

STABILITY STUDIES OF SHORT, IMPERFECT  
RNA DOUBLE HELICES

By



PAUL JOHN ROMANIUK, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

October, 1979

STABILITY STUDIES OF SHORT, IMPERFECT  
RNA DOUBLE HELICES

DOCTOR OF PHILOSOPHY (1979)  
(Biochemistry)

McMASTER UNIVERSITY  
Hamilton, Ontario

TITLE: Stability Studies of Short, Imperfect RNA Double  
Helices

AUTHOR: Paul John Romaniuk, B.Sc. (McMaster University)

SUPERVISOR: Dr. Thomas Neilson

NUMBER OF PAGES: xvi, 182

## ABSTRACT

The stability of short, imperfect RNA double helices was investigated by proton magnetic resonance spectroscopy, since this physical technique can give conformational information of solution species at the individual nucleoside level. The short oligoribonucleotides used in these studies were chemically synthesized by the phosphotriester method of Neilson and associates, which is capable of preparing a wide variety of sequences in the quantities required for NMR spectroscopic experiments. The work reported in this thesis was the first to combine chemical synthesis of oligoribonucleotides with pmr spectroscopic techniques to investigate systematically the effect of imperfections on the stability of short RNA duplexes:

1. The mechanism by which a dangling base region stabilizes an adjacent duplex region was studied. The results of this study demonstrated, for the first time, that a dangling base stabilizes a double helix by increasing favourable base stacking interactions without reducing fraying of terminal base pairs.
2. The duplex set AGGA:UCCU was studied as a model of the Shine-Dalgarno mRNA-16 S rRNA binding site. The AGGA:UCCU duplex was more stable than two other duplexes with identical base composition, CAUG and AGCU. Unlike AGCU, the AGGA:UCCU double helix does not fray at the terminal A·U base pairs. This result was the first to demonstrate that fraying is a sequence related phenomenon.

3. The effect of small loops on the stability of short duplexes was investigated by the pentanucleotides CAXUG (X = A,G,C,U). These pentamers did not form double helices with a non-bonded base opposition under conditions where CAUG exists as a stable duplex. This result suggests that short duplex regions separating small loops may be more susceptible to transient opening than other duplex regions in native RNAs.
4. The study of the duplex set CAGUG:CAUUG was the first to investigate the formation of a G·U wobble base pair within a regular Watson-Crick base paired double helix. The use of pmr spectroscopy allowed the observation that the G·U base pair was less stable than the surrounding Watson-Crick base pairs.
5. Certain sequences, for example CGCG, are capable of forming both a perfect helix and a staggered helix. The results of a pmr study of CGCG, which can only form a staggered duplex, demonstrated that the staggered duplex formed in solution must have a special conformation different from the conformation of a perfect helix. The pmr study of CGCG was probably complicated by the existence of a complex series of equilibria between random coil, perfect double helix and staggered duplex. Sequences which can only form a perfect duplex should be chosen for study to keep the helix-coil transition under study relatively simple.

Some of the results presented in this thesis have appeared in the literature:

1. P.J. Romaniuk, D.W. Hughes, R.J. Gregoire, T. Neilson and R.A. Bell.  
"Contribution of a G·U Base Pair to the Stability of a Short RNA

Helix." (1979) J. Chemical Soc. Chemical Commun., in press.

2. P.J. Romaniuk, D.W. Hughes, R.J. Gregoire, T. Neilson and R.A. Bell.

"Stabilizing Effect of Dangling Bases on a Short RNA Double Helix as Determined by Proton Nuclear Magnetic Resonance Spectroscopy."

(1978) J. American Chemical Soc. 100, 3971.

3. P.J. Romaniuk, T. Neilson, D.W. Hughes and R.A. Bell. "Proton Magnetic Resonance Studies on Short Duplexes. .II. The Self-Complementary Oligoribonucleotide CpApUpG." (1978) Can. J. Chem.

56, 2249.

#### ACKNOWLEDGEMENTS

I would like to thank Dr. Neilson for his guidance and enthusiasm. I am grateful to the members of my supervisory committee, Dr. R. Eband and Dr. R. Bell for their advice.

I particularly wish to thank Mr. R. Gregoire for preparing some of the oligoribonucleotides used in these studies. Special thanks also go to Drs. D. Hughes and J. Everett for their collaboration on the pmr work.

Many thanks to Dr. M. Henrichs of Eastman Kodak, Rochester, NY for providing some 270 MHz pmr spectra of the CAAUG:CAUUG duplex.

I thank the Ministry of Colleges and Universities for the financial support of an Ontario Graduate Scholarship.

## TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGEMENTS	vi
ABBREVIATIONS	x
LIST OF FIGURES	xii
LIST OF TABLES	xv
1. INTRODUCTION	1
1.1. Conformational Properties of Single Stranded Oligo- and Polyribonucleotides	1
1.1.1. Solid State Structure of Oligoribo- nucleotides	1
1.1.2. Torsion Angle Conformations Derived from Minimum Energy Calculations	10
1.1.3. NMR Studies on the Solution Conformations of Oligonucleotides	15
1.2. Model Studies on the Factors Which Affect the Stability of an RNA Double Helix	21
1.2.1. Studies on Short, Perfect RNA Double Helices	21
1.2.1.1. Studies Employing Optical Methods	22
1.2.1.2. NMR Studies	28
1.2.2. Model Studies on Loop Formation	35
1.2.2.1. Bulge Loops	35
1.2.2.2. Internal Loops	38
1.2.2.3. Hairpin Loop Formation	40
1.2.3. The Effect of Dangling Bases on Duplex Stability	41
1.2.4. The Formation of G-U Base Pairs Within the Secondary Structure of RNA	42
2. The Chemical Synthesis of Oligonucleotides	44
2.1. Blocking Groups for Oligonucleotide Synthesis	45
2.1.1. The 2'-Hydroxyl Blocking Group	46
2.1.2. Blocking of the 2',3' Cis Diol System	47



	Page
2.1.3. Blocking of the Primary Amino Groups of the Nucleobases	47
2.1.4. Protection of the Terminal 5'-Hydroxyl	48
2.1.5. The Phosphotriester, Blocking Group	48
2.2. Chemical Synthesis of Oligodeoxynucleotides Corresponding to the Lactose Operator Sequence of <u>E. Coli</u>	51
2.3. The Use of Chemically Synthesized Oligoribonucleotides in Model Studies of Translational Initiation	56
3. EXPERIMENTAL PROCEDURES	62
3.1. Materials and Reagents	62
3.1.1. Protected Nucleosides and Coupling Reagents	63
3.2. Preparation of Oligoribonucleotides	63
3.2.1. N <sup>4</sup> -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-cytidyl (3'-2,2,2-trichloroethyl-5') N <sup>2</sup> -benzoyl-2'-O-tetrahydropyranylguanosine	67
3.2.2. N <sup>4</sup> -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-cytidyl (3'-2,2,2-trichloroethyl-5') N <sup>2</sup> -benzoyl-2'-O-tetrahydropyranylguanyl (3'-2,2,2-trichloroethyl-5')-benzoyl-2'-O-tetrahydropyranylguanosine	68
3.2.3. N <sup>4</sup> -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-cytidyl (3'-2,2,2-trichloroethyl-5') N <sup>2</sup> -benzoyl-2'-O-tetrahydropyranylguanyl (3'-2,2,2-trichloroethyl-5') N <sup>4</sup> -benzoyl-2'-O-tetrahydropyranylcytidine	69
3.2.4. Deprotection of Protected Oligoribonucleotides	70
3.3. Pmr Studies of Duplex Formation	72
4. RESULTS AND DISCUSSION	74
4. Stability Studies of Short, Imperfect RNA Double Helices	74
4.1. The Self-Complementary Tetramer CAUG	76
4.1.1. Results	77
4.1.2. Discussion	82

	Page
4.2. The Effect of a Dangling Base Adjacent to Non-Fraying Terminal Base Pairs	86
4.2.1. Results	87
4.2.2. Discussion	94
4.3. The Effect of a Dangling Base on Fraying Terminal Base Pairs	99
4.3.1. Results	100
4.3.2. Discussion	108
4.4. The AGGA:UCCU Duplex: A Model of the Shine-Dalgarno mRNA-16 S rRNA Interaction	113
4.4.1. Results	114
4.4.2. Discussion	121
4.5. The Effects of Internal Non-Bonded Bases and a G·U Base Pair on the Stability of the CAUG Duplex	125
4.5.1. Results	127
4.5.2. Discussion	142
4.6. A Model Study on the Formation of a "Staggered" Duplex	146
4.6.1. Results	150
4.6.2. Discussion	160
5. CONCLUSION	165
6. REFERENCES	170

# ABBREVIATIONS

A <sup>a</sup>	adenosine
Å	angstrom (10 <sup>-8</sup> cm)
C <sup>a</sup>	cytidine
c	total strand concentration
CD	circular dichroism
Cu/Zn	copper-zinc couple
dA	2'-deoxyadenosine
dC	2'-deoxycytidine
dG	2'-deoxyguanosine
DMF	N,N-dimethylformamide
DNA	deoxyribonucleic acid
f[ <sup>35</sup> S]Met-tRNA	<sup>35</sup> S-formylmethionyl transfer RNA
G <sup>a</sup>	guanosine
g	gauche
ΔH <sub>i</sub>	enthalpy of single base pair formation
Hz	hertz (cycles/second)
J	coupling constant
MST	mesitylenesulfonyl 1,2,4-triazole
mRNA	messenger RNA
pNBST	paranitrobenzylsulfonyl 1,2,4-triazole
NMR	nuclear magnetic resonance
ORD	optical rotatory dispersion

pmr	proton nuclear magnetic resonance
ppm	parts per million
$R_F$	ratio of distance travelled by solute to that of solvent
RNA	ribonucleic acid
rRNA	ribosomal RNA
T	thymidine
t	trans
thp	tetrahydropyranyl
tlc	thin layer chromatography
$T_m$	melting temperature of a duplex at which 50% of the strands exist in the duplex form
trac	trityloxacetyl
tRNA	transfer RNA
U <sup>a</sup>	uridine
UV	ultraviolet

---

<sup>a</sup> mononucleotides and oligoribonucleotides are abbreviated in the standard format (IUPAC-IUB Commission on Biochemical Nomenclature, 1970). CpA represent 5'-adenylic acid, Cp represents 3'-cytidylic acid and CpA represents cytidyl (3'-5') adenosine. Unless otherwise noted, oligoribonucleotides are numbered sequentially from the 5' terminus.

## LIST OF FIGURES

Figure		Page
1	The seven dihedral torsion angles of rotation in a nucleotidyl unit	2
2	The two preferred conformations for the ribose ring	2
3	Schematic representation of the bond-angles	5
4	The preferred conformations about the phosphodiester group	12
5	The observable resonances of the bases and base pairs in RNA	30
6	Nomenclature of RNA loops	36
7	The chemical synthesis of oligodeoxynucleotides corresponding to fragments of the <u>lac</u> operator sequence	52
8	A general scheme for the phosphotriester synthesis of oligodeoxynucleotides corresponding to sequence fragments of <u>lac</u> operator DNA	54
9	The block synthesis scheme used for the preparation of oligoribonucleotides, by the phosphotriester method of Neilson and associates, for biological studies	59
10	The stepwise synthesis of CGGC	64
11	The 70°C spectrum of CAUG	79
12	The variable temperature plots of the aromatic base and anomeric proton resonances of CAU	80
13	The variable temperature plots of the aromatic base and anomeric proton resonances of AUG	81
14	Results of the variable temperature study of CAUG	83
15	Complete spectral analysis of CAUGU at 70°C	88

Figure		Page
16	Complete spectral analysis of CAUGA at 70°C	91
17	Results of the variable temperature study of CAUGU	92
18	Results of the variable temperature study of CAUGA	93
19	Diagrammatic representation of the base stacking interaction between the dangling base and terminal G-C base pair in (a) CAUGA and (b) CAUGU	98
20	The variable temperature plots of the aromatic base and anomeric proton resonances of AGC	102
21	Results of the variable temperature study of AGCU	104
22	Results of the variable temperature study of AGCUA	106
23	The effect of interdplex end-to-end aggregation on the melting transitions of resonances from the terminal and internal base pairs of short duplexes	109
24	The variable temperature plots for the aromatic base and anomeric proton resonances of AGGA	117
25	The variable temperature plots of the aromatic base proton resonances of UCCU	119
26	Results of the variable temperature study of the duplex formed by the complementary strands AGGA and UCCU	120
27	The 90 MHz pmr spectra of (a) CAAUG and (b) CAUUG in D <sub>2</sub> O at 70°C	128
28	The variable temperature plots for the aromatic base proton resonances of (a) CAUUG and (b) CAGUG	132
29	The variable temperature plots for the aromatic base proton resonances of (a) CACUG and (b) CAAUG	133
30	Results of the variable temperature study of the duplex formed by the complementary strands CAAUG and CAUUG	135
31	Assignment of the resonances of the 70°C spectrum of the duplex set CAAUG:CAUUG	136

Figure		Page
32	Results of the variable temperature study of the duplex formed by the self-complementary hexamer CAUAUG	139
33	Results of the variable temperature study of the duplex formed by the complementary strands CACUG and CAGUG	140
34	Results of the variable temperature study of the duplex formed by the complementary strands CAGUG and CAUUG	141
35	Illustration of the double helices formed by (a) CCGG, (b) CGGC and (c) CGCG	147
36	Various base paired structures possible for CCGG	149
37	The variable temperature plots of the aromatic base and anomeric proton resonances of CCGG	154
38	The variable temperature plots of the aromatic base proton resonances of CGGC	156
39	The temperature dependent line broadening of the CCGC spectra	157
40	The variable temperature plots of the aromatic base proton resonances of CGGA	158
41	The variable temperature plots of the aromatic base proton resonances of CGCG	159

# LIST OF TABLES

Table		Page
1	Sequences prepared by phosphotriester synthesis for use in the biological studies of Ganoza <u>et al.</u> (1978)	60
2	Summary of the preparation of protected oligoribonucleotides	65
3	Summary of the preparation of protected oligoribonucleotides	66
4	Experimental data of free oligoribonucleotides	71
5	Chemical shifts of the oligoribonucleotides in D <sub>2</sub> O at 70°C	78
6	Coupling constants for the oligoribonucleotides at 70°C	78
7	Chemical shifts of the CAUGX oligoribonucleotides in D <sub>2</sub> O at 70°C	90
8	Coupling constants of the CAUGX oligoribonucleotides at 70°C	90
9	Melting temperatures and concentrations of the CAUGX base paired duplexes	95
10	Incremental analysis of AGCU and AGCUA at 70°C in D <sub>2</sub> O	101
11	Determination of the melting temperature of AGCU and AGCUA	107
12	Incremental analysis of the AGGA:UCCU duplex	115
13	Comparison of the AGGA:UCCU duplex with other short, perfect double helices	125
14	Chemical shifts of the CAXUG oligoribonucleotides at 70°C	130



## Table

## Page

15	Incremental analysis of CACUG at 70°C	131
16	Melting temperatures and concentrations of the base paired duplexes	138
17	The chemical shift assignments of CCGG and its sequence isomers at 70°C	151

## INTRODUCTION

### 1.1. Conformational Properties of Single Stranded Oligo- and Polyribo-nucleotides

Elucidation of the conformational properties of single stranded polynucleotides has resulted from solid state (X-ray crystallography), semiempirical energy calculations and solution state studies employing UV, CD, fluorescence and NMR techniques. Figure one shows the seven dihedral torsion angles with rotational freedom in a nucleotidyl unit. By comparison, polysaccharides and polypeptides have only two bonds with rotational freedom in their backbone structures.

#### 1.1.1. Solid State Structure of Oligoribonucleotides

The use of X-ray crystallography in the determination of the conformational properties of single and multistranded nucleic acids is divided into two branches on the basis of sample preparation: single crystals or fibres. If the molecules within the single crystal form a three dimensional array, mathematical analysis of the diffraction pattern yields the unique position of each atom in the crystal. The resolution of this technique is dependent upon the size of the molecule: The resolution for a dinucleoside monophosphate may reach 0.01 Å, and resolution decreases as the size of the molecules increases (Kallenbach and Berman, 1977).

Long chain oligo- and polynucleotides do not form single crystals, but rather fibres. The fibres may be highly ordered

FIGURE 1. The seven dihedral torsion angles of rotation in a nucleotidyl unit.

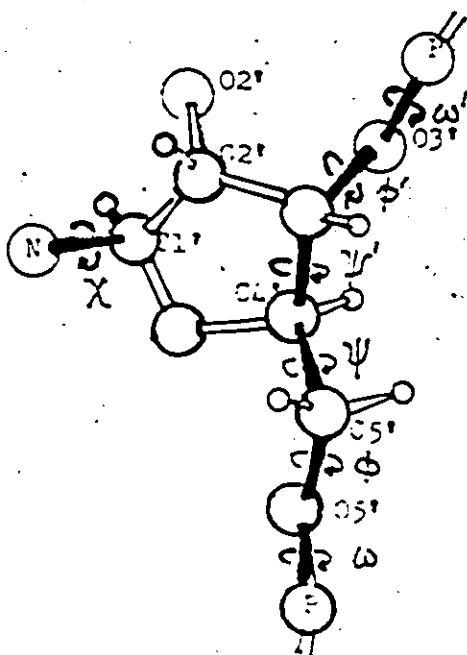
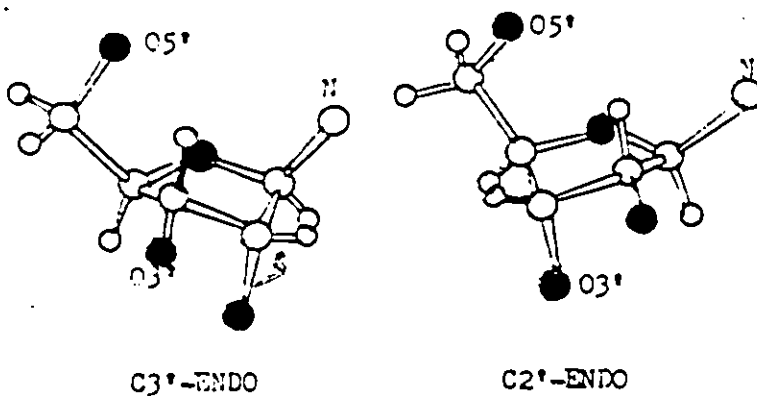


FIGURE 2. The two preferred conformations for the ribose ring.



structures, which will give results similar in quality to single crystals, or the fibres may be less ordered. In most cases, the minimum amount of information realized from fibre crystallography defines helical parameters of pitch, radius, rise per residue, and helical rotation. By combining this data with the more intimate details derived from studies on dinucleotides, it is often possible to generate a more detailed conformation from the polymer, using model building and Fourier transform techniques (Kallenbach and Berman, 1977).

Both single strand and double stranded nucleic acid helices belong to one of two classes: A or B. Although most of the torsion angle differences between the helical classes are small, the two helix types are in fact quite different. The base tilt and base overlaps, important components of the base stacking, are different for A and B. The A helix contains the C3'-endo sugar pucker and the B helix the C2'-endo pucker (Figure 2). The glycosyl torsion angle is also quite different for the two helices.

Studies on double stranded RNAs showed that the ribose ring is in the C3'-endo structure (Arnott et al., 1969) and in an RNA-DNA hybrid, an A type double helix is also formed even though DNA can exist in B type duplexes (Milman et al., 1967). X-ray crystallographic evidence of B type RNA helices has not been found, and it seems likely that the 2'-hydroxyl is partly responsible for the observed C3'-endo preference (Kallenbach and Berman, 1977).

X-ray studies of nucleosides and nucleotides first suggested the idea of a rigid conformation for the nucleoside torsion angles (Sundaralingam, 1969). The sugar pucker can exist only as C3'-endo or

4

C2'-endo, and this requirement necessarily means that there is little flexibility about the C3'-C4' bond ( $\psi'$ ). It appears that the other torsion angles are set and show very little variance. Sundaralingam (1975) has proposed that most of the different conformations for ribodinucleotides result from rotation about the P-O bonds ( $\omega, \omega'$ ) and the C3'-C4' ( $\psi'$ ) bond. Studies on nucleosides and nucleotides also suggested that there were three allowable conformations about the C4'-C5' bond (Figure 3) (Sundaralingam, 1969). Early studies on the crystal structure of UpA suggested that in a dinucleoside monophosphate, the degrees of freedom about several bonds were even more restricted. Although there are three possible conformations for the C4'-C5' bond, the C5'-O5', and the C3'-O3' bonds, these bonds were found to occupy exclusively gauche, trans, trans conformations respectively (Rubin et al., 1972; Sussman et al., 1972). However, the tenability of the rigid nucleoside concept, particularly in solution, remains unresolved at the present time (Kallenbach and Berman, 1977).

Several crystal studies on dinucleoside monophosphates have been reported (for early work see Ts'o, 1974b and references therein; Rosenberg et al., 1976; Seeman et al., 1976; Camerman et al., 1976; Wilson and Al-Mukhtar, 1976). The dinucleoside monophosphate A2'-p-5'U was the first to be studied by X-ray crystallography (Shefter et al., 1964, 1969). The bases were found to be in the anti conformation and the sugar puckers differed between the two nucleoside moieties: the pU moiety had a C3'-endo sugar pucker and the A2'p moiety had the C2'-endo sugar pucker. Presumably the 2'-phosphate substituent introduces steric hindrance when the sugar pucker is C3'-endo resulting in

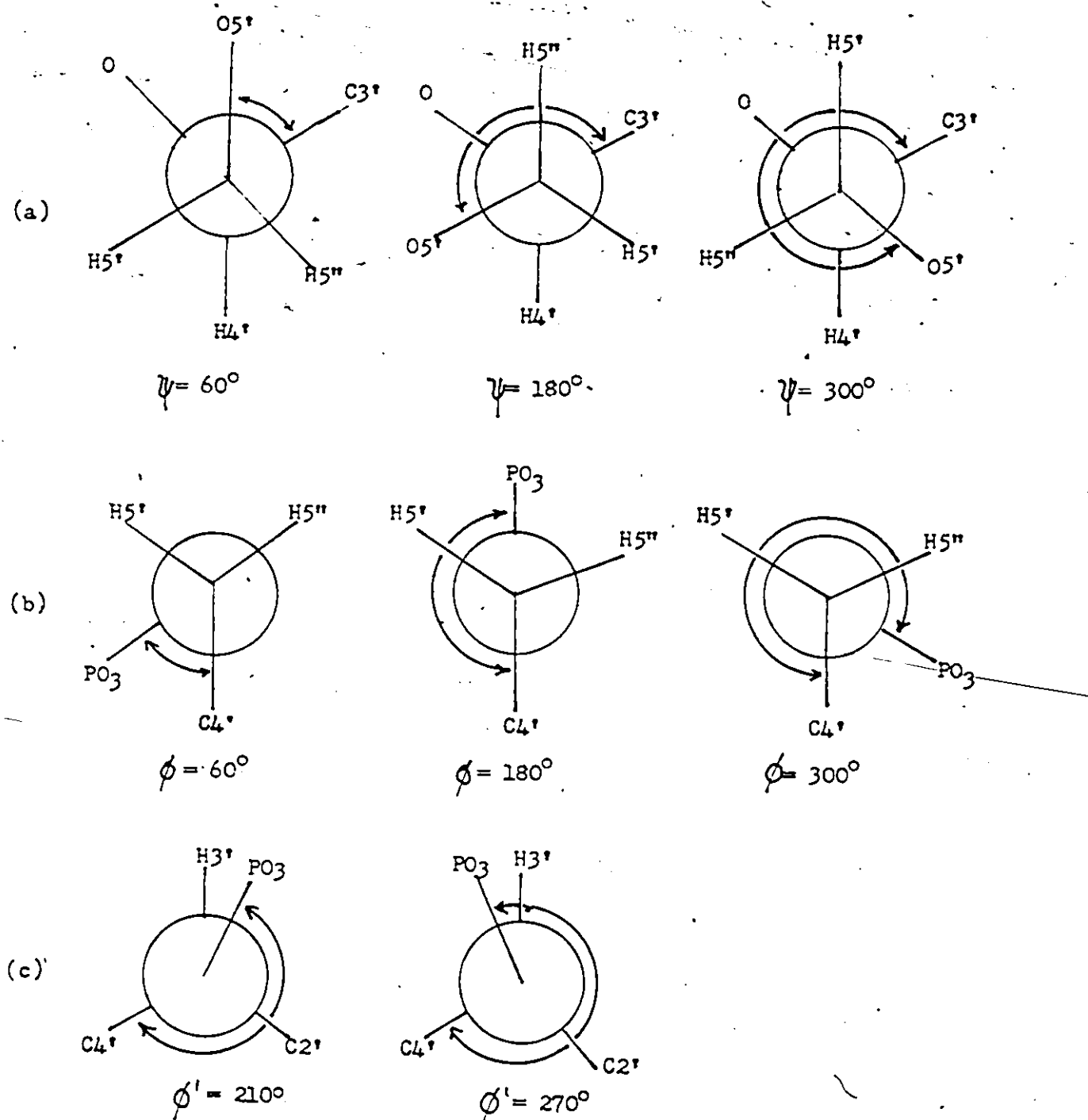


FIGURE 3. Schematic representation of the bond angles around the (a)  $C(4')-C(5')$  bond, along the (b)  $C(5')-O(5')$  bond, and along the (c)  $O(3')-C(3')$  bond. (Altona, 1975)

the shift to C2'-endo.

Two independent studies on UpA crystals showed very interesting results (Sussman et al., 1972; Rubin et al., 1972). The unit cells of these crystals were asymmetric and actually consisted of two different conformations. The UpA(1) conformation (nomenclature of Rubin et al., 1972) represents an open conformation of helical nature with the phosphate inside the helix ( $\omega = -90^\circ$ ). The UpA(2) conformation represents an interesting structure as well: if introduced into a regular right handed helix, it would effectively cause a reverse or turn in the helical direction because of the phosphate conformation ( $\omega = 90^\circ$ ). In the single crystal, the adenine bases are protonated and lie in the same plane, forming hydrogen bonding structures. Similarly, the uracils also lie ~~in~~ the same plane and form hydrogen bonded pairs. It appears that this type of hydrogen bonding in effect leads to these peculiar UpA(1) and UpA(2) structures, since rotations about the P-O bonds will convert each structure into a right-handed helix which could form an antiparallel Watson-Crick base paired duplex (Ts'o, 1974). This observation supported Sundaralingam's (1975) concept of a rigid nucleotide: that the various conformations of dinucleotides, trinucleotides, etc. could be interconverted solely by changes in  $\omega, \omega'$ .

Camerman and coworkers reported recently on the crystal structure of d(pTpT) (Camerman et al., 1976). They found that the bases were in an anti conformation and the sugars were C2'-endo (B-type DNA helix). The phosphodiester conformation was gauche, trans ( $\omega', \omega$ ) which results in an extended conformation. Extrapolation of this geometry to a polymer chain gives a right-handed helix with no

base stacking since the base planes are  $39^\circ$  from parallel and have very little overlap.

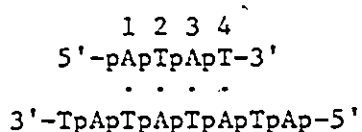
The crystal structure of protonated  $\text{ApA}^+\text{pA}^+$  has also been elucidated (Saenger *et al.*, 1975; Such *et al.*, 1976). The trinucleoside diphosphate is arranged in the crystal in a two molecule complex with an antiparallel orientation. The  $\text{A}^+\text{pA}^+$  portion of the complex is hydrogen bonded to other bases, and is a looped coil conformation similar to  $\text{UpA}(1)$  (Rubin *et al.*, 1972). The  $\text{ApA}^+\text{p}$  is a regular helical type conformation with the unprotonated adenosine not involved in base pairing (Saenger *et al.*, 1975).

The two ribonucleoside monophosphates GpC and ApU when crystallized from neutral solution, formed regular Watson-Crick base paired duplexes (Day *et al.*, 1973; Rosenberg *et al.*, 1973, 1976; Seeman *et al.*, 1976). The sugar puckers in GpC are both C3'-endo (A-RNA type helix) and the G-C base pairs are Watson-Crick with three hydrogen bonds of length 2.86 to 2.95 Å. Although the bases have a slight tilt, they are overlapped in a stacking fashion, and the base planes are separated by about 3.4 Å. The X-ray structure of ApU is similar to GpC, but the excellent resolution does point out some differences (Seeman *et al.*, 1976). The helix is slightly more compact, with the distances between the C-1' atoms slightly shorter in the A-U base pair. The bases are again heavily stacked with about a 60 to 90% overlap. One interesting feature is the difference in sodium ion binding. In GpC, each sodium ion binds to a phosphate and coordinates to other phosphates, bridging the rods of GpC and organizing them into sheets within the crystal. In ApU, one sodium ion binds to the phosphate group but another sodium



ion binds in the minor groove and is coordinated between the two uracil carbonyls protruding into the groove. This binding is specific to ApU and perhaps could signal ApU recognition sites (Seeman et al., 1976).

The results of crystal studies on deoxy(pApTpApT) are of particular interest (Viswamitra et al., 1978). Although the sequence is self complementary and could form a simple four base paired duplex, it forms a splint-type arrangement:



The sugar puckers are C3'-endo for the purine nucleotides and C2'-endo for the pyrimidine nucleotides, and this structure belongs to neither the A or B class of DNA helices. The bases are anti, although each has a different value for  $\chi$ , and the base pairs are Watson-Crick types. The phosphodiester linkage within the d(ApT) halves of the molecule are gauche, gauche giving rise to a regular right-handed helix. However, the T(2)-A(3) internucleotide linkage is trans-gauche, giving rise to an extended conformation, similar to that found in UpA (Sussman et al., 1972; Rubin et al., 1972) and d(pTpT) (Camerman et al., 1976). It was felt that this special conformation could play a role in the binding of lac repressor to poly (dA-T) regions of DNA (Viswamitra et al., 1978).

Complexes between intercalative drugs and complementary dinucleoside monophosphates have also been the subject of several X-ray crystallographic studies (Tsai et al., 1975; 1977; Sakore et al., 1977;

Jain et al., 1977; Neidle et al., 1977; Wang et al., 1978). The results of these studies can be briefly summarized:

1) There are to date two classes of intercalators: (a) proflavine, which intercalates between the base pairs, and forms hydrogen bonds between its amino groups and the phosphate oxygens. (b) compounds which intercalate without forming hydrogen bonds to the phosphate: ethidium bromide, 9-amino-acridine, terpyridine platinum, ellipticine and acridine orange.

2) The base pairs are 6.8 Å apart in order to accomodate intercalation of the drug.

3) In order to effect separation of the base pairs, the glycosidic torsion angles change, with the purine base adopting a high anti conformation at times ( $\chi \sim 100^\circ$ ).

4) The O-P-O torsion angles also undergo changes to allow greater separation of the base pairs.

5) When class b type drugs intercalate, the sugar pucker is C3'-endo-3'p5'-C2'-endo. When proflavine binds, the sugar pucker remains C3'-endo in both ribonucleotidyl moieties.

6) The double helix untwists upon intercalation of class b drugs. The normal twist angle in DNAB is  $36^\circ$  and in RNA-11 it is  $33^\circ$ . The twist angle is reduced to  $8-13^\circ$  upon binding of the drug. The only exception is proflavine, which has no effect on the double helical twist.

The changes in helix structure upon intercalation are thought to be involved in the mechanism of frameshift mutation caused by these drugs (Jain et al., 1977).

### 1.1.2. Torsion Angle Conformations Derived from Minimum Energy

#### Calculations

Calculations of the minimum energy conformations of the sugar-phosphate backbone are semiempirical in nature (Olson, 1978). The total potential energy is usually calculated as the sum of nonbonded, electrostatic and torsional contributions (Thiyagarajan and Ponnuswamy, 1978). While the torsion angle  $\chi, \psi, \psi', \phi, \phi', \omega, \omega'$  are varied, values for the other bond lengths, bond angles and atomic coordinates are derived from X-ray and previous minimum energy calculations. The minimum energy conformations are derived from starting conformations arrived at by using various combinations of the preferred values for the seven dihedral angles, known from previous studies (Thiyagarajan and Ponnuswamy, 1978). The energy calculations proceed and the torsion angles are varied from their starting values until a minimum energy conformation is arrived at.

In base stacking interactions, the bases should be approximately parallel and the strength of the interactions depends upon the interplanar distance between the bases and the degree of overlap of their rings. Calculations have been carried out on the most productive overlaps for base stacking between free purine and pyrimidine bases (Gupta and Sasisekharan, 1978a). It was found that including hydrogen bonding effects into the calculations demonstrated that the most productive overlaps from a base stacking viewpoint would be altered in order to provide the strongest possible hydrogen bonds. Certain base-base stacking interactions (T-T, G-G, C-C for example) are not favoured, but in order to form base pairs, these combinations will stack since

the stacking interaction is stabilized by hydrogen bond formation (Gupta and Sasisekharan, 1978a). The strongest interactions occur for those minimum energy stacking geometries which also allow for the strongest hydrogen bonding in base pair formation. A further study of the effect stacking interactions had on the backbone conformation showed that stacking can occur in either a right or left-handed helix and the sequence of bases in an oligonucleotide could determine which type of a helix is formed (Gupta and Sasisekharan, 1978b).

Yathindra and Sundaralingam (1975) used minimum energy calculations to demonstrate that a correlation exists between the sugar pucker and the internucleotide P-O bond conformations. These calculations neglected energy contributions from bases and thus explored the inherent restrictions in conformation of the sugar-phosphate backbone. Dinucleoside monophosphates favour  $g^-g^-$ ,  $g^+g^+$  conformations about  $\omega, \omega'$  (see Figure 4) particularly for conformations which have the same sugar pucker in both rings (Yathindra and Sundaralingam, 1975). The  $g^-g^-$  conformation is most favourable regardless of sugar pucker, and is the only conformation which will easily allow base stacking. Only the C3'-endo sugar puckers have favourable van der Waals interactions in the  $g^-g^-$  conformation and this becomes the lowest energy conformation. Other favoured conformations include:  $t, g^-$ ;  $t, g^+$ ;  $g^+g^-$  and  $g^-t$  (Figure 4), all being unstacked conformations. The  $t, g^-$  and  $t, g^+$  represent extended conformations, and the  $g^-, t$  conformation could provide bends in a nucleic acid.

Yathindra and Sundaralingam (1975) also found that pNpNp type backbones were restricted to  $g^-g^-$  conformations and  $g^+g^+$  conformations.

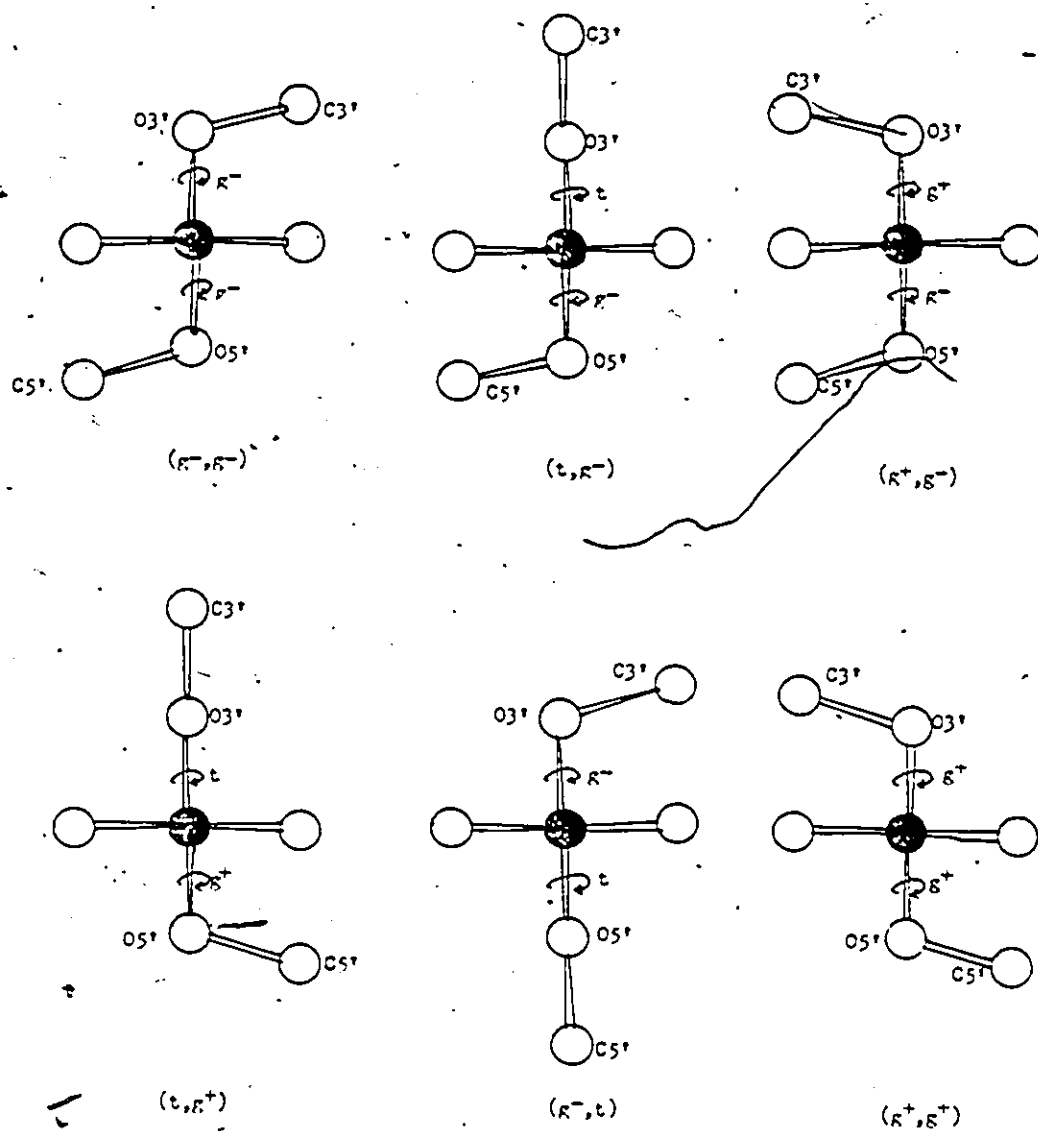


Figure 4. The preferred conformations about the phosphodiester group (Sundaralingam, 1975).

were unlikely because of a steric interaction involving the terminal phosphate. When the base energies are included in the calculations, it is found that the NpN backbone becomes restricted in its ability to adopt  $g^+, g^+$  conformations (Broyde et al., 1978).

Recently a thorough conformational study of ApA was carried out (Thiyagarajan and Ponnuswamy, 1978). When both sugars adopt the C3'-end sugar pucker, minimum energy calculations generate seventeen low energy conformations, differing by 5 kcal/mole in energy between the lowest and highest energy conformers. The lowest energy form was an unstacked extended structure which would be capable of forming loops in polynucleotides. The two conformations which showed stacking interactions were the A-RNA conformer (fourth lowest in energy) and a Watson-Crick type helix conformer (fifth lowest in energy, 0.5 kcal/mole higher than A-RNA). The calculations also verified earlier findings that conformations with C3'-endo sugars have less backbone flexibility than the more extended conformations containing C2'-endo sugars. The authors found that several of the conformations generated by their calculations are components of the tRNA crystal structure (Thiyagarajan and Ponnuswamy, 1978), demonstrating the utility of this theoretical approach.

Minimum energy calculations of the conformations of ApA, CpC, GpG, UpU, GpC and UpA show that helical geometric parameters are base sequence dependent (Broyde and Hingerty, 1978). Generally, the single stranded helices generated from the stacked conformation of each dinucleoside monophosphate are narrower and more tightly wound in comparison to the RNA-11 duplex. The stacking enthalpies in a RNA-11

duplex are about 30% of the stacking enthalpies found for the single strand helices. Therefore it requires energy to generate a double helix from single stranded helices (from 9 to 12 kcal/mole). The helical parameters of the single strands vary from sequence to sequence: significant differences exist in the number of residues per turn, rise per residue and helix radius. Although the torsion angles calculated for the A forms of the single stranded sequences are very similar to those for RNA-11, the small changes in torsion angles leads to overall geometric conformations which differ largely in energy (Broyde and Hingerty, 1978). A similar result was obtained for calculations on deoxydinucleoside monophosphates (Hingerty and Broyde, 1978).

Another minimum energy study also demonstrates that a variety of conformations are possible for both single and double stranded helices (Olson, 1978). In agreement with Broyde and Hingerty (1978), this study shows that single stranded helices have increased base stacking interactions and are more tightly wound in comparison to double stranded structures. There are many possible duplex conformations, including duplexes which have bases in the high anti conformation (anti,  $\chi = 30^\circ$ ; high anti,  $\chi = 120^\circ$ ). Chains with intermediate base anti conformations ( $30^\circ < \chi < 120^\circ$ ) can form duplexes with from ten to twenty-five bases per turn. Since the entire helix structure responds to base dependent changes in  $\chi$ , this suggests that similar structural changes in a double helix could act as a recognition site (Olson, 1978; Viswamitra *et al.*, 1978).

It is apparent then that varying the sugar pucker and all the dihedral angles in minimum energy calculations gives rise to a variety

of possible conformations. The range of angles generated by these calculations illustrates the potential flexibility of the oligonucleotides in solution. Some of these generated conformations can be related back to known crystal structures (Thiyagarajan and Ponnuswamy, 1978).

### 1.1.3. NMR Studies on the Solution Conformations of Oligonucleotides

Although it had been expected that the orientation state about the seven dihedral angles in a nucleotidyl unit would be quite flexible, data from X-ray studies on nucleosides, nucleotides and dinucleotides led to the concept of the "rigid nucleoside" (Sundaralingam, 1975): (1) the sugar pucker must be C2'-endo or C3'-endo, (2) the bases are oriented in an anti fashion, (3) the C3'-O3' and C5'-O5' bonds ( $\phi'$ ,  $\phi$ ) are in a trans conformation, (4) the C4'-C3' bond is in either a gauche (C3'-endo) or trans (C2'-endo) conformation depending upon the sugar pucker, and (5) the highest degrees of flexibility reside in the C4'-C5' bond and the P-O bonds, with rotation about these bonds responsible for the generation of a variety of conformations.

Minimum energy calculations suggest that a variety of stacked and unstacked conformations may exist in solution (Thiyagarajan and Ponnuswamy, 1978) and that helical parameters and stacking will vary according to sequence (Gupta and Sasisekharan, 1978a & b; Broyde and Hingerty, 1978) and whether a single or double stranded helix is formed (Broyde and Hingerty, 1978; Olson, 1978). Of the various techniques available to study the conformation of oligonucleotides, NMR spectroscopy has perhaps been the most useful (Ts'o, 1974). The



magnetic resonance of  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{31}\text{P}$  and associated chemical shift and coupling constant data allow the researcher to monitor changes in the sugar-phosphate backbone and base stacking under various conditions. Since the various solution equilibria are in a rapid state of exchange in comparison to the NMR time scale, the data from an NMR experiment represent an average of the various existing conformations.

Much of the early effort involved the study of the properties, interaction and conformation of bases, nucleosides and nucleotides (for reviews see Ts'o et al., 1969a; Ts'o, 1974a). Ts'o and coworkers demonstrated the magnetic anisotropic shielding effect exerted by neighbouring bases on the various base protons within the vertical stack formed by purine bases (Ts'o, 1974a). The main driving force for stacking appears to be a preference of the aromatic bases to form a relatively nonaqueous environment via a vertical stack (Kallenbach and Berman, 1977 and references therein). The work by Schweizer and coworkers (1968) resulted in the elucidation of the conformation of the base in nucleoside monophosphates. They showed that phosphorylation at the 3' or 5' position tends to decrease the amount of stacking to about 30 to 40% of that for the corresponding nucleoside. Studying the effect of pH upon the NMR spectra of nucleotides showed that a 5'-phosphoryl group had a specific deshielding effect only on the H-8 proton of purine bases and the H-6 proton of pyrimidine bases (Schweizer et al., 1968; Danyluk and Hruska, 1968). This deshielding effect was greatest when phosphate was dianionic, less when phosphate was monoanionic, and least for the monomethyl ester (Schweizer et al., 1968). Such a pH effect was not observed for 3'-nucleotides and

demonstrated that the bases had adopted the anti conformation as described by Donohue and Trueblood (1960).

Dinucleoside mono and diphosphates have been particularly useful compounds for studies on the conformational backbone of single stranded oligonucleotides in solution. The approach used to solve the solution conformational dynamics of the dinucleotide has relied upon careful comparison with the appropriate mononucleotides. This gives some indication of the manner by which a neighbouring unit exerts an influence on its linked partner. A dinucleotide may be considered to be the building block of a polymer in the sense that a trinucleotide consists of two dinucleotides sharing a common neighbouring group, and so on. This approach will only be valid if the properties observed are the result of interactions mainly between neighbouring mononucleotidyl moieties (Ts'o et al., 1969b). Calculations of ring currents for the bases and shift effects as a function of interplanar distance between bases confirm that shielding effects are a nearest neighbour phenomenon (Giessner-Prettre and Pullman, 1976; Giessner-Prettre et al., 1976, 1977). An observation of small end to end shielding effects in trinucleoside diphosphates (Kan et al., 1973) has been explained as the result of a looped out conformation which brings the terminal bases close together and slightly overlapped (Cheng et al., 1978). Only a small percent population of this conformation, within the conformational blend, is required to explain the observed shielding.

Ts'o's group was the first to undertake a definitive NMR study of the solution conformation of dinucleoside mono- and diphosphates (Ts'o et al., 1969b). This large work contains the high resolution

pmr data for fifteen nucleosides and nucleotides, and twenty-five dinucleoside mono- and diphosphates. The spectra of the compounds were studied for temperature and concentration dependence and data were obtained for all the base and anomeric protons. A general conformational model, which included all the dinucleotides studied, was constructed. In this model, the nucleoside units are in the anti conformation and the turn of the 3'-5' screw axis of the stack is right-handed.

Through  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{31}\text{P}$  NMR spectroscopy one can measure various coupling constants, whose magnitude vary with the conformation about a covalent bond. It is possible to use coupling constant values to calculate, in either a direct or indirect fashion, the population of various conformations about all the torsion angles except those about the P-O bonds ( $\omega, \omega'$ ). In the years following Ts'o's initial study of the dinucleotides, there was an intensive effort to understand the solution conformational properties of the dinucleoside monophosphates. Various  $^1\text{H}$  (Altona, 1975; Evans et al., 1975; Wood et al., 1975; Lee et al., 1975, 1976; Ezra et al., 1977),  $^{13}\text{C}$  (Lapper et al., 1973; Govil and Smith, 1973; Schleich et al., 1975, 1976; Alderfer and Ts'o, 1977) and  $^{31}\text{P}$  (Evans et al., 1975; Lee et al., 1975; Gorenstein et al., 1976) NMR approaches were used, and the culmination of this research came when Sarma's group reported the complete pmr spectral details for all the protons of all the dinucleoside monophosphates except GpG.

The results from all this work can be summarized in the following model for a dinucleoside monophosphate in solution:

- 1) The molecule has a flexible backbone in solution.
- 2) There is a preference for the ribose ring to adopt the C3'-endo pucker. A rapid C2'-endo  $\rightleftharpoons$  C3'-endo equilibrium exists and the equilibrium shifts towards C3'-endo with decreasing temperature.
- 3) The C4'-C5' torsion angle ( $\psi$ ) exhibits a preference for the g,g conformation.
- 4) The dominant conformation about the C5'-O5' ( $\phi$ ) is g',g'.
- 5) The conformations of C4'-C5' and C5'-O5' are not coupled to changes in the ribose sugar pucker.
- 6) The dinucleoside monophosphates can form right-handed ( $\omega, \omega' = g^-, g^-$ ) or left-handed ( $\omega, \omega' = g^+, g^+$ ) base stacked helices depending upon their sequence. This has been confirmed by a minimum energy study (Gupta and Sasisekharan, 1978b).
- 7) The stacking, unstacking process occurs primarily as a rotation about the O3'-P-O5' bonds. As stacking decreases and  $\chi$  increases, the ribose pucker equilibrium shifts towards the C2'-endo conformation.

Lee and Tinoco (1977) have studied the solution conformations of modified dinucleoside monophosphates containing either 1,N<sup>6</sup>-etheno-adenosine or 2'-O-methylcytidine. In order to rationalize the selective shielding effects they observed, it was necessary to propose two conformations which differed in  $\omega, \omega'$  values from the seven known possible conformations of ribodinucleoside monophosphates (Kim *et al.*, 1973; Yathindra and Sundaralingam, 1975). Sarma, Pullman and coworkers (Dhingra *et al.*, 1978) used a variety of techniques to analyze these conformations. These techniques included computer simulation of the

three dimensional arrays of a variety of backbone conformations, analyses of unallowed steric interactions, computation of new ring current shielding maps for 1,N<sup>6</sup>-ethenoadenosine and analyses of the various expected shielding effects. They concluded that the two new conformations proposed by Lee and Tinoco (1977) were unlikely to exist in solution because of steric interactions and also that these two conformations were not base stacked, as had been suggested. The predicted shielding trends for these conformations did not match the experimentally observed trends, which were in much better agreement with shielding effects predicted by a mixture of several of the accepted ribodinucleoside monophosphate conformations. In some cases, the dimerization data of Lee and Tinoco (1977) was best represented by a blend of conformations which varied by only  $\pm 5^\circ$  in  $\omega, \omega', \chi_1, \chi_2$ . This result illustrates the fact that only small angle changes are necessary in order to observe a difference in the pmr chemical shift data, and also points out the degree of sophistication available for analyzing the solution conformations of the sugar phosphate backbone in oligonucleotides.

Dhingra and Sarma (1978) have further used this technique in conjunction with experimental data, to propose that nature specifically chose 3'-5' phosphodiester linkages for RNA. Several groups have studied the effect a 2'-5' phosphodiester linkage has on base stacking in dinucleoside monophosphates (Brahms et al., 1967; Kondo et al., 1970; Schleich et al., 1976; Ezra et al., 1976) using a variety of physical techniques. Apparently a X2'p5'Y dinucleotide can only stack when the X base adopts a syn conformation, and will not base stack

16

at all if both bases adopt an anti conformation. A X3'p5'Y dinucleotide however can always form a right-handed, stacked helix for all anti values of  $\chi$  (Dhingra and Sarma, 1978). Since the storage and transfer of genetic information appears to require polynucleotide helical conformers, a 3'-5' phosphodiester linkage is a more versatile choice.

## 1.2. Model Studies on the Factors Which Affect the Stability of an RNA Double Helix

Many RNAs have secondary and tertiary structures (Fiers *et al.*, 1975 & 1976; Gross *et al.*, 1978; Sitz *et al.*, 1978; Rich, 1977; Rordorf *et al.*, 1976; for examples). Secondary structure is usually composed of short duplex regions separated by hairpin and bulge loops. The duplex sections may not be perfectly double stranded, but may also have several small looped out regions. In order to evaluate proposed secondary structures based on a known primary sequence, it will be necessary to understand how the sequence of duplex regions, the placement and size of loops within a double helix, and the presence of neighbouring looped regions contribute to the overall stability of an RNA secondary structure (Tinoco Jr. *et al.*, 1971 & 1973).

### 1.2.1. Studies on Short, Perfect RNA Double Helices

The coil to helix transition in the duplex formation of nucleic acids requires nucleation initially, followed by propagation of base pair formation to complete the double helical structure. A nucleus duplex must necessarily form at a faster rate than it dissociates in order for propagation to become possible. It is estimated that two or three base pairs must form a nucleus between complementary strands

(Craig et al., 1971). Propagation is simply the sequential closing of base pairs immediately adjacent to the duplex region. The stability constants for the formation of base pairs differ for nucleation and propagation. For nucleation the base pairs are formed in a concentration dependent process and are not stabilized by extensive base stacking. Once a stable nucleus is formed, the closing of the helix quickly occurs, since base pair formation is stabilized by the formation of favourable stacking interactions.

#### 1.2.1.1. Studies Employing Optical Methods

The presence of purine and pyrimidine base chromophores account for the UV absorption properties of nucleic acids. Optical studies on the duplexes formed between oligonucleotides of defined sequence have been carried out using UV hypochromicity, CD spectroscopy and temperature jump methods. Derivation of kinetic and thermodynamic parameters is possible using a combination of these techniques.

Initial studies on duplex formation used homopolynucleotides (for review, see Ts'o, 1974b). The field was considerably advanced once oligoribonucleotides synthesized by enzymatic techniques were used to study the helix to coil transition. Although less stable than homopolymer duplexes, the oligomers could be prepared in a greater variety of sequences.

The first study to employ complementary sets of trinucleoside diphosphates was not particularly encouraging (Jaskunas et al., 1968). Of the many complementary sets studied, only the GGC:GCC set showed evidence of association. The ORD spectrum for the mixture was different from a calculated spectrum for a non-interacting 1:1 mixture

of the sequences. The stoichiometry of the complex formed was found to be 2GGC:1GCC. The trinucleoside diphosphates AGC and GGC showed evidence of self aggregation, with as many as forty molecules forming a GGC complex. No simple double helix formation was observed (Jaskunas et al., 1968).

Sequences of the type  $A_mU_n$  were found to form double helices when  $m = n$  (Martin et al., 1971). These sequences are self-complementary, and complications from forming multi-molecular aggregates do not arise. Duplex formation by these sequences (6 to 14 base pairs long) was initially studied by UV hypochromicity (Martin et al., 1971). It was concluded that for  $m = n$ , the oligomers formed double stranded duplexes and not hairpin loops, since their  $T_m$ s were concentration dependent. The heat of single ~~base pair~~ formation,  $\Delta H_1$ , was calculated from the slope at the midpoint of the melting curve, and from the slope of a plot of the reciprocal of  $T_m$  versus log concentration. It was observed that  $\Delta H_1$  increased slightly as chain length increased, and that  $\Delta H_1$  values were lower than those determined from polymer studies. It was proposed that these results could be explained by fraying (a transient opening) of terminal base pairs, since this effect would be proportionately more important in short duplexes.

Studies on the  $A_nU_n$  oligomers were extended using temperature jump techniques (Craig et al., 1971). Experiments of this type allow derivation of thermodynamic and kinetic parameters of the helix-coil transition. Although the results are in agreement with an 'all or none' equilibrium model, the kinetic model must include both nucleation



and propagation. It was determined that nucleation requires three consecutive base pairs be formed, since the rate of association for this intermediate will be faster than the rate of dissociation, thereby allowing propagation. The propagation is so fast that it cannot be observed by temperature jump, implying that an equilibrium exists between a fully base paired double helix and single strands. Some evidence for fraying terminal base pairs also existed in these studies.

The helix to coil transition of  $A_7U_7$  was extensively studied by UV temperature jump and calorimetry (Breslauer *et al.*, 1975). Comparison of the  $T_m$  determined by calorimetry and UV hypochromicity demonstrated for the first time that accurate  $T_m$  measurements can be made by UV. The total enthalpy of the helix to coil transition was determined by calorimetry and compared to van't Hoff  $\Delta H$  values calculated using several different treatments of the UV and temperature jump data. Calculations which depend upon the integrated melting curve of UV hypochromicity had large associated errors because of the uncertainty in determining the fully duplexed state. The  $\Delta H$  values calculated from a log concentration vs. inverse  $T_m$  plot was found to be the most reliable. The van't Hoff  $\Delta H$  values differed by as much as 20% depending upon the analysis used, and all of the calculated  $\Delta H$  values were much lower than the calorimetric value. This result implies that a two state, 'all or none' model is not a valid representation of the helix-coil equilibrium for  $A_7U_7$ , and that fraying ends and intermediate states must also contribute to the equilibrium (Breslauer *et al.*, 1975).

Doty's group also studied the effect one or two G-C base pairs

has on the stability of short oligomers (Uhlenbeck et al., 1971).

This study was carried out using complementary sets  $A_m G U_n : A_n C U_m$  and self-complementary  $A_n G C U_n$  sequences as model compounds and UV hypochromicity to follow the helix-coil transition. They determined that one G·C pair was about equivalent to two A·U pairs in stability contributed to a double helix. In these short oligomers, there was about a  $0.9^\circ\text{C}$  increase in  $T_m$  per % increase in G + C content. This value was greater than that found for RNA ( $0.75^\circ\text{C}/\% \text{ G + C}$ ) and the authors suggested that there were greater stacking interactions for internal base pairs compared to terminal base pairs as a result of fraying. This would mean that the stabilizing effect of G·C pairs in these oligomers is anomalously large (cf. RNA) because they occupy central positions in the duplex.

The effect of sequence on  $T_m$  was studied using a series of AU block ~~oligomers~~ containing G·C pairs in various positions along the duplex (Poerschke et al., 1973). The  $T_m$ s determined by UV hypochromicity melting studies demonstrated that sequence effects can be quite large. It was demonstrated using temperature jump that nucleation occurs preferentially at G·C pairs. Calculations of the activation enthalpy suggested that one or two G·C pairs could nucleate a double helix. Thermodynamic data for sequences of the type  $A_n G C U_n$  could be evaluated by an all or none model, but sequences with long A·U blocks (e.g.  $A_4 G C U_4$  vs.  $A_8 C G : C G U_8$ ) had an appreciable population of intermediate bonded states. The rate of recombination also depended upon sequence and chain length: the rate of recombination decreased with increasing chain length, and sequences with G·C pairs in the middle

recombined at much slower rates.

The  $A_nGCU_n$  ( $n = 1 - 4$ ) were studied further by temperature jump (Ravetch et al., 1974). Observable  $T_m$  values could be determined only for the  $n = 2, 3, 4$  oligomers. The free energy lost upon decreasing the chain length of the duplex was calculated from  $\log C$  vs.  $T_m^{-1}$  plots. The results for  $A_2GCU_2$  were found to be anomalously large. The rate constants for association and dissociation were significantly different for the  $n = 2$  helix in comparison to expected values extrapolated from data for the  $n = 3, 4$  duplexes. These anomalies were explained either as the result of a unique conformation for very short duplexes or as the result of a peculiar conformation for blocks of A·U base pairs (Ravetch et al., 1974). This was an important point, since many of the model sequences that had been studied contained blocks of A·U base pairs.

Analysis of the temperature jump data for the duplex formed by GGGC:GCCC was best fit to the 'all or none' model (Podder, 1971). Nucleation involves the formation of a single G·C base pair, and the rate of helix closure depends only on the rate of forming the second base pair. The GGGC single strand formed multistranded aggregates irreversibly at low temperatures.

A comprehensive study of fifteen  $A_nU_n$  and  $A_nGCU_n$  type duplexes demonstrated the sensitivity of CD to the geometry and sequence of bases in an oligonucleotide (Borer et al., 1973). The sequences contained, between them, all of the ten nearest neighbour interactions of two base pairs. By using statistical methods, the contribution of each nearest neighbour pair to a CD spectrum was calculated. Although

the calculated oligomer spectra agreed well with corresponding experimental spectra, calculated polymer spectra did not reasonably agree with the experimental spectra. This failing was thought to result for at least two reasons. The first drawback was primarily assigned to the presence of a high percentage of A-U base paired blocks in the sequences. Other nearest neighbours were not very highly represented, and the errors associated in calculating their spectral contributions would be high. The other major explanation may rest in the inherent difference in conformation of the polymer in comparison to a short oligomer duplex with fraying ends. The experimental polymer spectra exhibited much less absorption than predicted (Borer et al., 1973).

The sequence dependence of duplex stability was thoroughly investigated by a UV hypochromicity study of nineteen oligomer sequences (Borer et al., 1974). The most important factor for stability of a perfect double helix is the base stacking interaction between nearest neighbours. The thermodynamic parameters for closing a base pair during propagation do reflect the importance of the stacking interaction. Some nearest neighbour pairs occurred less frequently in the sequences studied than others, and the less frequent ones were grouped together in order to provide a more accurate value for each parameter. The total  $\Delta G$  and  $\Delta H$  values were determined for each duplex and the  $\Delta G^\circ$ ,  $\Delta H^\circ$  and  $\Delta S^\circ$  values were determined for each of the six groups of nearest neighbour pairs, in an extension of earlier work (Gralla and Crothers, 1973a & b). An equation was derived to calculate the  $T_m$  of any duplex of known sequence.

The use of the data presented and the  $T_m$  equation is limited in scope (Borer et al., 1974). All ten nearest neighbour pairs were not evaluated individually, and next nearest neighbour effects could not be evaluated. The sequences studied contained blocks of A-U double helix, which could have a particular conformation. Re-evaluation of the data using other, more random sequences would be desirable. The thermodynamic parameters are most accurate for describing  $T_m$ s from 0 to 40°C. The end pairs of short helices may in some cases be 90% frayed, and the effect of this on  $T_m$  could not be taken into account. The results of this study were used in the development of a method to predict the secondary structure of RNA from the primary sequence (Tinoco Jr. et al., 1973).

#### 1.2.1.2. NMR Studies

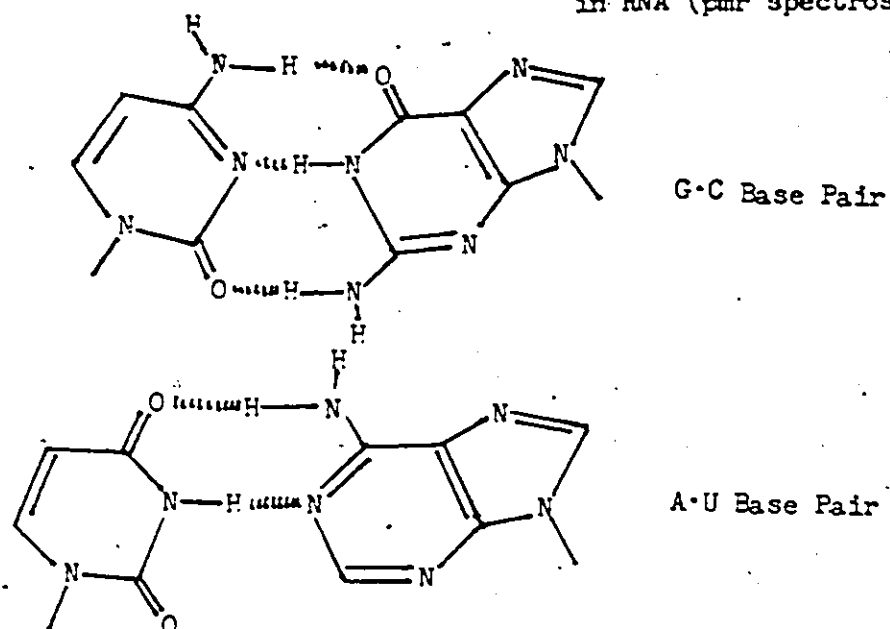
The considerable number of NMR studies of model duplexes have investigated both deoxy (Cross & Crothers, 1971; Crothers et al., 1973; Patel & Hilbers, 1975; Hilbers & Patel, 1975; Patel & Canuel, 1976; Patel, 1976, 1977, 1979; Young & Krugh, 1975; Early et al., 1977, 1978; Kearns, 1977; Kallenbach et al., 1976; Pless & Ts'o, 1977; Selsing et al., 1978) and ribonucleotide sequences (Arter et al., 1974; Ts'o et al., 1975; Borer et al., 1975; Kan et al., 1975; Krugh et al., 1976; England, 1976; Hughes et al., 1978). In sufficiently simple sequences, each nucleotide is identifiable by several resonances, and by the nature of an NMR experiment, the helix to coil transition of each base pair in the duplex can be monitored. This can be accomplished by observing the temperature dependent behaviour of the exchangeable, hydrogen bonded ring N-H resonances, or the

nonexchangeable aromatic ring protons (Figure 5).

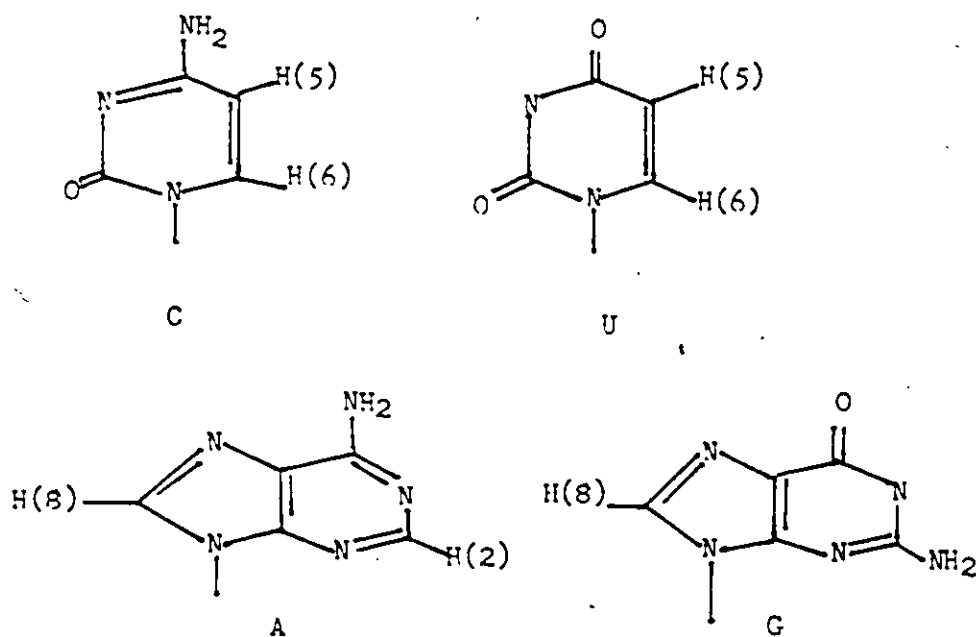
The duplex formed between the complementary pentadeoxynucleotides d(TTGTT):d(AACAA) was the first successfully studied by NMR (Cross & Crothers, 1971). The helix to coil transition was monitored by the change in chemical shift of the thymine methyl group resonance as a function of temperature. The other resonances were too broad to assign accurately. In a follow-up study, the exchangeable NH resonances of the base pairs were observed to broaden with increasing temperature, and disappear at temperatures well below the  $T_m$  of the duplex (Crothers et al., 1973). The dissociation rate constant was calculated from the temperature dependence of the line widths by assuming that the exchange rate of the proton with water was limited by the dissociation of the duplex. At the  $T_m$ , the association-dissociation process was no longer limiting, and the exchangeable ring nitrogen hydrogens were in rapid exchange with the solvent.

The helix to coil transition of the hexaribonucleotide  $A_2GCU_2$  was studied by both the exchangeable and the nonexchangeable resonances (Ts'o et al., 1975; Borer et al., 1975; Kan et al., 1975). Complete assignment of the nonexchangeable base and ribose anomeric proton resonances was realized through the incremental analysis technique (Borer et al., 1975). The plots of chemical shift versus temperature were sigmoidal (similar to UV melting curves) and the midpoints of the curves were used as a measure of  $T_m$ . The average  $T_m$  value calculated from the data agreed well with a value extrapolated from previous optical studies. This result demonstrated that the same helix to coil transition was being monitored by UV and NMR.

Figure 5. The Observable Resonances of the Bases and Base Pairs in RNA (pmr spectroscopy)



(a) The Exchangeable N-H Resonances of the Base Pairs.



(b) The Non-exchangeable Proton Resonances of the Bases.

techniques. However, the NMR melting data was complex:  $T_m$  values determined from different proton resonances of the same base pair sometimes differed drastically, and a pattern could not be discerned that explained this (Borer et al., 1975).

The temperature dependent behaviour of the exchangeable NH resonances showed that the helix to coil transition was not best explained by an all or none theory (Kan et al., 1975). The resonance assigned to the terminal A·U base pairs melted sooner than the rest and the data indicated that there was extensive fraying of the terminal base pairs.

Studies on a similar sequence, d(ATGCAT), also produced evidence of sequential 'end-in' melting and fraying of the terminal A·T base pair (Patel & Hilbers, 1975). Data from the temperature dependence of the NH chemical shifts suggested that the  $T_m$  of the terminal base pair was significantly lower than the  $T_m$  of the other base pairs. The resonance assigned to the G·C base pairs broadened as temperature increased, but showed very little shift change, while the A·T resonances broadened and shifted upfield as temperature increased (Hilbers & Patel, 1975). The line width of the G·C resonance was thought to monitor the lifetime of the double helix and therefore nucleation. The rate of base pair formation was consistent with a bimolecular process, in agreement with earlier results that nucleation will occur at G·C base pairs preferentially (Poerschke et al., 1973; Borer et al., 1974).

Krugh and coworkers have used NMR to monitor complementary association of both deoxyribo and ribodinucleotides (Young and Krugh,



1975; Krugh et al., 1976). The association process was monitored by the large concentration dependent shifts of the base primary amino group proton resonances. The strongest duplexes contained two G-C base pairs and the strength of the interaction was sequence dependent. Several of the dinucleotides that duplex have been used to study drug-nucleic acid interactions in solution (Patel, 1976; Krugh et al., 1977; Chiao & Krugh, 1977; Reinhardt & Krugh, 1978).

The self-complementary tetranucleotide CCGG was also studied by NMR techniques (Arter et al., 1974). A  $T_m$  of  $51 \pm 2^\circ\text{C}$  was determined from the midpoint of the chemical shift vs. temperature transitions for the nonexchangeable base aromatic proton resonances. In contrast to the results for  $A_2GCU_2$  (Borer et al., 1975), there was very close agreement between the  $T_m$ s determined from the plot for each resonance. The  $T_m$  was also close to the value predicted by extrapolation of UV data ( $48 \pm 2^\circ\text{C}$ ), again demonstrating that valid  $T_m$  measurements can be made by NMR studies. The exchangeable NH resonances were assigned both by their temperature sensitive behaviour, and by predictions of chemical shift values using anisotropic shielding contours. There was little evidence for fraying of the terminal base pairs (Arter et al., 1975).

Patel has studied the helix-coil transition of the self-complementary sequence isomers of d(CCGG) (Patel, 1976, 1977, 1979). Full proton NMR studies were carried out on d(GGCC) and d(CCGG). The  $T_m$  values determined from the transition profiles of the nonexchangeable proton resonances were in very good agreement with values extrapolated from UV data (Patel, 1977). The nonexchangeable proton

resonances exhibited an interesting trend. In a pair of resonances (eg. the two GH-8 resonances) there would be two distinctly different chemical shift vs. temperature curves. Although one curve was the typical sigmoidal curve expected, the other curve would have a slightly lower  $T_m$  and would not plateau at low temperatures, but rather would be shielded further. Patel reasoned that these resonances should be assigned to the terminal base and that the further shielding was attributed to end on end interduplex stacking. The same trend was observed for the CH-5 and CH-6 resonances. In a study of d(GCGC), Patel demonstrated that this extra shielding of one resonance was concentration dependent, thereby supporting his interpretation (Patel, 1979). Confirmation of these assignments by incremental analysis was not made because the necessary dinucleotide and trinucleotide sequence building blocks were not studied.

The chemical shift vs. temperature trends for the nonexchangeable resonances of d(CGCG) and d(CGCGCG) were the opposite of those for the other sequences studied (Patel, 1976b). In the tetranucleotide, one CH-6 resonance experienced a large shielding upon lowering temperature, and the other resonance was only slightly shielded. In the hexanucleotide, two CH-6 resonances experienced identical, large shieldings and one resonance was only slightly shielded. This observation allowed the resonance with the small shielding to be assigned to the terminal cytosine. The assignment of the CH-5 resonances was made in the same fashion. The transitions for the terminal resonances had lower  $T_m$ s, which possibly resulted from fraying in this particular sequence. These sequences have been used

in drug intercalation studies (Patel & Canuel, 1976, 1977; Kastrup et al., 1978).

The phosphate backbones of these compounds were investigated by  $^{31}\text{P}$  NMR spectroscopy (Patel, 1976, 1977, 1979). The resonances were well resolved in each case and experienced temperature dependent upfield chemical shift changes upon duplex formation. It was proposed that the shielding results from changes in the torsion angles about the O-P-O bonds.

The duplex formed by the complementary tetranucleotide set GAGC:GCUC has been studied by pmr (England, 1976; Hughes et al., 1978). The temperature dependent behaviour of the proton resonances of each single strand could be studied in the absence of duplexing, since a self-complementary sequence was not used. The resonances experienced linear upfield shielding effects when the temperature was lowered. Upon mixing, the observed temperature dependent chemical shift changes were nonlinear and sigmoidal in nature, indicative of base pairing. The  $T_m$  values determined for all of the resonances were in good agreement, unlike the results for  $\text{A}_2\text{GCU}_2$  (Borer et al., 1975). The average  $T_m$  was in good agreement with a value calculated by extrapolating data from optical studies (England & Neilson, 1977). The oligonucleotides used in this study were chemically synthesized in a stepwise fashion by a phosphotriester method (England, 1976; England & Neilson, 1977). It was possible to accumulate the dinucleotide and trinucleotide building blocks for each tetramer, and their high temperature spectra were used to assign unambiguously the aromatic base and anomeric ribose protons of each tetramer and in the

duplex. Definitive assignments of the resonances greatly increases the usefulness of the NMR data.

#### 1.2.2. Model Studies on Loop Formation

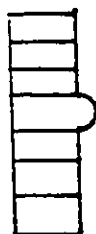
The secondary structure of naturally occurring RNA molecules contains various single stranded loop regions closed by hydrogen bonded, base paired double helices. The duplex regions stabilize the secondary structure (negative contribution to the  $\Delta G$ ) and the loops have a destabilizing influence. There are essentially three different types of loops (Figure 6) and model studies of the formation of each are discussed below.

##### 1.2.2.1. Bulge Loops

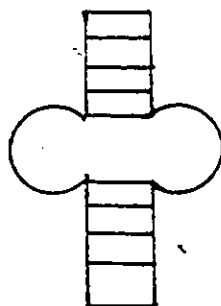
The work on bulge loop formation has been presented in an excellent review (Lomant & Fresco, 1975). Two conformations are possible for nonbonding oppositions: intrahelical or extrahelical (Figure 6). The lowest energy conformation will be determined on the basis of several criteria:

- a) A purine-purine opposition can adopt an intrahelical conformation only by an appreciable distortion of the backbone. An extrahelical conformation would require very little backbone distortion other than a rotation about the P-O bonds, flipping out the nonbonding bases. Conformations of this type are thought to exist for single stranded trinucleoside diphosphates (Cheng & Sarma, 1978).
- b) The entropy of an extrahelical conformation should be greater than the entropy of intrahelical conformations. Extrahelical conformations also have an enthalpy gain from increased hydrogen

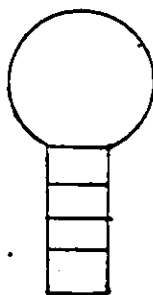
FIGURE 6. Nomenclature of RNA Loops.



(a) BULGE LOOP



(b) INTERNAL LOOP



(c) HAIRPIN LOOP

bonding interactions with the solvent.

- c) The extrahelical conformation has a localized concentration of negative charges in the backbone and must be stabilized by a suitable counter ion concentration.
- d) In a 'looped out' conformation, the adjacent base pairs are separated by the normal plane to plane distance (ca. 3.0 - 3.2 Å for RNA-11), creating a new stacking interaction, something that had been predicted by model building studies (Lomant & Fresco, 1975). Poor stackers, such as uracil, will preferentially loop out in order to create a more favourable stacking interaction between the adjacent base pairs.

Many different non-complementary base oppositions have been tested by homopolymer-copolymer titration mixing experiments (see Lomant & Fresco for a compilation). The bulge loops formed had an extrahelical conformation, regardless of the copolymer sequence or the type of nonbonding opposition (purine-purine, pyrimidine-purine, pyrimidine-pyrimidine). A thermodynamic analysis of the data, based simply on the reduction of  $T_m$  caused by nonbonding oppositions, was carried out (Lomant & Fresco, 1975). Several trends were observed:

- a) The observed destabilization (effect for A > G = C > U) reflected the strength of the stacking interaction lost when the base loops out.
- b) The degree of destabilization is diminished when the new stacking interaction created between the adjacent base pairs is strong:

eg.



In summary, the main source of duplex destabilization by bulge loops arises from the loss of a base stacking interaction when the mismatched base adopts an extrahelical conformation. The destabilization is less if the new stacking interaction formed is strong.

#### 1.2.2.2. Internal Loops

Information on the conformation of internal loops has come from homopolymer-copolymer mixing experiments (for review see Lomant & Fresco, 1975). The summary of the conformational considerations for bulge loops (section 1.2.2.1.) applies equally well to internal loops. However, the homopolymer-copolymer titration mixing experiments were not capable of quantitatively studying the thermodynamics of the formation of internal loops (Lomant & Fresco, 1975). An improved model system for studying the thermodynamics of loop formation utilized enzymatically synthesized oligonucleotides of defined sequence (Uhlenbeck et al., 1971).

Sequences of the type  $A_nCU_n$  were used in this initial study of internal loops. The effect on  $T_m$  of two non-complementary C residues opposite one another was found to be similar to the loss of two A-U base pairs. Comparison of the  $\Delta H$  for the helix-coil transition of  $A_nU_n$  and  $A_nCU_n$  demonstrated that the base pairs in  $A_nCU_n$  melted less cooperatively than those in the uninterrupted duplex. The conformation of the mismatched C residues could not be determined from this study, but an extrahelical conformation was suggested as a possibility (Uhlenbeck et al., 1971).

Sequences of the type  $A_4GC_nU_4$  have also been studied for internal loop formation (Gralla & Crothers, 1973b). Interior loops.

formed when  $n$  was less than five bases long: longer sequences preferentially formed hairpin loops. The instability caused by loop formation increased with the size of the loop. However, three one base loops would be more destabilizing to the duplex than one three base loop, based on the free energy calculations. Comparison of this data with the previous data on internal loops (Uhlenbeck *et al.*, 1971) demonstrated that the free energy required to close a loop by a G-C base pair was considerably less than the energy required when an A-U base pair closed a loop. The loop closure energy calculations provided some evidence for an extrahelical conformation for an internal loop containing only one nonbonding opposition (Gralla & Crothers, 1973b).

Internal loop formation has been recently investigated by melting and single strand nuclease susceptibility studies on homopolymer-cooligomer complexes (Dodgson & Wells, 1977a & b). The oligomers were synthesized enzymatically and contained specific lengths of bonding and nonbonding bases: both  $d(G)_n:C_xA_mC_x$  and  $d(G)_n:C_xG_mC_x$  systems were studied. For  $m = 1$ , a bulge loop was most likely formed, since these duplexes were found to be highly resistant to single strand nuclease. For larger loops, nuclease susceptibility was increased, and the possibility of internal loop formation is increased. It is interesting to note that the  $T_m$ s of the heteroduplex DNAs were very close to  $T_m$  values expected if the mismatched bases were replaced with dA-dT base pairs. This is in direct contrast to the results of a study on  $A_nCU_n$  duplexes (Uhlenbeck *et al.*, 1971), which demonstrated that a nonbonding interaction decreased the  $T_m$  similar to the loss of



two A·U base pairs. This suggests that the destabilizing effect of a loop is perhaps sequence dependent (Gralla & Crothers, 1973b). This concept will have to be tested by further model studies.

#### 1.2.2.3. Hairpin Loop Formation

Hairpin loop formation of  $A_6C_mU_6$  ( $m = 4, 5, 6, 8$ ) oligoribonucleotides was studied by UV and CD (Uhlenbeck et al., 1973). The  $T_m$  values for each of the oligomers were independent of concentration, which indicated that the duplex formed was intramolecular. A plot of  $T_m$  vs. chain length demonstrated that the  $m = 6$  loop was the most stable formed. The smaller loop structures would be strained (there may be some base pair disruption in order to increase the loop size) while for larger loops, there would be a diminished probability of loop closure. In fact, the CD spectra for  $A_6C_5U_6$  were significantly different from those of  $A_6C_6U_6$  and  $A_6C_8U_6$ , implying that the smaller loop must have had a conformation different from the larger loops.

Thermodynamic analysis of the data showed that there was a high positive enthalpy of initiation, probably the result of enthalpy lost upon destacking of the loop bases. This suggested that the stability of hairpin loops may in part depend upon the sequence of bases in the loop (Uhlenbeck et al., 1973).

A similar type of loop system was studied but the loop was closed by G·C base pairs rather than A·U base pairs (Gralla & Crothers, 1973a). Hairpin loop formation of  $A_4GC_nU_4$  ( $n = 5, 6$ ) oligonucleotides was studied using temperature jump techniques. Kinetic analysis showed that nucleation of hairpin loop closure is  $10^3$  fold faster than intermolecular nucleation. The free energy of loop closure is

less when G·C base pairs close the loop rather than A·U base pairs. The increased free energy of stability gained from increasing the size of a small loop to a preferred size, will offset the energy lost when base pairing and stacking are disrupted. The authors concluded from this observation that the size of helical and loop regions in RNA should be flexible and respond to process dependent changes in the tertiary structure (Gralla & Crothers, 1973a).

The sequence dependence of the stability of hairpin loops was studied using oligonucleotides of the type  $A_nUGU_m$  (Wickstrom & Tinoco Jr., 1974). Looped structures of this sequence were more stable than the looped structure of  $A_6C_8U_6$ , when each hairpin had a comparable length of duplex. This was explained on the basis that there is less stacking in AUG than in oligo C. These results are consistent with the idea proposed by Uhlenbeck *et al.* (1973) that the enthalpy of loop closure is derived from the enthalpy lost from disrupting base stacking. In principle, sequences which are poor stackers should have lower enthalpies of closure. Unfortunately the sequences used in this model study also formed other duplexed structures thought to contain G·U base pairs and dangling base residues. However, the study did indicate that the destabilizing effect of a loop is dependent upon the sequence of the loop (Wickstrom & Tinoco Jr., 1974).

### 1.2.3. The Effect of Dangling Bases on Duplex Stability

Dangling bases simply comprise a single stranded region covalently joined to a duplex:

```

XXXX
....
YYYYYYYY...

```

Dangling Base Region

Studies on  $A_mU_n$  oligoribonucleotides demonstrated that for  $m > n$ , the  $T_m$  of the duplex region was increased (Martin et al., 1971). The increase in  $T_m$  was approximately additive for one to three dangling bases.

More recently, two groups have demonstrated, using temperature jump techniques, the stabilizing effect longer dangling base sequences have on small RNA duplexes (Grosjean et al., 1976; Yoon et al., 1976). It has been demonstrated that the duplex formed by yeast tRNA<sup>Phe</sup> and E. Coli tRNA<sup>Glu</sup> at their complementary anticodon regions was stronger, by a factor of  $10^6$ , than expected for a three base pair duplex (Grosjean et al., 1976). It was concluded, from the results of fragmentation studies, that a  $10^3$  fold portion of this increased stability was derived from the additional base stacking in the unpaired regions adjacent to the double helix. Kinetic studies on the binding of UUCA to a dodecanucleotide anticodon loop fragment from yeast tRNA<sup>Phe</sup> showed that the binding was stronger than predicted for the UUCA:UGAA duplex region (Yoon et al., 1976). The increased strength of the binding was attributed to the dangling base regions. It has been proposed that dangling base regions contribute to the stability of the tRNA-mRNA interaction during protein synthesis (Grosjean et al., 1976).

#### 1.2.4. The Formation of G·U Base Pairs Within the Secondary Structure of RNA

Crick first proposed the G·U base pair as part of the wobble hypothesis (Crick, 1966), which allowed for the formation of non Watson-Crick base pairs at the terminus of a codon-anticodon double

helix, thereby explaining the degeneracy of the genetic code. This base pair has since been included within double helical regions of the secondary structure of several tRNAs (Rich, 1977; Rordorf et al., 1976; Robillard et al., 1976), as well as the proposed secondary structures of other native RNAs (Gross et al., 1978; Fiers et al., 1975 & 1976).

Evidence for the formation of G·U base pairs at non wobble positions in RNA double helices has been conflicting to date. Although physical evidence for the formation of G·U wobble base pairs has been implied from certain resonances in the NMR spectra of several tRNAs (Rordorf et al., 1976; Robillard et al., 1976), such assignments have been disputed (Kearns, 1976). Attempts to demonstrate G·U base pair formation using model compounds have been inconclusive for the most part. Studies on poly r(GU) using CD measurements found no evidence of duplex formation (Gray et al., 1972). Krugh has studied the deoxynucleotide d(pGpT) by NMR and found that there was no base pairing under conditions where other self-complementary dinucleotides duplex (Krugh & Young, 1975). Two independent CD studies on poly d(GT) demonstrated that the polymer forms an inter-molecular structure, although there was insufficient proof for a base paired double helix (Lezius & Domin, 1973; Gray & Ratliff, 1977). More recently, poly d(GT) has been the subject of both NMR and ethidium bromide binding studies, which demonstrated that a wobble G·T base paired duplex is formed (Early et al., 1978).

The relative stability of a G·U wobble base pair within a Watson-Crick hydrogen bonded double helix is not completely understood.

It is known that a terminal G·U base pair in a codon-anticodon interaction can be as stable as an A·U base pair in the same position (Uhlenbeck et al., 1970). Uhlenbeck and coworkers also showed, by optical methods, that the addition of an internal G·U base pair does not increase the stability of an  $A_nU_n$  self-complementary duplex (Uhlenbeck et al., 1971). However, there was no evidence from these experiments which would explain the process by which a G·U base pair could form without contributing to the overall duplex stability.

## 2. The Chemical Synthesis of Oligonucleotides

Chemically synthesized oligonucleotides have played a role in a variety of biologically oriented studies. Synthetic triribonucleoside diphosphates were used in experiments that elucidated the triplet nature of the genetic code (Nirenberg & Leder, 1964). Enzymatic joining of chemically synthesized oligodeoxynucleotide blocks resulted in the synthesis of two tRNA genes (Agarwal et al., 1970; Khorana et al., 1976) and the somatostatin gene (Itakura et al., 1977). A similar mixed synthesis method was used to prepare lac operator sequences for insertion into plasmid DNA (Itakura et al., 1975; Bahl et al., 1976; Mariani et al., 1976; Heynecker et al., 1976; Scheller et al., 1977).

Successful chemical synthesis of oligonucleotides remains a challenging problem. Before a specific 3'-5' internucleotide phosphodiester linkage can be formed, other reactive centres on the nucleosides must be chemically masked by blocking groups. After the internucleotide linkages are formed, these blocking groups must be removed without disrupting the nucleosides and the phosphodiester linkages.

The work in this field has been divided between the development of appropriately blocked nucleoside precursors and of suitable phosphorylating methods (for reviews, see Koessel & Seliger, 1975; Reese, 1978).

There are basically two approaches to the chemical synthesis of oligonucleotides: the phosphodiester method, where the phosphate bridges of the blocked oligomer are monoanionic, and the phosphotriester method, where a neutral phosphate ester is prevented by a blocking group from involvement in undesired side reactions. The phosphodiester approach has been used successfully, particularly for the synthesis of oligodeoxynucleotides, but it is thought to have several pitfalls:

- 1) The unprotected phosphate linkage of blocked oligonucleotides can be subject to electrophilic attacks during chain extension, leading to branching or cleavage of the chain.
- 2) The charged phosphodiester intermediates are only water soluble and must be purified by cellulose or sephadex columns, which limit the size of the scale of preparation (cf. organic purification techniques).
- 3) The phosphodiester linkage is sensitive to acid and base and care must be taken when working up reactions.

Protection of the phosphodiester linkage provides a solution to all of these problems.

### 2.1. Blocking Groups for Oligonucleotide Synthesis

The various functional groups that are required to be blocked at various times during synthesis are the primary 5'-hydroxyl, the secondary 3' (and in ribose, 2')-hydroxyl, the primary amino groups of

cytosine, adenine and guanine, and the phosphodiester linkage. A blocking group should be easily introduced onto the nucleoside and removable under conditions which leave other blocking groups intact. The conditions for removal of the blocking groups should not alter the phosphodiester or glycosidic linkages of the product oligonucleotide. Intermediates in the synthesis of blocked nucleosides should be identifiable by NMR analysis as a check for the specificity of blocking group introduction (Reese, 1978).

#### 2.1.1. The 2'-Hydroxyl Blocking Group

The 2'-hydroxyl function of oligoribonucleotides is usually the last to be deprotected, and it is essential that the deblocking conditions do not perturb the phosphodiester linkages. Studies on diribonucleoside monophosphates demonstrated that 0.1 M HCl at 80°C causes up to 50% acid catalyzed migration of the 3' phosphodiester linkage to the 2' hydroxyl as well as some chain cleavage (Brown *et al.*, 1956). This type of phosphodiester isomerization must be avoided in oligonucleotide synthesis, since it is not possible to separate the desired product from those containing some 2'-5' linkages.

The tetrahydropyranyl group has been a popular choice as a 2' blocking group since its introduction over fifteen years ago (Smrt & Sorm, 1962; Smith *et al.*, 1962). This group is easily deprotected in three to four hours upon treatment with 0.01 M HCl at room temperature, conditions which do not lead to phosphodiester isomerization. The point of attachment on the thp ring is asymmetric, giving rise to a diastereomeric mixture of the modified nucleoside. The methoxytetrahydropyranyl group is attached at an achiral centre,

giving only one product, but it is also more sensitive to acidic hydrolysis (Reese et al., 1967) making it tedious to work with. In any case, the thp diastereomeric mixtures are easily resolved by silica gel chromatography and can be recrystallized.

The use of base labile acyl blocking groups for 2'-hydroxyl protection are a poor choice in comparison to thp. Acyl groups will migrate rapidly between the 2' and 3' hydroxyls in a pyridine solution with only traces of water present, as is obtained during work up of reaction mixtures (Reese & Trentham, 1965; Griffin et al., 1966). As well, the equilibrium favours the 3'-isomer over the 2'-isomer, unsuitable for use as a protecting group.

A variety of other groups have been investigated as potential 2'-hydroxyl blocking groups (Koessel & Seliger, 1975). However, the thp groups are the preferred blocking groups. It has been demonstrated that under conditions of mixed anhydride phosphorylation, the thp group is sufficiently bulky to prevent 3'-3' phosphodiester formation without the need for separate 3'-hydroxyl blocking on the incoming nucleoside (Neilson, 1969).

#### 2.1.2. Blocking of the Terminal 2',3' Cis Diol System

In some synthetic schemes, it is necessary to protect the 2',3' cis diol of the incoming nucleoside. Orthoformate esters are the blocking groups of choice (Reese, 1978). Mild acid treatment at room temperature opens the ring to give a mixture of formate esters, which are removed upon very mild base hydrolysis.

#### 2.1.3. Blocking of the Primary Amino Groups of the Nucleobases

The unprotected amino groups can form phosphoramidates via



nucleophilic attack on the phosphorylating agent during oligonucleotide chain extension (Schaller et al., 1963). Acyl groups, whose use were introduced by Khorana (Lohrmann et al., 1966), are widely used to protect the primary amino function. The acyl group is stable to neutral, acidic and mildly basic pH, conditions used to deprotect other functions selectively. The acyl group is removed by ammonolysis, although more extensive treatment is required to deprotect N<sup>2</sup>-benzoyl-guanine (England, 1976).

#### 2.1.4. Protection of the Terminal 5'-Hydroxyl

The blocking group chosen to protect the 5'-hydroxyl plays a pivotal role in the block synthesis of longer oligonucleotides. In a block synthesis, two trimers for example are joined together to yield a hexanucleotide. The incoming block must have a free 5'-hydroxyl and all other reactive groups protected. It is necessary to remove the 5' blocking group selectively, leaving all of the other protecting groups intact. The aryloxyacetyl group is ideal, since it can be selectively introduced to the 5'-hydroxyl of N-benzoyl, 2'-thp nucleoside derivatives, and is removed under mild base conditions which do not deprotect the N-benzoyl groups (Reese & Stewart, 1968). The trityloxyacetyl (trac) group developed by Neilson and coworkers is of special interest (Werstiuk & Neilson, 1973). Trac carrying oligonucleotides are readily identified as yellow spots on tlc plates sprayed with ceric sulfate, which allows one to monitor the phosphorylation and coupling reactions conveniently.

#### 2.1.5. The Phosphotriester Blocking Group

The phosphotriester approach to the chemical synthesis of

oligonucleotides was first introduced by Michelson and Todd (1955). They used benzyl as the phosphotriester blocking group in the synthesis of TpT. The benzyl group is too easily removed under mild conditions and is not suitable for general use in oligonucleotide synthesis (Reese, 1978).

The phosphotriester synthesis was further investigated by Letsinger and coworkers, who introduced the use of 2-cyanoethyl as a blocking group (Letsinger & Mahadevan, 1965 & 1966). This blocking group was used in the phosphotriester synthesis of T(pT)<sub>n</sub> sequences (Letsinger & Ogilvie, 1967 & 1969). The 2-cyanoethyl group is removed by mild ammonia, and unfortunately it may also be removed during work up, by silica gel catalyzed hydrolysis and also under the mild basic conditions used to deprotect the 5'-hydroxyl. Consequently it is not suitable for general use in oligonucleotide synthesis.

The next triester group to be investigated was the 2,2,2-trichloroethyl group (Eckstein & Rizk, 1967a & b). This group is stable to the mild base conditions used to selectively deprotect 5'-aryloxacetyl groups. It is removed by reductive cleavage of the O-alkyl bond by treatment with copper-zinc couple in anhydrous dimethylformamide. Reese has recently questioned the ability of the copper-zinc couple to remove the trichloroethyl group completely (Reese, 1978). However, Neilson and associates have developed a versatile phosphotriester synthesis of oligoribonucleotides which uses 2,2,2-trichloroethyl as the phosphate protecting groups (Neilson, 1969; Werstiuk & Neilson, 1973, 1976; Neilson & Werstiuk, 1974). Carefully prepared copper-zinc couple was found to be capable of

completely deprotecting the trichloroethyl groups from sequences as long as a nonaribonucleotide in reasonable yields (Neilson & Werstiuk, 1974; Werstiuk & Neilson, 1976). It would appear that the trichloroethyl group is suitable for use as phosphotriester protection.

The phenyl group has been extensively investigated as a potential phosphate protecting group. This group is deprotected under basic conditions, by cleavage of the O-phosphoryl bond. Although phenoxide is the preferred leaving group, there is 2 to 3% inter-nucleotide bond cleavage as well at each phosphotriester linkage. This dramatically lowers the deblockings for longer sequences (van Boom et al., 1974). Deblockings with fluoride ion (Ogilvie et al., 1976) or aqueous ammonia (Adanuak et al., 1977) also suffer this side reaction (Reese, 1978). The use of aldoximate ions as deblocking agents may improve the specificity of the deblocking (Reese et al., 1978).

The 5' terminal hydroxyl is usually protected by an aryloxacetyl group which is removed by mild base. A careful strategy must be used when deblocking oligonucleotides which carry phenyl as the phosphotriester protecting group and base labile terminal hydroxyl protecting groups (van Boom et al., 1971). The 5'-terminal blocking group must first be selectively removed by mild base hydrolysis. The terminal hydroxyl is then reprotected with a base stable group such as thp, before deprotection of the phosphate. If this procedure is not followed, simple base hydrolysis will lead to 5'-5' and 3'-3' inter-nucleotide isomerization at the terminal positions (van Boom et al., 1971). Sequences containing isomerized bonds cannot be separated

from the desired product. This deprotection-reprotection step must necessarily decrease the overall yield in the deblocking procedure.

The phosphotriester approach to the chemical synthesis of oligonucleotides perhaps is best illustrated by considering the following two examples. In the first example oligodeoxynucleotides corresponding to the lactose operator of E. coli were synthesized both by the phosphodiester and the phosphotriester approach (Itakura et al., 1975). In the second example, oligoribonucleotide synthesis, more difficult because of the presence of the 2'-hydroxyl group, is illustrated by the phosphotriester synthesis of a oligoribonucleotides for use in model studies of translational initiation in E. coli (Ganoza et al., 1978).

## 2.2. Chemical Synthesis of Oligodeoxynucleotides Corresponding to the Lactose Operator Sequence of E. coli

The sequence of the repressor binding site of the lac operator of E. coli had been elucidated by Gilbert and Maxam (1973). The goal of the chemical synthesis of the lac operator was the preparation of the twenty-one base pair duplex in amounts sufficient for in vitro studies of the operator-repressor binding (Itakura et al., 1975). Both a phosphodiester and a phosphotriester approach were investigated (Figure 7).

The following blocking groups were used in the phosphotriester synthesis:

- (i) The acid labile dimethoxytrityl group was used to block the 5'-hydroxyl.
- (ii) The phosphate groups were blocked by the base labile

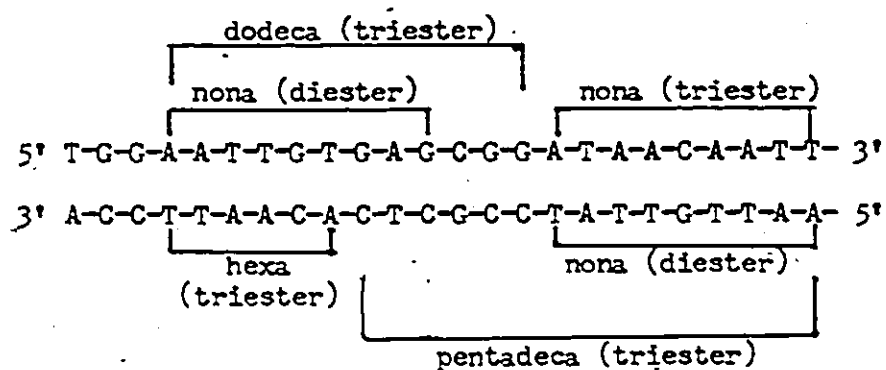


FIGURE 7. The chemical synthesis of oligodeoxynucleotides corresponding to fragments of the lac operator sequence (Itakura et al., 1975)

4-chlorophenyl group.

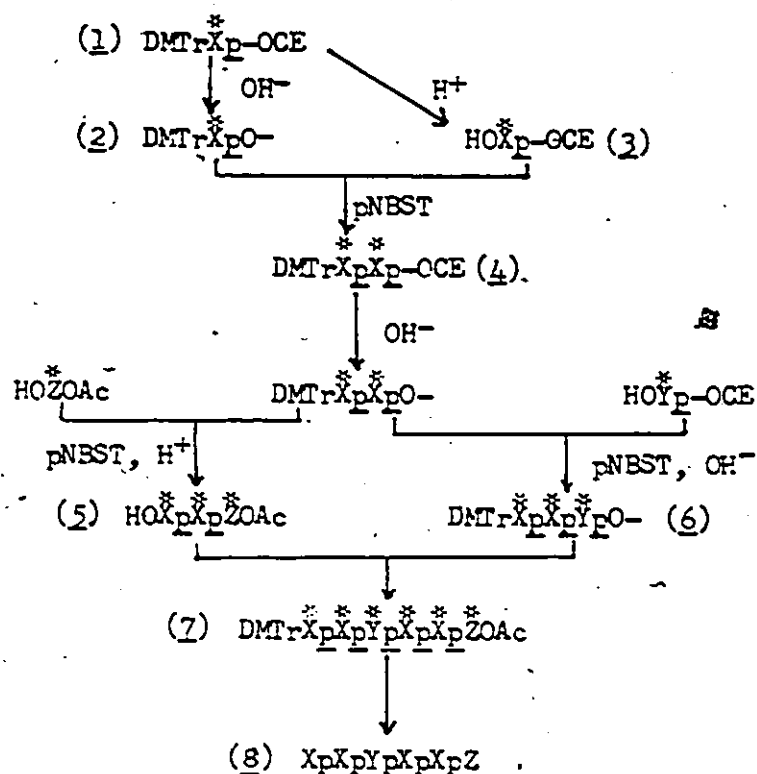
(iii) The primary amino groups of adenine and cytidine were protected by benzoyl; the acetyl group was used to mask the same function in guanine (less stable to ammonolysis than N<sup>2</sup>-benzoyl-guanosine).

(iv) The new coupling reagents MST and pNBST were introduced for use in internucleotide bond formation.

Compound (1) (Figure 8) can be converted by treatment with 0.1 N NaOH in dioxane (30 seconds) to (2) by selective removal of the 2-cyanoethyl group. Treatment of (1) with 80% acetic acid for twenty minutes at room temperature yields (3) by removal of the dimethoxytrityl group. Care must be taken to avoid depurination of the N-acyl purine during the acid treatment.

The coupling between (2) and (3) is effected by activation of the phosphate of (2) in anhydrous pyridine with an excess of par-nitrobenzenesulfonyl triazolidine. After two days, the reaction is complete, and after work up, the uncharged, fully protected dinucleotide can be purified by silica gel chromatography. Treatment of the purified product with base removes the terminal 2-cyanoethyl group, and the chain can then be extended (Figure 8). Synthesis proceeds in a 5' to 3' fashion.

Longer sequences such as (7) are formed by joining two blocks, (5) and (6). The terminal 3'-hydroxyl is protected by an acetyl group, and the specific internucleotide phosphotriester linkage is formed between an equimolar mixture of the two blocks in the same fashion as outlined above for the stepwise synthesis. These longer



Abbreviations: DMTr = dimethoxytrityl

$\text{X}^*, \text{Y}^*, \text{Z}^* = \text{G}^{\text{ac}}, \text{A}^{\text{bz}}, \text{C}^{\text{bz}}, \text{T}$

CE = 2-cyanoethyl

Ac = acetate

$\underline{\text{p}}$  = phosphate protected by  $\underline{\text{p}}$ -chlorophenyl

pNBST = para-nitrobenzenesulfonyl triazolidine

FIGURE 8. A general scheme for the phosphotriester synthesis of oligodeoxynucleotides corresponding to sequence fragments of lac/operator DNA (Iwakura et al., 1975).

sequences are again purified by silica gel chromatography. A suitable block synthesis was carried out for the four sequences indicated in Figure 7. Yields of the block couplings ranged from 35%, for forming the pentadecanucleotide from a hexanucleotide block and a nonanucleotide block, to 62% for the block synthesis of the protected hexanucleotide ACAATT (Figure 7).

Complete deprotection of the sequence was effected by the following scheme:

- (i) Treatment with 0.5 N NaOH in dioxane overnight to remove the phosphotriester groups.
- (ii) 3 hour ammonolysis at 50°C to remove the amino protecting groups.
- (iii) Followed by treatment for twenty minutes at room temperature with 80% acetic acid to remove the dimethoxytrityl group.
- (iv) Purification of the free oligonucleotide by Sephadex G-25 gel filtration and on Avicel-cellulose tlc plates in two solvent systems. Isolated yields ranged from 45 to 71% (Itakura et al., 1975).

The two nonanucleotide sequences synthesized by a phosphodiester approach (Figure 7) were synthesized by joining the appropriate trinucleotide blocks. Yields for the block couplings ranged from 10 to 25%, much lower than yields for the phosphotriester block couplings. This is the result of the participation of the unprotected phosphates in side reactions during the coupling.

The phosphotriester approach was found to be superior for the following reasons:

- (i) Much better yields for block couplings.



- (ii) Larger scale syntheses are possible.
- (iii) Block couplings can be made with an equimolar mixture of the two blocks: in the diester synthesis, large excesses of one block are required.

In a subsequent communication, the completed chemical synthesis of the two complementary, 21 nucleotide long strands was reported (Bahl et al., 1976). The duplex formed by the unprotected oligomer strands bound the lac repressor protein in a specific fashion. The chemically synthesized DNA was cloned in E. coli, and could function biologically to 'titrate out' the naturally occurring repressor protein, making the bacteria constitutive for  $\beta$ -galactosidase (Marians et al., 1976; Heyneker et al., 1976). Chemically synthesized nucleic acid sequences can be used for a variety of biophysical and biological studies, as illustrated by the next example.

### 2.3. The Use of Chemically Synthesized Oligoribonucleotides in Model Studies of Translational Initiation

The chemical synthesis of oligoribonucleotides has not developed as rapidly as the chemical synthesis of oligodeoxynucleotides. The presence of the 2'-hydroxyl on ribose complicates the preparation of specifically blocked nucleosides for incorporation into a synthetic scheme. Several groups have been working to develop phosphotriester syntheses of oligoribonucleotides (for review see Reese, 1978). One successful approach has been that of Neilson and associates.

Neilson was the first to report a phosphotriester synthesis of oligoribonucleotides, with the preparation of UpU and UpUpU (Neilson, 1969). Suitably protected derivatives of adenosine,

guanosine and cytidine have been developed since (Neilson & Werstiuk, 1971a; Neilson et al., 1973; Werstiuk & Neilson, 1973; Neilson et al., 1975; Gregoire & Neilson, 1978) along with their incorporation into both stepwise and block synthesis (Neilson & Werstiuk, 1971b; Werstiuk & Neilson, 1972; Neilson & Werstiuk, 1974). The elegance of this synthetic approach was demonstrated by the synthesis of a nonaribonucleotide corresponding to the anticodon loop sequence of E. coli tRNA<sup>Met</sup><sub>F</sub> (Neilson & Werstiuk, 1974; Werstiuk & Neilson, 1976). This sequence included the cytidine 2'-methyl modified nucleoside.

The basic features of the synthesis are:

- (i) Terminal 5'-hydroxyl is blocked by the triphenylmethoxyacetyl group, specifically removed by mild base for block synthesis purpose.
- (ii) 2,2,2-trichloroethyl phosphate protecting group, removed by carefully prepared copper-zinc couple via O-alkyl bond cleavage with no chain cleavage.
- (iii) The tetrahydropyranyl group is used to block the 2'-hydroxyl, and can be removed by treatment at pH 2.0 at room temperature with no appreciable phosphate migration. This group is sufficiently bulky to prevent 3'-3' internucleotide bonds without 3'-hydroxyl protection (Neilson, 1969).
- (iv) The primary amino groups are protected by benzoyl, removed by ammonolysis.
- (v) Phosphorylations and couplings are effected using the pyridinium salt of 2,2,2-trichloroethyl phosphate and mesitylene-1,2,4-triazolide as an activating agent.

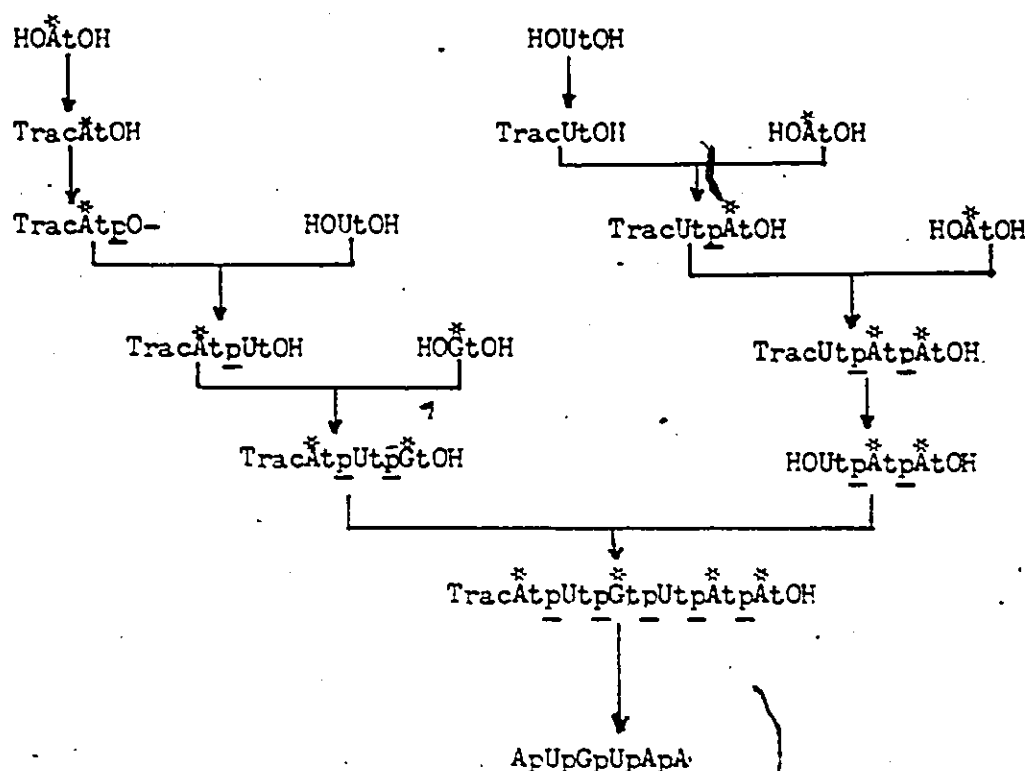
The versatility of this synthesis perhaps is best illustrated by the use of chemically prepared oligoribonucleotides in an investigation of the sequence dependence of translational initiation (Ganoza et al., 1978). The oligoribonucleotides used in this study were prepared using a block synthesis scheme, similar to that shown for AUGUAA (Figure 9). Yields for the block couplings ranged from 29 to 51%. This synthetic method is the only one that is capable of providing satisfactory coupling to guanosine residues in sequences longer than a trimer for oligoribonucleotides (Ganoza et al., 1978).

A three step deblocking procedure was used to obtain the free oligonucleotides (England & Neilson, 1976):

- (i) Treatment for one hour at 60°C in DMF with Cu/Zn couple to remove the trichloroethyl groups.
- (ii) Treatment for two days with 50% methanolic ammonia to remove the trac and benzoyl groups, followed by purification by Whatman 1 paper chromatography.
- (iii) Treatment for two days at pH 2.0 and room temperature to remove the thp groups.

The free oligoribonucleotides were purified twice by descending paper chromatography on Whatman 40. The sequences used in the study that were prepared chemically are shown in Table 1.

The ability of these sequences to form initiation complexes with 70S ribosomes and f[<sup>35</sup>S]Met-tRNA was investigated. Sequences 3' to a 5' terminal AUG were found to have no effect on the ability to form an initiation complex. The XAUG tetramers were quite different: PyAUG sequences were significantly more effective



Abbreviations: \* = N-benzoyl  
 t = tetrahydropyranyl  
 Trac = triphenylmethoxyacetyl  
 p = trichloroethyl phosphotriester

FIGURE 9. The block synthesis scheme used for the preparation of oligoribonucleotides, by the phosphotriester method of Neilson and associates, for biological studies (Canoza et al., 1978).

TABLE 1. Sequences prepared by phosphotriester synthesis for use in the biological studies of Ganoza et al. (1978).

<u>SEQUENCE</u>	<u>yield(%) for complete deblocking</u>
AUG	40
AAAAUG	23
UAAAUG	25
UUA AUG	27
AUGUUU	33
AUGUUA	32
AUGUAA	29
CAUG	29
AAUG	37
UAUG	35
GAUG	31
UUAUG	30
AUAUG	31
CAUAUG	21

initiator sequences than PuAUG sequences. The UAAAUG sequence was a very poor initiator and is perhaps a forbidden configuration. The results of the study, while pointing out some surprising nuances, are more consistent with the existence of a highly flexible initiating code. The versatility of the study arises mainly from the use of the chemical synthetic technique (Ganoza et al., 1978).

The phosphotriester synthesis developed by Neilson and associates is equally suited for the preparation of a variety of sequences for biophysical model studies of duplex formation (England, 1976; England & Neilson, 1977; Hughes et al., 1978), and it was used to prepare all oligoribonucleotides used in the studies reported in this thesis.

## EXPERIMENTAL

### 3. Experimental Procedures

#### 3.1. Materials and Reagents

All reagents and solvents were commercial reagent grade and were used without further purification except for the following: mesitylenesulfonyl chloride (Aldrich) was recrystallized from hot petroleum ether (30-60°C); pyridine, dimethyl formamide, and p-dioxane (Baker Analyzed reagents) were dried over Fischer 4 Å molecular sieves prior to use:

It is necessary to exclude all moisture from condensation reactions. After evaporation (at 30-40°C) in vacuo of the pyridine solution, normal pressure was restored with dry nitrogen.

Saturated sodium chloride (1-2 mL) was used to break the emulsions which frequently occur when aqueous pyridine solutions are extracted with methylene chloride.

Prescored Silica Gel G plates (Analtech) were used for thin layer chromatography of protected nucleosides and coupling reaction mixtures. The tlc solvent system was 10% methanol in methylene chloride. Detection was accomplished by spraying the plates with 1%  $\text{H}_4\text{Ce}(\text{SO}_4)_4$  in 10% sulfuric acid and heating to 150-200°C. All compounds appeared as brown spots, except those containing trityl groups which appeared as a bright yellow spot.

Silica gel (40-140 mesh, Baker Analyzed reagent) for column

chromatography was dried at 80°C over NaOH for two days prior to use. Columns were prepared in methylene chloride and eluted using step gradients of methanol in methylene chloride.

Whatman #1 and #40 papers were used for the descending paper chromatography of deprotected sequences. The solvent system was ethanol/NH<sub>4</sub>OAc (1 M, pH 7.3), 1:1 (V/V). A short wave UV lamp was used to detect deblocked oligonucleotides.

### 3.1.1. Protected Nucleosides and Coupling Reagents

The nucleoside derivatives were synthesized and characterized by published procedures: HOUtOH (Griffin et al., 1968); HO<sup>bz</sup>CtOH (Neilson & Werstiuk, 1971a); HO<sup>bz</sup>GtOH and HO<sup>bz</sup>AtOH (Gregoire & Neilson, 1978); TracUtOH (Werstiuk & Neilson, 1972); Trac<sup>bz</sup>AtOH and Trac<sup>bz</sup>CtOH (Werstiuk & Neilson, 1973).

The coupling reagents also were synthesized and characterized by published procedures: 2,2,2-trichloroethylphosphate (England & Neilson, 1976); MST (Katagiri et al., 1974).

### 3.2. Preparation of Oligoribonucleotides

Complete details for the synthesis of sequences prepared by the author are summarized in Table 2; synthetic details for sequences prepared by R. Gregoire appear in Table 3. All of the protected oligoribonucleotides were made using published techniques (England & Neilson, 1976; Werstiuk & Neilson, 1976). The stepwise synthesis of fully protected CGGC (Figure 10) is described below to illustrate the general procedure.



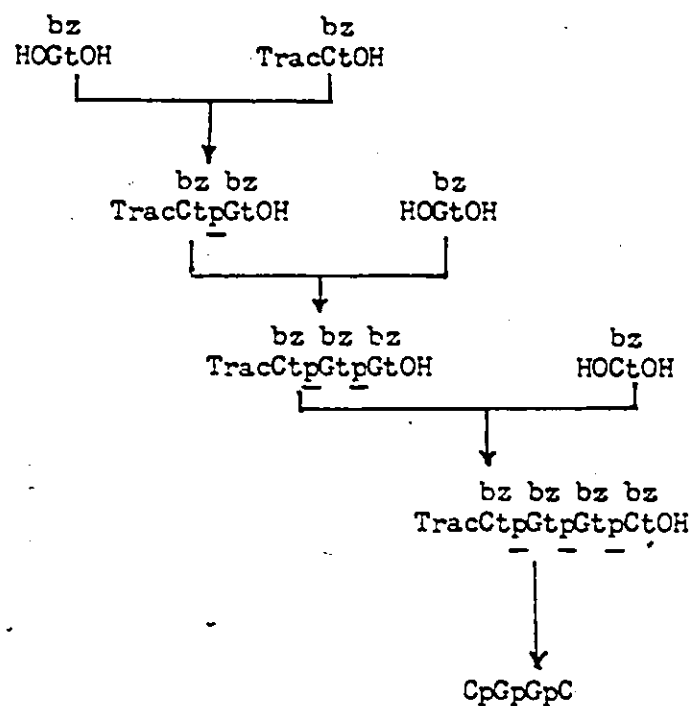


FIGURE 10. The stepwise synthesis of CGGC.

TABLE 2. Summary of the Preparation of Protected Oligoribonucleotides.<sup>a</sup>

REACTANTS						PRODUCTS		
COMPD	QUANTITY		COMPD	QUANTITY		COMPD	YIELD	
	mg	mmol		mg	mmol		mg	%
C	960	1.30	A	895	1.95	CA	1185	66
CA	985	0.71	U	350	1.07	CAU	610	46
CAU	410	0.22	G	160	0.33	CAUG	298	53
C	850	1.20	A	830	1.80	CA	765	49
CA	665	0.48	U	240	0.72	CAU	604	67
CAU	500	0.27	G	200	0.41	CAUG	295	43
CAUG	295	0.12	A	55	0.12	CAUGA	143	39
C	800	1.09	A	735	1.60	CA	462	30
CA	460	0.33	U	160	0.50	CAU	475	76
CAU	475	0.25	G	122	0.25	CAUG	380	60
CAUG	300	0.12	U	60	0.18	CAUGU	174	47
C	1200	1.80	G	820	1.70	CG	1330	59
CG	430	0.31	G	160	0.33	CGG	385	61
CGG	300	0.15	C	70	0.16	CGGC	165	43
C	1000	1.35	C	640	1.50	CC	700	39
CC	650	0.48	G	260	0.53	CCG	430	49
CCG	430	0.22	G	110	0.23	CCGG	325	58
C	1280	1.75	G	930	1.90	CG	1120	50
CG	500	0.36	C	175	0.40	CGC	390	54
CGC	350	0.17	G	95	0.19	CGCG	270	59
CG	500	0.36	G	195	0.40	CGG	430	58
CGG	400	0.20	A	95	0.21	CGGA	190	36

<sup>a</sup>Column 1 contains the 5'-trityloxacetyl reactants, and C stands for TracBz-CtOH; column 4 contains the incoming nucleosides and A stands for HOBz-AtoH; Column 7 contains the trityloxacetyl product and CA stands for TracBz-Ct-p-Bz-AtoH. Two equivalents of pyridinium mono-2,2,2-trichloroethylphosphate activated by 4 equiv. of MST in anhydrous pyridine is used in each phosphorylation step. The coupling step is driven by 1.2 equiv. MST.

TABLE 3. Summary of the Preparation<sup>a</sup> of Protected Oligoribonucleotides<sup>b</sup>

REACTANTS			PRODUCTS					
COMPD	QUANTITY		COMPD	QUANTITY		COMPD	YIELD	
	mg	mmol		mg	mmol		mg	%
C	1350	1.85	A	842	1.85	CA	1460	78
U	1000	1.58	G	745	1.58	UG	1263	62
CA	377	0.27	A	124	0.27	CAA	328	59
CAA	218	0.11	UG	106	0.11	CAAUG	95	28
CA	377	0.27	U	107	0.33	CAU	304	59
CAU	202	0.11	UG	105	0.11	CAUUG	135	41
CA	730	0.53	C	228	0.53	CAC	910	84
CAC	475	0.24	U	102	0.31	CACU	375	62
CACU	175	0.07	G	32	0.07	CACUG	135	62
CA	730	0.53	G	250	0.53	CAG	820	76
CAG	415	0.20	U	85	0.26	CAGU	300	58
CAGU	150	0.06	G	27	0.06	CAGUG	110	59
A	1230	1.63	G	770	1.63	AG	1597	69
AG	700	0.49	G	232	0.49	AGG	510	50
AGG	260	0.13	A	57	0.13	AGGA	145	42
U	1000	1.59	C	823	1.91	UC	1300	65
UC	586	0.47	C	244	0.57	UCC	532	60
UCC	200	0.11	U	112	0.34	UCCU	169	66
AG	300	0.21	C	90	0.21	AGC	200	47
AGC	200	0.10	U	40	0.12	AGCU	125	48
AGCU	120	0.05	A	21	0.05	AGCUA	63	43

<sup>a</sup>These sequences were synthesized by Mr. Rene Gregoire<sup>b</sup>For synthetic details, see footnote to Table 1.

3.2.1. N<sup>4</sup>-Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-  
cytidyl (3'-2,2,2-trichloroethyl-5') N<sup>2</sup>-benzoyl-2'-O-  
tetrahydropyranylguanosine

A sample of 2,2,2-trichloroethylphosphate (England & Neilson, 1976) (700 mg, 3.2 mmol, 2 equiv.) was evaporated in vacuo from anhydrous pyridine (3 x 30 mL) to convert it to the pyridinium salt. MST (1.6 g, 6.4 mmol, 4 equiv.) was added to the pyridine solution (ca. 25 mL) and the solution was warmed (35°C, 1 h) under dry nitrogen. The solution of activated phosphate was added to a pyridine solution (ca. 10 mL) of TracC<sup>bz</sup>OH (1.2 g, 1.6 mmol, 1 equiv.), the total volume was reduced in vacuo ca. 20 mL, sealed under dry N<sub>2</sub> and let stand at room temperature in the dark. Tlc indicated the phosphorylation was ca. 70% complete (R<sub>F</sub> 0.85 + 0.30) after 1 day. Additional MST (80 mg, 0.32 mmol) was added, and the reaction was complete after 2 days. The reaction was quenched by the addition of ice (ca. 5 g). After 30 minutes, the reaction mixture was poured into ice water (ca. 100 mL) and extracted with methylene chloride (5 x 50 mL). The combined organic extracts were washed with water (1 x 50 mL) and evaporated in vacuo to a light yellow oil.

TracC<sup>bz</sup>tpO<sup>-</sup> (ca. 1.6 mmol) was evaporated from anhydrous pyridine (3 x 30 mL, final volume ca. 25 mL) and MST (440 mg, 1.76 mmol, 1.1 equiv.) was added. The solution was warmed gently (35°C) for 1 h. A pyridine solution (ca. 10 mL) of high R<sub>F</sub> isomer of <sup>bz</sup>HOGtOH (820 mg, 1.76 mmol, 1.1 equiv.) was then added, the final volume reduced in vacuo ca. 20 mL and the reaction mixture sealed under dry nitrogen. After 2 days of reaction at room temperature in

the dark, tlc indicated the reaction was ca. 90% complete ( $R_f$  0.30 + 0.68), and 50 mg of MST were added. Tlc indicated no further reaction had occurred after 2 more days. The reaction was quenched with ice (ca. 5 g), poured into ice water (ca. 100 ml) and extracted with methylene chloride (5 x 50 ml). The combined extracts were washed with water (1 x 50 ml) and evaporated in vacuo to dryness. The last traces of pyridine were removed by co-distillation with toluene (3 x 15 ml) to give a yellow foam. The foam was purified by chromatography on 20 g silica gel. Elution with 2% methanol-methylene chloride yielded TracCtp<sup>bz bz</sup>GtOH (1.33 g, 0.94 mmol, 59%).

3.2.2. N<sup>4</sup>-Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-cytidylyl (3'-2,2,2-trichloroethyl-5') N<sup>2</sup>-benzoyl-2'-O-tetrahydropyranylguanlylyl (3'-2,2,2-trichloroethyl-5') N<sup>2</sup>-benzoyl-2'-O-tetrahydropyranylguanosine

A sample of 2,2,2-trichloroethylphosphate (145 mg, 0.62 mmol, 2 equiv.) was converted to its pyridinium salt by repeated evaporation in vacuo from anhydrous pyridine (3 x 10 mL) and activated by MST (320 mg, 1.24 mmol, 4 equiv.). After 1 h, the activated solution was added to a pyridine solution of TracCtp<sup>bz bz</sup>GtOH (430 mg, 0.31 mmol, 1 equiv.) and the volume was reduced in vacuo ca. 15 mL. The reaction stood at room temperature under dry nitrogen. After 2 days, tlc indicated ca. 70% reaction ( $R_f$  0.68 + 0.30) and MST (30 mg) was added. After an additional day, ice (ca. 2 g) was added, the reaction was poured into ice water (ca. 50 mL) and repeatedly extracted with methylene chloride (4 x 30 mL). The combined extracts were washed with water (1 x 50 mL) and evaporated in vacuo to a yellow oil.

<sup>bz bz</sup> TracCtpGtpO<sup>-</sup> (ca. 0.31 mmol) was evaporated from anhydrous pyridine (3 x 15 mL, final volume ca. 10 mL) and activated with MST (85 mg, 0.33 mmol, 1.1 equiv.). After 1 h, a pyridine solution of the high  $R_F$  isomer of <sup>bz</sup> HOGtOH (160 mg, 0.33 mmol, 1.1 equiv.) was added, the final volume reduced in vacuo ca. 10 mL and the reaction was sealed under dry nitrogen. After 3 days, tlc indicated the reaction was complete ( $R_F$  0.30  $\rightarrow$  0.60); ice (ca 2 g) was added, the reaction was poured into ice water (ca. 50 mL) and extracted with methylene chloride (4 x 30 mL). The combined extracts were washed with water (1 x 50 mL) and evaporated to a yellow foam. Purification of the foam on a 10 g silica gel column yielded <sup>bz bz bz</sup> TracCtpGtpGtOH (385 mg, 0.20 mmol, 61%) upon elution with 3% methanol-methylene chloride.

- 3.2.3. N<sup>4</sup>-Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-cytidylyl (3'-2,2,2-trichloroethyl-5') N<sup>2</sup>-benzoyl-2'-O-tetrahydropyranylguanlylyl (3'-2,2,2-trichloroethyl-5') N<sup>2</sup>-benzoyl-2'-O-tetrahydropyranylguanlylyl (3'-2,2,2-trichloroethyl-5') N<sup>4</sup>-benzoyl-2'-O-tetrahydropyranyl cytidine.

A sample of 2,2,2-trichloroethylphosphate (65 mg, 0.29 mmol, 2 equiv.) was converted to its pyridinium salt by repeated evaporation in vacuo from anhydrous pyridine (3 x 10 mL) and activated by MST (145 mg, 0.58 mmol, 4 equiv.). After 1 h, the activated solution was added to a pyridine solution of <sup>bz bz bz</sup> TracCtpGtpGtOH (300 mg, 0.145 mmol, 1 equiv.) and the volume was reduced in vacuo ca. 10 mL. The reaction stood at room temperature under dry nitrogen. After two days, tlc indicated ca. 80% reaction ( $R_F$  0.60  $\rightarrow$  0.30) and MST (30 mg) was added. The following day, ice (ca. 2 g) was added, the reaction was poured

into ice water (ca. 25 mL) and extracted with methylene chloride (5 x 25 mL). The combined extracts were washed with water (1 x 30 mL) and evaporated in vacuo to a brown oil.

<sup>bz bz bz</sup>TracCtpGtpGtpO<sup>-</sup> (ca. 0.145 mmol) was evaporated from anhydrous pyridine (3 x 15 mL, final volume ca. 10 mL) and activated with MST (40 mg, 0.16 mmol, 1.1 equiv.). After 1 h, a pyridine solution of the high R<sub>F</sub> isomer of HO<sup>bz</sup>CtOH (70 mg, 0.16 mmol, 1.1 equiv.) was added, the final volume reduced in vacuo ca. 10 mL and the reaction was sealed under dry nitrogen. After two days, tlc indicated ca. 20% reaction (R<sub>F</sub> 0.30 → 0.54), and MST (20 mg) was added. After an additional two days, tlc indicated the reaction was complete. Ice (ca. 2 g) was added, the reaction was poured into ice water (ca. 25 mL) and extracted with methylene chloride (5 x 25 mL). The combined extracts were washed with water (1 x 30 mL) and evaporated in vacuo to a light brown foam. Purification on a 10 g silica gel column yielded <sup>bz bz bz bz</sup>TracCtpGtpGtpCtOH (165 mg, 0.06 mmol, 43%) upon elution with 4% methanol-methylene chloride.

#### 3.2.4. Deprotection of Protected Oligoribonucleotides

All of the protected oligoribonucleotides were completely deblocked by the same procedure (England & Neilson, 1976).

Protected oligonucleotide (10 or 20 mg) was dissolved in 0.5 mL DMF and Cu/Zn couple (ca. 10-20 mg) was added. After stirring 1-2 h at 50°C, reaction was judged to be complete by tlc (R<sub>F</sub> → 0.0). Methanolic ammonia [methanol: c. NH<sub>3</sub>, 1:1 (v/v)] was added, the reaction vessel was tightly sealed, and the reaction left to stir at room temperature for 2 days. The Cu/Zn couple was filtered and washed with 1 N ammonia. Chelex-100 (NH<sub>4</sub><sup>+</sup> form) was added to the combined

TABLE 4. Experimental Data<sup>a</sup> of Free Oligoribonucleotides.

COMPOUND	R <sub>F</sub> <sup>b</sup>	% YIELD FOR DEPROTECTION <sup>c</sup>
CA	0.60	48
CAU	0.46	45
CAUG	0.35	29
CAUGA	0.22	33
CAUGU	0.26	26
CG	0.47	67
CGC	0.30	35
CGCG	0.21	19
CGG	0.28	54
CCGG	0.13	27
CGG	0.29	33
CGGC	0.32	27
CCGA	0.22	26
CAA	0.49	63
CAG	0.45	64
AGC	0.41	50
CAC	0.41	34
CACU	0.29	27
CACUG	0.15	22
CAAUG	0.16	55
CAGUG	0.15	26
CAUUG	0.21	54
AG	0.41	28
AGG	0.31	27
AGGA	0.16	29
UCC	0.45	49
UCCU	0.36	41
AGCU	0.31	33
AGCUA	0.12	40

<sup>a</sup>Sequences in the top half of the table were prepared by the author; those in the bottom half by Mr. R. Gregoire.

<sup>b</sup>Whatman #40 in absolute ethanol-1M ammonium acetate (pH 7.3) (1:1, v/v)

<sup>c</sup>Calculated from UV data assuming a 90% hypochromicity factor.



filtrate and washings, and stirred for 1 h. The Chelex was filtered and washed with 1 N ammonia. The combined filtrate and washings were evaporated in vacuo to dryness, and the partially deprotected oligomer was purified by descending paper chromatography on Whatman #1 in ethanol: 1 M  $\text{NH}_4\text{OAc}$  (pH 7.3) (1:1, v/v). The desired band ( $R_F \sim 0.8 - 0.9$ ) was cut out, desalted by soaking in absolute ethanol (1 h) and anhydrous diethyl ether (15 min), and eluted with glass distilled water. The pH of the eluate was adjusted to 2.0 with 2 N HCl, and left to stand at room temperature for 2 days. The solution was then neutralized with concentrated ammonia, evaporated in vacuo to dryness, and purified on Whatman #40 in the solvent system described above. The desired band was cut out, desalted as described and eluted with glass distilled water. The data for the deprotection of the oligoribonucleotides are presented in Table 4. The sequences were characterized by their  $70^\circ\text{C}$  pmr spectra.

### 3.3. Pmr Studies of Duplex Formation

The 90 MHz pmr spectra were obtained in the Fourier transform mode of a Bruker WH-90 spectrometer equipped with quadrature detection. Probe temperatures were maintained to within  $\pm 1^\circ\text{C}$  by a Bruker B-ST 100/700 variable temperature unit and were calibrated by thermocouple measurements. ~~Two~~ samples were lyophilized twice from  $\text{D}_2\text{O}$  and then dissolved in 100%  $\text{D}_2\text{O}$  (Aldrich) which contained 0.01 M sodium phosphate buffer (pD ca. 7.2) and 1.0 M sodium chloride. Sample concentrations were  $10^{-3}$  to  $10^{-2}$  M. tert-Butyl alcohol-OD was used as an internal reference and the chemical shifts are reported in ppm relative to sodium 2,2-dimethyl-2-silapentane-5-sulfonate. The field/frequency

lock was provided by the deuterium signal of  $D_2O$ . Spectra were recorded over a 1200 Hz sweep width in 8 K data points (3.411s acquisition time). The pulse width was 3  $\mu s$  (67.5° pulse angle). High temperature spectra were obtained in 200-300 scans, while those at lower temperatures required 1000 scans.

## RESULTS AND DISCUSSION

### 4. Stability Studies of Short, Imperfect RNA Double Helices

One of the interests arising from the sequencing of natural RNA molecules has been the determination of possible secondary structures for these molecules (Fiers et al., 1975, 1976; Gross et al., 1978; Sitz et al., 1978). The secondary structure of an RNA could play an important role in the initiation of translation (Fiers et al., 1975, 1976; Steitz & Jakes, 1975; Shine & Dalgarno, 1975), the interaction of proteins with RNA (Krol et al., 1978), and RNA-RNA interactions (Sitz et al., 1978). In the case of tRNAs, tertiary structure is also important to functionality.

The ability to predict and verify secondary structures for naturally occurring RNAs other than tRNAs is not well developed at present. Two main problems are the larger size of most RNAs of interest (cf. tRNA) and their diversity of functions. Present methods for determining the secondary structure of RNAs include clever digestion techniques (Fiers et al., 1975, 1976), maximizing base pairing of the primary structure (Gross et al., 1978), and employing rules derived from model studies (Delisi & Crothers, 1971; Tinoco Jr. et al., 1971, 1973; Studnicka et al., 1978). Further information about the factors which affect the stability of RNA secondary structure is required in order to improve these predictive methods (Tinoco Jr. et al., 1973).

The results from previous model studies of factors affecting RNA duplex stability have been summarized in section 1.2. Several interesting phenomena have been observed, [for example: the dangling base effect (Martin et al., 1971; section 1.2.3.) and the formation of internal G-U base pairs (Uhlenbeck et al., 1971; section 1.2.4.)] yet the underlying molecular mechanisms of these phenomena have not been elucidated. These phenomena have been further investigated by a series of pmr studies of short, imperfect RNA double helices.

Earlier studies used optical methods to monitor the helix-coil transition. These optical studies provided valuable thermodynamic and kinetic information about duplex formation, but were not capable of providing detailed information about the base pairs. NMR studies (section 1.2.1.2.) are superior in this respect: for sufficiently short sequences, individual aromatic resonances can be assigned for each base. Consequently pmr spectroscopy was used to investigate the behaviour of individual base pairs during the helix-coil transition of the short RNA double helices in the present studies.

Greater quantities (ca. 1-3 mg) of oligoribonucleotides will be required for the pmr experiments than would be required for optical studies. A convenient method for the large scale preparation of oligoribonucleotides is the phosphotriester synthesis of Neilson and associates. This synthetic system allows the preparation of a wide variety of sequences containing any combination of all four nucleosides (England & Neilson, 1976; Ganoza et al., 1978). Sequences of chain length up to five may be prepared in a stepwise fashion, an important consideration for NMR studies since the high temperature spectrum of

each deblocked precursor sequence can be used to aid in the unambiguous assignment of the resonances of the final sequence (Borer et al., 1975). Incremental analysis constitutes a useful extension of the NMR data.

Another important consideration is the choice of sequences to be studied. In most of the previous studies on RNA duplex formation, self-complementary sequences have been used. The use of self-complementary sequences greatly simplifies the synthetic problem, since to study a duplex of six base pairs, only one hexamer is synthesized rather than two complementary strands. The NMR spectra are also simplified because the duplex formed by a self-complementary strand has a two-fold axis of symmetry. Since the spectrometer available for these studies operated at 90 MHz for proton, it was important to keep the spectra as simple as possible.

#### 4.1. The Self-Complementary Tetramer CAUG (Romaniuk et al., 1978a)

A reference RNA double helix was required to allow measurements of the relative stability effects of various helix imperfections. A stepwise preparation of the oligoribonucleotides was preferred for the purposes of incremental analysis, and is more convenient for short chain lengths ( $\leq 5$ ). The choice of a self-complementary tetramer simplified both the NMR analysis and the synthetic requirements.

The sequence CAUG was chosen as the reference duplex. Duplexes  $n = 4$  studied previously have contained only G-C base pairs, except for the GAGC:GCUC duplex (England & Neilson, 1977; Hughes et al., 1978). Studies on the chain length dependence of thermodynamic parameters for the  $A_nGCU_n$  duplexes had suggested that A-U duplex blocks

may have a peculiar structure (Ravetch et al., 1974). Similar observations have been made as well about the G-C block of the  $d(C_{15}A_{15}):d(T_{15}G_{15})$  duplex (Early et al., 1977). The CAUG sequence is sufficiently random to avoid the problem of a 'special' conformation. As well, the CAUG duplex is the first to be studied consisting of terminal G-C base pairs and internal A-U base pairs.

The CAUG oligomer can be used as a reference only if it forms a stable double helix. A  $T_m$  of  $-3^\circ\text{C}$  was calculated for CAUG at 10 mM strand concentration using the data and equation of Borer et al. (1974). A comparison of the  $T_m$ s for CCGG (Arter et al., 1974) and GAGC:GCUC (Hughes et al., 1978) at similar concentrations indicates a decrease of  $9^\circ\text{C}$  in  $T_m$  when an A-U base pair is substituted for a G-C base pair ( $T_m$ s:  $51^\circ\text{C}$  vs.  $42^\circ\text{C}$  respectively). Although  $T_m$  values are dependent upon both base composition and sequence (Borer et al., 1973, 1974), it seemed reasonable to expect that a  $T_m$  for CAUG could be measured at 10 mM strand concentration.

#### 4.1.1. Results

The low field proton resonances of CAUG were assigned by comparison to the  $70^\circ\text{C}$  spectra of CA, CAU, and AUG. These data are presented in Tables 5 and 6. Figure 11 shows the  $70^\circ\text{C}$  spectrum of CAUG illustrating the complete assignment of the aromatic base and anomeric resonances.

The results of variable temperature studies of CAU and AUG are summarized in Figures 12 and 13. Most of the resonances were shifted upfield, in essentially a linear fashion, as temperature was decreased. At high temperatures, the UH-1' and GH-1' doublets of AUG were

TABLE 5. Chemical shifts<sup>a</sup> of the oligoribonucleotides in D<sub>2</sub>O<sup>b</sup> at 70 °C

Proton	CA	CAU	CAUG	AUG
CH-6	7.660	7.687	7.662	
AH-8	8.377	8.382	8.346	8.276
AH-2	8.260	8.250	8.196	8.198
UH-6		7.744	7.692	7.730
GH-8			7.962	7.963
CH-1'	5.779	5.768	5.765	
AH-1'	6.093	6.088	6.039	6.024
UH-1'		5.856	5.813	5.841
GH-1'			5.845	5.841
CH-5	5.958	5.925	5.912	
UH-5		5.746	5.738	5.772

<sup>a</sup>Chemical Shifts are in ppm relative to DSS using tert-butyl alcohol-OD as an internal reference and are accurate to  $\pm 0.005$  ppm.

<sup>b</sup>pD=7.0; concentrations: CA, 16 mM; CAU, 10 mM; AUG, 13 mM; CAUG, 9.2 mM.

TABLE 6. Coupling constants<sup>a</sup> for the oligoribonucleotides at 70 °C.

Proton	$J_{1',2'}$			
	CA	CAU	CAUG	AUG
CH-1'	4.4	4.0	3.8	
AH-1'	4.6	3.7	4.4	5.0
UH-1'		4.6	4.8	4.6
GH-1'			5.3	4.6

<sup>a</sup>Accurate to  $\pm 0.2$  Hz.

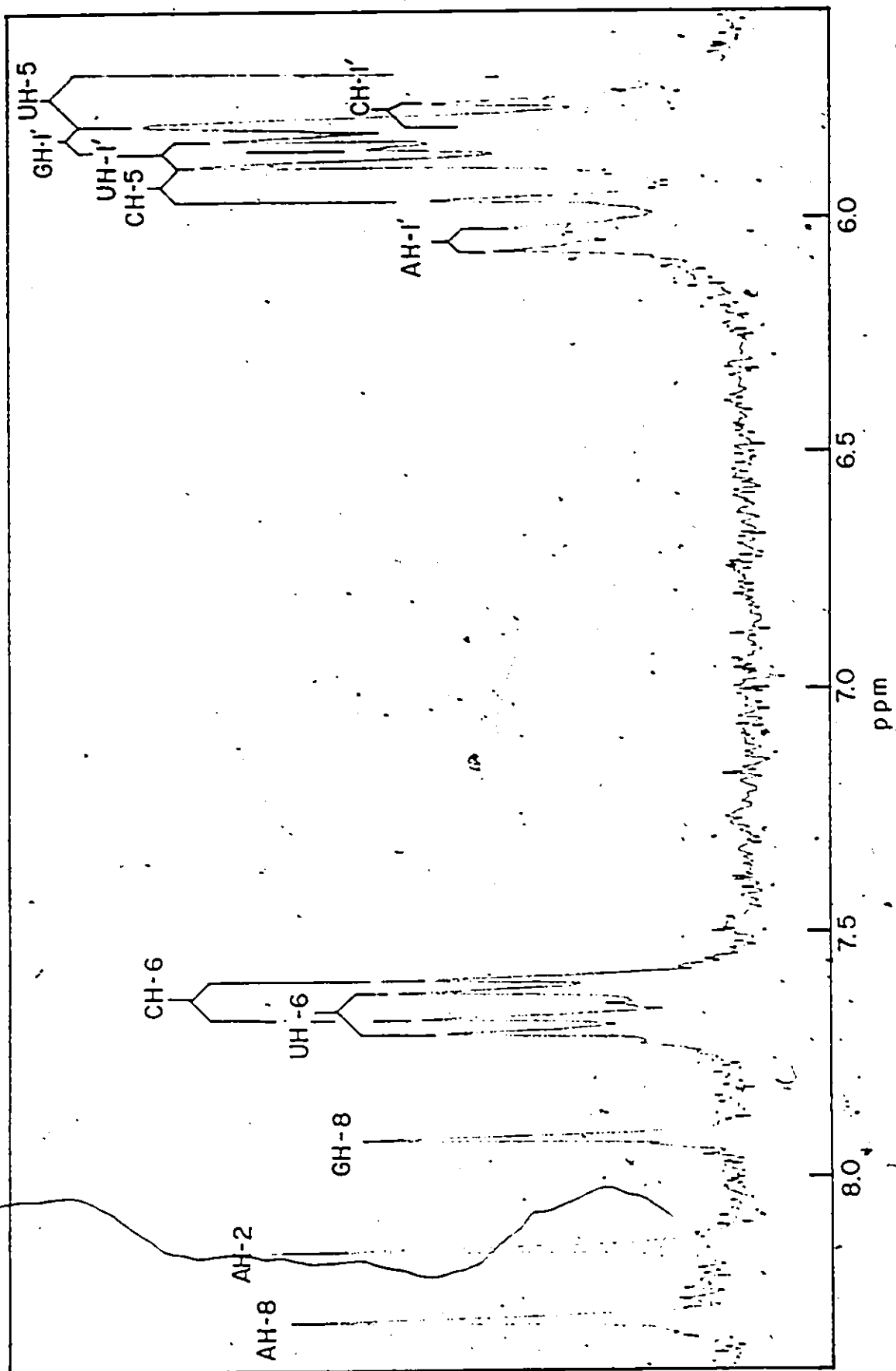


FIGURE 11. The 70 °C spectrum of CAUG.



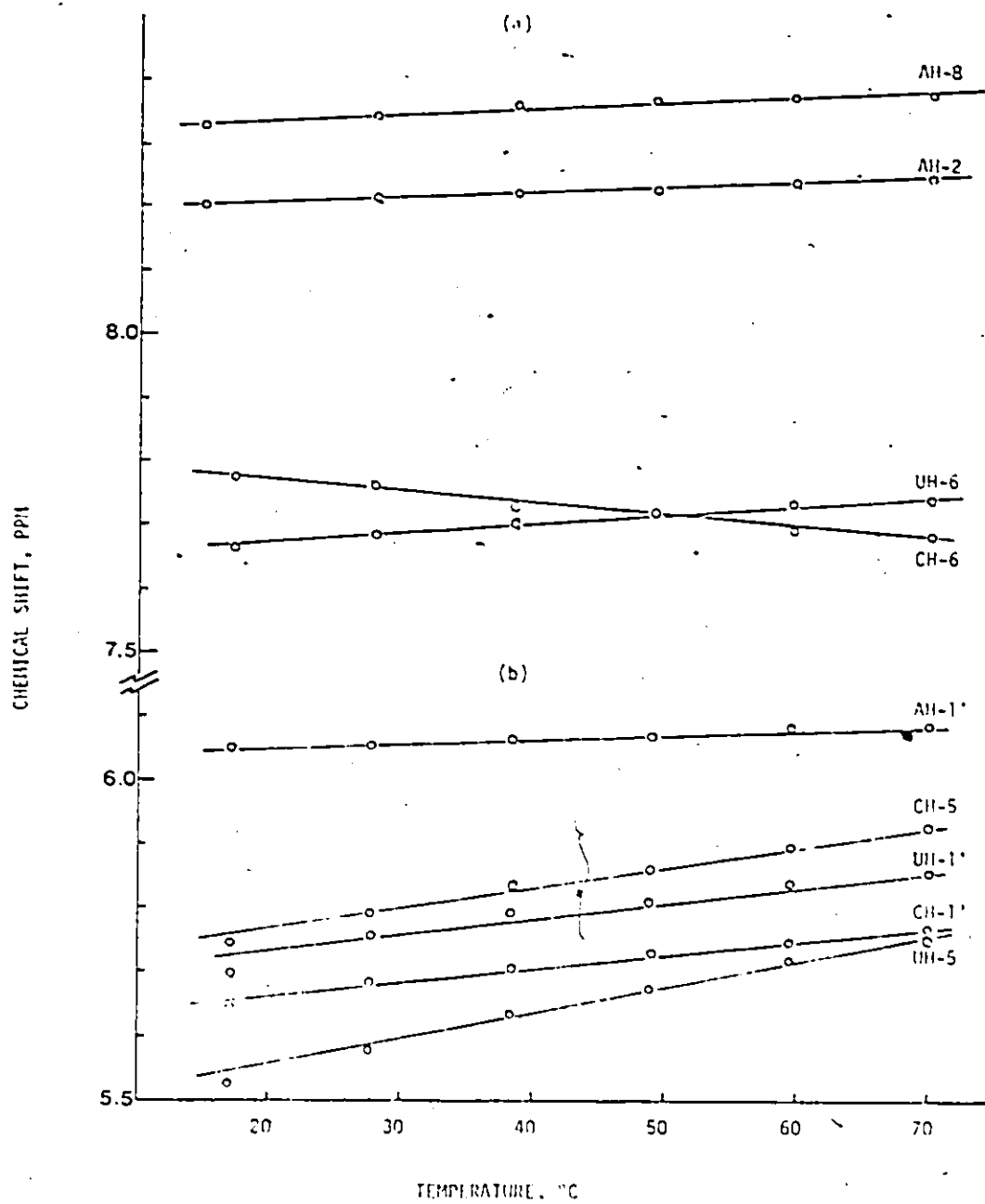


FIGURE 12. The variable temperature plots of the aromatic base and anomeric proton resonances of CAU.

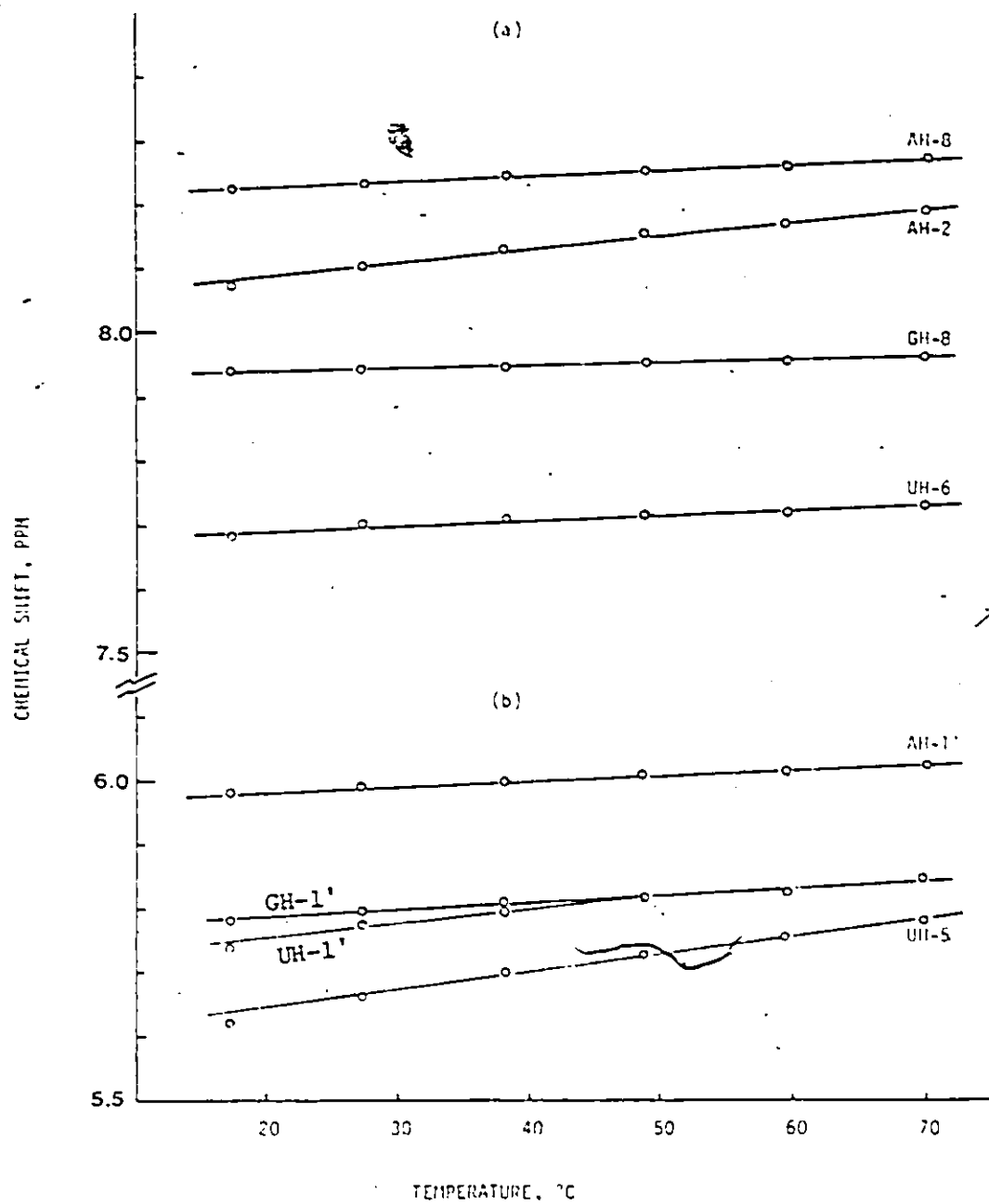


FIGURE 13. The variable temperature plots of the aromatic base and anomeric proton resonances of AUG.

equivalent, but began to separate at 50°C, and the UH-1' was assigned to the more shielded resonance.

The results of the variable temperature study of CAUG were significantly different from those of the trimer studies. The melting curves show a high degree of co-operativity (Figure 14). There was a rapid broadening of the resonances as the temperature was decreased below 35°C. Certain resonances could not be distinguished at these lower temperatures. The AH-2 signal was easily distinguished since it remained relatively sharp throughout the entire temperature range studied. The  $J_{1',2'}$  coupling constant values for the anomeric proton doublets of CAUG decreased as the temperature was lowered, and these resonances were broad singlets at temperatures below 20°C.

The  $T_m$  of CAUG was determined from the sigmoidal melting curves of AH-8, AH-2, GH-8 and GH-1' to be  $24 \pm 1^\circ\text{C}$ .

#### 4.1.2. Discussion

Chemical shift assignments were determined by the standard technique of incremental analysis (Borer *et al.*, 1975) in which the base and ribose anomeric proton resonances of the 70°C spectra of CA, CAU, and AUG were compared with those of CAUG. The chemical shift values for CA obtained in this work agreed well with previously published values (Ezra *et al.*, 1977). The 70°C spectrum of CAU contained three additional doublets appearing at 7.744 ppm (UH-6,  $J_{5,6} = 8.2$  Hz), 5.746 ppm (UH-5,  $J_{5,6} = 8.2$  Hz) and 5.856 ppm (UH-1',  $J_{1',2'} = 4.6$  Hz). The cytidine and uridine aromatic protons could be distinguished by a difference in the magnitude of the H-5, H-6 coupling constant (7.6 Hz for C, 8.2 Hz for U). Assignment of the

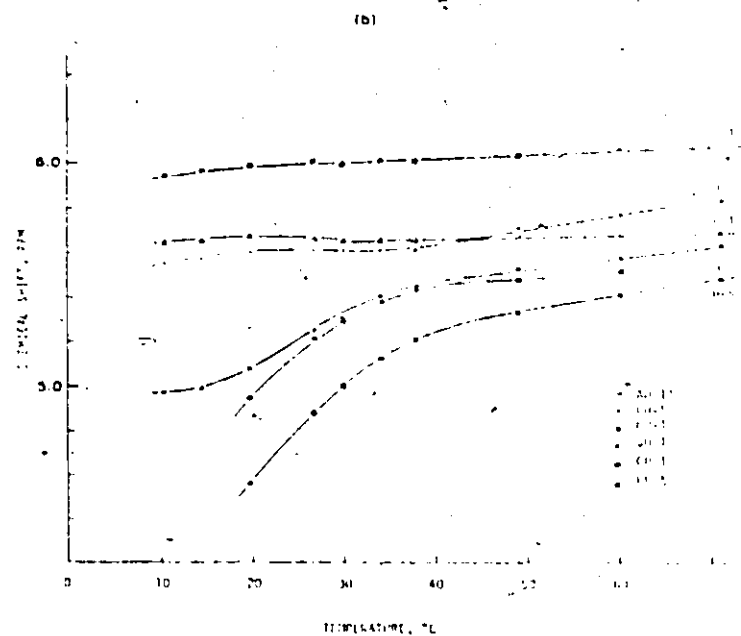
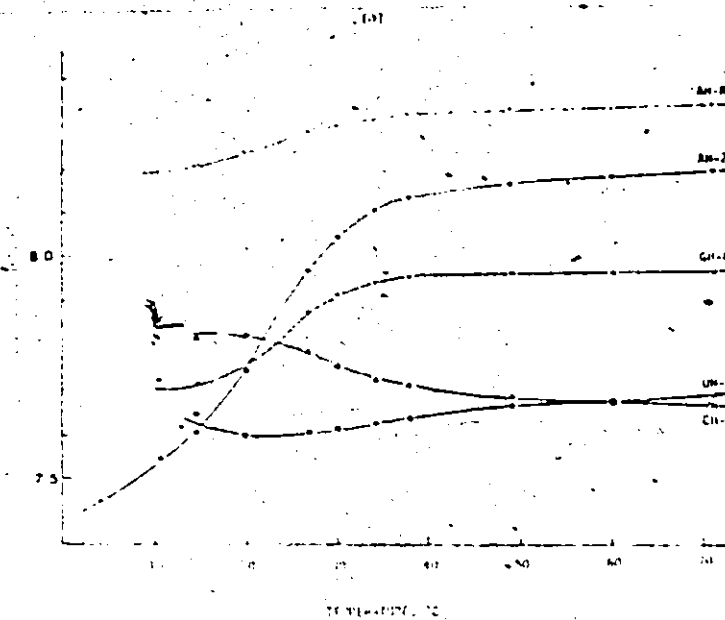


FIGURE 14. Results of the variable temperature study of CAUG.

POOR COPY

uridine H-5 was confirmed by irradiation of the H-6 resonance which resulted in the collapse of the doublet at 5.746 ppm.

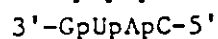
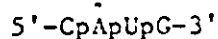
The lowfield signals of AUG were assigned by comparison to published chemical shift data for AU and UG (Ezra *et al.*, 1977). The UH-1' and GH-1' doublets were overlapped at higher temperatures and separated upon lowering the temperature below 50°C. The signal at lower field was tentatively assigned to GH-1' since it should be poorly shielded by a neighbouring uridine (Ezra *et al.*, 1977).

The low field aromatic resonances of the CAUG 70°C spectrum were assigned by comparison with the trimer data (Table 5 and Figure 11). The anomeric region of the spectrum was complicated by the overlap of five doublets. At 60°C the pyrimidine H-6 resonances were equivalent; simultaneous irradiation of these resonances resulted in the collapse of the H-5 doublets which facilitated the assignment of the anomeric protons. The GH-1' and UH-1' resonances of CAUG were nonequivalent at these temperatures, unlike the situation in AUG.

The chemical shift vs. temperature plots for CAU and AUG show that the resonances underwent essentially linear and upfield shifts as the temperature was lowered (Figures 12 and 13). These shift changes are the result of intramolecular base stacking and are not characteristic of base pairing (Hughes *et al.*, 1978). Reduction in the value of  $J_{1',2'}$  is a rough approximation of base stacking (Altona, 1975) and corresponds to a shift in the ribose conformational equilibrium towards the C3'-endo sugar pucker (Lee *et al.*, 1976; Ezra *et al.*, 1977). CAU experienced the largest conformational changes as evidenced by the collapse of the CH-1' and AH-1' resonances to broad singlets with

$J_{1',2'}$ , less than 1.0 Hz at 17°C. In contrast, the AUG anomeric couplings were still well resolved with a range of 2.0 to 4.5 Hz at this temperature.

The variable temperature behaviour of the CAUG resonances was dramatically different to that of the trimers. The high temperature (70 - 45°C) portion of each curve is linear, indicative of intramolecular base stacking, while at lower temperatures the shape of the curves reflect a co-operative melting process associated with the formation of the following base paired duplex:



The average  $T_m$  was determined to be  $24 \pm 1^\circ\text{C}$  from the well defined sigmoidal curves of the AH-8, AH-2, CH-8 and CH-1' proton resonances. It is interesting to note that the pyrimidine H-6 resonances follow essentially the same trend that was observed in CAU. This appears to be a result of the similarity in base stacking interactions in CAU and CAUG. In the anomeric region of the spectrum the AH-1', CH-5 and UH-1' resonances undergo relatively small shielding changes in comparison to CH-1', CH-1' and UH-5. Collapse of the anomeric resonances to broad singlets ( $J_{1',2'} < 1.0$  Hz) at temperatures of 20°C and lower is indicative of a high percentage of the C3'-endo ribose conformer, consistent with the formation of an A-RNA double helix.

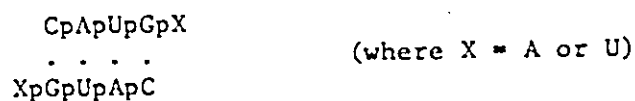
One interesting aspect of the duplex formation is the similarity in  $T_m$  values for resonances of both the terminal and interior base pairs. The small deviation ( $\pm 1^\circ\text{C}$ ) can largely be accounted for by the difficulty in making an exact determination of the transition midpoint

of the melting curve (Breslauer et al., 1975). The close agreement of the  $T_m$  values indicate an absence of significant fraying of the terminal G-C base pairs in the CAUG duplex, as was observed for CCGG (Arter et al., 1974). The  $T_m$  of the duplex was observed to be much higher than the calculated value of  $-3^\circ\text{C}$ . (Borer et al., 1974), demonstrating the need for further thermodynamic studies in order to improve the predictive methods (Tinoco Jr. et al., 1973). Increasing the % A+U content of a short duplex results in a lowering of the  $T_m$  value, as illustrated by comparing the  $T_m$ s of CAUG and GAGC:GCUC ( $24^\circ\text{C}$  vs.  $42^\circ\text{C}$  (Hughes et al., 1978)).

#### 4.2. The Effect of a Dangling Base Adjacent to Non-Fraying Terminal Base Pairs (Romaniuk et al., 1978b)

The dangling base effect was first noted in model studies on the duplex formation of  $A_mU_n$  oligoribonucleotides when  $m > n$  (Martin et al., 1971). The authors proposed that a dangling base increased favourable stacking interactions between the bases involved in base pairing. However, the terminal A-U base pairs of these and similar duplexes were subject to fraying (Uhlenbeck et al., 1971; Borer et al., 1973, 1975; Pohl, 1974) and this observation lead to the proposal that a dangling base increased the stability of a duplex by reducing fraying of the terminal base pairs (Kallenbach & Berman, 1977).

The hypothesis of Martin et al. (1971) was tested by studying the helix-coil transition of CAUGU and CAUGA. These oligoribonucleotides were expected to form a duplex of the type:



Since the CAUG duplex does not have fraying ends, any increase in  $T_m$  resulting from the presence of the dangling base must arise by some mechanism other than that proposed by Kallenbach and Berman (1977). If the alternate explanation of the dangling base effect is correct (Martin *et al.*, 1971), the measured  $T_m$ s of duplexes could be expected to follow the series CAUGA > CAUGU  $\geq$  CAUG, since A is a better base stacker than U (Lomant & Fresco, 1975). The pmr study yielded additional information about the dangling base effect that could not have been obtained from optical studies.

#### 4.2.1. Results

The procedure of incremental analysis (Borer *et al.*, 1975) was used to make the chemical shift assignments. The validity of this procedure relies upon the condition that adding a new residue to the sequence does not result in large chemical shift changes for proton resonances already assigned. Three additional doublets were observed in the 70°C CAUGU spectrum when compared to the CAUG spectrum: H-6 (7.783 ppm,  $J_{5,6} = 8.2$  Hz), H-5 (5.813 ppm) and H-1' (5.853 ppm,  $J_{1',2'} = 3.2$  Hz). The remaining resonances of the CAUGU spectrum can be assigned to the CAUG portion of the sequence by direct comparison to the CAUG data (Table 7). The complete spectral analysis of CAUGU at 70°C is illustrated by Figure 15.

For CAUGA the addition of the A(5) nucleotide resulted in a shielding of all of the base protons relative to CAUG at 70°C (Table 7). This did not complicate the assignment of the pyrimidine H-6s or the GH-8, but it was difficult to assign the A(2) and A(5) aromatic proton resonances. The assignment of these signals was made by comparing



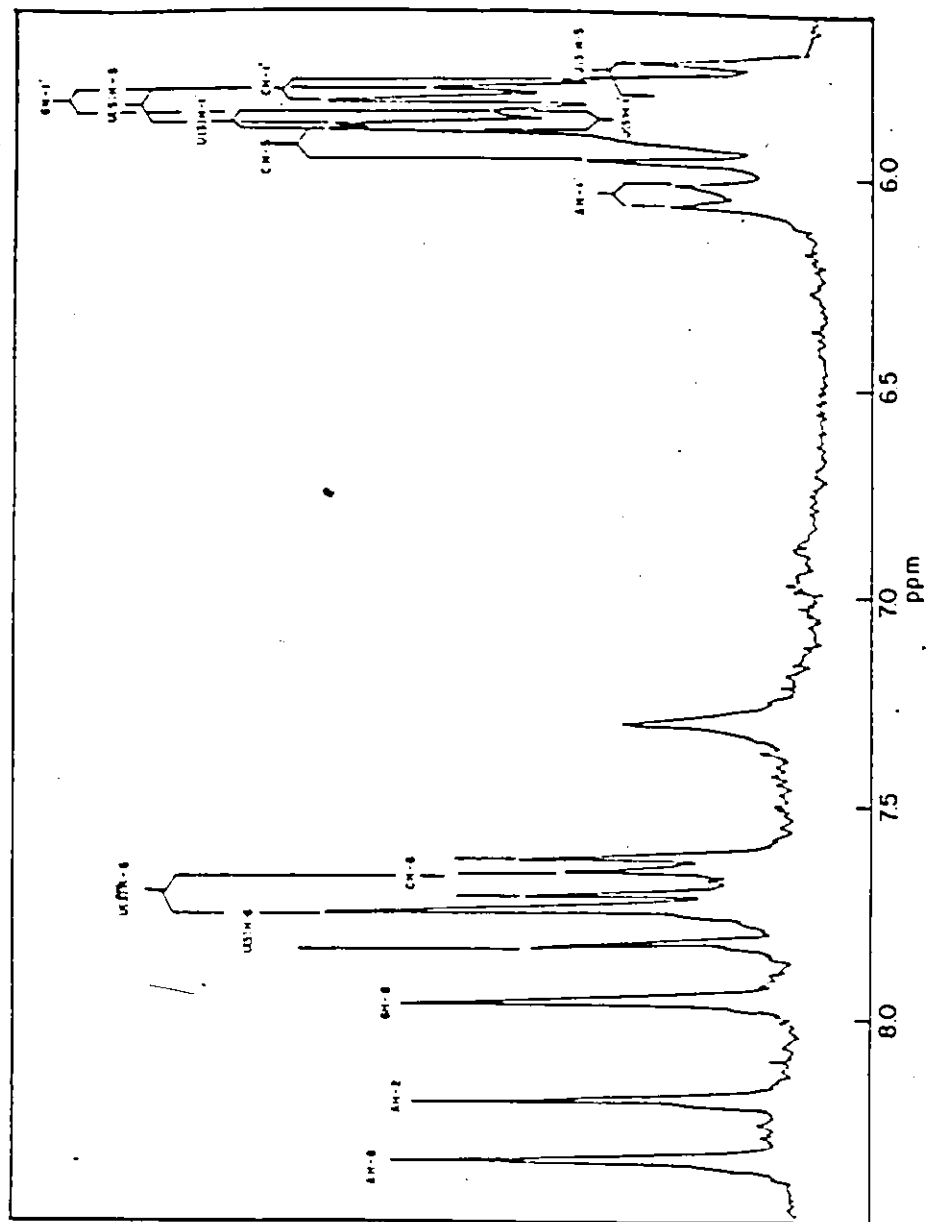


FIGURE 15. Complete spectral analysis of CAUGU at  $70^\circ\text{C}$ .

their variable temperature plots (Figure 18) with those for the A(2) of CAUG (Figure 14). The AH-2 resonance of CAUGA which undergoes the largest chemical shift change was assigned to the A(2) moiety, since this behaviour is typical of the A·U base pair of GAGC:GCUC (Hughes *et al.*, 1978) and CAUG (Figure 14). The AH-8 resonances were assigned on the same basis, and the complete assignments for CAUGA are summarized in Table 7. Figure 16 shows the 70°C spectral analysis of CAUGA.

Assignment of the ribose anomeric proton resonances of CAUGA was complicated by the overlap of the UH-1', GH-1', UH-5 and CH-1' (Figure 16). The CH-6 and U(3)H-6 resonances were equivalent at 60°C, similar to the situation for CAUG, and the assignments of the H-5 and anomeric proton resonances were made by simultaneously irradiating both H-6 resonances. Complete assignment of the CAUGU resonances required the use of similar decoupling experiments. The anomeric coupling constant data for both pentaribonucleotides are contained in Table 8.

The variable temperature data are shown in Figure 17 for CAUGU and Figure 18 for CAUGA. The melting curves are highly co-operative for the majority of resonances, indicative of a helix-coil transition. The H-8 and H-2 resonances of A(5) in CAUGA and the H-6 and H-5 resonances of U(5) in CAUGU underwent sigmoidal chemical shift changes as a function of temperature. This is of particular significance since these bases do not participate in hydrogen bonded base pair formation. Line broadening did not appear until the temperature was lowered below 30°C, and was not as severe as the line broadening

TABLE 7. Chemical shifts<sup>a</sup> of the oligoribonucleotides in D<sub>2</sub>O<sup>b</sup> at 70 °C

Proton	CAUG	CAUGA	CAUGU
CH-6	7.662	7.649	7.662
A(2)H-8	8.346	8.339	8.352
A(2)H-2	8.196	8.167	8.199
U(3)H-6	7.692	7.699	7.692
GH-8	7.962	7.890	7.958
A(5)H-8		8.316	
A(5)H-2		8.180	
U(5)H-6			7.783
CH-1'	5.765	5.721	5.772
A(2)H-1'	6.039	6.053	6.037
U(3)H-1'	5.845	5.779	5.849
GH-1'	5.813	5.764	5.796
A(5)H-1'		6.053	
U(5)H-1'			5.853
CH-5	5.912	5.897	5.912
U(3)H-5	5.783	5.741	5.749
U(5)H-5			5.813

<sup>a</sup>Chemical shifts are in ppm downfield from DSS using *t*-butyl alcohol-OD as an internal reference and are accurate to  $\pm 0.005$  ppm.

<sup>b</sup>pD = 7.0; Concentrations: CAUG, 9.2 mM; CAUGA, 9.2 mM; CAUGU, 7.7 mM.

TABLE 8. Coupling constants ( $J_{1',2'}$ )<sup>a</sup> of the oligoribonucleotides at 70 °C.

Proton	CAUG	CAUGA	CAUGU
CH-1'	3.8	4.1	3.8
A(2)H-1'	4.4	4.7	4.4
U(3)H-1'	4.8	5.0	4.1
GH-1'	5.3	3.5	5.3
A(5)H-1'		4.7	
U(5)H-1'			3.2

<sup>a</sup>Accurate to  $\pm 0.2$  Hz.

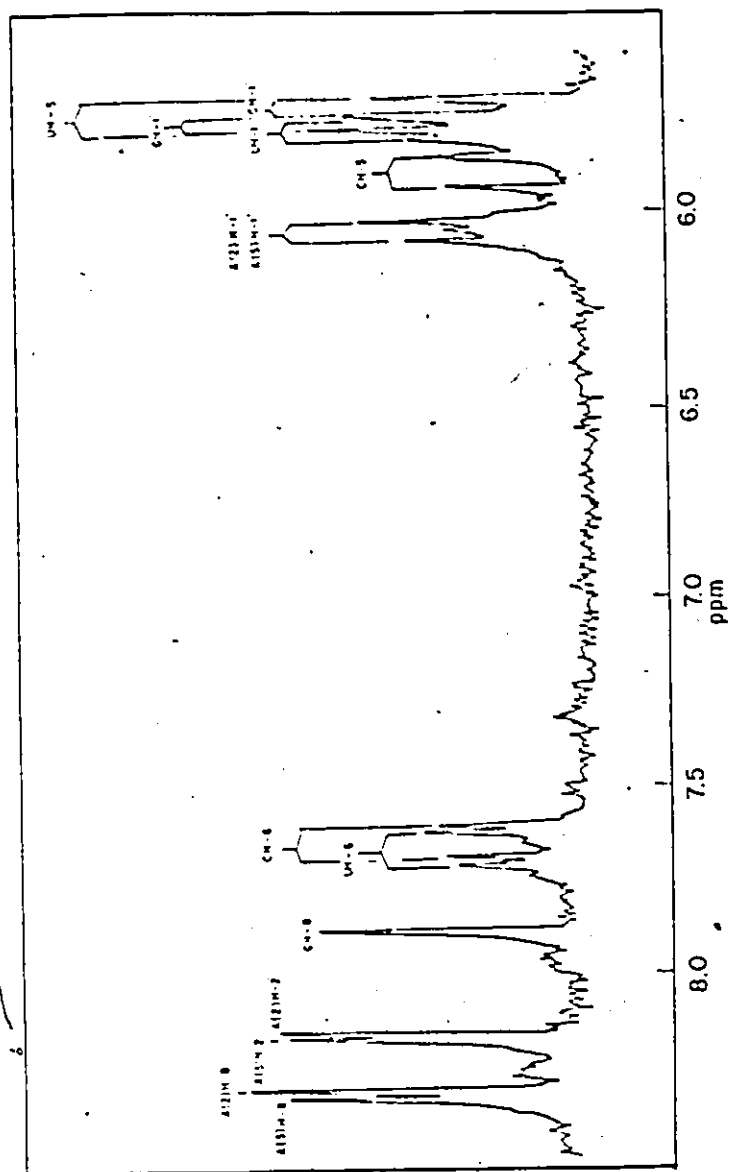


FIGURE 16. Complete spectral analysis of CAUGA at 70 °C.

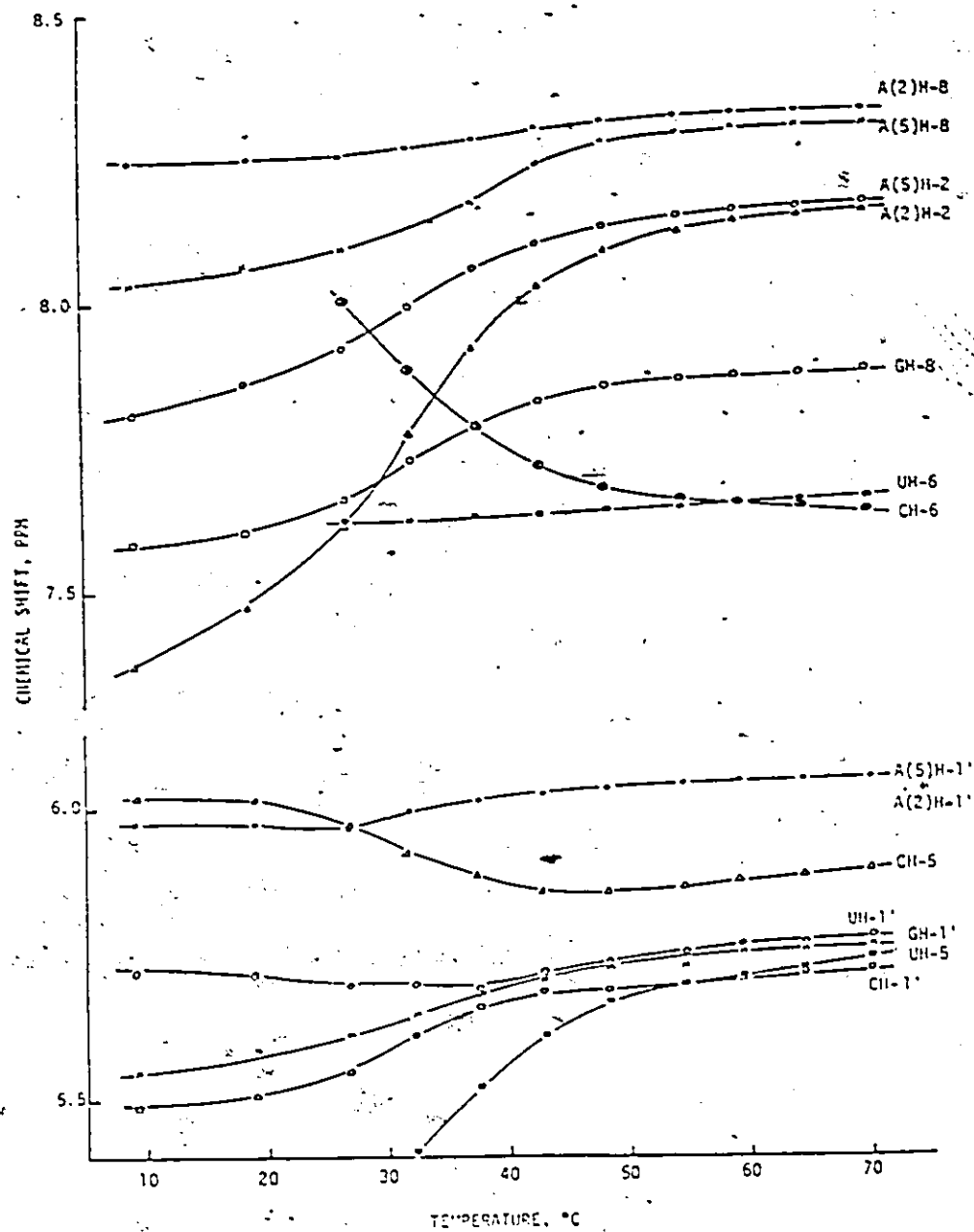


FIGURE 17. Results of the variable temperature study of CAUGU.

Poor Copy

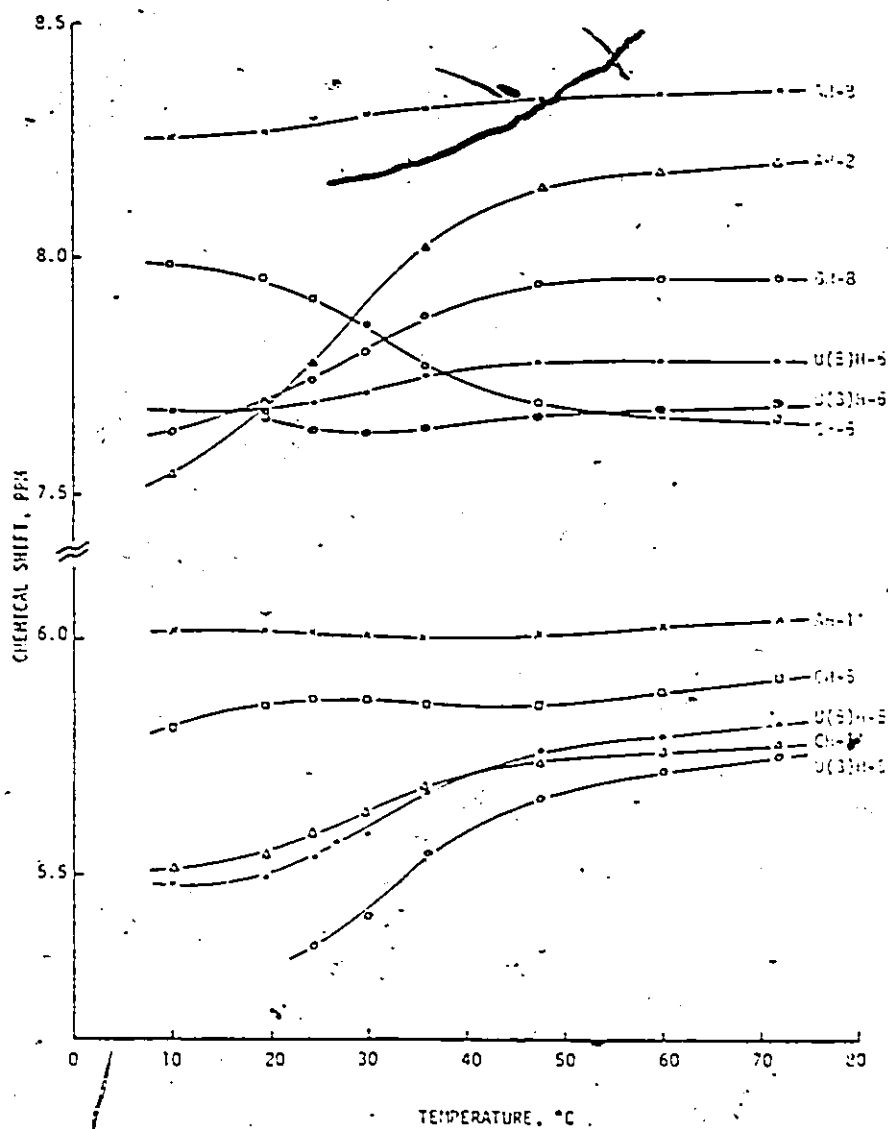


FIGURE 18. Results of the variable temperature study of CAUGA.

of the CAUG spectra. A temperature dependent reduction in the magnitude of the anomeric coupling constants occurred and these resonances were broad singlets below 27°C.

An average  $T_m$  of  $35.0 \pm 1^\circ\text{C}$  was measured for CAUGA, and a  $T_m$  of  $29.5 \pm 1^\circ\text{C}$  was determined for CAUGU.

#### 4.2.2. Discussion

The average melting temperatures of CAUG, CAUGU and CAUGA are summarized in Table 9. The  $T_m$  values for the pentanucleotides were determined only from the sigmoidal upfield curves of the residues involved in base pairing. The results indicate that a dangling base does increase the stability of a duplex without fraying ends, the effect being greater for a dangling A. This study was the first to demonstrate that a dangling U will increase duplex stability, a somewhat surprising result since U is considered to be a relatively poor stacker (Lomant & Fresco, 1975; Lee *et al.*, 1976; Ezra *et al.*, 1977). Studies on the duplex formed by the complementary set CAAUG: CAUUG ( $T_m = 28.5 \pm 2^\circ\text{C}$ ) indicate that the increase in  $T_m$  resulting from an additional internal A\*U base pair (cf. CAUG) is similar to the increase that would result from having a dangling U at each terminus, but less than the result from having a dangling A.

The results reported here are consistent with the proposal that a dangling base lends further stability to a duplex by increasing the favourable stacking interactions between the bases involved in base pairing (Martin *et al.*, 1971). This is best understood if one considers duplexing of short oligonucleotides to occur between single strands that are close to a fully stacked conformation (Appleby &

TABLE 9. Melting Temperatures and Concentrations of the Base Paired Duplexes.

Duplex	$T_m(^{\circ}\text{C})$	Concentration (mM)
$\begin{array}{c} \text{CAUG} \\ \vdots \quad \vdots \quad \vdots \\ \text{GUAC} \end{array}$	$24.0 \pm 1$	9.2
$\begin{array}{c} \text{CAUGU} \\ \vdots \quad \vdots \quad \vdots \\ \text{UGUAC} \end{array}$	$29.5 \pm 1$	7.7
$\begin{array}{c} \text{CAUGA} \\ \vdots \quad \vdots \quad \vdots \\ \text{AGUAC} \end{array}$	$35.0 \pm 1$	9.2
$\begin{array}{c} \text{CAAUG} \\ \vdots \quad \vdots \quad \vdots \\ \text{GUUAC} \end{array}$	$28.5 \pm 2.1$	11.0



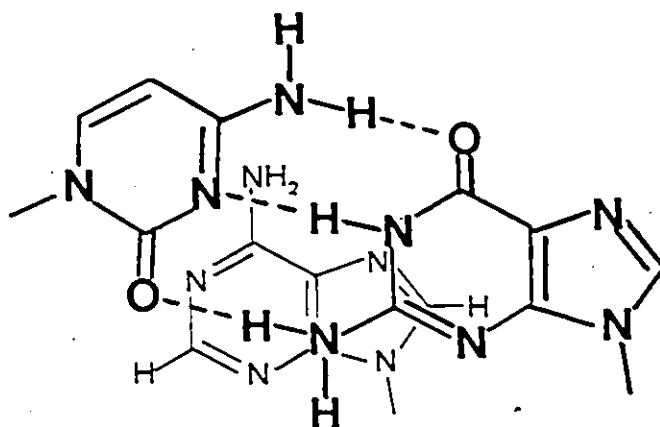
Kallenbach, 1973; Kallenbach & Berman, 1977), while longer oligomers probably have a high population of stacked conformers prior to and concomitant with duplexing. The use of NMR in this study has provided direct evidence of the base stacking interaction between duplex and dangling base, an interaction that could not be readily investigated by optical techniques.

The chemical shift vs. temperature plots for the aromatic proton resonances of the dangling bases in CAUGA and CAUGU are sigmoidal and shift upfield with decreasing temperature (Figures 17 and 18). In the past, such behaviour has been interpreted to be the result of direct involvement of the bases in base pairing (Borer *et al.*, 1975; Arter *et al.*, 1974). For example, one dangling base could hydrogen bond to another dangling base, increasing duplex stability via the formation of end-to-end aggregates. Such an interaction is unlikely however, as demonstrated by the variable temperature study of CAAUG and CAUUG. When each of these pentanucleotides was studied separately, the temperature dependent chemical shift changes were not sigmoidal, indicating the absence of U·U or A·A hydrogen bonding. An A·A interaction is unlikely at neutral pH under any circumstances (Ts'o, 1974b). The aggregate proposal can also be discounted on the basis of the temperature dependent broadening of the resonance line widths. As the temperature was lowered, the signals in CAUGA and CAUGU did not broaden as rapidly as the resonances of the CAUG spectra. It would appear that the dangling base reduces intermolecular end-to-end aggregation.

The temperature vs. chemical shift plots for the dangling base

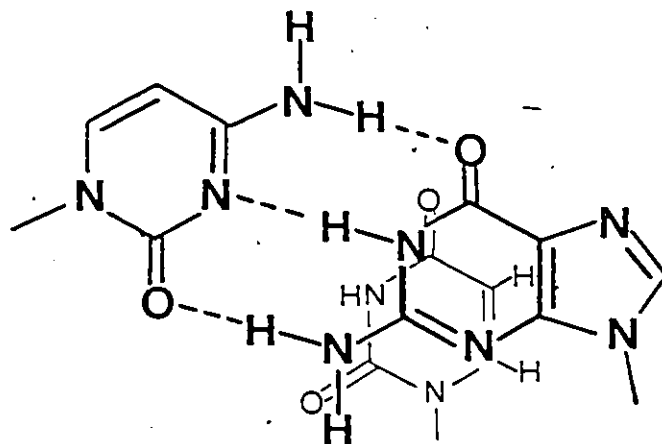
resonances of CAUGA and CAUGU are sigmoidal in nature even though these bases are not directly involved in base pairing. The dangling bases are experiencing the rapid conformational changes associated with single strand stacking and duplex formation through their base stacking interactions with the neighbouring bases in the duplex. This intimate relationship between the dangling base and the double helix is reflected in the highly co-operative melting transitions of the dangling base aromatic resonances. Additional evidence for the high degree of base stacking in these duplexes is the rapid reduction of  $J_{1',2'}$  for all the anomeric proton doublets upon lowering the temperature. As discussed earlier, such a reduction is a rough approximation of the increase in the percentage of C3'-endo ribose conformation arising from base stacking (Altona, 1975; Lee *et al.*, 1976; Ezra *et al.*, 1977).

Through a stacking interaction, the dangling base influenced the temperature dependent chemical shift behaviour of resonances of the adjacent G·C base pair. The CH-6 resonances of CAUGA and CAUGU were observed to experience a more rapid deshielding (Figures 17 and 18) than the corresponding resonance of CAUG (Figure 14). Such deshielding was also experienced for the CH-5 of CAUGA, while the same resonance in CAUG was shielded as the temperature decreased. The spatial relationship between the dangling base and the terminal G·C base pair was examined by building models of A'-RNA geometry (Borer *et al.*, 1975), and the results are illustrated by Figure 19. The cytidine H-5 and H-6 of CAUGA are approximately equidistant from the centre of the six membered ring of the dangling A, apparently in the



(a)

FIGURE 19. Diagrammatic representation of the base stacking interaction between the dangling base and terminal G-C base pair in (a) CAUGA (estimated distance of CH-5 and H-6 from the centre of the six membered ring of A(5): 6.5 Å) and (b) CAUGU (estimated distances of CH-5 and H-6 from the centre of the U(5) ring are 7.7 and 8.4 Å, respectively).



(b)

deshielding region of the adenine ring current (Giessner-Prettre et al., 1976). The result is similar for the dangling U (Figure 19) but the deshielding effect is weaker because of the smaller uridine ring current. The relatively larger deshielding of the cytidine aromatic resonances of the pentanucleotides may also reflect a decrease in end-to-end duplex aggregation in the presence of the dangling base.

Kallenbach and Berman (1977) suggested that the stacking of a dangling base minimizes fraying of the terminal base pair and consequently stabilizes the double helical region. However, fraying of the terminal G-C base pairs was not observed in CAUG, and the increased stability in these shorter sequences must arise from a higher degree of single strand stacking. The stability follows the series CAUGA > CAUGU > CAUG, which reflects the stacking capabilities of a dangling A vs. a dangling U. These results support the concept that base stacking is an integral process in double helix formation and makes a significant contribution to the helix stability (Appleby & Kallenbach, 1973). Further study of the role of dangling bases adjacent to fraying terminal base pairs were required in order to examine the proposal of Kallenbach and Berman (1977). Dangling base regions are important stabilizers of the binding interaction between various single stranded regions of RNAs (Grosjean et al., 1976; Yoon et al., 1976).

#### 4.3. The Effect of a Dangling Base on Fraying Terminal Base Pairs

An investigation of the effect of a dangling base on fraying required a reference duplex with fraying terminal ends. Terminal A-U base pairs are known to be more susceptible to fraying than terminal

G-C base pairs (Pohl, 1974). Pmr investigations of the  $A_2GCU_2$  duplex (Borer *et al.*, 1975; Kan *et al.*, 1975) produced direct evidence of terminal base pair fraying. Since it was desirable to use shorter sequences in this study, an investigation of the helix-coil transition of AGCU was warranted.

An optical study of the duplex formation of  $A_nGCU_n$  sequences found that AGCU did not form a double helix at  $10^{-6}$  M strand concentration (Ravetch *et al.*, 1974). A  $T_m$  of  $6^\circ\text{C}$  has been measured for the helix-coil transition of the related oligoribonucleotide AGCUC at 0.1 mM concentration (England & Neilson, 1977). Uhlenbeck proposed that internal G-C base pairs could contribute more to duplex stability than terminal G-C base pairs (Uhlenbeck *et al.*, 1971). The AGCU duplex might be expected to have a higher  $T_m$  than CAUG at comparable strand concentration. The results discussed below demonstrate that AGCU forms a duplex with fraying ends that is more stable than the CAUG duplex.

The sequence AGCUA was used to investigate the effect of a dangling base on the stability of the AGCU duplex. The primary interest of this study was in testing the hypothesis of Kallenbach and Berman (1977) that a dangling base can reduce the fraying of terminal base pairs. Based on the results from the CAUGX series, a dangling A could be expected to have the greater stabilizing effect.

#### 4.3.1. Results

The procedure of incremental analysis (Borer *et al.*, 1975) was used to make the chemical shift assignments, which are summarized in Table 10. The resonances of AG were assigned by comparison to

TABLE 10. Incremental analysis of AGCU and AGCUA at 70 °C in D<sub>2</sub>O<sup>a</sup>

Proton	AG	AGC	AGCU	AGCUA
CHEMICAL SHIFTS <sup>b</sup>				
AH-8	8.238	8.242	8.248	8.229
AH-2	8.186	8.180	8.176	8.144
GH-8	7.942	7.926	7.929	7.929
CH-6		7.744	7.737	7.704
UH-6			7.768	7.698
A(5)H-8				8.343
A(5)H-2				8.186
AH-1'	5.967	5.974	5.978	5.965
GH-1'	5.842	5.812	5.792	5.798
CH-1'		5.886	5.892	5.869
UH-1'			5.892	5.817
A(5)H-1'				6.063
CH-5		5.876	5.850	5.839
UH-5			5.821	5.793
COUPLING CONSTANTS <sup>c</sup>				
AH-1'	5.0	4.4	4.7	4.7
GH-1'	4.7	3.8	3.4	3.2
CH-1'		3.1	4.4	5.5
UH-1'			4.4	3.8
A(5)H-1'				4.7

<sup>a</sup>pD = 7.0; Concentrations: AG, 15.5 mM; AGC, 6.2 mM; AGCU, 11.1 mM; AGCUA, 11.6 mM.

<sup>b</sup>Chemical shifts are in ppm relative to DSS using *t*-butyl alcohol-OD as an internal reference and are accurate to ±0.005 ppm.

<sup>c</sup>Accurate to ±0.2 Hz.

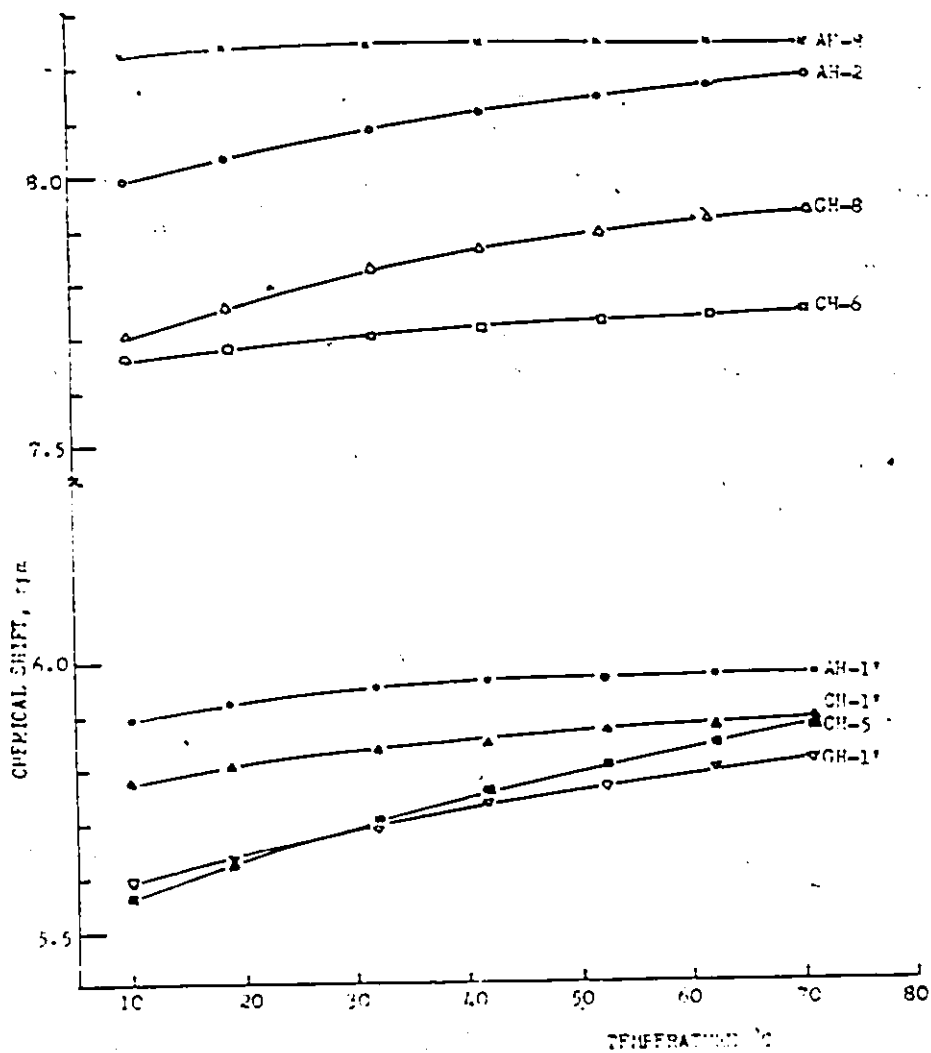


FIGURE 20. The variable temperature plots of the aromatic base and anomeric proton resonances of AGC.

Poor Copy

published data (Lee et al., 1976). Extension of the sequence to AGC resulted in the appearance of three additional doublets in the 70°C spectrum: CH-6 (7.744 ppm), CH-5 (5.876 ppm) and CH-1' (5.886 ppm). The results of a variable temperature study of AGC are presented in Figure 20. Although the plots are not strictly straight lines, the transitions show very little co-operativity when compared to those for AGCU or AGCUA (cf. Figures 21 and 22) and can be interpreted as resulting from single strand stacking.

Although the aromatic base resonances of the 70°C AGCU spectrum could be readily assigned by comparison to the AGC data (Table 1), the anomeric region of the spectrum was complicated by the overlap of five doublets. Simultaneous irradiation of both pyrimidine H-6 resonances and subsequent collapse of the two pyrimidine H-5 resonances aided in the assignment of the anomeric ribose resonances. The UH-1' and CH-1' resonances were equivalent at 70°C, but separated at lower temperatures. The higher field signal was tentatively assigned to CH-1', since this proton should experience greater shielding from the nearest neighbour guanine residue (Giessner-Prettre et al., 1976). Decoupling experiments were used at several lower temperatures as an aid in assigning the H-1' and H-5 signals. At temperatures of 30°C and lower, the ribose anomeric proton resonances were observed to be broad singlets ( $J_{1',2'} < 1.0$  Hz). Such a reduction in the coupling constant reflects a high degree of base stacking (Altona, 1975). The spectra were observed to undergo significant line broadening at temperatures of 31°C and lower. The results of a variable temperature study of the helix-coil transition of AGCU are presented in Figure 21.



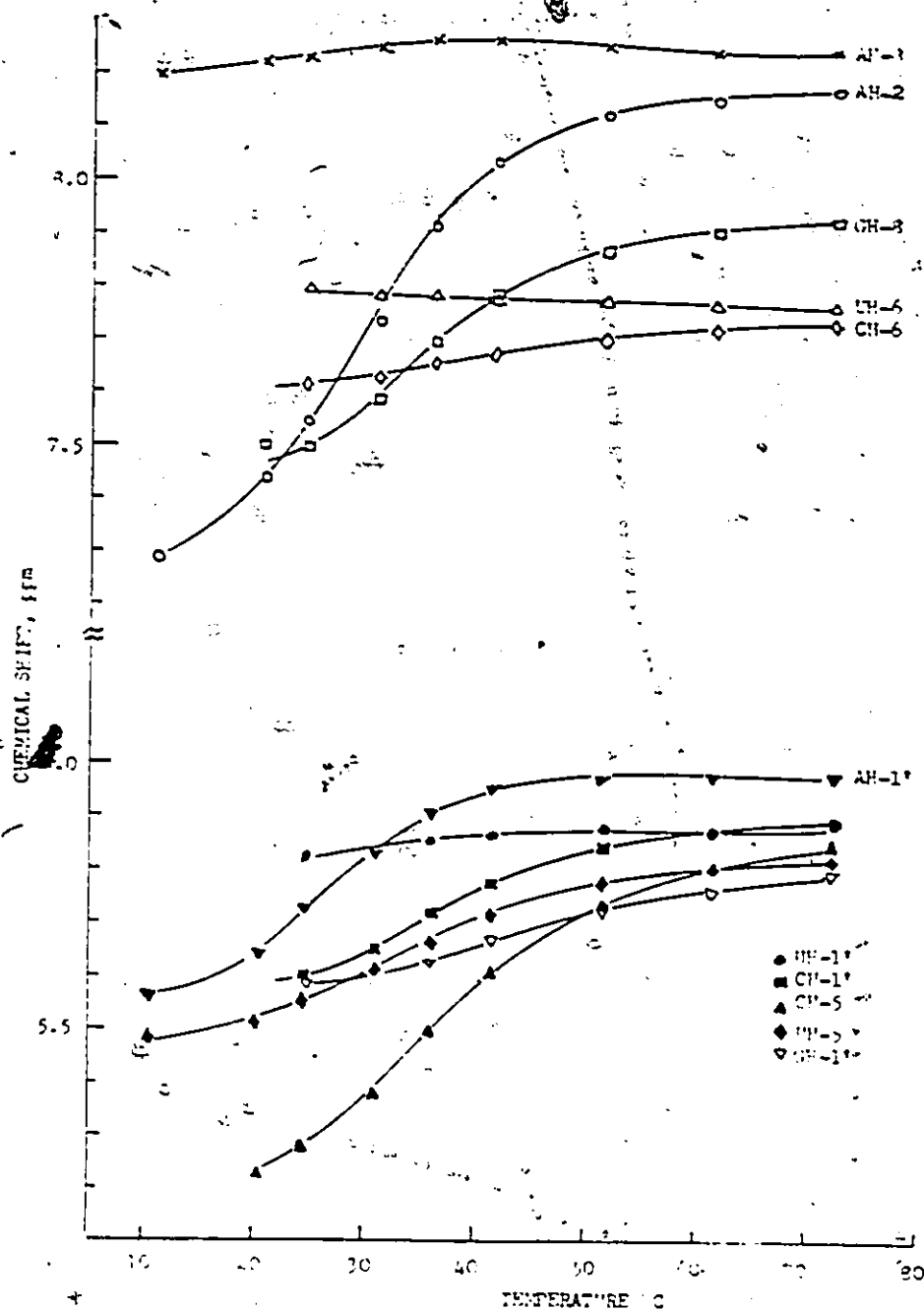


FIGURE 21. Results of the variable temperature study of AGCU.

Poor Copy

Comparison of the 70°C spectrum of AGCUA with that of AGCU shows the appearance of three new signals: A(5)H-8 (8.343 ppm), A(5)H-2 (8.186 ppm) and AH-1' (6.063 ppm,  $J_{1',2'} = 4.7$  Hz). Addition of the terminal A resulted in a general shielding of most of the resonances (Table 10). The only difficulty this shielding caused was in the assignment of the AH-2 resonances. The more shielded resonance was assigned to A(1)H-2, which reflects the general shielding trend throughout the sequence. A reversal of the AH-2 assignments would not alter the results of the variable temperature study. At 62.5°C, the pyrimidine H-6 signals were equivalent, and simultaneous irradiation of these resonances was used to confirm the H-5 and H-1' assignments. It is interesting to note that the UH-1' and CH-1' signals are not equivalent at 70°C, in contrast to the result observed for AGCU. The ribose anomeric coupling constants decreased as the temperature was lowered, until all but one of these resonances were observed to be broad singlets at temperatures of 41°C and lower. The exception to this trend was the A(5)H-1', which has a well resolved coupling constant at all temperatures studied. The dangling base must not stack as completely as the other bases, a result which is discussed on the basis of fraying phenomenon. Line broadening of the AGCUA spectra was not as severe as that observed for AGCU, and begins to occur at lower temperatures (20°C vs. 31°C). The results of a variable temperature study of the helix-coil transition of AGCUA are presented in Figure 22.

Determination of average  $T_m$  values for AGCU and AGCUA was complicated by the wide range of individual melting transition

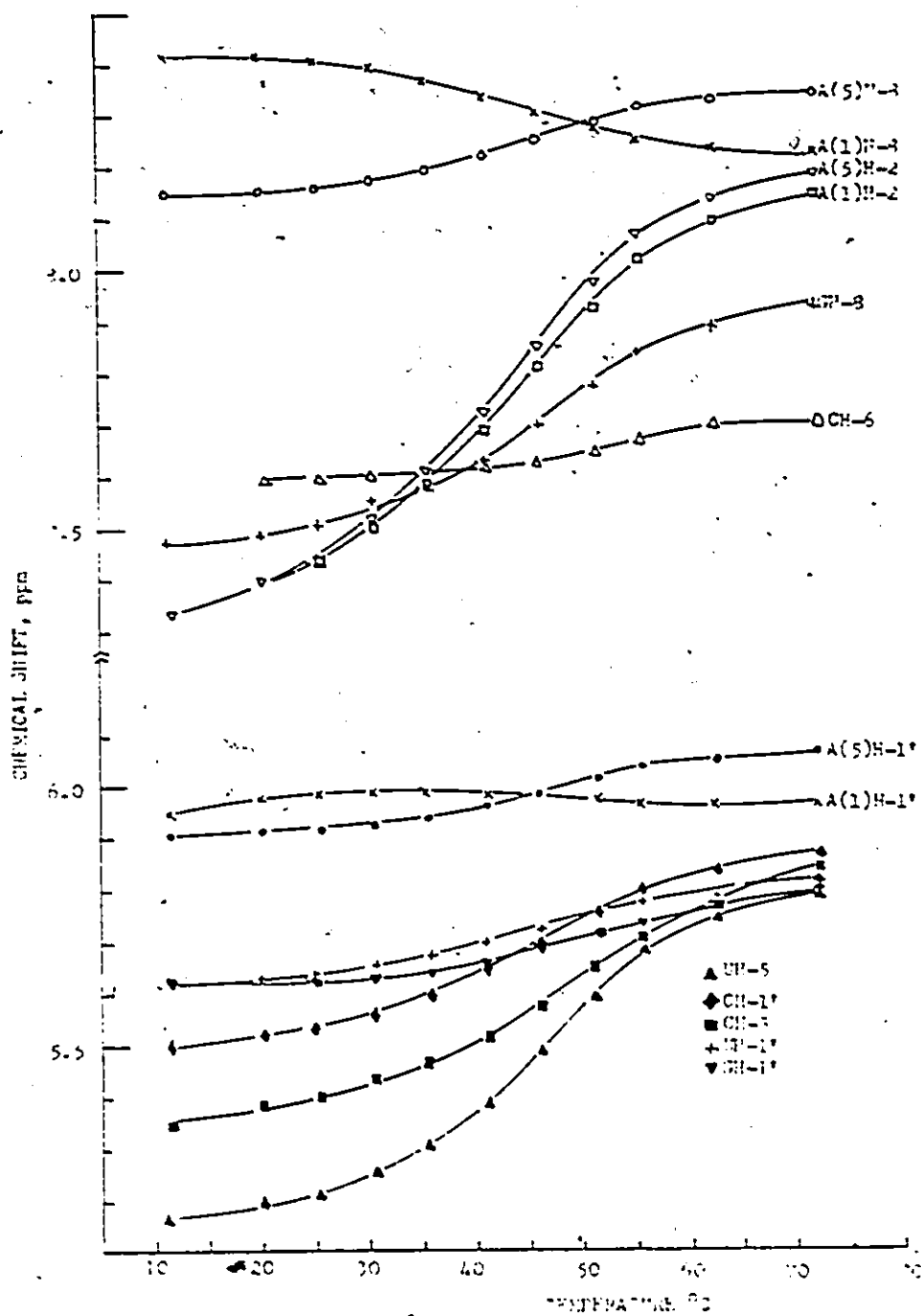


FIGURE 22. Results of the variable temperature study of AGCUA.

TABLE 11. Determination of the Melting Temperatures of AGCU and AGCUA.

Resonance	T <sub>m</sub> (°C)	Average T <sub>m</sub> / Base Pair	Average T <sub>m</sub> for Duplex
(1) AGCU			
AH-2	28.0	30.5 ± 2.2	33.6 ± 4.4
AH-1'	32.0		
UH-5	31.5		
GH-8	36.5	36.7 ± 0.3	
CH-6	37.1		
CH-5	36.5		
CH-1'	36.8		
(2) AGCUA			
A(5)H-8	45.7	44.8 ± 2.3	47.3 ± 3.6
A(5)H-2	44.8		
A(5)H-1'	44.5		
A(1)H-2	43.2	47.9 ± 3.0	
U(4)H-5	46.4		
G(2)H-8	47.0		
C(3)H-5	49.7	53.0	
C(3)H-6	53.0		

temperatures (Table 11). The  $T_m$  values could be more closely grouped according to A·U or G·C base pair formation. An average  $T_m$  of  $33.6 \pm 4.4^\circ\text{C}$  was determined for AGCU, and comparison with the  $T_m$  value of  $47.3 \pm 3.6^\circ\text{C}$  for AGCUA illustrates the stabilizing effect a dangling A exerts on the AGCU duplex. The results are consistent with a fraying model for both of these short duplexes (Borer *et al.*, 1975; Patel, 1976).

#### 4.3.2. Discussion

Determination of the  $T_m$  values for the AGCU and AGCUA duplexes is summarized in Table 11. The individual transitions were grouped according to the base pair being formed, and an average  $T_m$  value for each base pair was determined. For both duplexes, the G·C base pair had a significantly higher  $T_m$  than the A·U base pair. Such a result may arise by two possible mechanisms: intramolecular fraying of the terminal base pairs, or intermolecular end-to-end aggregation of the duplexes.

In his studies on the self-complementary sequence isomers of d(CCGG), Patel observed evidence of end-to-end aggregation of the short double helices at 20 mM strand concentration (Patel, 1976b, 1977, 1979). This aggregation was easily distinguished by comparing the variable temperature plots of the non-exchangeable aromatic base resonances for terminal and internal G·C base pairs (Patel, 1977). The plots for internal base pair resonances were sigmoidal in nature, containing a region of rapid shielding in the middle temperature range, followed by a gradual leveling off at lower temperatures (Figure 23). By contrast, the plots for resonances assigned to

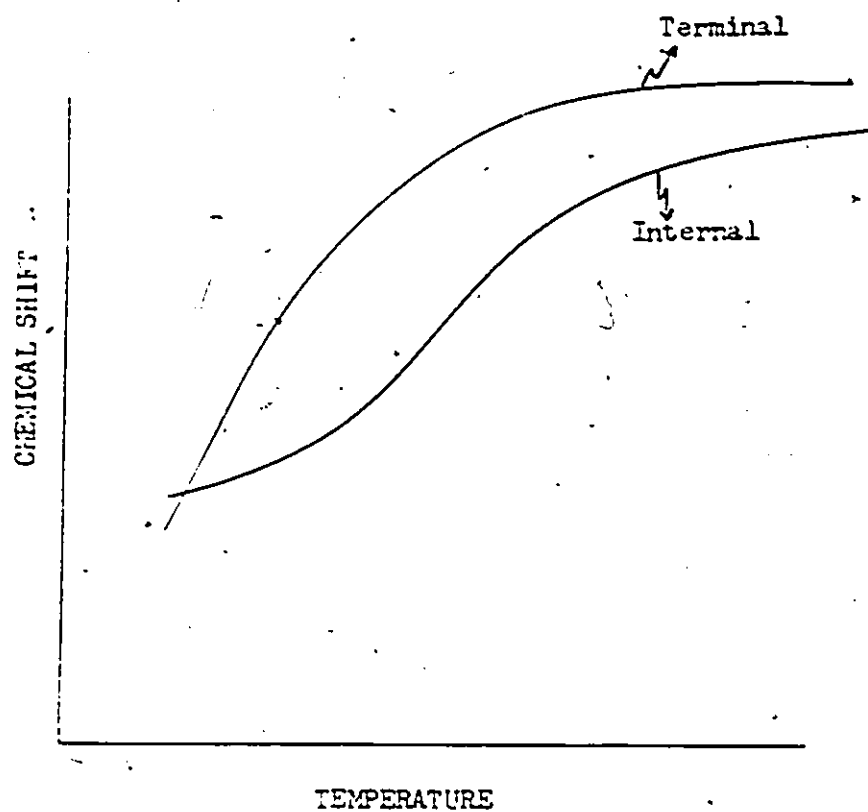


FIGURE 23. The effect of interduplex end-to-end aggregation on the melting transitions of resonances from the terminal and internal base pairs of short duplexes.

Poor Copy

terminal base pairs had lower  $T_m$  values and showed a continuous upfield shift at lower temperatures without a plateau region (Figure 23). A variable concentration study of the helix-coil transition of d(GGCC) demonstrated that reducing the strand concentration from 20 mM to 5 mM reduced the end-to-end aggregation of duplexes (Patel, 1979). All of the transitions were sigmoidal, with little temperature dependence of chemical shifts at lower temperatures, and had  $T_m$  values in close agreement.

Examination of the variable temperature plots for the AGCU (Figure 21) and AGCUA (Figure 22) duplexes reveals that the difference in  $T_m$  values for the A·U and G·C base pairs can not be accounted for on the basis of end-to-end aggregation of the duplexes. All of the melting curves are sigmoidal in nature and exhibit a decrease in the temperature dependence of the shielding at lower temperatures. These results do not give good agreement with Patel's criteria for end-to-end aggregation. Additional evidence against the explanation of the base pair  $T_m$  difference by an aggregation model comes from the results for AGCUA. Line broadening of the AGCUA spectra was not as severe and occurred at lower temperatures than line broadening of the AGCU spectra. Although the  $T_m$  of the AGCUA duplex is considerably higher than the  $T_m$  for the AGCU duplex, the line broadening results indicate that there is a considerable reduction of end-to-end interduplex aggregation in the presence of a dangling base. However, as the results of Table 11 illustrate, there is not a significant reduction in the  $\Delta T_m$  between the A·U and G·C base pairs. Apparently this difference can not be accounted for on the basis of interduplex

end-to-end aggregation.

NMR studies of the helix-coil transition of  $A_2GCU_2$  provided direct evidence of the fraying of the terminal A·U base pairs and an end-in melting of the duplex (Borer *et al.*, 1975; Kan *et al.*, 1975). A wide variance of  $T_m$  values was observed for the melting curves of individual resonances of non-exchangeable protons. The CCGG duplex, in contrast, did not fray and had a well defined average  $T_m$  value of  $51 \pm 2^\circ\text{C}$  (Arter *et al.*, 1974). Even though there were large discrepancies in  $T_m$  values for resonances of the same base pair (eg. CH-8,  $51^\circ\text{C}$ ; CH-6,  $60^\circ\text{C}$ ), the G·C base pairs were significantly more stable than the A·U base pairs, which was attributed to fraying of the terminal A·U base pairs (Borer *et al.*, 1975). Studies of the hydrogen bonded ring N-H protons provided additional evidence for a fraying model (Kan *et al.*, 1975).

The results summarized in Table 11 are consistent with a fraying duplex model for both AGCU and AGCUA. The melting curves are sigmoidal in nature, and the  $T_m$  of the A·U base pair is significantly lower than the  $T_m$  of the G·C base pair in each duplex. Pmr studies of the related deoxy sequence d(AGCT) indicate a much higher degree of fraying in the terminal A·T base pairs of this duplex than observed for AGCU (N.R. Kallenbach, personal communication). The temperature dependent behaviour of the anomeric coupling constants provide additional support for fraying of the AGCUA duplex. In the AGCU duplex, all H-1' doublets are reduced to broad singlets ( $J_{1',2'} < 1.0 \text{ Hz}$ ) at temperatures of  $31^\circ\text{C}$  and lower. This result is consistent with a high degree of base stacking (Altona, 1975), necessary for

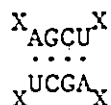


double helix formation. The A(5)H-1' doublet of AGCUA, however, remained well resolved at temperatures far below the  $T_m$  of the duplex (eg.  $J_{1',2'} = 2.4$  Hz at  $20^\circ\text{C}$ ), while the rest of the anomeric signals were singlets ( $J < 1.0$  Hz) below  $41^\circ\text{C}$ . This result implies that the dangling A residue has a relatively high degree of rotational freedom with respect to the bases involved in base pairing similar to the orientation of the CCA arm of tRNAs (Rich, 1977). All of the bases in the CAUGX duplexes ( $X = U, A$ ) were highly stacked, including the dangling bases. Such rotational freedom of the dangling A of the AGCUA double helix is consistent with fraying of the terminal A-U base pairs.

Since fraying of the terminal base pairs is not reduced by the presence of the dangling base, the increased stability of the duplex region must arise by a different mechanism. The  $\Delta T_m$  value for AGCUA-AGCU ( $14^\circ\text{C}$ ) is similar to the  $\Delta T_m$  CAUGA-CAUG ( $11^\circ\text{C}$ ). The dangling base must increase the overall stacking in such a way that the stability of each base pair is affected to the same extent, and the relative stabilities of internal vs. terminal base pairs remain the same, as illustrated by the results of this study (Table 11). It is interesting to note that a high percentage of stacked single strands is reached at  $41^\circ\text{C}$  for AGCUA and  $31^\circ\text{C}$  for AGCU. These results support the concept that duplex formation from short oligoribonucleotides requires that the single strands be close to a fully stacked conformation (Appleby & Kallenbach, 1973; Kallenbach & Berman, 1977). Additional support for a stacking role for the dangling base comes from two sources. First, the A(1)H-8 of AGCUA is more greatly deshielded upon duplex formation

(Figure 22) than the A(1)H-8 of AGCU (Figure 21). As the results for CAUGA have already demonstrated, this cross-strand deshielding arises from a stacking of the dangling A. Additional support comes from the  $T_m$  values of the A(5) resonances of AGCUA (Table 11). These values agree closely with the  $T_m$  values for the A·U base pair, probably because of the stacking interaction between the terminal A·U pair and the dangling base.

The results for AGCUA suggest a mechanism by which relatively large loops could stabilize adjacent duplex regions. A slight fraying of the base pairs which close the loop could allow a dangling base interaction to occur between the single strands of the loop and the double helical region, leading to a greater stability for the duplexes. Such a mechanism could be investigated using the appropriate short oligonucleotide sequences which would form a duplex of the type:



#### 4.4. The AGGA:UCCU Duplex: A Model of the Shine-Dalgarno mRNA-16 S rRNA Interaction

An extensive search continues for a mechanism to explain how a 30 S ribosomal subunit selectively forms an initiation complex with the correct AUG codon of an mRNA. The ribosomal binding sites of several mRNAs have been isolated by a ribonuclease digestion technique and subsequently base sequenced (for example, Maizels, 1974; Contreras *et al.*, 1973; Shine & Dalgarno, 1975 and references therein). There is a purine rich region to the 5' side of the initiator AUG codon on

these mRNAs which is complementary to a pyrimidine rich region at the 3' terminus of 16 S rRNA (Shine & Dalgarno, 1975). An appropriate section of the secondary structure of the 16 S rRNA could open up in order to base pair with the purine rich region of the mRNA (Steitz & Jakes, 1975). The short duplex AGGA:UCCU is common in many of the mRNA-rRNA interactions.

The AGGA:UCCU duplex is a sequence isomer of the CAUG and AGCU double helices. One question that can be asked about the mRNA-rRNA initiation complex is whether selection of an inherently more stable duplex from other duplexes with the same base composition has occurred. A study of the Helix-coil transition of the complementary oligoribonucleotides AGGA and UCCU, with comparison to the CAUG and AGCU results can begin to answer this question. The AGGA:UCCU duplex was found to be the most stable of the three duplexes. A surprising result was the observation that there was no detectable fraying of the terminal A·U base pairs in the AGGA:UCCU duplex. Fraying appears to be a sequence related phenomenon.

#### 4.4.1. Results

The procedure of incremental analysis (Borer et al., 1975) was used to make the chemical shift assignments. These data are summarized in Table 12. The resonances of AG were assigned by comparison with published data (Lee et al., 1976). Extension of AG to AGG was reflected by the appearance of two additional signals in the pmr spectrum: G(3)H-8 (7.906 ppm) and G(3)H-1' (5.785 ppm,  $J_{1',2'} = 5.0$  Hz). A general shielding of all resonances was observed in the 70°C AGGA spectrum, but this did not seriously complicate the chemical shift

TABLE 12. Incremental analysis of the AGGA:UCCU duplex in D<sub>2</sub>O<sup>a</sup> at 70 °C

proton	12 AG	123 AGG	1234 AGGA	AGGA 1234 UCCU	4321 UCCU	432 UCC	43 UC
CHEMICAL SHIFTS <sup>b</sup>							
A(1)H-8	8.239	8.212	8.196	8.193			
A(1)H-2	8.186	8.167	8.141	8.137			
G(2)H-8	7.942	7.935	7.883	7.883			
G(3)H-8		7.906	7.874	7.874			
A(4)H-8			8.294	8.290			
A(4)H-2			8.157	8.154			
U(4)H-6				7.806	7.806	7.806	7.806
C(3)H-6				7.837	7.837	7.832	7.832
C(2)H-6				7.819	7.814	7.832	
U(1)H-6				7.806	7.806		
A(1)H-1'	5.967	5.946	5.933	5.933			
G(2)H-1'	5.842	5.837	5.726	5.726			
G(3)H-1'		5.785	5.726	5.726			
A(4)H-1'			6.037	6.037			
U(4)H-1'				-	- <sup>c</sup>	5.837	5.864
C(3)H-1'				-	-	5.927	5.934
C(2)H-1'				-	-	5.927	
U(1)H-1'				-	-		
U(4)H-5				5.917	5.919	5.858	5.883
C(3)H-5				6.037	6.037	6.045	6.058
C(2)H-5				6.005	6.005	6.018	
U(1)H-5				5.949	5.949		
<sup>d</sup> J <sub>1',2'</sub>							
A(1)H-1'	5.0	5.3	5.3	5.3			
G(2)H-1'	4.7	5.0	5.0	5.0			
G(3)H-1'		5.0	5.0	5.0			
A(4)H-1'			4.7	4.7			
U(4)H-1'				-	- <sup>c</sup>	4.4	4.7
C(3)H-1'				-	-	4.1	3.2
C(2)H-1'				-	-	4.1	

<sup>a</sup>pD=7.0; concentrations: AG, 15.5 mM; AGG, 6.3 mM; AGGA, 7.0 mM; UC, 23 mM; UCC, 9.8 mM; UCCU, 7.0 mM; Duplex, 7.0 mM each strand, 14 mM total.

<sup>b</sup>Chemical shifts are in ppm relative to DSS using *t*-butyl alcohol-OD as an internal reference and are accurate to  $\pm 0.005$  ppm.

<sup>c</sup>These resonances could not be assigned because of the high degree of overlap in this region of the spectrum.

<sup>d</sup>Accurate to  $\pm 0.2$  Hz.

assignments (Table 12).

The resonances of UC were assigned by comparison with published data for this dinucleoside monophosphate (Lee et al., 1976). Comparison of the UCC spectrum with that of UC showed the presence of three additional doublets: C(2)H-6 (7.832 ppm, overlapped with C(3)H-6), C(2)H-5 (6.018 ppm) and C(2)H-1' (5.927 ppm,  $J_{1',2'} = 4.1$  Hz; overlapped with C(3)H-1'). Extension of the sequence to UCCU resulted in an appearance of new signals attributed to the U(1) nucleotide, and resulted in the two CH-5s being non-equivalent in the tetramer (Table 12). Absolute assignment of the H-1' resonances of UCCU was not possible because of the overlap of 8 doublets in this region of the spectrum.

Each of the tetramers was studied separately by pmr spectroscopy in D<sub>2</sub>O (no salt) before the mixing experiment was performed. The results of the variable temperature study on AGGA are presented in Figure 24. As the temperature is lowered, each of the resonances is slightly shielded in essentially a linear fashion. Such behaviour is analogous to that displayed by CAU and AUG, and is attributed to a single strand base stacking process. Line broadening of the spectra became evident at temperatures of 30°C and was quite pronounced at 20°C and lower. Such a result is unusual for a single strand study, in which the spectra remain very sharp and well resolved at all temperatures. However, oligonucleotides with a nearest neighbour GpC interaction in their sequence have been known to form multi-stranded aggregates (Jaskunas et al., 1968; Podder, 1971). Such an aggregation appears to be a likely explanation of the observed line broadening of

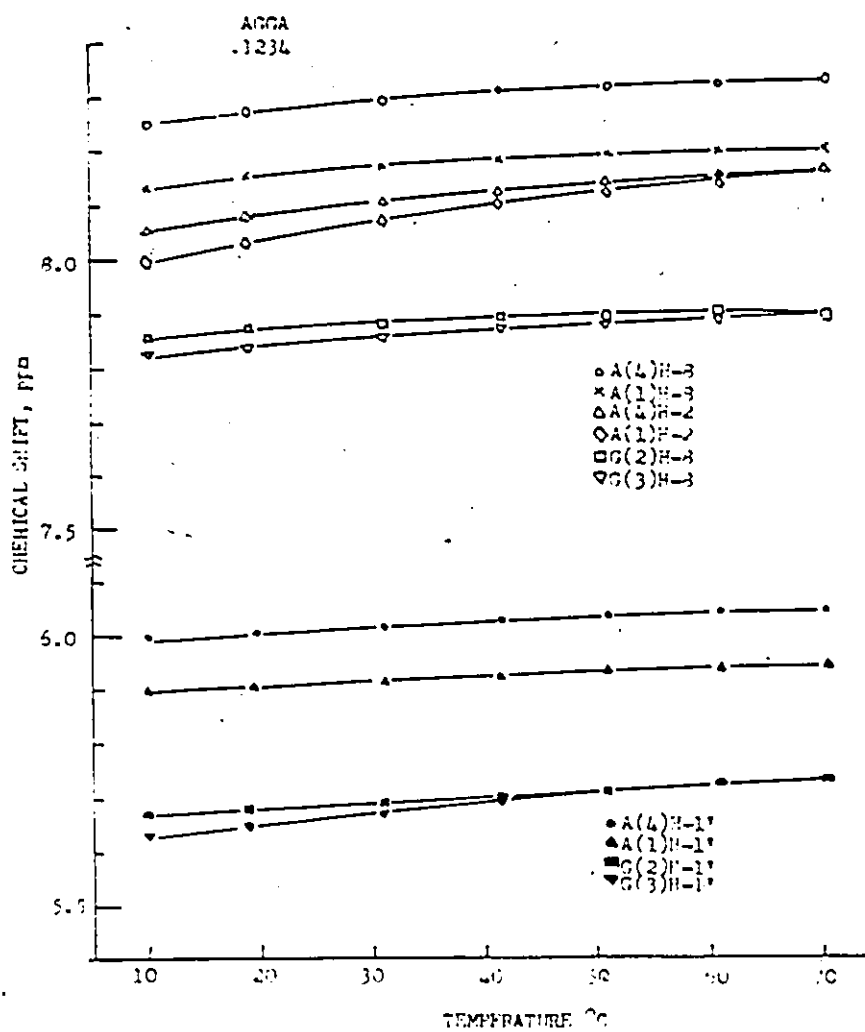


FIGURE 24. The variable temperature plots for the aromatic base and anomeric proton resonances of AGGA.

Page 104

the AGGA spectra. The temperature dependent behaviour of the anomeric coupling constants was also peculiar for this sequence. At 19°C, the four  $J_{1',2'}$  coupling constants are all well resolved: A(4)H-1',  $J = 3.8$  Hz; A(1)H-1',  $J = 5.0$  Hz; G(2)H-1',  $J = 5.0$  Hz; G(3)H-1',  $J = 4.4$  Hz.

The results of the variable temperature study of UCCU are presented in Figure 25. All of the H-6 resonances undergo a slight deshielding as the temperature is decreased, while the H-5 resonances experience a slight shielding. This difference reflects changes in the base overlap geometries upon stacking of the single strand. The spectra were well resolved at all temperatures, unlike the situation observed for AGGA. Although the H-1' resonances could not be unequivocally assigned, at 20°C and lower, these resonances were observed to be broad singlets, an indication of the existence of a high percentage of base stacked conformers (Altona, 1975). This relatively high degree of base stacking was not expected for the oligo-pyrimidine strand.

The two single strand samples were recovered, mixed, lyophilized to dryness, and dissolved in the nmr buffer used throughout these studies. A variable temperature study of the AGGA:UCCU mixtures (14 mM total strand concentration) yielded the results presented in Figure 26. The chemical shift vs. temperature plots are quite different from those for either of the single strands (Figures 24, 25). The co-operative melting profiles monitor the formation of the duplex between the two complementary strands. Because a self-complementary sequence could not be used to study this duplex, the spectra are

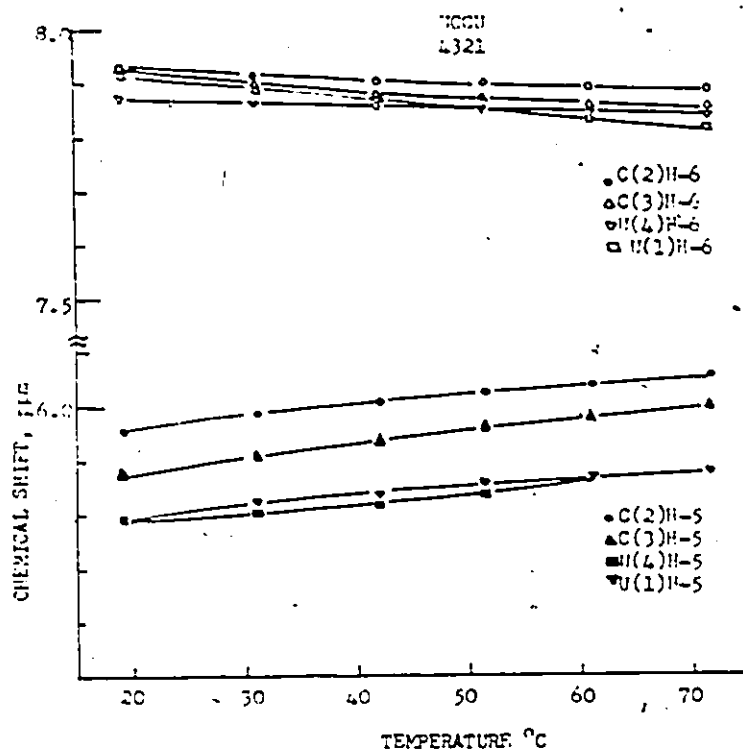


FIGURE 25. The variable temperature plots of the aromatic base proton resonances of UCCU.

Poor Copy



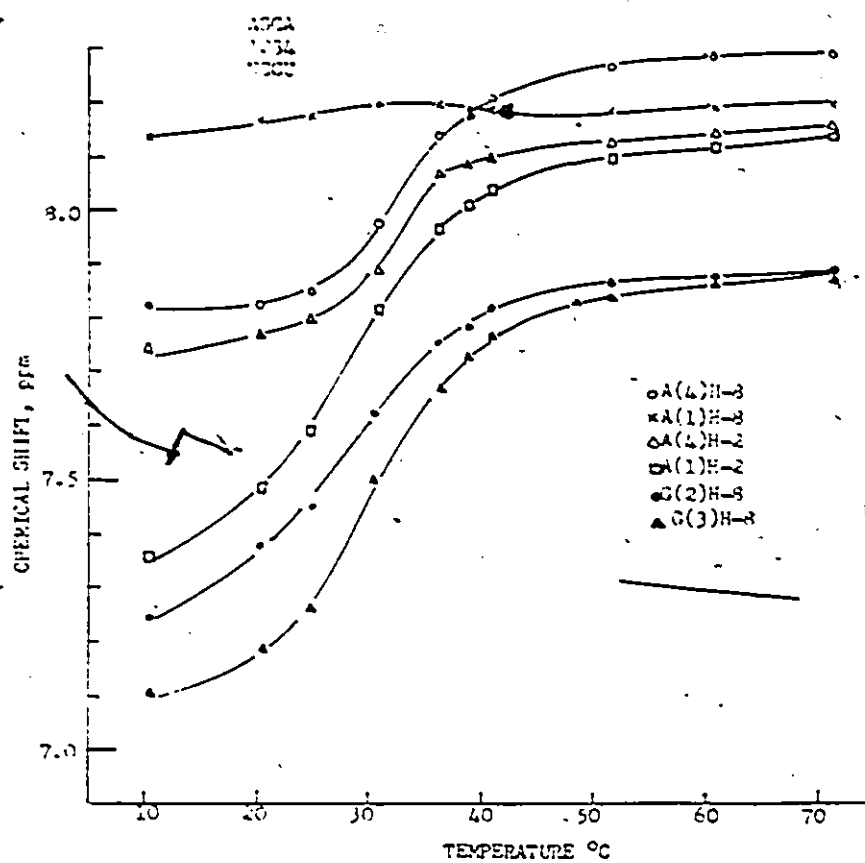


FIGURE 26. Results of the variable temperature study of the duplex formed by the complementary strands AGCA and UCCU.

Page 104

considerably more complex, and the pyrimidine signals could not be unambiguously monitored over the temperature range. An average  $T_m$  was determined using the melting transitions of A(4)H-8, A(4)H-2, A(1)H-2, G(2)H-8 and G(3)H-8. The AGGA:UCCU duplex has an average  $T_m$  of  $30.9 \pm 1.5^\circ\text{C}$  at 14 mM total strand concentration. Fraying of the terminal A-U base pairs of this duplex was not observed.

#### 4.4.2. Discussion

A comparison of the variable temperature plots for AGC (Figure 20) and AGGA (Figure 24) shows that the resonances in AGC are more shielded during the high-low temperature transition than the resonances of AGGA. The AGGA, purine rich tetramer, should show a stronger preference for a stacked conformation than AGC (Lomant & Fresco, 1975); but the results indicate a low stacking preference for AGGA. The large values of  $J_{1',2'}$  for the four nucleotide moieties of this sequence are consistent with a high percentage population of C2'-endo (Altona, 1975), a sugar pucker not found in a normal single stranded RNA stack conformer (Broyde & Hingerty, 1978). The AGGA single strand apparently does not stack to the degree expected for this sequence, a particularly surprising observation.

Short oligonucleotides with a ...GpG... sequence form multi-stranded aggregates in solution (Jaskunas et al., 1968; Podder, 1971). A consideration of the resonance line widths of AGGA spectra support an aggregation-type behaviour for this sequence. At temperatures below  $30^\circ\text{C}$ , a rapid broadening of the lines is observed. This line broadening could result from an increase in the correlation time for molecular motion of the oligomer as it forms a higher-molecular weight

aggregate in solution. X-ray crystallography studies of poly G indicated the formation of a triple helical structure with C2'-endo sugar puckers (Arnott et al., 1974). The AGGA oligonucleotide demonstrated a preference for the C2'-endo sugar pucker at low temperatures, and all results for the single strand study are consistent with an aggregation model for this sequence.

All of the H-6 resonances of UCCU undergo a slight deshielding as the temperature is decreased, while the H-5 resonances experience a slight shielding (Figure 25). This difference reflects changes in the base overlap geometry upon stacking: each H-5 proton is shielded by its 5'-neighbour, and each H-6 proton lies just outside the shielding cone of the ring current effect (Arter & Schmidt, 1976). The shielding and deshielding effects observed are quite small because the ring currents of pyrimidines are relatively weak, particularly uridine (Giessner-Prettre et al., 1976). At temperatures of 20°C and lower, the H-1' resonances were observed to be broad singlets ( $J_{1',2'} < 1.0$  Hz), consistent with the presence of a high percentage of base stacked conformers in solution (Altona, 1975).

A comparison of the variable temperature plot for the AGGA:UCCU mixture (Figure 26) with those of AGGA (Figure 24) and UCCU (Figure 25) demonstrates that a duplex is formed between the two complementary strands. The curves are sigmoidal in nature and an individual  $T_m$  could readily be measured for five of the resonances. The AGGA single strand apparently forms a double helix with UCCU rather than self-aggregating. The overlap of a large number of resonances in the aromatic region of the spectra prevented collection of the variable temperature data for

the pyrimidine H-6 signals.

An average  $T_m$  of  $30.9 \pm 1.5^\circ\text{C}$  was determined for the AGGA:UCCU duplex from the five sigmoidal curves of Figure 26. Table 13 contains the  $T_m$  values for CAUG, AGCU and AGGA:UCCU. A comparison of these values shows that the AGGA:UCCU duplex is more stable than CAUG and approximately as stable as AGCU. This result is slightly misleading. The  $T_m$  of a duplex is concentration dependent. In order to compare the  $T_m$  of a self-complementary duplex, with that of a duplex formed between two complementary strands, the total strand concentration of the latter duplex set should be double the concentration of the self-complementary strand (Uhlenbeck *et al.*, 1971). Unfortunately this condition could not be met for the AGGA:UCCU duplex, since there was not enough material to raise the concentration further. The results in Table 13 indicate however that at a comparable concentration, the AGGA:UCCU duplex would probably be more stable than the AGCU duplex.

The individual  $T_m$  data provide evidence that the terminal A·U base pairs of the AGGA:UCCU duplex do not fray. There is very close agreement between the  $T_m$  values for all of the resonances, similar to the situation for CAUG, but unlike the situation for AGCU, where fraying ends were observed. This short duplex is the first to be studied to have terminal A·U base pairs which do not fray. Fraying appears to be a sequence related phenomenon.

The AGGA:UCCU duplex plays a role in many of the mRNA-16 S rRNA interactions studied so far (Steitz & Jakes, 1975; Shine & Dalgarno, 1975). The results of this study suggest that this duplex was selected for this important RNA-RNA interaction because it is inherently a more

TABLE 13. Comparison of the AGGA:UCCU duplex with other short perfect double helices.

Duplex	$T_m(^{\circ}\text{C})$	Concentration (mM)
CAUG GUAC	$24.0 \pm 1$	9.2
AGCU UCCA	$33.6 \pm 4.4$	11.1
AGGA UCCU	$30.9 \pm 1.5$	14.0

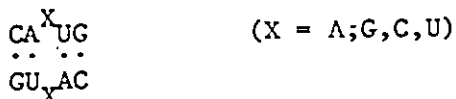
stable duplex than others which contain the same base composition. Certainly this double helix is the only short duplex known to have non-fraying terminal A·U base pairs, which might aid in reducing premature melting of the mRNA-16 S rRNA binding site.

#### 4.5. The Effects of Internal Non-Bonded Bases and a G·U Base Pair on the Stability of the CAUG Duplex

Optical studies on the formation of internal loops within short RNA double helices have been limited to loops containing cytidine (Uhlenbeck et al., 1971; Gralla & Crothers, 1973b). In both studies, the duplex region on each side of the loop contained four or more base pairs. A more complete review of model studies on the formation of all three kinds of loops is found in Section 1.2.2. Very little is known about the formation of internal loops within very short duplex regions.

Oligoribonucleotides of sequence CAXUG (X = A, G, C or U) were synthesized for use in studies on the formation of an internal loop within the CAUG duplex. Extrahelical conformations have been observed for mismatched bases in a pmr study of the complex formed between ethidium bromide and CUG (Lee & Tinoco Jr., 1978). The (CUG)<sub>2</sub> minihelix was formed only in the presence of ethidium bromide, and the complex represented a model for frameshift mutation. The purpose of the CAXUG study was to investigate free formation of internal loops.

Each of the CAXUG pentaribonucleotides can theoretically form a duplex containing a 'looped out' region:



Such a duplex would be an ideal model to probe the effect small

nonbonding regions have on the formation of adjacent base pairs. This series of pentamers also allows for the investigation of sequence related effects pertaining to loop closure. Pmr evidence of loop formation would be a deshielding of the nonbonding base resonances upon duplexing, similar to the effect on the GH-8 resonance for the 'looped-out' G of AGA<sub>n</sub> single stranded oligoribonucleotides (Shum, 1977).

The pmr studies on the CAXUG oligoribonucleotides indicated that these sequences did not duplex under the conditions in which CAUG had a T<sub>m</sub> of 24°C. Direct assignment of the X base resonances was made for each sequence, and there was no evidence of an extrahelical conformation for the mismatched bases.

The CAXUG pentamers can also be used to study the following duplex sets: CAAUG:CAUUG, CACUG:CAGUG and CAGUG:CAUUG. The first two duplex sets represent an extension of the CAUG duplex by the addition of an internal A·U and G·C base pair respectively. Although the effect of additional A·U and G·C base pairs has been studied previously (Martin et al., 1971; Uhlenbeck et al., 1971) it is important to use these duplex sets to calibrate the model system currently under investigation. Such a calibration allows the measurement of the relative magnitude of two different effects, for example the stabilizing effects of a dangling A vs. an A·U base pair. These duplexes can also serve as references for the CAGUG:CAUUG duplex. The self-complementary CAUAUG hexamer (a gift from Dr. T. Neilson) was studied to allow a comparison of the relative effect of two internal A·U base pairs (CAUAUG - CAUG) to that of two terminal A·U base pairs [AAGCUU (Borer et al., 1975) - AGCU].

Previous studies on the formation of G·U base pairs have been summarized in Section 1.2.4. The pmr study of the CAGUG:CAUUG duplex was the first to observe directly the formation of a G·U base pair within a regular Watson-Crick hydrogen bonded double helix. This study was also the first to measure the stability of a G·U base pair relative to the stability of the other base pairs in the same duplex. The results indicate that a G·U base pair may be a point of local instability within a double helical region of RNA. The importance of model studies of G·U base pairs, similar to the one reported here, has been mentioned in the literature (Early *et al.*, 1978). This study was conceived independently and completed several months before the work of Early *et al.* on poly d(GT) appeared in print.

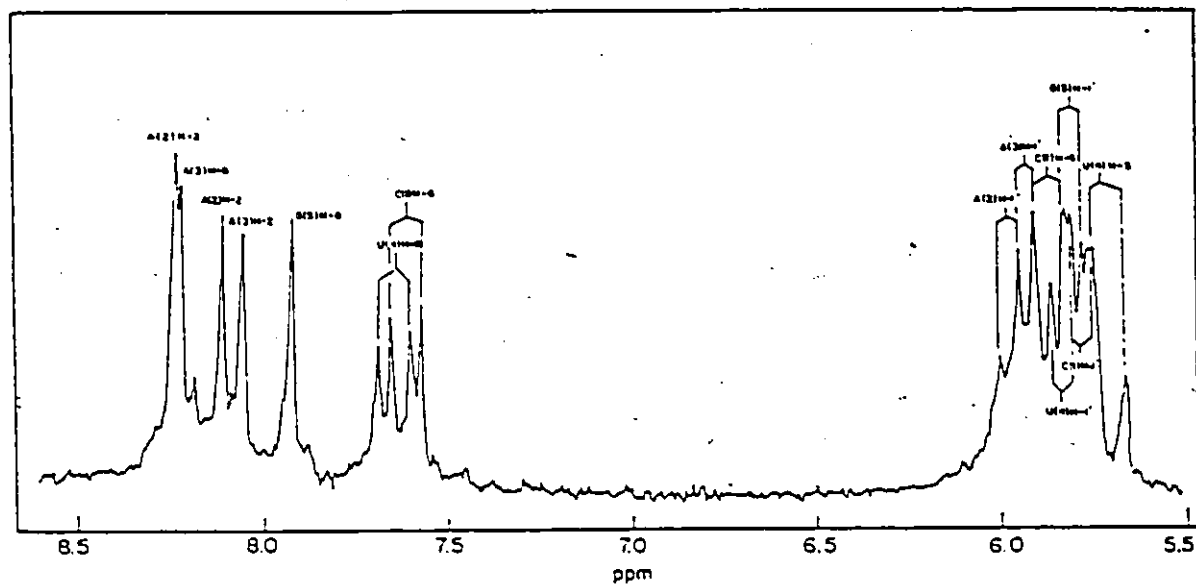
#### 4.5.1. Results

The low field nonexchangeable proton chemical shift assignments were determined by comparison to the CAUG data. Unfortunately, the resolution at 90 MHz was not sufficient to allow for the unequivocal assignment of the H-5 and H-1' resonances at 70°C for all of the pentamers, and the temperature dependent shift changes of these resonances could not be distinguished because of the high degree of overlap in this region of the spectra (Figure 27).

In CAAUG the presence of A(3) was indicated by two singlets at 8.255 ppm (H-8) and 8.079 ppm (H-2). The fact that adenine has a strong ring current shielding effect (Giessner-Prettre *et al.*, 1976) would account for the upfield shifts (cf. CAUG) experienced by the neighbouring aromatic protons (Table 14). The two GH-8 resonances in CAGUG were readily distinguished since the G(5)H-8 signal at 7.958 ppm



(a)



(b)

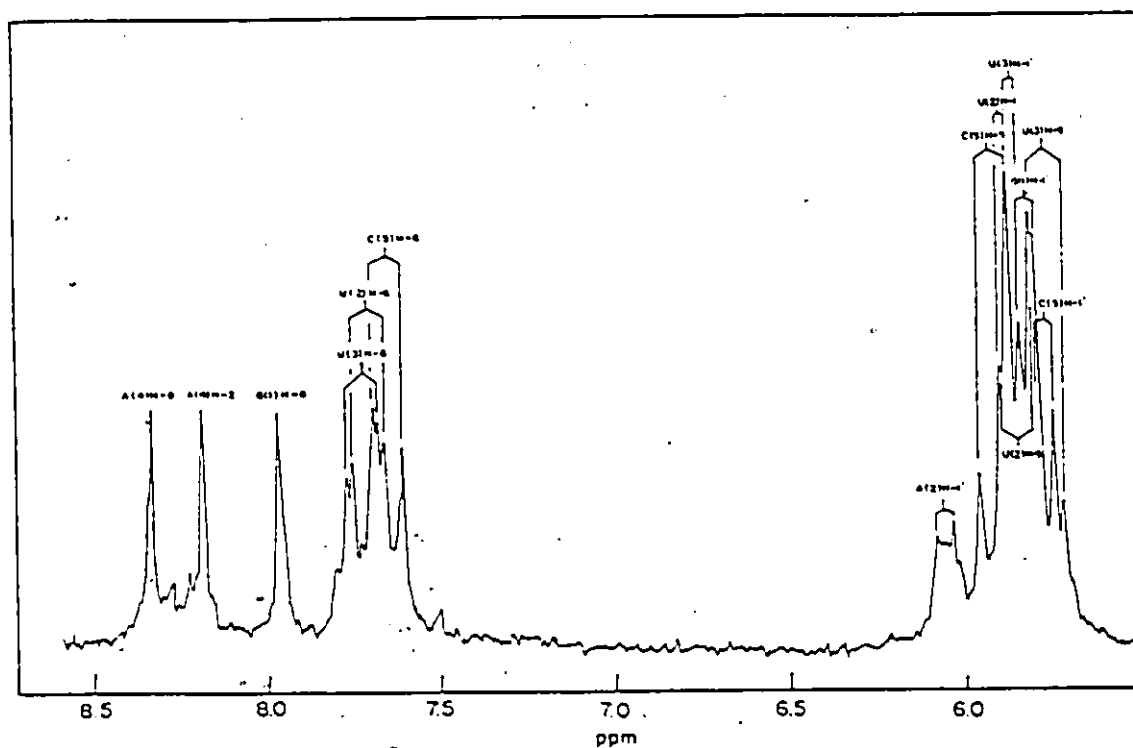


FIGURE 27. The 90 MHz PMR spectra of (a) CAAUG and (b) CAUUG in D<sub>2</sub>O at 70 °C.

Poor Copy

shows very little change from that in CAUG (Table 14). The appearance of a new doublet in the spectrum of CAUUG is attributed to the U(3)H-6 resonance at 7.716 ppm ( $J_{5,6} = 8.2$  Hz). In CACUG the U(4) and C(3)H-6 doublets are overlapped, and confirmation of the chemical shift assignments was provided by incremental analysis (Table 15). The conversion of CAC into CACU causes a shielding of the C(3)H-6 resonance to 7.711 ppm. This proton remains unaffected when G is added to CACU, although the UH-6 is shielded upfield by 0.060 ppm.

The chemical shifts of the CH-6 and GH-8 resonances in CAUAUG were assigned by comparison to CAUG (Table 14). The U(3) and U(5)H-6 doublets were equivalent at 70°C. Differentiation of the A(2) and A(4)H-2 signals was achieved by comparison with the data on the trinucleotides CAU and AUG (Table 5, Section 4.1.1.). In CAU the AH-2 signal is at lower field (8.250 ppm) than that in AUG (8.198 ppm) and this trend is retained in the hexanucleotide (Table 14). Although the same chemical shift trend is observed for the H-8 signals, further support for assigning the lower field H-8 resonance to A(2) was provided by its variable temperature behaviour, which was similar to the AH-8 in CAUG.

The results of the full variable temperature experiments on each pentaribonucleotide are shown in Figures 28 and 29. For most resonances, the chemical shift vs. temperature plot is characterized by a linear region at high temperatures and a gentle upfield curve at temperatures below 30°C. The plots do not display the distinct sigmoidal line shape characteristic of duplex formation (e.g. Figure 30). The oligoribonucleotides do not form a stable duplex of the type:

TABLE 14 - Chemical Shifts<sup>a</sup> of the Oligoribonucleotides in D<sub>2</sub>O<sup>b</sup> at 70°C

Proton	CAUG	CAAUG	CAGUG	CAUUG	CACUG	CAUAUG
C(1) H-6	7.662	7.623	7.641	7.659	7.672	7.643
A(2) H-8	8.346	8.268	8.294	8.359	8.352	8.326
A(2) H-2	8.196	8.134	8.160	8.206	8.183	8.154
A(3) H-8		8.255				
A(3) H-2		8.079				
G(3) H-8			7.906			
U(3) H-6				7.731		7.685
C(3) H-6					7.710	
U(4) H-6	7.692	7.654	7.703	7.716	7.710	
G(5) H-8	7.962	7.942	7.958	7.978	7.978	
A(4) H-8						8.310
A(4) H-2						8.131
U(5) H-6						7.685
G(6) H-8						7.945

<sup>a</sup> Chemical shifts are in ppm relative to DSS using tert-butyl alcohol-OD as an internal reference, and are accurate to  $\pm 0.005$  ppm.

<sup>b</sup> pH = 7.0; concentrations: CAUG, 9.2 mM; CAAUG, 11 mM; CAGUG, 3.2 mM, CAUUG, 11 mM; CACUG, 5.3 mM; CAUAUG, 9.9 mM.

TABLE 15 - Incremental Analysis<sup>a</sup> of CACUG in D<sub>2</sub>O<sup>b</sup> at 70°C

Proton	CA	CAC	CACU	CACUG
C(1) H-6	7.660	7.679	7.688	7.672
A(2) H-8	8.377	8.369	8.356	8.352
A(2) H-2	8.260	8.222	8.199	8.183
C(3) H-6		7.741	7.711	7.710
U(4) H-6			7.770	7.710
G(5) H-8				7.978

<sup>a</sup> Chemical shifts are in ppm relative to DSS using tert-butyl alcohol-OD as an internal reference and accurate to  $\pm 0.005$  ppm.

<sup>b</sup> pH = 7.0; concentrations: CA, 16 mM; CAC, 23 mM; CACU, 11 mM; CACUG, 5.3 mM.

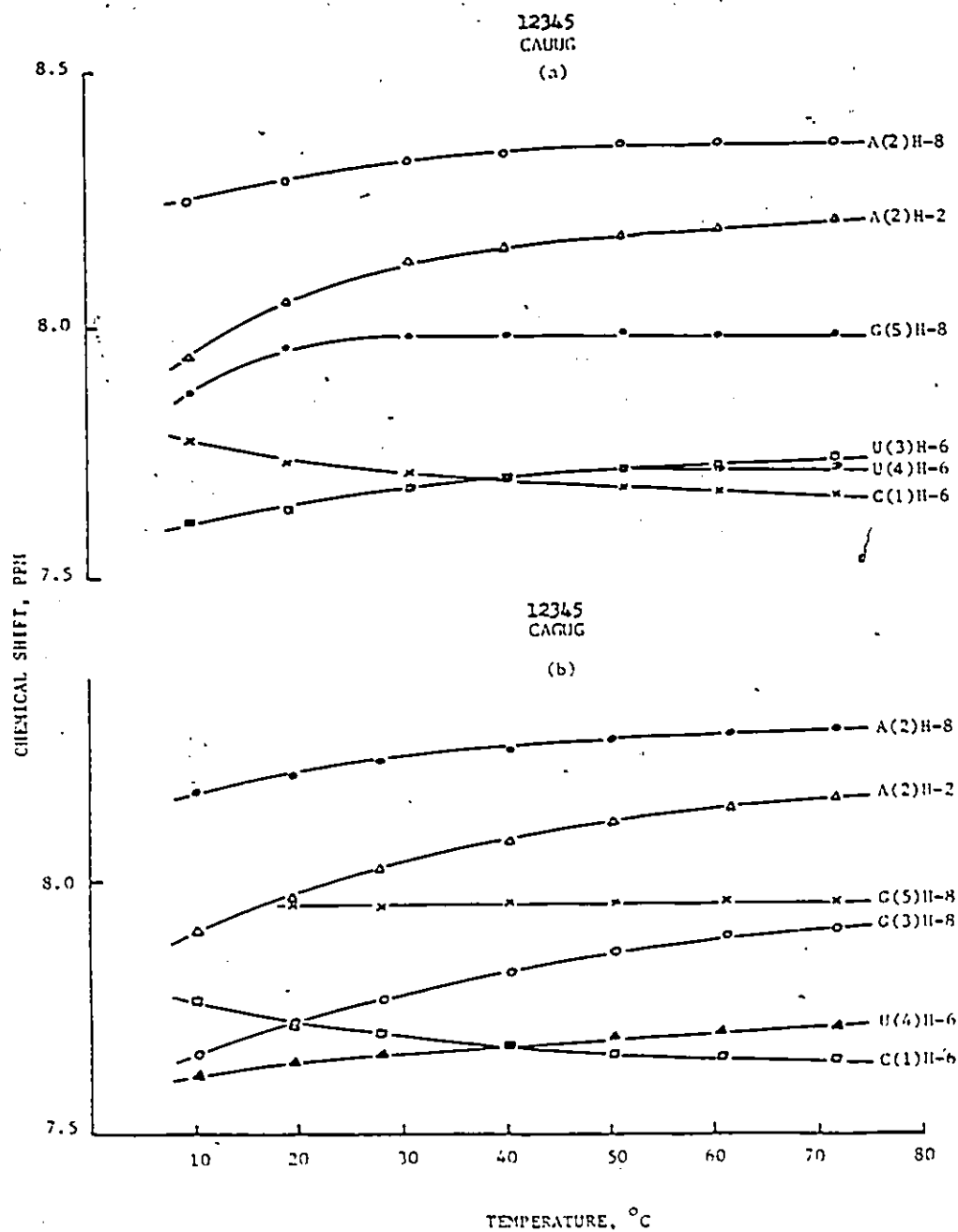


FIGURE 28. The variable temperature plots for the aromatic base proton resonances of (a) CAUUG and (b) CAGUG.

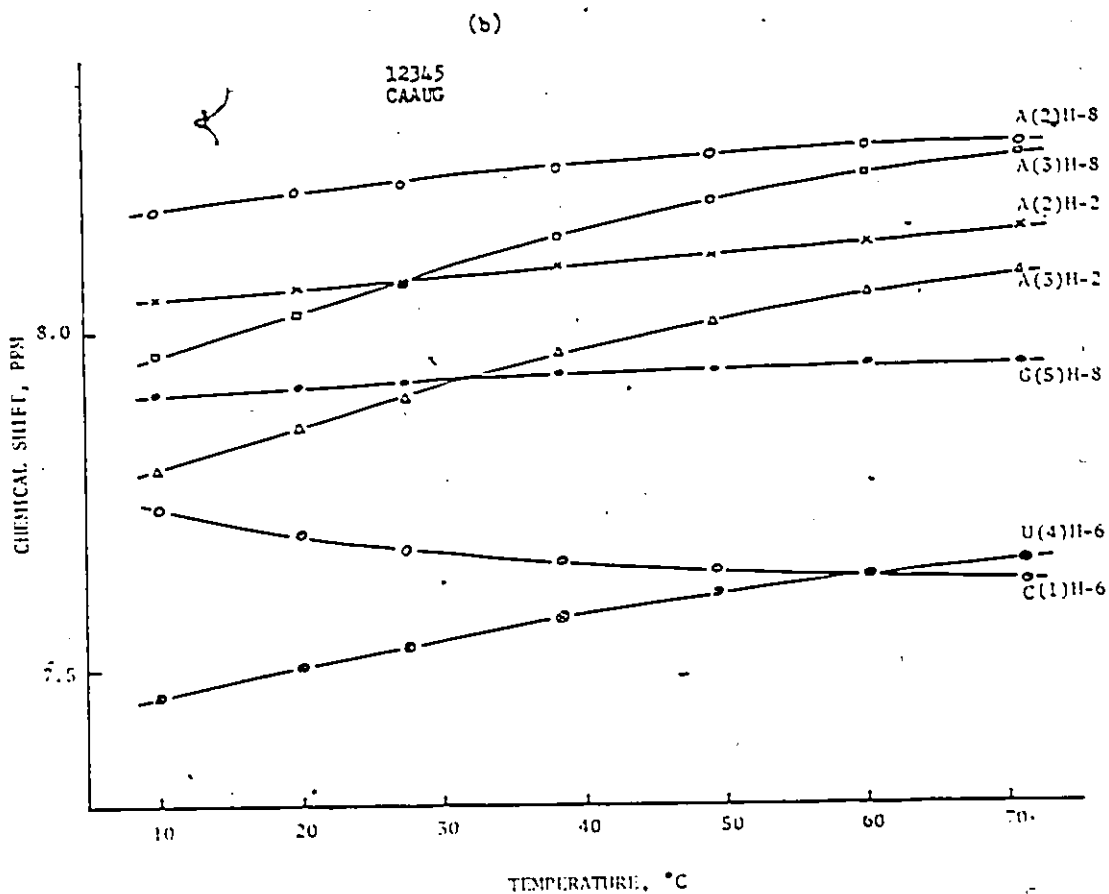
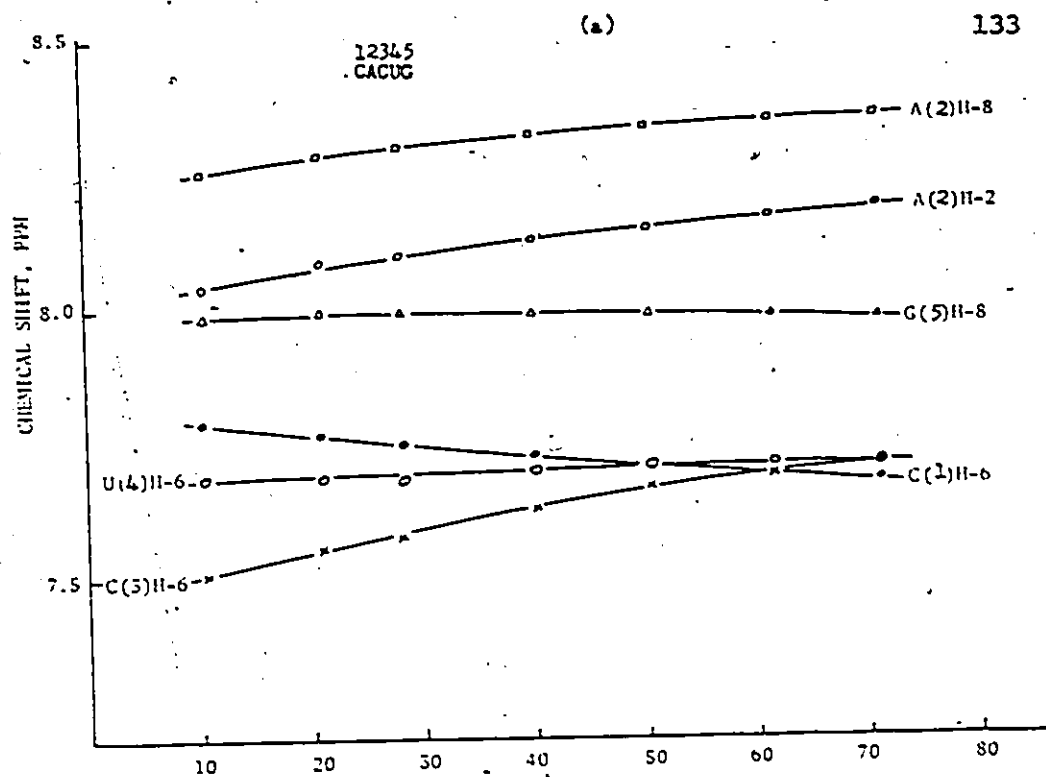
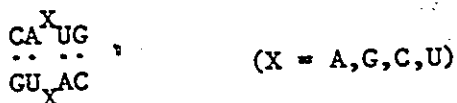


FIGURE 29. The variable temperature plots for the aromatic base proton resonances of (a) CACUG and (b) CAAUG.



under the conditions in which CAUG formed a duplex with a  $T_m$  of  $24 \pm 1^\circ\text{C}$ .

Before investigating the formation and stability of a G·U base pair, the model system under study was calibrated by examining the effect an additional Watson-Crick base pair has on the  $T_m$  of CAUG. Thus the effect of an additional A·U base pair was illustrated by the duplex formed by CAAUG:CAUUG. The results of the variable temperature experiment are shown in Figure 30. The low temperature chemical shift values were determined at 270 MHz since excessive broadening of the 90 MHz spectra at these temperatures made it difficult to assign individual resonances. The average  $T_m$  for the duplex, as determined from the eight purine aromatic resonances is  $28.5 \pm 2.1^\circ\text{C}$  which represents an increase of almost  $5^\circ\text{C}$  over the  $T_m$  of CAUG.

When the spectrum of the mixed complementary sequences CAAUG and CAUUG was recorded at  $70^\circ\text{C}$ , the purine base protons displayed chemical shifts which were essentially identical to those in the single stranded spectra at the same temperature. However, the pyrimidine H-6 signals could not be assigned directly because of the overlap of these resonances. This problem was solved by the technique of spectral subtraction and is illustrated in Figure 31 by the sequence CAAUG. Computer subtraction of the CAUUG  $70^\circ\text{C}$  spectrum from that of the mixture CAAUG:CAUUG produced the difference spectrum of CAAUG, allowing its chemical shifts to be determined by comparison to the original  $70^\circ\text{C}$  spectrum of CAAUG. Spectral subtraction of mixtures of complementary

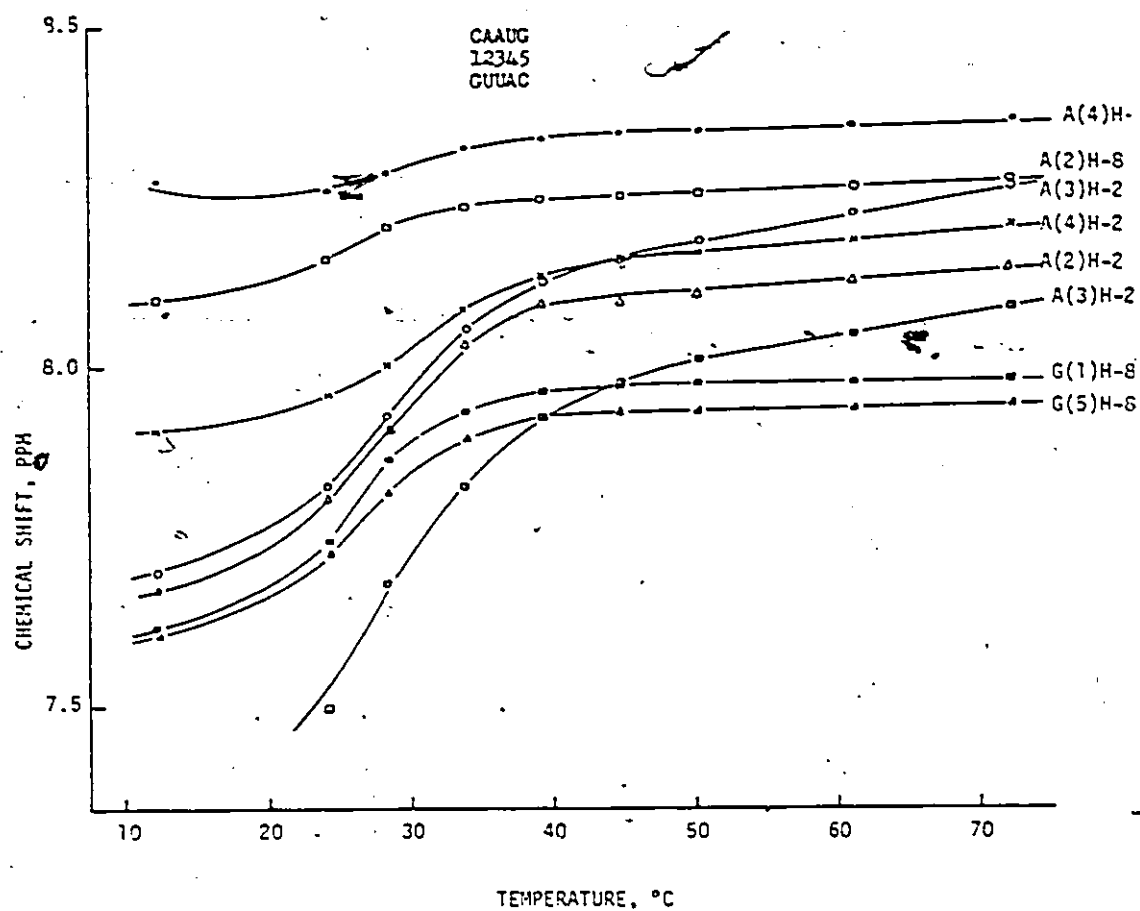


FIGURE 30. Results of the variable temperature study of the duplex formed by the complementary strands CAAUG and CAUUG.



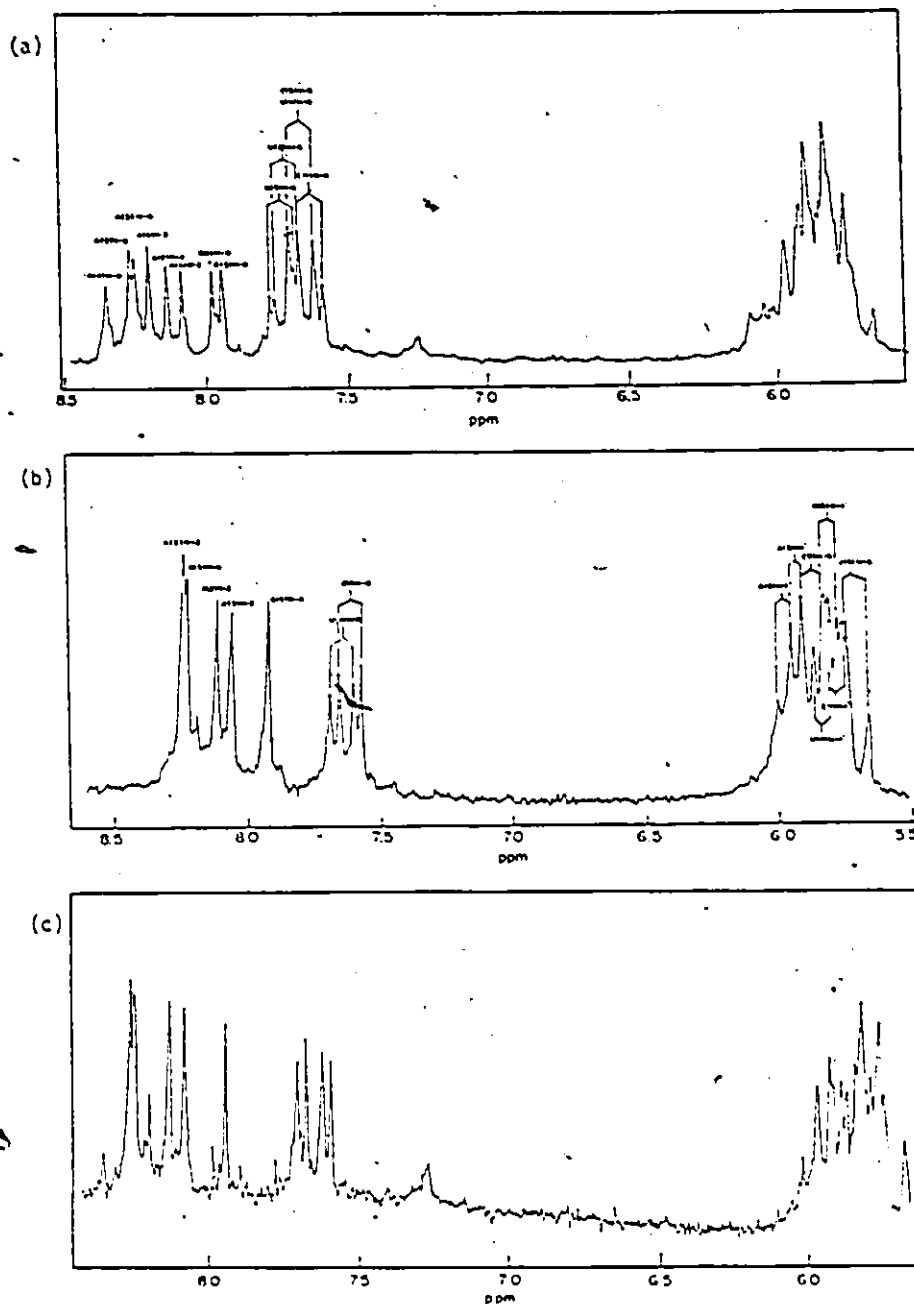


FIGURE 31. Assignment of the resonances of the 70 °C spectrum of the duplex set CAAUG:CAUUG (a) was carried out by the technique of spectral subtraction. The 70 °C spectrum of CAUUG (not shown) was subtracted from (a) to give spectrum (c). Comparison with the 70 °C spectrum of CAAUG (b) allowed assignment of these resonances in the duplex spectrum.

oligoribonucleotides is limited to only the high temperature spectra since both the intrastrand base stacking and interstrand base pairing are at a minimum under these conditions.

The self-complementary hexaribonucleotide CAUAUG was also studied as it represents an extension of the CAUG duplex by the incorporation of two internal A·U base pairs. The chemical shift vs. temperature plots for the purine aromatic resonances are shown in Figure 32. The average  $T_m$ , as determined from the sigmoidal lines, is  $41.5 \pm 1^\circ\text{C}$ , representing a  $17^\circ\text{C}$  increase over the  $T_m$  of CAUG at about the same concentration (Table 16). The difference in  $T_m$  between AAGCUU (Borer *et al.*, 1975) and AGCU is also  $17^\circ\text{C}$  (Table 16). Evidently the stability of two A·U base pairs does not depend upon whether they occupy internal or terminal positions. Evidence of fraying of the terminal G·C base pairs of the CAUAUG duplex was not observed, analogous to CAUG.

The effect of an additional internal G·C base pair on the duplex stability of CAUG was studied using the complementary set CAGUG:CACUG. The results of this experiment are shown in Figure 33. The average  $T_m$  of the duplex was  $38.4 \pm 0.6^\circ\text{C}$ , representing a large increase over the  $T_m$  of CAUG.

By mixing the two single strands CAGUG and CAUUG together, it was possible to observe, using pmr spectroscopy, the formation of a RNA double helix containing an internal G·U base pair surrounded by regular Watson-Crick A·U and G·C base pairs. The chemical shift vs. temperature plots are shown in Figure 34. The average  $T_m$  of the duplex formed is  $23.4 \pm 2.0^\circ\text{C}$ , and represents neither an increase or

TABLE 16. Melting Temperatures and Concentrations of the Base Paired Duplexes.

DUPLEX	T <sub>m</sub> (°C)	CONCENTRATION (mM)
CAUG ..... GUAC	24.0 ± 1.0	9.2
CAGUG ..... GUUAC	23.4 ± 2.0	18.0
CAAUG ..... GUUAC	28.5 ± 2.1	11.0
CAGUG ..... GUCAC	38.4 ± 0.6	3.2
CAUAUG ..... GUAUAC	41.5 ± 1.1	9.9
AGCU ..... UCCA	33.9 ± 4.4	11.0
AAGCUU <sup>a</sup> ..... UUCGAA	51.3 ± 6.0	10.0

<sup>a</sup>Data of Borer et al. (1975).

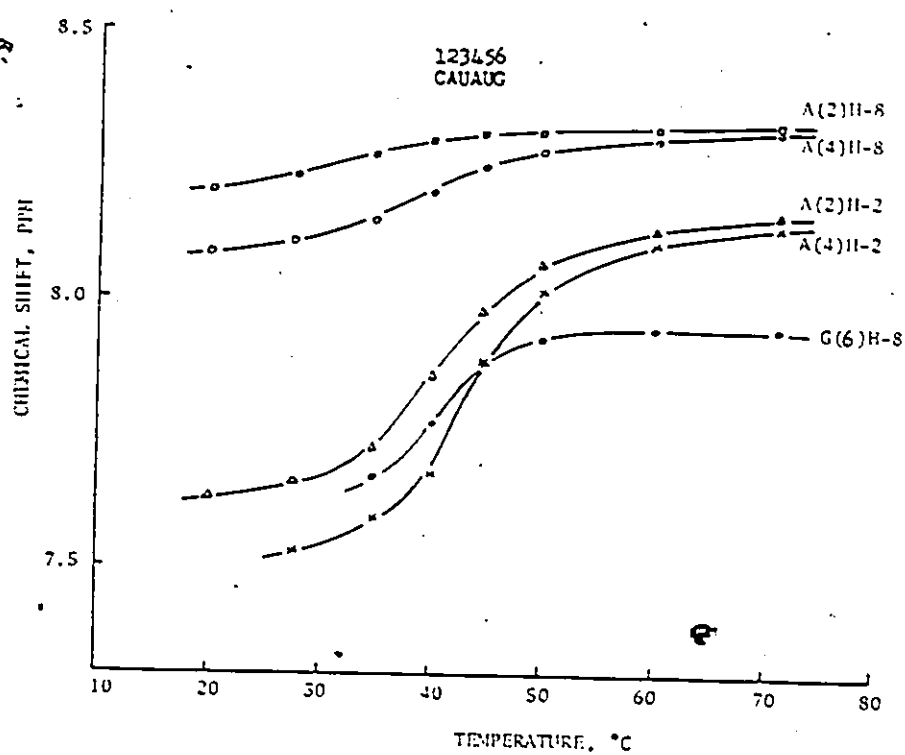


FIGURE 32. Results of the variable temperature study of the duplex formed by the self-complementary hexamer CAUAUG.

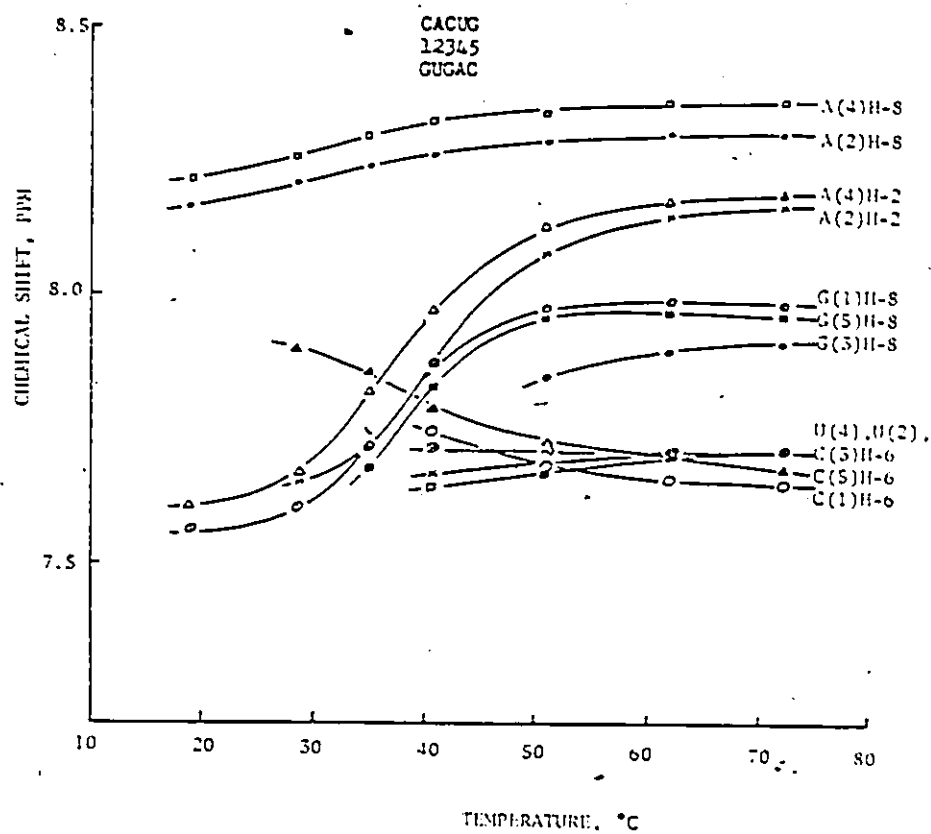


FIGURE 33. Results of the variable temperature study of the duplex formed by the complementary strands CACUG and GUGAC.

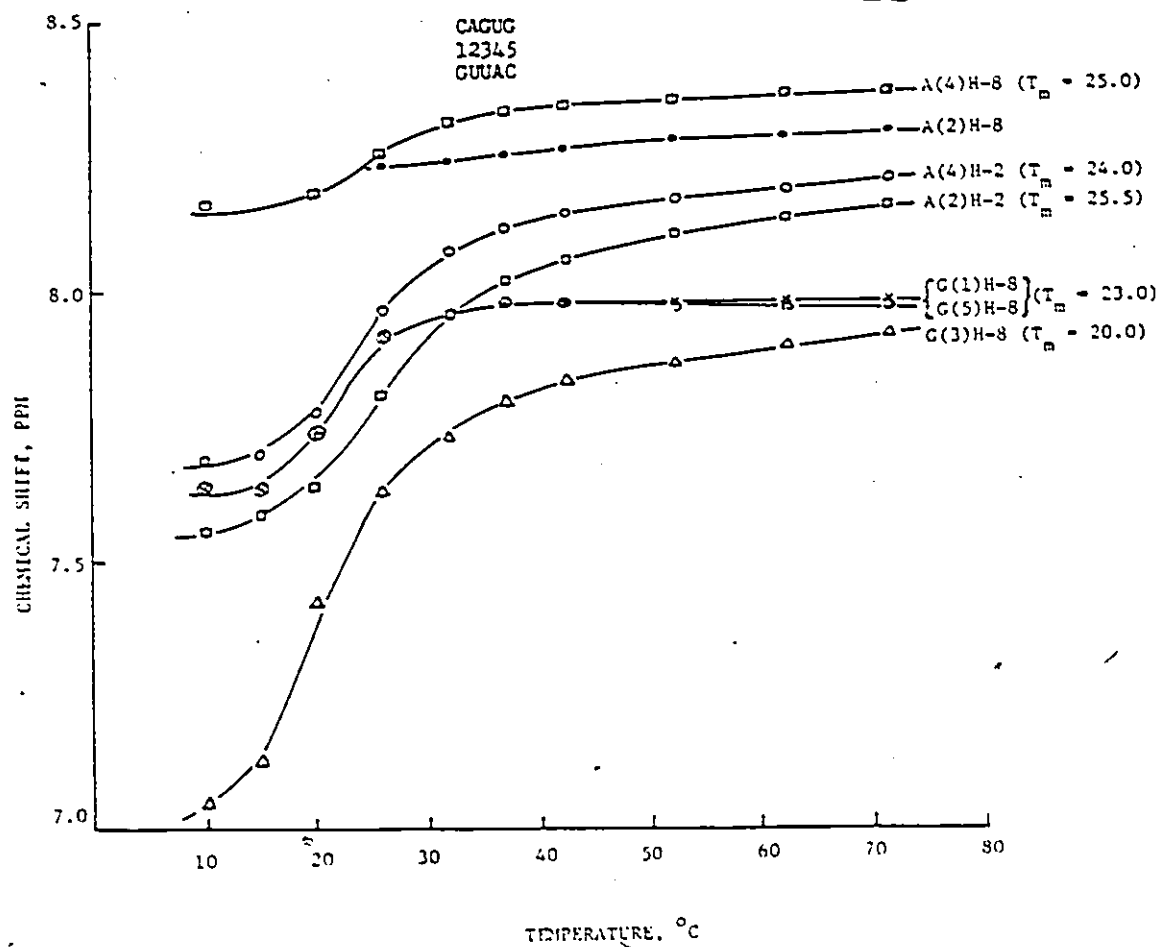


FIGURE 34. Results of the variable temperature study of the duplex formed by the complementary strands CAGUG and CAUG.

decrease in stability with respect to CAUG, the reference duplex (Table 16). In fact, the  $T_m$  values for the aromatic resonances of the A and G bases involved in Watson-Crick base pairs (Figure 34) are in very good agreement with those for the same base pairs in CAUG. By using NMR techniques, it was possible to monitor the G(3)H-8 separately, and the  $T_m$  of the guanine of the G·U base pair is 20°C, which is significantly lower than the  $T_m$ s of the other base pairs in this duplex. The G·U base pair appears to be a point of local instability within this duplex.

#### 4.5.2. Discussion

The studies reported in this section were designed to examine systematically, using a model system, the effect very small loops have on short neighbouring duplex regions and the relative stability of a G·U base pair within a RNA double helix containing regular A·U and G·C base pairs. The elegance of this study lies in the fact that a maximum number of experiments were generated from relatively small amounts of synthetic material. Seven experiments were conducted on diverse aspects of RNA secondary structure using only four pentaribonucleotides. These studies were possible because of the chemical synthesis methods described here, in combination with NMR techniques.

Each single pentaribonucleotide allowed for the study of duplex formation of a CAUG type double helix containing a nonbonding base opposition in the centre of the duplex. Previous studies on loops of this type employed longer  $A_nU_n$  type duplexes and optical techniques, which could give little conformational information (Uhlenbeck et al., 1971). It is now generally accepted that even a

U-U nonbonding opposition will take an extrahelical conformation, allowing for stacking of the adjacent duplex regions, and for most loops studied there is evidence of an extrahelical conformation (Lomant & Fresco, 1975). NMR evidence of looping out is a downfield shift with decreasing temperature as compared to the upfield shifts usually observed under these conditions (Shum, 1977). In Figures 28 and 29, the chemical shift vs. temperature plots for the aromatic base protons are not the sharp sigmoidal shapes associated with duplex formation, although the upfield shifts are somewhat larger than those usually associated with single stranded base stacking (Hughes *et al.*, 1978). Even more important however, is the observation that the plot for the aromatic proton of the middle base of each sequence shows an upfield trend with decreasing temperature. These results are interpreted to mean that the base which would be involved in a nonbonding base opposition is not taking an extrahelical conformation. Under the conditions where CAUG had a  $T_m$  of  $24 \pm 1^\circ\text{C}$ , these pentanucleotides do not form a stable duplex with a small loop, but prefer to maintain a single stranded stacked conformation. This result is particularly surprising in the case of CAUUG, since looping out of the U(3) would only break a U-U stacking interaction, and an A-U stacking interaction would be maintained in the duplex conformation (Lomant & Fresco, 1975). Perhaps in naturally occurring RNA molecules, short duplex regions separating very small loops (1-2 bases) may open up under suitable conditions to give larger loops which are stabilized by increased base stacking interactions, or by the formation of new base pairing interactions with other RNAs (Steitz & Jakes, 1975).



The effect of adding additional A·U or G·C base pairs was studied using the following duplexes: CAAUG:CAUUG, CAUAUG, and CAGUG:CACUG. Average  $T_m$  values were determined from the sigmoidal chemical shift vs. temperature curves of the aromatic base protons, and the results are summarized in Table 16. In each case, the duplex  $T_m$  is increased relative to CAUG, a result which is not surprising. The concentrations of the CAAUG:CAUUG and CAGUG:CACUG duplex sets are too low for a quantitative comparison to be made to the CAUG duplex. The total strand concentration of the CAGUG:CACUG duplex would need to be increased by a factor of 5 to allow a direct comparison to CAUG. Since other experiments required CAGUG, most notably a study of the CAGUG:CAUUG duplex, it was not possible to conduct a second study of the CAGUG:CACUG duplex at 18.4 mM total strand concentration. However, this concentration imbalance serves to emphasize the stabilizing effect of an additional G·C base pair: even at a 1 to 5 disadvantage in concentration, the  $T_m$  of CAGUG:CACUG is 14°C greater than the  $T_m$  of CAUG! At a comparable concentration the  $T_m$  of this duplex would certainly be greater than that of CAUAUG, illustrating the exaggerated importance of % G+C content to short double helices, an effect previously noted (Uhlenbeck et al., 1971).

The increase in  $T_m$  for CAAUG:CAUUG is comparable to the  $T_m$  for CAUGU (Table 9): a dangling U adjacent to the CAUG duplex stabilizes the duplex as much as an additional internal A·U base pair. The CAAUG:CAUUG, CAUAUG, and CAGUG:CACUG duplexes act as a second set of references in the model system under study, and allow for further comparisons to be made between various features of secondary structure.

Figure 34 shows the chemical shift vs. temperature plots for the CAGUG:CAUUG duplex set. The sharp sigmoidal nature of the curves is indicative of duplex formation and a  $T_m$  can be determined. It is important to remember that each single strand of this duplex has been studied in two different ways. As a single strand study, neither strand showed evidence of formation of a self-complementary duplex containing a loop, but rather maintained a stacked structure. Each strand has demonstrated its ability to form a Watson-Crick double helix with the appropriate complementary strand. On the basis of this evidence, it is apparent that the set CAGUG:CAUUG has formed a double helix which contains a G·U base pair. The stability of this base pair can be measured from the sigmoidal curve for the G(3)H-8 (Figure 34). This study is the first on the formation and relative stability of a G·U base pair within a double helix containing regular Watson-Crick A·U and G·C base pairs.

The results from the duplex studies reported here are summarized in Table 16. Each single strand in the CAGUG:CAUUG duplex has the same stacking interactions as in the corresponding CAGUG:CACUG and CAAUG:CAUUG duplexes. The G·U pair most likely being formed is the wobble pair with two hydrogen bonds (Rordorf et al., 1976; Early et al., 1978). A comparison of the duplexes in Table 16 shows that the G·U pair is less stable than the A·U pair, and there is no increase of stability throughout the duplex, as there is for the CAAUG:CAUUG and CAGUG:CACUG duplexes when compared to CAUG. In fact, the  $T_m$  values for the A·U and G·C base pairs of the CAGUG:CAUUG duplex are identical to those found for CAUG (Section 4.1.) and the G·U base pair is

significantly less stable ( $T_m = 20.0^\circ\text{C}$ ) than the rest of the duplex ( $T_m = 24 \pm 1^\circ\text{C}$ ).

Several mechanisms could account for this decreased stability. For hydrogen bonding to occur in a wobble G·U base pair a shift is required in the glycosyl torsion angles from the angles usually associated with Watson-Crick base pairing (Mizuno & Sundaralingam, 1978). This displacement of the bases can be easily accommodated with little perturbation of the backbone conformation (Mizuno & Sundaralingam, 1978). However, although stacking interactions are only slightly changed, this change might lead to an overall decrease in stability throughout the duplex. An alternate explanation arises from the observation that a G·U pair in the wobble position of a codon-anticodon interaction is as stable as an A·U pair in the same position (Uhlenbeck *et al.*, 1970). It is possible that the G·U pair cannot adopt its preferred conformation within a duplex and is therefore a region of local instability, perhaps affecting the nucleation and closing of the entire duplex. In all probability both situations exist in solution and contribute to the observed effect.

#### 4.6. A Model Study on the Formation of a "Staggered" Duplex

The self-complementary oligonucleotides CCGG (Arter *et al.*, 1974) d(CCGG) and d(GGCC) (Patel, 1977) form short, perfect double helices, as shown by pmr studies. The helix-coil transitions of the sequence isomers d(CGCG) (Patel, 1976b) and d(GCGC) (Patel, 1979) have also been investigated by NMR. These sequences can form either a perfect duplex or a "staggered" duplex (Figure 35), a fact not considered in earlier studies (Patel, 1976b, 1979).

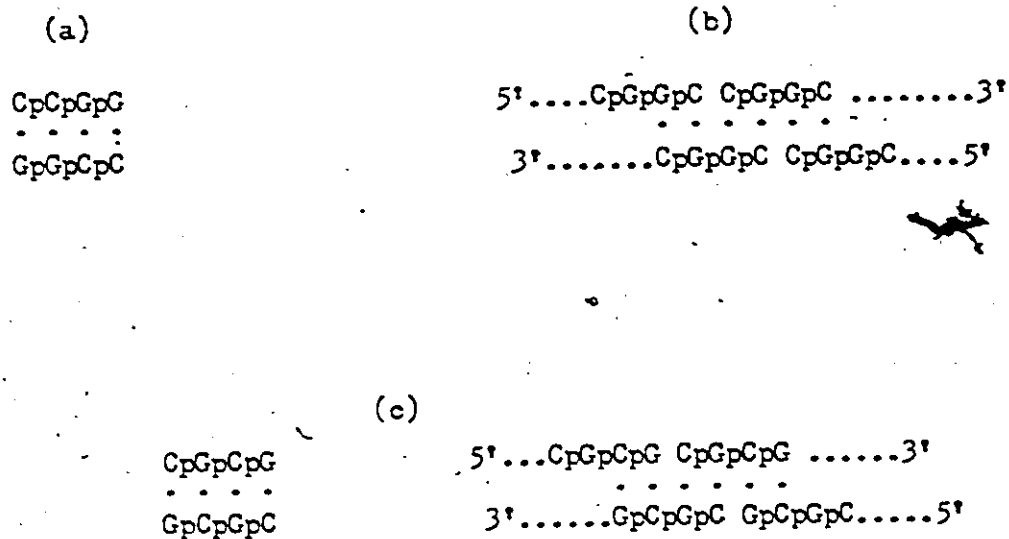


FIGURE 35. Illustration of the double helices formed by (a) CCGG - perfect helix; (b) CGGC - staggered helix; (c) CGCG - perfect helix and staggered helix.

A study of the staggered duplex is of interest for several reasons. The propagation of a staggered helix is an intermolecular event, whereas intramolecular propagation occurs during the formation of a perfect duplex (Figure 36). An investigation of staggered duplexing may indicate how important intramolecular propagation is to the stability of a perfect double helix. The staggered duplex may also act as a model of "butt-end" stacking of perfect duplexes in concentrated solution. The shielding data derived from an experiment on staggered duplex formation possibly could be used to estimate the amount of "butt-end" stacking in experiments on perfect duplexes.

The formation of a staggered duplex was studied using the three sequences shown in Figure 35. The CCGG oligoribonucleotide acted as a control for the formation of a perfect helix. This sequence has been studied previously by pmr (Arter et al., 1974). The present work represents an extension of this earlier work. The use of Fourier transform NMR techniques greatly improved the signal to noise ratio of the spectra, and all of the non-exchangeable aromatic base and anomeric ribose protons have now been identified in the CCGG spectra. Comparison of the 70°C spectrum of CCGG with those of sequence precursors and sequence isomers allowed the complete assignment of the CCGG resonances. As a result of this study, more information is available about the perfect double helix formed by CCGG.

The CCGC sequence can only form a staggered duplex (Figure 34) and therefore is a control for the formation of such a duplex. Complete analysis of the pmr spectra was possible by using data obtained for sequence precursors. The results of a variable temperature study

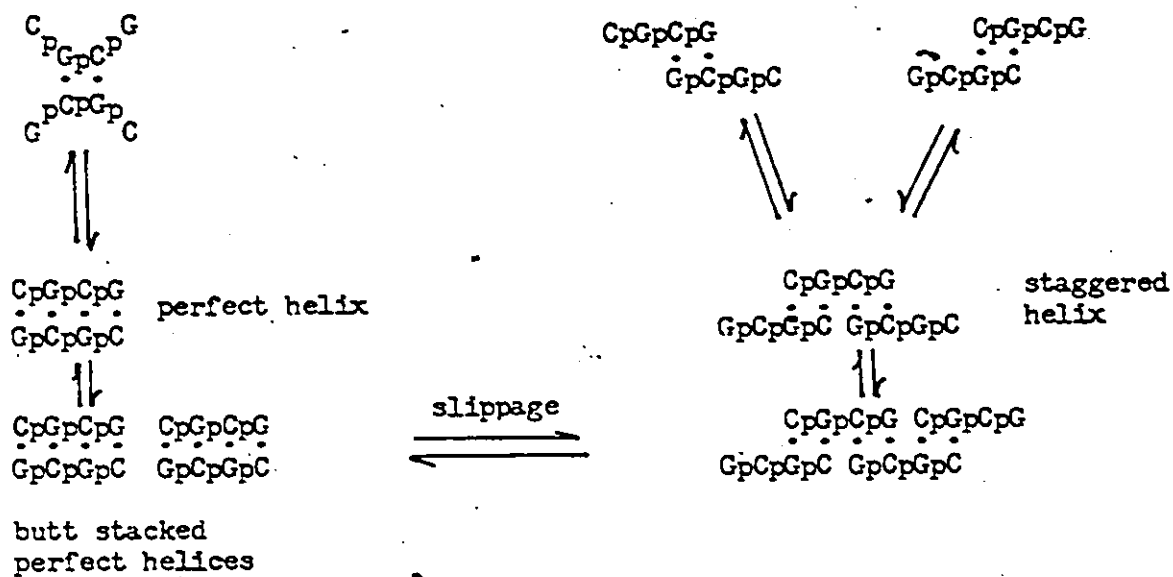


FIGURE 36. Various base paired structures possible for CGCG.

indicated that CGGC formed a staggered duplex in a reversible manner. The sequence CGGA was studied in an attempt to investigate the nucleation of a staggered duplex. Aggregation of this oligomer was observed, similar to other ...GpG... type sequences that are not self-complementary (Jaskunas *et al.*, 1968; Podder, 1971)..

The final sequence of this series is CGCG, analogous to the deoxy sequence studied by Patel (1976b). Comparison of the data for this sequence with those for CCGG and CGGC should allow an elucidation of which duplex is formed by CGCG. The results of the CGCG study were complex, and it is possible that a mixture of the two duplex types co-exist in solution.

#### 4.6.1. Results

Incremental analysis is a valid assignment technique only when extension of a sequence does not result in large shielding changes in the resonances previously assigned (Borer *et al.*, 1975). Such a condition was not met for the 70°C spectra of many of the sequences studied (Table 17) and a comparative analysis was required in order to assign the chemical shifts of some protons.

The assignment of the resonances of CG was straight forward and agreed well with published data (Ezra *et al.*, 1977). Extension of this sequence to CGC lead to several changes in the spectrum. The C(3) resonances were assigned to doublets at 7.787 ppm (CH-6), 5.930 ppm (CH-5, equivalent to C(1)H-5) and 5.911 ppm (CH-1'). The G(2)H-1' was shielded 0.03 ppm from its position in the CG spectrum. Comparison of the CGG spectrum to CG revealed the following differences:

TABLE 17. The Chemical Shift<sup>a</sup> Assignments of CCGG and its Sequence Isomers and Precursors in D<sub>2</sub>O<sup>b</sup> at 70 °C.

RES.	CG	CGG	CGC	CCG	CCGG	CGGC	CGCG	CGGA
C(1)H-6	7.686	7.662	7.693	7.750	7.747	7.666	7.701	7.652
C(2)H-6				7.750	7.711			
C(3)H-6			7.787				7.701	
C(4)H-6						7.767		
G(2)H-8	7.999	7.968	7.992			7.955	7.968	7.913
G(3)H-8		7.942		7.984	7.916	7.935		7.896
G(4)H-8					7.948		7.968	
A(4)H-8								8.316
A(4)H-2								8.170
C(1)H-5	5.978	5.946	5.930	5.976	5.975	5.940	5.923	5.927
C(2)H-5				5.954	5.940			
C(3)H-5			5.930				5.846	
C(4)H-5						5.912		
C(1)H-1'	5.817	5.803	5.805	5.819	5.817	5.806		5.801
C(2)H-1'				5.884	5.847			
C(3)H-1'			5.911					
C(4)H-1'						5.891		
G(2)H-1'	5.891	5.811	5.865			5.792		5.751
G(3)H-1'		5.873		5.884	5.798	5.858		5.751
G(4)H-1'					5.868			
A(4)H-1'								6.052

<sup>a</sup>Chemical shifts are in ppm relative to DSS using tert-butyl alcohol-OD as an internal reference, and are accurate to  $\pm 0.005$  ppm.

<sup>b</sup>pD=7.0; Concentrations: CG, 17.0 mM; CGG, 14.6 mM; CGC, 12.7 mM; CCG, 12.7 mM; CCGG, 9.2 mM; CGGC, 9.2 mM; CGCG, 9.2 mM; CGGA, 17.2 mM.



- i) the C(1)H-6 was shielded 0.025 ppm
- ii) the G(2)H-8 was shielded 0.03 ppm to 7.968 ppm
- iii) the G(3)H-8 was then assigned to the signal at 7.942 ppm
- iv) the C(1)H-5 was shielded 0.03 ppm
- v) the two GH-1' signals were assigned to the doublets at 5.811 ppm and 5.873 ppm. The shielding resulting from extension of the CG sequence suggested the following assignments: CG, 5.891 ppm; CGC, 5.865 ppm and CCG, 5.811 ppm. The G(3) should provide a stronger shielding of the G(2)H-1' than the C(3) (Arter & Schmidt, 1976). The G(3)H-1' was assigned to the doublet at 5.873 ppm.

The chemical shift assignments of CCG were made with the aid of published data for CC (Lee et al., 1976). The G(3) resonances were readily assigned: 7.984 ppm (GH-8) and 5.884 ppm (GH-1'). The CH-6 resonances were equivalent at 70°C, but the CH-5 resonances were not and the more shielded signal was assigned to the C(2) base. Assignment of the CH-1' resonances was resolved by comparison to the CCGG data: one of the H-1' resonances has the same chemical shift in both spectra, and this resonance (5.819 ppm) is assigned to C(1)H-1'. The other H-1' resonance is more shielded in CCGG than in CCG, which supports the assignment of this resonance (5.884 ppm) to C(2)H-1'.

~~Comparison~~ of the data of CCGG with that of CCG led to the assignment of G(4)H-8 to the signal at 7.948 ppm (Table 17). Unlike the situation in CCG, the CH-6s of CCGG were not equivalent, and the doublet to higher field was assigned to G(2)H-6 (7.711 ppm), since this base should be shielded more than the C(1). Assignment of the other resonances of CCGG was readily made by appropriate comparison to the

CGG or CCG data (Table 17).

The assignment of the CCGC resonances was made in essentially an incremental fashion by comparison to the data for CGG. Some shielding of the resonances was observed, but this did not complicate the analysis. The situation for CGGA was essentially the same. The data for these two sequences are summarized in Table 17.

The CH-6 proton resonances of CCGG at 70°C were equivalent, as were the two GH-8 resonances. At lower temperatures, the two H-8 resonances separated, and the signal to higher field was assigned to the G(2)H-8 resonance, since this proton will be shielded by two nearest neighbour bases. The two CH-6 resonances also became non-equivalent at lower temperatures, and the resonance which was deshielded as temperature decreased was assigned to the C(1)H-6. This deshielding trend is analogous to similar trends observed for the CH-6 of the CAUG duplex (Section 4.1.). The CH-5 resonances were not equivalent at 70°C: the resonance at 5.923 ppm was assigned to C(1)H-5 by comparison to CGC (5.930 ppm) and the more shielded resonance, 5.846 ppm was assigned to C(3)H-5. A high degree of overlap in the anomeric region of the spectrum prevented the assignment of the H-1' resonances.

The results of the study of the helix-coil transition of CCGG (9.2 mM) are presented in Figure 37. This data extends the previous study of CCGG (Arter *et al.*, 1974) by the inclusion of the pyrimidine H-5 and all four H-1' plots. A more rigorous assignment of the resonances was possible in this study. Each of the four anomeric proton doublets were clearly distinguishable at all temperatures, and a reduction of the H-1', H-2' coupling constants to less than 1.0 Hz

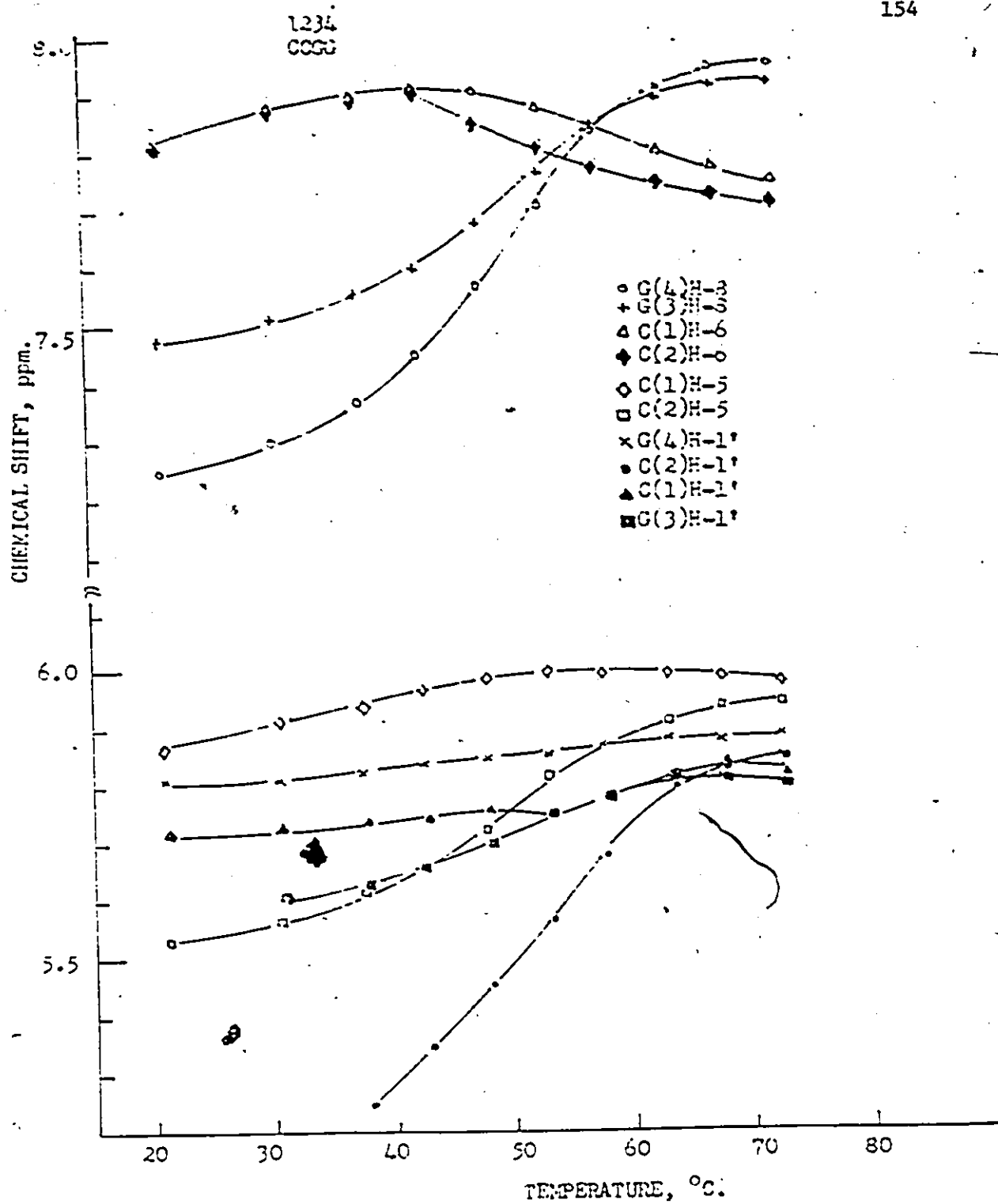


FIGURE 3D. The variable temperature plots of the aromatic base and anomeric proton resonances of CCGG.

occurred as the temperature was decreased to 48°C. Broadening of the line widths became noticeable at temperatures of 38°C and lower. The average  $T_m$  determined from the sigmoidal curves of C(2)H-5, G(4)H-8 and G(3)H-8 was  $50.1 \pm 0.8^\circ\text{C}$ . This value is in good agreement with the  $T_m$  of  $51 \pm 2^\circ\text{C}$  determined for CCGG at a concentration of 13 mM (Arter et al., 1974).

The results of a variable temperature study of CGGC are shown in Figure 38. Several of the plots are sigmoidal in nature and the average  $T_m$  for the duplex formed is  $33.5 \pm 0.5^\circ\text{C}$ . As the temperature was decreased past the  $T_m$ , excessive broadening of the lines was observed (Figure 39). This increase in the correlation time for molecular motion could be related to an increase in molecular weight of the duplex structure. A similar temperature dependent increase in line width was observed for CGGA, but the variable temperature plots do not indicate a helix-coil transition for this sequence (Figure 40). The CGGC duplex formed in a reversible manner: repeating the melting experiment several times with the same sample did not alter the shape of the individual plots or the  $T_m$  values.

Figure 41 presents the results of the variable temperature study of CGCG. The sigmoidal nature of the curves indicated that double helix formation had occurred, but a wide variation of the individual  $T_m$  values was observed. The average  $T_m$  value determined from the GH-8 plots is  $41 \pm 2^\circ\text{C}$ , while the average  $T_m$  calculated from the C(3)H-6 and H-5 plots is  $47.4 \pm 0.6^\circ\text{C}$ . The results do not clearly indicate whether CGCG forms a perfect or staggered duplex in solution.

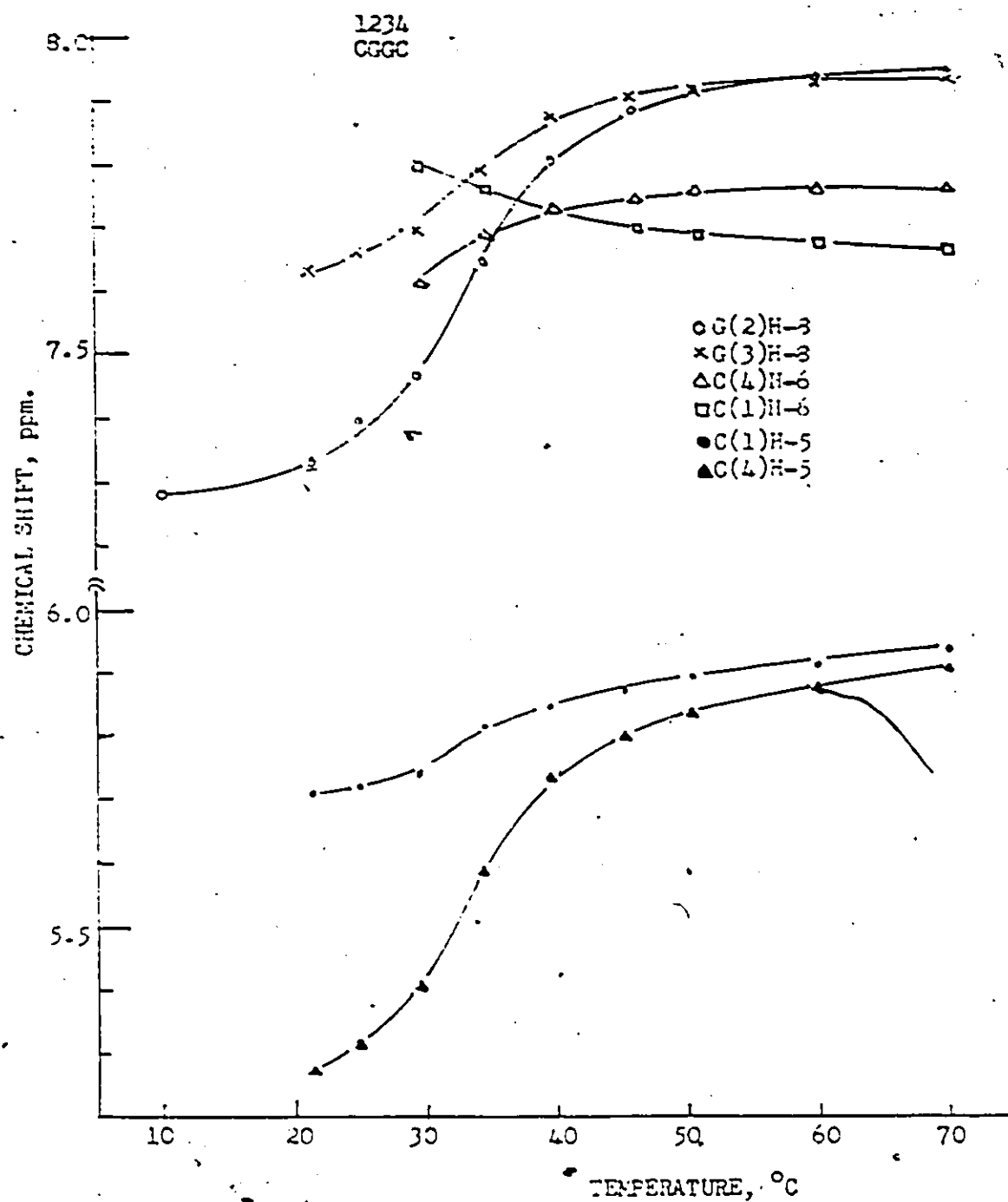


FIGURE 38. The variable temperature plots of the aromatic base proton resonances of CGGC.

Poor Copy

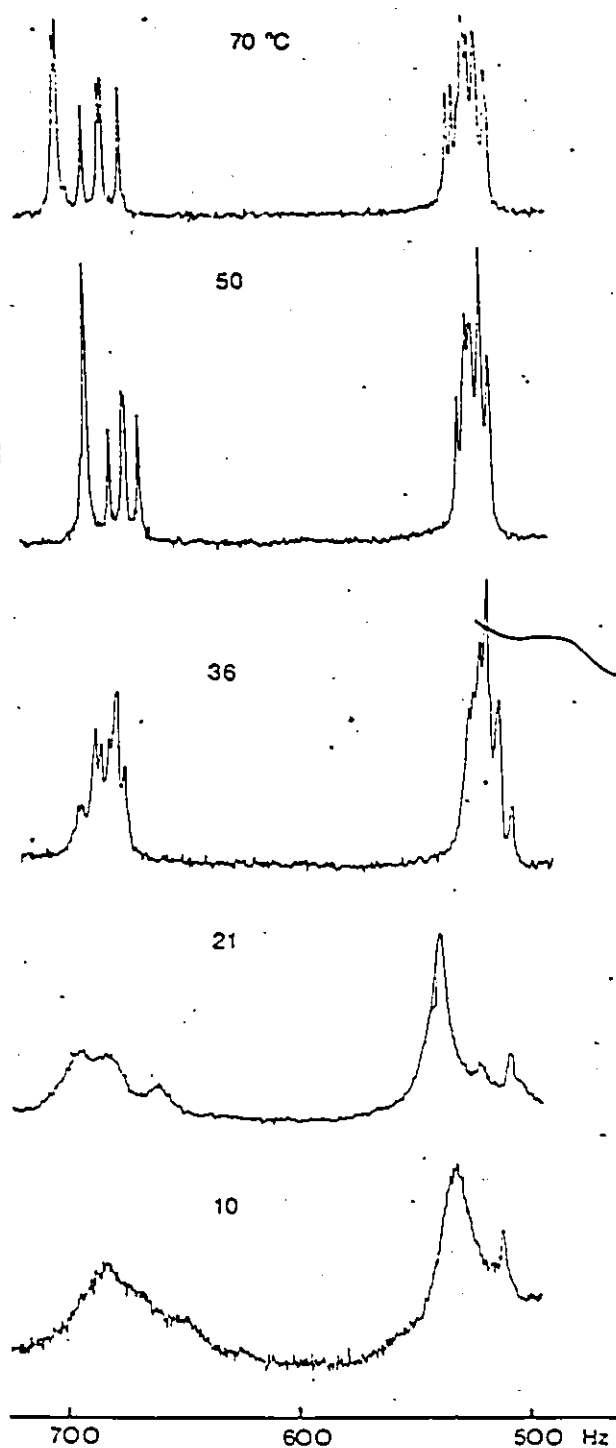


FIGURE 39. The temperature dependent line broadening of the CGGC spectra.

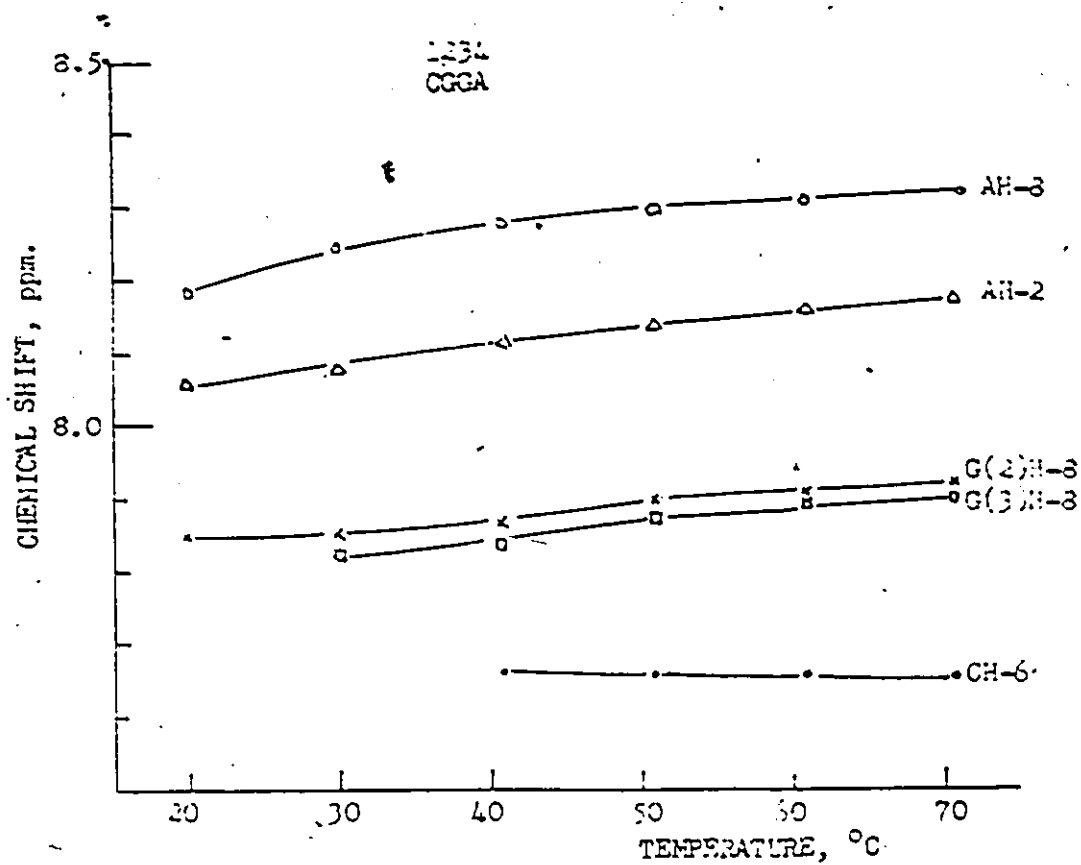


FIGURE 40. The variable temperature plots of the aromatic base proton resonances of CGGA.

Poor Copy

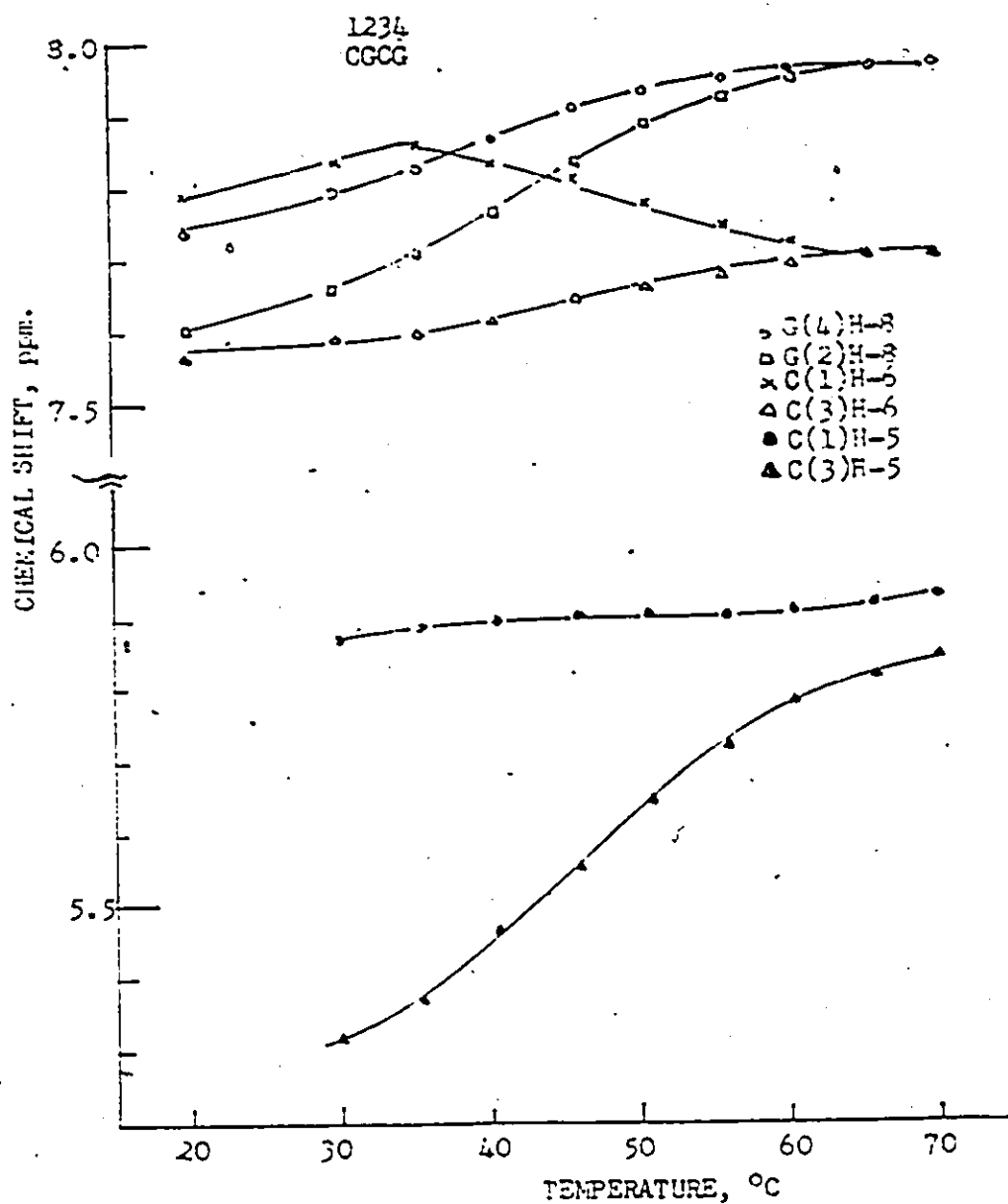


FIGURE 41. The variable temperature plots of the aromatic base proton resonances of CGCG.

O. P. A.



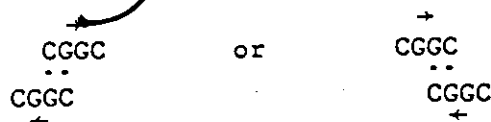
#### 4.6.2. Discussion

The sequence CCGG can form only a perfect double helix. The results of Figure 37 indicate that the CCGG duplex has a  $T_m$  of  $50.1 \pm 0.8^\circ\text{C}$  at 9.2 mM strand concentration. Preliminary pmr studies of this sequence yielded a  $T_m$  of  $51 \pm 2^\circ\text{C}$  at 13.1 mM concentration (Arter et al., 1974), in good agreement with the present results. Even though these earlier studies were conducted at 220 MHz for proton, the resolution available from a computer-average of transients technique only allowed the identification of the two CH-6 and GH-8 resonances. In the present study, the use of Fourier transform acquisition, coupled with careful comparison of the  $70^\circ\text{C}$  chemical shift data of CCGG with those for sequence precursors allowed the identification and assignment of all the aromatic base and anomeric ribose proton resonances. This improved resolution of the spectra allowed further observations of the helix-coil transition of CCGG. The C(1)H-5 signal follows a shielding pattern similar to the two CH-6 signals. As the temperature is initially decreased, these resonances experience a deshielding as duplex formation begins. Decreasing the temperature further results in a shielding trend for the same resonances. This shielding occurs at temperatures below  $T_m$ , and might result from an end-to-end aggregation of the duplex species in solution.

The H-1' melting curves are more complex and are not understood. Two of the resonances, C(2)H-1' and G(3)H-1', undergo a sigmoidal transition, with a large shielding observed for the C(3)H-1' proton. The H-1', H-2' coupling constants for these four resonances steadily decreased as the temperature was lowered, until at  $48^\circ\text{C}$  and lower,

these resonances were observed as broad singlets ( $J < 1.0$  Hz). This reduction in the coupling constants reflects a high population of stacked conformers (Altona, 1975) at temperatures close to the  $T_m$  of the duplex. This result supports the concept that for duplexing to occur with short oligoribonucleotides, each single strand is probably close to a fully stacked conformation (Appleby & Kallenbach, 1973; Kallenbach & Berman, 1977).

The sequence CGGC is only capable of forming a staggered double helix. The results of several variable temperature studies of this oligomer indicate that a duplex is formed in a reversible manner. The average  $T_m$  for the duplex, as determined from the melting curves for the GH-8 and CH-5 resonances, is  $33.5 \pm 0.5^\circ\text{C}$ . This  $T_m$  value is anomalously low for a duplex of G-C base pairs, even when compared to CCGG, which is only four base pairs long. This low  $T_m$  value raised the question of whether the following duplexes were being formed:

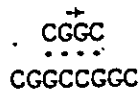


This possibility was ruled out by the results of a variable temperature study of the tetramer CGGA (Figure 40). Under the conditions used for the CGGC study, CGGA is not capable of forming a two base pair duplex with dangling ends.

Why does the staggered duplex formed by CGGC have such a low  $T_m$ ? A repeating unit of the staggered duplex can be represented by the perfect helix CGGC:GCCG. This perfect double helix has a calculated  $T_m$  of  $60^\circ\text{C}$  at 9.2 mM strand concentration, determined using the data

and equation of Borer et al. (1974). The CGGC staggered duplex has a much lower  $T_m$ , even though its chain length is probably longer than four G-C base pairs. Every two base pairs in a staggered helix, one of the two strands is interrupted by a "gap" in the backbone: the absence of a phosphodiester linkage. The oligomer d(ATAT) forms a staggered duplex, as demonstrated by X-ray studies (Viswamitra et al., 1978). The results of this study showed that the phosphodiester linkage opposite a gap is trans, gauche, giving rise to an extended conformation. In the CGGC staggered duplex, the trans, gauche conformation would occur in each GpG phosphodiester linkage. The base stacking interactions, important to the stability of a double helix, are decreased in a staggered duplex because of the greater separation of base planes, and disruption of the right handed helix at the gap sites. The  $T_m$  results of the variable temperature study of CGGC are consistent with the special conformation observed for the staggered duplex of d(ATAT) (Viswamitra et al., 1978).

How long is the staggered duplex? A staggered duplex formed by three CGGC molecules would contain four G-C base pairs and two single strand dangling regions:



Dangling bases decrease the amount of end-to-end stacking of duplex species in solution, leading to a decrease in low temperature line broadening in comparison to the line broadening observed for a perfect duplex. Figure 39 shows that a rapid and large line broadening effect

is observed for CGGC as temperature is decreased, inconsistent with the formation of a short staggered duplex. The line broadening result suggests that the staggered duplex has a much longer chain length. Consideration of the nature of the helix-coil transition of a staggered duplex would also support a longer chain length model. The nucleation complex must be stable enough to allow duplex formation to occur in a forward direction (Craig et al., 1971). Each 'intermediate' form of a staggered duplex can therefore act as a 'nucleation' complex for intermolecular propagation of the duplex. The results of the CGGA study show that a simple two base pair duplex is very unstable; however, CGGC does duplex and this simple duplex must also be capable of being a nucleation complex. Presumably propagation will be able to occur at least equally well with all of the various chain length 'nucleation' complexes, and the chain lengths will continue to grow past four or six base pairs. Such an interpretation of the helix-coil transition is supported by the observed broadening of the spectral lines at low temperatures.

The CGCG tetramer can form both a perfect and a staggered duplex. The results of a variable temperature study on this sequence are presented in Figure 41. The  $T_m$  values determined from the sigmoidal curves can be separated into two groups:  $41 \pm 2^\circ\text{C}$  for the GH-8 resonances, and  $47.4 \pm 0.6^\circ\text{C}$  for the C(3)H-6 and H-5 resonances. This discrepancy in  $T_m$  values occurs for both C(3) and G(2), which form a base pair together in a perfect helix, and also for both C(3) and G(4), which form a base pair in a staggered duplex. This result can not therefore be clearly attributed to either a perfect or staggered

double helix. The observed low temperature broadening of the spectra resonances could be accounted for by either the formation of a staggered duplex, or the end-to-end aggregation of perfect duplexes. Duplex formation of this oligomer is probably a complex series of equilibria:

- 1) at relatively high temperatures, the main equilibrium is coil  $\rightleftharpoons$  perfect duplex. Nucleation complexes formed at these temperatures would be stable enough to form duplexes by intramolecular propagation but not by intermolecular propagation.
- 2) as the temperature is further decreased, it is probable that the following equilibrium mixture exists:

staggered duplex  $\rightleftharpoons$  coil  $\rightleftharpoons$  perfect duplex

with shifts in equilibria to respond to the most stable products.

- 3) at lower temperatures, the perfect duplex species can form butt end-to-end aggregations, which can form a staggered duplex by chain slippage:



Evaluation of the results to quantify the population of the various species and equilibrium positions is not possible at present.

The results of the study arise from either a staggered duplex or end-to-end aggregation of a perfect duplex, but can not distinguish between these two possibilities. However, the  $T_m$  data suggest that the helix-coil transition for CGCG is more complex than for CCGG or GCGC, each of which can only form a single duplex type.

## CONCLUSION

The stability of short, imperfect RNA double helices was investigated by proton magnetic resonance spectroscopy, since this physical technique can give conformational information of solution species at the individual nucleoside level. The short oligoribonucleotides used in these studies were chemically synthesized by the phosphotriester method of Neilson and associates, which is capable of preparing a wide variety of sequences in the quantities required for NMR spectroscopic experiments. Stepwise synthesis of the oligomers yielded small amounts of precursor sequences, whose pmr data was used in making unambiguous assignments of the proton resonances of longer sequences. The work reported in this thesis was the first to combine chemical synthesis of oligoribonucleotides with pmr spectroscopic techniques to investigate systematically the effect of imperfections on the stability of short RNA duplexes.

The stabilizing effect of single stranded regions adjacent to a double helix is important to the intermolecular interaction between two RNA molecules, for example the anticodon-codon duplex formed between tRNA and mRNA (Grosjean et al., 1976; Yoon et al., 1976). The mechanism by which a dangling base region stabilizes the duplex was not known. The dangling base was thought to introduce this additional stability by either increasing favourable base stacking interactions (Uhlenbeck et al., 1971) or by reducing the fraying of the terminal

base pairs (Kallenbach & Berman, 1977). The pmr study of CAUG demonstrated that this oligomer forms a stable duplex which does not fray at the G-C base pairs. Investigation of the dangling base effect by the oligoribonucleotides CAUGA and CAUGU demonstrated an increase in the stability of the duplex, the effect being greater for a dangling A. Additional pmr evidence supported a base stacking role for the dangling base. The oligomer AGCU was investigated by pmr spectroscopy, and found to form a double helix with fraying terminal A-U base pairs. Extension of the sequence to AGCUA allowed an examination of the proposal of Kallenbach and Berman (1977). The dangling A increased the duplex stability of AGCU but did not significantly reduce the fraying of the terminal A-U base pairs. The results of these two studies are consistent with the proposal that a dangling base stabilizes a double helix by increasing favourable base stacking interactions.

The duplex set AGGA:UCCU was studied in an attempt to determine whether the proposed Shine-Dalgarno mRNA-16 S rRNA binding site was selected by nature because it is inherently more stable than other duplexes with the same base composition. The results of the experiment demonstrated that the AGGA:UCCU duplex was more stable than the CAUG and AGCU duplexes. Unlike AGCU, the AGGA:UCCU double helix does not fray at the terminal A-U base pairs. Apparently fraying is a sequence related phenomenon.

The four pentanucleotides CAXUG (X = A,G,C,U) were used in seven experiments designed to investigate diverse aspects of RNA secondary structure. Each separate pentanucleotide was used to

examine the formation of a small internal loop within a short duplex. The results of the pmr study indicated that these pentamers did not form double helices under conditions where CAUG exists as a stable duplex. The double helix sets CAAUG:CAUUG and CAGUG:CACUG were used to further calibrate the model system to allow useful comparisons to be made. The results of the pmr study of CAGUG:CACUG illustrate the exaggerated stabilizing effect of G-C base pairs within short double helices. The increased stability of CAAUG:CAUUG with respect to CAUG was similar to the increase in stability observed for CAUGU. A dangling U increases duplex stability approximately as much as an additional A-U base pair.

The final duplex set to be studied was CAGUG:CAUUG. This study was the first to investigate the formation of a G-U wobble base pair within a regular Watson-Crick base paired double helix. The use of pmr spectroscopy allowed the observation that the G-U base pair was less stable than the surrounding Watson-Crick base pairs. This decreased stability may result from the inability of the G-U base pair to adopt a true wobble geometry.

Recognition sites in RNA may be partially masked in a duplex region, which must open to allow interaction with other RNA species, for example the mRNA-16 S rRNA interaction (Steitz & Jakes, 1975). The results of the experiments described here can possibly suggest certain criteria which may be useful in evaluating potential recognition sites on the basis of their ability to open:

- 1) The results of the CAXUG study suggest that short duplex regions separating small loops may have an increased susceptibility to



transient opening.

- 2) The presence of G·U base pairs may also increase the susceptibility of short double helical regions to transient opening. This possibility was suggested by the observation that the G·U base pair is a point of local instability within the CAGUG:CAUUG duplex.
- 3) The results of the AGCU, AGCUA study suggest that once opening has occurred, the larger loop formed may possibly stabilize neighbouring duplex regions by a dangling base effect if the base pairs which close the loop fray. The ability of the base pairs to fray will be sequence related, as illustrated by the AGGA:UCCU duplex.
- 4) Once formed, the intermolecular RNA-RNA binding complex will be stabilized by the adjacent single stranded regions. This stabilization reflects an increase in base stacking interactions, and the strength of the stability will necessarily be dependent upon the sequence of the single strand region (i.e. CAUGA > CAUGU > CAUG).

The final study reported in this thesis illustrates the care which must be exercised in selecting sequences for study. Certain sequences, for example CGCG, are capable not only of forming a perfect double helix, but can also form a staggered duplex. The results of the pmr study of CGCG, which can only form a staggered duplex, demonstrate that in solution the staggered duplex must have a special conformation similar to the one proposed by the X-ray study of the staggered duplex of d(ATAT) (Viswamitra *et al.*, 1978). A pmr study of CGCG was probably complicated by the existence of a complex series of equilibria between random coil, perfect double helix and staggered duplex. Earlier studies on similar sequences (Patel, 1976b, 1979) did

not recognize this complicating factor, and treated the data on the basis of a simple perfect helix  $\rightleftharpoons$  coil equilibrium. Sequences which can only form a perfect duplex should be chosen to keep the helix-coil transition under study relatively simple.

#### REFERENCES

- Adamiak, R.W., Arentzen, R.W. and Reese, C.B. (1977) Tetrahedron Letters, 1431.
- Agarwal, K.L., Buechi, H., Caruthers, M.H., Gupta, N., Khorana, H.G., Kleppe, K., Kumar, A., Ohtsuka, E., RajBhandary, U.L., van de Sande, J.H., Sgaramella, V., Weber, H. and Yamada, T. (1970) Nature 277, 27.
- Alderfer, J.L. and Ts'o, P.O.P. (1977) Biochemistry 16, 2410.
- Altona, C. (1975) in "Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions." Sundaralingam, M. and Rao, S.T., Eds., Baltimore, University Park Press, p. 613.
- Appleby, D.W. and Kallenbach, N.R. (1973) Biopolymers 12, 2093.
- Arnott, S., Chandrasekaran, R. and Marttilla, C.M. (1974) Biochem. Journal 141, 537.
- Arnott, S., Dover, S.D. and Wonacott, A.J. (1969) Acta Crystallogr. B. 25, 2192.
- Arter, D.B. and Schmidt, P.G. (1976) Nucleic Acids Research 3, 1437.
- Arter, D.B., Walker, G.C., Uhlenbeck, O.C. and Schmidt, P.G. (1974) Biochem. Biophys. Res. Commun. 61, 1089.
- Bahl, C.P., Wu, R., Itakura, K., Katagiri, N. and Narang, S.A. (1976) Proceedings Natl. Acad. Sci. U.S.A. 73, 91.
- Borer, P.N., Kan, L.S. and Ts'o, P.O.P. (1975) Biochemistry 14, 4847.

Borer, P.N., Dengler, B., Tinoco Jr., I. and Uhlenbeck, O.C. (1974)

J. Molecular Biology 86, 843.

Borer, P.N., Uhlenbeck, O.C., Dengler, B. and Tinoco Jr., I. (1973) J.

Molecular Biology 80, 759.

Brahms, J., Maurizot, J.C. and Michelson, A.M. (1967) J. Molecular

Biology 25, 481.

Breslauer, K.J., Sturtevant, J.M. and Tinoco Jr., I. (1975) J. Molecular

Biology 99, 549.

Brown, D.M., Magrath, D.I., Neilson, A.H. and Todd, A.R. (1956) Nature

177, 1124.

Broyde, S. and Hingerty, B. (1978) Nucleic Acids Research 5, 2729.

Broyde, S., Wartell, R.M., Stellman, S.D. and Hingerty, B. (1978)

Biopolymers 17, 1485.

Camerman, N., Fawcett, J.K. and Camerman, A. (1976) J. Molecular

Biology 107, 161.

Cheng, D.M., Dhingra, M.M. and Sarma, R.H. (1978) Nucleic Acids

Research 5, 4399.

Cheng, D.M. and Sarma, R.H. (1977) J. Am. Chem. Soc. 99, 7333.

Chiao, Y.-C. and Krugh, T.R. (1977) Biochemistry 16, 747.

Contreras, R., Ysebaert, M., Min Jou, W. and Fiers, W. (1973) Nature

New Biol. 241, 99.

Craig, M.E., Crothers, D.M. and Doty, P. (1971) J. Molecular Biology

62, 383.

Crick, F.H.C. (1966) J. Molecular Biology 19, 548.

Cross, A.D. and Crothers, D.M. (1971) Biochemistry 10, 4015.

Crothers, D.M., Hilbers, C.W. and Schulman, R.G. (1973) Proc. Natl. Acad. Sci. U.S.A. 70, 2899.

Danyluk, S.S. and Hruska, F.E. (1968) Biochemistry 7, 1038.

Day, R.O. Seeman, N.C., Rosenberg, J.M. and Rich, A. (1973) Proc. Natl. Acad. Sci. U.S.A. 70, 849.

Delisi, C. and Crothers, D.M. (1971) Proc. Natl. Acad. Sci. U.S.A. 68, 2682.

Dhingra, M.M. and Sarma, R.H. (1978) Nature 272, 798.

Dhingra, M.M., Sarma, R.H., Giessner-Prettre, C. and Pullman, B. (1978) Biochemistry 17, 5815.

Dodgson, J.B. and Wells, R.D. (1977a) Biochemistry 16, 2367.

Dodgson, J.B. and Wells, R.D. (1977b) Biochemistry 16, 2374.

Donohue, J. and Trueblood, K.N. (1960) J. Molecular Biology 2, 363.

Early, T.A., Kearns, D.R., Burd, J.F., Larson, J.E. and Wells, R.D. (1977) Biochemistry 16, 541.

Early, T.A., Olmsted III, J., Kearns, D.R. and Lezius, A.G. (1978) Nucleic Acids Research 5, 1955.

Eckstein, F. and Rizk, I. (1967a) Angew. Chem. Internat. Edit. 6, 695.

Eckstein, F. and Rizk, I. (1967b) Angew. Chem. Internat. Edit. 6, 949.

England, T.E. (1976) Ph.D. Thesis, McMaster University.

England, T.E. and Neilson, T. (1976) Can. J. Chem. 54, 1714.

England, T.E. and Neilson, T. (1977) Can. J. Chem. 55, 365.

Evans, F.E., Lee, C.-H. and Sarma, R.H. (1975) Biochem. Biophys. Res. Commun. 63, 106.

Ezra, F.S., Kondo, N.S., Ainsworth, C.F. and Danyluk, S.S. (1976) Nucleic Acids Research 3, 2549.

Ezra, F.S., Lee, C.-H., Kondo, N.S., Danyluk, S.S. and Sarma, R.H.

(1977) *Biochemistry* 16, 1977.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J.,

Min Jou, W., Raeymaekers, A., Volckaert, G., Ysebaert, M.,

Van de Kerckhove, J., Nolf, F. and Van Montagu, M. (1975)

*Nature* 256, 273.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Tserentant, D.,

Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A.,

van den Berghe, A., Volckaert, G. and Ysebaert, M. (1976)

*Nature* 260, 500.

Ganoza, M.C., Fraser, A.R. and Neilson, T. (1978) *Biochemistry* 17, 2769.

Giessner-Prettre, C. and Pullman, B. (1976) *Biochem. Biophys. Res.*

*Commun.* 70, 578.

Giessner-Prettre, C., Pullman, B., Borer, P.N., Kan, L.-S. and Ts'o,

P.O.P. (1976) *Biopolymers* 15, 2277.

Giessner-Prettre, C., Pullman, B. and Caillet, J. (1977) *Nucleic Acids*

*Research* 4, 99.

Gilbert, W. and Maxam, A. (1973) *Proc. Natl. Acad. Sci. U.S.A.* 70, 3581.

Gorenstein, D.G., Findlay, J.B., Momii, R.K., Luxon, B.A. and Kar, D.

(1976) *Biochemistry* 15, 3796.

Govil, G. and Smith, I.C.P. (1973) *Biopolymers* 12, 2589.

Gralla, J. and Crothers, D.M. (1973a) *J. Mol. Biol.* 73, 497.

Gralla, J. and Crothers, D.M. (1973b) *J. Mol. Biol.* 78, 301.

Gray, D.M. and Ratliff, R.L. (1977) *Biopolymers* 16, 1331.

Gray, D.M., Tinoco Jr., I. and Chamberlin, M.J. (1972) *Biopolymers* 11,

1235.

- Gregoire, R.J. and Neilson, T. (1978) Can. J. Chem. 56, 487.
- Griffin, B.E., Jarman, M., Reese, C.B., Sulston, J.E. and Trentham, D.R. (1966) Biochemistry 5, 3638.
- Grosjean, H., Soell, D.G. and Crothers, D.M. (1976) J. Mol. Biol. 103, 499.
- Gross, H.J., Domdey, H., Lossow, C., Jank, P., Raba, M., Alberty, H. and Saenger, H.L. (1978) Nature 273, 203.
- Gupta, G. and Sasisekharan, V. (1978a) Nucleic Acids Research 5, 1639.
- Gupta, G. and Sasisekharan, V. (1978b) Nucleic Acids Research 5, 1655.
- Heynecker, H.L., Shine, J., Goodman, H.M., Boyer, H.W., Rosenberg, J., Dickerson, R.E., Narang, S.A., Itakura, K., Lin, S. and Riggs, A.D. (1976) Nature 263, 748.
- Hilbers, C.W. and Patel, D.J. (1975) Biochemistry 14, 2656.
- Hingerty, B. and Broyde, S. (1978) Nucleic Acids Research 5, 127.
- Hughes, D.W., Bell, R.A., England, T.E. and Neilson, T. (1978) Can. J. Chem. 56, 2243.
- Itakura, K., Hirose, T., Crea, R., Riggs, A., Heyneker, H., Bolivar, F. and Boyer, H. (1977) Science 198, 1056.
- Itakura, K., Katagiri, N., Narang, S.A., Bahl, C.P., Mariani, K.J. and Wu, R. (1975) J. Biol. Chem. 250, 4592.
- Jain, S.C., Tsai, C.-C. and Sobell, H.M. (1977) J. Mol. Biol. 114, 317.
- Jaskunas, S.R., Cantor, C.R. and Tinoco Jr., I. (1968) Biochemistry 7, 3164.
- Kallenbach, N.R. and Berman, H.M. (1977) Quarterly Review of Biophysics 10, 138.

Kallenbach, N.R., Daniel Jr., W.E. and Kaminker, M.A. (1976) *Biochemistry* 15, 1218.

Kan, L.S., Barrett, J.C. and Ts'o, P.O.P. (1973) *Biopolymers* 12, 2409.

Kan, L.S., Borer, P.N. and Ts'o, P.O.P. (1975) *Biochemistry* 14, 4864.

Kan, L.S., Ts'o, P.O.P., van der Haar, F., Sprinzl, M. and Cramer, F.  
(1974) *Biochem. Biophys. Res. Commun.* 59, 22.

Kastrup, R.V. and Schmidt, P.G. (1978) *Nucleic Acids Research* 5, 257.

Kastrup, R.V., Young, M.A. and Krugh, T.R. (1978) *Biochemistry* 17, 4855.

Katagiri, N., Itakura, K. and Narang, S.A. (1974) *J. Chem. Soc. Chem. Commun.*, 325.

Kearns, D.R. (1976) *Prog. Nucl. Acid. Res. Mol. Biol.* 18, 91.

Kearns, D. (1977) *Ann. Rev. Biophys. Bioeng.* 6, 477.

Khorana, H.G., Agarwal, K.L., Besmer, P., Buechi, H., Caruthers, M.H.,  
Cashion, P.J., Fridkin, M., Jay, E., Kleppe, K., Kleppe, R.,  
Kumar, A., Loewen, P.C., Miller, R.C., Minamoto, K., Panet, A.,  
RajBhandary, U.L., Ramamoorthy, B., Sekiya, T., Takaya, T. and  
van de Sande, J.H. (1976) *J. Biol. Chem.* 251, 565 *et seq.*

Kim, S.H., Berman, H.M., Seeman, N.C. and Newton, M.D. (1973) *Acta Crystallogr., Sect. B* 29, 703.

Koessel, H. and Seliger, H. (1975) *Fortschr. Chem. Org. Naturstoffe* 32, 298.

Kondo, N.S., Holmes, H.M., Stempel, L.M. and Ts'o, P.O.P. (1970)  
*Biochemistry* 9, 3479.

Krol, A., Machatt, M.A., Branlant, C. and Ebel, J.P. (1978) *Nucleic Acids Research* 5, 4933.



- Krugh, T.R., Laing, J.W. and Young, M.A. (1976) *Biochemistry* 15, 1224.
- Krugh, T.R., Mooberry, E.S. and Chaio, Y.-C.C. (1977) *Biochemistry* 16, 740.
- Krugh, T.R. and Young, M.A. (1975) *Biochem. Biophys. Res. Commun.* 62, 1025.
- Lapper, R.D., Mantsch, H.H. and Smith, I.C.P. (1973) *J. Am. Chem. Soc.* 95, 2878.
- Lee, C.-H., Evans, F.E. and Sarma, R.H. (1975) *FEBS Letters* 51, 73.
- Lee, C.-H., Ezra, F.S., Kondo, N.S., Sarma, R.H. and Danyluk, S.S. (1976) *Biochemistry* 15, 3627.
- Lee, C.-H. and Tinoco Jr., I. (1977) *Biochemistry* 16, 5403.
- Lee, C.-H. and Tinoco Jr., I. (1978) *Nature* 274, 609.
- Letsinger, R.L. and Mahadevan, V. (1965) *J. Am. Chem. Soc.* 87, 3526.
- Letsinger, R.L. and Mahadevan, V. (1966) *J. Am. Chem. Soc.* 88, 5319.
- Letsinger, R.L. and Ogilvie, K.K. (1967) *J. Am. Chem. Soc.* 89, 4801.
- Letsinger, R.L. and Ogilvie, K.K. (1969) *J. Am. Chem. Soc.* 91, 3350.
- Lezius, A.G. and Domin, E. (1973) *Nature New Biol.* 244, 169.
- Lohrmann, R., Soell, D., Hayatsu, H., Ohtsuka, E. and Khorana, H.G. (1966) *J. Am. Chem. Soc.* 88, 819.
- Lomant, A.J. and Fresco, J.R. (1975) *Prog. Nucl. Acid Res. Mol. Biol.* 15, 185.
- Maizels, N. (1974) *Nature* 249, 647.
- Marians, K.J., Wu, R., Stawinski, J., Hozumi, H. and Narang, S.A. (1976) *Nature* 263, 744.
- Martin, F.H., Uhlenbeck, O.C. and Doty, P. (1971) *J. Mol. Biol.* 57, 201.

- Michelson, A.M. and Todd, A.R. (1955) J. Chem. Soc., 2632.
- Milman, G., Langridge, R. and Chamberlin, M.J. (1976) Proc. Natl. Acad. Sci. U.S.A. 57, 1804.
- Mizuno, H. and Sundaralingam, M. (1978) Nucleic Acids Research 5, 4451.
- Neidle, S., Achari, A., Taylor, G.L., Berman, H.M., Carrell, H.L., Glusker, J.P. and Stallings, W.C. (1977) Nature 269, 304.
- Neilson, T. (1969) Chem. Comm., 1139.
- Neilson, T., Deugau, K.V., England, T.E. and Werstiuk, E.S. (1975) Can. J. Chem. 53, 1093.
- Neilson, T., Wastrodowski, E.V. and Werstiuk, E.S. (1973) Can. J. Chem. 51, 1068.
- Neilson, T. and Werstiuk, E.S. (1971a) Can. J. Chem. 49, 493.
- Neilson, T. and Werstiuk, E.S. (1971b) Can. J. Chem. 49, 3004.
- Neilson, T. and Werstiuk, E.S. (1974) J. Am. Chem. Soc. 96, 2295.
- Nirenberg, N.W. and Leder, P. (1964) Science 145, 1399.
- Ogilvie, K.K., Beaucage, S.L. and Entwistle, D.W. (1976) Tetrahedron Letters, 1255.
- Olson, W.K. (1978) Biopolymers 17, 1015.
- Patel, D.J. (1976a) Biochimica et Biophysica Acta 442, 98.
- Patel, D.J. (1976b) Biopolymers 15, 533.
- Patel, D.J. (1977) Biopolymers 16, 1635.
- Patel, D.J. (1979) Biopolymers 18, 553.
- Patel, D.J. and Canuel, L. (1976a) Proc. Natl. Acad. Sci. U.S.A. 73, 674.
- Patel, D.J. and Canuel, L.L. (1976b) Proc. Natl. Acad. Sci. U.S.A. 73, 3343.

- Patel, D.J. and Canuel, L.L. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 2624.
- Patel, D.J. and Hilbers, C.W. (1975) Biochemistry 14, 2651.
- Pless, R.C. and Ts'o, P.O.P. (1977) Biochemistry 16, 1239.
- Podder, S.K. (1971) Eur. J. Biochem. 22, 467.
- Poerschke, D., Uhlenbeck, O.C. and Martin, F.H. (1973) Biopolymers 12, 1313.
- Pohl, F.M. (1974) FEBS Letters 38, 202.
- Ravetch, J., Gralla, J. and Crothers, D. (1974) Nucleic Acids Research 1, 109.
- Reese, C.B. (1978) Tetrahedron 34, 3143.
- Reese, C.B., Saffhill, R. and Sulston, J.E. (1967) J. Am. Chem. Soc. 89, 3366.
- Reese, C.B. and Stewart, J.C.M. (1968) Tetrahedron Letters, 4273.
- Reese, C.B., Titmas, R.C. and Yau, L. (1978) Tetrahedron Letters, 2727.
- Reese, C.B. and Trentham, D.R. (1965) Tetrahedron Letters, 2467.
- Reinhardt, C.G. and Krugh, T.R. (1978) Biochemistry 17, 4845.
- Rich, A. (1977) Accts. of Chemical Research 10, 388.
- Robillard, G.T., Hilbers, C.W., Reid, B.R., Gangloff, J., Dirheimer, G. and Schulman, R.G. (1976) Biochemistry 15, 1883.
- Romaniuk, P.J., Neilson, T., Hughes, D.W. and Bell, R.A. (1978a) Can. J. Chem. 56, 2249.
- Romaniuk, P.J., Hughes, D.W., Gregoire, R.J., Neilson, T. and Bell, R.A. (1978b) J. Am. Chem. Soc. 100, 3971.
- Rordorf, B.F., Kearns, D.R., Hawkins, E. and Chang, S.H. (1976) Biopolymers 15, 325.

Rosenberg, J.M., Seeman, N.C., Day, R.O. and Rich, A. (1976) J. Mol. Biol. 104, 145.

Rosenberg, J.M., Seeman, N.C., Kim, J.J.P., Suddath, F.L., Nicholas, H.B. and Rich, A. (1973) Nature 243, 150.

Rubin, J., Brennan, T. and Sundaralingam, M. (1972) Biochemistry 11, 3112.

Saenger, W., Mazumdar, S.K., Suck, D. and Manor, P.C. (1975) in "Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions." Sundaralingam, M. and Rao, S.T., Ed., Baltimore, University Park Press, p. 537.

Sakore, T.D., Jain, S.C., Tsai, C.-C. and Sobell, H.M. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 188.

Schaller, H., Weimann, G., Lerch, B. and Khorana, H.G. (1963) J. Am. Chem. Soc. 85, 3821.

Schechter, E., Barlow, M. Sparks, R. and Trueblood, K.N. (1969) Acta Crystallogr., Sect. B. 25, 895.

Schechter, E., Barlow, M., Sparks, R. and Trueblood, K.N. (1964) J. Am. Chem. Soc. 86, 1872.

Scheller, R.H., Dickerson, R.E., Boyer, H.W., Riggs, A.D. and Itakura, K. (1977) Science 196, 177.

Schleich, T., Cross, B.P., Blackburn, B.J. and Smith, I.C.P. (1975) in "Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions." Sundaralingam, M. and Kao, S.T., Eds., Baltimore, University Park Press, p. 223.

Schleich, T., Cross, B.P. and Smith, I.C.P. (1976) Nucleic Acids Research 3, 355.

Schweizer, M.P., Broom, A.D., Ts'o, P.O.P. and Hollis, D.P. (1978) J. Am. Chem. Soc. 90, 1042.

Seeman, N.C., Rosenberg, J.M., Suddath, F.L., Kim, J.J.P. and Rich, A.

(1976) J. Mol. Biol. 104, 109.

Selsing, E., Wells, R.D., Early, T.A. and Kearns, D.R. (1978) Nature  
275, 249.

Shine, J. and Dalgarno, L. (1975) Nature 254, 34.

Shum, B.W.-K. (1977) Ph.D. Thesis, Yale University.

Sitz, T.O., Kuo, S.-C. and Nazar, R.N. (1978) Biochemistry 17, 5811.

Smith, M., Rammler, D.H., Goldberg, I.H. and Khorana, H.G. (1962)  
J. Am. Chem. Soc. 84, 430.

Smrt, J. and Sorm, F. (1962) Coll. Czech. Chem. Comm. 27, 73.

Steitz, J.A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. U.S.A. 72,  
4734.

Studnicka, G.M., Rahn, G.M., Cummings, I.W. and Salser, W.A. (1978)  
Nucleic Acids Research 5, 3365.

Suck, D., Manor, P.C. and Saenger, W. (1976) Acta crystallogr. Sect. B.  
32, 1727.

Sundaralingam, M. (1969) Biopolymers 7, 821.

Sundaralingam, M. (1975) in "Structure and Conformation of Nucleic  
Acid Interactions." Sundaralingam, M. and Rao, S.T., Eds.,  
Baltimore, University Park Press, p. 487.

Sussman, J.L., Seeman, N.C., Kim, S.H. and Berman, H.B. (1972) J.  
Mol. Biol. 66, 403.

Thiyagarajan, P. and Ponnuswamy, P.K. (1978) Biopolymers 17, 2143.

Tinoco Jr., I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck,  
O.C., Crothers, D.M. and Gralla, J. (1973) Nature New Biol.  
246, 40.

Tinoco Jr., I., Uhlenbeck, O.C. and Levine, M.D. (1971) *Nature* 230, 362.

Tsai, C.-C., Jain, S.C. and Sobell, H.M. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 628.

Tsai, C.-C., Jain, S.C. and Sobell, H.M. (1977) *J. Mol. Biol.* 114, 301.

Ts'o, P.O.P. (1974a) in Vol. I of "Basic Principles in Nucleic Acid Chemistry." Ts'o, P.O.P., Ed., New York, NY, Academic Press, p. 453.

Ts'o, P.O.P. (1974b) in Vol. II of "Basic Principles in Nucleic Acid Chemistry." Ts'o, P.O.P., Ed., New York, NY, Academic Press, p. 305.

Ts'o, P.O.P., Alderfer, J.L., Borer, P.N. and Kan, L.S. (1975) in "Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions." Sundaralingam, M. and Rao, S.T., Eds., Baltimore, University Park Press, p. 193.

Ts'o, P.O.P., Schweizer, M.P. and Hollis, D.P. (1969a) *Annals of the New York Academy of Sciences* 158, 256.

Ts'o, P.O.P., Kondo, N.A., Schweizer, M.P. and Hollis, D.P. (1969b) *Biochemistry* 8, 997.

Uhlenbeck, O.C., Baller, J. and Doty, P. (1970) *Nature* 225, 509.

Uhlenbeck, O.C., Borer, P.N., Dengler, B. and Tinoco Jr., I. (1973) *J. Mol. Biol.* 73, 483.

Uhlenbeck, O.C., Martin, F.H. and Doty, P. (1971) *J. Mol. Biol.* 57, 217.

van Boom, J.H., Burgers, P.M.J., Owen, G.R., Reese, C.B. and Saffhill, R. (1971) *Chem. Comm.*, 869.

van Boom, J.H., Burgers, P.M.J., van Deursen, P.H., Arentzen, R. and Reese, C.B. (1974) *Tetrahedron Letters*, 3785.

Viswamitra, M.A., Kennard, O., Jones, P.G., Sheldrick, G.M., Salisbury,

S., Falvello, L. and Shakked, Z. (1978) *Nature* 273, 687.

Wang, A.H.J., Nathans, J., van der Marel, G., van Boom, J.H. and Rich,

A. (1978) *Nature* 276, 471.

Werstiuk, E.S. and Neilson, T. (1972) *Can. J. Chem.* 50, 1283.

Werstiuk, E.S. and Neilson, T. (1973) *Can. J. Chem.* 51, 1889.

Werstiuk, E.S. and Neilson, T. (1976) *Can. J. Chem.* 54, 2689.

Wickstrom, E. and Tinoco Jr., I. (1974) *Biopolymers* 13, 2367.

Wilson, H.R. and Al-Mukhtar, J. (1976) *Nature* 263, 171.

Wood, D.J., Ogilvie, K.K. and Hruska, F.E. (1975) *Can. J. Chem.* 53,

2781.

Yathindra, N. and Sundaralingam, M. (1975) in "Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions."

Sundaralingam, M. and Rao, S.T., Eds., Baltimore, University Park Press, p. 649.

Yoon, K., Turner, D.H., Tinoco Jr., I., von der Haar, F. and Cramer, F.

(1976) *Nucleic Acids Research* 3, 2233.

Young, M.A. and Krugh, T.R. (1975) *Biochemistry* 14, 4841.