Jones, P. A., & Taylor, S. M. (1981) *Nucleic Acids Res. 9*, 2933.
Khan, M. S. N., Salim, M., & Maden, B. E. H. (1978) *Biochem. J. 169*, 531.
Klagsbrun, M. (1973) *J. Biol. Chem. 248*, 2612.
Lee, T. L., & Karon, M. R. (1976) *Biochem. Pharmacol. 25*, 1737.
Lehrach, H., Diamond, D., Wozney, J. M., & Boedtker, H. (1977) *Biochemistry 16*, 4743.
Lin, H.-L., & Glazer, R. I. (1981) *Mol. Pharmacol. 20*, 644.
Locker, J. (1979) *Anal. Biochem. 98*, 358.
Lu, L. W., Chiang, G. H., Medina, D., & Randerath, K. (1976) *Biochem. Biophys. Res. Commun. 68*, 1094.
Lu, L.-J. W., Tseng, W.-C., & Randerath, K. (1979) *Biochem. Pharmacol. 28*, 489.
Peterkofsky, B., & Tomkins, G. M. (1968) *Proc. Natl. Acad. Sci. U.S.A. 60*, 222.
Razin, A., & Friedman, J. (1981) *Prog. Nucleic Acid Res. Mol. Biol. 25*, 33.
Reichman, M., & Penman, S. (1973) *Biochim. Biophys. Acta 324*, 282.
Salim, M., & Maden, B. E. H. (1981) *Nature (London) 291*, 205.
Santi, D. V., Garrett, C. E., & Barr, P. J. (1983) *Cell (Cambridge, Mass.) 33*, 9.
Siev, M., Weinberg, R., & Penman, S. (1969) *J. Cell Biol. 41*, 510.
Snyder, A. L., Kann, H. E., Jr., & Kohn, K. W. (1971) *J. Mol. Biol. 58*, 555.
Srinivasan, S., & Jaspars, E. M. J. (1982) *Biochim. Biophys. Acta 696*, 260.
Stahl, D. A., Walker, T. A., Meyhack, B., & Pace, N. R. (1979) *Cell (Cambridge, Mass.) 18*, 1133.
Stoyanova, B. B., & Hadjiolov, A. A. (1979) *Eur. J. Biochem. 96*, 349.
Stoyanova, B. B., & Dabeva, M. D. (1980) *Biochim. Biophys. Acta 608*, 358.
Swann, P. F., Peacock, A. C., & Bunting, S. (1975) *Biochem. J. 150*, 335.
Szer, W., & Shugar, D. (1966) *J. Mol. Biol. 17*, 174.
Tscherne, J. S., & Wainfan, E. (1978) *Nucleic Acids Res. 5*, 451.
Van Charldorp, R., Heus, H. A., & Van Knippenberg, P. H. (1981) *Nucleic Acids Res. 9*, 267.
Weiss, J. W., & Pitot, H. C. (1974a) *Arch. Biochem. Biophys. 160*, 119.
Weiss, J. W., & Pitot, H. C. (1974b) *Cancer Res. 34*, 581.
Weiss, J. W., & Pitot, H. C. (1974c) *Arch. Biochem. Biophys. 165*, 588.
Weiss, J. W., & Pitot, H. C. (1975) *Biochemistry 14*, 316.
Wilkinson, D. S., & Pitot, H. C. (1973) *J. Biol. Chem. 248*, 63.
Wolf, S. F., & Schlessinger, D. (1977) *Biochemistry 16*, 2783.
Yannarell, A., Niemann, M., Schumm, D. E., & Webb, T. E. (1977) *Nucleic Acids Res. 4*, 503.
Zar, J. H. (1974) *Biostatistical Analysis*, pp 151–163, Prentice-Hall, Englewood Cliffs, NJ.

Relative Stability of Guanosine-Cytidine Diribonucleotide Cores: A ^{1}H NMR Assessment†

A. Sinclair, D. Alkema, R. A. Bell,* J. M. Coddington, D. W. Hughes, T. Neilson,* and P. J. Romaniuk

ABSTRACT: Proton NMR was used to study the secondary structure and melting behavior of six self-complementary oligoribonucleotide tetramers, each containing two guanosine and two cytidine residues (GGCC, CCGG, GCCG, CGGC, GCGC, and CGCG). GGCC and CCGG formed perfect duplexes containing four G·C base pairs with T_ms of 54 and 47.8 °C, respectively; GCCG and CGGC formed staggered duplexes with two G·C base pairs and four 3′ double-dangling bases, with T_ms of 35.5 and 29.2 °C, respectively; GCGC formed a perfect duplex with a T_m of 49.9 °C, while CGCG formed a staggered duplex with a T_m of 36.9 °C. From these results, an order of stability of the cores containing two G·C base pairs was proposed: GC:GC is more stable than GG:CC which is more stable than CG:CG. The RY model for secondary structure stability prediction was applied to the above tetramers with reasonable success. Suggestions for refinements are discussed.

Duplex formation occurs in several important RNA cellular interactions including the maintenance of the native secondary structures of transfer (Rich & RajBhandary, 1976) and ribosomal (Noller & Woese, 1981) RNA, the control of transcriptional termination (Rosenberg & Court, 1981) and attenuation (Yanofsky, 1981), and tRNA–mRNA recognition in translation. Double-stranded RNA adopts the A helix, which in contrast to the B helix of DNA has 11 base pairs per turn, a shorter pitch (2.3 Å), and a 20° base pair tilt relative to the perpendicular to the helix axis. The ribofuranoside rings in the A helix are 3′-endo in conformation (Dickerson et al., 1982). Short self-complementary oligoribonucleotide sequences have been used as models for double-stranded RNA in the calculation of thermodynamic parameters of formation of the double helix from the single strands, and for the derivation of empirical rules for prediction of secondary structure stability. To date, the most comprehensive set of parameters available is that derived by Borer et al. (1974), applying a nearest-neighbor approximation to the results of optical melting

† From the Departments of Biochemistry and Chemistry, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. *Received September 26, 1983; revised manuscript received December 28, 1983.* Supported by the NSERC and MRC of Canada.

0006-2960/84/0423-2656$01.50/0

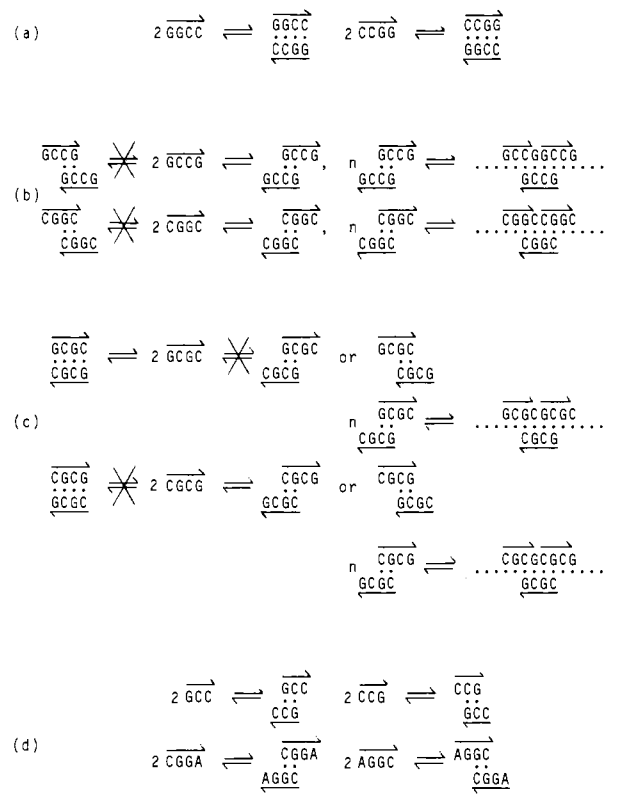


FIGURE 1: Possible secondary structures of the six tetramers containing guanosine and cytidine residues and the four model sequences: (a) GGCC and CCGG; (b) GCCG and CGGC; (c) GCGC and CGCG; and (d) GCC, CCG, CGGA, and AGGC. Equilibria which are not observed are indicated.

studies of 19 perfect oligoribonucleotide duplexes. According to this approximation, the principal contributors to the stability of a duplex are the interactions between adjacent base pairs, and the overall stability can be estimated by summation of the contributions of all two base pair units, or "dinucleotide cores". From their data, Borer et al. (1974) derived thermodynamic parameters for the 10 different two base pair "cores". The three cores containing two guanosine-cytidine base pairs were treated individually; the free energies of formation were determined to be the following: GG:CC, −4.8 kcal/mol; GC:GC, −4.3 kcal/mol; and CG:CG, −3.0 kcal/mol.[1] Although used extensively, these values have not been updated since their first publication.

In the present paper, we further assess the sequence dependence of the relative stability of oligomers containing guanosine and cytidine residues. Adjacent dinucleotide cores containing guanosine and cytidine were lacking in the original work. The six self-complementary tetraribonucleotides, GGCC, CCGG, GCCG, CGGC, GCGC, and CGCG, comprise a complete set containing two guanosine and two cytidine residues, with representation of all dinucleotide cores containing G and C. There are several possibilities for secondary structure formation: GGCC and CCGG may anneal to form perfect helices containing four G·C base pairs (Figure 1a); GCCG and CGGC may anneal to form staggered duplexes containing two G·C base pairs and four dangling bases, which may be either 5′ or 3′ dangling (Figure 1b); GCGC and CGCG may anneal to form either perfect duplexes or staggered duplexes, again with the option of having either 5′ or 3′ dangling bases (Figure 1c). Variable-temperature proton NMR spectroscopy of the nonexchangeable protons (Romaniuk et al., 1979) has been used in this study of the secondary structure and melting behavior of short oligomers. Comparison of the chemical shift–temperature plots and their associated T_ms with those of suitable model sequences allows deduction of secondary structure. A relative order of stability for the three dinucleotide cores GG:CC, GC:GC, and CG:CG is derived.

Materials and Methods

Materials

Nucleosides and 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane were obtained from Terochem Laboratories, Ltd., Edmonton, Alberta, Canada; mesitylenesulfonyl chloride, 1,2,4-triazole, 2,2,2-trichloroethanol, and 100% D_2O were from Aldrich, Milwaukee, WI; silica gel (60–200 mesh) was from J. T. Baker, Philipsburg, NJ; silica gel G (250-μm) thin-layer chromatography plates were from Analtech, Inc., Newark, DE.

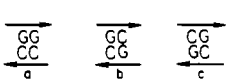
Methods

Chemical Synthesis of Oligoribonucleotides. The tetramers were synthesized by the phosphotriester method developed in this laboratory by T. Neilson and E. S. Werstiuk. The protected cytidine nucleoside N^4-benzoyl-2′-*O*-tetrahydropyranylcytidine was synthesized with 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane (Markiewicz & Wiewiórowski, 1978; Markiewicz, 1979) to block the 5′- and 3′-ribose positions of cytidine, followed by specific benzoylation of the 4-amino group with benzoic anhydride, addition of the tetrahydropyranyl group to the 2′-position, and removal of the disiloxane blocking group with tetrabutylammonium fluoride. After full deblocking, the sequences were characterized at 70 °C by ^1H NMR.

NMR Methodology. The deblocked sequences were lyophilized twice from 99.9% D_2O and dissolved in 400 μL of buffer (1.0 M NaCl, 0.01 M sodium phosphate, pD 7.4, and 100% D_2O). The concentrations were determined from optical density measurements at 260 nm (see Table III). The ^1H NMR spectra were obtained on a Bruker WM-250 spectrometer. Probe temperatures were maintained to within ±1 °C by a Bruker variable-temperature unit and calibrated with a thermocouple. The internal reference was *tert*-butyl alcohol (1.231 ppm), and the resonances were reported relative to DSS (sodium 4,4-dimethyl-4-sila-1-pentanesulfonate). Individual melting temperatures were obtained from computer-fitted polynomial curves which were generated from the plots of chemical shift vs. inverse temperature data and the calculated inflection point used as the T_m value. The T_m values of each of the nonexchangeable protons showing a clearly defined sigmoid curve were then averaged, and this mean value was taken as the overall T_m of the duplex. The values of the T_ms of individual protons lay in a narrow temperature range (normally 2–5 °C, dependent on sequence) and reflected the magnetic environment of each proton within a given duplex. In the present duplexes, ca. 90% of the individual proton temperature curves comprised the mean, and consequently protons through the length of the duplex were represented. This mean T_m value has been shown to approximate the T_m derived from optical studies (Borer et al., 1975).

Initial proton assignments were made by comparison of the observed chemical shifts with those predicted by the RNA

[1] The notation G·C refers to a guanosine-cytidine base pair, GG:CC to form a, GC:GC to form b, and CG:CG to form c.



Both strands are written 5′ to 3′, and base pairing is indicated with a colon, e.g., GG:CC.

Table I: Summary of the Preparation of Oligoribonucleotides[a]

reactants						protected products			free oligomers	
	quantity			quantity			yield			
compd	mg	mmol	compd	mg	mmol	compd	mg	%	yield (%)[c]	R_f[b]
G	1500	1.95	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	7	0.06
C	1000	1.35	C	640	1.50	CC	700	39		
CC	650	0.48	G	260	0.53	CCG	430	49	54	0.28
CCG	430	0.22	G	110	0.23	CCGG	325	58	27	0.13
G	1500	1.95	C	760	1.76	GC	1650	62		
GC	360	0.25	C	111	0.26	GCC	320	66	37	
GCC	280	0.15	G	61	0.15	GCCG	178	47	12	0.21
C	1200	1.60	G	820	1.70	CG	1330	59		
CG	430	0.31	G	160	0.33	GCC	385	61		
CGG	300	0.15	C	70	0.16	CGGC	165	43	27	0.32
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	12	0.07
C	1280	1.75	G	930	1.90	CG	1120	50		
CG	500	0.36	C	175	0.40	CGC	390	54		
CGC	350	0.17	G	95	0.19	CGCG	270	59	19	0.21
CG	500	0.36	G	195	0.40	CGG	430	58		
CGG	400	0.20	A	95	0.21	CGGA	190	36	26	0.22

[a] G, C, and A in all compounds in columns 1, 4, and 7 stand for the fully protected residues N^2-benzoyl-2′-*O*-tetrahydropyranylguanosine, N^4-benzoyl-2′-*O*-tetrahydropyranylcytidine, and N^6-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 G*p*G with two protected guanosine residues linked by a protected phosphate group, the phosphate protection being 2,2,2-trichloroethyl. The coupling reagent is mesitylenesulfonyltriazolide. [b] Chromatography system: 1 M ammonium acetate–ethanol (1:1) on Whatman 40 paper. [c] Calculated from UV spectrophotometric data by assuming a 90% hypochromicity factor.

parameter set devised by Hader et al. (1982). The assignments were then further refined by application of the incremental analysis technique (Borer et al., 1975; Everett et al., 1980) by using the assigned shifts of the precursor dimers, trimers, and, where appropriate, tetramers. The comparative behavior of the chemical shift vs. temperature plots of individual protons served to further confirm assignments.

Results

The six tetramers GGCC, CCGG, GCCG, CGGC, GCGC, and CGCG were synthesized stepwise; experimental data at the various stages of preparation are given in Table I. The sequences were purified after the second stage of deblocking and again at the end by paper chromatography. Deblocking data are given in Table I.

The NMR assignments of the tetramers (Table II) were based essentially on incremental analysis methods (Borer et al., 1975; Everett et al., 1980), and the assignments were confirmed with a computer program designed to predict RNA chemical shifts (Hader et al., 1982).

The T_ms of the six tetramers are shown in Table III. GGCC, CCGG, and GCGC form perfect duplexes with T_ms around 50 °C; GCCG, CGGC, and CGCG form staggered duplexes with T_ms around 33 °C. The family of chemical shift vs. temperature curves for perfect duplexes is different in appearance from those for staggered duplexes, as illustrated by those for GCGC in Figure 2 and those for CGCG in Figure 3. For the perfect duplexes, the curves are obviously sigmoidal, whereas for the staggered duplexes this is not readily apparent and, in fact, the lower temperature points in the curves generally are not easily assigned because of the broadening and loss of the signals associated with end to end aggregation, giving a high molecular weight, loose duplex (Figure 1b,c).

Four other sequences, GCC, CCG, CGGA (see experimental data in Table I), and AGGC (Alkema et al., 1982a), were assembled to elucidate the nature of the staggered du-

Table II: Chemical Shift Assignments for the Tetramers at 70 °C[a]

proton[b]	GGCC	CCGG	GCCG	CGGC	GCGC	CGCG
G(1)H-8	7.910		7.952		7.936	
C(1)H-6		7.746		7.666		7.692
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′		5.814		5.798		5.800
G(2)H-1′	5.851			5.783		5.816
C(2)H-1′		5.840	5.868		5.893	
G(3)H-1′		5.795		5.854	5.851	
C(3)H-1′	5.870		5.868			5.885
G(4)H-1′		5.868	5.868			5.856
C(4)H-1′	5.887			5.895	5.893	

[a] Chemical shifts are in parts per million relative to DSS, with *tert*-butyl alcohol-*d* as an internal reference. [b] Residues are numbered consecutively from the 5′ to the 3′ termini.

Table III: T_m Data for the Six Self-Complementary Tetramers and the Model Sequences

sequence	concn (mM)	av T_m (°C)
GGCC	1.7	54.0
CCGG	1.9	47.8
GCGC	1.9	49.9
CGCG	1.2	36.9
GCCG	2.1	35.5
CGGC	2.1	29.2
GCC	9.0	20.0
CCG	7.8	no T_m
CGGA	3.6	19.2
AGGC	2.2	no T_m

Table IV: Chemical Shift Assignments for the Model Compounds GCC, CCG, CGGA, and AGGC at 70 °C[a]

GCC		CCG		CGGA		AGGC	
proton[b]	shift	proton	shift	proton	shift	proton	shift
G(1)H-8	7.964	C(1)H-6	7.750	C(1)H-6	7.648	A(1)H-8	8.210
C(2)H-6	7.776	C(2)H-6	7.750	G(2)H-8	7.908	A(1)H-2	8.159
C(3)H-6	7.805	G(3)H-8	7.984	G(3)H-8	7.891	G(2)H-8	7.902
C(2)H-5	5.857	C(1)H-5	5.976	A(4)H-8	8.314	G(3)H-8	7.926
C(3)H-5	5.998	C(2)H-5	5.954	A(4)H-2	8.179	C(4)H-6	7.742
G(1)H-1′	5.899	C(1)H-1′	5.884	C(1)H-5	5.929	C(4)H-5	5.875
C(2)H-1′	5.860	C(2)H-1′	5.819	C(1)H-1′	5.792	A(1)H-1′	5.940
C(3)H-1′	5.899	G(3)H-1′	5.884	G(2)H-1′	5.738	G(2)H-1′	5.762
				G(3)H-1′	5.751	G(3)H-1′	5.812
				A(4)H-1′	6.052	C(4)H-1′	5.881

[a] Chemical shifts are in parts per million relative to DSS, with *tert*-butyl alcohol-*d* as an internal reference. [b] Residues are numbered consecutively from the 5′ to the 3′ termini.

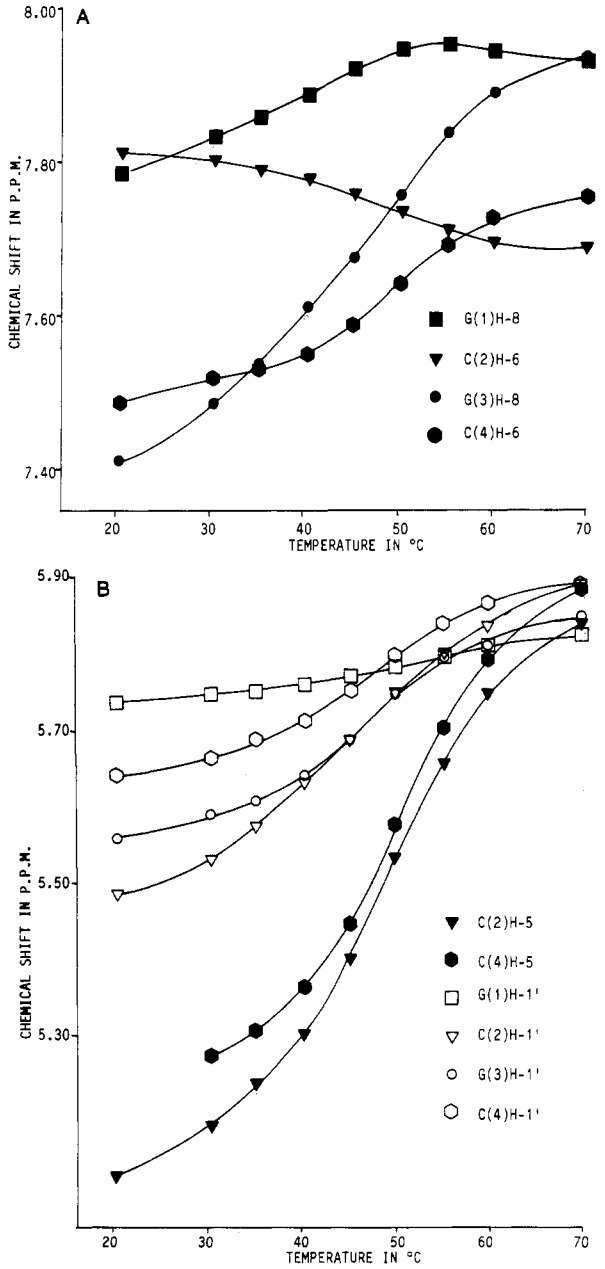


FIGURE 2: Chemical shift vs. temperature plots for GCGC, which forms perfect duplexes.

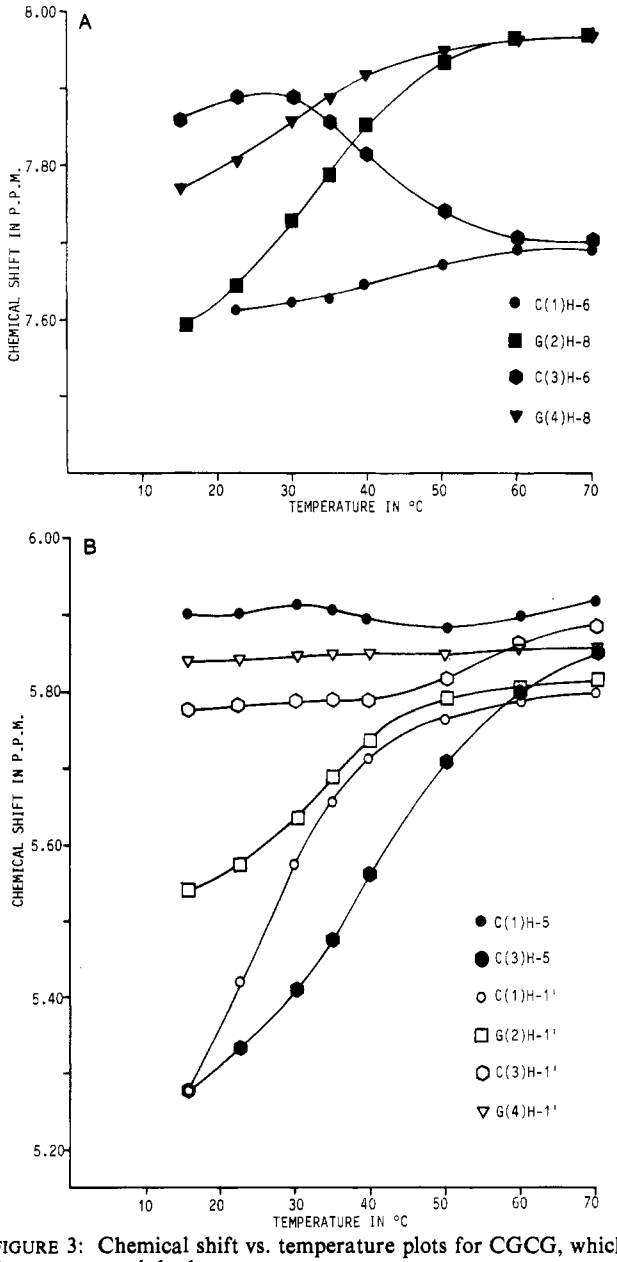


FIGURE 3: Chemical shift vs. temperature plots for CGCG, which forms staggered duplexes.

plexes formed by GCCG and CGGC (Figure 1b,d). Their T_ms, also shown in Table III, indicate a preference for staggered duplexes with 3′ double-dangling bases. The chemical shift assignments of GCC, CCG, CGGA, and AGGC are shown in Table IV.

Discussion

The varied secondary structures adopted by these oligomers provide a data set for GC dinucleotide cores. The melting temperatures (T_ms) of the tetramers monitor the stability of the duplex and, as such, reflect the relative stabilities of the

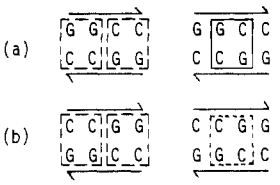


FIGURE 4: Dinucleotide core composition of the perfect duplexes formed by (a) GGCC and (b) CCGG. GG:CC cores, common to both duplexes, are boxed by a dashed line; the GC:GC core of GGCC is boxed by a solid line and the CG:CG core of CCGG by a dotted line.

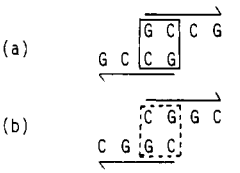


FIGURE 5: Dinucleotide core composition of the 3′ double-dangling staggered duplexes formed by (a) GCCG and (b) CGGC. The GC:GC core of GCCG is boxed by a solid line and the CG:CG core of CGGC by a dotted line.

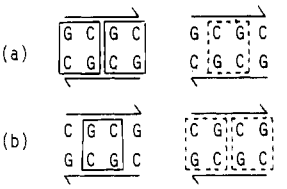


FIGURE 6: Dinucleotide core composition of (a) the perfect duplex formed by GCGC and (b) the perfect duplex which would be formed by CGCG. The GC:GC cores are boxed by a solid line and the CG:CG cores by a dotted line.

dinucleotide cores. Our earlier studies indicate that the GC:GC core is more stable than the CG:CG core. GCA is seen to form a stable duplex with a GC:GC core and two 3′-dangling A's, T_m 34 °C (Alkema et al., 1981), while CGA has a T_m below 0 °C. The data from the tetramers presented here support this conclusion. Unfortunately, the duplex GGA:CCA does not form: GGA prefers self-aggregation which precludes comparisons with the GG:CC core. The present data, however, allow comparison of the stability of the GG:CC core with the other two cores. The T_ms obtained for the six self-complementary sequences are shown in Table III. The tetramers are best considered in pairs, since options for secondary structures differ among pairs. The first pair, GGCC and CCGG, may form only perfect duplexes (Figure 1a). The second pair, GCCG and CGGC, may form only staggered duplexes, consisting of a two base pair core and four double-dangling bases; however, they have the option of forming duplexes with 5′ double-dangling bases or 3′ double-dangling bases (Figure 1b). The 5′ double-dangling duplex of GCCG has a CG:CG core and the 3′ double-dangling duplex a GC:GC core. The 5′ double-dangling duplex of CGGC has a GC:GC core and the 3′ double-dangling duplex a CG:CG core. Overlap of the dangling bases may then lead to formation of an extended loose duplex (Figure 1b). The third pair, GCGC and CGCG, is the one for which the situation is more complicated: either perfect or staggered duplexes are possible (Figure 1c). Either type may give rise to aggregation: the staggered duplexes by direct overlap of dangling bases and the perfect duplexes by end to end stacking.

Perfect Duplexes, GGCC and CCGG. Spectra of both these tetramers show well-resolved resonances over the 70–15 °C range. The chemical shift vs. temperature curves are sigmoidal. This behavior is typical of simple duplex formation without aggregation. The average T_ms of GGCC and CCGG differ by 6 °C. The duplex formed by GGCC contains two GG:CC cores and one GC:GC core (Figure 4a) whereas the duplex formed by CCGG contains two GG:CC cores and one CG:CG core (Figure 4b). When the nearest-neighbor approximation (Borer et al., 1974) is used, according to which the overall duplex stability can be described as a sum of two base pair core stabilities, the primary contributor to the difference in T_ms is the composition of the central core, GC:GC, in GGCC vs. CG:CG in CCGG, with the former being more stable.

Staggered Duplexes, GCCG and CGGC. To determine which, if either, of the two possibilities for staggered duplexes was favored by GCCG and CGGC, comparison was made with four other sequences. The two GCCG staggered duplexes were modeled by GCC and CCG; GCC formed a GC:GC core with two 3′ dangling bases while CCG formed a CG:CG core with two 5′ dangling bases (see Figure 1b,d). The two CGGC staggered duplexes were modeled by CGGA, which forms a CG:CG core with four 3′ double-dangling bases, and AGGC, which formed a GC:GC core with four 5′ double-dangling bases (see Figure 1b,d). GCC has a T_m of 20 °C while CCG had no T_m within the working range of the experiment (0–70 °C); CGGA has a T_m of 19.2 °C, while AGGC also has no T_m. Therefore, for both GCCG and CGGC, it would appear that the formation of a two base pair core with 3′ dangling bases is favored, irrespective of the nature of the core. The stability imparted by the presence of a 3′ dangling base was originally demonstrated by the comparison of GCA (T_m 34 °C) with AGC (no T_m) (Alkema et al., 1981) and AGCUA (T_m 45 °C) with AAGCU (T_m 29 °C) (Alkema et al., 1982b).

The T_ms for individual duplexes formed from GCCG and CGGC follow the same trend as the T_ms obtained for the perfect duplexes, GGCC and CCGG. GCCG, which forms a duplex containing a GC:GC core, has a T_m 6 °C higher than that for CGGC, which forms a duplex with a CG core (Figure 5a,b).

Duplexes of CGCG and GCGC. Of the third pair of sequences, GCGC appears to favor the formation of perfect duplexes and CGCG to favor the formation of staggered duplexes. The temperature response curves for GCGC (Figure 2) are clearly defined sigmoids, and the T_m is high (49.9 °C) and readily discernible. The line widths of the resonances remained reasonably narrow over the temperature range (25–70 °C) of the experiment. In contrast, the curves CGCG (Figure 3) resemble those of the sequences which form staggered duplexes, GCCG and CGGC, and its T_m is close to the T_ms of those duplexes (ca. 33 °C). In addition, the line widths of the signals broadened more rapidly below the T_m than those of the discrete duplexes. This phenomenon arises from the combination of the formation of a high molecular weight, loose duplex resulting from cohesive end aggregation and a slower net exchange rate of the monomer with a duplex of increasing length as the temperature is lowered.

This preference can be rationalized by the following argument. The perfect duplex formed by GCGC can be described as containing two strong GC:GC cores and one weak CG:CG core (Figure 6a) whereas the perfect duplex formed by CGCG would contain one strong GC:GC core and two weak CG:CG cores (Figure 6b). If aggregation can be taken as the driving force for formation of the staggered duplexes, the overall core composition of the aggregates is the same, but a splint junction (the two base pairs flanking the nicks in the aggregate backbone) in GCGC is formed by a CG:CG core (Figure 7a) and in CGCG by a GC:GC core (Figure 7b). Thus, the aggregate

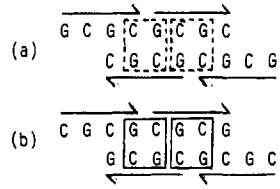


FIGURE 7: Dinucleotide core composition of the splint junctions of the aggregates formed by (a) GCGC and (b) CGCG. The GC:GC cores at the splint junction of the CGCG aggregate are boxed by a solid line and the CG:CG cores at the splint junction of the GCGC aggregate by a dotted line.

formed by CGCG would be more cohesive than any aggregate formed by GCGC. In summary, the duplex formed by GCGC is more stable than that formed by CGCG, but the aggregate formed by GCGC is less stable; the net result is that GCGC forms perfect duplexes and CGCG forms staggered duplexes with aggregation.

Relative Dinucleotide Stabilities. Comparison of the T_m of the perfect duplex formed by GCGC (Figure 6a) with that formed by CCGG (Figure 4a) leads to the conclusion that the *GC:GC core is stronger than the GG:CC core.* These duplexes have a common CG:CG central core, which in CCGG is flanked by two GG:CC cores and in GCGC by two GC:GC cores; the difference in stabilities of the individual cores gives rise to the difference in overall duplex stability.

The difference in T_m due to the replacement of two GC:GC cores in GCGC by two GG:CC cores in CCGG is less than that due to the replacement of a single GC:GC core in GGCC by a single CG:CG core in CCGG. For the former, the difference is 3 °C and for the latter 5 °C, which suggests an order of stability for the three CG cores; *GC:GC is more stable than GG:CC which is more stable than CG:CG*, in contrast to the order proposed from the early optical measurements (Borer et al., 1974) in which the first two cores were interchanged.

Application of the RY Model. Two recent publications propose a model to account for the sequence dependence of duplex stability in terms of both intrastrand and interstrand overlap between adjacent base pairs (Cruz et al., 1982; Bubienko et al., 1983). Instead of the regular 30° winding angle between base pairs in the A′ helix, the authors postulate that the winding angle varies according to the sequence and, using NMR data, have determined the best fit of the winding angle to the data. The winding angles obtained were 50° ± 10° for RR:YY, 20° ± 10° for RY:RY, and 45° ± 10° for YR:YR, where Y is pyrimidine and R is purine. For RR:YY and YR:YR, the 3′ residues of each strand are stacked upon each other (interstrand stacking) while in RY:RY cores the bases in the same strand stack upon each other (intrastrand stacking). The length of the stacks in a secondary structure determines its stability. The authors have not yet determined the relative contributions of the length of the stack, the base composition of the stack, and the amount of cross-stacking in stabilizing the duplex.

This model has been applied to the tetramers under discussion. The perfect duplex formed by GGCC (T_m 54.0 °C) has two three-base stacks with two interstrand stackings (Figure 8a), that formed by CCGG (T_m 47.8 °C) has three interstrand two-base stacks (Figure 8a), and that formed by GCGC (T_m 49.9 °C) has one four-base stack and two two-base stacks with only one interstrand interaction (Figure 8c). According to the RY model, a perfect duplex of CGCG (Figure 8c) should have the same two three-base stacks with two interstrand stackings as the perfect duplex, GGCC (Figure 8a). In fact, CGCG exists as a 3′ double-dangling staggered duplex with a T_m of 36.9 °C, indicative of much lower stability

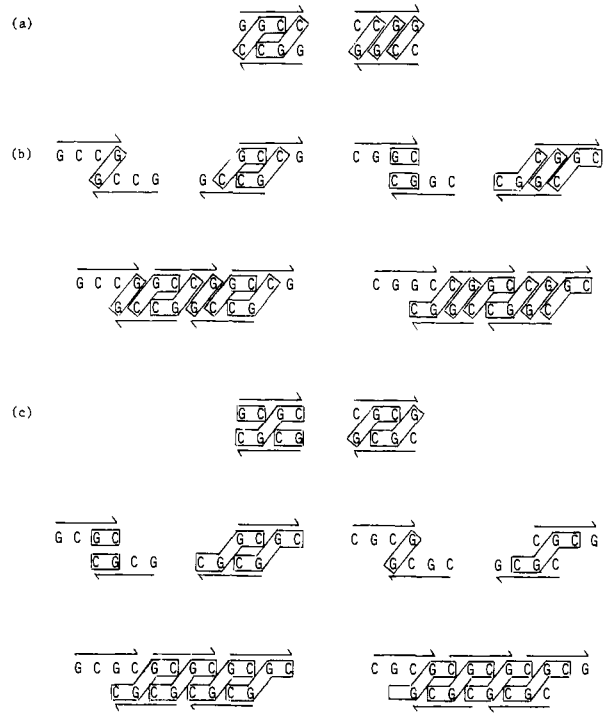


FIGURE 8: Application of the RY model to all possible secondary structures of the six tetramers: (a) perfect duplexes GGCC and CCGG; (b) 5′ and 3′ double-dangling staggered duplexes and aggregates of GCCG and CGGC; (c) perfect duplexes, 5′ and 3′ double-dangling staggered duplexes, and aggregates of GCGC and CGCG. Solid lines enclose those bases which are stacked on each other.

than the GGCC duplex (T_m 54.0 °C). Apparently the RY model requires refinement with regard to the base composition of any stack. Our results suggest that the C-G interstrand stacking is stronger than the G-G interstrand stacking. The model correctly predicts the preference for 3′ double-dangling bases in the formation of staggered duplexes. The first 3′ dangling base extends the stack, either by intrastrand stacking with its 5′ neighbor or by interstrand stacking with the 5′ terminus of the other strand. For GCCG, the 5′ double-dangling duplex would have only one interstrand two-base stack, whereas the 3′ double-dangling duplex, which is the one that forms, has two three-base stacks and two interstrand stacking interactions (Figure 8b). For CGGC, the 5′ double-dangling duplex would have two intrastrand two-base stacks and no interstrand interactions, whereas the 3′ double-dangling duplex has three interstrand two-base stacks (Figure 8b). In this example, the longer stacks do correlate with stability. The model also predicts the greater stability of the CGCG aggregate over the GCGC aggregate, as the four-base stacks common to both aggregates involve bases from two and four different CGCG strands alternately in the CGCG aggregate, as opposed to involving only two GCGC strands in the GCGC aggregate (Figure 8c).

Conclusions

^{1}H NMR studies of six self-complementary oligoribonucleotide tetramers show that GGCC, CCGG, and GCGC form perfect duplexes with four G·C base pairs and GCCG, CGGC, and CGCG form staggered duplexes with two G·C base pairs and four 3′ double-dangling bases. Comparison of the secondary structures and the numerical values of the T_ms of the duplexes leads to the following conclusions about RNA secondary structure stability:

(1) In contrast to previously reported work, the order of stability of the three dinucleotide cores containing two gua-

nosine and two cytidine residues is GC:GC is more stable than GG:CC which is more stable than CG:CG.

(2) The greater stability of a GC:GC core over a CG:CG core can be outweighed by the stabilization imparted by four 3′ double-dangling bases.

(3) The application of the RY model for prediction of RNA secondary structure stability to the tetramers studied was reasonably successful. Base composition of the interstrand stacks is obviously important and is a logical extension of the model.

Registry No. GGCC, 56399-78-1; CCGG, 55048-62-9; GCGC, 89873-22-3; CGCG, 89873-23-4; GCCG, 73942-16-2; CGGC, 89435-89-2; GCC, 3184-24-5; CCG, 3960-32-5; CGGA, 67147-82-4; AGGC, 56399-93-0.

References

Alkema, D., Bell, R. A., Hader, P. A., & Neilson, T. (1981) *J. Am. Chem. Soc. 103*, 2866–2868.

Alkema, D., Hader, P. A., Bell, R. A., & Neilson, T. (1982a) *Biochemistry 21*, 2109–2117.

Alkema, D., Bell, R. A., Hader, P. A., & Neilson, T. (1982b) in *Biomolecular Stereodynamics* (Sarma, R. H., Ed.) Adenine Press, New York.

Borer, P. N., Dengler, B., Tinoco, I., Jr., & Uhlenbeck, O. C. (1974) *J. Mol. Biol. 86*, 843–853.

Borer, P. N., Kan, L. S., & Ts'o, P. O. P. (1975) *Biochemistry 14*, 4847–4869.

Bubienko, E., Cruz, P., Thompson, J. F., & Borer, P. N. (1983) *Prog. Nucleic Acid Res. Mol. Biol. 30*, 41–90.

Cruz, P., Bubienko, E., & Borer, P. N. (1982) *Nature (London) 298*, 198–200.

Dickerson, R. E., Drew, H. R., Conner, B. N., Kopla, M. L., & Pjura, P. E. (1982) *Cold Spring Harbor Symp. Quant. Biol. 48*, 13–24.

Everett, J. R., Hughes, D. W., Bell, R. A., Alkema, D., Neilson, T., & Romaniuk, P. J. (1980) *Biopolymers 19*, 557–573.

Hader, P. A., Alkema, D., Bell, R. A., & Neilson, T. (1982) *J. Chem. Soc., Chem. Commun.*, 10–12.

Markiewicz, W. T. (1979) *J. Chem. Res., Synop.*, 24–25.

Markiewicz, W. T., & Wiewiôrowski, M. (1978) *Nucleic Acids Res., Spec. Publ. No. 4*, S185–S188.

Noller, H. F., & Woese, C. R. (1981) *Science (Washington, D.C.) 212*, 403–411.

Rich, A., & RajBhandary, U. L. (1976) *Annu. Rev. Biochem. 45*, 805–860.

Romaniuk, P. J., Hughes, D. W., Gregoire, R. J., Bell, R. A., & Neilson, T. (1979) *Biochemistry 18*, 5109–5116.

Rosenberg, M., & Court, D. (1979) *Annu. Rev. Genet. 13*, 319–353.

Yanofsky, C. (1981) *Nature (London) 289*, 751–758.

Solid-State ^{13}C NMR Studies of Retinal in Bacteriorhodopsin†

Gerard S. Harbison, Steven O. Smith, Johannes A. Pardoen, Patrick P. J. Mulder, Johan Lugtenburg, Judith Herzfeld,* Richard Mathies, and Robert G. Griffin

ABSTRACT: Solid-state ^{13}C magic-angle sample spinning (MASS) NMR has been used to study lyophilized dark-adapted purple membrane containing ^{13}C-labeled retinals. C-10-, C-11-, and C-12-labeled derivatives each showed two lines, assigned to the coexisting 13-cis and all-trans isomers. The isotropic chemical shifts, particularly of C-11, indicate that the Schiff base is protonated. Shift anisotropies are also similar to those of model compounds, indicating that this part of the chromophore is rigid and immobile and possesses the same degree of in-plane bending as crystalline retinal derivatives. Purple membrane samples labeled on the C-19- and C-20-methyl groups both give single lines from the retinal, upfield shifted by 2.1 and 1.0 ppm, respectively, from model compounds. In all cases, high-quality spectra were obtained from ~50-mg samples in modest signal-averaging times. These results suggest that it is now practical to exploit the enormous potential of MASS NMR for structural studies of ^{13}C-labeled membrane proteins.

Bacteriorhodopsin (bR),[1] the single protein of the purple membrane of *Halobacterium halobium* (Oesterhelt & Stoeckenius, 1971, 1973) is a source of considerable interest to biochemists as an integral membrane protein, an ion pump, a transducer of light, and an analogue of mammalian visual pigments. A wide variety of spectroscopic techniques have been employed to examine the structure and function of bR. Electron and X-ray diffraction (Henderson & Unwin, 1975; Henderson, 1975) have provided a low-resolution structure of the protein, and neutron diffraction combined with specific ^{2}H labeling (Trewhella et al., 1983) has permitted some amino acid residues to be localized within that structure. In addition, flash spectroscopy has been widely used to study photointermediate absorption and kinetics [see, for example, Nagle et al. (1982)], and resonance Raman has been used to study the structure of the retinal prosthetic group in many of the

† From the Department of Physiology and Biophysics, Harvard Medical School, Boston, Massachusetts 02115 (G.S.H. and J.H.), the Department of Chemistry, University of California, Berkeley, California 94720 (S.O.S. and R.M.), the Department of Chemistry, Rijksuniversiteit te Leiden, 2300 RA Leiden, The Netherlands (J.A.P., P.P.J.M., and J.L.), and the Francis Bitter National Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 (R.G.G.). *Received October 21, 1983.* This research was supported by the National Institutes of Health (GM-22316, GM-23289, EY-02051, and RR-00995) and the National Science Foundation (C-670 and CHE-8116042) and by the Netherlands Foundations for Chemical Research (SON) and the Netherlands Organization for the Advancement of Pure Research (ZWO). J.H. and R.M. are recipients of an American Cancer Society Faculty Research Award and an NIH Research Career Development Award, respectively.

[1] Abbreviations: bR, bacteriorhodopsin; CP, cross-polarization; HPLC, high-pressure liquid chromatography; MASS, magic-angle sample spinning; NOE, nuclear Overhauser enhancement; PM, purple membrane; TFA, trifluoroacetic acid.

0006-2960/84/0423-2662$01.50/0