# Structural Optimization of Non-Nucleotide Loop Replacements for Duplex and Triplex DNAs

**Squire Rumney, IV, and Eric T. Kool***

*Contribution from the Department of Chemistry, University of Rochester,
Rochester, New York 14627*

*Received February 1, 1995*<sup>®</sup>

**Abstract:** Described are studies systematically exploring structural effects in the use of ethylene glycol (EG) oligomers as non-nucleotide replacements for nucleotide loops in duplex and triplex DNAs. The new structurally optimized loop replacements are more stabilizing in duplexes and triplexes than previously described EG-based linkers. A series of compounds ranging in length from tris(ethylene glycol) to octakis(ethylene glycol) are derivatized as mono-dimethoxytrityl ethers on one end and phosphoramidites on the other, to enable their incorporation into DNA strands by automated methods. These linker molecules span lengths ranging from 13 to 31 Å in extended conformation. They are incorporated into a series of duplex-forming and triplex-forming sequences, and the stabilities of the corresponding helixes are measured by thermal denaturation. In the duplex series, results show that the optimum linker is the one derived from heptakis(ethylene glycol), which is longer than most previous loop replacements studied. This affords a helix with greater thermal stability than one with a natural $T_4$ loop. In the triplex series, the loop replacements were examined in four separate situations, in which the loop lies in the 5′ or 3′ orientation and the central purine target strand is short or extends beyond the loop. Results show that in all cases the loop derived from octakis(ethylene glycol) ($EG_8$) gives the greatest stability. In the cases where the target strand is short, the $EG_8$-linked probe strands bind with affinities in some cases greater than those with a natural pentanucleotide ($T_5$) loop. For the cases where the target strand extends beyond the linker, the $EG_8$-linked strand is destabilized relative to an optimum $T_5$-bridged strand and the stabilities with the EG-linked strands are much lower in the 5′ loop orientation than in the 3′ loop orientation. It is found that extension by one additional nucleotide in one of the binding domains in the EG-linked series can result in considerably greater stabilities with long target strands. Overall, the data show that optimum loop replacements are longer than would be expected from simple distance analysis. The results are discussed in relation to expected lengths and geometries for double and triple helixes. The findings will be useful in the design of synthetically modified nucleic acids for use as diagnostic probes, as biochemical tools, and as potential therapeutic agents.

## Introduction

There has been growing interest in recent years in the development of synthetically altered nucleic acid structure, for use in diagnostics applications, as tools in the study of gene expression and as potential therapeutic agents.[1] Synthetic modification of DNA and RNA structure can lead to potentially useful properties such as increased lifetime in biological media, improved ability to be taken up into cells, or improved hybridization properties.[2]

Nucleic acid loops are a ubiquitous part of RNA and DNA secondary structure. Nucleotide loops arise in folded domains such as occur in intrastrand duplexes and triple helixes. There has been increasing recent interest in synthetic DNAs and RNAs which are designed to contain hairpin loops. Studies in modification of these structures have led to the development of non-nucleotide linking groups which replace several nucleotides bridging a folded duplex or triplex structure.[3–12] Nonnucleotide

groups have also been used as linkers in nonfolded structures as well.[13–17] Such linking groups are potentially useful replacements of natural nucleotides for several reasons. First, they can shorten the synthesis of a given structure by several steps, since one relatively long linking group replaces several indi-

® Abstract published in *Advance ACS Abstracts,* May 1, 1995.

(1) Uhlmann, E.; Peyman, A. *Chem. Rev.* **1990,** *90,* 543.

(2) Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1993,** *49,* 6123–6194.

(3) Durand, M.; Chevrie, K.; Chassignol, M.; Thuong, N. T.; Maurizot, J. C. *Nucleic Acids Res.* **1990,** *18,* 6353–6359.

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

(5) Rumney, S.; Kool, E. T. *Angew. Chem.* **1992,** *104,* 1686–1689; *Angew. Chem., Int. Ed. Engl.* **1992,** *31,* 1617–1619.

(6) Salunkhe, M.; Wu, T.; Letsinger, R. L. *J. Am. Chem. Soc.* **1992,** *114,* 8768–8772.

(7) Durand, M.; Peloille, S.; Thuong, N. T.; Maurizot, J. C. *Biochemistry* **1992,** *31,* 9197–9204.

(8) Ma, M. Y.-X.; Reid, L. S.; Climie, S. C.; Lin, W. C.; Kuperman, R.; Sumner-Smith, M.; Barnett, R. W. *Biochemistry* **1993,** *32,* 1751–1758.

(9) Ma, M. Y.-X.; McCallum, K.; Climie, S. C.; Kuperman, R.; Lin, W. C.; Sumner-Smith, M.; Barnett, R. W. *Nucleic Acids Res.* **1993,** *21,* 2585–2589.

(10) Doktycz, M. J.; Paner, T. M.; Benight, A. S. *Biopolymers* **1993,** *33,* 1765–1777.

(11) Thomson, J. B.; Tuschl, T.; Eckstein, F. *Nucleic Acids Res.* **1993,** *21,* 5600–5603.

(12) Benseler, F.; Fu, D.; Ludwig, J.; McLaughlin, L. W. *J. Am. Chem. Soc.* **1993,** *115,* 8483–8484.

(13) Cload, S. T.; Schepartz, A. *J. Am. Chem. Soc.* **1991,** *113,* 6324–6326.

(14) Horne, D. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990,** *112,* 2435–2437.

(15) Kessler, D. J.; Pettitt, B. M.; Cheng, Y.-K.; Smith, S. R.; Jayaraman, K.; Vu, H. M.; Hogan, M. E. *Nucleic Acids Res.* **1993,** *21,* 4810–4815.

(16) Dolinnaya, N. G.; Blumenfeld, M.; Merenkova, I. N.; Oretskaya, T. S.; Krynetskaya, N. F.; Ivanovskaya, M. G.; Vasseur, M.; Shabarova, Z. A. *Nucleic Acids Res.* **1993,** *21,* 5403–5407.

(17) Nilsson, M.; Malmgren, H.; Samiotaki, M.; Kwiatkowski, M.; Chowdhary, B. C.; Landegren, U. *Science* **1994,** *265,* 2085–2088.

(18) Jäschke, A.; Fürste, J. P.; Nordhoff, E.; Hillenkamp, F.; Cech, D.; Erdmann, V. A. *Nucleic Acids Res.* **1994,** *22,* 4810–4817.

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

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

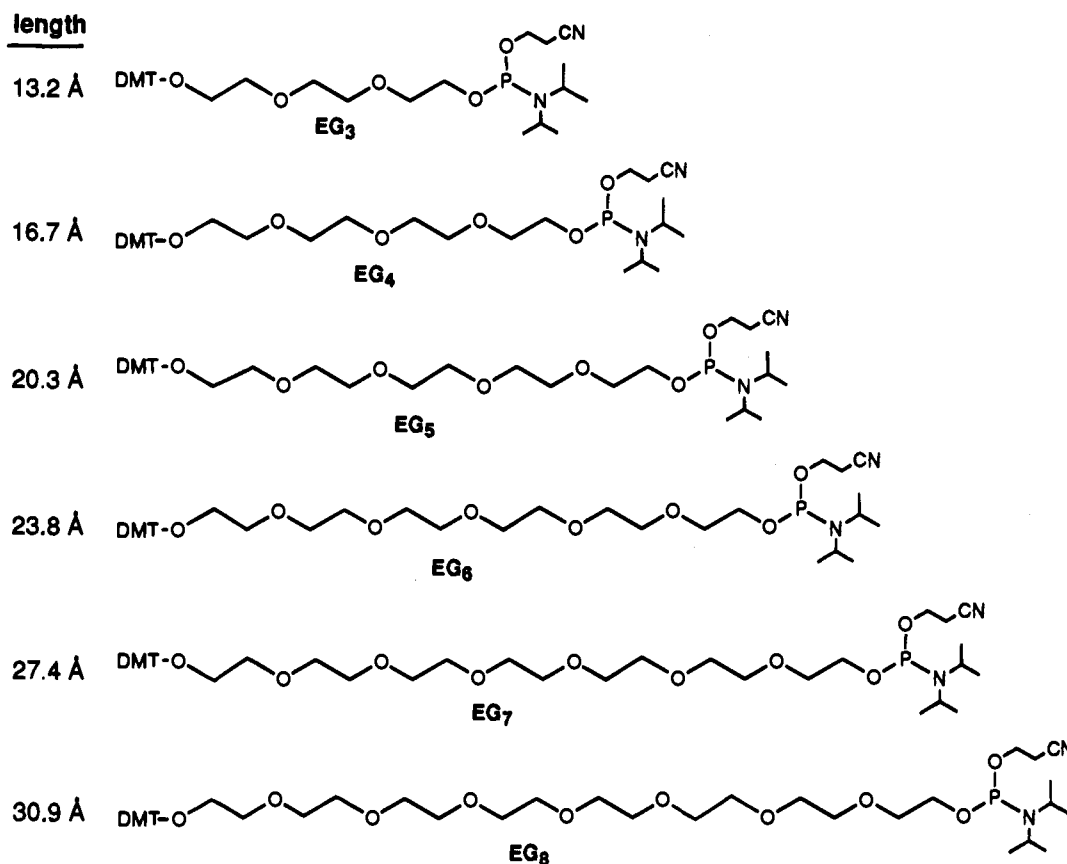(21) Petersheim, M.; Turner, D. H. *Biochemistry* **1983,** *22,* 256–263.

**Figure 1.** Structures and lengths of the six oligo(ethylene glycol)-based loop replacements in this study. The compounds are shown as the dimethoxytrityl (DMT) phosphoramidite derivatives which are used for incorporation into the DNA.

vidual nucleotides which normally constitute a loop. Second, such nonnatural loops can confer resistance to degradation by nucleases which would ordinarily act on a natural loop structure in biological applications.[5] Careful design of a loop replacement also has the potential to give more stable folded structure than occurs with the natural loop nucleotides.

Several nucleic acid loop replacements have been documented in the literature.[3-12] Maurizot et al. reported the use of hexakis-(ethylene glycol) as a replacement for a T4 loop in duplex DNA and found that the replacement led to slightly higher thermal stability than the natural loop.[3] Hélène reported the use of the same linker in a folded triplex-forming oligodeoxynucleotide, although its binding stability was not compared to that with a nucleotide loop.[4] We reported the use of pentakis- and hexakis-(ethylene glycol) chains in the linking of a triplex-forming circular DNA molecules.[5] In that preliminary study we found that the longer linker of the two was the most stabilizing in triple helical complexes, although the stability with a nucleotide loop was somewhat higher. Letsinger at al. reported a bis-amide-based linker for double- and triple-helical complexes but did not test nucleotide loops for comparison.[6] For loop replacements in duplex DNAs, Benight and co-workers described the use of dodecyl chains, which were found to be more stabilizing than nucleotide loops.[10] In duplex RNAs, Ma et al. reported the use of tris(ethylene glycol) and hexakis(ethylene glycol) loop replacements[8,9] and found the latter to be more stabilizing than the former. Similar linkers have been used recently to replace loops in catalytic RNAs;[11,12] in those cases catalytic activity, rather than thermodynamic stability, was studied. Ethylene glycol oligomers, and in particular hexakis-(ethylene glycol), have been the predominant structures used as such linkers. Ethylene glycols have the advantage of good water solubility, which can be a problem with hydrophobic

chains,[10] and ethylene glycol oligomers are available in multiple lengths and are easily derivatized in the same manner as nucleosides for incorporation into DNA.

Despite the growing number of laboratories utilizing synthetic loop replacements, there has not yet been a systematic study of simple unmodified ethylene glycol linkers of several different lengths. Such a study would be useful in evaluating the optimum lengths for such linkers, as well as in comparing stabilities with standard nucleotide loops. In an effort to characterize length and structural effects of such linkers both in duplex and triplex DNAs, we have synthesized and studied a series of modified oligodeoxynucleotides incorporating ethylene glycol oligomers containing three to eight monomer units, and thus spanning 13 to 31 Å in distance. We find that the hexakis(ethylene glycol) linker—the most commonly used loop replacement—is generally not the optimum structure for such complexes. The results of stability studies of 65 intra- and intermolecular DNA complexes are described here and are discussed in light of the structures, sequences, and geometries involved.

## Results and Discussion

To test the effects of linker structure to be used as a substitution for natural nucleotide loops in DNA, we studied the six ethylene glycol-derived compounds shown in Figure 1. We abbreviate the linkers $EG_n$, where EG represents ethylene glycol and $n$, the number of ethylene glycol monomers in the chain. Aside from our own work with the $EG_{4,5}$ chains,[5] only the $EG_3$ and $EG_6$ cases have been previously reported as linker groups; the other two ($EG_{7,8}$) have been conjugated to the end of oligonucleotides but not linking two DNA sequences. These last two were obtained in that study by separation from a poly-(ethylene glycol)-derived mixture.
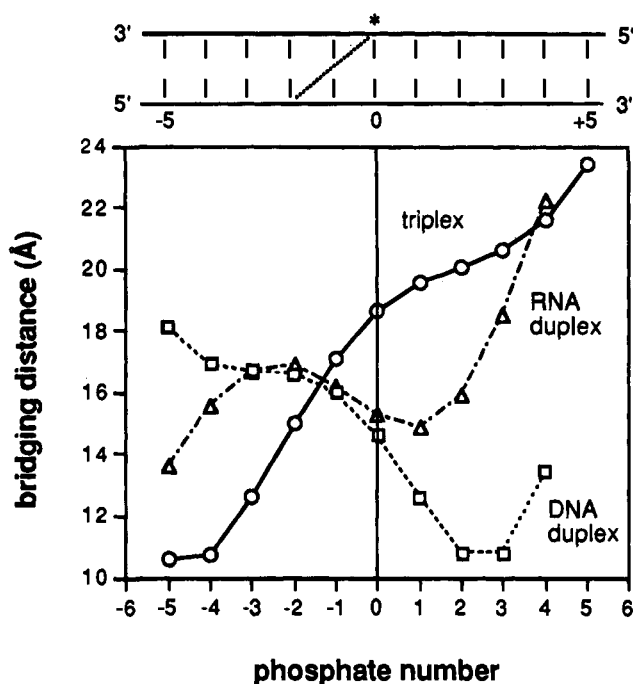
*Loop Replacements for DNA*

*J. Am. Chem. Soc., Vol. 117, No. 21, 1995* 5637



**Figure 2.** Haasnoot-type plots[25] of the bridging distance between 5'-oxygens as a function of the base steps involved for B-duplex DNA, A-duplex RNA, and an A'-regular triplex.[26] For duplexes, the distances are from one strand to the other; for pyr·pur·pyr triplex, the distance shown is from one pyrimidine strand to the other. The measurement starts from the * position in one strand and bridges to various 5'-O positions in the opposite strand. As an example, in a triplex, the distance from * to the −2 position (shown as a dashed line) is 15 Å.

**Synthetic Strategy.** In the present case we pursued a higher-yield approach to these compounds. The precursor alcohols tris-, tetrakis-, pentakis-, and hexakis(ethylene glycol) were all obtained from commercial sources. Mono-tritylated derivatives of heptakis(ethylene glycol) and octakis(ethylene glycol) were prepared by starting with mono-tritylated tetrakis(ethylene glycol), which was extended with tris- or tetrakis(ethylene glycol) by $S_N2$ displacements of tosylates. We thus obtained the mono-trityl derivatives of all six bis-alcohols and then derivatized the remaining hydroxyl group with a phosphoramidite moiety. In this manner, all six compounds were prepared for automated DNA synthesis, and they were incorporated into DNA strands in good yields using standard coupling protocols.

**Experimental Design.** Our aim was to study the effects of these loop replacements on helix stability both in double- and triple-helical DNA (see Figure 3). In the duplex DNA series we studied a hairpin-type duplex containing a six-base-pair stem bridged either by a natural $T_4$ loop or by one of the $EG_n$ series loop replacements. The sequence is 5'-dCGAACG-loop-CGT-TCG. This stem sequence was studied previously by Breslauer,[22] who showed that a $T_4$ loop was more stabilizing to the hairpin structure than the other natural loops $A_4$, $C_4$, and $G_4$. Altogether we studied one natural hairpin and six with non-nucleotide loop replacements.

In the triple-helix series we studied the effects of loop structure on bimolecular complexes formed from pyrimidine-rich strands having the sequence 5'-dTTCTTTTC-loop-CTT-TTCTT and purine-rich target strands containing the binding site 5'-dAAGAAAAG or the reverse, 5'-dGAAAAGAA.[23] We also examined two closely related pyrimidine strands which

(22) Senior, M. M.; Jones, R. A.; Breslauer, K. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 6242−6246.

(23) Wang, S.; Booher, M. A.; Kool, E. T. *Biochemistry* **1994**, *33*, 4639−4644.



**Figure 3.** Sequences and structures of complexes in this study (curved lines indicate ethylene glycol-derived loop replacements of varied length): (A) DNA duplexes with natural (left) and nonnatural (right) loops; (B) DNA triplexes formed between hairpin-type pyrimidine ligands and minimal eight-base purine target strands; (C) similar DNA triplexes formed with longer targets which extend beyond the loop.

contain one extra natural nucleotide (underlined): 5'-dT-TCTTTTC-loop-$\underline{T}$CTTTTCTT and 5'-dTTCTTTTC$\underline{T}$-loop-CTTTTCTT. For all the triplex-forming pyrimidine sequences we substituted the loop with non-nucleotide $EG_n$ linkers and we compared the results to those with a natural loop of sequence -TTTTT-.

The four specific target sequences studied were 5'-dAA-GAAAAG, 5'-dGAAAAGAA, 5'-dAAGAAAAGACCCCC, and 5'-dCCCCCAGAAAAGAA. These are broken down into two sets of categories: the short and extended targets, and the 5' loop and 3' loop targets. The short targets represent minimal binding sites for the ligands and allow loops to bridge the pyrimidine domains without interference from the central purine strand. The extended targets were designed to test specifically the interactions of a given loop/linker with the central strand as it extends outward from the complex,[24] as might be expected to occur in many practical DNA-binding applications. The two 5' loop targets are bound by the ligands in an orientation which places the loop near the 5' end of the central target strand; conversely, 3' loop targets orient the loop near the 3' end of the target strand. Figure 3 shows the structures of the 12 types of complexes examined using the three types of pyrimidine-rich ligands and four types of purine-rich target sequences. In all, 48 such "hairpin triplex"-type complexes were examined using natural loops or non-nucleotide linkers.

In addition, to test the effects of the nonnatural linkers on triple helical complexes in a different sequence and structure, we synthesized four circular pyrimidine-rich oligodeoxynucleotides which contain two $EG_n$ linkers (Figure 7). These circular ligands were designed to bind a minimal purine target sequence

(24) Booher, M. A.; Wang, S.; Kool, E. T. *Biochemistry* **1994**, *33*, 4645−4651.

5′-dAAGAAAAGAAAG, as well as a longer 24-mer containing this target sequence in the center.[5]

**Structural Considerations.** Two important structural features to be considered in the design of loop replacements in DNA are the distances being spanned by the linking group and the geometries required to bridge from one side to the other. In the length analysis one would prefer a linker which is just long enough to span the required distance; a bridge which is too short will cause enthalpic strain in the complex, while one which is too long will be less than optimum from an entropic standpoint.

In the present linker structure, a bridging group carries its own 3′-phosphate group (as do the natural nucleotides in automated DNA synthesis) and then presents its own 5′-OH group to be linked to the 3′-phosphate on the next natural nucleotide. One can use models of duplex and triplex nucleic acids to examine the 5′-O to 5′-O distances involved in a bridge. Figure 2 shows these distances plotted for a B-DNA duplex, an A-RNA duplex, and an A′-type triple helix; the distances are shown from a specific starting point (marked *) on one chain to the various possible linked positions in 5′ and 3′ directions on the other strand. In the triplex case, a linker may in principle either bridge straight across a base triad or bridge in diagonal fashion between nucleotides not in the same triad.

For double-helical structures, results of this distance analysis[25,26] show that the bridging distance from one nucleotide to its pairing partner is about 15 Å and is almost the same for either DNA or RNA. Because of the right-handed turn of the helixes, the actual positions of closest approach to the * position on one strand are those two or three nucleotides in the 3′ direction on the opposite strand in DNA and one nucleotide in the 3′ direction in an RNA duplex. In the present case we are concerned initially with loops which span a blunt-end duplex with no overhanging nucleotides (i.e., from * to position 0).

For DNA triplexes, distance analysis shows that at the *-to-0 step (a blunt-end triplex), the distance is about 19 Å, or about 4 Å longer than the corresponding duplex distances. The actual point of closest approach to the * position is the one on the opposite strand four or five nucleotides in the 5′ direction.

The linker lengths in the EG$_n$ series range from 13 to 31 Å in fully extended conformation (Figure 1). Thus, in blunt-ended DNA or RNA duplex, one might expect a linker of length ≥ 15 Å (corresponding to EG$_4$ or EG$_5$) to be ideal and the EG$_6$–

EG$_8$ linkers to be perhaps overly long. In the blunt-end triplex case, the EG$_3$ and EG$_4$ linkers would be predicted to be too short, the EG$_5$ and EG$_6$ linkers to be nearly optimum, and the EG$_7$ and EG$_8$ linkers to be overly long. However, this analysis does not take into account the starting vectors of the groups being linked, which may not be aimed directly at each other; thus, longer lengths might be required to reorient the chain and aid in reversing the chain direction from one strand to the other. In addition, there may be steric barriers between the linked points that must be bypassed. Thus, it would not be surprising if longer lengths were required than predicted from simple distance analysis.

The analysis does indicate that one may expect optimum linker lengths to be similar for RNA and DNA duplexes; in addition, triple-helical structures are likely to require about 4−5 Å of additional length for an optimum bridge. For the EG$_n$ series, this 4−5 Å of length corresponds to one to two additional ethylene glycol units relative to the optimum duplex loop replacement. One additional structural aspect to take into consideration in triple-helical structures is the interaction of the loop/linker with the central purine target strand as it extends past the complex. We have previously shown that such interactions can be destabilizing,[24] and therefore, longer linkers might aid in passing around the steric bulk of this central strand. Alternatively, alteration of the linking geometry might aid in avoiding this problem (see below).

**Duplex DNA Loop Replacements.** Six EG$_n$-linked duplexes were studied at pH 7.0 in the presence of 100 mM NaCl and 10 mM MgCl$_2$. Stabilities of the duplexes were measured by thermal denaturation experiments monitored at 260 nm in a UV−vis spectrophotometer. Melting temperature ($T_m$) values were determined from the first derivative of the temperature versus absorbance data, and the free energies were determined by nonlinear least squares fitting of the data to a two-state model with linear sloping baselines. We studied the six duplexes over a concentration range from 0.2 to 10 μM, and the $T_m$ values were all independent of concentration over this range, confirming the hairpin-type unimolecular structure of the duplexes.

**Effect of Linker Length.** Figure 4A shows representative denaturation curves for these six complexes, which contain the EG$_3$ through EG$_8$ linkers. Examination of the curves shows that the duplexes with the four shortest linkers are similar in
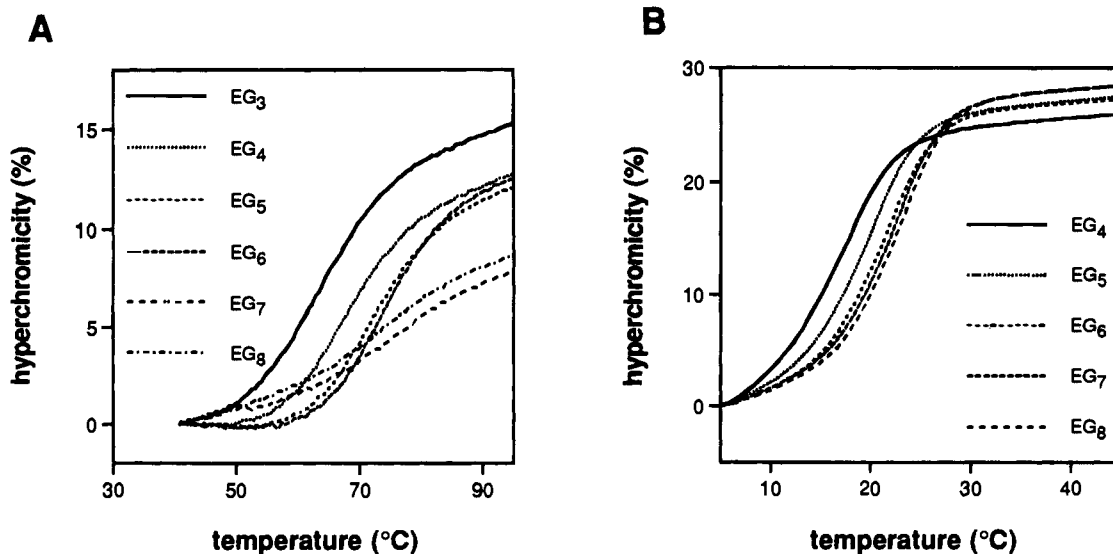


**Figure 4.** Representative thermal denaturation curves of hyperchromicity (percent increase in absorbance at 260 nm) vs temperature for two types of complexes in this study: (A) DNA duplexes with EG$_n$ linkers where $n = 3−8$ (see Figure 3A); (B) DNA triplexes formed between the probe strands 5′-dTTCTTTTC-EG$_n$-CTTTTCTT (where $n = 4−8$) and the 8-mer target 5′-dGAAAAGAA.

*Loop Replacements for DNA*

*J. Am. Chem. Soc., Vol. 117, No. 21, 1995* 5639

**Table 1.** Melting Temperatures ($T_m$) and Free Energies ($-\Delta G°_{37}$) for Duplex Hairpins Bridged by Oligo(ethylene glycol) (EG) Loops and by a Natural $T_4$ Loop at pH 7.0

| duplex hairpins | $T_m$ [a,b] (°C) | $-\Delta G°_{37}$ [a,b] (kcal) |
|---|---|---|
| 5' C G A A C G ⌐EG₃ 3' G C T T G C ⌐ | 62.5 | 3.4 |
| 5' C G A A C G ⌐EG₄ 3' G C T T G C ⌐ | 66.4 | 4.0 |
| 5' C G A A C G ⌐EG₅ 3' G C T T G C ⌐ | 71.1 | 4.8 |
| 5' C G A A C G ⌐EG₆ 3' G C T T G C ⌐ | 72.9 | 4.9 |
| 5' C G A A C G ⌐EG₇ 3' G C T T G C ⌐ | 74.9 | 3.9 |
| 5' C G A A C G ⌐EG₈ 3' G C T T G C ⌐ | 73.8 | 3.7 |
| 5' C G A A C G$^T$T 3' G C T T G C$_T$T | 67.8 | 4.6 |

[a] Conditions: 2.0 μM total strand concentration, 100 mM NaCl, 10 mM MgCl₂, 10 mM Na·PIPES buffer. [b] Error limits for individual measurements are estimated at ±0.5 °C in $T_m$ and ±10% in free energy. Data shown are the averages from three measurements.
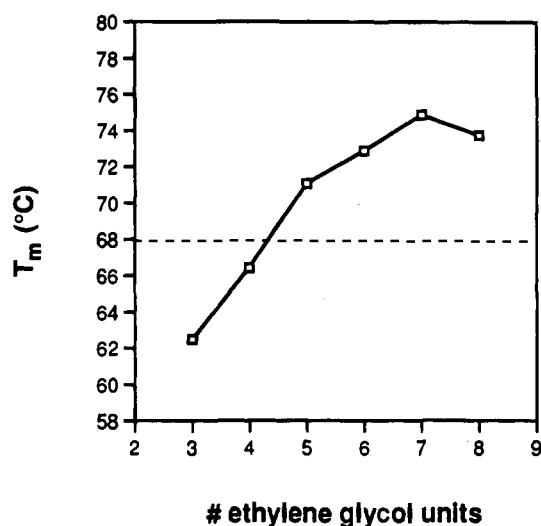


**Figure 5.** Thermal melting temperature ($T_m$) versus the number of ethylene glycol units, $n$, for DNA hairpins formed from the sequence 5'-dCGAACG-EG$_n$-CGTTCG. The horizontal line (at 67.8 °C) indicates $T_m$ for the same sequence but bridged with a natural $T_4$ loop. See Table 1 for conditions.

appearance but with the curves shifting toward higher temperatures as the linkers get longer. The duplexes with the EG₇ and EG₈ linkers show curves with relatively high $T_m$'s but with lower cooperativity and hyperchromicity. Table 1 displays the numerical data, some of which is plotted in Figure 5. The plot of $T_m$ values as a function of linker length shows that the $T_m$'s rise from a low of 62.5 °C with the shortest loop to 74.9 °C with the heptakis(ethylene glycol)-derived loop. This is the point of highest thermal stability; on lengthening the chain further to EG₈, the thermal stability then drops somewhat. The $T_m$ values in this case offer both the most accurate and precise measurement of stabilities; individual $T_m$ values for a given complex generally fall within 0.5−1.0 °C of each other, and our reported values are averaged from repeated experiments with standard deviations generally less than ±0.5 °C. Examination

of free energy data (calculated at 37 °C) shows that it follows a similar profile, but with a maximum at EG₅ and EG₆, or shorter than the $T_m$ data would indicate. The free energy numbers, however, are considerably less accurate because they are extrapolated from melting which occurs at much higher temperatures. In addition, the precision for $\Delta G°$ is less, at ±5−10%. It is therefore difficult to judge the stabilities on the basis of these free energy measurements. At higher temperatures (60−70 °C), the free energy data agree with the $T_m$ optimum at the EG₇-linked structure.

Thus, the heptakis(ethylene glycol) linker gives the highest thermal stability of the series for this DNA duplex. It would appear that, if higher thermal stability is desired, then the EG₇ linker may be superior to the commonly-used EG₆ linker for DNA duplexes. Although we have not studied RNA duplexes with this linker series, the geometric analysis leads to the prediction that heptakis(ethylene glycol) may be the optimum length for RNA as well. It is interesting that this optimum (a linker length of 27.4 Å) exceeds the minimal straight-line distance being bridged in the DNA duplex by 12 Å. This is apparently the extra length required to "turn the corner", that is, reverse the strand orientation by ∼180° and avoid any repulsive steric interactions as well.

**Comparison to a Natural Loop.** Since the EG$_n$ linkers are used as loop replacements, it is reasonable to compare the stabilities of those hairpins with one bridged by a natural $T_4$ loop. The $T_4$ loop in this context has been shown to be the most stable of the natural homo-tetraloops,[22] and four-nucleotide loops are thought to be the optimum length for natural loops in duplex DNA.[27]

The results show that the EG₇-linked hairpin is significantly more thermally stable than the hairpin with the $T_4$ loop. Under the same conditions, the latter has a $T_m$ of 67.8 °C, or 7 °C lower than with the optimum non-nucleotide linker. In fact, all the linkers from EG₅ through EG₈ are more stabilizing than the natural loop in this sequence context. It is not yet clear why this is the case; the high flexibility of the ethylene glycol chains would normally be expected to entropically disfavor stability relative to more rigid nucleotide loops, although it is possible that the EG chain retains some entropic freedom even in the complex.

**Comparison to Previous Results.** To our knowledge, only one previous study has reported the use of ethylene glycol oligomers in duplex DNA.[3] In that case, hexakis(ethylene glycol) was studied and the results were compared with the same sequence bridged by a $T_4$ loop. That study found a 3 °C increase in $T_m$ with the non-nucleotide loop. Our findings are therefore consistent with that result.

Two studies have focused on the effects of ethylene glycol and related linkers on the stability of RNA duplexes.[8,9] In those studies it was found that a hexakis(ethylene glycol) linker was more thermally stabilizing than a short tris(ethylene glycol) linker; comparison with nucleotide loops found the EG₆ linker to be more stabilizing than a hexanucleotide loop in one case, but less stabilizing than a tetranucleotide loop in another. In this case our results are not directly comparable, since our study does not involve RNA helices.

One previous study described the use of dodecyl chains, rather than ethylene glycols, as loop replacements in duplex DNA structures;[10] comparisons to a tetranucleotide loop showed that

the non-nucleotide linkers were substantially more stabilizing. The authors suggested, based on model building, that hexakis-(ethylene glycol) may be too long to be optimum as a loop replacement. Interestingly, we find that this is not the case; as noted above, an extra ~12 Å in length beyond the modeled bridging length is apparently needed for highest stability. It remains to be seen whether longer alkyl chains would give greater loop stability; however, hexadecyl chains have been reported to cause solubility problems when attached to DNA.[10]

**Loop Replacements in pyr·pur·pyr Triplexes.** There has been considerable and growing interest recently in the use of triplex-forming oligonucleotides to bind to single-stranded DNA[29] and RNA[28i,28m,29] by triple-helix formation. This new general strategy often involves the construction of an oligo-nucleotide which contains two binding domains—Watson—Crick and Hoogsteen—bridged by one or two loops. In a pyri-midine·purine·pyrimidine triplex, this involves linking two pyrimidine domains by a loop and using the resulting ligand to bind a purine-rich complementary site.

We have shown that, using natural nucleotides, the optimum loop length for such a triplex appears to be five nucleotides,[28c] or one longer than optimum for duplex. In a preliminary study,[5] we have also replaced the two pentanucleotide loops in a circular triplex-forming ligand with pentakis- and hexakis(ethylene glycol) linkers, and it was found that the EG$_6$-linked case gave higher stability in binding a minimal target sequence than the EG$_5$-linked case, and both of these were somewhat less stable than the complex with a cyclic ligand bridged by pentanucleotide loops. Additional experiments showed that the use of the non-nucleotide linkers conferred higher resistance to degradation by nuclease enzymes.

In the present study we wished to explore a wide range of linker lengths, to additionally explore any effects of 5' or 3' loop orientation, and finally to investigate any interactions between the loops and an extended target strand. We therefore studied binding to four target sequences, both short and extended, and using both 5' and 3' loop orientations, and tested the effect of linker length on these complexes. We initially studied the series using the sequence-symmetrical probe strand 5'-dTTCTTTTC-loop-CTTTTCTT. For comparison we also measured the stability of the triplex using the same sequence bridged by a T$_5$ loop.

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

(26) (a) Arnott, S.; Campbell Smith, P. J.; Chandrasekaran, R. In *CRC Handbook of Biochemistry and Molecular Biology: Nucleic Acids*; Fasman, G. D., Ed.; CRC Press: Cleveland, OH, 1976; Vol. II, p 411. (b) Arnott, S.; Selsing, E. *J. Mol. Biol.* **1974**, *88*, 509–521.

(27) Groebe, D. R.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1988**, *16*, 11725–11735.

(28) (a) Kool, E. T. *J. Am. Chem. Soc.* **1991**, *113*, 6265–6266. (b) Prakash, G.; Kool, E. T. *J. Chem. Soc., Chem. Commun.* **1991**, 1161–1162; **1994**, 646. (c) Prakash, G.; Kool, E. T. *J. Am. Chem. Soc.* **1992**, *114*, 3523–3528. (d) D'Souza, D. J.; Kool, E. T. *J. Biomol. Struct. Dyn.* **1992**, *10*, 141–152. (e) Perkins, T. A.; Goodman, J. L.; Kool, E. T. *J. Chem. Soc., Chem. Commun.* **1993**, 215–217. (f) Giovannangeli, C.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10013–10017. (f) Gryaznov, S. M.; Lloyd, D. H. *Nucleic Acids Res.* **1993**, *21*, 5909–5915. (g) Hudson, R. H. E.; Damha, M. J. *Nucleic Acids Symp. Ser.* **1993**, *29*, 97–99. (h) D'Souza, D. J.; Kool, E. T. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 965–970. (i) Wang, S.; Kool, E. T. *J. Am. Chem. Soc.* **1994**, *116*, 8857–8858. (j) Kandimal, E. R.; Agrawal, S. *Gene* **1994**, *149*, 115–121. (k) Noll, D. M.; O'Rear, J. L.; Cushman, C. D.; Miller, P. S. *Nucleosides Nucleotides* **1994**, *13*, 997–1005. (l) Trapane, T. L.; Christopherson, M. S.; Roby, C. D.; Ts'o, P. O. P.; Wang, D. *J. Am. Chem. Soc.* **1994**, *116*, 8412–8413. (m) Reynolds, M. A.; Arnold, L. J.; Almazan, M. T.; Beck, T. A.; Hogrefe, R. I.; Metzler, M. D.; Stoughto, S. R.; Tseng, B. Y.; Trapane, T. L.; Ts'o, P. O. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2433–2437. (n) Bandaru, R.; Hashimoto, H.; Switzer, C. *J. Org. Chem.* **1995**, *60*, 786–788.

(29) Wang, S.; Kool, E. T. *Nucleic Acids Res.* **1994**, *22*, 2326–2333.

**Table 2.** Melting Temperatures ($T_m$) and Free Energies ($-\Delta G°_{37}$) for Triplex Hairpins Bridged by Oligo(ethylene glycol) (EG) Loops at pH 7.0

| 5' loops | $T_m$[a,b] (°C) | $-\Delta G°_{37}$[a,b] (kcal) | 3' loops | $T_m$[a,b] (°C) | $-\Delta G°_{37}$[a,b] (kcal) |
|---|---|---|---|---|---|
| # ethylene glycol units (CTTTTCTT 5' / GAAAAGAA 3' / CTTTTCTT 3') | | | # ethylene glycol units (5'TTCTTTTC / 5'AAGAAAAG / 3'TTCTTTTC) | | |
| 4 | 17.7 | 3.3 | 4 | 17.8 | 5.0 |
| 5 | 20.3 | 3.6 | 5 | 19.5 | 5.0 |
| 6 | 22.1 | 4.2 | 6 | 21.5 | 5.0 |
| 7 | 23.0 | 4.4 | 7 | 22.4 | 5.2 |
| 8 | 23.5 | 4.5 | 8 | 23.3 | 5.1 |
| (CTTTTCTT 5' / GAAAAGAA 3' / TCTTTTCTT 3') | | | (5'TTCTTTTC– / 5'AAGAAAAG / 3'TTCTTTTCT) | | |
| 4 | 23.6 | 4.4 | 4 | 22.5 | 5.1 |
| 5 | 24.4 | 5.0 | 5 | 23.0 | 5.4 |
| 6 | 24.2 | 4.7 | 6 | 23.2 | 5.4 |
| 7 | 23.8 | 4.7 | 7 | 23.2 | 5.0 |
| 8 | 23.4 | 4.5 | 8 | 22.8 | 4.8 |
| (TCTTTTCTT 5' / GAAAAGAA 3' / CTTTTCTT 3') | | | (5'TTCTTTTCT / 5'AAGAAAAG / 3'TTCTTTTC–) | | |
| 6 | 23.2 | 4.4 | 6 | 22.0 | 5.3 |

[a] Conditions: 2.0 μM total strand concentration, 100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na·PIPES buffer. [b] Error limits for individual measurements are estimated at ±0.5 °C in $T_m$ and ±10% in free energy. Data shown are the averages from two to three separate measurements.

**Effect of Linker Length.** The effect of varied linker length was first examined with minimal eight-base targets in the 5' and 3' loop orientations. The thermal denaturation data for the bimolecular complexes are listed in Table 2 and are plotted graphically in Figure 6. The experiments were carried out at pH 7.0, with 100 mM NaCl and 10 mM MgCl$_2$. Figure 4B also shows representative melting plots for one set from the EG$_n$ series in the 5' loop orientation with the minimal target. The curves indicate that the complexes are well-behaved and give apparently two-state, all-or-none denaturation. All of the complexes have similar hyperchromicity and curve shape.

Results with the short targets show a significant dependence on linker length, with $T_m$ values ranging from 17.7 to 23.5 °C in the 5' loop series. The results for the 3' loop series are experimentally indistinguishable from those with the 5' loop orientation. In both cases, the $T_m$ increases monotonically with increasing linker length and is highest at the EG$_8$-linked point. Examination of the overall trends (Figure 6A) shows that, although the series has not passed beyond an optimum, it appears to be approaching it, with only small improvements evident from EG$_7$ to EG$_8$. The $T_m$ differences are statistically significant even in this smaller last step. The free energy values for the EG$_6$ through EG$_8$ complexes fall within experimental error of each other and so are not helpful in distinguishing differences.

Thus, it would appear that in general, with blunt-ended triplexes, the linker which gives highest thermal stability in this series is the octakis(ethylene glycol)-derived linker. Interest-ingly, we cannot rule out the possibility that a nonakis(ethylene glycol) linker may be slightly more stabilizing, although the trend approaching this length would indicate that this longer loop is not likely to give large improvements. The data also
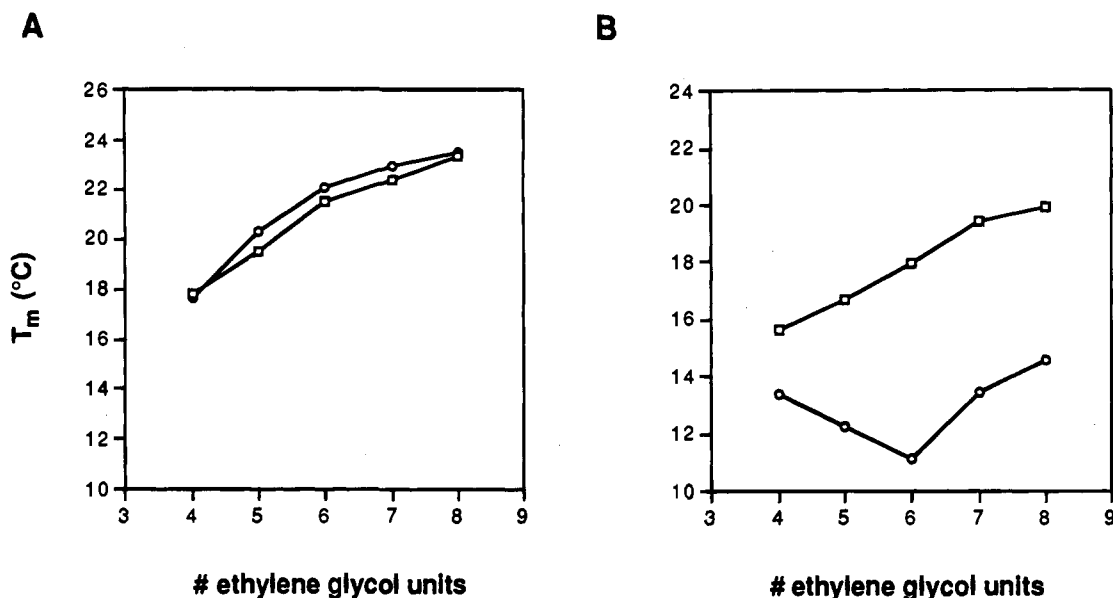
**Figure 6.** $T_m$ versus number of ethylene glycol (EG) units for triplexes formed between the probe strands 5'-dTTCTTTTC-EG$_n$-CTTTTCTT (where $n = 4-8$) and 8-mer target sequences at pH 7.0: (A) short target sequences, 5'-dAAGAAAAG (□) and dGAAAAGAA (O); (B) extended target sequences, 5'-dAAGAAAAGACCCCC (□) and 5'-dCCCCCAGAAAAGAA (O).

show that there is no significant difference in loop stabilities for the 5' and 3' loop orientations.

As seen for the duplex series above, the optimum linker is considerably longer than straight-line distance analysis would indicate. The EG$_8$ linker has a length of 30.9 Å, or about 12 Å longer than the distance being bridged. It is interesting to note that this is precisely the extra length which is optimum for the duplex linkers and which was ascribed to the length need to reverse the strand orientation. It does seem, as predicted by the distance analysis, that the optimum for triplex linkers is indeed one to two ethylene glycol units longer than that for duplex.

**Effect of the Target Strand Length.** In most practical applications, such triplex-forming ligands would bind to purine-rich target sites within a much longer strand of DNA. To test the effect of interactions between loops or loop replacements and the target strand as it passes beyond the loop, we studied two longer target sequences which contain six extra nucleotides, an A followed by five C residues. In one case this extension is placed at the 5' end of the target site, and in the other, the 3' end. This allows for the testing of loop interactions in either loop orientation.[23] The same EG$_n$-looped ligand series was hybridized to these two longer targets and evaluated again by thermal denaturation.

The data from these experiments are listed in Table 3 and are plotted graphically in Figure 6B. Results show that the stabilities depend strongly on which loop orientation is being studied: the range of $T_m$ values in the 5' orientation is 11.1−14.5 °C, whereas in the 3' orientation the range is higher overall, 15.6−20.0 °C. Thus, the least thermostable 3' loop is more stable than the strongest 5'-type loop; in addition, both cases show lower thermal stability than do complexes with the minimal eight-base binding sites (compare parts A and B of Figure 6). From these two observations, we can conclude that there are destabilizing interactions between the loop and the central strand and that the magnitude of the interaction depends on the strand orientation relative to the loop.

Also interesting are the overall trends as a function of length for the two orientations (Figure 6B): for the 3' loop series, the $T_m$ data increases monotonically with length to the EG$_8$-linked compound, as was seen with the short target strands. However,

**Table 3.** Melting Temperatures ($T_m$) and Free Energies ($-\Delta G°_{37}$) for Triplex Hairpins Bridged by Oligo(ethylene glycol) (EG) Loops at pH 7.0

| 5' loops | $T_m^{a,b}$ (°C) | $-\Delta G°_{37}^{a,b}$ (kcal) | 3' loops | $T_m^{a,b}$ (°C) | $-\Delta G°_{37}^{a,b}$ (kcal) |
|---|---|---|---|---|---|
| # ethylene glycol units 5'CCCCC⟨CTTTTCTT5' AGAAAAGAA3' CTTTTCTT3'⟩ | | | # ethylene glycol units 5'TTCTTTTC AAGAAAAGA CCCCC3' TTCTTTTC | | |
| 4 | 13.4 | 4.9 | 4 | 15.6 | 4.6 |
| 5 | 12.3 | 5.1 | 5 | 16.7 | 4.5 |
| 6 | 11.1 | 5.0 | 6 | 18.0 | 4.6 |
| 7 | 13.5 | 4.2 | 7 | 19.5 | 4.6 |
| 8 | 14.6 | 3.9 | 8 | 20.0 | 4.6 |
| 5'CCCC⟨CTTTTCTT5' CAGAAAAGAA3' TCTTTTCTT3'⟩ | | | 5'TTCTTTTC AAGAAAAGAC CCCC3' TTCTTTTCT | | |
| 4 | 15.1 | 5.5 | 4 | 23.4 | 5.5 |
| 5 | 15.0 | 5.2 | 5 | 24.0 | 5.4 |
| 6 | 16.4 | 4.8 | 6 | 24.3 | 5.6 |
| 7 | 17.8 | 4.5 | 7 | 24.4 | 5.4 |
| 8 | 18.5 | 4.1 | 8 | 23.8 | 5.8 |
| 5'CCCC⟨TCTTTTCTT5' CAGAAAAGAA3' CTTTTCTT3'⟩ | | | 5'TTCTTTTCT AAGAAAAGAC CCCC3' TTCTTTTC | | |
| 6 | 18.1 | 4.5 | 6 | 22.5 | 5.3 |

[a] Conditions: 2.0 μM total strand concentration, 100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na·PIPES buffer. [b] Error limits for individual measurements are estimated at ±0.5 °C in $T_m$ and ±10% in free energy. Data shown are the averages from 2−3 measurements.

with the 5'-type loops, the $T_m$ actually drops initially with the longer target, reaching a low at the hexakis(ethylene glycol) point, and then it increases again, to the highest point at the EG$_8$ case. Again, these effects are statistically significant, since the standard deviations for the $T_m$ values are generally less than ±0.5 °C, much smaller than the magnitude of the actual variation seen.

These effects are likely to be caused by steric interactions between the linker/loop and the target strand. Models show that, in a DNA triple helix, a straight line drawn from one pyrimidine backbone to the other one (at the same base step) must pass through the steric space of the central purine strand and that the specific group involved seems to be the next base beyond the target site in the central strand. Such a steric interference would explain the lower $T_m$ with the longer target strands for both loop orientations; in addition, since the two loop orientations cause differences in the local structure, this might explain why the $T_m$ behavior as a function of length is different in the two cases. Examination of the models does not suggest an obvious explanation for the bimodal behavior seen for the 5' loop series, and so it is not yet clear what causes this effect.

In any case, we find that the optimum loop for bridging triplexes in either orientation is the octakis(ethylene glycol)-derived loop. We cannot rule out the possibility in this case that longer loops might give yet higher stability; indeed, this might be expected since the loop may have to travel a greater distance to avoid destabilizing steric interactions with the extended central strand. In addition, we find that the commonly-used linker derived from hexakis(ethylene glycol) is a relatively poor choice, especially if the 5' loop orientation is desired. Finally, we conclude that, with extended target strands, all the loop replacements in this series are significantly destabilized by interaction with the central strand.

**Comparison to Previous Results in Triplex DNA.** Aside from our preliminary report,[5] few reports exist in the literature on the use of non-nucleotide linkers in triplex DNA. Two groups have used hexakis(ethylene glycol) linkers bridging pyrimidine strands in triple helixes,[4,7] but neither of these studies evaluated other linker lengths or compared the stabilities to those with nucleotide loops. A study by Letsinger et al. reported the use of a bis-amide linker to bridge triplex structure,[6] but the stability relative to that with a nucleotide loop was not evaluated.

**Comparison to Naturally Bridged Loops.** The comparison of our loop replacement series with an optimized $T_5$ nucleotide loop[23] in the same triplexes reveals that, although none of the above synthetic replacements leads to stabilization relative to the natural-looped case, the best replacements approach this value. When the pyrimidine probe strand dTTCTTTTC-$T_5$-CTTTTCTT is hybridized to the two short targets under identical conditions, we find $T_m$ values of 25.2 and 24.4 °C for the 5' and 3' loops, respectively, and corresponding free energies (37 °C) of −5.3 and −5.7 kcal/mol. Thus, this natural loop gives only slightly higher thermal stability (by 1.7 and 1.1 °C for 5' and 3' loops) than is shown by the most stable $EG_8$-linked compounds.

In the case where the target strand is extended, the difference between natural and nonnatural loops becomes more pronounced. The $T_m$ values for binding of the $T_5$-looped sequence to the longer target strands is actually higher than with the short targets. The $T_m$ values in the extended case are 29.5 °C in the 5' loop orientation and 26.3 °C in the 3' loop orientation. These values are 4.3 and 1.9 °C higher than those for the same ligand with the short targets; thus, the interaction between the natural loop and extended strand bases is a stabilizing one. As seen above, however, the corresponding interaction with the non-nucleotide linkers is significantly destabilizing. The $T_m$ difference between the $T_5$-looped ligand and the optimized $EG_8$-linked oligomer with these longer target strands is 15.2 °C in the 5' loop orientation and 6.3 °C in the 3' orientation.

Possible explanations for this difference lie in structural differences between these natural and nonnatural linkers. First,

the natural pentanucleotide loop may be longer; this structure could span as long a distance as 33−34 Å in an extended conformation, while the octakis(ethylene glycol) linker spans a shorter ~31 Å. Second, the nucleotide backbone is more rigid than the ethylene glycol chain, which favors the former entropically in binding. Third, the nucleotide loop is likely to benefit significantly from favorable stacking interactions involving the loop bases. Finally, in the natural loop, there are probably stabilizing base pairing interactions between the first/last loop bases and the base adjacent to the minimal binding site.[24]

**Circular DNAs as a Second Test of Linker Length Effects in Triplexes.** To test the effects of structural variation in the $EG_n$ series in a second triplex system, we incorporated these linkers as loop replacements in circular triplex-forming oligonucleotides. These ligands consist of two pyrimidine-rich binding domains linked at both ends by two separate loops or loop replacements, and they are designed to bind purine-rich targets by sandwiching the target between the two pyrimidine domains. A preliminary study with the $EG_5$- and $EG_6$-linked circles in binding short 12-mer targets showed the latter to bind with higher affinity.[5] In the present study, we wished to study additionally the $EG_7$- and $EG_8$-linked cases and to test the ability to bind longer targets as well. The data for these experiments, carried out at pH 7.0 with 100 mM NaCl and 10 mM $MgCl_2$, are presented in Table 4 and graphically in Figure 7.

Comparison of linkers of varied length in this new structure shows length effects which parallel those seen for the hairpin-type triplex studies. Results in the binding of the minimal 12-base targets show that binding affinity increases with increasing linker length up to the maximum length of the $EG_8$-linked compound. $T_m$ values vary strongly from 37.3 °C for the $EG_5$-doubly-linked case to 53.7 °C for the $EG_8$-linked case. The differences in free energy are also substantial and in this case are statistically significant. We find that free energies vary correspondingly from −8.8 to −17.0 kcal/mol (37 °C) from the shortest to longest chain. This is a similar result to the observations made in the singly-linked hairpin-type triplex-forming ligands, in which there was a preference for the longest loop replacement. Interestingly, a version of this cyclic ligand bridged by pentanucleotide loops binds with somewhat lower affinity, with a $T_m$ of 52.2 °C and a free energy (37 °C) of −14.1 kcal/mol. Thus, in this case, the best loop replacements are superior to an optimized five-nucleotide loop, and this is the first example in triplex DNA wherein loop replacements have outperformed natural loops.

When the same ligands are hybridized to the same binding site embedded in a longer 24-mer sequence, the same length-dependent trend is seen, with a preference for the longest $EG_8$ linker. Again, as seen for the previous triplex-forming EG-linked ligands, the overall stabilities are lower than with the short target sequence. Thus, this destabilization by unfavorable loop−central strand interaction is consistently found in all cases with non-nucleotide linkers.

**Extension of One Binding Domain Leads to Added Stabilization.** The above results indicate that there is a general destabilizing interaction between the non-nucleotide loop replacements and the central strand as it passes beyond the loop. Examination of models of triplex structures suggests a strategy to alleviate such an unfavorable interaction: extension of one of the two binding domains at the end of the triplex near the loop, so that one strand is lengthened in the 5' direction. In such a case, the linker bridges two nucleotides not in the same base step, but instead in adjacent (or further) steps. The distance analysis (Figure 2) shows that such a 5' extension actually brings

**Table 4.** Melting Temperatures $(T_m)$ and Free Energies $(-\Delta G°_{37})$ for Triple Helical Complexes Bridged by Oligo(ethylene glycol) (EG) Loops or by a Pentanucleotide Loop at pH 7.0

| loop structure | $T_m$[a,b] (°C) | $-\Delta G°_{37}$[a,b] (kcal) |
|---|---|---|
| $\binom{\text{TTCTTTTCTTTC}}{\substack{\text{AAGAAAAGAAAG}\\\text{TTCTTTTCTTTC}}}$ | | |
| EG$_5$ | 37.3 | 8.8 |
| EG$_6$ | 51.0 | 15.6 |
| EG$_7$ | 52.9 | 16.6 |
| EG$_8$ | 53.7 | 17.0 |
| -CACAC- | 52.2 | 14.1 |
| ₅ GGACTCTATCA$\binom{\text{TTCTTTTCTTTC}}{\substack{\text{GAAGAAAAGAAAGG}\\\text{TTCTTTTCTTTC}}}$GACTCTATCAG ₃ | | |
| EG$_5$ | 36.8 | 8.8 |
| EG$_6$ | 43.5 | 11.7 |
| EG$_7$ | 46.5 | 13.0 |
| EG$_8$ | 47.6 | 13.9 |
| -CACAC- | 52.5 | 14.6 |

[a] Conditions: 2.0 μM total strand concentration, 100 mM NaCl, 10 mM MgCl$_2$, 10 mM Na·PIPES buffer. [b] Error limits for individual measurements are estimated at ±0.5 °C in $T_m$ and ±10% in free energy.

**A**

$$\binom{\text{TTCTTTTCTTTC}}{\substack{\text{AAGAAAAGAAAG}\\\text{TTCTTTTCTTTC}}}$$

₅'GG ACTCTATCA$\binom{\text{TTCTTTTCTTTC}}{\substack{\text{GAAGAAAAGAAAGG}\\\text{TTCTTTTCTTTC}}}$GACTCTATCAG ₃'
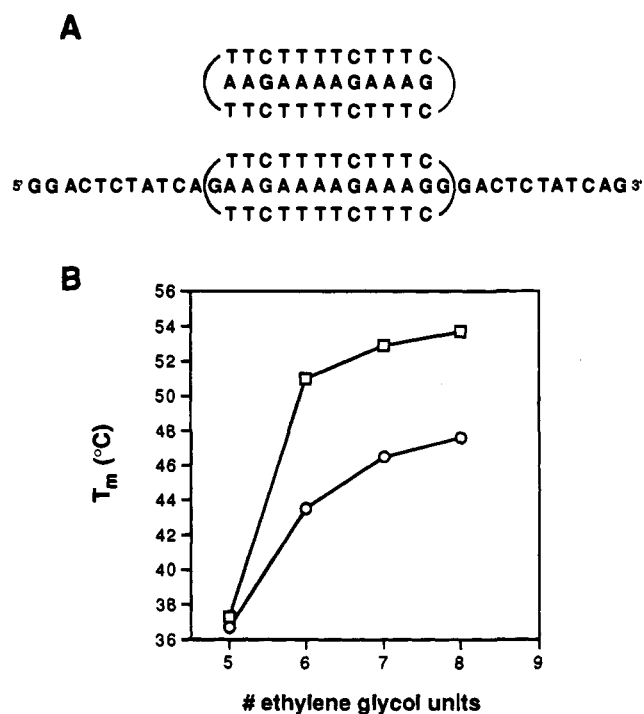
**B**



**Figure 7.** Effect of linker length (EG$_n$) on the binding of purine-rich targets by loop-modified circular DNAs at pH 7.0: (A) sequences of two complexes formed with short (top) and extended (bottom) target strands; (B) $T_m$ versus number of EG units $(n)$ for the two complexes (short targets, □; long targets, ○). See Table 4 for conditions.

the two groups being linked closer together than when they are in the same base step, from a *-to-position-0 link (18.6 Å) to a *-to-position-−1 link (17.1 Å). In addition, the models show that, because of the right-handed turn of the helix, this extension

moves a straight-line linking group further away from the interfering central strand.

To test this possibility we constructed several additional hairpin-type ligands targeted to the sequences dAAGAAAAG or dGAAAAGAA. These differ from the previous ligands by having one extra nucleotide (a T) as an extension adjacent to the loop. An EG$_n$-linked series, where $n$ = 4−8, was constructed with a 5′ extension; when binding in 5′ loop orientation, this makes the extension part of the Watson−Crick binding domain, and in the 3′ loop orientation, the extension is part of the Hoogsteen domain. In addition, one ligand with a 3′ extension was constructed having an EG$_6$-derived linker; this reverses the position of the extension with respect to loop orientation. These new ligands were hybridized to the same short and long target sequences as was previously done. The results are tabulated in Tables 2 and 3 and are presented graphically in Figure 8, where they are compared with results with nonextended ligands.

Results with the 5′-extended and 3′-extended ligands with the minimal 8-mer targets show that lengthening one strand by one base on either side of the loop increases the stability of all complexes up to the longest, the EG$_8$-linked case (see Figure 8A,B). With this last exception, the $T_m$ values are higher with the two extended ligand types than they are with the blunt, nonextended ligand. For these short targets, the behavior is almost identical regardless of 5′ or 3′ loop orientation (compare parts A and B of Figure 8). With the 5′-extended series, the dependence on linker length is small relative to the nonextended series, and the stability peaks broadly at the EG$_{5-7}$ cases and then begins to drop. As pointed out previously, the nonextended ligands markedly increase their binding affinity with increasing loop length all the way to the EG$_8$-linked case. One 3′-extended ligand was tested for comparison, and results show (Figure 8A,B) that this extension is for both loop orientations less stabilizing than the 5′ extension.

Results are considerably different when the extended ligands are hybridized to longer strands which interact with the loops. Figure 8C,D shows these results, again compared to the binding of the nonextended (blunt-ended) ligands. As discussed above, for the 5′ loop orientation, the blunt-ended ligands bind very poorly. The 5′-extended ligands, by comparison, all bind more tightly, with $T_m$ values 2−5 °C higher depending on length. The strongest binding case is that with the EG$_8$-derived linker. Interestingly, the 3′-extended ligand binds better than the 5′-extended one at the EG$_6$ case (Figure 8C). This is the only example where the 3′ extension has better success than the 5′-extended version, although it remains to be seen what the effect of longer linkers would be in this case.

Finally, for the 3′ loop orientation, we find that the 5′-extended ligands are quite successful at stabilizing the complexes relative to the unextended ligands. Results show that they bind the longer target even more strongly than they do the short target (by a margin of about 1 °C in $T_m$), indicating that there are new favorable interactions with the central strand. The $T_m$ versus length profile is very flat, with a broad maximum from EG$_5$ to EG$_7$. Comparison to the unextended ligands shows that the extension adds considerable affinity, with $T_m$ values increasing by 4−8 °C. In this case we see that the 3′-extended ligand binds less strongly than the 5′-extended ones do.

In general, then, these single-nucleotide extensions allow for considerable alleviation of the strained interactions between the non-nucleotide linkers and the longer target strands. The general finding of increased stabilities by extension of one strand supports the argument that the extended nucleotide undergoes pairing interactions with the purine strand, rather than simply
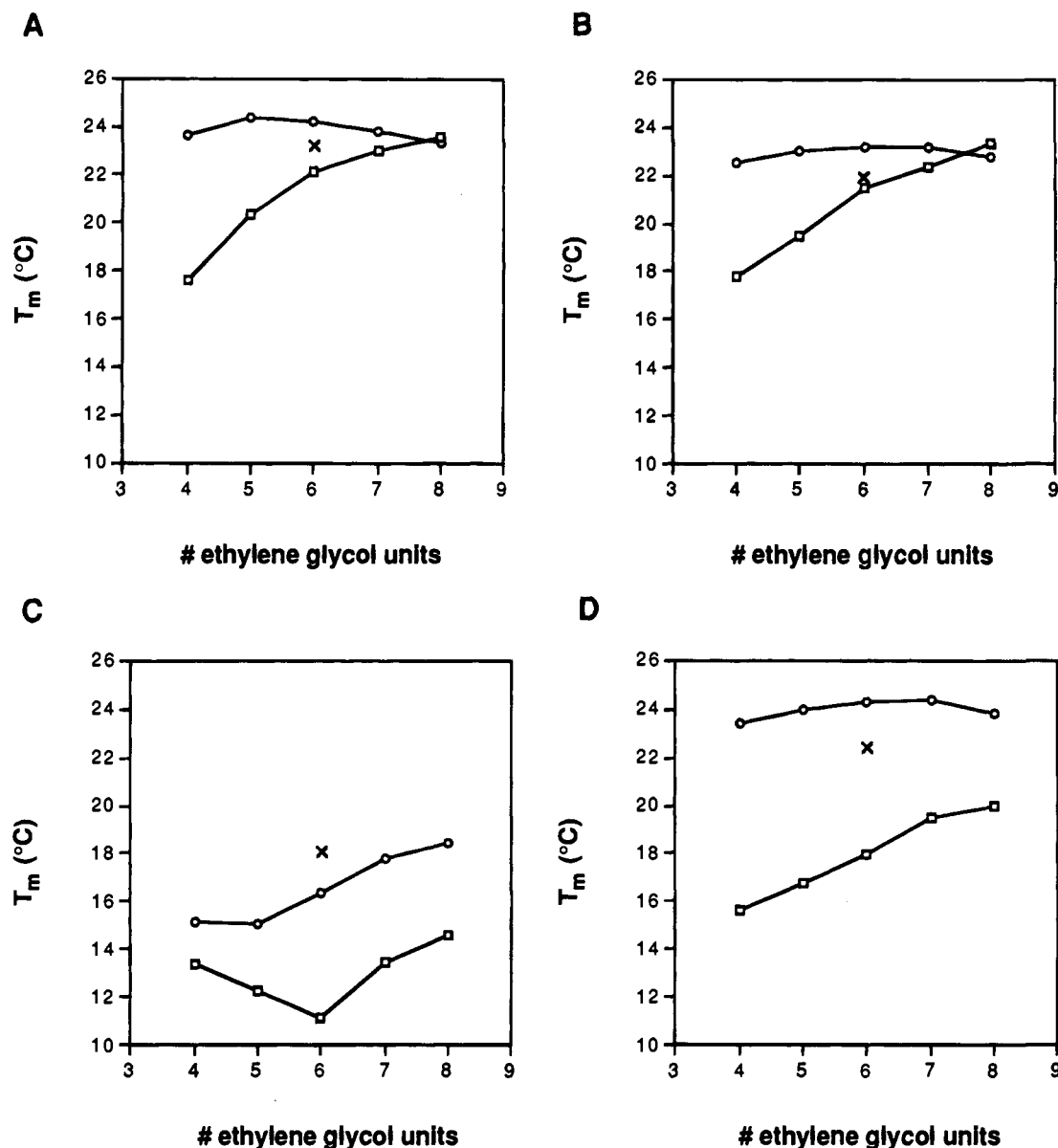
**Figure 8.** Melting temperature versus number of ethylene glycol (EG) units for complexes formed between triplex-forming probe strands and short or extended target sequences at pH 7.0: (A) the binding of 5'-dGAAAAGAA by probe strands 5'-dTTCTTTTC-EG$_n$-CTTTTCTT (□), 5'-dTTCTTTTC-EG$_n$-T-CTTTTCTT (○), and 5'-dTTCTTTTC-T-EG$_n$-CTTTTCTT (×); (B) the binding of 5'-dAAGAAAAG by the same three probe strands; (C) the binding of 5'-dCCCCCAGAAAAGAA by probe strands 5'-dTTCTTTTC-EG$_n$-CTTTTCTT (□), 5'-dTTCTTTTC-EG$_n$-T-CTTTTCTT (○), and 5'-dTTCTTTTC-T-EG$_n$-CTTTTCTT (×); (D) the binding of 5'-dAAGAAAAGACCCCC by the same three probe strands.

serving as part of the linker. Although the affinities do not quite rise to the level of optimized pentanucleotide loops, this strategy represents a considerable improvement over the use of blunt-end ligands bridged by non-nucleotide linkers.

**General Recommendations for Loop Design.** With short target sequences which do not extend beyond the triplex, the findings with the ligands which are extended by one nucleotide in the 5' direction near the loop can be summarized as follows: the extension gives some added benefit with suboptimum linkers, but simple use of the longest (octakis(ethylene glycol)) linker makes this unnecessary. Loop orientation makes little or no difference. In general, we can therefore recommend the octakis(ethylene glycol) linker as the best available non-nucleotide loop with short targets. In some cases this non-nucleotide loop gives complexes more stable than those with nucleotide loops.

If the target site is found within a longer strand of DNA, the findings can be summarized as follows: if the loop is to be oriented to the 3' end of such a triple-helical complex, then the

best loop design (of those studied here) will include an added base at the 5' end of the Watson–Crick domain next to the loop and will be linked by EG$_5$, EG$_6$, or EG$_7$ loop replacements. The resulting affinity can approach that of an optimized pentanucleotide loop. However, if a 5' loop orientation is necessary and a non-nucleotide loop replacement is to be used, then considerably lower binding affinity will result. The best such design in this case will be a ligand which is extended in the 5' direction in the Hoogsteen domain near the loop, with a linker derived from octakis(ethylene glycol). However, an optimized pentanucleotide loop[24] will give considerably higher binding affinity.

Hairpin-type triplex-forming ligands which are closed on one end by a loop can be designed to bind a target with the loop in either 5' or 3' orientation.[23] Our results show that, if one is using non-nucleotide linkers, there is a clear binding advantage to ligands with the 3' loop orientation. The results seem to indicate that this may be due to a greater steric interference at the 5' end of the triplex, although more detailed structural studies

*Loop Replacements for DNA*

*J. Am. Chem. Soc., Vol. 117, No. 21, 1995* 5645

will be necessary to confirm this. One possible way to alleviate this problem (aside from the use of natural nucleotide loops) may be the use of linkers which are yet longer than those in this study.

It should be noted that, considerations of binding affinity aside, non-nucleotide loop replacements have other potentially valuable properties. They shorten the synthesis of a ligand quite substantially; for example, replacement in a duplex gives a sequence three steps shorter than would be the case for a commonly-used tetranucleotide loop. In triplexes, where pentanucleotide loops are commonly used, such a replacement shortens the synthesis by four steps each time it occurs in a strand. In addition, since ethylene glycol oligomers are relatively inexpensive, they may cost less than nucleotides even on a one-for-one basis. Finally, such non-nucleotide loops can increase the resistance of an oligonucleotide to degradation in biological media.[5]

## Conclusions

It is clear that non-nucleotide linkers can be efficient replacements for natural nucleic acid loops, but it is equally clear that the stability of the folded helical structures depends strongly on linker length and structure. The following summarizes the conclusions from this study:

(1) The optimum linker for duplex is heptakis(ethylene glycol); this gives complexes considerably more stability than those bridged by natural tetranucleotide loops.

(2) Of the structures studied here, the optimum linker for a blunt-ended triplex is derived from octakis(ethylene glycol); this gives complexes which are in some cases more stable than structures with natural pentanucleotide loops.

(3) In both double- and triple-helical cases the optimum linker is ~12 Å longer than simple straight-line distances derived from models would indicate.

(4) For triplexes where the target is embedded in a longer strand, the optimum linker length is again octakis(ethylene glycol), but the stability depends greatly on 5' or 3' loop orientation, with the latter being much more stable; both cases are less stable, probably because of steric interference, than the case with a natural pentanucleotide loop.

(5) It is possible to bypass most of this destabilization due to central strand steric effects, by simply extending one end of one of the binding domains near the loop in the 5' direction.

We anticipate that these findings will be useful in the design of modified DNA structures, with significant practical applications possible in biological systems.

## Experimental Section

**General Remarks.** $^1$H and $^{13}$C NMR spectra were obtained with a GE 300 MHz instrument, and chemical shifts are reported in ppm on the $\delta$ scale with the solvent as an internal reference. For $^{31}$P NMR spectra an external reference of 85% phosphoric acid was used. Column chromatography was performed with EM Science silica gel 60 230−400 mesh. Thin layer chromatography (TLC) was performed on silica gel 60 (Merck) F-254 precoated 0.25 mm plates. Mass spectral analyses were performed by University of California at Riverside Mass Spectrometry Facility.

Dichloromethane and triethylamine were dried by distillation from calcium hydride, and pyridine was distilled from barium oxide. Pentakis(ethylene glycol) was obtained from Lancaster Synthesis. All other solvents and chemicals were purchased from Aldrich, Sigma, J. T. Baker, or Fisher unless otherwise noted.

The phosphoramidite derivatives of (dimethoxytrityl)tetrakis-, (dimethoxytrityl)pentakis-, and (dimethoxytrityl)hexakis(ethylene glycol) were prepared as described previously.[5] (Dimethoxytrityl)tris-(ethylene glycol) 2-(cyanoethyl) *N,N*-diisopropylphosphoramidite was obtained from Glen Research.

**(4,4′-Dimethoxytrityl)tetrakis(ethylene glycol).** Tetrakis(ethylene glycol) (40 mmol) was coevaporated twice with anhydrous pyridine (10 mL), and the residue was dissolved in dry CH$_2$Cl$_2$ (10 mL). To this were added dry Et$_3$N (14.4 mmol) and 4-(*N,N*-dimethylamino)-pyridine (DMAP) (0.40 mmol). Next, 2.71 g of 4,4′-dimethoxytrityl chloride (DMTCl) (8.9 mmol) was added in small portions under argon. The mixture was stirred at room temperature for 1−3 h and monitored by TLC (EtOAc). The solution was diluted with CH$_2$Cl$_2$ (100 mL) and washed twice with 5% NaHCO$_3$ (50 mL) and once with H$_2$O (50 mL). The organic phase was dried over anhydrous sodium sulfate, filtered, and concentrated to an oily residue under reduced pressure. This residue was purified by chromatography on silica gel. The column was equilibrated with EtOAc containing 1% Et$_3$N and the product eluted with EtOAc. The product was obtained as a pale yellow oil in 63% yield (2.49 g, 5.0 mmol): $R_f$ 0.35 (EtOAc); $^1$H NMR (CDCl$_3$, ppm) 2.55 (br s, 1H, O*H*), 3.25 (t, 2H, $J$ = 5 Hz, DMTOC*H$_2$*), 3.60−3.63 (m, 2H, C*H$_2$*CH$_2$OH), 3.67−3.78 (m, 12H, OC*H$_2$*CH$_2$O), 3.80 (s, 6H, OC*H$_3$*), 6.82−6.85 (m, 4H, arom H), 7.21−7.49 (m, 9H, arom H); $^{13}$C NMR (CDCl$_3$, ppm) 55.8, 62.4, 63.8, 71.1, 71.4, 71.7, 73.1, 86.6, 113.7, 127.3, 128.4, 128.7, 128.9, 130.7, 131.0, 131.1, 137.0, 145.7, 159.0.

**(4,4′-Dimethoxytrityl)tetrakis(ethylene glycol) *p*-Toluenesulfonate.** The monotritylated tetrakis(ethylene glycol) (5.0 mmol) was dissolved in CHCl$_3$ (5 mL) which had been filtered through a plug of alumina. To this was added dry pyridine (10.0 mmol) and DMAP (0.25 mmol). The mixture was cooled in an ice bath (0 °C) and *p*-toluenesulfonyl chloride (7.52 mmol) was added in small portions under argon with constant stirring over a period of 1 h. After 2.5 h, the reaction had reached completion (monitored by TLC). The solution was diluted with CH$_2$Cl$_2$ (50 mL) and washed twice with 5% NaHCO$_3$ (40 mL) and once with H$_2$O (40 mL). The organic phase was dried over anhydrous sodium sulfate, filtered, and concentrated to an oily residue under reduced pressure. This residue was purified by flash chromatography on silica gel. The column was equilibrated with EtOAc:hexanes (1:1) containing 1% Et$_3$N and the product eluted with EtOAc:hexanes (1:1). The product was obtained as a yellow oil in 85% yield (2.77 g, 4.3 mmol): $R_f$ 0.31 (EtOAc:hexanes, 1:1); $^1$H NMR (CDCl$_3$, ppm) 2.45 (s, 3H, ArC*H$_3$*), 3.24 (t, 2H, $J$ = 5 Hz, DMTOC*H$_2$*), 3.61−3.75 (m, 12H, OC*H$_2$*CH$_2$O), 3.80 (s, 6H, OC*H$_3$*), 4.15 (t, 2H, $J$ = 5 Hz, C*H$_2$*OTs), 6.82−6.85 (m, 4H, arom H), 7.23−7.49 (m, 11H, arom H), 7.79−7.82 (m, 2H, arom H); $^{13}$C NMR (CDCl$_3$, ppm) 21.3, 54.9, 62.9, 68.4, 69.0, 70.3, 70.5, 85.7, 112.8, 126.4, 127.5, 127.6, 127.9, 129.6, 129.8, 132.9, 136.1, 144.5, 144.9, 158.2.

**(4,4′-Dimethoxytrityl)heptakis(ethylene glycol).** (4,4′-Dimethoxytrityl)tetrakis(ethylene glycol) *p*-toluenesulfonate (0.513 mmol) was lyophilized from anhydrous dioxane (2.5 mL) and then dissolved in anhydrous DMF (5.0 mL) and HMPA (0.45 mL). To this was added tris(ethylene glycol) (2.57 mmol) followed by Cs$_2$CO$_3$ (0.565 mmol) and NaH (0.565 mmol). The mixture was stirred for 40 h at room temperature under an argon atmosphere. The solution was diluted with CH$_2$Cl$_2$ (30 mL) and washed three times with H$_2$O (30 mL), twice with 5% NaHCO$_3$ (30 mL), and then once with saturated NaCl solution (30 mL). The organic phase was dried over anhydrous sodium sulfate, filtered, and concentrated to an oil under reduced pressure. This residue was purified by flash chromatography on silica gel using methanol:EtOAc (1:9) as the eluent. The product was obtained as an oil in 42% yield (137 mg, 0.218 mmol): $R_f$ 0.47 (methanol:EtOAc, 1:9); $^1$H NMR (CDCl$_3$, ppm) 2.98 (br s, 1H, O*H*), 3.23 (t, 2H, $J$ = 5 Hz, DMTOC*H$_2$*), 3.57−3.60 (m, 2H, C*H$_2$*CH$_2$OH), 3.64−3.68 (m, 24H, OC*H$_2$*CH$_2$O), 3.77 (s, 6H, OC*H$_3$*), 6.80−6.83 (m, 4H, arom H ortho of OCH$_3$), 7.20−7.48 (m, 9H, arom H); $^{13}$C NMR (CDCl$_3$, ppm) 54.9, 61.3, 62.8, 70.0, 70.2, 70.3, 70.4, 70.6, 72.2, 85.6, 112.7, 126.3, 127.4, 127.9, 129.7, 136.0, 144.8, 158.1.

**(4,4′-Dimethoxytrityl)octakis(ethylene glycol).** This compound was synthesized in a manner similar to that for formation of (4,4′-dimethoxytrityl)heptakis(ethylene glycol) above. The coupling of (4,4′-dimethoxytrityl)tetrakis(ethylene glycol) *p*-toluenesulfonate (513 mmol) with tetrakis(ethylene glycol) (2.57 mmol) followed by purification with silica gel chromatography (methanol:EtOAc, 3:97) gave the product as an oil in 51% yield (0.176 g, 0.262 mmol): $R_f$ 0.43 (methanol:EtOAc, 1:9); $^1$H NMR (CDCl$_3$, ppm) 2.68 (br s, 1H, O*H*), 3.24 (t, 2H, $J$ = 5 Hz, DMTOC*H$_2$*), 3.62−3.73 (m, 30H, OC*H$_2$*CH$_2$O, CH$_2$CH$_2$-OH), 3.81 (s, 6H, OC*H$_3$*), 6.82−6.85 (m, 4H, arom H), 7.20−7.49 (m,

9H, arom H); $^{13}$C NMR (CDCl$_3$, ppm) 54.8, 61.2, 62.9, 69.9, 70.2, 70.4, 72.4, 85.6, 112.7, 126.3, 127.4, 127.9, 129.7, 136.0, 144.8, 158.1.

**(4,4′-Dimethoxytrityl)heptakis(ethylene glycol) 2-Cyanoethyl *N,N*-Diisopropylphosphoramidite.**[18] To (4,4′-dimethoxytrityl)heptakis-(ethylene glycol) (0.318 mmol) dissolved in dry CH$_2$Cl$_2$ (0.75 mL) were added *N,N*-diisopropylethylamine (DIPEA) (1.27 mmol) and 2-cyano-ethyl *N,N*-diisopropylchlorophosphoramidite (0.477 mmol) via syringe. The mixture was stirred at room temperature for 20 min under argon. The solution was diluted with EtOAc (10 mL) and Et$_3$N (0.5 mL) and washed twice with 10% Na$_2$CO$_2$ (8 mL) and then twice with saturated NaCl solution (8 mL). The organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure. The residue was purified by flash chromatography on silica gel using EtOAc:hexanes (3:1). The product was obtained as a clear oil in 67% yield (0.213 mmol): *R$_f$* 0.37 (EtOAc:hexanes, 3:1); $^1$H NMR (CDCl$_3$, ppm) 1.15−1.19 (2d, 12 H, CH(C*H*$_3$)$_2$), 2.63 (t, 2H, *J* = 6 Hz, C*H*$_2$CN), 3.21 (t, 2H, *J* = 5 Hz, DMTOC*H*$_2$), 3.56−3.86 (m, 40 H, OC*H*$_3$, OC*H*$_2$C*H*$_2$O, C*H*$_2$OP, POC*H*$_2$CH$_2$CN, NC*H*(CH$_3$)$_2$), 6.79−6.82 (m, 4H, arom H), 7.20−7.48 (m, 9H, arom H); $^{13}$C NMR (CDCl$_3$, ppm) 20.0, 20.1, 24.3, 42.7, 42.9, 54.9, 58.1, 58.4, 62.2, 62.4, 62.9, 70.3, 70.4, 70.7, 70.9, 71.0, 85.6, 112.7, 126.3, 127.4, 127.9, 129.8, 136.1, 144.8, 158.1; $^{31}$P NMR (CDCl$_3$, ppm) 149.1; HRFAB (CHCl$_3$/NBA/PPG matrix) (M − H)$^+$ calcd for C$_{44}$H$_{65}$N$_2$O$_{11}$P 827.4251, found 827.4245.

**(4,4′-Dimethoxytrityl)octakis(ethylene glycol) 2-Cyanoethyl *N,N*-Diisopropylphosphoramidite.**[18] This compound was prepared in a manner similar to that for formation of the (dimethoxytrityl)heptakis-(ethylene glycol) phosphoramidite above. The reaction of (4,4′-dimethoxytrityl)octakis(ethylene glycol) (0.297 mmol) with 2-(cyano-ethyl) *N,N*-diisopropylchlorophosphoramidite (0.446 mmol) followed by flash chromatography on silica gel using EtOAc:hexanes (3:1) gave the product as an oil in 80% yield (0.238 mmol): *R$_f$* 0.32 (EtOAc:hexanes, 3:1); $^1$H NMR (CDCl$_3$, ppm); $^{13}$C NMR (CDCl$_3$, ppm) 1.18−1.22 (2d, 12 H, CH(C*H*$_3$)$_2$), 2.66 (t, 2H, *J* = 6 Hz, C*H*$_2$CN), 3.24 (t, 2H, *J* = 5 Hz, DMTOC*H*$_2$), 3.59−3.89 (m, 40 H, OC*H*$_3$, OC*H*$_2$C*H*$_2$O, C*H*$_2$OP, POC*H*$_2$CH$_2$CN, NC*H*(CH$_3$)$_2$), 6.82−6.85 (m, 4H, arom H), 7.23−7.51 (m, 9H, arom H); $^{13}$C NMR (CDCl$_3$, ppm) 20.0, 20.1, 24.2, 24.3, 24.4, 42.7, 42.9, 54.9, 58.1, 58.4, 62.2, 62.4, 62.9, 70.29, 70.36, 70.42, 70.46, 70.68, 70.9, 71.0, 85.6, 112.7, 126.3, 127.4, 127.9, 129.8, 136.1, 144.8, 158.1; $^{31}$P NMR (CDCl$_3$, ppm) 149.1; HRFAB (CHCl$_3$/NBA/PPG matrix) (M − H)$^+$ calcd for C$_{46}$H$_{68}$N$_2$O$_{12}$P 871.4513, found 871.4548.

**Oligonucleotide Synthesis.** DNA oligonucleotides were synthesized on an Applied Biosystems 392 synthesizer using standard β-cyanoethyl phosphoramidite chemistry.[19] Stepwise coupling yields for the non-natural residues were all greater than 95% as determined by trityl cation monitoring. Oligomers were purified by preparative 20% denaturing polyacrylamide gel electrophoresis, isolated by the crush and soak method, and quantitated by absorbance at 260 nm. Molar extinction coefficients were calculated by the nearest neighbor method.[20] Oligodeoxynucleotides were obtained after purification as the sodium salt.

**Thermal Denaturation Studies.** For the bimolecular complexes, solutions for the thermal denaturation studies contained a one-to-one ratio of two complementary oligomers. Concentrations for given experiments are listed in the text and figure legends. The buffer used was 100 mM NaCl, 10 mM MgCl$_2$, and 10 mM Na·PIPES (1,4-piperazinebis(ethanesulfonate)). The buffer pH is that of a 2× stock solution at 25 °C containing the buffer and salts. After the solutions were prepared, they were heated to 90 °C and allowed to cool slowly to room temperature prior to the melting experiments.

The melting studies were carried out in teflon-stoppered 1 cm pathlength quartz cells under nitrogen atmosphere on a Varian Cary 1 UV−vis spectophotometer equipped with thermoprogrammer. Absorbance was monitored at 260 nm while temperature was raised at a rate of 0.5 °C/min; a slower heating rate with this apparatus does not affect the results. In all cases the complexes displayed sharp, apparently two-state transitions, with all-or-none melting from bound complex to free oligomers. Melting temperatures ($T_m$) were determined by computer fit of the first derivative of absorbance with respect to $1/T$. Uncertainty in $T_m$ is estimated at ±0.5 °C on the basis of repetitions of experiments. Free energy values were derived by computer fitting the denaturation data with an algorithm employing linear sloping baselines, using a two-state model for melting.[21] Uncertainty in individual free energy measurements is estimated at ±5−10%.