Thermodynamics of triple helix formation: spectrophotometric studies on the $d(A)_{10} \cdot 2d(T)_{10}$ and $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triple helices

Daniel S. Pilch¹, Roland Brousseau^{3,*} and Richard H. Shafer^{1,2,*}

¹Graduate Group in Biophysics and ²Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143 and ³Cetus Corporation, Emeryville, CA 94608, USA

Received June 18, 1990; Revised and Accepted August 30, 1990

ABSTRACT

We have stabilized the $d(A)_{10} \cdot 2d(T)_{10}$ and $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triple helices with either NaCl or $MgCl_2$ at pH 5.5. UV mixing curves demonstrate a 1:2 stoichiometry of purine to pyrimidine strands under the appropriate conditions of pH and ionic strength. Circular dichroic titrations suggest a possible sequence-independent spectral signature for triplex formation. Thermal denaturation profiles indicate the initial loss of the third strand followed by dissociation of the underlying duplex with increasing temperature. Depending on the base sequence and ionic conditions, the binding affinity of the third strand for the duplex at 25°C is two to five orders of magnitude lower than that of the two strands forming the duplex. Thermodynamic parameters for triplex formation were determined for both sequences in the presence of 50 mM $MgCl_2$ and/or 2.0 M NaCl. Hoogsteen base pairs are 0.22 – 0.64 kcal/mole less stable than Watson-Crick base pairs, depending on ionic conditions and base composition. $C^+ \cdot G$ and $T \cdot A$ Hoogsteen base pairs appear to have similar stability in the presence of Mg^{2+} ions at low pH.

INTRODUCTION

Renewed interest has been generated recently in the triple-helical form of nucleic acids containing homopurine-homopyrimidine (pur-pyr) sequences due to evidence showing its potential biological role as a regulator of eukaryotic gene expression (reviewed in 1). Observations that pur-pyr tracts in natural and recombinant plasmid DNA can adopt intramolecular triple-stranded structures *in vitro* have kindled much of this interest (2–6). Frank-Kamenetskii and coworkers have proposed a model for these structures, termed H-DNA, which postulates formation of pur-pur-pyr triple helical regions and the concomitant looping out of a homopurine strand (7,8). Their model provides a structural basis for the S1 nuclease hypersensitivity ascribed to

these sequences by identifying the single-stranded homopurine loop as the substrate. Pur-pyr sequences are widespread in eukaryotic genomes and often occur in positions that flank transcribed genes (1). In many cases, these sequences adopt the H form of DNA *in vitro* (9–11) and have also been shown to be a prerequisite for optimal expression of the *hsp26* gene *in vitro* (12).

In addition, Moser and Dervan (13) and Hélène and coworkers (14,15) have synthesized homopyrimidine oligodeoxynucleotides covalently linked to DNA cleaving agents, which cut duplex DNA in a sequence specific fashion via triplex formation. These studies establish the usefulness of such sequences as probes for chromosome mapping (16) and as antisense DNA for chemotherapeutic applications, in which gene expression is regulated by triplex formation. This application has been addressed in studies on the inhibition of *c-myc* oncogene expression *in vitro* by pur-pur-pyr triplex formation (17).

The existence of triplexes in RNA and DNA homopolymers has been known for over thirty years (18,19). Polynucleotide fiber diffraction studies (20–22) resulted in a model of the triple helix in which the third strand resides in the major groove of the underlying duplex, and interacts with the purine strand of the duplex via Hoogsteen base pairing. The base triplets resulting from interaction of a third homopyrimidine strand with a complementary duplex are shown in Figure 1. Furthermore, the duplex portion of the triplex assumes an A-like conformation with a C3'-endo sugar pucker. However, recent NMR work by Rajagopal and Feigon (23) on the $d(GA)_4 \cdot 2d(TC)_4$ triplex indicates that only the two pyrimidine strands have C3'-endo sugar pucker. The third strand can be either homopurine or homopyrimidine. In the latter case, the third strand is parallel to the purine strand of the duplex (13,14). Other NMR studies probing the solution structures of triple-helical oligodeoxynucleotides have been recently reported (24–26).

Very little information is currently known about the thermodynamics of triplex formation. The results of a few studies in the middle 1960s, in which the enthalpy changes associated

* To whom correspondence should be addressed

* Present address: Biotechnology Research Institute, Montreal, Quebec H4P 2R2, Canada

with the formation of the RNA triple helix, poly(rA)·2poly(rU), were determined (27–29), comprise the majority of this information. Recently, however, Quadrifoglio and coworkers have reported the enthalpy and entropy changes associated with dissociation of DNA triplexes composed of two single strands and either a loop (30) or a hairpin (31). In order to fully realize the chemotherapeutic potential and gain a better understanding of the biological significance of triple helices, it will be important to determine and, if possible, predict the effects that base sequence, temperature, solvent, and ions have on the stability and structural energetics of these unusual nucleic acid forms. The need for a detailed understanding of the thermodynamics associated with triplex formation is therefore apparent. In this work, we present circular dichroic and thermodynamic analyses of the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ and $d(A)_{10} \cdot 2d(T)_{10}$ triple helices, stabilized by Na^+ and/or Mg^{2+} ions at pH 5.5. Here, C^+ refers to a cytosine residue that is protonated at its N3 position. UV mixing curves establish the 1:2 purine to pyrimidine stoichiometry, and circular dichroic titrations suggest a possible sequence-independent spectral signature for triplex formation. Thermal denaturation profiles reveal the initial loss of the third strand followed by dissociation of the underlying duplex. Thermodynamic analysis indicates that Hoogsteen (HG) base pairs are 0.22–0.64 kcal/mole less stable than Watson-Crick (W-C) base pairs, depending on ionic conditions and base composition. Furthermore, $C^+ \cdot G$ and $T \cdot A$ Hoogsteen base pairs appear to have similar stability in the presence of Mg^{2+} ions and low pH.

MATERIALS AND METHODS

Chemicals and Oligodeoxynucleotides. Synthesis and purification of the $d(A)_{10}$ and $d(T)_{10}$ oligomers were done as previously described (26). Synthesis of the $d(G_3A_4G_3)$ and $d(C_3T_4C_3)$ oligomers was also carried out as previously described (26). However, these oligomers did not require HPLC purification. They were dialyzed against 2 mM sodium phosphate (pH 7.0) and their purity was checked electrophoretically and by 1H -NMR. Concentrations of all oligomers were determined spectrophotometrically. The reported extinction coefficients for poly(dA) [$\epsilon_{257} = 8600 \text{ cm}^{-1} (\text{mole base/liter})^{-1}$] (32) and poly(dT) [$\epsilon_{265} = 8700 \text{ cm}^{-1} (\text{mole base/liter})^{-1}$] (19) were used for $d(A)_{10}$ and $d(T)_{10}$, respectively. The extinction coefficients used for $d(G_3A_4G_3)$ and $d(C_3T_4C_3)$ were $\epsilon_{255} = 11,500 \text{ cm}^{-1} (\text{mole base/liter})^{-1}$ and $\epsilon_{271} = 8,300 \text{ cm}^{-1} (\text{mole base/liter})^{-1}$, respectively, and were determined as described by Griswold *et al.* (33). All buffer reagents were obtained from Sigma Chemical Company. All oligomer concentrations, except where noted otherwise, are reported on a strand basis.

UV Mixing Curves. Equimolar (10 μM or 12 μM) solutions of $d(G_3A_4G_3)$ and $d(C_3T_4C_3)$ were made in 10 mM sodium cacodylate, 0.1 mM EDTA and either 100 mM NaCl, pH 7.1 or 2.0 M NaCl, pH 5.5. Mixing experiments were carried out as previously described for $d(A)_{10}$ and $d(T)_{10}$ (26) with the exceptions that the mixing experiment with pH 7.1 was run at 12°C, while that with pH 5.5 was run at 4°C.

Circular Dichroic (CD) Titrations. For the $d(G_3A_4G_3) + d(C_3T_4C_3)$ system, CD measurements were done on a solution of $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ (15 μM duplex), into which aliquots from a 1.5 mM solution of $d(C_3T_4C_3)$ were added to a final concentration of 15 μM . For the $d(A)_{10} + d(T)_{10}$ system, separate solutions containing either $d(A)_{10} \cdot d(T)_{10}$ (10 μM

duplex) or $d(A)_{10} \cdot 2d(T)_{10}$ (10 μM triplex), but equimolar with respect $d(A)_{10}$, were made. These duplex and triplex solutions were mixed together at differing ratios and their ellipticities were measured. The buffer conditions for all the CD experiments were 10 mM sodium cacodylate, 50 mM $MgCl_2$, 0.1 mM EDTA, pH 5.5. All solutions were preheated at 70°C for 5 minutes and slowly cooled prior to CD analysis. Ellipticities were measured from 200 to 350 nm on a Jasco J-500A spectropolarimeter equipped with a Fisher 9000 circulating water bath. All CD spectra were averaged over 16 acquisitions, and the scan rate was 100 nm/min. The temperature was held at 6.0°C and the cell path length was 1.0 cm.

Helix-Coil Transitions. For both the $d(A)_{10} + d(T)_{10}$ and $d(G_3A_4G_3) + d(C_3T_4C_3)$ systems, the UV absorbance of solutions containing 10 mM sodium cacodylate, 50 mM $MgCl_2$ or 2.0 M NaCl, 0.1 mM EDTA, pH 5.5, and differing concentrations (5–50 μM) of either triplex or duplex were measured once per minute at 260, 284, and 300 nm on a Gilford 2600 Spectrophotometer interfaced to an Optima Systems AT-10 computer for data collection and analysis. The duplex and triplex solutions were always equimolar in the concentration of purine strand. In addition, all DNA solutions were preheated at 70°C for 5 minutes and slowly cooled prior to UV analysis. The temperature was increased at a rate of 0.25°C/min with a Gilford 2527 thermoprogrammer, and the total run time was 320 min. The cell path length was 1 cm for solutions containing 5–35 μM duplex or triplex, and 0.1 cm for solutions containing higher concentrations. The cuvette-holding chamber was flushed with N_2 gas for the duration of the run. First derivatives were calculated over a window of $\pm 2.5^\circ\text{C}$.

Thermodynamic Analysis. We have presented evidence (26) that the biphasic helix-coil transition of the $d(A)_{10} \cdot 2d(T)_{10}$ triplex at 260 nm reflects the following two helix-to-coil equilibria:

We present evidence below showing that the biphasic helix-coil transition of the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex at 260 nm reflects the same two conformational equilibria. Thus, both systems can be analyzed in a similar manner. In order to extract thermodynamic parameters from the biphasic melting curves, we analyze each phase (HG and W-C) separately, using a two-state model. Several lines of evidence are presented to support the validity of this approach.

The equilibrium constant for the first reaction (HG) may be written as

$$K_{HG} = \frac{\theta'}{C_A(1-\theta')^2} \quad (1)$$

where C_A is the total concentration of $d(A)_{10}$ strands and θ' is the fraction of $d(A)_{10}$ strands in the triplex form, while that for the second reaction (W-C) may be written as

$$K_{W-C} = \frac{\theta}{2C_A(1-\theta)(1-\theta/2)} \quad (2)$$

where θ is the fraction of $d(A)_{10}$ strands in the duplex form. Using a derivation similar to that presented by Gralla and Crothers (34), we determined the values for θ' and θ at which the $d(A)/d(1/T)$ vs. T curves reach their maxima ($T = T_{\max}$) to be 0.414 and 0.486, respectively. In general, K can be expressed as

$$K = \exp(-\Delta G^\circ/RT) = \exp(-\Delta H^\circ/RT + \Delta S^\circ/R) \quad (3)$$

Combining equation (3) with equations (1) and (2) at $T = T_{\max}$ yields the following two equations for the Hoogsteen and W-C reactions, respectively:

$$\text{HG: } 1/T_{\max} = (2.3R/\Delta H^\circ)\log C_A + (\Delta S^\circ - 0.188R)/\Delta H^\circ \quad (4)$$

$$\text{W-C: } 1/T_{\max} = (2.3R/\Delta H^\circ)\log C_A + (\Delta S^\circ + 0.468R)/\Delta H^\circ \quad (5)$$

If the standard enthalpy (ΔH°) and standard entropy (ΔS°) changes are assumed to be independent of temperature, they can be determined from plots of $1/T_{\max}$ vs. $\log C_A$ (35,36). All the $1/T_{\max}$ vs. $\log C$ data were analyzed by linear regression. The errors for the thermodynamic parameters reported in tables I and II reflect the magnitudes of the 95% confidence limits calculated by the linear regression analysis.

The standard free energy changes as a function of $T_{\max}$ for the Hoogsteen and W-C reactions, respectively, may be given by

$$\text{HG: } \Delta G^\circ(T_{\max}) = 2.3RT_{\max}\log(0.828C_A) \quad (6)$$

$$\text{W-C: } \Delta G^\circ(T_{\max}) = 2.3RT_{\max}\log(1.60C_A) \quad (7)$$

Hence, ΔG° values at 25°C were determined by either extrapolating the least-squares fit for the $1/T_{\max}$ vs. $\log C_A$ data to the C_A value for which $T_{\max}$ is 25°C and then solving equations (6) or (7) or by using the ΔH° and ΔS° values derived from the plots of $1/T_{\max}$ vs. $\log C_A$ and combining them to give $\Delta G^\circ_{25} = \Delta H^\circ - (298)\Delta S^\circ$. The ΔG°_{25} values calculated by either method were virtually identical.

In solutions containing 1:1 stoichiometries of purine to pyrimidine strands, the monophasic helix-coil transitions of both the $d(A)_{10} \cdot d(T)_{10}$ and $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplexes reflect a single helix-coil equilibrium, for which the equilibrium constant is given by

$$K = \frac{2\theta''}{C_T(1-\theta'')^2} \quad (8)$$

where C_T is total concentration of all single strands and θ'' is the fraction of total strands in the duplex form. Given that $\theta'' = 0.414$ at $T = T_{\max}$ (34), $1/T_{\max}$ may be written as

$$1/T_{\max} = (2.3R/\Delta H^\circ)\log C_T + (\Delta S^\circ - 0.880R)/\Delta H^\circ \quad (9)$$

Thus, ΔH° and ΔS° values can be determined from $1/T_{\max}$ vs. $\log C_T$ plots. Knowing the values for ΔH° and ΔS° , we then determined ΔG° at 25 °C by the same two methods described above.

RESULTS

UV Mixing Curves. We have previously shown (26) that sufficiently high concentrations of Na^+ or Mg^{2+} ions will stabilize the $d(A)_{10} \cdot 2d(T)_{10}$ triple helix at neutral pH, as evidenced in part by a resultant $1d(A)_{10}:2d(T)_{10}$ stoichiometry in UV mixing curves. The stoichiometry associated with the interaction of $d(G_3A_4G_3)$ and $d(C_3T_4C_3)$ under varying conditions of ionic strength and pH has been determined by similar methods. Figure 2 shows the changes in UV absorbance associated with continuous addition of $d(C_3T_4C_3)$ to a fixed amount of $d(G_3A_4G_3)$ in either 100 mM NaCl, pH 7.1 (Figure 2A) or 2.0 M NaCl, pH 5.5 (Figure 2B). The point of minimum absorbance at 100 mM NaCl, pH 7.1 corresponds to an approximate 1:1 molar ratio [48% $d(C_3T_4C_3)$] of $d(G_3A_4G_3)$ to $d(C_3T_4C_3)$ (Figure 2A). Hence, only duplex

Figure 1. $\text{C}^+ \cdot \text{G} \cdot \text{C}$ (top) and $\text{T} \cdot \text{A} \cdot \text{T}$ (bottom) base triplets indicating both Watson-Crick and Hoogsteen hydrogen bonds.

$d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ formation occurs under these conditions. In contrast, the point of minimum absorbance at 2.0 M NaCl, pH 5.5 corresponds to an approximate 1:2 molar ratio [65% $d(C_3T_4C_3)$] of $d(G_3A_4G_3)$ to $d(C_3T_4C_3)$ (Figure 2B). These conditions are therefore sufficient to promote formation and stabilization of the $d(\text{C}^+T_4\text{C}^+) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex structure. Neither a pH of 5.5 nor a high salt concentration individually gave rise to a mixing curve with a $1d(G_3A_4G_3):2d(C_3T_4C_3)$ stoichiometry at the DNA concentrations that were used (data not shown). Unlike the requirements for stabilization of the $d(A)_{10} \cdot 2d(T)_{10}$ triplex (26), low pH is a prerequisite for formation of the $d(\text{C}^+T_4\text{C}^+) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex. This dependence on low pH supports the $\text{C}^+ \cdot \text{G} \cdot \text{C}$ base triplet model, which postulates that one of the two Hoogsteen hydrogen bonds involves the N3 proton of cytosine and the N7 atom of guanine (see Figure 1). When coupled with pH 5.5, suitable concentrations of MgCl_2 (e.g. 50 mM) also give rise to UV mixing curves which exhibit a $1d(G_3A_4G_3):2d(C_3T_4C_3)$ stoichiometry (data not shown). **Circular Dichroic Titrations.** CD spectra of solutions (pH 5.5) containing 50 mM MgCl_2 and differing ratios of triplex to duplex are shown for the $d(G_3A_4G_3) + d(C_3T_4C_3)$ system in Figure 3A and for the $d(A)_{10} + d(T)_{10}$ system in Figure 3B. These spectra are weighted such that they are all equimolar with respect to purine strand. The spectra of each of the two sequences exhibit changes in the 240–300 nm region with increasing triplex:duplex ratios. However, the nature of these changes appears to be unique to the specific sequence. Unlike the spectral changes in the 240–300 nm region, those in the 210–240 nm region are more independent of sequence. With increasing triplex:duplex ratios, the spectra of both sequences show a marked

Figure 2. UV mixing curves for the reactions between $d(G_3A_4G_3)$ and $d(C_3T_4C_3)$ in either 100 mM NaCl, pH 7.1 (A) or 2.0 M NaCl, pH 5.5 (B). The buffer was 10 mM sodium cacodylate, 0.1 mM EDTA for both experiments.

amplitude decrease, broadening, and red-shift of the positive band at 218 nm and an amplitude decrease and red-shift of the negative band at 210 nm. The similarity in spectral variations that both sequences display in the 210–240 nm region upon alteration of the triplex:duplex ratio suggests that this spectral region presents a useful circular dichroic signature of triplex formation irrespective of base sequence.

The mathematical sums of the appropriately weighted CD spectra for duplex $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ and triplex $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ are virtually identical to the experimentally-derived spectra for triplex:duplex ratios of 2:1 and 1:2, which are presented in Figure 3A (data not shown). Furthermore, the slight differences that do exist between the mathematically and experimentally-derived spectra do not occur at wavelengths where single-stranded $d(C_3T_4C_3)$ has a large positive or negative ellipticity under identical buffer conditions. In addition, the four spectra in Figure 3A show isoelliptic points at 236 and 269 nm, through which an appropriately normalized spectrum of single-stranded $d(C_3T_4C_3)$ does not pass. These results indicate an absence of single-stranded $d(C_3T_4C_3)$ in the duplex or triplex solutions under these conditions. Similar results were obtained for the $d(G_3A_4G_3) + d(C_3T_4C_3)$ system in 2.0 M NaCl, pH 5.5.

Helix-Coil Transitions. Melting of the third $d(C^+_3T_4C^+_3)$ strand from the underlying $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex coincides with a hypochromic UV absorbance change at 300 nm and no change at 284 nm, while denaturation of the duplex coincides

Figure 3. CD spectra of solutions containing different ratios of duplex to triplex for both the $d(G_3A_4G_3) + d(C_3T_4C_3)$ (A) and $d(A)_{10} + d(T)_{10}$ (B) systems. The ratios of duplex to triplex are 1:0 (short dashed line), 2:1, (medium dashed line), 1:2 (long dashed line), and 0:1 (solid line). The spectra are normalized with respect to purine strand concentration. Both experiments were done in 10 mM sodium cacodylate, 50 mM $MgCl_2$, 0.1 mM EDTA, pH 5.5 at a temperature of 6°C.

with a hyperchromic UV absorbance change at 284 nm and no change at 300 nm (data not shown). Dissociation of triplex $d(A)_{10} \cdot 2d(T)_{10}$ into duplex $d(A)_{10} \cdot d(T)_{10}$ and single stranded $d(T)_{10}$ is accompanied by a hyperchromic absorbance change at 284 nm, while denaturation of the duplex shows no change at this wavelength (26). However, denaturation of both Hoogsteen and W-C base paired strands in either sequence is accompanied by a hyperchromic UV absorbance change at 260 nm. Hence, for either sequence, dissociation of the third strand from the underlying duplex or of the duplex itself can be represented by a monophasic melting profile with the appropriate choice of wavelength.

First-derivative plots of the thermal denaturation profiles at differing DNA concentrations and 260 nm are shown in Figures 4A and B for the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex in 2.0 M NaCl and 50 mM $MgCl_2$, respectively, and in Figure 4C for the $d(A)_{10} \cdot 2d(T)_{10}$ triplex in 50 mM $MgCl_2$. All experiments were carried out at pH 5.5. In either 2.0 M NaCl (Figure 4A) or 50 mM $MgCl_2$ (Figure 4B), the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ transitions are biphasic in nature. Thermal denaturation studies on both the $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex and the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex at 284 and 300 nm (data not shown) confirm that the first phase of the transition (with the lower T_{max}) corresponds to the melting of the third strand from the underlying duplex (Hoogsteen phase), whereas the second phase of the transition corresponds to denaturation of the remaining duplex (W-C phase). Control studies under similar conditions indicate that single-stranded

Figure 4. First derivative plots of helix-coil transitions at 260 nm for differing concentrations of the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex in 2.0 M NaCl (A) or 50 mM $MgCl_2$ (B) and of the $d(A)_{10} \cdot 2d(T)_{10}$ triplex (C) in 50 mM $MgCl_2$. In each case, the triplex concentrations from bottom to top are 10, 15, 20, 25, and 30 μM . The buffer for all the melting experiments was 10 mM sodium cacodylate, 0.1 mM EDTA, pH 5.5.

$d(C_3T_4C_3)$ self-associates at temperatures below 30°C and exhibits a hypochromic absorbance change at 284 nm upon dissociation (data not shown). However, the thermal denaturation profiles of both the $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex and the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex show no appreciable absorbance change at 284 nm over this temperature range (data not shown). An absence of the self-associated form of $d(C_3T_4C_3)$ in the duplex or triplex solutions is therefore indicated. This result confirms that determined from the circular dichroic studies described in the preceding section. Single-stranded $d(G_3A_4G_3)$ does not self-associate under the conditions and over the temperature range used in these studies (data not shown).

The helix-coil transitions of the $d(A)_{10} \cdot 2d(T)_{10}$ triple helix are also biphasic in nature (Figure 4C), and this structure has been shown to undergo thermal dissociation by a similar mechanism (26). Separation between the Hoogsteen and W-C phases is greater in the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ transitions (Figures 4A and B) than in the $d(A)_{10} \cdot 2d(T)_{10}$ transitions (Figure 4C). Since the T_{max} values for the Hoogsteen phases in 50 mM $MgCl_2$ are essentially the same, this increased separation reflects the additional thermal stability afforded the $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex by the six G·C base pairs.

The Hoogsteen phases of both the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ and $d(A)_{10} \cdot 2d(T)_{10}$ transitions in 50 mM $MgCl_2$ (Figures 4B and C) are considerably broader [by roughly 6.5°C in the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ transitions] than the W-C phases. In contrast, the Hoogsteen phases of the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ transitions in 2.0 M NaCl (Figure 4A) are approximately 2.5°C sharper than the W-C phases and roughly 6.8°C sharper than the Hoogsteen phases of the transitions in 50 mM $MgCl_2$. Transition breadth differences between those Hoogsteen or W-C phases in the presence of NaCl and those in the presence of $MgCl_2$ may result in part because of differing binding affinities and/or activities of Na^+ and Mg^{2+} ions for the triplex, duplex, and single-stranded structures.

In the melting experiments, upper baselines for the Hoogsteen phases and lower baselines for the W-C and Hoogsteen phases were not always accurately measurable. Hence, thermal denaturation profiles were evaluated by determining the temperatures (T_{max} 's) at which $d(A)/d(1/T)$ vs. T plots reached their maxima. Obtaining T_{max} in this way does not require knowledge of upper and lower baselines. T_{max} is the temperature at which the fraction of purine strands in the helical form is 0.486 for the duplex portion of the triplex and 0.414 for the triplex (see Materials and Methods), and thus differs from the more commonly used melting temperature, T_m , at which the fraction of strands in helical form is defined to be 0.50. The T_{max} values of the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ and $d(A)_{10} \cdot 2d(T)_{10}$ helix-coil transitions at a purine strand concentration (C_{pur}) of 10 μM are listed in Tables I and II, respectively. In accordance with the melting mechanisms proposed above for both triplex sequences, the Hoogsteen transitions have consistently lower T_{max} values than the corresponding W-C transitions for each of the DNA concentrations that were examined.

Thermodynamic Analysis. Plots of $1/T_{max}$ vs. $\log C_{pur}$ are shown in Figures 5A and B for the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ helix-coil transitions in 2.0 M NaCl and 50 mM $MgCl_2$, respectively, and in Figure 5C for the $d(A)_{10} \cdot 2d(T)_{10}$ transitions in 50 mM $MgCl_2$. The Hoogsteen and W-C phases of the transitions for both triplex sequences, regardless of ionic conditions, have $1/T_{max}$ values which decrease linearly with increasing $\log C_{pur}$ over the entire DNA concentration range that was studied. This linear correlation suggests that both phases reflect bimolecular melting processes, supporting the notion that the dissociation of the third strand from the duplex occurs as an all-or-none process. Thermodynamic parameters derived from the plots in Figure 5 as described in the Materials and Methods section as well as equilibrium binding constants at 25°C are listed in Tables I and II. The values of ΔG° , ΔH° , and ΔS° refer to a standard condition in which one molar concentrations of two different single strands or of duplex and single strand react to form one molar concentrations of duplex or triplex, respectively.

The free energy change at 25°C of Hoogsteen base pair formation is 0.22–0.64 kcal/mole bp formed less negative than that for W-C base pair formation for either sequence in 2.0 M NaCl and/or 50 mM $MgCl_2$. Hoogsteen base pairs are thus 0.22–0.64 kcal/mole less stable than the corresponding W-C base pairs. In 50 mM $MgCl_2$, the ΔG°_{25} of Hoogsteen base pairing in the $d(C^+{}_3T_4C^+{}_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex was only 0.02 kcal/mole base pair formed less negative than that in the $d(A)_{10} \cdot 2d(T)_{10}$ triplex. Therefore, excluding the possible contribution from nearest neighbor effects, which are not currently known for Hoogsteen base pairs, $C^+ \cdot G$ and $T \cdot A$

Figure 5. Plots of $1/T_{\max}$ vs. $\log(C_{\text{pur}})$ for the $d(C^+3T4C^+3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex in 2.0 M NaCl (A) or 50 mM $MgCl_2$ (B) and for the $d(A)_{10} \cdot 2d(T)_{10}$ triplex in 50 mM $MgCl_2$ (C). Data from the Hoogsteen phases of the thermal denaturation profiles for the triplexes are presented as triangles, while those from the Watson-Crick phases are presented as circles. C_{pur} is the total concentration of purine strands, and K^{-1} is reciprocal degrees Kelvin. See Materials and Methods for definition of $T_{\max}$.

Hoogsteen base pairs appear to have similar stability under these conditions.

Three sets of control experiments were done in order to confirm the validity of our method of analysis, and the resulting thermodynamic parameters were in good agreement with each other. First, for both the $d(A)_{10} + d(T)_{10}$ and $d(G_3A_4G_3) + d(C_3T_4C_3)$ systems, the thermodynamic parameters for duplex formation in solutions containing 1:1 stoichiometries of purine to pyrimidine strands and the same concentrations of purine strands as in the solutions containing 1purine:2pyrimidine stoichiometries were determined (data not shown). In each case, discrepancies between these values and those calculated for the duplex portion of the triplex were within experimental error. In contrast to the results of Krakauer and Sturtevant (29) for the poly(rA) + poly(rU) system, duplex formation in the presence or absence of a third strand is thus thermodynamically equivalent for these oligomer sequences. Second, thermodynamic parameters for duplex and triplex formation derived from monophasic melting profiles at 284 and 300 nm of solutions containing 1purine:2pyrimidine stoichiometries were also determined (data not shown). Discrepancies between these values and those determined for both phases of the melting profiles at 260 nm (Figure 3) were also within experimental error.

Third, 'differential' thermal denaturation experiments were carried out on solutions containing either 1:1 or 1:2 stoichiometric ratios of $d(G_3A_4G_3)$ to $d(C_3T_4C_3)$, but equimolar with respect to concentration of $d(G_3A_4G_3)$. In these experiments, which were run in 2.0 M NaCl at pH 5.5, the UV absorbance of the 1:2 solution was referenced to that of the 1:1 solution. The resulting profile therefore corresponds to the dissociation of the third strand. Discrepancies between the thermodynamic parameters derived from such experiments and those derived from the Hoogsteen phases of the helix-coil transitions for solutions containing 1 $d(G_3A_4G_3)$:2 $d(C_3T_4C_3)$ stoichiometric ratios were within experimental error as well. Hence, reliable thermodynamic parameters for Hoogsteen base pair formation could be derived from the Hoogsteen phases in the thermal denaturation profiles

Table I. Thermodynamic Parameters for the $d(G_3A_4G_3) + d(C_3T_4C_3)$ System

Base Pair Interaction ^a	Cation ^d	$T_{\max}$ (°C) ^c	ΔH° (kcal/mole) ^f	ΔS° (cal/mole deg) ^f	ΔG°_{25} (kcal/mole) ^f	K (25°C) (M^{-1})
HG ^b	Na ⁺	29.9 ± 1.0	-34.4 ± 1.5	-90 ± 5	-7.5 ± 0.5	3.2 × 10 ⁵
W-C ^c	Na ⁺	51.9 ± 1.0	-62.5 ± 5.1	-171 ± 16	-11.7 ± 0.5	3.8 × 10 ⁸
HG ^b	Mg ²⁺	20.6 ± 1.0	-42.0 ± 4.1	-120 ± 14	-6.4 ± 0.5	4.9 × 10 ⁴
W-C ^c	Mg ²⁺	52.9 ± 1.0	-72.9 ± 1.9	-202 ± 6	-12.8 ± 0.5	2.4 × 10 ⁹

^a HG is Hoogsteen; W-C is Watson-Crick

^b Reaction is duplex + single strand → triplex

^c Reaction is single strand + single strand → duplex

^d [Na⁺] = 2.0 M; [Mg²⁺] = 50 mM

^e $T_{\max}$ at a purine strand concentration of 10 μ M

^f Mole refers to mole structure formed (duplex for W-C interactions or triplex for HG interactions)

Table II. Thermodynamic Parameters for the $d(A)_{10} + d(T)_{10}$ System*

Base Pair Interaction ^a	Cation ^d	$T_{\max}$ (°C) ^c	ΔH° (kcal/mole) ^f	ΔS° (cal/mole deg) ^f	ΔG°_{25} (kcal/mole) ^f	K (25°C) (M^{-1})
HG ^b	Mg ²⁺	20.6 ± 1.0	-23.3 ± 0.9	-56 ± 3	-6.6 ± 0.5	6.9 × 10 ⁴
W-C ^c	Mg ²⁺	36.6 ± 1.0	-59.9 ± 2.4	-171 ± 8	-8.8 ± 0.5	2.8 × 10 ⁶

*Footnotes are the same as in Table I.

for solutions containing 1:2 stoichiometries of purine to pyrimidine strands, even when separation between the W-C and Hoogsteen phases was incomplete.

DISCUSSION

Two features emerge from the triplex helix-coil transitions presented in Figure 4. First, the transition breadths of the Hoogsteen phases for both the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ and $d(A)_{10} \cdot 2d(T)_{10}$ transitions in 2.0 M NaCl and/or 50 mM $MgCl_2$ differ from those of the W-C phases. These distinctions suggest differences in the mechanism of denaturation of W-C and Hoogsteen base paired strands. A statistical mechanical consideration of helix-coil transitions in oligonucleotides views duplex formation as an initial nucleation reaction followed by a series of growth reactions (37). The overall binding constant for complete strand association is thus a function of the binding constants for the nucleation and growth reactions. High equilibrium binding constants for the growth reactions (38), linear correlations between $1/T_m$ and $\log C_T$ values (34–36), agreement between calorimetrically and spectrophotometrically-derived enthalpy values (35), and separate kinetic studies (39,40) indicate that duplex formation in oligonucleotides may be approximated as an all-or-none (two-state) process. Binding constants for the nucleation and/or growth reactions of the Hoogsteen base pairing part of triplex formation may differ from those for the W-C base pairing portion, and thereby contribute to the observed differences in transition breadth. In support of this notion, the overall binding constant at 25°C for the Hoogsteen base pairing of the third strand to the duplex is two to five orders of magnitude smaller than that for the W-C base pairing of the duplex strands, depending on sequence and ionic conditions (Tables I and II). However, despite this difference in the magnitude of the overall binding constant, the linear correlation between $1/T_{max}$ and $\log C_{pur}$ and agreement between thermodynamic data derived from melting profiles taken at different wavelengths suggest that triplex (Hoogsteen base pair) formation in these oligonucleotide systems may be approximated by an all-or-none process as well. Additional equilibrium binding, kinetic and calorimetric studies on these and other oligomer sequences with the potential of forming triple helices would allow determination of the binding constants for the nucleation and growth reactions of Hoogsteen base pairing.

Second, the nature of the differences in transition breadth between the Hoogsteen and W-C phases of the helix-coil transitions depends on the type of cationic environment. In $MgCl_2$, the Hoogsteen phase for each sequence is broader than the W-C phase, whereas it is comparatively sharper for the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex in NaCl. Mg^{2+} ions bind duplex DNA much more tightly than do Na^+ ions, as evidenced by the former having an activity coefficient of approximately 0.07 and latter having one of roughly 0.37 at similar DNA concentrations (41).

There are conflicting views as to whether Mg^{2+} ions bind duplex or single-stranded DNA more tightly and whether binding is site specific or diffuse in nature (37,41,42). However, more counterions are thought to associate with duplex than with single-stranded DNA because the former has a higher charge density (37). Fiber diffraction studies indicate that the axial rise per residue is 0.11 Å smaller in the $poly(dA) \cdot 2poly(dT)$ triplex than in standard B-DNA (20–22,43). Given this structural difference, coupled with the additional negative charges introduced by the

third strand, the charge density of the triplex should be even greater than that of the duplex. This potential variation in charge density is supported by the work of Krakauer and Sturtevant (29) on the $poly(rA) + poly(rU)$ system in which they determined the fraction of Na^+ ions released per base pair upon denaturation of the $poly(rA) \cdot poly(rU)$ duplex or dissociation of the third strand [$poly(rU)$] from the $poly(rA) \cdot poly(rU)$ duplex. The fraction of Na^+ ions released per base pair for both types of dissociation reactions was quite close (0.31 and 0.29, respectively), suggesting that more Na^+ ions were bound to the triplex than to the duplex form. It is therefore likely that Na^+ and Mg^{2+} ions have differing activities and binding affinities for triplex structures, which may differ still from those they have for duplex structures. These differences might affect the manner by which Hoogsteen and W-C base paired strands denature in the presence of either cation.

The ΔG°_{25} for Hoogsteen base pair formation was less negative than that for W-C base pair formation by 0.42 and 0.64 kcal/mole base pair for the $d(G_3A_4G_3) + d(C_3T_4C_3)$ system in 2.0 M NaCl and 50 mM $MgCl_2$, respectively, and by 0.22 kcal/mole base pair for the $d(A)_{10} + d(T)_{10}$ system in 50 mM $MgCl_2$. Hence, in each of these cases, Hoogsteen base pairs are less stable than the corresponding W-C base pairs. The enthalpy and entropy changes for Hoogsteen base pair formation in both sequences were also less negative than those for W-C base pair formation, irrespective of ionic conditions (Tables I and II). This enthalpy difference thus contributes to the observed reduction in stability of Hoogsteen relative to the corresponding W-C base pairs. In contrast, the entropic barrier is lower for the formation of Hoogsteen base pairs. Solvent effects may contribute in part to the reduction of the entropic barrier. More water molecules may be freed upon triplex formation than upon duplex formation. Since structural studies on triplexes (20–23) indicate that the duplex portion of the triplex is conformationally distinct from standard B-DNA, it is likely that the conformational change which the duplex undergoes upon binding of the third strand contributes to the observed thermodynamic parameters for Hoogsteen base pair formation. The less negative ΔH° values of Hoogsteen base pair formation relative to those of W-C base pair formation agree with the pattern of enthalpy changes observed by others for these types of interactions between strands of $poly(rA)$ and $poly(rU)$ (27–29).

Quadrifoglio and coworkers have recently used both a two and a three-state model to describe formation of DNA triplexes containing either a loop (30) or a hairpin (31), respectively. Including contributions from cytosine protonation, they report an average enthalpy change of 6.2–7.5 kcal/mole base pair for Hoogsteen base pair formation. These values are consistently higher than our values for either sequence (2.3–4.2 kcal/mole base pair) and those of others (1.3–4.3 kcal/mole base pair) (27–29) for the $poly(rA) + poly(rU)$ system. This difference may arise from variations in the method and model used for analysis, pH, types and concentrations of cations, end effects, type and pucker of the sugar moiety, and base composition (stacking interactions).

Comparison of the thermodynamic parameters derived from the W-C and Hoogsteen phases of the $d(C^+_3T_4C^+_3) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ and $d(A)_{10} \cdot 2d(T)_{10}$ helix-coil transitions in 50 mM $MgCl_2$ (Tables I and II, respectively) reveals two key characteristics. First, the ΔG°_{25} of formation for the $d(A)_{10} \cdot d(T)_{10}$ duplex is 0.40 kcal/mole base pair less negative than that for formation of the $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex. An

obvious contribution to this free energy difference is the additional stability afforded the $d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ duplex by the extra hydrogen bonds introduced upon substituting A·T with G·C base pairs. However, base sequence has been shown to markedly affect the thermodynamics of duplex formation via nearest-neighbor (base stacking) interactions (44). Furthermore, since the conformation of the poly(dA)·poly(dT) duplex is of a non-standard B-form (43), the conformation of the $d(A)_{10} \cdot d(T)_{10}$ duplex may be similar in nature. It is therefore likely that base-stacking interactions and conformational differences are contributing to the observed thermodynamic variations as well.

Second, the ΔG°_{25} for Hoogsteen base pair formation in the $d(A)_{10} \cdot 2d(T)_{10}$ triplex is only 0.02 kcal/mole base pair less negative than that for Hoogsteen base pair formation in the $d(C^+T_4C^+) \cdot d(G_3A_4G_3) \cdot d(C_3T_4C_3)$ triplex. Despite this similarity in ΔG°_{25} values, the ΔH° and ΔS° values differ markedly. Hence, the enthalpic and entropic changes associated with the replacement of six T·A with six $C^+ \cdot G$ Hoogsteen base pairs essentially compensate for each other. As a result, T·A and $C^+ \cdot G$ Hoogsteen base pairs appear equivalent in regard to overall stability under these conditions. A possible caveat to this notion may reside in potential energetic contributions from nearest-neighbor interactions. Additional studies on various oligonucleotides containing all four possible pyrimidine/purine nearest-neighbor interactions and evaluation of the free energy changes associated with the nucleation and growth reactions of Hoogsteen base pairing will be necessary to fully address this issue.

ACKNOWLEDGEMENTS

We thank Dr. Corey Levenson of the Cetus Corporation, Emeryville, CA, for help with the synthesis and purification of the oligonucleotides. This work was supported by U. S. Public Health Service Grant CA27343, awarded by the National Cancer Institute, and Training Grant GM08284, awarded by the National Institute for General Medical Sciences, Department of Health and Human Services.

REFERENCES

- Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., and Wohlrab, F. (1988) *FASEB J.* **2**, 2939–2949.
- Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V., and Vassart, G. (1985) *Nucleic Acids Res.* **13**, 5127–5144.
- Htun, H., and Dahlberg, J. E. (1988) *Science* **241**, 1791–1796.
- Hanvey, J. C., Shimizu, M., and Wells, R. D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6292–6296.
- Johnston, B. H. (1988) *Science* **241**, 1800–1804.
- Kohwi, Y. and Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3781–3785.
- Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., and Frank-Kamenetskii, M. D. (1987) *Nature (London)* **330**, 495–497.
- Voloshin, O. N., Mirkin, S. M., Lyamichev, V. I., Belotserkovskii, B. P., and Frank-Kamenetskii, M. D. (1988) *Nature (London)* **333**, 475–476.
- Larsen, A., and Weintraub, H. (1982) *Cell* **29**, 609–622.
- Dybvig, K., Clark, C. D., Aliperti, G., and Schlesinger, M. J. (1983) *Nucleic Acids Res.* **11**, 8495–8508.
- Caddle, M. S., Lussier, R. H., and Heintz, N. H. (1990) *J. Mol. Biol.* **211**, 19–33.
- Glaser, R. L., Thomas, G. H., Siegfried, E., Elgin, S. C. R., and Lis, J. T. (1990) *J. Mol. Biol.* **211**, 751–761.
- Moser, H. E., and Dervan, P. B. (1987) *Science* **238**, 645–650.
- Praseuth, D., Perrouault, L., Le Doan, T., Chassignol, M., Thuong, N., and Hélène, C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1349–1353.
- Perrouault, L., Asseline, U., Rivalle, C., Thuong, N. T., Bisagni, E., Giovannangeli, C., Le Doan, T., and Hélène, C. (1990) *Nature (London)* **344**, 358–360.
- Strobel, S. A., Moser, H. E., and Dervan, P. B. (1988) *J. Am. Chem. Soc.* **110**, 7927–7929.
- Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., and Hogan, M. E. (1988) *Science* **241**, 456–459.
- Felsenfeld, G., Davies, D. R., and Rich, A. (1957) *J. Am. Chem. Soc.* **79**, 2023–2024.
- Riley, M., Maling, B., and Chamberlin, M. J. (1966) *J. Mol. Biol.* **20**, 359–389.
- Arnott, S., and Selsing, E. (1974) *J. Mol. Biol.* **88**, 509–521.
- Arnott, S., Chandrasekaran, R., Hukins, D. W. L., Smith, P. J. C., and Watts, L. (1974) *J. Mol. Biol.* **88**, 523–533.
- Arnott, S., Bond, P. J., Selsing, E., and Smith, P. J. C. (1976) *Nucleic Acids Res.* **3**, 2459–2470.
- Rajagopal, P., and Feigon, J. (1989) *Biochemistry* **28**, 7859–7870.
- Rajagopal, P., and Feigon, J. (1989) *Nature (London)* **339**, 637–640.
- Santos, C. D. L., Rosen, M., and Patel, D. (1989) *Biochemistry* **28**, 7282–7289.
- Pilch, D. S., Levenson, C., and Shafer, R. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1942–1946.
- Ross, P. D., and Scruggs, R. L. (1965) *Biopolymers* **3**, 491–496.
- Neumann, E., and Ackermann, T. (1967) *J. Phys. Chem.* **71**, 2377–2379.
- Krakauer, H., and Sturtevant, J. M. (1968) *Biopolymers* **6**, 491–512.
- Xodo, L. E., Manzini, G., and Quadrifoglio, F. (1990) *Nucleic Acids Res.* **18**, 3557–3564.
- Manzini, G., Xodo, L. E., Gasparotto, D., Quadrifoglio, F., Van Der Marel, G. A., and Van Boom, J. H. (1990) *J. Mol. Biol.* **213**, 833–843.
- Chamberlin, M. J. (1965) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **24**, 1446–1457.
- Griswold, B. L., Humoller, F. L., and McIntyre, A. R. (1951) *Anal. Chem.* **23**, 192–194.
- Gralla, J., and Crothers, D. M. (1973) *J. Mol. Biol.* **78**, 301–319.
- Albergo, D. D., Marky, L. A., Breslauer, K. J., and Turner, D. H. (1981) *Biochemistry* **20**, 1409–1413.
- Aboul-ela, F., Koh, D., Martin, F. H., and Tinoco, I. Jr. (1985) *Nucleic Acids Res.* **13**, 4811–4824.
- Bloomfield, V. A., Crothers, D. M., and Tinoco, I. Jr. (1974) *Physical Chemistry of Nucleic Acids*, Harper & Row, New York, pp. 294–360.
- Gralla, J. and Crothers, D. M. (1973) *J. Mol. Biol.* **73**, 497–511.
- Pörschke, D., and Eigen, M. (1971) *J. Mol. Biol.* **62**, 361–381.
- Craig, M. E., Crothers, D. M., and Doty, P. (1971) *J. Mol. Biol.* **62**, 383–401.
- Lyons, J. W., and Kotin, L. (1964) *J. Am. Chem. Soc.* **86**, 3634–3640.
- Dove, W. F., and Davidson, N. (1962) *J. Mol. Biol.* **5**, 467–478.
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, pp. 234–236.
- Breslauer, K. J., Frank, R., Blöcker, H., and Marky, L. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3746–3750.