Base Pairing and Steric Interactions between Pyrimidine Strand Bridging Loops and the Purine Strand in DNA Pyrimidine·Purine·Pyrimidine Triplexes†

Mark A. Booher, Shaohui Wang, and Eric T. Kool*

Department of Chemistry, University of Rochester, Rochester, New York 14627

Received December 6, 1993; Revised Manuscript Received February 9, 1994®

ABSTRACT: Bimolecular triple-helical DNA complexes recently have found use in a new strategy for the recognition of single-stranded nucleic acids, in which circular (Kool, 1991; Prakash & Kool, 1992) or hairpin-shaped (Giovannangeli et al., 1991; D'Souza & Kool, 1992) oligonucleotides bind these single strands by triplex formation. Bimolecular triplexes may also be formed in vivo as H-DNA, where this structure may potentially play a role in gene expression and recombination (Belotserkovskii et al., 1990; Hanvey et al., 1989; Shimizu et al., 1989). In all of these complexes, the central strand of the triplex must pass beyond the loop that bridges the outer two strands, and models and preliminary experiments have indicated that there may be important interactions between this central strand and the loop (Prakash & Kool, 1992). We now report thermal denaturation studies carried out specifically to investigate these interactions in detail, using as a model the 5′-loop and 3′-loop complexes formed between 14 pyrimidine oligodeoxynucleotides having the sequence 5′-dTTCTTTTCL$_1$TTTL$_5$CTTTTCTT, where L$_1$ and L$_5$ represent varied nucleotides in the loop (which is underlined), and eight target strands having the sequence 5′-dCCCCCFAAGAAAAG-3′ or 5′-dGAAAAGAAFCCCCC-3′, where F is a varied nucleotide flanking the triplex in the central strand. Results correlated from 64 different sequence combinations show that there is wide variation in the stabilities of the complexes, indicating specific and substantial interactions between the nucleotides at the L$_1$, F, and L$_5$ positions. Melting temperatures at pH 7.0 range from 17.0 °C to 34.6 °C, and free energies (37 °C) range from −3.2 to −7.8 kcal mol^{-1}. Several general conclusions are drawn from the 64 combinations studied: (1) Extra stability is gained when one of the first (L$_1$) or fifth (L$_5$) nucleotides in the loop is complementary to the F nucleotide in the purine strand, with an average advantage of 1.9–3.0 °C in T_m and 0.6 kcal in free energy. (2) Even higher stability is gained when *both* the L$_1$ and L$_5$ nucleotides are complementary to the flanking F nucleotide. The advantage (relative to no complementarity) is 3.2–4.7 °C in T_m and 1.1–1.5 kcal in free energy, on average; however, evidence indicates that this interaction is not the result of standard triple-helix pairing. (3) The correct choice of loop nucleotides can add both binding affinity and sequence selectivity to bimolecular triplexes. Binding studies of two circular oligodeoxynucleotides constructed as a test of the loop studies shows that the results allow semiquantitative predictions of the stabilities of bimolecular triplexes that involve these loop interactions. In the design of synthetic oligonucleotides as triplex-forming agents, the results suggest optimal choices of loop nucleotides for binding a given target sequence. The results may also aid in the understanding of the relative stabilities of H-DNA complexes.

Bimolecular and unimolecular DNA triple-helical complexes recently have found use both as subjects for structural study (Rajagopal & Feigon, 1989; de los Santos et al., 1989) and as new and potentially useful motifs in nucleic acid recognition. For example, bimolecular triple helices are involved in a new strategy for nucleic acid recognition involving the binding of single strands by triplex formation (Kool, 1991; Giovannangeli et al., 1991; D'Souza & Kool, 1992). In this approach, two pyrimidine-rich strands are connected on one or both ends by loops, and these molecules form strong and highly selective complexes with single-stranded nucleic acids by sandwiching a purine-rich complementary sequence in between. In an earlier report, we found preliminary evidence that, in such a complex, the pyrimidine-bridging loop interacts with the purine strand as it passes across (Prakash & Kool, 1992). This was observed to change the overall binding affinity relative to the binding of short strands that cannot interact with the loop. Model building indicates that the most likely source of the energetic differences is steric interaction between the first and/or last nucleotide in the loop and the first flanking nucleotide adjacent to the triplex in the central targeted strand.

Another bimolecular triple-helical complex is that formed in H-DNA, and the proposed formation of this structure in vivo may potentially have biological consequences in gene expression, replication, or recombination (Mirkin et al., 1987; Kohwi & Kohwi-Shigematsu, 1988; Htun & Dahlberg, 1989; Shimizu et al., 1990; Ussery & Sinden, 1993). As is true for the synthetically designed cases above, the accepted models for the structure of H-DNA also involve a nucleotide loop bridging the central purine-rich strand, and so it seems likely that such loop–central strand interactions will occur in this complex as well. Although there is one report concerning loop sizes in these complexes (Shimizu et al., 1989), there are, to our knowledge, no studies to date on the effect of loop sequence on H-DNA stability.

In order to investigate the nature of these possible interactions, we have constructed 14 21-nucleotide triplex-forming DNA oligomers with varied combinations of loop sequences.

† This work was supported by the National Institutes of Health (GM46625). E.T.K. is the recipient of a Camille and Henry Dreyfus Teacher–Scholar Award and awards from the Young Investigator Programs of the Army Research Office, the Arnold and Mabel Beckman Foundation, and the Office of Naval Research.

* Author to whom correspondence should be addressed.

® Abstract published in *Advance ACS Abstracts,* April 1, 1994.

0006-2960/94/0433-4645$04.50/0

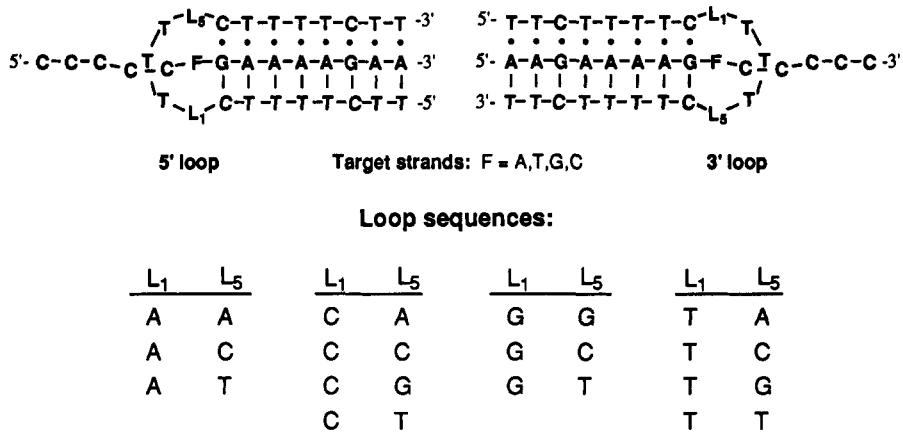


FIGURE 1: Illustration of complexes formed in this study, along with specific sequences of pyrimidine-rich probe strands and the eight purine-rich target strands. The complexes of the probes with targets containing the sequence 5′-GAAAAGAA result in 5′-loop orientation, while complexes with targets containing the sequence 5′-AAGAAAAG result in 3′-loop formation.

These are hybridized to eight 14-mer oligodeoxynucleotides with eight-base purine target sites and a variable flanking nucleotide, and the interactions are compared in terms of binding affinity. The results, combined with the results from the previous study of the intrinsic loop stabilities in the absence of these interactions (Wang et al., 1994), are expected to aid in understanding of the stability and structure of H-DNAs and to make possible the choice of optimal loop sequences in bimolecular and unimolecular triplexes.

EXPERIMENTAL PROCEDURES

Linear and Circular Oligodeoxynucleotide Synthesis. DNA oligomers were synthesized on a Pharmacia LKB automated synthesizer using the standard phosphoramidite method (Beaucage & Caruthers, 1981). The linear precursors of circles were also 5′-phosphorylated on the synthesizer using a commercially available reagent purchased from Cruachem (Horn & Urdea, 1986). Oligomers were purified by preparative 20% denaturing polyacrylamide gel electrophoresis and quantitated by absorbance at 260 nm. Extinction coefficients for the oligomers were calculated by the nearest-neighbor method (Borer, 1985).

The cyclization of the 5′-phosphorylated precursors of the circular ligands was carried out essentially as described previously (Prakash & Kool, 1991a). The 34-nucleotide precursors (50 μM) were allowed to form a complex around a 12-nucleotide complementary template oligomer (55 μM) having the sequence 5′-dAAGAAAGAAAAG, bringing the 5′-phosphate and 3′-hydroxyl ends in close proximity. Closure of the nick was carried out with BrCN (125 mM), imidazole hydrochloride (200 mM, pH 7), and $NiCl_2$ (100 mM) at 25 °C for 12 h (Kanaya & Yanagawa, 1986; Luebke & Dervan, 1989). The circular products were isolated by preparative denaturing gel electrophoresis; they migrated at a rate 0.9 times that of the linear starting materials.

Thermal Denaturation Studies. Solutions for the thermal denaturation studies contained a 1:1 molar ratio of variable-sequence pyrimidine 21-nucleotide oligomer and complementary 14-nucleotide purine oligomer (1.0 μM each). Buffers and solutions were prepared as previously reported (Wang et al., 1994). After the solutions were prepared, they were heated to 90 °C and allowed to cool slowly to room temperature prior to the melting experiments.

Thermal denaturation experiments were carried out as previously described (Wang et al., 1994). In nearly all cases (62 of 64), the complexes displayed sharp, apparently two-state transitions, with monophasic melting from bound complex to free oligomers. Cases that clearly were not two-state are noted in the text. Melting temperatures (T_m) are reported as the midpoint of transition, as determined by a computer fit of the first derivative of absorbance with respect to $1/T$. The uncertainty in T_m is estimated at ±0.3 °C on the basis of standard deviations of repeated experiments.

Each reported value for T_m and free energy is an average of 2–4 separate experiments; the repeated T_m values generally agreed within 0.5 °C, and the uncertainty in the averaged value is estimated at ±0.3 °C. Free energy values were derived by computer-fitting the denaturation data, using the two-state approximation for melting (Petersheim & Turner, 1983). Fits to the curves were excellent, with χ^2 values typically 5×10^{-6} or better. For the experiments in which the denaturation curves clearly were not two-state (see above), free energy values were not calculated. The temperature of 37 °C for free energy calculations was chosen both because of its physiological relevance and because it is near the T_m of the complexes, thus ensuring only a small extrapolation from the melting data (Petersheim & Turner, 1983). The uncertainty in individual free energy measurements is estimated at ±10–15%.

RESULTS

Structural Considerations. The formation of bimolecular triplexes between purine target strands and pyrimidine strands containing both Watson–Crick and Hoogsteen complementarity is well-precedented, with several examples appearing in the literature since 1990 (Xodo et al., 1990; Kool, 1991; Giovannangeli et al., 1991; D'Souza & Kool, 1992). To test specific interactions between triplex pyrimidine strand bridging loops and the central strand, we constructed eight target 14-nucleotide oligodeoxynucleotides with purine triplex target sites and one variable flanking base (F) directly adjacent to the triplex site (Figure 1). The target sequences were extended with five cytosines beyond the F nucleotide to mimic the structure of long substrates, which would occur in vivo. Also synthesized were 14 21-nucleotide pyrimidine-rich probe oligomers, which bind the target strands by triplex formation. Four of the target strands contain the sequence 5′-AA-GAAAAG, and the second four contain the reverse sequence, 5′-GAAAAGAA; triplex binding of the probes to the first group results in loop placement near the 5′-end of the target strand, while binding of the second group necessarily places the loop near the 3′-end of the target. The 14 probe oligomers,

Table 1: Melting Transition Temperatures and Free Energies of Complexation for Triplex-Forming Hairpin 21-mer Oligodeoxynucleotides Containing Variable Nucleotides L_1 and L_5, with 5′-dCCCCCFGAAAAGAA (5′-Loop Series) and 5′-dAAGAAAAGFCCCCC (3′-Loop Series) and Variable Nucleotide F, at pH 7.0 with 100 mM Na^+ and 10 mM Mg^{2+}

5′-loop			3′-loop		
L_1,F,L_5	T_m (°C)	$-\Delta G°_{37}$ (kcal mol^{-1})	L_1,F,L_5	T_m (°C)	$-\Delta G°_{37}$ (kcal mol^{-1})
T,A,A	22.5	4.7	T,A,A	26.5	5.8
T,A,C	*a*	*a*	T,A,C	27.6	6.0
T,A,G	21.8	4.7	T,A,G	25.7	5.8
T,A,T	28.9	7.6	T,A,T	27.2	6.1
A,A,T	24.4	5.4	A,A,T	25.6	5.7
C,A,T	29.2	6.4	C,A,T	26.9	5.9
G,A,T	24.9	6.1	G,A,T	26.9	6.2
G,A,G	24.8	5.3	G,A,G	29.8	6.0
T,C,A	22.3	3.4	T,C,A	27.3	5.4
T,C,C	*a*	*a*	T,C,C	24.4	5.6
T,C,G	23.9	3.3	T,C,G	33.6	7.7
T,C,T	20.4	5.8	T,C,T	22.9	5.4
A,C,T	19.5	4.9	A,C,T	24.5	4.4
C,C,T	20.7	5.7	C,C,T	22.0	4.9
G,C,T	23.5	6.1	G,C,T	26.7	5.4
G,C,G	26.9	5.9	G,C,G	34.6	7.8
C,G,A	25.8	5.4	C,G,A	28.5	5.8
C,G,C	22.9	5.7	C,G,C	30.2	6.7
C,G,G	25.4	6.1	C,G,G	28.0	5.9
C,G,T	22.8	6.1	C,G,T	27.7	6.5
A,G,C	19.2	3.5	A,G,C	30.4	6.6
T,G,C	32.3	6.6	T,G,C	27.0	6.0
G,G,C	21.8	3.8	G,G,C	33.7	7.6
A,G,A	20.0	4.0	A,G,A	25.8	5.5
C,T,A	29.8	6.4	C,T,A	20.1	3.2
C,T,C	22.1	4.7	C,T,C	18.2	3.3
C,T,G	28.2	6.2	C,T,G	24.6	4.3
C,T,T	21.2	4.7	C,T,T	21.2	5.5
A,T,C	22.5	4.5	A,T,C	20.3	4.5
T,T,C	22.5	5.3	T,T,C	17.0	3.5
G,T,C	25.3	5.4	G,T,C	21.3	3.2
A,T,A	29.6	6.3	A,T,A	21.0	4.4

[a] Melting transitions were not two-state.

described in the preceding article (Wang et al., 1994), bind eight-base purine complements in hairpin fashion, with five-base loops bridging the pyrimidine domains. The probe strands vary only in the loop sequence at the first (L_1) and fifth (L_5) positions. Models indicate that the most likely interstrand interactions occur between these two positions in the loop and the flanking base F in the target strand.

A test of all possible combinations of L_1, F, and L_5 nucleotides in both the 5′ and 3′ orientation (128 in total) was outside the scope of this study. We tested approximately one-half this number; however, to increase the accuracy of the data, each experiment was carried out multiple times and the results were averaged.

Overall Results of Thermal Denaturation Studies. Thermal denaturation experiments showed that the complexes varied considerably in melting temperatures and free energies of complexation (Table 1). With two exceptions, the melting curves showed apparently two-state denaturation behavior, allowing good fits to the curves for estimation of free energies. The range of T_m values at pH 7.0 is 17.0–34.6 °C, and the range of free energies ($-\Delta G°_{37}$) is 3.2–7.8 kcal mol^{-1}, which are similar in magnitude to the ranges of the complexes in the absence of loop interactions with the central strand (Wang et al., 1994). The strongest complex overall involves the L_1,F,L_5 combination G,C,G with the loop in the 3′ orientation. The weakest complex is that involving the combination C,T,A in the 3′ orientation. The global average T_m is 25.1 °C, and the free energy is −5.4 kcal mol^{-1}. These can be compared to the values of 28.6 °C and −5.8 kcal mol^{-1} for similar complexes with short target strands that cannot interact with the loop (Wang et al., 1994); this indicates that destabilizing interactions between the loop and the extended strand are common.

Three lines of evidence indicate that most or all of the complexes in the study are similar in overall structure. First, while the focus of this study is on data taken at pH 7.0, many complexes were also studied at pH 5.5 (data not shown), and they were all stabilized by this lower pH to a similar degree (another ~10 °C in T_m and 2–3 kcal in free energy). This is consistent with the expected triplex structure (Xodo et al., 1990). Second, the T_m and free energy values for the complexes vary quite consistently in parallel fashion, and the overall conclusions are the same using either of these values. This indicates that cooperativity (slope of the melting transition) is similar for the various sequences. Finally, the fact that the stabilities vary even among the weakest complexes indicates that secondary interactions are likely to be occurring throughout the series; if, for example, the Hoogsteen strand had melted off, interactions with the loop, and the resulting variations in stability, would be expected to disappear.

Effects of Complementarity between Loop and Target Strand. In order to obtain evidence on the nature of the interactions between loop nucleotides L_1 and L_5 and nucleotide F flanking the target site, the 64 data points were checked for correlations with specific structural features. For example, to investigate whether the loop interactions involve specific base pairing patterns, we averaged the data points that contain at least one Watson–Crick complementarity between the nucleotides L_1 or L_5 and the nucleotide F (Table 2). The average T_m and $-\Delta G°_{37}$ values were compared to those obtained for the cases in which no complementarity occurs.

The results show a clear and substantial positive correlation of stability with complementarity between the loop nucleotides L_1 and L_5 and the flanking base F (Table 2). The average values for noncomplementary cases are 23.8 °C and −5.1 kcal

Table 2: Averaged Values for Melting Transitions and Free Energies of Complexation Associated with Structural Features in the Loops of 21-mer Hairpin Sequences Containing Variable Nucleotides L_1 and L_5, When Bound to 14-mer Target Strands with Variable Nucleotide F

	5′-loop		3′-loop		all loops	
structural feature	T_m (°C)	$-\Delta G°_{37}$ (kcal mol^{-1})	T_m (°C)	$-\Delta G°_{37}$ (kcal mol^{-1})	T_m (°C)	$-\Delta G°_{37}$ (kcal mol^{-1})
overall averages	24.2	5.3	25.9	5.5	25.1	5.4
number of purines in loop						
0	24.5	5.9	24.3	5.5	24.4	5.6
1	23.8	5.0	26.5	5.5	25.1	5.2
2	25.3	5.4	27.8	5.9	26.6	5.7
complementarity of L_1 and L_5 to F						
noncomplementary	22.4	4.9	25.1	5.2	23.8	5.1
L_1 or L_5 complementary	25.4	5.5	27.0	5.8	26.2	5.7
L_1 and L_5 complementary	27.1	6.4	28.3	6.3	27.7	6.4

mol^{-1}. The values for at least one pairing complementarity are 26.2 °C and −5.7 kcal mol^{-1}. Thus, on average, a pairing arrangement is stabilizing to the complex, adding 2.4 °C to the T_m and −0.6 kcal to the free energy.

The 5′- and 3′-loop orientations bring about different geometric relationships between the L_1, F, and L_5 nucleotides. In the 5′-loop, the strands containing the L_1 and F nucleotides are parallel (Figure 1), while those containing the F and L_5 are antiparallel. In the 3′-loop orientation, the situation is reversed, with L_1 and F being antiparallel and F and L_5 parallel. Thus, if there were a single Watson–Crick pairing between the loop and the central strand, it would occur between the F and L_5 nucleotides in the 5′-loop and the L_1 and F nucleotides in the 3′-loop. To determine whether any correlation exists between stability and which side of the loop has complementarity, we averaged the data from Table1 accordingly. The results show that there appears to be no correlation of this sort, with base complementarity giving similar stabilization regardless of whether it occurs between parallel or antiparallel strands or on the 3′- or 5′-end of the complex. Parallel versus antiparallel cases show only a relatively small range of variation, with T_m's within 1.2 °C of each other and $\Delta G°$ values within 0.1 kcal. These ranges are small relative to the differences in the averaged data for complementary versus noncomplementary cases.

Interestingly, the correlated results show that there is even greater stability when *both* L_1 and L_5 nucleotides are complementary to the central F nucleotide (Table 2). For example, the case that has L_1,F,L_5 as T,A,T (on the 5′-end) gives a complex that is more stable than any of the other loops interacting with an A. Perhaps more surprisingly, the cases where L_1,F,L_5 is G,C,G are also among the most stable complexes, and similarly, the A,T,A loop interaction is among the more stable ones as well. When the total data are averaged, we find that cases in which both L_1 and L_5 nucleotides are complementary to F are more stable than singly matched cases by 1.5 °C in T_m and −0.7 kcal mol^{-1} in free energy, and the doubly matched cases overall are more stable than unmatched cases by 3.8 °C in T_m and −1.3 kcal mol^{-1}, on average. Thus, it appears that pairing complementarity on either or both sides of the loop adds to the affinity of these complexes.

Effects of Purine Substitution and Loop Orientation. In contrast to the results for these loops in the absence of interactions with the central strand (Wang et al., 1994), the data in the present study show no strong correlation of stability with the number of purines in the L_1 and L_5 positions. The total averaged values for T_m are 24.4, 25.1, and 26.6 °C for cases with zero, one, or two purines, respectively; however, the average corresponding values for $\Delta G°_{37}$ are −5.6, −5.2, and −5.7 kcal mol^{-1}. Thus, while purines stabilize these loops by themselves (Wang et al., 1994), the interaction of the loops with the central strand appears to negatively compensate for this advantage.

The inherent stabilities of the loops in the absence of interactions were previously found to be affected by the loop orientation, with 5′-loops being more stable than 3′-loops by a small amount, averaging 0.4 kcal (Wang et al., 1994). When interactions with the central strand are included, however, no such difference is seen. The average values for 5′-loops in the present study are a T_m of 24.2 °C and a free energy of −5.3 kcal mol^{-1}. For 3′-loops, the values are 25.9 °C and −5.5 kcal mol^{-1}; therefore, the two cases are within error limits of one another in free energy, and if there is any preference, it appears to be relatively small.

Table 3: Free Energies of Loop Interactions[a] with Extended Central Strands in a Bimolecular pyr·pur·pyr Triplex, as Measured by Subtracting the Free Energies of Complexes with Short Target Strands from Those with the Extended Strands[b]

5′-loop		3′-loop	
L_1,F,L_5	$\Delta\Delta G°_{37}$ (kcal)	L_1,F,L_5	$\Delta\Delta G°_{37}$ (kcal)
T,A,A	−2	T,A,A	−1
T,A,C	*c*	T,A,C	1
T,A,G	−1	T,A,G	0
T,A,T	2	T,A,T	1
A,A,T	0	A,A,T	0
C,A,T	1	C,A,T	1
G,A,T	0	G,A,T	1
G,A,G	−2	G,A,G	−1
T,C,A	−4	T,C,A	−1
T,C,C	*c*	T,C,C	0
T,C,G	−3	T,C,G	2
T,C,T	0	T,C,T	0
A,C,T	−1	A,C,T	−1
C,C,T	0	C,C,T	0
G,C,T	0	G,C,T	0
G,C,G	−1	G,C,G	1
C,G,A	0	C,G,A	0
C,G,C	1	C,G,C	2
C,G,G	−2	C,G,G	−1
C,G,T	0	C,G,T	1
A,G,C	−2	A,G,C	1
T,G,C	1	T,G,C	1
G,G,C	−3	G,G,C	2
A,G,A	−3	A,G,A	0
C,T,A	1	C,T,A	−2
C,T,C	0	C,T,C	−2
C,T,G	−2	C,T,G	−2
C,T,T	−1	C,T,T	0
A,T,C	−1	A,T,C	−1
T,T,C	0	T,T,C	−2
G,T,C	−1	G,T,C	−2
A,T,A	−1	A,T,A	−2

[a] To the nearest kilocalorie. [b] L_1, F, and L_5 are the varied nucleotides, and positive numbers indicate stabilizing interactions. [c] Melting transitions were not two-state.

Stabilization and Destabilization by Loop Interactions. While there appears to be a correlation of stability with base pairing interactions between the loop and the central strand, this correlation does not address whether these interactions are stabilizing or destabilizing overall relative to cases with no interactions. Since we have previously measured the stabilities of complexes with short target strands that do not pass beyond the loop, a direct comparison of the two sets of data (with interactions and without) may give information about this question. For example, if a probe strand with a given loop binds a long target strand more strongly than a short target that lacks flanking bases, then this presents evidence that the interactions between the loop and the flanking nucleotide are favorable.

Since comparison of the two data sets involves the subtraction of the separately measured free energies, it should be noted that the error limits are greater than those for the two individual sets of data. Table 3 shows the subtraction data, which are rounded to the nearest whole kilocalorie. Results show that the "energies of loop interaction" measured in this way vary from a destabilization by 4 kcal to a stabilization by 2 kcal. This is consistent with specific pairing interactions that are proposed to occur with the central strand; thus, a correctly matched base would be expected to be stabilizing, while mismatches would be destabilizing. It should also be noted, however, that in some cases the loops "without interactions" are probably involved in specific stabilizing interactions across the loop (Wang et al., 1994). For these

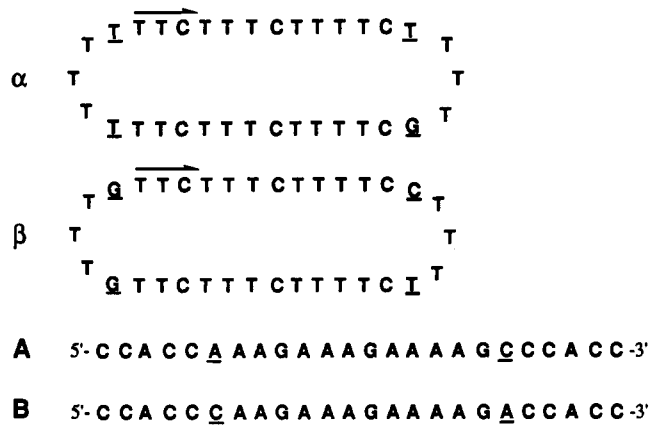


FIGURE 2: Sequences of the two circular DNAs (α and β) and two target sequences (A and B) constructed to test loop interaction predictions. Underlined bases correspond to the L_1, F, and L_5 nucleotides interacting in the loop study. Arrows show 5′–3′ orientation.

cases, the subtraction data in Table 3 will give energies of interaction that are too small. For example, the data involving the especially stable CTTTG loop are all negative numbers, which are influenced by the 1–2 kcal of extra stability in the complex with the short target strand.

Examination of the stabilizing and destabilizing trends in Table 3 shows that, of the 23 cases whose loops are not complementary to the flanking nucleotide F, all result in destabilizing (14 cases) or neutral (9 cases) interactions. By contrast, the cases that have at least one complementary base are about equally split between stabilizing interactions (15 cases) and destabilizing ones (14 cases). Thus, the results show that while complementarity does not guarantee stabilizing loop interactions, complete noncomplementarity is quite consistently neutral or destabilizing.

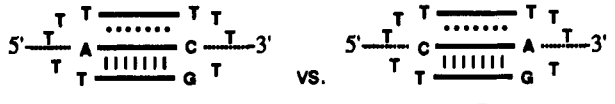
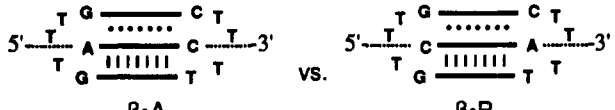
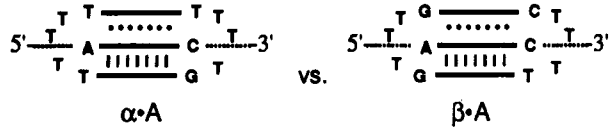
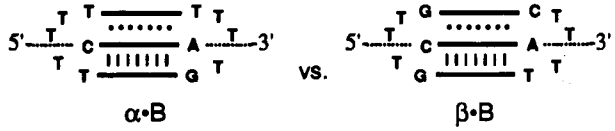
Testing Loop Interaction Predictions in a Second System. To test whether the present data concerning loop stabilities (listed in Table 1) have predictive value for other related triplexes, we constructed two circular oligodeoxynucleotides (circles "α" and "β" in Figure 2), which are designed to bind the 12-base sequence 5′-dAAGAAAGAAAAG by bimolecular triplex formation. We constructed two longer target sequences ("A" and "B" in Figure 2) that have flanking C and A nucleotides in both possible (5′ and 3′) orientations. The two circles are the same except for the nucleotides at the first and fifth loop positions, so that a given complex involving a circle simultaneously tests two loop interactions, such as those in the hairpin model system.

There are four possible complexes between the circles and target strands: α·A, α·B, β·A, and β·B. The data in Table1 allow predictions of the relative stabilities of the circle complexes, by combining the stabilities of the two loop interactions and using the initial assumption that the loop stabilities will be additive in the circle complexes. For example, the complex α·A involves a T,A,T arrangement for L_1,F,L_5 on the 5′-end, and a T,C,G arrangement on the 3′-end. Table 1 gives a total of 7.6 + 7.7 = 15.3 kcal from the model loops. The α·B complex, similarly, gives a total of 11.6 kcal. The *relative* stability of these two loop combinations is then the difference, with the α·A complex predicted to be 3.7 kcal more stable than the α·B complex.

The four complexes were measured three times each, and the T_m and free energy data were averaged. The results are as follows: α·A, T_m = 54.8 °C, $-\Delta G°_{37}$ = 17.2 kcal mol^{-1}; α·B, T_m = 49.6 °C, $-\Delta G°_{37}$ = 14.5 kcal mol^{-1}; β·A, T_m = 51.6 °C, $-\Delta G°_{37}$ = 16.1 kcal mol^{-1}; β·B, T_m = 54.0 °C, $-\Delta G°_{37}$ = 16.8 kcal mol^{-1}. The average standard deviations for these measurements were ±0.1 °C for T_m and ±0.6 kcal for $\Delta G°$.

Results show (Table 4) that there is good qualitative agreement between predicted and experimental values for the relative stabilities of the complexes. The data in Table 1 predict that, for the α circle, the preferred target should be the A sequence over the B sequence, and this is confirmed experimentally. Similarly, for the β circle, the experimentally preferred target is the B sequence over the A sequence, as predicted. From the point of view of the target strands, similar success in prediction is found: the A strand prefers to be bound by the α circle over the β as predicted, and the B strand

Table 4: Pairwise Comparison of Predicted and Measured Values of Relative Stabilities Resulting from Differing Loop Interactions in the Complexes of Two Circular Oligodeoxynucleotides with Two Target Sequences (Sequences in Figure 2)[a]

comparison ($\Delta G°_{\text{first complex}}$) − ($\Delta G°_{\text{second complex}}$)	experimental relative stability ($\Delta\Delta G°_{37}$, kcal)	predicted relative stability ($\Delta\Delta G°_{37}$, kcal)
 α·A vs. α·B	−2.7	−3.7
 β·A vs. β·B	+0.7	+1.6
 α·A vs. β·A	−1.1	−5.1
 α·B vs. β·B	+2.3	+0.2

[a] See text for the derivation of the predicted values.

prefers to be complexed by the β circle, also correctly predicted by the model experiments.

When the four complexes are taken together, the predicted and experimentally measured orders of stability agree reasonably well, although not perfectly. The predicted order of stability is α·A > β·B > α·B > β·A. The order actually found is α·A > β·B > β·A > α·B. The predicted overall range of stabilities is 5.1 kcal mol^{-1}, and the range found is somewhat smaller, 2.7 kcal mol^{-1}.

DISCUSSION

Structural Considerations of Complementarity between Loop and Target Strand. The results described above show significant correlations of stability with base complementarity between the loop nucleotides and the central strand. Since complementarity appears to be favorable on either side, and even simultaneously on both sides of the central strand, one is tempted to propose that the first and last loop bases may be involved in a triplex-type interaction with the central strand. This would imply the formation of a triplex one base triad longer than was originally designed, with a three-nucleotide loop bridging the pyrimidine strands. However, two lines of evidence argue against this possibility. First, an earlier study (Prakash & Kool, 1992) showed that five-nucleotide loops are preferred over shorter loops, even when the first and last nucleotides are complementary in triplex fashion with the central strand. This argues that such a triplex is unfavored energetically, possibly because the loops are too short to bridge the necessary distance without strain. Models support this premise: a trinucleotide in the B conformation only spans 17–18 Å, whereas the triplex interphosphate distance in a given step is approximately 20 Å (Harvey et al., 1988). The second line of evidence that argues against a standard triplex interaction between positions L_1, F, and L_5 is the fact that two-sided complementarity appears to be favorable even in cases that cannot form a canonical pyr·pur·pyr base triad; for example, the cases where L_1,F,L_5 is A,T,A and G,C,G are among the more stable interactions in this study, whereas these are strongly destabilizing triads in standard DNA triplexes (Griffin & Dervan, 1989; Roberts & Crothers, 1991; Mergny et al, 1991; Fossella et al., 1993).

It is unclear at present how this type of complementarity would be energetically favored, since pyrimidines in the central strand presumably can only pair with a single purine at a given time. It may be that the actual complementarity on the back side of the pyrimidine is not favored so much as the presence of a purine at this position. Although the overall stability in this study does not correlate well with the number of purines in loop positions, the previous study (Wang et al., 1994) did show clearly that purines are stabilizing to such a loop in the absence of central strand interactions. Further structural studies will be necessary to characterize this interaction in greater detail.

It is also interesting that cases with single-sided complementarity show no preference as to which side the complementarity occurs. This implies that pairing interactions may occur between strands oriented either parallel or antiparallel with each other. Recent studies have demonstrated that DNA is quite multifunctional in its ability to pair, and parallel Watson–Crick-paired and Hoogsteen-paired DNA duplexes recently have been observed (Ramsing & Jovin, 1988; Liu et al., 1993).

Potential Effects in H-DNA. The present studies indicate that complementarity on both sides of the central flanking F position adds to the stability of the overall complex. In a five-nucleotide loop, the optimal sequence then has the form XNNNX, since the first and last nucleotides will be the same for complementarity to F. In H-DNA, the triplex form is derived by the structural rearrangement of duplex DNA, so that in the H-DNA form involving a 5′-loop [the only observed isomer (Shimizu et al., 1989)], the first nucleotide will necessarily always be complementary to the central strand flanking nucleotide F (Figure 3). The present study implies that this favors H-DNA structurally, relative to mismatches. The fifth nucleotide in this structure, however, will not necessarily be complementary to F, and this study indicates that this complementarity would be expected to add to H-DNA stability. To date, we know of no reports comparing the stabilities of H-DNAs that have this complementarity; thus, this prediction must await further studies.

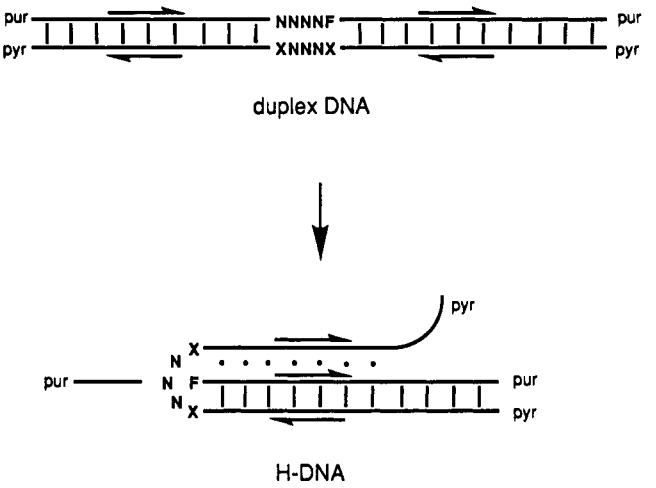


FIGURE 3: Schematic diagram illustrating the transition from duplex to H-form DNA and showing the disposition of loop nucleotides prior to H-form. The X nucleotides within the pyrimidine strand of the duplex become the first and last loop nucleotides in the H-form and ideally are complementary to the F nucleotide.

Loop Choices for Optimum Affinity and Selectivity. In the use of hairpin or circular oligonucleotides for binding a strand by triplex formation, the formation of the desired complex involves interactions with flanking nucleotides on one or both sides of the triplex site. To achieve optimum binding affinity, the best choice of loop nucleotides for a given flanking nucleotide F can be taken from Table 1. The present study shows that, in some of the central strand (F) cases, more than one loop sequence can interact with reasonably high affinity. The choice of sequence may be somewhat arbitrary in these cases; however, affinity is not the only property to be considered in DNA complexes. Sequence selectivity is also crucially important in DNA recognition, and since specific pairing interactions are proposed here, the correctly chosen interactions would be expected to add selectivity to a given complex. For example, in the binding of a hairpin-like molecule to a DNA strand, the loop interaction adds one nucleotide of target recognition to the site specificity, whereas an arbitrarily chosen loop would not (and, in fact, may select for an undesired nucleotide at this site). In a circular DNA triplex, the loop interactions add at least two nucleotides of recognition to the target site length.

In order to choose loop sequences from the present study for the highest selectivity, one can compare loop sequences and relative affinities in Table 1. Some loops show high affinity for more than one central nucleotide, and in these cases it may be desirable to choose a different loop sequence which may have the same or slightly lower affinity, but which has higher

selectivity for the desired flanking nucleotide. For example, if the flanking nucleotide is a 3′-A, the highest affinity loops would appear to be either TTTTT or TTTTC, which result in similar affinity. However, the TTTTT loop shows greater selectivity for the desired 3′-A than does the TTTTC loop, which binds a 3′-C with an affinity only 0.4 kcal lower.

Predictive Value of the Model System Data. The pairwise predictions for the relative stabilities of the circle complexes (Table IV) are quite successful on a qualitative level, with the stronger complex of the two predicted successfully in each of the four comparisons. On a quantitative level, the predicted range of free energies for the four complexes (5.1 kcal) is larger than the range found experimentally, indicating that the stabilities of the two loops in a given circle complex may not be entirely independent of each other. Even so, the measured range of stabilities is 4.8 °C in T_m and 2.7 kcal mol^{-1} in free energy, confirming that loop interactions are substantial relative to the total energies of complexation, which are ~15–17 kcal.

The present results offer evidence of significant and specific loop–central strand interactions, and they allow predictions for other related complexes as well. It is likely, however, that the effects will vary somewhat with context, depending on triplex sequence and direct nearest-neighbor effects. It is also possible that the effects may not be the same in triplexes containing RNA, although the present findings offer a useful starting point for study. In addition, further studies will be needed to investigate whether the central three nucleotides in the loop, which were not varied in this study, make significant contacts with the central strand.

CONCLUSIONS

We conclude that nucleotide loops that bridge the purine strand in DNA pyr·pur·pyr triple helices significantly interact with the purine strand and that substantial amounts of interaction arise from contacts between the first and last nucleotides in the loop and the first nucleotide flanking the triplex in the central strand. The present data support the possibility that base pairing interactions may occur between these positions and that these pairing interactions are stabilizing by up to 2–3 kcal relative to "mismatched" cases. Finally, the data allow qualitative predictions concerning the optimal loop nucleotide choices for a given central strand sequence in bimolecular triplexes.

REFERENCES

Beaucage, S. L., & Caruthers, M. H. (1981) *Tetrahedron Lett. 22*, 1859–1862.

Belotserkovskii, B. P., Veselkov, A. G., Filippov, S. A., Dobrynin, V. N., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1990) *Nucleic Acids Res. 18*, 6621–6624.

Borer, P. N. (1985) in *Handbook of Biochemistry and Molecular Biology*, 3rd ed. (Fasman, G. D., Ed.) Vol. I, p 589, CRC Press, Cleveland, OH.

de los Santos, C., Rosen, M., & Patel, D. J. (1989) *Biochemistry 28*, 7282–7289.

D'Souza, D. J., & Kool, E. T. (1992) *J. Biomol. Struct. Dyn. 10*, 141–152.

Fossella, J. A., Kim, Y. J., Shih, H., Richards, E. G., & Fresco, J. R. (1993) *Nucleic Acids Res. 21*, 4511–4515.

Giovannangeli, C., Montenay-Garestier, T., Rougée, M., Chassignol, M., Thuong, N. T., & Hélène, C. (1991) *J. Am. Chem. Soc. 113*, 7775–7776.

Hanvey, J. C., Shimizu, M., & Wells, R. D. (1989) *J. Biol. Chem. 264*, 5950–5956.

Harvey, S. C., Luo, J., & Lavery, R. (1988) *Nucleic Acids Res. 16*, 11795–11809.

Horn, T., & Urdea, M. (1986) *Tetrahedron Lett. 27*, 4705–4708.

Htun, H., & Dahlberg, J. E. (1989) *Science 243*, 1571–1576.

Kanaya, E., & Yanagawa, H. (1986) *Biochemistry 25*, 7423–7430.

Kohwi, Y., & Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3781–3785.

Kool, E. (1991) *J. Am. Chem. Soc. 113*, 6265–6266.

Liu, K., Miles, H. T., Frazier, J., & Sasisekharan, V. (1993) *Biochemistry 32*, 11802–11809.

Luebke, K. J., & Dervan, P. B. (1989) *J. Am. Chem. Soc. 111*, 8733–8734.

Mergny, J.-L., Sun, J.-S., Rougée, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J., & Hélène, C. (1991) *Biochemistry 30*, 9791–9798.

Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., & Frank-Kamenetskii, M. D. (1987) *Nature 330*, 495–497.

Moser, H. E., & Dervan, P. B. (1987) *Science*, 238, 645–650.

Petersheim, M., & Turner, D. H. (1983) *Biochemistry 22*, 256–263.

Plum, G. E., Park, Y.-W., Singleton, S. F., Dervan, P. B., & Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9436–9440.

Prakash, G., & Kool, E. T. (1992) *J. Am. Chem. Soc. 114*, 3523–3528.

Rajagopal, P., & Feigon, J. (1989) *Nature 339*, 637–640.

Ramsing, N. B., & Jovin, T. M. (1988) *Nucleic Acids Res. 16*, 6659–6676.

Roberts, R. W., & Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397–9401.

Shimizu, M., Hanvey, J. C., & Wells, R. D. (1989) *J. Biol. Chem. 264*, 5944–5949.

Ussery, D. W., & Sinden, R. R. (1993) *Biochemistry 32*, 6206–6213.

Voloshin, O. N., Mirkin, S. M., Lyamichev, V. I., Belotserkovskii, B. P., & Frank-Kamenetskii, M. D. (1988) *Nature 333*, 475–476.

Wang, S., Booher, M. A., & Kool, E. T. (1994) *Biochemistry* (preceding article in this issue).

Xodo, L. E., Manzini, G., & Quadrifoglio, F. (1990) *Nucleic Acids Res. 18*, 3557–3564.