Oligodeoxynucleotide Folding in Solution: Loop Size and Stability of B-Hairpins†

Luigi Emilio Xodo,‡ Giorgio Manzini,‡ Franco Quadrifoglio,* Gijs A. van der Marel,§ and Jacques H. van Boom§

Department of Biochemistry, Biophysics and Macromolecular Chemistry, University of Trieste, I-34127 Trieste, Italy, Institute of Biology, Faculty of Medicine, University of Udine, I-33100 Udine, Italy, and Gorlaeus Laboratory, State University, P.O. Box 9502, 2333 RA Leiden, The Netherlands

Received October 7, 1987; Revised Manuscript Received February 25, 1988

ABSTRACT: The secondary structures of the synthetic DNA fragments d(CGCGCGTTTTTCGCGCG) (**T5**), d(CGCGCGAAAAACGCGCG) (**A5**), d(CGCGCGTACGCGCG) (**TA**), and d(CGCGCGATCGCGCG) (**AT**) were investigated in a combined electrophoretic and spectroscopic study. All the oligomers exist, at low temperature and over a wide range of ionic strength (0.5–100 mM salt) and of nucleotide concentration [0.1–2.0 mM (phosphate)], as a mixture of two slowly interconverting species, identified as the dimeric duplex and the monomeric hairpin structure. The thermodynamic parameters for hairpin denaturation of **T5**, **A5**, **TA**, and **AT** and for duplex denaturation of d(CGCGCG) show that (a) the hairpins are more stable than the reference hexamer duplex at all accessible nucleotide concentrations; (b) the loop contributes favorably to the enthalpy change of hairpin denaturation in the four DNA fragments; (c) the base composition of the loop (A vs T) and the size of the loop (A_5/T_5 vs TA/AT) do not appreciably influence the enthalpic contents of the hairpins; (d) hairpins **TA** and **AT**, with two AT bases intervening in the CG self-complementary part of the molecule, exhibit a markedly higher thermal stability than hairpins **T5** and **A5**, which is entropic in origin. These findings are consistent with the presence of two-residue loops in the tetradecamers **TA** and **AT**.

Inverted repeat sequences in DNA often occur in regions known to have a particular function such as recognition sites for proteins in many biological processes (Rosenberg & Court, 1979; Wells et al., 1980; Muller & Fitch, 1982). In principle, inverted repeats can either exist as double-stranded helices or rearrange in more complex secondary structures such as cruciforms. It has been demonstrated that various plasmids, in response to topological stresses, develop cruciforms (Panayotatos & Wells, 1981; Lilley, 1980; Vologodskii et al., 1979) and that specific endonucleases cleave supercoiled plasmids just in the middle of the palindromic sequence, i.e., in a possible loop (Lilley, 1980, 1981). These findings suggest that the presence of these peculiar secondary structures may have an important role in the biological processes of DNA. As a consequence of the increasing importance attributed to hairpins and cruciforms, many biophysicists have drawn their attention to conformational and thermodynamic studies of DNA segments with a defined sequence potentially capable to originate hairpins. Earlier studies on RNA fragments forming loops (Tinoco et al., 1971; Wickstrom & Tinoco, 1974; Gralla & Crothers, 1973a,b; Uhlenbeck et al., 1973) combined with the recent investigations of the homologous series of molecules d(ATCCTAT$_n$TAGGTA), n = 0–7 (Hilbers et al., 1985; Haasnoot et al., 1986), each of which can exist as monomeric hairpin, contribute to build up the general opinion that the maximal stability displayed by hairpins is obtained when the loop portion of the molecule encompasses four to five nucleotide residues and that a loop of only two residues is unlikely to occur. Thus, on the basis of this hairpin model structure, several partly self-complementary sequences have been investigated by different techniques including optical spectroscopy, electrophoresis, and NMR (Germann et al., 1985; Roy et al., 1986; Wemmer et al., 1985; Nadeau & Gilham, 1985; Summers et al., 1985, Marky et al., 1983).

However, in a recent NMR study of the mismatched octamer d(m^{5}CGm^{5}CGTGm^{5}CG) new aspects of oligonucleotide loop formation were disclosed (Orbons et al., 1987). It has been shown, in fact, that, in suitable experimental conditions, this 8-mer adopts the monomeric hairpin structure with a stem of three Watson–Crick base pairs and a loop of two residues. A subsequent model building study revealed that the two-nucleotide loop should not cause severe steric strain in the oligomer backbone (Orbons et al., 1987). This is in keeping with the results we have so far obtained on the influence of loop size and loop composition on the thermodynamic stability of hairpin structures with an alternating purine–pyrimidine stem, which we report in this paper, which follows a previous one from this laboratory on the conformational dynamics of the 17-mer d(CGCGCGTTTTTCGCGCG) (Xodo et al., 1986). In this work, indeed, spectroscopic and electrophoretic evidences show that the fully self-complementary 14-mers d(CGCGCGTACGCGCG) and d(CGCGCGATCGCGCG) exist in a slowly interconverting equilibrium between the dimeric full duplex and the monomeric hairpin and suggest that the hairpin loop is formed by only two-base residues, namely, the central thymine and adenine.

Materials and Methods

Oligodeoxyribonucleotides. The sequences d(CGCGCGTTTTTCGCGCG) (**T5**), d(CGCGCGAAAAACGCGCG) (**A5**), d(CGCGCGTACGCGCG) (**TA**), d(CGCGCGATCGCGCG) (**AT**), and d(CGCGCG) were synthesized following a modified phosphotriester method as described elsewhere (van der Marel et al., 1981; van Boom et al., 1982). Purifications were performed by gel permeation chromatography using a Sephadex G-50 resin and as eluent a solution of 0.05 M tetraethylammonium hydrogen carbonate. Purity was checked by analytical reverse-phase chromatography and 20% polyacrylamide gel electrophoresis in dena-

† This project has been carried out with the financial contribution of the Italian Ministero della Pubblica Istruzione and of the Netherlands Organization for the Advancement of Pure Research.

* Author to whom correspondence should be addressed at the University of Udine.

‡ University of Trieste.

§ State University.

0006-2960/88/0427-6321$01.50/0

Table I

Oligodeoxynucleotides Studied	
d(CGCGCGTTTTTCGCGCG)	**T5**
d(CGCGCGAAAAACGCGCG)	**A5**
d(CGCGCGTACGCGCG)	**TA**
d(CGCGCGATCGCGCG)	**AT**
Reference Sequences for PAGE	
d(GCGCGCG)	I
d(CGCGCGCGC)	II
d(CCTATATAGG)	III
d(ATACGCGCGTAT)	IV

turing conditions (Frank & Koster, 1979).

UV Spectroscopy. Absorption spectra and absorbance versus temperature profiles, measured at 270 and 249 nm, were obtained with a Cary 219 spectrophotometer, equipped with a thermostated cell holder and a Haake F3 thermostat provided with a Haake PG 10 temperature programmer. Melting profiles were obtained by increasing the temperature at the rate of 0.1 °C/min. The oligomer concentrations were spectrophotometrically determined by using an extinction coefficient of 7000 M^{-1} cm^{-1} for pyrimidine bases and 14 000 M^{-1} cm^{-1} for purine bases.

Buffers. UV measurements were performed in 0.5 mM $NaClO_4$, 0.5 mM Tris, and 0.1 mM EDTA, pH 7.4 (buffer A). Such a low ionic strength and buffer concentration were imposed by the marked thermal stability of the double-helical tetradecamers, in order to keep their T_M's sufficiently low to allow a reliable determination of the temperature dependence of ϵ_C (see Thermodynamic Analysis). However, the low buffer capacity of this solvent was sufficient to avoid significant variations of the pH (pH = 7.4 at 25 °C, apparent pH = 6.6 at 80 °C) as also confirmed by the quite normal B-type CD behavior exhibited by all samples at temperatures below denaturation. Melting profiles were also obtained at higher ionic strength (0.1 M NaCl, 0.1 M Tris·HCl, pH 7.4): in this case the T_M for both 17-mers resulted to be 83 ± 1 °C [data not shown for **A5**, Xodo et al. (1986) for **T5**], whereas the denaturation of the 14-mers was far from being completed at 97 °C, pointing to a T_M certainly higher than 90 °C.

Electrophoretic experiments were performed either in 0.1 M NaCl, 0.1 M Tris·HCl, and 1 mM EDTA, pH 7.4 (buffer B), or in 0.01 M NaCl, 0.01 M Tris·HCl, and 1 mM EDTA, pH 7.4 (buffer C), in which the samples exhibited B-type CD spectra.

Electrophoresis. PAGE was carried out on gels (10 × 15 × 0.15 cm) obtained from a solution containing 20% acrylamide, 3.3% bis(acrylamide), 100 mM (or 10 mM) NaCl, 100 mM (or 10 mM) Tris·HCl, 0.07% ammonium persulfate, and 1 mM EDTA, pH 7.4. The stacking gel contained 5% acrylamide. The electrophoresis was performed at low temperature (5 °C) by means of a Bio-Rad thermostatable apparatus. The gels were run at constant voltage of 100 V, and bromphenol blue was used as marker. Electrophoresis bands were stained with "stains all" dye in formamide/water (1:1).

Results

Melting Experiments. All the oligomers (Table I) were melted in buffer A at concentrations varying in the range 0.1–2 mM (phosphate). Figure 1 shows two typical absorbance versus temperature profiles. All the oligomers present a biphasic behavior. For simplicity we call "transition 1" and "transition 2" those occurring at lower and at higher temperature, respectively. The melting temperature (T_M) of transition 2 is always independent of oligomer concentration, whereas the T_M of transition 1 does show a dependence in all

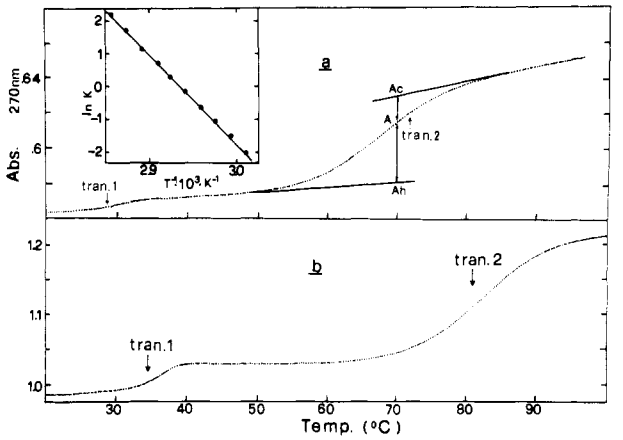

FIGURE 1: Absorbance (at 270 nm) versus temperature profile for (a) d(CGCGCGAAAAACGCGCG) in buffer A and nucleotide concentration of 1.3 × 10^{-4} M (phosphate) (the insert shows the van't Hoff plot); (b) d(CGCGCGTACGCGCG) in buffer A at nucleotide concentration of 2.5 × 10^{-4} M (phosphate).

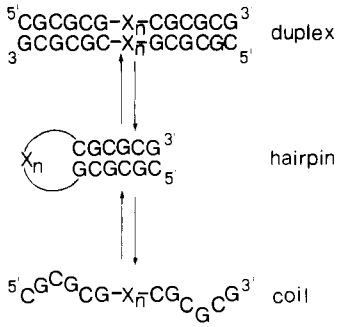

FIGURE 2: Duplex–hairpin and hairpin–coil equilibria occurring in d(CGCGCGTTTTTCGCGCG), d(CGCGCGAAAAACGCGCG), d(CGCGCGTACGCGCG), and d(CGCGCGATCGCGCG).

cases. Both melting temperatures are positively affected by the ionic strength.

When the melted samples are cooled down at a rate of 0.1 °C/min, transition 2 is fully reversible. On the contrary, transition 1 in the same experiment shows hysteresis, more pronounced for compounds **TA** and **AT**, so that the original absorbance level is restored only after several hours. A similar melting behavior was found by Scheffler et al. (1968) studying a series of $(dA\text{-}dT)_n$ oligomers. These findings clearly suggest (a) an intrastrand process for transition 2 and an interstrand process for transition 1 (Martin et al., 1971); (b) a fast equilibrium for transition 2 and a slow interconversion for transition 1; (c) a pretransition structure, for both transformations, with a higher electrostatic energy than the post-transition one (Record & Lohman, 1978a,b).

Considering the base sequence of these DNA fragments, which are fully (14-mers **TA** and **AT**) or partly (17-mers **T5** and **A5**) self-complementary, the spectroscopic results support the assignment of transition 1 to the duplex–hairpin interconversion and of transition 2 to the denaturation of the hairpin form, as illustrated in Figure 2.

Electrophoresis Experiments. The molecular size of the four oligodeoxynucleotides (Table I), at nucleotide concentrations comparable to those used for spectroscopic experiments [0.1–2 mM (phosphate)], was determined by polyacrylamide gel electrophoresis. For this purpose we used as reference samples oligomers of defined length, reported in Table I and numbered from I to IV. Figure 3A shows the electrophoretic profile of the two 17-mers, **T5** and **A5**, at 5 °C in buffers B. As shown in lane c, the 17-mer **A5** migrates in two well-resolved bands, in accord with the presence of two species dif-

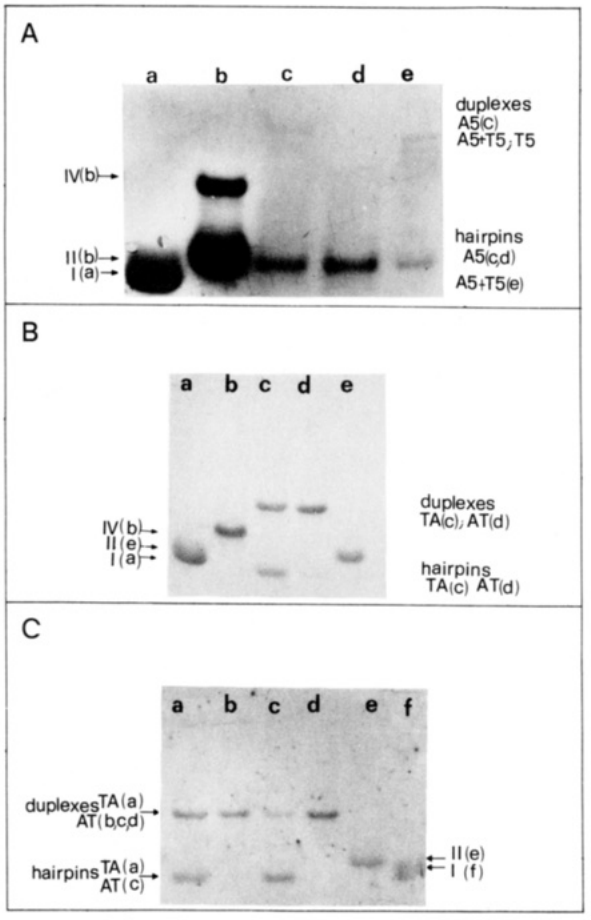

FIGURE 3: (A): Electrophoretic analysis, in buffer B, of d-(GCGCGCG) (lane a), d(CGCGCGCGC) (lane b), d-(CGCGCGAAAAACGCGCG) before being heated (lane c) and after being heated at 95 °C (lane d), and d(CGCGCGAAAAACGCGCG) + d(CGCGCGTTTTTCGCGCG) (lane e). (B) Electrophoretic analysis, in buffer C, of d(GCGCGCG) (lane a), d(CCTATATAGG) (lane b), d(CGCGCTACGCGCG) (lane c), d-(CGCGCGATCGCGCG) (lane d), and d(CGCGCGCGC) (lane e). (C) Electrophoretic analysis, in buffer B, of 1.3 mM (phosphate) d(CGCGCGTACGCGCG) (lane a), 0.1 mM (phosphate) d-(CGCGCGATCGCGCG) (lane b), 0.5 mM (phosphate) d-(CGCGCGATCGCGCG), heated at 95 °C just prior to loading (lane c), the same sample as in (c) but not preheated (lane d), d-(CGCGCGCGC) (lane e), and d(GCGCGCG) (lane f). All the experiments used 20% polyacrylamide gels, run at 5 °C and at a constant voltage of 150 V.

fering in size, whose interconversion, suggested by the absorption versus temperature experiments, is slow on the electrophoresis time scale (a few hours). Lane d, which contained the same sample, but heated at 95 °C just prior to loading, shows only the band corresponding to the fast migrating species of lane c. A nearly equimolar amount of the two complementary 17-mers (**T5** and **A5**), first heated at 95 °C and then kept at room temperature for several hours, was loaded in lane e to see whether the duplex would have been better stabilized with respect to the hairpin. Again, each compound gave a unique fast migrating hairpin band and a close doublet of slowly moving bands. The comparison of these electrophoretic mobilities with those of the reference oligomers, 7-mer I (lane a) and 9-mer II + 12-mer IV (lane b), confirms the interpretation previously provided for the first step in the biphasic thermal profiles observed for these 17-mers: i.e., the presence of a slow interconversion between the full duplex and hairpin structures at near-room temperatures. This is clear for the 17-mer **A5** in lane c, whereas the presence of a slow doublet in lane e for the mixture of the two 17-mers is likely to be ascribed to the simultaneous presence of fully Watson–Crick (WC) duplexes of **T5** + **A5** and of partly WC-paired duplex of **T5**.

Figure 3B shows the electrophoretic result for the two 14-mers **TA** and **AT**, at 5 °C in buffer C. The fragments **TA** and **AT** [at nucleotide concentration of 2 mM (phosphate)] are in lane c and d, respectively, the other lanes being loaded with reference samples. This analysis confirms that these two fully self-complementary oligomers do not behave differently from the 17-mers **T5** and **A5**: they are present at low temperature in two structures corresponding to the hairpin and the full duplex, respectively, since the fast moving bands exhibit nearly the same mobility as the 7-mer reference duplex I (lane a), while the slow moving ones migrate at a markedly lower speed than the reference 10-mer III (lane b). From a semiquantitative point of view it is worth noting that in these conditioins the amount of hairpin **AT** present is markedly lower (not much above the limit of detection) than in the case of **TA**. This lower tendency of **AT** to adopt the hairpin form with respect to **TA**, which could be thermodynamic or kinetic in origin, was checked by another experiment (Figure 3C) in which the electrophoretic mobility of **AT** was measured as a function of nucleotide concentration [lane d, 0.5 mM (phosphate); lane b, 0.1 mM (phosphate)] and of thermal treatment (in lane c heated at 95 °C prior to loading). The results of these two experiments show that dilution does not enhance the hairpin content but, on the other hand, heating is effective in promoting a heavy conversion of the duplex into hairpin. Therefore, it appears that the high propensity of **AT** to assume the full duplex form is mainly due to kinetic factors rather than to thermodynamic reasons, although careful kinetic studies would be required to define this point. On the basis of the electrophoretic and UV absorption–temperature experiments it is possible to state that, at low temperature, in the range of 0–20 °C, all the oligodeoxynucleotides considered, in principle capable of folding back to originate a loop, are present in a mixture of slowly interconverting duplexes and hairpins, although the relative amounts of the two forms may be different for each individual oligomer.

Thermodynamic Analysis. On the basis of the equilibria illustrated in Figure 2 we have performed a thermodynamic analysis of transition 2 in order to estimate the enthalpic and entropic changes involved in this intramolecular transition. Thus assuming a two-state model (Borer et al., 1974), the following equilibrium can be written for transition 2:

$$\underset{\text{coil}}{d(CG)_3\text{–}X_n\text{–}(CG)_3} \rightleftharpoons \underset{\text{hairpin}}{d(CG)_3\text{–}X_n\text{–}(CG)_3}$$

whose equilibrium constant K is equal to

$$K = f/(1-f) = \exp(-\Delta H/RT + \Delta S/R) \qquad (1)$$

where f is the hairpin molar fraction. The thermodynamic parameters were evaluated in three different ways. After correction of the absorbance data for solvent volume expansion due to the increase of the temperature, we used the following procedures:

ln K vs 1/T. Estimates of ΔH and ΔS were obtained by the slope and y intercept of these plots. The method requires the calculation of K at each temperature; thus the procedure with upper and lower base line subtraction was used (Albergo et al., 1981). Figure 1 illustrate a sample case, where the values of K at each temperature are given by the quantity $(A - A_h)/(A_c - A)$, A being the effective absorbance at T and A_h and A_c the extrapolated temperature-dependent values of the hairpin and coil forms. Obviously, the values of ΔH and ΔS depend, to a certain extent, on the way the linear base lines,

Table II: ΔH and ΔS for d(CGCGCGTACGCGCG) Hairpin Denaturation Calculated from Individual Melting Profiles

concn (nM P)	T_M (°C)	ln K vs $1/T$ ΔH (kcal/mol)	ln K vs $1/T$ ΔS (eu)	$4RT_M^2(df/dT)_{T_M}$ ΔH (kcal/mol)	$4RT_M^2(df/dT)_{T_M}$ ΔS (eu)	simulation ΔH (kcal/mol)	simulation ΔS (eu)
0.103	82.5	55	155	54	152	53.6	151.0
0.135	83.5	56	157	57	160	57.0	159.8
0.249	83.0	53	149	55	154	52.4	147.8
0.270	83.0	54	151	52	146	51.5	145.0
mean	83.0	54.5	153	54.5	153	53.6	150.9
SD	±0.4	±1.3	±4	±2.2	±6	±2.4	±6.5

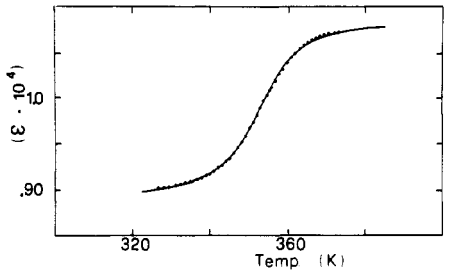

FIGURE 4: Absorbance versus temperature profile for d(CGCGCGTACGCGCG) in buffer A and nucleotide concentration 0.03×10^{-4} M (phosphate). The symbols (●) indicate the experimental UV values, whereas the solid line represents the nonlinear best fit (ΔH = 53.6 kcal/mol, ΔS = 151 eu).

above and below the transition range, are drawn. Therefore, for each sample independent melting profiles were recorded at different nucleotide concentrations, both on heating and cooling. Table II, column 3, reports several ΔH values for **TA**. The error estimate for ΔH is at most ±5% while for T_M it is ±1 °C, analogously to what already has been found in this type of analysis (Freier et al., 1985).

Slope at T_M. The fraction of hairpin at each temperature is given by the quantity $(A - A_h)/(A_c - A_h)$ (Figure 1). Therefore, the ΔH can be calculated from the following van't Hoff relation (Breslauer et al., 1975):

$$\Delta H = 4RT_M^2(df/dT)_{T_M}$$

Table II, column 5, reports the ΔH's for TA obtained from individual melting profiles. Error estimate on ΔH is again in the order of ±5%.

Curve Simulation. On the assumption of a two-state model, the molar absorptivity of the melting profiles as a function of the temperature, $\epsilon(T)$, can be written as

$$\epsilon(T) = \epsilon_h/(1 + K) + \epsilon_c/(1 + K^{-1}) \quad (2)$$

where ϵ_h and ϵ_c are the molar absorptivities of the hairpin and coil forms, respectively, and K is the equilibrium constant as defined by eq 1. We fitted eq 2 to the experimental melting profiles by using a nonlinear least-squares program (Bevington, 1969). The best fits were shown to be very good when ϵ_h and ϵ_c were allowed to vary linearly with the temperature (Freier et al., 1983). The differences between computed and experimental $\epsilon(T)$ values were always within 0.4%. Figure 4 shows a representative fit for **TA**. Table II, column 7, reports the computed values for ΔH.

The methods, although correlated (all are based on the shape analysis), allow us to obtain several estimates of ΔH, which in the case of the 14-mer **TA** give a mean value of 54 kcal/mol and a standard deviation not greater than ±2.4 kcal/mol, i.e., in the order of ±5%. The mean values of the thermodynamic parameters for all the sequences, obtained following the above-mentioned procedures, are collected in Table III. A number of conclusions can be made: (a) The melting experiments clearly show that the thermal stability of the hairpins

Table III: Thermodynamic Parameters for Hairpin Denaturation in 0.5 mM $NaClO_4$ and 0.5 mM Tris, pH 7.4

sample	ΔH^a (kcal/mol)	ΔS^b (eu)	T_M^c (°C)
d(CGCGCGTTTTTCGCGCG)	57	166	69
d(CGCGCGAAAAACGCGCG)	55	162	68
d(CGCGCGTACGCCG)	54	152	83
d(CGCGCGATCGCGCG)	54	151	82
d(CGCGCG)[d]	51		

[a] Error estimate ±2 kcal/mol (see text). [b] Error estimate ±6 eu (see text). [c] Semitransition temperature; error estimate ±1 °C. [d] Average values from $1/T$ vs ln C_T and ln K vs $1/T$ plots (Manzini et al., 1987); in very good accord with data from literature (Ikuta et al., 1986).

is higher than that of the duplex of d(CGCGCG), corresponding to the stem of the hairpins. The T_M of the hexamer is concentration dependent and always lower than the values found for the hairpins [at 0.1 M NaCl $d(CGCGCG)_2$ should reach a T_M comparable to those of **TA** and **AT**, i.e., >90 °C, at a physically unaccessible concentration of about 1 M/strand (Ikuta et al., 1986)]; (b) The enthalpy change associated with the process of loop formation can be considered very similar for the four hairpins and slightly higher than the ΔH of duplex formation of the reference hexamer d(CGCGCG). This finding indicates that the presence of the loop, as already observed (Hilbers et al., 1985; Haasnoot et al., 1986; Ikuta et al., 1986), slightly increases the enthalpy change of hairpin formation in the four sequences. The origin of this enthalpic stabilization and the structural implications suggested by the enthalpic data will be analyzed in the next section; (c) Hairpins **TA** and **AT** have a higher T_M than **T5** and **A5** (83 vs 68 °C). The similar enthalpic content of these hairpins shows that the higher stability of **TA** and and **AT** with respect to **T5** and **A5** is entropic in origin.

Discussion

The results presented in this paper are based on accurate spectroscopic and electrophoretic experiments. The biphasic melting profiles of **TA** and **AT**, combined with their electrophoretic behavior, offer strong experimental evidence that such fully self-complementary tetradecamers can adopt both the hairpin and the duplex structures. This finding confirms that the hairpin structures are common features not only of partly self-complementary (Hilbers et al., 1985; Haasnoot et al., 1986; Germann et al., 1985) or mismatched (Orbons et al., 1986) sequences but also of sequences that can afford a perfect WC dimeric duplex structure, as previously found spectrophotometrically for d(CGCGAATTCGCG) (Marky et al., 1983), d(CGCGATATCGCG) (Wemmer et al., 1985), and oligo(dA-dT) (Scheffler et al., 1968). In the same way, the two heptadecamers **T5** and **A5** exist at room and lower temperatures in slow duplex–hairpin equilibrium since the melting profile of such molecules is biphasic and the premelting transition (at around 30 °C) is concentration dependent. This conclusion, based on spectroscopic evidences and already proposed for the d(CGCGCGTTTTTCGCGCG) (Xodo et al., 1986), is fully supported by the results of the electrophoretic analysis of **A5**. Indeed, we have seen that d(CGCGCGAAAAACGCGCG) exists (Figure 3) as a mixture of two slowly interconverting species, identified as the duplex and hairpin forms, whose amounts strongly depend on the temperature: at 5 °C both species can be evidenced, while above 35 °C the duplex disappears in favor of the hairpin.

The data of Table II suggest that the presence of the loop slightly stabilizes enthalpically the hairpin with respect to the unconstrained stem, namely, $d(CGCGCG)_2$, by a few kcal/mol. A similar behavior has already been reported for different

hairpin systems (Hilbers et al., 1985; Haasnoot et al., 1986; Ikuta et al., 1986). Recently, the three-dimensional structure of the hairpin d(CGCGTTTTTCGCG) as determined from proton NMR data by using distance geometry methods was described (Hare & Reid, 1986). Their resulting refined hairpin structure shows that the bases of the loop region are stacked; in particular, stackings occur between the thymidine residues in the loop and the flanking cytosine and guanine base pair in the top of the stem. Similar conclusions were drawn by Haasnoot et al. (1986) in studying their homologous series. Thus, it is reasonable to expect that the enthalpic stabilization of the hairpin structure is due to the extension of the stacking interactions from the 3'-side of the stem helix into the first residues of the loop according to the model of Haasnoot et al., (1986), or from both sides according to the model of Hare and Reid (1986). However, the geometry of the loop in the 14-mer hairpins may not correspond to either model, as shown by the energy-refined model for two-residue loops of Orbons et al. (1987).

The thermodynamic data for **T5** and **A5** indicate that the base composition of the loop (A in place of T) does not influence significantly the overall hairpin stability, although a destabilization caused by the bulkier adenine in the loop could have been expected a priori. Conversely, the size of the loop plays an important role in the thermodynamic stability of hairpins. The data of Table III offer a clear picture regarding this aspect: the hairpins **TA** and **AT**, with two intervening bases, exhibit a remarkably greater stability (T_M = 83 °C) with respect to hairpins **T5** and **A5**, with five intervening bases (T_M = 68 °C). This finding is in qualitative accord with the results (Hilbers et al., 1985; Haasnoot et al., 1986) concerning the homologous series d(ATCCTAT$_n$TAGGAT) (n = 0–7) for which the stability (i.e., the T_M) of these hairpins has been found to decrease as the loop size increases.

Moreover, the comparison of the enthalpy changes on denaturation of **TA** and **AT** (54 ± 2 kcal/mol) with that of the unconstrained hexamer duplex, d(CGCGCG)$_2$ (51 ± 2 kcal/mol) (Ikuta et al., 1986; Manzini et al., 1987) suggests that the stem part of the hairpins **TA** and **AT** could include six Watson–Crick base pairs, and consequently the loop could be limited to two residues. This suggestion is strongly supported by the remarkably higher thermal stability of the **TA** and **AT** hairpins with respect to that of the **T5** and **A5** ones, whose ΔH and T_M values are, in similar solvent conditions, almost coincident with those of the 16-mer d-(CGCGCGTTTTCGCGCG), for which a 6-bp stem and a 4-residue loop has been established by NMR (Ikuta et al., 1986). In fact, were the dC·dG base pair closing the two-residue loop (of the 14-mers) to be sacrified in favor of a loop size of four residues, it would be very difficult to explain the quite higher thermal stability (ΔT_M about 15 °C) of a hairpin with a loop size similar (to that of **T5** and **A5**) or even equal [to that of d(CGCGCGTTTTCGCGCG)] but with a shorter stem, i.e., with five dC·dG base pairs instead of six. Yet, it is worth noting that Hare and Reid (1986) found in their NMR hairpin refined structure of d(CGCGTTTTCGCG) that the two pyrimidines adjacent to the stem stack to the terminal dC·dG and point inward one in front of the other, leaving room for the possibility of a hydrogen-bonded wobble pairing. A similar arrangement should be better expected in **TA** and **AT**, the proposed dT-dT wobble being replaced by a stronger dC-dG base pairing. Moreover, in case a dC-dG base pairing of the stem would be missing, one should expect, for **TA** and **AT**, a markedly lower enthalpy change on denaturation (about 10 kcal/mol less than what actually found) and an inversion of the thermal stability; i.e., **T5** and **A5** should be more stable than **TA** and **AT**. This corresponds exactly to what we have found for the 14-mers, studying the thermodynamics of loop formation in 4.6 M $NaClO_4$ [see following paper (Xodo et al., 1988)], where a loop of two bases does not seem to be tolerated by a stem in the left-handed conformation. Hence, the occurrence in **TA** and **AT** hairpin structures of a miniloop of two bases is strongly supported by experimental evidences. This conclusion is in keeping with the recent discovery of Orbons et al. (1986, 1987), on the basis of NMR measurements, that the mismatched octadeoxynucleotide d(m^{5}CGm^{5}CGTGm^{5}CG) also forms a very stable hairpin structure with a two-residue loop and three base pair stem.

The finding that hairpins with two base loops not only are feasible but also can be very stable is apparently at variance with the predictions of Hilbers et al. (1985), on the basis of the conformational analysis of their homologous series. On the basis of UV and T-jump experiments they proposed a thermodynamic picture of oligonucleotide loop formation, by which the maximal hairpin stability is attained when the loop is composed of four or five residues, and they also hinted that loops consisting of only two bases are unlikely to exist. However, it has to be noted that the stem base composition of the hairpins of this work and those studied by Hilbers et al. (1985) is different, in particular with respect to the base pair closing the loop: the CG versus the AT pair. The lower stability of AT compared to a CG base pair could help to explain why the first hairpins (n = 0–3) in the cited homologous series are seen to sacrifice one or two base pairs to form a loop of at least four residues. This picture is more akin to the already well-established pattern of RNA loop formation, where maximal stability is achieved with loops including six/seven residues, loops with three residues being generally considered destabilizing (Gralla & Crothers, 1973a,b). Thus it appears evident that the geometry of the double-helical stem [see also following paper (Xodo et al., 1988)] assumes a critical role in determining the best loop size for hairpin stability and it is not trivially connected with the interstrand distance between the phosphates of two paired bases. Although this parameter is similar for A and B helices (about 17.5 Å) (Wang et al., 1979), they appear to require different loop sizes to attain maximal hairpin stability. The Z helix, with an interstrand P–P distance of only about 15 Å (Wang et al., 1979), according to the results presented in the companion paper, seems to require loops of length closer to those found for A-RNA than for B-DNA.

The greater thermodynamic stability of hairpins **TA** and **AT** with respect to **T5** and **A5** is entropic in origin, the enthalpic change on melting being approximately the same for all these molecules. Indeed, it is reasonable to expect that upon loop formation the loss in entropy for **T5** and **A5**, with a five-base loop, is greater (experimentally by about 15 eu) with respect to hairpins **TA** and **AT** with a small loop of two bases. This is in keeping with the expectation that the nucleation entropy in the process of loop formation is more favorable for hairpins with a small loop than for hairpins with a larger one.

Conclusion

We have reported experimental evidence, based on UV and electrophoretic measurements, that the palindromic sequences d(CGCGCGTTTTTCGCGCG), d(CGCGCGAAAAACGCGCG), d(CGCGCGTACGCGCG), and d-(CGCGCGATCGCGCG), in appropriate conditions of nucleotide concentration, ionic strength, and temperature, exist as a mixture of two species in slow exchange with each other. In the range of ionic strengths (up to 10^{-1} M salt) and oli-

gonucleotide concentrations (up to millimolar phosphate) explored, the four oligomers adopt exclusively the monomeric hairpin form above about 40 °C, as demonstrated by the independence of the helix–coil semitransition temperature from the oligonucleotide concentration and by the electrophoretic mobility of these samples which, after being heated at 95 °C prior to loading, are comparable to that of a 7-mer duplex. The analysis of the cooperative helix–coil transition, in terms of a two-state model, suggests the following: (a) The loop slightly stabilizes enthalpically the hairpins with respect to the unconstrained $d(CGCGCG)_2$ stem. (b) The base composition of the loop does not influence the overall stability of the hairpins, while the loop size has a marked effect on the thermodynamic stability, mainly through the entropic term. (c) The enthalpic content [as compared to the ΔH of duplex formation of d(CGCGCG)] and the unexpected higher thermal stability of the 14-mers as compared to the 17-mers strongly suggest the presence of a two-base loop in the 14-mers. Although this is in contrast with the current views on DNA hairpin loop size, it agrees with the recent discovery on DNA loop structure by Orbons et al. (1986, 1987). (d) These findings, added to the results already reported in the literature, indicate that a great deal of caution has to be taken in developing a unique oligonucleotide loop formation model, since both loop size and stem composition play an important role in the loop structure.

Registry No. **A5**, 115185-40-5; **T5**, 105417-61-6; **AT**, 115185-39-2; **TA**, 115185-38-1; d(CGCGCG), 58927-26-7.

References

Albergo, D. D., Marky, L. A., Breslauer, K. J., & Turner, D. H. (1981) *Biochemistry 20*, 1409–1413.

Bevington, P. R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York.

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

Breslauer, K., Sturtevant, J., & Tinoco, I., Jr. (1975) *J. Mol. Biol. 99*, 549–465.

Frank, R., & Koster, H. (1979) *Nucleic Acids Res. 6*, 2069–2087.

Freier, S. M., Albergo, D. D., & Turner, D. H. (1983) *Biopolymers 22*, 1107–1131.

Freier, S. M., Alkema, D., Sinclair, A., Neilson, T., & Turner, D. H. (1985) *Biochemistry 24*, 4533–4539.

Germann, M. W., Schoenwälder, K.-H., & van de Sande, J. H. (1985) *Biochemistry 24*, 5698–5702.

Gralla, J., & Crothers, D. M. (1973a) *J. Mol. Biol. 73*, 497–511.

Gralla, J., & Crothers, D. M. (1973b) *J. Mol. Biol. 78*, 301–319.

Haasnoot, C. A. G., Hilbers, C. W., van der Marel, G. A., van Boom, J. H., Singh, U. C., Pattabiraman, N., & Kollman, P. A. (1986) *J. Biomol. Struct. Dyn. 3*, 843–857.

Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.

Hilbers, C. W., Haasnoot, C. A. G., de Bruin, S. H., Joordens, J. J. M., van der Marel, G. A., & van Boom, J. H. (1985) *Biochimie 67*, 685–695.

Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E., & Kearns, D. R. (1986) *Biochemistry 25*, 4840–4849.

Lilley, D. M. J. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6468–6472.

Lilley, D. M. J. (1981) *Nucleic Acids Res. 9*, 1271–1289.

Manzini, G. Xodo, L., Quadrifoglio, F., van Boom, J. H., & van der Marel, G. A. (1987) *J. Biomol. Struct. Dyn. 4*, 651–662.

Marky, L. A., Blumenfeld, K. S., Kozlowski, S., & Breslauer, K. J. (1983) *Biopolymers 22*, 1247–1257.

Martin, F. H., Uhlenbeck, O. C., & Doty, P. (1971) *J. Mol. Biol. 57*, 201–215.

Muller, U. R., & Fitch, W. M. (1982) *Nature (London) 289*, 582–585.

Nadeau, J. G., & Gilham, P. T. (1985) *Nucleic Acids Res. 13*, 8259–8274.

Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1986) *Nucleic Acids Res. 14*, 4187–4195.

Orbons, L. P. M., van Benzekorn, A. A., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 965–967.

Panayotatos, N., & Wells, R. D. (1981) *Nature (London) 289*, 466–470.

Record, M. T., Jr., & Lohman, T. M. (1978a) *Biopolymers 17*, 159–166.

Record, M. T., Jr., & Lohman, T. M. (1978b) *Q. Rev. Biophys. 11*, 103–178.

Rosenberg, M., & Court, D. (1979) *Annu. Rev. Genet. 13*, 319–351.

Roy, S., Weinstein, S., Borah, B., Nickol, J., Appella, E., Sussman, J. L., Miller, M., Shindo, H., & Cohen, J. S. (1986) *Biochemistry 25*, 7417–7423.

Scheffler, I. E., Elson, E. L., & Baldwin, R. L. (1968) *J. Mol. Biol. 36*, 291–304.

Summers, M. F., Byrd, R. A., Gallo, K. A., Samson, C. J., Zon, G., & Egan, W. (1985) *Nucleic Acids Res. 13*, 6375–6386.

Tinoco, I., Uhlenbeck, O. C., & Levine, M. D. (1971) *Nature (London) 230*, 362–367.

Uhlenbeck, O. C., Borer, P. N., Dengler, B., & Tinoco, I. Jr. (1973) *J. Mol. Biol. 73*, 483–496.

van Boom, J. H., van der Marel, G. A., van Boeckel, C. A. A., Wille, G., & Hoyng, C. F. (1982) *Chemical and Enzymatic Synthesis of Gene Fragments* (Gassen, H. G., & Lang, A. Eds.) pp 53–70, Verlag Chemie, Weinheim.

van der Marel, G. A., van Boeckel, C. A. A., Wille, G., & van Boom, J. H. (1981) *Tetrahedron Lett. 22*, 3887–3890.

Vologodskii, A. V., Lukashin, A. V., Anshelevich, V. V., & Frank-Kamenetskii, M. D. (1979) *Nucleic Acids Res. 6*, 967–982.

Wang. A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A., & Rich, A. (1979) *Nature (London) 282*, 680–686.

Wemmer, D. E., Hare, D. R., & Reid, B. R. (1985) *Nucleic Acids Res. 13*, 3755–3771.

Wells, R. D., Goodman, T. C., Hillen, W. Horn, G. T., Klein, R. D., Larson, J. E., Muller, U. R., Neundorf, S. K., Panayotatos, N., & Stirdivant, S. M. (1980) *Prog. Nucleic Acids Res. Mol. Biol. 25*, 167–267.

Wickstrom, E., & Tinoco, I. Jr. (1974) *Biopolymers 13*, 2367–2383.

Xodo, L. E., Manzini, G., Quadrifoglio F., van der Marel, G. A., & van Boom, J. H. (1986) *Nucleic Acids Res. 14*, 5389–5398.

Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. H. (1988) *Biochemistry* (following paper in this issue).