be expected to be ~2–3 × 10^9 s^{-1} at 10°C (i.e. an average residence time of ~0.3–0.5 ns) (11). The diffusion limited association rate constant will be of similar magnitude, i.e. 10^9–10^{10} M^{-1} s^{-1}, so that the effective dissociation constant is ~0.2–2 M. The free energy change is comparable with the free energy of mixing, and represents the case where water is not thermodynamically bound. However, at 55 M water, the occupancy of each exposed site will be >96%. Thus, water can be considered to be bound if the effective correlation time is significantly longer than ~0.5 ns.

A particularly important structural role for water has been proposed in the spine of hydration found in the narrowed minor groove of dAnTn sequences (6,7) which has been proposed to be responsible for the relatively slow exchange of NMR-visible water molecules from the neighbourhood of the adenine C2H (1,2). Because an RNA duplex has a very wide, shallow minor groove, the groove-width hypothesis predicts that there should be no slowly exchanging water molecules in the minor groove. However, it is possible that other factors such as hydrogen bonding are important. Thus, the minor groove of RNA is lined with C2′-OH groups, and is therefore relatively more hydrophilic than the minor groove of DNA. It would therefore be expected that the minor groove of RNA should show extensive contact with water, particularly around the C2′ positions. We have therefore examined the hydration of r(CGCAAAUUUGCG)$_2$. The hydration properties of this molecule can be directly compared with the DNA analogue d(CGCAAATTTGCG)$_2$, which has been extensively studied by X-ray diffraction (12) and NMR methods (13).

MATERIALS AND METHODS

r(CGCAAAUUUGCG) was synthesised using phosphoramidite chemistry and purified by reversed-phase HPLC (14). A quantity of 112 A_{260} units of the purified and annealed dodecamer were

* To whom correspondence should be addressed

dissolved in 0.6 ml 90% H_2O:10% D_2O containing 0.01 M sodium phosphate, 100 mM KCl, 0.2 mM EDTA and 0.1 mM DSS, pH 7.

^1H NMR spectra were recorded at 14.1 T on a Varian Unity NMR spectrometer and at 11.75 T on a Varian UnityPlus spectrometer. 2D NMR spectra were recorded in the phase-sensitive mode (15). Spectra in H_2O were recorded using the Watergate pulsed gradient method for solvent suppression (16) with acquisition times of 0.4 s in t_2 and 0.05 s in t_1. NOESY spectra were obtained using mixing times of 25, 50, 100 and 250 ms. A ROESY spectrum was obtained with a mixing time of 25 ms, and a spin-lock field strength of 4.5 kHz.

NOESY spectra in D_2O were recorded at 30°C with acquisition times of 0.7 s in t_2 and 0.06 s in t_1, with mixing times of 50, 100 and 250 ms. A 2-quantum COSY experiment (17) was recorded with a mixing pulse of 135°, acquisition times of 0.5 s in t_2 and 0.04 s in t_1, with a 2-quantum creation time of 40 ms. In this experiment, cross-peaks are disposed either side of the two-quantum diagonal, which is free of peaks. Further, the 135° pulse suppresses the remote connectivities. The spectrum produced in this experiment was phased to pure absorption along F2, and absolute value along F1. Data matrices were transformed as 16384 by 2048 complex points, using a Gaussian function for apodisation in both dimensions.

Hydration was assessed by observing both NOEs and ROEs from water to different protons in cross-sections along F2 at the water frequency. In the absence of spin diffusion, ROEs are positive, and the cross-peaks have the opposite sign to the diagonal. Chemical exchange peaks in ROESY spectra are negative, and have the same sign as the diagonal. In NOESY spectra, exchange cross-peaks are negative, and in general, NOEs are also negative unless the effective correlation time is short (less than ~0.3 ns). The ratio of the cross-relaxation rate constants in the laboratory and rotating frames, R, depends only on correlation times and (known) Larmor frequencies. Thus, for a tightly bound water molecule (i.e. for which the residence time is longer than the rotational correlation time), the cross relaxation rate constants in the two experiments are given by:

$$\sigma_1 = a[6J(2\omega) - J(0)]/r^6 \quad 1$$

$$\sigma_r = a[3J(\omega) + 2J(0)]/r^6 \quad 2$$

$$R = [6J(2\omega) - J(0)]/[3J(\omega) + 2J(0)] \quad 3$$

where a is a nuclear constant and $J(\omega)$ are the spectral density functions.

Equation 3 predicts that the ratio R, will approach –0.5 for long correlation times, as in macromolecules. If the bound water molecule can undergo rapid, large amplitude motions, the spectral density functions become more complex. For sufficiently rapid internal motions, each spectral density function can be decomposed into two parts, corresponding to overall rotation of the complex, and internal mobility as (18):

$$J(\omega,\tau) = S^2J(\omega,\tau_0) + (1 - S^2)J(\omega,\tau_e) \quad 4$$

where S^2 is the order parameter, $\tau_e = \tau_0\tau_i/(\tau_0+\tau_i)$ and τ_0, τ_i are the correlation times for global rotation and internal motion, respectively. If the internal motion is of large amplitude, such that $S^2 = 0$, the form of the spectral density function reduces to that of a rigid rotor, except that it is determined by τ_e rather than τ_0. In this case, the value of R can vary between unity (very short correlation times) and –0.5 (long correlation times). This allows NOEs to become positive. An alternative is that water molecules are not strongly bound, but diffuse in and out of the molecular surface on a time-scale comparable with the global correlation time. Models of translation diffusion give quite complex spectral density functions (19), but the ratio, R, still depends only on the correlation time and Larmor frequency (and not on distances) (19–21). For this model, the correlation time is related to the diffusion coefficient of the diffusing water molecule, which may be smaller than that of pure water if there is an interaction between the water molecules and the solute particle. Hence, these models can provide limits on the effective correlation time for water-solute interactions as probed by NMR.

Experimentally, the ratio R was determined by measuring cross peak areas in cross-sections at the water frequency, and normalising them to the intensity of Cyt or Uri H5–H6 cross peaks. This eliminates the effects of autorelaxation and non-uniform excitation with the Watergate pulse. Corrections were made for off-resonance effects in the ROESY spectra as described (22,23). Effective correlation times were then determined using equation 3 for both rotational and translational diffusion models. Where absolute volume ratios were used, corrections were also made for the Watergate excitation profile, which is particularly important for the H1′ resonances.

Magnetisation time courses were calculated either analytically for three-spin systems, or by numerical integration of the appropriate Bloch equations for complete spin-systems, using the program NUCFIT (24). Coordinates were obtained for energy-minimised RNA using DISCOVER (Molecular Simulations, San Diego). Calculations were carried out assuming a correlation time of 5 ns and a recycle time of 3.5 s. The complete spin-system calculations also provided values for the autorelaxation rate constants of each proton.

RESULTS

NMR assignments

Non-exchangeable protons were assigned using a combination of NOESY and DQF-COSY and 2-quantum COSY in D_2O solution. Figure 1 shows a NOESY spectrum in D_2O recorded at 14.1 T and 30°C. It was possible to trace the sequential base–H1′–base proton connectivities. In addition, the adenine C2H resonances (and see below) showed three cross-peaks. The weak cross-peak corresponds to the intraresidue interaction with H1′, and the two strong peaks correspond to the sequential H2(i)–H1′(i+1) and the cross-strand H2(i)–H1–(i–1) interactions. The latter cross-peak was more intense than expected for standard A-RNA, where this distance is ~4.5 Å, compared with the sequential H2–H1′ distance of ~3.6 Å. The H2′ resonances were assigned using the strong H1′–H2′ NOE cross-peaks, and also the sequential pathways H1′(i)–H2′(i)–H8/H6(i+1) pathway. H3′ peaks were assigned from both the NOESY, using the intraresidue H8/6(i)–H′(i) and sequential H3′(i)–H8/H6(i+1) cross-peaks, and the H2′–H3′ cross-peaks in the DQF-COSY and DQ-COSY spectra (not shown). The H4′ resonances were initially assigned from traces through H1′ of NOESY spectra. The most intense peak in this cross-section is the H2′, followed by the H4′ and H3′ peaks. The H4′ resonance was assigned by elimination, and by H3′–H4′ cross-peaks in the DQF-COSY spectrum. The H3′ and H4′ resonances were confirmed using the 2-quantum COSY experiment, which revealed scalar interactions between protons of similar chemical shifts (not shown). The 2Q-COSY and DQF-COSY spectra also revealed numerous cross-peaks that

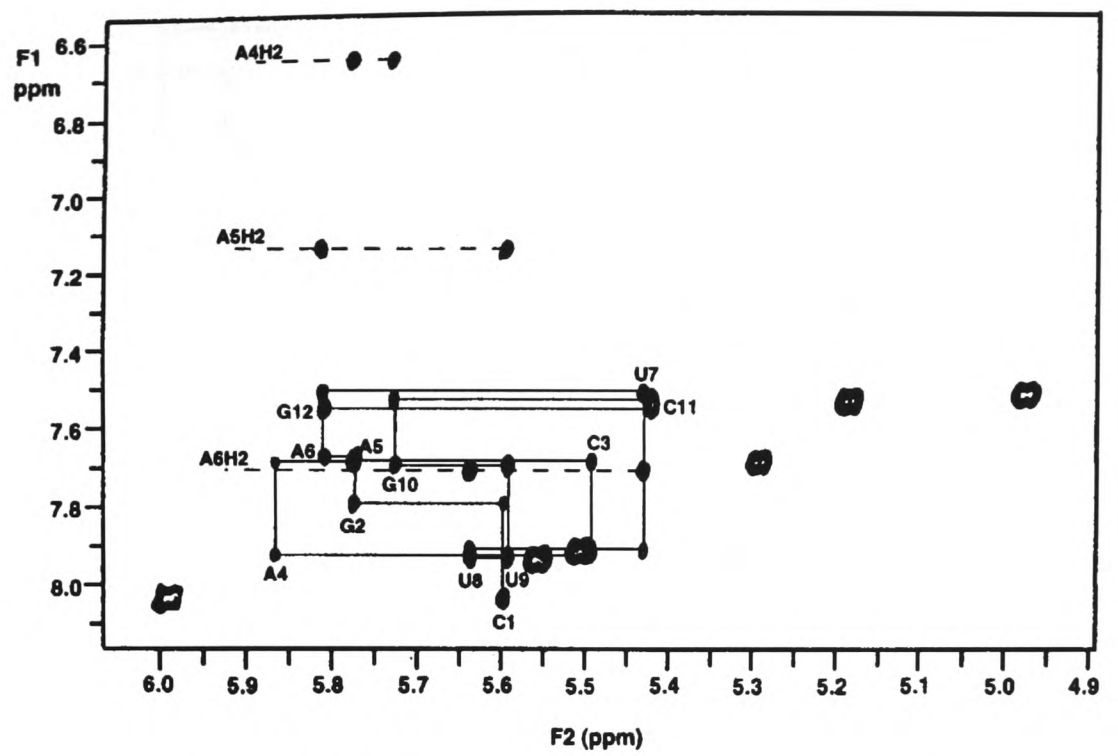


Figure 1. NOESY spectrum of r(CGCAAAUUUGCG)$_2$ in D_2O. The spectrum was recorded at 30°C, 14.1T as described in Materials and Methods, with a mixing time of 250 ms.

could arise only from inequivalent H5′/H5″, which in many instances had large differences in chemical shifts. In some instances, these could be independently verified from NOESY spectra by observing H6–H5/H5″ and H1′–H5′/H5″ (cf. U7 and C11). The assignments are given in Table 1.

Figure 2A shows a NOESY spectrum recorded in H_2O at 10°C and 14.1 T. The two-fold symmetry of the duplex was confirmed by the observation of six resonances in the 12–14 p.p.m. region of the spectrum. The broadest resonance at 12.47 p.p.m. was assigned to the terminal G12N1H. The remaining imino protons were assigned from the NOE between neighbouring imino protons, and in the case of UN3H, by strong NOEs to the base-paired AC2H, which were independently assigned from NOEs to H1′ in the D_2O NOESY spectra.

Cytosine amino protons were assigned according to the strong NOE between the hydrogen bonded amino proton and the N1H of the base-paired G. The non-H-bonded amino proton was then found from the strong NOE within the NH_2 group. These assignments were further verified by the strong NOE from the upfield amino proton to the H5 of the same nucleotide, and the weaker NOE to the downfield amino proton.

Table 1. ^1H assignments in r(CGCAAAUUUGCG)$_2$

Base	H8/6	H5/Me H2	H1′	H2′	H3′	H4′	NH	NH_2
C1	8.04	5.99	5.60	4.52	4.56	4.34	–	8.18, 7.00
G2	7.80	–	5.77	4.57	4.69	(4.50)	13.08	n.d.
C3	7.69	5.29	5.49	4.52	(4.56)	(4.44)	–	8.39, 6.79
A4	7.93	6.67	5.87	4.58	4.73	4.50	–	7.71, 6.35
A5	7.69	7.15	5.77	4.52	4.64	(4.50)	–	7.80, 6.75
A6	7.68	7.72	5.81	4.42	4.50	nd	–	7.96, 6.77
U7	7.52	4.97	5.43	4.34	4.38	(4.42)	13.98	–
U8	7.92	5.51	5.64	4.41	4.48	nd	13.77	–
U9	7.95	5.56	5.59	4.52	4.54	4.46	13.12	–
G10	7.70	–	5.73	4.42	4.57	4.52	12.44	7.95, 5.80
C11	7.54	5.19	5.42	4.26	4.42	4.34		8.31,6.75
G12	7.56	–	5.81	4.08	4.26	4.19	12.88	n.d.

C1 (5′,5″ 3.94,4.04) p.p.m.; C2′-OH 6.7, 6.9 p.p.m. Proton assignments are given for 30°C, except for exchangeable protons which are given for 10°C. Shifts in parentheses are tentative. n.d., not determined.

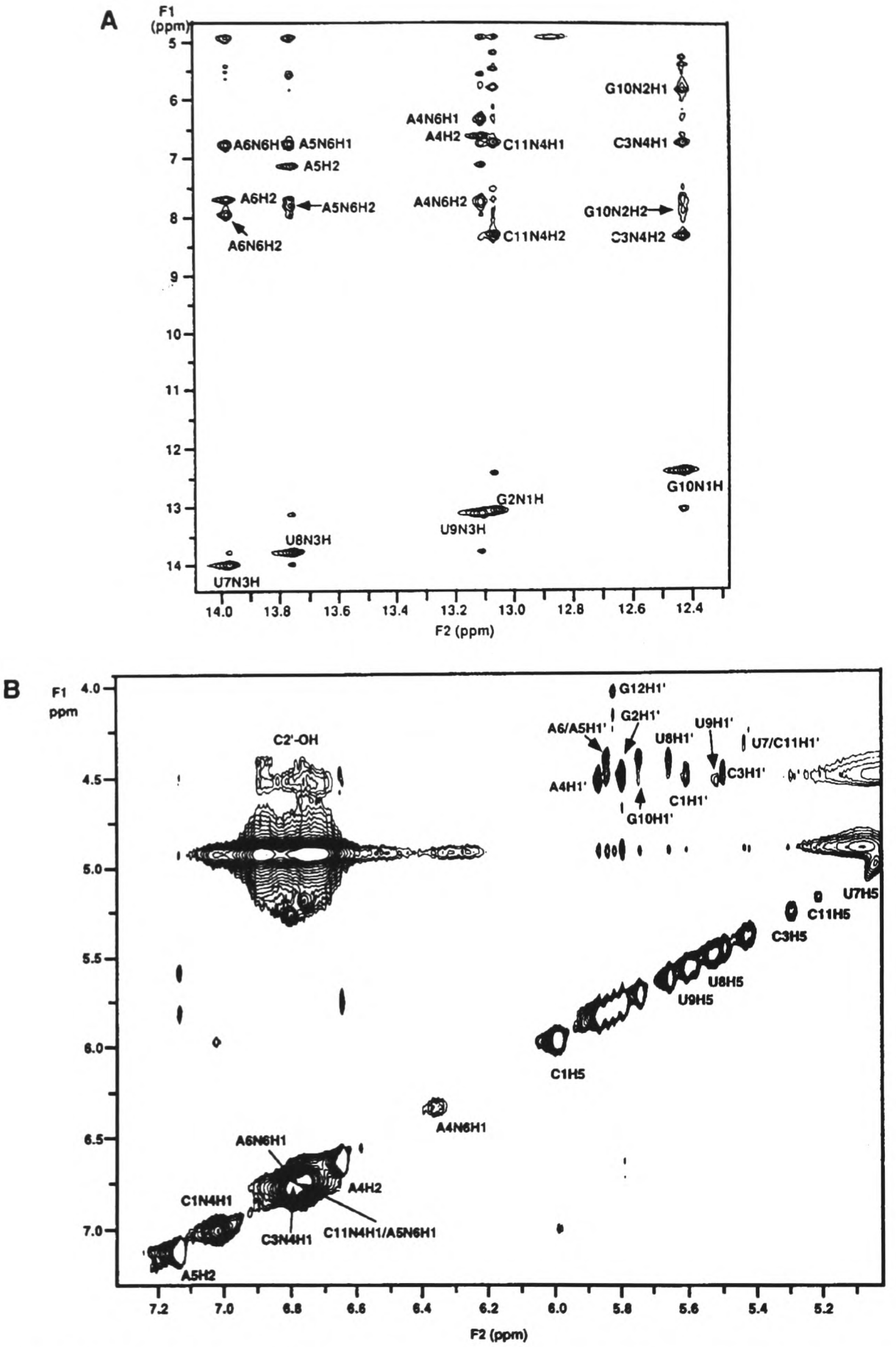


Figure 2. NOESY spectrum of r(CGCAAAUUUGCG)$_2$ in H_2O. The spectrum was recorded at 10°C as described in the Materials and Methods. (**A**) Imino and amino protons (mixing time = 250 ms, 14.1 T). (**B**) Hydroxyl protons (mixing time = 50 ms, 11.75 T).

Medium strength NOEs were also observed between pairs of exchangeable protons ($\delta \approx 8$ and 6.7) with UN3H. These resonances also showed a very intense mutual NOE, and can be identified as the N6 amino protons of the base-paired adenines. The ANH_2 were then assigned using the method described for the cytosine amino protons (see above). The amino protons of G10

were found in a similar manner, but the amino protons of G2 and G12 were not assigned. The assignments of the exchangeable protons are collected in Table 1.

As Table 1 and Figure 2B show, the non-hydrogen-bonded amino protons resonate between 6.3 and 7 p.p.m. In the 1D spectrum, there were two intense peaks at 6.75 and 6.85 p.p.m. The former contains the amino protons of A5, A6, C3 and C11, and the latter no amino protons. Hence, these two peaks should account for only four protons. However, the integral of the two peaks was at least 10 protons, which indicates that there must be at least another six exchangeable protons in this region. The only remaining candidate amino protons are from G2 and G12. Models of A-RNA showed that there are no amino protons within 5 Å of H2′ or H3′ within nucleotides in the A conformation and the closest amino proton is the non H-bonded N6H2 of G to the neighbouring cross-strand H2′ of the preceding residue, which is ~4.3 Å. The NOE calculated for the A-RNA structure, excluding chemical exchange of the amino proton with water, was very small. Also, in the DNA analogue of this sequence, there were no protons that exchange with water under similar conditions (Lane, Frenkiel and Jenkins, unpublished data). However, the C2′-OH is close to the C2′-H, and NOE cross-peaks were observed between the resonances near 6.85 p.p.m. and peaks at 4.3–4.5 p.p.m. (Fig. 2B), which correspond to H2′ and H3′ (Table 1). Further, it has been reported that the C2′-OH in RNA resonates at ~ 6.7 p.p.m. (25), and between 6.1 and 6.9 p.p.m. in simple carbohydrates, where they can be observed only at low temperatures (26). We conclude that the additional intensity near 6.7–6.8 p.p.m. arises from at least six C2′-OH. However, the resolution of the spectrum in this region and in the C2′-H region prevented us making firm sequence-specific assignments. Nevertheless, it is clear that the RNA OH exchange quite slowly under the conditions of these experiments, suggesting that these proton are involved in H-bonding interactions, such as to the nearby O3′ or to the O4′ of the next sugar (25,27–29).

Hydration

Figure 3 shows cross-sections at the water frequency from NOESY and ROESY spectra of r(CGCAAAUUUGCG)$_2$ recorded at 10°C and 14.1 T, using mixing times of 50 and 25 ms, respectively. Imino protons, which resonate between 12 and 14 p.p.m. (not shown) showed some exchange with water, in the order: G12N1H>U7N3H≈U8N3H>U9N3H≈G2N1H>G10N1H(≈0). The N4H2 (hydrogen bonded amino proton) of C3 and C11 showed essentially no exchange whereas that of C1 exchanged extensively with the water. The intense peaks between 6.6 and 6.9 p.p.m. arise from chemical exchange primarily of C2′-OH (see above) with water. The relatively strong peak at 7.58 p.p.m. is predominantly a direct NOE between U7H5 (δ = 4.97 p.p.m.) and U7H6 (cf. Table 1). However, it provides a useful internal intensity standard, as the distance between these protons is 2.42 Å.

The degree of hydration at different non-exchangeable protons was assessed from the sign and magnitude of water-solute NOE in cross-sections at the water frequency as described in Materials and Methods. Values of R for each resolved resonance are given in Table 2. At 14.1 T, some of the H8/H6 showed negative NOEs and positive ROEs (Fig. 3). In general the absolute values of R are smaller at 11.7 T than at 14.1 T, which is characteristic of correlation times close to 0.5 ns. These kinds of NOEs are similar to those found in DNA duplexes (1–4). The adenine C2H-water

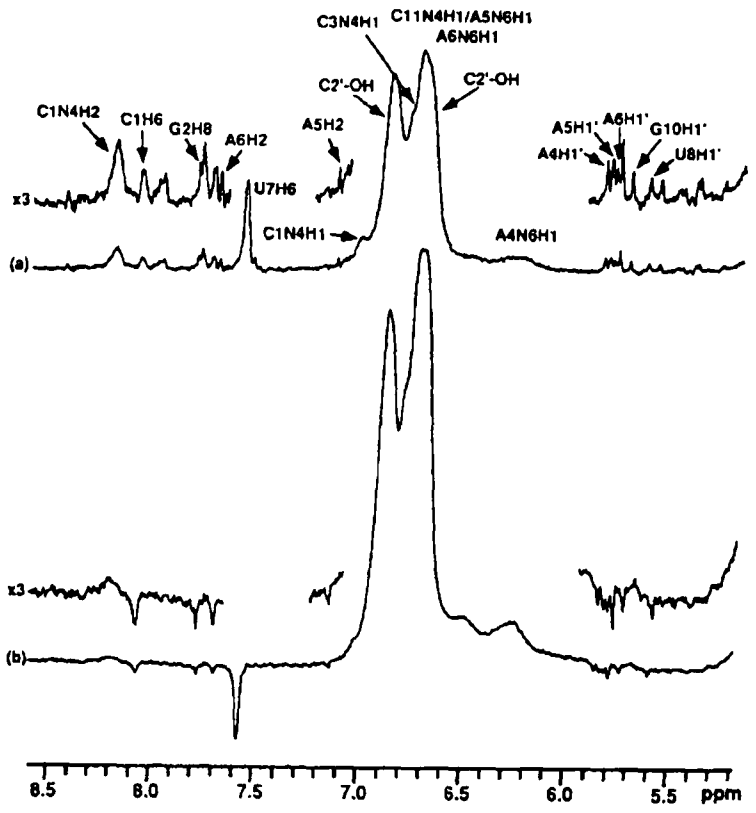


Figure 3. Cross-sections at the water frequency in NOESY and ROESY spectra of r(CGCAAAUUUGCG)$_2$. Upper trace NOESY (50 ms mixing time), lower trace ROESY (25 ms mixing time). The large peak at 7.58 p.p.m. arises from cross-relaxation of U7H6 with U7H5 (δ = 4.98 p.p.m.).

NOE cross-peaks were also very small indicating a short correlation time near 0.5 ns. The normalised cross-peak intensities in the NOESY spectrum (~0.06 at 600 MHz), and R values were quite small for these peaks compared with those of the analogous DNA duplex (Lane, Frenkiel and Jenkins, unpublished). The upper limit to correlation time for the water–C2H vector is equal to that for overall rotation of the molecule (i.e. when the water is irrotationally bound). For this case, as the observed NOE was 0.06 as intense as that of the cytosine H6–H5 NOE, we can calculate an upper limit to the water C2H distance of as NOE(water–C2H)/NOE(CytH6–H5)≈[(r(H6–H5)/r(water–C2H)]^6 for a short mixing time, assuming r(H6–H5) = 2.42 Å. However, under these conditions, the NOE would be independent of the magnetic field strength, in contradiction with the results (Table 2). Hence, the effective correlation time must be short, and which requires a distance shorter than 3.8 Å. The small effective correlation times for both AC2H and H8/H6 are consistent with rapid, large-scale internal motion of the water or a short residence time (or both). In either interpretation, these water molecules can be considered to be at most very weakly bound. This indicates that in RNA, both the major groove, and at least some aspects of the minor groove, are weakly hydrated. This is in contrast with the DNA analogue, where the minor groove is relatively strongly hydrated in the AnTn tract, as observed also by others for different sequences (1–4).

Cross-peaks to the H1′ resonances were observed, which are not usually seen in DNA except at the ends of the duplex or at much longer mixing times (3). Note that the resonances between 5.5 and 6 p.p.m. are significantly attenuated by the Watergate pulse, and so appear less intense than they should. In principle, these latter NOE peaks could arise via several mechanisms. First

there could be a water molecule with a long residence time close to H1′. Second there could be an NOE between the C2′-OH and the C1′-H with rapid exchange between the C2′-OH and water protons. Both mechanisms would give rise to a negative NOE and positive ROE. Other mechanisms involving two sequential NOE transfers from water to C2′-OH to C1′-H would give rise to a negative ROE, in contrast to the observations (Fig. 3 and Table 2). There are also no H2′ that coresonate with the water that could give rise to a direct NOE (Table 1). We have assigned at least six C2′-OH protons to the peaks at 6.7 and 6.85 p.p.m. (see above), which are therefore in slow exchange with water on the chemical shift time scale. It has been reported that the exchange rate of the secondary alcohol proton should be around 10–20 s^{-1} in RNA at 10°C (25). However, because there were only very weak NOEs from the hydroxyl protons to H1′ at a mixing time of 50 ms, and no detectable NOE at 25 ms (Fig. 2), most of the magnetisation transfer from water to H1′ cannot be via the route $H_2O \rightarrow$C2′-OH→H1′. We have calculated the magnitude of the exchange mediated NOE for different geometries. The OH protons can in principle be in Van der Waals contact with H1′, which would give rise to an extremely intense OH–H1′ NOE (and also the exchange mediated NOE from water), which was not observed. For an exchange rate constant up to at least 150 s^{-1}, calculations (not shown) indicate that the normalised water–H1′ NOE for this mechanism should increase markedly with increasing mixing time, whereas the observed NOE was essentially independent of the mixing time. Only if the exchange rate constant is very large (when the OH peak merges with the solvent peak) does the lag in the NOE build up curve become undetectable. In this case, the observed magnetisation transfer from water to H1′ depends exclusively on the cross-relaxation rate constant C2′OH–H1′. Indeed, the small observed NOE indicates that the C2′-OH must be pointing away from the H1′. Most significantly, whereas there are signifcant cross-peaks between H1′ and water, there are no cross-peaks of comparable or greater intensity between the C2′-OH and the H1′, as would be required for the exchange mediated pathway.

Table 2. Effective water–r(CGCAAAUUUGCG)$_2$ correlation times

	R		τ_{eff}/ns
	14.1 T	11.75 T	
G2H8	<0	≈0	0.3–0.8
G10H8	<0	≈0	0.3–0.8
A5H2	–0.2	≈0	0.3–0.8
A6H2	–0.3	–0.13	0.6–1
A4H1′	–0.5	–0.38	>0.5[a]
A5H1′	–0.47	[b]	>0.5[a]
A6H1′	–0.44	[b]	>0.5[a]
U8H1′	–0.5	[c]	>0.5[a]
G10H1′	–0.42	–0.28	>0.5[a]

[a]See text.
[b]Peaks overlap at 500 MHz.
[c]Peak too small to quantitate.
R is the ratio of the cross-relaxation rate constant in the laboratory frame to that in the rotating frame. τ_{eff} is the effective correlation time. R was determined at 10°C as described in Materials and Methods.

To reduce the contribution from exchange further, we have recorded a NOESY spectrum at 5°C with a mixing time of 30 ms: At the lower temperature, cross-relaxation rates should increase by ~17%, and decrease the OH exchange rate by a factor of ~2. Comparison of the intensity of the peaks with and without presaturation showed a reduction of ~60%, indicating an exchange rate constant of similar magnitude to the spin-lattice relaxation rate constant. As the latter is <10 s^{-1}, the exchange rate constant should also be < 10 s^{-1}. Water–H1′ cross-peaks were again present, and of an intensity ~30–40% of the U7 H5–H6 cross-peak intensity.

It is notable that whereas there were NOEs between the C2′-OH and C2′-H for most if not all nucleotides, there was no such NOE for the well-resolved G12, for which there is no 3′-phosphate or possible 3′ nucleotides to protect this terminal OH, which is not visible under these conditions. Hydrogen bonding of the (NMR visible) hydroxyl protons would place them ~3.4 Å from H1′, and give rise to a very small direct NOE, and therefore a very much smaller exchange mediated NOE. This suggests that a significant fraction of the observed transfer of magnetisation from water to H1′ in this RNA molecule is direct. As the observed NOE was negative, the effective correlation time for the water–H1′ interaction must be longer than ~0.5 ns (Table 1).

DISCUSSION

We have shown that C2′-OH can be observed in RNA under appropriate conditions, and that they are probably hydrogen bonded to groups within the RNA molecule. This indicates that additional information is available about RNA conformation and stability. Measurements of the NOEs involving the ribose hydroxyl protons should assist in the calculation of structures.

Both grooves of the RNA are clearly in contact with water, as has been observed by NMR for DNA. RNA has a deep major groove and a wide, shallow minor groove. Further, in RNA, the minor groove should be quite hydrophilic owing to the presence of the C2′-OH groups. However, the presence of water molecules close to H1′ occurs despite the wide and shallow minor groove of RNA. This is in contrast to the DNA analogue, d(CGCAAATTTGCG)$_2$, which is the B form and has a narrow minor groove in the A3T3 tract (12,13). Slowly exchanging water molecules have been detected in the minor groove in the region of the adenine residues of the related d(CGCGAATTCGCG)$_2$, and more rapidly exchanging water elsewhere (2,3). In DNA there is no nearby hydroxyl group to stabilise a water molecule, and under conditions similar to those used in this work, no NOEs were observed between water and H1′ (Lane, Jenkins and Frenkiel, unpublished). Because the hydroxyl proton exchanges with the solvent, it must be both accessible to, and interact with water, which indicates hydration at this site. Our results indicate that it is possible for water protons to be sufficiently close to H1′ for a direct NOE to be observed, and that hydrogen bonding to the hydroxyl oxygen may be responsible for the relatively long residence time, compared with DNA. This would be consistent with recent X-ray studies of RNA duplexes where water molecules interacting directly with the C2′-OH have been observed (27–29).

The reason for slow exchange in dAnTn tracts may be related more to the high propeller twist in these sequences, rather than the narrowed minor groove *per se*, which allows bridging of water molecules between neighbouring base pairs. This interaction is not possible in mixed-sequence DNA or RNA, where as we have shown the AdeC2H is only weakly hydrated. However, even in the narrowed minor groove of AnTn DNA tracts, hydration near to H1′ is very weak or non-existent, whereas in RNA it is substantial, and can be attributed to the presence of the OH group. The water close to this group is relatively long lived, which does

not correlate with a narrow minor groove. Although only a lower limit to the water residence time can be obtained by NMR, we imagine that even the long-lived molecules have residence times in nanoseconds rather than much longer as is sometimes observed in proteins (20). This is supported by the observed dependence of the ratio R on the magnetic field strength. For correlation times longer than ~2 ns, R becomes independent of the magnetic field strength, whereas we observed a significant magnetic field dependence for all protons (Table 2).

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council of the UK, by a Wellcome Travelling Research Fellowship to MRC and a Royal Society of Edinburgh Caledonian Research Fellowship to GC. We thank Dr J. Feeney for helpful discussions.

REFERENCES

1 Liepinsh, E., Otting, G. and Wüthrich, K. (1992) *Nucleic Acids Res.* **20,** 6549–6553.
2 Liepinsh, E., Leupin, W. and Otting, G. (1994) *Nucleic Acids Res.* **22,** 2249–2254.
3 Kubinec, M.G. and Wemmer, D.E. (1992) *J. Am. Chem. Soc.* **114,** 8739–8740.
4 Fawthrop, S.A., Yang, Ji-Chun and Fisher, J. (1993) *Nucleic Acids Res.* **21,** 4860–4866.
5 Chalikian, T.V., Plum, G.E., Sarvazyan, A.P. and Breslauer, K.J. (1994) *Biochemistry* **33,** 8629–8640.
6 Drew, H.R. and Dickerson, R.E. (1981) *J. Mol. Biol.* **151,** 535.
7 Privé, G.G., Heinemann, U., Chandrasegaran, S., Kan, L-S., Kopka, M.L. and Dickerson, R.E. (1987) *Science* **238,** 498–504.
8 Wang, S. and Kool, E.T. (1995) *Biochemistry* **32,** 4125–4132.
9 Berman, H.M. (1994) *Curr. Opin. Struct. Biol.* **4,** 345–350.
10 Radhakrishnan, I. and Patel, D.J. (1994) *Structure* **2,** 395–405.
11 Amdur, I. and Hammes, G.G. (1966) *Chemical Kinetics: Principles and Selected Topics*. Chapter 2. McGraw Hill, New York, USA.
12 Edwards, K.J., Brown, D.G., Spink, N., Skelly, J.V. and Neidle, S. (1992) *J. Mol. Biol.* **226,** 1161–1173.
13 Jenkins, T.C., Brown, D.G., Neidle, S. and Lane, A.N. (1993) *Eur. J. Biochem.* **213,** 1175–1184.
14 Ebel, S., Brown, T. and Lane, A.N. (1994) *Eur. J. Biochem.* **220,** 703–715.
15 States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) *J. Magn. Reson.* **48,** 286–292.
16 Piotto, M., Saudek, V. and Sklenar, V. (1992) *J. Biomol. Str.* **2,** 661–665.
17 Mareci, T.H. and Freeman, R. (1983) *J. Magn. Reson.* **51,** 531–535.
18 Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.* **104,** 4559–4570.
19 Ayant, Y., Belorizky, E., Fries, P. and Rosset, J. (1977) *J. Phys. (Paris)* **38,** 325–337.
20 Abragam, A. (1965) *Principles of Nuclear Magnetism*. Ch. VIII. Oxford University Press, Oxford, UK.
21 Otting, G., Liepinsh, E. and Wüthrich, K. (1991) *Science* **254,** 974–980.
22 Griesinger, C. and Ernst, R.R. (1987) *J. Magn. Reson.* **75,** 261–271.
23 Leeflang, B.R. and Kroon-Batenburg, L.M.J. (1992) *J. Biomol. NMR* **2,** 495–518.
24 Lane, A.N. (1990) *Biochim. Biophys. Acta* **1049,** 189–204.
25 Leroy, J-L., Broseta, D. and Guéron, M. (1985) *J. Mol. Biol.* **184,** 165–178.
26 Poppe, L. and van Halbeek, H. (1994) *Nature Struct. Biol.* **1,** 215–216.
27 Leonard, G.A., McAuley-Hecht, K.E., Ebel, S., Lough, D.M., Brown, T. and Hunter, W.N. (1994) *Structure* **2,** 483–494.
28 Egli, M., Portmann, S. and Usman, N. (1996) *Biochemistry* **35,** 8489–8484.
29 Wahl, M.C., Ban, C., Sekharudu, C., Ramakrishnan, B. and Sundaralingam, M. (1996) *Acta Cryst.* **D52,** 655–667.