The chemical end-ligation of homopyrimidine oligodeoxyribonucleotides within a DNA triple helix

Tianhu Li, David S Weinstein and KC Nicolaou

Background: Triple-helical nucleic acids, first reported in the late 1950s, are receiving attention for their possible involvement in controlling gene expression. Certain sequences of DNA are believed to form local triple-helical structures (H-form DNA), although this has not been directly observed *in vivo*. Studies carried out in our laboratories have suggested that self-replicating oligonucleotides could have been involved in chemical evolution via triple-helical intermediates. In addition to self-replication mechanisms, elucidating processes for the nonenzymatic elongation of biologically relevant polymers remains an important challenge in understanding the origin of life. To this end, we have studied a novel ligation of oligodeoxyribonucleotides that lie within a triple helix.

Results: The chemical end-ligation of homopyrimidine oligodeoxyribonucleotides on a triple helix is reported. This selective process, induced by cyanoimidazole, is facilitated by a template effect of the DNA aggregate and occurs between the 3′ end (hydroxyl) of the third minor-groove-bound strand and the 5′ end (phosphate) of the antiparallel oligopyrimidine strand.

Conclusions: Double-helical homopurine/homopyrimidine DNA can serve as a template for the elongation of oligonucleotides in a manner that has not been described previously. The end-ligation of homopyrimidine oligomers, a nonenzymatic process, proceeds via a requisite triple-helical intermediate and constitutes an efficient and selective method for the template-directed elongation of nucleic acids. Such a process could conceivably have been involved in the elongation of primordial information-bearing biopolymers.

Addresses: Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA and Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, California, USA.

Correspondence: KC Nicolaou

Key words: DNA triple helix, end-ligation, Hoogsteen, reverse Hoogsteen, template effect

Received: **29 November 1996**
Revisions requested: **20 December 1996**
Revisions received: **28 February 1997**
Accepted: **28 February 1997**

Electronic identifier: **1074-5521-004-00209**

Chemistry & Biology March 1997, **4**:209–214

© Current Biology Ltd ISSN 1074-5521

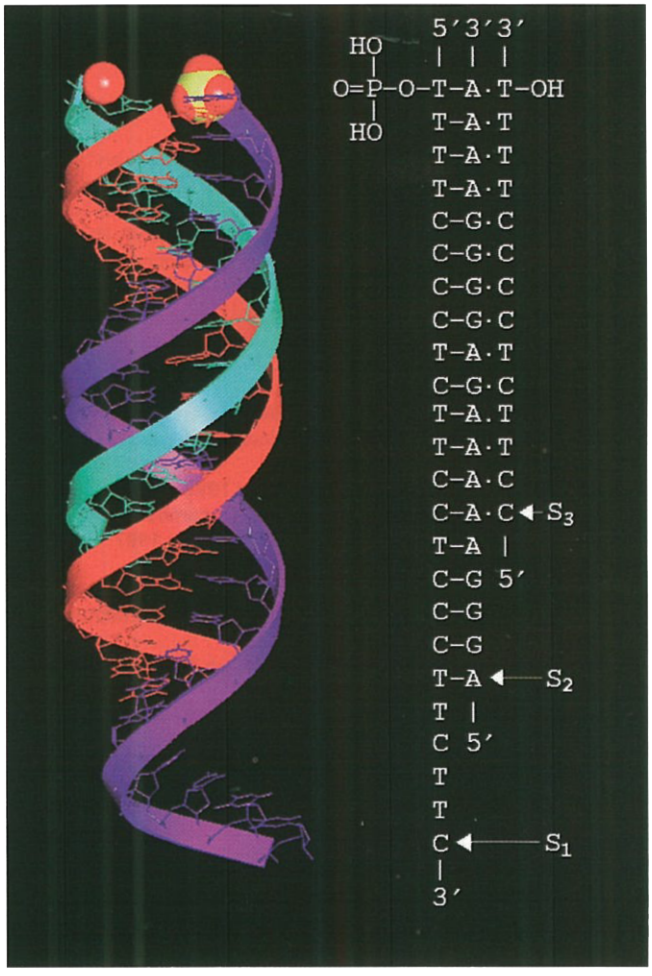
Introduction

Triple-helical DNA has received renewed interest in recent years for its possible involvement in the control of gene expression through the sequence-specific recognition of double-helical DNA [1–3]. Most recent studies have focused on an intramolecular or localized triplex in which a short homopyrimidine (pyr) oligodeoxyribonucleotide binds so that it is parallel to a homopurine (pur) sequence within the major groove of a pur•pyr tract of a Watson–Crick duplex under acidic conditions. More recently, third strand sequences that are rich in purine have been found to bind antiparallel to the homopurine strand in a relatively pH-independent manner [4]. In the course of studying the self-replication of palindromic duplex DNA [5], a triplex intermediate 24 nucleotides long was found to undergo what was believed to be the chemical ligation of the blunt ends of at least two of its complementary strands. Here, we report the investigation of this efficient end-ligation phenomenon, in which the 3′ end of the third (Hoogsteen hydrogen-bond mediated) oligopyrimidine condenses with the activated phosphorylated 5′ end [6] of the antiparallel homopyrimidine strand. The end-ligation has been demonstrated to occur with two randomly generated sequences, while both a third sequence rich in cytosine at the blunt-ended terminus and a triplex involving the reverse-Hoogsteen hydrogen-bonding mode gave no ligation products. The reaction constitutes a template-directed ligation of oligonucleotides that has not been described previously [7–9]. Unlike the work of Luebke and Dervan [7,8] and Rubin *et al.* [9], which involves the nonenzymatic ligation of DNA oligomers with localized triple helices, the end-ligation we describe occurs at the termini of oligomers that are not pre-aligned head-to-tail within a local triplex. In the context of chemical evolution, this type of self-directed elongation may have provided one mechanism for the nonenzymatic growth of stable DNA oligomers.

Results and discussion

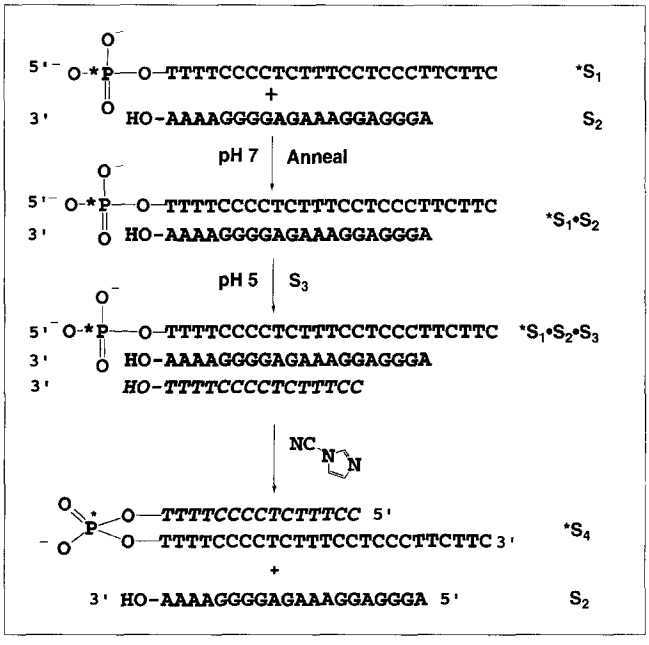
Initially, a triple-helical DNA composed of a 25-mer, 20-mer, and 15-mer strand (S_1, S_2, and S_3, respectively) was used in this study (Fig. 1). The oligopyrimidine S_1 and oligopurine S_2 were annealed at neutral pH to form a double helix, S1•S2 (Fig. 2). The association of oligopyrimidine strand S_3 with the duplex was then conducted at pH 5, to generate the triplex S_1•S_2•S_3. The subsequent addition of cyanoimidazole [7–10] led to condensation of S_1 and S_3 to give the ligation product S_4, presumably

Figure 1



Computer-generated model of DNA strands undergoing end-ligation within a triple helix. The participating hydroxyl and phosphate functionalities are shown (space filling models, top).

Figure 2



End-ligation of homopyrimidine oligomers within a DNA triple helix.

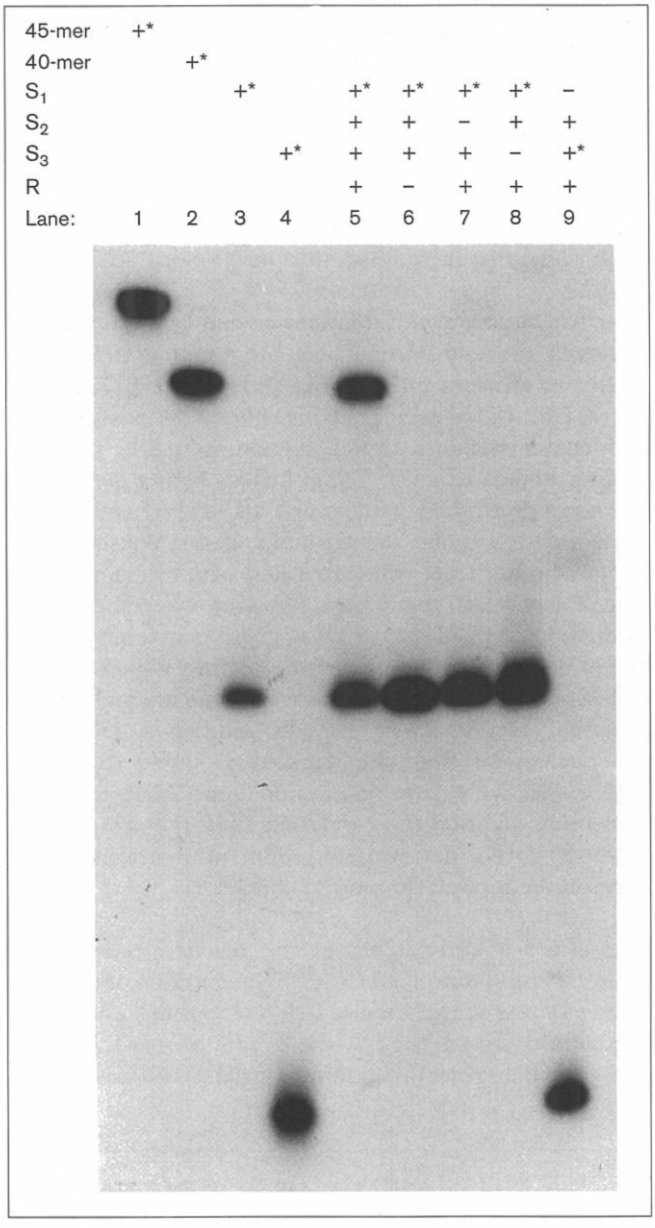
through activation of the monophosphoester at the 5′ end of S_1 followed by nucleophilic substitution by the 3′ hydroxyl of S_3. An analogous triplex in which the 5′ terminus of the homopurine strand was in contact with the 5′ phosphorylated terminus of the Hoogsteen-bound homopyrimidine strand was also briefly investigated, but this gave only a mixture of noncharacterized ligation products at a much slower rate.

The reaction mixtures were analyzed by denaturing gel electrophoresis and visualized by autoradiography (Fig. 3). The condensation product from the reaction mixture containing all components (lane 5) was identified as a 40-mer by comparing its electrophoretic mobility with that of a synthetic 5′-phosphorylated 40-mer (lane 2), indicating exclusive ligation between S_1 (25-mer) and S_3 (15-mer). The slight retardation of the known 40-mer (lane 2) relative to S_4 (lane 5) is due to the additional phosphate at its 5′ end. Ligation between the two strands of the Watson–Crick duplex (i.e. S_1–S_2, lane 1) was not detected in the reaction mixture (lane 5). The requirement for cyanoimidazole in the reaction mixture (lane 6) also suggests that a new covalent bond is formed [10], and the requirement for each of the three strands (lanes 7–9) is consistent with the model of a triplex-mediated process.

In order to explore the generality of the end-ligation phenomenon, experiments were also performed with a second triplex of random sequence (S_1'•S_2'•S_3'). Selective and efficient ligation between the homopyrimidine oligomers also occurred using this triplex (Fig. 4, lane 5). As expected, four control experiments analogous to those for triplex S_1•S_2•S_3 (Fig. 3, lanes 6–9) gave no ligation products (data not shown).

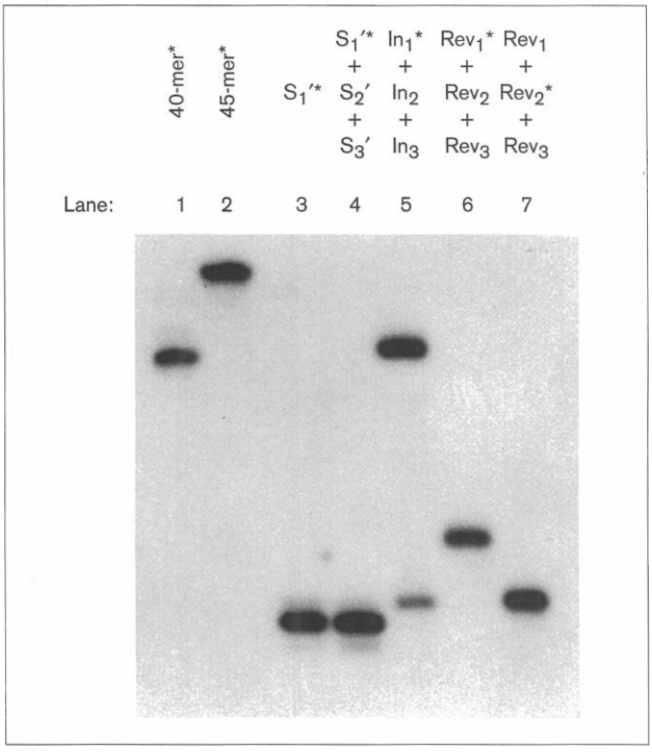
A third sequence was chosen in which each of the bases of triplex S_1•S_2•S_3 was substituted for its closest relative (i.e. C for T, T for C, A for G, and G for A). This 'inverse' sequence (In_1•In_2•In_3) gave no ligation products under the usual conditions (Fig. 4, lane 4). The failure of this particular sequence to participate in end-ligation may be attributed to the four adjacent cytosines at the 5′ and 3′ termini of the 25-mer S_1' and 15-mer S_3', respectively. It is known that adjacent cytosines, which must each be protonated to participate in both Watson–Crick and Hoogsteen hydrogen bonds in this type of triplex, are destabilizing to the overall structure [11]. Thus, structural and/or electronic effects imposed by these positively charged bases may be impeding the

Figure 3



End-ligation within triple helix $S_1 \cdot S_2 \cdot S_3$. (Asterisks imply radioactivity incorporated to the extent that unlabeled:labeled > 100.) Lane 1, standard 5′-^{32}P 45-mer (i.e. '*(S_1+S_3)'); lane 2, standard 5′-^{32}P 40-mer* (note: unlike end-ligation product *S_4 which contains ^{32}P mid-sequence, end-labeled standard has a slightly different motility than the reaction product); lane 3, standard 5′-^{32}P 25-mer (*S_1); lane 4, standard 5′-^{32}P 15-mer (*S_3); lane 5, 10 μM *$S_1 \cdot S_2$, 30 μM S_3, 50 mM cyanoimidazole (R), 100 mM MES (4-morpholineethanesulfonic acid) (pH 5), 50 mM NaCl and 10 mM $ZnCl_2$; lane 6, same as lane 5, without cyanoimidazole; lane 7, same as lane 5 without S_2; lane 8, same as lane 5 without S_3; lane 9, same as lane 5 without S_1, replacing S_3 with 5′-^{32}P *S_3.

Figure 4



Investigating the generality of the end-ligation phenomenon. Reactions were carried out as in Figure 3. Lane 1, 5′-^{32}P 40-mer; lane 2, standard 5′-^{32}P 45-mer*; lane 3, 5′-^{32}P S1′*; lane 4, reaction mixture of In_1*, In_2 and In_3 (In_1*=5′P*-CCCCTTTTCTCCCTTCTTTCCTCCT 3′; In_2=5′GAAAGAAGGGAGAAAAGGGG 3′; In_3=5′TTCCCTCTTTT-CCCC 3′); lane 5, reaction mixture of S_1′*, S_2′ and S_3′ (S_1′*=5′ P*-TCTTCTTTCTCTTTTCTTCTTCTTC 3′; S_2′=5′AGAAGAAAAGAGAA-AGAAGA 3′; S_3′=5′TTTTCTCTTTCTTCT 3′); lane 6, reaction mixture of 5′ P*-Rev_1-P 3′, Rev_2 and Rev_3 (5′ P*-Rev_1-P 3′=5′ P*-AGAGAGT-TTTTCCCTCCCCTCCCCTCCCT-P 3′; Rev_2=5′P-AGGGAGGGG-AGGGGAGGGAAAAAC 3′; Rev_3=5′TGGGTGGGGTGGGGTGG-GT 3′); lane 7, reaction mixture of Rev_1, 5′ P*-Rev_2 and Rev_3. (5′P*-Rev_2=5′P*-AGGGAGGGGAGGGGAGGGAAAAAC 3′).

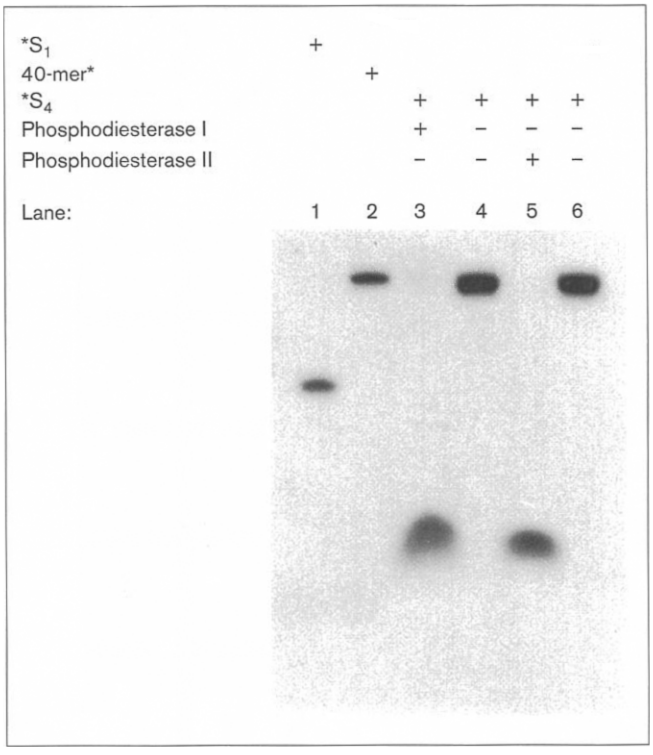
ligation reaction, although it is not clearly understood at which step in the reaction scheme or ligation mechanism this occurs. Alternatively, the failure to observe ligation in this particular case may simply be due to subtle sequence-dependent structural or conformational differences.

The end-ligation of a triplex similar to that which was used in an affinity-cleavage experiment of Beal and Dervan [4] was also investigated. The third (purine-rich) strand was expected to bind in an antiparallel orientation to the homopurine strand of the Watson–Crick duplex in the reverse-Hoogsteen hydrogen-bond mode. At neutral pH, no ligation between either of the two antiparallel strands of the putative triplex $Rev_1 \cdot Rev_2 \cdot Rev_3$ was observed (Fig. 4, lanes 6,7).

Characterization of the end-ligation product

To confirm the formation of a natural 5′–3′ phosphodiester linkage between S_1 and S_3, the ligation product, S_4, was purified by gel electrophoresis [12] and treated separately

Figure 5



Phosphodiesterase digestion of the end-ligation product (*S_4). Radiolabeling and reaction processing were carried out as in lane 5 of Figure 3. Lane 1, synthetic 40-mer*; lane 2, standard 25-mer (5′-^{32}P *S_1); lane 3, mixture of 100 mM CHES [2-(cyclohexylamino)-ethanesulfonic acid] (pH 10), 5 mM $MgCl_2$, 10 units phosphodiesterase I and purified ligation product *S_4 after incubation at 25°C for 1 h; lane 4, same as lane 3 but without phosphodiesterase I; lane 5, 50 mM sodium acetate (pH 6), 10 units phosphodiesterase II and purified ligation product *S_4 after incubation at 37°C for 1 h; lane 6, same as lane 5 but without phosphodiesterase II.

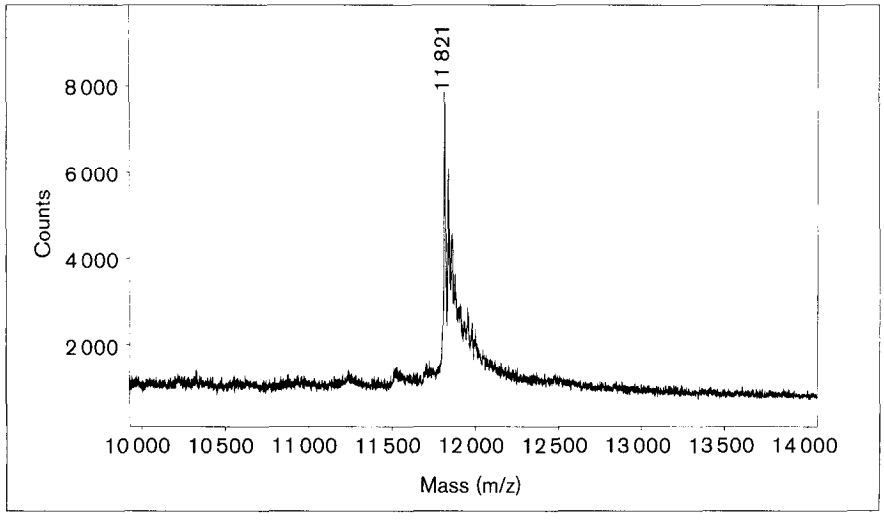
with either phosphodiesterase I [7–8] or phosphodiesterase II. As expected, the 40-mer was completely digested by phosphodiesterase I, a 5′-exonuclease, and by phosphodiesterase II, a 3′-exonuclease verifying the nature of the newly formed bond (Fig. 5, lanes 3,5). To examine further the constitution of the end-ligation product, purified S_4 was analyzed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Fig. 6) [13]. The expected mass of 11 821 Da was found experimentally, conclusively establishing that the product was the 40-mer.

Effects of structural perturbations on end-ligation

Although the end-ligation of a blunt-ended triplex was clearly an efficient process (Fig. 3, lane 5; Fig. 7, lane 2; Fig. 8, lane 4), we decided to investigate the possibility of end-ligation within a triplex containing one or two protruding strands (Fig. 7). Triple helices with a pyrimidine oligomer protruding either two (S_3+2) or one (S_3+1) nucleotide(s) beyond the terminus of the Watson–Crick duplex did not react when activated with cyanoimidazole (Fig. 7, lanes 3,4). In contrast, however, when the duplex extended one (S_3–1) or two (S_3–2) base pair(s) beyond the end of the third strand, the corresponding 39-mer (lane 5) and 38-mer (lane 6) were formed in moderate yield under the usual reaction conditions. The addition or deletion of two nucleotides from the oligopurine strand (S_2+2 and S_2–2, respectively) gave no reaction (lanes 7,10), and a one nucleotide addition to or deletion from the same strand (S_2+1, S_2–1) did not entirely inhibit the reaction, giving traces of the product 40-mer, S_4 (lanes 8,9).

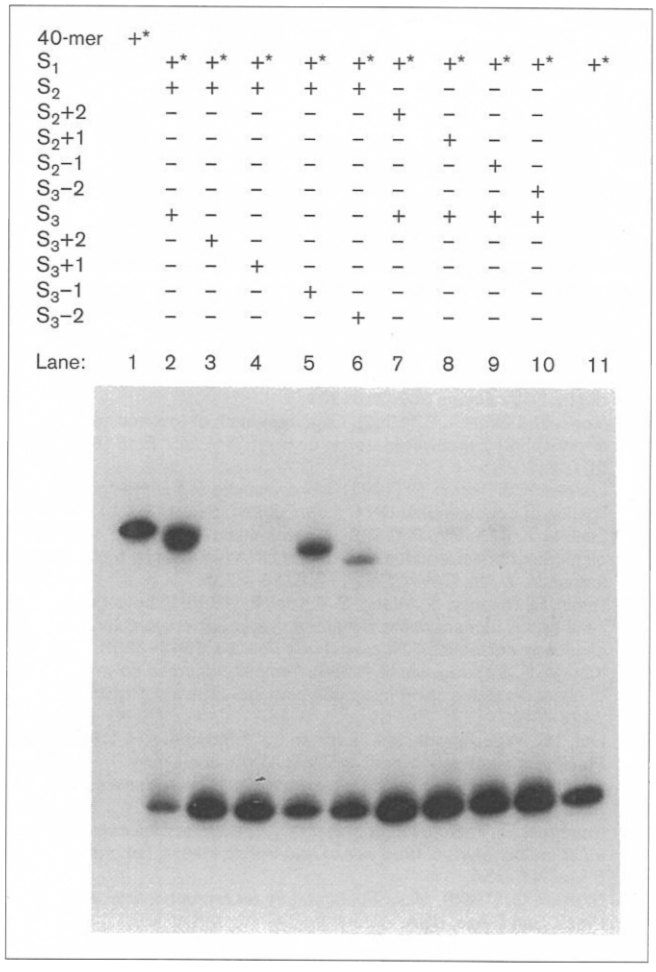
The effect of mismatches at the reactive ends of the triplex was also examined (Fig. 8). Substitution of thymine for two or one of the adenines closest to the 3′ end of the oligopurine strand (i.e. S_2-mis-2 and S_2-mis-1, respectively) gave no reaction under the usual conditions (Fig. 8,

Figure 6



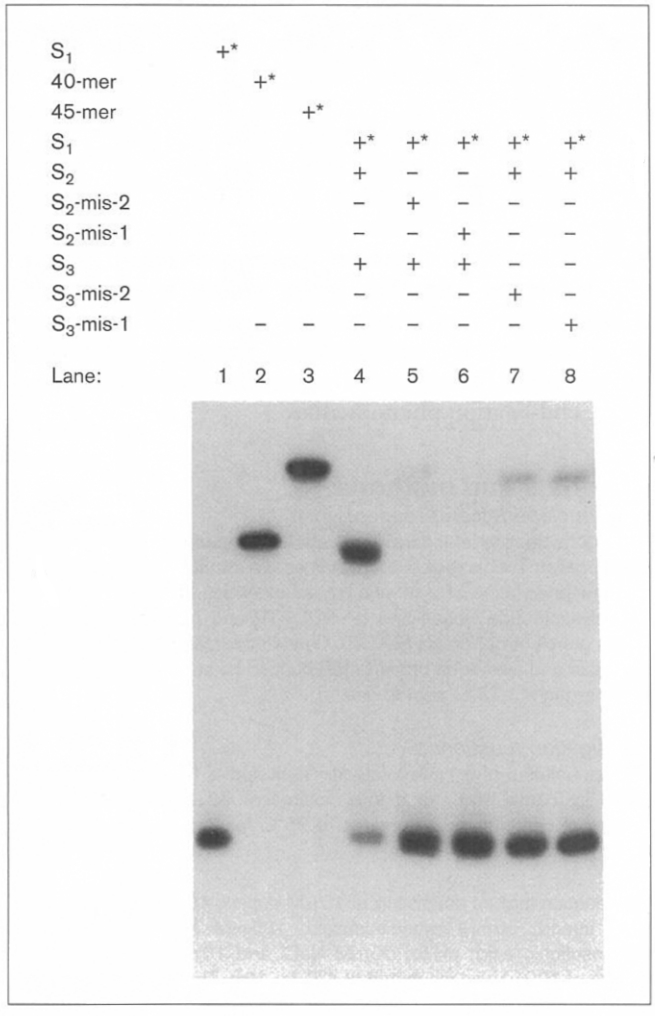
MALDI mass spectrum of the product 40-mer S_4.

Figure 7



End-ligation experiments with triple helices containing protruding ends. Reactions were carried out as in lane 5 of Figure 3. Lane 1, standard 5′-^{32}P 40-mer*; lane 2, 10 μM S_1•S_2, 30 μM S_3, 50 mM cyanoimidazole, 100 mM MES (4-morpholineethanesulfonic acid; pH 5), 50 mM NaCl and 10 mM $ZnCl_2$; lane 3, as lane 2 but replacing S_3 (5′ CTTTTCTCCCCTTTT 3′) with 17-mer S_3+2 (5′ CTTTTCTCCCCTTTTAA 3′); lane 4, as lane 2, but replacing S_3 with S_3+1 (5′ CTTTTCTCCCCTTTT A 3′); lane 5, as lane 2, but replacing S_3 with 14-mer S_3−1 (5′ CTTTTCTCCCCTTT 3′); lane 6, as lane 2, but replacing S_3 with S_3−2 (5′ CTTTTCTCCCCTT 3′); lane 7, as lane 2, but replacing S_2 (5′ AGGGAGGAAAGAGGGGAAAA 3′) with 22-mer S_2+2 (5′ AGGGAGGAAAGAGGGGAAAATT 3′); lane 8, as lane 2, but replacing S_2 with 21-mer S_2+1 (5′ AGGGAGGAAAGAGGGGAAAAT 3′); lane 9, as lane 2, but replacing S_2 with 19-mer S_2−1 (5′ AGGGAGGAAAGAGGGGAAA 3′); lane 10, as lane 2, but replacing S_2 with 18-mer S_2−2 (5′ AGGGAGGAAAGAGGGGAA 3′); lane 11, standard 5′-^{32}P 25-mer S_1*.

Figure 8



End-ligation experiments with triple helices containing mismatched ends. Lane 1, standard 5′-^{32}P 25-mer (S_1*); lane 2, standard 5′-^{32}P 40-mer (S_4*); lane 3, standard 5′-^{32}P 45-mer; lane 4, same as lane 5 of Figure 3; lane 5, as lane 4, but replacing S_2 (5′ AGGGAGGAAAGAGGGGAAAA 3′) with S_2-mis-2 (5′ AGGGAGGAAAGAGGGGAATT 3′); lane 6, as lane 4, but replacing S_2 with S_2-mis-1 (5′ AGGGAGGAAAGAGGGGAAAT 3′); lane 7, as lane 4, but replacing S_3 with S_3-mis-2 (5′ CTTTTCTCCCCTTAA 3′); lane 8, as lane 4, but replacing S_3 with S_3-mis-1 (5′ CTTTTCTCCCCTTTA 3′).

lanes 5,6). This was not very surprising since the purine bases of triple-helical DNA are essential, being involved in hydrogen bonding to both the parallel and antiparallel pyrimidine strands. Remarkably, substitution of adenine for thymine at both one and two of the 3′ terminal nucleotides of the oligopyrimidine S_3 gave traces of the 45-mer (lanes 7,8), corresponding to end-ligation within the Watson–Crick duplex (*i.e.* S_1+S_2).

Significance

Studies have suggested that DNA oligomers have roles in the origin of life beyond that of information-bearing templates (that were preceded by functional RNA catalysts). For example, DNA enzymes have recently been developed that cleave RNA [14,15], ligate DNA [16], catalyze porphyrin metallation [17], or cleave DNA [18]. Although nonenzymatic mechanisms for the elongation of oligonucleotides have been reported [6,19], they involve the formation of phosphodiester linkages on a growing chain in a stepwise fashion. Like the self-replication of DNA oligomers which has been demonstrated

previously [5], the end-ligation reaction described here may be another prebiotically important reaction that proceeds via a triple-helical intermediate. The end-ligation within DNA triple helices is an ordered and predictable process in that the DNA oligomer serves as a template for its own elongation, so the sequence of the product is predetermined by that of the template. It is conceivable that a prebiotic functional DNA oligomer could direct its own elongation via an intermediate triple-helical structure. A computer-generated model indicates that there are not obvious geometric constraints that would confine the reaction to the ends of the homopyrimidine strands. Studies are in progress to investigate the selectivity and thereby investigate further the scope of the end-ligation phenomenon.

Materials and methods

Oligonucleotides and reagents

All oligonucleotide standards and starting materials were purchased from Operon Technologies, Inc. Radioactive strands were prepared by the phosphorylation of synthetic oligonucleotides at their 5′ ends with T4 polynucleotide kinase and [γ-^{32}P] ATP and purified by polyacrylamide gel electrophoresis (PAGE). Cyanoimidazole was prepared from imidazole and cyanogen bromide [8], purified by sublimation and stored under argon at −78°C prior to use.

End-ligation reactions

A stock solution of 0.1 mM Watson–Crick duplex (e.g. S_1•S_2) was prepared in 20 mM NaCl at a total volume of 50 μl by incubating the mixture of complementary strands at 95°C for 3 min followed by slow cooling to 25°C over 1 h [15].

The reaction mixture consisting of 10 μM duplex, 40 μM third polypyrimidine strand, 50 mM cyanoimidazole, 100 mM MES (4-morpholineethanesulfonic acid; pH 5), 50 mM NaCl and 10 mM $ZnCl_2$ in a total volume of 100 μl was incubated at 4°C for 12 h. The reaction was terminated by the addition of loading buffer and the products were analyzed by PAGE.

For the reverse-Hoogsteen end-ligation attempt, stock solutions of 5′ P*-Rev_1-P 3′•Rev_2 and Rev_1•5′ P*-Rev_2 were prepared as for duplex S_1•S_2. The reaction mixture consisting of 10 μM duplex (5′ P*-Rev_1-P 3′•Rev_2 or Rev_1•5′ P*-Rev_2), 40 μM third strand (Rev_3), 100 mM imidazole (pH 7), 10 mM NaCl and 50 mM cyanoimidazole at a total volume of 100 μl was incubated at 4°C for 12 h; the reaction was terminated by the addition of loading buffer and the products analyzed by PAGE.

Enzymatic digestion of the end-ligation product

End-ligation product *S_4 was purified by PAGE. A reaction mixture consisting of 100 mM CHES [2-(cyclohexylamino)ethanesulfonic acid] (pH 10), 5 mM $MgCl_2$, 10 units phosphodiesterase I and purified ligation product *S_4 was incubated at 25°C for 1 h, quenched by addition of loading buffer and analyzed by PAGE.

Another reaction mixture consisting of 50 mM sodium acetate (pH 6), 10 units phosphodiesterase II and purified ligation product *S_4 was incubated at 37°C for 1 h, quenched by addition of loading buffer and analyzed by PAGE.

MALDI mass spectral analysis

The nonradiolabeled product S_4 was purified by PAGE and further desalted by elution through a Sephadex PD-10 column with deionized water. The MALDI spectrum was obtained on a PerSeptive Biosystems Voyager-Elite mass spectrometer with delayed extraction.

Acknowledgements

We thank Gary Siuzdak for his assistance with the mass spectrometry. This work was financially supported by the National Institutes of Health, USA, the ASLAM foundation, Merck, Pfizer, Hoffman La Roche, Amgen, and The Skaggs Institute of Chemical Biology.

References

1. Soyfer, V.N. & Potaman, V.N. (1996). *Triple-Helical Nucleic Acids.* Springer-Verlag, New York, USA.
2. Moser, H. & Dervan, P. (1987). Sequence-specific cleavage of double helical DNA by triple helix formation. *Science* **238**, 645–650.
3. Thuong, N. & Helene, C. (1993). Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.* **32**, 666–690.
4. Beal, P. & Dervan, P. (1991). Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science* **251**, 1360–1363.
5. Li, T. & Nicolaou, K.C. (1994). Chemical self-replication of palindromic duplex DNA. *Nature* **369**, 218–221.
6. Inoue, T. & Orgel, L.E. (1982). Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly(C). *J. Mol. Biol.* **162**, 201–217.
7. Luebke, K. & Dervan, P. (1991). Nonenzymatic sequence-specific ligation of double-helical DNA. *J. Am. Chem. Soc.* **113**, 7447–7448.
8. Luebke, K. & Dervan, P. (1989). Nonenzymatic ligation of oligodeoxyribonucleotides on a duplex DNA template by triple-helix formation. *J. Am. Chem. Soc.* **111**, 8733–8735.
9. Rubin, E., Rumney, S., Wang, S. & Kool, E. (1995). Convergent DNA synthesis: a non-enzymatic dimerization approach to circular oligodeoxynucleotides. *Nucleic Acids Res.* **23**, 3547–3553.
10. Kanaya, E. & Yanagawa, H. (1986). Template-directed polymerization of oligoadenylates using cyanogen bromide. *Biochemistry* **25**, 7423–7430.
11. Lee, J.S., Woodsworth, M.L., Latimer, L.J. & Morgan, A.R. (1984). Poly(pyrimidine)•poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res.* **12**, 6603–6614.
12. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). In *Molecular Cloning: a Laboratory Manual (2nd edn).* Cold Spring Harbor Laboratory Press, NY, USA.
13. Siuzdak, G. (1996). *Mass Spectrometry for Biotechnology.* Academic Press, San Diego, USA.
14. Breaker, R.R. & Joyce, G.F. (1995). A DNA enzyme with Mg^{2+}-dependent RNA phosphoesterase activity. *Chem. Biol.* **2**, 655–660.
15. Breaker, R.R. & Joyce, G.F. (1994). A DNA enzyme that cleaves RNA. *Chem. Biol.* **1**, 223–229.
16. Cuenoud, B. & Szostak, J.W. (1995). A DNA metalloenzyme with DNA ligase activity. *Nature* **375**, 611–614.
17. Li, Y. & Sen, D. (1996). A catalytic DNA for porphyrin metallation. *Nat. Struct. Biol.* **3**, 743–747.
18. Carmi, N., Shultz, L. & Breaker, R. (1996). *In vitro* selection of self-cleaving DNAs. *Chem. Biol.* **3**, 1039–1046.
19. Ferris, J., Hill, A., Liu, R. & Orgel, L. (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61.