

SEQUENCE DEPENDENCE OF STABILITY  
AND INTERCALATION STUDIES  
ON SHORT SYNTHETIC OLIGORIBONUCLEOTIDE DUPLEXES.

By

© ALISON SINCLAIR, 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

September 1986


Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-35897-1



OLIGORIBONUCLEOTIDE STABILITY AND INTERCALATION STUDIES

DOCTOR OF PHILOSOPHY (1986)  
(Biochemistry)

McMASTER UNIVERSITY  
Hamilton, Ontario.

TITLE: Sequence Dependence of Stability and Intercalation Studies on Short  
Oligonucleotide Duplexes.

AUTHOR: Alison Sinclair, B.Sc. (University of Victoria)

SUPERVISORS: Professor T. Neilson, Professor R.A. Bell.

NUMBER OF PAGES: xxv, 319

## ABSTRACT

A phosphotriester synthesis was used to prepare short oligoribonucleotides for study by variable temperature  $^1\text{H-NMR}$ . The duplex  $\rightleftharpoons$  single strand transition was monitored by the associated changes in chemical shift and coupling constant for the ribose base, imino, and anomeric protons. A measurable parameter of stability, the melting temperature ( $T_m$ ), was extracted from the chemical shift versus temperature data.

The sequence dependence of stability was studied for a series of tetramer sequences comprising exclusively guanosine and cytidine residues. These sequences formed perfect duplexes with stabilities:  $(\text{GGCC})_2 > (\text{GCGC})_2 > (\text{CCGG})_2 = \text{GCCG}:\text{CGGC} > (\text{CGCG})_2$ , which led to a ranking of stabilities of nearest neighbour interactions:  $\text{GC}:\text{GC} \geq \text{GG}:\text{CC} > \text{CG}:\text{CG}$ .

Study of complementary mixtures of GC-rich trimers in the presence and absence of salt revealed a highly salt dependent self-aggregation of the G-rich strand. A structure for the GCG aggregate is proposed.

Studies of stability and structure of base-base mismatches in GC-rich oligoribonucleotides were performed on six pentaribonucleotides derived from the GC tetramers. Two sequences, CCAGG and GCUGC, formed duplexes of comparable stability, containing AxA and UxU mismatches respectively. The mismatched bases stacked into the duplex. One sequence, GGACC, did not duplex; the others (GGUCC, CCUGG, GCAGC) formed structures not incorporating mismatched base pairs. Mixing experiments on the complementary pentamers (e.g., GGACC:GGUCC) allowed study of the effect of introduction of A•U base pairs on the GC tetramer stability.

Intercalation of proflavine into the mismatch containing duplexes CCAGG and GCUGC (and their parent duplexes CCGG and GCGC) was studied to compare intercalation into mismatch and non mismatch sites. Addition of increasing ratios of proflavine produced a steady increase in  $T_m$  and shielding for most protons, but there is no indication of site preference or looping out of mismatched bases. GCUGC tabled a threefold higher  $T_m$  increase than GCGC, but the starting  $T_m$ s are too widely separated to allow direct comparison of binding affinities.

## ACKNOWLEDGEMENTS.

I wish to thank the following people for their contributions to this work:

Dr. Dirk Alkema, Dr. Don Hughes and Dr. Jan Coddington for their assistance in learning techniques of synthesis and instrumentation in the early stages of this project; Mr. Garry Buchko, Mr. Brian Allore, Dr. John Orban for stimulating discussion in the latter stages; my committee of Dr. R.A. Bell, Dr. R. Epand and Dr. G.E. Gerber for their advice and attention; and especially my supervisor, Dr. Thomas Neilson, to whose memory this thesis is dedicated.

I also wish to thank the National Sciences and Engineering Research Council and the Medical Research Council of Canada for their financial support.

## CONTENTS

	Page
Abstract	iii
Acknowledgements	v
List of Figures	xii
List of Tables	xviii
List of Abbreviations	xxiv
1. Introduction.	1
1.1. Nucleic acid structure.	2
1.2. NMR theory.	8
1.3. <sup>1</sup> H-NMR on nucleic acids.	11
1.3.1. Assignment of single strands and duplexes.	11
1.3.2. Conformational analysis of the single strand.	16
1.3.3. Monitoring the single strand to duplex transition.	24
1.3.4. Conformational analysis of the duplex.	26
1.4. Phosphotriester synthesis of oligoribonucleotides.	28
2. Materials and methods.	30
2.1. Oligoribonucleotide synthesis.	30
2.1.1. Materials.	30
2.1.2. Preparation of N <sup>4</sup> -benzoyl-2'-O-tetrahydro- pyranilycytidine.	31
2.1.2.1. 3',5'-O-[1,1,3,3-tetraisopropyl-1,3- disiloxanediyl]-cytidine.	31
2.1.2.2. N <sup>4</sup> -Benzoyl-3',5'-O-[1,1,3,3-tetraisopropyl-1,3- disiloxanediyl]-cytidine.	33



	Page
2.1.2.3. N <sup>4</sup> -Benzoyl-3'-5'-O-[1,1,3,3-tetraisopropyl-1,3-disiloxanediyl]-2'-O-tetrahydropyranyl-cytidine.	34
2.1.2.4. N <sup>4</sup> -Benzoyl-2'-O-tetrahydropyranylcytidine.	35
2.1.3. Synthesis of oligoribonucleotides.	36
2.1.3.1. N <sup>2</sup> -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenyl-methoxy-acetylguanylyl[3'-(2,2,2-trichloroethyl)-5']-N <sup>4</sup> -benzoyl-2'-O-tetrahydropyranylcytidine.	36
2.1.3.2. N <sup>2</sup> -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenyl-methoxy-acetylguanylyl[3'-(2,2,2-trichloroethyl)-5']-N <sup>4</sup> -benzoyl-2'-O-tetrahydropyranylcytidylyl-[3'-(2,2,2-trichloroethyl)-5']-N <sup>2</sup> -benzoyl-2'-O-tetrahydropyranylguanosine.	41
2.1.3.3. N <sup>2</sup> -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenyl-methoxy-acetylguanylyl[3'-(2,2,2-trichloroethyl)-5']-N <sup>4</sup> -benzoyl-2'-O-tetrahydropyranylcytidylyl-[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-tetrahydropyranylcytidine.	42
2.1.4. Block Coupling.	42
2.1.5. Deprotection procedures for oligoribonucleotides.	44
2.1.6. Purification of sequences by High Performance Liquid Chromatography (HPLC).	48
2.1.7. <sup>1</sup> H-NMR methodology.	49
2.1.7.1. Verification of sequences.	50

	Page
2.1.7.2. Observation of melting behaviour and determination of melting temperature.	50
2.1.7.3. Observation of hydrogen-bonded imino protons.	50
3. Sequence dependence of stability studies on tri- and tetranucleotide duplexes containing guanosine-cytidine base pairs.	56
3.1. Introduction.	
3.1.1. Sequence dependence of stability.	56
3.1.2. Sequence dependence of structure.	60
3.1.3. Studies of the sequence dependence of stability of GC-containing tetramers.	63
3.2. Results.	
3.2.1. Assignment of proton spectra.	68
3.2.2. Assignment of CC by 2-D shift correlated spectroscopy (COSY).	68
3.2.3. Assignment of imino proton resonances.	72
3.2.4. Secondary structure and stability of the six self-complementary GC containing tetramers.	82
3.2.5. Relative stability of GC cores.	96
3.2.6. The mixing experiments: Duplexing vs. aggregation.	103
3.2.6.1. The duplexes: GCCG:CGGC and GGC:GCC.	104
3.2.6.2. The aggregates: GGA with CCA, GCG with CGC, and CGG with CCG in 1.0M salt; self-aggregation of GCG.	107
3.2.6.3. The duplexes: GGA:CCA, GCG:CGC, CGG:CCG and GCG in the absence of salt.	120

	Page
3.3 Discussion	130
3.3.1. Comparison of results for individual sequences with previously published oligoribonucleotide studies.	130
3.3.2. Relative stability of diribonucleotide cores.	131
3.3.3. Comparison of the duplexing behaviour of RNA and DNA tetramers.	132
3.3.4. Application of the RY model to the tetramer results.	137
3.3.5. Nature of the GCG aggregate.	138
3.3.6. Salt-dependence of aggregation.	144
3.3.7. Comparison of calculated and observed imino proton shifts.	145
3.3.7.1. Assignment of imino proton resonances for the AGXCU:AGYCU series.	148
3.3.7.2. Discussion of the trends in imino proton behaviour.	151
3.4. Conclusions.	154
4. The effect of the introduction of central base-base mismatches on the secondary structure and stability of the GC tetramers.	156
4.1. Introduction.	
4.1.1. Published studies on base-base mismatches.	156
4.1.2. Studies on mismatch-containing duplexes derived from the GC tetramers.	166
4.2. Results.	
4.2.1. Sequences derived from parent GGCC.	168

	Page
4.2.1.1. Stability and secondary structure of GGUCC and GGACC.	168
4.2.1.2. Effect of insertion of an A•U base-pair: GGACC:GGUCC.	179
4.2.2. Sequences derived from parent CCGG.	183
4.2.2.1. Stability and secondary structure of CCUGG and CCAGG.	183
4.2.2.2. Conformation of the AxA mismatch: CCAGG.	191
4.2.2.3. Effect of insertion of A•U base pairs: CCAGG:CCUGG and CCAUGG.	192
4.2.3. Sequences derived from parent GCGC.	200
4.2.3.1. Stability and secondary structure of GCUGC and GCAGC.	200
4.2.3.2. Conformation of the UxU mismatch: GCUGC.	217
4.2.3.3. Effect of insertion of A•U base pair: GCAGC:GCUGC and GCAUGC.	218
4.3. Discussion.	227
4.3.1. Effect of insertion of base-base mismatches upon duplex stability.	227
4.3.2. Effect of insertion of A•U base pairs upon duplex stability.	230
4.3.3. Neighbour effects on the chemical shifts of the central adenosine in the A•U mixtures.	234
4.4. Conclusions.	234
5. Intercalation of proflavine into mismatch-containing duplexes.	234

	Page
5.1. Introduction.	
5.1.1. Biological effects of intercalation.	234
5.1.2. Sequence preferences of intercalation.	238
5.1.3. Detail of the intercalation site: 1. The intercalator orientation.	239
5.1.4. Detail of the intercalation site: 2. The nucleic acid conformation.	241
5.1.5. Intercalation into imperfect duplexes.	242
5.1.6. Studies of the intercalation of proflavine into GCUGC and CCAGG.	244
5.2. Results.	245
5.2.1. Intercalation of proflavine into GCGC and GCUGC.	245
5.2.2. Intercalation of proflavine into CCGG and CCAGG.	260
5.2.3. Duplex stabilization.	287
5.2.4. Proflavine orientation.	288
5.2.5. Binding sites: GCGC and CCGG.	288
5.2.6. Binding sites: GCUGC.	293
5.3. Discussion.	298
5.4. Conclusions.	302
6. Conclusions	303
References	307

## LIST OF FIGURES

Figure	Page
1.1 Structure of the oligoribonucleotide GCGC.	3
1.2 Normal base pairing.	5
1.3 Structural parameters used to describe base pair conformation.	6
1.4a Exchangeable and nonexchangeable protons in the aromatic bases.	12
1.4b Nonexchangeable protons belonging to the ribose ring.	12
1.5 IUPAC-IUB nomenclature for the torsion angles in a dinucleotide unit.	17
1.6 Idealized Newman projections for the classical rotamers of backbone torsion angles.	19
1.7 Idealized conformers of the ribose ring.	21
2.1 Preparation of N <sup>4</sup> -benzoyl-2'-O-tetrahydropyranylcytidine.	32
2.2 Chemical synthesis of GCGC.	39
3.1 RY model overlaps.	64
3.2 COSY spectrum of CC.	71
3.3 Chemical shift versus temperature results for CC.	73
3.4 Temperature dependence of line widths of hydrogen-bonded imino protons of GGCC.	76
3.5 Temperature dependence of line widths of hydrogen-bonded imino protons of CCGG.	77
3.6 Temperature dependence of line widths of hydrogen-bonded imino protons of GCGC.	78
3.7 Chemical shift versus temperature results for GGCC in 1.0 M salt buffer.	83
3.8 Chemical shift versus temperature results for CCGG in 1.0 M salt buffer.	84

Figure	Page
3.9 Chemical shift versus temperature results for GCCG in 1.0 M salt buffer.	88
3.10 Chemical shift versus temperature results for CGGC in 1.0 M salt buffer.	89
3.11 Chemical shift versus temperature results for GCGC in 1.0 M salt buffer.	94
3.12 Chemical shift versus temperature results for CGCG in 1.0 M salt buffer.	97
3.13 Chemical shift versus temperature results for CGCG in 0.1 M salt buffer.	99
3.14 Chemical shift versus temperature results for GCCG:CGGC in 1.0 M salt buffer.	106
3.15 Chemical shift versus temperature results for GGC:GCC in 1.0 M salt buffer.	108
3.16 Variable temperature <sup>1</sup> H-NMR spectra of GCG:CGC in 1.0 M salt buffer.	110
3.17 Chemical shift versus temperature results for GGA:CCA in 1.0 M salt buffer.	111
3.18 Chemical shift versus temperature results for GCG:CGC in 1.0 M salt buffer.	112
3.19 Chemical shift versus temperature results for CGG:CCG in 1.0 M salt buffer.	113
3.20 Chemical shift versus temperature results for GCG in 1.0 M salt buffer.	118
3.21 Chemical shift versus temperature results for GGA:CCA in no salt buffer.	121
3.22 Chemical shift versus temperature results for GCG:CGC in no salt buffer.	123
3.23 Chemical shift versus temperature results for CGG:CCG in no salt buffer.	126
3.24 Chemical shift versus temperature results for GCG in no salt buffer.	128
3.25 Base pair overlaps in the A and B helices.	136

Figure	Page
3.26 Temperature dependence of the anomeric proton coupling constants of GGCC.	139
3.27 Temperature dependence of the anomeric proton coupling constants of CCGG.	139
3.28 Temperature dependence of the anomeric proton coupling constants of GCGC.	140
3.29 Temperature dependence of the anomeric proton coupling constants of CGCG.	140
3.30 Possible base pairing between guanosine residues.	142
3.31 Proposed structure for the GCG aggregate.	143
3.32 Trends in winding angles between terminal base pairs from imino proton data.	153
4.1 Chemical shift versus temperature results for GGCC in 0.1 M salt buffer.	169
4.2 Chemical shift versus temperature results for GGUCC in 0.1 M salt buffer.	171
4.3 Chemical shift versus temperature results for GGACC in 0.1 M salt buffer.	175
4.4 Chemical shift versus temperature results for GGAC in 0.1 M salt buffer.	177
4.5 Chemical shift versus temperature results for GGACC:GGUCC in 0.1 M salt buffer.	180
4.6 Chemical shift versus temperature results for CCGG in 0.1 M salt buffer.	184
4.7 Chemical shift versus temperature results for CCUGG in 0.1 M salt buffer.	186
4.8 Chemical shift versus temperature results for CCAGG in 0.1 M salt buffer.	189
4.9 Chemical shift versus temperature results for CCAGG:CCUGG in 0.1 M salt buffer.	194
4.10 Chemical shift versus temperature results for CCAUGG in 1.0 M salt buffer.	197



Figure	Page
4.11 Chemical shift versus temperature results for GCGC in 0.1 M salt buffer.	201
4.12 Chemical shift versus temperature results for GCUGC in 0.1 M salt buffer.	203
4.13 Chemical shift versus temperature results for GCAGC in 1.0 M salt buffer.	206
4.14 Chemical shift versus temperature results for GCUG in 0.1 M salt buffer.	210
4.15 Chemical shift versus temperature results for GCAG in 1.0 M salt buffer.	212
4.16 Chemical shift versus temperature results for GCA in 1.0 M salt buffer.	215
4.17 Temperature dependence of the anomeric proton coupling constants of GCAGC.	216
4.18 Temperature dependence of the anomeric proton coupling constants of GCUGC.	216
4.19 Chemical shift versus temperature results for GCAGC:GCUGC in 0.1 M salt buffer.	220
4.20 Chemical shift versus temperature results for GCAUGC in 0.1 M salt buffer.	223
5.1 Structures of common intercalators.	235
5.2 Changes in $T_m$ of GCGC upon addition of proflavine.	259
5.3 Changes in $T_m$ of GCUGC upon addition of proflavine.	259
5.4 Changes in chemical shift of G(1)H-8 of GCGC upon addition of proflavine.	261
5.5 Changes in chemical shift of C(2)H-6 of GCGC upon addition of proflavine.	261
5.6 Changes in chemical shift of G(3)H-8 of GCGC upon addition of proflavine.	262
5.7 Changes in chemical shift of C(4)H-6 of GCGC upon addition of proflavine.	262

Figure	Page
5.8 Changes in chemical shift of C(2)H-5 of GCGC upon addition of proflavine.	263
5.9 Changes in chemical shift of C(4)H-6 of GCGC upon addition of proflavine.	263
5.10 Changes in chemical shift of G(1)H-8 of GCUGC upon addition of proflavine.	264
5.11 Changes in chemical shift of C(2)H-6 of GCUGC upon addition of proflavine.	264
5.12 Changes in chemical shift of U(3)H-6 of GCUGC upon addition of proflavine.	265
5.13 Changes in chemical shift of G(4)H-8 of GCUGC upon addition of proflavine.	265
5.14 Changes in chemical shift of C(5)H-6 of GCUGC upon addition of proflavine.	266
5.15 Changes in chemical shift of C(2)H-5 of GCUGC upon addition of proflavine.	266
5.16 Changes in chemical shift of U(3)H-5 of GCUGC upon addition of proflavine.	269
5.17 Changes in chemical shift of C(5)H-5 of GCUGC upon addition of proflavine.	269
5.18 Changes in $T_m$ of CCGG upon addition of proflavine.	281
5.19 Changes in $T_m$ of CCAGG upon addition of proflavine.	281
5.20 Changes in chemical shift of C(1)H-6 in CCGG upon addition of proflavine.	282
5.21 Changes in chemical shift of C(2)H-6 in CCGG upon addition of proflavine.	282
5.22 Changes in chemical shift of G(3)H-8 in CCGG upon addition of proflavine.	283
5.23 Changes in chemical shift of G(4)H-8 in CCGG upon addition of proflavine.	283
5.24 Changes in chemical shift of C(1)H-5 in CCGG upon addition of proflavine.	284

Figure		Page
5.25	Changes in chemical shift of C(2)H-5 in CCGG upon addition of proflavine.	284
5.26	Characteristic chemical shift versus temperature curves for proflavine in oligoribonucleotide complex.	289
5.27	Potential intercalator binding sites in tetramers GCGC and CCGG.	290
5.28	Schematic diagram of possible imino proton spectra for intercalation complexes in slow exchange.	294
5.29	Potential intercalator binding sites in mismatch-containing pentamers GCUGC and CCAGG.	297

## LIST OF TABLES

Table		Page
1.1	Chemical shifts of non-exchangeable protons in the ribose monomers at 70°C.	14
2.1	Preparation of oligoribonucleotides.	37
2.2	Preparation of oligoribonucleotides.	38
2.3	Oligoribonucleotide deprotection yields.	45
2.4	Incremental assignments of proton NMR spectra of tetraribonucleotides GGCC, GCCG and GCGC.	51
2.5	Incremental assignments for tetraribonucleotides CCGG, CGGC and CGCG.	52
2.6	Incremental assignments for pentaribonucleotides GGUCC and GGACC.	53
2.7	Incremental assignments for pentaribonucleotides CCUGG and CCAGG.	54
2.8	Incremental assignments for pentaribonucleotides GCUGC and GCAGC.	55
3.1	Summary of assignments for tetraribonucleotides at 70°C.	69
3.2	Hydrogen-bonded imino proton chemical shift assignments for GGCC.	74
3.3	Hydrogen-bonded imino proton chemical shift assignments for CCGG.	75
3.4	Hydrogen-bonded imino proton chemical shift assignments for GCGC.	75
3.5	Hydrogen-bonded imino proton chemical shift assignments for CGCG.	80
3.6	Hydrogen-bonded imino proton chemical shift assignments for GCCG.	80

Table	Page
3.7 Hydrogen-bonded imino proton chemical shift assignments for GCGC:CGGC.	81
3.8 Chemical shift versus temperature data for GGCC in 1.0 M salt buffer.	85
3.9 Chemical shift versus temperature data for CCGG in 1.0 M salt buffer.	86
3.10 Chemical shift versus temperature data for GCCG in 1.0 M salt buffer.	90
3.11 Chemical shift versus temperature data for CGGC in 1.0 M salt buffer.	91
3.12 Chemical shift versus temperature data for GCGC in 1.0 M salt buffer.	95
3.13 Chemical shift versus temperature data for CGCG in 1.0 M salt buffer.	98
3.14 Chemical shift versus temperature data for CGCG in 0.1 M salt buffer.	100
3.15 Summary of Tms for tetranucleotides and trimer mixtures studied in section 1.	102
3.16 Chemical shift versus temperature data for GCCG:CGGC in 1.0 M salt buffer.	105
3.17 Chemical shift versus temperature data for GGC:GCC in 1.0 M salt buffer.	109
3.18 Chemical shift versus temperature data for GGA:CCA in 1.0 M salt buffer.	114
3.19 Chemical shift versus temperature data for GCG:CGC in 1.0 M salt buffer.	115
3.20 Chemical shift versus temperature data for CGG:CCG in 1.0 M salt buffer.	116
3.21 Chemical shift versus temperature data for GCG in 1.0 M salt buffer.	119
3.22 Hydrogen-bonded imino proton chemical shift assignments for GCG in 1.0 M salt buffer.	119

Table	Page
3.23 Chemical shift versus temperature data for GGA:CCA in no salt buffer.	122
3.24 Chemical shift versus temperature data for GCG:CGC in no salt buffer.	124
3.25 Chemical shift versus temperature data for CGG:CCG in no salt buffer.	127
3.26 Chemical shift versus temperature data for GCG in no salt buffer.	129
3.27 Comparison of Tms and chemical shift changes upon complexing for tetranucleotides and their deoxyribonucleotide analogues.	134
3.28 Comparison of observed and calculated hydrogen-bonded imino proton chemical shifts.	146
3.29 Hydrogen-bonded imino proton chemical shift assignments for AGGCU:AGCCU.	149
3.30 Hydrogen-bonded imino proton chemical shift assignments for AGGCU:AGUCU.	149
3.31 Hydrogen-bonded imino proton chemical shift assignments for AGACU:AGUCU.	150
4.1 Summary of published studies of base-base mispairing in oligonucleotide systems.	159
4.2 Chemical shift versus temperature data for GGCC in 0.1 M salt buffer.	170
4.3 Chemical shift versus temperature data for GGUCC in 0.1 M salt buffer.	172
4.4 Chemical shift versus temperature data for GGACC in 0.1 M salt buffer.	176
4.5 Chemical shift versus temperature data for GGAC in 0.1 M salt buffer.	178
4.6 Chemical shift versus temperature data for GGACC:GGUCC in 0.1 M salt buffer.	181
4.7 Hydrogen-bonded imino proton chemical shift assignments for GGACC:GGUCC.	182

Table	Page
4.8 Chemical shift versus temperature data for CCGG in 0.1 M salt buffer.	185
4.9 Chemical shift versus temperature data for CCUGG in 0.1 M salt buffer.	187
4.10 Chemical shift versus temperature data for CCAGG in 0.1 M salt buffer.	190
4.11 Chemical shift versus temperature data for CCAGG:CCUGG in 0.1 M salt buffer.	193
4.12 Hydrogen-bonded imino proton chemical shift assignments for CCAGG:CCUGG.	196
4.13 Chemical shift versus temperature data for CCAUGG in 0.1 M salt buffer.	198
4.14 Hydrogen-bonded imino proton chemical shift assignments for CCAUGG.	199
4.15 Chemical shift versus temperature data for GCGC in 0.1 M salt buffer.	202
4.16 Chemical shift versus temperature data for GCUGC in 0.1 M salt buffer.	204
4.17 Hydrogen-bonded imino proton chemical shift assignments for GCUGC.	205
4.18 Chemical shift versus temperature data for GCAGC in 1.0 M salt buffer.	207
4.19 Chemical shift versus temperature data for GCAGC in 0.1 M salt buffer.	208
4.20 Chemical shift versus temperature data for GCUG in 0.1 M salt buffer.	211
4.21 Chemical shift versus temperature data for GCAG in 1.0 M salt buffer.	213
4.22 Chemical shift versus temperature data for GCAGC:GCUGC in 0.1 M salt buffer.	221
4.23 Hydrogen-bonded imino proton chemical shift assignments for GCAGC:GCUGC.	222

Table	Page
4.24 Chemical shift versus temperature data for GCAUGC in 0.1 M salt buffer.	224
4.25 Hydrogen-bonded imino proton chemical shift assignments for GCAUGC.	225
5.1 Chemical shift versus temperature data for GCGC with proflavine, single strand:drug 15:1.	246
5.2 Chemical shift versus temperature data for GCGC with proflavine, 4:1.	247
5.3 Chemical shift versus temperature data for GCGC with proflavine, 2:1.	248
5.4 Chemical shift versus temperature data for GCGC with proflavine, 1:1.	249
5.5 Chemical shift versus temperature data for GCGC with proflavine, 1:2.	250
5.6 Chemical shift versus temperature data for GCUGC with proflavine, 20:1.	251
5.7 Chemical shift versus temperature data for GCUGC with proflavine, 6:1.	252
5.8 Chemical shift versus temperature data for GCUGC with proflavine, 4:1.	253
5.9 Chemical shift versus temperature data for GCUGC with proflavine, 2:1.	254
5.10 Chemical shift versus temperature data for GCUGC with proflavine, 1:1.	255
5.11 Chemical shift versus temperature data for GCUGC with proflavine, 1:2.	256
5.12 Summary of Tms of GCGC with proflavine.	257
5.13 Summary of Tms of GCUGC with proflavine.	258
5.14 Hydrogen-bonded imino proton chemical shift assignments of GCGC with proflavine.	268
5.15 Hydrogen-bonded imino proton chemical shift assignments of GCUGC with proflavine.	269



Table		Page
5.16	Chemical shift versus temperature data for CCGG with proflavine, 10:1.	270
5.17	Chemical shift versus temperature data for CCGG with proflavine, 4.6:1.	271
5.18	Chemical shift versus temperature data for CCGG with proflavine, 3.5:1.	272
5.19	Chemical shift versus temperature data for CCGG with proflavine, 2:1.	273
5.20	Chemical shift versus temperature data for CCGG with proflavine, 1:1.	274
5.21	Chemical shift versus temperature data for CCAGG with proflavine, 5:1.	276
5.22	Chemical shift versus temperature data for CCAGG with proflavine, 2:1.	277
5.23	Chemical shift versus temperature data for CCAGG with proflavine, 1:1.	278
5.24	Summary of Tms of CCGG with proflavine.	279
5.25	Summary of Tms of CCAGG with proflavine.	280
5.26	Hydrogen-bonded imino proton chemical shift assignments of CCGG with proflavine.	285
5.27	Hydrogen-bonded imino proton chemical shift assignments of CCAGG with proflavine.	286
5.28	Chemical shift changes upon intercalation for GCGC with proflavine, 2:1.	292
5.29	Chemical shift changes upon intercalation for CCGG with proflavine, 2:1.	292
5.30	Comparison of average Tms of internal and terminal protons for GCGC.	295
5.31	Comparison of average Tms of internal and terminal protons for CCGG.	295
5.32	Comparison of average Tms of internal and terminal protons for GCUGC.	299

## LIST OF ABBREVIATIONS

A <sup>3</sup>	adenosine
bz	benzoyl
C <sup>3</sup>	cytidine
COSY	2-D shift correlated NMR spectroscopy
Cu/Zn	copper-zinc couple
D <sub>2</sub> O	deuterium oxide
DHP	dihydropyran
DMF	N,N-dimethylformamide
DNA	deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulphonate
G <sup>3</sup>	guanosine
HPLC	high performance liquid chromatography
Hz	hertz (cycles/second)
J	coupling constant
me	methyl
mRNA	messenger RNA
MSNT	1-(2-mesitylenesulphonyl)-3-nitro-1,2,4-triazole
MST	1-(2-mesitylenesulphonyl)-1,2,4-triazole
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NSB	no sigmoidal behaviour

P	phosphate
P	2,2,2-trichloroethylphosphate
ppm	parts per million
pro	proflavine
pur	purine
pyr	pyrimidine
R <sub>f</sub>	ratio of distance travelled by solute to distance travelled by solvent
rRNA	ribosomal RNA
RNA	ribonucleic acid
T <sup>a</sup>	thymidine
t	tetrahydropyranyl
TIPS	1,1,3,3-tetraisopropylidisiloxane
TIPS-Cl	1,3-dichloro-1,1,3,3-tetraisopropylidisiloxane
tlc	thin layer chromatography
T <sub>m</sub>	melting temperature of a duplex, at which half the strands are in duplex form
TPS-NI	1-(2,4,6-triisopropylbenzenesulphonyl)-nitroimidazole
Trac	trityloxyacetyl
tRNA	transfer RNA
U <sup>a</sup>	uridine
UV	ultraviolet

<sup>a</sup> mononucleosides and oligoribonucleotides are abbreviated according to standard format. GpC represents guanylic(3'-5')cytidine. Oligoribonucleotides are written and numbered sequentially from the 5' terminus.

## 1. INTRODUCTION

In the forty two years since deoxyribonucleic acid was identified as the medium of transmission of genetic information, research into the structure and function of the nucleic acids has advanced until the major pathways and participants in the transmission and expression of genetic information have been identified. The structure of double helical DNA, the arrangement of its constituent aromatic bases, sugar moieties and inorganic phosphate groups, was elucidated in 1953 (Watson and Crick, 1953). The salient features of the model, with regard to the transmission of information, were the variability of the bases within an invariant framework--the double helical backbone. The specific geometry for hydrogen bonding between bases belonging to opposite strands allows each to be used as a template for the synthesis of a second set of information, which can then be divided between progeny during cell meiosis.

Genetic information was subsequently found to be encoded, linearly and contiguously, on sections of the chromosomal DNA; these coding units are termed genes. Expression of this information occurs through a transient intermediate messenger molecule, messenger ribonucleic acid (mRNA) (Jacob and Monod, 1961; Lewin, 1983). RNA differs from DNA only in that the sugar moiety is ribose, instead of deoxyribose, and the base uracil is methylated to form thymine; this permits transfer of information from DNA to RNA by a process analogous to that of replication, transcription. Three major types of RNA are produced, the aforementioned mRNA, information carrier, and ribosomal and transfer RNA (rRNA and tRNA), both of which are involved in the synthesis of proteins from mRNA, the final step in the expression of genetic information (Lewin, 1984).

Unlike DNA, RNA is not synthesized to have a complementary strand; it has regions of complementarity, and will self-associate and fold to form L-shaped tRNA (Rich, 1977, and other papers in same issue), and, in association with numerous proteins, globular ribosomes (Noller, 1984). The secondary and tertiary structure of mRNA may also be involved in regulation of processing and translation (Yanofsky, 1981; Gottesman, 1982; Ziff, 1980). In the absence of direct observation of the secondary structure, which has so far only been possible for the smallest natural RNA molecule, tRNA, various methods of prediction of secondary structures have been investigated, in the expectation that structure will give indications as to function. The study of short oligoribonucleotide single strands and duplexes, model sequences for the secondary structure elements in natural RNA, provides insight into the factors determining RNA secondary structure and stability.

### 1.1. Nucleic acid structure

A single strand of RNA consists of a backbone, with ribose residues connected to one another by phosphate diester linkages between the 5' hydroxyl of one sugar and the 3' hydroxyl of the next. The aromatic bases, purines adenine and guanine, pyrimidines cytosine and uracil, are attached by ring nitrogen N1 (pyrimidine) or N9 (purine) to the ribose 1' position (Figure 1.1). DNA differs from RNA in that the ribose moiety lacks a 2' hydroxyl group, and thymine replaces uracil in the bases.

The classic double helix, as proposed by Watson and Crick (1953), comprises two DNA molecules entwined to form a right-handed helix of diameter about 20Å. The

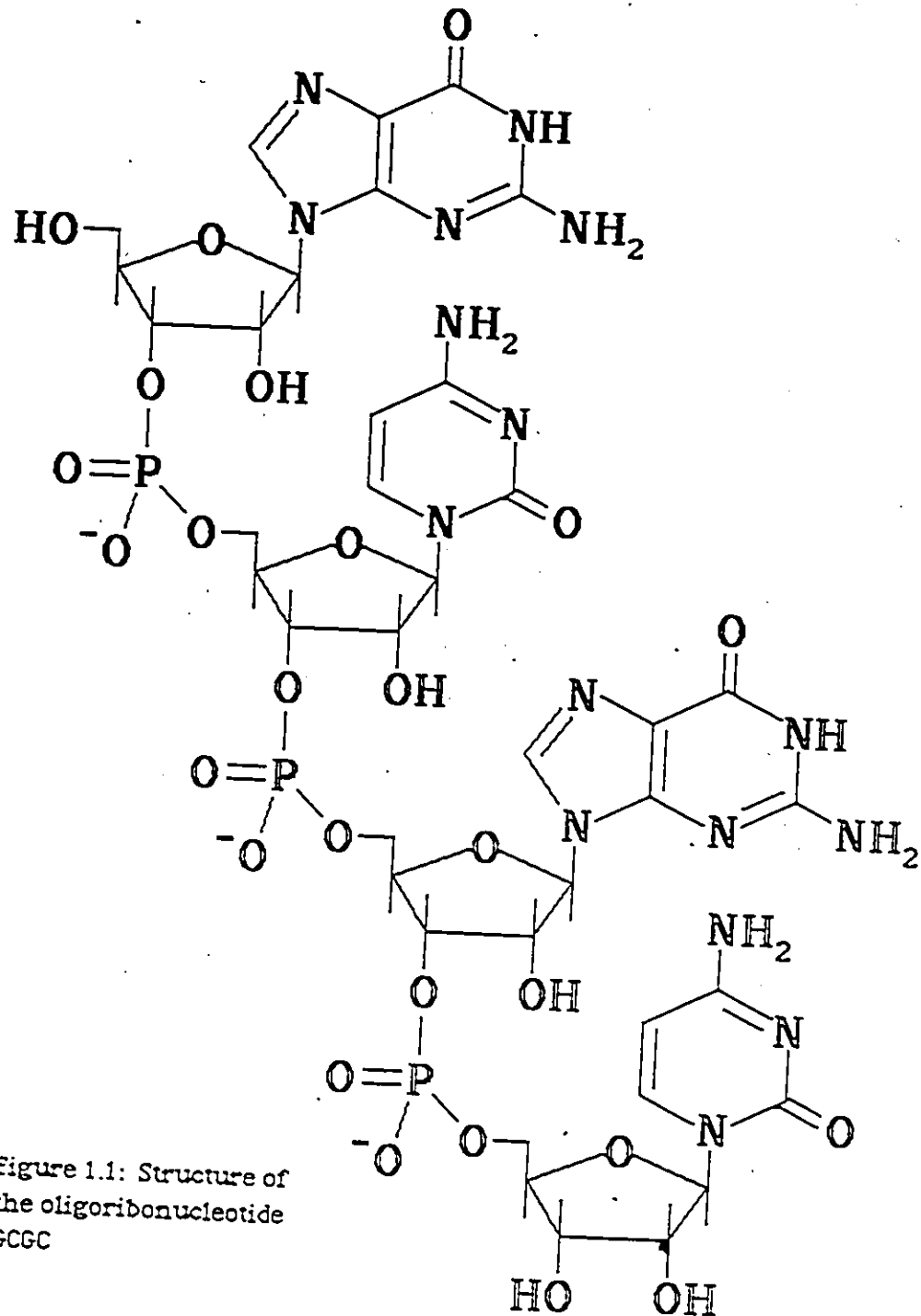


Figure 1.1: Structure of  
the oligoribonucleotide  
GCGC

sugar-phosphate backbones are on the helix exterior, and are aligned antiparallel to each other—the 3' terminus of one strand being opposite to the 5' terminus of the other (Watson and Crick, 1953; Saenger, 1984). The bases have their planes perpendicular to the long axis of the helix, and each is hydrogen-bonded to one on the opposite strand, with the normal pairings being G•C and A•U. (Figure 1.2) Ordering of the base pairs is unrestricted; it is in their order that the information resides and through the specificity of pairing during replication that it is propagated (Crick and Watson, 1953). About ten base pairs are necessary per turn, and base pairs are separated by 3.4Å. The helix coheres by hydrogen bonding and vertical hydrophobic stacking interactions between aromatic bases.

Fibre diffraction studies of DNA and synthetic polynucleotides and, later, single crystal structure determinations on tRNA and oligodeoxynucleotides, have resulted in the identification of three major classes of DNA helix, A, B (including the subclasses B', C, C', C'', D, E and T DNA) and Z, and two of RNA helix, A and A' (Saenger, 1984). Other structural parameters have also been introduced: Base pair tilt—inclination of the base pair long axis from the perpendicular to the helix axis; base pair roll—rotation of the base pair around its long axis; twist—rotation of the base pair in its plane around the helix axis; propeller twist—rotation of the two base planes relative to each other (Figure 1.3).

The DNA helix types are interconvertible through adjustment of humidity and salt conditions, but differ in backbone conformation, base-pair stacking, and, in one case, sense of winding, which results in different proportions for all three. By changing the sugar puckering from C2' endo, as in the B helix, to C3' endo, as in A-DNA and A- and A'- RNA, the phosphate groups move 1Å further apart, producing a squatter, thicker helix; the base pairs then tilt relative to the axis to improve

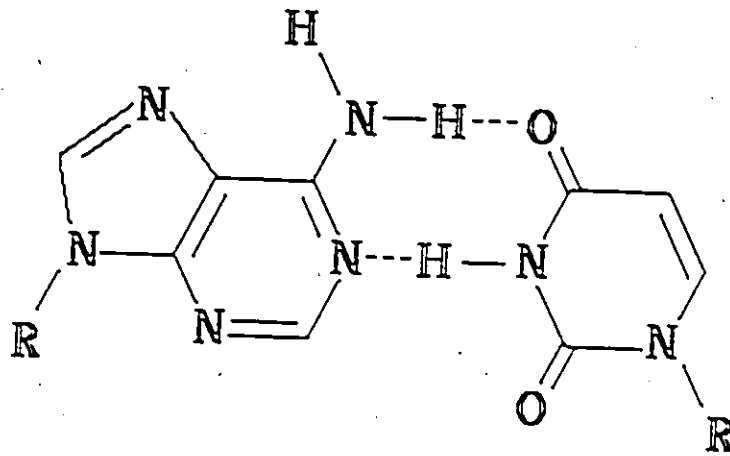
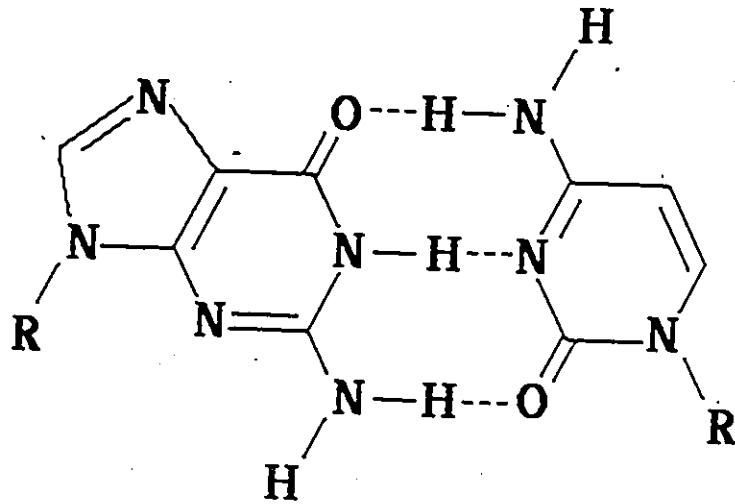


Figure 1.2: Normal base pairing: guanine-cytidine (top), adenine-uracil (bottom).



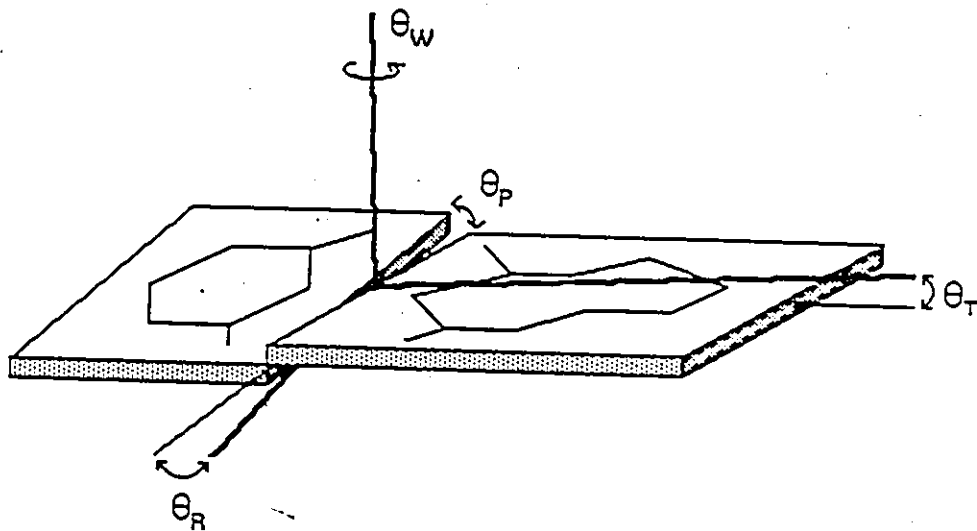


Figure 1.3: Structural parameters used to describe base pair conformation:  $\theta_T$  base pair tilt,  $\theta_R$  base pair roll,  $\theta_W$  base pair twist,  $\theta_P$  propellor twist.

stacking interactions, and the base pair separation is reduced to 2.8Å. The tilt of the base pairs in A-DNA (and A-RNA) results in a hollow centred helix, while the interior of the B-helix is filled in. The shift of the helix axis from between the base pairs in B-DNA to the major groove side in A-DNA (and A'-RNA), deepens the major groove, and reduces the depth of the minor groove. In both types of helix, the plane of each member of a base pair is twisted relative to that of its partner by 16 - 18° to improve stacking (Saenger, 1984).

In B- and A-DNA the conformation of the glycosidic linkage between base and sugar is *anti*, with the base pointing away from the sugar, whereas in left handed Z-DNA, a conformation peculiar to alternating pyrimidine-purine sequences in high salt solution, every purine residue assumes a *syn* conformation. Sugar puckering is mixed, with purine C3' endo and pyrimidine C2' endo, and the backbone forms a characteristic zig-zag pattern. The minor groove is deep and the major groove absent. Base pair tilt is smaller than in A-DNA, but present, but there is no propeller twist as it will not improve stacking (Saenger, 1984).

In contrast to the polymorphism of DNA, RNA exhibits two closely related double helical forms, A at low ionic strengths, A' at high. The A-RNA helix is similar to the A-DNA helix, with 11 base pairs per turn, a deep major groove and shallow minor groove. A'-RNA differs from A-RNA mainly in its number of base pairs per turn, 12, and its pitch (pitch being the height of a full turn of the helix), 36Å for A' RNA as opposed to 30Å for A-RNA. A-RNA has a smaller rise per residue, 2.73 - 2.81Å, and a larger base pair tilt angle, 16 - 19°, compared with 3.0Å and 10° in A'-RNA. The conformations of the individual nucleotides are the same in both (Saenger, 1984).

The previous descriptions are of average parameters; the single crystal structure solutions of deoxyoligonucleotides 2 - 12 base pairs in length (Berman and

Shieh, 1981; Kennard, 1984), have shown local variations from the average for the DNA helices in sugar puckering, turn angle, tilt and propeller twist, with apparent anticorrelation between members of a base pair: they deviate equally to opposite sides of the mean conformation (Dickerson, 1983). NMR studies have shown a similar, albeit subtler, variation of geometry in solution (Patel *et al.*, 1983; Clore and Gronenborn, 1985)

No oligoribonucleotides have as yet been crystallized. Details as to the solid-state helical forms of RNA have been extracted from crystallographic studies of tRNA, which has short (4 - 6bp) double helical sections (Rich, 1977).

## 1.2. NMR theory

Physical studies of nucleic acids have been concerned with two aspects, stability and structure, in the expectation that both might be related to function. Among the physical techniques available nuclear magnetic resonance (NMR) is unique in that it can provide information on local structure and stability of portions of the duplex.

The basis of NMR is the possession by certain nuclei of a magnetic moment. In free space the energy levels of this magnetic moment are degenerate, but upon application of a magnetic field,  $H$ , this degeneracy is removed, and they become separated by an energy of  $E = \gamma H/2\pi$ , in the case of nuclei with a nuclear spin of  $1/2$  ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{19}\text{F}$ ,  $^{31}\text{P}$ ), where  $\gamma$  is the magnetogyric ratio; at field strengths currently available this transition absorbs electromagnetic radiation in the radiofrequency region (Williams and Fleming, 1973; Jardetzky and Roberts, 1981). Proton NMR was the first to become well established on account of the high intrinsic sensitivity and

abundance of the magnetic isotope  $^1\text{H}$ .

The frequency at which a given nucleus absorbs can be modified by a number of factors (Williams and Fleming, 1973; Jardetzky and Roberts, 1981). Firstly, electrons have a shielding influence, and the electron density around the nucleus reduces the effective field seen by that nucleus, causing it to resonate at lower frequencies. For protons the electron density is primarily dependent upon the electronegativity of the atom to which the proton is bonded. Secondly, aromatic rings have strong shielding and deshielding effects, due to the induced ring currents in the  $\pi$ -bonding system. The effect above and below the ring is shielding, causing any protons within this volume to absorb at lower frequency. In the plane of the ring the effect is deshielding, causing a proton to resonate at higher frequency. Thus the transition from a random coil to a stacked coil may be monitored because the chemical shifts change with the changing influence of the neighbouring bases. Thirdly, protons separated by two to five bonds may interact with each other (through the electrons in the bonds) and affect the fields seen by each other, and therefore the frequency at which each resonates. For the pyrimidine ring system, for instance, there are two protons attached to adjacent carbon atoms. Depending upon the orientation of its spin, each nucleus will add or subtract a small component from the field experienced by the other, resulting in the splitting of both signals into doublets, with equal separation of both lines.

More detailed information in the interaction between magnetic nuclei has become available with the advent of Fourier Transform NMR (FT-NMR) spectroscopy. Instead of scanning through all frequencies successively and observing absorbance at all frequencies, a single strong radiofrequency pulse is applied at the start of the experiment, exciting all nuclei. As these nuclei return to their equilibrium state,

each re-emits its own characteristic frequency; these frequencies are collected as a decay signal over time, and Fourier transformed to yield a normal spectrum. Within the original decay signal—collected over time—is information as to the lifetimes of the magnetized states of each proton. These lifetimes depend upon, among other things, the size of the molecule, the flexibility of the moiety to which the proton is attached, the proximity of other magnetic nuclei, and exchange with the solvent; they can be interpreted to provide information on structure and flexibility (Jardetzky and Roberts, 1981; Kearns, 1984).

Fourier transform techniques also provide for specific excitation of one proton, with the intention of suppressing its signal (solvent suppression), suppressing its coupling to another nucleus to allow identification of that nucleus (decoupling), or perturbing another nucleus with which it exchanges magnetization (Nuclear Overhauser Effect, NOE). Since the efficiency of through space exchange of magnetization is dependent upon separation, the NOE has proven extremely useful in structural analysis since it can be used to identify close contacts ( $<3.4\text{\AA}$ ) between protons in the helix (Clare and Gronenborn, 1985).

A further level of sophistication is available through the use of multiple pulse experiments, whereby the nuclei are perturbed by a series of pulses separated by delay times which allow for interaction between the nuclei, transfer and exchange of magnetization (Feigon *et al.*, 1982a; Kearns, 1985). Variation of the lengths of these delays allows the time-course of these interactions to be followed, yielding intensity vs. time data in two dimensions, which may then be Fourier transformed to provide frequency information in two dimensions. 2-D shift-correlated spectroscopy (COSY) (Aue *et al.*, 1976; Feigon *et al.*, 1983a,b) and spin-echo correlated spectroscopy (SECSY) (Nagayama *et al.*, 1979; Lankhorst *et al.*

1985), which provide information on coupled spin systems, and 2-D NOE (NOESY) (Macura and Ernst, 1980; Jamin *et al.*, 1985) spectroscopy, which provides information on close contacts within the molecule, have all been applied to the assignment of resonances and the analysis of structure. Other potentially powerful techniques include relay coherence transfer spectroscopy (RELAY) (Bolton, 1982; Eich *et al.*, 1982), which ideally should allow identification of all signals belonging to a single ribose residue, and double quantum spectroscopy (DOUBTFUL) (Hore *et al.*, 1982; Hughes *et al.*, 1984), which is another technique for identification of coupled resonances. Pulse sequences have also been devised which allow suppression of solvent peaks, particularly for the observation of the hydrogen-bonding protons in water; the pulse may be manipulated so as to create a zero spectral density at the frequency of the undesired peak (Clare and Gronenborn, 1983a).

### 1.3. $^1\text{H}$ -NMR of nucleic acids

#### 1.3.1. Assignment of single strands and duplexes

The most immediate challenge faced in the analysis of  $^1\text{H}$ -NMR spectra of oligonucleotides is the accurate assignment of the signals, for on this depends correct interpretation of structure. Each aromatic base has one or two non-exchangeable protons and at least one exchangeable proton which may be observed in water when hydrogen bonding slows the exchange rate (Figure 1.4a): guanine H-8, H-1 (imino),  $\text{NH}_2$ , adenine H-8, H-2,  $\text{NH}_2$ , cytosine H-6, H-5,  $\text{NH}_2$ , uracil H-6, H-5, H-1 (imino),  $\text{NH}_2$ . The non-exchangeable and imino base protons were the first to be observed and monitored through the single strand to helix transition in r(AAGCUU) (Borer *et al.*,

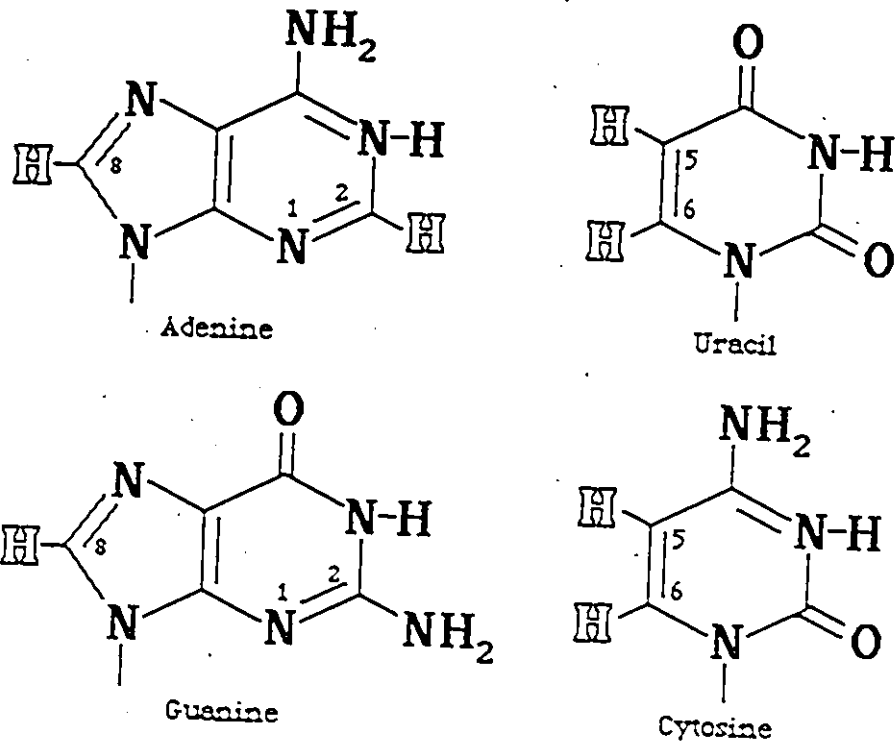


Figure 1.4a: Exchangeable (solid) and nonexchangeable (outline) protons in the aromatic bases.

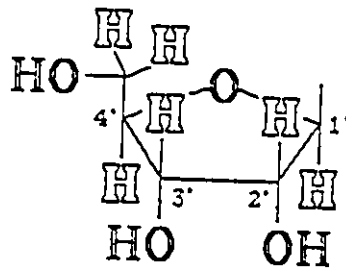


Figure 1.4b: Nonexchangeable protons belonging to the ribose ring (outline).

1975) and r(CCGG) (Arter *et al.*, 1974). Each ribose moiety contributes six additional signals, belonging to H1', H2', H3', H4', H5' and H5'' (Figure 1.4b); deoxyribose has an additional proton at H2''. The assignment of certain of the ribose protons has proven refractory due to extensive signal overlap, while those in deoxyribose are better dispersed; in part for this reason much of the studies on longer sequences have been performed on deoxyoligonucleotides.

Each of the base and ribose signals has a characteristic chemical shift and coupling pattern in the monomer (Table 1.1), but the addition of neighbouring bases changes the magnetic environment experienced by these protons (Borer *et al.*, 1974; Everett *et al.*, 1980). The signals of all protons shift, since even in a random coil conformation the protons will experience some shielding or deshielding effect from as far away as their third neighbours. The earliest assignments, generally of the base and anomeric (H1') protons alone, depended upon the comparison of related series of compounds, for instance the precursors in a stepwise chemical synthesis, to follow the movement of the resonances with addition of each new residue (Borer *et al.*, 1974; Everett *et al.*, 1980; Bell *et al.*, 1981). Later these effects were reduced to a general series of parameters, so that the chemical shift of a proton in any single strand could be calculated by computer summation of the effects of the neighbouring residues (Hader *et al.*, 1981). Selective deuteration has been employed for the assignment of the oligoribonucleotide AAA (Kondo *et al.*, 1975). Initially the NMR instruments were of too low field to permit resolution of the ribose region, but as instrumentation improved, small oligomers could be at least partially assigned by stepwise homodecoupling experiments, with verification by computer simulation of



Table 1.1: Chemical shift assignments and coupling constants for the nonexchangeable base and anomeric protons of the nucleosides at 70°C, 7.3mM, in 1.0M salt buffer, relative to DSS.

Adenosine			Guanosine		
proton	$\delta$ (ppm)	$J_{5'6'}$ or $J_{1'2'}$ (Hz)	proton	$\delta$ (ppm)	$J_{5'6'}$ or $J_{1'2'}$ (Hz)
AH-8	8.313	—	GH-8	7.961	—
AH-2	8.260	—			
AH-1'	6.065	5.6	GH-1'	5.888	5.6
Cytidine			Uridine		
proton	$\delta$ (ppm)	$J_{5'6'}$ or $J_{1'2'}$ (Hz)	proton	$\delta$ (ppm)	$J_{5'6'}$ or $J_{1'2'}$ (Hz)
CH-6	7.781	7.3	UH-6	7.798	8.1
CH-1'	5.894	3.7	UH-1'	5.887	4.9
CH-5	6.073	7.3	UH-5	5.907	8.1

the spectra (Sarma and Dhingra, 1981; Lee, 1983).

With the advent of high resolution NMR and multiple pulse techniques it became possible to assign an oligomer without reference to precursors. The members of a ribose spin system may be identified by spin correlation experiments, such as COSY, SECSY and RELAY (Mellema *et al.*, 1984; Lankhorst *et al.*, 1985). Their association with particular base signals may be determined by observation of NOEs between base and ribose protons (Petersheim and Turner, 1983a; Frechet *et al.*, 1983). The assignment of a particular set of base and ribose signals to a residue in the single strand can be achieved by the observation of inter-residue NOEs, which allow identification of residues adjacent to each other. The longest single stranded oligoribonucleotide to have been completely assigned is AUAUUAU (Lankhorst *et al.*, 1985) while other sequences to have been assigned are at the trimer to pentamer level (Hartel *et al.*, 1981; Lankhorst *et al.*, 1982; Doornbos *et al.*, 1983; Gronenborn *et al.*, 1983).

An alternative approach, one applied to the assignment of longer fully duplexed DNA sequences, is to assign the duplex rather than the single strand, again through spin correlation experiments, and, especially, NOEs, since the duplex provides an intricate network of direction-specific NOEs, which, ideally, allow one to "walk through" the molecule in question (Frechet *et al.*, 1983; Clore and Gronenborn, 1985; Gronenborn and Clore, 1985). DNA sequences assigned in this fashion include:  $d(\text{CGCGAATTCGCG})_2$  (Hare *et al.*, 1983),  $d(\text{CGCGTATACGCG})_2$ ,  $d(\text{CGCGATATCGCG})_2$  (Patel *et al.*, 1985a),  $d(\text{CGATTATAATCG})_2$  (Patel *et al.*, 1985b),  $d(\text{CGCGCG})_2$  (Frechet *et al.*, 1983),  $d(\text{ATATCGATAT})_2$  (Feigon *et al.*, 1982, 1983b),  $d(\text{ATATGCATAT})_2$  (Feigon *et al.*, 1983a),

d(CCAATTCC)<sub>2</sub> (Broido *et al.*, 1984); mixed RNA:DNA duplexes r(GCG)d(TATACGC)<sub>2</sub> (Mellema *et al.*, 1983) and d(TCACAT):r(AUGUGA)<sub>2</sub> (Reid *et al.*, 1983) and the RNA duplex r(CUGUG):r(CACAG) (Clare *et al.*, 1985). Observation of NOEs between the exchangeable (imino) protons, have enabled complete assignment of the low field proton spectra of tRNA<sup>phe</sup> (Hare *et al.*, 1985) and tRNA<sup>Val</sup> (Hyde and Reid, 1985).

Whether assignment is of the single or double strand, the assignments can then be applied at other temperatures by following the change of chemical shift with temperature for the individual protons.

### 1.3.2. Conformational analysis of the single strand.

The conformation of every residue in a nucleic acid strand is defined by six backbone torsion angles,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$  and  $\zeta$ , one base-torsion angle  $\chi$  (IUPAC, 1983) (Figure 1.5), and the sugar puckering geometry (five torsion angles). Some, though not all, of these angles may be determined through analysis of three bond <sup>1</sup>H-<sup>1</sup>H and <sup>31</sup>P-<sup>1</sup>H couplings. For <sup>1</sup>H-<sup>1</sup>H vicinal coupling, the relationship between  $J_{HH}$  and torsion angle  $\phi$  is given by  $J_{HH} = A \cos \phi + B$  (Karplus, 1969) where A and B are empirically determined constants. Various modifications to this equation have been proposed, incorporating an additional  $\cos \phi$  dependence (Altona and Sundgralingham, 1972, 1973), and the effects of electronegative substituents on coupling constants (Haasnoot *et al.*, 1980).

The angles  $\alpha$  and  $\zeta$  cannot be directly determined from coupling constants, but the <sup>31</sup>P chemical shift has been correlated with the values of these angles

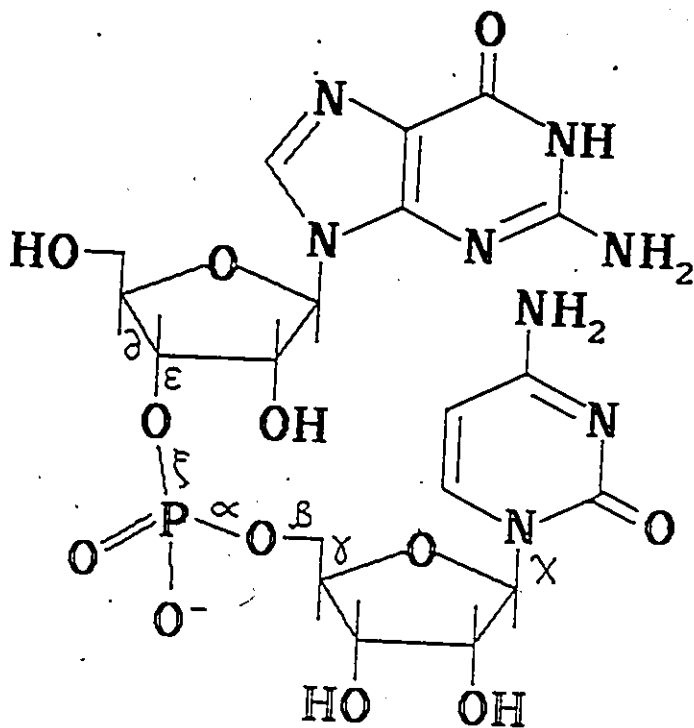


Figure 1.5: IUPAC-IUB nomenclature for the torsion angles in a dinucleotide unit.

(Gorenstein, 1975; Gorenstein *et al.*, 1976).  $\beta$  (C4'-C5'-O5'-P) is best monitored by  $^{31}\text{P}$ - $^1\text{H}$  couplings between ribose H5' and phosphorus attached to O5',  $^3J_{5'\text{P}}$  and  $^3J_{5'\text{P}}$ ,  $\gamma$  (C3'-C4'-C5'-O5') by  $^3J_{4'5'}$  and  $^3J_{4'5''}$ , and  $\epsilon$  (C3'-C4'-C5'-O5') by  $^3J_{3'\text{P}}$  (reviewed, Sarma and Dhingra, 1980). The conformations obtained from the coupling constants are generally described as mixtures of the populations of classical rotamers (Figure 1.6), by equations such as  $p(\gamma^-) = (13.3 - (J_{4'5'} + J_{4'5''}))/9.7$  (Lee and Sarma, 1976), where  $p(\gamma^-)$  is the percentage population of  $\gamma^-$  conformer (see Figure 1.6), and the constants have been experimentally determined. The results of classical conformational analysis are qualitative, and dependant upon the accuracy of the determination of the limiting values for the coupling constants. Particular combinations of conformations around two adjacent torsion angles allow for the observation of long range couplings between protons linked by a coplanar series of bonds: Coupling between P and H4' is observable when  $\gamma$  and  $\beta$  are  $g^+$  and  $t$  respectively, and coupling between P and H2' appears when the sugar ring is S (approximately 3' endo) and  $\epsilon$  is  $g^-$ . All three angles,  $\beta$ ,  $\gamma$  and  $\epsilon$ , demonstrate conformational preference (Figure 1.6):  $\beta^t \approx 180^\circ$ , over a narrow range, under all temperatures and sugar conformations,  $\gamma^+$  predominates in the stacked strand, with increasing populations of  $\gamma^-$  with increasing temperature (Altona, 1982; Lee, 1983). The conformation around  $\epsilon$  cannot be directly determined from the single coupling  $^3J_{3'\text{P}}$ , but by observation of a long-range coupling it is deduced to be  $\epsilon^t$  in the stacked state.

An accurate value for  $\delta$  may be obtained from its correlation with the sugar ring geometry (see below).

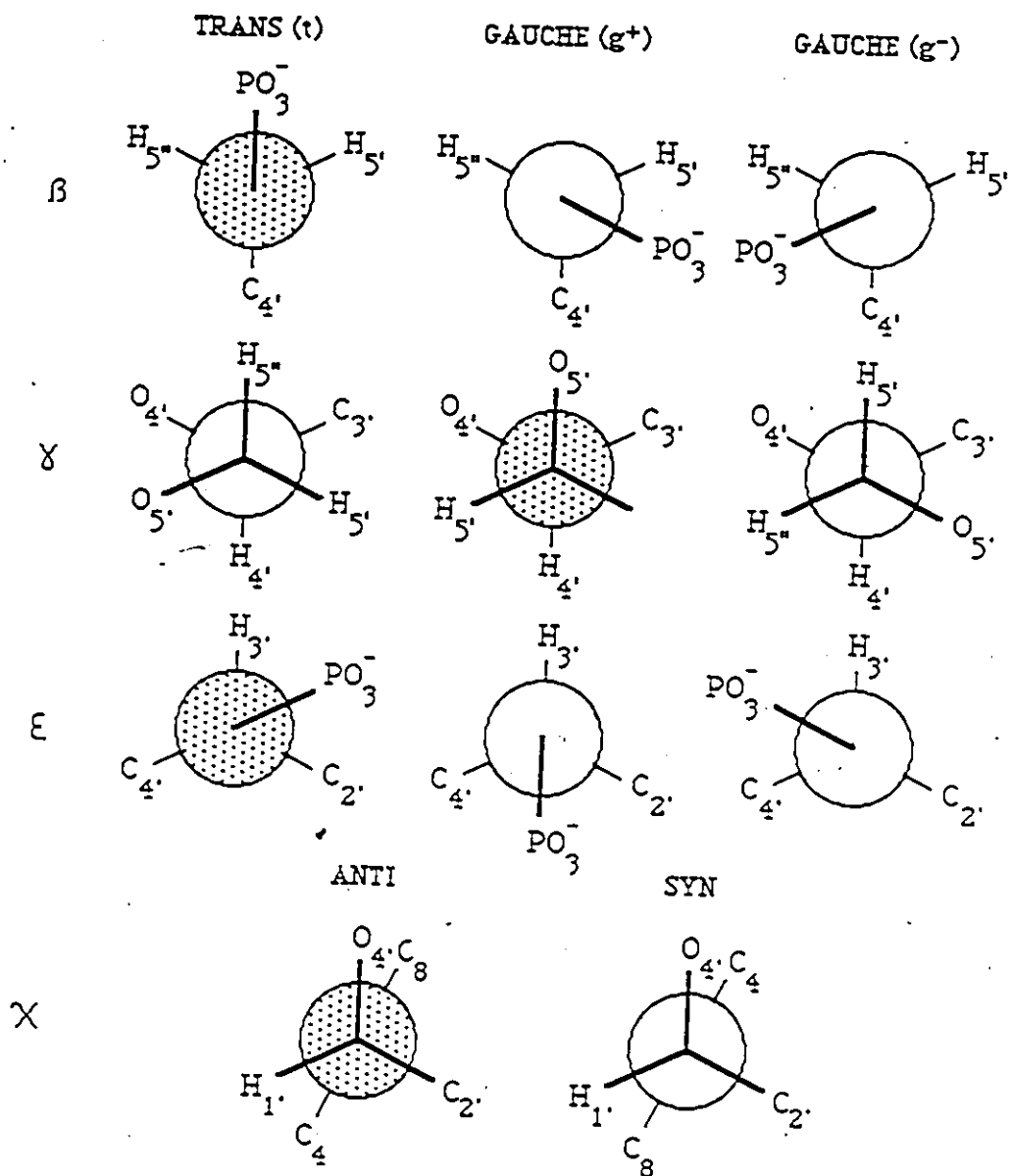


Figure 1.6: Idealized Newman projections for the classical rotamers for the backbone torsion angles  $\beta$ ,  $\gamma$ ,  $\epsilon$  and  $\chi$ . Conformers preferred in the right handed stack are shaded.

The base-base torsion angle  $\chi$  may theoretically be determined from the four-bond  $^1\text{H}$ - $^{13}\text{C}$  couplings  $^4J_{\text{H1}'\text{C6}}$  and  $^4J_{\text{H1}'\text{C2}}$  (for pyrimidine), and  $^4J_{\text{H1}'\text{C8}}$  and  $^4J_{\text{H1}'\text{C4}}$  (for purine) (Sarma and Dhingra, 1981; Altona, 1982). In certain pyrimidines, a five bond coupling between H6 and H1' has been observed. In the single stranded stacked conformation  $\chi$  is exclusively anti, while in the random coil, there is an appreciable population of the syn conformer, particularly for purines.

Various approaches have been developed for the analysis of sugar geometry, all of which have had to address the problem of defining five torsion angles from three coupling constants. Fortunately in the cyclic ribose system the torsion angles are interdependent, and knowledge of the values of three torsion angles permits calculation of the other two (Altona *et al.*, 1982). Ribose geometry is commonly described as an equilibrium between C2' endo and C3' endo puckered conformers (Figure 1.7), in which either the 2' or the 3' carbon is displaced towards C5' from the mean plane formed by the other atoms (Evans and Sarma, 1974; reviewed, Sarma, 1980), or alternately between S and N conformers (Altona and Sundaralingham, 1972; reviewed, Altona, 1982) (Figure 1.7). According to the latter approach the conformation of the ring is represented by a continuum of displacements of each successive ring carbon from the plane; this is commonly depicted as a wheel (Saenger, 1984). The two preferred conformations of the ribose ring fall into the northern (N) and southern (S) hemispheres. A pseudorotation phase angle  $P$  may be defined as

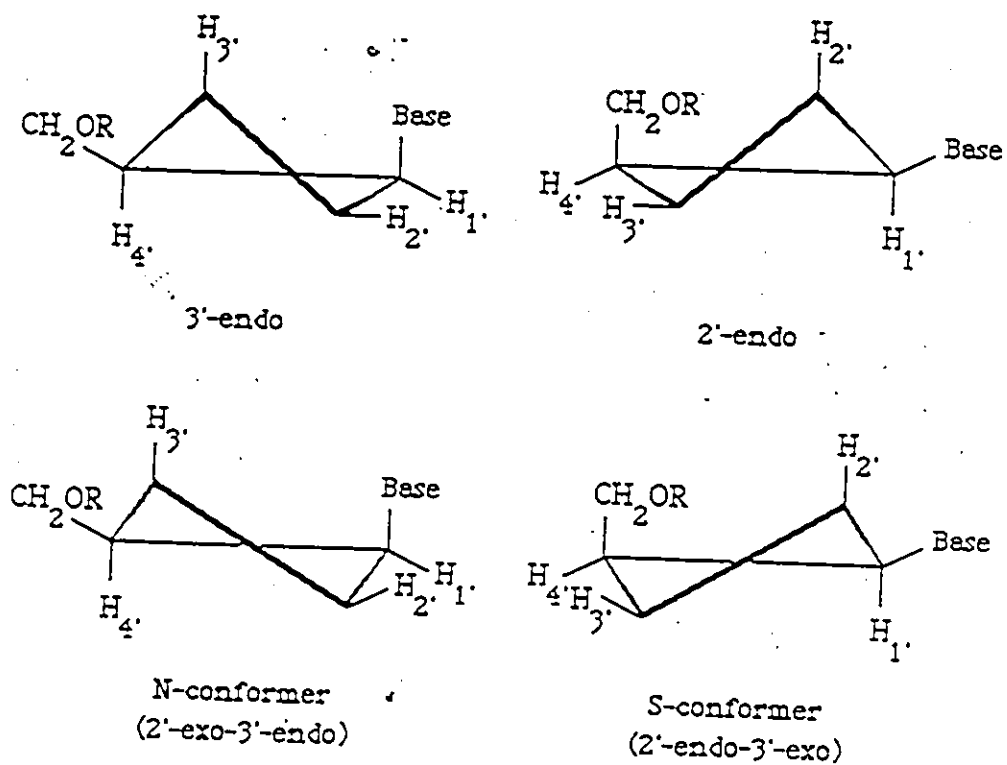


Figure 1.7: The two pairs of idealized conformers used in analysis of ribose conformation: 3'-endo and 2'-endo, and N- and S- conformers.



$$\tan P = \frac{(\tau_4 + \tau_1) - (\tau_3 + \tau_0)}{2\tau_2 (\sin 36^\circ + \sin 72^\circ)} \quad (+180 \text{ for negative } \tau_2).$$

$\tau_0 - \tau_4$  ring torsion angles.

The maximum puckering amplitude  $\tau_m$  is defined by  $\tau_2 = \tau_m \cos P$ .  
(Saenger, 1984)

The phase angle and amplitude of pucker may then be used to calculate the value of  $\vartheta$  from  $\vartheta = 120.6 + 1.1 \cos \varphi (P + 145.2)$  (Altona, 1982).

Early application of detailed conformational analysis included fifteen diribonucleotide monophosphates and fifteen deoxydiribonucleotide monophosphates (Sarma, 1981, and references therein). Stacked diribonucleotide diphosphates demonstrate a  ${}^3E$  preference, and stacked deoxydiribonucleotide diphosphates a  ${}^2E$  preference, which is the same preference as is observed in the double helix. Extension of these studies to the trimers AAA (Dhingra and Sarma, 1979), ACC and CCA (Cheng et al., 1980) determined that the conformational preference at low temperatures was a right handed stack, but, for CCA, species in which the central nucleotide in the trimer is excluded from stacking between the end units contribute to the conformational mix (Cheng et al., 1980). Similar "bulged" conformations have been observed for members of a set of thirteen oligoribonucleotides, particularly for those of the form pur-pyr-pur, in which the central pyrimidine is forced out from the pur-pur stack, with a concurrent increase in flexibility of the ribose ring, and higher populations of the "unstacked"  ${}^2E$  conformation (Lee and Tinoco, 1981; Lee, 1983).

Altona and coworkers (reviewed in Altona, 1982) have lately incorporated a generalized Karplus equation (one accounting for the effects of electronegative

substituents upon coupling constants) into a numerical algorithm which generates amplitudes and phases of pucker for each of the two extreme ribose conformations N and S, and calculates an equilibrium constant K for their interconversion. This method has been applied to the single stranded series  $m^6_2AUm^6_2A$  (Lankhorst *et al.*, 1982),  $m^6_2AUm^6_2AU$  (Hartel *et al.*, 1981),  $m^6_2AUm^6_2AUm^6_2A$  (Lankhorst *et al.*, 1983), and to AUAUUAU in its single stranded form (Lankhorst *et al.*, 1985), with confirmation of the earlier observation of a preferred 1,3 purine-purine stacking with bulged pyrimidine in oligoribonucleotides with alternating pyrimidine-purine sequence (Lee and Tinoco, 1980; Lee 1983). Analogous deoxyoligonucleotides d(ATAT) and d(TATA) show no such pattern (Mellema *et al.*, 1984).

Comparison of the results of conformational analysis of short oligomers of RNA and DNA reveal some further differences (Altona, 1982, review). For RNA oligomers the amplitude of sugar pucker for the N conformer is the same for stacked and unstacked strands, whereas this is not the case for the S conformer of stacked and unstacked DNA oligomers. Stacked DNA shows a 5' - 3' transmission effect in ribose conformation: pyrimidine residues 3' to purines favour the S conformer. Stacked RNA shows no apparent sequence effects. RNA backbone angles seem to undergo minimal changes upon formation of the duplex from the stacked single strand, whereas for DNA larger changes seem to be necessary, at least for the compounds studied (Altona, 1982).

### 1.3.3. Monitoring the single strand to duplex transition.

Any observable parameter whose value is different for single strand and duplex may be used to monitor the single strand to duplex transition,  $2M \rightleftharpoons M_2$ . UV absorbance, circular dichroism, fluorescence and proton chemical shifts have all been used. If a proton resonates at  $\delta M$  in the single strand and at  $\delta M_2$  in the helix then the observed frequency  $\delta_{obs}$  in the limit of fast exchange will be:

$$\delta_{obs} = XM\delta M + XM_2\delta M_2 = \delta M + (\delta M_2 - \delta M)XM_2$$
 $XM$  and  $XM_2$  being the mole fractions of the proton in the helix and single strand respectively. If the equilibrium constant for the process is  $K$  and the initial concentration  $M_0$ , then substitution for  $XM_2$  yields:

$$\delta_{obs} = \delta M + (\delta M_2 - \delta M) \left[ \frac{1 + 4KM_0 + \sqrt{(1 + 8KM_0)^2 - 4KM_0}}{4KM_0} \right]$$
 (Cantor and Schimmel, 1981).

$K$  is temperature dependent:  $K = -RT \ln \Delta G$ ; this dependence results in sigmoidal variation of  $\delta_{obs}$  with temperature between the extrema represented by  $\delta M$  and  $\delta M_2$ . If, by analogy with a true phase change, a melting temperature  $T_m$  is defined to be that temperature at which  $\Delta G = 0$ ,  $K = 1$ , then  $T_m = \Delta H^\circ / \Delta S^\circ$ , and may be determined as being the inflection point of the plot of  $\delta_{obs}$  vs  $1/T$ , and is an indicator of stability (Borer et al., 1974; Cantor and Schimmel, 1981). Thermodynamic data may be determined directly from the concentration dependence of the melting temperature as being

$$(T_m)^{-1} = (2.303R / \Delta H^\circ) \log C_T + \Delta S^\circ / \Delta H^\circ$$
 where  $C_T$  is total oligomer concentration (Borer et al., 1974). This approach has been applied to the

determination, largely from optical data, of the thermodynamic parameters for perfect and imperfect duplexes (Borer *et al.*, 1974; Freier *et al.*, 1985a), and of their constituent structural elements (nearest neighbour elements, mismatches) (Borer *et al.*, 1974; Freier *et al.*, 1985b; Hickey and Turner, 1985; Freier *et al.*, 1984). Real systems deviate from the simple two state model by fraying, particularly of terminal A•U and A•T base pairs (Borer *et al.*, 1974; Patel and Hilbers, 1975), multistranded aggregation (Sagan and Walmsley, 1985; Hopkins, 1984), particularly of GC rich sequences, premelting transitions (Patel *et al.*, 1982a), and base stacking prior to duplexing (Alkema, 1982). Thermodynamic parameters calculated from NMR data for a single sequence can show a wide variation in the processes observed by each individual proton (Cantor and Schimmel, 1981).

Until recently measurement of the thermodynamics of the helix-coil transition relied upon measurement of a wide range of concentrations, a time-consuming procedure for the NMR spectroscopist, for whom the available concentrations are limited by instrument sensitivity at the lower end and by aggregation effects at the upper. Variable temperature NMR spectroscopy has been used primarily to monitor the structural details of the transition and to correlate assignments made for the single strand to the duplex resonances and vice-versa (Romaniuk *et al.*, 1979; Alkema *et al.*, 1982; Patel *et al.*, 1982a). This may change with increasing access to high-field instruments and computer systems, leading to developments such as the differential concentration profiles method (DCPM), which uses computer least squares analysis of the chemical shift behaviour to isolate thermodynamic parameters for both stacking and self-association, each process

being described by a two-state model (Hartel *et al.*, 1982). Computer fitting of models of binding site affinities are being used to study the binding of intercalating drugs to duplexes with multiple binding sites (Nelson and Tinoco, 1984, 1985) (See section 5).

#### 1.3.4. Conformational analysis of the duplex.

Analysis of coupling constants is the preferred approach to conformational analysis of the single strand, but duplexing is frequently accompanied by broadening of NMR signals, due to chemical exchange and slowing of tumbling, resulting in loss of coupling information. Although analysis of coupling constants has been applied to some systems (Sanderson *et al.*, 1983), the most usual approaches to analyses of duplex conformation rely upon analysis of chemical shifts and NOE measurements.

Analysis of conformation by chemical shift involves the testing of shifts computed for a given geometry against observed shifts. Giessner-Prettre and Pullman (1965, 1969, 1970) calculated the ring currents in the four bases, and from these derived isoshielding contours at 3.4Å above and below the rings (Giessner-Prettre and Pullman, 1976a,b). Borer *et al.* and Arter *et al.* used the ring current results and helix parameters derived from fibre diffraction data to calculate the expected values for chemical shifts upon duplexing in AAGCUU (Borer *et al.*, 1974) and CCGG (Arter *et al.*, 1974), and to determine that both duplexes were A-type. Arter and Schmidt subsequently published extensive tables of parameters for the calculation of chemical shifts for any proton in an A, A' or B helix (Arter and Schmidt, 1976). When tested against the available data, these parameters showed -

reasonable to poor agreement--poor when applied to GC rich sequences, which was tentatively explained by aggregation, and by the contributions of the anisotropies of the base functional groups and the phosphate (Arter and Schmidt, 1976). Patel employed a similar approach to determine that the three deoxytetranucleotides d(GGCC), d(CCGG), and d(GCGC) formed B-type duplexes (Patel, 1977, 1979).

Chemical shift changes upon duplexing have been used to argue for the stacking in or bulging out of mismatched or non-paired bases (Alkema *et al.*, 1982; Patel *et al.*, 1982b; Haasnoot *et al.*, 1979): The H-8 and H-2 base protons of unpaired A(4) in the duplex formed by d(CGCAATTCGCG) show comparable upfield shifts and melting temperatures to the base paired adenosines A(6) and A(7) (Patel *et al.*, 1982b).

Computer fitting of chemical shift data has been used to adjust the base pair overlaps in the A-RNA helix to derive a model in which the turn angle varies with the sequence (Cruz *et al.*, 1982; Bubenko *et al.*, 1983) (See section 3.1.2).

The relative intensities of intra- and intermolecular NOEs, indicative of interproton distance relationships, may be used to determine the type of helix adopted. In particular it has been used to distinguish between right and left handed conformations, to monitor the B to Z transition in DNA oligomers (Tran-Dinh *et al.*, 1984; Feigon *et al.*, 1984) and to resolve whether the unusual circular dichroism spectrum of r(CGCGCG) results from a low salt Z form (Uesugi *et al.*, 1984; Westerink *et al.*, 1984)--which it does not. In Z-DNA, where the purine residues have adopted a *syn* conformation with 3' endo sugar pucker, the relative intensities of the intranucleotide NOEs are  $N_{H1'-HS} \gg N_{H2'-HS} = N_{H2''-HS}$ , whereas in B-DNA, with anti and the sugar 2' endo  $N_{H2'-HS} \gg N_{H1'-HS}$  (Gronenborn and Clore, 1985). The two

right handed helices may be distinguished from each other by the relative intensities of the NOEs between the base protons and the H2' and H2'' protons on adjacent nucleotides. Thus the duplexes formed by d(ATATCGATAT) (Feigon *et al.*, 1982), d(CGTACG) and d(ACGCGCGT) (Gronenborn and Clore, 1985) have been shown to be B-form in solution while r(CGCGCG) (Uesugi *et al.*, 1984; Westerink *et al.*, 1984) and r(CACAG):r(CUGUG) (Clore *et al.*, 1985) are A-form.

More detailed conformational analysis is possible if a reference distance is available for calibration of the NOEs, for nucleic acids, the distances between members of the pyrimidine H-5 - H-6 system, and the deoxyribose H-2' - H-2'' may be used. Then the distance  $r_{ij}$  between any two atoms  $i, j$ , is given by  $\langle r_{ij}^6 \rangle^{-1/6} = (\sigma_{H5H6} / \sigma_{ij})^{-1/6} r_{H5H6}$  where  $\sigma_{H5H6}, \sigma_{ij}$  are the cross relaxation rates between H5, H6 and  $i, j$  (Gronenborn and Clore, 1985). Direct structure determination by NOEs is not possible because they are far fewer in number than the degrees of freedom of the duplex. The favoured approach has been to use NOE data to computer refine structures derived from classical geometry. Detailed structural studies of this type have been performed on d(CGTACG) (Clore *et al.*, 1985), d(ACGCGCGT) (Clore and Gronenborn, 1984a, 1984b) and CACAG:CUGUG (Clore *et al.*, 1985). Local variations in torsion angles, propeller twist, base roll, are all observed (See section 3.1.2).

#### 1.4. Phosphotriester synthesis of oligoribonucleotides.

With the advent of automated DNA synthesis, and the promise of automated

RNA synthesis, chemical synthesis of oligonucleotides has become the preferred method of preparing short oligonucleotides for structural and biological studies, largely superceding enzymatic synthesis. Chemical synthesis of oligoribonucleotides has lagged behind DNA synthesis largely because of the difficulties entailed in protecting the 2' hydroxyl group in such a way that, while the blocking group is stable under the phosphorylation and coupling conditions, it may be removed without cleavage or isomerization of the adjacent 3' and 5' linkage. The synthesis used for preparation of the oligoribonucleotides in this work was developed by Dr. Neilson and coworkers (England and Neilson, 1976; Werstiuk and Neilson, 1976), and the strategy is as follows:

1. Protection of the reactive sites on the monomers, using blocking groups that may be removed independently of each other. The amino groups on adenosine, cytidine and guanosine are protected by base-sensitive groups, the ribose 2' hydroxyl by an acid sensitive tetrahydropyranyl group, and the ribose 5' hydroxyl of the 5' terminal residue by a triphenylmethoxyacetyl (TRAC) group, which may be removed by very mild base.
2. Coupling 5', 3': Phosphorylation of the 3' hydroxyl position of the 5' terminal residue with  $\text{BBB-trichloroethylphosphate}$ , followed by coupling to the 5' hydroxyl group of the incoming blocked monomer. Repetition of as many cycles of phosphorylation and coupling as required.
3. Deprotection of the final sequence: Removal of the trichloroethyl blocking group with zinc/copper couple, followed by removal of the base-protecting groups with methanolic ammonia and the ribose protecting group with aqueous acid.



## 2. MATERIALS AND METHODS

### 2.1 Oligoribonucleotide synthesis

#### 2.1.1 Materials

Fully protected monomers N<sup>2</sup>-benzoyl-2'-O-tetrahydropyranylguanosine (HOG<sup>bz</sup><sub>1</sub>OH), N<sup>6</sup>-benzoyl-2'-O-tetrahydropyranyladenosine (HOA<sup>bz</sup><sub>1</sub>OH) and 2'-O-tetrahydropyranyluridine (HOU<sub>1</sub>OH) were synthesized according to previously published procedures (HOU<sub>1</sub>OH, Griffin *et al.*, 1968; HOA<sup>bz</sup><sub>1</sub>OH and HOG<sup>bz</sup><sub>1</sub>OH, Gregoire and Neilson, 1978; TracU<sub>1</sub>OH, Werstiuk and Neilson, 1972; TracA<sup>bz</sup><sub>1</sub>OH and TracG<sup>bz</sup><sub>1</sub>OH, Werstiuk and Neilson, 1973) from monomers obtained from Terochem Laboratories Ltd., Edmonton, Alberta, Canada. N<sup>4</sup>-benzoyl-2'-O-tetrahydropyranylcytidine was synthesized using 1,3-dichloro-1,1,3,3-tri-isopropylidisiloxane as a temporary blocking group for the 3' and 5' positions (Markiewicz and Wiewiórowski, 1978; Markiewicz, 1979) during benzylation of N<sup>4</sup> and addition of the tetrahydropyranyl group at the 2' position as described in detail below. Cytidine and 1,3-dichloro-1,1,3,3-tri-isopropylidisiloxane were also purchased from Terochem Laboratories Ltd.

Blocking group triphenylmethoxyacetate (Trac), coupling reagents triisopropylsulphonyl nitroimidazole (TPS-NI) (Van Boom *et al.*, 1977) and mesitylenesulphonyl triazole (MST) (Katagiri *et al.*, 1974), and phosphorylating agent BBB-trichloroethylphosphate (BBB) (England and Neilson, 1977) were synthesized as previously described. Starting materials triisopropylsulphonyl chloride,

mesitylenesulphonyl chloride, 1,2,4-triazole, nitroimidazole and 2,2,2-trichloroethanol were obtained from Aldrich, Milwaukee, WI. Coupling reagent mesitylenesulphonyl-4-nitrotriazole (MSNT) was synthesized according to Jones et al. (1980); 4-nitrotriazole was also obtained from Aldrich.

Reactions were monitored by silica gel thin layer chromatography, Pre-scored silica gel G plates from Analteck, Newark, DE. Plates were run in 5% or 10% v/v methanol/methylene chloride and developed by ceric sulphate spray (1% w/v  $Ce(SO_4)_2$  in 10% sulphuric acid) followed by heating to 350°C on a hot plate.

Blocked sequences were purified after each coupling step on silica gel (60-200 mesh) obtained from J.T. Baker, Philipsburg, NJ or by flash column chromatography on Merck Kieselgel 60 (230 - 400 mesh) obtained from BDH Chemicals, Ontario. For conventional chromatography stepwise gradient elution was used, 0 - 10% methanol in methylene chloride. For flash chromatography a linear gradient of methanol/methylene chloride was employed.

$D_2O$ , 99.9% and 100%, was obtained from Aldrich. 99.9%  $D_2O$  was distilled from EDTA (40 mg) before use.

2.1.2. Preparation of  $N^4$ -benzoyl-2'-O-tetrahydropyranylecytidine: (Hughes, Alkema, Sinclair and Neilson, unpublished procedure) (Figure 2.1).

2.1.2.1 3',5'-O-[1,1,3,3-Tetraisopropyl-1,3-disiloxanedivyl]-cytidine

1,3-Dichloro-1,1,3,3-tetraisopropylidisiloxane (12.28 g, 0.042 mol (315 g/mol; 1.001 g/ml)) was added to a stirred suspension of cytidine (10 g, 0.042 mol (243 g/mol))

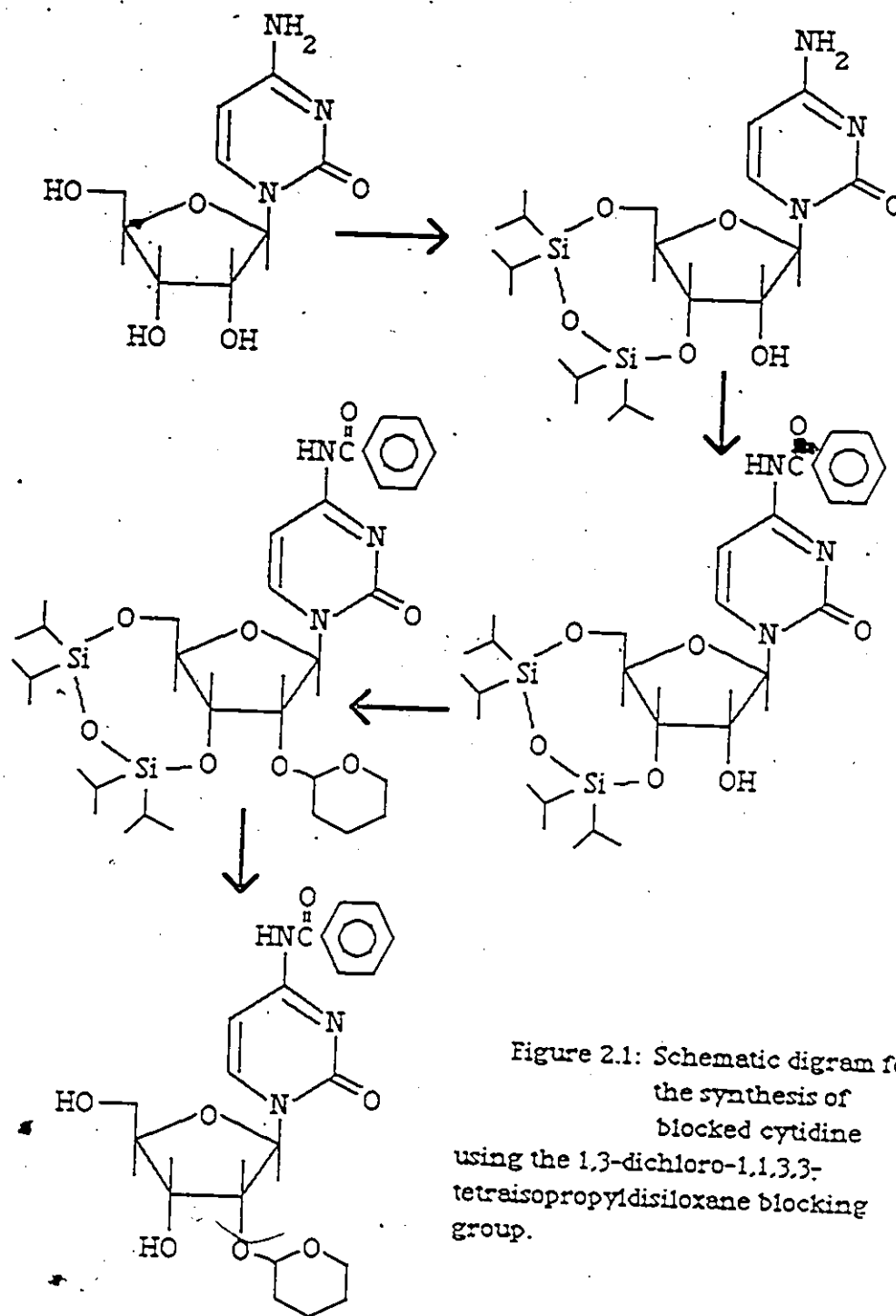


Figure 2.1: Schematic diagram for the synthesis of blocked cytidine using the 1,3-dichloro-1,1,3,3-tetraisopropylidisiloxane blocking group.

in dry pyridine (10 ml/g; 100 ml). Clearing of the suspension occurred after 45 min, and the reaction was complete in 2 h, as monitored by silica gel tlc with 10% methanol/methylene chloride mobile phase. Cytidine appeared at the origin, and the product at  $R_f$  0.4 - 0.5. Upon completion, the reaction was quenched with ice and the mixture extracted three times with  $\text{CH}_2\text{Cl}_2$ . The solvents were evaporated to dryness at reduced pressure, and toluene was used to azeotrope the residual pyridine. The product was purified by silica gel chromatography with methanol/methylene chloride as a mobile phase, the solvent was removed at reduced pressure, and the residue redissolved in  $\text{CH}_2\text{Cl}_2$ . The product precipitated overnight as a white powder. The product was verified by  $^1\text{H-NMR}$ .

Yield 16 g; 88%.

$^1\text{H-NMR}$  ( $\text{CD}_3\text{OH}$ , 250.13MHz):  $\delta$  7.953 (H-6), 5.844 (H-5), 5.897 (H-1',  $J_{1'2}$  0.5 Hz), 0.99 - 1.15 (isopropyl) (D.W. Hughes, 1982).

#### 2.1.2.2 $\text{N}^4$ -Benzoyl-3',5'-O-[1,1,3,3-tetraisopropyl-1,3-disiloxanediy]l-cytidine

3',5'-O-[1,1,3,3-Tetraisopropyl-1,3-disiloxanediy]l-cytidine (16 g, 0.036 mol (441.1 g/mol)) was dissolved in a minimum volume of absolute ethanol (100 ml). The flask was equipped with a stirrer and fitted with a reflux condenser. Benzoic anhydride (24 g, 0.108 mol (226 g/mol)) was added to the refluxing mixture in three equal portions at one hour intervals. The progress of the reaction was monitored by silica gel tlc, run in 10% methanol/methylene chloride. The starting material appeared at about  $R_f$  0.4, the product about  $R_f$  0.8. In the case of incomplete reaction a further 1 eq benzoic anhydride was added and the reflux continued for another hour.

The reaction was quenched with ice, and the mixture extracted with methylene chloride three times. The combined extracts were reduced to dryness in vacuo. The remaining ethanol was removed by azeotropeing with toluene two to three times under reduced pressure. The residue was chromatographed on silica gel, with methanol/methylene chloride as the mobile phase. Early fractions of the product tended to be contaminated with ethyl benzoate, and so were collected separately and rechromatographed. The product was verified by  $^1\text{H-NMR}$ .

Yield: 17.2g; 84%.

$^1\text{H-NMR}$  (acetone- $d_6$ , 250.13MHz)  $\delta$  8.283 (H-6), 7.408 (H-5), 5.759 (H-1'), 0.99 - 1.15 (isopropyl), 8.150, 7.653, 7.551 (benzoyl) (D.W. Hughes, 1982).

2.1.2.3.  $\text{N}^4$ -Benzoyl-3',5'-O-[1,1,3,3-tetraisopropyl-1,3-disiloxanediy]-2'-O-tetrahydropyranylcytidine

$\text{N}^4$ -Benzoyl-3',5'-O-[1,1,3,3-tetraisopropyl-1,3-disiloxanediy]-cytidine (14 g, 0.025 mol (545 g/mol)) and p-toluenesulphonic acid (3.6 g, 0.019 mol (190 g/mol)) were dissolved in dry dioxane (10 ml/g nucleoside; 140 ml) with added molecular sieves (4A) (1 - 2g). The mixture was stirred in an icebath until the dioxane began to freeze, and cold dihydropyran (10.95 mmol/ml; 10 eq; 20 ml) added portionwise, slowly, with stirring or agitation. Upon completion of addition, the reaction mixture was allowed to warm slowly to room temperature, and the progress of the reaction monitored by silica gel tlc, run in 5% methanol/ $\text{CH}_2\text{Cl}_2$ . The starting material appears at  $R_f$  0.7, the product at 0.9, and a brownish granular spot due to products of dihydropyran side reactions at the solvent front. Completion of the reaction took a few hours, sometimes

overnight. When complete the reaction mixture was cooled for 5 min, carefully neutralized with conc. aqueous ammonia (if made alkaline, precipitated p-toluenesulphonate will redissolve), filtered to remove precipitated salt and sieves, and evaporated to dryness in vacuo, using toluene to azeotrope dioxane. The resultant sticky foam was chromatographed on silica gel using methanol/methylene chloride as the mobile phase; the product eluted at 1 - 2% methanol and dried to a white or pale yellow foam.

Yield: 14.5 g; 88%.

#### 2.1.2.4. N<sup>4</sup>-benzoyl-2'-O-tetrahydropyranylcytidine

Tetrabutylammonium fluoride, 0.1 M, in tetrahydrofuran (stock solution 1.0M diluted 1/10 with THF; 1 eq; 17 ml) was added to dry N<sup>4</sup>-benzoyl-3',5'-O-[1,1,3,3-tetraisopropyl-1,3-disiloxanediyl]-2'-O-tetrahydropyranylcytidine ( 11 g, 0.017 mol ( 645 g/mol)), and the solution stirred at room temperature for 15 minutes, after which silica gel tlc showed the reaction to be complete. The starting material ran just behind the solvent front in 10% methanol/methylene chloride, and the product appeared as two isomers in approx. equal quantities at R<sub>f</sub> 0.55 and 0.45.

The reaction volume was doubled with methylene chloride, the mixture poured onto water and extracted with methylene chloride three to four times. The combined extracts were reduced to dryness in vacuo, and the THF azeotroped with toluene. The residue was dissolved in methylene chloride, and any product which precipitated after sitting overnight was filtered before the supernatant was chromatographed on silica gel.

The isomer with low  $R_f$  was recrystallized from methylene chloride and the one with high  $R_f$  from methylene chloride upon addition of hexanes, or else used as a foam. The product was verified by  $^1\text{H-NMR}$ .

Yield: (total, both isomers) 5.6 g, 77%.

$^1\text{H-NMR}$  (DMSO- $d_6$  + 1 drop  $\text{H}_2\text{O}$ ) high  $R_f$  (400.13 MHz)  $\delta$  8.552 (H-6), 7.338 (H-5), 5.830 (H-1',  $J_{1'2'}$  2.1 Hz), 7.960, 7.616, 7.498 (benzoyl), 4.935, 1.36 - 1.86, 3.3 (THP).

low  $P_f$  (250.13 MHz)  $\delta$  8.423 (H-6), 7.320 (H-5), 6.035 (H-1',  $J_{1'2'}$  4.8 Hz), 7.978, 7.619, 7.500 (benzoyl), 4.832, 1.33 - 1.74, 3.313 (THP) (D.W. Hughes, 1982).

### 2.1.3. Synthesis of oligoribonucleotides

Quantities and yields for preparation of all oligonucleotides used in this study are given in Tables 2.1 and 2.2. The procedures used have been previously published (England and Neilson, 1976; Werstiuk and Neilson, 1976); the synthesis of protected GpCpGpC will be used to illustrate them (Figure 2.2).

All reactions were carried out under dry nitrogen.

#### 2.1.3.1. $\text{N}^2$ -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxycarbonyl-guanyl-[3'(2,2,2-trichloroethyl)-5'-N $^4$ -benzoyl-2'-O-tetrahydropyranyl-cytidine (TracG $_1^{\text{bz}}$ C $_1^{\text{bz}}$ OH)

BBB-Trichloroethylphosphate (0.9 g, 0.39 mmol (230 g/mol)) was converted to the pyridinium salt by reducing to dryness from pyridine three times. On the last evacuation, the final volume was left at about 5 ml, and 1.95 g MST added. The solution was activated by heating to 40 - 50°C for 1.5 h, until the MST had dissolved.

Table 2.1: Summary of the Preparation of Oligoribonucleotides<sup>a</sup>, quantities used and yields of product.

reactants						products		
cmpd	mg	mmol	cmpd	mg	mmol	cmpd	mg	%
G	1500	1.195	G	850	1.80	GG	1070	38
GG	500	0.35	C	150	0.35	GGC	270	40
GGC	270	0.14	C	61	0.14	GGCC	167	47
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 <sup>b</sup>	325	58
G	1500	1.95	C	760	1.76	GC	1650	62
GC	360	0.25	C	111	0.26	GCC	320	66
GCC	280	0.15	G	61	0.14	GCCG	178	47
C	1200	1.80	G	820	1.70	CG	1330	59
CG	430	0.34	G	160	0.33	GCC	385	61
CGG	300	0.15	C	70	0.16	CGGC <sup>b</sup>	165	43
G	1500	1.95	C	760	1.76	GC	1650	62
GC	400	0.29	G	152	0.32	GCG	200	36
GCG	300	0.16	C	69	0.16	GCGC	170	42
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 <sup>b</sup>	270	59
CG	500	0.36	G	195	0.40	CGG	430	58
CGG	400	0.20	A	95	0.21	CGGA <sup>c</sup>	190	36

<sup>a</sup>G, C and A in all compounds in columns 1, 4 and 7 stand for the fully protected residues N<sup>2</sup>-benzoyl-2'-O-tetrahydropyranopyranylguanosine, N<sup>4</sup>-benzoyl-2'-O-tetrahydropyranopyranylcytidine and N<sup>6</sup>-benzoyl-2'-O-tetrahydropyranyladenosine, respectively; in addition, the 5' terminal 5'-ribose hydroxyl position of the compounds in columns 1 and 7 is blocked by a trityloxyacetyl group. GG represents GpG with two protected guanosine residues linked by a protected phosphate group, the phosphate blocking group being 2,2,2-trichloroethyl. The coupling reagent in all cases was mesitylenesulphonyltriazole.

<sup>b</sup>Compounds prepared by Dr. P. Romaniuk (P.J. Romaniuk, 1979)

<sup>c</sup>Compounds prepared by Dr. D. Alkema.



Table 2.2: Summary of the Preparation of Oligoribonucleotides<sup>3</sup>, quantities used and yields of products.

reactants						products		
cmpd	mg	mmol	cmpd	mg	mmol	cmpd	mg	%
G	2340	3.30	G	1400	3.30	GG	2420	55
GG	370	0.25	U	84	0.25	GCU	236	47
GGU	220	0.11	CC	118	0.11	GGUCC	130	36
GG	2000	1.46	A	632	1.49	GGA	800	30
GGA	400	0.19	C	93	0.19	GGAC	270	56
GGA	400	0.19	CC	202	0.19	GGACC	70	12
G	4140	5.40	C	2300	5.2	CC	3300	45
GC	1300	0.93	U	305	0.82	GCU	940	56
GCU	940	0.49	G	235	0.5	GCUG	630	51
GCUG	600	0.23	C	104	0.24	GCUGC	340	46
CC	2000	1.46	A	756	1.46	GCA	1380	47
GCA	1300	0.65	G	300	0.64	GCAG	770	45
GCAG	385	0.14	C	161	0.14	GCAGC	210	44
C	2700	3.69	C	1590	3.69	CC	2940	59
CC	1000	0.76	U	252	0.76	CCU	860	59
CCU	850	0.45	G	212	0.45	CCUG	570	48
CCUG	550	0.22	G	108	0.22	CCUGG	210	30
CC	1800	1.34	A	302	1.34	CCA	1470	53
CCA	1000	0.48	G	234	0.5	CCAG	520	41
CCAG	450	0.17	G	79	0.17	CCAGG	180	33

<sup>3</sup>G, C and A in all compounds in columns 1, 4 and 7 stand for the fully protected residues N<sup>2</sup>-benzoyl-2'-O-tetrahydropyranylguanosine, N<sup>4</sup>-benzoyl-2'-O-tetrahydropyranylcytidine and N<sup>6</sup>-benzoyl-2'-O-tetrahydropyranyladenosine, respectively; in addition, the 5' terminal 5'-ribose hydroxyl position of the compounds in columns 1 and 7 is blocked by a trityloxyacetyl group. GG represents GpG with two protected guanosine residues linked by a protected phosphate group, the phosphate blocking group being 2,2,2-trichloroethyl. The coupling reagent in all cases was mesitylenesulphonyltriazone.

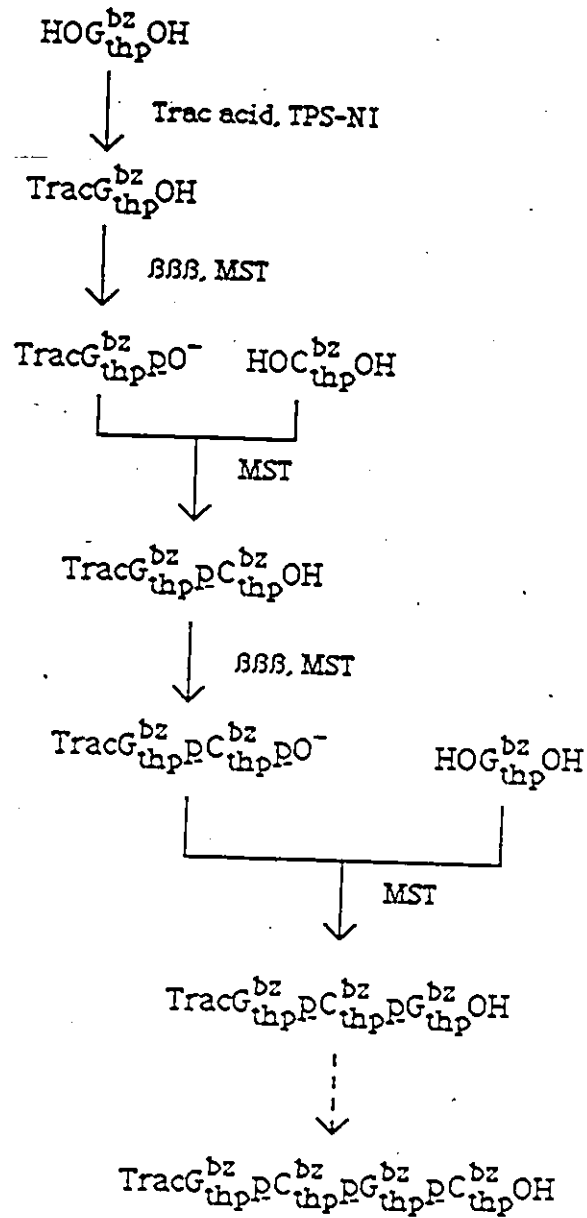


Figure 2.2: Stepwise synthesis of protected GCGC by the phosphotriester method.

TracG ( 1.5 g, 0.19 mmol (771 g/mol)) was dissolved in pyridine (15 ml), and dried azeotropically by evaporation of the pyridine at reduced pressure. This was repeated twice. Dry nitrogen was admitted to the flask after each evaporation. The activated 888-MST complex was added by pipette, using dry pyridine as rinse and the whole submitted to evaporation at reduced pressure to give an oil. The flask was then filled with dry nitrogen, sealed and left in the dark at room temperature

The progress of the reaction was monitored daily by silica gel tlc. At 10% methanol/methylene chloride TracGOH appeared as a yellow spot at  $R_f$  0.9, and TracGpO<sup>-</sup> as a yellow cusp at  $R_f$  0.2 - 0.4. After three days the reaction was over 95% complete and was worked up. A small quantity--about equal to the reaction volume--of ice was added, and the flask let stand for fifteen minutes. Methylene chloride (10 ml) was added to dissolve the oil and the mixture let stand a further fifteen minutes. Then the mixture was poured into water (20 ml) and extracted with methylene chloride (4 x 20 - 30 ml). The combined extracts were washed once with water and reduced to dryness, first on a water aspirator rotary evaporator, and then on a vacuum rotary evaporator. The residue was redried by reducing to dryness from pyridine three times, MST (0.5 g) was added to reactivate the phosphate, and the flask was let stand at room temperature for 1 h, until the MST had dissolved.

HOC<sub>t</sub><sup>bz</sup>OH ( 0.76 g, 0.17 mol(431 g/mol)) was dried by reducing to dryness from pyridine three times, and added to the reactivated TracG<sub>t</sub><sup>bz</sup>pO<sup>-</sup>. The volume was reduced and the flask resealed under dry nitrogen and left at room temperature in the dark.

The progress of the reaction was monitored by the disappearance of the yellow cusp belonging to TracG<sub>t</sub><sup>bz</sup>pO<sup>-</sup>, and of the blackish spot belonging to the incoming nucleotide, and the appearance of a yellow spot around  $R_f$  0.6, above the cusp.

The tlc showed all the phosphorylated material to have been consumed in five days, although some  $\text{HOC}_t^{\text{bz}}\text{OH}$  remained; at this point the oil was a very dark brown.

The workup of the reaction was identical to that for the phosphorylation: the reaction was quenched with ice, the oil redissolved in methylene chloride, poured into water and extracted with methylene chloride. The residual pyridine was azeotroped with toluene, and the toluene with methylene chloride.

The residue was dissolved in 100 ml methylene chloride and loaded onto a 10 g silica gel column. The elution was performed with 100 ml each of 0 - 2% methanol/ methylene chloride, 250 ml 3 - 4%, 100 ml 5%. Certain of the product fractions appeared contaminated with small quantities of the incoming nucleotide. The final yield was 0.65 g of  $\text{TracG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{OH}$  (a single spot by tlc) and 1.2 g  $\text{TracG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{OH} + \text{HOC}_t^{\text{bz}}\text{OH}$  (4:1 by tlc); approximately 75%.

2.1.3.2.  $\text{N}^2$ -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-guanylvl[3'(2,2,2-trichloroethyl)-5']  $\text{N}^2$ -benzoyl-2'-O-tetrahydropyranyl-cyridylvl [3'(2,2,2-trichloroethyl)-5']  $\text{N}^2$ -benzoyl-2'-O-tetrahydropyranyl-guanosine ( $\text{TracG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{G}_t^{\text{bz}}\text{OH}$ )

BBB-Trichloroethylphosphate (110 mg, 0.48 mmol) were converted to the pyridinium salt by evaporating from pyridine *in vacuo*, leaving about 5 ml solvent in the final evaporation. MST (240 mg, 0.96 mmol) were added, and the mixture heated to 40 - 50°C until the MST had dissolved. Then the activated complex was added to a predried pyridine solution of  $\text{TracG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{OH}$  300 mg, 0.24 mmol), and the flask sealed.

The progress of the reaction was monitored by silica gel tlc plates run in 10%

methanol/methylene chloride. The reaction was complete in three days, as indicated by the conversion of a yellow spot at  $R_f$  0.7 to a yellow cusp,  $r_f$  0.2 - 0.5. The reaction was quenched with ice (2 g), let stand for 15 min, mobilized with methylene chloride, poured into water (20 ml), and extracted with methylene chloride (3 x 20-30 ml). The combined extracts were washed once with water and evaporated to an oil. The oil was redried in preparation for the coupling step by reducing to dryness from pyridine three times under vacuum. MST (60 mg, 0.24 mmol) were added and the flask let stand until the MST had dissolved (1 h), then a pre-dried pyridine solution of 114 mg  $\text{HOC}_t^{\text{bz}}\text{OH}$  was added to the activated  $\text{TracG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{P}_2$ , the volume of solution reduced and the flask resealed under nitrogen.

The reaction was complete by tlc after four days standing in the dark. The yellow cusp at  $R_f$  0.2 - 0.5, and the black spot for incoming nucleotide at  $r_f$  0.4, had been replaced by a yellow spot at  $R_f$  0.55. The reaction was quenched with ice, the oil dissolved in methylene chloride, poured into water and extracted with methylene chloride. The combined residues were reduced to an oil, and the residual pyridine removed by azeotroping with toluene. The residue was chromatographed on silica gel (15 g) with methanol/methylene chloride as the mobile phase, to give 130 g  $\text{HOG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{G}_t^{\text{bz}}\text{OH}$  (32%).

The preparation had to be repeated once, on a comparable scale, to obtain sufficient starting material for the next step.

2.1.3.3.  $N^2$ -Benzoyl-2'-O-tetrahydropyranyl-5'-O-triphenylmethoxyacetyl-guanlyl [3'(2,2,2-trichloroethyl)-5']  $N^4$ -benzoyl-2'-O-tetrahydropyranyl-cytidylyl [3'(2,2,2-trichloroethyl)-5']  $N^2$ -benzoyl-2'-O-tetrahydropyranyl-

guanylyl [3'(2,2,2-trichloroethyl)-5'] N<sup>4</sup>-benzoyl-2'-O-  
tetrahydropyranyl-cytidine (TracG<sub>t</sub><sup>bz</sup>C<sub>t</sub><sup>bz</sup>G<sub>t</sub><sup>bz</sup>C<sub>t</sub><sup>bz</sup>OH)

BBB-Trichloroethylphosphate (74 mg; 0.32 mmol) was converted to its pyridinium salt by repeated evaporations *in vacuo* from dry pyridine, to a final volume of 5 ml. MST (161 mg, 0.64 mmol) was added, and the mixture heated for 1 h to activate the complex. The activated complex was added to 300 mg HOG<sub>t</sub><sup>bz</sup>C<sub>t</sub><sup>bz</sup>G<sub>t</sub><sup>bz</sup>OH (1869 g/mol; 0.16 mmol), and the flask sealed under dry nitrogen and left in the dark, at room temperature. From the third to the fifth day the tlc indicated little change in the proportion of reactant ( $R_f$  0.5) to product ( $R_f$  0.2 - 0.4), and the reaction was therefore stopped.

The reaction was quenched with ice, the oil dissolved in methylene chloride, poured into water and extracted with methylene chloride (4 x 20 ml). The combined extracts were washed once with water and reduced to dryness, the resulting oil was redried by three times evaporating from pyridine *in vacuo*, and the flask sealed under nitrogen. MST (40 mg, 0.32 mmol) was added and the flask let stand at room temperature for 1 h to allow the MST to dissolve. A solution of HOC<sub>t</sub><sup>bz</sup>OH (69 mg, 0.16 mmol (431 g/mol)), predried by evaporation *in vacuo* from dry pyridine, was added, the volume reduced, the flask resealed under dry nitrogen, and the reaction left to sit at room temperature in the dark.

After four days the tlc indicated complete reaction of the phosphorylated material, with formation of a spot at about  $R_f$  0.45. The reaction was quenched by addition of ice, poured into water and extracted with methylene chloride (4 x 20 ml). The combined extracts were washed with water and reduced to dryness. Purification of

the residue by silica gel column chromatography, with methanol/methylene chloride as the mobile phase, yielded 170 mg  $\text{TracG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{G}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{OH}$  (43%).

#### 2.1.4. Block coupling

Certain sequences were prepared by block coupling (Werstiuk and Neilson, 1972). The 5' terminal protecting group (Trac) of the 3' fragment was removed using 1% ammonia:methanol, 1 ml/10 - 20 mg protected oligonucleotide. The progress of the reaction was monitored by tlc by the disappearance of the yellow spot belonging to the 5' protected nucleotide and the appearance of a black spot belonging to the 5' deprotected material at lower  $r_f$  and a yellow trityl spot at the solvent front. The reaction was quenched by reducing to dryness upon first appearance of lower  $R_f$  spots resulting from removal of base protecting benzoyl groups, and separated from unreacted starting material by column chromatography.

The deprotected material was used as an incoming nucleotide in a normal coupling reaction.

#### 2.1.5. Deprotection procedure for oligoribonucleotides

The deprotection procedure has been fully described elsewhere (England and Neilson, 1976). The deprotection of  $\text{TracG}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{G}_t^{\text{bz}}\text{C}_t^{\text{bz}}\text{OH}$  will be used as illustration of the method. Yields of deprotection of the oligomers are given in Table 2.3.

All glassware used in this procedure was washed with chromic acid to remove traces of ribonucleases.

Table 2.3: Chromatographic data and yields following deprotection for deprotected sequences.

Sequence <sup>a</sup>	R <sub>f</sub> <sup>b</sup>	Yield % <sup>c</sup>	Sequence <sup>a</sup>	R <sub>f</sub> <sup>b</sup>	Yield % <sup>c</sup>
GG	0.48	17	CC	0.68	46
GGC	0, 0.25	16	CCU	0.50	44
GGU	0, 0.26	15	CCA	0.49	38
GGA	0, 0.29	32	CCG	0.48	25
GGCC	0.6	7	CCUG	0.39	20
GGAC		13	CCAG	0.26	16
GGUCC	0.14	11	CCGG	0.13	27
GGACC	0.10	6	CCUGG	0, 0.25	8
			CCAGG	0.14	7
			CCAUGG	0.05	6
GC	0.51	34			
GCC	0.48	35			
GCU	0.35				
GCA	0.35	12			
GCG	0.2	25	CCG	0.43	10
GCCG	0.21	12	CCGG	0.21	19
GCUG	0.27	17	CCGC	0.32	27
GCAG	0.17	35	CCGA	0.22	26
GCGC	0.07	12			
GCUGC	0.18	34			
GCAGC	0.15	38			
GCAUGC	0.05	10			

<sup>a</sup>Refers to the fully deblocked oligoribonucleotide, written 5' - 3'.

<sup>b</sup>Watman #1, #40 paper, mobile phase 1:1 1M ammonium acetate:ethanol.

<sup>c</sup>Calculated from UV absorbance at 260nm, assuming a 90% hypochromicity factor.



1. Removal of the B88-trichloroethyl group (deprotection of the phosphate):

TracG<sub>t</sub><sup>bz</sup>C<sub>t</sub><sup>bz</sup>G<sub>t</sub><sup>bz</sup>C<sub>t</sub><sup>bz</sup>OH (15 mg) was dissolved in 0.5 ml dry dimethylformamide (DMF) in a screw-cap sample vial. 30 - 50 mg zinc/copper couple (preparation described in Legoff *et al.*, 1964) and a magnetic stirring flea were added, and the mixture stirred at 40 - 50°C until the tlc showed complete conversion to a charged species (one yellow spot at the origin).

2. Removal of the Trac and benzoyl groups. Methanol and concentrated ammonium hydroxide (2 ml each) were added to the cooled DMF solution, and the mixture left to stir at room temperature for 2 days.

The ammoniacal solution was filtered to remove the spent zinc/copper couple and the vial and filters were washed with 2 x 2 ml of 1M aqueous ammonia followed by washing with 2 x 2 ml methanol. Chelex-NH<sub>4</sub><sup>+</sup> was added to the filtrate, and the mixture was stirred for 1 h to remove the Cu<sup>2+</sup>, then the chelex was removed by filtration and the filtrate reduced to dryness and transferred to Whatman #1 chromatography paper.

The papers were eluted overnight in descending mode with a mobile phase of 1:1 1 M ammonium acetate: absolute ethanol. The desired band was located by its fluorescence under ultraviolet light, just behind the solvent front. The band was cut out, the acetate salts removed by a 1 h immersion in absolute ethanol, and the band dried by a 20 min immersion in ether, followed by air drying. The oligonucleotide was eluted from the band with distilled water.

3. Removal of the tetrahydropyranyl groups. The above eluent was acidified to pH2 with 2 M hydrochloric acid, and let stand for 2 days, then neutralized to pH7 - pH8 with 1 M aqueous ammonia and reduced to dryness *in vacuo*. The residue was

transferred to Whatman #40 chromatography paper (Whatman #1 was used for later syntheses), and the papers eluted under the same conditions as before, overnight in 1:1 1M ammonium acetate: absolute ethanol. Under ultraviolet light the product band appeared at an  $R_f$  of 0.07. The band was cut out, the acetate salts removed by soaking in absolute ethanol for two 1 h periods, and the band dried by soaking in ethyl ether, and finally the oligonucleotide product was removed from the band by elution with distilled water.

The volume of final eluent and  $A_{260}$  of a 1/10 dilution were measured and used to calculate the yield in optical density units, ODs. The % yield was calculated using the following extinction coefficients: Ap  $15.2 \times 10^3$ , Gp  $13.4 \times 10^3$ , Cp  $13.4 \times 10^3$ , Up  $10.1 \times 10^3$ , A  $14.9 \times 10^3$ , G  $13.6 \times 10^3$ , C  $9.1 \times 10^3$ , U  $10.1 \times 10^3$ . For GCGC % yield: 12%.

Certain sequences, particularly the longer ones, were further purified by repeat chromatography on Whatman #40 (#1) under the same conditions as the final step.

#### 2.1.6. Purification of sequences by High Performance Liquid Chromatography (HPLC)

A few sequences, most notably CCAGG and CCUGG, were purified by high performance liquid chromatograph, using a reverse phase C-18 column (Waters  $\mu$ Bondapack C-18) on a Waters HPLC with solvent gradient programmer. The mobile phase was 0.1M ammonium acetate (pH 6) and methanol. Detection was achieved by monitoring the UV absorbance at 254 nm, with a Waters Model 440 UV spectrophotometer.

The two gradients most commonly used were as follows:

<u>Time (min)</u>	<u>Flow rate</u> <u>(ml/min)</u>	<u>% buffer</u>	<u>% methanol</u>	<u>curve</u>
0.0	0.8	100	0	06
30	0.8	70	30	06
32	0.8	60	40	06
35	0.8	60	40	06
37	0.8	100	0	06
50	0.8	100	0	06

<u>Time (min)</u>	<u>Flow rate</u> <u>(ml/min)</u>	<u>% buffer</u>	<u>% methanol</u>	<u>curve</u>
0.0	0.8	100	0	06
30	0.8	75	25	06
32	0.8	60	40	06
35	0.8	60	40	06
37	0.8	100	0	06
50	0.8	100	0	06

The desired compounds eluted with retention times of approximately 12 - 14 min.

### 2.1.7. <sup>1</sup>H-NMR methodology

Most proton NMR spectra were obtained on a Bruker WM-250 NMR spectrometer with a double tuned <sup>1</sup>H/ <sup>19</sup>F probe and Aspect 2000 computer. Other spectra were obtained on a Bruker AM-500 spectrometer with a double tuned <sup>1</sup>H/ <sup>13</sup>C probe, operating at 500.14 MHz. Probe temperatures were maintained to ±1°C by a Bruker variable temperature unit and calibrated with a thermocouple. The internal reference was t-butyl alcohol (1.231 ppm) and the resonances were reported relative to DSS

(sodium 4,4-dimethyl-4-sila-1-pentansulphonate).

Suppression of the residual HOD resonance in samples dissolved in  $D_2O$  for observation of the nonexchangeable base and ribose protons was accompanied by homonuclear solvent suppression (PRESAT in the Bruker microprogram library) with selective irradiation of the water peak. Spectra were acquired in 16 K and zero-filled to 32 K.

Suppression of the  $H_2O$  peak in samples dissolved in 90/10  $H_2O/D_2O$  for observation of the exchangeable imino protons involved the use of a 1:1 hard pulse sequence to generate zero spectral density at the frequency of the water peak, as described by Clore and Gronenborn, 1983.

The COSY spectrum used to assign CC was acquired using the microprogram COSY.AU (Bruker microprogram library), with SI1 = 256W, SI2 = 128W, TD1 = TD2 = 512W.

#### 2.1.7.1 Verification of sequences

The final eluent from an oligonucleotide preparation was frozen and lyophilized to dryness, and twice (or three times) reconstituted in  $D_2O$ , frozen and lyophilized to remove residual water. Stock NMR buffer was prepared in deionized water, frozen in 1 ml aliquots, lyophilized and reconstituted in 100%  $D_2O$ . Buffers consisted of either 1.0M NaCl, 0.01M  $NaH_2PO_4$ , pH 7 (often referred to as 1.0 M salt buffer), or 0.1M NaCl, 0.01M  $NaH_2PO_4$ , pH 7 (referred to as 0.1 M salt buffer); the latter was adopted on account of the improvement in resolution of the GC sequences at low temperature in lower salt, and to increase the solubility of the intercalator complexes studied in Section 5. The dry oligonucleotide was dissolved in 400  $\mu$ l reconstituted buffer, with 3  $\mu$ l of 0.1%

*t*-butyl alcohol-OD as reference. A  $^1\text{H}$ -NMR spectrum was obtained at  $70^\circ\text{C}$ , and the chemical shifts and coupling constants compared to those of the sequence precursors, and those predicted by the parameter set derived from triribonucleotide data (Hader et al., 1981) (in the case of samples in 1.0 M buffer) (See Tables 2.4 to 2.8).

#### 2.1.7.2. Observation of melting behaviour and determination of melting temperature

Following satisfactory characterization, a series of variable temperature spectra were obtained for the sequence over the range  $70^\circ\text{C}$  (in some cases  $80^\circ\text{C}$ ) to at least below  $30^\circ\text{C}$  (lower limit depending upon the  $T_m$  and the resolution at low temperatures) at intervals of 5 or  $10^\circ\text{C}$ . The chemical shift and coupling constant vs. temperature data was plotted for inspection. Individual melting temperatures for all protons showing sigmoidal chemical shift vs. temperature behaviour were determined by fitting the experimental chemical shift versus inverse temperature (also sigmoidal) to a fifth order polynomial curve, and then differentiating to obtain an inflection point. The calculated inflection point was used as the individual proton  $T_m$  and the average of the individual proton  $T_m$ s was taken to represent the overall duplex  $T_m$ .

#### 2.1.7.3. Observation of imino protons

The sample was frozen, lyophilized, and reconstituted in 400  $\mu\text{l}$  90/10  $\text{H}_2\text{O}/\text{D}_2\text{O}$  with 5  $\mu\text{l}$  *t*-butyl alcohol-OD for reference. Spectra were obtained at 8 -  $15^\circ$  intervals from  $-5^\circ\text{C}$  until the imino protons broadened and disappeared. Assignment of imino proton spectra is discussed in Section 3.2.3.

Table 2.4: Chemical shift assignments for the base and anomeric protons of GGCC, GCCG and GCGC at 70°C by incremental analysis and computer prediction. All data and predicted values for 1.0M salt buffer, relative to DSS.

Sequence	Chemical shift (ppm)											
	G	GG	GC	GGC	GCC	GCG	GGCC	GCCG	GCGC			
Proton							obsd.	calc.	obsd.	calc.	obsd.	calc.
G(1)H-8	7.961	7.896	7.966	7.903	7.935	7.964	7.910	7.909	7.952	7.936	7.936	7.934
G(2)H-8		7.971		7.977			7.949	7.976				
C(2)H-6			7.780		7.766	7.712			7.739	7.748	7.698	7.724
G(3)H-8						7.965					7.942	7.970
C(3)H-6				7.774	7.804		7.740	7.763	7.733	7.745		
G(4)H-8									7.976	7.981		
C(4)H-6							7.787	7.815			7.760	7.787
C(2)H-5			5.910		5.857	5.761			5.841	5.867	5.842	5.842
C(3)H-5				5.895	5.998		5.828	5.894	5.943	5.972		
C(4)H-5							5.967	6.004			5.887	5.926
G(1)H-1'	5.888	5.798	5.874	5.820	5.860	5.834	5.825	5.807	5.852	5.860	5.827	5.873
G(2)H-1'		5.877		5.866			5.851	5.859				
C(2)H-1'			5.913		5.898	5.894			5.868	5.904	5.893	5.924
G(3)H-1'						5.855					5.851	5.839
C(3)H-1'				5.895	5.898		5.870	5.917	5.868	5.895		
G(4)H-1'									5.868	5.904		
C(4)H-1'							5.887	5.900			5.893	5.938

Table 2.5: Chemical shift assignments of the base and anomeric protons of CCGG, CGGC and CGCG at 70°C by incremental analysis and computer prediction. All data and predicted values for 1.0M salt buffer, relative to DSS.

Sequence	Chemical shift (ppm)											
	C	CC	CG	CCG	CCG	CGC	CCGG	CGGC	CGCG			
Proton							obsd.	calc.	obsd.	calc.	obsd.	calc.
C(1)H-6	7.781	7.786	7.691	7.750	7.662	7.693	7.746	7.734	7.666	7.675	7.692	7.710
C(2)H-6		7.832		7.750			7.710	7.730				
G(2)H-8			8.004		7.942	7.992			7.935	7.952	7.968	7.974
C(3)H-6						7.788					7.705	7.718
G(3)H-8				7.984	7.968		7.918	7.925	7.951	7.977		
C(4)H-6									7.764	7.770		
G(4)H-8							7.957	7.959			7.968	7.952
C(1)H-5	6.073	6.020	5.975	5.976	5.946	5.931	5.970	5.977	5.935	5.970	5.919	5.954
C(2)H-5		6.020		5.954			5.935	5.945				
C(3)H-5						5.931					5.851	5.894
C(4)H-5									5.904	5.944		
C(1)H-1'	5.894	5.900	5.823	5.884	5.803	5.805	5.868	5.881	5.798	5.845	5.856	5.836
C(2)H-1'		5.852		5.819			5.795	5.833				
G(2)H-1'			5.894		5.811	5.865			5.783	5.809	5.800	5.836
C(3)H-1'						5.911					5.885	5.932
G(3)H-1'				5.884	5.873		5.814	5.806	5.854	5.855		
C(4)H-1'									5.895	5.939		
G(4)H-1'							5.840	5.876			5.816	5.861

Table 2.6: Chemical shift assignments for the base and anomeric protons of GGUCC and GGACC at 70°C by incremental analysis and computer prediction. Experimental data in 0.1M salt buffer, predicted values in 1.0M salt buffer.

Sequence	Chemical shift (ppm)								
	G	GG	GGU	GGA	GGAC	GGUCC		GGACC	
Proton						obsd.	pred.	obsd.	pred.
G(1)H-8	7.956	7.893	7.895	7.862	7.865	7.895	7.909	7.865	7.965
G(2)H-8		7.971	7.982	7.919	7.925	7.989	7.888	7.865	7.869
U(3)H-6			7.797			7.797	7.778		
A(3)H-8				8.326	8.331			8.321	8.286
A(3)H-2				8.175	8.150			8.128	8.136
C(4)H-6					7.745	7.837	7.790	7.717	7.934
C(5)H-6						7.848	7.855	7.818	7.795
U(3)H-5			5.826			5.818	5.816		
C(4)H-5					5.862	6.030	5.996	5.977	5.851
C(5)H-5						6.013	6.044	5.797	5.973
G(1)H-1'	5.899	5.801	5.801	5.754	5.762	5.804	5.822	5.768	5.768
G(2)H-1'		5.872	5.888	5.768	5.755	5.871	5.888	5.762	5.761
U(3)H-1'			5.888			5.884	5.908		
A(3)H-1'				6.055	6.043			6.018	6.038
C(4)H-1'					5.842	5.920	5.910	5.793	5.848
C(5)H-1'						5.920	9.931	5.883	5.888



Table 2.7: Incremental and predicted assignments for the base and anomeric protons of CCUGG and CCAGG at 70°C. Experimental data at 2.5mM in 0.1 M salt buffer salt, predicted values for 1.0M salt buffer.

Sequence	Chemical shift (ppm)									
	C	CC	CCU	CCA	CCUG	CCAG	CCUGG	CCAGG		
Proton							obsd.	pred.	obsd.	pred.
C(1)H-6	7.777	7.806	7.825	7.759	7.782	7.731	7.771	7.759	7.736	7.736
C(2)H-6		7.838	7.856	7.736	7.834	7.722	7.825	7.804	7.722	7.709
U(3)H-6			7.787		7.767		7.730	7.740		
A(3)H-8				8.409		8.335			8.304	8.298
A(3)H-2				8.244		8.168			8.137	8.166
G(4)H-8					7.994	7.933	7.931	7.936	7.878	7.878
G(5)H-8							7.979	7.980	7.949	7.932
C(1)H-5	6.046	6.009	5.987	5.900	5.983	5.911	5.983	6.004	5.926	5.966
C(2)H-5		5.863	5.959	5.886	5.959	5.892	5.950	5.976	5.888	5.828
U(3)H-5				5.867		5.837	5.834	5.829		
C(1)H-1'	5.876	5.899	5.910	5.864	5.897	5.836	5.906	5.864	5.836	5.828
C(2)H-1'		5.853	5.820	5.800	5.849	5.810	5.836	5.901	5.836	5.846
U(3)H-1'			5.910		5.878		5.813	5.862		
A(3)H-1'				6.092		6.021			5.962	6.038
G(4)H-1'					5.878	5.829	5.813	5.825	5.749	5.763
G(5)H-1'							5.868	5.884	5.810	5.832

Table 2.8: Chemical shift assignments for the base and anomeric protons of GCUGC and GCAGC at 70°C by incremental analysis and computer prediction. Experimental data obtained for 0.1M salt, except for GCA, GCAG and GCAGC, which were in 1.0M salt. Predicted values for 1.0M salt.

Sequence	Chemical shift (ppm)									
	G	GC	GCU	GCA	GCUG	GCAG	GCUGC		GCAGC	
Proton							obsd.	calc.	obsd.	calc.
G(1)H-8	7.956	7.969	7.967	7.920	7.952	7.903	7.949	7.960	7.870	7.908
C(2)H-6		7.796	7.789	7.709	7.796	7.670	7.781	7.764	7.684	7.668
U(3)H-6			7.804		7.793		7.735	7.751		
A(3)H-8				8.348		8.275			8.261	8.288
A(3)H-2				8.191		8.133			8.099	8.139
G(4)H-8					7.983	7.913	7.977	7.984	7.901	7.927
C(5)H-6							7.821	7.810	7.737	7.750
C(2)H-5		5.879	5.885	5.895	5.837	5.844	5.895	5.892	5.845	5.825
U(3)H-5			5.841		5.819		5.845	5.845		
C(5)H-5							5.948	5.985	5.859	5.913
G(1)H-1'	5.899	5.873	5.913	5.812	5.858	5.799	5.839	5.837	5.793	5.814
C(2)H-1		5.919	5.913	5.883	5.839	5.920	5.910	5.839	5.849	5.880
U(3)H-1'			5.861		5.858		5.856	5.875		
A(3)H-1'				6.061		5.798			5.950	5.976
G(4)H-1'					5.895	5.819	5.961	5.877	5.773	5.815
G(5)H-1'							5.928	5.970	5.867	5.927

### 3. SEQUENCE DEPENDENCE OF STABILITY STUDIES ON TRI- AND TETRARIBONUCLEOTIDES CONTAINING GUANOSINE-CYTIDINE BASE PAIRS

#### 3.1. Introduction

##### 3.1.1. Sequence dependence of stability

Primary structure interests the modern nucleic acid chemist only in as much as it dictates secondary and higher order structure, the duplexing and folding of complementary strands or regions. Quiescent DNA is fully duplexed; functional tRNA contains four short double-helical regions, and further tertiary base pairing; ribosomal RNA is extensively duplexed, with formation of numerous regions of base-pairing (Woese et al., 1983; Noller, 1984). Further subtlety is introduced by those processes which involve more than a single secondary structure: Partial melting of initiation regions is required for initiation of transcription; in *E. coli* control of translation by attenuation involves alternative secondary structures in the mRNA (Yanofsky, 1981); the activity of self-splicing RNA from *Tetrahymena thermophila* is abolished at low temperature by stabilizing mutations in a weak putative helix (Tanner and Cech, 1985a,b), suggesting that melting of this helix is required for activity. Thus, especially in RNA, the individual units of secondary structure, the helices, are required to have a considerable range of stability, from stable helices which serve mainly to support and orientate functional sites, to unstable helices which undergo transition from the duplexed to the single stranded state as conditions around them—conformation of adjacent elements, solvation, binding of proteins—alter their equilibrium constant.

Length of helix alone would not permit such variation; it has gradually become understood that the stability of both the DNA and RNA double helices are highly sequence dependent, with gross modulation in stability being accomplished by the G-C to A-T(U) ratio, and finer modulation by the ordering of pyrimidine and purine bases, and, in RNA, the addition of unpaired or dangling bases to the end of the stack.

The first described effect of sequence on stability was the greater stability of DNAs rich in G-C base pairs. An expression can be derived expressing the melting temperature  $T_m$  for a series of helices of the same length according to their percentage composition of G-C (Cantor and Schimmel, 1981)

The earliest attempts to predict secondary structure stability from primary sequence relied upon the summation of contributions from individual base pairs and helix defects (hairpin loops, interior loops and bulges) (Tinoco *et al.*, 1971); this approach was abandoned as the evidence accumulated that stability depended upon order of base pairs as well as on composition, for instance, two G-C base pairs separated by an A-U pair proved to be less stable than two adjacent G-C pairs flanked by an A-U.

Most subsequent approaches to the prediction of secondary structure stability have employed the nearest neighbour approximation, which assumes that a property of any polymer may be approximated by the summation of nearest neighbour interactions of all its elements; in the case of a nucleic acid double helix, the elements being units of two neighbouring base pairs, for convenience referred to hereafter as "dinucleotide cores." Thus attempts were made to determine the free energy of formation of the ten "dinucleotide cores", in the expectation that summation of the free energies of the contributory dinucleotide cores would provide a reliable estimate of the free energy of formation of the structure as a whole.

Studies of the dinucleotide monophosphates themselves yielded little

satisfaction in this direction. Krugh and coworkers (Krugh *et al.*, 1976) found from NMR data that only rCpG and rGpC dimerized under their conditions, and they could not estimate an equilibrium constant for the latter dimer from their data. Results for 5' monophosphorylated deoxyribonucleotides were more complete, giving an order of stability  $d(pGpG) \cdot d(pCpC) > d(pGpC) \cdot d(pGpC) \gg d(pCpG) \cdot d(pCpG)$  (Young and Krugh, 1975; Krugh and Young, 1975):

Nearest neighbour thermodynamic parameters derived from oligoribonucleotide data were published by Crothers and coworkers (Gralla and Crothers, 1973a, 1973b; Uhlenbeck *et al.*, 1971), and then by Tinoco and Borer and coworkers (Tinoco *et al.*, 1973; Borer *et al.*, 1974). The latter publication contained the results from data for 19 oligomers, which, when fitted to an all-or-nothing model of helix melting, allowed determination of six values for free energies of formation of the ten dinucleotide cores—an average value had to be obtained for certain of the cores since the data set was insufficiently large to provide a unique value for all. Unique values of the free energy of formation were obtained for the three G-C containing cores, GC:GC, GG:CC and CG:CG, -4.3, -4.8 and -3.0 kcal/mol at 25°C, following initiation (Borer *et al.*, 1974).

Various algorithms were derived for the computer prediction of RNA secondary structure (Salser, 1977; Nussinov and Jacobson, 1980; Zuker and Stiegler, 1981; Auren *et al.*, 1982). Most of those applied used oligoribonucleotide data to determine the contributions of the various structure elements. The mediocre performance of these programs in predicting the cloverleaf tRNA structure prompted Ninio and coworkers to a radically different approach (Ninio, 1979; Papanicolaou *et al.*, 1984): They computer generated two sets of parameters, the first based on the assumption that statistical occurrence of dinucleotide cores was indicative of stability,

the second, empirically, by repeated iteration with small changes in each parameter, with intent to maximize prediction of the cloverleaf structure for their set of tRNAs.

They, too, predicted an order of stability of GG:CC > GC:GC > CG:CG, with  $\Delta G$ s of -4.1, -3.6 and -3.1 respectively.

Omstein and Fresco tested the predicted Tms from empirical potential calculations against published data for 16 RNA (Omstein and Fresco, 1983a) and 20 DNA (Omstein and Fresco, 1983b) polynucleotides of repeating sequence, and by adjusting contributions to their empirical potential function and by varying parameters such as base overlap, obtained a fit to the data with correlation coefficient > 0.99. For both RNA and DNA the order of GC containing dinucleotide cores was (descending) GC:GC > GG:CC > CG:CG --note the reversal of the first two cores from the previously established order of GG:CC > GC:GC > CG:CG. Interestingly, while in RNA the three GC containing cores ranked first, second and third in the overall list, in DNA, the corresponding cores ranked first, fourth and sixth, order unchanged, but now interspersed with mixed GC/AT cores, a demonstration of the greater relative stability of dA·dT with respect to dG·dC compared to rA·rU with respect to rG·rC, a difference the authors attributed to the van der Waals contribution of the thymidine methyl group (Omstein and Fresco, 1983a, 1983b).

Subsequent to the study described in this section, samples of certain sequences were used by S. Freier and coworkers in the preparation of perfect hexaribonucleotides containing only G-C base pairs. From the optical melting data for these sequences and others, they obtained free energies of formation (following initiation) for the three GC cores of GC:GC, GG:CC and CG:CG of -3.3, -3.2 and -2.3 kcal/mol respectively, the order of stability now being GC:GC > GG:CC > CG:CG (Freier *et al.*, 1985).

### 3.1.2 Sequence dependence of structure

As originally proposed, the double helix was a flawlessly invariant structure, with each unit of backbone identical in configuration to, and related by a simple rotation-translation to, each one of its neighbours; only the identity of the attached bases differed. As detailed information on structure has become available, through X-ray crystallographic studies (reviewed, Kennard, 1984) of deoxyoligonucleotides and NMR structures of both deoxy and ribonucleotides, it has become apparent that it is not so. Backbone conformation is sequence dependent with significant variation of all parameters around a mean close to the classical values.

Alternating C2' endo and C3' endo sugar puckering was observed in the first paired oligonucleotide structure, d(pATAT) (Viswamitra *et al.*, 1973), and gave rise to the suggestion that a helix could be assembled from repeating two base-pair units. The octamer d(ACGCGCGT)<sub>2</sub> in solution has an alternating structure with dinucleotide repeat, with alternation in the helical twist, and backbone conformation, as determined by NOE measurements complemented by manual model building (Clore and Gronenborn, 1984). The alternating hexamer d(CGTAACG)<sub>2</sub> shows sequence dependent variation, but the repeating backbone unit is a single nucleotide (Clore and Gronenborn, 1983b). A RNA-DNA hybrid, r(GCG)d(TATACGC) (Reid *et al.*, 1983a; Jamin *et al.*, 1985) also crystallizes as a helix with alternating pyrimidine-purine residues, but the difference between the A-type r(GCG):d(CGC) termini and the more flexible d(TATA) core obscures any alternating pattern (Wang *et al.*, 1982a). The flexibility of the d(TATA) region is also observed in the crystal (Shakked *et al.*, 1983) and solution (Jamin *et al.*, 1985) structures of d(GGTATACC)<sub>2</sub>; intriguingly, in the crystal the helix formed is A-type, while the solution NOE data is best fit by a B-type

helix.

The most extreme example of an alternating structure is Z-DNA, in which alternating pyrimidine-purine repeats, with purine  $\times$  *syn* and sugar 2' *endo*, and pyrimidine  $\times$  *anti* and sugar 3' *endo*, are assembled to form a left handed helix, with its characteristic zig-zag backbone (Saenger, 1984). The Z-helix has been seen in crystals of d(CGCGCG) (Wang *et al.*, 1979), d(CGCG) (Drew *et al.*, 1980; Crawford *et al.*, 1980), d(CGCAATGCG) (Fujui *et al.*, 1985), d(TGCGCG) (Kennard, 1985) and d(CGCGTG) (Ho *et al.*, 1985). The latter two contain G-T wobble mispairings.

The dodecamer d(CGCGAATTCGCG) forms a full turn of B-helix and has been studied intensively, both in crystal structure and solution. In the crystal structure (Wing *et al.*, 1982), the twist angle measured between C1' vectors of adjacent bases varies between 27.4° and 40.3° around a mean of 35.8°, with the lowest values being observed for the internal CG sequences. Sugar puckering, base pair tilt and propeller twist are also sequence dependent, obedient to the principle of anticorrelation between members of one base pair. It has been proposed (Callandine, 1983) that the observed reduction of the turn angle and propeller twist in pyr(3'-5')pur sequences relieves Van der Waals contact between purine rings and functional groups resulting from the propeller twist of the bases. Clashes occur in the major groove for pur(3'-5')pyr sequences, and in the minor groove for pyr(3'-5')pur. This proposal was developed into a series of empirical sum functions for the determination of dependence of winding angle, base roll, torsion angle  $\vartheta$  (sugar pucker) and propeller twist (Dickerson, 1983). Similar sequence variations have been observed in solution for d(CGATTATAATCG), where the distances between adjacent AH-2 protons were determined by steady-state NOE measurements to vary as predicted (Patel *et al.*, 1983) and in d(CGTACG)<sub>2</sub> (Clare and Gronenborn, 1983b) by restrained least-squares



refinement of the structure on the basis of interproton distance data (by NOE).

Results of these studies, and of a theoretical  $^1\text{H-NMR}$  study on  $d(\text{CGCGAATTCGCG})_2$  (Giessner-Prettre and Pullman, 1982) indicate that the local variations in structure are less pronounced in solution than in the crystal. Solution structures of two other sequences, however, do not conform to the sum rules (Clare and Gronenborn, 1985).

A number of DNA sequences crystallize in the A-form:  $d(\text{GGGGCCCC})$  (McCall *et al.*, 1984),  $d(\text{GGCCGGCC})$  (Conner *et al.*, 1982),  $d(\text{CCGG})$  (Wang *et al.*, 1982b),  $r(\text{GCG})d(\text{TATACGC})$  (Wang *et al.*, 1982a) and the helical stretches in tRNA. Greater uniformity of structure is expected—and has been observed—since both the displacement of the helix axis, and the  $20^\circ$  tilt of the base pairs eliminates major groove purine clash, leaving only the pyr(3'-5')pur clash in the minor groove (Saenger, 1984). Propeller twist and roll angle only can be predicted by the sum rules (Dickerson, 1983).

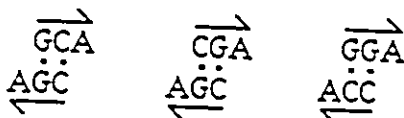
The solution structure of a single oligoribonucleotide duplex CUGUG:CACAG has been determined by restrained least squares refinement to fit NOE-determined distances. While there is local variation of structure it does not conform to the sum rules (Clare *et al.*, 1985).

Bubienko *et al.* (1983) have employed solution NMR data from oligoribonucleotide duplexes in an alternate approach: They adjusted the winding angle (the angle between the glycosidic bonds of neighbouring residues) so that the proton chemical shift values calculated (for A'-RNA geometry) using base isoshielding contours best fitted data published for assorted short oligoribonucleotide helices. The new winding angles were sequence dependent:  $20 \pm 10^\circ$  for RY:RY,  $50 \pm 10^\circ$  for RR:YY and  $45 \pm 10^\circ$  for YR:YR, where R is purine and Y, pyrimidine. (For YR:YR sequences this is in direct contradiction to the decreased winding angle

proposed for d(YR):d(YR) from the crystal structure data.) In consequence, the stacking interaction in RR:YY and YR:YR cores is exclusively interstrand—the two 3' residues stack upon each other—and that in RY:RY cores is exclusively intrastrand—the purine stacks upon the pyrimidine (Figure 3.1). By this model any RNA duplex coheres through the interaction of stacked segments of two (any of the dinucleotide cores in isolation) to four (in the sequence RYRY) bases, the length and nature of the stacks determining duplex stability. In addition, the variation of winding angle, and the resulting irregularity of the surface contours of the helix, was proposed as a feature for enzyme recognition.

### 3.1.3 Studies of the sequence dependence of stability of GC-containing tetramers

At the time at which this study commenced, the set of experimentally determined stabilities of the three GC oligoribonucleotides was over eight years old, and, moreover, had been derived from sequences primarily comprised of A•U base pairs (Borer *et al.* 1974). The initial approach to the determination of relative stabilities was to complete the set of duplexes,



the simplest system in which all three GC cores could be studied. GCA was already known to form a duplex, and CGA not do do so (Alkema *et al.* 1982), accordingly the sequences GGA and CCA were synthesized and their behaviour in a mixture studied. It emerged that in 1.0 M salt buffer, GGA self aggregation was preferred over duplex formation.

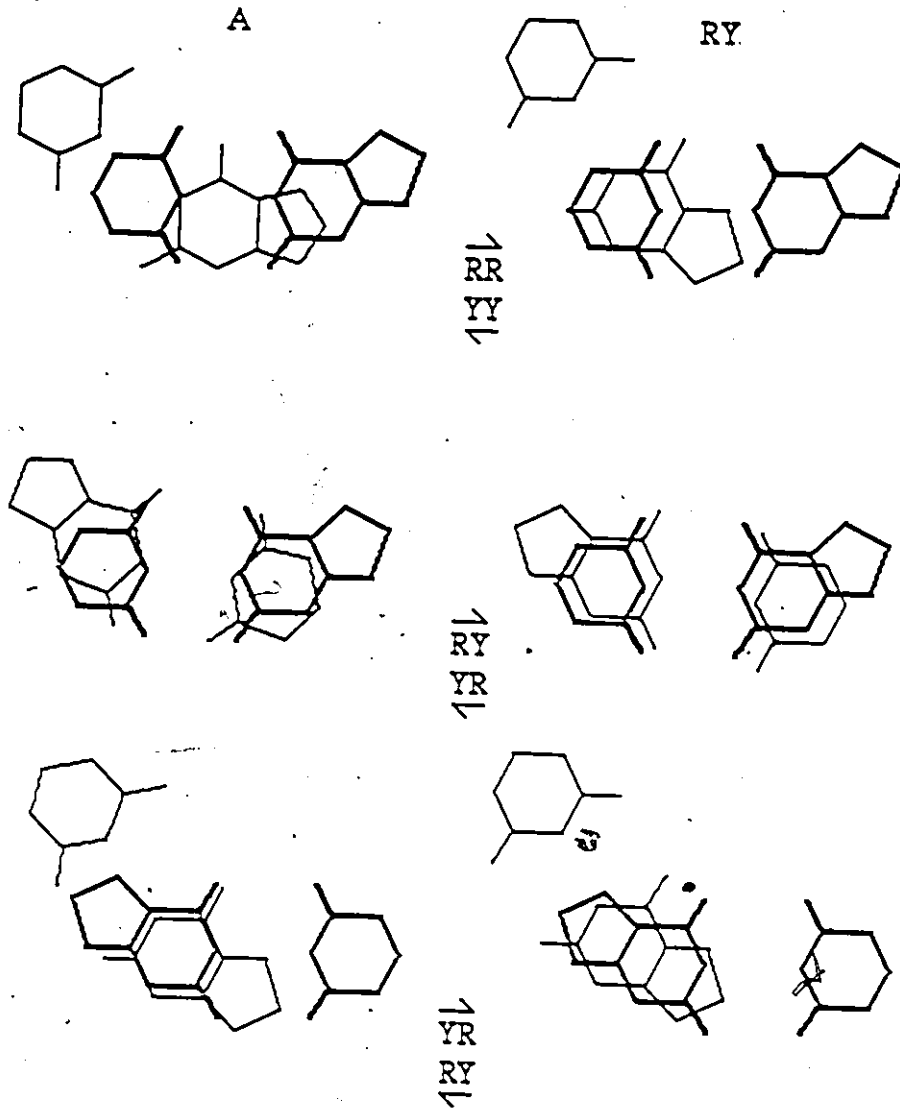
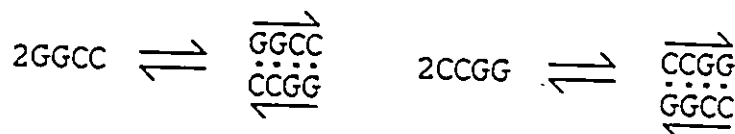


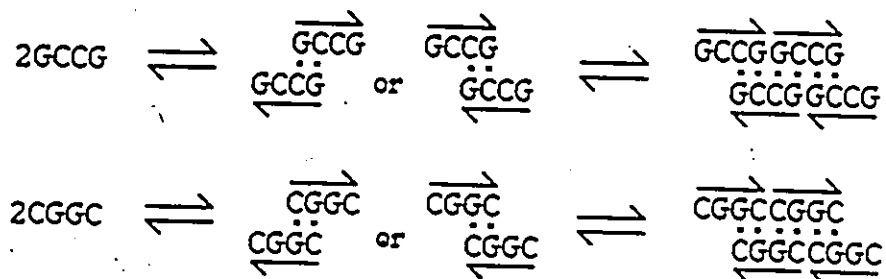
Figure 3.1: The base stacking for the three cores RR:YY, RY:RY and YR:YR in the classical A' helix and in the RY model (after Bubenko *et al.*, 1983)

Longer sequences were required; the core stabilities would then have to be extracted using the nearest neighbour approximation. The sequences CCGG, CGGC and CGCG had been synthesized and studied at low NMR resolution by Dr. Paul Romaniuk (Romaniuk, 1979). For this study the synthesis of GGCC, GCCG and GCGC completed the set of self-complementary tetramers each containing two guanosine and two cytidine residues, and the secondary structure of all six was studied by variable temperature  $^1\text{H-NMR}$  of the exchangeable and non-exchangeable protons at 250 MHz resolution. Subsequently deprotected samples of certain of these sequences were sent to Dr. S. Freier and Dr. D.H. Turner in Rochester, who determined the melting thermodynamics and used them as precursors in the enzymatic synthesis of self-complementary hexamers; their results for the thermodynamic parameters of the three GC cores were published as Freier *et al.*, 1985a.

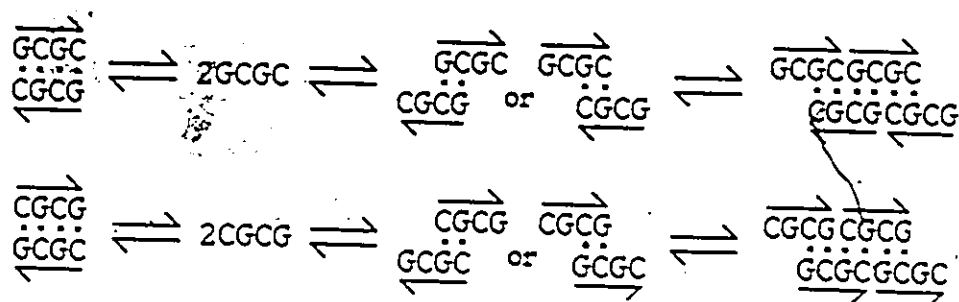
The sequences may be grouped for comparison according to the type of secondary structure available to them. GGCC and CCGG may form only discrete duplexes comprising four base pairs, three dinucleotide cores.



GCCG and CGGC may form only staggered duplexes, comprising two base pairs, and four double-dangling bases, which may be either 3' or 5' dangling, and which may then overlap to form an extended loose duplex.

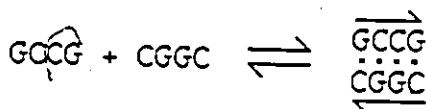


GCGC and CGCG may form either discrete duplexes, or staggered duplexes with 3' or 5' dangling bases, with aggregation.



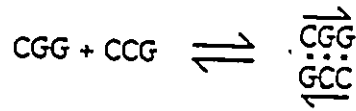
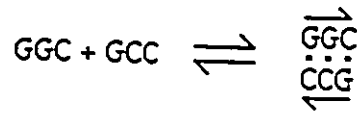
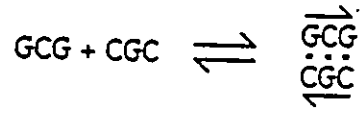
The stability of duplexes of the same type, i.e., perfect or staggered, can be compared, and the relative contribution of each of the three types of cores assessed.

The duplexing behaviour of the mixture of GCCG with CGGC was also studied; the stability of the duplex formed provided a check on the deduced order of stability for the constituent cores.



Mixing experiments were also performed on with the complementary trimers (tetramer precursors) GCG with CGC, GGC with GCC and CGG with CCG, which potentially form three base pair duplexes, and were expected to provide a further

check on the deduced order of stabilities.



## 3.2 Results

### 3.2.1 Assignment of proton spectra

The 70°C spectra were assigned by incremental analysis of the sequence precursors, and the assignment verified by computer prediction of single strand chemical shifts. Results for the two methods are shown in Tables 2.4 and 2.5, and the tetramer assignments summarized in Table 3.1. Agreement between the observed and the calculated shifts was good in most cases, but as the shielding parameters have been derived largely from trimer data, of which only two duplex with  $T_m$  greater than 0°C, the predictions at 70°C tended to err on the high (deshielded) side for sequences which form high-melting duplexes, as do GGCC, CCGG and GCGC. The chemical shift versus temperature curves of protons in those sequences have appreciable slope at 70°C, indicating that base stacking has already begun; it is not surprising that they should deviate from predicted values.

### 3.2.2 Assignment of CC by 2-D shift correlated spectroscopy (COSY)

Upon initial assignment of CCGG by incremental methods, the behaviour of the coupling constants and chemical shifts with temperature of C(1)H-1' and C(2)H-1' as assigned were found to be inconsistent with the assignments—related collapse of  $J_{1'2'}$  and non-sigmoidal curves being more characteristic of terminal H-1's than internal. The incremental analysis of CC-CCG-CCGG had been without anomaly, accordingly the assignments for CC itself were reexamined.

Table 3.1: Summary of chemical shift assignments for the base and anomeric protons of the GC tetramers at 70°C in 1.0 M salt buffer.

Sequence Proton	Chemical shift (ppm)					
	<u>GGCC</u>	<u>CCGG</u>	<u>GCCG</u>	<u>CGGC</u>	<u>GCGC</u>	<u>CGCG</u>
G(1)H-8	7.910		7.952		7.936	
C(1)H-6		7.746		7.666		7.792
G(2)H-8	7.949			7.935		7.968
C(2)H-6		7.710	7.739		7.698	
G(3)H-8		7.918		7.951	7.942	
C(3)H-6	7.740		7.733			7.705
G(4)H-8		7.957	7.976			7.968
C(4)H-6	7.787			7.764	7.760	
C(1)H-5		5.970		5.953		5.919
C(2)H-5		5.935	5.841		5.842	
C(3)H-5	5.828		5.943			5.851
C(4)H-5	5.967			5.904	5.887	
G(1)H-1'	5.825		5.852		5.827	
C(1)H-1'		5.868		5.798		5.856
G(2)H-1'	5.851			5.783		5.800
C(2)H-1'		5.795	5.868		5.893	
G(3)H-1'		5.814		5.854	5.851	
C(3)H-1'	5.870		5.868			5.885
G(4)H-1'		5.840	5.868			5.816
C(4)H-1'	5.887			5.895	5.893	



A spin correlated spectrum of CC (7 mM in 0.1M salt buffer) was obtained using a standard Bruker microprogram. In the COSY spectrum the normal spectrum is projected along the diagonal, and spin-coupled signals show cross peaks at the intersection of the horizontal and vertical perpendiculars from both members of the system. The COSY spectrum of CC is shown in Figure 3.2. The ribose protons are extensively overlapped; the interpretation proceeds as follows:

The H-5' and H-5'' of the 5' terminus appears most upfield, as a distinct multiplet pattern. The coupling between C(1)H-5' and C(1)H-5'' is responsible for the tight square in the upper right hand corner of the spectrum. Coupling to C(1)H-4' gives two cross-peaks, which place the C(4)H-1' within the largest group of peaks, second after the water peak. Extension of the perpendicular dropped from the H-5' to H-4' crosspeak bisects the crosspeak between the second system and the first, indicating coupling between H-4' in the second system and H-3' in the first. The only crosspeaks exhibited by that first system are between the H-1' signals (extreme left) and the second, indicating that the C(1)H-2' is overlapped with C(1)H-4' and belongs to that second system. Of the two H-1' signals, the downfield H-1' signal exhibits a connectivity to the second system after the water peak, while the upfield H-1' has a crosspeak to the first. It is the downfield H-1' therefore, that belongs to the ribose system containing the free 5' terminal hydroxyl group.

The assignments can be checked using the relative positions of the H-3' protons in the two residues. The H-3' of the carbon to which the phosphorus is attached, i.e., C(1), will be expected to appear more deshielded than the H-3' at the 3' terminus. C(1)H-3' has already been located within the first system upfield from the water signal. C(2)H-1' exhibits a connectivity to the first system of peaks, indicating C(2)H-2' is also

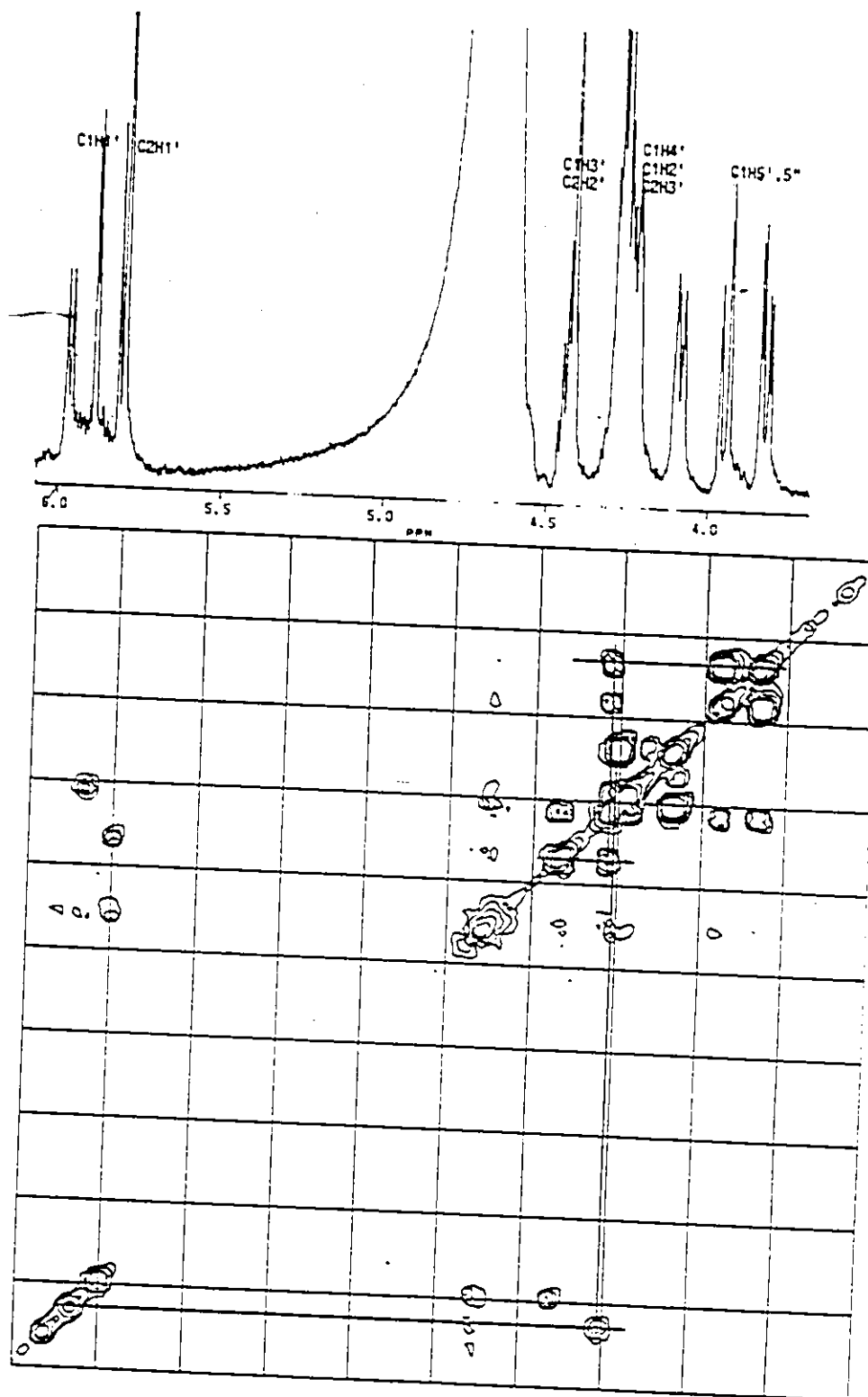


Figure 3.2 (top) Spectrum of ribose region of CC, 7mM in 0.1M salt buffer. (bottom) COSY spectrum for CC, with connectivities throughout C(1) ribose marked.

part of it. C(2)H-2' will be coupled to C(2)H-3', but the only crosspeak possible for this coupling appears between the first peak and the second, therefore C(2)H-3' is either superimposed upon C(2)H-2', which is not indicated by the area under the peaks in system one, or part of system two. Therefore C(1)H-3', as assigned, appears in system one and C(2)H-3' in system two, the first downfield of the second, as expected.

The H-1' assignments, made at 30°C, can be related by variable temperature data to the assignments at 70°C (Figure 3.3); there are no crossovers, so at 70°C the upfield peak is assigned to C(2)H-1' and the downfield to C(1)H-1'.

### 3.2.3 Assignments of imino proton resonances

The accepted convention for the assignment of the hydrogen-bonded imino protons observed for oligoribonucleotides in 90:10 H<sub>2</sub>O: D<sub>2</sub>O at low temperature is that the resonances which broaden first are either adjacent to the ends, or to a site of destabilization, such as a base-base mismatch, the rationale being that the broadening is due to exchange, the exchange to the transient opening of the base pair, and the opening occurs most readily at less constrained sites (Patel and Hilbers, 1975; Fritzsche *et al.*, 1983). Most of the sequences for which imino protons were observed had four base pairs related by an axis of rotation accordingly, two imino proton signals were observed, and one preceded the other in broadening; it was assigned to the terminal base pair, and the other to the internal base pair (Tables 3.2 to 3.4; Figures 3.4 to 3.6). Thus for GCGC (Table 3.4; Figure 3.6) at 2°C two signals were observed, one at 12.77 ppm which was already beginning to broaden at 9°C, and the other at 13.05 ppm, which did not begin to broaden until above 19°C; the first was assigned to the terminal base pair,

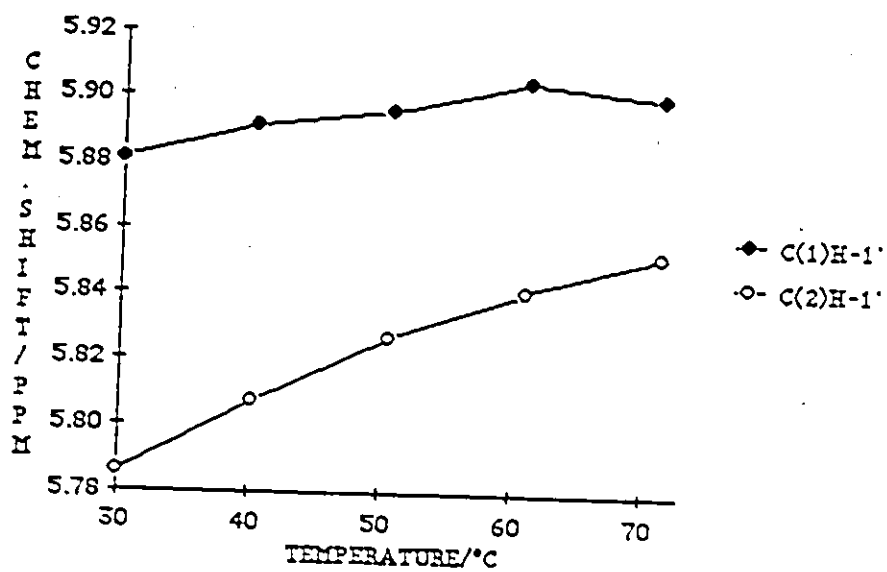


Figure 3.3: Chemical shift versus temperature curves for the anomeric protons of CC, showing correlation of assignments at 70°C with those at 30°C.

Table 3.2: Chemical shifts of hydrogen-bonded imino protons of GGCC, (a) at 1.5mM, in 1.0M salt buffer, (b), at 2.8mM, in 0.1M salt buffer, (c) at 2.5mM, in no salt buffer.

(a)	Chemical shift (ppm)				
Temperature	<u>1.0°</u>	<u>8.7°</u>	<u>19.4°</u>	<u>29.2°</u>	
Proton					
Internal	13.52	13.54	13.39	13.39	
Terminal	12.32	12.40	12.55	12.32	
(b)	<u>4.5°</u>	<u>6.1°</u>	<u>14.5°</u>	<u>25.0°</u>	<u>30.0°</u>
Internal	13.49	13.46	13.44	13.37	13.26
Terminal	12.40	12.44	12.47	12.47	--
(c)	<u>4.3°</u>	<u>5.7°</u>	<u>15.2°</u>	<u>25.1°</u>	
Internal	13.55	13.47	13.41	13.16	
Terminal	12.45	12.46	--	--	

Table 3.3: Assignments of hydrogen-bonded imino protons of CCGG, 2.5mM, in 0.1M salt buffer.

Temperature	Chemical shift (ppm)			
	<u>-3.4°</u>	<u>7.5°</u>	<u>17.0°</u>	<u>25.0°</u>
Proton				
Terminal	13.34	13.30	13.25	13.05
Internal	12.53	12.53	12.46	12.39

Table 3.4: Chemical shifts for the hydrogen-bonded imino protons of GCGC, 2.5mM, in 0.1M salt buffer.

Temperature	Chemical shift (ppm)			
	<u>2.0°</u>	<u>9.2°</u>	<u>19.0°</u>	<u>31.0°</u>
Proton				
Terminal	12.76	12.84	12.87	(13.0)
Internal	13.05	13.06	13.04	(13.0)

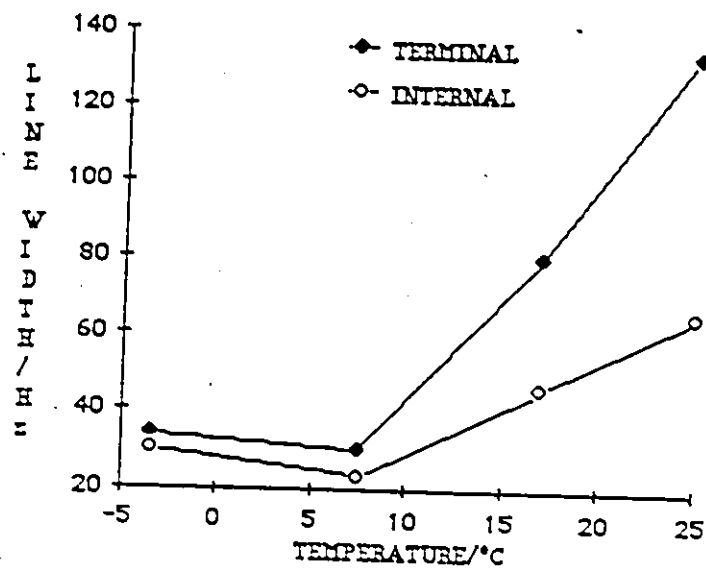


Figure 3.4: Line widths versus temperature of the hydrogen-bonded imino protons of GGCC, 2.5 mM, in 0.1 M salt buffer (0.1 M sodium chloride, 0.01 M sodium phosphate, pH 7.0).

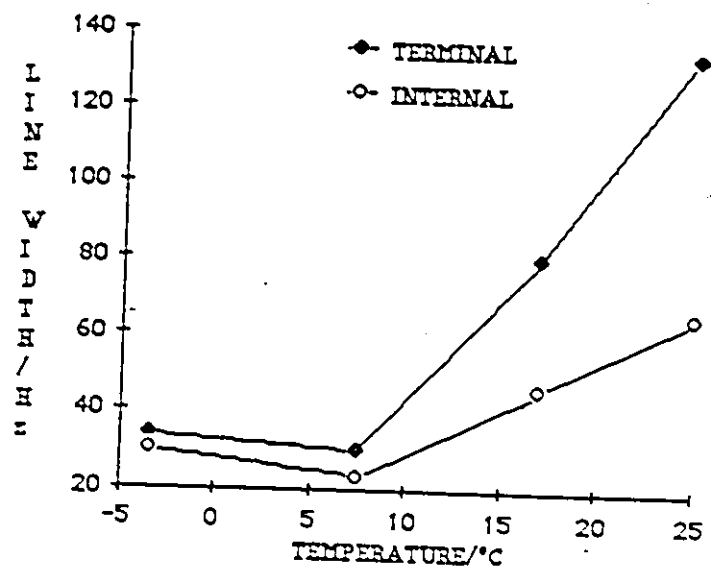


Figure 3.5: Line widths versus temperature of the hydrogen-bonded imino protons of CCGG, 2.5 mM, in 0.1 M salt buffer.



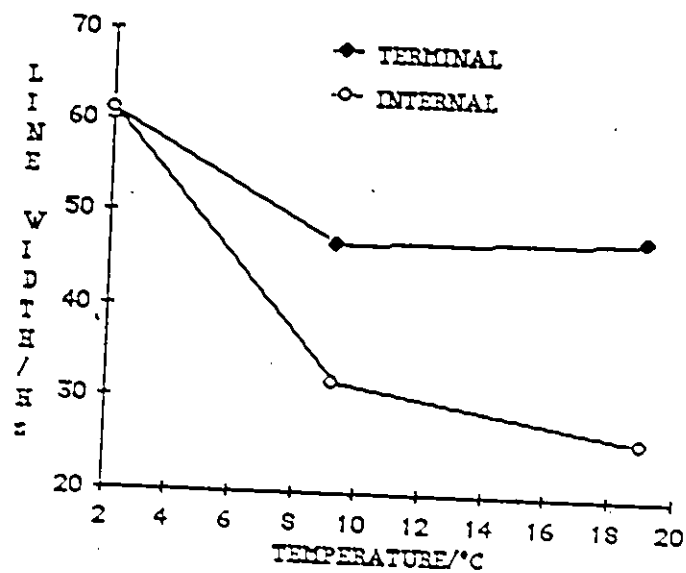
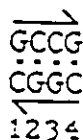


Figure 3.6: Line widths versus temperature of the hydrogen-bonded imino protons of GCGC, 2.5 mM, in 0.1 M salt buffer.

the second to the internal. Similar reasoning was applied to the results for GGCC, CCGG and CGCG. Assignment of resonances for the duplex formed by GCCG:CGGC proved more problematic. The molecule lacks an axis of rotation, and a distinct signal was expected for each of the four base pairs. Three resonances were observed (see Table 3.7), but the highest field signal had an area of approximately double the others, an area which decreased to unitary value upon warming, demonstrating that the peak comprised the overlapping signals from one terminal and one internal base pair. Signals at 12.84 and 12.63ppm were assigned to the terminal base pairs by their early broadening and disappearance (measurements of line widths not possible on account of overlap of signals) while the remaining two signals, at 13.37 and 12.63ppm, were assigned to the internal base pairs. The right-handed sense of the double helix results in the 3' neighbour being the most influential as regards shielding of the hydrogen bonded imino proton (Kearns, 1976; Arter and Schmidt, 1976). One terminal base pair has a cytidine residue 3' to it, the other a guanosine. Guanine has greater shielding capacity than cytosine; accordingly, the upfield terminal imino signal is assigned to the base pair with a 3' neighbouring guanosine (base pair 4 below), and the downfield signal that with a 3' neighbouring cytidine (base pair 1). The same reasoning was applied to the internal base pairs, one of which (base pair 3) is flanked by two guanosines, and is therefore expected to occur upfield of the other (base pair 2), which has two 3' cytidine neighbours.



Comparison of the observed imino proton chemical shifts with those predicted

Table 3.5: Chemical shift assignments for the hydrogen-bonded imino protons of CGCG, 2.5mM, 1.0M salt buffer, at 2.0°C

Chemical shift (ppm)	
<u>Temperature</u>	<u>2.0°</u>
Proton	
Terminal	12.87
Internal	13.15

Table 3.6: Chemical shift assignment for the hydrogen-bonded imino protons of GCCG, 1.9mM, in 1.0M salt buffer, over the temperature range -5 to 15°C.

Proton	Chemical shift (ppm)		
	<u>-4.5°</u>	<u>6.1°</u>	<u>15.0°</u>
Duplex	12.60	12.58	-
Overlap	13.36	13.30	-

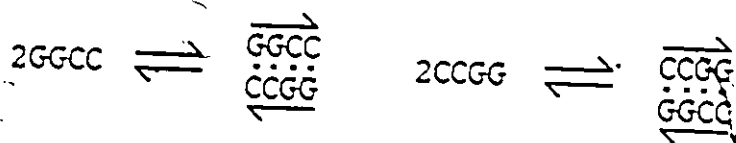
Table 3.7: Chemical shift assignments for the hydrogen-bonded imino protons of GCCG:CGGC, 1.7mM each strand, in 1.0M salt buffer, over the temperature range 1° to 31°C. Numbering is from the 5' end of GCCG.

Temperature	Chemical shift (ppm)			
	<u>1.0°</u>	<u>8.7°</u>	<u>19.4°</u>	<u>31.0°</u>
Proton				
G·C1	12.84	12.87	12.84	-
G·C2	13.34	13.38	13.34	-
G·C3	12.63	12.66	12.62	-
G·C4	12.63	12.66	12.49	-

based on isoshielding contours and classical helical geometry reveals some interesting differences. Section 3.3.7 is dedicated to discussion of these differences, in these and other sequences.

### 3.2.4. Secondary structure and stability of the six GC containing tetramers.

GGCC and CCGG have a single simple secondary structure available to them: discrete duplexes with four base pairs:



The chemical shift versus temperature curves of the nonexchangeable base and anomeric protons are sigmoidal for the majority of the ten such protons in each (see Figures 3.7 and 3.8); nonsigmoidal behaviour is most conspicuous in the curves of protons belonging to the 5' terminal base, which, according to the H1' data, stacks poorly (see section 3.3.4), and the H1' protons, some of which are outside the shielding influence of the base stack. For GGCC, individual proton Tms occur over the range 49.9 - 55.8°C, with an average of 54°C (see Table 3.8). For CCGG, five protons have Tms 45.3 - 49.8°C, while the sixth, C1H6, has a Tm of 54.2°C; the average Tm is 47.6°C (Table 3.9). Repetition of these studies in 0.1M NaCl buffer gives identical results if allowance is made for the depression of Tm by the lower ionic strength: at 2.5mM in 0.1M salt buffer, the Tm of GGCC is 52.1°C, and that of CCGG, 44.1°C (see Section 4.2). Low temperature spectra of the hydrogen bonded imino protons show two distinct resonances for each sequence, one of which disappears before the other with rising temperature. For GGCC

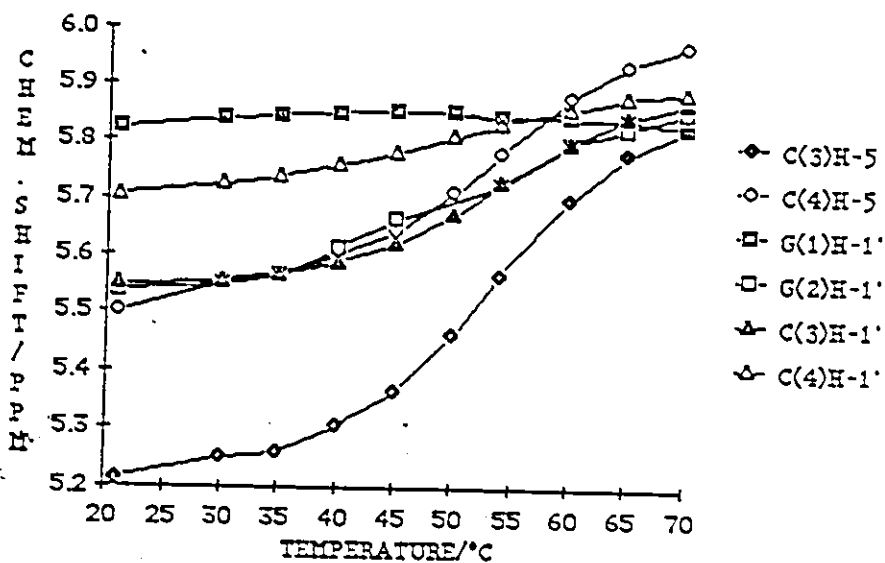
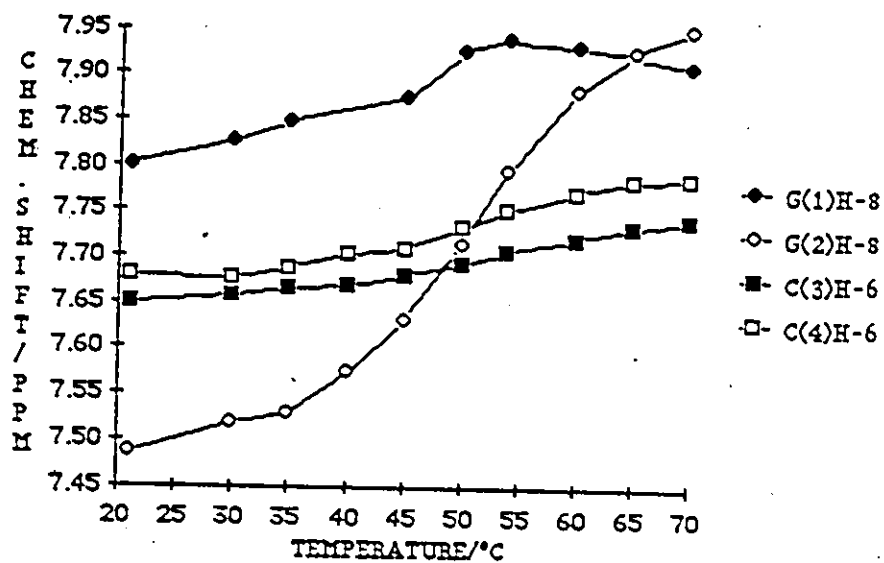


Figure 3.7: Chemical shift versus temperature curves for the base and anomeric protons of GGCC, 1.7 mM, in 1.0 M salt buffer (1.0 M sodium chloride, 0.01 M sodium phosphate, pH 7.0).

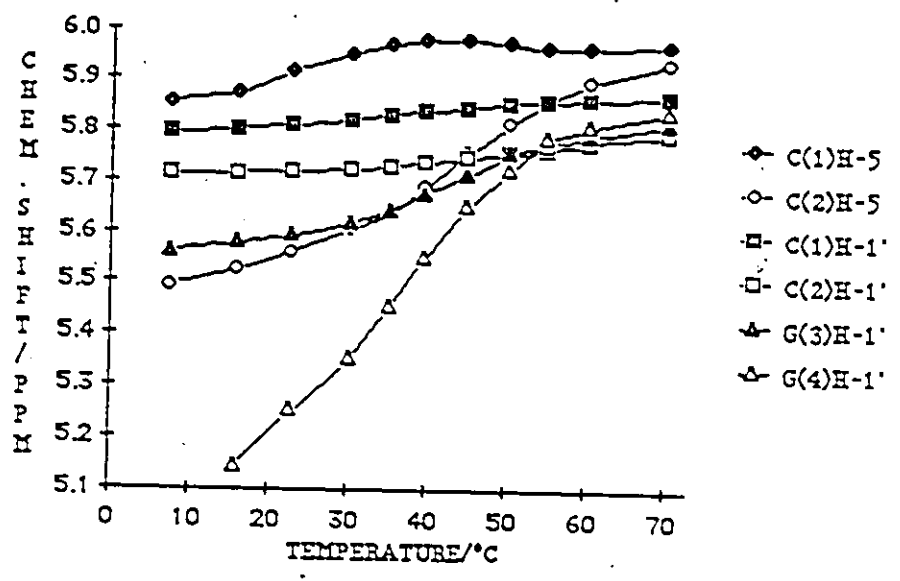
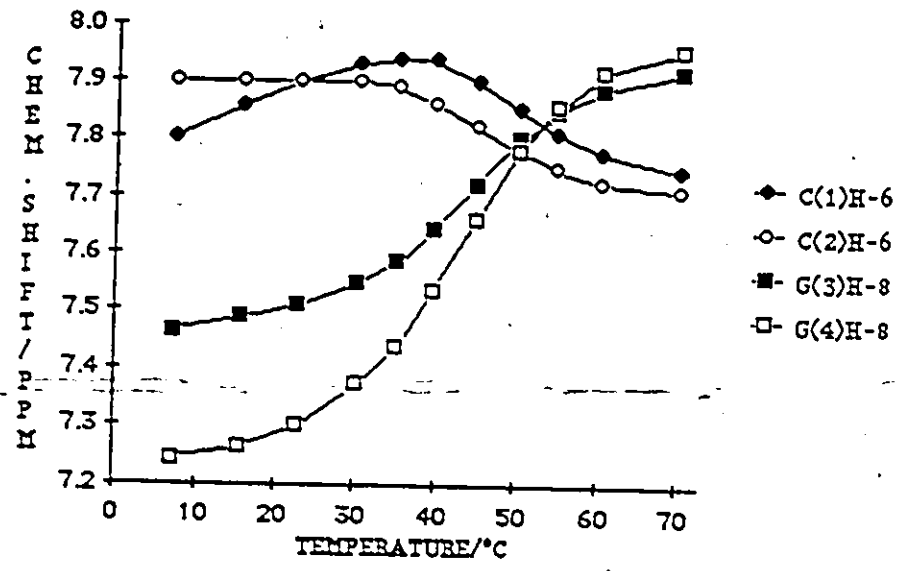


Figure 3.8: Chemical shift versus temperature curves for the base and anomeric protons of CCGG, 1.9 mM, in 1.0 M salt buffer.

Table 3.9: Chemical shift assignments and anomeric proton coupling constants versus temperature for protons of CCGG, 1.9mM, in 1.0M buffer over the temperature range 70 - 7°C. NSB denotes no sigmoidal behaviour.

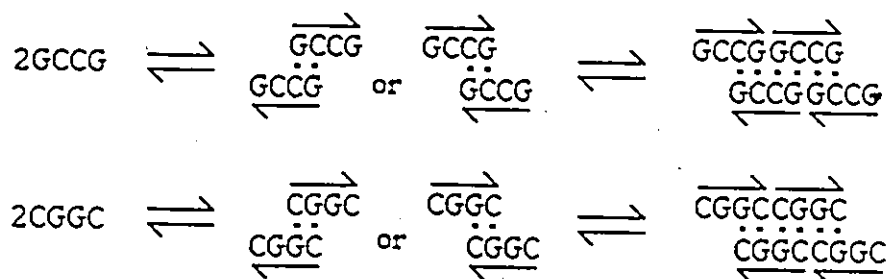
Proton	Chemical shift (ppm)											T <sub>m</sub>
	70.3°	60.3°	54.8°	50.3°	45.0°	39.7°	35.1°	30.2°	22.7°	15.7°	7.3°	
C(1)H-6	7.746	7.776	7.812	7.853	7.901	7.937	7.940	7.930	7.899	7.857	7.803	54.2
C(2)H-6	7.710	7.725	7.748	7.779	7.821	7.862	7.890	7.901	7.899	7.902	7.900	49.8
G(3)H-8	7.918	7.887	7.850	7.800	7.722	7.645	7.587	7.551	7.512	7.487	7.463	45.3
G(4)H-8	7.957	7.917	7.857	7.781	7.660	7.538	7.439	7.375	7.300	7.263	7.242	45.3
C(1)H-5	5.970	5.963	5.966	5.974	5.981	5.979	5.967	5.949	5.918	5.875	5.854	NSB
C(2)H-5	5.935	5.895	5.858	5.815	5.756	5.689	5.636	5.606	5.561	5.526	5.494	45.3
C(1)H-1'	5.868	5.863	5.857	5.854	5.846	5.837	5.829	5.822	5.810	5.803	5.797	NSB
C(2)H-1'	5.795	5.777	5.764	5.758	5.747	5.737	5.729	7.724	5.718	5.715	5.712	NSB
G(3)H-1'	5.814	5.790	5.772	5.758	5.712	5.674	5.642	5.619	5.594	5.579	5.560	NSB
G(4)H-1'	5.840	5.811	5.784	5.723	5.652	5.555	5.453	5.355	5.250	5.142	--	45.8
											Average T <sub>m</sub>	47.6
Coupling constant (Hz)												
C(1)H-1'	5.1	4.9	4.4	4.1	2.8	1.3	1.7	--	--	--	--	
C(2)H-1'	5.5	4.5	2.3	3.3	2.5	--	--	--	--	--	--	
G(3)H-1'	4.3	3.2	2.1	3.3	2.0	--	--	--	--	--	--	
G(4)H-1'	4.8	4.5	3.8	2.7	1.2	--	--	--	--	--	--	





at 2.5mM in 0.1M salt, the signals occur at 13.49 and 12.39ppm (-4.5°C), with the latter being the first to broaden, and therefore belonging to the terminal base pairs. For CCGG at 2.5mM in 1.0M salt, the resonances appear at 12.53 (internal) and 13.34ppm (terminal) at -5°C. This is consistent with formation of a four-base-pair structure with an axis of rotation; the assignments and variable temperature data are provided in Tables 3.2 and 3.3. The high average  $T_m$ s, cooperative melting behaviour as indicated by the narrow range of individual proton  $T_m$ s, and the imino proton data, support the formation of the expected four-base-pair perfect duplex by these two sequences.

GCCG and CGGC may only form staggered duplexes, comprising two G-C base pairs and four double dangling bases, with subsequent overlap of the dangling bases to form an extended loose duplex. There are, however, two alternatives for the staggered duplexes for each structure, a duplex with 3' dangling bases (left), or one with 5' dangling bases (right):



The chemical shift versus temperature plots for the protons of GCCG and CGGC are shown in Figures 3.9 and 3.10 respectively. The curves are incomplete, owing to broadening at low temperatures due to formation of high molecular weight aggregates; an average of computer-picked inflection points for five of the ten curves give a  $T_m$  for GCCG of 35.5°C, and one for CGGC of 26.4°C (see data in Tables 3.10 and 3.11). It is possible to discern the preferred direction of staggering through comparison with similar

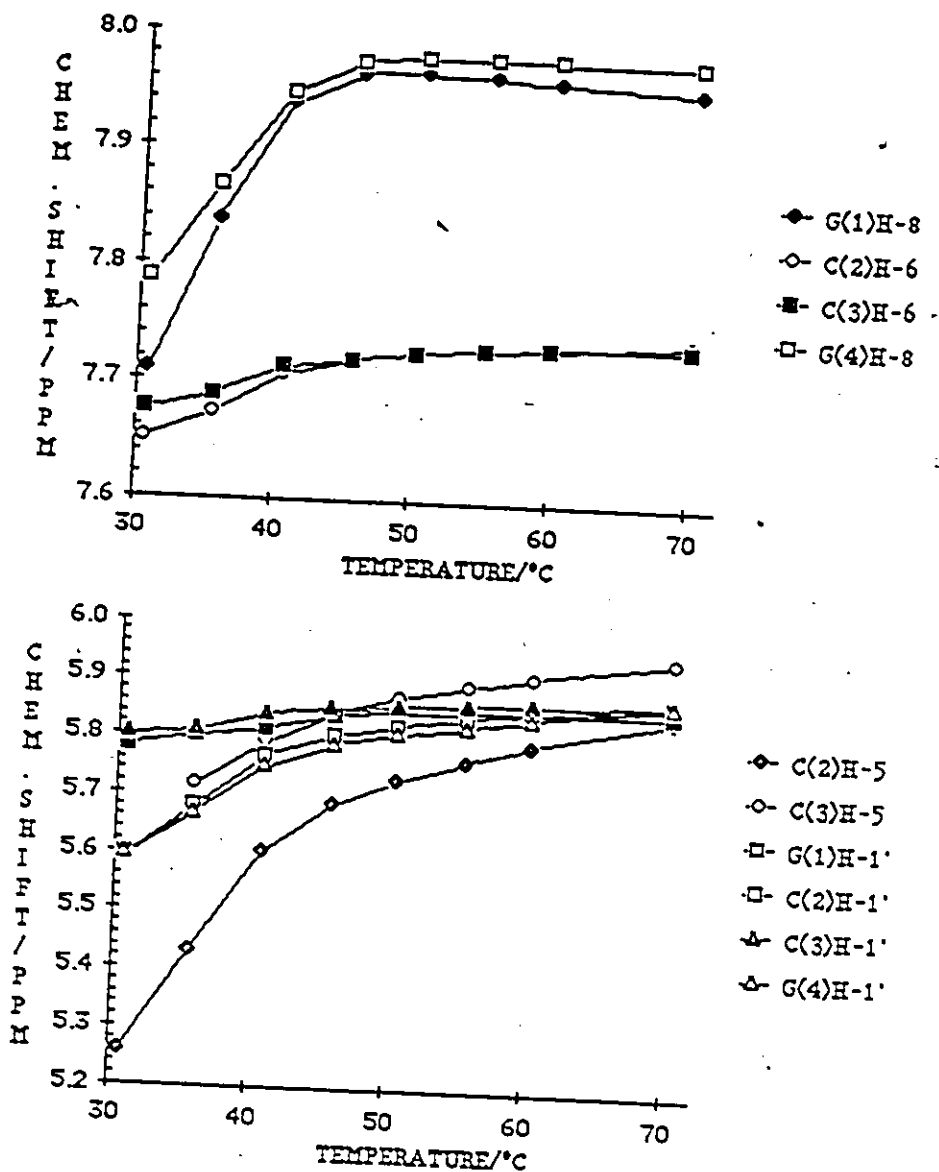


Figure 3.9: Chemical shift versus temperature curves for the base and anomeric protons of GCCG, 2.1 mM, 1.0 M salt buffer.

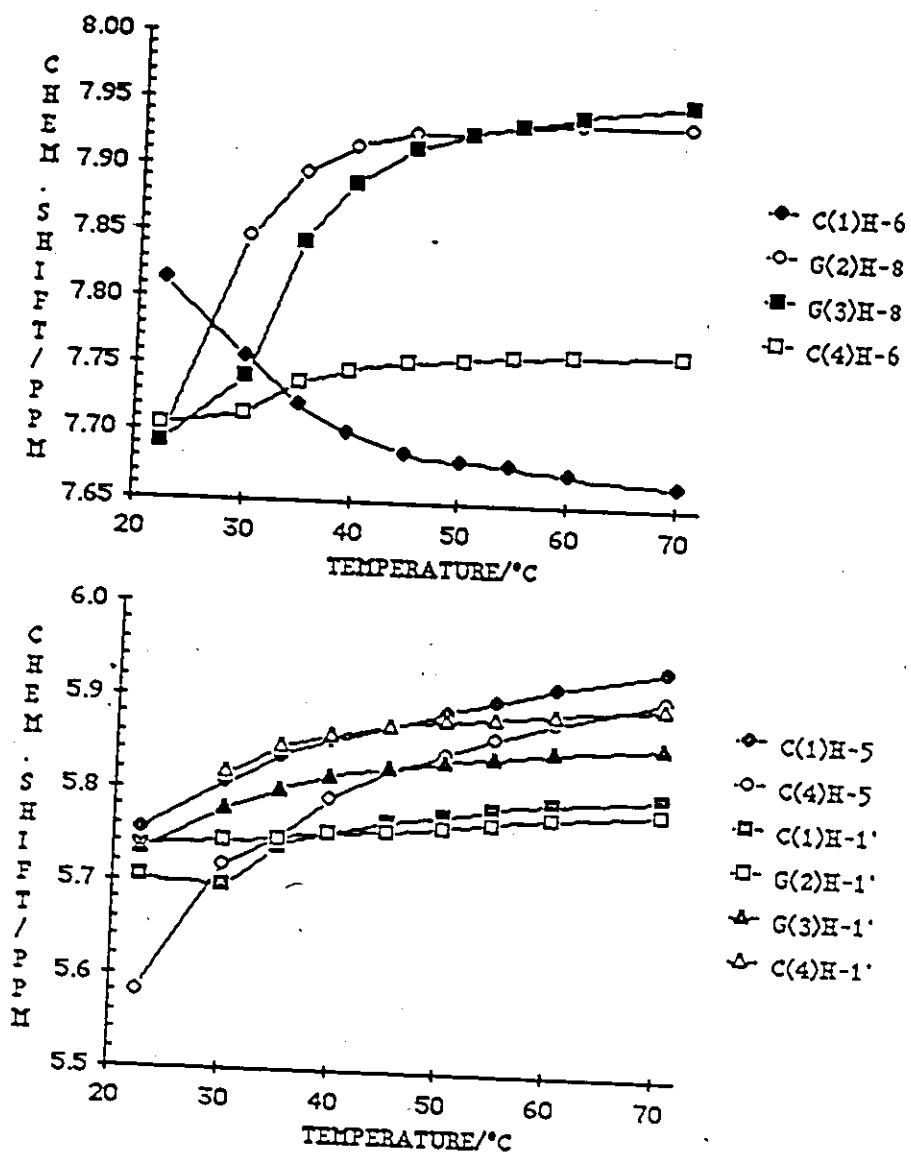


Figure 3.10: Chemical shift versus temperature curves for the base and anomeric protons of CGGC, 2.1 mM, in 1.0 M salt buffer.

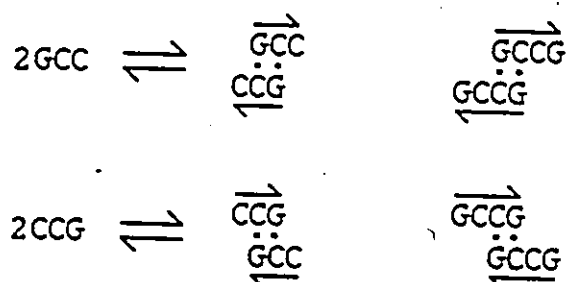
Table 3.10: Chemical shift assignments and anomeric proton coupling constants for the protons of GCCG, 2.1mM. in 1.0M salt buffer over the temperature range 70 - 30°C.

Temp.	Chemical shift (ppm)								T <sub>m</sub>
	70.2°	60.0°	55.3°	50.3°	45.6°	40.7°	35.6°	30.8°	
Proton									
G(1)H-8	7.952	7.960	7.963	7.965	7.964	7.937	7.837	7.708	32.9
C(2)H-6	7.739	7.732	7.729	7.725	7.720	7.706	7.672	7.650	36.4
C(3)H-6	7.733	7.732	7.729	7.725	7.720	7.713	7.688	7.675	36.6
G(4)H-8	7.976	7.978	7.978	7.977	7.973	7.947	7.866	7.786	34.6
C(2)H-5	5.841	5.792	5.764	5.730	5.686	5.603	5.429	5.259	35.0
C(3)H-5	5.943	5.911	5.893	5.871	5.844	5.791	5.714	--	36.5
G(1)H-1'	5.852	5.849	5.847	5.845	5.837	5.814	5.801	5.780	39.5
C(2)H-1'	5.868	5.849	5.839	5.823	5.804	5.768	5.678	5.592	34.7
C(3)H-1'	5.868	5.862	5.859	5.856	5.851	5.840	5.810	5.780	37.3
G(4)H-1'	5.868	5.839	5.823	5.807	5.787	5.749	5.666	5.592	35.3
								Average T <sub>m</sub>	35.5
	Coupling constants (Hz)								
G(1)H-1'	4.4	4.4	4.5	4.9	4.2	3.5	--	--	
C(2)H-1'	4.9	4.4	4.5	4.0	3.9	3.7	3.1	--	
C(3)H-1'	4.9	5.4	5.6	5.3	5.2	5.1	--	--	
G(4)H-1'	4.9	4.2	3.3	3.6	3.4	2.7	3.1	--	

Table 3.11: Chemical shift assignments and anomeric proton coupling constants for the protons of CGGC, 2.1mM, in 1.0M salt buffer over the temperature range 70 - 15°C.

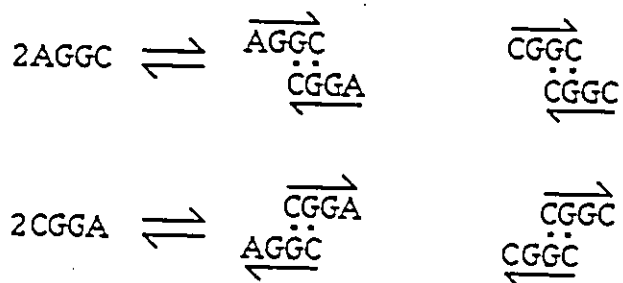
Temp.	Chemical shift (ppm)								T <sub>m</sub>
	70.3°	60.3°	50.5°	39.7°	35.1°	30.2°	22.7°	15.7°	
Proton									
C(1)H-6	7.666	7.673	7.681	7.701	7.721	7.756	7.813	--	28.1
G(2)H-8	7.935	7.934	7.926	7.914	7.894	7.846	7.707	--	NSB
G(3)H-8	7.951	7.941	7.926	7.887	7.843	7.741	7.691	--	31.5
C(4)H-6	7.764	7.762	7.756	7.747	7.737	7.714	7.664	--	24.5
C(1)H-5	5.935	5.914	5.885	5.853	5.835	5.806	5.755	5.723	25.5
C(4)H-5	5.904	5.878	5.842	7.792	5.748	5.717	5.703	5.689	NSB
C(1)H-1'	5.798	5.790	5.775	5.753	5.739	5.696	5.581	--	NSB
G(2)H-1'	5.783	5.774	5.763	5.753	5.746	5.743	5.736	5.711	NSB
G(3)H-1'	5.854	5.845	5.833	5.814	5.799	5.779	5.733	5.695	23.6
C(4)H-1'	5.895	5.86	5.877	5.860	5.845	5.817	5.756	5.720	25.0
							Average T <sub>m</sub>		26.4
	Coupling constants (Hz)								
C(1)H-1'	3.7	3.5	3.8	3.7	3.5	2.4	--	--	
G(2)H-1'	4.7	4.5	4.4	3.7	2.8	1.3	--	--	
G(3)H-1'	4.6	4.0	3.2	3.4	3.4	2.6	--	--	
C(4)H-1'	3.0	3.1	3.6	3.5	3.1	2.8	--	--	

sequences which may form only one of the two possible structures: GCC, which can form a GC:GC core with two 3' dangling cytidines, is a model for the GCCG staggered duplex with 3' dangling bases, while CCG, which can form a CG:CG core with two 5' dangling cytidines, is a model for the staggered duplex with 5' dangling bases.



GCC has a  $T_m$  of 19°C at 7.3mM, in 1.0M salt, while CCG at 8.1mM does not duplex; the more stable structure is that with 3' dangling bases, a not unexpected result (Sinclair *et al.*, 1984).

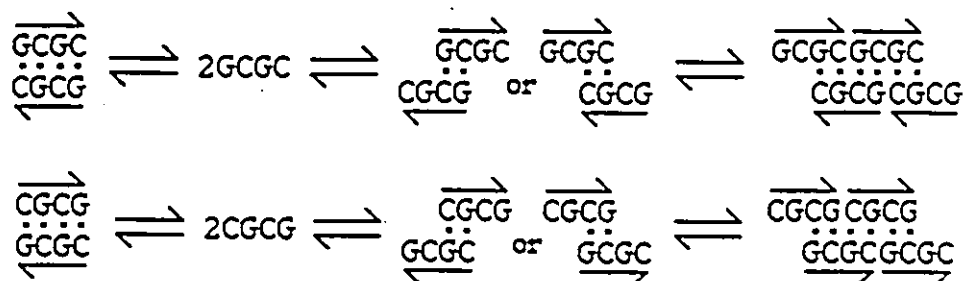
Similar reasoning can be applied for CGGA and AGGC as models for the possible structures of CGGC: CGGA has a  $T_m$  of 20°C at 2mM, while AGGC has no  $T_m$  above 0°C (Sinclair *et al.*, 1984).



Thus for both GCCG and CGGC the secondary structure having four 3'

double-dangling bases is preferred over that having four 5' double-dangling bases.

GCGC and CGCG may (on paper) form either perfect four-base-pair duplexes (left), or staggered duplexes with aggregation.



The chemical shift versus temperature curves of the base and anomeric protons of GCGC show a well defined average  $T_m$  of 49.9°C in 1.0M salt at 1.9mM, for eight of ten curves with a range of  $T_m$ s of 44.5 - 54.7°C (Table 3.12; Figure 3.11). For GCGC at 2.5mM in 0.1M salt, imino proton signals are observed at low temperature, at 13.05 and 12.77ppm (2°C); both persist to at least 19°C, and the downfield to 31°C, though poorly defined. (The  $T_m$  in 0.1M salt is 48.1°C at 2.5mM). The downfield resonance is thus assigned to the internal base pairs. Variable temperature data are given in Table 3.6. The high  $T_m$ —between those of the perfect duplexes GGCC and CCGG—the completeness of the sigmoidal curves, in contrast with those of the staggered duplexes GCCG and CGGC, and the imino proton data are all consistent with formation of the perfect duplex (left, above)

CGCG was originally thought to form a staggered duplex: its  $T_m$  in 1.0M salt is 36.4°C, low compared to the  $T_m$ s of the perfect duplexes GGCC, CCGG and GCGC, and its curves incomplete due to broadening of the signals which made the lower temperatures



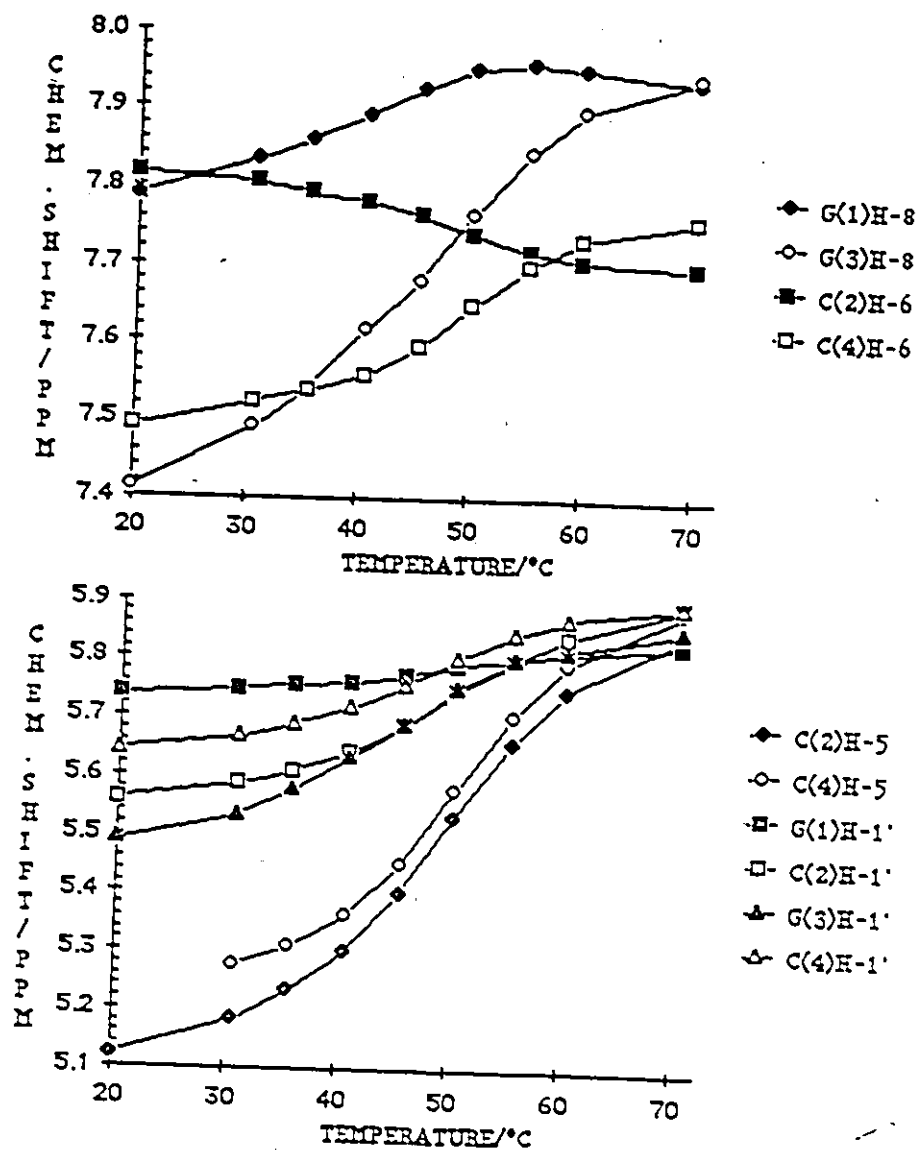


Figure 3.11: Chemical shift versus temperature curves for the base and anomeric protons of GCGC, 1.9 mM, in 1.0 M salt buffer.



unassignable (Figure 3.12; Table 3.13). Published reports which described discrete duplex secondary structure for CGCG in 0.1M salt prompted study of CGCG at low ionic strength. In 0.1M salt resolution improved, allowing more reliable assignment of the lower portions of the curves; the  $T_m$  obtained was 37.2°C, an average of the individual values for seven curves, with a range of 29.5 - 43.5°C (Figure 3.13; Table 3.14). The exchangeable proton spectrum of CGCG shows two distinct signals, at 13.15 (internal) and 12.87ppm (terminal) (Table 3.5). This tends to suggest perfect duplex formation, although the  $T_m$  is still low enough that competition from the aggregate may be significant—GCCG and CGGC, which form only the staggered duplex with aggregation, have  $T_m$ s of 35.5 and 29.2°C respectively in 1.0M salt.

#### 3.2.5. Relative stability of GC cores

The relative stability of the dinucleotide cores may be deduced by comparison of the stabilities of the secondary structures formed by the six self-complementary tetramers with reference to their core composition.

The order of stability of the cores GC:GC and CG:CG, previously determined as being GC:GC more stable than CG:CG, is confirmed: 1. The discrete duplexes formed by GGCC and CCGG have two GG:CC cores in common; they differ, however, in the composition of their central core. GGCC, which has a central GC:GC core, is 6° more stable than CCGG, which has a central CG:CG core (Table 3.15).

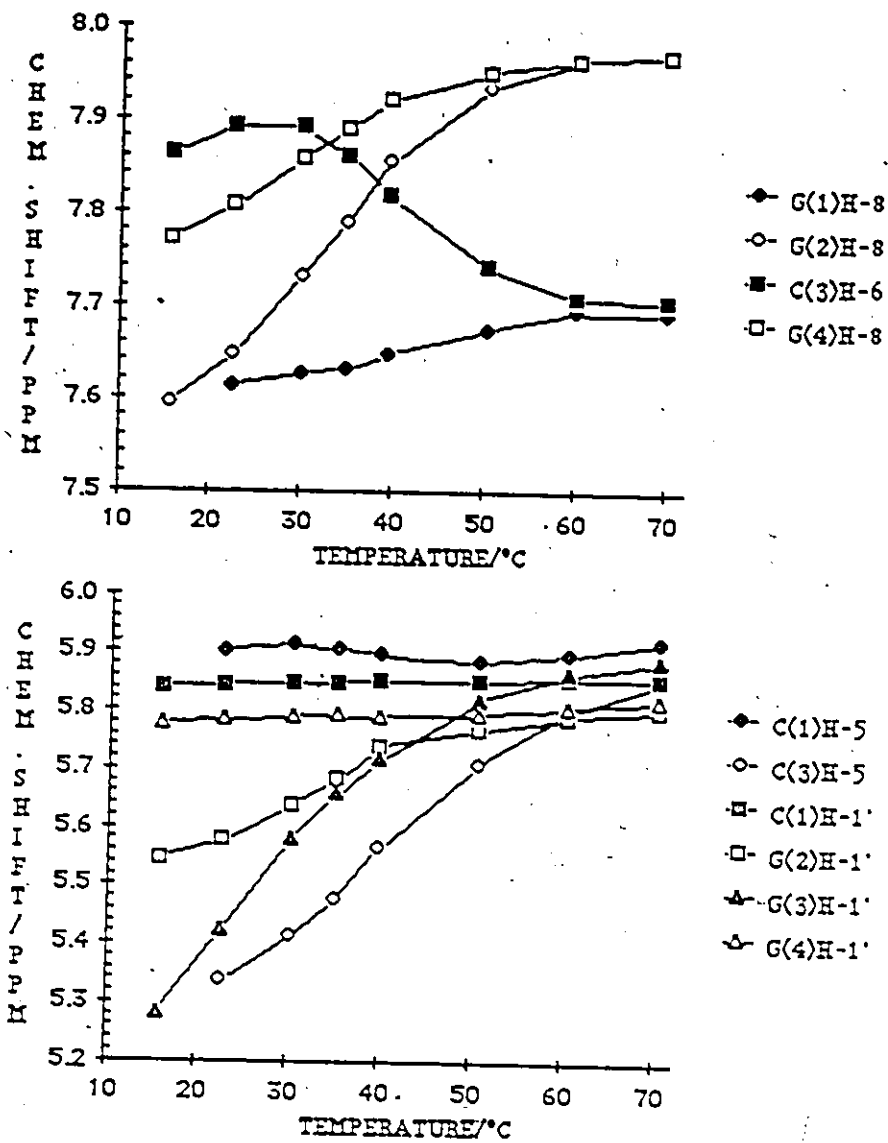


Figure 3.12: Chemical shift versus temperature curves of the base and anomeric protons of CGCG, 1.2 mM, in 1.0 M salt buffer.

Table 3.13: Chemical shift assignments and anomeric proton coupling constants for protons of CGCG, 1.2 mM, in 1.0M salt buffer over the temperature range 70 - 15°C.

Temp.	Chemical shifts (ppm)								T <sub>m</sub>
	70.3°	60.3°	50.5°	39.7°	35.1°	30.2°	22.7°	15.7°	
Proton									
C(1)H-6	7.692	7.693	7.673	7.647	7.630	7.625	7.613	7.578	45.2
G(2)H-8	7.968	7.963	7.933	7.852	7.788	7.730	6.645	7.593	33.5
C(3)H-6	7.705	7.708	7.742	7.817	7.859	7.890	7.890	7.861	41.4
G(4)H-8	7.968	7.963	7.950	7.919	7.888	7.858	7.806	7.771	28.5
C(1)H-5	5.919	5.898	5.883	5.895	5.906	5.913	5.901	5.901	NSB
C(3)H-5	5.851	5.800	5.709	5.563	5.478	5.412	5.335	5.279	38.5
C(1)H-1'	5.856	5.855	5.850	5.849	5.847	5.846	5.842	5.839	NSB
G(2)H-1'	5.800	5.789	5.766	5.737	5.681	5.637	5.575	5.542	30.1
C(3)H-1'	5.885	5.862	5.818	5.713	5.657	5.576	5.422	5.279	23.0
G(4)H-1'	5.816	5.804	5.792	5.788	5.790	5.789	5.783	5.777	NSB
									Average T <sub>m</sub>
									36.2
	Coupling constants (Hz)								
C(1)H-1'	5.11	5.60	4.92	4.01	2.50	2.72	--	--	
G(2)H-1'	3.89	3.62	3.50	2.04	1.08	--	--	--	
C(3)H-1'	4.55	4.03	3.43	2.62	<1	--	--	--	
G(4)H-1'	3.75	4.00	3.27	2.26	--	--	--	--	

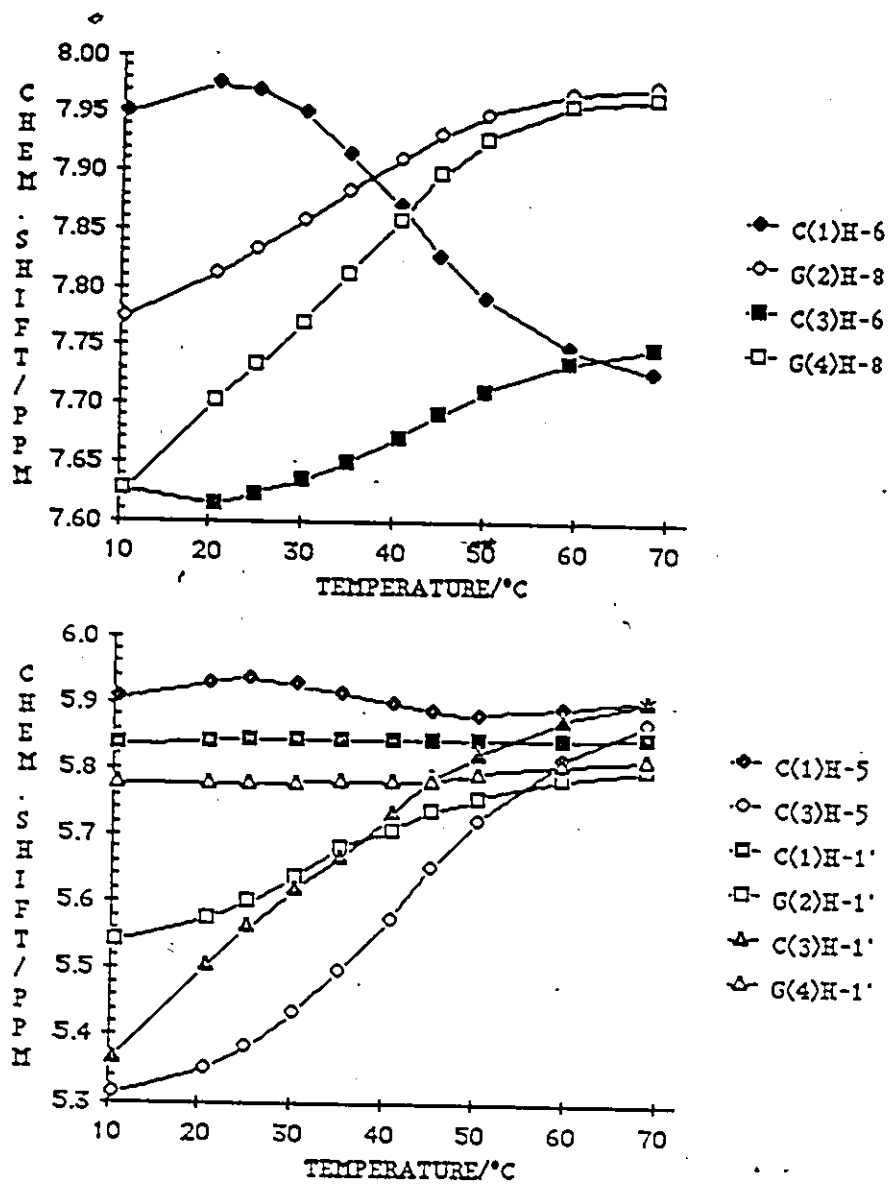
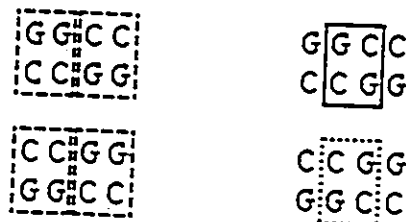


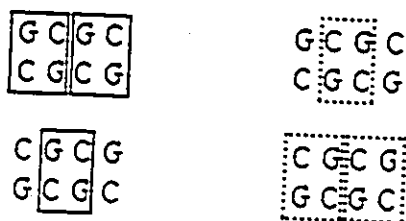
Figure 3.13: Chemical shift versus temperature curves of the base and aromatic protons of CGCG, 2.5 mM, in 0.1 M salt buffer.

Table 3.14: Chemical shifts and anomeric proton coupling constants for protons of CGCG, 2.5mM, in 0.1M salt buffer over the temperature range 70 - 10°C.

Proton	Chemical shift (ppm)										T <sub>m</sub>
	68.5°	59.4°	50.1°	45.1°	40.7°	35.0°	30.1°	25.1°	20.7°	10.7°	
C(1)H-6	7.728	7.748	7.791	7.827	7.871	7.915	7.950	7.970	7.976	7.950	41.3
G(2)H-8	7.974	7.967	7.949	7.932	7.911	7.884	7.858	7.833	7.811	7.774	33.6
C(3)H-6	7.749	7.735	7.710	7.691	7.671	7.650	7.635	7.723	7.614	7.627	43.5
G(4)H-8	7.963	7.956	7.927	7.897	7.857	7.811	7.769	7.733	7.702	7.627	36.4
C(1)H-5	5.908	5.890	5.882	5.887	5.899	5.915	5.927	5.934	5.930	5.906	NSB
C(3)H-5	5.868	5.813	5.721	5.652	5.576	5.496	5.435	5.383	5.349	5.312	42.0
C(1)H-1'	5.849	5.845	5.842	5.842	5.842	5.843	5.843	5.843	5.840	5.836	NSB
G(2)H-1'	5.799	5.784	5.757	5.735	5.706	5.681	5.639	5.602	5.575	5.540	29.5
C(3)H-1'	5.901	5.873	5.823	5.789	5.734	5.665	5.620	5.563	5.503	5.366	35.6
G(4)H-1'	5.815	5.805	5.791	5.781	5.782	5.780	5.779	5.779	5.779	5.777	NSB
Average T <sub>m</sub>											37.4
Proton	Coupling constant (Hz)										
	C(1)H-1'	5.1	5.5	5.2	4.8	4.3	3.1	3.1	2.5	2.4	--
	G(2)H-1'	3.7	3.7	3.1	2.7	2.2	1.6	1.0	--	--	--
	C(3)H-1'	4.2	4.3	4.3	2.8	2.3	1.5	--	--	--	--
	G(4)H-1'	4.2	3.4	3.1	2.8	2.3	1.7	--	--	--	--



2. The discrete duplex formed by GCGC in 0.1M salt is more stable than that formed by CGCG by 11° (Table 3.15). GCGC forms two GC:GC cores, and one CG:CG core, while CGCG forms one GC:GC and two CG:CG.



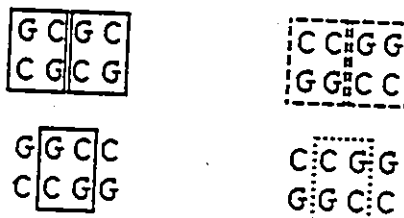
3. The staggered duplex formed by GCCG, which has a single GC:GC core and four 3' double dangling bases, is more stable by 6° than that formed by CGGC, which has a similar structure, but with a CG:CG core (Table 3.15).

The stability of GG:CC appears to be between that of GC:GC and that of CG:CG. GCGC and CCGG both have a central CG:CG core; GCGC has two terminal GC:GC cores, and its  $T_m$  is 2° higher than CCGG, which has two terminal GG:CC cores. Substitution of two GG:CC cores for two GC:GC cores (GCGC to CCGG) causes a smaller decrease in  $T_m$  than substitution of a single CG:CG core for a GC:GC core (GGCC to CCGG).



Table 3.15:  $T_m$  data for the six self-complementary tetramers, the model sequences and the mixing experiments. 1.0M salt buffer unless specified otherwise.

<u>Sequence</u>	<u>Concn (mM)</u>	<u>Average <math>T_m</math> (<math>^{\circ}</math>C)</u>
GGCC	1.7	54.0
CCGG	1.9	47.8
GCCG	2.1	35.5
CGGC	2.1	29.2
GCGC	1.9	49.9
CGCG	1.2	36.9
GCC	9.0	20.0
CCG	7.8	no $T_m$
CGGA	3.6	19.2
AGGC	2.2	no $T_m$
GCCG:CGGC	2.1:2.1	47.7
GGA:CCA, 1.0 M salt	5.1:5.1	no $T_m$
no salt	2.5:2.5	21.7
GGC:GCC	1.7:1.7	24.0
GCG:CGC, 1.0 M salt	4.3:4.3	no $T_m$
no salt	2.5:2.5	21.4
GCG, 1.0 M salt	7.3	no $T_m$
no salt	2	25.8
CGG:CCG, 1.0 M salt	1.7:1.7	no $T_m$
no salt	1.0:1.0	<10



Thus the order of stabilities is deduced to be GC:GC more stable than GG:CC more stable than CG:CG. This differs from the published results of 1974, which gave free energies of formation for the cores as GG:CC -4.8kCal/mol, GC:GC -4.3kCal/mol and CG:CG -3.0kCal/mol (Borer *et al.*, 1974) but is in agreement with thermodynamic parameters calculated from GC containing sequences only: GC:GC -3.3kCal/mol, GG:CC -3.2 kCal/mol and CG:CG -2.3 kCal/mol (Freier *et al.*, 1985a).

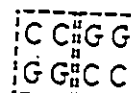
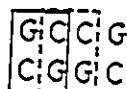
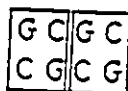
### 3.2.6 The mixing experiments: Duplexing versus aggregation

Five pairs of complementary mixtures were studied in 1.0 M salt buffer: GGA with CCA, GCCG with CGGC, GGC with GCC, GCG with CGC and CGG with CCG. Of these, only two duplex normally to give a measurable  $T_m$ : GCCG:CGGC and GGC:GCC. In the other three mixtures self aggregation of one of the strands is preferred over duplexing with the complementary strand. This self-aggregation also occurs in isolation; particular attention was given to the behaviour of GCG. Upon removal of the salt from these aggregate-forming mixtures, two, GGA:CCA and GCG:CGC, formed duplexes with a measurable  $T_m$ , while the other, CGG:CCG, exhibited the upper portion of sigmoidal curves, but had a  $T_m$  too low to be measured. GCG also duplexed in the absence of salt.

### 3.2.6.1 The duplexes: GCCG:CGGC and GGC:CCG

GCCG and CGGC are self-complementary, forming staggered duplexes with 5' dangling bases,  $T_m$ s 35.5 and 26.4°C respectively (see section 3.2.4). The chemical shift versus temperature curves are incomplete, with loss of signal at low temperature due to end-to-end overlap and aggregation. The chemical shift versus temperature curves for the base protons of the 1:1 mixture are sigmoidal (seven out of eight), complete, and have  $T_m$ s ranging from 43.5 - 49.7°C, with an average of 47.7°C (see Figure 3.14, Table 3.16). Three imino proton resonances are observed at 1.0°C, occurring in approximately a 1:1:2 ratio at 13.34, 12.84 and 12.66ppm, corresponding to four individual base pairs in the non-centrosymmetric helix. Their assignment is discussed in section 3.2.3. All signals persist to at least 19°C (see Table 3.7). The high  $T_m$ s of the nonexchangeable protons in the mixture compared to the single strand, and the persistence of the exchangeable protons to > 19°C, compared to the disappearance of those observed for GCCG alone below 15°C (Table 3.6), supports annealing of complementary strands to form a perfect four-base-pair duplex.

From the order of relative stability for the GC cores, GC:GC > GG:CC > CG:CG, one would have predicted a  $T_m$  for the duplex GCCG:CGGC less than that of GCGC, since, compared to GCGC, a GC:GC core has been replaced by a less stable GG:CC core, but more than that of CCGG in which both GC:GC cores have been replaced by GG:CC cores.



The  $T_m$  of GCCG:CGGC (2.1mM:2.1mM, 1.0M salt) is virtually identical to that of

Table 3.16: Chemical shift assignments for the aromatic base protons of GCCG:CGGC at 2.1 mM each strand in 1.0M salt buffer over the temperature range 70 to 30°C. Numbering scheme:

1 4  
GCCG  
CGGC  
8 5

Proton	Chemical shift (ppm)								T <sub>m</sub>
	Temp. 70.2°	60.4°	55.8°	51.1°	45.2°	40.0°	35.3°	30.3°	
G(1)H-8	7.955	7.970	7.974	7.975	5.957	7.930	7.903	7.876	NSB
C(2)H-6	7.744	7.748	7.756	7.771	7.796	7.814	7.823	7.815	49.7
C(3)H-6	7.734	7.732	7.726	7.719	7.709	7.700	7.696	7.693	49.2
G(4)H-8	7.971	7.956	7.930	7.885	7.827	7.758	7.705	7.678	47.4
C(5)H-6	7.672	7.698	7.726	7.778	7.843	7.881	7.898	7.893	47.2
G(6)H-6	7.941	7.892	7.827	7.735	7.567	7.449	7.382	7.337	47.9
G(7)H-6	7.935	7.930	7.920	7.901	7.861	7.824	7.789	7.753	43.5
C(8)H-8	7.760	7.740	7.711	7.667	7.589	7.533	7.493	7.464	47.4
C(5)H-5	5.893	5.824	5.747	5.621	5.471	5.352	—	—	47.4
C(8)H-5	5.939	5.890	5.846	5.780	5.680	5.632	5.601	5.572	49.6
C(2)H-5	5.835	5.764	5.705	5.621	5.497	5.402	—	—	47.4
C(3)H-5	5.939	5.924	5.924	5.928	5.934	5.929	5.915	5.894	NSB
Average T <sub>m</sub>									47.7

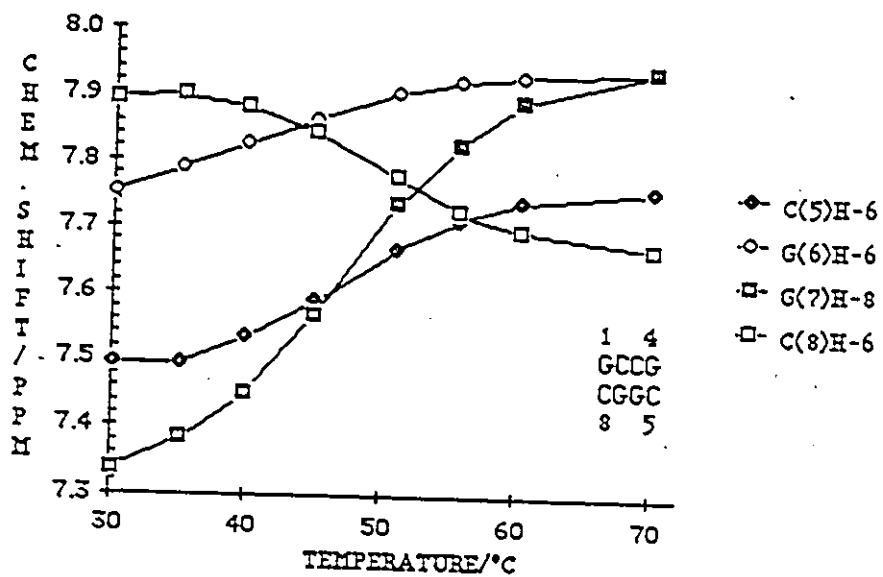
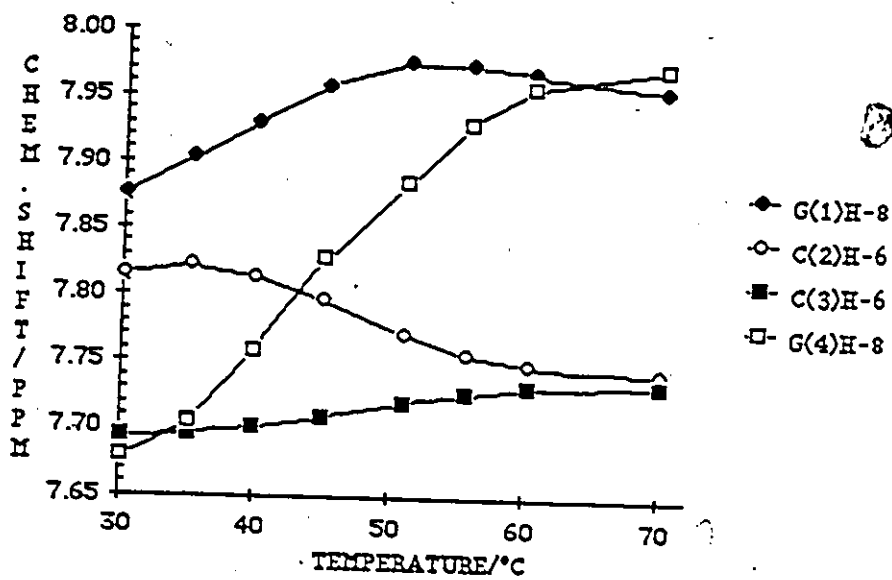


Figure 3.14: Chemical shift versus temperature curves for the base protons of GCCG:CGGC, 2.1 mM each strand, in 1.0 M salt buffer.

CCGG (1.9mM, 1.0M salt), 47.7°C compared with 48.0°C for base protons alone. This suggests, as confirmed by Freier *et al.* (1985a), that the difference in the stabilities between GC:GC and GG:CC is small.

Of the four complementary trimer mixtures studied, only GGC with GCC produced a set of sigmoidal chemical shift versus temperature curves, of which thirteen out of fifteen were sigmoidal (the exceptions being G1H8 of GGC and G3H1' of GCC; see Figure 3.15) with  $T_m$ s of 22.0 - 26.4°C, average 24.0°C (Table 3.17). No imino proton spectrum could be observed for this duplex; the  $T_m$  was too low. GCC in isolation at 8.1mM has a  $T_m$  of about 20°C, while GGC does not duplex. The narrow range of the individual proton  $T_m$ s supports formation of a complementary duplex.

Addition of a GC base pair to this duplex raises the  $T_m$  by 24 - 30°C: The  $T_m$  of the GGCC duplex, formed by addition of the base pair to the 3' end of GGC, is 54.0°C, while the  $T_m$  of GCCG:CGGC duplex, where the base pair is added 5' to GGC, is 47.7°C (Table 3.15).

3.2.6.2. The aggregates: GGA with CCA, GCG with CGC and CGG with CCG in 1.0 M salt buffer: self-aggregation of GCG

Equimolar mixtures of GGA with CCA, GCG with CGC and CGG with CCG do not form complementary duplexes; instead, in all three mixtures, the resonances of the G-rich strand (GGA, GCG and CGG) lose intensity with decreasing temperature until at low temperature only the signals for the complementary strand (CCA, CGC and CCG) remain, little perturbed from their position in isolation (see Figure 3.16, for example of spectra; Figures 3.17, 3.18 and 3.19 for chemical shift versus temperature plots; Tables 3.18, 3.19 and 3.20 for data). Study of the "vanishing" sequences alone revealed that the loss of intensity and disappearance is due to considerable broadening of the signals

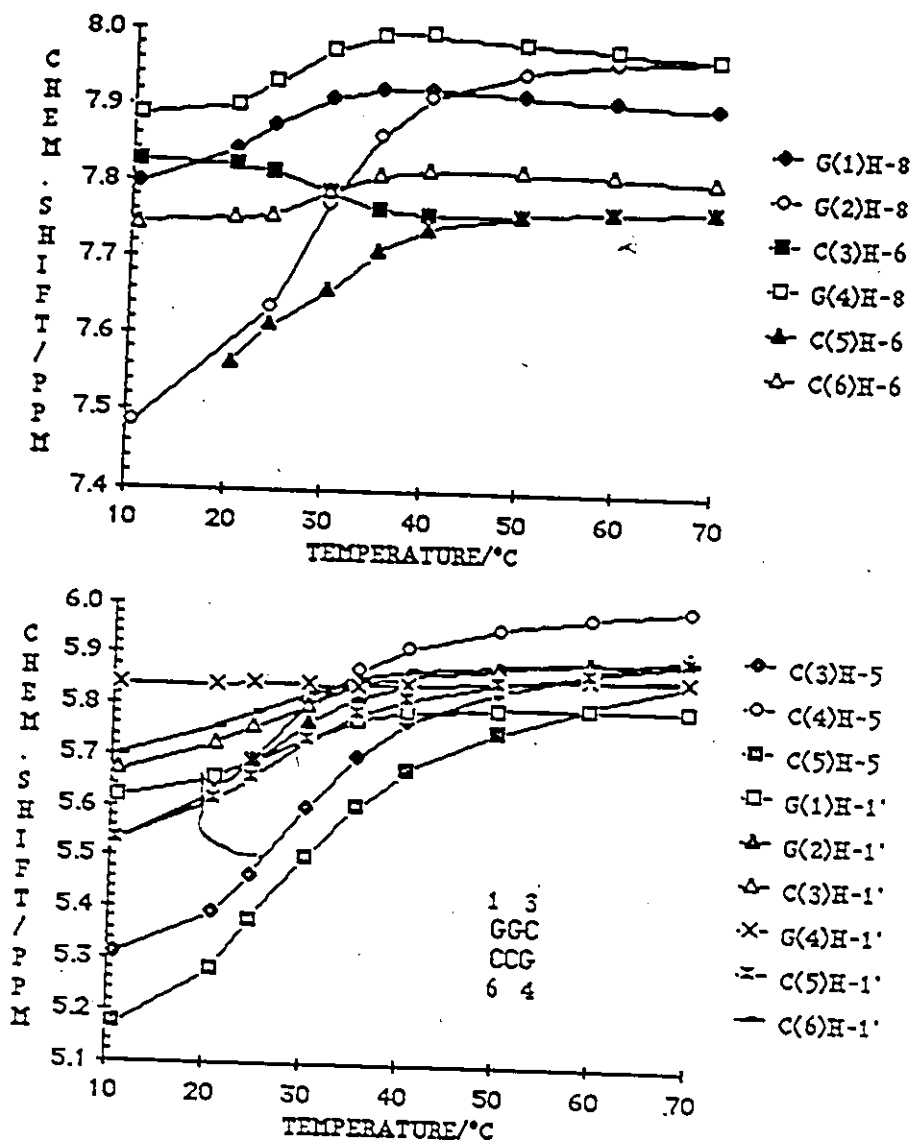


Figure 3.15: Chemical shift versus temperature curves for the base and anomeric protons of GGC:GCC, 1.7 mM each strand, in 1.0 M salt buffer.

Table 3.17: Chemical shift assignments and anomeric proton coupling constants of the protons of GGC:GCC at 1.7:1.7 mM in 1.0 M salt buffer over the range 70 to 10°C. Numbering scheme:

1 3  
GGC  
CCG  
6 4

Temp.	Chemical shift (in ppm)									T <sub>m</sub>
	69.7°	59.4°	50.1°	40.4°	35.3°	30.3°	24.5°	20.6°	10.7°	
Proton										
G(1)H-8	7.903	7.908	7.914	7.922	7.923	7.908	7.873	7.842	7.795	20.7
G(2)H-8	7.967	7.959	7.945	7.912	7.860	7.772	7.637	—	7.486	23.9
C(3)H-6	7.768	7.764	7.758	7.758	7.765	7.786	7.813	7.821	7.826	26.4
G(4)H-8	7.967	7.975	7.984	7.994	7.993	7.973	7.930	7.900	7.887	22.7
C(5)H-6	7.768	7.764	7.758	7.739	7.711	7.659	7.612	7.563	—	24.3
C(6)H-6	7.807	7.812	7.815	7.816	7.808	7.786	7.755	7.751	7.743	26.4
C(3)H-5	5.906	5.875	5.836	5.773	5.702	5.597	5.465	5.392	5.312	26.2
C(5)H-5	5.855	5.807	5.753	5.677	5.605	5.502	5.379	5.282	5.175	24.7
C(6)H-5	5.999	5.980	5.956	5.917	5.873	5.805	5.688	5.626	5.530	23.4
G(1)H-1'	5.800	5.799	5.795	5.790	5.773	5.738	5.688	5.657	5.616	23.4
G(2)H-1'	5.865	5.857	5.850	5.843	5.818	5.767	5.688	5.626	5.530	22.0
C(3)H-1'	5.897	5.890	5.876	5.857	5.841	5.803	5.758	5.726	5.669	22.7
G(4)H-1'	5.860	5.857	5.850	5.843	5.841	5.842	5.843	5.838	5.837	NSB
C(5)H-1'	5.897	5.870	5.850	5.816	5.785	5.731	5.655	5.611	5.530	23.7
C(6)H-1'	5.897	5.890	5.883	5.871	5.853	5.824	5.785	5.758	5.701	22.0
									Average T <sub>m</sub>	24.0
	Coupling constants (Hz)									
G(1)H-1'	4.7	3.8	4.5	4.3	3.6	—	—	—	—	
G(2)H-1'	4.4	4.3	3.8	3.9	3.0	2.0	—	—	—	
C(3)H-1'	3.3	4.4	3.4	3.4	3.1	2.1	—	—	—	
G(4)H-1'	4.7	4.3	3.8	3.8	3.1	2.2	—	—	—	
C(5)H-1'	3.3	5.5	3.8	2.8	2.7	3.1	—	—	—	
C(6)H-1'	3.3	4.4	3.7	3.2	2.8	2.1	—	—	—	



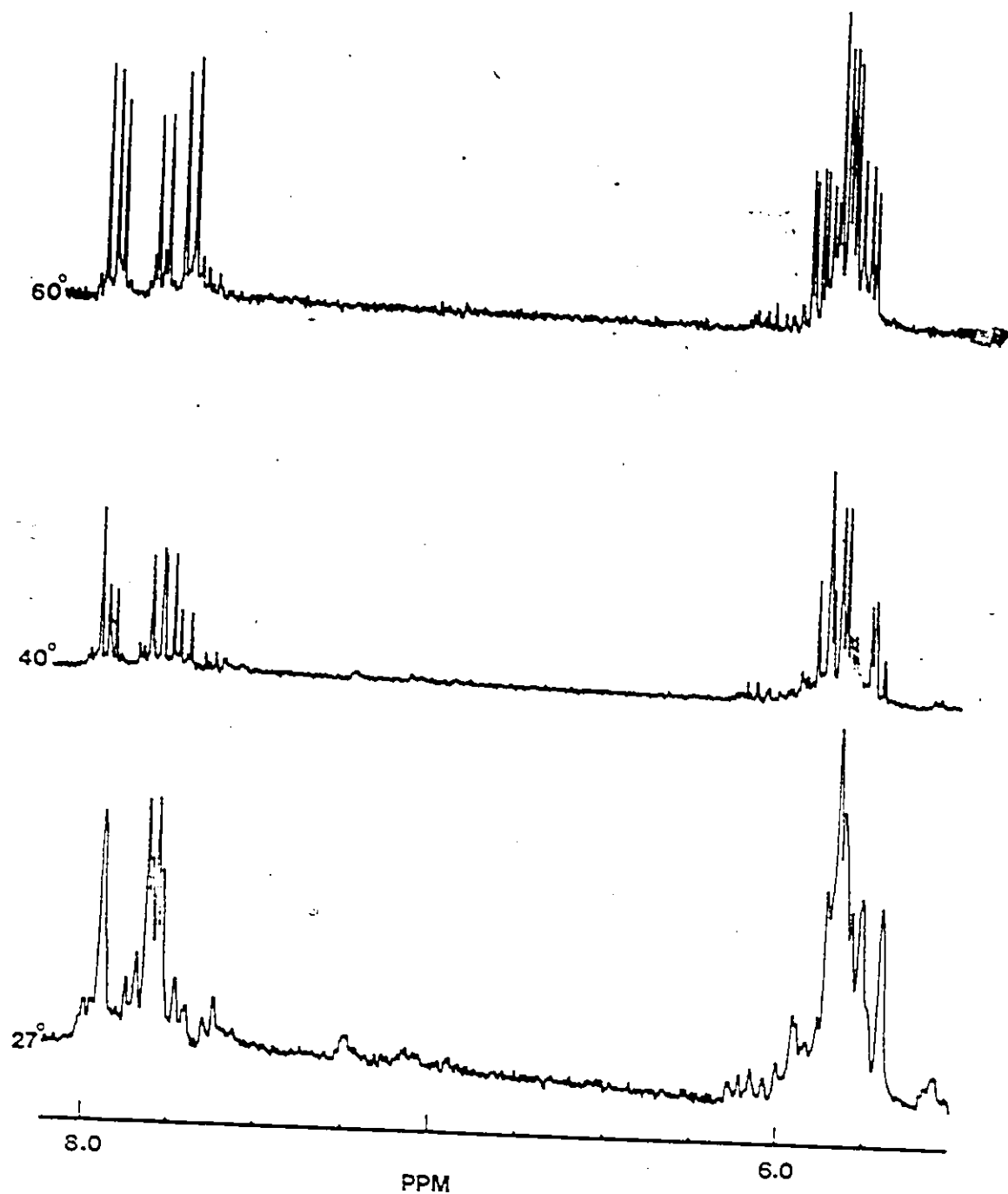


Figure 3.16:  $^1\text{H-NMR}$  spectra of GCG with CGC at (top to bottom) 60°, 40° and 27°C, showing disappearance of resonances belonging to GCG due to aggregation.

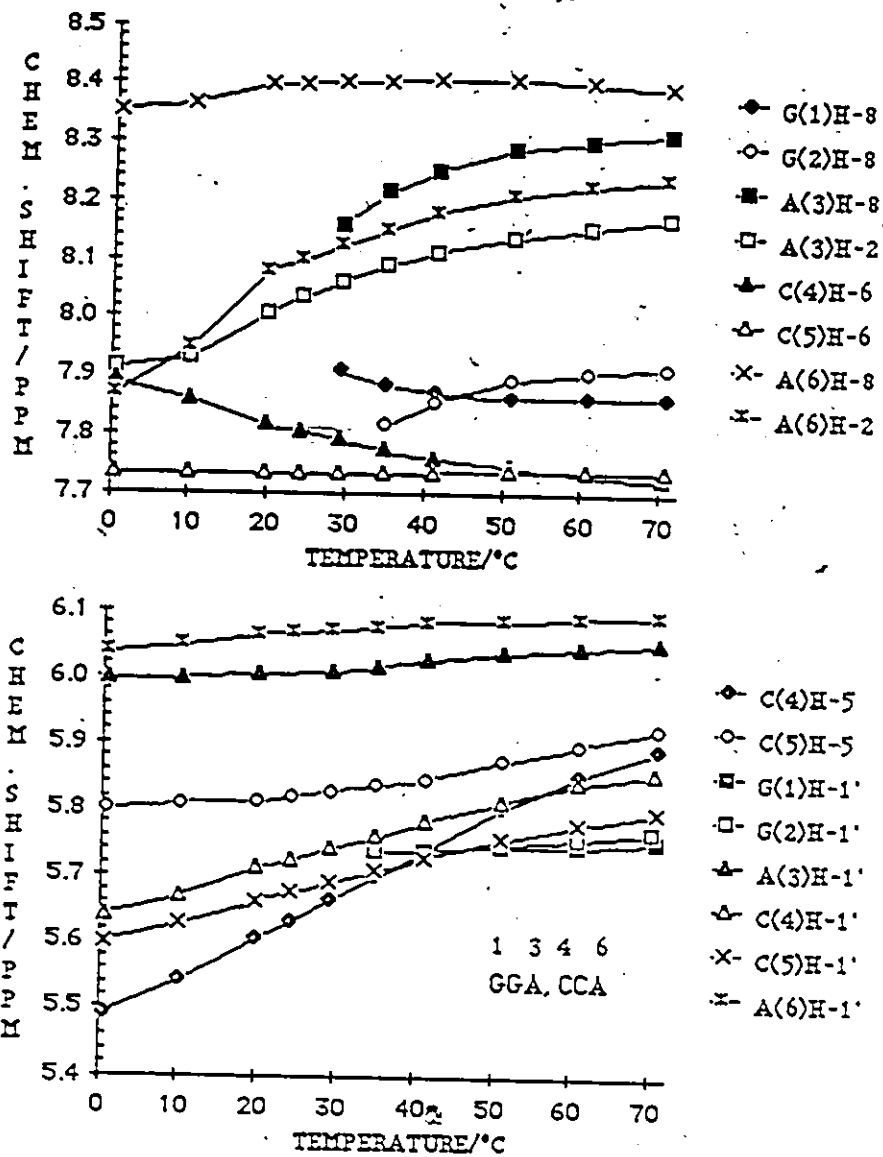


Figure 3.17: Chemical shift versus temperature curves of the base and anomeric protons of GGA with CCA, 4.5 mM, in 1.0 M salt buffer.

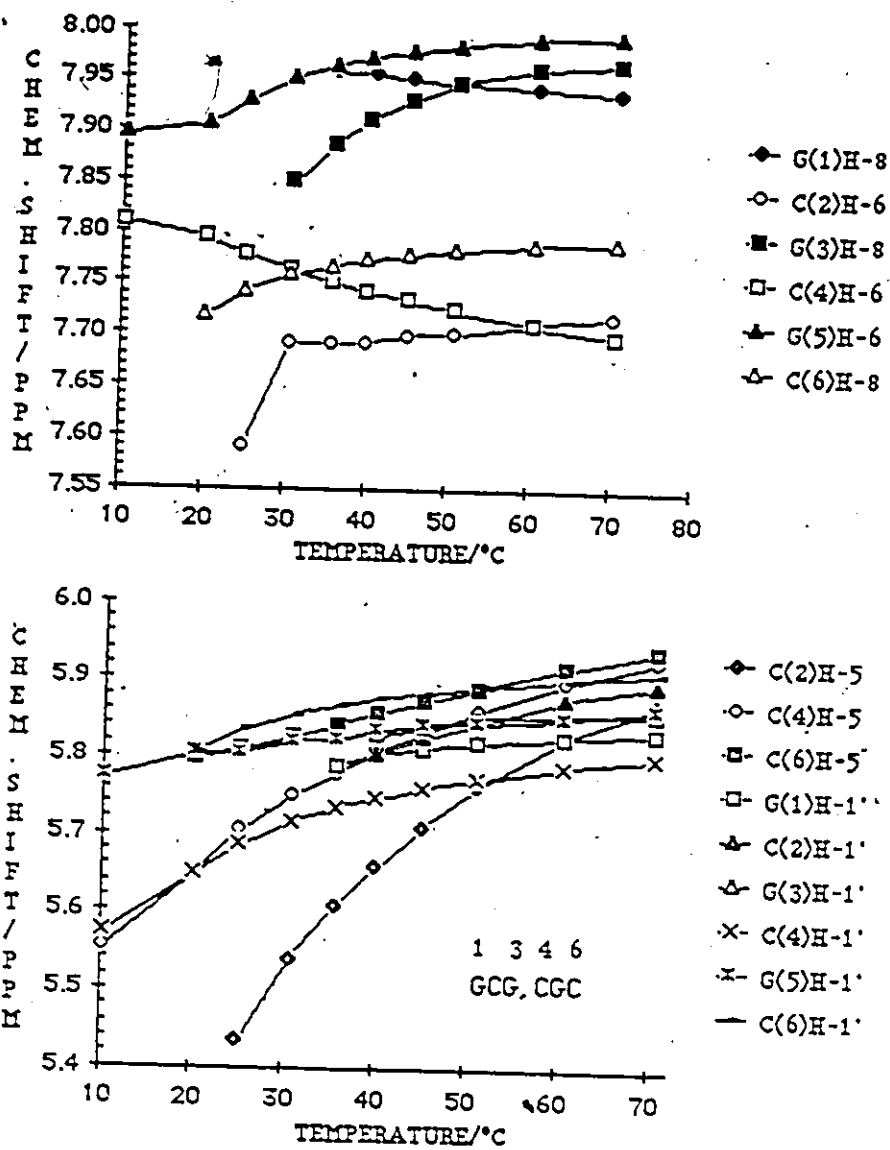


Figure 3.18: Chemical shift versus temperature curves for the base and anomeric protons of GCG with CGC, 4.4 mM each strand, in 1.0 M salt buffer.

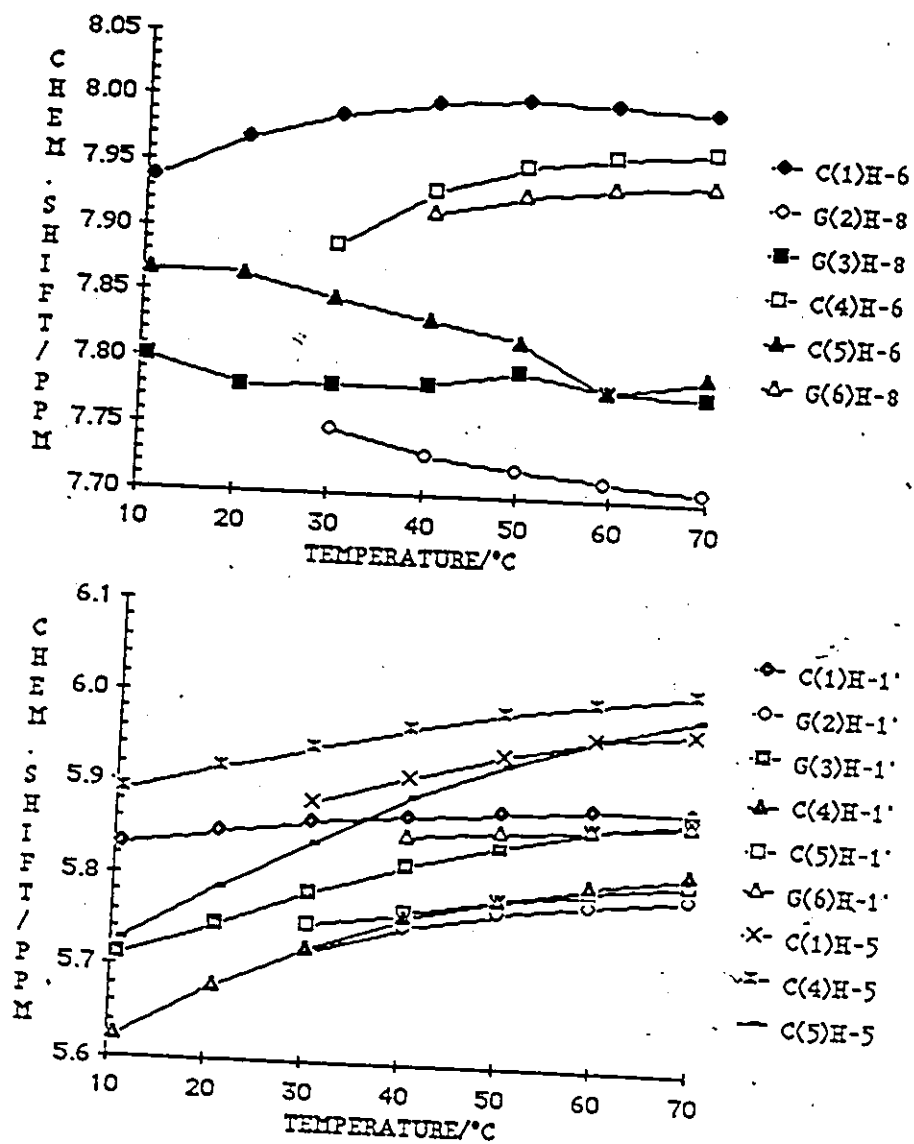


Figure 3.19: Chemical shift versus temperature curves for the base and anomeric protons of CGG:CCG, 1.7 mM each strand, in 1.0 M salt buffer.

Table 3.18: Chemical shift and coupling constants for GGA:CCA, 3.5mM each strand, in 1.0M salt buffer, over the temperature range 70 to 0°C.  
Numbering scheme: 1 3 4 6  
GGA:CCA

Temp.	Chemical shifts (ppm)									
	70.9°	60.7°	51.0°	41.3°	35.2°	29.5°	24.5°	20.0°	10.2°	0.7°
Proton										
G(1)H-8	7.831	7.861	7.863	7.871	7.884	7.907	--	--	--	--
G(2)H-8	7.911	7.903	7.890	7.855	7.815	--	--	--	--	--
A(3)H-8	8.313	8.302	8.286	8.251	8.214	8.156	--	--	--	--
A(3)H-2	8.171	8.154	8.137	8.112	8.090	8.061	8.033	8.006	7.930	7.911
C(4)H-6	7.724	7.731	7.743	7.759	7.774	7.790	7.804	7.816	7.859	7.890
C(5)H-6	7.738	7.736	7.735	7.733	7.732	7.731	7.730	7.731	7.731	7.732
A(6)H-8	8.393	8.400	8.405	8.407	8.403	8.400	8.396	8.395	8.362	8.351
A(6)H-2	8.238	8.225	8.206	8.177	8.149	8.122	8.096	8.076	7.946	7.868
C(4)H-5	5.893	5.851	5.803	5.739	5.702	5.663	5.631	5.604	5.541	5.491
C(5)H-5	5.921	5.896	5.873	5.844	5.836	5.825	5.818	5.810	5.809	5.799
G(1)H-1'	5.752	5.744	5.743	5.736	5.736	--	--	--	--	--
G(2)H-1'	5.766	7.754	5.743	5.736	5.736	--	--	--	--	--
A(3)H-1'	6.052	6.044	6.036	6.024	6.015	6.006	--	6.002	5.995	5.994
C(4)H-1'	5.875	5.841	5.812	5.783	5.760	5.740	5.723	5.710	5.666	5.638
C(5)H-1'	5.798	5.778	5.756	5.727	5.708	5.689	5.674	5.660	5.626	5.597
A(6)H-1'	6.090	6.087	6.083	6.080	6.073	6.069	6.064	6.062	6.046	6.034
	Coupling constants (Hz)									
G(1)H-1'	5.3	5.2	3.8	2.4	<1.5	--	--	--	--	--
G(2)H-1'	5.2	5.1	3.8	2.4	<1.5	--	--	--	--	--
A(3)H-1'	4.9	4.9	4.7	4.4	4.0	2.7	--	2.7	2.3	1.2
C(4)H-1'	4.3	3.9	3.3	3.0	2.9	2.5	2.1	2.0	1.1	--
C(5)H-1'	3.8	3.5	3.0	1.9	2.0	2.0	1.7	1.8	1.2	--
A(6)H-1'	5.0	5.1	5.0	4.9	4.6	4.1	3.9	4.3	4.1	3.9

Table 3.19: Chemical shift assignments and anomeric proton coupling constants for GCG:CGC at 4.4 mM each strand in 1.0 M salt buffer over the temperature range 70 - 10°C. Numbering scheme:  $\begin{matrix} 1 & 3 & 4 & 6 \\ \text{GCG:CGC} \end{matrix}$

Temp.	Chemical shifts (ppm)									
	70.6°	60.6°	51.0°	45.0°	40.0°	35.6°	30.7°	25.1°	20.1°	10.3°
Proton										
G(1)H-8	7.936	7.941	7.947	7.951	7.955	7.957	--	--	--	--
C(2)H-6	7.718	7.711	7.700	7.698	7.692	7.691	7.692	7.589	--	--
G(3)H-8	7.966	7.961	7.947	7.929	7.910	7.887	7.850	--	--	--
C(4)H-6	7.699	7.711	7.725	7.734	7.742	7.752	7.763	7.778	7.794	7.810
G(5)H-8	7.992	7.990	7.983	7.977	7.970	7.963	7.952	7.929	7.905	7.895
C(6)H-6	7.790	7.788	7.783	7.777	7.772	7.766	7.758	7.742	7.717	--
C(2)H-5	5.869	5.826	5.761	5.709	5.657	5.606	5.537	5.432	--	--
C(4)H-5	5.927	5.898	5.859	5.830	5.805	5.779	5.749	5.704	5.648	5.553
C(6)H-5	5.937	5.915	5.889	5.872	5.856	5.842	5.826	5.810	5.795	--
G(1)H-1'	5.831	5.825	5.817	5.808	5.802	5.787	--	--	--	--
C(2)H-1'	5.893	5.875	5.843	5.824	5.802	--	--	--	--	--
G(3)H-1'	5.855	5.853	5.850	5.839	5.827	--	--	--	--	--
C(4)H-1'	5.800	5.788	5.770	5.758	5.745	5.733	5.715	5.686	5.648	5.572
G(5)H-1'	5.862	5.851	5.844	5.842	5.834	5.821	5.818	5.804	5.804	5.771
C(6)H-1'	5.909	5.902	5.891	5.881	5.875	5.865	5.854	5.834	5.804	--
	Coupling constant (Hz)									
G(1)H-1'	4.7	4.7	4.5	3.5	4.5	3.7	--	--	--	--
C(2)H-1'	4.5	4.0	4.2	2.4	4.5	--	--	--	--	--
G(3)H-1'	5.4	3.9	4.9	5.0	3.5	--	--	--	--	--
C(4)H-1'	4.1	4.0	3.4	3.5	3.2	2.9	2.5	1.5	2.9	--
G(5)H-1'	4.1	5.1	3.2	3.8	3.6	2.8	4.0	2.6	2.9	--
C(6)H-1'	3.7	3.6	3.4	3.1	2.3	3.4	3.0	2.8	2.9	--

Table 3.20: <sup>1</sup>H chemical shifts at 250 MHz for CGG:CCG at 1.3 mM each strand concentration in 1.0M salt buffer for the range 70 to 10°C.

Numbering scheme:      1 3 4 6  
                                  CGG:CCG

Temp.	Chemical shift (ppm)							
	69.7°	59.4°	50.1°	40.3°	30.3°	20.6°	10.7°	
Proton								
C(1)H-6	7.707	7.713	7.720	7.730	7.748	--	--	
G(2)H-8	7.966	7.961	7.952	7.932	7.889	--	--	
G(3)H-8	7.940	7.936	7.929	7.915	--	--	--	
C(4)H-6	7.794	7.782	7.818	7.832	7.847	7.864	7.865	
C(5)H-6	7.780	7.782	7.795	7.783	7.782	7.780	7.800	
G(6)H-8	7.996	8.000	8.002	7.998	7.748	7.968	7.936	
C(1)H-5	5.967	5.959	5.939	5.912	5.882	--	--	
C(4)H-5	6.012	5.998	5.983	5.964	5.941	5.918	5.891	
C(5)H-5	5.984	5.959	5.929	5.889	5.837	5.786	5.726	
C(1)H-1'	5.789	5.777	5.766	5.748	5.720	--	--	
G(2)H-1'	5.804	5.793	5.782	5.765	5.749	--	--	
G(3)H-1'	5.869	5.860	5.856	5.847	--	--	--	
C(4)H-1'	5.873	5.860	5.840	5.815	5.784	5.745	5.711	
C(5)H-1'	5.815	5.800	5.782	5.759	5.723	5.679	5.622	
G(6)H-1'	5.881	5.881	5.876	5.868	5.860	5.847	5.832	
			Coupling constant (Hz)					
C(1)H-1'	4.5	3.4	5.0	4.0	--	--	--	
G(2)H-1'	5.0	4.7	3.8	4.5	3.6	--	--	
G(3)H-1'	5.4	3.7	4.6	5.2	5.9	--	--	
C(4)H-1'	3.7	3.6	3.7	3.5	3.1	--	--	
C(5)H-1'	3.9	3.7	3.8	1.9	1.8	--	--	
G(6)H-1'	5.4	3.9	5.0	5.4	4.8	4.6	--	

which is associated with formation of high molecular weight aggregates.

GCG has been the best studied of these sequences, since the duplexing of sequences of the type GCX have established a norm from which it conspicuously deviates. As described in the introduction, GCA, GCC and GCG\*, where G\* is guanosine modified at the N1, O6, and N2 and O6 positions so as to preclude normal hydrogen bonding, form minihelices with a GC:GC core stabilized by two 3' dangling bases, and Tms (at 7.3mM in 1.0M salt) of 34 - 42°C (Alkema *et al.*, 1981; D'Andrea *et al.*, 1983). The chemical shift versus temperature curves of GCG, though similar in shape to those of the well-behaved analogues over the upper portions, are non-sigmoidal. Beginning around 40°C the resonances broaden and lose intensity, until they become unassignable below 20°C (Figure 3.20; Table 3.21). Coincidentally, a new signal appears at 5.95ppm, at lower field than any of the H-1' or H-5 resonances, and increases in intensity with decreasing temperature, until by 20°C it dominates the spectrum. Neither the marked broadening of signals, nor the appearance of the new peak have any parallel in the spectra of GCA or the modified GCGs.

Low temperature spectra (6.6°C) of the exchangeable resonances have been obtained for GCA and GC<sup>1Me</sup>G (spectra by G.W.Buchko). As expected, they show a single signal within the hydrogen-bonded imino region. In GCA the signal is at 12.2ppm and in GC<sup>1Me</sup>G, at 12.9ppm, the difference being accountable by the greater shielding capacity of adenosine.

The imino region of GCG at 6.6°C shows three resonances, at 13.43, 11.54 and 11.22ppm. The first appears in the region in which GN1H imino proton of a G-C hydrogen bond might be expected, while the latter two appear where a GN1H bonded to a carbonyl group (as in a G-U base pair) would be observed. All three resonances persist until 30°C; it is hard to determine which, if any, disappear first (Table 3.22).



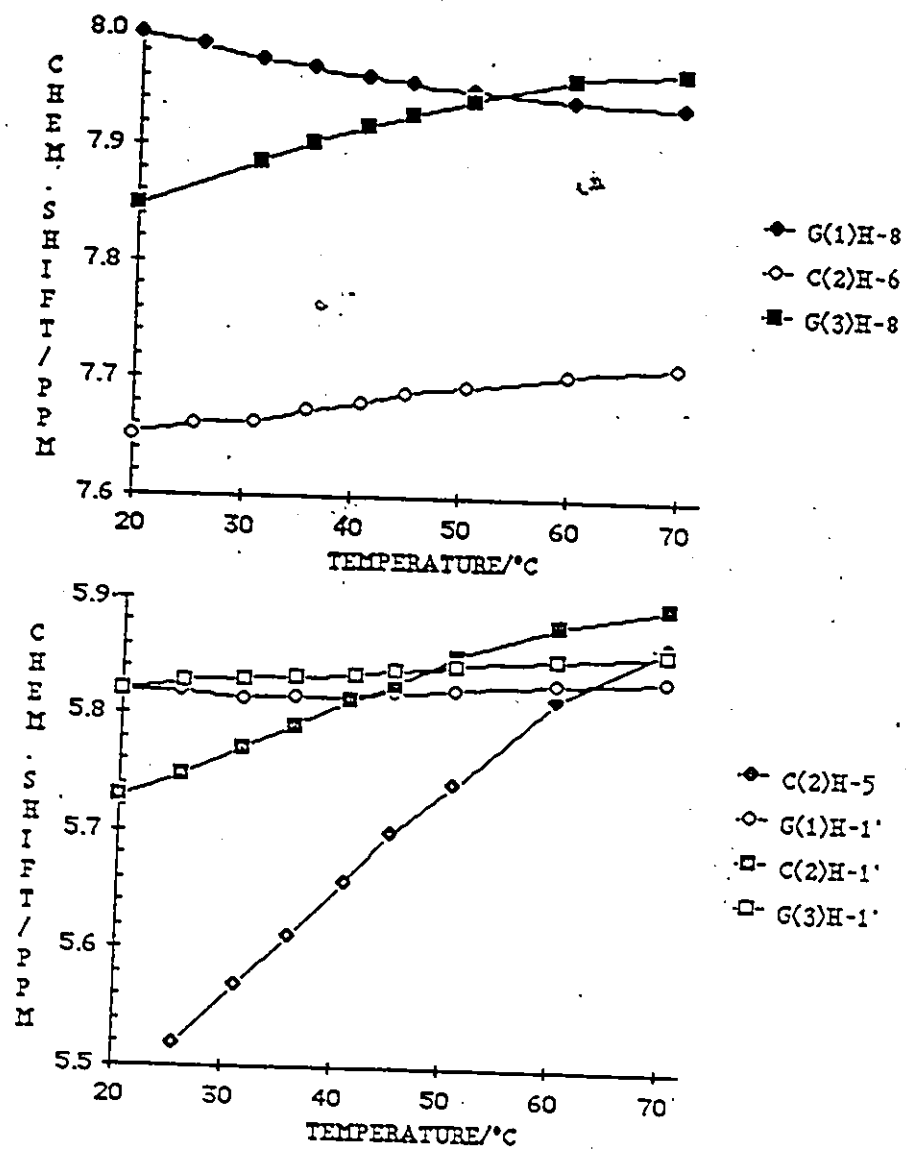


Figure 3.20: Chemical shift versus temperature curves for the base and anomeric protons of GCG, 5 mM, in 1.0 M salt buffer.

Table 3.21: Chemical shifts and anomeric proton coupling constants for the protons of GCG, 7.3mM, in 1.0M salt buffer over the temperature range 70 to 20°C.

Temp.	Chemical shift (ppm)								
	<u>70.0°</u>	<u>60.0°</u>	<u>50.8°</u>	<u>45.0°</u>	<u>41.0°</u>	<u>36.0°</u>	<u>31.2°</u>	<u>25.9°</u>	<u>20.0°</u>
Proton									
G(1)H-8	7.935	7.940	7.949	7.954	7.960	7.967	7.974	7.974	7.995
C(2)H-6	7.712	7.705	7.693	7.688	7.680	7.673	7.665	7.660	7.650
G(3)H-8	7.965	7.959	7.941	7.928	7.916	7.902	7.886	7.914	7.847
C(2)H-5	5.862	5.815	5.741	5.698	5.656	5.611	5.569	5.519	--
G(1)H-1'	5.832	5.827	5.821	5.818	5.815	5.814	5.813	5.820	5.818
C(2)H-1'	5.895	5.877	5.850	5.826	5.812	5.789	5.771	5.748	5.728
G(3)H-1'	5.885	5.848	5.842	5.837	5.834	5.831	5.829	5.827	5.818
	Coupling constant (Hz)								
G(1)H-1'	4.8	5.1	4.9	4.4	4.1	3.8	3.5	3.1	1.5
C(2)H-1'	4.5	5.3	3.8	4.0	2.6	3.0	2.3	1.5	1.5
G(3)H-1'	4.8	4.6	5.4	5.1	5.2	4.4	4.1	3.2	1.5

Table 3.22: Chemical shifts of signals observed in the imino proton spectra of GCG, 5mM, in 1.0M salt buffer, over the temperature range 6 to 31°C.

Temperature	Chemical shift (ppm)				
	<u>6.6°</u>	<u>16.1°</u>	<u>20.0°</u>	<u>30.1°</u>	<u>35.0°</u>
	13.43	13.42	13.32	13.30	--
	11.54	11.51	11.45	11.41	--
	11.22	11.22	11.15	11.13	--

Section 3.3.4 is dedicated to the discussion of these results and of a postulated structure for the aggregate.

3.2.6.3. The duplexes: GGA:CCA, GCG:CGC, CGG:CCG and GCG in the absence of salt

In no salt buffer (no sodium chloride, 0.01 M sodium phosphate, pH 7), GGA:CCA (2 mM each strand) forms a duplex with an average  $T_m$  of 21.7°C, for eight of sixteen protons with individual proton  $T_m$ s ranging from 14° to 33.1°C (Figure 3.21; Table 3.23). The sigmoidal chemical shift versus temperature curves are evenly distributed between the two strands.

For GCG:CGC (no salt buffer, 2 mM each strand), seven of fifteen chemical shift versus temperature curves are sigmoidal (Figure 3.22), with  $T_m$ s ranging from 14.0° to 30.1°C, giving an average of 21.4°C (Table 3.24). However, GCG by itself at similar concentration, in absence of salt, gives a higher  $T_m$  of 25.8°C (Table 3.26), and the resolution of the lowest temperature spectra for the GCG:CGC mixture is extremely poor, suggesting either aggregation, or competition for formation between the GCG:CGC and the GCG duplexes.

The chemical shift versus temperature curves of GCG (1.5 mM, no salt buffer) (Figure 3.24), are identical in appearance and in the distribution of sigmoidal and nonsigmoidal behaviour to those of GCA (see Figure 4.16). Five of the seven curves have well-defined individual  $T_m$ s ranging from 20.1° to 30.1°C (Table 3.26), with an average  $T_m$  of 25.8°C. Thus GCG appears to be forming a duplex comprising a GC core stabilized by two 3' dangling guanosine residues.

The low temperature hydrogen-bonded imino proton spectrum of GCG in the absence of salt at -5°C is, however, identical to that of GCG in 1.0 M salt (though much

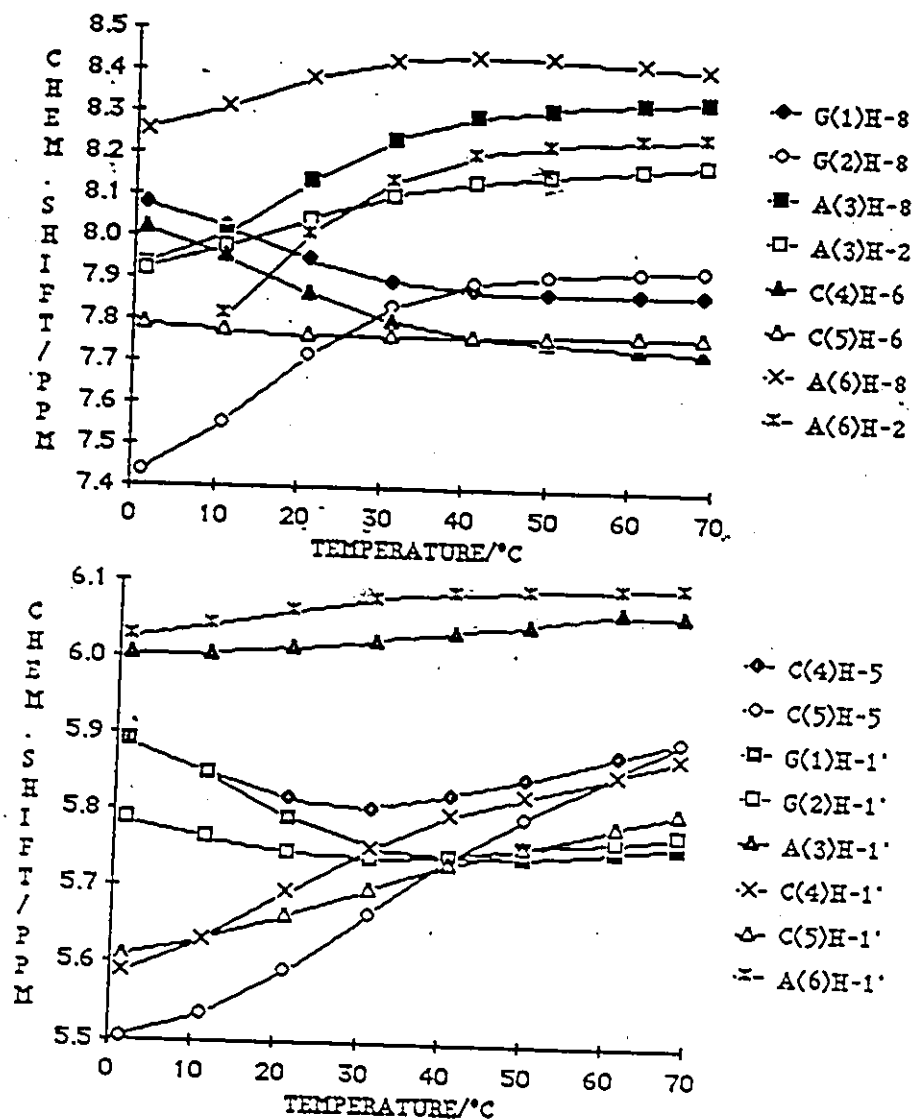


Figure 3.21: Chemical shift versus temperature curves for the base and anomeric protons of GGA:CCA, 2.0 mM each strand, in no salt buffer.

Table 3.23: Chemical shifts and anomeric proton coupling constants for GGA:CCA, 2mM each strand, no salt buffer, over the temperature range 70° to 0°C.

Numbering scheme:  
 1 3  
 GGA  
 ACC  
 6 4

Temp.	Chemical shift (ppm)								Tm
	68.8°	61.2°	50.0°	41.1°	31.4°	21.2°	11.0°	1.3°	
Proton									
G(1)H-8	7.866	7.866	7.867	7.872	7.894	7.945	8.020	8.074	14.0
G(2)H-8	7.924	7.920	7.909	7.891	7.832	7.713	7.550	7.435	14.8
A(3)H-8	8.333	8.327	8.312	8.291	8.235	8.134	8.002	7.929	16.3
A(3)H-2	8.182	8.168	8.149	8.130	8.095	8.041	7.974	7.918	24.8
C(4)H-6	7.730	7.736	7.748	7.762	7.798	7.863	7.951	8.017	14.6
C(5)H-6	7.769	7.767	7.764	7.762	7.760	7.764	7.775	7.787	NSB
A(6)H-8	8.410	8.420	8.430	8.434	8.424	8.382	8.309	8.252	NSB
A(6)H-2	8.245	8.236	8.220	8.197	8.132	8.004	7.807	—	NSB
C(4)H-5	5.900	5.876	5.844	5.822	5.804	5.815	5.848	5.889	NSB
C(5)H-5	5.895	5.854	5.794	5.742	5.666	5.592	5.533	5.501	33.1
G(1)H-1'	5.759	5.753	5.743	5.744	5.755	5.791	5.848	5.889	NSB
G(2)H-1'	5.775	5.764	5.755	5.744	5.739	5.746	5.765	5.786	NSB
A(3)H-1'	6.059	6.062	6.043	6.035	6.023	6.011	6.004	6.002	NSB
C(4)H-1'	5.874	5.852	5.823	5.797	5.752	5.694	5.630	5.587	24.8
C(5)H-1'	5.802	5.784	5.755	5.732	5.697	5.662	5.630	5.609	31.4
A(6)H-1'	6.094	6.093	6.090	6.086	6.077	6.061	6.040	6.024	NSB
							Average Tm		21.7
	Coupling constant (Hz)								
G(1)H-1'	5.7	5.8	3.0	5.1	5.1	3.8	2.0	—	
G(2)H-1'	5.5	5.5	2.7	5.1	4.1	3.1	—	—	
A(3)H-1'	5.0	4.9	4.8	4.6	4.5	4.4	2.6	1.8	
C(4)H-1'	4.3	5.0	3.5	3.1	2.1	1.7	—	—	
C(5)H-1'	3.6	3.3	2.7	2.2	2.0	1.6	—	—	
A(6)H-1'	5.3	5.3	4.8	5.1	4.3	2.0	3.6	—	

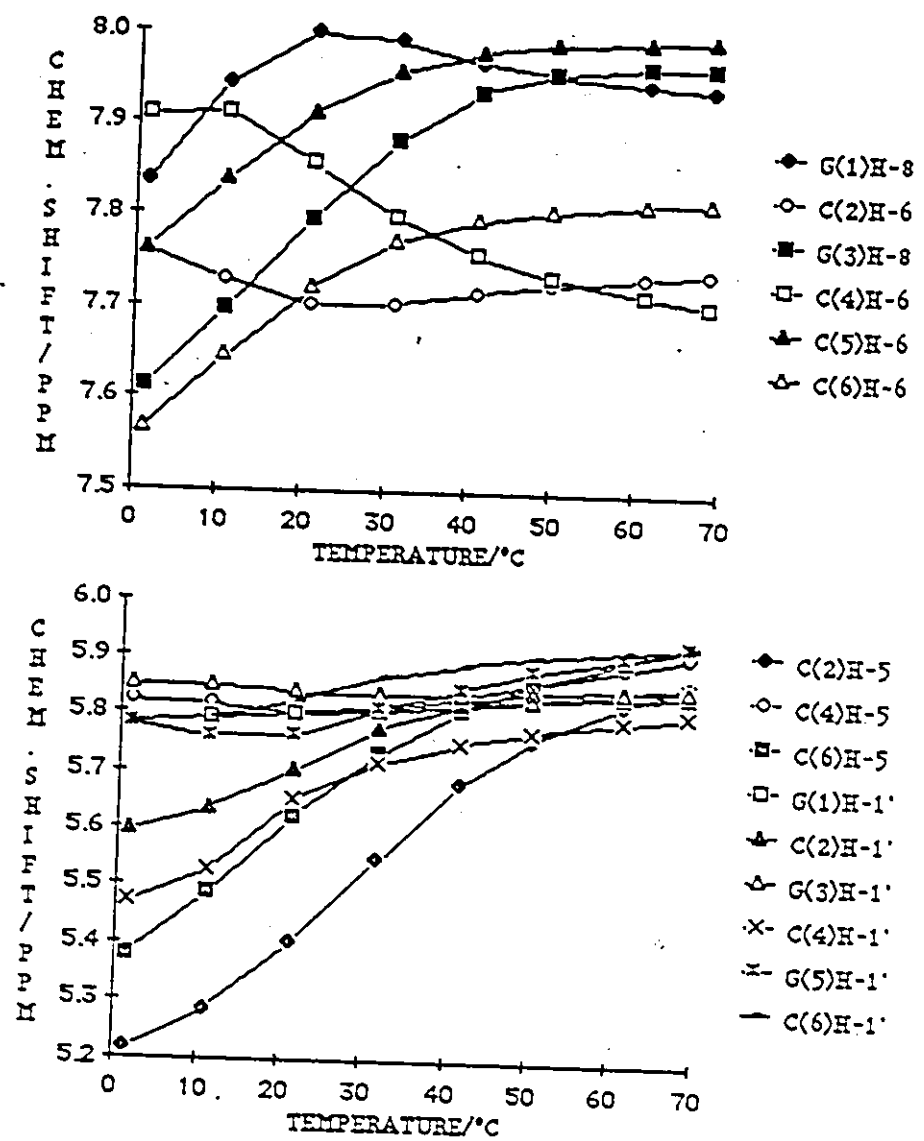


Figure 3.22: Chemical shift versus temperature curves for the base and anomeric protons of GCG:CGC, 2.0 mM each strand, in no salt buffer.

Table 3.24: Chemical shifts and anomeric proton coupling constants for GCG:CGC, 2mM each strand, in no salt buffer, over the temperature range 70° to 0°C.

Numbering scheme:

1 3  
GCG  
CGC  
6 4

Temp.	Chemical shifts (ppm)								Tm
	68.8°	61.2°	50.0°	41.2°	31.4°	21.2°	11.0°	1.3°	
Photon									
G(1)H-8	7.940	7.946	7.955	7.968	7.991	7.998	7.944	7.835	NSB
C(2)H-6	7.740	7.735	7.726	7.716	7.703	7.703	7.730	7.762	NSB
G(3)H-8	7.963	7.963	7.955	7.935	7.880	7.796	7.696	7.611	17.3
C(4)H-6	7.704	7.716	7.735	7.757	7.799	7.857	7.912	7.908	24.5
G(5)H-8	7.993	7.992	7.987	7.978	7.956	7.911	7.839	7.762	NSB
C(6)H-6	7.816	7.815	7.807	7.796	7.771	7.722	7.645	7.565	NSB
C(2)H-5	5.861	5.823	5.760	5.683	5.548	5.402	5.285	5.214	30.1
C(4)H-5	5.909	5.886	5.855	5.830	5.808	5.803	5.820	5.823	NSB
C(6)H-5	5.930	5.901	5.855	5.809	5.732	5.623	5.488	5.377	15.4
G(1)H-1'	5.839	5.834	5.828	5.819	5.805	5.800	5.791	5.782	NSB
C(2)H-1'	5.858	5.849	5.844	5.819	5.777	5.702	5.635	5.594	23.4
G(3)H-1'	5.848	5.844	5.828	5.834	5.834	5.838	5.848	5.847	NSB
C(4)H-1'	5.804	5.791	5.771	5.751	5.714	5.652	5.526	5.470	14.0
G(5)H-1'	5.926	5.903	5.877	5.843	5.812	5.761	5.759	7.782	NSB
C(6)H-1'	5.926	5.916	5.903	5.888	5.864	5.822	5.791	5.782	24.8
							Average Tm		21.4
	Coupling constants (Hz)								
G(1)H-1'	4.7	4.4	3.8	2.9	3.7	2.3	-	-	
C(2)H-1'	4.4	4.3	3.9	3.6	2.8	-	-	-	
G(3)H-1'	5.0	5.3	4.7	5.2	4.4	3.5	-	-	
C(4)H-1'	4.0	3.8	3.7	3.2	-	1.7	-	-	
G(5)H-1'	4.6	4.0	4.7	2.9	-	-	-	-	
C(6)H-1'	3.8	3.8	3.6	3.5	2.8	2.3	-	-	

weaker), suggesting that the removal of salt depresses the temperature of onset aggregation more than it does that of duplexing, but does not prevent aggregation entirely. This conclusion is strengthened by the behaviour of CGG:CCG in the absence of salt. The beginnings of sigmoidal curves are observed (Figure 3.23; Table 3.25) at low temperature, but the curves are incomplete, with  $T_{ms}$  in the region of 10°C or lower. Between 8° and 0°C the signals belonging to C(1)H-1' and G(2)H-1', both on the CGG strand, disappear into the baseline, suggesting the onset of aggregation.



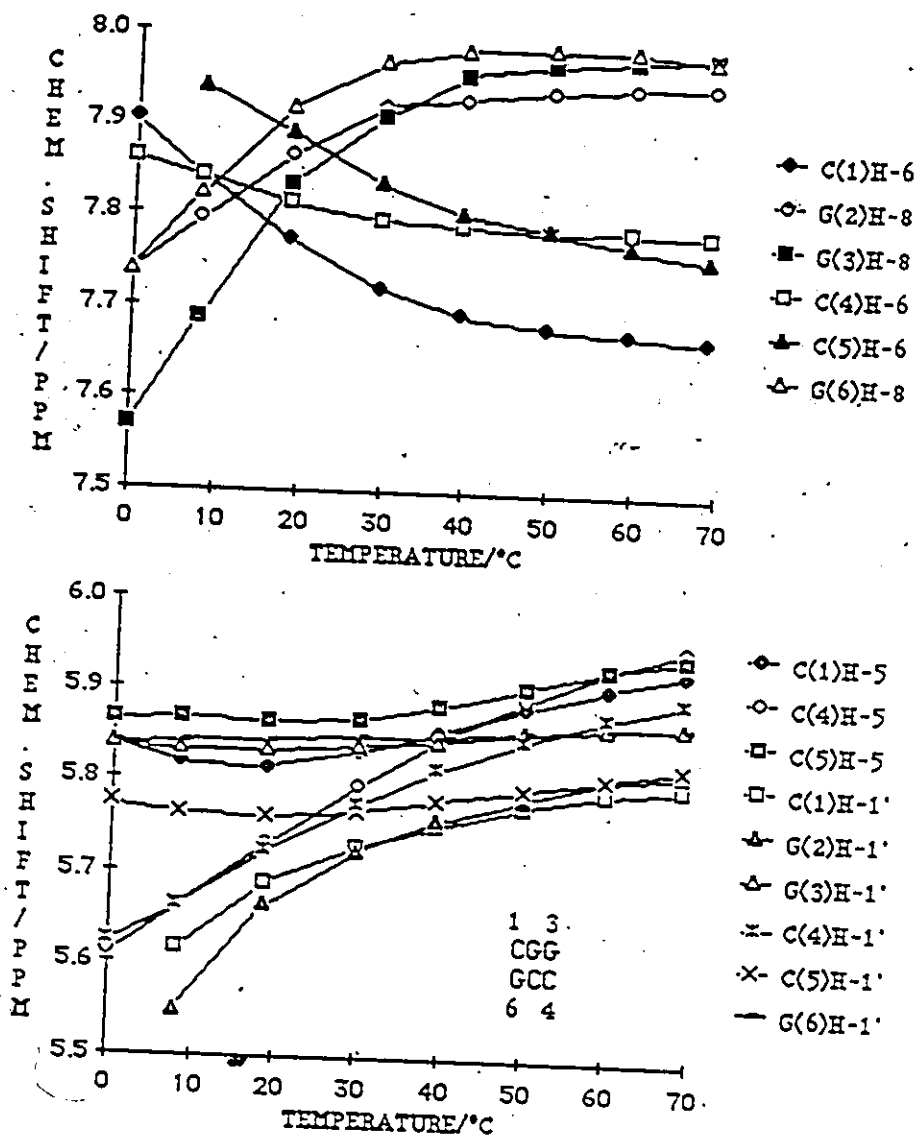


Figure 3.23: Chemical shift versus temperature curves for the base and anomeric protons of CGG:CCG, 1.0 M each strand, in no salt buffer.

Table 3.25: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CGG:CCG, 1.0 M each strand, in no salt buffer over the temperature range 70° to 0°C. Numbering scheme: 1 3  
CGG  
GCC  
6 4

Temp.	Chemical shift (ppm)							
	69.0°	59.5°	49.7°	39.3°	29.5°	18.5°	8.0°	0.0°
Proton								
C(1)H-6	7.670	7.675	7.680	7.693	7.720	7.774	7.842	7.903
G(2)H-8	7.945	7.943	7.937	7.927	7.920	7.865	7.795	7.737
G(3)H-8	7.978	7.973	7.967	7.954	7.904	7.833	7.685	7.567
C(4)H-6	7.786	7.787	7.784	7.790	7.797	7.815	7.842	7.861
C(5)H-6	7.759	7.772	7.788	7.804	7.835	7.889	7.938	-
G(6)H-8	7.974	7.982	7.983	7.981	7.966	7.916	7.822	7.737
C(1)H-5	5.923	5.906	5.883	5.854	5.832	5.815	5.820	5.839
C(4)H-5	5.950	5.926	5.891	5.947	5.796	5.731	5.663	5.610
C(5)H-5	5.939	5.926	5.906	5.884	5.868	5.864	5.868	5.864
C(1)H-1'	5.798	5.789	5.775	5.753	5.728	5.690	5.617	-
G(2)H-1'	5.815	5.806	5.783	5.757	5.724	5.664	5.547	-
G(3)H-1'	5.864	5.862	5.856	5.843	5.838	5.832	5.834	5.839
C(4)H-1'	5.892	5.873	5.847	5.814	5.774	5.724	5.663	5.623
C(5)H-1'	5.819	5.806	5.793	5.780	5.767	5.760	5.764	5.774
G(6)H-1'	5.864	5.862	5.856	5.850	5.848	5.844	5.842	5.839
	Coupling constant (Hz)							
C(1)H-1'	4.2	4.2	6.0	3.6	3.6	2.4	-	--
G(2)H-1'	5.5	4.4	5.2	-	-	-	-	--
G(3)H-1'	5.5	5.8	5.6	4.6	4.6	5.5	4.1	--
C(4)H-1'	4.1	4.7	3.7	3.4	2.7	-	-	--
C(5)H-1'	1.2	4.4	5.2	4.6	4.2	3.1	-	--
G(6)H-1'	5.5	5.8	5.6	3.0	5.5	5.8	4.1	--

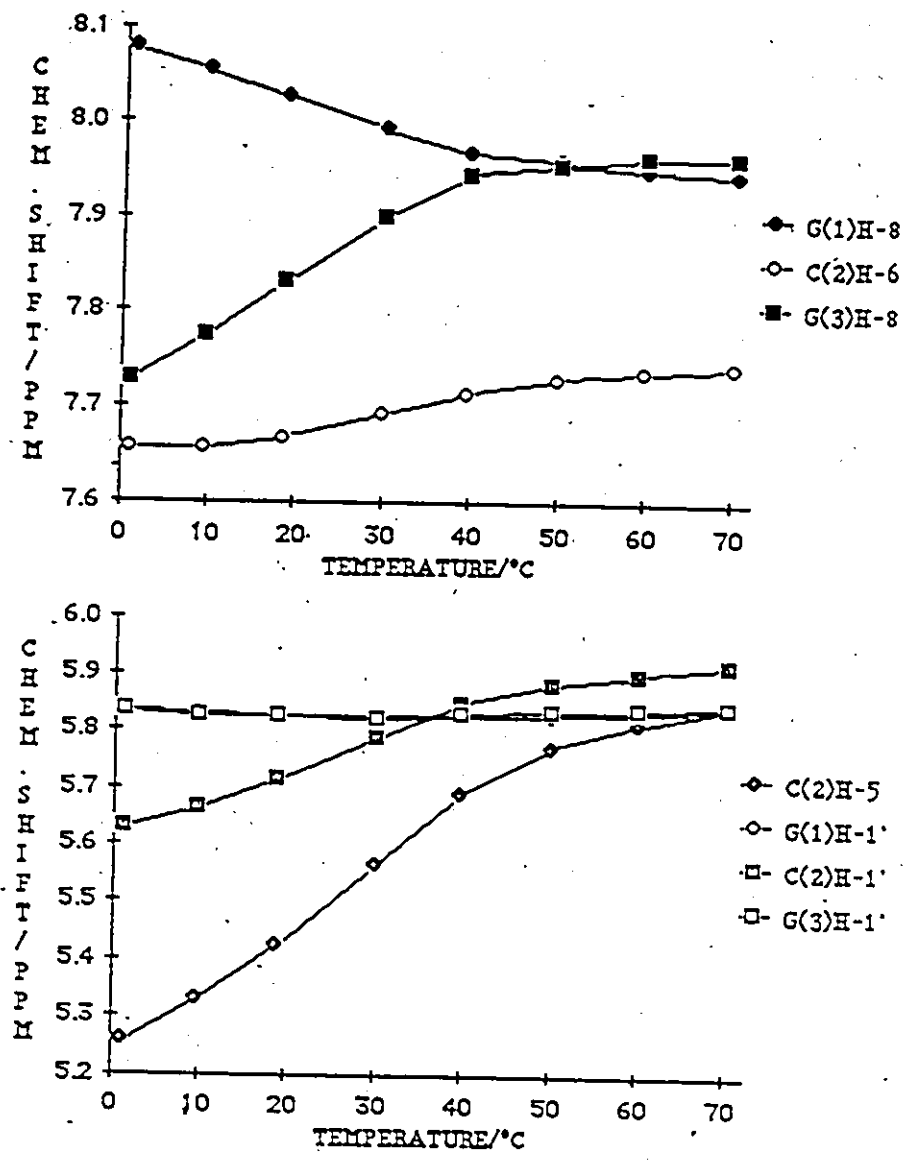


Figure 3.24: Chemical shift versus temperature curves of the base and anomeric protons of GCG, 1.5 mM, in no salt buffer.

Table 3.26: Chemical shifts and anomeric proton coupling constants for the protons of GCG, 1.5mM, in no salt buffer, over the temperature range 70° to 0°C.

Temp.	Chemical shifts (ppm)								T <sub>m</sub>	
	70.1°	60.1°	50.1°	39.8°	30.0°	18.8°	9.7°	1.4°		
Proton										
G(1)H-8	7.943	7.949	7.958	7.967	7.992	8.027	8.055	8.078	22.3	
C(2)H-6	7.741	7.734	7.728	7.714	7.691	7.667	7.656	7.555	30.7	
G(3)H-8	7.962	7.962	7.954	7.944	7.899	7.831	7.774	7.728	20.1	
C(2)H-5	5.855	5.812	5.770	5.692	5.566	5.318	5.329	5.256	28.7	
G(1)H-1'	5.839	5.833	5.829	5.824	5.823	5.826	5.832	5.837	NSB	
C(2)H-1'	5.917	5.901	5.884	5.850	5.789	5.716	5.664	5.629	27.3	
G(3)H-1'	5.845	5.840	5.836	5.830	5.825	5.826	5.826	5.837	NSB	
								Average T <sub>m</sub>	25.8	
				Coupling constants (Hz)						
G(1)H-1'	4.9	4.4	4.6	3.6	3.0	2.0	--	--		
C(2)H-1'	4.5	4.3	4.1	3.5	2.7	--	--	--		
G(3)H-1'	5.6	5.2	5.7	5.3	5.1	2.0	--	--		

### 3.3. Discussion

#### 3.3.1 Comparison of results with previously published oligoribonucleotide studies.

CCGG was one of the first sequences to be synthesized and studied by variable temperature proton NMR, by Arter *et al.*, in 1974. The chemical shift versus temperature data they obtained are identical to those depicted in Figure 3.8, except that the latter are displaced to lower temperatures, due to the lower strand concentration, with an average of 48.6°C (1.7mM) (Table 3.9) as opposed to 51°C (13mM) for the GH8 and CH6 protons. The hydrogen bonded imino protons were observed by Arter *et al.* at 12.45ppm (internal) and 13.18ppm (terminal) in 1.0M salt (Arter *et al.*, 1974); in 0.1M salt, at the lower concentration, they appear at 12.53ppm (internal) and 13.34ppm (terminal) (Table 3.3).

CCGGp was studied by NMR by Petersheim and Turner, who found identical curves, and a  $T_m$ , at 0.8mM in 1.0M salt, of around 38°C (Petersheim and Turner, 1983b).

The inter-residue NOEs of CGCG and CGCGCG were recently studied to resolve whether the duplexes in 1.0M salt were A- or Z- form, as the circular dichroism spectrum of each at low temperature shows a large negative band hitherto believed unique to the Z-form (Westerink *et al.*, 1984; Uesugi *et al.*, 1984). The published chemical shift versus temperature curves are identical in shape to those in Figure 3.13; the quoted  $T_m$  was 40°C (2.9mM, at 0.1M salt), with no indications as to which proton  $T_m$ s were averaged, which is comparable to the value of 37.2°C (2.5mM, 0.1M salt) averaged over all protons, or 39.4°C, for base protons only (Table 3.14). The magnitude and pattern of the inter-residue NOEs verified that the duplex was A-form (Uesugi *et al.*, 1984).

CCGG, CGGC and CGCG were also studied by P.R. Romaniuk (Romaniuk, 1979). All

three gave similar results to those obtained here; curve shapes were identical, and the differences in  $T_m$ s were accountable by the concentration difference. CCGG (9.2mM, 1.0M salt) formed a perfect duplex with a  $T_m$  of 50°C, averaged for the CH6 and GH8 protons—compared to a  $T_m$  of 48.6°C, for CCGG at 1.7mM. CGGC displayed the same marked broadening indicative of formation of an extended loose duplex,  $T_m$  33.5° at 9.2mM—compared to a  $T_m$  of 29.2°C at 2.1 mM. The results of CGCG did not allow a decision to be made regarding whether the duplex was perfect or staggered with aggregation. The curves for the GH8s had an average  $T_m$  of 41°C, while those of the CH6s averaged 47°C; there was no indication that some base pairs might be forming before others. The quoted  $T_m$  was 44°C, for base protons, compared to 39.4°C.

According to an early study which involved monitoring of the interaction between GGC and GCC by optical rotary dispersion, these two trimers formed a  $(GGC)_{2n}(GCC)_n$  complex (total single strand concentration 9.3 mM), with considerable GGC self aggregation (Jaskunas *et al.*, 1968) particularly in the presence of magnesium ion. In the NMR spectra the signals of GGC and GCC displayed similar broadening, but it was not possible, in the absence of observable imino protons, or without a continuous variation experiment at NMR concentrations and conditions, to exclude the possibility of a 2:1 complex.

### 3.3.2 Relative stability of diribonucleotide cores: Results versus literature

The ranking for the stability of the GC diribonucleotides cores obtained in this study was  $GC:GC \geq GG:CC > CG:CG$ . This differs from the ranking of  $GG:CC > GC:GC > CG:CG$  in the early published results of Borer and coworkers (Borer *et al.*, 1974), derived from oligoribonucleotide data, and from the results of Ninio and coworkers, derived from

computer refinement (Ninio, 1981; Papanicolaou *et al.*, 1984). It is in agreement with the thermodynamic results of Freier *et al.*, 1985, derived from measurements on sequences which contained exclusively GC cores, and with those of Ornstein and Fresco, who used computer fitting of estimated parameters to observed oligoribonucleotide stabilities (Ornstein and Fresco, 1983a). The stabilities of GC:GC and GG:CC are very similar, and may be modulated further by second neighbour effects.

The instability of CG:CG relative to the other two cores is a common observation. Its deoxy analog, d(CG):d(CG), is also less stable than the other two dG-dC cores, and three mixed GG/AT cores, according to the results of Ornstein and Fresco (1983b). The instability of CG:CG and d(CG):d(CG) might be attributed to purine clash (Callandine, 1981), in both major and minor grooves in the B-form helix, and in the minor groove in the A-form, and to dipole-dipole interactions between the stacked guanosines.

### 3.3.3. Comparison of the duplexing behaviour of RNA and DNA tetramers

The duplexing behaviour of the deoxyribonucleotide sequence analogues of four of the tetramers, d(GGCC), d(CCGG) (Patel, 1977a), d(GCGC) (Patel, 1979) and d(CGCG) (Patel, 1976), has been studied by proton NMR by D.J. Patel. Table 3.27 shows in detail the comparisons of individual proton  $T_m$ s and chemical shifts upon duplexing.

In comparing the proton  $T_m$ s the large differences in concentration should be noted. Even with this difference, which should be to the advantage of the DNA sequences, the  $T_m$ s of the protons in the DNA duplexes are lower than those in the RNA duplexes in three of the four cases, while in the fourth (CGCG) the higher  $T_m$  in the DNA sequence is probably accountable for by the sixteenfold concentration difference. There is also the possibility that the perfect duplex/staggered aggregate equilibrium is

Table 3.27: Comparison of  $T_m$ s and observed and predicted chemical shift changes upon duplexing for the aromatic base protons of the oligoribonucleotide duplexes GGCC, CCGG, GCGC and GCGC, and their deoxy analogues (from the literature)

	RNA		DNA		PREDICTED <sup>a</sup>	
	Shift	$T_m$	Shift	$T_m$	A-RNA	B-DNA
GGCC <sup>b</sup>						
G(1)H-8	+0.12		+0.12		-0.03	+0.02
G(2)H-8	>+0.45	51.4	-0.08	37.5	+0.34	+0.11
C(3)H-6	≈ 0.12	54.2	+0.33	40	+0.31	+0.08
C(4)H-6	≈ 0.12		>+0.39		+0.14	+0.02
C(3)H-5	>+0.5	53.7	>+0.56	41.5	+1.12	+0.55
C(4)H-5	>+0.7		>+0.67		+0.5	+0.23
CCGG <sup>b</sup>						
C(1)H-6	-0.12		≈ 0.0		-0.03	+0.01
C(2)H-6	-0.17	47.2	+0.12	53	-0.05	+0.01
G(3)H-8	+0.27	44.7	+0.09*	35.6*	+0.15	+0.04
G(4)H-8	+0.55	44.2	-0.04*	35*	+0.42	+0.11
C(1)H-5	+0.06		+0.14		-0.04	+0.01
C(2)H-5	+0.48	44.5	+0.27	42	+0.37	+0.22
GCGC <sup>c</sup>						
G(1)H-8	+0.08		-0.09*		-0.03	+0.01
C(2)H-6	-0.12	46.5	+0.18*	39.5*	+0.24	+0.02
G(3)H-8	+0.33		-0.05*		+0.24	+0.05
C(4)H-6	+0.25	48.9	+<0.4*		+0.32	+0.08
C(2)H-5	>+0.60		+0.41*	39.5*	+0.98	+0.52
C(4)H-5	>+0.65	47.6	<+0.42*		+1.10	+0.52

(Continued on next page)



Table 3.27 (continued)

	RNA		DNA		PREDICTED <sup>a</sup>	
	Shift	T <sub>m</sub>	Shift	T <sub>m</sub>	A-RNA	B-DNA
CGCG <sup>d</sup>						
C(1)H-6	-0.22		+0.05		-0.03	+0.01
G(2)H-8	+0.2		0.0*		+0.09	+0.03
C(3)H-6	+0.14	43.5	+0.3	55	+0.28	+0.08
G(4)H-8	+0.3		+0.05*		+0.19	+0.05
C(1)H-5	= 0.0		>+0.2		-0.04	+0.03
C(2)H-5	>+0.55	42.0	+0.45	50	+1.06	+0.52

<sup>a</sup>Calculated from tables in Arter and Schmidt, 1976.

<sup>b</sup>RNA data: GGCC, 2.8mM single strand concentration, 0.1M salt buffer, CCGG, 2.5mM single strand, 0.1M salt buffer. DNA data from Patel, 1977: d(GGCC) and d(CCGG) 20mM single strand, 0.1M phosphate buffer.

<sup>c</sup>RNA data: GCGC, 2.5mM single strand, 0.1M salt buffer. DNA data from Patel 1979: d(GCGC), 20mM single strand, 0.1M phosphate buffer.

<sup>d</sup>RNA data: CGCG, 2.5mM single strand, 0.1M salt buffer. DNA data from Patel 1976: d(CGCG), 40mM single strand, 0.1M cacodylate buffer

\*Protons assigned as to type, but not individually, therefore any two GH-8s, CH-6s, or CH-5s can be interchanged.

different for deoxy sequences, and the duplex formation is more favoured in d(CGCG). The overall result, that DNA duplexes are less stable than RNA duplexes of the same length and base composition, does not agree with that of Pardi *et al.*, who found no difference in stability of sequences of the type d/r(CA<sub>5</sub>G: CU<sub>5</sub>G); these sequences, it must be noted, are predominantly A•U(T) (Pardi *et al.*, 1981).

Changes in chemical shift upon duplexing for individual protons showed some interesting differences, particularly in the shifts observed for the GH8s, which were substantially larger in the ribonucleotide sequences (Table 3.27). If the overlaps predicted by the classical A and B helices are examined (Figure 3.25), it is apparent that the B-helix, being more tightly wound, forces adjacent base-pairs to stack across each other in such a way that the GH8 is always poorly shielded by the neighbouring base pairs and this will be unaffected by adjustments to the winding angle. In the A helix the bases lie along each other, so that where the geometry of the bases and turn angle allows, the GH8 proton experiences shielding by one of the neighbouring bases. 5' GH8 in GC:GC and GG:CC will always be poorly shielded by its 3' neighbour because of the length of the purine base and the positive turn angle, but 3' GH8 in CG:CG and CC:GG are shielded by 5' G and 5' C respectively, to an observed value of 0.2 - 0.3ppm.

Differences between the changes in the chemical shifts upon duplexing of the CH6s tend to be small but in all cases the shifts upon duplexing are 0.1 - 0.2 ppm more upfield in the deoxy tetramers than in the ribo tetramers. Inspection of the overlaps cannot immediately account for this, since such inspection would place the ribo CH6s in a marginally better position as regards shielding. It may be that the deviations from the classical overlaps observed in actual sequences shield CH6 to advantage in deoxyoligomers without affecting GH8.

Due to the still appreciable slopes of the CH5s at 70°C—the sequences are not

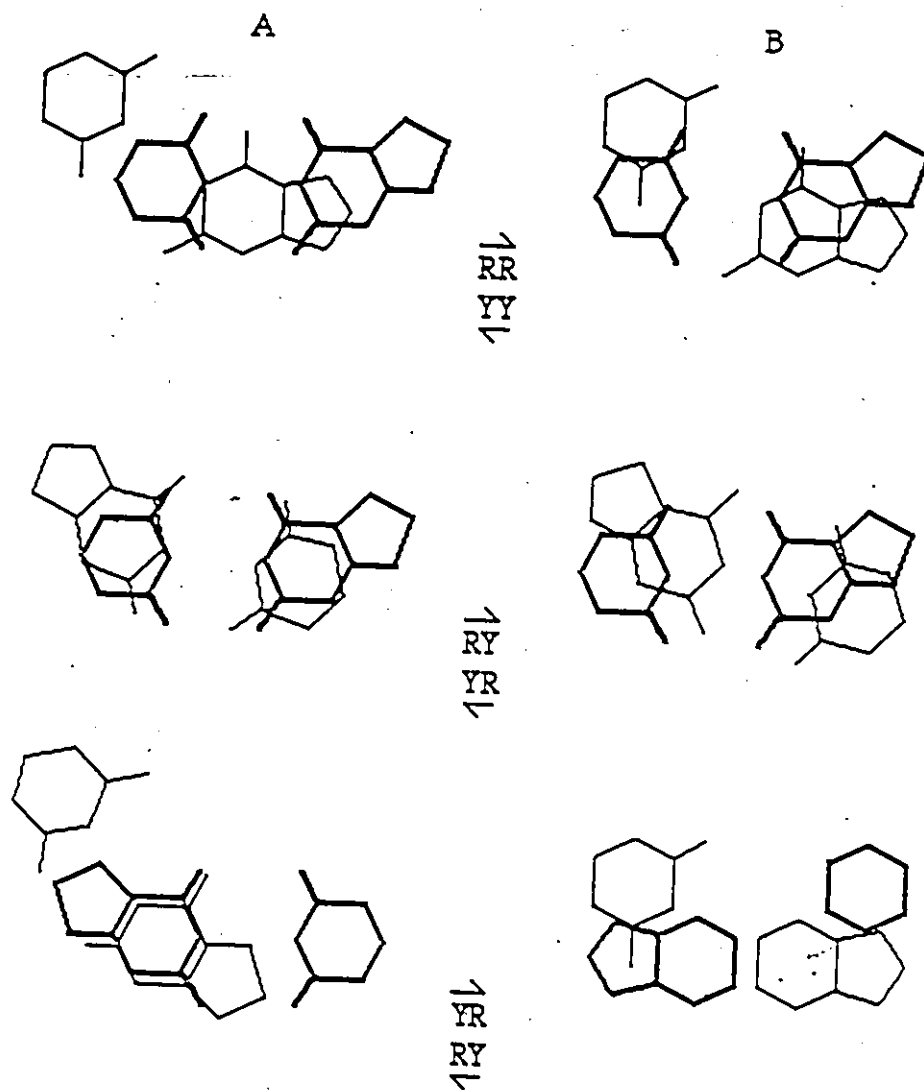


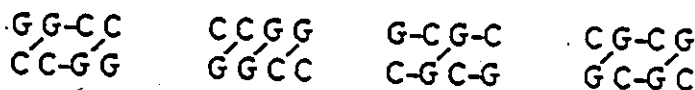
Figure 3.25: Comparison of the base pair overlaps according to the classical A- and B-helix models for the three cores (after Arnott et al., 1973; Arnott and Hulkins, 1973).

completely in the random coil—it is possible only to put a lower bound on the chemical shift change upon duplexation of the ribose CH5s. In six of the eight cases where substantial changes are seen the chemical shift changes on upon formation of the ribo duplexes are larger than than the chemical shift changes upon formation of the deoxy duplexes—how much larger, it is not possible to tell. In the remaining two cases, both CH5s occur on a 5' terminal cytidine (CCGG and CGCG), and the shift changes are small, but slightly larger in the deoxy sequences. The larger chemical shift changes observed for most CH5s in the ribo sequences arise—as did the larger GH8 shifts—for the more advantageous overlap in the A-type helix, which places CH5 directly over its shielding neighbour, as compared to the B-type helix, where the base pairs are skewed relative to each other, so that CH5 is outside the zone of maximum shielding.

#### 3.3.4 Application of the RY model to the tetramer results

Using NMR data to adjust base overlaps, Bubenko et al. (1983), proposed a modification to the A'-RNA helix according to which the winding angles become sequence dependent. The normal 33° winding angle of A'-RNA is replaced by 20° for RY:RY, 50° for RR:YY and 45° for YR:YR (R denoting purine, Y, pyrimidine) with the result that in RY:RY, the purine and pyrimidine form an intrastrand stack, and in RR:YY and YR:YR the two 3' residues in the dinucleotide unit form an interstrand stack (Figure 3.1).

The stacking interactions according to the RY model are shown below; dashes indicate bases which are stacked upon each other.



Length of stack alone cannot be used as a criterion of stability: GGCC, with a  $T_m$  of  $54^\circ\text{C}$  has the same number and length of stacks as CGCG, with a  $T_m$  of  $36.4^\circ\text{C}$ ; composition of stacks must also be taken into account. The stacking in GGCC is C-G-C, while that in CGCG is G-G-C. In an earlier paper by the same authors, they mentioned the CG:CG appears to have an anomalous winding angle of  $35^\circ$ , as opposed to a general YR:YR value of  $45^\circ$ , which results in poor G-G stacking, (Cruz *et al.*, 1983).

The RY model, does, however, predict the apparent greater flexibility of the 5' terminal residue in GGCC, CCGG and CGCG, compared to the G1 in GCGC. The 5' residue in GGCC, CCGG and CGCG are not predicted to be part of any stack, and the  $J_{1,2}$  couplings for these residues are the last to collapse with decreasing temperature (Figures 3.26, 3.27 and 3.29), and remain resolved to at least  $15^\circ$  below the average molecule  $T_m$ . In GCGC, the 5' terminal guanosine is stacked upon its 3' neighbour, and the coupling constants of G1 and G3 collapse before those of C2 and C4, and all four are fully collapsed by  $40^\circ\text{C}$ ,  $8^\circ$  below the  $T_m$  (Figure 3.28).

### 3.3.5. A proposed structure for the GCG aggregate

A structure for the GCG aggregate was sought which satisfied the following criteria:

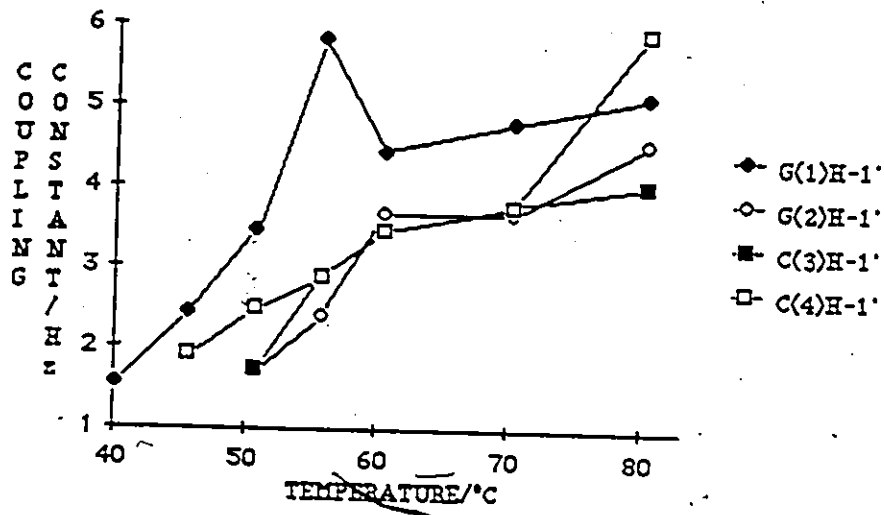


Figure 3.26: Coupling constants versus temperature for the anomeric protons of GGCC, 2.5 mM, in 0.1 M salt buffer.

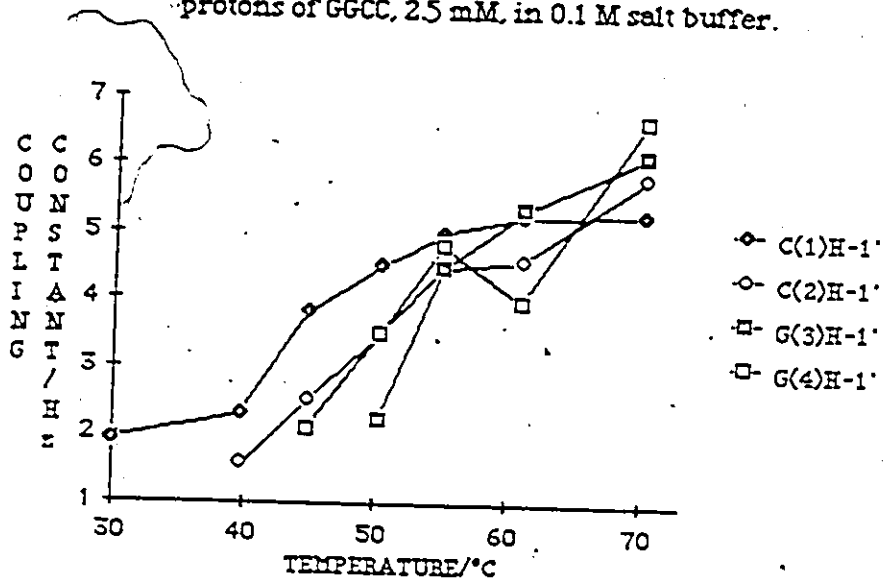


Figure 3.27: Coupling constants versus temperature for the anomeric protons of CCGG, 2.5 mM, in 0.1 M salt buffer.

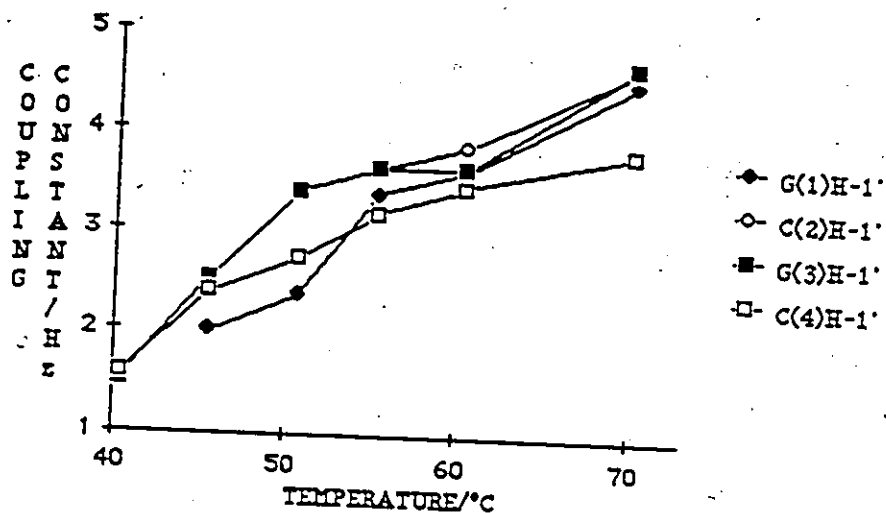


Figure 3.28: Coupling constants versus temperature for the anomeric protons of GCGC, 2.4 mM, in 0.1 M salt buffer.

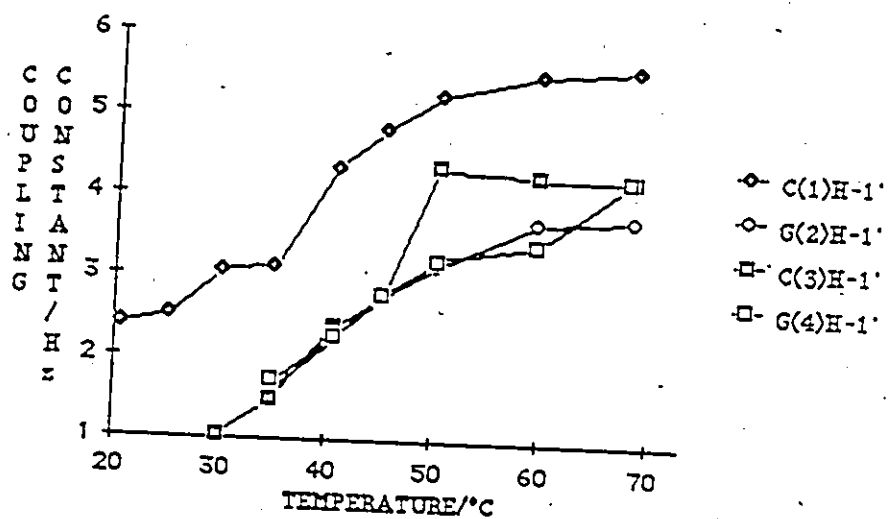


Figure 3.29: Coupling constants versus temperature for the anomeric protons of CGCG, 2.5 mM, in 0.1 M salt buffer.

1. The basic unit is the miniduplex comprising two G-C base pairs with two 3' dangling guanosine residues, since the upper portions of the chemical shift versus temperature curves of GCG (Figure 3.20; Table 3.21) are identical to those forming a duplex of this type [GCA (Alkema et al., 1981), GCG\* (D'Andrea et al., 1983) and GCI(D.W. Hughes, unpublished results)]; there is, moreover, only one imino proton observed in the G-C hydrogen bonding region (Table 3.22), thereby suggesting identical environments for all G-C imino protons.

2. Hydrogen bonding to and from G3 must make a substantial contribution to the forces holding the aggregate together, since in the absence of available H-bonding sites (as in GC<sup>1-Me</sup>G), aggregation does not occur.

3. The aggregate must contain a single environment for G-C imino protons, poorly shielded (one signal appears at 13.43ppm) and two non-equivalent environments for guanosine imino protons hydrogen-bonded to carbonyl oxygens.

Study of the properties of guanosine monophosphate solutions have led to the postulation of G-G self pairing of the type shown in Figure 3.30a (Gellert et al. 1962; Sagan and Walmsley, 1985). This pairing can be extended to include Watson-Crick base pairing to cytidine (Figure 3.30b), and this kind of structure has been proposed to account for the 2:1 stoichiometry observed in such systems as mixtures of polyG and polyC (Cantor and Schimmel, 1981). The G-G bonding in these structures would produce imino protons in the upfield region.

A structure has been devised comprising two extended loose duplexes of GCG dimers which form the front and back faces of a cuboid arrangement, held together by G-G bonds along the side faces (Figure 3.31). Adjacent strands run antiparallel to each other. Two levels of a three level repeating unit consist of G-G-G-C bonding (which is possible at least on paper—see Figure 3.30c), while the third level consists of two



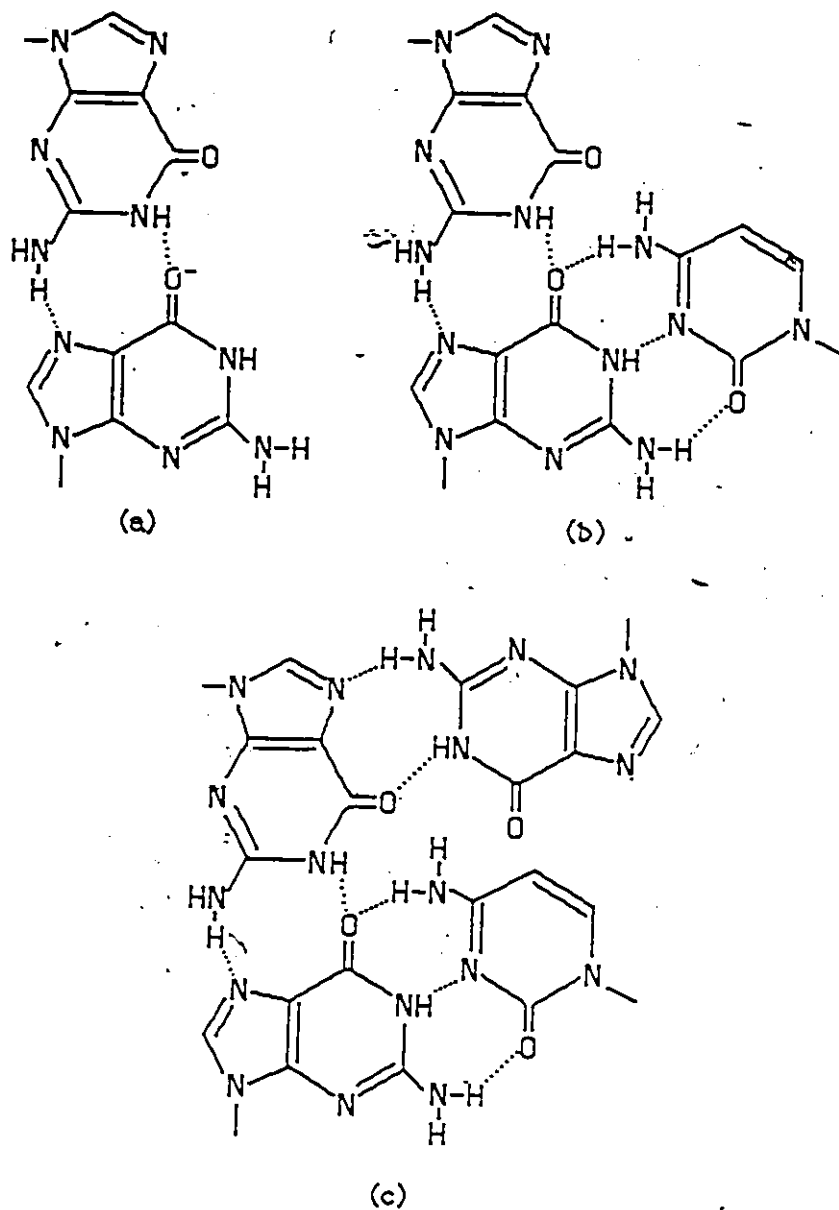
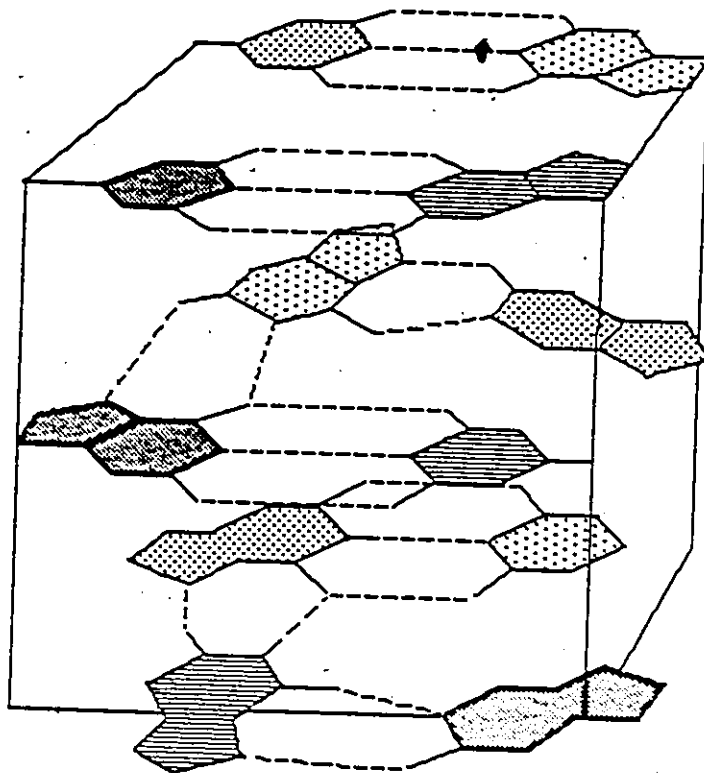


Figure 3.30: Bonding arrangements of GCG aggregates (a) G⋅G base-pairing, (b) G⋅G⋅C base pairing and (c) G⋅G⋅G⋅C base pairing.



Key to shading:





- |   |  |   |                                 |
|---|--|---|---------------------------------|
|  | 5'-3' top to bottom,<br>back face, W-C bonded to:  |  | 5'-3' bottom to top, back face  |
|  | 5'-3' top to bottom,<br>front face, W-C bonded to: |  | 5'-3' bottom to top, front face |

Figure 3.31: Proposed aggregate structure for GCG in 1.0 M salt.

separate G-C base pairs. There is no strand crossover between front and back faces; only G-G bonds hold the front and back faces together. If this kind of structure is being formed in solution, at least part of the stability must be provided by solvent exclusion and hydrophobic interaction in the core of the aggregate. The structure satisfies the requirement for two separate environments for the G-G hydrogen-bonded imino protons, but it requires an extreme winding angle and syn configuration of certain of the G3 residues (Figure 3.31). Molecular modelling studies will be required to see whether both the G-G-G-C arrangement and the syn conformation of the G3 would be possible.

### 3.3.6. Salt dependence of aggregation

In the presence of 1.0 M salt, GGA, GCG and CGG aggregate, both in isolation and in a 1:1 mixture with their complementary sequences, CCA, CGC and CCG respectively (Section 3.2.6.2). Upon removal of the salt, GGA, GCG:CGC and GCG alone all duplex with measurable  $T_m$ s (Section 3.2.6.3), but there is indication that aggregation is beginning at lower temperatures. Increasing cation concentration increases duplex stability on account of cation shielding of the phosphate groups—which must approach each other more closely in the duplex. According to simple theory this dependence is linear with the logarithmic concentration of the counterion concentration; in actuality, the dependence is  $1/T_m$  proportional to  $\ln[M]$ , but over narrow ranges of temperature the two expressions are approximately equivalent (Cantor and Schimmel, 1981). Given that aggregation involves the close association of numerous strands, then counterion shielding of the phosphate groups will be even more significant.

### 3.3.7. Comparison of calculated and observed imino protons shifts

Parameters for the prediction of imino proton chemical shifts, based on ring current calculations and classical helix geometries, have been published by Kearns (1976) and Arter and Schmidt (1976), and applied to results for tRNA. Agreement was reasonable, although the method was eventually superseded by the development of NOE techniques for the assignment of larger molecules. Comparison of the predicted imino protons spectra with those obtained for the four, five and six base base pair duplexes are shown in Table 3.28; a consistent pattern of differences between predicted and observed values is observed for most of the duplexes:

1. The terminal imino proton appears upfield of the predicted value, in all cases except CCGG, and related sequences, and CGCG. In duplexes with terminal GC:GC or GG:CC cores, the difference between predicted and observed chemical shifts is 0.5 - 0.6ppm, and in those with terminal AG:CU, the difference is 0.3 - 0.4 ppm.
2. The imino proton of the first internal base pair appears 0.2 - 0.4 ppm downfield of its predicted position.

The assignment of some of these sequences is discussed in the relevant sections in this thesis: GGCC, CCGG, GCGC and CGCG in Section 3.2.3; GGACC:GGUCC in Section 4.2.1; CCAGG:CCUGG and CCAUGG in Section 4.2.2; GCAGC:GCUGC and GCAUGC in Section 4.2.3. The component sequences of the complementary mixtures AGGCU:AGCCU, AGGCU:AGUCU and AGACU:AGUCU were synthesized, and their duplexing behaviour studied by variable temperature proton NMR of the nonexchangeable protons by Dr. Dirk Alkema (Alkema, 1982; Alkema et al., 1982); the imino proton data was obtained for samples already deblocked and mixed.

Table 3.28: Comparison of observed and predicted chemical shifts for the imino protons of the perfect duplexes studied in Sections 3 and 4. Numbering of base pairs from left to right as shown.

		Observed	Predicted	
			Kearns <sup>a</sup>	A&S <sup>b</sup>
GGCC*	G•C1	12.40	12.9	12.99
CCGG	G•C2	13.49	13.2	13.08
1				
GGACC*	G•C1	12.29	12.9	12.93
GGUCC	G•C2	12.63	12.3	12.36
1 5	A•U3	14.49	14.2	13.91
	G•C4	13.41	13.3	13.09
	G•C5	12.29	12.9	12.99
CCGG*	G•C1	13.34	13.35	13.32
GGCC	G•C2	12.53	12.2	12.57
1				
CCAGG*	G•C1	13.31	13.4	13.20
GGUCC	G•C2	12.34	11.9	12.37
1 5	A•U3	13.31	13.1	13.61
	G•C4	13.12	12.7	12.92
	G•G5	13.31	13.4	13.25
CCAUGG*	G•C1	13.31	13.4	13.20
GGUACC	G•C2	12.36	11.9	12.31
1	A•U3	13.31	13.8	13.68
GCGC*	G•C1	12.76	13.4	13.23
CGCG	G•C2	13.05	12.7	12.85
1				
GCAGC*	G•C1	12.64	13.4	13.19
CGUCG	G•C2	12.64	12.4	12.52
1 5	A•U3	13.59	13.3	13.63
	G•C4	13.59	13.3	13.16
	G•C5	12.64	13.4	13.30
GCAUGC*	G•C1	12.74	13.4	13.19
CGUACG	G•C2	12.74	12.4	12.53
1	A•U3	13.37	13.8	13.56

(Continued on next page)

(Table 3.28 continued)

CGCG* GCGC 1	G•C1	12.87	12.9	13.00
	G•C2	13.15	12.7	12.81
GCCG+ CGGC 1 4	G•C1	12.84	13.4	13.25
	G•C2	13.37	13.2	13.08
	G•C3	12.63	12.2	12.57
	G•C4	12.63	12.9	12.98
AGGCU+ UGCGA 1 5	A•U1	13.55	13.9	14.0
	G•C2	12.97	12.8	12.9
	G•C3	13.55	13.15	12.9
	G•C4	13.55	13.3	13.2
	A•U5	13.55	13.9	14.02
AGGCU+ UCCGA 1 5	A•U1	13.56	13.9	14.00
	G•C2	12.97	12.8†	12.9†
	G•U3	11.49/10.74	—	—
	G•C4	13.56	13.4†	13.2†
	A•U5	13.56	14.0	14.02
AGACU+ UCUGA 1 5	A•U1	13.42	13.9	13.95
	G•C2	12.61	12.4	12.42
	A•U3	14.45	14.1	13.9
	G•C4	13.42	13.4	13.17
	A•U5	13.42	13.9	14.11

<sup>a</sup>Calculated using values from Kearns, 1976.

<sup>b</sup>Calculated using values from Arter and Schmidt, 1976.

\*Data in 0.1 M salt buffer.

†Data in 1.0 M salt buffer.

‡Calculated assuming same ring current effect for neighbouring G•C and G•U.

### 3.3.7.1. Assignment of imino proton resonances for AGXCU:AGYCU series

At the lowest temperature,  $-2^{\circ}\text{C}$  for the AGGCU:AGCCU and AGGCU:AGUCU mixtures, and  $1^{\circ}\text{C}$  for AGACU:AGUCU (Table 3.29, 3.30 and 3.31), hydrogen-bonded imino protons are visible for all five base pairs in all three mixtures, five signals for AGACU:AGUCU and AGGCU:AGCCU and six for the AGGCU:AGUCU mixture. By  $10^{\circ}\text{C}$ , two of the signals occurring about 13.5ppm in all mixtures are no longer visible. The other signals persist until at least  $20^{\circ}\text{C}$  in all sequences.

Loss of an hydrogen-bonded imino proton signal is generally interpreted to mean opening of the hydrogen bonds allowing exchange with solvent. As these sequences terminate in A•U base pairs, fraying of the ends prior to melting might be expected; the first signals to be lost would be those due to the terminal A•U base pairs. In each sequence, two signals disappear at  $10\text{-}20^{\circ}\text{C}$  below the others, suggesting that they are due to the terminal UN3H protons. In addition, these signals are at the same position in the spectra of each of the three mixtures, as might be expected, since the three differ only in their second neighbours (Tables 3.29, 3.30, 3.31).

The resonances due to the central base pair protons may be assigned by their uniqueness when the spectra are compared: Only the spectrum of the AGACU:AGUCU mixture has a signal at 14.45ppm, which then may be assigned to the central A•U base pair.

The spectrum of the AGGCU:AGUCU mixture shows three unique resonances, at 12.28, 11.49 and 10.74ppm. The first is very sharp compared to the other signals in the spectrum, and since the latter two are in better agreement with published chemical shifts for G•U and G•T base pairs (Robillard *et al.*, 1976; Kearns and Schulman, 1976; Patel *et al.*, 1982a), they are tentatively assigned to the G•U base pair.

Table 3.29: Chemical shifts of hydrogen-bonded imino protons of AGGCU:AGCCU, 2.5 mM each strand, in 1.0M salt buffer, over the temperature range -2 to 29°C.

Numbering scheme:     1     5  
                          AGGCU  
                          UCCGA

Temperature	Chemical shift (ppm)			
	<u>-1.9°</u>	<u>10.6°</u>	<u>20.6°</u>	<u>29.2°</u>
Proton				
A•U1	13.55	—	—	—
G•C2	12.97	12.95	12.91	—
G•C3	13.55	13.50	13.45	13.40
G•C4	13.55	13.30	13.32	—
A•U5	13.55	—	—	—

Table 3.30: Chemical shifts of hydrogen-bonded imino protons of AGGCU:AGUCU, 2.5 mM each strand, in 1.0M salt buffer, over the temperature range -2 to 20°C.

Numbering scheme:     1     5  
                          AGGCU  
                          UCUGA

Temperature	Chemical shift (ppm)		
	<u>-1.9°</u>	<u>10.6°</u>	<u>20.6°</u>
Proton			
A•U1	13.56	—	—
G•C2	12.97	13.04	12.80
G•U3	11.49	11.53	11.60
	10.74	—	—
G•C4	13.56	13.65	13.67
A•U5	13.56	—	—



Table 3.31: Chemical shifts of hydrogen-bonded imino protons of AGACU:AGCCU, 3 mM each strand, in 1.0M salt buffer, over the temperature range -2 to 20°C.

Numbering scheme:  $\begin{array}{c} 1 \quad 5 \\ \text{AGGCU} \\ \text{UCCGA} \end{array}$

Temperature	Chemical shift (ppm)		
	<u>1.0°</u>	<u>8.7°</u>	<u>19.4°</u>
Proton			
A•U1	13.42	—	—
G•C2	12.61	12.68	12.58
A•U3	14.45	14.46	14.41
G•U4	13.42	13.48	13.42
A•U5	13.42	—	—

Spectra of all three compounds have one GN3H resonance appearing at or below 13ppm, and one appearing above. The AGGCU:AGCCU mixture has an additional GN3H resonance as part of the broad peak centred at 13.55ppm; this, then, is assigned to the central G-C base pair.

The signals due to the internal G-C base pairs appear at 13.55 and 12.97 ppm (AGGCU:AGCCU), 13.56 and 12.97ppm (AGGCU:AGUCU) and 13.42 and 12.61 ppm (AGACU:AGUCU). Following the reasoning that the substitution of an A•U base pair for a G-C base pair in the centre position would increase the shielding of the imino proton on the 5' neighbouring guanosine residue more than that of the 3' neighbouring residue, the resonances are assigned as: GC4 (numbered for the strand containing a central purine) 13.56 ppm (AGGCU:AGCCU and AGGCU:AGUCU) and 13.42 ppm (AGACU:AGUCU) and GC2, 12.97 ppm (AGGCU:AGCCU and AGGCU:AGUCU) and 12.61 ppm (AGACU:AGUCU) (Tables 3.29, 3.30, 3.31).

### 3.3.7.2. Discussion of the trends in imino proton behaviour

The opposing direction of the two trends of divergence of predicted and observed imino proton chemical shift--upfield for the terminal protons and downfield for the internal ones--precludes simple adjustment of the zero positions of the base pairs to best fit the data. Both sets of shielding parameters have been calculated for 11-RNA geometry and refined against tRNA data. The minihelices in tRNA close hairpin loops, or have dangling bases, whereas the synthetic oligoribonucleotide

duplexes have blunt ends. Imino proton spectra were obtained for certain available members of the series AGCUX and XAGCU, which form minihelices with dangling bases, but the A•U imino protons could not be observed at  $-5^{\circ}\text{C}$ .

Possible explanation of the deviations of observed from predicted values include end-to-end aggregation, fraying of the ends, and altered base overlap at the terminal base pairs.

Stacking of the duplexes upon each other at low temperature would result in shielding of the terminal base-pair imino protons; it might be expected, however, that as the aggregate was dispersed upon warming, the imino proton signals would move downfield towards their predicted value before broadening and disappearing, and this movement is not observed.

Fraying of the ends of the duplex with exchange between duplexed and non-duplexed states would cause an upfield shift of the observed resonance towards the position of the free GN1H. Such an explanation would be more credible for the sequences ending in A•U, where the terminal imino protons disappear before  $8^{\circ}\text{C}$ , than for those ending in G•C, where the terminal imino protons persist to  $20^{\circ}\text{C}$  and above, and where the linewidths—which are also indicative of the rate of exchange—are similar for the terminal and the internal imino protons at low temperature (See Figures 3.4, 3.5 and 3.6). The shifts before broadening and disappearance tend to be small— $0.14\text{ppm}$  being the largest, for CCGG—and can be either upfield or downfield (See Table 3.3).

Examination of the base overlaps for the 11-RNA geometries used to calculate the shielding coefficients shows that for RR:YY and RY:RY, the shielding of the imino protons of the first base pair may be improved by increasing the winding angle, as shown in (Figure 3.32). For RR:YY, increasing the winding angle places the five

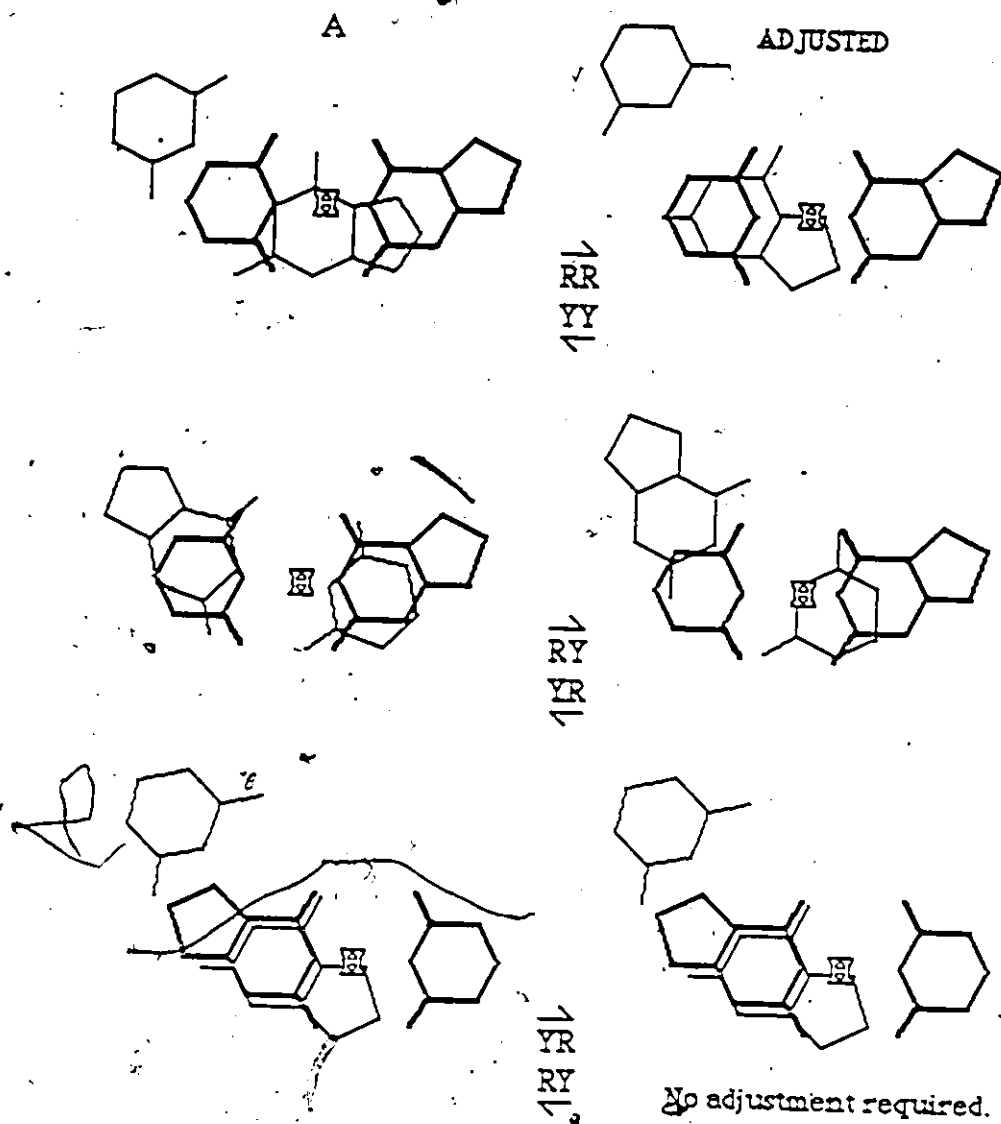


Figure 3.32: Trends of winding angle adjustment required to fit the chemical shift data for imino protons belonging to terminal base pairs.

membered ring of the 3' purine, the region of greatest shielding effect, directly under the imino proton involved in the terminal bond. The change also moves the imino proton of the second (internal) base pair into a region of lesser shielding, which is also consistent with observation. This opening of winding angle is according to the RY model (Figure 3.1).

To improve the shielding of the imino proton in the RY:RY unit, the winding angle must be opened substantially to move the 3' pyrimidine from beneath the 5' purine, to a position between the two bases, beneath the imino proton (Figure 3.32). This should, however, also result in more advantageous shielding to the internal base pair proton. The RY model predicts a decreased winding angle of RY:RY sequences—to 20°. Crystal structures of deoxyoligonucleotides show an increased winding angle for RY:RY units, accounted for by the avoidance of "purine clash" (Callendine, 1981).

Crystal structure data shows no drastic changes in winding angle from terminal relative to internal dinucleotide cores, although all crystal structures published to date are of deoxyoligonucleotides.

#### 3.4. Conclusions

The duplexing behaviour of self-complementary tetramers and complementary mixtures of tetramers and trimers which formed duplexes consisting of G-C base pairs was studied with the intention of using the stabilities of these duplexes to determine the relative stabilities of the three GC diribonucleotide cores, GC:GC, GG:CC and CG:CG. In 1.0M salt buffer the sequences which formed perfect duplexes were GGCC, CCGG, GCGC, CGCG and the mixtures GCCG:CGGC and GGC:GCC. The helixes were A-type; observed chemical shifts for certain protons are larger than those observed for the deoxy

analogues, for which data had been previously published (Patel, 1977a, 1979), owing to the better overlaps in the RNA helix. The order of stability deduced for the dinucleotide cores was:  $GC:GC \geq GG:CC > CG:CG$ , which disagrees in the ordering of the first two cores with earlier thermodynamic parameters (Borer *et al.*, 1974) but agrees with more recent ones (Freier *et al.*, 1985a).

The sequences GCCG and CGGC formed staggered duplexes, with a single dinucleotide core, and four 3' double-dangling bases—sticky ends which promote overlap aggregation and formation of an extended loose duplex.

In the remaining mixtures at 1.0M salt buffer, GGA with CCA, GCG with CGC and CGG with CCG, the G-rich strand formed a self-aggregate in preference to forming a duplex. This behaviour also occurred for these sequences in isolation, and the observation of a well-defined imino proton spectrum for GCG allowed a proposal to be made for the structure of that aggregate.

The aggregation is salt-dependent: two of the three mixtures, GGA:CCA and GCG:CGC and GCG alone formed duplexes with measurable  $T_m$ s when the salt was removed. Evidence of aggregation occurring at lower temperatures was observed for GCG and CCG:CCG.

#### 4. THE EFFECT OF THE INTRODUCTION OF CENTRAL BASE-BASE MISMATCHES ON THE SECONDARY STRUCTURE AND STABILITY OF THE GC TETRAMERS.

##### 4.1. Introduction

Base mismatches in nucleic acid double helices have been studied largely from the standpoint of their identity in DNA, as replication errors giving rise to mutation. The questions addressed pertain to the mechanism of their introduction, recognition and excision by repair enzymes. Mutations arising from single-base replication errors are of two types, point, or frameshift, with point mutations being further subdivided as transitions or transversions—in the former case purine substitutes for purine and pyrimidine for pyrimidine, and in the latter, pyrimidine replaces purine and vice versa (Stryer, 1981).

Two hypotheses have been proposed to account for the base-base mispairing required for point mutation:

1. The wobble hypothesis, according to which, pairing occurs between major tautomeric forms, with the bases sliding relative to one another to make the possible hydrogen bonds (Crick, 1976).
2. The tautomer hypothesis, which holds that pairing between mismatched pairs occurs with one base in a minor tautomeric form to maximize hydrogen bonding and preserve the orientation of the bases within the helix (Watson and Crick, 1953; Topal and Fresco, 1976). Tautomer formation can be enhanced by chemical modification of the bases by mutagens.

Results from most of the recent studies on deoxyoligoribonucleotides tend to

support the wobble model of base pairing for unmodified bases.

Frameshift mutation occurs by deletion or insertion of a base, or bases, with the result that, should the mutation occur in a coding region, that portion of the gene downstream from the mutation will be out of phase from that upstream, resulting in mistranslation. According to the Streisinger hypothesis (Streisinger *et al.*, 1966), such mutation occurs through transient stabilization of a structure with "looped out" bases on one strand, possibly by an intercalating molecule.

Mismatches in RNA have been given less attention than those in DNA, perhaps because the best characterized RNA until recently has been tRNA, whose double helices were too short to support internal mismatches aside from G•U (which forms a base pair of comparable stability to A•U) (Romaniuk *et al.*, 1979a,b). Accumulating information on the primary sequence and secondary structure of rRNAs and mRNAs suggests that those structures do contain helices with internal base mispairings (Noller, 1974; Woese *et al.*, 1983; Thompson and Hearst, 1983), that those mismatches are conserved between species, and that they are necessary for protein binding—for instance, binding studies of R17 coat protein to an RNA fragment containing its binding site have shown that a single unpaired A residue is necessary for binding (Carey *et al.*, 1983). Thus base mispairing in RNA may serve as a further modulator of stability of "active" helices, and as a recognition or binding site for protein.

The earliest studies of mispairing were performed on mixtures of complementary polyribonucleotides in which a certain fraction of the strand were non-pairing, having either been introduced during synthesis, or chemically modified afterwards (Fink and Crothers, 1972a, 1972b). Maximum base pairing was observed at a 1:1 stoichiometry of complementary residues; it was concluded that normal base



pairing occurred around the defect. In a mixture of poly(A,U):poly(U), the mispaired U residues demonstrated greater sensitivity to photochemical modification, and this was taken as an indication that they were, for at least part of the time, outside the stack (Lomar and Fresco, 1973). "Bulge defects" of this type are part of the Streisinger hypothesis.

Since the development of chemical synthesis of oligoribonucleotides and deoxyoligoribonucleotides base-mispairings have been studied in a variety of systems, particularly by NMR, since NMR allows examination of local conformation. Table 4.1 contains a summary of type of defect, system and reference.

Certain common conclusions emerge from these studies:

1. Mismatches are destabilizing.

Replacement of G-C 3 in the duplex formed by d(CGCGAATTCGCG) ( $T_m$  72°C, 0.7mM, 0.1M phosphate) by GxT or GxA destabilizes the helix by about 20°C, and by AxC or TxC by 30°C (Patel *et al.*, 1984a, b, c). Theoretical calculations concur with this order of effect (Keepers *et al.*, 1984).

The duplex formed by the mixture of d(CAACTTGATATTAATA) with d(TATTAATATCAAGTTG) is destabilized by 10°C (0.1M salt) by an AxA or AxC mismatch at position 8, while GxT mismatches at position 6, 7, and 12 or a TxC mismatch at position 9 lowers the  $T_m$  by 6°C (Tibayenda *et al.*, 1984).

```

CAACTTGATATTAATA
GTTGAACTATAATTAT
1       6       12

```

Aboul-ela *et al.* have studied all possible mismatches in the system d(CA<sub>3</sub>XA<sub>3</sub>G):d(CT<sub>3</sub>YT<sub>3</sub>G), X=Y=A,C,G,T (Aboul-ela *et al.*, 1985), and could therefore supply an overall ranking of the mismatches (for this system). In descending order of

Table 4.1: Published studies on base-base mismatches in short oligomers.

## (a) OLIGORIBONUCLEOTIDES

<u>Mismatch</u>	<u>System</u>	<u>Reference</u>
CxC, GxG	$A_n X U_n$ $n=4,5$ ; $X=C,G$	Uhlenbeck <i>et al.</i> , 1971
CxC	$A_4 G C_n U_4$ $n \leq 4$	Gralla and Crothers, 1973a,b
CxC, UxU, AxA, GxG	CAXUG $X=C,U,A,G$	Romaniuk <i>et al.</i> , 1979a
GxU	CAGUG:CAUUG	Romaniuk <i>et al.</i> , 1979b
CxC, UxU, AxA, GxG, GxU	AGXCU $X=C,U,A,G$	Alkema <i>et al.</i> , 1982

## (b) DEOXYOLIGORIBONUCLEOTIDES

(In the longer sequences, the bases involved in the mismatch are underlined.)

<u>Mismatch</u>	<u>System</u>	<u>Reference</u>
$A_n \times G_n, G_n \times G_n$	(G):(C <sub>12</sub> A <sub>m</sub> C <sub>x</sub> ) $m=1-6$ (G):(C <sub>12</sub> G <sub>m</sub> C <sub>x</sub> ) $m=1,3-5$	Dodgson and Wells, 1977
TxT	ATCCTATTAGGAT CGTCG	Haasnoot <i>et al.</i> , 1979, 1980 Mellema <i>et al.</i> , 1984b
TxG	CGTGAATT <u>CGCG</u> GGGGCTCC GGGGTCCC TGCGCG	Patel <i>et al.</i> , 1982b, 1984c Brown <i>et al.</i> , 1985 (X-ray) Kennard, 1985 (X-ray) Kennard, 1985 (X-ray)
CxA	CGCGAATT <u>CACG</u>	Patel <i>et al.</i> , 1984a,c

(Continued on next page)

(Table 4.1 continued)

AxG	CGAGAATTCCGG CCAAGATTGG CGCGAATTAGCG	Patel <i>et al.</i> , 1984b,c Kan <i>et al.</i> , 1983 Kennard, 1985 (X-ray)
UxG, AxI, IxA*, UxC	trimer fragments	Chuprina and Poltev, 1983 (theoretical)
TxC, TxG, CxA, AxA	(CAACTTGATATTAATA): (TATTAATATCAAGTTG) assorted positions	Tibayenda <i>et al.</i> , 1984
TxC, TxG, CxA, AxG	CGXGAATTCGCG and CGCGAATTCYCG X,Y=T,A	Keppers <i>et al.</i> , 1984 (theoretical)
all mismatches	CA <sub>3</sub> XA <sub>3</sub> G:CT <sub>3</sub> YT <sub>3</sub> G X,Y=A,G,C,U	Aboul-ela <i>et al.</i> , 1985

stability:  $GxT > GxG > GxA > CxT > AxA = TxT > AxC = CxC$  (Aboul-ela *et al.*, 1985).

The four-base-pair duplex formed by  $r(\text{CAUG})$  has a  $T_m$  of  $24^\circ\text{C}$  (9.2mM, 1.0M salt); but none of the pentamers of the  $r(\text{CAXUG})$  series, where  $X=A, G, C, U$ , form duplexes (Romaniuk *et al.*, 1979a).

$r(\text{AGCU})$  forms a duplex with  $T_m$   $34^\circ\text{C}$ .  $r(\text{AGACU})$  and  $r(\text{AGCCU})$  both duplex to form helices with a central base-base mismatch; the  $T_m$  of  $\text{AGACU}$ , with its central  $\text{AxA}$  mismatch, is  $26^\circ\text{C}$ , and that of  $\text{AGCCU}$ , with a  $\text{CxC}$  mismatch, is  $25^\circ\text{C}$ .  $\text{AGUCU}$  does not duplex and  $\text{AGGCU}$  forms a slipped secondary structure with two  $\text{G}\cdot\text{C}$  and two  $\text{GxU}$  base pairs and two  $5'$  dangling adenosine residues (Alkema *et al.*, 1982).

2. The effect of the mismatch upon the enthalpy of transition is salt dependent.

Patel and coworkers have studied the enthalpies of the helix-coil transition of  $d(\text{CGTGAATTCGCG})_2$  in 0.1M salt and find that both the calorimetric and Van t'Hoff enthalpies are little changed from those of  $d(\text{CGCGAATTCGCG})_2$  (Patel *et al.*, 1982b). They ascribed the destabilization to entropic effects—stronger geometric constraints on the formation of a  $\text{GxT}$  base pair—but added the caveat that the  $20^\circ\text{C}$   $T_m$  difference translates to an energy value within the uncertainty of the experiment. Similar results were obtained for the duplex formed by  $d(\text{CGCAGAATTCGCG})$ , which contains two single unpaired adenosine residues (Patel *et al.*, 1982c).

Tibayenda *et al.* measured enthalpies for the mismatch containing variants of  $d(\text{TATTAATATCAAGTTG}):d(\text{CAACTTGATATTAATA})$ , in salt concentrations ranging from 0.015M to 1.0M, and observed that in low salt the enthalpies of the mismatch containing duplexes were similar to those of the parent, but that they became lower relative to the parent as ionic strength increased (Tibayenda *et al.*, 1984). Increasing divergence between calorimetric and Van t'Hoff enthalpies indicated decreasing

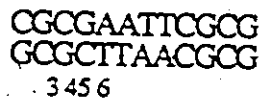
cooperativity of the transition, which was ascribed to the more effective screening of the backbone phosphates by counterions.

Aboul-ela *et al.*, in their study of mismatches in  $d(CA_3XA_3G):d(CT_3YT_3G)$  in 1.0M salt buffer, found that both  $\Delta G^\circ$  and  $\Delta S^\circ$  were dependent upon the nature of the mismatch.

### 3. Mismatches cause kinetic destabilization of adjacent base-pairs.

Patel *et al.*, and Haasnoot *et al.*, monitored the hydrogen-bonded imino protons around the mismatched sites of the variants of  $d(CGCGAATTCGCG)_2$  (Patel *et al.*, 1984c) and the TxT containing duplex  $d(ATCCATTAGGAT)_2$  (Haasnoot *et al.*, 1979, 1980) and observed a consistent increase of line-width in these imino protons relative to the parent duplexes, indicating greater lability towards opening and exchange.

Saturation recovery experiments on the variants of  $d(CGCGAATTCGCG)_2$ , in



which G-C 3 was replaced by GxT, GxA, AxC and TxC mispairs, showed that kinetic destabilization at G-C 4 accompanied introduction of all mispairs, while introduction of AxC and TxC also destabilized A-T 5 and 6. The TxC mismatch proved the most destabilizing (Patel *et al.*, 1984).

In the low temperature spectra of  $d(ATCCATATGGAT)_2$ , the hydrogen bonded imino protons disappear in the order A-T 1, A-T 2, G-C 3, followed by the central three together. In the spectra of  $d(ATCCATTATGGAT)_2$ , which contains a central TxT mismatch, A-T 1 and A-T 6 disappear simultaneously. The mismatch has introduced an additional melting site into the double helix (Haasnoot *et al.*, 1979).

### 4. Wobble-type pairing prevails for those mismatches capable of

base-pairing.

The definitive proof of hydrogen bonding between bases is the observation of the hydrogen-bonding protons themselves.  $r(\text{GxU})$  imino protons have been observed in the spectra of tRNA (Robillard *et al.*, 1976; Johnston and Redfield, 1978) and short synthetic oligoribonucleotide duplexes (Section 3.3.7.1), and  $d(\text{GxT})$  protons in  $d(\text{CGTGAATTCGCG})_2$  (Patel *et al.*, 1982) and  $d(\text{CAACTTGATATTAATA})$ :  $d(\text{TATTAATATTAAGTTG})$  (Tibayenda *et al.*, 1984). Two imino protons are seen, since both bases in the major tautomeric form have one, shifted upfield compared to their position in the Watson-Crick base pairs, since the acceptors are carbonyl oxygens instead of ring nitrogens. This is consistent with wobble base pair formation but not with tautomer pairing.

Crystal structures of tRNAs and of deoxyoligonucleotides duplexes containing GxT base pairs,  $d(\text{GGGGCTCC})_2$  (Brown *et al.*, 1985),  $d(\text{GGGGTCCC})_2$  and  $d(\text{TGCGCG})_2$  (Kennard, 1985), and  $d(\text{CGCGTG})_2$  (Ho *et al.*, 1985) all show wobble base pair formation, the latter two in Z-DNA duplexes.

A number of possible pairing schemes have been proposed for the GxA mismatch, involving syn conformations of both bases and tautomerization (Topal and Fresco, 1976). Intra base pair NOEs, particularly between the imino and purine nonexchangeable protons can be used to differentiate them, and the conformation apparently preferred in  $d(\text{CGAGAATTCGCG})_2$  (Patel *et al.*, 1984b, c) and  $d(\text{CCAAGATTGG})_2$  (Kan *et al.*, 1983) has both bases anti, and hydrogen bonding occurring between GNIH and AN1 and ANH<sub>2</sub> and GO6. One imino proton is observed.

This geometry is confirmed by the X-ray crystal structure of  $d(\text{CGCGAATTAGCG})_2$  (Kennard, 1985), which shows that the C1-C1 distance for this configuration is barely 1 Å larger than that for a Watson-Crick base pair, a difference

readily accommodated by the backbone.

Detection of CxA base pairing in  $d(\text{CGCGAATTCACG})_2$  (Patel *et al.* 1984a,c) is more problematic, particularly since in the absence of tautomerization of either base an imino proton will not be observed, which proves to be the case. On the basis of the absence of NOEs between the imino protons of G-C 2 and 4, and A3H2 (in contrast to the results for the duplex containing a GxA mismatch), and the very small shift of A3H2 upon duplexing, a structure is proposed whereby the adenosine residue slides to permit pairing of ANH<sub>2</sub> with CN3, pushing A3H2 into the minor groove.

5. "Non pairing" mismatches are accommodated within the duplex stackings.

The small changes in enthalpy upon introduction of a mismatch, have been interpreted to mean that the mismatch does not radically alter base-stacking. Tibayenda *et al.* estimated the penalties for formation of an interior loop with unstacking of bases, or bulging of both bases, and concluded that since the  $> 23$  kJ/mol reduction in  $\Delta G$  due to such conformations is so much larger than that observed (12 kJ/mol), the mismatched bases remain stacked to some degree (Tibayenda *et al.* 1984).

Pyrimidine-pyrimidine mismatches should, by their geometry, be non-pairing. Patel *et al.* have proposed a hydrogen-bonded structure for the CxT mismatch, based on their finding of a labile imino resonance at 11.72ppm in the low temperature spectra of  $d(\text{CGCGAATTCTCG})_2$ , which they assigned to TN3H, binding through a bridging water molecule to the acceptor groups on C. (Patel *et al.* 1984a,c).

Haasnoot *et al.* have also observed an imino proton signal for a pyrimidine-pyrimidine mismatch (Haasnoot *et al.* 1979): In the imino proton spectrum of  $d(\text{ATCCATTATCCAT})_2$ , the resonance at 10.5ppm is assigned to TN3H, shifted upfield from its single strand position in  $d(\text{TTTT})$  of 11.2ppm. This upfield shift is ascribed to the stacking in of the two thymidine residues.

Alkema *et al.* observe sigmoidal upfield movement of chemical shifts upon cooling of protons in AGACU, indicative of formation of a duplex with a central AxA mismatch (Alkema *et al.*, 1982). Upfield shifts were observed for the protons of the central mismatched A residue, suggesting that the As were stacked as opposed to being looped out. Results for AGCCU, which forms a duplex containing a central CxC mismatch, are less definitive: Although there appears to be partial stacking in, there may be several competing structures, with varying conformations around the mismatch (Alkema *et al.*, 1982).

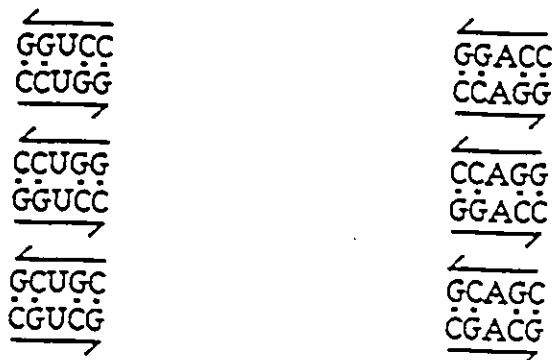
Three studies have been published on systems containing single unpaired bases, on d(CGCAGAATTCGCG) (Patel *et al.*, 1982b), which duplexes to give a double helix containing two unpaired adenosine residues, and on the mixtures of d(CA<sub>3</sub>CA<sub>3</sub>G) with d(CT<sub>6</sub>G) (Morden *et al.*, 1983) and d(CA<sub>6</sub>G) with d(CT<sub>3</sub>CT<sub>3</sub>G) (Aboul-ela *et al.*, 1985), both of which form a duplex with a central unpaired cytidine residue. Similar destabilization of the sequence containing the mismatch relative to the parent sequence is observed in these systems as in those containing base-base mismatches: The T<sub>m</sub> of d(CGCAGAATTCGCG)<sub>2</sub> (in 0.1 M salt) is 57°C, compared to 72°C for d(CGCGAATTCGCG)<sub>2</sub> (Patel *et al.*, 1982c) while the T<sub>m</sub>s of d(CA<sub>3</sub>CA<sub>3</sub>G):d(CT<sub>6</sub>G) and d(CA<sub>6</sub>G):d(CT<sub>6</sub>G) are 22° and 37°C respectively (at 0.4 mM, 1.0M salt) (Morden *et al.*, 1983; Aboul-ela *et al.*, 1985). The striking difference between the results of these two studies is in the conformations of the unstacked bases. Unpaired adenosine in d(CGCAGAATTCGCG)<sub>2</sub> is stacked into the duplex, according to the large upfield shifts observed for A3H8 and A3H2 upon duplexing (Patel *et al.*, 1982c). There is no decrease in the enthalpy of the helix-coil transition upon introduction of the mismatch, indicating no disruption of base-stacking. In contrast, the cytidine in



$d(CA_3CA_3G):d(CT_6G)$  is looped out (Morden *et al.*, 1985). An NOE can be seen between the imino protons of the A·T base pairs on either side, which indicates they are less than 4Å apart. Stacking in of the cytidine would cause them to be separated by 7Å. There is a 2.9kcal/mol destabilization of this duplex relative to the parent, which is ascribed to the loss of stacking interactions. According to these results, the behaviour of a single mismatched base depends upon its identity, and the nature of the neighbouring sequences.

#### 4.1.2. Studies on mismatch containing oligoribonucleotides derived from the GC tetramers

Duplexes containing mismatches were of interest both in their own right, and as suitable systems for the study of drug intercalation into imperfect duplexes. The three high-melting perfect GC tetramers, GGCC, CCGG and GCGC, seemed candidates for such studies, potentially sufficiently stable to support the introduction of internal mismatches. Studies of GxG and CxC mismatches in these systems were precluded by the likelihood of competing structures, but introduction of a central uridine or a central adenosine into the tetramer sequences gave a series of pentamers which might duplex around a UxU (GGUCC, CCUGG and GCUGC) or an AxA (GGACC, CCAGG and GCAGC) mismatch.



Mixing of the A-containing with the corresponding U-containing sequences would provide a series of pentameric duplexes in which the central GC core was interrupted by an A·U base pair, thereby allowing comparison of the effect that insertion of an A·U base pair would have on the stability of these duplexes.



In addition, the sequences GCAUGC and CCAUGG had already been synthesized but not studied, by Dr. D. Alkema, and as these sequences were derived from the duplexes GCGC and CCGG by the insertion of two A·U base pairs into the interior, they were included to extend the comparisons: GCGC, GCAGC:GCUGC and GCAUGC, and CCGG, CCAGG:CCUGG and CCAUGG.

## 4.2 Results

### 4.2.1 Sequences derived from parent GGCC

#### 4.2.1.1 Stability and secondary structure of GGUCC and GGACC

GGCC at 2.8 mM in 1.0M salt forms a four-base-pair perfect duplex with an average  $T_m$  of 52.1°C derived from the curves for eight of ten protons with an individual  $T_m$  range of 47.3 - 54.2°C (Figure 4.1; Table 4.2). At low temperature, in 90% water, the expected two imino proton signals are observed at 13.41 and 12.40 ppm, assigned to the internal and the terminal base pairs respectively. The downfield (internal) resonance persists to at least 30°C (Table 3.2).

GGUCC was first studied in 1.0M salt but extreme broadening of the signals renders the lower portion of the curves unassignable at the point at which they were beginning to promise sigmoidal behaviour. Repetition of the experiment with 0.1M salt buffer improved the resolution at low temperature, but points below 25°C quoted in Table 4.3 and plotted in Figure 4.2 must be considered tentative. The behaviour of the curves above 25°C, where firm assignment is possible, is characteristic of the upper portions of the sigmoidal curves, so the  $T_m$  of 24.8°C, averaged from the curves of seven protons which could be followed throughout their transition (individual  $T_m$ s 18.3 - 29.1°C), is reasonable, albeit less accurate than previously quoted  $T_m$ s.

Alkema *et al.* (1982) observed that AGGCU forms a duplex comprising two G-C and two G-U base pairs, with two 5' dangling As, in preference to a duplex with an internal GxG mismatch. In view of these results, a competing secondary structure may be written for GGUCC:

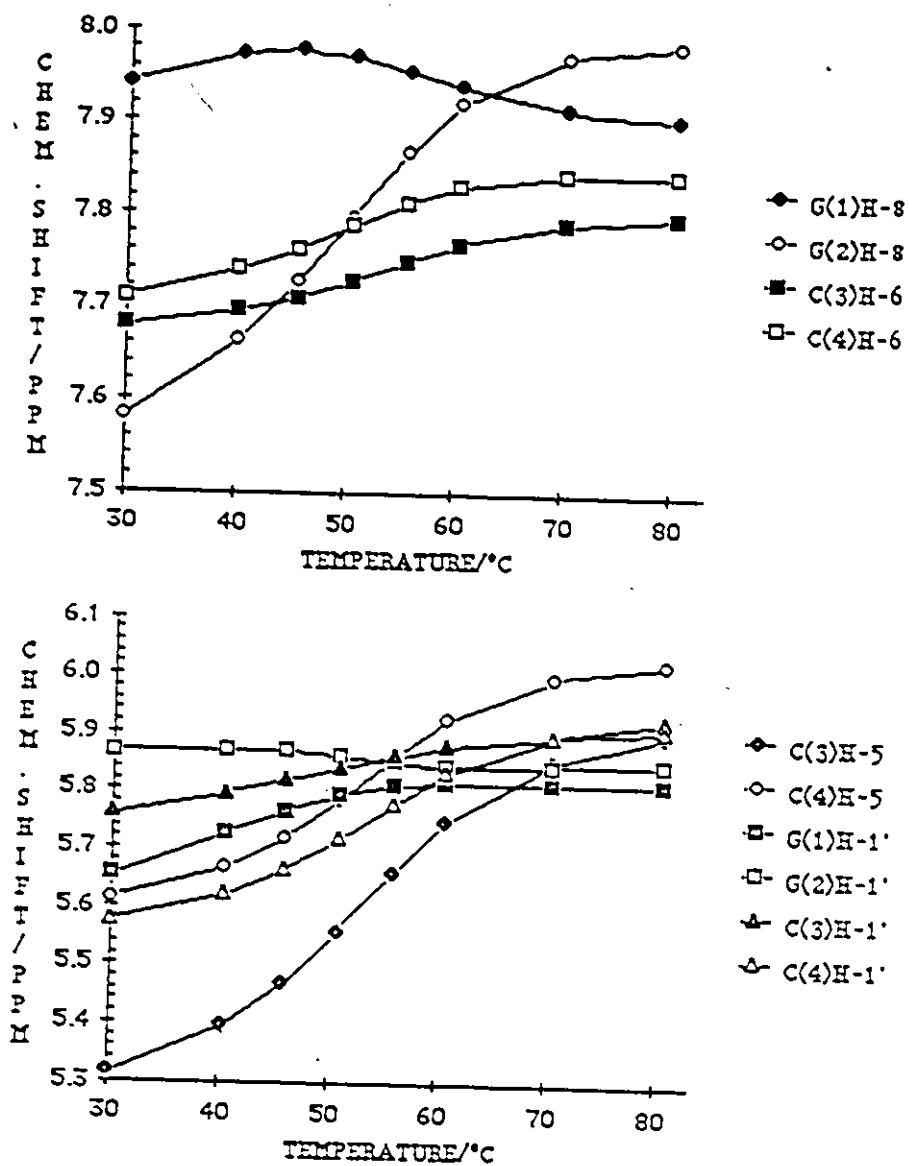


Figure 4.1: Chemical shift versus temperature curves for the base and anomeric protons of GGCC, 2.5 mM, in 0.1 M salt buffer.

Table 4.2: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GGCC, 2.5mM, in 0.1M salt buffer, over the temperature range 80 to 30°C.

Temp.	Chemical shift (ppm)								T <sub>m</sub>
	80.5°	70.3°	60.5°	55.8°	50.8°	45.8°	40.2°	30.1°	
Proton									
G(1)H-8	7.903	7.913	7.937	7.955	7.971	7.978	7.972	7.941	NSB
G(2)H-8	7.982	7.970	7.919	7.866	7.796	7.725	7.662	7.581	51.4
C(3)H-6	7.798	7.789	7.766	7.748	7.727	7.709	7.694	7.678	54.2
C(4)H-6	7.843	7.844	7.829	7.811	7.786	7.762	7.739	7.709	51.2
C(3)H-5	5.901	5.854	5.748	5.659	5.558	5.467	5.395	5.314	53.8
C(4)H-5	6.021	5.996	5.879	5.862	5.785	5.718	5.665	5.612	47.3
G(1)H-1'	5.816	5.816	5.815	5.809	5.792	5.763	5.727	5.652	NSB
G(2)H-1'	5.85	5.846	5.845	5.852	5.860	5.869	5.869	5.865	NSB
C(3)H-1'	5.923	5.896	5.879	5.862	5.840	5.817	5.795	5.760	50.7
C(4)H-1'	5.906	5.896	5.831	5.778	5.718	5.664	5.621	5.575	54.2
								Average T <sub>m</sub>	52.1
	Coupling constant (Hz)								
G(1)H-1'	5.1	4.8	4.4	4.1	3.4	2.4	1.5	--	
G(2)H-1'	4.5	3.6	3.6	2.4	1.7	--	--	--	
C(3)H-1'	4.0	3.7	3.6	2.6	1.7	--	--	--	
C(4)H-1'	4.2	3.7	3.4	2.9	2.5	1.9	1.3	--	

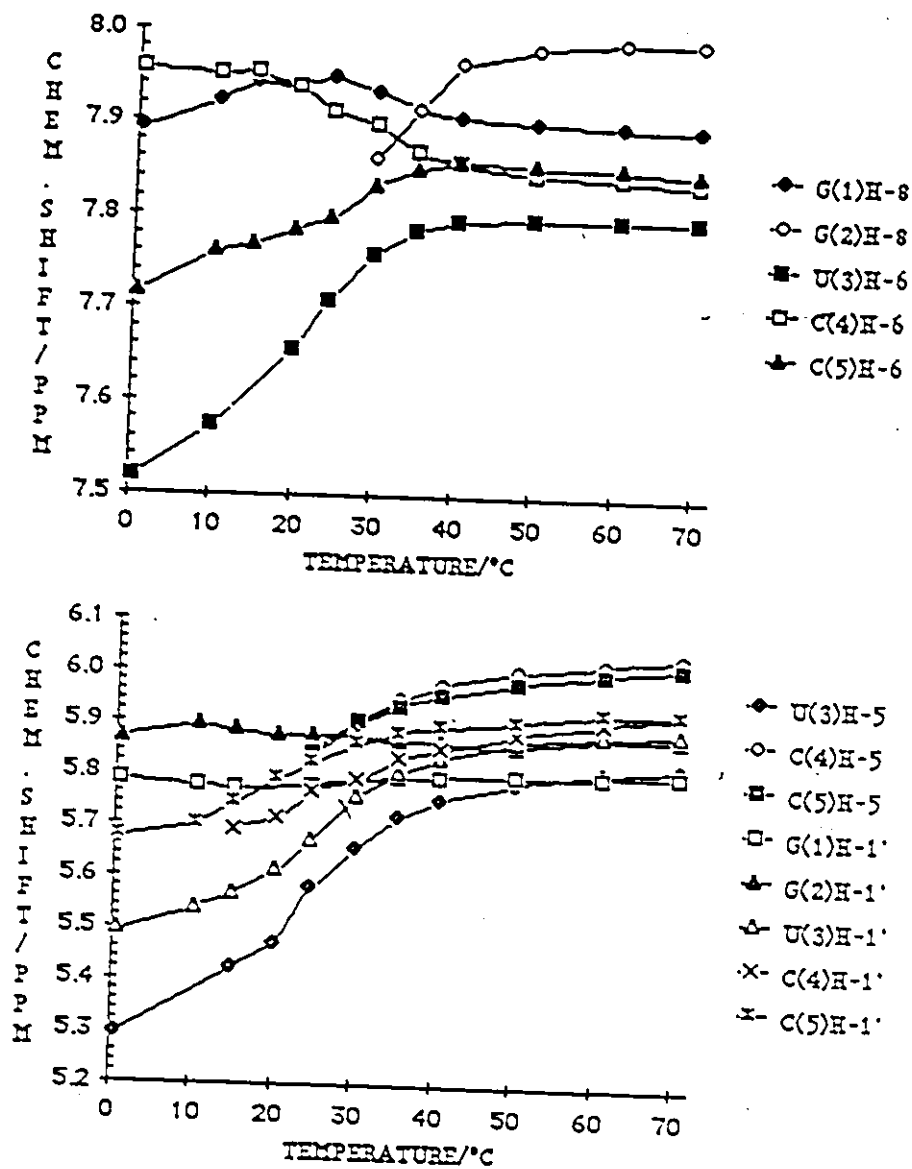
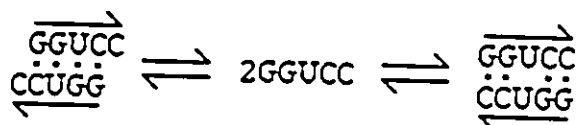


Figure 4.2: Chemical shift versus temperature curves for the base and anomeric protons of GGUCC, 2.5 mM, in 0.1 M salt buffer.

Table 4.3: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GGUCC, 2.5mM, in 0.1M salt buffer, over the temperature range 70 to 1°C.

Temp.	Chemical shift (ppm)											T <sub>m</sub>
	70.3°	60.7°	50.0°	40.6°	35.1°	29.9°	24.5°	20.3°	15.0°	10.3°	0.9°	
G(1)H-8	7.895	7.897	7.900	7.907	7.917	7.933	7.949	7.897	7.943	7.922	7.892	NSB
G(2)H-8	7.989	7.987	7.981	7.964	7.913	7.862	--	--	--	--	--	NSB
U(3)H-6	7.797	7.798	7.798	7.795	7.784	7.759	7.708	7.655	--	7.572	7.518	21.8
C(4)H-6	7.837	7.842	7.846	7.859	7.869	7.897	7.912	7.939	7.953	7.950	7.956	27.3
C(5)H-6	7.848	7.853	7.855	7.859	7.850	7.833	7.797	7.784	7.768	7.760	7.716	29.1
U(3)H-5	5.818	5.805	5.783	7.752	7.717	7.657	7.580	7.468	5.421	--	5.294	27.3
C(4)H-5	6.030	6.018	6.001	5.976	5.945	5.904	5.853	--	--	--	--	NSB
C(5)H-5	6.013	5.999	5.980	5.955	5.933	5.904	5.841	--	--	--	--	NSB
G(1)H-1'	5.803	5.800	5.798	5.795	5.792	5.786	5.776	--	5.770	5.775	5.785	NSB
G(2)H-1'	5.872	5.873	5.857	5.866	5.866	5.873	5.873	5.875	5.883	5.892	5.865	NSB
U(3)H-1'	5.884	5.879	5.865	5.833	5.803	5.756	5.669	5.610	5.563	5.537	5.494	27.3
C(4)H-1'	5.920	5.903	5.881	5.852	5.831	5.786	5.765	5.710	5.690	--	--	22.3
C(5)H-1'	5.920	5.920	5.904	5.891	5.880	5.861	5.822	5.787	5.739	5.697	5.672	18.3
											Average T <sub>m</sub>	24.8
Coupling constant (Hz)												
G(1)H-1'	5.4	5.5	5.0	4.5	4.7	4.5	--	--	--	--	--	--
G(2)H-1'	5.1	5.5	5.2	3.4	4.1	3.2	--	--	--	--	--	--
U(3)H-1'	5.7	5.9	4.6	5.7	4.5	4.8	--	--	--	--	--	--
C(4)H-1'	4.1	4.1	3.5	3.7	3.2	2.7	--	--	--	--	--	--
C(5)H-1'	4.1	3.7	3.5	3.5	3.2	2.7	--	--	--	--	--	--



The low temperature imino proton spectrum suggests contribution from at least two species. Five weak signals are observed at  $-4.5^{\circ}\text{C}$ . at 13.69, 12.46, 12.06 (small), 11.02 (broad) and 10.33 ppm. G•U imino proton signals are observable in spectra of tRNA between 11.7 and 10.4 ppm (Robillard *et al.*, 1976; Johnston and Redfield, 1978), and have been observed in spectra of AGGCU:AGUCU at 11.49 and 10.74 ppm (Section 3.3.7). The low temperature spectrum of AGGCU in isolation does not show any signals, G•U or otherwise. The duplex formed by GCUGC (Section 4.2.5) contains a central UxU opposition, and at  $2^{\circ}\text{C}$  in 90% water the UN3H appears at 10.25 ppm, but there are no other peaks upfield from the G•C imino region, 12.87 - 13.34 ppm. It seems likely that there is substantial contribution from the G•U containing slipped duplex, but it alone cannot account for the multiplicity of signals assignable to G•C base pairs. There should be a single G•C imino proton signal from this structure, and since it is flanked by a dangling 3' cytidine, and a stacked, internal 3' guanosine, it should appear upfield from 13.0 ppm. The experimental spectrum shows two signals, at 13.69 and 12.48 ppm. The G•C imino protons in the mismatch structure of GGUCC are predicted to appear at 12.9 (terminal) and 13.3 ppm (internal) (Kearns 1976), but if these values are adjusted according to the trends described in Section 3.3.7, and in GCUGC, where peaks predicted at 13.4 (terminal) and 13.3 ppm (internal), appear at 12.9 and 13.4 ppm respectively, then they should be closer to 12.5 (terminal) and 13.4 ppm (internal), which concurs well with the upfield signal, but less so with the the downfield signal. The signals, however, were broad and weak, and noise and errors in phasing could produce errors



in the assignments of their maxima.

From the rapid broadening of the spectra of the non-exchangeable protons, and the number of peaks observed in the imino proton spectrum, it would seem that there is an equilibrium between the two possible structures at low temperature.

The chemical shift versus temperature curves obtained for the protons of GGACC at 2.5mM, in 0.1M salt are shown in Figure 4.3, while the data itself appears in Table 4.4. On paper, GGACC can form a duplex with an internal A•A mismatch, flanked by two GG:CC cores, but it apparently does not do so in solution in the temperature range of the experiment. Although some of the protons show striking changes in chemical shift with temperature, especially those belonging to C4, the majority of curves are devoid of sigmoidal character (Figure 4.3). The H1' coupling constants gradually decrease, indicating stacking. The intensity and resolution begin to degrade rapidly below room temperature, leading to the incomplete curves shown. The imino proton spectrum at -4.5°C shows a single low, very broad peak around 11.5 ppm, too far upfield from the expected position for G•C protons, and possibly due to G-G aggregation.

The tetramer precursor GGAC, is expected to show no duplexing behaviour, and does not. Its chemical shift versus temperature curves (Figure 4.4; Table 4.5) are flat, or have a slight slope, this slope being most pronounced, again, for protons belonging to C4, although it is less than that observed in GGACC. The anomeric proton coupling constants steadily decrease, again indicating stacking. Although the resolution becomes poor for the last two or three spectra, it remains good at lower temperatures than in GGACC. No signals are observable in the imino proton spectrum at -4.5°C.

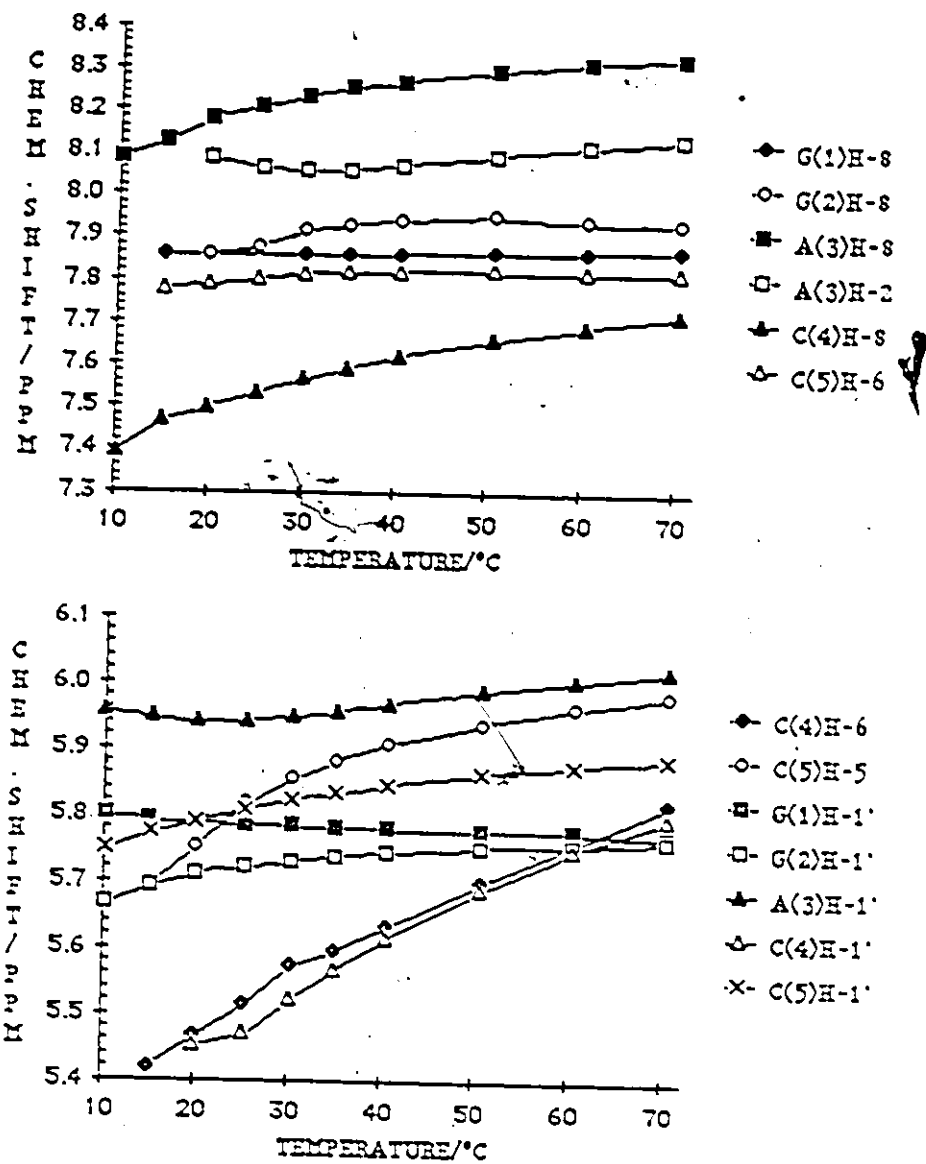


Figure 4.3: Chemical shift versus temperature curves for the base and anomeric protons of GGACC, 2.5 mM, in 0.1 M salt buffer.

Table 4.4: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GGACC, 2.5mM, in 0.1M salt buffer over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)									
	70.6°	60.6°	50.8°	40.6°	35.0°	30.3°	25.3°	20.0°	15.1°	10.3°
Proton										
G(1)H-8	7.865	7.864	7.862	7.857	7.855	7.854	7.855	--	7.856	--
G(2)H-8	7.934	7.938	7.949	7.936	7.927	7.912	7.876	7.858	--	--
A(3)H-8	8.321	8.310	8.292	8.266	8.249	8.229	8.204	8.177	8.121	8.081
A(3)H-2	8.128	8.110	8.089	8.066	8.056	8.052	8.058	8.085	--	--
C(4)H-6	7.717	7.689	7.656	7.616	7.591	7.565	7.532	7.498	7.464	7.390
C(5)H-6	7.818	7.818	7.820	7.817	7.814	7.809	7.800	7.786	7.776	--
C(4)H-5	5.818	5.757	5.697	5.630	5.595	5.572	5.514	5.465	5.418	--
C(5)H-5	5.979	5.960	5.937	5.904	5.881	5.853	5.815	5.749	5.692	--
G(1)H-1'	5.770	5.775	5.775	5.778	5.780	5.782	5.783	5.789	5.791	5.800
G(2)H-1'	5.762	5.753	5.751	5.741	5.735	5.729	5.721	5.711	5.692	5.665
A(3)H-1'	6.018	6.003	5.986	5.965	5.955	5.947	5.940	5.939	5.945	5.953
C(4)H-1'	5.793	5.748	5.687	5.613	5.565	5.522	5.469	5.449	--	--
C(5)H-1'	5.882	5.871	5.860	5.843	5.832	5.821	5.807	5.789	5.771	5.745
	Coupling constant (Hz)									
G(1)H-1'	5.6	3.9	5.6	4.6	4.2	3.8	4.0	2.0		
G(2)H-1'	5.4	4.2	5.6	5.0	4.7	4.5	4.3	4.3		
A(3)H-1'	4.0	3.7	3.2	2.8	2.5	2.2	--	--		
C(4)H-1'	4.2	2.6	2.6	2.4	2.3	--	--	--		
C(5)H-1'	3.9	3.8	3.6	3.6	3.2	3.0	3.0	2.0		

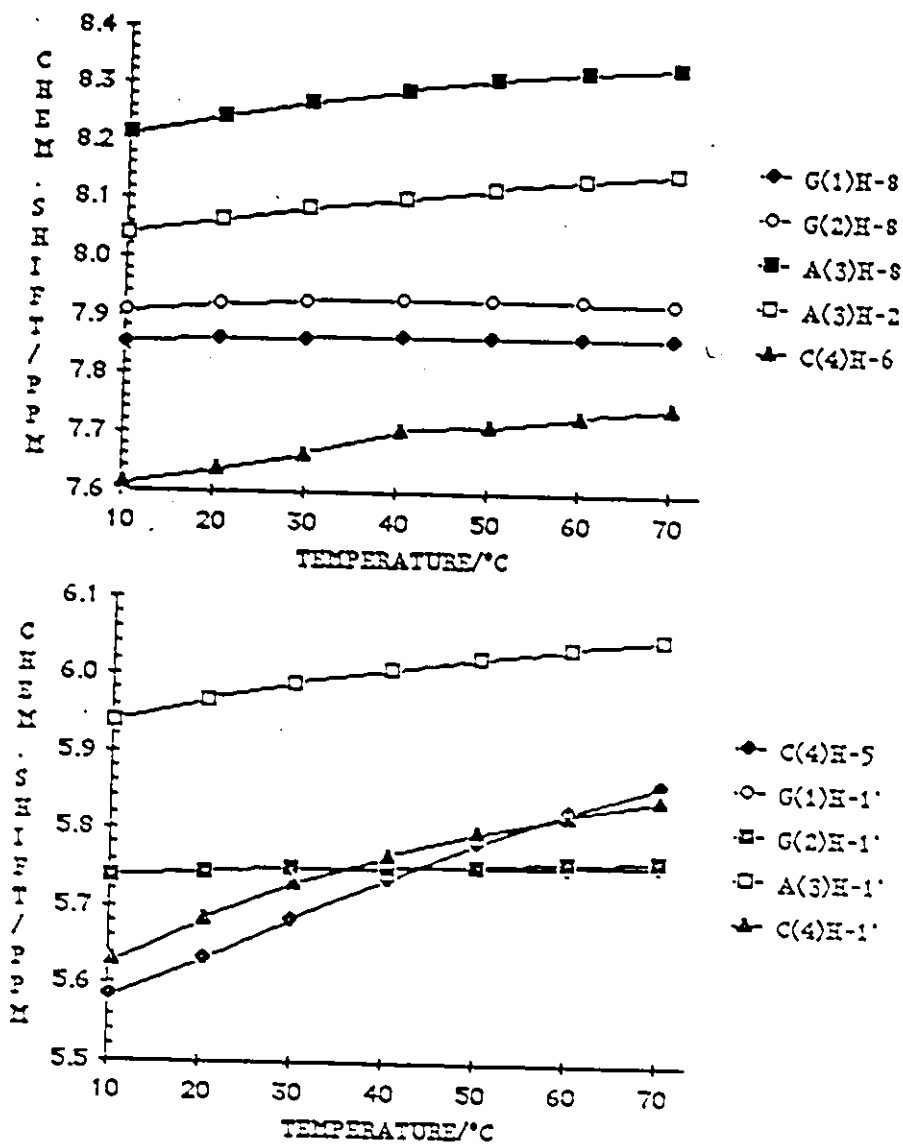


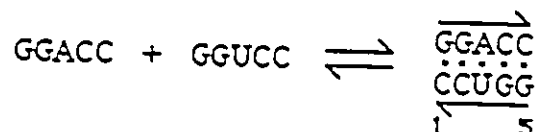
Figure 4.4: Chemical shift versus temperature curves for the base and anomeric protons of GGAC, 1.9 mM, in 0.1 M salt buffer.

Table 4.5: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GGAC, 1.7mM, in 0.1M salt buffer, over the temperature range 70 to 10°C.

	Chemical shift (ppm)						
	70.3°	60.3°	50.3°	40.6°	30.0°	20.5°	10.3°
G(1)H-8	7.865	7.865	7.864	7.864	7.862	7.859	7.853
G(2)H-8	7.925	7.927	7.928	7.928	7.925	7.920	7.907
A(3)H-8	8.33	8.321	8.308	8.290	8.268	8.242	8.210
A(3)H-2	8.150	8.135	8.119	8.102	8.083	8.062	8.037
C(4)H-6	7.745	7.730	7.711	7.704	7.663	7.636	7.609
C(4)H-5	5.862	5.826	5.784	5.736	5.684	5.632	5.583
G(1)H-1'	5.755	5.752	5.752	5.750	5.729	5.744	5.738
G(2)H-1'	5.763	7.758	5.753	5.750	5.729	5.744	5.738
A(3)H-1'	6.047	6.035	6.022	6.006	5.987	5.965	5.938
C(4)H-1'	5.842	5.820	5.796	5.766	5.729	5.684	5.629
	Coupling constant (Hz)						
G(1)H-1'	5.6	5.3	4.6	5.1	5.1	4.2	3.7
G(2)H-1'	5.4	5.4	5.3	5.1	5.1	4.2	3.7
A(3)H-1'	4.1	4.2	3.8	3.4	2.9	2.5	—
C(4)H-1'	3.0	4.5	3.3	3.3	3.3	2.7	—

#### 4.2.1.2 Effect of insertion of A·U base pairs, GGACC:GGUCC

A 1:1 mixture of GGACC with GGUCC can form a complementary duplex comprising two GG:CC cores separated by a single A·U base pair:



At 2.5mM concentration of each strand, the chemical shift versus temperature curves of nine of eleven base protons are sigmoidal (Figure 4.5) having  $T_m$ s ranging from 45.4 - 54.7°C (Table 4.6), with an average  $T_m$  of 51.8°C. The imino proton spectra were obtained for a sample at 1.5mM, since the sample was originally run in a microtube to achieve the desired concentration, and had to be diluted to improve the signal to noise for the imino protons. Previous experience suggested that this dilution would only reduce the  $T_m$  by 2 - 3°C, and should not affect the imino proton positions at low temperature. Four imino proton signals were observed at -5.0°C (Table 4.7) 14.49, 13.41, 12.63 and 12.29 ppm; the areas are approximately 1:1:1:2. The peak at 12.29 ppm loses intensity first and is gone by 15°C, while the others persist to above 25°C. From this, and from comparison with the parent duplex GGCC (terminal G·C 12.32 ppm), the signals at 12.29 ppm may be assigned to the terminal G·C base pairs. The other three signals can be assigned as follows—A·U 3, which is an A·U base pair flanked by pyrimidines, 14.49 ppm; G·C 2, 12.63 ppm and GC 4, 13.41 ppm, since G·C 2 has a 3' neighbouring adenosine, and is therefore expected to be at higher field than G·C 4, with 3' neighbouring pyrimidines.

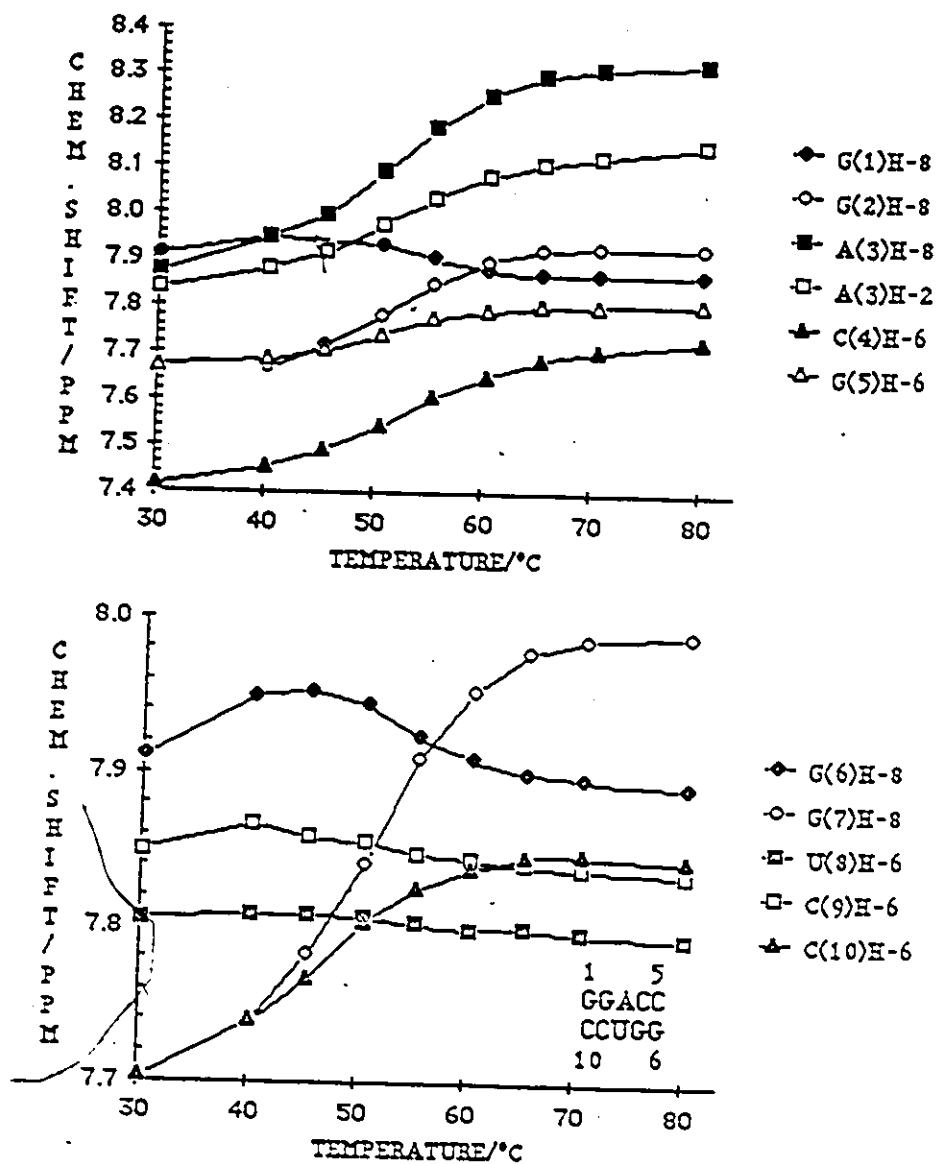


Figure 45: Chemical shift versus temperature curves for the base protons of GGACC:GGUCC, 2.5 mM each strand, in 0.1 M salt buffer.

Table 4.6: Chemical shifts for the low field aromatic protons of GGACC:GGUCC, 2.5mM each strand, in 0.1M salt buffer over the temperature range 80 - 30°C.

Numbering scheme:  $\begin{array}{cc} 1 & 5 \\ \text{GGACC} & \\ \text{CCUGG} & \\ 10 & 6 \end{array}$

Proton	Chemical shift (ppm)									T <sub>m</sub>
	80.1°	70.6°	65.5°	60.5°	55.5°	50.7°	45.5°	40.7°	30.9°	
G(1)H-8	7.866	7.866	7.870	7.880	7.903	7.933	7.951	7.949	7.910	54.7
G(2)H-6	7.929	7.929	7.922	7.896	7.946	7.780	7.717	7.665	-	51.5
A(3)H-8	8.326	8.313	8.296	8.257	8.187	8.089	8.996	8.949	8.875	52.7
A(3)H-2	8.149	8.122	8.106	8.078	8.031	7.976	7.916	7.880	7.835	52.2
C(4)H-6	7.726	7.702	7.681	7.648	7.603	7.538	7.487	7.451	7.413	53.1
C(5)H-6	7.805	7.802	7.799	7.788	7.771	7.738	7.706	7.681	7.666	49.9
G(6)H-8	7.892	7.896	7.900	7.909	7.923	7.943	7.951	7.949	7.910	55.0
G(7)H-8	7.988	7.985	7.977	7.952	7.909	7.840	7.782	7.739	7.702	51.8
U(8)H-6	7.793	7.796	7.799	7.798	7.802	7.805	7.808	7.807	7.804	NSB
C(9)H-6	7.834	7.838	7.840	7.843	7.847	7.855	7.858	7.866	7.849	NSB
C(10)H-6	7.844	7.847	7.846	7.838	7.824	7.803	7.766	7.739	7.702	45.4
Average T <sub>m</sub>										51.8



Table 4.7: Chemical shifts of hydrogen-bonded imino protons of GGACC:GGUCC, 1.5 mM each strand, in 0.1M salt buffer, over the temperature range -5 to 25°C.

Numbering scheme:  $\begin{array}{cc} 1 & 5 \\ \text{GGACC} \\ \text{CCUGG} \end{array}$

Temp.	Chemical shift (ppm)			
	<u>-5.0°</u>	<u>5.7°</u>	<u>15.0°</u>	<u>25.0°</u>
Proton				
G•C1	12.29	12.36	(12.4)	—
G•C2	12.63	12.60	12.51	12.50
A•U3	14.49	14.46	14.41	14.36
G•C4	13.41	13.43	13.42	13.38
G•C5	12.29	12.36	(12.4)	—

The average  $T_m$  calculated for the duplex GGACC:GGUCC from base protons only is 51.8°C, compared to an average of 52.3°C for the base protons of the parent duplex GGCC. This difference of 0.5°C is within experimental error; the two duplexes are of similar stability.

#### 4.2.2. Sequences derived from parent CCGG

##### 4.2.2.1. Stability and secondary structure of CCUGG and CCAGG

CCGG, at 2.5mM in 0.1M salt, has a measured  $T_m$  of 44.1°C, calculated from eight sigmoidal curves with a  $T_m$  range of 37.8 - 47.2°C (Table 4.8; Figure 4.6). As expected, two imino proton signals are observed, the terminal one at 13.34 ppm, and the internal one at 12.53 ppm, assigned by the order of broadening (Table 3.3; Figure 3.5); these values, unlike those for other sequences, agree with those predicted from ring current calculations and classical geometry (see Section 3.3.7).

The spectra of both CCUGG and CCAGG broaden appreciably below room temperature, and therefore large uncertainties attend the assignments of the base protons. The H-1' protons can be best followed through their transition since their singlet signals (following collapse) remain of good intensity, and resolved from each other. Sigmoidal curves can be observed for the majority of protons for both sequences (Figures 4.7 and 4.8).

The curves for the protons of CCUGG (seven protons) have individual  $T_m$ s ranging from 12.1 to 19.5°C, with an average of 17.3 (Figure 4.7; Table 4.9). The exchangeable proton spectrum of CCUGG at -4.6°C shows one strong signal at 13.18 ppm, with perhaps two broad and much weaker signals at 11.2 and 10.2 ppm.

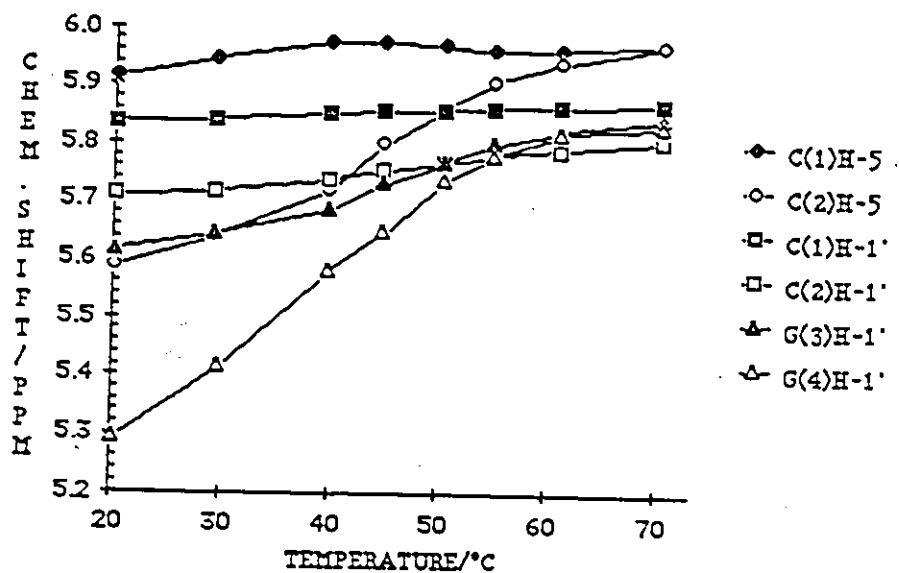
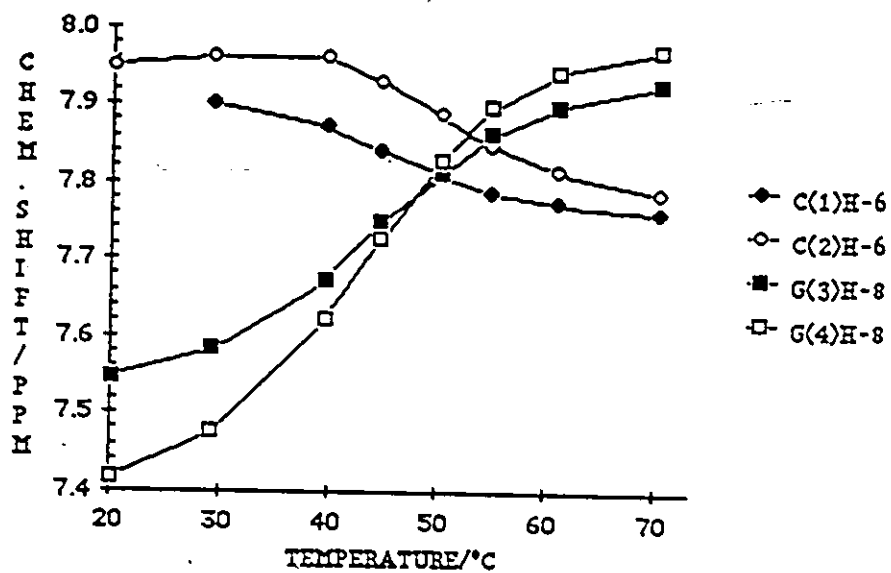


Figure 4.6: Chemical shift versus temperature curves for the base and aromatic protons of CCGG, 2.5 mM, in 0.1 M salt buffer.

Table 4.8: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CCGG, 2.5mM, in 0.1M salt buffer, over the temperature range 70 - 20°C

Temp.	Chemical shift (ppm)								T <sub>m</sub>
	70.3°	61.1°	55.1°	50.4°	44.9°	39.8°	29.5°	20.3°	
Proton									
C(1)H-6	7.762	7.772	7.787	7.811	7.840	7.871	7.900	--	45.8
C(2)H-6	7.786	7.813	7.846	7.886	7.930	7.962	7.960	7.948	47.2
G(3)H-8	7.920	7.898	7.863	7.812	7.748	7.673	7.583	7.546	44.7
G(4)H-8	7.969	7.943	7.896	7.826	7.724	7.621	7.475	7.416	44.2
C(1)H-5	5.967	5.960	5.961	5.968	5.974	5.971	5.944	5.912	NSB
C(2)H-5	5.967	5.937	5.906	5.860	5.799	5.718	5.639	5.586	44.5
C(1)H-1'	5.867	5.861	5.858	5.856	5.853	5.848	5.839	5.835	NSB
C(2)H-1'	5.802	5.789	5.780	5.767	5.752	5.736	5.716	5.709	45.5
G(3)H-1'	5.842	5.821	5.798	5.767	5.732	5.683	5.641	5.614	45.5
G(4)H-1'	5.827	5.815	5.780	5.736	5.674	5.580	5.413	5.292	37.8
									Average T <sub>m</sub> 45.3
	Coupling constant (Hz)								
C(1)H-1'	5.3	5.2	5.0	4.5	3.8	2.3	1.9	--	
C(2)H-1'	5.8	4.6	4.5	3.5	2.5	1.6	--	--	
G(3)H-1'	6.1	5.3	4.5	2.2	--	--	--	--	
G(4)H-1'	6.7	4.0	4.8	3.5	2.1	--	--	--	

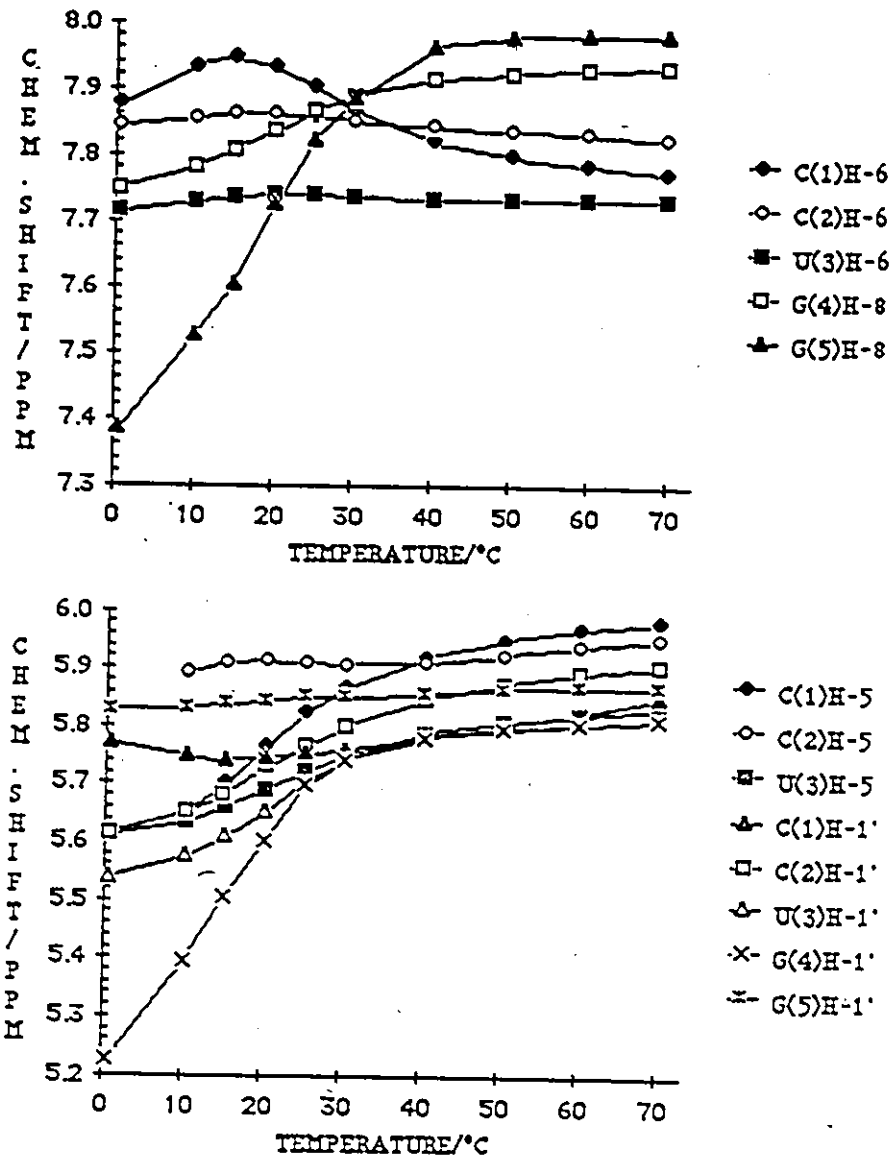
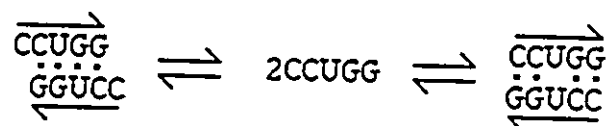


Figure 4.7: Chemical shift versus temperature curves of the base and anomeric protons of CCUGG, 2.5 mM, in 0.1 M salt buffer.

Table 4.9: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CCUGG, 2.5mM, in 0.1M salt buffer, over the temperature range 70 - 0°C.

Temp.	Chemical shift (ppm)										T <sub>m</sub>
	70.0°	60.1°	50.4°	40.4°	30.3°	25.3°	20.4°	15.3°	10.3°	0.9°	
Proton											
C(1)H-6	7.771	7.784	7.798	7.82	7.866	7.903	7.932	7.946	7.933	7.876	30.5
C(2)H-6	7.825	7.831	7.836	7.843	7.850	7.859	7.861	7.862	7.753	7.841	NSB
U(3)H-6	7.730	7.730	7.730	7.731	7.736	7.740	7.740	7.736	7.728	7.712	NSB
G(4)H-8	7.931	7.927	7.922	7.912	7.887	7.865	7.834	7.805	7.78	7.748	17.2
G(5)H-8	7.797	7.980	7.977	7.963	7.882	7.821	7.724	7.603	7.523	7.748	15.5
C(1)H-5	5.983	5.967	5.947	5.919	5.867	5.824	5.766	5.702	5.649	5.612	16.7
C(2)H-5	5.950	5.937	5.923	5.909	5.905	5.910	5.913	5.908	5.891	-	NSB
U(3)H-5	5.834	5.821	5.806	5.785	5.750	5.723	5.690	5.661	5.634	5.612	19.5
C(1)H-1'	5.906	5.891	5.871	5.846	5.801	5.766	5.730	5.680	5.649	5.612	18.1
C(2)H-1'	5.848	5.819	5.799	5.779	5.758	5.752	5.746	5.742	5.747	5.770	NSB
U(3)H-1'	5.813	5.803	5.796	5.785	5.742	5.705	5.651	5.608	5.573	5.538	19.5
G(4)H-1'	5.813	5.803	5.793	5.779	5.743	5.697	5.602	5.501	5.392	5.225	12.1
G(5)H-1'	5.868	5.865	5.861	5.856	5.850	5.848	5.842	5.837	5.829	5.825	NSB
	Coupling constants (Hz)										
C(1)H-1'	4.5	4.4	4.1	3.8	3.1	2.7	4.4	-	-	-	
C(2)H-1'	4.7	3.2	4.5	4.9	4.1	4.4	3.2	-	-	-	
U(3)H-1'	5.2	4.8	6.0	4.9	4.4	3.8	-	-	-	-	
G(4)H-1'	5.2	4.8	4.5	4.9	4.4	5.0	-	-	-	-	
G(5)H-1'	5.4	5.4	5.4	5.2	4.9	4.3	-	2.3	-	-	

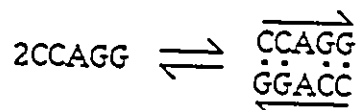
Competing secondary structures similar to those for GGUCC (Section 4.2.1.) may be written for this duplex. CCUGG may form a duplex containing a central UxU mismatch, or one containing G•U base pairs, with two 5' dangling Cs.



For the slipped structure (left), one G•C imino proton, at around 13.4 ppm, and two GxU protons, between 11.5 and 10.7 ppm, would be expected, while in the mismatch-containing structure (right) two G•C protons, at 13.4 and 12.8 ppm, and one UNH3 proton at around 10.2 ppm would be expected. Neither pattern is observed in full at -5°C; the duplexes are too weak.

The individual proton Tms of CCAGG span the range 11.9 - 24.6°C (eleven of thirteen protons), with an average of 17.3°C (Figure 4.8; Table 4.10). The scatter of Tms is suggestive of the uncertainties associated with the lower temperature signals. Imino proton signals are observed at 13.38 and 12.22 ppm, with a very low, broad signal at around 11 ppm:

CCAGG can form a four base pair duplex comprising two GG:CC core separated by an AxA opposition.



No competing simple secondary structures can be written for this duplex, although the deteriorating low temperature resolution suggests aggregation is

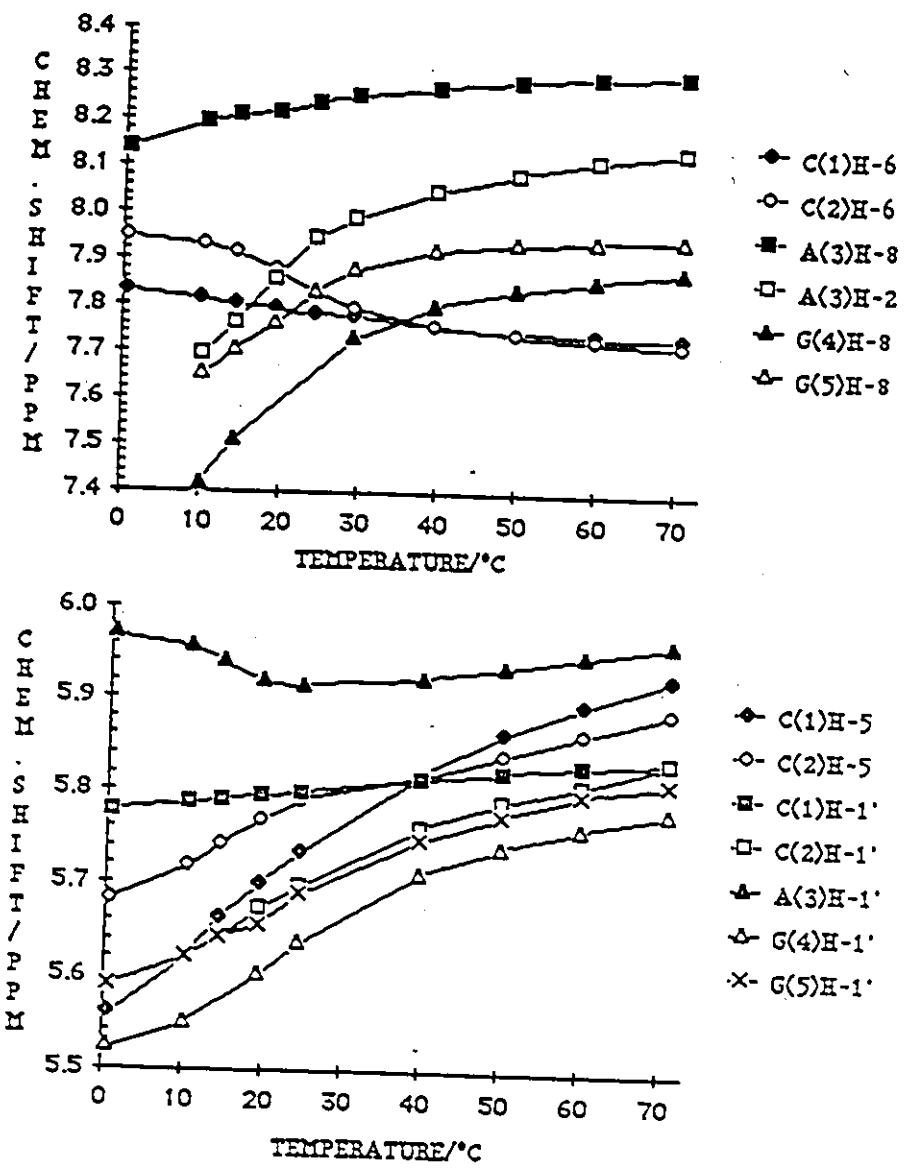


Figure 4.8: Chemical shift versus temperature curves of the base and anomeric protons of CCAGG, 2.5 mM, in 0.1 M salt buffer.



Table 4.10: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CCAGG, 2.5mM, in 0.1M salt buffer, over the temperature range 70 - 0°C.

	Chemical shift (ppm)										T <sub>m</sub>
	71.0°	60.0°	50.0°	39.7°	29.5°	24.4°	19.4°	14.4°	10.3°	1°	
C(1)H-6	7.736	7.740	7.746	7.756	7.774	7.779	7.793	7.805	7.814	7.831	NSB
C(2)H-6	7.722	7.729	7.739	7.756	7.796	7.833	7.879	7.917	7.934	7.950	20.4
A(3)H-8	8.304	8.296	8.287	8.273	8.253	8.241	8.219	8.212	8.195	8.183	18.8
A(3)H-2	8.137	8.155	8.088	8.051	7.992	7.946	7.860	7.764	7.696	—	15.9
G(4)H-8	7.878	7.860	7.837	7.802	7.730	7.665	—	7.509	7.415	—	NSB
G(5)H-8	7.949	7.945	7.938	7.922	7.890	7.833	7.762	7.704	7.649	—	17.4
C(1)H-5	5.926	5.896	5.863	5.819	—	5.734	5.700	5.664	5.621	5.559	13.6
C(2)H-5	5.888	5.864	5.840	5.816	—	5.790	5.768	5.742	5.718	5.682	11.6
C(1)H-1'	5.836	5.829	5.822	5.813	—	5.798	5.794	5.790	5.786	5.777	NSB
C(2)H-1'	5.836	5.809	5.790	5.761	—	5.697	5.673	5.641	5.621	5.590	18.8
A(3)H-1'	5.962	5.948	5.934	5.922	—	5.915	5.920	5.940	5.957	5.970	11.9
G(4)H-1'	5.779	5.761	5.740	5.771	—	5.637	5.602	—	5.550	5.523	20.5
G(5)H-1'	5.810	5.797	5.775	5.747	—	5.688	5.654	5.641	5.621	5.590	24.6
											Average T <sub>m</sub> 17.3
	Coupling constants (Hz)										
C(1)H-1'	5.3	5.4	5.5	5.5	—	5.1	4.9	4.0			
C(2)H-1'	5.3	5.4	5.2	4.1	—	2.7	5.3	—			
A(3)H-1'	5.3	5.1	4.8	—	—	2.7	—	—			
G(4)H-1'	5.3	5.3	5.2	4.9	—	—	—	—			
G(5)H-1'	4.0	5.5	4.4	3.0	—	1.7	—	—			

occurring. The aggregation may also be responsible for the detection of the imino protons belonging to this weak duplex; the broad resonance at around 11 ppm could be assigned to the G-G base pairs which are forming. The G-C imino protons do, however, conform to trends expected in the series that CCAGG forms with related sequences CCGG, CCAGG:CCUGG and CCAUGG, all of which are discrete, high-melting duplexes. The terminal G-C protons appear around 13.3- 13.4 ppm for all sequences, while the internal G-C protons, which appear at 12.53 ppm in the CCGG duplex, shift upfield to 12.22 ppm in the CCAGG spectrum, when a G-C base pair adjacent to the imino proton is replaced by an AxA opposition, and downfield to 12.34 ppm in both CCAGG:CCUGG and CCAUGG, when the AxA opposition is itself replaced by an A-U base pair. The expected order of effectiveness of shielding would be AxA > A-U > G-C, which is as observed.

#### 4.2.2.2. Conformation around the AxA mismatch

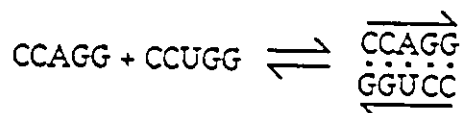
The broadening of the nonexchangeable proton signals obscures the H-1' coupling constants before their collapse, preventing any deduction of order of base stacking from the sugar conformation at the mismatch site. Both the A3H-8 and the A3H-2 protons move upfield, the former by 0.1 ppm, and the latter by > 0.4 ppm (Figure 4.8), which suggests some degree of stacking into the duplex, although the upfield shifts are by no means as pronounced as those in the related duplex CCAGG:CCUGG (Figure 4.9), in which the A residue is involved in an A-U base pair, and therefore fully stacked. A3H-8 in CCAGG:CCUGG moves upfield by 0.4 ppm, and A3H-2 by 1.1 ppm. The best evidence for stacking of the adenosine residues is provided by the position of the imino proton of the G-C base pair adjacent to the mismatch, which is 0.3 ppm upfield from its position in CCGG (G-C neighbour), and 0.12 ppm upfield from its

position in CCAGG:CCUGG and CCAUGG (A•U neighbour). Not only is this imino proton experiencing the effect of the 3' neighbouring adenosine, it is also experiencing some shielding due to the 5' neighbouring adenosine—both residues are stacking in to the duplex.

The above is in agreement with published results. Alkema *et al.* concluded that the central adenosine residues in the duplex formed by AGACU duplex were stacking in to the duplex (Alkema *et al.*, 1982), and the calculations by Tibayenda *et al.* for their deoxyoligonucleotide system show that the formation of two bulged bases reduces the  $\Delta G$  of the duplex by  $> 29$  kJ/mol, compared to  $< 12$  kJ/mol when both mismatched bases are stacked (Tibayenda *et al.*, 1984).

#### 4.2.3. Effect of insertion of A•U base pairs. CCAGG:CCUGG and CCAUGG

A 1:1 mixture of CCAGG and CCUGG leads to formation of a duplex comprising two GG:CC cores separated by an A•U base pair.



Poor signal to noise (the sample was run in a microtube to achieve the desired concentration), and extensive signal overlap in the anomeric proton region restricted analysis to the low field aromatic proton signals; of these, nine of eleven showed sigmoidal behaviour (Figure 4.9), with  $T_{ms}$  ranging from 49.2 - 58.1°C (Table 4.11), and an average of 52.7°C, at 2.5mM of each strand, in 0.1M salt. At -4°C, the five expected imino proton signals are observed as three signals, one, representing three

Table 4.11: Chemical shifts for the base protons of CCAGG:CCUGG, 2.5mM each strand, in 0.1M salt buffer over the temperature range 70 - 20°C.

Numbering scheme:

1	5
CCAGG	
GGUCC	
10	6

Temp.	Chemical shift (ppm)									T <sub>m</sub>
	<u>70.5°</u>	<u>66.1°</u>	<u>60.2°</u>	<u>55.0°</u>	<u>50.5°</u>	<u>44.9°</u>	<u>39.5°</u>	<u>30.2°</u>	<u>20.3°</u>	
Proton										
C(1)H-6	7.727	7.733	7.748	7.765	7.771	7.775	7.762	7.779	7.779	NSB
C(2)H-6	7.716	7.724	7.48	7.772	7.800	-	7.867	7.889	7.896	49.7
A(3)H-8	8.300	8.292	8.276	8.249	8.205	8.164	8.115	8.04	-	49.6
A(3)H-2	8.128	8.101	8.037	7.890	7.691	-	7.332	7.170	7.078	49.7
G(4)H-8	7.872	7.851	7.801	7.701	7.600	-	7.444	7.360	7.310	54.4
G(5)H-8	7.940	7.928	7.920	7.862	7.799	7.756	7.724	7.663	7.614	54.4
C(6)H-6	7.773	7.785	7.816	7.848	7.885	-	7.941	7.979	7.983	57.0
C(7)H-6	7.826	7.834	7.877	7.901	7.941	-	8.008	8.028	-	58.1
U(8)H-6	7.731	7.733	7.743	7.740	7.739	7.739	7.740	7.744	7.746	NSB
G(9)H-8	7.929	7.921	7.901	7.862	7.829	7.782	7.756	7.727	-	51.6
G(10)H-8	7.972	7.960	7.932	7.890	7.838	7.775	7.735	7.688	7.648	49.6
										Average T <sub>m</sub> 52.7

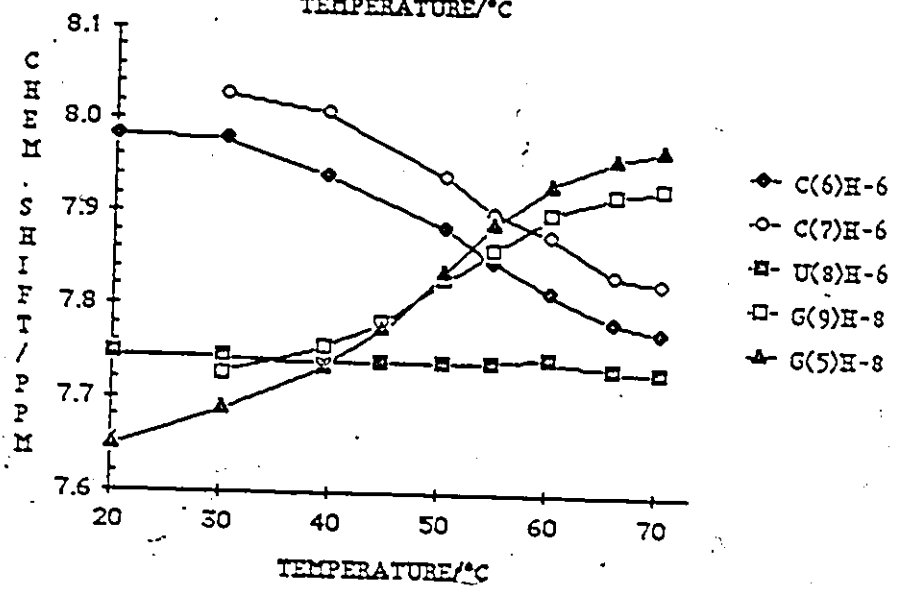
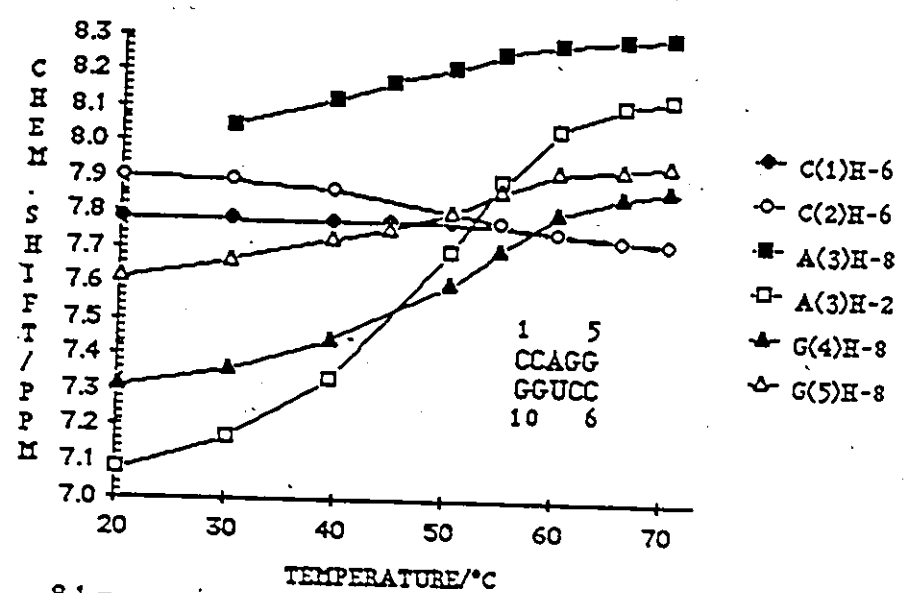
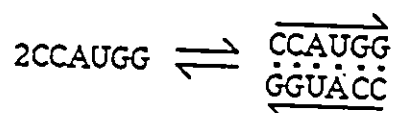


Figure 4.9: Chemical shift versus temperature curves for the base protons of CCAGG:CCUGG, 2.5 mM each strand, in 0.1 M salt buffer.

protons, at 13.31 ppm, a shoulder on that peak belonging to a single proton at 13.12 ppm, and a singlet at 12.34 ppm (Table 4.12). By 25.1°C the signals appear at 13.26, 13.1 and 12.32 ppm respectively, with areas 1:1:1; the terminal protons are thereby assigned to the peak at 13.31 ppm. The most upfield signal can be assigned to G-C2, which has a 3' neighbouring adenosine and a 3' neighbouring guanosine, and the remaining two signals are tentatively assigned as A-U3 13.31 ppm and G-C4 13.12 ppm. CCAUGG self-associates to form a perfect, six base pair duplex in which two GG:CC cores are separated by two A-U base pairs (an AU:AU core).



Due to sample limitations, the final concentration of this sequence was 1.5mM. Sigmoidal curves were obtained for fifteen of sixteen protons, over a range of 43.7 - 56.4°C, with an average  $T_m$  of 49.4°C (Figure 4.10; Table 4.13). The  $T_m$  of the seven low field aromatic protons alone is 48.0°C. Three imino proton signals would be expected in the spectrum of this self-complementary hexamer in water. Two are observed, at 13.31 and 12.36 ppm, the former being double the area of the latter (Table 4.14). By 25.1°C the two signals are of equal intensity, allowing assignment of the terminal G-C proton signals to 13.31 ppm. The internal G-C protons have two pyrimidine 3' neighbours, whereas the A-U protons have only one, and occur intrinsically at lower field, therefore the 12.36 ppm signal is assigned to G-C2 and the other 13.31 ppm signal to A-U3.

The four low field aromatic protons of CCGG have an average  $T_m$  of 45.3°C,

Table 4.12: Chemical shifts of the hydrogen-bonded imino protons of CCAGG:CCUGG, 2.5mM each strand, in 1.0M salt buffer, over the temperature range -4° to 30°C.

Numbering scheme:       1     5  
                              CCAGG  
                              GGUCC

Temperature	Chemical shift (ppm)				
	<u>-4.0</u>	<u>5.7</u>	<u>14.8</u>	<u>25.1</u>	<u>29.5</u>
Proton					
G•C1	13.31	13.29	13.27	--	--
G•C2	12.34	12.35	12.33	12.32	12.32
A•U3	13.12	13.15	13.11	13.10	13.13
G•C4	13.31	13.29	13.27	13.26	13.32
G•C5	13.31	13.29	13.27	--	--

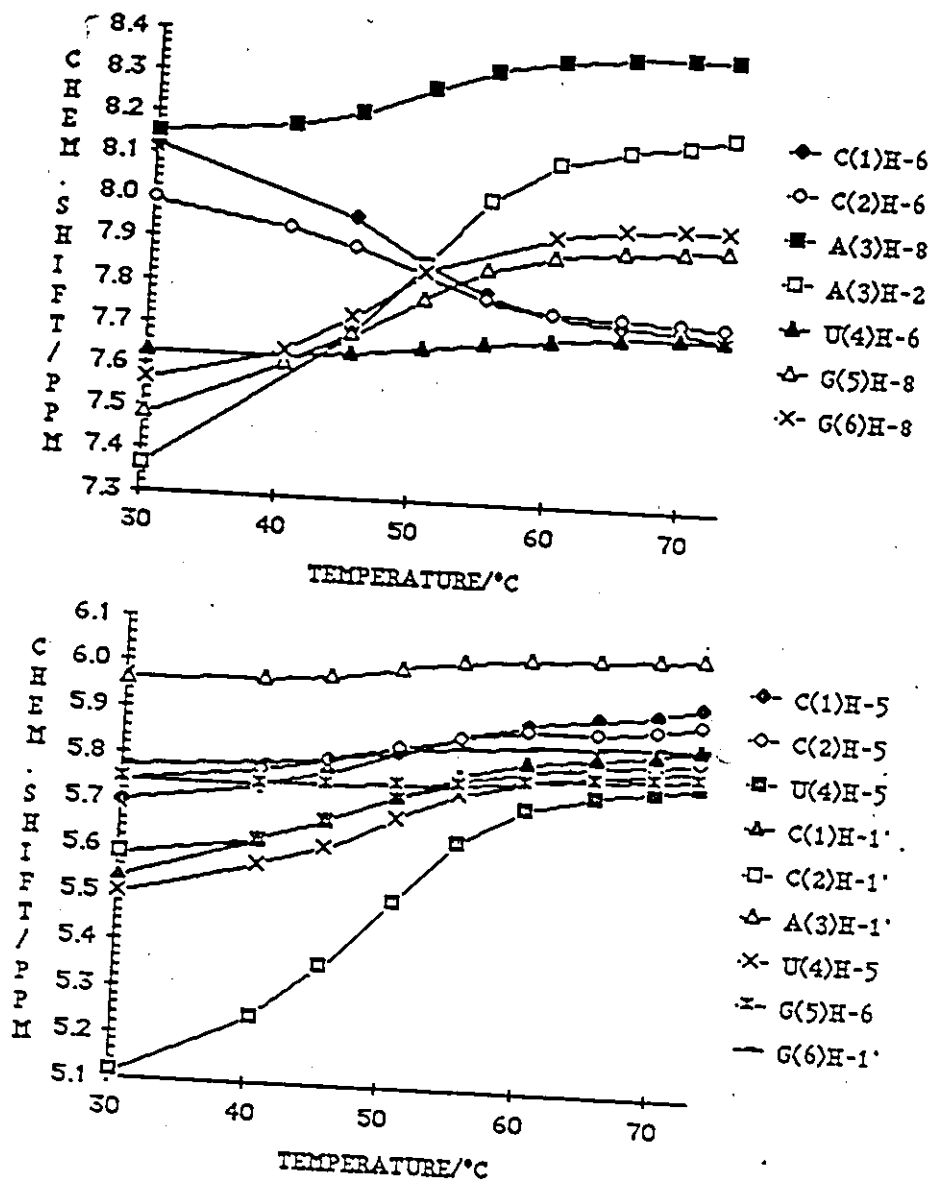


Figure 4.10: Chemical shift versus temperature curves for the base and anomeric protons of CCAUGG, 1.5 mM, in 0.1 M salt buffer.



Table 4.13: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CCAUGG, 1.3mM, in 0.1M salt buffer over the temperature range 80 - 30°C.

Temp	Chemical shift (ppm)									Tm
	80.0°	69.7°	65.3°	60.2°	55.2°	50.7°	45.4°	40.4°	30.3°	
Proton										
C(1)H-6	7.713	7.720	7.729	7.747	7.790	7.859	7.958	—	8.120	NSB
C(2)H-6	7.731	7.737	7.742	7.753	7.780	7.830	7.889	7.934	7.988	48.8
A(3)H-8	8.369	8.367	8.362	8.350	8.322	8.275	8.210	8.174	8.149	49.4
A(3)H-2	8.180	8.161	8.143	8.104	8.014	7.856	7.685	—	7.364	47.5
U(4)H-6	7.706	7.702	7.698	7.690	7.677	7.659	7.639	7.627	7.632	51.1
G(5)H-8	7.913	7.908	7.901	7.885	7.848	7.773	7.685	7.610	7.481	47.3
G(6)H-8	7.961	7.960	7.953	7.931	—	7.840	7.729	7.643	7.565	43.9
C(1)H-5	5.942	5.925	5.912	5.892	5.862	5.821	5.773	5.739	5.698	50.3
C(2)H-5	5.909	5.893	5.882	5.879	5.862	5.833	5.802	5.774	5.742	49.5
U(4)H-5	5.774	5.758	5.743	5.709	5.631	5.500	5.537	5.241	5.112	49.2
C(1)H-1'	5.856	5.841	5.830	5.810	5.777	5.725	5.667	5.621	5.534	49.6
C(2)H-1'	5.826	5.815	5.806	5.791	5.766	5.725	5.667	5.621	5.584	46.4
A(3)H-1'	6.052	6.046	6.040	6.034	6.022	6.005	5.982	5.971	5.961	50.5
U(4)H-1'	5.810	5.797	5.787	5.772	5.736	5.679	5.610	5.557	5.500	48.5
G(5)H-1'	5.790	5.784	5.778	5.772	5.761	5.754	5.746	5.743	5.742	NSB
G(6)H-1'	5.856	5.853	5.850	5.846	5.837	8.824	5.803	5.790	5.773	47.7
									Average Tm	49.4
	Coupling constant (Hz)									
C(1)H-1'	5.3	3.1	4.5	4.8	3.5	2.1	—	—		
C(2)H-1'	3.8	3.4	3.4	4.5	2.0	2.1	—	—		
A(3)H-1'	4.4	5.1	4.9	4.7	4.4	3.2	—	—		
U(4)H-1'	3.4	5.6	6.2	5.4	4.5	2.5	—	—		
G(5)H-1'	3.2	5.1	5.3	5.4	4.2	2.9	2.2	—		
G(6)H-1'	5.3	5.4	5.5	5.1	5.1	4.0	1.7	—		

Table 4.14: Chemical shifts of the hydrogen-bonded imino protons of CCAUGG, 1.5mM, in 0.1M salt buffer over the temperature range -4° to 33°C.

Numbering scheme:

1  
CCAUGG  
GGUACC

Proton	Chemical shift (ppm)				
	<u>-4.3°</u>	<u>-5.7°</u>	<u>15.2°</u>	<u>25.1°</u>	<u>32.5°</u>
G•C1	13.31	13.33	13.32	--	--
G•C2	12.36	12.38	12.38	12.36	12.33
A•U3	13.31	13.33	13.32	13.31	13.31

compared to 52.7°C for the low field protons of CCAGG:CCUGG and 48°C for CCAUGG. The concentration difference would account for a drop in  $T_m$  of 2 - 3°C, given previous observations, bringing the  $T_m$  to 50 - 51°C. This gives a 7°C  $T_m$  increase for insertion of the first A•U base pair, and little effect upon the insertion of a second.

#### 4.2.3. Sequences derived from parent GCGC

##### 4.2.3.1. Stability and secondary structure of GCUGC and GCAGC

In 0.1M salt, at 2.5mM single strand concentration the parent duplex GCGC forms a perfect four-base-pair duplex with  $T_m$  48.1°C, averaged over seven curves with  $T_m$  range 44.9 - 54.7°C (Table 4.15; Figure 4.11). Two distinct hydrogen bonded imino proton resonances are observed at 13.05 and 12.77 ppm, which are assigned to the internal and terminal base pairs respectively since the upfield resonance is the first to broaden with increasing temperature (Table 3.4; Figure 3.6).

The chemical shift versus temperature curves of GCUGC (2.5mM, 0.1M salt) are sigmoidal for ten of thirteen curves (Figure 4.12; Table 4.16), with individual proton  $T_m$ s occurring over the range 17.2 - 22.3°C. The imino proton spectrum at 2.0°C shows three distinct resonances, at 13.38, 12.87 and 10.25 ppm. There is no clear order of melting with increasing temperature (Table 4.17).

For GCAGC, the  $T_m$  in 1.0M salt (2.0mM single strand concentration) is 36.6°C, over ten of thirteen curves with a range of 33.8 - 40.6°C (Table 4.18; Figure 4.13), and the  $T_m$  in 0.1M salt (2.5mM strand concentration) is 35.1°C, over ten curves with a 31.8 - 37.1°C range (Table 4.20). The imino proton spectrum of GCAGC at 6°C shows a single resonance at 12.37 ppm.

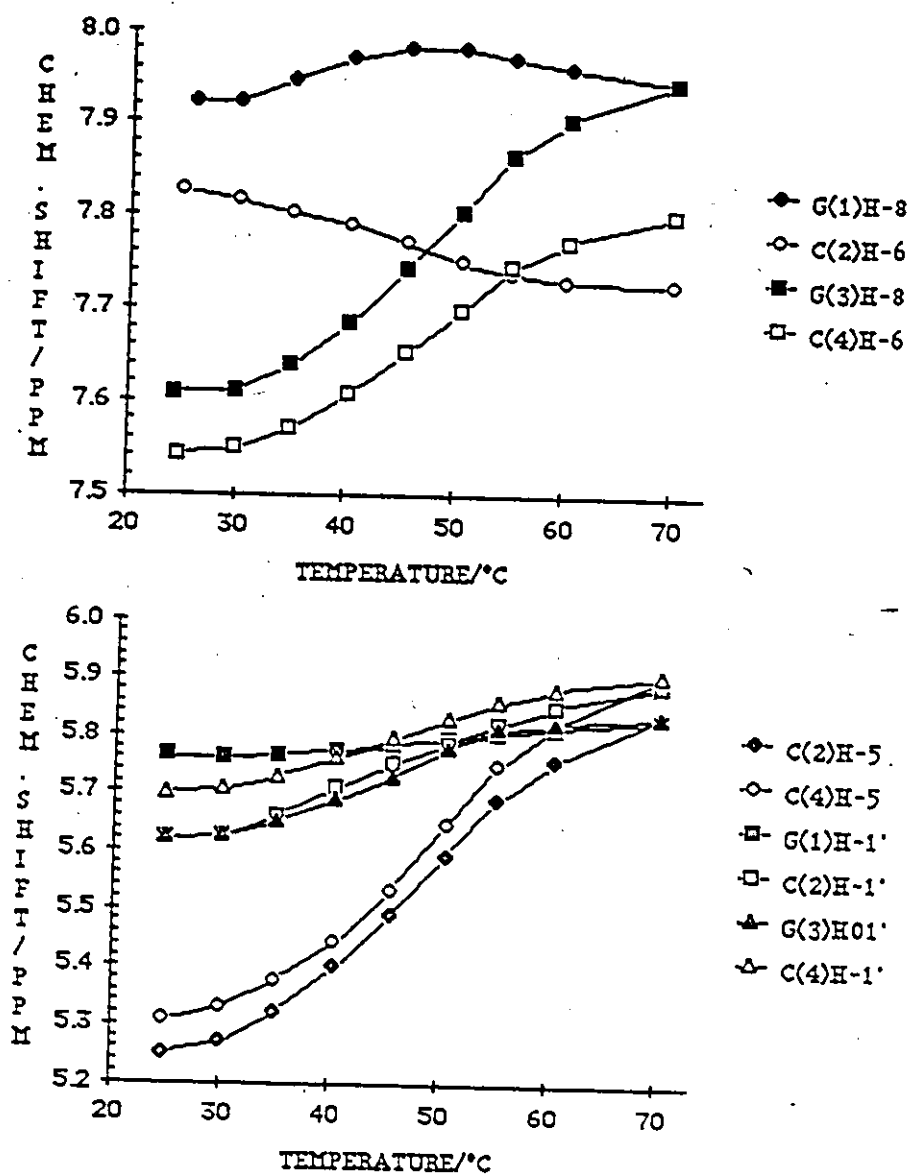


Figure 4.11: Chemical shift versus temperature curves for the base and anomeric protons of GCGC, 2.4 mM, in 0.1 M salt buffer.

Table 4.15: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GCGC, 2.4mM, in 0.1M salt buffer over the temperature range 70 to 25°C.

Proton	Chemical shift (ppm)										
	Temp.	<u>70.2°</u>	<u>60.4°</u>	<u>55.3°</u>	<u>50.7°</u>	<u>45.6°</u>	<u>40.5°</u>	<u>35.0°</u>	<u>30.1°</u>	<u>24.9°</u>	T <sub>m</sub>
G(1)H-8	7.942	7.958	7.970	7.981	7.981	7.969	7.946	7.923	7.921	NSB	
C(2)H-6	7.725	7.729	7.738	7.751	7.770	7.789	7.804	7.815	7.828	46.5	
G(3)H-8	7.942	7.904	7.865	7.807	7.742	7.684	7.738	7.609	7.607	48.5	
C(4)H-6	7.800	7.772	7.745	7.696	7.651	7.608	7.571	7.584	7.541	48.9	
C(2)H-5	5.839	5.755	5.687	5.592	5.489	5.400	5.322	5.249	5.251	50.9	
C(4)H-5	5.906	5.821	5.749	5.646	5.534	5.442	5.376	5.331	5.309	49.9	
G(1)H-1'	5.827	5.813	5.805	5.794	5.784	5.774	5.767	5.762	5.765	NSB	
C(2)H-1'	5.887	5.850	5.819	5.788	5.752	5.709	5.661	5.625	5.620	38.5	
G(3)H-1'	5.833	5.821	5.810	5.777	5.729	5.687	5.651	5.625	5.620	45.0	
C(4)H-1'	5.904	5.880	5.859	8.829	5.793	5.759	5.727	5.704	5.698	46.2	
Average T <sub>m</sub>										48.1	
Proton	Coupling constant (Hz)										
	G(1)H-1'	4.5	3.6	3.4	2.4	2.0	--	--	--	--	
	C(2)H-1'	4.7	3.9	3.6	3.4	2.5	1.5	--	--	--	
	G(3)H-1'	4.6	4.4	4.1	2.5	1.7	--	--	--	--	
	C(4)H-1'	3.8	3.4	3.2	2.7	2.4	1.6	--	--	--	

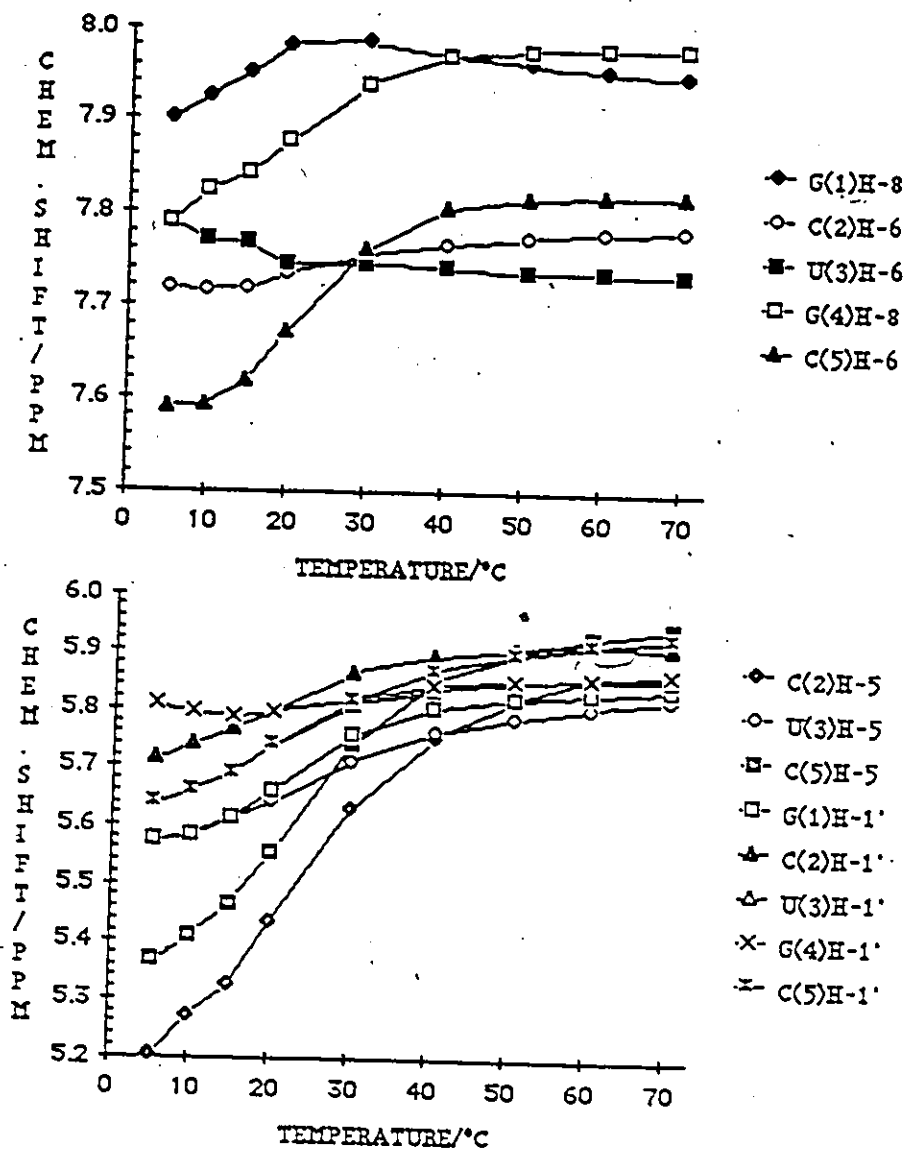


Figure 4.12: Chemical shift versus temperature curves for the base and anomeric protons of GCUGC, 2.5 mM, in 0.1 M salt buffer.



Table 4.17: Chemical shifts of the hydrogen-bonded imino protons of GCUGC, 2.5mM strand concentration, in 0.1M salt buffer, over the temperature range 2° to 9°C.

Proton	Chemical shift (ppm)		
	<u>2.0°</u>	<u>4.3°</u>	<u>9.2°</u>
TERMINAL	12.87	13.00	12.88
INTERNAL	13.38	13.43	13.27
UNH3	10.25	10.33	10.28



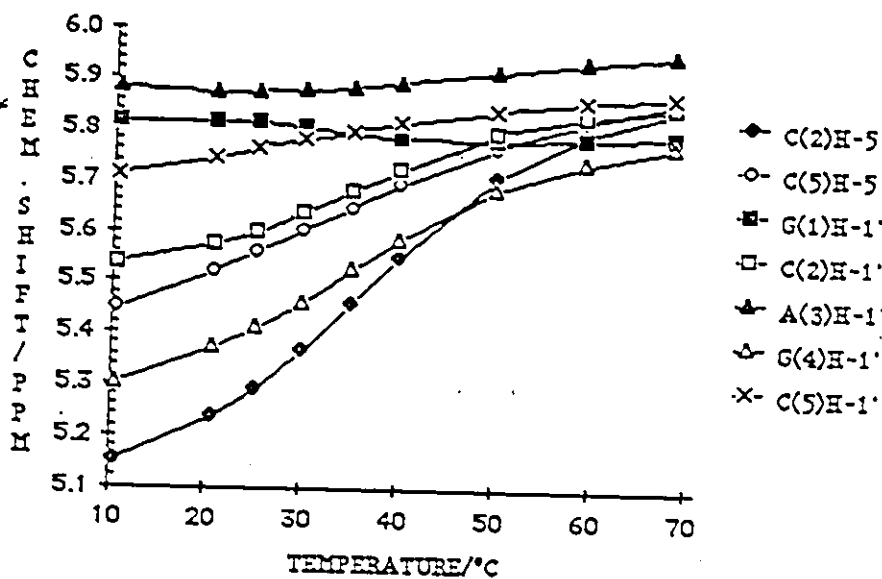
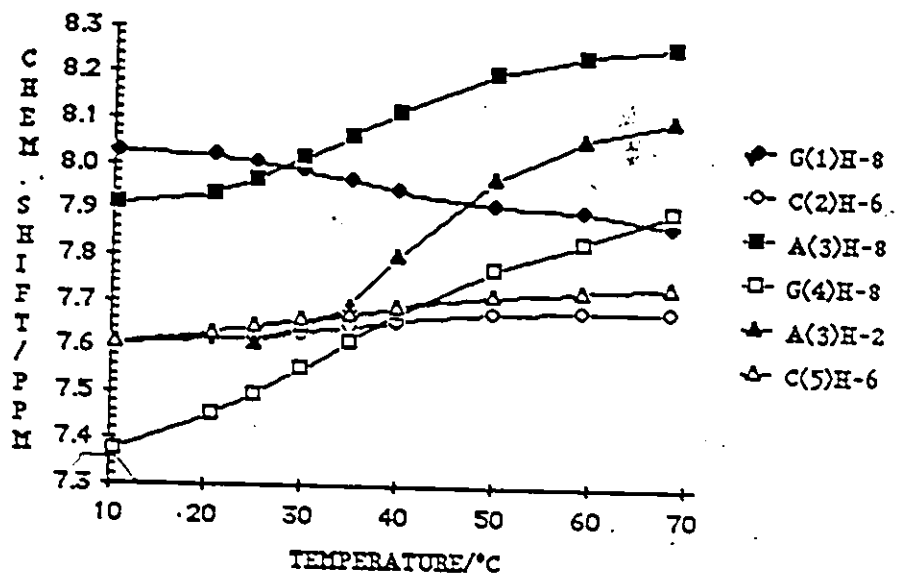


Figure 4.13: Chemical shift versus temperature curves for the base and anomeric protons of GCAGC, 2.5 mM, in 1.0 M salt buffer.

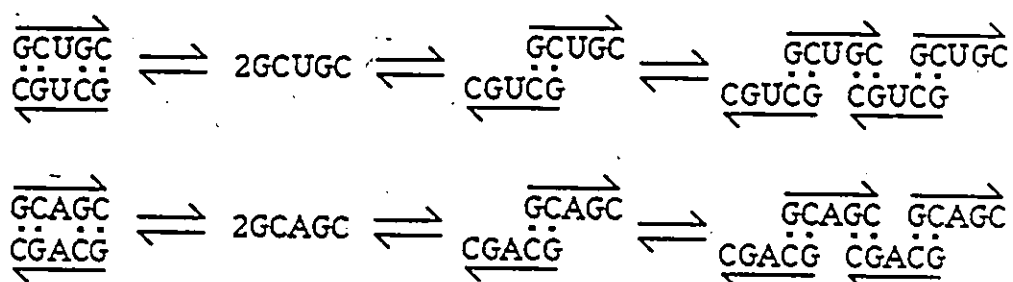
Table 4.18: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GCAGC, 2.5mM, in 1.0M salt buffer over the temperature range 70 to 10°C.

Temp	Chemical shift (ppm)									T <sub>m</sub>
	68.5°	59.4°	50.1°	40.1°	35.0°	30.0°	25.1°	20.7°	10.7°	
Proton										
G(1)H-8	7.870	7.902	7.913	7.943	7.965	7.986	8.005	8.017	8.024	35.6
C(2)H-6	7.684	7.682	7.675	7.658	7.643	7.631	7.621	7.617	7.604	36.6
A(3)H-8	8.261	8.241	8.199	8.116	8.063	8.012	7.965	7.935	7.910	36.0
A(3)H-2	8.099	8.057	7.970	7.799	7.688	7.646	7.610	-	-	38.9
G(4)H-8	7.901	7.835	7.776	7.675	7.612	7.553	7.495	7.453	7.737	35.7
C(5)H-6	7.737	7.728	7.714	7.690	7.673	7.660	7.643	7.629	7.604	NSB
C(2)H-5	5.845	5.796	5.707	5.550	5.458	5.368	5.291	5.235	5.152	40.3
C(5)H-5	5.859	5.821	5.769	5.694	5.648	5.605	5.559	5.521	5.447	35.6
G(1)H-1'	5.793	5.786	5.779	5.787	5.796	7.803	7.816	5.815	5.815	33.8
C(2)H-1'	5.869	5.829	5.797	5.723	7.678	5.638	5.559	5.574	5.536	36.9
A(3)H-1'	5.950	5.935	5.915	5.893	5.884	5.877	5.874	5.873	5.883	NSB
G(4)H-1'	5.773	5.739	5.683	5.586	5.524	5.461	5.412	5.371	5.301	36.7
C(5)H-1'	5.867	5.858	5.841	5.814	5.796	5.780	5.762	5.742	5.709	NSB
										Average T <sub>m</sub> 36.6
					Coupling constant (Hz)					
G(1)H-1'	5.0	4.7	3.1	5.6	3.2	2.2	-	-		
C(2)H-1'	5.4	3.7	2.6	3.0	2.4	-	-	-		
A(3)H-1'	4.4	4.1	3.3	2.3	1.8	-	-	-		
G(4)H-1'	4.7	4.1	4.0	3.2	2.7	-	-	-		
C(5)H-1'	3.6	3.6	3.6	3.3	3.2	3.0	2.7	2.6		

Table 4.19: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GCAGC, 2.5 mM, in 0.1 M salt buffer, over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)								T <sub>m</sub>
	68.5°	59.4°	50.1°	40.0°	30.0°	25.1°	20.7°	10.7°	
Proton									
G(1)H-8	7.908	7.910	7.921	7.950	7.993	8.014	8.028	8.050	33.8
C(2)H-6	7.735	7.735	7.726	7.702	7.669	7.656	7.651	7.641	37.2
A(3)H-8	8.296	8.278	8.240	8.163	8.064	8.014	7.990	7.943	36.0
A(3)H-2	8.112	8.070	7.984	7.816	7.595	7.499	7.430	—	34.9
G(4)H-8	7.885	7.854	7.800	7.705	7.581	7.523	7.475	7.384	34.3
C(5)H-6	7.796	7.789	7.776	7.752	7.720	7.702	7.685	7.641	NSB
C(2)H-5	5.873	5.832	5.752	5.607	5.428	5.352	5.289	5.211	36.0
C(5)H-5	5.910	5.881	5.840	5.777	5.697	5.655	5.616	5.552	NSB
G(1)H-1'	5.804	5.797	5.794	5.798	5.806	5.814	5.815	5.823	31.9
C(2)H-1'	5.869	5.850	5.813	5.742	5.654	5.617	5.592	5.552	36.9
A(3)H-1'	5.965	5.950	5.932	5.910	5.892	5.889	5.888	5.901	NSB
G(4)H-1'	5.787	5.760	5.710	5.623	5.510	5.457	5.415	5.335	34.8
C(5)H-1'	5.880	5.867	5.849	5.822	5.784	5.764	5.746	5.705	NSB
								Average T <sub>m</sub>	35.1
	Coupling constant (Hz)								
G(1)H-1'	4.6	4.7	4.8	2.8	2.4	—	—	—	
C(2)H-1'	4.8	4.7	5.1	3.0	1.2	—	—	—	
A(3)H-1'	4.8	4.3	3.7	2.7	1.5	—	—	—	
G(4)H-1'	3.8	4.1	3.9	3.4	3.0	—	—	—	
C(5)H-1'	3.9	3.6	3.2	3.3	3.0	2.6	2.3	—	

Both sequences have the same available secondary structures. A four base pair duplex with central base-base mismatch, or two-base pair duplex with six triple-dangling bases—on the basis of previous results (Section 3.2.4), 3' dangling bases would in all probability be preferred—with, possibly, overlap of the sticky ends and formation of an extended loose duplex.



The tetramer precursors of GCUGC and GCAGC, GCUG and GCAG, were studied as models of the staggered structure:



GCUG gives a  $T_m$  of 13.5°C—computer identified because of the incompleteness of the curves (Figure 4.14); individual proton  $T_m$ s range from 11.6 - 14.5°C, for seven of ten curves (Table 4.20). The only imino proton resonance observed occurs at 11.2 ppm, which is the same as that observed for single-stranded CUG (Buchko, 1985).

GCAG (2.5mM, 1.0M salt) has a  $T_m$  of 33.1°C, calculated for eight of ten curves with individual  $T_m$ s of 32.3 - 37.1°C (Figure 4.15; Table 4.21). A single resonance appears in the low temperature imino proton spectrum, at 12.47 ppm.

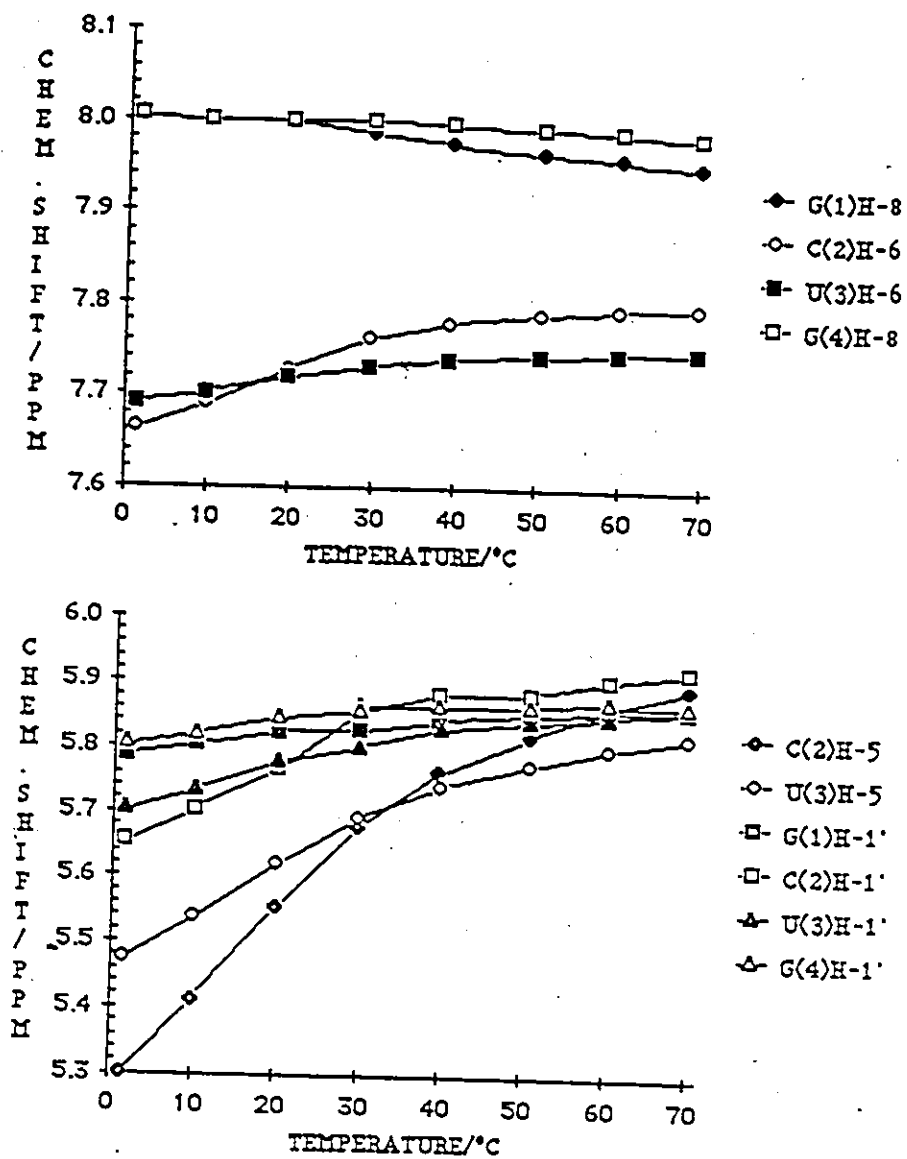


Figure 4.14: Chemical shift versus temperature curves for the base and anomeric protons of GCUG, 2.5 mM, in 0.1 M salt buffer.

Table 4.20: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GCUG, 2.5 mM, in 0.1 M salt buffer, over the temperature range 70 to 0°C.

Temp.	Chemical shift (ppm)								T <sub>m</sub>
	69.7°	60.0°	50.4°	39.5°	29.8°	19.8°	9.9°	1.6°	
Proton									
G(1)H-8	7.952	7.959	7.966	7.975	7.987	8.001	7.999	8.004	9.5
C(2)H-6	7.746	7.794	7.788	7.778	7.761	7.728	7.689	7.662	13.9
U(3)H-6	7.748	7.746	7.743	7.738	7.731	7.718	7.700	7.69	13.9
G(4)H-8	7.983	7.988	7.992	7.996	7.999	8.001	7.999	8.004	NSB
C(2)H-5	5.895	5.863	5.820	5.761	5.678	5.553	5.412	5.300	14.2
U(3)H-6	5.818	5.799	5.773	5.739	5.692	5.620	5.537	5.476	14.5
G(1)H-1	5.858	5.857	5.853	5.842	5.827	5.824	5.802	5.785	NSB
C(2)H-1'	5.920	5.905	5.884	5.856	5.833	5.766	5.704	5.654	13.6
U(3)H-1'	5.858	5.848	5.840	5.828	5.801	5.777	5.734	5.703	12.2
G(4)H-1'	5.869	5.870	5.864	5.863	5.856	5.843	5.820	5.804	11.6
								Average T <sub>m</sub>	13.5
	Coupling constant (Hz)								
G(1)H-1'	5.1	4.5	4.6	3.2	5.9	--	--	--	
C(2)H-1'	4.8	4.5	4.5	3.8	2.6	--	--	--	
U(3)H-1'	5.1	4.1	5.7	6.0	4.6	2.0	--	--	
G(4)H-1'	5.6	4.1	5.9	3.2	3.0	1.4	--	--	

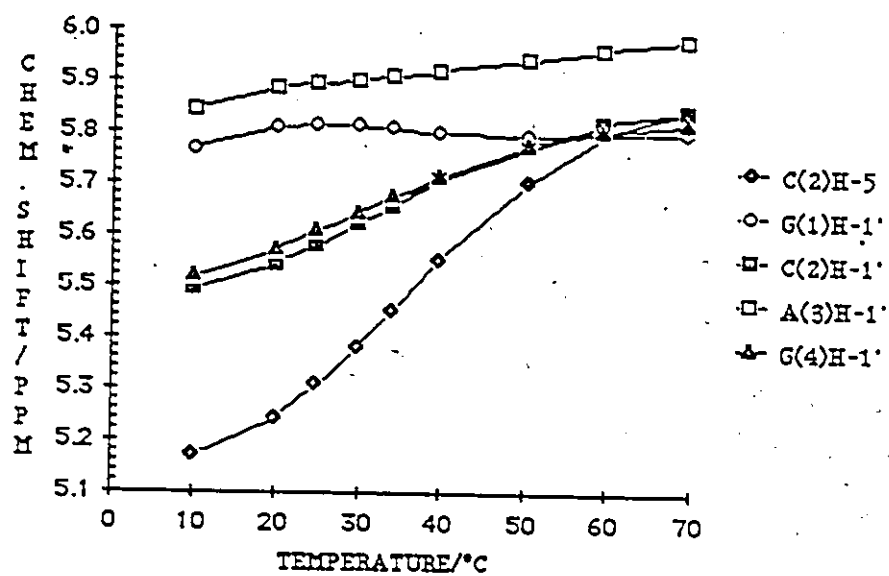
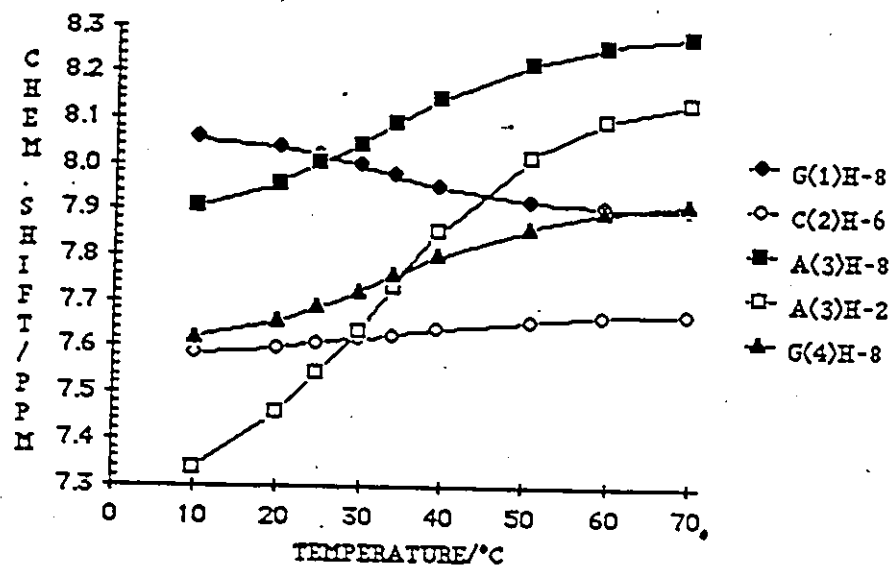


Figure 4.15: Chemical shift versus temperature curves for the base and anomeric protons of GCAG, 2.5 mM, in 1.0 M salt buffer.

Table 4.21: Chemical shifts and anomeric proton coupling constants for base and anomeric protons of GCAG, 2.5mM, in 1.0M salt buffer, over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)									T <sub>m</sub>
	69.7°	59.5°	50.4°	39.5°	34.0°	28.9°	24.8°	19.8°	9.9°	
Proton										
G(1)H-8	7.903	7.907	7.919	7.952	7.978	7.996	8.017	8.034	8.055	34.3
C(2)H-6	7.670	7.667	7.657	7.626	7.616	7.606	7.597	7.586	7.586	35.3
A(3)H-8	8.275	8.255	8.217	8.141	8.087	8.042	8.001	7.857	7.915	33.8
A(3)H-2	8.133	8.095	8.015	7.852	7.731	7.632	7.545	7.457	7.335	34.6
G(4)H-8	7.913	7.895	7.861	7.801	7.758	7.721	7.688	7.654	7.617	37.1
C(2)H-5	5.844	5.794	5.703	5.551	5.450	5.379	5.305	5.243	5.170	37.1
G(1)H-1'	5.779	5.797	5.791	5.800	5.808	5.812	5.811	5.805	5.766	NSB
C(2)H-1'	5.839	5.819	5.771	5.705	5.656	5.621	5.578	5.542	5.494	32.3
A(3)H-1'	5.978	5.961	5.942	5.919	5.906	5.898	5.891	5.883	5.842	NSB
G(4)H-1'	5.819	5.801	5.776	5.714	5.674	5.641	5.605	5.571	5.516	33.8
										Average T <sub>m</sub> 33.1
	Coupling constant (Hz)									
G(1)H-1'	5.0	6.4	2.5	3.6	3.1	2.4	--	--	--	
C(2)H-1'	4.9	4.4	4.8	2.8	1.8	--	--	--	--	
A(3)H-1'	4.8	4.5	3.9	3.1	2.3	1.5	--	--	--	
G(4)H-1'	5.1	4.4	5.0	4.3	4.1	3.2	2.5	3.4	--	



The closeness of the  $T_m$ s of the duplexes formed by GCA ( $T_m$  33.7°C, 7.3mM; Alkema et al., 1981), GCAG and GCAGC, two of which can only form duplexes comprising a GC:GC core and 3' dangling bases, and the observation of a single imino proton at a similar position for all three—12.24 ppm for GCA, 12.47 ppm for GCAG and 12.37 ppm for GCAGC—suggests that the preferred conformation for GCAGC is the staggered one. This suggestion is further confirmed by the similarity of the curve-shapes for individual protons, especially those of the adenosine residue (See Figures 4.13, 4.15 and 4.16) and the behaviour of the 3' terminal residue of GCAGC, C5.

From the results described in section 3.3.3, it seems that the  $J_{1-2}$  coupling constants of residues involved in GC:GC cores tend to lead the others slightly in their collapse, and certainly do not lag. In Figure 4.17, the coupling constants of GCAGC are plotted; the coupling constant of C5H1' is conspicuously slower than the others in collapsing. Comparison of the  $T_m$ s of the individual protons of C2 and C5 add to the evidence of delayed stacking. The average  $T_m$  of C2H6 and C2H5 is 38.6°C, while the curves of C5H6 and C5H1' are nonsigmoidal, and the  $T_m$  of C2H5 is 35.6°C.

The staggered duplexes apparently are not inclined to end-to-end aggregation in the temperature range studied, since only a single imino proton signal is observed.

In contrast to the single imino proton signal observed for GCAGC, three signals are observed in the GCUGC spectrum (Table 4.17). For a four-base-pair, self-complementary duplex, two signals from G-C protons would be expected, and these are observed at 13.38 and 12.87, positions qualitatively predictable from the results for GCGC. Introduction of the central UxU mismatch should affect the terminal imino proton less, and should cause the internal one to move downfield, since a 3' guanosine is being replaced by a 3' uridine. The terminal resonance, which appeared at 12.77

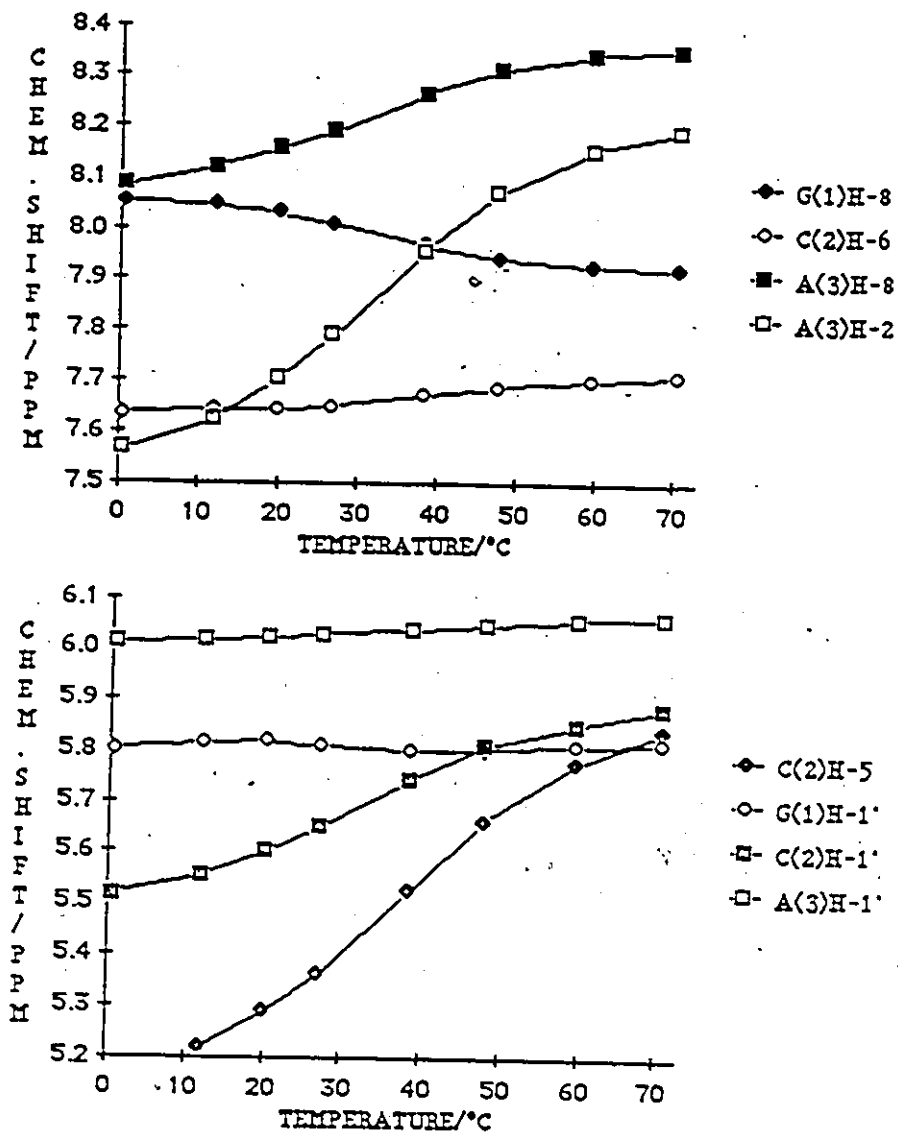


Figure 4.16: Chemical shift versus temperature curves for the base and anomeric protons of GCA, 7.3 mM, in 1.0 M salt buffer. (From Alkema et al., 1981)

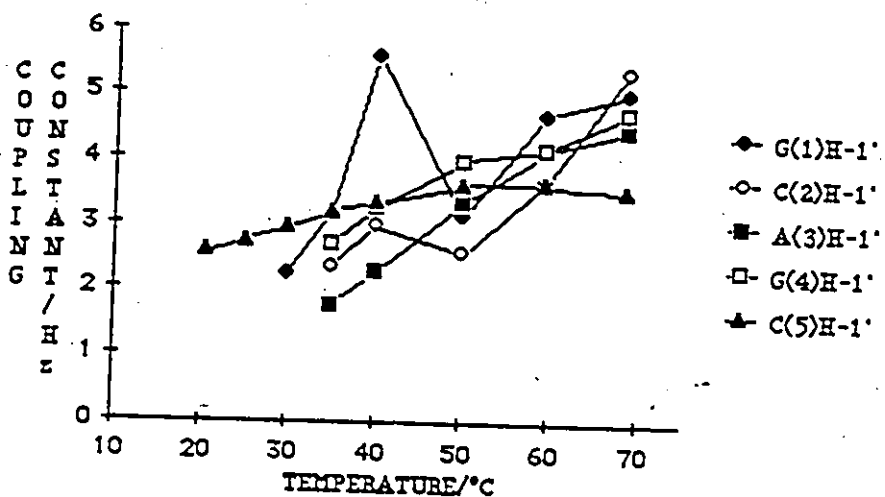


Figure 4.17: Coupling constants versus temperature for the anomeric protons of GCAGC, 2.5 mM, in 1.0 M salt buffer.

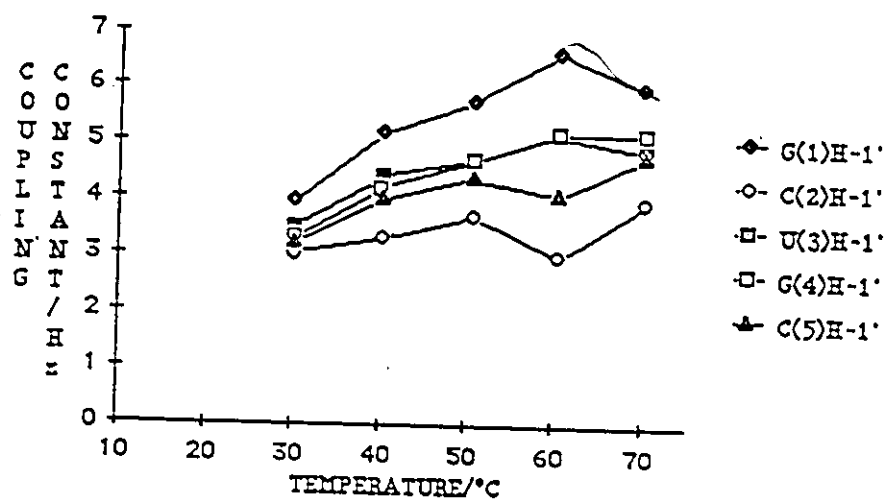


Figure 4.18: Coupling constants versus temperature for the anomeric protons of GCUGC, 2.5 mM, in 0.1 M salt buffer.

ppm in GCGC, now appears at 12.87, and the internal one has moved from 13.05 ppm to 13.38 ppm. The third signal, at 10.25 ppm, can be assigned to the N3H, shifted upfield from its single strand position of 11.2 (as in single stranded CUG (Buchko, 1985)). The H1' data for GCUGC (Figure 4.18) is less satisfactory than that for GCAGC, as the broadening of the resonances at low temperature obscures the coupling constants below 30°C, at which point they are still > 2Hz. The average T<sub>m</sub> of the C2 base protons is 20.7°C, and the average for the C5 base protons is 21.0°C, with all curves sigmoidal—C2 and C5 appear to be stacking at similar rates.

In conclusion, the evidence favours formation of the staggered duplex by GCAGC, and the duplex containing a mismatch by GCUGC.

#### 4.2.3.2. Conformation at the mismatch site in GCUGC

The conformation around the uridine-uridine mismatch is of interest, since the duplex geometry can accommodate a stacked-in uridine-uridine opposition, but uridine stacks poorly, and there might be an incentive to force the uridine out. The base protons of a looped out uridine might be expected to show a downfield shift with duplex formation, as the random coil, in which uridine would be partially shielded, changes to a configuration in which uridine moves outside the stack. U3H6 moves downfield by a small amount, 0.055 ppm, but U3H5 moves upfield by 0.245 ppm (Figure 4.12). There is no point in attempting to compare the behaviour of these uridines with that of uridines which are hydrogen-bonded to adenosine, as the deshielding effect due to the adenosine would obscure the chemical shift changes due to stacking.

The looping out of the uridine resonances might be expected to result in a different conformation, or greater flexibility, of the uridine ribose ring, however,

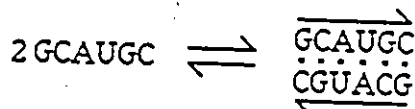


H5 protons were too extensively overlapped. Eight of eleven chemical shift versus temperature curves for the downfield base protons (AH8, AH2, GH8, C and U H6) are sigmoidal, with  $T_m$ s of 47.7 - 52.7°C, average 51.0°C (Figure 4.19; Table 4.22).

Comparable  $T_m$  values are observed in both strands, and in any case these values are 15 - 30°C higher than those in the single strands, sufficient to justify the interpretation that they belong to the GCA<sup>+</sup>GC:GCUGC duplex.

The five expected imino proton signals are observed overlapped as two broad peaks. G-C 1, G-C 5 and G-C 2 appear at 12.64 ppm and G-C 4 and A-U 3 at 13.59 ppm (see Section 3.3.7 for assignments.) Intensity is lost upon warming first at 12.64 ppm (Table 4.23); the sequence frays from the ends inwards.

GCAUGC was synthesized by Dr Dirk Alkema, but its duplexing behaviour has not been previously studied. It is self complementary and forms a duplex consisting of two GC:GC cores separated by one AU:AU core:



At 2.5mM, in 0.1M salt, twelve of fifteen curves are sigmoidal (Figure 4.20) with  $T_m$ s of 51.7 - 54.9°C, overall average, 53.4°C, downfield base protons only, 53.9°C (Table 4.24). The signal of the sixteenth proton, C6H5, loses intensity and moves rapidly underneath the H1' signals.

This sample was in 250µl in a microtube, and provided very poor signal to noise in the imino proton spectra while in the microtube, and poor resolution when transferred to an ordinary tube. The sample was subsequently diluted to 400µl, since previous experience suggested that such dilution should lower the  $T_m$  by 2 - 3°C at

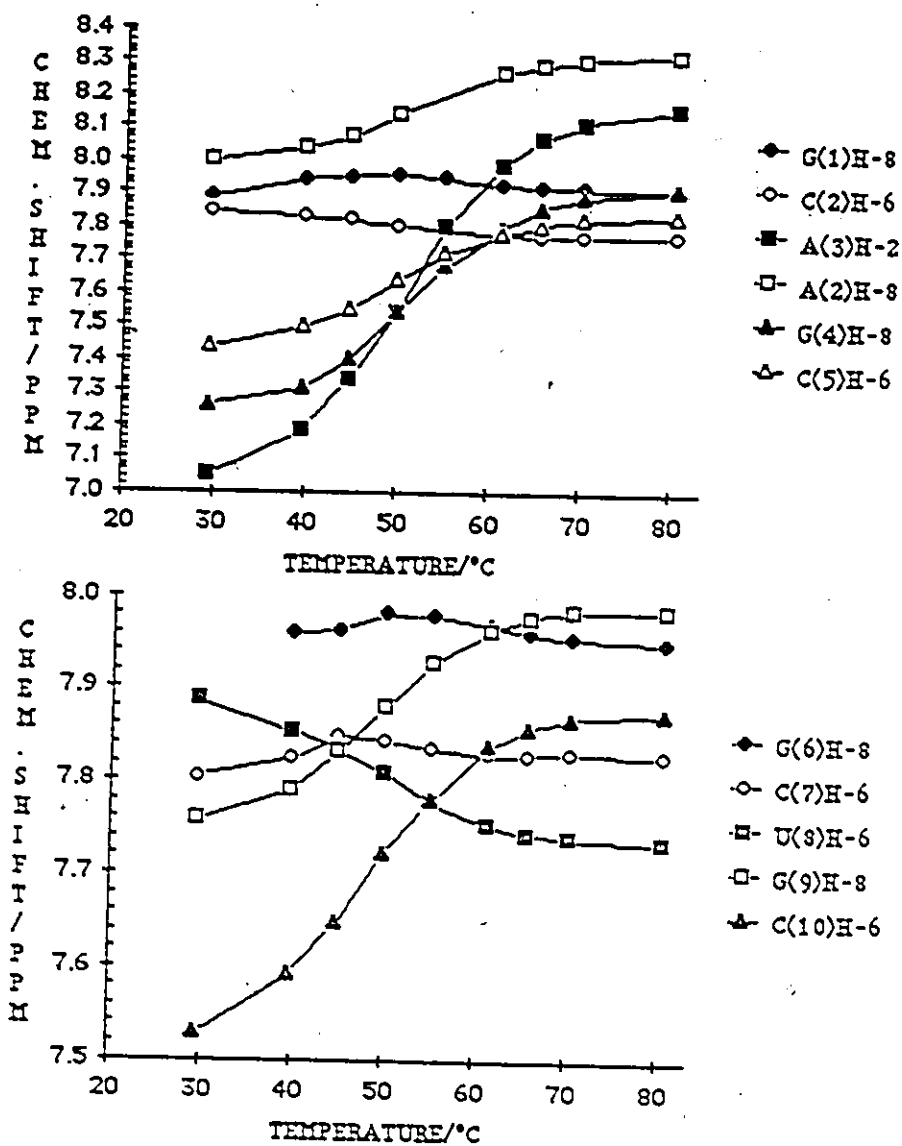


Figure 4.19: Chemical shift versus temperature curves for the base protons of GCAGC:GCUGC, 2.5 mM each strand, in 0.1 M salt buffer.

Table 4.22: Chemical shifts for the base protons of GCAGC:GCUGC, 2.5mM each strand, in 0.1M salt buffer over the temperature range 80 to 30°C.

Numbering scheme:  $\begin{array}{cc} 1 & 5 \\ \text{GCAGC} & \\ \text{CGUCG} & \\ 10 & 6 \end{array}$

Proton	Chemical shifts (ppm)									T <sub>m</sub>
	80.5°	70.2°	65.5°	61.2°	55.1°	50°	44.9°	39.8°	29.5°	
G(1)H-8	7.911	7.912	7.914	7.925	7.942	7.953	7.946	7.941	7.887	NSB
C(2)H-6	7.765	7.766	7.766	7.774	7.787	7.799	7.816	7.823	7.739	52.7
A(3)H-8	8.314	8.300	8.285	8.260	—	8.140	8.073	8.033	7.994	51.7
A(3)H-2	8.150	8.109	8.064	7.983	7.793	7.536	7.336	7.183	7.047	51.7
G(4)H-8	7.911	7.882	7.850	7.792	7.677	7.536	7.393	7.306	7.256	51.3
C(5)H-6	7.827	7.817	7.800	7.774	7.714	7.633	7.547	7.496	7.436	50.6
G(6)H-8	7.948	7.954	7.958	7.966	7.978	7.979	7.963	7.960	—	NSB
C(7)H-6	7.827	7.830	7.828	7.830	7.834	7.843	7.849	7.823	7.804	NSB
U(8)H-6	7.735	7.739	7.742	7.753	7.776	7.808	7.834	7.854	7.887	51.7
G(9)H-8	7.984	7.982	7.975	7.962	7.928	7.880	7.833	7.791	7.757	48.7
C(10)H-6	7.871	7.867	7.857	7.838	7.780	7.722	7.646	7.592	7.527	49.6
										Average T <sub>m</sub> 51.0



Table 4.23: Chemical shifts of the hydrogen-bonded imino protons of GCAGC:GCUGC, 2.5mM each strand, in 0.1M salt buffer, over the temperature range -2° to 37°C.

Numbering scheme:  $\begin{array}{c} 1 \quad 5 \\ \text{GCAGC} \\ \text{CGUCG} \end{array}$

Proton	Chemical shift (ppm)				
	<u>-2.1°</u>	<u>8.2°</u>	<u>17.4°</u>	<u>27.0°</u>	<u>37.2°</u>
G•C1	12.64	12.63	12.69	—	—
G•C2	12.64	12.63	12.61	12.60	12.56
A•U3	13.59	13.54	13.51	13.50	13.46
G•C4	13.59	13.54	13.51	13.50	13.46
G•C5	12.64	12.63	12.69	—	—

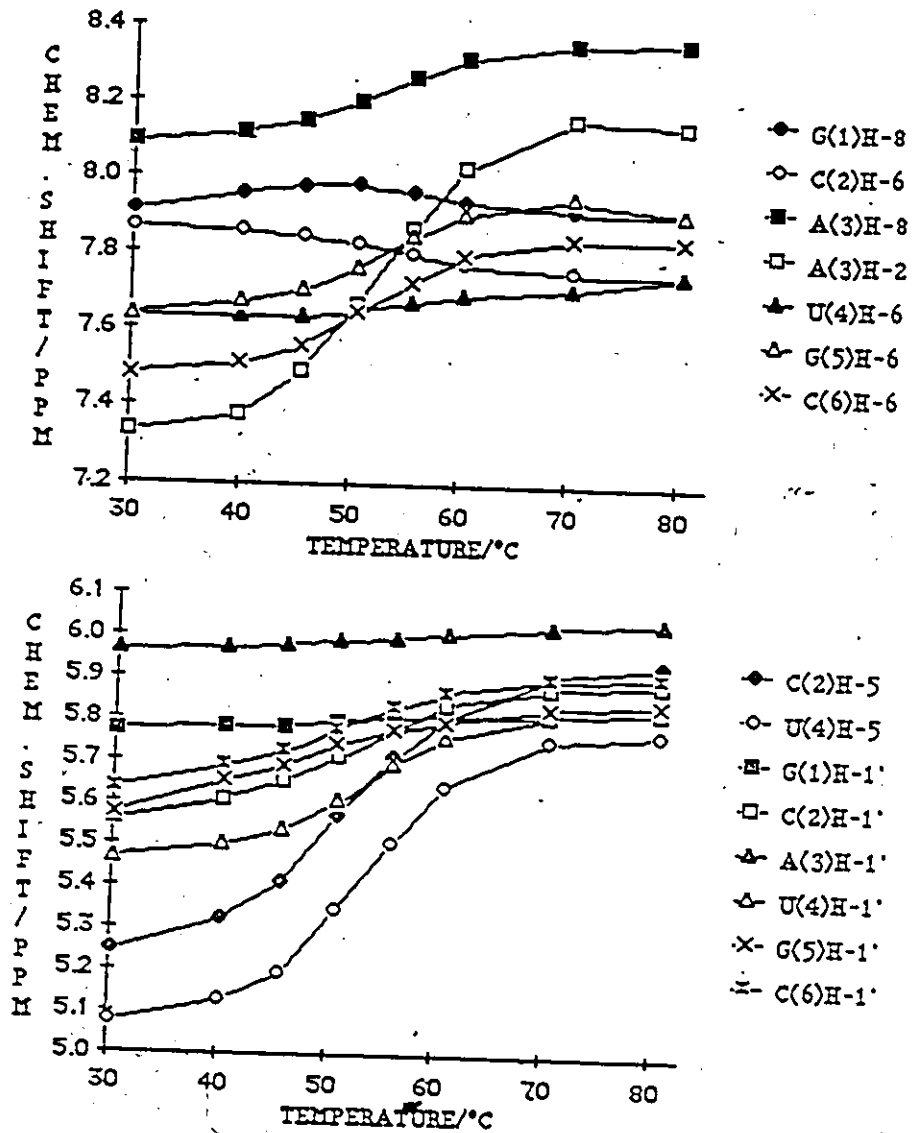


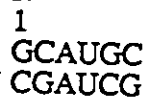
Figure 4.20: Chemical shift versus temperature curves for the base and anomeric protons of GCAUGC, 2.5 mM, in 0.1 M salt buffer.

Table 4.24: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of GCAUGC, 2.5 mM, in 0.1 M salt buffer over the temperature range 80 to 30°C.

Temp.	Chemical shift (ppm)								T <sub>m</sub>
	80.5°	70.3°	60.5°	55.8°	50.8°	45.8°	40.2°	30.1°	
Proton									
G(1)H-8	7.912	7.913	7.935	7.961	7.978	7.973	7.957	7.908	NSB
C(2)H-6	7.752	7.757	7.774	7.799	7.829	7.747	7.859	7.867	53.8
A(3)H-8	8.361	8.354	8.312	8.260	8.198	8.147	8.188	8.088	53.8
A(3)H-2	8.183	8.156	8.026	7.863	7.665	7.492	7.380	7.332	53.4
U(4)H-6	7.715	7.711	7.690	7.670	7.348	7.635	7.636	7.642	54.9
G(5)H-8	7.912	7.950	7.905	7.845	7.763	7.704	7.669	7.634	54.9
C(6)H-6	7.839	7.836	7.796	7.726	7.648	7.559	7.512	7.481	52.5
C(2)H-5	5.932	5.906	5.803	5.708	5.568	5.410	5.323	5.245	52.3
U(4)H-5	5.771	5.748	7.640	5.503	5.345	5.192	5.127	5.072	53.8
C(6)H-5	5.983	5.956	5.863	--	--	--	--	--	--
G(1)H-1'	5.819	5.811	5.812	5.801	5.791	5.782	5.778	5.771	52.7
C(2)H-1'	5.883	5.872	5.833	5.775	5.707	5.649	7.605	5.557	52.7
A(3)H-1'	6.033	6.022	6.004	5.884	5.985	5.976	5.970	5.961	59.6
U(4)H-1'	5.819	5.811	5.755	5.691	5.602	5.533	5.497	5.465	53.4
G(5)H-1'	5.837	5.829	5.794	5.775	5.736	5.688	5.648	5.568	47.5
C(6)H-1'	5.902	5.896	5.863	5.825	5.773	5.720	5.683	5.626	51.7
							Average T <sub>m</sub>		53.9
	Coupling constant (Hz)								
G(1)H-1'	4.0	5.0	4.3	3.2	2.2	--			
C(2)H-1'	5.9	5.0	2.2	3.3	--	--			
A(3)H-1'	5.0	5.0	4.3	3.4	2.2	--			
U(4)H-1'	4.0	5.0	5.1	1.4	--	--			
G(5)H-1'	5.0	4.0	4.3	3.3	3.2	--			
C(6)H-1'	3.6	2.6	3.3	3.0	1.4	--			

Table 4.25: Chemical shifts of the hydrogen-bonded imino protons of GCAUGC, 2.5mM single strand concentration, in 0.1M salt buffer, over the temperature range  $-5^{\circ}$  to  $35^{\circ}\text{C}$ .

Numbering scheme:



Proton	Chemical shift (ppm)				
	<u><math>-5.5^{\circ}</math></u>	<u><math>-5.4^{\circ}</math></u>	<u><math>15.1^{\circ}</math></u>	<u><math>25.2^{\circ}</math></u>	<u><math>35.3^{\circ}</math></u>
G•C1	12.74	12.71	(12.8)	--	--
G•C2	12.74	12.71	12.69	12.68	12.64
A•U3	13.37	13.37	13.39	13.40	13.40

most, and should not affect the positions of the imino protons. At  $-5.5^{\circ}\text{C}$  two signals were observed, at 13.37 ppm and at 12.74 ppm, in a 1:2 ratio (Table 4.25). By  $15^{\circ}\text{C}$  the ratio was 1:1, and the resonance belonging to the terminal G-C base pair can accordingly be assigned as one of the two at 12.74 ppm. The remaining pair of resonances disappear together, but can be assigned as 13.37 ppm, A-U, and 12.74 ppm, G-C, on the reasoning that the imino proton belonging to an A-U base pair (A-U<sub>0</sub> 14.5 ppm) with two 3' neighbouring pyrimidines will be more deshielded than that belonging to a G-C base pair (G-C<sub>0</sub> 13.6 ppm) with two 3' neighbouring purines (Table 4.25).

The  $T_m$  of the parent duplex, GCGC, for downfield base protons only, is  $48.0^{\circ}\text{C}$ . For the non-self-complementary duplex GCAGC:GCUGC, the average  $T_m$  for the same set of protons is  $51.0^{\circ}\text{C}$ , and for GCAUGC,  $53.9^{\circ}\text{C}$ . This gives an increase in  $T_m$  of about  $3^{\circ}\text{C}$  for the insertion of each of the first two A-U base pairs into the centre of the GCGC duplex.

### 4.3. Discussion

#### 4.3.1. Effect of insertion of base-base mismatches upon stability

Of the six pentaribonucleotides studied, each of which could potentially form a duplex containing a UxU or an AxA mismatch, two, GCUGC and CCAGG, formed mismatch containing duplexes, so identified by the cooperative sigmoidal transitions of the aromatic and anomeric protons, and the observation of hydrogen-bonded imino proton signals at positions consistent with a four base pair duplex with a central mismatch. The  $T_m$ s were 20.0° and 17.3°C for GCUGC and CCAGG respectively. One sequence, GCAGC, formed a two-base-pair staggered duplex comprising a GC:GC core with six triple-dangling bases, showing very similar behaviour to its precursors, GCA and GCAG, the  $T_m$ s being 33.4°, 33.1° and 36.7°C for GCA, GCAG and GCAGC respectively. Two sequences, GGUCC and CCUGG, had an alternative secondary structures available to them, involving G-U base pairing, and seemed to form equilibrium mixtures of the two. The sixth sequence, GGACC, underwent base stacking, but did not duplex above 20°C. Below this temperature deterioration of resolution indicated aggregation.

The destabilization due to the introduction of a UxU mismatch into the perfect duplex GCGC, in GCUGC, is 28°C. That due to the introduction of an AxA mismatch into CCGG, in CCAGG, is 27°C. Lower bounds may be placed upon the destabilization due to these mismatches when introduced into the other systems according to the structure which does form. An AxA mismatch introduced into GCAGC lowers the  $T_m$  by at least 15°C, 33.7°C being the  $T_m$  of the staggered structure which is preferred. A central UxU mismatch introduced into GGCC and CCGG, as in GGUCC and CCUGG, lowers the  $T_m$  by at least 28°C in both cases. And a central AxA mismatch in GGACC lowers the  $T_m$  by 35°C

or more.

Inclusion of the mismatched pairs in the base-stack is preferred in the duplexes formed by GCUGC and CCAGG, according to the generally upfield movements of the aromatic protons of the mismatched residues upon duplexing, and the chemical shifts of the hydrogen-bonded imino protons.

Most published studies of the effect of the introduction of internal UxU/TxT and AxA mismatches on oligonucleotide duplex stability and conformation have involved DNA oligomers, with the exception on the two series  $r(\text{CAXUG})$  and  $r(\text{AGXCU})$ , X=A,G,C,U. None of the CAXUG series were seen to duplex, the  $T_m$  of the parent duplex being only 24°C—leading to an estimate of at least 24°C destabilization for each of the four mismatches in this system (Romaniuk *et al.* 1979a). In the AGXCU series, derived from the parent duplex AGCU,  $T_m$  34°C, only AGACU and AGCCU formed mismatch-containing duplexes, with  $T_m$ s of 26° and 25°C respectively (Alkema *et al.* 1982). The UxU mismatch containing duplex  $(\text{AGUCU})_2$  does not form, introduction of this mismatch must therefore lower the  $T_m$  by at least 34°C.

In their temperature-jump studies of the association of tRNAs with complementary anticodon regions, Grosjean *et al.* found that a UxU mismatch in the central (2/2) position proved unexpectedly stable, compared to CxC, AxA, GxG or CxA oppositions (Grosjean *et al.*, 1978).

The deoxy analogue of a UxU opposition, dTxdT, has been studied in  $d(\text{CGTCG})$  (Mellema *et al.*, 1984b),  $d(\text{TATCCTATTAGGATA})$  (Haasnoot *et al.*, 1979) and  $d(\text{CA}_3\text{TA}_3\text{G}):d(\text{CT}_7\text{G})$  (Aboul-ela *et al.*, 1985).  $d(\text{CGTCG})$  in D<sub>2</sub>O shows detectable duplexing below 310 K (37°C), with observation of a (G•C) hydrogen bonded imino proton resonance at around 12.5ppm. No data is given for the parent duplex CGCG, so it is not possible to assess destabilization relative to CGCG. The thymidine methyl groups

are strongly shielded, indicating stacking in of these residues (Mellema *et al.*, 1985).

The duplex formed by d(TATCCTATTAGGATA) is destabilized by about 11°C relative to its parent d(TATCCTATAGGATA)<sub>2</sub>. The imino proton of U3NH is observed at 10.3ppm, relative to 11.2ppm for the non-duplexing d(T<sub>4</sub>); this is interpreted to mean stacking in of the T residues (Haasnoot *et al.*, 1979).

The TxT mismatch in the system d(CA<sub>3</sub>XA<sub>3</sub>G):d(CT<sub>3</sub>YT<sub>3</sub>G) (Aboul-ela *et al.*, 1985) is one of the most destabilizing, lowering the T<sub>m</sub> by 15°C (at 0.4mM total strand concentration, 1.0M salt) from d(CA<sub>5</sub>G):d(CT<sub>6</sub>G) and by 23°C from d(CA<sub>7</sub>G):d(CT<sub>7</sub>G). Comparable destabilization is produced by an AxA mismatch (Aboul-ela *et al.*, 1985).

AGACU, as already mentioned, forms a duplex with T<sub>m</sub> 26°C, this being a destabilization of only 8°C from the parent duplex AGCU (Alkema *et al.*, 1982), compared to a 27°C reduction in T<sub>m</sub> upon the insertion of AxA into CCGG.

A dAx<sub>2</sub>dA pairing has been studied in the duplex

CAACTTGATATTAATA GTTGAACAATAATTAT 8	derived from	CAACTTGATATTAATA GTTGAACTATAATATA II
---	--------------	--

in this case, substituted for an dA•dT base pair, as opposed to being inserted.

At optical concentrations in 1M NaCl, the dAx<sub>2</sub>dA mismatch lowers the T<sub>m</sub> by 10°C.

According to calculations performed by these authors, the change in measured ΔG, ΔΔG, is insufficiently large to allow unstacking or bulging of these bases (Tibayenda *et al.*, 1984).

According to the results for GCUGC and CCAGG, the introduction of the UxU and Ax<sub>2</sub>A mismatches into GCGC and CCGG respectively produces comparable destabilization. In both these systems the mismatch is being inserted into a CG:CG core. In the AGXCU series, in which the mismatch is being introduced into a GC:GC core, the UxU mismatch is substantially more destabilizing than the Ax<sub>2</sub>A mismatch (Alkema *et al.*,



1982). In  $d(\text{CA}_3\text{XA}_3\text{G}):d(\text{CT}_3\text{YT}_3\text{G})$  (Aboul-ela *et al.*, 1985) TxT and AxA mismatches have similar destabilizing effect. A dTxdC mismatch—another pyrimidine-pyrimidine opposition—has been introduced into duplex II (above) at positions 9 and 12, by substitution of a normal base pair; in this case the dTxdC mismatches prove less destabilizing than the dAxdA, although they all have different neighbours (Tibayenda *et al.*, 1984). There is insufficient data to assess the effect of site of insertion and/or the nature of the neighbouring base pairs on mismatch stability, which will probably prove important, since the mismatched base pairs are generally stacked in, and therefore interacting with their neighbours.

#### 4.3.2 Effect of insertion of A•U base pairs upon stability

Systems containing a single A•U base pair inserted into the central core of the perfect duplexes GGCC, CCGG and GCGC may be prepared by mixing the complementary mismatch containing pentamers derived for each sequence:



The changes in average  $T_m$  for the low field aromatic protons for the tetramer to the pentamer mixtures are  $-0.5$ ,  $+7.5$  and  $+2.9^\circ\text{C}$  for GGCC, CCGG and GCGC respectively.

A second A•U base pair has been introduced into GCGC in GCAUGC, and into CCGG and CCAUGG. The  $T_m$  of GCAUGC increases by a further  $3^\circ\text{C}$  over CCAGG:CCUGG, to a total of  $5.9^\circ\text{C}$  over CCGG. CCAUGG had to be studied at a lower concentration through lack of availability of sample, and at half the concentration tabled a  $T_m$  decrease of  $4^\circ\text{C}$  from the pentamer mixture.

From the limited data available it would seem that insertion of an A·U base pair produces a greater increment in  $T_m$  when it interrupts a CG:CG core (CCGG  $\Delta T_m$  +7.5°C, GCGC  $\Delta T_m$  +2.9°C) than when it interrupts a GC:GC core (GGCC  $\Delta T_m$  -0.5°C). Since the GC:GC core is the strongest of the ten dinucleotide cores, a greater penalty is incurred by replacing one GC:GC core by two mixed G·C/A·U cores than by replacing one CG:CG core. The difference in  $\Delta G$  at the site of insertion may be estimated using the values for dinucleotide cores measured by Freier *et al.* (Freier *et al.*, 1985a,b):

GC	GAC	$\Delta G$ -1.07 kcal/mol	CG	CAG	$\Delta G$ -1.53 kcal/mol
CG	CUG		GC	GUC	
		$\Delta\Delta G$ -0.46 kcal/mol			

These figures are for oligonucleotides in 1.0M salt, but although the figures themselves will be somewhat changed in lower ionic strengths the order of stabilities will likely remain the same—reduction of the salt concentration from 1.0M to 0.1M did not affect the order of stabilities of the GC tetramers (See results in section 3.2 and 4.2).

Mixing experiments have been conducted on the members of the CAXUG and AGXCU series (in 1.0M salt). Insertion of a single A·U base pair into CAUG raises the  $T_m$  from 24°C to 28.8°C in CAAUG:CAUUG (Romaniuk *et al.*, 1979a) while the insertion of a single A·U base pair into AGCU raises the  $T_m$  from 34°C to 45°C in AGACU:AGUCU (Alkema *et al.*, 1982) a difference of 11°C. Insertion in AGCU is occurring into a GC:GC core, yet there is a substantial increase in  $T_m$  compared to the transition from GGCC to GGACC:GGUCC, which also involves insertion into a GC:GC core. This magnitude of difference in  $T_m$  increase could correspond to a small difference in increase of  $\Delta H$  over  $\Delta S$ ; moreover, from the form of the relation between  $T_m$  and the thermodynamic parameters, (Freier *et al.*, 1985b)  $T_m = \Delta H / (\Delta S + R \ln(C)) - 273.15$  °C, it is apparent that any increment in  $\Delta H$  and  $\Delta S$  will produce a greater change when the initial values of

these quantities are small—as in the less stable helices.

#### 4.3.3 Neighbour effects on the chemical shifts of the central adenosine residues in the A•U series

In a right-handed helix the base overlaps are such that the most influential neighbours as regards shielding of protons on the interior of the helix are the 3' neighbours. The central A•U base pair in the A•U series is flanked by two 3' guanine bases in two of the three pentamer duplexes, CCAGG:CCUGG and GCAGC:GCUGC, and by two 3' cytosine bases in the third, GGACC:GGUCC. Purine is more shielding than pyrimidine, and this difference is reflected in the magnitude of the upfield chemical shifts of the hydrogen-bonded imino proton UH3 and the nonexchangeable base proton AH2. An unshielded UH3 in an A•U base pair appears around 14.5ppm. In the spectrum of CCAGG:CCUGG UH3 appears at 13.10ppm, in that of GCAGC:GCUGC at 13.59ppm and in that of GGACC:GGUCC at 14.49ppm, upfield shifts of 1.4ppm, 0.9ppm and 0ppm respectively (Tables 4.7, 4.12 and 4.23). The nonexchangeable proton A(3)H2 is also in the middle of the stack and very sensitive to neighbour effects. In CCAGG:CCUGG, GCAGC:GCUGC and GGACC:GGUCC, A(3)H-2 moves upfield by >1.05ppm, >1.15ppm and 0.25ppm respectively (Figures 4.5, 4.9 and 4.19; Tables 4.6, 4.11 and 4.22).

#### 4.4. Conclusions

Insertion of a potential central UxU or AxA mismatch into the very stable tetranucleotide duplexes GGCC, CCGG and GCGC resulted in appreciable destabilization: of the six sequences studied, only two, GCUGC and CCAGG, formed

mismatch-containing duplexes, which tabled decreases of 27° and 28°C from the Tms of the parent duplexes. The other sequences either failed to duplex (GGACC) or formed slipped competing structures (GGUCC, CCUGG and GCAGC).

Stacking in of mismatched bases appears to be the preferred conformation for both UxU in GCUGC and AxA in CCAGG.

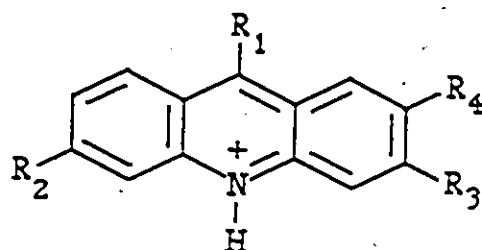
The introduction of a central A•U base pair (in duplexes formed by the mixing of the U-containing and A-containing duplexes) into the tetramer duplexes produces greater stabilization when the A•U base pair is inserted into the weaker CG:CG core in CCGG or GCGC than into the stronger GC:GC core in GGCC.

## 5. INTERCALATION OF PROFLAVINE INTO MISMATCH CONTAINING DUPLEXES.

### 5.1 Introduction

Intercalation in a biological system is understood to mean the insertion of a planar aromatic moiety between the base-pairs of a nucleic double helix. The model first proposed by Lerman (Lerman, 1961), to describe the interaction of ethidium cation with DNA, has the following features: preservation of the hydrogen bonding and covalent structure of the DNA; increase in length by  $3.4\text{\AA}$  per bound drug molecule, as adjacent base pairs at the binding site separate, with concomitant adjustments of backbone conformation; accommodation of the intercalator in the cavity, lying parallel to the base pairs (Lerman, 1961; Berman and Young, 1981; Wilson and Jones, 1982; Neidle and Abraham, 1985). The intercalating moiety must thus be planar, and of suitable dimensions—usually three fused aromatic rings. Common intercalators are pictured in Figure 5.1; these, in addition to fulfilling the size criterion, often are charged or have hydrogen bonding substituents capable of stabilizing interactions with the backbone or adjacent bases. Stacking, however, appears to contribute most to the interaction, given the similar enthalpies measured for different molecules (Berman and Young, 1981; Neidle and Abraham, 1985).

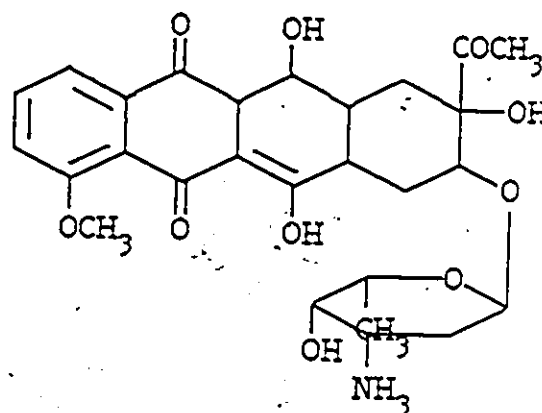
Studies of binding to natural DNA have established that intercalation is anticooperative, due to electrostatic repulsion between drug molecules, and that saturation is achieved at occupancy of half the available sites: "neighbour exclusion". Bisintercalator studies have verified this: for intercalation of both chromophores, the linker joining them must be  $10.2\text{\AA}$  or longer. Whether neighbour exclusion is



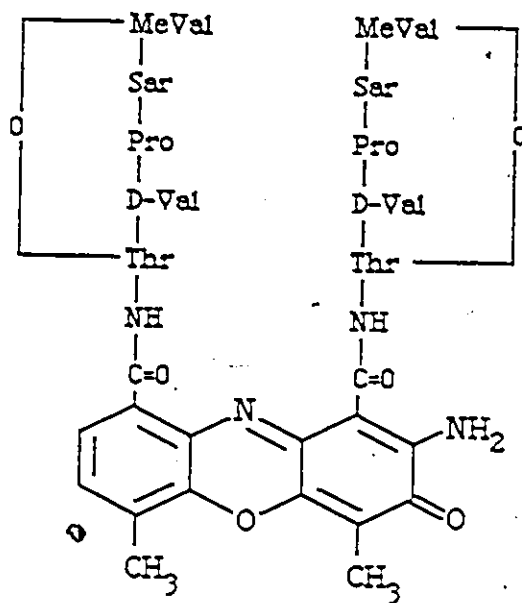
$R_1 = \text{NH}_2, R_2 = R_3 = R_4 = \text{H}$  9-aminoacridine

$R_1 = R_4 = \text{H}, R_2 = R_3 = \text{NH}_2$  proflavine

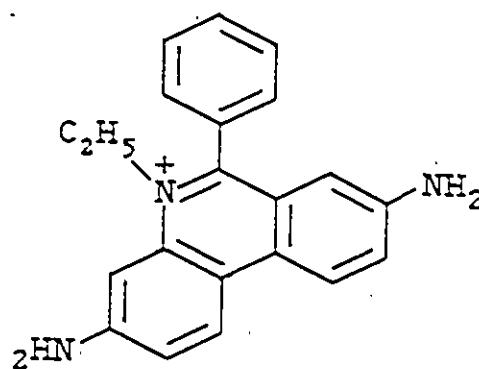
$R_1 = R_4 = \text{H}, R_2 = R_3 = \text{N}(\text{CH}_3)_2$  acridine orange



daunomycin



actinomycin D



ethidium

Figure 5.1: Some common intercalators

thermodynamic or structural in origin has not yet been established. A second, weaker, unsaturating binding mode has been identified, involving association of the intercalator with the backbone. Both modes, the intercalative and outside binding mode, are ionic strength dependent, decreasing with increasing salt concentration (Wilson and Jones, 1981; Berman and Young, 1981).

Association and dissociation kinetics are fast for all intercalators studied, with the exception of Actinomycin D and echinomycin, which dissociate slowly from DNA. Proflavine intercalation is preceded by a very fast non-intercalative binding, probably to the backbone, whereas ethidium intercalation does not involve a prior step, but is accomplished by direct ligand transfer (Wilson and Jones, 1982; Neidle and Abraham, 1985).

Both the length increase and the helix-unwinding can be accurately measured, the former through viscometry and sedimentation, and the latter using the relaxation of supercoiling of negatively closed circular DNA. The actual length increase is somewhat less than the model value in some cases, ranging from 2.0 - 3.4 Å (Berman and Young, 1981). The maximum unwinding measured for a monointercalator is 26°, for ethidium. There is no correlation of either parameter with biological activity (Berman and Young, 1981).

### 5.1.1 Biological effects of intercalation

DNA intercalating drugs have seen clinical trial and use as antiviral, antitrypanosomal and antineoplastic agents (Wilson and Jones, 1981), as they inhibit the growth, and promote the death and transformation of their target cells. Biological consequences of the binding of intercalators to DNA include the inhibition of

DNA-dependent enzymes, DNA strand breakage and frameshift mutagenesis (Wilson and Jones, 1981; Berman and Young, 1981). These effects--particularly the latter two--may be interrelated: Binding of the intercalator, with distortion in the DNA structure and alteration in supercoiling, leads to nicking of the DNA by topoisomerases (Ralph *et al.*, 1983) or repair enzymes. The intercalator subsequently inhibits the resealing of the nick--leading to strand breakage--or stabilizes transient looped out intermediates--resulting in erroneous repair, and frameshift mutation (Streisinger *et al.*, 1966). No clear correlation is apparent between strength of binding by intercalation and mutagenicity; certain non-intercalating drugs can also produce frameshift mutations (Berman and Young, 1981).

Drugs which inhibit gene expression by inhibition of DNA-dependent enzymes act at different stages of the process. Ethidium bromide and proflavine impede both replication and transcription, possibly by stabilizing the promoter region against the melting required for initiation, or by preventing recognition of the the region by DNA polymerase. Actinomycin-D prevents elongation during transcription, since its slow dissociation forces pausing of the RNA polymerase at the drug binding sites (Berman and Young, 1981).

Biological consequences of binding of intercalating drugs to native RNA have been less well studied. Binding of intercalators has long been known to alter the secondary structure of RNAs (Liebman *et al.*, 1977; Jones and Kearns, 1975). tRNA<sup>Phe</sup> is thought to contain at least one strong ethidium binding site. According to the results of solutions studies, it is an intercalative binding site between A•U5 and A•U6 of the acceptor stem (Wells and Cantor, 1977; Jones *et al.*, 1978; Goldfield *et al.*, 1983), but in the crystal structure ethidium is bound nonintercalatively in a cavity formed by the tertiary folding (Leibman *et al.*, 1977).



The iron (II) complex of methylpropidium-EDTA catalyses strand cleavage adjacent to its intercalation site and has been used to map intercalation sites in tRNA and in the self-splicing intervening sequence of *Tetrahymena* (Tanner and Cech, 1985b). The self-cyclization of the latter is inhibited by proflavine *in vivo* (Neilsen et al., 1984) and by ethidium bromide *in vitro* (Tanner and Cech, 1985a), and digestion with MPE-EDTA Fe(II) identifies three major intercalative binding sites, two of which are within regions previously implicated (by site-directed mutagenesis) in self-cyclization (Tanner and Cech, 1985b).

#### 5.1.2. Sequence preferences in intercalation

As data accumulated, the binding and unwinding of most intercalators were found to be composition or sequence depended to some degree. Actinomycin D seems to have the most stringent requirements: it does not bind to RNA or A.T rich DNA, and in model systems intercalates preferentially into the sequence d(GC):d(GC), with hydrogen bonding to guanosine NH<sub>2</sub> and the backbone (Patel, 1976; Reid et al., 1983). Further intricacies of sequence preference in natural DNA have been revealed by restriction enzyme inhibition studies (in which the binding of the drug is assayed by the immunity of the sites to digestion) and nick-translation inhibition (in which the progress of a nick around a double stranded plasmid incubated with DNA polymerase is arrested at binding sites). Some, but not all, GC rich, d(GC):d(GC) containing recognition sites are protected, indicating long-range effects which are, as yet, unclear (Neidle and Abraham, 1985).

Simple intercalators, lacking extensively binding side chains, such as proflavine and ethidium cation, exhibit a slight preference for intercalation into

pyrimidine(3'-5')purine sites in model sequences (Berman and Young, 1981; Neidle and Abraham, 1985). Most of the oligonucleotide-intercalator complexes crystallized have involved self-complementary pyr(3'-5')pur dinucleotides (Neidle, 1981; Sobell *et al.*, 1982); in NMR solution studies of the interaction of 9-aminoacridine and ethidium bromide with dinucleotides, the strongest binding was observed to r(CG) and d(CG) (Young and Kallenbach, 1981; Krugh and Reinhardt, 1975; Krugh *et al.*, 1975; Reinhardt and Krugh, 1978); in NMR and optical studies of deoxytetranucleotides with ethidium and proflavine, the stoichiometry and strength of binding can be roughly correlated with the availability of d(CG):d(CG) sites (Patel and Camuel, 1976; Kastrup *et al.*, 1978; Patel and Camuel, 1977). Sequence preferences of these intercalators have been largely unstudied by biochemical techniques.

Associated theoretical studies have addressed the reason for this preference (Pack and Loew, 1978; Nuss *et al.*, 1978; Ornstein and Rein, 1979). The consensus seems to be that it arises from the difference in energy required to open the site for intercalation, and not from interaction with the intercalator, but results of different methods disagree as to the major contributor to the difference—according to empirical potential function calculations, the largest difference arises in stacking energies of the two sequences (Nuss *et al.*, 1978), according to molecular orbital studies, it arises from energy of unwinding (Ornstein and Rein, 1979).

### 5.1.3. Detail of the intercalation site: 1. The intercalator orientation

The simple intercalator model proposes that the intercalator should lie parallel with the base pairs, perpendicular to the helix axis (Lerman, 1961; Berman and

Young, 1981). This has proven to be an oversimplification. Transient electric dichroism studies reveal twisting and buckling of base pairs around the intercalation site, and a nonparallel drug (Hogan *et al.*, 1979).

Crystal structures of ethidium and acridines with dinucleotides offer conformations more in accord with the classical model, with maximum possible overlap between base pairs and drug (Sobell, 1980; Neidle, 1981; Sobell *et al.*, 1982; Neidle, 1983; Neidle and Abraham, 1985). In the only crystal structure so far obtained for an intercalator complex with a sequence larger than a dimer, the daunomycin-d(CGATCG) complex, daunomycin inserts into the d(CG):d(CG) sequence, with the chromophore at a near perpendicular orientation to the long axis, a conformation probably imposed by the binding of sugar substituent in the minor groove (Quigley *et al.*, 1980). Chemical shift changes for the drug resonances in solution NMR studies have been used to describe solution geometry for daunomycin (Nuss *et al.*, 1980; Patel, 1979), proflavine (Patel and Camuel, 1977; Patel, 1977b) and 9-aminoacridine complexes (Young and Kallenbach, 1981; Reuben *et al.*, 1978; Young and Kallenbach, 1980). The solution chemical shifts of daunomycin are small, consistent with the observed crystallographic orientation (Nuss *et al.*, 1980); those for proflavine are uniformly large (Patel and Camuel, 1977; Patel 1977), also consistent with the crystal structures, with the long axis lying along the base pairs (Neidle, 1981; Shieh *et al.*, 1980; Aggarwal *et al.*, 1984). 9-Aminoacridine also shows large chemical shifts in the 2:1 complexes with dinucleotides of alternating sequences (Young and Kallenbach, 1981), while in the crystal structure of its complex with iodoCpG, 9-aminoacridine has two distinct configurations, one symmetric, with the 9-amino group protruding into the minor groove, and one asymmetric, with the amino group in the major groove (Sakore *et al.*, 1979).

The direction of intercalation becomes more significant for intercalation with bulky side chains, such as ethidium or daunomycin; in these, the substituent resides in the minor groove (Quigley *et al.*, 1980; Jain *et al.*, 1977; Tsai *et al.*, 1977). The form of helix, whether A or B, would be expected to affect binding, since the groove dimensions in the two helices are quite different.

#### 5.1.4. Detail of the intercalation site 2. The nucleic acid conformation

Accommodation of the intercalator involves the opening of two base pairs to twice their normal separation. According to the dinucleotide crystal structures, this is accomplished primarily by an increase in the O5 - C5 torsion angle, which increases the base separation, and in the 3' glycosidic angle, which preserves the parallel orientation of the base pairs (Neidle, 1981; Neidle, 1983; Neidle and Abraham, 1985). Model building studies verify that these changes are sufficient (Neidle and Abraham, 1985).

The sugar geometries have been the subject of some controversy. In most crystals C3'endo(3'-5')C2'endo sugar puckering is observed, proflavine being the major exception (Neidle, 1981; Neidle, 1983). Sobell and coworkers have proposed that mixed sugar puckering is a common feature of intercalation, and that hydrogen bonding between the amino groups and the backbone are responsible for its absence in the proflavine complexes (Sobell, 1980; Sobell *et al.*, 1982), while Berman and coworkers argue that the changes in backbone torsion angles are sufficient to open the site adequately, and that, out of a continuum of stereochemically plausible geometries, the exact geometry depends upon the intercalator (Berman *et al.*, 1978; Sheih *et al.*, 1980; Berman and Young, 1981).

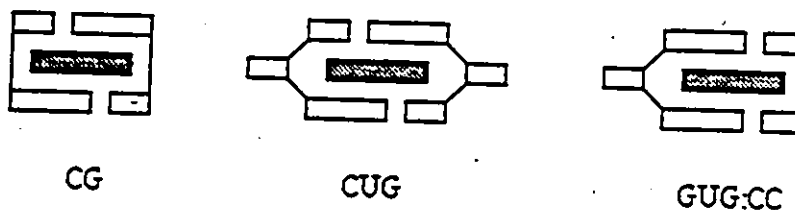
Model building studies involving insertion of the crystallographically derived dinucleotide geometry into a polynucleotide to generate a plausible intercalated polynucleotide structure, show an altered geometry at adjacent sites such as might be responsible for neighbour exclusion (Berman and Young, 1981).

#### 5.1.5. Intercalation into imperfect duplexes

According to the Streisinger hypothesis, intercalator-induced frameshift mutagenesis occurs through the stabilization of looped out intermediates by interaction with the intercalator, and subsequent "fixing" of these looped out structures by repair enzymes, with insertion or deletion of a base (Streisinger et al., 1966). The results of studies of the interaction of ethidium with model systems containing unpaired bases indicate that intercalation at or near the mismatch site stabilizes the structure and is preferred.

The binding affinity of ethidium to the mismatch IxA sites in poly(I):poly(C,A<sub>x</sub>) (x = mole fraction of A, 0.1, 0.2, 0.3) is estimated by Scatchard analysis to be twenty times that for binding to the I-C sites (Hefgott and Kallenbach, 1979).

Variable temperature <sup>1</sup>H-NMR studies have been published on the behaviour of 1:1 mixtures of ethidium bromide and CG, CUG and GUG:CC (Lee and Tinoco, 1978). Sigmoidal transitions and large upfield shifts are observed for the phenanthridium ring protons with decreasing temperature, and these are interpreted to mean that the ethidium is intercalated between two G-C base pairs, with looping out of the uridine residues.



C3' endo(5'-3')C2' endo preferences are observed for the dimers involved in the intercalation complexes, but the sugar conformations in the trimers are not substantially different in the intercalated complex from their values in the absence of intercalator (Lee and Tinoco, 1978).

A proportion of 1:1 complexes of 9-aminoacridine with pur-pur:pyr-pyr deoxydinucleotides have the intercalator stacked crosswise between the purine residues (Young and Kallenbach, 1981), with room on the other strand to accommodate a mismatched base stacked in to the duplex. This has been advanced as a possible model for frameshift mutagenesis (Young and Kallenbach, 1981).

Ethidium has been found to intercalate equally well into all sites within the deoxyoligoribonucleotide duplex formed by  $(CA_6G):(CT_6G)$  (Nelson and Tinoco, 1984). Insertion of a looped out cytidine into the centre of this duplex, in  $(CA_3CA_3G):(CT_6G)$  (Morden *et al.*, 1984) results in the production of two strong binding sites, believed to be those on either side of the mismatch site (Nelson and Tinoco, 1985). Fitting of the model was equally good if there was assumed to be no intercalation whatsoever into the mismatch site itself. The stabilization from binding to both sites effectively cancels the destabilization due to the mismatch (Nelson and Tinoco, 1985).

Strong intercalation sites around a single bulged adenosine have been observed in MPE-EDTA Fe(II) footprinting studies of 16S RNA (Kean *et al.*, 1985) and of the self-splicing intervening sequence of *Tetrahymena* (Tanner and Cech, 1985a,b).

### 5.1.6 Intercalation of proflavine into GCUGC and CCAGG

Mismatch sites appear to have greater affinity for intercalators than non-mismatch sites, but studies of intercalation into short oligonucleotide duplexes containing mismatched bases have been few. Of the sequences studied in section 4, two, CCAGG and GCUGC, form mismatch-containing duplexes, containing an AxA and a UxU mismatch respectively. GCUGC also incorporates the CUG sequence studied in mixtures with ethidium by Lee and Tinoco (1978), but, unlike CUG, GCUGC duplexes in the absence of intercalator, permitting comparison of duplex behaviour before and after addition of intercalator. The behaviour of the CAG site in CCAGG, with its central AxA mismatch, may be compared with the CUG site in GCUGC, with its central UxU mismatch.

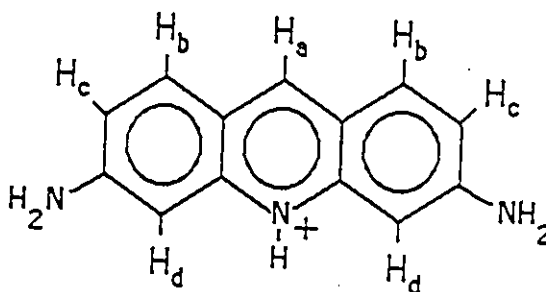
Studies of the intercalation of proflavine into CCAGG and GCUGC and their respective parent duplexes, CCGG and GCGC, were undertaken to address the following questions:

1. Was there greater stabilization of mismatch-containing duplexes upon addition of intercalator, compared to non-mismatch containing duplexes?
2. Was intercalation into a mismatch site preferred to intercalation into an intact site?
3. Was intercalation into the mismatch site accompanied by the looping out of the mismatched bases to accommodate the intercalator, or by other striking conformational changes?
4. Was there a difference between intercalation into a UxU and into an AxA mismatch, given the different stacking properties of the two bases, uracil and adenine?

## 5.2. Results

### 5.2.1. Intercalation of proflavine into GCGC and GCUGC

Variable temperature series of spectra were obtained for GCGC in  $D_2O$  (2.4mM, 0.1M salt buffer) with proflavine at single strand:intercalator ratios of 15:1, 4:1, 2:1, 1:1 and 1:2 (Tables 5.1 to 5.5), and of GCUGC with proflavine at ratios of 20:1, 6:1, 4:1, 2:1 1:1 and 1:2 (Tables 5.6 to 5.11). Assignments of the sequence resonances were as described in Sections 3.1, 4.2.2 and 4.2.3. The proflavine signals at high temperature were assigned as follows: pro-a, singlet, unit area, downfield from the aromatic base area at  $\delta = 8.5$  ppm; pro-b, doublet,  $J = 9$  Hz, area of two, starting overlapped with the pyrimidine H6 protons at  $\delta = 7.6$  ppm, ortho coupled to pro-c, doublet of doublets,  $J = 9, 1$  Hz, area of two, at  $\delta = 7$ . Pro-c is meta coupled to pro-d,  $J = 1$  Hz, a doublet at  $\delta = 6.3$  ppm. Single strand:drug ratios were checked by integration of the proflavine resonances against the CH-6 and UH-6 signals in the 70°C spectra.



Addition of the intercalator to GCGC produced an increase in  $T_m$  by a maximum of 5.3°C at 2:1 single strand: drug (Table 5.12; Figure 5.2). Addition of the intercalator to GCUGC produced a  $T_m$  increase of 15.6°C at 2:1 single strand:drug (Table 5.13; Figure 5.3). Upon addition of further proflavine the  $T_m$  of both GCGC and GCUGC declined, GCGC more



Table 5.1: Chemical shift assignments and coupling constants for the base and anomeric protons of GCGC and the aromatic protons of proflavine for GCGC, 2.4 mM, with proflavine, 15:1 single strand: proflavine, in 0.1M salt, over the temperature range 70 to 20°C. NSB denotes no sigmoidal behaviour.

Temp.	Chemical shift (ppm)								T <sub>m</sub>
	70.2°	60.1°	55.5°	50.1°	45.7°	40.0°	30.3°	20.2°	
Proton									
G(1)H-8	7.932	7.950	7.963	7.980	7.990	7.991	7.983	7.949	NSB
C(2)H-6	7.716	7.717	7.720	7.729	7.742	7.758	7.787	7.812	42.5
G(3)H-8	7.932	7.890	7.851	7.802	7.745	7.692	7.628	7.587	48.5
C(4)H-6	7.793	7.761	7.731	7.693	7.647	7.605	7.554	7.519	49.4
C(2)H-5	5.830	5.740	5.670	5.590	5.503	5.422	5.336	5.269	50.6
C(4)H-5	5.896	5.807	5.734	5.643	5.539	5.450	5.331	5.262	49.4
G(1)H-1'	5.818	5.798	5.787	5.776	5.765	5.765	5.753	5.748	NSB
C(2)H-1'	5.889	5.848	5.815	5.776	5.735	5.696	5.647	5.613	50.4
G(3)H-1'	5.825	5.812	5.800	5.784	5.765	5.735	5.692	5.641	46.5
C(4)H-1'	5.904	5.881	5.861	5.836	5.804	5.772	5.729	5.691	43.5
							Average T <sub>m</sub>		48.3
pro-a	8.565	8.438	8.376	8.331	8.293	8.267	8.244	8.212	62.0
pro-b	7.729	7.597	7.525	7.467	7.419	7.384	7.352	7.302	65.2
pro-c	6.947	6.797	6.705	6.626	6.562	6.503	6.447	6.382	59.9
pro-d	6.670	6.528	6.440	6.373	6.312	6.256	6.197	6.135	60.7
							Average T <sub>m</sub>		61.9
							Coupling constant (Hz)		
G(1)H-1'	4.6	2.3	2.7	2.6	1.8	1.3	--	--	
C(2)H-1'	4.4	3.9	3.4	2.6	1.4	--	--	--	
G(3)H-1'	4.8	4.7	4.0	2.4	1.8	1.6	--	--	
C(4)H-1'	3.6	3.4	3.2	2.8	2.3	2.0	--	--	

Table 5.2: Chemical shift assignments and coupling constants for the base and anomeric protons of GCGC and the aromatic protons of proflavine for GCGC, 2.4 mM, with proflavine, 4:1 single strand: proflavine, in 0.1M salt, over the temperature range 70 to 20°C.

Temp	Chemical shift (ppm)								T <sub>m</sub>
	70.0°	60.4°	55.2°	50.2°	45.3°	40.2°	30.3°	20.2°	
Proton									
G(1)H-8	7.925	7.943	7.957	7.971	7.979	7.981	7.960	7.927	NSB
C(2)H-6	7.715	7.710	7.710	7.716	7.726	7.739	7.763	7.779	38.0
G(3)H-8	7.920	7.865	7.822	7.771	7.716	7.668	7.594	7.555	49.1
C(4)H-6	7.786	7.744	7.710	7.677	7.623	7.582	7.522	7.489	46.4
C(2)H-5	5.819	5.709	5.630	5.547	5.463	5.391	5.284	5.221	52.4
C(4)H-5	5.887	5.780	5.697	5.599	5.496	5.411	5.284	5.221	52.2
G(1)H-1'	5.811	5.779	5.764	5.750	5.737	5.730	5.718	5.713	NSB
C(2)H-1'	5.883	5.832	5.792	5.750	5.707	5.668	5.610	5.574	51.7
G(3)H-1'	5.815	5.794	5.778	5.761	5.737	5.709	5.649	5.594	37.6
C(4)H-1'	5.900	5.872	5.849	5.822	5.790	5.758	5.704	5.661	45.7
								Average T <sub>m</sub>	49.6
pro-a	8.508	8.381	8.321	8.277	8.843	8.219	8.192	8.158	66.7
pro-d	7.684	7.552	7.483	7.427	7.383	7.348	7.295	7.263	64.9
pro-c	6.916	9.762	7.675	6.599	6.535	6.483	6.401	6.338	62.2
pro-d	6.622	6.473	6.393	6.324	6.266	6.216	6.132	6.061	63.7
								Average T <sub>m</sub>	64.4
									Coupling constants (Hz)
G(1)H-1'	31	3.4	3.3	2.5	2.4	1.5	--	--	
C(2)H-1'	5.1	3.7	3.1	2.5	1.7	--	--	--	
G(3)H-1'	5.0	4.0	3.8	2.9	2.4	1.8	--	--	
C(4)H-1'	3.7	3.4	3.2	2.9	2.5	2.1	--	--	

Table 5.3: Chemical shift assignments and coupling constants for the base and anomeric protons of GCGC and the aromatic protons of proflavine for GCGC, 2.4 mM, with proflavine, 2:1 single strand: proflavine, in 0.1M salt, over the temperature range 70 to 20°C.

Temp	Chemical shift (ppm)									T <sub>m</sub>
	68.5°	59.4°	50.1°	45.1°	40.0°	35.0°	30.0°	20.7°	10.7°	
Proton										
G(1)H-8	7.992	7.930	7.955	7.964	7.967	7.965	7.958	7.936	7.909	NSB
C(2)H-6	7.708	7.696	7.692	7.698	7.702	7.711	7.717	7.737	7.740	NSB
G(3)H-8	7.895	7.827	7.734	7.684	7.642	7.606	7.578	7.537	7.500	54.9
C(4)H-6	5.768	7.716	7.635	7.591	7.553	7.519	7.489	7.456	7.431	53.0
C(2)H-5	5.791	5.655	5.493	5.420	5.357	5.303	5.263	5.193	5.144	58.2
C(4)H-5	5.861	5.731	5.549	5.452	5.366	5.287	5.222	5.160	5.144	53.2
G(1)H-1'	5.786	5.748	5.718	5.692	5.679	5.672	5.667	5.656	5.664	NSB
C(2)H-1'	5.871	5.802	5.805	5.774	5.743	5.710	5.683	5.641	5.596	46.4
G(3)H-1'	5.796	5.764	5.724	5.701	5.679	5.652	5.629	5.585	5.545	NSB
C(4)H-1'	5.887	5.857	5.710	5.672	5.634	5.602	5.576	5.538	5.507	57.0
								Average T <sub>m</sub>		53.8
pro-a	8.444	8.318	8.223	8.188	8.160	8.137	8.117	8.081	8.046	65.3
pro-b	7.636	7.506	7.392	7.349	7.314	7.283	7.258	7.217	7.176	63.9
pro-c	6.880	6.732	6.587	6.528	-	6.438	6.401	6.254	6.289	64.2
pro-d	6.567	6.418	6.281	6.224	6.174	6.127	6.088	6.008	5.931	63.1
								Average T <sub>m</sub>		64.1
	Coupling constant (Hz)									
G(1)H-1'	4.8	4.0	-	2.6	1.5	-	-	-	-	
C(2)H-1'	4.8	3.4	3.0	2.7	2.4	-	-	-	-	
G(3)H-1'	4.6	4.0	2.9	2.1	1.5	-	-	-	-	
C(4)H-1'	4.1	3.4	4.2	1.4	<1	-	-	-	-	

Table 5.4: Chemical shift assignments and coupling constants for the base and anomeric protons of GCGC and the aromatic protons of proflavine for GCGC single strand concentration, 2.4 mM, with proflavine, 1:1 single strand: proflavine, in 0.1M salt, over the temperature range 80 to 20°C.

Proton	Chemical shift (ppm)										T <sub>m</sub>		
	79.5°	69.4°	65.3°	59.4°	55.0°	50.0°	45.3°	40.3°	35.4°	30.0°		20.6°	
G(1)H-8	7.890	7.893	7.911	7.933	7.947	7.962	7.976	7.982	7.985	7.978	7.953	NSB	
C(2)H-6	7.713	7.705	7.702	7.704	7.703	7.705	7.712	7.721	7.733	7.743	7.771	NSB	
G(3)H-8	7.911	7.877	7.870	7.853	7.816	7.776	7.727	7.680	7.637	7.603	7.568	45.9	
C(4)H-6	7.786	7.763	7.751	7.736	7.703	7.672	7.630	7.589	7.552	7.524	7.477	46.1	
C(2)H-5	5.857	5.785	5.745	5.694	5.623	5.552	5.475	5.404	5.344	5.298	5.233	49.6	
C(4)H-5	5.915	5.884	5.818	5.771	5.695	5.614	5.521	5.422	5.344	5.284	5.196	48.3	
G(1)H-1'	5.800	5.771	5.766	5.763	5.751	5.738	5.718	5.713	5.70	5.705	5.695	NSB	
C(2)H-1'	5.901	5.868	5.847	5.823	5.788	5.753	5.723	5.671	5.637	5.609	5.569	46.6	
G(3)H-1'	5.800	5.779	5.781	5.779	5.767	5.753	5.735	5.713	5.691	5.665	5.616	NSB	
C(4)H-1'	5.901	5.884	5.878	5.867	5.848	5.825	5.797	5.767	5.736	5.705	5.666	41.9	
												Average T <sub>m</sub>	46.4
pro-a	8.491	8.401	8.405	8.399	8.351	8.307	8.265	8.234	8.206	8.183	8.183	--	
pro-b	7.691	7.611	7.598	7.578	7.520	7.467	7.417	7.376	7.342	7.313	7.267	49.6	
pro-c	6.945	6.868	6.845	6.806	6.728	6.658	6.591	6.534	6.481	6.445	6.377	49.1	
pro-d	6.628	6.535	6.516	6.499	6.406	6.369	6.305	6.246	6.199	6.154	6.075	50.7	
												Average T <sub>m</sub>	49.8
	Coupling constants (Hz)												
G(1)H-1'	4.6	4.8	2.9	4.1	4.3	4.1	2.6	2.1	--	--			
C(2)H-1'	4.3	4.3	4.3	3.7	3.2	3.2	3.1	1.2	--	--			
G(3)H-1'	4.6	3.6	4.5	3.9	4.0	3.2	2.8	2.1	--	--			
C(4)H-1'	4.4	4.8	3.7	3.4	3.2	2.9	2.5	2.3	1.3	--			

Table 5.5: Chemical shift assignments and coupling constants for the base and anomeric protons of GCGC and the aromatic protons of proflavine for GCGC single strand concentration, 2.4 mM, with proflavine, 1:2 single strand: proflavine, in 0.1M salt, over the temperature range 80 to 20°C.

Temp	Chemical shift (ppm)							T <sub>m</sub>
	80.1°	70.1°	60.5°	50.8°	40.4°	31.4°	21.1°	
Proton								
G(1)H-8	7.917	7.910	7.910	7.937	7.971	7.993	7.988	46.4
C(2)H-6	7.718	7.716	7.709	7.722	7.708	7.719	7.722	NSB
G(3)H-8	7.949	7.924	7.886	7.832	7.693	7.550	7.456	35.8
C(4)H-6	7.802	7.794	7.765	7.705	7.601	7.507	7.456	40.3
C(2)H-5	5.873	5.845	5.766	5.651	5.495	5.387	5.175	48.1
C(4)H-5	5.935	5.903	5.835	5.723	5.526	5.354	5.256	40.3
G(1)H-1'	5.822	5.811	5.773	5.768	5.721	5.702	5.684	NSB
C(2)H-1'	5.909	5.897	5.869	5.803	5.707	5.640	5.575	46.2
G(3)H-1'	5.828	5.803	5.789	5.746	5.733	5.693	5.659	NSB
C(4)H-1'	5.909	5.900	5.879	5.851	5.780	5.748	5.695	45.5
							Average T <sub>m</sub>	42.7
pro-a	8.713	8.639	8.568	8.533	8.442	8.351	8.246	31.3
pro-b	7.852	7.799	7.741	7.699	7.598	7.505	7.400	33.1
pro-c	7.058	7.021	6.971	6.916	6.780	6.688	6.554	43.5
pro-d	6.802	6.731	6.658	6.601	6.491	6.370	6.218	30.2
							Average T <sub>m</sub>	34.5
	Coupling constants (Hz)							
G(1)H-1'	3.3	4.7	4.0	4.3	2.6			
C(2)H-1'	5.5	4.8	5.1	2.8	2.4			
G(3)H-1'	3.0	4.7	4.2	3.1	2.7			
C(4)H-1'	5.4	3.2	4.0	3.5	2.7	2.2		

Table 5.6: Chemical shift assignments and coupling constants for the base and anomeric protons of GCUGC and of the aromatic protons of proflavine for GCUGC, 2.5 mM, with proflavine, 20:1 single strand: proflavine, in 0.1M salt buffer over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)									T <sub>m</sub>
	70.0°	60.1°	50.0°	40.0°	30.1°	25.2°	20.2°	15.1°	10.1°	
G(1)H-8	7.943	7.948	7.955	7.965	7.981	7.983	7.978	7.950	7.928	NSB
C(2)H-6	7.780	7.778	7.771	7.756	7.739	7.736	7.736	7.727	7.727	NSB
U(3)H-6	7.733	7.735	7.738	7.739	7.739	7.736	7.736	7.746	7.761	NSB
G(4)H-8	7.972	7.971	7.967	7.953	7.928	7.896	7.867	7.833	7.806	19.3
C(5)H-6	7.819	7.817	7.811	7.792	7.748	7.710	7.659	7.614	7.589	20.8
C(2)H-5	5.855	5.854	5.808	5.730	5.628	5.523	5.427	5.324	5.267	20.8
U(3)H-5	5.817	5.802	5.784	5.754	5.712	5.672	5.636	5.596	5.588	25.0
C(5)H-5	5.945	5.921	5.889	5.828	5.712	5.656	5.546	5.443	5.400	21.1
G(1)H-1'	5.833	5.828	5.819	5.802	5.765	5.721	5.666	5.617	5.588	20.3
C(2)H-1'	5.909	5.913	5.900	5.884	5.860	5.832	5.801	5.763	5.739	19.0
U(3)H-1'	5.835	5.848	5.839	5.827	5.815	-	5.757	5.719	-	NSB
G(4)H-1'	5.855	5.848	5.839	5.812	5.787	5.774	5.757	5.763	5.763	NSB
C(5)H-1'	5.924	5.913	5.894	5.860	5.808	5.774	5.735	5.686	5.658	20.3
										Average T <sub>m</sub>
pro-a	8.559	8.487	8.390	8.268	8.194	8.159	-	-	-	46.5
pro-b	-	-	7.584	7.450	7.363	7.320	-	-	-	-
pro-c	6.967	6.917	6.819	6.664	6.552	6.467	6.454	6.436	6.407	42.1
pro-d	6.680	6.615	6.511	6.365	6.253	6.205	6.171	6.138	6.111	46.0
										Average T <sub>m</sub>
										45.0
										Coupling constant (Hz)
G(1)H-1'	6.1	5.1	5.3	5.4	4.4	-	-	-	-	
C(2)H-1'	4.1	3.5	2.2	3.4	3.1	2.3	-	-	-	
U(3)H-1'	3.9	4.7	4.4	4.1	-	-	-	-	-	
G(4)H-1'	5.3	4.7	4.4	3.8	3.8	-	-	-	-	
C(5)H-1'	6.2	3.5	4.8	4.0	3.4	-	-	-	-	

Table 5.7: Chemical shift assignments and coupling constants for the base and anomeric protons of GCUGC and of the aromatic protons of proflavine for GCUGC, 2.5 mM, with proflavine, 6:1 single strand: proflavine, in 0.1M salt buffer over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)										T <sub>m</sub>	
	70.0°	60.2°	50.2°	40.2°	34.8°	30.0°	25.0°	20.2°	15.1°	10.1°		
G(1)H-8	7.935	7.939	7.946	7.956	7.964	7.966	7.965	7.954	7.937	7.911	NSB	
C(2)H-6	7.778	7.773	7.763	7.745	7.735	7.724	7.716	7.711	7.720	7.720	NSB	
U(3)H-6	7.732	7.734	7.736	7.736	7.735	7.724	7.716	7.711	7.720	7.720	NSB	
G(4)H-8	7.964	7.961	7.952	7.927	7.913	7.882	7.856	7.831	7.813	7.781	29.1	
C(5)H-6	7.815	7.812	7.800	7.768	7.745	7.702	7.659	7.590	7.591	7.616	25.8	
C(2)H-5	5.853	5.847	5.788	5.689	5.628	5.540	5.456	5.370	5.289	5.246	24.8	
U(3)H-5	5.814	5.799	5.779	5.745	5.719	5.684	5.648	5.616	5.591	—	31.3	
C(5)H-5	5.941	5.915	5.871	5.791	5.735	5.640	5.560	5.466	5.405	5.371	26.1	
G(1)H-1'	5.842	5.837	5.826	5.799	5.775	5.739	—	5.654	5.616	5.603	23.5	
C(2)H-1'	5.907	5.907	5.893	5.872	5.858	5.835	5.809	5.781	5.753	5.727	20.0	
U(3)H-1'	5.829	5.827	5.818	5.815	5.795	5.775	5.754	—	5.730	5.727	27.3	
G(4)H-1'	5.850	5.837	5.818	5.782	5.766	5.752	—	5.694	5.671	5.644	NSB	
C(5)H-1'	5.920	5.910	5.891	5.846	5.821	5.784	5.740	5.707	5.671	5.644	26.5	
										Average T <sub>m</sub>	25.4	
pro-a	8.502	8.421	8.316	8.208	8.178	8.139	8.121	8.111	8.100	8.097	55.0	
pro-b	7.689	7.624	7.521	7.397	7.354	7.308	7.287	7.265	7.252	7.265	51.8	
pro-c	6.939	6.876	6.767	6.656	6.555	6.493	6.458	6.430	6.415	6.362	41.1	
pro-d	6.634	6.557	6.443	6.298	6.248	6.193	6.163	6.134	6.115	6.090	49.3	
										Average T <sub>m</sub>	49.3	
				Coupling constant (Hz)								
G(1)H-1'	5.4	4.7	4.4	3.6	4.9	3.6	—	—	—	—		
C(2)H-1'	4.5	3.6	3.6	3.4	3.1	2.8	2.1	—	—	—		
U(3)H-1'	4.5	5.7	4.9	3.9	3.7	2.1	—	—	—	—		
G(4)H-1'	3.7	4.7	4.9	3.3	3.6	2.9	—	—	—	—		
C(5)H-1'	6.7	4.9	4.3	3.8	3.4	2.6	—	—	—	—		

Table 5.8: Chemical shift assignments and coupling constants for the base and anomeric protons of GCUGC and of the aromatic protons of proflavine for GCUGC, 2.5 mM, with proflavine, 4:1 single strand: proflavine, in 0.1M salt buffer over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)									T <sub>m</sub>	
	70.8°	62.2°	50.7°	40.7°	36.1°	30.0°	25.0°	20.4°	10.0°		
G(1)H-8	7.929	7.931	7.938	7.947	7.953	7.958	7.958	7.949	7.913	NSB	
C(2)H-6	7.775	7.770	7.755	7.734	7.721	7.708	7.703	7.698	7.705	NSB	
U(3)H-6	7.730	7.733	7.736	7.734	7.721	7.719	7.710	7.713	7.705	NSB	
G(4)H-8	7.959	7.955	7.938	7.911	7.889	7.860	7.833	7.813	7.779	30.5	
C(5)H-6	7.812	7.807	7.787	7.752	7.729	7.678	7.636	7.590	7.558	25.2	
C(2)H-5	5.855	5.849	5.769	5.668	5.598	5.509	5.432	5.368	5.235	31.8	
U(3)H-5	5.813	5.800	5.776	5.723	5.700	5.680	5.642	5.612	5.591	35.1	
C(5)H-5	5.939	5.914	5.853	5.767	5.716	5.609	5.518	5.445	5.308	26.5	
G(1)H-1'	5.832	5.828	5.820	5.814	5.783	5.745	5.717	5.689	5.647	28.0	
C(2)H-1'	5.911	5.907	5.888	5.865	5.848	5.826	5.803	5.776	5.730	21.4	
U(3)H-1'	5.839	5.831	5.814	5.786	5.767	5.733	5.709	5.682	5.647	25.8	
G(4)H-1'	5.841	5.828	5.799	5.766	5.739	5.709	5.694	5.655	5.607	26.8	
C(5)H-1'	5.939	5.911	5.881	5.841	5.812	5.772	5.737	5.706	5.670	29.0	
										Average T <sub>m</sub>	
pro-a	8.480	8.402	8.265	8.177	8.142	8.114	8.098	8.093	8.075	58.4	
pro-b	7.679	7.613	7.482	7.375	7.330	7.291	7.273	7.252	7.226	54.6	
pro-c	6.930	6.872	6.736	6.606	6.542	6.484	6.449	6.425	6.392	48.5	
pro-d	6.622	6.544	6.397	6.277	6.223	6.174	6.143	6.121	6.076	54.1	
										Average T <sub>m</sub>	
										53.9	
				Coupling constant (Hz)							
G(1)H-1'	5.9	5.9	5.0	2.2	1.8	2.5	2.0	--	--		
C(2)H-1'	6.6	3.7	3.3	3.3	3.3	2.9	2.1	--	--		
U(3)H-1'	4.5	4.7	4.2	1.7	--	--	--	--	--		
G(4)H-1'	5.8	5.9	3.1	3.5	3.2	2.3	--	--	--		
C(5)H-1'	4.9	5.8	4.6	3.8	3.3	2.7	<1	--	--		



Table 5.9: Chemical shift assignments and coupling constants for the base and anomeric protons of GCUGC and of the aromatic protons of proflavine for GCUGC, 2.5 mM, with proflavine, 2:1 single strand: proflavine, in 0.1M salt buffer over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)									T <sub>m</sub>	
	68.5°	59.4°	50.1°	40.1°	35.0°	30.0°	25.1°	20.7°	10.7°		
G(1)H-8	7.909	7.914	7.923	7.928	7.930	7.929	7.928	7.927	7.895	NSB	
C(2)H-6	7.769	7.760	7.739	7.707	7.684	7.669	7.659	—	—	38.4	
U(3)H-6	7.730	7.733	7.737	7.730	7.721	7.705	7.693	7.676	7.665	NSB	
G(4)H-8	7.939	7.932	7.914	7.867	7.835	7.805	7.778	7.763	7.738	38.5	
C(5)H-6	7.801	7.791	7.769	7.708	7.662	7.619	7.576	7.539	7.476	33.1	
C(2)H-5	5.836	5.824	5.738	5.596	5.507	5.433	5.364	5.315	5.262	39.1	
U(3)H-5	5.804	5.792	5.771	5.748	5.698	5.644	5.609	—	—	33.1	
C(5)H-5	5.924	5.887	5.824	5.693	5.600	5.517	5.422	5.414	—	38.4	
G(1)H-1'	5.831	5.825	5.821	5.795	5.770	5.737	5.699	5.672	5.622	32.8	
C(2)H-1'	5.890	5.894	5.871	5.839	5.814	5.791	5.763	5.749	5.711	32.8	
U(3)H-1'	5.819	5.807	5.791	5.748	5.718	5.683	5.659	5.631	5.587	33.6	
G(4)H-1'	5.819	5.807	5.774	5.711	5.670	5.644	5.609	5.603	5.571	39.4	
C(5)H-1'	5.920	5.901	5.874	5.812	5.773	5.737	5.699	5.672	5.622	35.2	
									Average T <sub>m</sub>	35.6	
pro-a	8.364	8.288	8.201	8.103	8.068	8.047	8.039	8.038	8.028	52.9	
pro-b	7.590	7.519	7.427	7.310	7.263	7.234	7.216	7.205	7.187	49.5	
pro-c	6.817	6.799	—	6.506	6.488	6.440	6.408	6.386	6.348	—	
pro-d	6.521	6.441	6.339	6.200	6.146	6.105	6.077	6.060	6.026	47.8	
									Average T <sub>m</sub>	50.1	
				Coupling constant (Hz)							
G(1)H-1'	5.5	4.3	5.6	4.9	4.7	1.5	—	—	—		
C(2)H-1	3.9	3.9	3.4	3.2	3.2	2.7	—	—	—		
U(3)H-1'	4.8	4.3	4.2	3.6	2.3	—	—	—	—		
G(4)H-1'	4.8	4.3	4.2	2.2	1.2	—	—	—	—		
C(5)H-1'	4.8	4.6	5.2	3.4	3.1	1.5	—	—	—		

Table 5.10: Chemical shift assignments and coupling constants for the base and anomeric protons of GCUGC and of the aromatic protons of proflavine for GCUGC, 2.5 mM, with proflavine, 1:1 single strand: proflavine, in 0.1M salt buffer over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)									Tm
	69.7°	59.4°	50.0°	40.3°	35.2°	30.0°	24.4°	20.6°	10.7°	
G(1)H-8	7.898	7.909	7.920	7.927	7.927	7.927	7.919	7.917	7.901	NSB
C(2)H-6	7.767	7.745	7.738	7.718	7.679	7.656	7.646	7.638	—	37.6
U(3)H-6	7.729	7.733	7.737	7.749	7.722	7.706	7.685	7.668	7.654	NSB
G(4)H-8	7.930	7.927	7.912	7.866	7.828	7.796	7.765	7.742	7.740	34.8
C(5)H-6	7.797	7.789	7.768	7.718	7.654	7.604	7.550	7.521	7.467	34.0
C(2)H-5	5.836	5.824	5.739	5.593	5.499	5.416	5.336	5.277	5.180	38.3
U(3)H-5	5.804	5.791	5.771	5.734	5.700	5.680	5.639	5.612	5.567	27.1
C(5)H-5	5.912	5.988	5.824	5.691	5.591	5.501	5.406	—	—	35.5
G(1)H-1'	5.831	5.824	5.789	5.745	5.711	5.680	5.639	5.612	5.567	26.7
C(2)H-1'	5.896	5.891	5.887	5.838	5.810	5.784	5.751	5.733	5.703	35.0
U(3)H-1'	5.810	5.791	5.799	5.798	—	—	—	—	—	—
G(4)H-1'	5.810	5.795	5.772	5.710	5.661	5.629	5.587	5.572	—	37.6
C(5)H-1'	5.919	5.900	5.881	5.813	5.771	5.726	7.688	5.663	5.608	36.6
									Average Tm	35.1
pro-a	8.347	8.300	8.223	8.119	8.070	8.045	8.024	8.015	8.009	47.4
pro-b	7.578	7.531	7.448	7.329	7.271	7.235	7.204	7.188	7.167	45.6
pro-c	6.862	6.810	6.716	6.573	6.499	6.449	6.404	6.379	6.349	44.9
pro-d	6.511	6.450	6.356	6.219	6.148	6.107	6.061	6.039	6.003	43.9
									Average Tm	45.5
	Coupling constant (Hz)									
G(1)H-1'	5.8	3.6	4.7	2.5	2.7	—	—	—	—	—
C(2)H-1'	6.5	3.7	3.7	3.3	3.5	2.9	—	—	—	—
U(3)H-1'	4.9	3.3	5.3	4.7	—	—	—	—	—	—
G(4)H-1'	4.9	3.0	3.7	3.6	3.0	—	—	—	—	—
C(5)H-1'	5.3	4.6	6.9	3.2	3.1	—	—	—	—	—

Table 5.11: Chemical shift assignments and coupling constants for the base and anomeric protons of GCUGC and of the aromatic protons of proflavine for GCUGC, 2.5 mM, with proflavine, 1:2 single strand: proflavine, in 0.1M salt buffer over the temperature range 70 to 10°C.

Temp.	Chemical shift (ppm)									T <sub>m</sub>	
	70.1°	60.5°	50.8°	45.1°	40.4°	36.1°	31.6°	21.1°	10.0°		
G(1)H-8	7.800	7.904	7.916	7.923	7.928	7.931	7.931	7.940	—	NSB	
C(2)H-6	7.769	7.762	7.746	7.734	7.726	7.696	7.671	—	—	35.4	
U(3)H-6	7.727	7.731	7.735	7.734	7.734	7.728	7.713	7.672	—	24.5	
G(4)H-8	7.935	7.928	7.916	7.901	7.880	7.854	7.796	7.692	—	28.7	
C(5)H-6	7.800	7.791	7.776	7.755	7.716	7.696	7.639	7.552	—	31.1	
C(2)H-5	5.841	5.829	5.761	5.705	5.639	5.568	5.412	5.336	—	31.7	
U(3)H-5	5.803	5.791	5.775	5.761	5.741	5.719	5.684	5.649	—	32.1	
C(5)H-5	5.925	5.891	5.839	5.797	5.728	7.661	5.567	—	—	NSB	
G(1)H-1'	5.830	5.819	5.823	5.816	5.807	5.788	5.764	—	5.722	34.7	
C(2)H-1'	5.899	5.891	5.877	5.866	5.484	5.828	5.799	5.754	7.722	33.7	
U(3)H-1'	5.812	5.804	5.793	5.797	5.763	5.741	5.700	5.649	5.635	34.2	
G(4)H-1'	5.812	5.804	5.777	5.755	5.726	5.695	5.646	5.588	5.575	35.4	
C(5)H-1'	5.921	5.905	5.862	5.854	5.829	5.795	5.758	5.688	5.635	30.8	
									Average T <sub>m</sub>	32.8	
pro-a	8.434	8.373	8.316	8.264	8.206	8.156	8.102	8.059	8.030	39.5	
pro-b	7.697	7.594	7.533	7.477	7.415	7.359	7.297	7.235	7.198	39.4	
pro-c	6.918	6.867	6.798	6.736	6.663	6.599	6.523	6.432	6.377	39.5	
pro-d	6.578	6.511	6.440	6.378	6.304	6.245	6.170	6.087	6.033	39.4	
									Average T <sub>m</sub>	39.5	
				Coupling constant (Hz)							
G(1)H-1'	5.3	5.2	5.5	5.0	6.4	3.7	—				
C(2)H-1'	5.6	3.1	3.8	1.8	3.4	3.2	2.3				
U(3)H-1'	3.4	5.4	4.1	4.6	3.3	2.9	—				
G(4)H-1'	3.4	5.5	3.9	4.9	4.1	3.8	3.3				
C(5)H-1'	5.1	4.3	3.9	4.1	4.1	—	—				

Table 5.12:  $T_m$ s of the individual protons of GCGC and proflavine for GCGC:proflavine mixtures in 0.1 M salt buffer. GCGC, 2.4 mM single strand concentration, with proflavine at single strand:drug ratios given.

Proton	$T_m$ (°C)					
	GCGC	15:1	4:1	2:1	1:1	1:2
G(1)H-8	NSB	NSB <sup>a</sup>	NSB	NSB	NSB	NSB
C(2)H-6	46.5	42.4	38.0	NSB	NSB	NSB
G(3)H-8	48.5	48.5	49.1	54.9	45.9	35.8
C(4)H-6	48.9	49.3	46.4	53.0	46.1	40.3
C(2)H-5	54.7	50.6	52.4	58.2	49.6	48.1
C(4)H-5	47.6	49.6	52.2	53.2	48.3	40.3
G(1)H-1'	NSB	NSB	NSB	NSB	NSB	NSB
C(2)H-1'	38.5	50.4	51.7	46.4	46.6	46.2
G(3)H-1'	45.0	46.5	37.6	NSB	NSB	NSB
C(4)H-1'	46.2	43.5	45.7	57.0	41.9	45.5
Average	48.1	48.3	49.6	53.8	46.4	42.7
pro-a	—	62.0	66.7	65.3	—	31.1
pro-b	—	65.2	64.9	63.9	49.6	33.1
pro-c	—	59.8	62.2	64.2	49.1	43.5
pro-d	—	60.7	63.7	63.1	50.7	30.2
Average	—	61.9	64.4	64.1	49.8	34.5

Table 5.13:  $T_m$ s of the individual protons of GCUGC and proflavine for GCGC:proflavine mixtures in 0.1 M salt buffer. GCGC, 2.5 mM single strand concentration, with proflavine at single strand:drug ratios given.

Proton	$T_m$ ( $^{\circ}\text{C}$ )						
	GCUGC	20:1	6:1	4:1	2:1	1:1	1:2
G(1)H-8	NSB	NSB	NSB	NSB	NSB	NSB	NSB
C(2)H-6	NSB	NSB	NSB	NSB	38.4	37.6	35.4
U(3)H-6	NSB	NSB	NSB	NSB	NSB	NSB	NSB
G(4)H-8	19.1	19.3	29.1	30.5	38.5	34.8	28.7
C(5)H-6	20.2	20.8	25.8	25.2	33.1	34.0	31.1
C(2)H-5	21.0	20.8	24.8	31.8	39.1	38.3	31.7
U(3)H-5	22.3	25.0	31.3	35.1	32.1	27.1	32.1
C(5)H-5	21.8	21.1	26.1	26.5	38.4	35.5	NSB
G(1)H-1'	20.7	20.3	32.5	28.1	27.4	26.7	34.7
C(2)H-1'	19.5	19.0	20.0	21.4	32.8	35.0	33.7
U(3)H-1'	17.2	NSB	27.3	25.8	33.6	NSB	34.2
G(4)H-1'	NSB	NSB	NSB	26.8	39.4	37.6	35.4
C(5)H-1'	19.9	20.3	26.5	29.0	35.2	36.6	30.8
Average	20.0	21.0	25.4	27.2	35.6	35.1	32.8
pro-a	--	46.5	55.0	58.4	52.9	47.4	39.5
pro-b	--	--	51.8	54.6	49.5	45.6	39.4
pro-c	--	42.0	41.1	48.5	--	44.9	39.5
pro-d	--	46.0	49.3	54.1	47.8	43.9	39.4
Average	--	45.0	49.3	53.9	50.1	45.5	39.5

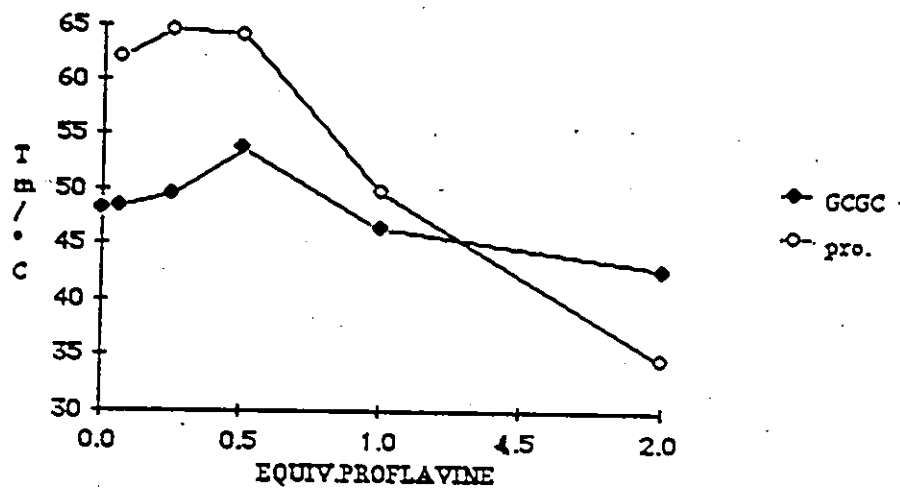


Figure 5.2: Changes in average  $T_m$  for GCGC and proflavine upon addition of proflavine to GCGC.

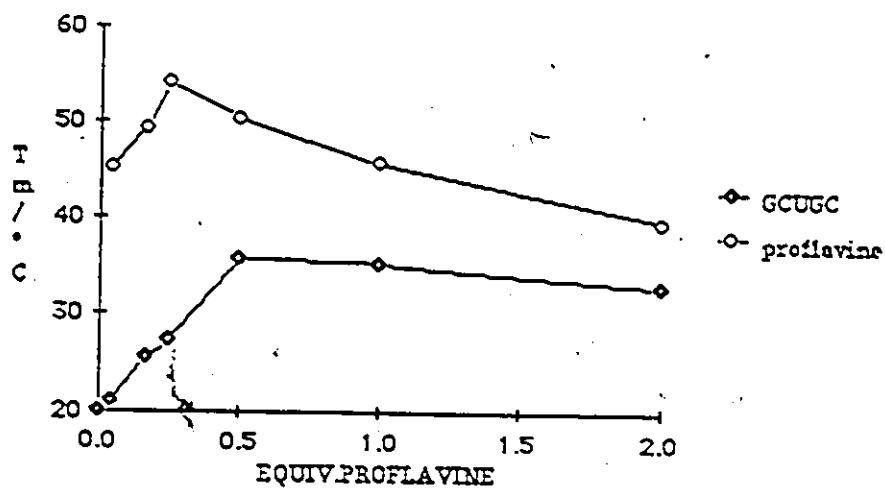


Figure 5.3: Changes in average  $T_m$  for GCUGC and proflavine upon addition of proflavine to GCUGC.

markedly than GCUGC (Figure 5.2) A gradual shift of all chemical shift versus temperature curves to higher temperature and higher field was apparent, the base protons of GCGC at differing proflavine concentrations are plotted individually in Figures 5.4 to 5.9, and those of GCUGC in Figures 5.10 to 5.17. Certain curves which initially displayed sigmoidal behaviour lost their sigmoidal character (e.g., C2H6 in GCGC) and others, nonsigmoidal at the outset, lost it (e.g., G4H1' in GCUGC).

The proflavine signals showed sigmoidal behaviour with upfield shifts of  $> 0.35 - > 0.55$  ppm (see Figure 5.26), compared with gradual nonsigmoidal upfield drift of the signals of proflavine in isolation. The  $T_m$  measured for the proflavine signals starts, and continues to be, considerably higher than the duplex  $T_m$ , and peaks at lower proflavine concentrations (Figure 5.2; Figure 5.3).

Variable temperature spectra of the exchangeable protons were obtained for GCGC and GCUGC in the absence of proflavine (Section 4.2.3) and with 6:1 and 2:1 single strand:drug; the results are given in Tables 5.14 and 5.15. Assignments were as previously discussed. At the lower proflavine concentration (6:1) the spectra appeared much as they did in the absence of proflavine, but in the 2:1 single strand: drug, the UNH3 signal in GCUGC has almost disappeared, and the spectrum of GCGC has acquired an extra peak, upfield from the others. All resonances shift upfield upon addition of proflavine, but this is most pronounced for the terminal resonances.

### 5.2.2 Intercalation of proflavine into CCGG and CCAGG

Variable temperature series for the nonexchangeable and exchangeable protons of CCGG and CCAGG with proflavine were obtained for single strand:drug ratios of 10:1, 4.6:1, 3.5:1, 2:1 and 1:1 for CCGG (2.5 mM, 0.1M salt) (Tables 5.16 to 5.20), and for

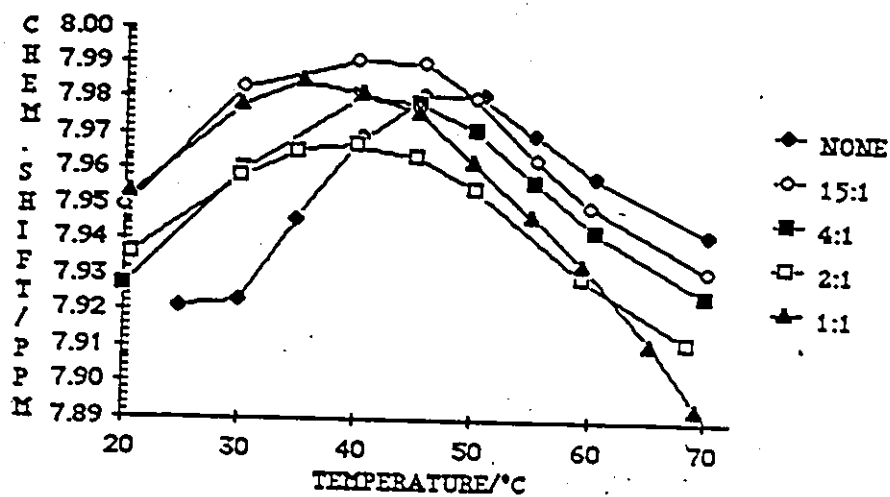


Figure 5.4: Changes in chemical shift for G(1)H-8 of GCGC, 2.4 mM in 0.1 M salt buffer, upon addition of proflavine.

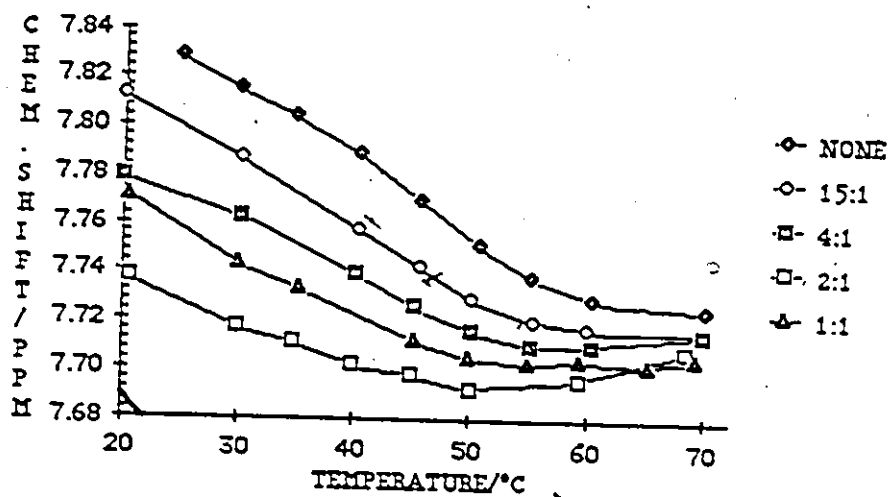


Figure 5.5: Changes in chemical shift for C(2)E-6 of GCGC upon addition of proflavine at ratios shown.



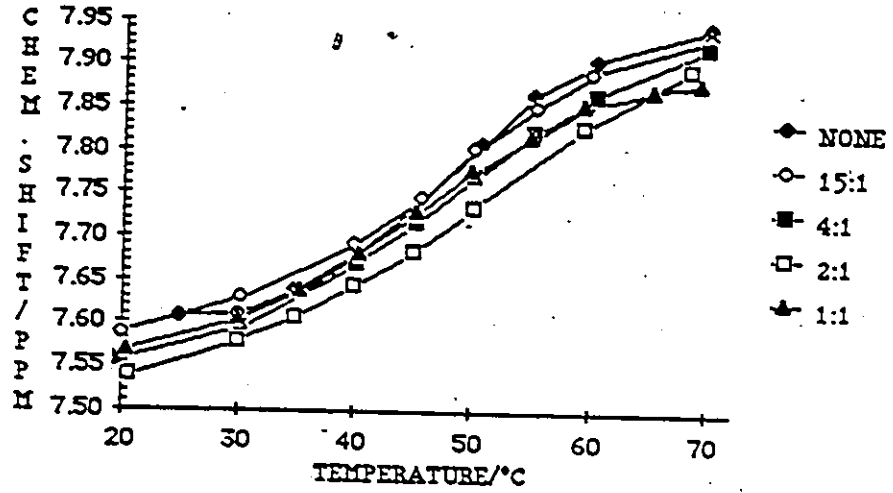


Figure 5.6: Changes in chemical shift for G(3)H-8 of GCGC, upon addition of proflavine at ratios shown.

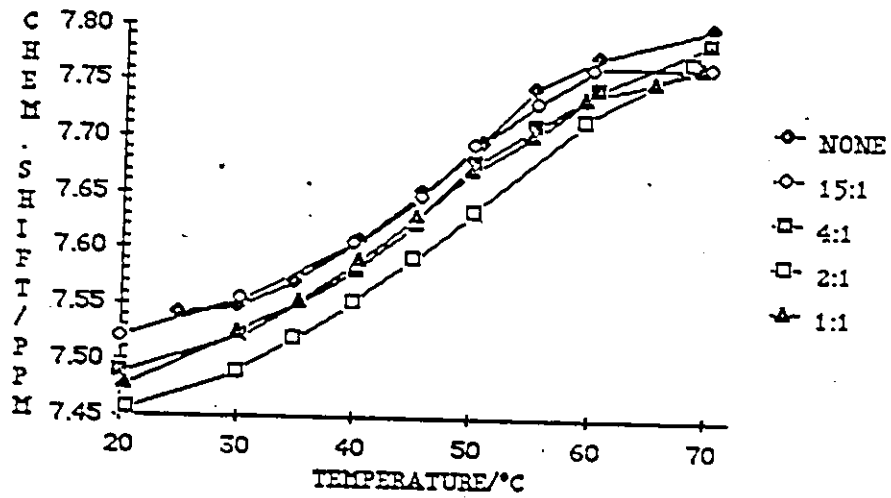


Figure 5.7: Changes in chemical shift for C(4)H-6 of GCGC upon addition of proflavine at ratios shown.

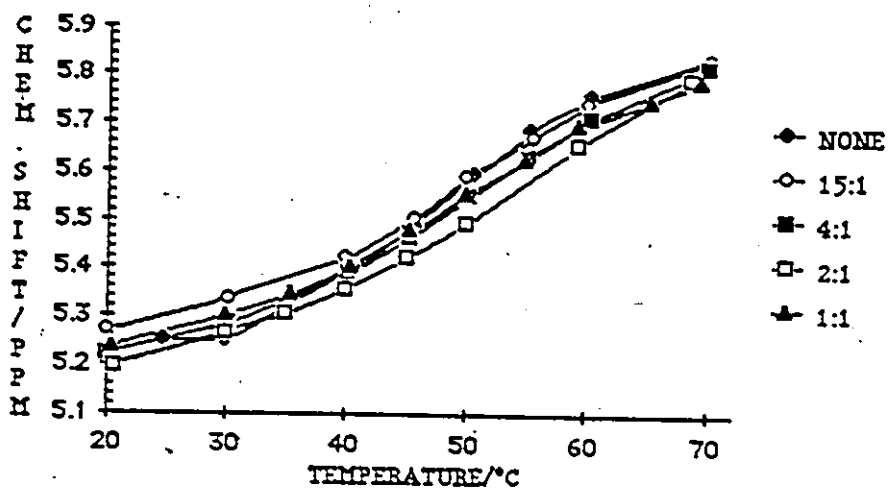


Figure 5.8: Changes in chemical shift for C(2)H-5 of GCGC upon addition of proflavine at ratios shown.

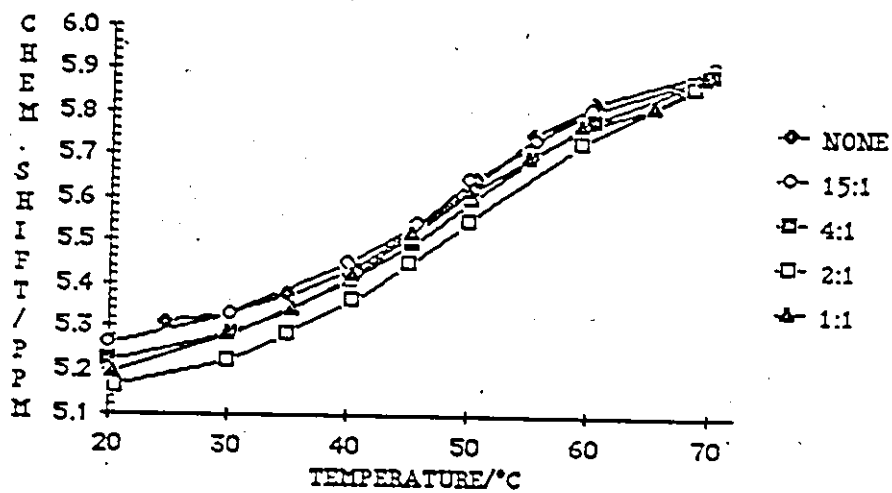


Figure 5.9: Changes in chemical shift for C(4)H-5 of GCGC upon addition of proflavine at ratios shown.

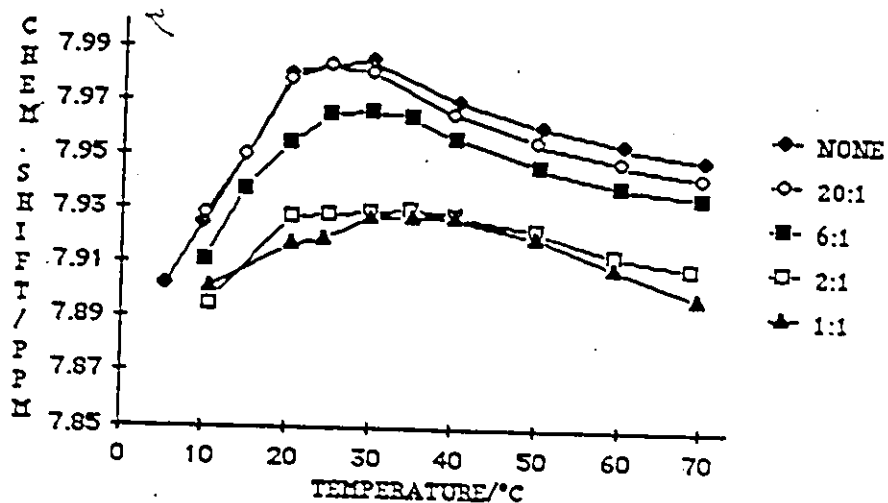


Figure 5.10: Changes in chemical shift for G(1)H-8 in GCUGC upon addition of proflavine at ratios shown.

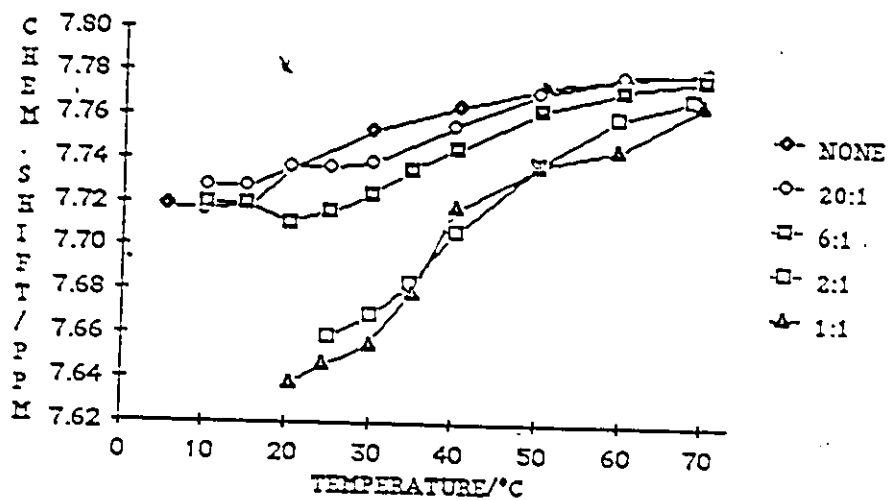


Figure 5.11: Changes in chemical shift with C(2)H-6 in GCUGC upon addition of proflavine at ratios shown.

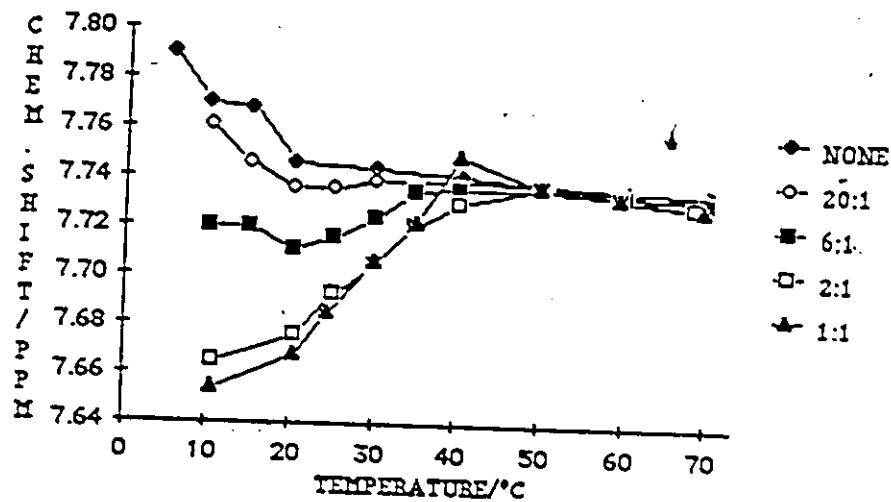


Figure 5.12: Changes in chemical shift for U(3)H-6 in GCUGC upon addition of proflavine at ratios shown.

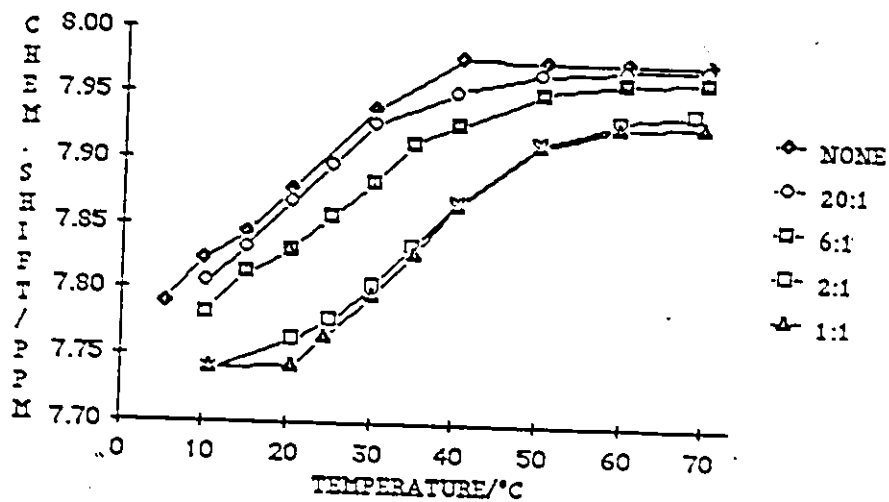


Figure 5.13: Changes in chemical shift for G(4)H-8 in GCUGC upon addition of proflavine at ratios shown.

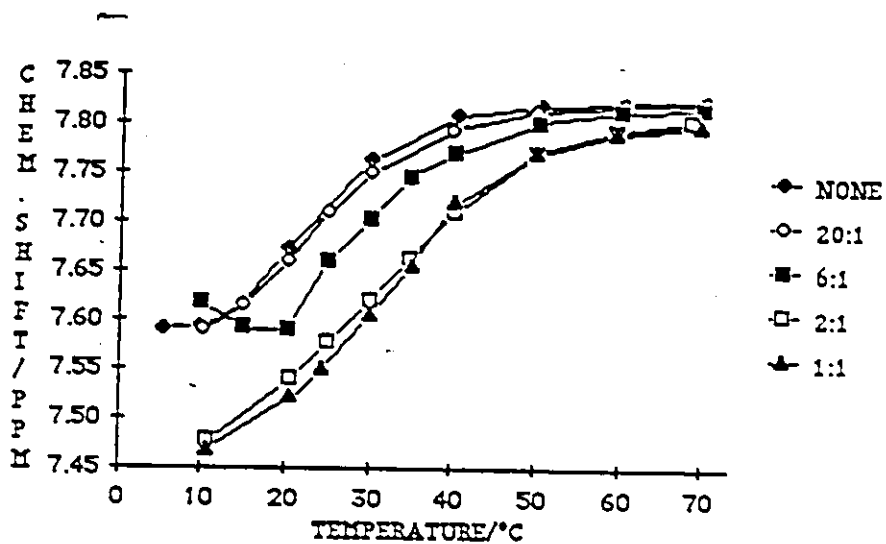


Figure 5.14: Changes in chemical shift for C(5)H-6 in GCUGC upon addition of proflavine at ratios shown.

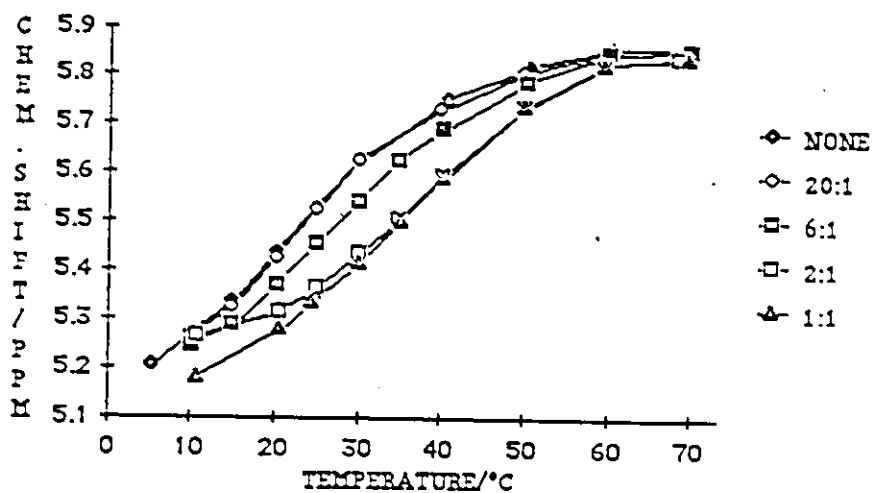


Figure 5.15: Changes in chemical shift for C(2)H-5 in GCUGC upon addition of proflavine at the ratios shown.

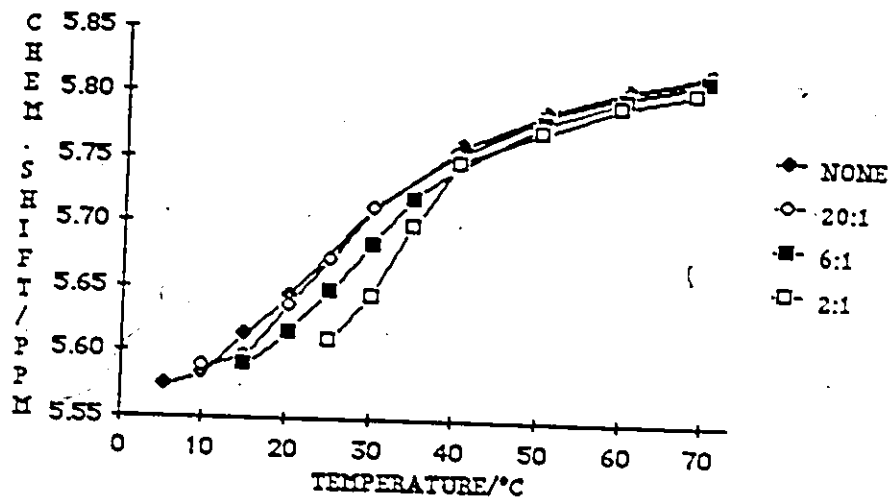


Figure 5.16: Changes in chemical shift for U(3)H-5 in GCUGC upon addition of proflavine at ratios shown.

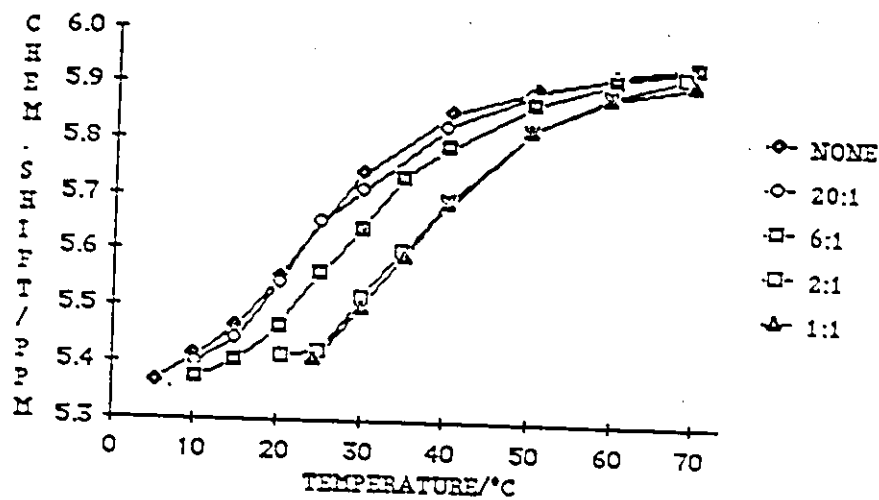


Figure 5.17: Changes in chemical shift for C(5)E-5 in GCUGC upon addition of proflavine at ratios shown.

Table 5.14: Chemical shifts of the hydrogen-bonded imino protons of GCGC, 2.4mM in 0.1M salt buffer, with proflavine at single strand:drug ratios of 6:1 and 2:1.

GCGC	Chemical shift (ppm)				
	<u>2.0°</u>	<u>9.2°</u>	<u>19.0°</u>	<u>31.0°</u>	
Proton					
TERMINAL	12.77	12.84	12.87	(12.8)	
INTERNAL	13.05	13.06	13.05	12.94	
<u>6:1</u>	<u>1.0°</u>	<u>8.7°</u>	<u>19.4°</u>	<u>29.2°</u>	
TERMINAL	12.60	12.65	12.68	(12.9)	
INTERNAL	12.97	13.00	12.92	13.04	
<u>2:1</u>	<u>-3.4°</u>	<u>2.9°</u>	<u>7.5°</u>	<u>17.6°</u>	<u>24.5°</u>
TERMINAL	12.34	12.87	12.36	12.38	12.41
INTERNAL	12.85	12.85	12.83	12.79	12.74
	12.15	12.19	12.21	12.24	12.24

Table 5.15: Chemical shifts of the hydrogen-bonded imino protons of GCUGC, 2.5mM in 0.1M salt buffer, with proflavine at single strand:drug ratios of 6:1 and 2:1.

GCUGC Proton	Chemical shift (ppm)			
	<u>2.0°</u>	<u>3.3°</u>	<u>9.2°</u>	
TERMINAL	12.87	13.00	12.88	
INTERNAL	13.38	13.43	13.27	
UNHS	10.25	10.33	10.28	
<u>6:1</u>	<u>1.0°</u>		<u>8.7°</u>	
TERMINAL	12.86		-	
INTERNAL	13.41		13.14	
UNHS	10.29		10.22	
<u>2:1</u>	<u>-3.4°</u>	<u>3.0°</u>	<u>7.5°</u>	<u>17.6°</u>
TERMINAL	12.46	12.41	12.34	12.29
INTERNAL	13.30	13.15	12.80	12.66



Table 5.16: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CCGG, 2.5 mM, with proflavine, 10:1 single strand:drug, in 0.1 M salt buffer, over the temperature range 80° to 20°C

Temp.	Chemical shift (ppm)										T <sub>m</sub>
	80.4°	70.0°	60.2°	55.1°	50.4°	45.4°	40.4°	35.1°	30.3°	20.3°	
Proton											
C(1)H-6	7.768	7.785	7.816	7.842	7.875	7.910	7.934	7.943	7.940	7.903	51.1
C(2)H-6	7.759	7.761	7.768	7.777	7.793	7.810	7.828	7.841	7.852	7.841	45.6
G(3)H-8	7.919	7.906	7.873	7.842	7.797	7.740	7.682	7.628	7.592	7.547	45.3
G(4)H-8	7.694	7.953	7.914	7.873	7.809	7.728	7.640	7.559	7.504	7.414	44.8
C(1)H-5	5.972	5.961	5.951	5.948	5.946	5.943	5.934	5.919	5.902	5.860	NSB
C(2)H-5	5.981	5.961	5.915	5.880	5.834	5.779	5.721	5.668	5.623	5.562	50.1
C(1)H-1'	5.863	5.855	5.846	5.840	5.833	5.826	5.819	5.810	5.805	5.794	NSB
C(2)H-1'	5.805	5.791	5.769	5.754	5.737	5.719	5.702	5.687	5.682	5.673	46.3
G(3)H-1'	5.861	5.841	5.801	5.772	5.737	5.699	5.662	5.627	5.607	5.573	50.1
G(4)H-1'	5.829	5.814	5.784	5.754	5.713	5.656	5.585	5.505	5.433	5.294	NSB
										Average T <sub>m</sub>	47.1
pro-a	—	8.477	8.357	8.296	8.243	8.208	8.180	8.163	8.149	8.127	63.5
pro-b	7.734	7.649	7.506	7.426	7.353	7.297	7.256	7.224	7.203	7.172	62.1
pro-c	6.960	6.874	6.705	6.601	6.500	6.418	6.353	6.303	6.270	6.212	58.9
pro-d	6.677	6.586	6.444	6.362	6.285	6.223	6.174	6.141	6.107	6.056	61.3
										Average T <sub>m</sub>	61.5
	Coupling constant (Hz)										
C(1)H-1'	4.5	5.4	—	4.9	4.6	4.0	3.0	2.3	—	—	
C(2)H-1'	5.4	5.2	4.9	4.3	3.7	3.1	2.0	—	—	—	
G(3)H-1'	5.6	4.7	4.1	4.3	3.7	2.6	1.9	—	—	—	
G(4)H-1'	4.0	3.8	2.6	4.3	2.4	1.7	—	—	—	—	

Table 5.17: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CCGG, 2.5 mM, with proflavine, 4.6:1 single strand:drug, in 0.1 M salt buffer, over the temperature range 80° to 20°C

Temp.	Chemical shift (ppm)									Tm	
	80.1°	70.3°	60.4°	55.2°	50.7°	45.6°	40.7°	30.2°	20.5°		
Proton											
C(1)H-6	7.759	7.774	7.804	7.828	7.858	7.887	7.909	7.916	7.887	51.5	
C(2)H-6	7.750	7.750	7.753	7.757	7.767	7.780	7.793	7.810	7.825	47.5	
G(3)H-8	7.910	7.894	7.855	7.820	7.776	7.723	7.670	7.586	7.537	46.5	
G(4)H-8	7.956	7.941	7.895	7.851	7.790	7.715	7.636	7.511	7.431	45.8	
C(1)H-5	5.968	5.948	5.934	5.926	5.920	5.911	5.900	5.866	5.825	NSB	
C(2)H-5	5.961	5.941	5.889	5.850	5.802	5.749	5.694	5.606	5.562	47.3	
C(1)H-1'	5.862	5.847	5.834	5.825	5.816	5.805	5.795	5.776	5.763	NSB	
C(2)H-1'	5.797	5.781	5.752	5.733	5.712	5.692	5.674	5.652	5.643	52.5	
G(3)H-1'	5.854	5.835	5.785	5.750	5.708	5.670	5.631	5.574	5.539	48.5	
G(4)H-1'	5.825	5.808	5.769	5.736	5.695	5.636	5.571	5.436	5.316	NSB	
									Average Tm	49.1	
pro-a	8.516	8.424	8.308	8.251	8.207	8.174	8.150	8.118	8.092	67.2	
pro-b	7.698	7.607	7.466	7.389	7.322	7.270	7.230	7.179	7.146	63.2	
pro-c	6.936	6.843	6.672	6.570	6.475	6.397	6.335	6.251	6.198	59.7	
pro-d	6.637	6.541	6.397	6.315	6.248	6.190	6.143	6.078	6.023	62.5	
									Average Tm	63.2	
				Coupling constant (Hz)							
C(1)H-1'	5.0	5.3	5.1	4.8	4.5	4.1	3.1	2.6	--		
C(2)H-1'	5.3	5.3	4.9	4.6	3.8	3.4	2.5	--	--		
G(3)H-1'	5.4	4.6	4.4	3.9	4.3	2.7	2.0	--	--		
G(4)H-1'	3.9	3.7	3.5	3.1	5.7	1.7	--	--	--		



Table 5.19: Chemical shifts and anomeric proton coupling constants for the base and anomeric protons of CCGG, 2.5 mM, with proflavine, 2:1 single strand:drug, in 0.1 M salt buffer, over the temperature range 80° to 20°C.

Temp	Chemical shift (ppm)									T <sub>m</sub>
	80.0°	70.0°	60.2°	55.0°	50.5°	45.5°	40.3°	30.1°	20.2°	
Proton										
C(1)H-6	7.747	7.760	7.790	7.813	7.836	7.858	7.873	7.878	7.855	53.8
C(2)H-6	7.740	7.736	7.737	7.738	7.737	7.738	7.740	7.744	7.745	NSB
G(3)H-8	7.901	7.880	7.845	7.811	7.762	7.710	7.660	7.588	7.535	47.6
G(4)H-8	7.946	7.927	7.887	7.843	7.780	7.710	7.638	7.530	7.463	46.4
C(1)H-5	5.950	5.933	5.918	5.907	5.890	5.873	5.853	5.811	5.767	NSB
C(2)H-5	5.957	5.924	5.872	5.830	5.773	5.717	5.662	5.581	5.528	48.4
C(1)H-1'	5.861	5.836	5.825	5.814	5.797	5.779	5.761	5.730	5.712	NSB
C(2)H-1'	5.790	5.769	5.741	5.719	5.693	5.663	5.640	5.614	5.600	50.1
G(3)H-1'	5.846	5.826	5.776	5.736	5.684	5.632	5.585	5.518	5.490	49.4
G(4)H-1'	5.822	5.800	5.760	5.724	5.645	5.617	5.553	5.439	5.339	NSB
								Average T <sub>m</sub>		49.3
pro-a	8.493	8.402	8.318	8.261	8.200	8.157	8.124	8.082	8.053	57.0
pro-b	7.686	7.597	7.486	7.501	7.327	7.265	7.218	7.155	7.118	56.3
pro-c	6.935	6.845	6.703	6.599	6.493	6.493	6.337	6.245	6.185	55.1
pro-d	6.620	6.526	6.413	6.332	6.247	6.181	6.124	6.048	5.988	54.5
								Average T <sub>m</sub>		55.7
								Coupling constant (Hz)		
C(1)H-1'	4.6	4.9	5.2	5.0	4.5	4.2	3.4	3.0	--	
C(2)H-1'	5.3	5.3	5.3	4.8	2.2	3.6	3.0	--	--	
G(3)H-1'	5.4	5.4	4.2	3.5	2.2	2.7	2.2	--	--	
G(4)H-1'	3.8	4.0	3.9	2.2	2.6	1.6	--	--	--	

Table 5.20: Chemical shifts and anomeric proton coupling constants of the base and anomeric protons of CCGG, 2.5 mM, with proflavine, 1:1 single strand:drug, in 0.1 M salt buffer, over the temperature range 80° to 20°C

Temp.	Chemical shift (ppm)									Tm
	80.8°	71.1°	60.8°	55.8°	50.9°	45.8°	40.7°	30.6°	20.3°	
Proton										
C(1)H-6	7.741	7.752	7.780	7.801	7.824	7.847	7.859	7.869	7.854	54.0
C(2)H-6	7.736	7.731	7.729	7.728	7.725	7.727	7.720	7.716	7.725	NSB
G(3)H-8	7.899	7.878	7.846	7.816	7.770	7.720	7.668	7.593	7.548	46.2
G(4)H-8	7.946	7.926	7.890	7.851	7.793	7.720	7.657	7.555	7.468	46.5
C(1)H-5	5.946	5.921	5.911	5.900	5.883	5.868	5.839	5.797	5.760	NSB
C(2)H-5	5.954	5.928	5.870	5.828	5.775	5.720	5.658	5.583	5.531	49.0
C(1)H-1'	5.843	5.832	5.824	5.813	5.798	5.778	5.752	5.714	5.697	42.7
C(2)H-1'	5.789	5.767	5.740	5.718	5.692	5.662	5.633	5.598	5.586	48.3
G(3)H-1'	5.864	5.832	5.778	5.740	5.682	5.626	5.577	5.507	5.474	49.0
G(4)H-1'	5.823	5.801	5.763	5.727	5.682	5.626	5.563	5.453	5.368	42.7
									Average Tm	48.8
pro-a	8.516	8.424	8.355	8.297	8.235	8.187	8.137	8.089	8.051	--
pro-b	7.707	7.622	7.520	7.451	7.368	7.302	7.238	7.166	7.122	52.3
pro-c	6.954	6.874	6.751	6.655	6.544	6.450	6.365	6.266	6.195	52.6
pro-d	6.639	6.552	6.451	6.375	6.286	6.213	6.142	6.053	6.989	51.8
									Average Tm	52.3
	Coupling constants (Hz)									
C(1)H-1'	5.8	4.7	5.3	5.0	3.7	4.2	3.9	--	--	
C(2)H-1'	5.4	5.4	5.1	4.9	3.0	3.8	3.6	--	--	
G(3)H-1'	4.3	4.7	4.2	2.6	1.7	--	--	--	--	
G(4)H-1'	4.3	3.9	3.3	--	1.7	--	--	--	--	

ratios of 5:1, 2:1 and 1:1 for CCAGG (2.0mM, 0.1M salt) (Tables 5.21 to 5.23).

The increase in duplex  $T_m$  upon addition of proflavine to CCGG was comparable to that upon addition of proflavine to GCGC (previous section), 4.8°C at a ratio of 3:5:1 single strand:drug (Table 5.24; Figure 5.18). The stabilization of CCAGG upon addition of proflavine reached a maximum of 5°C (Table 5.25; Figure 5.19). With addition of proflavine the sigmoidal curves are shifted to higher field and temperature, one loses its sigmoidal character, and certain of the nonsigmoidal curves are shifted upfields. Plots for individual base protons are presented in Figures 5.20 to 5.25 for CCGG.

The proflavine signals were sigmoidal for both duplexes; in the CCGG experiments the proflavine  $T_m$  reached a maximum at 4.6:1, whereas in the CCAGG experiments, there was no defined maximum (Tables 5.24 and 5.25).

Broadening of the signals in the CCAGG experiments was a problem, as was degradation of this sample. The  $T_m$  values derived for this duplex are attended by appreciable uncertainty. The proflavine signals pro-b, pro-c and pro-d can be readily followed throughout a sigmoidal transition (pro-a moves into the AH-2 and AH-8 regions and cannot be followed) and are determined to have  $T_m$ s of around 40°C at low proflavine ratios. This is a full 10°C lower than the corresponding curves for the GCUGC:proflavine complex.

Chemical shifts for the hydrogen bonded imino protons of CCGG and CCAGG with proflavine are given in Tables 5.26 and 5.27. The trends observed were similar to those described in Section 5.2.1: The terminal imino protons were more shielded by addition of proflavine, and the tetramer developed an extra resonance at high proflavine concentrations.

Table 5.21: Chemical shifts and anomeric coupling constants for CCAGG, 2.5 mM, with proflavine, 5:1 single strand:drug, in 0.1 M salt buffer over the temperature range 70° to 4°C.

Temp.	Chemical shift (ppm)										T <sub>m</sub>
	69.3°	59.7°	50.4°	40.4°	30.1°	24.6°	20.3°	15.3°	10.8°	3.8°	
Proton											
C(1)H-6	7.725	7.727	7.726	7.724	7.725	7.735	7.746	7.759	7.783	7.799	25.5
C(2)H-6	7.710	7.716	7.726	7.745	7.783	7.714	7.840	7.867	7.881	7.913	13.4
A(3)H-8	8.290	8.282	8.271	8.256	8.231	8.213	8.195	8.178	8.172	8.185	NSB
A(3)H-2	8.111	8.086	8.051	8.007	7.938	7.896	7.853	7.793	7.731	7.681	15.5
G(4)H-8	7.867	7.850	7.824	7.784	7.705	7.635	7.559	7.452	7.373	-	NSB
G(5)H-8	7.936	7.930	7.916	7.888	7.826	7.770	7.717	7.634	7.579	7.505	15.8
C(1)H-5	5.912	5.884	5.843	5.787	5.714	5.672	5.636	5.595	5.572	5.535	28.2
C(2)H-5	5.878	5.855	5.829	5.804	5.789	5.778	5.772	5.760	5.755	5.723	NSB
C(1)H-1'	5.833	5.816	5.789	5.748	5.690	5.662	5.623	5.582	5.558	5.519	19.3
C(2)H-1'	5.806	5.789	5.765	5.735	7.681	5.647	5.623	5.582	5.558	5.519	20.6
A(3)H-1'	5.950	5.937	5.929	5.910	5.906	5.910	5.919	5.933	-	-	NSB
G(4)H-1'	5.771	5.754	5.729	5.693	5.637	5.602	5.571	5.522	5.500	5.484	20.1
G(5)H-1'	5.828	5.819	5.809	5.798	5.782	5.782	5.760	5.760	5.755	5.755	24.6
										Average T <sub>m</sub>	20.3
pro-a	8.467	8.387	-	-	-	-	-	-	-	-	-
pro-b	7.667	7.599	7.501	7.363	7.218	7.147	7.117	7.083	7.075	7.059	40.9
pro-c	6.919	6.859	6.761	6.608	6.421	6.336	6.283	6.234	6.208	6.16	37.6
pro-d	6.604	6.530	6.426	6.292	6.142	6.070	6.035	-	-	-	42.1
										Average T <sub>m</sub>	40.2
	Coupling constant (Hz)										
C(1)H-1'	5.5	4.1	4.5	6.0	5.3	-	-	-	-	-	-
C(2)H-1'	4.5	3.6	3.3	5.1	5.6	6.3	-	-	-	-	-
A(3)H-1'	5.9	5.1	4.9	4.5	4.1	3.5	2.6	-	-	-	-
G(4)H-1'	5.3	5.3	5.2	4.8	4.1	3.7	3.7	-	-	-	-
G(5)H-1'	5.6	5.5	5.5	4.8	4.6	-	4.1	-	-	-	-





Table 5.23: Chemical shifts and anomeric coupling constants for CCAGG, 2.5 mM, with proflavine, 1:1 single strand:drug, in 0.1 M salt buffer over the temperature range 70° to 0.6°C.

Temp	Chemical shift (ppm)										T <sub>m</sub>	
	71.0°	60.8°	49.7°	39.5°	30.3°	25.1°	20.3°	15.2°	10.0°	0.6°		
Proton												
C(1)H-6	7.721	7.720	7.716	7.710	7.689	7.687	-	-	7.664	7.657	NSB	
C(2)H-6	7.704	7.710	7.722	7.743	7.774	7.800	7.823	7.841	7.845	7.809	27.2	
A(3)H-8	8.282	8.272	8.263	8.253	8.234	8.219	8.204	8.190	8.173	8.170	20.0	
A(3)H-2	8.092	8.062	8.030	7.993	7.934	7.892	7.804	7.732	7.764	7.608	29.1	
G(4)H-8	7.861	7.841	7.813	7.778	7.703	7.645	7.574	7.470	7.405	-	NSB	
G(5)H-8	7.926	7.916	7.900	7.871	7.810	7.764	7.711	7.643	7.594	7.501	16.9	
C(1)H-5	5.909	5.877	5.828	5.767	5.667	5.626	5.570	5.518	5.471	-	27.8	
C(2)H-5	5.874	5.848	5.820	5.796	5.775	5.776	5.755	5.767	5.739	5.735	NSB	
C(1)H-1'	5.835	5.811	5.797	5.786	5.767	5.757	5.755	7.745	5.739	5.735	NSB	
C(2)H-1'	5.817	5.808	5.781	5.743	5.698	-	5.644	5.625	5.594	-	33.7	
A(3)H-1'	5.943	5.928	5.919	5.910	5.907	5.909	5.911	5.921	5.935	-	NSB	
G(4)H-1'	5.766	5.747	5.719	5.683	5.630	5.594	5.570	5.519	5.471	-	NSB	
G(5)H-1'	5.801	5.783	5.755	5.718	5.670	5.648	5.613	5.577	5.552	5.522	20.7	
											Average T <sub>m</sub>	25.1
pro-a	8.411	8.329	8.233	8.127	8.015	7.973	7.940	-	-	-	39.5	
pro-b	7.629	7.555	7.458	7.328	7.183	7.122	7.079	7.055	7.044	-	38.0	
pro-c	6.898	6.828	6.723	6.578	6.399	5.312	6.246	6.216	6.187	6.142	35.8	
pro-d	6.567	6.485	6.380	6.251	6.104	6.048	6.004	-	-	-	37.6	
											Average T <sub>m</sub>	37.7
C(1)H-1'	4.5	5.6	3.9	5.1	4.6	4.3	-	-	-	-		
C(2)H-1'	4.4	4.1	4.4	4.5	3.6	-	-	-	-	-		
A(3)H-1'	5.3	5.3	4.9	4.6	4.4	5.0	-	-	-	-		
G(4)H-1'	5.4	5.3	5.2	4.7	4.0	3.4	-	-	-	-		
G(5)H-1'	3.5	3.7	3.3	2.9	3.9	3.7	-	-	-	-		

Table 5.24:  $T_m$ s of CCGG and proflavine for CCGG:proflavine mixtures with single strand:drug ratios of 10:1 to 3:2. CCGG, 2.5mM single strand in 0.1M salt buffer.

Ratio	$T_m$ (°C)					
	CCGG	10:1	4.6:1	3.5:1	2:1	3:2
Proton						
C(1)H-6	45.8	51.1	51.5	52.4	53.8	54.0
C(2)H-6	47.2	45.6	47.5	42.8	NSB	NSB
G(3)H-8	44.7	45.3	46.5	47.0	47.6	46.2
G(4)H-8	44.2	44.8	45.8	44.0	46.4	46.5
C(1)H-5	NSB	NSB	NSB	NSB	NSB	NSB
C(2)H-5	44.5	44.8	47.3	48.2	48.4	49.0
C(1)H-1'	NSB	NSB	NSB	NSB	NSB	42.7
C(2)H-1'	45.5	46.3	52.5	53.4	50.1	48.3
G(3)H-1'	45.1	50.1	51.0	53.7	49.4*	49.0
G(4)H-1'	NSB	NSB	NSB	NSB	45.4	42.7
Average	45.0	47.1	49.1	49.8	49.3	48.8
pro-a	-	63.5	67.3	60.3	57.0	-
pro-b	-	62.1	63.2	65.3	56.3	52.3
pro-c	-	58.9	59.7	61.8	55.1	52.6
pro-d	-	61.3	62.5	59.6	54.5	51.8
Average	-	61.5	63.2	61.8	55.7	52.3

Table 5.25:  $T_m$ s for CCAGG and proflavine for CCAGG:proflavine mixtures with single strand:drug ratios of 6:1 to 2:1. CCAGG, 2.0mM in 0.1M salt buffer.

Ratio	CCAGG	$T_m$ (°C)		
		6:1	2:1	1:1
Proton				
C(1)H-6	NSB	NSB	NSB	NSB
C(2)H-6	20.2	13.4	22.3	27.2
A(3)H-8	18.8	NSB	21.2	20.0
A(3)H-2	15.9	15.8	15.8	29.1
G(4)H-8	NSB	NSB	19.2	NSB
G(5)H-8	17.4	15.8	20.0	16.9
C(1)H-5	13.6	28.2	27.0	27.8
C(2)H-5	NSB	NSB	NSB	NSB
C(1)H-1'	NSB	19.3	17.5	NSB
C(2)H-1'	18.8	20.6	NSB	33.7
A(3)H-1'	NSB	NSB	NSB	NSB
G(4)H-1'	20.5	20.1	23.7	NSB
G(5)H-1'	24.6	24.5	18.9	20.7
Average	18.7	19.2	19.9	25.1
pro-a	--	--	--	39.5
pro-b	--	40.9	40.2	38.0
pro-c	--	37.6	37.1	35.8
pro-d	--	42.1	38.6	37.6
Average	--	40.2	38.6	37.7

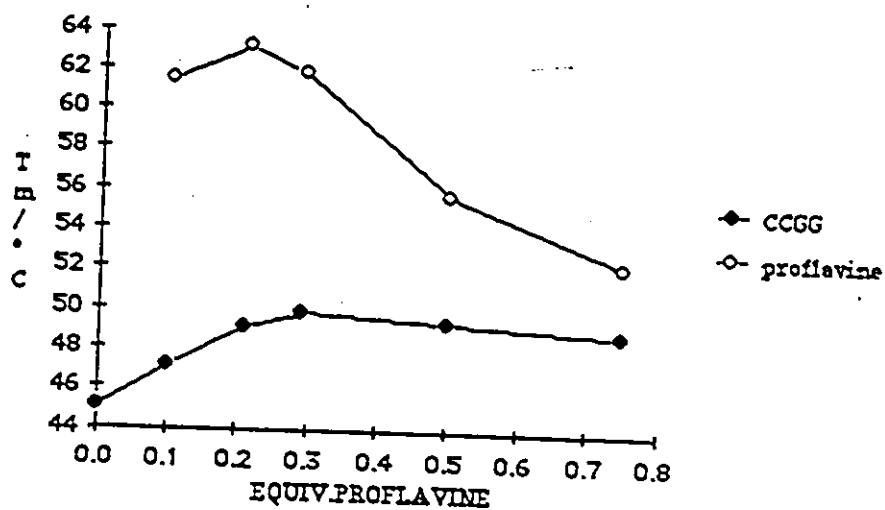


Figure 5.18: Changes in  $T_m$  for CCGG and proflavine for CCGG upon addition of proflavine.

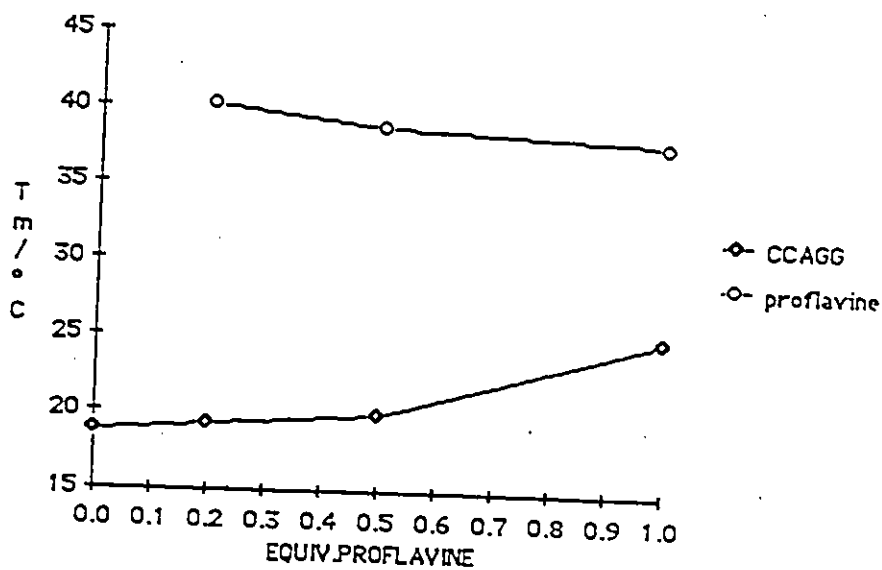


Figure 5.19: Changes in  $T_m$  of CCAGG and proflavine upon addition of proflavine to CCAGG.

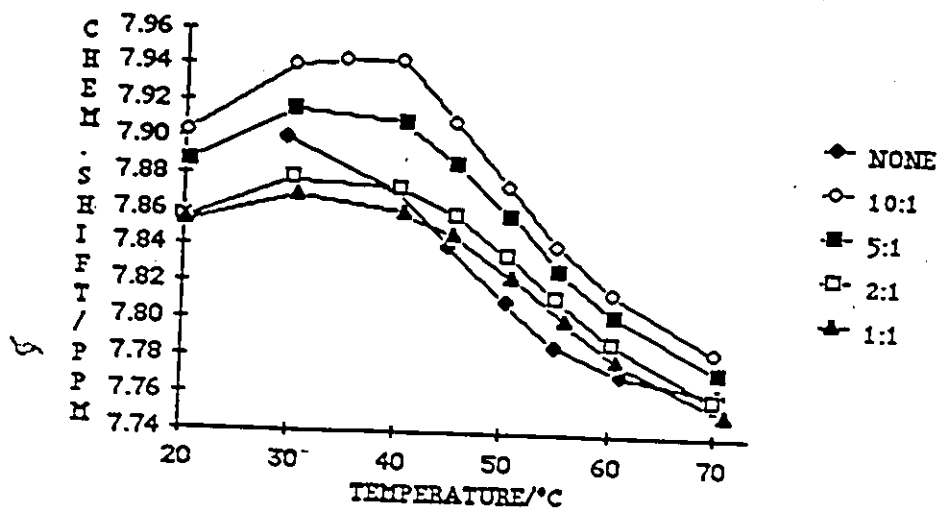


Figure 5.20: Changes in chemical shift for C(1)H-6 in CCGG upon addition of proflavine at single strand:drug ratios shown.

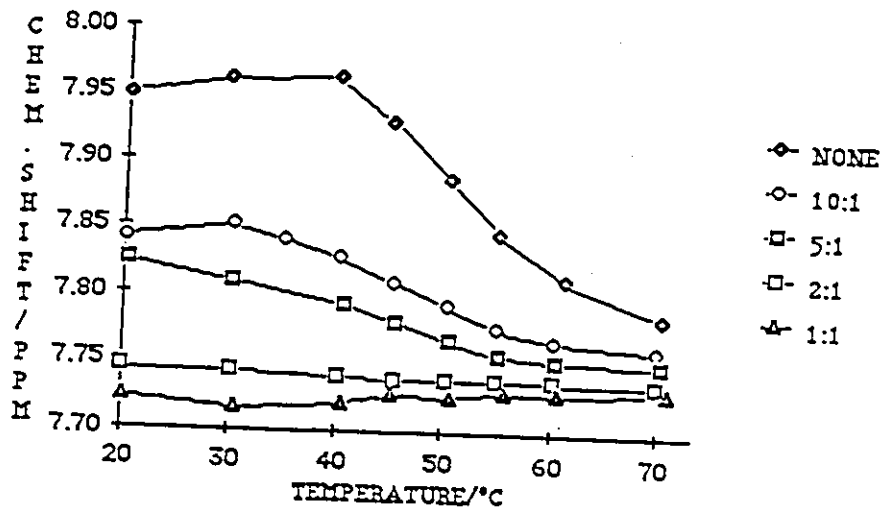


Figure 5.21: Changes in chemical shift for C(2)H-6 in CCGG upon addition of proflavine at ratios shown.

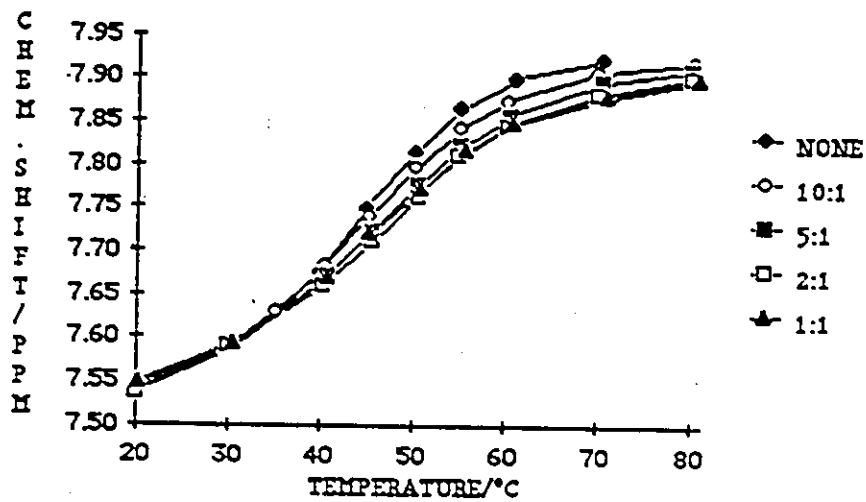


Figure 5.22: Changes in chemical shift for G(3)H-8 in CCGG upon addition of proflavine at single strand:drug ratios shown.

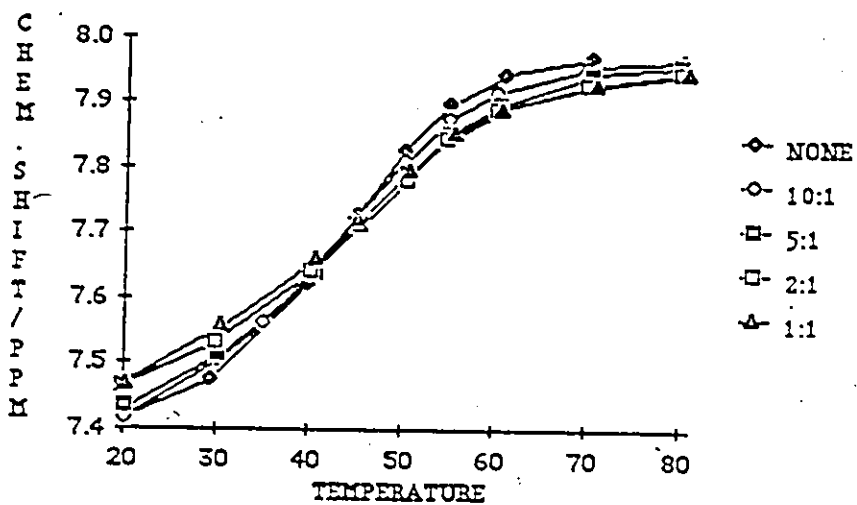


Figure 5.23: Chemical shift changes for G(4)H-8 in CCGG upon addition of proflavine at ratios shown.

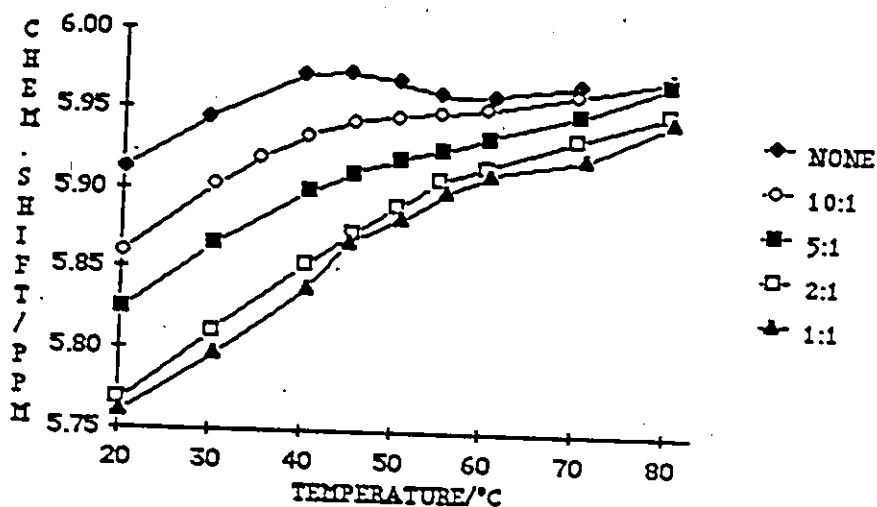


Figure 5.24: Changes in chemical shift for C(1)H-5 in CCGG upon addition of proflavine at ratios shown.

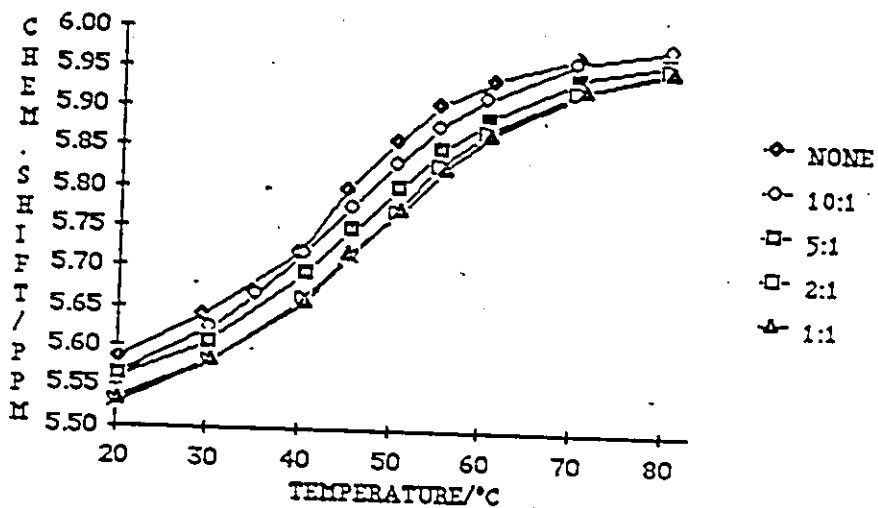


Figure 5.25: Changes in chemical shift for C(2)H-5 in CCGG upon addition of proflavine at ratios shown.

Table 5.26: Chemical shifts of the hydrogen-bonded imino protons of CCGG, 2.5mM in 0.1M salt, with proflavine, at ratios 10:1 to 3:2 single strand:drug.

	Chemical shift (ppm)				
	<u>-3.4°</u>	<u>7.5°</u>	<u>17.0°</u>	<u>25.0°</u>	
CCGG					
TERMINAL	13.34	13.30	13.25	13.05	
INTERNAL	12.53	12.53	12.46	12.39	
10:1					
TERMINAL	<u>-4.6°</u>	<u>5.4°</u>	<u>14.8°</u>	<u>24.9°</u>	
INTERNAL	13.21	13.17	13.10	12.79	
	12.49	12.49	12.41	12.33	
4.6:1					
TERMINAL	<u>-4.8°</u>	<u>6.1°</u>	<u>15.0°</u>	<u>25.3°</u>	<u>30.2°</u>
INTERNAL	13.07	13.07	13.06	12.79	--
	12.44	12.46	12.42	13.39	12.38
				12.23	12.21
3.5:1					
TERMINAL	<u>-4.5°</u>	<u>5.2°</u>	<u>15.5°</u>	<u>25.0°</u>	<u>34.8°</u>
INTERNAL	13.04	12.98	12.86	--	--
	12.45	12.40	12.38	12.39	12.36
	--			12.21	11.98
2:1					
TERMINAL	<u>-4.6°</u>	<u>5.1°</u>	<u>15.0°</u>	<u>24.8°</u>	<u>32.0°</u>
INTERNAL	12.90	12.89	12.82	--	--
	12.42	12.40	12.37	12.36	12.36
				12.16	12.05
3:2					
TERMINAL	<u>-4.3°</u>	<u>5.6°</u>	<u>14.9°</u>	<u>25.1°</u>	<u>31.4°</u>
INTERNAL	12.85	12.83	12.62	--	--
	12.42	12.39	12.36	12.32	12.36
				12.15	11.99



Table 5.27: Chemical shifts of the hydrogen-bonded imino protons of CCAGG, 2.5mM in 0.1M salt buffer, with proflavine, at single strand:drug ratios of 6:1 to 1:1.

Chemical shift (ppm)			
<u>CCAGG</u>	<u>-4.5°</u>	<u>2:1</u>	<u>-4.6°</u>
Proton			
TERMINAL	13.38	TERMINAL	13.16
INTERNAL	12.22	INTERNAL	12.18
<u>6:1</u>	<u>-4.5°</u>	<u>1:1</u>	<u>-4.5°</u>
TERMINAL	13.25	No visible spectrum.	
INTERNAL	12.19		

### 5.2.3. Duplex stabilization

Addition of proflavine to GCGC ( $T_m$  48.1°C), CCGG ( $T_m$  44.1°C) and GCUGC ( $T_m$  20.0°C) produces  $T_m$  increases of up to 5.3°, 4.8° and 15.6°C respectively. Maximum stabilization is reached at single strand:drug ratios of 4:1 to 2:1, and thereafter the  $T_m$  starts to decline. The proflavine protons also show sigmoidal behaviour and a  $T_m$  which is persistantly 15 - 25°C higher than the duplex  $T_m$ .

No conclusion may be reached from these data as to the relative affinity of binding to the parent duplex GCGC and to the mismatch-containing duplex GCUGC. The  $T_m$ s are too widely separated. Estimates of  $\Delta T_m$  obtained by entering plausible increments of  $\Delta S$  and  $\Delta H$  (resulting from intercalator binding) into the equation  $T_m = \Delta H / (\Delta S + \ln CT)$ , where  $CT$  is the total strand concentration, show that a twofold difference in stabilization of a duplex with  $T_m = 20^\circ\text{C}$  and a duplex with  $T_m = 48^\circ\text{C}$  can result for binding with the same  $\Delta S$  and  $\Delta H$ . Introduction of the stabilizing element produces greater effect on the less stable duplex.

The fourth sequence studied, CCAGG, experienced up to 5°C stabilization upon addition of intercalator. There were two major difficulties on the studies of CCAGG: pronounced broadening of the resonances at low temperature and degradation of the sequence.  $T_m$ s measured for the proflavine protons for CCAGG are around 40°C, compared to the maximum of 53°C measured for GCUGC; the duplex formed by CCAGG upon intercalation is less stable than that formed by GCUGC.

#### 5.2.4. Proflavine orientation.

The upfield chemical shifts observed for the proflavine protons in all these oligonucleotide:proflavine complexes are large: pro-a > 0.4 ppm, pro-b > 0.5 ppm, pro-c > 0.55 ppm and pro-d > 0.55 ppm. These values are lower limits, since the chemical shift versus temperature curves already have appreciable slope even at 70 - 80°C (Figure 5.26). These shifts support an intercalation geometry in which the long axis of the proflavine molecule is lying parallel to the long axis of the base pairs between which it is intercalated. Any appreciable stacking perpendicular to the long axis would result in the proton on the central ring being more shielded than those on the other two—which would be protruding into solution. This is not the case. This result concurs with the geometry observed in the crystal structures of the complexes with proflavine of CG (Neidle, 1981), dCG (Shieh *et al.*, 1980) and CA:UG (Aggarwal *et al.*, 1984), and that deduced for the solution structures of complexes of proflavine with d(GGCC), d(CCGG) (Patel and Camuel, 1977) and poly(dA-dT) (Patel, 1977b).

#### 5.2.5. Binding sites: GCGC and CCGG

The duplexes formed by the tetramers GCGC and CCGG each have three sites for intercalation. An internal CG:CG site is common to both, while the terminal sites of GCGC are GC:GC and the terminal sites of CCGG are GG:CC. Plausible binding sites and combinations of binding sites are depicted in Figure 5.27. The degeneracy of each configuration due to the self-complementarity of the duplex is noted beside the structure.

Maximum duplex stabilization of GCGC occurs around 2:1 single strand:drug, or

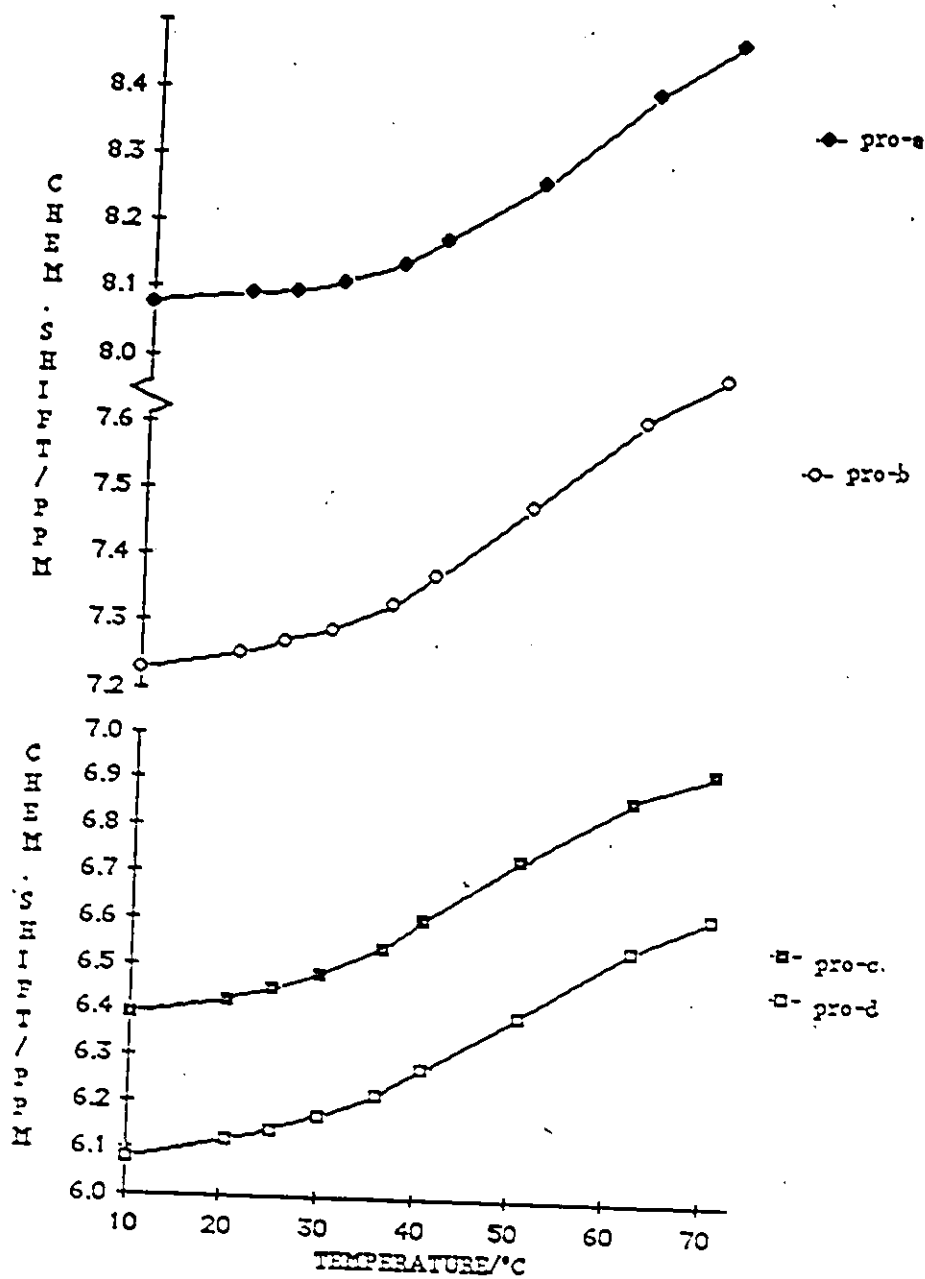


Figure 5.26: Representative chemical shift versus temperature curves for the protons of proflavine. For GCUGC with proflavine, 4:1 single strand:drug.

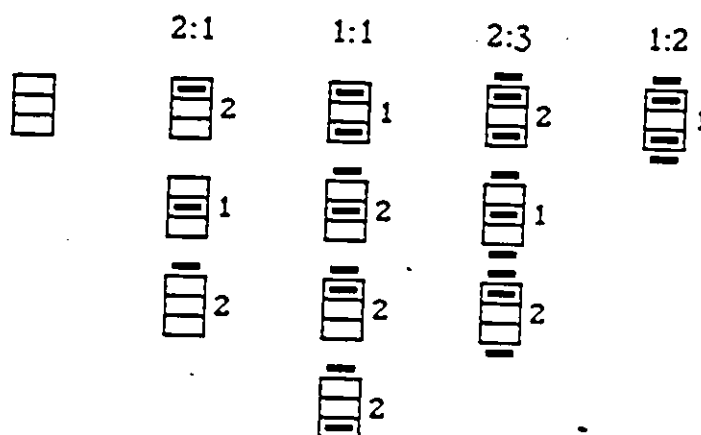


Figure 5.27: Possible intercalation sites and combinations of intercalation sites in GCGC and CCGG, for single strand:drug ratios shown. Numbers beside structures represent the statistical occurrence of that particular structure.

1:1 duplex:drug, suggesting that the optimum binding mode occurs at one drug molecule per duplex (Table 5.12). The maximum  $T_m$  of the proflavine protons is reached at even lower single-strand:drug ratios, about 4:1.

Maximum duplex stabilization of CCGG occurs at slightly lower ratios, between 3.5:1 and 2:1 for the duplex, and at 4.6:1 for the drug (Table 5.24)

According to the ring current calculations of Geissner-Prettre and Pullman (Geissner-Prettre and Pullman, 1982), the three membered ring of proflavine has greater shielding capacity than the two-membered purine ring or one-membered pyrimidine ring. Thus, intercalation might be expected to produce an upfield shift in the protons of the adjacent base-pairs. A gradual upfield shift is noted for all protons in GCGC and CCGG upon addition of intercalator (Figures 5.4 to 5.9 and 5.20 to 5.24; Tables 5.28 and 5.29), with the largest shifts occurring in the 5' terminal residues, G(1) and C(2) in GCGC and C(1) and C(2) in CCGG. The behaviour of C(2)H-6 is similar in both sequences: the sigmoidal behaviour of the curve is gradually lost as the curve moves upfield. However, these shift changes are not matched by shifts in the protons of the bases to which they are base paired, and the nonexchangeable protons of guanosine and cytidine lie to the outside of the stack; it is possible that these shift changes result from conformational changes occurring upon intercalation, and are not directly due to the intercalator.

The imino protons, lying as they do in the centre of the stack, might be expected to provide a better indication of the shielding properties of the base pair neighbours. In both cases, the hydrogen-bonded imino protons belonging to the terminal base pair show greater upfield shifts than those belonging to the internal imino protons (Tables 5.14 and 5.26). This trend establishes itself even at low proflavine ratios, where substantial end-stacking would not be expected; it might be that intercalator is trapped

Table 5.28: Chemical shift changes for the aromatic base protons of GCGC with proflavine at 2:1 single strand:drug ratio, relative to GCGC in the absence of proflavine. (Upfield shifts negative).

Proton	Chemical shift (ppm)		
	$\Delta\delta$ at 70°C	$\Delta\delta$ at 30°C	$\Delta\delta(30^\circ) - \Delta\delta(70^\circ)$
G(1)H-8	+0.05	-0.037	-0.042
C(2)H-6	-0.017	-0.098	-0.081
G(3)H-8	-0.047	-0.031	+0.016
C(4)H-6	-0.032	-0.059	-0.027
C(2)H-5	-0.048	-0.068	-0.020
C(4)H-5	-0.045	-0.047	-0.002

Table 5.29: Chemical shift changes for the aromatic base protons of CCGG with proflavine, 2:1 single strand:drug ratio, relative to CCGG in the absence of proflavine. (Upfield shifts negative).

Proton	Chemical shift (ppm)		
	$\Delta\delta$ at 70°C	$\Delta\delta$ at 20°C	$\Delta\delta(20^\circ) - \Delta\delta(70^\circ)$
C(1)H-6	+0.002	-0.070	-0.072
C(2)H-6	-0.050	-0.215	-0.165
G(3)H-8	-0.040	-0.048	-0.008
G(4)H-8	-0.042	-0.012	+0.030
C(1)H-5	-0.034	-0.177	-0.143
C(2)H-5	-0.043	-0.111	-0.068

by end-to-end aggregation, although the trend persists to higher temperatures, at which the aggregate would be expected to be beginning to break up.

♦ Addition of high proflavine ratios results in the appearance of an additional peak, upfield from the others. The intensity of this peak appears to be dependent upon proflavine ratio: it is absent from the spectra of GCGC with proflavine at 6:1 single-strand:drug, and of equal intensity to the others at 2:1 single strand:drug, and in the spectra of CCGG it steadily increases in intensity from 3.5:1 to 1:1 single strand:drug. Slow exchange between species has been observed to give rise to extra peaks in spectra of oligoribonucleotides (Reid *et al.*, 1985) and oligoribonucleotide-drug mixtures (Tran-Dinh *et al.*, 1984). Figure 3.28 shows the imino proton pattern expected for the three interior intercalation sites in slow exchange. Superposition of peaks seems likely, given the observation of only one new peak, and it is not possible to deduce where the drug intercalates.

Comparison of the Tms belonging to the interior base protons and to the exterior base protons shows a slight trend towards stabilization of the interior bases (Tables 5.30 and 5.31). These averages, however, are generated from the results of two, or at most, three protons, and the uncertainty associated with the Tms is of similar order to the difference between them.

From the uniformity of effect of addition of proflavine throughout the duplex, it would seem that these duplexes exhibit no site-specificity: The intercalator enters all sites equally, at least in the concentration range used in these experiments.

#### 5.2.6. Binding sites: GCUGC.

The CUG sequence in GCUGC, which replaces the CG site in GCGC, might be



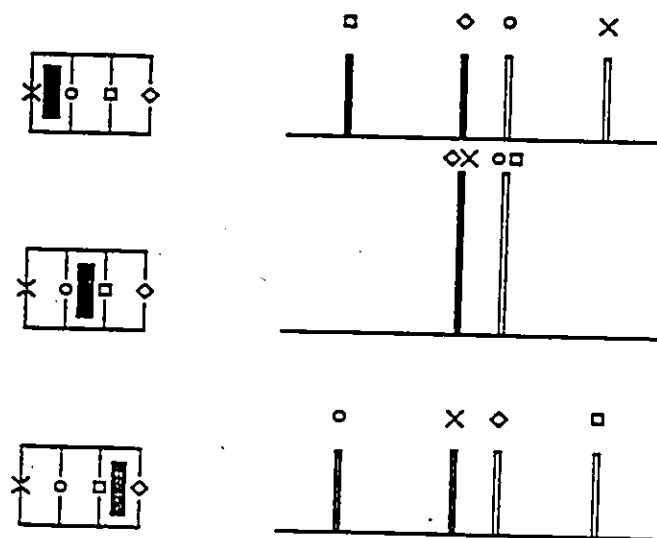


Figure 5.28: Pattern of imino proton spectra expected for intercalation into the three sites of GCGC. (For CCGG reverse the order of the internal and terminal resonances.)

Table 5.30: Comparison of average  $T_m$ s for the base protons belonging to the terminal (G(1) and C(4)) and internal (C(2) and G(3)) residues of GCGC with proflavine.

Ratio Proton	$T_m$ ( $^{\circ}\text{C}$ )					
	15:1	4:1	2:1	1:1	1:2	2:1
Terminal	48.2	49.2	49.3	53.1	47.2	40.3
Internal	48.5	49.6	50.7	56.5	47.8	42.0
$\Delta T_m$	0.3	0.7	1.4	3.4	0.6	1.7

Table 5.31: Comparison of average  $T_m$ s for the base protons belonging to the terminal (C(1) and G(4)) and internal (C(2) and G(3)) residues of CCGG, with proflavine.

Ratio Proton	$T_m$ ( $^{\circ}\text{C}$ )					
	CCGG	10:1	4.6:1	3.5:1	2:1	1:1
Terminal	45.0	48.0	48.7	48.2	50.1	50.3
Internal	44.6	45.2	46.9	47.6	48.0	47.6
$\Delta T_m$	-0.4	-2.8	-1.7	-0.5	-2.1	-2.7

expected to act as one or two possible intercalation sites depending upon whether the uridines are stacked in or looped out of the duplex. If the uridines are stacked in, the effect is to introduce an additional intercalation site, or portion thereof (Figure 5.29a). While if the uridines are looped out, the number of internal intercalation sites will be the same as in GCGC, but the affinity of the central site will be different (Figure 5.29b). The affinities of the two terminal sites may also be affected by changes in geometry introduced by the uridine-uridine mismatch. In this instance, the number of sites is further increased by the provision of more external binding sites—the intercalator may stack on the looped out base, giving rise to a great many possible structures.

The increase in the  $T_m$  of GCUGC reaches a maximum between 2:1 and 1:1 single strand:drug (Table 5.13), at slightly higher ratios than in the parent duplex GCGC (Table 5.12). The proflavine  $T_m$  increase is at a maximum around 2:1.

An examination of the behaviour of the individual proton curves and  $T_m$ s show no changes which might be indicative of a change in conformation of the duplex from the duplex in the absence of proflavine.  $T_m$ s all increase, some more rapidly than others (Table 5.27). A general trend towards greater shielding is observed for all protons (Figures 5.10 to 5.17), with the curves of three protons, C2H5 and two of the H1's, which were nonsigmoidal in the duplex without proflavine, becoming sigmoidal at high proflavine concentrations. The uridine base protons show the same magnitudes of increase in shielding as other protons—though it should be noted that proflavine stacking on a bulged uridine would produce an increase in shielding in that residue's protons. However, the coupling constants continue to collapse in similar order and at similar relative rates as in the duplex without proflavine, in which the uridines stack in to the helix.

Selective stabilization of the internal base pairs is observed in this sequence

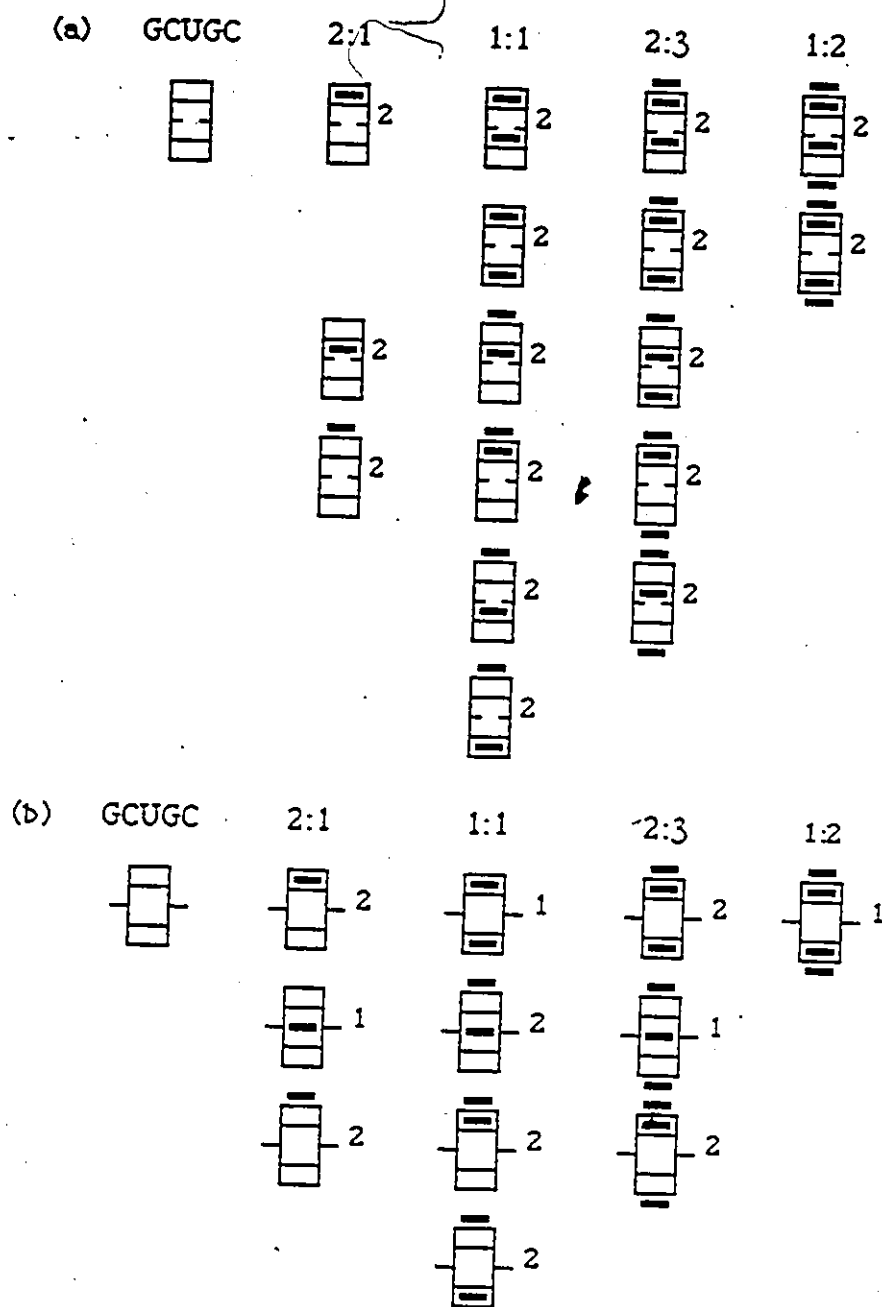


Figure 5.29: Intercalation sites in GCUGC, (a) mismatched bases stacked in to the helix (b) mismatched bases looped out. Statistical occurrence shown beside each structure.

as well, when the average  $T_m$ s of the internal base pairs are compared to those of the terminal base pairs (Table 5.32). Again, the averages are drawn from a small number of curves.

The imino protons show similar trends to those in GCGC in going from no proflavine to 6:1 single strand:drug--greater shielding of terminal than internal imino protons (Table 5.15), but substantial changes occur from 6:1 to 2:1. The signal around 13.4 - 13.3ppm, belonging to the internal G-C imino base pair, has lost intensity relative to that at 12.5ppm, belonging to the terminal G-C base pair. The UNH3 proton signal is greatly reduced in intensity. Apparently addition of proflavine promotes exchange of imino protons from the interior of the duplex, or slow exchange is occurring between species with very similar imino proton shifts.

Loss of the UNH3 signal might be expected if the proflavine were intercalating into the central site and forcing the uridine residues to bulge out into solution. However the apparent stoichiometry--with the peak of the stabilization versus concentration of intercalator curve occurring at higher ratios in GCUGC than in the parent duplex GCGC--the increased shielding seen by the non-exchangeable protons of the mismatched uridines, and the unchanged pattern of collapse of coupling constants, all suggest that the uridines are also stacking in to provide an extra intercalation site. There does not appear to be one intercalation geometry or intercalated species predominating.

### 5.3. Discussion

The increase of duplex  $T_m$  upon addition of proflavine is ascribable to preferential binding of the intercalator to the duplex which shifts the equilibrium to

Table 5.32: Comparison of average  $T_m$ s for the base protons belonging to the terminal (G(1) and C(5)) and internal (C(2) and G(4)) residues of GCUGC with proflavine.

Ratio Proton	$T_m$ ( $^{\circ}\text{C}$ )						
	GCUGC	20:1	6:1	4:1	2:1	1:1	1:2
Terminal	21.0	21.0	25.9	25.9	35.7	34.8	0.8
Internal	20.2	20.1	26.9	31.2	38.7	36.9	31.9
$\Delta T_m$	-0.8	-0.9	+1.0	+5.3	+3.0	+2.4	+0.8

favour the duplex (Cantor and Schimmel, 1981). At low ratios of proflavine, the difference between drug and duplex  $T_m$ s is to be expected, since all proflavine will be intercalating, but only a fraction of the base pairs will be associated with intercalator—and therefore subject to stabilization—at any given time. However, this difference persists to higher ratios, which leads to the supposition that intercalation is being accomplished by partial melting of the duplex and capture of the intercalator, rather than by insertion of the intercalator into an intact duplex—if the intercalator is entering a partially denatured duplex, then its  $T_m$ , which monitors its interaction with the duplex, will be higher than the duplex  $T_m$ , which monitors duplex self-association. The levelling off or subsequent decline of the proflavine  $T_m$  is to be ascribed to greater contribution by lower-stability, non-intercalative types of binding.

No intercalation site preference is determinable from the NMR data for GCGC, GCUGC, CCGG and CCAGG with proflavine. The current state of knowledge regarding intercalation into mismatch containing sequences suggests that the preferred site for intercalation may be the site adjacent to the mismatch, i.e., a fully base paired site—since intercalators do not appear to interact with single strands—next to a site of destabilization, which will allow for capture of the intercalator. The same effect is liable to be operative at the duplex ends, fraying of which would allow capture of the intercalator. In short duplexes, therefore, end effects would obscure sequence effects.

The binding of proflavine to the deoxy-analogue of CCGG, d(CCGG) has been studied by  $^1\text{H-NMR}$  and optical methods (Patel and Canuel, 1977b). The  $T_m$ s and complexation shifts of d(CCGG) show similar trends to those of CCGG: For d(CCGG) ( $T_m = 37^\circ\text{C}$  in absence of proflavine), at 7:1 single strand:duplex, the proflavine  $T_m$  was  $54^\circ\text{C}$ , compared to  $62^\circ\text{C}$  for CCGG ( $T_m$  in absence of proflavine,  $44^\circ\text{C}$ ) (Patel and Canuel,

1977b). No data was given for the duplex in presence of proflavine. The complexation shifts for the individual base protons of  $d(\text{CCGG})$  are of similar magnitude to those seen for CCGG, and in both sequences the largest shifts occur in the terminal residues. In  $d(\text{CCGG})$  the descending order of magnitude is:  $\text{C}(1)\text{H}-5 > \text{C}(1)\text{H}-6 > \text{C}(2)\text{H}-6 > \text{C}(2)\text{H}-5 > \text{G}(3)\text{H}-8$  and  $\text{G}(4)\text{H}-8$ , while in CCGG the order is:  $\text{C}(2)\text{H}-6 > \text{C}(1)\text{H}-5 > \text{C}(1)\text{H}-6 > \text{C}(2)\text{H}-5 > \text{G}(3)\text{H}-8 > \text{G}(4)\text{H}-8$ . The proflavine chemical shifts quoted for  $d(\text{CCGG})$  are approximately double those found for CCGG, however, the curves in the former study are more complete, the variable temperature runs having been started at  $95^\circ\text{C}$  as opposed to  $80^\circ\text{C}$ . No imino proton data was quoted. A comparison of binding of proflavine to  $d(\text{CCGG})$  and  $d(\text{GGCC})$  was carried out using optical binding curves, and it was from this data that a preference from the  $d(\text{CG}):d(\text{CG})$  core was deduced (Patel and Canuel, 1977b).

Intercalated species in slow exchange have been observed for  $d(\text{CGm}^5\text{CG})$  and  $d(\text{m}^5\text{CGCG})$  with daunomycin. Discrete methyl signals for each of the two discrete intercalated species (intercalation into the terminal and central sites) are visible, leading to the conclusion that intercalation is occurring into all sites equally (Tran-Dinh *et al.*, 1984).

Ethidium binds to CUG to form a species in which the central  $\text{UxU}$  mismatch is excluded from the stack to accommodate the intercalator (Lee and Tinoco, 1978). The imino proton results for GCUGC suggest that there is contribution of such a species to the conformational mix in the GCUGC:proflavine complex, but the overall results indicate that it is not the only conformation.

From comparison of the results for GCGC and GCUGC it is evident that proflavine stabilizes the duplex containing the mismatch to a greater degree. Whether this indicates preferential binding, or is simply due to the greater



contribution of addition of a stabilizing element to a weak duplex had yet to be established.

#### 5.4. Conclusions

Stabilization of the mismatch-containing duplex GCUGC upon addition of proflavine was greater than stabilization of either of the perfect parent duplexes GCGC or CCGG; whether this indicated stronger binding could not be determined, as the greater stabilization of GCUGC might have been a consequence simply of its being initially much less stable—introduction of a stabilizing element has more effect on a less stable duplex. CCAGG apparently was not stabilized by intercalation, but this sequence was susceptible to cleavage, and the sample degraded with handling.

Binding site preferences in the parent duplex were not readily observable; changes in  $T_m$  and chemical shift upon addition of proflavine were distributed equally throughout the duplex. In the imperfect duplexes there was no evidence of preference for the central base-base mismatches, or for a conformation in which the mismatched bases loop out. The rapid movement of the intercalator into and out of the duplex, combined with the lack of strong site preferences, results in averaging of the observed parameters over all conformations.

The duplex containing a UxU mismatch, GCUGC, appeared more accommodating of the intercalator than that containing the AxA mismatch, according to the 10°C difference measured for the proflavine  $T_m$ .

## 6. CONCLUSIONS

The study of biological macromolecules by physical methods frequently requires isolation of particular structural features in a model system so that their individual properties may be studied and the effect of perturbations and substitutions monitored. Physical studies of synthetic oligonucleotides incorporating features important in native nucleic acids are a necessary complement to physical and biological studies of the native molecules themselves.

Nuclear magnetic resonance is a particularly powerful means of studying nucleic acid model systems, since information on stability and details of structure of nucleic acid secondary structures may be obtained from chemical shift and coupling constant data. Moreover, samples for study by NMR are dissolved in aqueous buffer solutions constituted so as to approximate physiological conditions.

One of the earliest issues addressed in synthetic oligoribonucleotide studies was sequence dependence of stability. The nearest neighbour approximation states that any physical property of a polymer may be described by summation of the interactions between adjacent constituents, in the case of nucleic acids, adjacent base pairs. This approximation has been employed in the prediction of the stability of proposed secondary structures of rRNAs and mRNAs, which cannot yet be directly observed, from the stabilities of the individual two base pair units, "diribonucleotide cores". At the time of development of the first project of this thesis (described in section 3), the thermodynamic values for the stabilities of the three GC diribonucleotide cores, GC:GC, GG:CC and CG:CG, had been derived from limited data, and were outdated. Study of the

duplexing behaviour of six tetramers containing exclusively guanosine and cytidine residues allowed deduction of a relative ranking of stabilities  $GC:GC \geq GG:CC > CG:CG$ , which differed from those previously published in the ordering of the first two cores. Thermodynamic studies on these sequences and longer sequences derived from them were performed by Drs. Freier and Turner in Rochester and provided quantitative results that could be directly applicable in structure prediction (Freier *et al.*, 1985b). Comparison of the duplex stabilities according to NMR with literature values and results for the deoxy analogues showed greater stability for the ribo tetramers.

NMR data also contains structural information. The larger chemical shift changes upon duplexing observed for certain protons in the RNA duplexes compared to the literature results for their DNA analogues were qualitatively consistent with the base pair overlap differences between classical A- and B- helices. Deviations in the imino proton data from predicted values from literature parameters derived from data for native tRNA suggested that the conformation at the blunt ends differ from the classical values. Application of the RY model for sequence dependent winding in RNA helices to the tetramer data allowed correlation of the rate of collapse of the anomeric coupling constants of the 5' terminal residue with the degree of overlap of the 5' terminus with adjacent base.

The extension of the study to mixtures of complementary trimers, the precursors of the tetramers and related sequences, produced unexpected results. For three of these four mixtures in 1.0 M salt buffer the G-rich strand aggregated, preventing duplex formation. Upon removal of salt three of the four mixtures duplexed to give a measurable  $T_m$ , and the fourth was beginning to duplex; aggregation was beginning at much lower temperatures.

The observation of a distinct imino proton spectrum for GCG, and the availability of information regarding the behaviour of related sequences, led to the proposal of a structure for the GCG aggregate. Investigation of aggregate structures by NMR is hampered by the extreme broadening of the signals due to chemical shift anisotropy and exchange, and this aspect of the study would benefit from application of another method; it may be possible to crystallize GCG, for instance.

Mispaired bases are a common feature of the secondary structure of native RNAs, perhaps required to facilitate the melting of a helix when a transition between secondary structures is required, or to provide a site for protein binding. Mismatched bases are usually introduced as a perturbation in model systems, and their effect on stability and conformation are described relative to the related perfect duplex. Studies on six pentaribonucleotides (section 4) derived from the three most stable tetramers established that mismatches were appreciably destabilizing. Of the six sequences studied, only two produced duplexes containing the desired central mismatch, CCAGG and GCUGC. Introduction of both these mismatches (AxA and UxU) produced comparable destabilization, and in both cases, the mispaired bases stacked into the duplex. This is common to all systems described in the literature; it appears that continuation of the stacking through the mismatch is more stable than forcing the mismatch out.

No comparison could be made of the effect of disruption of the central GC core in the tetramers by the mismatched bases since only two of the six potential duplexes formed. However, mixing experiments performed with the complementary pentamers allowed comparison of the effect of introduction of a single A•U base pair into the centre core of the three GC tetramers. Greater stabilization was observed for insertion into the weaker CG:CG core than into the GC:GC core.

The mismatch-containing duplexes were used in a study of intercalation into mismatch-containing duplexes (section 5). According to the Streisinger hypothesis, frameshift mutagens arise from fixing of transient unpaired structures by repair enzymes, and mutagens exert their effects by stabilizing these mispairings prior to repair. Literature results suggested stronger intercalator binding to duplexes containing mismatch sites. The intercalation of proflavine into CCAGG and GCUGC (and their parent duplexes CCGG and GCGC) was studied to determine whether there was a demonstrable preference for intercalation into a mismatch site, greater stabilization of an imperfect duplex, or conspicuous alteration in the conformation of the mismatch site—for instance looping out of mismatched bases to accommodate the intercalator. Addition of increasing amounts of proflavine produced a steady increase in  $T_m$  for the systems studied, and a slight increase in shielding for most protons, but no indication of site preference. The mismatch-containing duplex GCUGC tabled a threefold higher  $T_m$  increase than GCGC, but the starting  $T_m$ s were too widely separated to allow a direct comparison. There was no indication that mismatched bases were looped out into solution.

From these results, and from the most recent results emerging in the literature, it appears that the preferred intercalation site may in fact be adjacent to a site of destabilization, such as a mismatch, or an end. The mismatch site itself may have too much single stranded character to have affinity for an intercalator, but it promotes intercalation at the adjacent sites by destabilizing the adjacent base pairs, facilitating opening and capture of the intercalator. The inconclusive results for the tetramers and pentamers studied is probably as a result of their being too short to provide a normal site for comparison with the mismatch sites.

## REFERENCES

- Aboul-ela, F., Koh, D. and Tinoco, L, Jr. (1985) *Nucleic Acids Res.*, **13**, 4811.
- Aggarwal, A., Islam, S.A., Kurod, R. and Neidle, S. (1984) *Biopolymers*, **23**, 1025.
- Alkema, D. (1982) Ph.D. thesis. McMaster University.
- Alkema, D., Bell, R.A., Hader, P.A. and Neilson, T. (1981) *J. Am. Chem. Soc.*, **103**, 2866.
- Alkema, D., Hader, P.A., Bell, R.A. and Neilson, T. (1982) *Biochemistry*, **21**, 2109.
- Altona, C. (1982) *Recl. Trav. Chim. Pays. Bas.*, **101**, 413.
- Altona, C. and Sundaralingham, M. (1972) *J. Am. Chem. Soc.*, **94**, 8205.
- Altona, C. and Sundaralingham, M. (1973) *J. Am. Chem. Soc.*, **95**, 2333.
- Arnott, S. and Hulkins, D.W.L. (1973) *J. Mol. Biol.*, **81**, 93.
- Arnott, S., Hulkins, D.W.L., Dover, S.D., Fuller, W. and Hodgson, A.R. (1973) *J. Mol. Biol.*, **81**, 107.
- Arter, D.B. and Schmidt, P.G. (1976) *Nucleic Acids Res.*, **3**, 1437.
- Arter, D.B., Walker, G.C., Uhlenbeck, O.C. and Schmidt, P.G. (1974) *Biochem. Biophys. Res. Comm.*, **61**, 1089 (1974)
- Aue, W.P., Bartholdi, E. and Ernst, R.R. (1976) *J. Chem. Phys.*, **64**, 2229.
- Auron P.E., Ridone, P.E., Vary, W.P., Celentano, J.J. and Vournakis, J.N. (1982) *Nucleic Acids Res.*, **10**, 3517.
- Bell, R.A., Everett, J.R., Hughes, D.W., Alkema, D., Hader, P., Neilson, T. and Romaniuk, P.J. (1981) *Biopolymers*, **20**, 1383.
- Berman, H.M., Neidle, S. and Slodola, R.K. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 828.
- Berman, H.M. and Shieh, H.-S. (1981) In "Topics in Nucleic Acid Structure", S. Neidle, ed. Wiley and Sons, NY. pp 17 - 32.
- Berman, H.M. and Young, P.R. (1981) *Ann. Rev. Biophys. Bioeng.*, **10**, 87.

- Berman, H.M., Stallings, W., Carnell, H.L., Glusker, J.P., Neidle, S., Taylor, G. and Achari, A. (1978) *Biopolymers*, 18, 2405.
- Bolton, P.H. (1982) *J. Magn. Reson.*, 61, 306.
- Borer, P.N., Dengler, B., Tinoco, I., Jr. and Uhlenbeck, O.C. (1974) *J. Mol. Biol.*, 86, 843.
- Borer, P.N., Kan, L.S. and T'so, P.O.P. (1974) *Biochemistry*, 14, 4847.
- Broido, M.S., Zon, G. and James, T.L. (1984) *Biochem. Biophys. Res. Comm.*, 119, 663.
- Brown, T., Kennard, O., Kneale, G. and Rabinovich, D. (1985) *Nature*, 315, 604.
- Bubienko, E., Cruz, P., Thompson, J.F. and Borer, P.N. (1983) *Prog. Nucl. Acids Res. Mol. Biol.*, 30, 41.
- Buchko, G. (1985) M.Sc. Thesis, McMaster University.
- Callandine, D.R. *J. Mol. Biol.*, 161, 343.
- Cantor, C.R. and Schimmel, P.R. (1981) "Biophysical Chemistry: III The Behaviour of Biological Macromolecules." W.H. Freeman and Company, San Francisco.
- Carey, J., Lowary, P.T. and Uhlenbeck, O.C. (1983) *Biochemistry*, 22, 4723.
- Cheng, D.M., Danyluk, S.S., Dhingra, M.M., Ezra, F.S., MacCross, M., Mitra, C.K. and Sarma, R.H. (1980) *Biochemistry*, 19, 2491.
- Chuprina, V.P. and Poltev, V.I. (1983) *Nucleic Acids Res.*, 11, 5205.
- Clore, G.M. and Gronenborn, A.M. (1983a) *J. Magn. Res.*, 53, 423.
- Clore, G.M. and Gronenborn, A.M. (1983b) *EMBO J.*, 2, 2109.
- Clore, G.M. and Gronenborn, A.M. (1984) *Eur. J. Biochem.*, 141, 119.
- Clore, G.M. and Gronenborn, A.M. (1985) *FEBS Lett.*, 179, 187.
- Clore, G.M., Gronenborn, A.M. and McLauchlin, L.W. (1985) *Eur. J. Biochem.*, 151, 153.
- Conner, B.N., Takano, T., Tanaka, S., Ikatura, K. and Dickerson, R.E. (1982) *Nature*, 295, 294.
- Crawford, J.L., Kolpak, F.J., Wang, A.H.J., Quigley, C.G., van Boom, J.H., van der Marel, G. and Rich, A. (1980) *Proc. Natl. Acad. Sci. USA*, 77, 4016.

- Crick, F.H. (1976) *J. Mol. Biol.*, 19, 548.
- Crick, F.H. and Watson, J.D. (1953). *Proc. Roy. Soc (London) Ser. A.*, 223, 80.
- Cruz, P., Bubienko, E. and Borer, P.N. (1982) *Nature*, 298, 198.
- D'Andrea, P., Alkema, D., Bell, R.A., Coddington, J.M., Hader, P.A., Hughes, D.W. and Neilson, T. (1983) *J. Am. Chem. Soc.*, 105, 636.
- Dhingra, M.M. and Sarma, R.H. In "Stereodynamics of Biomolecular Systems," R.H. Sarma, ed. Pergamon Press, Oxford. p. 3.
- Dickerson, R.E. (1983) *J. Mol. Biol.*, 166, 419.
- Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.H., Fratani, A.V., Kopka, M.L. (1982) *Science*, 216, 475.
- Dodgson, J.B. and Wells, R.D. (1977) *Biochemistry*, 16, 2367.
- Doornbos, J., Wreesman, C.T.G., van Boom, J.H. and Altona, C.A. (1983) *Eur. J. Biochem.*, 131, 571.
- Drew, H., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R.E. (1980) *Nature*, 286, 567.
- Eich, G., Bodenhauser, G., and Ernst, R.R. (1982). *J. Am. Chem. Soc.*, 104, 3731.
- 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.G. and Sarma, R.H. (1974) *J. Biol. Chem.*, 249, 4745.
- Everett, J.R., Hughes, P.W. and Bell, R.A. (1980) *Biopolymers*, 19, 557.
- Feigon, J., Wright, J.M., Leupin, W., Denny, W.A. and Kearns, D.R. (1982) *J. Am. Chem. Soc.*, 104, 5540.
- Feigon, J., Denny, W.A., Leupin, W. and Kearns, D.R. (1983a) *Biochemistry*, 22, 5930.
- Feigon, J., Leupin, W., Denny, W.A. and Kearns, D.R. (1983b) *Biochemistry*, 22, 5943.
- Feigon, J., Wang, A.H.J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1984) *Nucleic Acids Res.*, 12, 1243.



- Fink, T.R. and Crothers, D.M. (1972) *J. Mol. Biol.*, 66, 1.
- Frechet, D., Cheng, D.M., Kan, L.S. and T'so, P.O.P. (1983) *Biochemistry*, 22, 5194.
- Freier, S., Petersheim, M., Hickey, D.R. and Turner, D.H. (1984) *J. Biomol. Struct. Dynam.*, 1, 1229.
- Freier, S., Sinclair, A., Neilson, T. and Turner, D.H. (1985a) *J. Mol. Biol.*, 185, 645.
- Freier, S., Alkema, D., Sinclair, A., Neilson, T. and Turner, D.H. (1985b) *Biochemistry*, 24, 4533.
- Fujii, S., Wang, A.H.-J., Quigley, G.J., Westerink, H., van der Marel, G., van Boom, J.H. and Rich, A. (1985) *Biopolymers*, 24, 243.
- Gellert, M., Lipset, M.N. and Davies, D.R. (1962) *Proc. Natl. Acad. Sci. USA*, 48, 2013.
- Giessner-Prettre, C. and Pullman, B. (1965) *C.M. hebdomadaire Seances Acad. Sci. Paris*, 261, 2521.
- Giessner-Prettre, C. and Pullman, B. (1969) *C.M. hebdomadaire Seances Acad. Sci. Paris*, 268, 1115.
- Giessner-Prettre, C. and Pullman, B. (1970) *J. Theor. Biol.*, 27, 89.
- Giessner-Prettre, C. and Pullman, B. (1976a) *Biochem. Biophys. Res. Comm.*, 90, 578.
- Giessner-Prettre, C. and Pullman, B. (1976b) *C.R. Hebd. Seances Acad. Sci. Ser. D*, 283, 675.
- Giessner-Prettre, C. and Pullman, B. (1982) *Biochem. Biophys. Res. Comm.*, 107, 1539.
- Goldfield, E.M., Luxton, B.A., Bowie, V. and Gorenstein, D.G. (1983) *Biochemistry*, 22, 3336.
- Gorenstein, D.G. (1975) *J. Am. Chem. Soc.*, 97, 898.
- Gorenstein, D.G., Findlay, J.M., Kommi, R.K., Luxon, B.A. and Kan, D. (1976) *Biochemistry*, 15, 3796.
- Gottesman, M., Oppenheimer, A. and Court, D. (1982) *Cell*, 29, 727.
- Gralla, J. and Crothers, I.M. (1973a) *J. Mol. Biol.*, 73, 497.

- Gralla, J. and Crothers, J.M. (1973b) *J. Mol. Biol.*, 78, 301.
- Gregoire, R.J. and Neilson, T. (1978) *Can. J. Chem.*, 56, 487.
- Griffin, B.C., Jarman, M., Reese, C.B., Sulston, J.E. and Trentham, J.R. (1968) *Biochemistry*, 5, 3638.
- Gronenborn, A.M. and Clore, G.M. (1985) *Prog. NMR Spectroscopy*, 17, 1.
- Gronenborn, A.M., Bimblrier, B.J., Clore, G.M. and McLaughlin, L.W. (1983) *Nucleic Acids Res.*, 11, 5691.
- Grosjean, H.J., de Henau, S. and Crothers, D.M. (1978) *Proc. Natl. Acad. Sci. USA*, 75, 610.
- Hader, P.A., Neilson, T., Alkema, D., Kofoid, E.C. and Ganoza, C. (1981) *FEBS Lett.*, 136, 65.
- Hare, D.R., Weimer, D.E., Chou, S.H., Drobney, G. and Reid, B.R. (1983) *J. Mol. Biol.*, 171, 319.
- Hare, D.R., Ribeiro, N.S., Wemmer, D.E. and Reid, B.R. (1985) *Biochemistry*, 24, 4300.
- Hartel, A.J., Wille-Haezelegen, G., van Boom, J.H. and Altona, C. (1981) *Nucleic Acids Res.*, 9, 1405.
- Hartel, A.J., Lankhorst, P.P. and Altona, C. (1982) *Eur. J. Biochem.*, 129, 343.
- Haasnoot, C.A.G., den Hartog, J.H.J., de Rooij, J.F.T., van Boom, J.H. and Altona, C. (1979) *Nature*, 281, 235.
- Haasnoot, C.A.G., de Leeuw, F.A.A.M and Altona, C. (1980) *Tetrahedron*, 36, 2783.
- Hefgout, D.C. and Kallenbach, N.R. (1979) *Nucleic Acids Res.*, 4, 1011.
- Hickey, D.R. and Turner, D.H. (1985) *Biochemistry*, 24, 3987.
- Ho, P.S., Frederick, C.A., Quigley, G.J., van der Marel, G.A., van Boom, J.H., Wang, A.H.-J. and Rich, A. *EMBO J.*, 4, 3617.
- Hogan, M., Dattagupta, N. and Crothers, D.M. (1979) *Biochemistry*, 18, 280.
- Hopkins, R.C. (1984) *Comments Mol. Cell. Biol.*, 2, 153.

- Hore, P.J., Zuiderweg, E.R.P., Nicoley, K., Dukstra, K. and Kaptein, R. (1982) *J. Am. Chem. Soc.*, 104, 4286.
- Hughes, D.W., Neilson, T., Coddington, J.M. and Bell, R.A. (1984) *Can. J. Chem.*, 62, 1214.
- Hyde, E.D. and Reid, B.R. (1985) *Biochemistry*, 24, 4307.
- IUPAC-IUB Nomenclature Commission. (1983) *Eur. J. Biochem.*, 131, 9.
- Jacob, F. and Monod, J. (1961) *J. Mol. Biol.*, 3, 318.
- Jain, S.C., Tsai, C.C. and Sobell, H.M. (1977) *J. Mol. Biol.*, 114, 317.
- Jamin, N., James, T.L. and Zon, G. (1985) *Eur. J. Biochem.*, 152, 157.
- Jardecky, O., and Roberts, G.C.K. (1981) "NMR in Molecular Biology", Academic Press, NY.
- Jaskunas, S.R., Cantor, C.R. and Tinoco, L, Jr. (1968) *Biochemistry*, 7, 3164.
- Johnston, P.D. and Redfield, A.G. (1978) *Nucleic Acids Res.*, 5, 3913.
- Jones, C.R. and Kearns, D.R. (1975) *Biochemistry*, 14, 2660.
- Jones, C.R., Bolton, P.H. and Kearns, D.R. (1978) *Biochemistry*, 17, 601.
- Katagiri, M., Itakura, K. and Narang, S.A. (1974) *J. Am. Chem. Soc. Comm.*, 325.
- Kan, L.-S., Chandrasegaran, S., Pulford, S.M. and Miller, P.S. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 4263.
- Karplus, M. (1969) *J. Chem. Phys.*, 33, 1842.
- Kastrup, R.V., Young, M.A. and Krugh, T.R. (1978) *Biochemistry*, 17, 4855.
- Kean, J.M., White, S.A. and Draper, D.E. (1985) *Biochemistry*, 24, 5026.
- Kearns, D.R. (1976) *Prog. Nucl. Acids Res. Mol. Biol.*, 18, 91.
- Kearns, D.R. (1984) *CRC Crit. Rev. Biochem.*, 15, 237.
- Keepers, J.W., Schmidt, P., James, T.L. and Kollman, P.A. (1984) *Biopolymers*, 23, 2901.
- Kennard, O. (1984) *Pure and Appl. Chem.*, 56, 989.
- Kennard, O. (1985) *J. Biomol. Struct. Dynam.*, 3, 205.

- Kondo, N.S., Ezra, F. and Danyluk, S.S. (1975) *FEBS Lett.*, 35, 213.
- Krugh, T.R. and Reinhardt, C.G. (1975) *J. Mol. Biol.*, 97, 133.
- Krugh, T.R. and Young, M.A. (1975) *Biochem. Biophys. Res. Comm.*, 62, 1025.
- Krugh, T.R., Wittlin, F.N. and Kramer, S.R. (1975) *Biopolymers*, 14, 197.
- Krugh, T.R., Liang, J.W. and Young, M.A. (1976) *Biochemistry*, 15, 1224.
- Krugh, T.R. and Kallenbach, N.R. (1981) *J. Mol. Biol.*, 145, 785.
- Lankhorst, P.P., Gronenveld, G.M., Wille, G., van Boom, J.H., Altona, C. and Haasnoot, G.A.G. (1982) *Recl., J.R. Neth. Chem. Soc.*, 101, 253.
- Lankhorst, P.P., Wille, G., van Boom, J.H., Altona, C. and Haasnoot, C.A.G. (1983) *Nucleic Acids Res.*, 11, 2839.
- Lankhorst, P.P., van der Marel, G.A., Wille, G., van Boom, J.H. and Altona, C. (1985) *Nucleic Acids Res.*, 13, 3317.
- Lee, C.-H. (1983) *Eur. J. Biochem.*, 137, 347.
- Lee, C.-H. and Sarma, R. (1976) *J. Am. Chem. Soc.*, 98, 3541.
- Lee, C.-H. and Tinoco, L., Jr. (1978) *Nature*, 274, 609.
- Lee, C.-H. and Tinoco, L., Jr. (1980) *Biophys. Chem.*, 11, 283.
- Lehrman, L.S. (1961) *J. Mol. Biol.*, 3, 18.
- Lewin, B. (1983) "Genes". John Wiley and Sons, NY.
- Liebman, M., Rubin, J. and Sundaralingham, M. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 4821.
- Lomat, A.J. and Fresco, J.R. (1973) *J. Mol. Biol.*, 77, 345.
- McCall, M., Brown, T., and Kennard, O. (1985) *J. Mol. Biol.*, 183, 385.
- Macura, S. and Ernst, R.R. (1980) *Mol. Phys.*, 41, 95.
- Markeiwicz, W.T. (1979) *J. Chem. Res. (S)* 24.
- Markeiwicz, W.T. and Wiewiórowski, M. (1978) *Nucleic Acids Res. Special Publ. No. 4*, S185.

- Mellema, J.R., Haasnoot, C.A.G., van der Marel, G.A., Wille, G., van der Marel, C.A., van Boom, J.H. and Altona, C. (1983) *Nucleic Acids Res.*, 11, 5717.
- Mellema, J.R., Peters, J.M.L., van der Marel, G.A., van Boom, J.H., Haasnoot, C.A.G. and Altona, C. (1984a) *Eur. J. Biochem.*, 143, 285.
- Mellema, J.R., van der Woerd, R., van der Marel, G.A., van Boom, J.H. and Altona, C. (1984b) *Nucleic Acids Res.*, 12, 5061.
- Morden, K.M., Chu, Y.G., Martin, F.H. and Tinoco, I., Jr. (1983) *Biochemistry*, 21, 428.
- Nagazama, K., Wulrick, K. and Ernst, R.R. (1979) *Biochem. Biophys. Res. Comm.*, 90, 305.
- Neidle, S. (1981) In "Topics in Nucleic Acid Structure", Neidle, S., ed. J. Wiley and Sons, NY, p 177.
- Neidle, S. (1983) *Prog. Biophys. Mol. Biol.*, 41, 43.
- Neidle, S. and Abraham, Z. (1985) *C.R.C. Crit. Rev. Biochem.*, 17, 73.
- Neidle, S., Achari, A., Taylor, G.L., Berman, H.M., Glusker, J.P. and Stallings, W.C. (1979) *Nature*, 269, 304.
- Neilson, O.F., Carin, M. and Westergaard, O. (1984) *Nucleic Acid Res.*, 12, 873.
- Nelson, J. and Tinoco, I., Jr. (1984) *Biochemistry*, 23, 213.
- Nelson, J. and Tinoco, I., Jr. (1985) *Biochemistry*, 24, 6416.
- Ninio, J. (1979) *Biochimie*, 61, 1133.
- Noller, H.F. (1984) *Ann. Rev. Biochem.*, 10, 119.
- Nuss, M.E., Marsh, F.J. and Kollman, P.A. (1978) *J. Am. Chem. Soc.*, 101, 825.
- Nuss, M.E., James, T.L., Apple, M.A. and Kollman, P.A. (1980) *Biochim. Biophys. Acta.*, 136, 147.
- Nussinov and Jacobson. (1980) *Proc. Natl. Acad. Sci. USA*, 77, 6309.
- Ornstein, R.L. and Fresco, J.R. (1983a) *Biopolymers*, 22, 1979.
- Ornstein, R.L. and Fresco, J.R. (1983b) *Biopolymers*, 22, 2001.

- Ornstein, R.L. and Rein, R. (1979) *Biopolymers*, 18, 1277.
- Pack, G. and Loew, G. (1978) *Biochim. Biophys. Acta*, 519, 163.
- Papanicolaou, C., Gouy, M. and Ninio, J. (1984) *Nucleic Acids Res.*, 12, 31.
- Pardi, A., Martin, F.H. and Tinoco, L., Jr. (1981) *Biochemistry*, 20, 3986.
- Pardi, A., Morden, K.M., Patel, D.J. and Tinoco, L., Jr. (1983) *Biochemistry*, 21, 6567.
- Patel, D.J. (1976) *Biopolymers*, 15, 533.
- Patel, D.J. (1977a) *Biopolymers*, 16, 1635.
- Patel, D.J. (1977b) *Biopolymers*, 16, 2739.
- Patel, D.J. (1979) *Biopolymers*, 18, 553.
- Patel, D.J. and Hilbers, C.W. (1975) *Biochemistry*, 14, 2651.
- Patel, D.J. and Camiel, L.J. (1976) *Proc. Natl. Acad. Sci. USA*, 73, 3343.
- Patel, D.J. and Camiel, L.J. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 2624.
- Patel, D.J., Kozlowski, S.A., Marky, L.A., Broka, C., Rice, J.A., Ikatura, K. and Breslauer, K.J. (1982a) *Biochemistry*, 21, 428.
- Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Ikatura, K. and Breslauer, K.J. (1982b) *Biochemistry*, 21, 437.
- Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Ikatura, K. and Breslauer, K.J. (1982c) *Biochemistry*, 21, 445.
- Patel, D.J., Kozlowski, S.A., and Bhatt, R. (1983) *Proc. Natl. Acad. Sci. USA*, 80, 3908.
- Patel, D.J., Kozlowski, S.A., Ikuta, S. and Itakura, K. (1984a) *Biochemistry*, 22, 3218.
- Patel, D.J., Kozlowski, S.A., Ikuta, S. and Itakura, K. (1984b) *Biochemistry*, 23, 3218.
- Patel, D.J., Kozlowski, S.A., Ikuta, S. and Itakura, K. (1984c) *Federation Proc.*, 43, 2663.
- Patel, D.J., Kozlowski, S.A., Hare, D.R., Reid, B.R., Ikuta, S., Landen, N. and Ikuta, S. (1985a) *Biochemistry*, 24, 926.
- Patel, D.J., Kozlowski, S.A., Weiss, M. and Bhatt, R. (1985b) *Biochemistry*, 24, 936.
- Petersheim, M. and Turner, D. (1983a) *Biochemistry*, 22, 264.

- Petersheim, M. and Turner, D. (1983b) *Biochemistry*, 22, 269.
- Quigley, G.J., Wang, A., Urgetto, G., van der Marel, G., van Boom, J. and Rich, A. (1980)  
*Proc. Natl. Acad. Sci. USA*, 77, 6453.
- Ralph, R.K., Marshall, B. and Darkin, S. (1983) *Trends in Biol. Sci.*, 8, 212
- Reid, D.R., Salisbury, S.A., Bellard, S.A., Shakked, S., and Williams, D.H. (1983a)  
*Biochemistry*, 22, 2109.
- Reid, D.R., Salisbury, S.A., Brown, T., Williams, D.H., Vasseur, J.J., Rayner, B. and  
Imbach, J.L. (1983b) *Eur. J. Biochem.*, 135, 307.
- Reid, D.R., Salisbury, S.A., Brown, T. and Williams, D.H. (1985) *Biochemistry*, 24, 4325.
- Reinhardt, C.R. and Krugh, T.R. (1978) *Biochemistry*, 14, 4845.
- Rich, A. (1977) *Accounts Chem. Res.*, 10, 1.
- Robillard, G.T., Tarr, C.E., Vosman, F. and Berendsen, H.J.C. (1976) *Nature*, 262, 363.
- Romaniuk, P.J., Hughes, D.W., Gregoire, R.J., Bell, R.A. and Neilson, T. (1979a)  
*Biochemistry*, 18, 5109.
- Romaniuk, P.J., Hughes, D.W., Gregoire, R.J., Bell, R.A. and Neilson, T. (1979b) *J.C.S.*  
*Chem. Comm.*, 222, 559.
- Saenger, W. (1984) "Principles of Nucleic Acid Structure." Springer-Verlag, NY,  
Berlin, Heidelberg, Tokyo.
- Sagan, D.L. and J.A. Walmsley. (1985) *Biochem. Biophys. Res. Comm.*, 128, 980.
- Sakore, T.D., Reddy, B.S. and Sobell, H.M. (1979) *J. Mol. Biol.*, 135, 763.
- Salsler, W. (1977) *Cold Spring Harbour Symp. Quant. Biol.*, 42, 985.
- Sarma, R.H. (1980) In "Nucleic Acid Geometry and Dynamics", R.H. Sarma, ed.  
Adenine Press, NY. pp 1 - 45.
- Sarma, R.H. and Dhingra, M.M. (1981) In "Topics in Nucleic Acid Structure", S.  
Neidle, ed., Wiley and Sons, NY. pt 1, pp 33 - 64.
- Shakked, Z., Rabinovich, D., Cruse, W.B.T., Egert, E., Kennard, O., Salisbury, S.A., and

- Viswamitra, M.A. (1983) *J. Mol. Biol.*, 166, 183.
- Sheih, H.-S., Berman, H.M., Dabnow, M. and Neidle, S. (1980) *Nucleic Acids Res.*, 8, 85.
- Sobell, H.M. (1980) In "Nucleic Acid Geometry and Dynamics", Sarma, R.H., ed. Academic Press, NY, p 289.
- Sobell, H.M., Sakore, T.D., Jain, S.C., Banerjee, A., Bhandary, K.K., Reddy, B.S. and Lozansky, E.D. (1982) *Cold Spring Harbour Symposium Quant. Biol.*, 43, 293.
- Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Tenzaghi, E., and Inouye, M. (1966) *Cold Spring Harbour Symp. Quant. Biol.*, 31, 77.
- Stryer, L. (1981) "Biochemistry". W.H. Freeman and Coy, San Fransisco.
- Summers, M.F., Byrd, R.A., Gallo, K.A., Samson, C.J., Zon, G. and Egan, W. (1985) *Nucleic Acids Res.*, 13, 6375.
- Tanner, N.K. and Cech, T.R. (1985a) *Nucleic Acids Res.*, 13, 7741.
- Tanner, N.K. and Cech, T.R. (1985b) *Nucleic Acids Res.*, 13, 7759.
- Thompson, J.R. and Hearst, J.E. (1983) *Cell*, 33, 19.
- Tibayenda, N., de Bruin, S.M., Haasnoot, C.A.G., van der Marel, G.A., van Boom, J.H. and Hilbers, C.W. (1984) *Eur. J. Biochem.*, 139, 19.
- Tinoco, I. Jr., Uhlenbeck, O.C. and Levine, M.C. (1971) *Nature*, 230, 362.
- Tinoco, I. Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biol.*, 246, 40.
- Topal, M.D. and Fresco, J.R. *Nature*, 263, 285.
- Tran-Dinh, S., Taboury, J., Neumann, J.M., Huynh-Dinh, Y., Genissel, B., d'Estaintot, B.L. and Igolen, J. (1984) *Biochemistry*, 23, 1362.
- Tsai, C.C., Jain, S.C. and Sobell, H.M. (1977) *J. Mol. Biol.*, 114, 307.
- Uesugi, S., Ohkubo, M., Ohtsuka, E., Ikehara, M., Kabayashi, Y., Kyogukii, Y., Westerink, H.P., van der Marel, G.A., van Boom, J.H. and Haasnoot, C.A.G. (1984) *J. Biol. Chem.*, 259, 1390.



- Uhlenbeck, O.C., F.H. Martin and Doty, P. (1971) *J. Mol. Biol.*, 57, 217.
- van Boom, J.H., Burgens, P.M.J., van Deuran, P.H., Arentzen, R. and Reese, C.B. (1974) *Tetrahedron Lett.*, 3785.
- Viswamitra, W.A., Kennard, O., Jones, P.G., Sheldrick, G.M., Salisbury, S.A. and Falvello, L. (1978) *Nature*, 273, 697.
- Wang, A.H.-J., Fujii, S., van Boom, J.H., and Rich, A. (1982a) *Proc. Natl. Acad. Sci. USA*, 79, 3968.
- Wang, A.H.-J., Fujii, S., van Boom, J.H., van der Marel, G.A., van Boekel, S.A.A. and Rich, A. (1982b) *Nature*, 299, 601.
- Wang, A.H.-J., Quigley, C.J., Kolpak, F.J., Crawford, J.L., van Boom, J.L., van der Marel and Rich, A. (1979) *Nature*, 282, 680.
- Watson, J.D. and Crick, F.H. (1953) *Nature*, 171, 737.
- Wallace, R.B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose and Ikatura, I. (1979) *Nucleic Acids Res.*, 6, 3543.
- Wells, B.D. and Cantor, C.R. (1977) *Nucleic Acids Res.*, 4, 1667.
- 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.
- Westerink, H.P., van der Marel, G.A., van Boom, J.H. and Haasnoot, C.A.G. (1984) *Nucleic Acids Res.*, 12, 4323.
- Williams, M.A. and Fleming, I. (1973) "Spectroscopic Methods in Organic Chemistry", McGraw-Hill, London.
- Wilson, W.D. and Jones, R.L. (1981) *Adv. Pharmacol. Chemother.*, 18, 21.
- Wilson, W.D. and Jones, R.L. (1982) In "Intercalation Chemistry", Whittingham, M.S. and Jacobson, A.J., eds. Academic Press, NY, p445.
- Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E.

(1980) *Nature*, 287, 755.

Woese, C.R., Gutell, R., Gupta, R. and Noller, H.F. (1983) *Microbiol. Rev.*, 47, 621.

Yanofsky, C. (1981) *Nature*, 289, 491.

Young, P.R. and Kallenbach, N.R. (1980) *Proc. Natl. Acad. Sci. USA*, 77, 6453.

Young, P.R. and Kallenbach, N.R. (1981) *J. Mol. Biol.*, 145, 785.

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

Ziff, E.B. (1980) *Nature*, 287, 491.

Zucker, M. and Stiegler, P. (1981) *Nucleic Acids Res.*, 9, 133.